



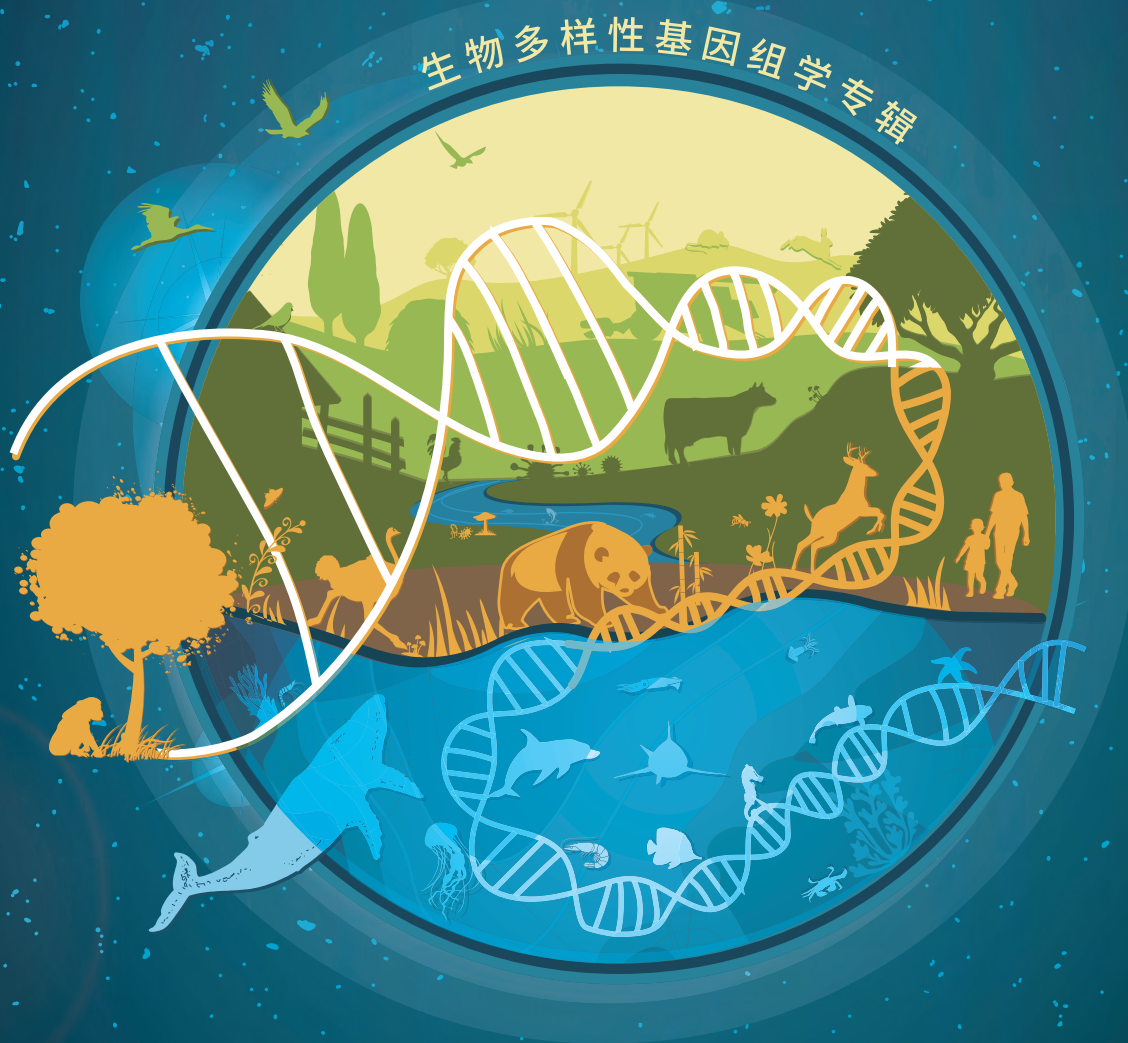
ISSN 1005-0094  
CODEN SHDUEM

# BIODIVERSITY SCIENCE

# 生物多样性

第27卷 第5期  
2019年5月

Vol. 27 No. 5  
May 2019



主 办

中国科学院生物多样性委员会

中国植物学会

中国科学院植物研究所

中国科学院动物研究所

中国科学院微生物研究所

Biodiversity Committee, CAS

Botanical Society of China

Institute of Botany, CAS

Institute of Zoology, CAS

Institute of Microbiology, CAS

<http://www.biodiversity-science.net>

# 生物多样性

## SHENGWU DUOYANGXING

第27卷 第5期 2019年5月

### 目次

#### 编者按

- 475 运用基因组学方法探究生物多样性: 大数据时代的“胡克显微镜”  
周欣

#### 综述

- 480 环境DNA技术在淡水底栖大型无脊椎动物多样性监测中的应用  
李萌 尉婷婷 史博洋 郝希阳 徐海根 孙红英
- 491 基于环境DNA-宏条形码技术的水生生态系统入侵生物的早期监测与预警  
李晗溪 黄雪娜 李世国 战爱斌
- 505 模式动物在共生微生物研究中的作用  
肖雅倩 刘传 肖亮
- 516 洞察景观环境影响蜜蜂之新视角: 肠道微生物  
唐敏 邹怡 苏秦之 周欣
- 526 DNA条形码参考数据集构建和序列分析相关的新兴技术  
刘山林

- 534 基于三代测序技术的微生物组学研究进展  
许亚昆 马越 胡小茜 王军

#### 研究报告

- 543 基于粪便DNA及宏条形码技术的食肉动物快速调查及食性分析  
邵昕宁 宋大昭 黄巧雯 李晟 姚蒙
- 557 用于蜜蜂和熊蜂肠道微生物分类的细菌16S rRNA数据库优化  
张雪 李兴安 苏秦之 曹棋钠 李晨伊 牛庆生 郑浩

#### 方法

- 567 基于Nextflow构建的宏条形码自动化分析流程EPPS  
李诣远 David C. Molik Michael E. Pfrender
- 576 本期作者团队介绍

---

**封面:** 核酸是地球生命组成的基本元件之一, 包含了具有物种特异性的遗传信息, 为我们从分子水平认知生物多样性提供了全新的角度。基于高通量测序技术的生物多样性基因组学方法, 可以同时对接海量的核酸序列进行测序, 使混合生物样本的多样性分析成为可能。在新方法的推动下, 针对海洋、陆地、淡水生态系统的生物多样性解码工作迅速展开。在准确快速的物种鉴定基础上, 生物多样性的认知得以深入到多样性动态监测、生态网络构建、宿主-微生物关系、入侵物种检测等领域。(封面设计: 胡婷婷)

# BIODIVERSITY SCIENCE

Vol. 27 No. 5 May 2019

## CONTENTS

### Editorial

- 475 **Understanding biodiversity using genomics: Hooke's microscope in the era of big data**  
Xin Zhou

### Reviews

- 480 **Biodiversity monitoring of freshwater benthic macroinvertebrates using environmental DNA**  
Meng Li, Tingting Wei, Boyang Shi, Xiyang Hao, Haigen Xu and Hongying Sun
- 491 **Environmental DNA (eDNA)-metabarcoding-based early monitoring and warning for invasive species in aquatic ecosystems**  
Hanxi Li, Xuena Huang, Shiguo Li and Aibin Zhan
- 505 **The role of model animals in the study of symbiotic microorganisms**  
Yaqian Xiao, Chuan Liu and Liang Xiao
- 516 **A new perspective on landscape impact in bee populations: Considering the bee gut microbiome**  
Min Tang, Yi Zou, Qinzhi Su and Xin Zhou

- 526 **DNA barcoding and emerging reference construction and data analysis technologies**  
Shanlin Liu
- 534 **Analysis of prospective microbiology research using third-generation sequencing technology**  
Yakun Xu, Yue Ma, Xiaoxi Hu and Jun Wang

### Original Papers

- 543 **Fast surveys and molecular diet analysis of carnivores based on fecal DNA and metabarcoding**  
Xinning Shao, Dazhao Song, Qiaowen Huang, Sheng Li and Meng Yao
- 557 **A curated 16S rRNA reference database for the classification of honeybee and bumblebee gut microbiota**  
Xue Zhang, Xing'an Li, Qinzhi Su, Qina Cao, Chenyi Li, Qingsheng Niu and Hao Zheng

### Methodology

- 567 **EPPS, a metabarcoding bioinformatics pipeline using Nextflow**  
Yiyuan Li, David C. Molik and Michael E. Pfrender

---

**Cover Illustration:** Nucleic acids are one of the essential elements of global lives, encoding taxon-specific genetic information, which creates a whole new angle for understanding biodiversity using a molecular approach. Biodiversity genomics methods based on high throughput sequencing are capable of charactering massive amounts of nucleotide sequences in a parallel way, therefore enabling biodiversity analyses for mixed biology samples. Facilitated by this new technology, global initiatives of biodiversity registration for marine, terrestrial, and freshwater systems have been rapidly coordinated. Eventually, on the basis of accurate and efficient taxonomic identification, biodiversity genomics will see extensive applications in biodiversity monitoring, ecological network construction, host-symbiont interactions, invasive species detection, among many others. (Cover design: Tingting Hu)



•编者按•

# 运用基因组学方法探究生物多样性： 大数据时代的“胡克显微镜”

周 欣\*

(中国农业大学植物保护学院昆虫学系, 北京 100193)

## Understanding biodiversity using genomics: Hooke's microscope in the era of big data

Xin Zhou\*

Department of Entomology, College of Plant Protection, China Agricultural University, Beijing 100193

对生物多样性的好奇和欣赏是人类认识自然的重要动力,而以探究和描述生物多样性为主要内容的分类学也是生命科学发展史上最早的分支学科之一。生物多样性为人类发展提供了基础和资源,为生命机理的认知提供了模型,也为维持可持续发展的生态系统提供了重要的服务功能。

分类学是生物多样性研究的基础,是我们分析多样性组成、度量多样性变化、评价生态系统质量的核心工具。分类学使用生物固有的特征(traits)来进行物种的界定(species delineation),同时根据物种的定义进行物种鉴定(species identification)。自林奈建立现代分类学体系后(Linnaeus, 1758),研究者得以在一个稳定有序的方法系统内对物种的定义进行验证和修订,成就了近300年来分类学以及生物多样性知识体系的持续性积累。应该说,林奈分类学系统是一个经历了时间考验的科学框架,实现了跨类群、跨学科的分类名称标准化。例如,在林奈系统下,发育生物学家基于模式昆虫黑腹果蝇(*Drosophila melanogaster*)来理解微小RNA在发育中的作用机理,而进化生物学家得以研究这一机理在果蝇属甚至昆虫纲中的普适性和演化历史(Belles, 2017)。更重要的是,林奈系统对新方法、新证据具有强大的兼容性。随着科学技术的迅速发展,可用于分类学的特征从单一的外部形态学扩展到包括

内部器官结构、染色体特征、核酸/蛋白/小分子序列、生理学等生物学或生态学特征的综合性证据。而这些整合分类学特征均可以统一在林奈分类系统下,使得研究者不需要各自独创一套新的分类体系,而是可以在同一系统中不断完善物种的定义。

在此大背景之下,研究者认识到各种生物特征在界定物种和认知生物多样性方面均存在自身的优势和局限性。经典的形态分类学在特定的场景下遇到了较大的挑战,例如同一种的不同形态(如真菌的孢子和子实体,昆虫的不同生活史阶段)或生物体残片等。针对上述问题,分子分类学(DNA taxonomy)应运而生。核酸序列既是生物遗传信息的载体,也是新性状产生的重要基础,因而在演化过程中既具有一定的保守性,也具有较强的类群特异性。这个特性很自然地使分类学家作为新的分类学特征应用于物种鉴定当中。耐人寻味的是,微生物(真菌、细菌、病毒等)的形态分类本身存在较大的困难,这可能是促成分子分类及分子系统发育方法(如16S rRNA基因)最早在该类群应用的主要原因(Woese & Fox, 1977)。

近十几年来,在真核生物分子分类学领域最突出的进展是国际DNA条形码项目(International Barcode of Life)的建立和实施(Hebert et al, 2003)。这项由全球科学家参与的大型国际合作使用标准

收稿日期: 2019-05-13; 接受日期: 2019-05-13

基金项目: 科技基础资源调查专项(2018FY100403)

\* 通讯作者 Author for correspondence. E-mail: xinzhoucaddis@icloud.com



化的DNA序列(DNA条形码)进行跨类群的物种研究,其核心数据库BOLD (Barcode of Life Databases) (Ratnasingham & Herbert, 2007)已包含700多万条条形码序列,涉及近60万物种级别的分类单元(www.boldsystems.org,截至2019年5月)。在统一的国际标准下,DNA条形码允许研究者对有研究基础的、生态系统中常见的动植物物种进行鉴定,并且对分类学和基于分类数据的应用科学产生了巨大的推动作用。我国DNA条形码研究的发展与国际保持同步,近年来在相关应用方面也取得了长足的进步(参见《生物多样性》2015年第3期“DNA条形码应用进展专题”的总结)。作为新的特征证据,DNA条形码在为物种鉴定提供方法的同时,也起到了搭建分类学知识桥梁的关键作用。它通过林奈双名法系统,将具有物种特异性的分子序列与该物种相关的生物学知识关联起来。在构建条形码参考数据库的过程中,这个关键步骤是在分类学家的支撑和监督指导下完成的,这也为分子分类学走向高通量混合体系生物分类奠定了重要基础。

尽管以DNA条形码为代表的分子分类学取得了重要进展,但该方法在生物多样性的实际研究和应用中存在明显的局限性。一方面,生物多样性研究中采集的大部分样本是由多物种组成的,而DNA条形码的核心技术则建立在单一生物样本的分析基础之上,包括样本的分拣,DNA的提取、扩增和测序等。而生物群落的自然分布及定植状态通常不利于单个样本的分离(如土壤动物群落,肠道菌群等),这就极大地限制了条形码技术在复杂生态样本分析中的应用(Hajibabaei et al, 2011)。另一方面,环境中存在大量游离DNA,这些核酸物质包含了该环境的生物多样性组成信息,而核酸样本的分离在操作上则更加困难(Creer et al, 2016)。相比之下,高通量测序技术(high-throughput sequencing)可以同时大量的核酸片段进行测序,并通过生物信息分析对测序的结果进行序列的拆分和物种鉴定,较好地应对了环境混合生物多样性研究的挑战(Zhou et al, 2013)。

为了及时地反映我国在本领域的研究进展和科研热点,《生物多样性》期刊特组织了本期“生物多样性基因组学专辑”,由9篇论文组成,包括6篇综述和3篇研究论文。

目前,使用环境游离DNA (environmental DNA,

eDNA)进行群落多样性分析最能体现多样性组学技术的优势,也是最早开展的系列研究之一。由于eDNA主要是由生物释放到环境中的DNA组成的,检测eDNA样品具有非损伤性采样的优势(即无需采集生物样本本身),在很大程度上避免了对生物群落的干扰,也有利于检测较为隐蔽的类群。基于eDNA的宏条形码(eDNA metabarcoding)被广泛应用于水体(Fronhofer et al, 2018)和土壤群落(Bahram et al, 2018)的多样性监测研究中,并可以与已有的环境监测(environment monitoring)方法较好地契合。在淡水生态系统中,利用大型底栖无脊椎动物多样性的组成和动态变化进行水体质量评估,是水生态研究与应用中的关键步骤。在本专辑中,李萌等(2019)综述了环境DNA技术在这一领域的应用潜力及进展,强调了eDNA技术对于摸清我国水体生态系统多样性组成的重要性。由于高通量测序技术具有较高的灵敏性,该方法对于目标物种的检测具有明显的优势,比如从环境样本中排查是否存在入侵或检验检疫有害物种等。李晗溪等(2019)针对水生生态系统入侵生物的早期监测与预警论述了eDNA和宏条形码技术结合的应用前景,以及引物选择、数据准确性、参考数据库完整性等的影响。李诣远等(2019)建立了宏条形码分析的自动化流程EPPS,将基于eDNA的多样性分析进一步简化,降低了操作分析的难度,并有利于在不同研究之间进行结果共享。

高通量分子分类方法的另一个重要的应用领域是生态网络的分析和构建。在动物(如肉食动物、草食动物、吸血节肢动物等)的食性分析中(Calvignac-Spencer et al, 2013; De Barba et al, 2014),肠道内容物的准确鉴定是构建食物网络的关键环节。多样性组学分析方法可以提供可靠的食物组成信息,很好地解决了消化过程对食物形态特征破坏的问题。与此类似,多样性组学分析方法也有助于研究传粉动物的携粉组成(Bell et al, 2018; 郎丹丹等, 2018),为传粉网络的构建提供关键信息。在本专辑中,邵昕宁等(2019)使用宏条形码技术分析了川西高原的7种食肉动物的食性组成,发现狼(*Canis lupus*)、狗(*C. lupus familiaris*)、棕熊(*Ursus arctos*)、豹猫(*Prionailurus bengalensis*)等在捕食种类上存在共性与一定的分化,积累了基础数据,并有针对性地讨论了与保护生物学相关的问题,如野生动物与

人类社会发展的相关性等。

微生物多样性研究的进展在很大程度上得益于技术方法的进步。300多年前,胡克基于显微镜的观察结果发表了经典著作*Micrographia* (Hooke, 1665),将生命科学带入微观世界。而分子分类学,特别是高通量测序方法对于微生物研究的意义,可以类比为显微镜对于微生物学的贡献,让我们从核酸水平重新认知生物多样性。近期基于高通量测序技术的环境微生物研究揭示,在海洋(Sunagawa et al, 2015)、土壤(Thompson et al, 2017; Bahram et al, 2018)以及无脊椎动物宿主(Shi et al, 2016)中均存在大量无法用经典方法检测到的细菌和病毒多样性。围绕人类肠道菌群的研究也取得了突飞猛进的发展,组学研究揭示了大量未知的菌类多样性,并逐步证明肠道微生态系统对宿主具有重要的功能作用(Qin et al, 2010)。与此类似,小鼠(Xiao et al, 2015)、猪(Xiao et al, 2016)和蜜蜂(Zheng et al, 2018)等模式动物的肠道菌群研究也取得了突破,为理解宿主-共生微生物的关系提供了更广泛的模式平台。在本专辑中,肖雅倩等(2019)综述了模式动物中共生微生物研究的进展,包括斑马鱼、小鼠、猪和猕猴等,并强调了宏基因组等新方法对于本领域的推动作用。唐敏等(2019)则探讨了景观环境通过肠道微生物系统对传粉蜜蜂产生影响的可能性,提出肠道宏基因组学将有助于揭示环境-宿主之间新的互作机制。

目前大部分基于基因组学技术的多样性分析方法依赖于较短的DNA参考序列进行物种鉴定,比如线粒体或叶绿体标准条形码序列的部分片段,以及16S rRNA基因的部分区域等。这些方法的优势是可以充分利用现有的参考数据库(如DNA条形码)进行多样性分析,并通过针对靶标片段的扩增获取大量扩增子产物,减少了对样本量的要求。显而易见,标准条形码数据库的构建对于高通量分子分类的准确性至关重要,准确到物种的多样性组成信息使生物群落的组成单位不再是人为定义的可操作分类单元(operational taxonomic unit, OTU),而是具备生物学特征的进化单元。随之而来的是针对关键类群的DNA参考数据库的不断完善。在本专辑中,张雪等(2019)构建了更为完整的蜜蜂核心肠道菌16S rRNA基因数据库,补充了东方蜜蜂(*Apis cerana*)核心菌*Apibacter*等的序列,对蜜蜂肠道微生物数据库

进行了优化。同时,基于高通量测序技术构建标准DNA条形码数据库的方法(刘山林, 2019)在降低分析成本的同时,也克服了二代测序技术读长(read length)短的缺点,有助于进一步提高标准DNA条形码参考数据库的完整性。另一方面,近期的相关研究扩展了高通量分子分类研究中可用于分类的参考序列的范围,例如使用全长16S基因(Karst et al, 2018)、全线粒体(Crampton-Platt et al, 2016)、全叶绿体(Lang et al, 2018)甚至全基因组进行物种分类(Hollingsworth et al, 2016)。在本专辑中,许亚昆等(2019)总结了近期基于三代测序技术的微生物组学的技术和研究进展,指出长读长高通量测序可提高菌群多样性的分析效率,甚至可以揭示表观修饰等遗传特征。

一系列新方法特别是基于高通量测序技术的新流程,大大降低了基因组级别数据库构建的门槛,为更高效地利用DNA序列进行分类研究创造了条件。值得关注的是,我国学者近期在动植物体基因组研究领域取得了一系列令人瞩目的进展,基于线粒体基因组的昆虫系统发育研究(Song et al, 2016; Wang et al, 2016; Nie et al, 2017; Tang et al, 2019)和基于叶绿体基因组的植物系统发育研究(Li et al, 2019)都为未来分子分类学的升级提供了重要的数据和方法学基础。而基于宏基因组(metagenomics)的分析将有助于在功能基因的遗传多样性维度上来理解生物多样性组成。

从生物类群和生境类型的研究范畴来看,我国目前的生物多样性组学应用主要集中在水体多样性、动物食性分析、模式动物肠道菌群研究和物种互作的工作中。我们期待,随着相关研究方法的成熟和推广,基于高通量测序技术的多样性分析将成为群落生态学研究、生态网络构建、生态环境监测、检验检疫应用、与健康相关的微生物学等领域的常规分析手段。而本领域未来的主要挑战可能来自生物样本采集和保存方法的优化,以及生物大数据分析方法的流程化、标准化等。这些都为生物多样性科学研究在大数据时代产生新的概念、理论和方法体系提供了新的发展机遇。

## 参考文献

- Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, Bengtsson-Palme J,

- Anslan S, Coelho LP, Harend H (2018) Structure and function of the global topsoil microbiome. *Nature*, 560, 233–237.
- Bell KL, Burgess KS, Botsch JC, Dobbs EK, Read TD, Brosi BJ (2018) Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. *Molecular Ecology*, 28, 431–455.
- Belles X (2017) MicroRNAs and the evolution of insect metamorphosis. *Annual Review of Entomology*, 62, 111–125.
- Calvignac-Spencer S, Leendertz FH, Gilbert MTP, Schubert G (2013) An invertebrate stomach's view on vertebrate ecology. *BioEssays*, 35, 1004–1013.
- Crampton-Platt A, Yu DW, Zhou X, Vogler AP (2016) Mitochondrial metagenomics: Letting the genes out of the bottle. *GigaScience*, 5, 15.
- Creer S, Deiner K, Frey S, Porazinska D, Taberlet P, Thomas WK, Potter C, Bik HM (2016) The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, 7, 1008–1018.
- De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, Taberlet P (2014) DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources*, 14, 306–323.
- Fronhofer EA, Chler EMA, Walser JC, Deiner K, Altermatt F (2018) Environmental DNA reveals that rivers are conveyor belts of biodiversity information. *Nature Communications*, 7, 12544.
- Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, 6, e17497.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270, 313–321.
- Hollingsworth PM, Li DZ, Van der Bank M, Twyford AD (2016) Telling plant species apart with DNA: From barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20150338.
- Hooke R (1665) *Micrographia: Or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses, with Observations and Inquiries Thereupon*. Courier Dover Publications. John Martyn and James Allestry, London.
- Karst SM, Dueholm MS, McIlroy SJ, Kirkegaard RH, Nielsen PH, Albertsen M (2018) Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nature Biotechnology*, 36, 190–195.
- Lang D, Tang M, Hu J, Zhou X (2018) Genome-skimming provides accurate quantification for pollen mixtures. *bioRxiv*, doi: <https://doi.org/10.1101/408039>.
- Lang DD, Tang M, Zhou X (2018) Qualitative and quantitative molecular construction of plant–pollinator network: Application and prospective. *Biodiversity Science*, 26, 445–456. (in Chinese with English abstract) [郎丹丹, 唐敏, 周欣 (2018) 传粉网络构建的定性定量分子研究: 应用与展望. *生物多样性*, 26, 445–456.]
- Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, Gitzen-danner MA, Fritsch PW, Cai J, Luo Y, Wang H, Bank M, Zhang SD, Wang QF, Wang J, Zhang ZR, Fu CN, Yang J, Hollingsworth PM, Chase MW, Soltis DE, Soltis PS, Li DZ (2019) Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants*, 5, 461–470.
- Li HX, Huang XN, Li SG, Zhan AB (2019) Environmental DNA (eDNA)-metabarcoding-based early monitoring and warning for invasive species in aquatic ecosystems. *Biodiversity Science*, 27, 491–504. (in Chinese with English abstract) [李晗溪, 黄雪娜, 李世国, 战爱斌 (2019) 基于环境DNA-宏条形码技术的水生生态系统入侵生物的早期监测与预警. *生物多样性*, 27, 491–504.]
- Li M, Wei TT, Shi BY, Hao XY, Xu HG, Sun HY (2019) Biodiversity monitoring of freshwater benthic macroinvertebrates using environmental DNA. *Biodiversity Science*, 27, 480–490. (in Chinese with English abstract) [李萌, 尉婷婷, 史博洋, 郝希阳, 徐海根, 孙红英 (2019) 环境DNA技术在淡水底栖大型无脊椎动物多样性监测中的应用. *生物多样性*, 27, 480–490.]
- Li YY, Molik DC, Pfrender ME (2019) EPPS, a metabarcoding bioinformatics pipeline using Nextflow. *Biodiversity Science*, 27, 567–575. (in Chinese with English abstract) [李诣远, Molik DC, Pfrender ME (2019) 基于Nextflow构建的宏条形码自动化分析流程EPPS. *生物多样性*, 27, 567–575.]
- Linnaeus C (1758) *Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis* (in Latin), 10th edn. Laurentius Salvius, Stockholm.
- Liu SL (2019) DNA barcoding and emerging reference construction and data analysis technologies. *Biodiversity Science*, 27, 526–533. (in Chinese with English abstract) [刘山林 (2019) DNA条形码参考数据集构建和序列分析相关的新兴技术. *生物多样性*, 27, 526–533.]
- Nie RE, Breeschoten T, Timmermans MJTN, Nadein K, Xue HJ, Bai M, Huang Y, Yang XK, Vogler AP (2017) The phylogeny of Galerucinae (Coleoptera: Chrysomelidae) and the performance of mitochondrial genomes in phylogenetic inference compared to nuclear rRNA genes. *Cladistics*, 34, 113–130.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O,



- Parkhill J, Weissenbach J, Antolin M, Artiguenave F, Blottiere H, Borruel N, Bruls T, Casellas F, Chervaux C, Cultrone A, Delorme C, Denariatz G, Dervyn R, Forte M, Friss C, van de Guchte M, Guedon E, Haimet F, Jamet A, Juste C, Kaci G, Kleerebezem M, Knol J, Kristensen M, Layec S, Le Roux K, Leclerc M, Maguin E, Melo Minardi R, Oozeer R, Rescigno M, Sanchez N, Tims S, Carlsen T, Varela E, de Vos W, Winogradsky Y, Zoetendal E, Bork P, Ehrlich SD, Wang J (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464, 59–65.
- Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System ([www.barcodinglife.org](http://www.barcodinglife.org)). *Molecular Ecology Notes*, 7, 355–364.
- Shao XN, Song DZ, Huang QW, Li S, Yao M (2019) Fast surveys and molecular diet analysis of carnivores based on fecal DNA and metabarcoding. *Biodiversity Science*, 27, 543–556. (in Chinese with English abstract) [邵昕宁, 宋大昭, 黄巧雯, 李晟, 姚蒙 (2019) 基于粪便DNA及宏条形码技术的食肉动物快速调查及食性分析. *生物多样性*, 27, 543–556.]
- Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu JG, Holmes EC, Zhang YZ (2016) Redefining the invertebrate RNA virosphere. *Nature*, 540, 539–543.
- Song F, Li H, Jiang P, Zhou X, Liu J, Sun C, Vogler AP, Cai W (2016) Capturing the phylogeny of Holometabola with mitochondrial genome data and Bayesian site-heterogeneous mixture models. *Genome Biology and Evolution*, 8, 1411–1426.
- Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A (2015) Structure and function of the global ocean microbiome. *Science*, 348, 1261359.
- Tang M, Zou Y, Su QZ, Zhou X (2019) A new perspective on landscape impact in bee populations: Considering the bee gut microbiome. *Biodiversity Science*, 27, 516–525. (in Chinese with English abstract) [唐敏, 邹怡, 苏秦之, 周欣 (2019) 洞察景观环境影响蜜蜂之新视角: 肠道微生物. *生物多样性*, 27, 516–525.]
- Tang P, Zhu JC, Zheng BY, Wei SJ, Sharkey M, Chen XX, Vogler AP (2019) Mitochondrial phylogenomics of the Hymenoptera. *Molecular Phylogenetics and Evolution*, 131, 8–18.
- Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 551, 457–463.
- Wang Y, Liu X, Garzón-Orduña IJ, Winterton SL, Yan Y, Aspöck U, Aspöck H, Yang D (2016) Mitochondrial phylogenomics illuminates the evolutionary history of Neuropterida. *Cladistics*, 33, 617–636.
- Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences, USA*, 74, 5088–5090.
- Xiao L, Estellé J, Kiilerich P, Ramayo-Caldas Y, Xia Z, Feng Q, Liang S, Pedersen AO, Kjeldsen NJ, Liu C, Maguin E, Doré J, Pons N, Le Chatelier E, Prifti E, Li J, Jia H, Liu X, Xu X, Ehrlich SD, Madsen L, Kristiansen K, Rogel-Gaillard C, Wang J (2016) A reference gene catalogue of the pig gut microbiome. *Nature Microbiology*, 1, 16161.
- Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, Li X, Long H, Zhang J, Zhang D, Liu C, Fang Z, Chou J, Glanville J, Hao Q, Kotowska D, Colding C, Licht TR, Wu D, Yu J, Sung JJY, Liang Q, Li J, Jia H, Lan Z, Tremaroli V, Dworzynski P, Nielsen HB, Bäckhed F, Doré J, Le Chatelier E, Ehrlich SD, Lin JC, Arumugam M, Wang J, Madsen L, Kristiansen K (2015) A catalog of the mouse gut metagenome. *Nature Biotechnology*, 33, 1103–1108.
- Xiao YQ, Liu C, Xiao L (2019) The role of model animals in the study of symbiotic microorganisms. *Biodiversity Science*, 27, 505–515. (in Chinese with English abstract) [肖雅倩, 刘传, 肖亮 (2019) 模式动物在共生微生物研究中的作用. *生物多样性*, 27, 505–515.]
- Xu YK, Ma Y, Hu XX, Wang J (2019) Analysis of prospective microbiology research using third-generation sequencing technology. *Biodiversity Science*, 27, 534–542. (in Chinese with English abstract) [许亚昆, 马越, 胡小茜, 王军 (2019) 基于三代测序技术的微生物组学研究进展. *生物多样性*, 27, 534–542.]
- Zhang X, Li XA, Su QZ, Cao QN, Li CY, Niu QS, Zheng H (2019) A curated 16S rRNA reference database for the classification of honeybee and bumblebee gut microbiota. *Biodiversity Science*, 27, 557–566. (in Chinese with English abstract) [张雪, 李兴安, 苏秦之, 曹棋钠, 李晨伊, 牛庆生, 郑浩 (2019) 用于蜜蜂和熊蜂肠道微生物分类的细菌16S rRNA数据库优化. *生物多样性*, 27, 557–566.]
- Zheng H, Steele MI, Leonard SP, Motta EVS, Moran NA (2018) Honey bees as models for gut microbiota research. *Lab Animal*, 47, 317–325.
- Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, 2, 4.

(责任编辑: 周玉荣)





•综述•

# 环境DNA技术在淡水底栖大型无脊椎动物多样性监测中的应用

李萌<sup>1</sup> 尉婷婷<sup>1</sup> 史博洋<sup>1</sup> 郝希阳<sup>1</sup> 徐海根<sup>2</sup> 孙红英<sup>1\*</sup><sup>1</sup> (南京师范大学生命科学学院江苏省生物多样性与生物技术重点实验室, 南京 210023)<sup>2</sup> (生态环境部南京环境科学研究所, 南京 210042)

**摘要:** 环境DNA (eDNA)是指生物有机体在环境中(例如土壤、沉积物或水体)遗留下的DNA片段。eDNA技术是指从环境中提取DNA片段进行测序以及数据分析来反映环境中的物种或群落信息。与传统方法相比, eDNA技术具有高灵敏度、省时省力、无损伤等优点。目前, eDNA技术已成为一种新的水生生物监测方法, 主要应用于水生生物的多样性研究、濒危和稀有动物的物种状态及外来入侵物种扩散动态的监测等。本文从eDNA技术在水生生物多样性监测应用领域的发展历程、eDNA技术的操作流程以及其在监测淡水底栖大型无脊椎动物方面的应用进展、技术优势和局限性五个方面进行了综述。最后, 本文对eDNA技术在淡水底栖大型无脊椎动物多样性监测应用的发展趋势和前景作出展望。

**关键词:** 宏条形码; 第二代测序技术; 外来入侵物种; 淡水生态系统; 细胞色素c氧化酶亚基I (COI)

## Biodiversity monitoring of freshwater benthic macroinvertebrates using environmental DNA

Meng Li<sup>1</sup>, Tingting Wei<sup>1</sup>, Boyang Shi<sup>1</sup>, Xiyang Hao<sup>1</sup>, Haigen Xu<sup>2</sup>, Hongying Sun<sup>1\*</sup><sup>1</sup> Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023<sup>2</sup> Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment, Nanjing 210042

**Abstract:** Environmental DNA (eDNA) refers to DNA fragments that organisms leave behind in their surrounding environment (such as soil, sediment and water). eDNA technology sequences these DNA fragments and can provide information on taxonomic composition of benthic macroinvertebrate communities. Compared with traditional biological survey methods, eDNA technology is more sensitive, efficient and noninvasive. As a novel method for surveying aquatic organisms, eDNA techniques have been widely used in biodiversity assessments of aquatic organisms, including monitoring of endangered, rare and invasive species. In this review, we summarize recent developments in eDNA technology and focus primarily on the operational procedure and its application for freshwater benthic macroinvertebrate analyses. Finally, we discuss the advantages and potential caveats of current eDNA practices.

**Key words:** metabarcoding; next-generation sequencing; invasive species; freshwater ecosystem; cytochrome c oxidase subunit I (COI)

中国淡水水域广阔, 淡水生态系统物种丰富。其中, 淡水底栖大型无脊椎动物是物种多样性和生物量最为丰富的类群之一。在国际上, 淡水底栖大型无脊椎动物被公认为水域生态系统的指示生物,

在维持淡水水域的生态平衡、响应人类活动胁迫、指示淡水生态系统健康状况等方面具有关键作用 (Smith et al, 2007; Yeom & Adams, 2007)。

长期以来, 淡水底栖大型无脊椎动物受到国内

收稿日期: 2018-08-17; 接受日期: 2018-12-27

基金项目: 生态环境部生物多样性保护专项

\* 通讯作者 Author for correspondence. E-mail: sunhongying@njnu.edu.cn

外学者们的广泛关注(Rosenberg & Resh, 1993; 段学花等, 2010; 王艳杰等, 2012), 相关研究丰富了对底栖动物多样性的认识, 同时积累了大量相关基础数据(徐海根, 2013)。近半个世纪以来, 人类活动的增加, 特别是淡水资源的过度开发和利用, 使淡水生态系统遭受严重破坏甚至面临退化的风险。此外, 环境污染、外来物种入侵与气候变化等因素的影响加剧了淡水底栖无脊椎动物所面临的威胁。以淡水蟹为例, 中国是全球淡水蟹物种多样性最高的国家, 所记录物种数与整个新热带界的物种总数相当(楚克林等, 2018)。然而, 其中将近75%的物种的生存面临威胁, 或因数据不足其受威胁状况需要更为深入的调查研究。更为严重的是, 近20年来, 受水域生态环境变化的影响, 一些先前被评估为无危的物种其多样性资源衰退明显, 以至于临近濒危(曹亮等, 2016)。但是, 面临如此危机, 我国仍未针对淡水底栖大型无脊椎动物生存现状建立行之有效的监测和评估体系, 特别是某些生态域狭窄的物种, 其受威胁状况与保护力度尚未得到有关管理部门和公众的广泛关注。在采用传统野外调查方法对底栖大型无脊椎动物进行监测时, 常受限于水域的复杂环境条件、各类群多样化的生活史以及隐秘的生活习性等。尤其是在激流或深水水体采样时, 需投入大量人力并辅以重型采样设备, 使区域物种多样性的全面监测难以在短时间内完成。因此, 目前亟需高效、快速、简便的方法, 以对淡水水域底栖大型无脊椎动物的多样性进行全面系统地监测。

环境DNA (environmental DNA, eDNA)是指从生物体生活环境中(例如土壤、沉积物或水体等)直接提取到的DNA片段的总和, 主要包含不同生物体释放到环境中的胞内DNA和细胞裂解或死亡后的胞外DNA (Taberlet et al, 2012; Rees et al, 2015)。近年来出现的eDNA条形码调查技术(以下简称eDNA技术)从环境样品中直接提取DNA, 并利用DNA测序技术对环境中的目标物种进行定性或定量分析(Ficetola et al, 2008)。国际上已经将该技术应用于监测各种淡水水生生物, 例如鱼类(Takahara et al, 2012, 2013)、两栖类(Ficetola et al, 2008)、爬行类(Piaggio et al, 2014), 以及腹足类(Goldberg et al, 2013)、甲壳类(Tréguier et al, 2014; Dougherty et al, 2016; Geerts et al, 2018)等无脊椎动物。然而, 利用eDNA技术监测淡水水生生物的研究, 尤其是涉及

淡水底栖大型无脊椎动物多样性的研究在我国还鲜有报道(Cai et al, 2017)。

本文综述了eDNA技术及其在水生生物多样性监测领域的发展历程、基于eDNA条形码及宏条形码技术的操作流程, 以及该技术在监测淡水底栖大型无脊椎动物方面的应用进展、技术优势和局限性, 最后指出利用eDNA开展针对我国淡水底栖大型无脊椎动物多样性监测的应用前景, 以期为我国淡水生态系统的多样性保护、管理和决策提供必要的理论依据和技术支撑。

## 1 eDNA技术在淡水生物多样性监测领域的发展历程

### 1.1 定性研究

eDNA技术在定性研究中主要通过常规PCR方法与产物片段分析对研究水域中的目标物种进行监测, 来判断某目标物种在研究水域内是否存在及其分布特征。随着eDNA技术在淡水生态系统多样性监测中的定性研究不断推广, eDNA研究对象从脊椎动物逐渐拓展到无脊椎动物类群。Ficetola等(2008)通过对水样中提取的微量eDNA进行PCR反应与片段测序成功监测到入侵物种美洲牛蛙(*Rana catesbeiana*)的存在, 并由此开启了eDNA技术在淡水动物物种监测中的研究。Jerde等(2011)将eDNA技术用于监测连接密西西比河和大湖盆地的芝加哥地区水道中入侵的两类物种——鳙鱼(*Hypophthalmichthys nobilis*)和鲢鱼(*H. molitrix*), 结果表明该地区将面临物种入侵的威胁, 且eDNA技术检测灵敏度明显高于传统渔业监控手段。随后, Deiner和Altermatt (2014)针对淡水无脊椎动物长刺溞(*Daphnia longispina*)和淡水珠蚌(*Unio tumidus*)设计物种特异性引物, 通过PCR反应和测序确认了瑞士境内河流中这两个物种eDNA的存在。

### 1.2 定量研究

通过定量PCR (qPCR), eDNA技术不仅可以确定水样中目标生物的存在, 同时可以通过量化eDNA数据对物种的丰富度和生物量进行表征。数字PCR技术(digital PCR或dPCR)与经典qPCR技术相比, 更适合检测微量的DNA, 同时对抑制剂耐受性好, 不需要依赖参考标准(Whale et al, 2012), 这将改进DNA检测和定量(Vogelstein & Kinzler, 1999; Hindson et al, 2011), 从而实现物种eDNA的绝对

定量。

Takahara等(2012)首次利用eDNA技术对日本淡水湖中的普通鲤鱼(*Cyprinus carpio*)进行生物量的量化分析,研究表明水体中eDNA浓度与鲤鱼的生物量正相关,并且基于eDNA浓度所得的生物量数据能推测自然环境中鲤鱼的潜在分布情况。Doi等(2017)采用eDNA技术对日本岛内栖息在湿地生态系统中的一种小型濒危水生半翅类昆虫的分布状况进行了成功预测,但未发现eDNA检出率与捕获物种数量之间具有显著的统计学关系。Dougherty等(2016)将传统诱捕小龙虾的方法与eDNA技术相结合,对其eDNA技术在监测美国中西部地区内陆湖中入侵种锈斑小龙虾(*Orconectes rusticus*)及估测其相对丰度的能力进行研究,结果表明虽然eDNA技术检测物种的成功率随该物种相对丰度的增加而增加,但是水样中eDNA的拷贝数与该物种相对丰度的相关性较差。随后,Larson等(2017)采用相同方法对美国加州和内华达州大湖中的锈斑小龙虾和信号小龙虾(*Pacifastacus leniusculus*)进行研究,结果表明水样中的eDNA拷贝数与通过诱捕龙虾估计的相对丰度之间相关性较差。以上研究表明,带有坚硬外骨骼或者贝壳的淡水底栖大型无脊椎动物可能因其eDNA捕获难度较大,从而导致其eDNA浓度与实际生物量之间的关系并不明显。另有研究表明水温对鱼类等的生长、代谢具有一定的影响,进而影响其eDNA的释放速率,以至于影响eDNA技术对生物量的估计(Takahara et al, 2011)。水体pH值与光照也会对eDNA的释放速率和降解速率产生影响,进而影响生物量估测的准确性(Strickler et al, 2015)。

因此,目前eDNA技术尚不能完全实现对任意物种的定量估计,尤其是对甲壳类和贝类生物量的评估仍需进一步深入研究。除此之外,由于不同生物类群之间的差异,其释放的eDNA总量和速度各不相同(Geerts et al, 2018),因而难以实现跨类群的定量分析。总之,利用eDNA技术估测生物量的准确度以及在未来的适用性还有待进一步验证。

### 1.3 多物种生物多样性监测

随着测序技术的发展,第二代测序(next-generation sequencing, NGS)技术应运而生,该技术能实现对几十万到百万条DNA同时进行序列测定。NGS技术与eDNA技术相结合所产生的

eDNA宏条形码(eDNA metabarcoding)技术,主要用于生物多样性监测,同时对资源量具备一定的定量描述能力。

近年来,eDNA宏条形码技术已广泛应用于生态学研究领域。该技术通过从环境样品中提取DNA,借助第二代高通量测序获取其物种组成,进一步探究群落水平上的生物多样性。Valentini等(2016)利用eDNA宏条形码技术对河流、溪流和湖泊中的两栖类和硬骨鱼类进行生物多样性监测,结果表明,与传统调查相比eDNA宏条形码技术具有更高的检测成功率:在两栖动物中,eDNA宏条形码技术的检测概率为97%,而传统调查方法为58%;在鱼类中,eDNA宏条形码技术在89%的研究地点中检测到的分类群数量与传统方法相同或更多。因此,eDNA宏条形码技术有望成为集高效、精准和标准化为一体的新一代生物多样性监测工具,具有对环境样品中生物物种进行有效监测的潜力。目前,该技术在淡水无脊椎动物的生物多样性研究中已有成功应用。Klymus等(2017)基于线粒体16S rRNA基因对劳伦提亚大湖和周围水域中的入侵双壳类和腹足类进行eDNA宏条形码研究,结果表明该方法对软体动物物种的鉴别效果比形态学分析更加快速准确。

总体而言,eDNA技术在淡水生态系统的研究和监测中经历了最初的定性研究(物种监测)到定量研究(生物量估测),再到全水域的生物多样性监测研究。其研究对象从单一物种扩展到多物种甚至整个群落,从大型脊椎动物(如鱼类、两栖类、爬行类等)扩展到大型底栖无脊椎动物(如贝类、淡水甲壳类和水生昆虫等)。

## 2 eDNA技术操作流程

eDNA技术的操作流程主要包括4个环节,即样品采集、eDNA捕获、eDNA提取和eDNA分析等(图1)。考虑到淡水底栖大型无脊椎动物自身的特点,本文综合了涉及淡水生物特别是淡水底栖大型无脊椎动物eDNA研究和应用的成熟经验,对操作流程中各个环节需要注意的细节进行了简要说明和讨论。

### 2.1 样品采集与保存

淡水底栖大型无脊椎动物类群主要包括扁形动物门、环节动物门、软体动物门和节肢动物门四大类。这些动物类群主要是通过粪便释放脱落的上



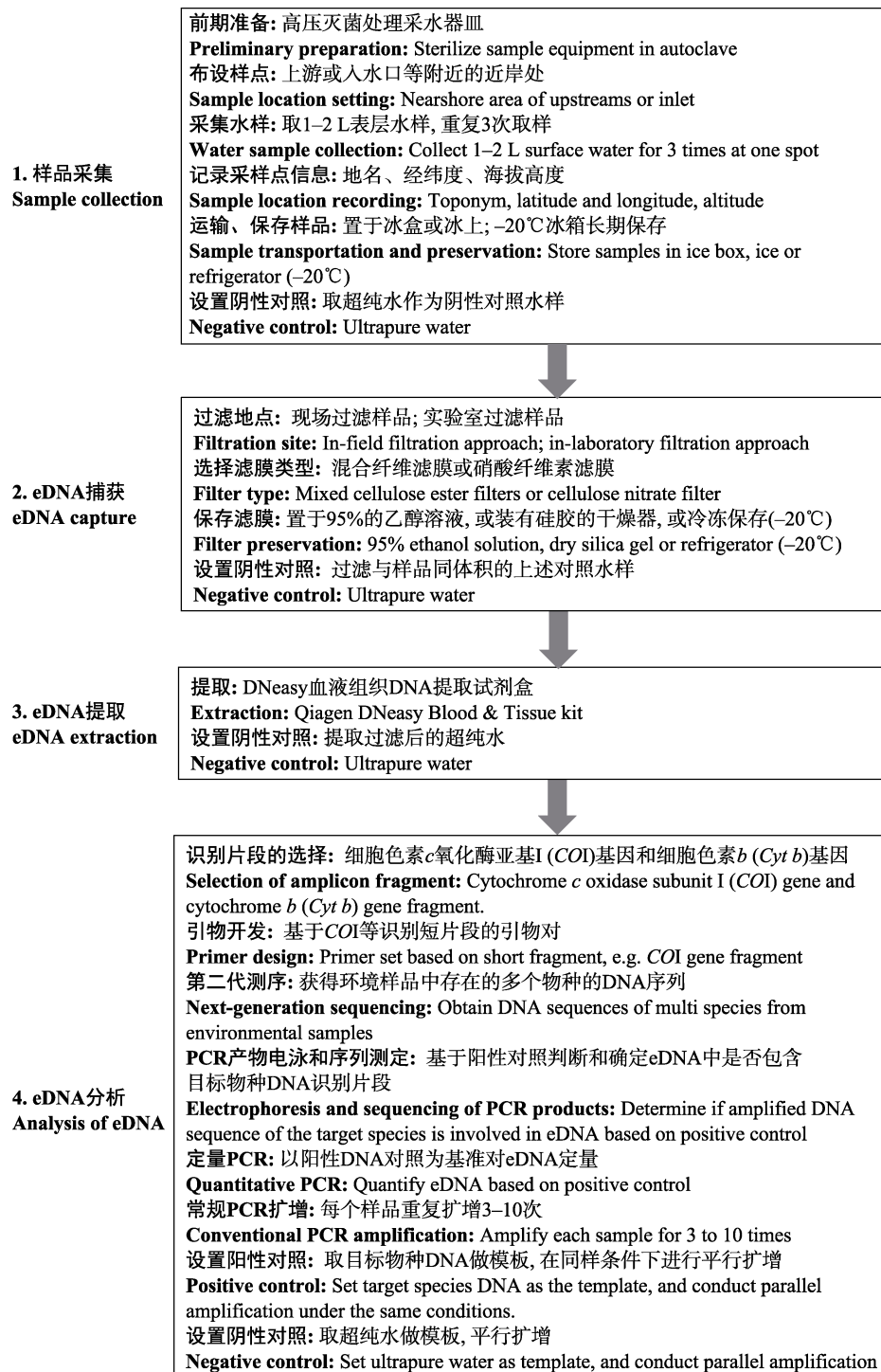


图1 利用eDNA宏条形码技术开展淡水底栖大型无脊椎动物多样性监测的主要流程

Fig. 1 The framework of biodiversity monitoring for freshwater benthic macroinvertebrates using eDNA metabarcoding

皮组织细胞内的DNA; 其次是通过死亡个体的腐烂分解释放DNA (Merkes et al, 2014)。

考虑到动物机体遗留的eDNA在水中的分布具有异质性, 水样采集时不宜对水体造成过大的扰动

(比如行船等, 以免对eDNA的获取造成不利影响), 因此一般在水体的边缘近岸处采集。对于取样点的布设因水体而异, 在湖泊等静水水体一般依据湖泊形状和周边的出入水口位置等, 可每隔一定距离设



置1个取样点;而在河流、溪流等流水水体,多在上游设置取样点。考虑到底栖动物栖息地的多样化,其释放的eDNA在沉积物 and 水中均有分布。已有的研究显示,来自湖泊或河流表层水体中的eDNA反映了最近时期内的生物多样性,而沉积物中的eDNA反映的生物多样性在时间跨度上有很大程度的不确定性,需要谨慎对待或进行进一步的研究(Deiner et al, 2017)。Shaw等(2016)比较了水体和沉积物表层样本中获得的鱼类物种丰富度估计值,他们在两类样本中都检测到了目标物种,但是与沉积物样本相比,水样的物种丰度估计值与水体实际存在的物种丰度更为一致。Moyer等(2014)针对不同的取样深度对检测率的影响进行比较,发现表层和底层水样的检测概率较高,但在低生物密度的情况下,均不具备100%的检测成功率。因此,现行的大多数研究仍将采集表层水样作为获取淡水底栖大型无脊椎动物eDNA的主要途径。

为了避免外来污染以及交叉污染,在采样前期需预先对采水器皿进行高压灭菌处理,操作人员在采样过程中须始终佩戴一次性手套,每采样一次更换一次手套。为了提高水样中物种的覆盖度,应在每一个采样点重复3次取样(Thomsen et al, 2012a)。

## 2.2 eDNA捕获

目前,用于eDNA捕获的方法包括沉淀法(Thomsen et al, 2012b)和过滤法(Goldberg et al, 2011)。沉淀法通常采集15 mL的水样,而过滤法采集水样量为1–2 L。相对于沉淀法,过滤法可以捕获到更多物种的DNA(Deiner et al, 2015; Majaneva et al, 2018)。研究发现,对于淡水底栖无脊椎动物来说,使用过滤法捕获DNA可以提高eDNA的检测成功率(Deiner et al, 2015; Eichmiller et al, 2016)。其中,过滤法还分为现场过滤和实验室过滤,前者在野外即时过滤水样,将获得的滤膜低温保存,从而减缓样品中eDNA的降解,这对于远距离的野外采集非常重要;后者无需野外进行过滤,可以减少现场操作的时间,水样置于冰上保存,并保证运输时间最长不超过5 h(Deiner et al, 2015),带回实验室过滤后将滤膜–20℃冷冻保存。

目前,对于滤膜规格和类型的选择并没有统一的标准,多数研究采用0.45 μm硝酸纤维素滤膜,个别研究采用0.75 μm或1.5 μm玻璃纤维滤膜或者聚碳酸酯(Wilcox et al, 2013; Deiner et al, 2015)。不

同的滤膜具有不同的DNA捕获率,其中使用混合纤维滤膜或硝酸纤维素滤膜的DNA捕获效果最佳(Majaneva et al, 2018)。

## 2.3 eDNA提取

目前有多种不同的eDNA提取方法,包括氯仿提取(Renshaw et al, 2015)、裂解细胞的物理破裂(Jerde et al, 2011)和二氧化硅提取法(Goldberg et al, 2011)。然而,选择不同的捕获和提取方法可能极大地影响目标生物体的eDNA检测结果。Deiner等(2015)测试了不同的eDNA捕获和提取方案对检测淡水生物多样性的影响,结果表明采用过滤法与DNeasy血液组织DNA提取试剂盒相结合的方法提取效果更好。

## 2.4 eDNA分析

### 2.4.1 宏条形码选择和通用引物设计

水样中的eDNA含有可用于分辨物种的基因片段。基于eDNA的宏条形码分析方法,是通过选择合理的宏条形码和设计通用引物以及第二代高通量测序对环境样品中存在的群落进行分类鉴别。研究表明线粒体DNA (mitochondrial DNA, mtDNA)相比于细胞核DNA (nuclear DNA)更适合作为eDNA宏条形码靶标,尤其是细胞色素c氧化酶亚基I (COI)基因和细胞色素b (Cyt b)基因(Andújar et al, 2018)。其原因有三:(1)细胞中mtDNA的拷贝数要比核DNA的多;(2) mtDNA具有相对保守性,含有更多区分物种的高变区及易于设计通用引物的保守区,有利于靶标基因的扩增;(3)基因组数据库中含有大量mtDNA的序列信息,便于对获取的eDNA进行序列比对、搜索和快速识别,以确认其物种组成。

对于淡水底栖大型无脊椎动物来说,COI条形码片段具有良好的分类学辨识度,可参考的数据库已涵盖最常见的淡水动物类群(Fernández et al, 2018)。因此,设计基于COI条形码片段的通用引物成为利用eDNA宏条形码技术监测淡水大型无脊椎动物群落的明智选择。最近,Elbrecht和Leese (2017)开发了专门针对淡水大型无脊椎动物的通用引物(上、下游引物名称分别为BF和BR),特别是引物BF2和BR2可以很好地降低引物偏倚,但该引物对高度降解的eDNA并不适用。Vamos等(2017)针对淡水大型无脊椎动物开发了基于COI基因短片段(178 bp)的1对引物fwh1 (包括fwhF1和fwhR1)用于宏条形码研究,它对高度降解的eDNA有较好的扩

增效果。

### 2.4.2 结果分析

eDNA宏条形码技术利用通用引物扩增水样中提取的eDNA,经高通量测序后得到大量包含分类学信息的数据。由于得到的序列片段较短,而对于环境样品通常很难确定某条序列来自哪个物种,因此对于海量的短片段序列,要进行拼接以及进一步分析具有巨大的挑战。一般来说,对于数据结果的分析过程主要包括:(1)质量过滤和序列拼接。首先对原始reads数据进行测序噪声去除、移除嵌合体、掺杂序列和核线粒体假基因等过滤处理。随后根据reads之间的重叠区进行拼接组装成供精准分析的优化序列数据。(2)OTU (operational taxonomic units,可操作分类单元)聚类和物种注释。对所得优质序列按照大于97%的相似度进行OTU聚类,即大于97%的序列相似度视作同种,然后选取每个OTU中最长的序列为代表序列,将该序列与参考数据库相比对进行物种注释分析,获得各样品中OTU的丰度表和物种注释结果。(3)多样性分析。根据OTU注释结果和各样品中OTU的丰度表,得到各个样品的物种丰度表,并对研究区域内的群落物种组成、 $\alpha$ 多样性和 $\beta$ 多样性进行分析。

### 2.5 阳性和阴性对照

在水样采集、eDNA捕获和提取以及PCR扩增过程中都需要设置阴性对照,以检测实验操作过程中是否受到污染以及确定污染的来源。同时在对目标物种进行eDNA分析时还需设置阳性对照作为定量分析、电泳以及测序结果的标准参考依据,避免出现非目标物种eDNA的扩增,以确保实验结果的准确性和可靠性。这也有助于及时监控实验过程中是否受到交叉污染。

## 3 应用eDNA技术监测淡水底栖大型无脊椎动物

### 3.1 物种多样性监测

近年来,人们对eDNA宏条形码技术在淡水底栖大型无脊椎动物的生物多样性监测的可行性进行了大量探索。据报道,在河流生态系统的生物监测中,传统采样方法仅能反映局部的群落结构变化,而eDNA宏条形码技术可以对大尺度区域内的生物多样性甚至整个景观的生物多样性进行监测,但严格的生物信息学阈值标准与参考数据库数据的不完整导致eDNA宏条形码技术无法检出部分类群

(Deiner et al, 2016)。

与传统监测方法相比,eDNA技术对物种的监测具有更高的检出率。Fernández等(2018)将室内群落模拟实验、基于eDNA技术的野外实验,与基于传统监测方法的野外实验在对河流内大型无脊椎动物监测中的应用进行了对比分析,结果表明基于COI宏条形码的eDNA技术的物种检出率明显高于传统监测方法和基于18S rDNA宏条形码的eDNA技术。因此在监测不同的生物类群时,标记基因的恰当选择对eDNA技术在生物多样性监测中的应用具有重要作用。

总之,eDNA宏条形码技术在淡水底栖大型无脊椎动物多样性的监测中具有巨大潜力,但在淡水水域中底栖大型无脊椎动物的全面准确鉴定方面面临巨大挑战。

### 3.2 外来入侵物种监测

早期预警和快速监测对于入侵物种的控制和清除十分重要,但因在入侵早期阶段物种数量较低,传统的监测方法因耗时长与效率低而难以及时发现,eDNA技术以其高的灵敏度极大地增加了入侵物种早期发现的可能性(Goldberg et al, 2013; Tréguier et al, 2014)。克氏原螯虾(*Procambarus clarkii*)已被列入全球百大外来入侵物种名单(Lowe et al, 2000)。在我国长江中下游地区,克氏原螯虾已成为当地常见的一种经济食用动物。但不容忽视的是,这种入侵动物已对我国淡水水域生态系统及生物多样性造成了严重威胁(蔡凤金等, 2010)。Cai等(2017)对克氏原螯虾在云南元阳哈尼梯田中分布情况的监测研究发现,采用eDNA技术的检出率明显高于传统诱捕法。此外,与传统的诱捕法相比,eDNA技术作为一种非侵入性方法,对水生生物无附带伤害且易于现场实施操作。另一个全球百大入侵生物之一——斑马贻贝(*Dreissena polymorpha*),原产于里海和黑海,目前已扩散至英国、西欧、加拿大和美国,对当地的软体动物区系、淡水生态系统造成危害,而货船压舱水是该物种传播的重要途径之一(Lowe et al, 2000)。美国研究人员将eDNA技术结合光透射光谱(light transmission spectroscopy, LTS) DNA检测技术实现对船只压舱水所携带外来入侵物种的快速检测,降低了外来入侵物种的引入和传播的风险(Egan et al, 2015)。

## 4 eDNA在淡水底栖无脊椎动物研究中的优势与局限性

### 4.1 优势

相对于传统的野外调查监测方法, eDNA技术具有其独特的优势。主要包括以下几个方面: (1)无损伤、无公害采样。相对于传统的捕获式采样, eDNA技术只采集动物机体栖息地的水样, 并不会危害到目标物种或者采集区域内的其他物种, 对采样点周围的生态系统干扰较少。(2)不依赖动物样本。克服了传统的动物野外调查方法对动物样品自身的依赖性, 避免了未遇见或未捕获样本而造成对物种多样性及其分布状况的低估或误判。(3)采样方法简单, 耗时少、成本低。相对于传统的采样方法, 比如使用采泥器、踢网以及诱捕等, 采集水样既容易又耗时少、成本低。该优势在监测外来入侵的亚洲鲤鱼*Hypophthalmichthys molitrix*和*H. nobilis*等水生动物调查工作中得到佐证(Jerde et al, 2011)。(4)采样受限小。由于eDNA主要采集水样, 与传统的调查监测方法相比, 在采样环节对环境要求低, 可以最大限度地降低因天气状况、水位、物候等环境条件变化对采样的影响和限制。(5)灵敏度高。研究表明对于监测淡水水生生物, 尤其是外来入侵或者濒危乃至稀有动物, 其灵敏度要高于传统的监测方法(Dejean et al, 2012)。(6)快速和高效检测群落多样性。结合了新一代高通量测序技术的eDNA宏条形码方法可以在短时间内对混合的环境样品包含的所有物种进行识别和检测, 具有大规模、大尺度生物多样性监测的潜力。(7)有利于制定规范的鉴定标准, 避免了传统方法受样品条件限制(如非成熟个体或严重受损的样品)、操作者实践经验(比如在野外调查采集、样品筛选和分拣, 物种分类鉴定等工作经验)和专业知识等限制的不利影响。(8)数据的标准化程度高, 便于质量控制。

可见, eDNA技术的出现为淡水底栖大型无脊椎动物的多样性监测与研究开辟了一条全新的思路。

### 4.2 局限性

虽然eDNA技术在多个方面优于传统调查方法, 但目前该技术仍存在一系列不足, 比如: (1)尚未形成一套通用的标准化操作流程。近年来多项研究表

明, 不同的采样方案和eDNA提取方法可导致不同的检测结果(Deiner et al, 2015)。因此, 有必要对该技术流程的各个环节进行优化, 建立统一、规范的标准化方案。(2)时间和空间的不一致性。eDNA的浓度和分布受水生动物生理活动、水流运动的影响(Eichmiller et al, 2014)。同时, eDNA在环境中的降解速率也受到多种非生物因素的影响。这就需要模拟并建立监测结果与物种时空分布之间的相关性, 从而提高监测结果的精确度。(3)污染问题。从水样采集到eDNA分析的每一步操作都可能存在交叉污染的风险。尤其是PCR扩增中容易发生交叉污染, 而高通量测序技术可将污染效果进一步放大, 从而导致监测结果出现错误(Willerslev & Cooper, 2005; Champlot et al, 2010)。解决方法包括: 对实验操作流程进行严格的消毒灭菌, 设置阴性对照来检测污染问题等。(4)假阳性及假阴性问题。假阳性和假阴性对于监测濒危物种以及稀有物种、甚至被认为已灭绝的物种来说十分棘手。假阳性是指检测到目标物种的环境DNA但实际研究区域内并不存在该物种。这种结果有可能是样品中的目标物种DNA来源于捕食者的粪便、排放的污水或废水(Jerde et al, 2011), 或者是由于引物的特异性不高, 出现非目标DNA的扩增现象(Kristine et al, 2014)。此问题的解决办法是: 要避免出现污染, 首先在采样期间使用的设备必须经过严格消毒以及在PCR反应中设置阳性对照来检验扩增结果是否存在假阳性。其次是根据研究目的设计合适的引物。对于研究的某一目标物种来说, 需要设计特异性高的引物, 避免非目标物种的扩增。而对于研究的整个群落来说, 需要设计通用性引物。而假阴性是指研究区域内实际存在目标物种, 但却检测不到该物种的环境DNA。这种现象可能是由于引物的偏倚性或者抑制PCR扩增的杂质、腐殖质等的存在(McKee et al, 2015; Balasingham et al, 2016)。此问题的解决方法是: 提高引物的适配性或通用性, 对样品进行稀释以降低抑制作用或者使用抑制剂去除试剂盒(Tsai & Olson, 1992; McKee et al, 2015)。

由此可见, eDNA技术无法完全替代传统的生物多样性监测方法, 但却可以作为一种重要的补充工具, 用于快速监测目标物种的分布和水域内底栖大型无脊椎动物群落的多样性。



## 5 eDNA技术在我国淡水底栖大型无脊椎动物多样性监测中应用的前景与展望

中国是全球生物多样性最丰富的国家之一(Xu et al, 2008)。但由于缺乏深入持久的调查监测, 中国的淡水生物多样性, 尤其是淡水底栖大型无脊椎动物多样性的基础数据依然十分匮乏。许多底栖无脊椎动物类群的数据仍停滞在20世纪中、晚期, 鲜有更新。相比其他水生动物, 如鱼类和鲸类等, 公众对底栖无脊椎动物多样性的认知还十分薄弱、甚至贫乏。同时因疏于监测和量化、缺乏管理依据等, 淡水底栖大型无脊椎动物的多样性保护已然成为水生生物乃至整个生物多样性保护严重的“短板”之一。此外, 中国是全球遭受外来生物入侵危害较为严重的“重灾区”之一(万方浩等, 2002; 万方浩, 2009; 李博和马克平, 2010; 丁晖等, 2015)。据最新统计, 已知的外来入侵生物有667种(徐海根和强胜, 2018), 其中包括一些危害严重的外来入侵淡水底栖大型无脊椎动物, 如克氏原螯虾、尖膀胱螺(*Physella acuta*) (郭云海等, 2009; 杨海芳等, 2014)等, 给我国淡水水域的生态系统安全带来严峻挑战。因此, 我们提倡将eDNA技术尽快应用于我国淡水底栖大型无脊椎动物的多样性监测, 包括对外来入侵的底栖大型无脊椎动物等水生生物的早期预警, 从而快速、准确、高效地掌握整个淡水水域底栖大型无脊椎动物的多样性状况, 包括群落和生态系统层面, 尽早发现并确认外来生物入侵的地区、分布及扩散情况等。这些监测数据和信息可以为政府和其他相关决策部门针对水域生态环境改造、大型水利或基建工程建设、工农业开发区建设、经济动植物引种和生态修复工程等项目提供科学依据, 从而最大限度地避免或降低水域生态系统的健康状况遭受危害。

eDNA技术, 尤其是eDNA宏条形码技术可以对整个区域的淡水底栖动物群落进行多样性监测, 极大地改善了对水域生态系统的监测模式。在今后的eDNA宏条形码研究中, 还需要不断加强和完善底栖大型无脊椎动物的DNA条形码数据库, 特别是针对当地本土物种的数据库构建, 以期为本地区淡水底栖大型无脊椎动物的生物多样性监测提供更准确的数据支撑。除此之外, 针对淡水底栖大型无

脊椎动物的特点, 深入研究并阐明温度、光照等环境因子对于其eDNA释放以及降解速率的影响, 使eDNA更能准确反映水域环境中底栖大型无脊椎动物的相对丰度和生物量的变化等。同时, 在进一步研究中制定标准化的eDNA监测技术并与传统监测技术相结合将有利于规范水域生态环境的现状调查和评估, 以及准确了解水域水质情况或变化趋势, 为制定切实可行、具体有效的保护措施提供依据。并且, 可以客观反映生态修复工程的改善效果, 有力推动我国淡水生态系统的生物多样性保护和可持续发展。

### 参考文献

- Andújar C, Arribas P, Yu DW, Vogler AP, Emerson BC (2018) Why the COI barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology*, 27, 3968–3975.
- Balasingham KD, Walter RP, Heath DD (2016) Residual eDNA detection sensitivity assessed by quantitative real time PCR in a river ecosystem. *Molecular Ecology Resources*, 17, 523–532.
- Cai FJ, Wu ZJ, He N, Ning L, Huang CM (2010) Research progress in invasion ecology of *Procambarus clarkia*. *Chinese Journal of Ecology*, 29, 124–132. (in Chinese with English abstract) [蔡凤金, 武正军, 何南, 宁蕾, 黄乘明 (2010) 克氏原螯虾的入侵生态学研究进展. *生态学杂志*, 29, 124–132.]
- Cai W, Ma Z, Yang C, Wang L, Wang W, Zhao G, Geng Y, Yu DW (2017) Using eDNA to detect the distribution and density of invasive crayfish in the Honghe-Hani Rice Terrace World Heritage Site. *PLoS ONE*, 12, e0177724.
- Cao L, Zhang E, Zang CX, Cao WX (2016) Evaluating the status of China's continental fish and analyzing their causes of endangerment through the red list assessment. *Biodiversity Science*, 24, 598–609. (in Chinese with English abstract) [曹亮, 张鄂, 臧春鑫, 曹文宣 (2016) 通过红色名录评估研究中国内陆鱼类受威胁现状及其成因. *生物多样性*, 24, 598–609.]
- Champlot S, Berthelot C, Pruvost M, Bennett EA, Grange T, Geigl EM (2010) An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS ONE*, 5, e13042.
- Chu KL, Ma XP, Zhang ZW, Wang PF, Lv LN, Zhao Q, Sun HY (2018) A checklist for the classification and distribution of China's freshwater crabs. *Biodiversity Science*, 26, 274–282. (in Chinese with English abstract) [楚克林, 马晓萍, 张泽伟, 王鹏飞, 吕琳娜, 赵强, 孙红英 (2018) 中国淡水蟹分类与分布名录(十足目: 拟地蟹科, 溪蟹科). *生物多样性*, 26, 274–282.]



- Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, Vere ND (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26, 5872–5895.
- Deiner K, Fronhofer EA, Mächler E, Walser JC, Altermatt F (2016) Environmental DNA reveals that rivers are conveyor belts of biodiversity information. *Nature Communications*, 7, 12544.
- Deiner K, Walser JC, Mächler E, Altermatt F (2015) Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, 183, 53–63.
- Deiner K, Altermatt F (2014) Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE*, 9, e88786.
- Dejean T, Valentini A, Miquel C, Taberlet P, Bellemain E, Miaud C (2012) Improved detection of an alien invasive species through environmental DNA barcoding: The example of the American bullfrog *Lithobates catesbeianus*. *Journal of Applied Ecology*, 49, 953–959.
- Ding H, Ma FZ, Wu J, Lei JC, Le ZF, Xu HG (2015) Considerations of building up a supervision and management system for prevention and control of invasive alien species posing environmental hazards in China. *Journal of Ecology and Rural Environment*, 31, 652–657. (in Chinese with English abstract) [丁晖, 马方舟, 吴军, 雷军成, 乐志芳, 徐海根 (2015) 关于构建我国外来入侵物种环境危害防控监督管理体系的思考. *生态与农村环境学报*, 31, 652–657.]
- Doi H, Katano I, Sakata Y, Souma R, Kosuge T, Nagano M, Ikeda K, Yano K, Tojo K (2017) Detection of an endangered aquatic heteropteran using environmental DNA in a wetland ecosystem. *Royal Society Open Science*, 4, 170568.
- Dougherty MM, Larson ER, Renshaw MA, Gantz CA, Egan SP, Erickson DM, Lodge DM (2016) Environmental DNA (eDNA) detects the invasive rusty crayfish *Orconectes rusticus* at low abundances. *Journal of Applied Ecology*, 53, 722–732.
- Duan XH, Wang ZY, Xu MZ (2010) Benthic Macroinvertebrate and Application in the Assessment of Stream Ecology. Tsinghua University Press, Beijing. (in Chinese) [段学花, 王兆印, 徐梦珍 (2010) 底栖动物与河流生态评价. 清华大学出版社, 北京.]
- Egan SP, Grey E, Olds B, Feder JL, Ruggiero ST, Tanner CE, Lodge DM (2015) Rapid molecular detection of invasive species in ballast and harbor water by integrating environmental DNA and light transmission spectroscopy. *Environmental Science & Technology*, 49, 4113–4121.
- Eichmiller JJ, Bajer PG, Sorensen PW (2014) The relationship between the distribution of common carp and their environmental DNA in a small lake. *PLoS ONE*, 9, e112611.
- Eichmiller JJ, Miller LM, Sorensen PW (2016) Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Molecular Ecology Resources*, 16, 56–68.
- Elbrecht V, Leese F (2017) Validation and development of *COI* metabarcoding primers for freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science*, 5, 11.
- Fernández S, Rodríguez S, Martínez JL, Borrell YJ, Ardura A, García-Vázquez E (2018) Evaluating freshwater macroinvertebrates from eDNA metabarcoding: A river Nalón case study. *PLoS ONE*, 13, e0201741.
- Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology Letters*, 4, 423–425.
- Geerts AN, Boets P, Heede SVD, Goethals P, Heyden CVD (2018) A search for standardized protocols to detect alien invasive crayfish based on environmental DNA (eDNA): A lab and field evaluation. *Ecological Indicators*, 84, 564–572.
- Goldberg CS, Pilliod DS, Arkle RS, Waits LP (2011) Molecular detection of vertebrates in stream water: A demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. *PLoS ONE*, 6, e22746.
- Goldberg CS, Sepulveda A, Ray A, Baumgardt J, Waits LP (2013) Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science*, 32, 792–800.
- Guo YH, Wang CM, Luo J, He HX (2009) *Physa acuta* found in Beijing, China. *Chinese Journal of Zoology*, 44(2), 127–128. (in Chinese with English abstract) [郭云海, 王承民, 罗静, 何宏轩 (2009) 北京发现尖膀胱螺. *动物学杂志*, 44(2), 127–128.]
- Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC (2011) High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical Chemistry*, 83, 8604–8610.
- Jerde CL, Mahon AR, Chadderton WL, Lodge DM (2011) “Sight-unseen” detection of rare aquatic species using environmental DNA. *Conservation Letters*, 4, 150–157.
- Klymus KE, Marshall NT, Stepien CA (2017) Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. *PLoS ONE*, 12, e0177643.
- Kristine B, Alice E, Gilbert MTP, Carvalho GR, Simon C, Michael K, Yu DW, Mark DB (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29, 358–367.
- Larson ER, Renshaw MA, Gantz CA, Umek J, Chandra S, Lodge DM, Egan SP (2017) Environmental DNA (eDNA) detects the invasive crayfishes *Orconectes rusticus* and *Pacifastacus leniusculus* in large lakes of North America. *Hydrobiologia*, 800, 1–13.
- Li B, Ma KP (2010) Biological invasions: Opportunities and challenges facing Chinese ecologists in the era of

- translational ecology. *Biodiversity Science*, 18, 529–532. (in Chinese) [李博, 马克平 (2010) 生物入侵: 中国学者面临的转化生态学机遇与挑战. *生物多样性*, 18, 529–532.]
- Lowe S, Brown M, Boujdelas S, De PM (2000) 100 of the World's Worst Invasive Alien Species. A selection from the Global Invasive Species Database. The IUCN Invasive Species Specialist Group, USA.
- Majaneva M, Diserud OH, Eagle SHC, Boström E, Hajibabaei M, Ekrem T (2018) Environmental DNA filtration techniques affect recovered biodiversity. *Scientific Reports*, 8, 4682.
- McKee AM, Spear SF, Pierson TW (2015) The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biological Conservation*, 183, 70–76.
- Merkes CM, Mccalla SG, Jensen NR, Gaikowski MP, Amberg JJ (2014) Persistence of DNA in carcasses, slime and avian feces may affect interpretation of environmental DNA data. *PLoS ONE*, 9, e113346.
- Moyer GR, Díaz-Ferguson E, Hill JE, Shea C (2014) Assessing environmental DNA detection in controlled lentic systems. *PLoS ONE*, 9, e103767.
- Piaggio AJ, Engeman RM, Hopken MW, Humphrey JS, Keacher KL, Bruce WE, Avery ML (2014) Detecting an elusive invasive species: A diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. *Molecular Ecology Resources*, 14, 374–380.
- Rees HC, Maddison BC, Middleditch DJ, Patmore James RM, Gough KC (2015) The detection of aquatic animal species using environmental DNA—A review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51, 1450–1459.
- Renshaw MA, Olds BP, Jerde CL, McVeigh MM, Lodge DM (2015) The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol chloroform isoamyl alcohol DNA extraction. *Molecular Ecology Resources*, 15, 168–176.
- Rosenberg DM, Resh VH (1993) Introduction to freshwater biomonitoring and benthic macroinvertebrates. In: *Freshwater Biomonitoring and Benthic Macroinvertebrates* (eds Rosenberg DM, Resh VH), pp. 1–9. Chapman & Hall, New York.
- Shaw JLA, Clarke LJ, Wedderburn SD, Barnes TC, Weyrich LS, Cooper A (2016) Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation*, 197, 131–138.
- Smith AJ, Bode RW, Kleppel GS (2007) A nutrient biotic index (NBI) for use with benthic macroinvertebrate communities. *Ecological Indicators*, 7, 371–386.
- Strickler KM, Fremier AK, Goldberg CS (2015) Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation*, 183, 85–92.
- Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Environmental DNA. *Molecular Ecology*, 21, 1789–1793.
- Takahara T, Yamanaka H, Suzuki AA, Honjo MN, Minamoto T, Yonekura R, Itayama T, Kohmatsu Y, Ito T, Kawabata ZI (2011) Stress response to daily temperature fluctuations in common carp, *Cyprinus carpio* L. *Hydrobiologia*, 675, 65–73.
- Takahara T, Minamoto T, Doi H (2013) Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS ONE*, 8, e56584.
- Takahara T, Minamoto T, Yamanaka H, Doi H, Kawabata ZI (2012) Estimation of fish biomass using environmental DNA. *PLoS ONE*, 7, e35868.
- Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MTP, Orlando L, Willerslev E (2012a) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21, 2565–2573.
- Thomsen PF, Kielgast J, Iversen LL, Møller PR, Rasmussen M, Willerslev E (2012b) Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, 7, e41732.
- Tréguier A, Paillisson J, Dejean T, Valentini A, Schlaepfer MA, Roussel J, Crispo E (2014) Environmental DNA surveillance for invertebrate species: Advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. *Journal of Applied Ecology*, 51, 871–879.
- Tsai YL, Olson BH (1992) Detection of low numbers of bacterial cells in soils and sediments by polymerase chain reaction. *Applied & Environmental Microbiology*, 58, 754–757.
- Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25, 929–942.
- Vamos E, Elbrecht V, Leese F (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. *Metabarcoding and Metagenomics*, 1, e14625.
- Vogelstein B, Kinzler KW (1999) Digital PCR. *Proceedings of the National Academy of Sciences, USA*, 96, 9236–9241.
- Wan FH (2009) *Research on Biological Invasions in China*. Science Press, Beijing. (in Chinese) [万方浩 (2009) *中国生物入侵研究*. 科学出版社, 北京.]
- Wan FH, Guo JY, Wang DH (2002) Alien invasive species in China: Their damages and management strategies. *Biodiversity Science*, 10, 119–125. (in Chinese with English abstract) [万方浩, 郭建英, 王德辉 (2002) 中国外来入侵生物的危害与管理对策. *生物多样性*, 10, 119–125.]
- Wang YJ, Li FY, Fan ZP, Cheng ZH, Zhang JQ (2012) The application of benthic macroinvertebrates in aquatic ecosystem health assessment. *Journal of Meteorology and*

- Environment, 28, 90–96. (in Chinese with English abstract) [王艳杰, 李法云, 范志平, 程志辉, 张建祺 (2012) 大型底栖动物在水生态系统健康评价中的应用. 气象与环境学报, 28, 90–96.]
- Whale AS, Huggett JF, Cowen S, Speirs V, Shaw J, Ellison S, Foy CA, Scott DJ (2012) Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *Nucleic Acids Research*, 40, e82.
- Wilcox TM, McKelvey KS, Young MK, Jane SF, Lowe WH, Whiteley AR, Schwartz MK (2013) Robust detection of rare species using environmental DNA: The importance of primer specificity. *PLoS ONE*, 8, e59520.
- Willerslev E, Cooper A (2005) Ancient DNA. *Proceedings of the Royal Society B: Biological Sciences*, 272, 3–16.
- Xu HG, Wu J, Liu Y, Ding H, Zhang M, Wu Y, Xi Q, Wang LL (2008) Biodiversity congruence and conservation strategies: A national test. *BioScience*, 58, 632–639.
- Xu HG (2013) Introduction to the Biological Species Resources Monitor. Science Press, Beijing. (in Chinese) [徐海根 (2013) 生物物种资源监测概论. 科学出版社, 北京.]
- Xu HG, Qiang S (2018) *China's Invasive Alien Species* (revised edition). Science Press, Beijing. (in Chinese) [徐海根, 强胜 (2018) 中国外来入侵生物(修订版). 科学出版社, 北京.]
- Yang HF, Zhou WC, Qian ZX, Wang P (2014) An important invasive snail, *Physa acuta* Draparnaud 1805, found firstly in Zhejiang Province. *Plant Quarantine*, 28(3), 63–65. (in Chinese with English abstract) [杨海芳, 周卫川, 钱周兴, 王沛 (2014) 浙江发现重要外来入侵物种——尖膀胱螺. 植物检疫, 28(3), 63–65.]
- Yeom DH, Adams SM (2007) Assessing effects of stress across levels of biological organization using an aquatic ecosystem health index. *Ecotoxicology & Environmental Safety*, 67, 286–295.

(特邀责任编辑: 周欣 责任编辑: 闫文杰)



•综述•

# 基于环境DNA-宏条形码技术的水生生态系统入侵生物的早期监测与预警

李晗溪<sup>1,2</sup> 黄雪娜<sup>1</sup> 李世国<sup>1,2</sup> 战爱斌<sup>1,2\*</sup>

1 (中国科学院生态环境研究中心环境生物技术重点实验室, 北京 100085)

2 (中国科学院大学资源与环境学院, 北京 100049)

**摘要:** 外来生物入侵是继生境破坏后造成生物多样性丧失的第二大威胁因素, 已对入侵地的生态安全、经济和社会发展及人类健康等造成严重负面影响, 成为21世纪五大全球性环境问题之一。作为水产养殖、航运和水生宠物交易大国, 我国水生生态系统的生物入侵问题尤为严重。研究表明, 系统地构建并应用早期监测预警技术是防控水生生态系统生物入侵最有效的途径。和陆生生物相比, 水生生物群落的物种繁多、群落结构复杂、生物形体微小且在入侵初期群体规模极小、隐匿于水下、可用于物种鉴定的外部形态缺乏, 使得在水生生态系统中构建并应用早期监测和预警体系在技术层面更具挑战。随着高通量测序技术的快速发展, 环境DNA-宏条形码技术成为构建水生生态系统入侵生物早期监测与预警技术的首选。本文主要综述了基于环境DNA-宏条形码技术的水生生态系统入侵生物的早期监测与预警技术方法; 解析了环境DNA-宏条形码监测系统的应用现状、技术优势; 着重探讨了影响监测结果准确性的I型和II型错误及其产生原因, 并为避免两类错误提供了可行的优化/改进方案; 最后对该方法在水生入侵生物监测中的应用前景进行了展望。

**关键词:** 生物入侵; DNA-宏条形码; 生物多样性; 水生生态系统; 早期监测与预警; I型错误; II型错误

## Environmental DNA (eDNA)-metabarcoding-based early monitoring and warning for invasive species in aquatic ecosystems

Hanxi Li<sup>1,2</sup>, Xuena Huang<sup>1</sup>, Shiguo Li<sup>1,2</sup>, Aibin Zhan<sup>1,2\*</sup>

1 Key Laboratory of Environmental Biotechnology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085

2 College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049

**Abstract:** Biological invasion is a major threat to multiple ecosystems across the globe, causing severe damages to ecological integrity, loss of biodiversity, economic and social development and even human health. With the rapid development in aquaculture, shipping and aquarium and ornamental trades in the past several decades, China has become one of the countries most influenced by invasive species. Studies have clearly shown that the development and application of robust early monitoring and warning is one of the most effective ways to prevent and possibly control invasive species in aquatic ecosystems. Compared to terrestrial ecosystems, there remain several technical difficulties for developing early monitoring and warning in aquatic habitats. The technical challenges are mainly due to several features of aquatic biological communities such as high biodiversity and complex structure, a large number of microscopic species, extremely low population density and lack of available taxonomic keys for species identification. With the rapid development of high-throughput sequencing techniques, environmental DNA (eDNA)-metabarcoding has become the top priority method for developing the early monitoring and warning programs in aquatic ecosystems. In this review, we aim to synthesize the research progress on eDNA-metabarcoding and its application to early monitoring and warning of invasive species in aquatic ecosystems. In addition, we briefly discuss the technological advantages of eDNA-metabarcoding for the early monitoring and warning

收稿日期: 2018-09-04; 接受日期: 2019-05-07

基金项目: 国家自然科学基金优秀青年科学基金(31622011)

\* 通讯作者 Author for correspondence. E-mail: azhan@cees.ac.cn



programs. Finally, we propose research perspectives for solutions to technical issues for false positive and false negative errors in the eDNA-metabarcoding process.

**Key words:** biological invasion; DNA metabarcoding; biodiversity; aquatic ecosystems; early monitoring and warning; type I error; type II error

## 1 引言

近一个世纪以来,生物入侵、环境污染、气候变化等生态环境问题给全球生态系统带来巨大威胁。其中,生物入侵已导致严重生态灾害,成为五大全球性环境问题之一(鞠瑞亭等,2012)。随着全球化进程的加快,更加频繁的航运、水产品和宠物交易以及水产养殖引种等人类活动把长期地理隔离的物种带到同一水域,导致外来水生生物在新的生态系统中快速传播,引发生物入侵(Lockwood et al, 2013)。水生生物入侵由于隐蔽性强、所造成的灾害易扩散、治理手段缺乏等原因,对入侵地生态系统具有很强的破坏性,正严重威胁着水生生态系统的健康和稳定(Dudgeon et al, 2006; Hambler et al, 2011; Holland et al, 2012)。

我国国土面积广阔、海岸线长、港口众多,沿海地区人口密度大、人类活动频繁,加之长期实行的对外开放战略(建立自由贸易区和“一带一路”政策等),为水生入侵生物的引入、定殖、扩散、灾害暴发等提供了便利条件,成为受生物入侵影响最为严重的国家之一(Chen et al, 2017)。我国水生生态系统外来物种总数量已达553种,包括429个淡水种和124个海水种,其中66个入侵物种(淡水物种23个,海水物种43个)已经引发了巨大的生态灾害(Chen et al, 2017)。入侵生物引发灾害的频率、程度和受灾面积在短时间内剧增,严重破坏了入侵地生态环境,改变了当地生物群落结构,降低了入侵水域的生物多样性。其中,最为典型的灾害事件即是由水生入侵藻类引发的赤潮。赤潮藻类生存能力极强,是我国水生入侵物种中的第二大门类(目前共计发现15种入侵赤潮藻类),这些藻类入侵后迅速成为区域优势种,导致赤潮灾害频发(Chen et al, 2017)。据不完全统计,我国每年由外来入侵生物造成的直接经济损失约2,000亿元人民币(万方浩等,2009),而其中由赤潮造成的经济损失高达5亿元人民币(王朝晖等,2010)。

入侵生物一旦在水生生态系统中入侵成功,将

非常难以清除,入侵生物的预防和治理已经成为相关领域科学家们关注的焦点。传统的治理手段包括物理法、化学法和生物法等(万方浩等,2009; Lin et al, 2015; Chen et al, 2017)。然而,这些方法在使用过程中都存在一定的局限性:物理清除方法虽然简单方便,但无法彻底根除入侵生物;化学法见效较快,但会使入侵生物产生耐药性,并且容易造成水体环境的二次污染;生物法作用温和,但会一定程度上增加新物种入侵的风险(刘芳明等,2007)。通过长期的实践探索发现,以防为主、防治结合是最为有效的方式(Lin et al, 2015; Chen et al, 2017)。通过完善入侵生物档案库并构建早期监测预警技术体系和平台,可对生物入侵的各阶段进行综合监控和管理,并将危害损失控制在最低水平,成为防控生物入侵的最直接有效的方式。与陆生生物相比,水生生物的种类繁多、群落结构复杂,很多生物类群具有形体微小、在入侵初期群体规模极小、隐匿于水下、可用于物种鉴定的外部形态缺乏等特点,导致在水生生态系统中构建早期检/监测和预警体系在技术层面更具挑战(Xiong et al, 2016)。

由于入侵初期外来生物群体规模极小,因此从复杂的群落中检测出这些低丰度生物便成为构建早期检/监测和预警体系的关键(Xiong et al, 2016)。水生群落中繁多的低丰度物种(生物量小于万分之一)是水生生态系统的重要组成部分。Sogin等(2006)提出“稀有生物区系(rare biosphere)”的概念,即群落中所有稀有生物的集合。稀有生物区系主要包含两大类:低丰度土著种(一部分为濒危种)和入侵早期的外来物种。低丰度土著种通常对环境胁迫更敏感,在物质循环和能量传递中发挥着重要作用(杨江华,2017; 陈义永等,2018);外来入侵种在引入初期虽然种群密度极低,但随之而来的快速繁殖可使生态系统退化并加速本地种的灭绝进程(Pysek & Richardson, 2010)。

水生生物鉴定的传统手段依赖于显微镜或流式细胞技术(Shi et al, 2011; Weber & Pawlowski, 2013),该方法工作量大,无法满足快速大量监测的

需求,且灵敏度低,难以监测和识别数量庞大的稀有生物区系(Galand et al, 2009; Cheung et al, 2010)。此外,采样方式和物种自身形态特征也对水生生物物种鉴定工作产生了较大影响。水体的流动和季节性水文变化使取样的随机性大大增加(McDonald, 2004; Jerde et al, 2011)。样品保存不当会导致形态特征缺失,给准确的分类鉴定造成困难(Briski et al, 2011; Darling & Mahon, 2011)。许多类群有变态发育的特点,不同生活史阶段的形态结构和习性会发生很大变化;部分类群在同一生活史时期的形态学特点极为相似,也为物种鉴定带来了极大困难(Briski et al, 2011; Darling & Mahon, 2011)。上述这些因素直接影响了基于形态的水生入侵物种鉴定的准确度和效率,因此亟需建立集快速精确、高灵敏度、高分辨率于一体的检测方法。环境DNA-宏条形码技术检测灵敏度高、耗时短、成本低、通量高,成为鉴定外来水生入侵物种的强有力工具(Zhan & Maclsaac, 2015; Xiong et al, 2016),在复杂群落稀有种的监测方面优势明显,具有极大的应用前景(Zhan & Maclsaac, 2015; Xiong et al, 2016)。

本文系统总结了环境DNA-宏条形码技术的发展历程、基本特征及其在水生生物入侵早期监测与预警过程中的应用,探讨了该技术应用局限性并提出了可能的解决方案,同时还对该技术在水生入侵物种监测和预警中的应用前景进行了展望,期望为水生生物入侵的科学防控提供技术支撑。

## 2 环境DNA-宏条形码技术

### 2.1 环境DNA-宏条形码技术简介

环境DNA (environmental DNA, eDNA)是指直接从环境样品(例如土壤、空气、水体)中提取到的DNA,是不同物种遗留在环境中的DNA的总和(Levy-Booth et al, 2007; Pietramellara et al, 2009)。水体中的环境DNA包含生物体经由皮肤、尿液、粪便、粘液等释放到环境中的DNA和细胞死亡裂解后释放到环境中的DNA (Taberlet et al, 2012; Rees et al, 2014)。

环境DNA-宏条形码技术主要操作流程为:采集环境样本,对样本中的环境DNA进行提取、基于分子标记基因设计引物对环境DNA进行PCR扩增及高通量测序,将得到的大量序列直接与数据库中的序列进行比对,或者先将序列聚类为可操作分类

单元(operational taxonomic units, OTUs)后再进行比对,达到同时对环境样本中多个物种(或高级分类单元)进行鉴定的目的。

### 2.2 环境DNA-宏条形码技术的关键环节

入侵物种监测包括对靶标和非靶标生物的监测。靶标物种的监测相对简单,通过设计物种特异性引物监测靶标物种是否存在,技术要点为保证引物的特异性、避免近缘或共存物种的错配造成非特异性扩增(郝雅宾等, 2018)。而对非靶标生物的监测对技术要求相对较高,在选择高分辨率的分子标记基因的基础上,设计高效率的通用引物,通过分析群落组成和结构的动态变化从而实现早期监测与预警。检测稀有生物区系的组成是监测非靶标入侵物种的技术关键。基于环境DNA-宏条形码的监测体系目前的研究主要集中在两个方面:(1)分子标记基因的选择和通用引物的设计与评估;(2)高通量测序数据的处理及对关键参数的优化与评估。

#### 2.2.1 高分辨率分子标记基因选择

选择适当的分子标记基因并设计高分辨率的通用引物,是提高检测灵敏度和稀有物种检出率的关键。分子标记基因具有以下两方面特征:有相对保守的区域用来设计通用性强的引物,便于PCR扩增;扩增区兼具种间差异大、种内差异小的特点,能够有效区分不同物种(Zhan & Maclsaac, 2015; 吴彪等, 2018)。常见的真核生物分子标记基因包括线粒体细胞色素C氧化酶亚基I (COI)基因、内转录间隔区(ITS)、核糖体18S rRNA基因、28S rRNA基因和线粒体16S rRNA基因等。

不同分子标记基因适用范围不同,需根据待检测生物类群以及研究目的综合选择。COI基因具有进化速度快和功能保守性强的特点,常用来分析浮游动物、甲壳动物、微型后生动物等(Bourlat et al, 2013; Leray et al, 2013; Zhou et al, 2013; Avó et al, 2017)。但COI基因较快的进化速度给设计群落水平的通用引物带来很大困难(Zhan et al, 2014a)。海洋生态系统中的海绵动物、部分刺胞动物和水母类的线粒体DNA的进化速率比其他后生动物慢10-20倍,进化速度不一致引起的DNA序列差异,使得难以针对COI基因设计出能够将混合DNA样品中全部物种进行有效扩增的通用引物,因而对海洋生态系统部分类群进行研究时不宜选用(Tang et al, 2012; Zhan et al, 2014a)。18S rRNA基因存在保守区和可变区且

序列进化速度较慢,适用于研究亲缘关系较远的生物(Brannock et al, 2016)。16S rRNA基因对刺胞动物具有较高的识别率(Bucklin et al, 2010)。Zhan等(2014a)比较了COI、16S rRNA和18S rRNA等分子标记基因在浮游动物环境DNA-宏条形码研究中的差别,发现其所用的COI引物都无法提供高质量的PCR产物,并建议使用18S rRNA基因和线粒体16S rRNA基因进行条形码研究以获得更多的物种信息。

### 2.2.2 高灵敏度通用引物设计

通用引物设计是环境DNA-宏条形码技术应用的基础,也是对分析结果具有决定性影响的环节(Zhan & Maclsaac, 2015)。扩增引物的设计一般应遵从通用性强、灵敏度高等特点。通用性强意味着可以将复杂群落中的环境DNA进行有效扩增,实现对绝大多数物种进行检测,进而提高物种覆盖度(Zhan & Maclsaac, 2015);高灵敏度通用引物可降低监测成本,仅需进行一次PCR扩增即可完成对所有类群DNA片段的收集,为检测复杂群落中低丰度物种提供了便利(Zhan et al, 2014a; Zhan & Maclsaac, 2015)。

高灵敏度引物的获得离不开对引物进行全面的评价和测试(Leray, 2013; Zhan et al, 2013)。引物测试需要基于从单一生物类群、复杂生物群落到稀有生物区系层层推进(Zhan et al, 2014a)。首先是对单一类群水平引物扩增稳定性和覆盖度的检测,进而推广到复杂群落水平,对引物进行覆盖度和PCR偏向性的检测。Leray等(2013)在研究天竺鲷科 *Nectamia savayensis*、金鳞鱼科 *Myripristis berndti* 和 *Sargocentron microstoma* 的胃内容物时,对COI基因设计了5组通用引物,结果表明其中一对引物(mlCOIintF/jgHCO2198)比传统通用引物(LCO1490/HCO2198)在后生动物中扩增效率更高,能够获得更多的物种信息。最后是针对稀有生物区系的检测。目前较为常用的方法是内标法,即通过在提取DNA之前加入群落中不存在的物种作为内标,然后利用内标的绝对和相对丰度计算其他物种的绝对丰度。Zhan等(2014a)分析了淡水生态系统和海洋生态系统的典型群落,成功设计出基于核糖体小亚基rDNA (SSU)的高分辨率通用引物,通过内标法对基于通用引物Uni18S/Uni18SR新构建的宏条形码体系进行测试,发现该体系灵敏度极高,比已有方法提高了5个数量级。对复杂水生群落的PCR扩增结果显示,此通用引物可以扩增出几乎所有水生真核

生物类群(动物、原生生物、藻类、真菌等),且扩增偏向性极小;此通用引物将对低丰度入侵生物的检出率提高了5倍,即仅存在一个幼虫的条件下,检出率达到100% (Zhan et al, 2014a)。

### 2.2.3 重现性和检出率

重现性(reproducibility)指同一序列在各平行样本间同时出现的概率,是衡量数据准确度和可靠性的重要指标(Prosser, 2010; Zhou et al, 2011)。Zhan等(2014c)对此问题进行了一系列研究,通过对海洋和淡水两个生态系统浮游生物群落检测的结果进行比较分析发现,只有39.8% (海洋)和36.4% (淡水)的OTU在几次重复中被同时检测到,但与土壤微生物等其他复杂群落相比检出率已属较高(Zhan et al, 2014c)。研究同时发现,低丰度OTU的重现性远低于高丰度OTU,对于包含100条序列的OTU重现性可达到100%,而单体(singletons, OTU中仅包含一个序列)的重现性则低于25% (Zhan et al, 2014c)。研究同时指出,实验中各阶段的随机过程是重现性低的重要原因,主要包括随机取样、DNA提取及PCR扩增的偏向性,以及高通量测序深度不足和数据滤错(质量控制)环节过滤大量原始序列等(Zhan et al, 2014b, c; Zhan & Maclsaac, 2015)。

提高入侵种检出率是早期监测关注的核心问题。虽然针对高分辨率分子标记基因设计高灵敏度通用引物在一定程度上提高了检出率,但尚有一些错综复杂的因素同时制约着检出率。目前已有通过内标和外标相结合的方法,研究高通量测序及数据滤错环节对入侵种检出率的影响。结果表明,数据滤错会影响稀有种检出率,滤错严格度与检出率呈负相关关系,造成此种结果的主要原因是低丰度生物对应的序列中存在测序质量不高等问题,滤错过程将这些低丰度序列删除(Zhan et al, 2014b)。借助稀疏曲线(species rarefaction curve)可以更好地确定测序深度,在一定范围内测序深度的增加会显著增加OTU检出数目,特别是提高低丰度OTU的检出率。对序列丰度与检出率关系的研究表明,虽然序列丰度与检出率总体呈正相关关系,但对不同群落加入相同指示生物,测序数据也并不能可靠地量化物种相对丰度(Sun et al, 2015)。近年来,利用高通量测序技术结合无PCR扩增(PCR-free genome-skimming)的“超级条形码”技术,可以解决由于PCR扩增偏向性造成的物种偏向性以及混合物种相对丰度



难以分析等问题, 获得较准确的相对丰度信息, 有望以更高的灵敏度和准确性对混合物种样品进行定性及相对定量的监测(郎丹丹等, 2018)。增加生物学重复可以有效降低随机取样导致的误差, 同时, 结合实验重复(DNA提取、PCR扩增等)并增加测序深度, 谨慎处理原始数据和OTU聚类结果, 对于检出率的提高有极大的帮助(Zhan & Maclsaac, 2015)。

### 3 环境DNA-宏条形码技术在水生入侵物种早期监测和预警中的应用

环境DNA-宏条形码技术以其高效、低成本的特点极大拓展了条形码技术在生物多样性分析、大范围物种检测以及入侵种监测等方面的应用(表1)。水生生态系统生物入侵的研究已从单一靶标生物的检出朝复杂非靶标物种、大尺度的方向发展, 在对靶标入侵种和非靶标入侵种的监测上都取得了良好的效果。

#### 3.1 靶标入侵种的监测

入侵种检测的主要环节是针对靶标生物设计特异性引物, 通过检验环境DNA中是否含有目标种释放的DNA来判断是否存在靶标物种的入侵。Ficetola等(2008)基于美国牛蛙(*Rana catesbeiana*) *Cyt-b*基因设计特异性引物, 对法国18个自然水体的环境DNA样本进行扩增, 并在传统方法未检出的池塘中成功检测到该物种, 首次将环境DNA-宏条形码技术引入入侵物种监测领域。环境DNA与遥感、光谱等技术的联合使用为外来入侵物种的监测预警提供了更广阔的发展空间。Egan等(2015)结合现场准备光透射光谱(LTS)监测了压舱水中两种入侵贝类: 斑驴贻贝(*Dreissena bugensis*)和斑马贻贝(*D. polymorpha*)的入侵情况。目前, 这项技术被广泛应用于鲤鱼(*Cyprinus carpio*; Takahara et al, 2012)、新西兰泥螺(*Potamopyrgus antipodarum*; Goldberg et al, 2013)、克氏原螯虾(*Procambarus clarkii*; Tréguier et al, 2014)、鲢鱼(*Hypophthalmichthys molitrix*; Song et al, 2017)等入侵生物的监测, 监测物种类型涉及鱼、两栖、爬行、昆虫等多个门类。与国外相比, 我国环境DNA-宏条形码技术起步较晚, 目前仅在克氏原螯虾<sup>①</sup>、鱼类(姜维等,

2016; 徐念和常剑波, 2016; 郝雅宾等, 2018)等几个入侵物种的监测上有报道。随着环境DNA-宏条形码技术的不断发展, 其应用领域会逐步扩展到各种水生生态系统的多个物种(吴昀晟等, 2019)。

#### 3.2 非靶标入侵种的监测

与靶标入侵种不同, 对非靶标入侵生物的监测需要设计高灵敏度通用引物, 通过比较群落物种组成变化达到监测预警的目的(单秀娟等, 2016)。Saunders (2005)通过对加拿大海区白蒙藻科的COI条形码设计通用引物, 发现了3个本来不存在于该海区的物种, 进而确定为入侵物种。Zaiko等(2015a)对航行不同时间的船舶压舱水进行采集, 然后利用环境DNA-宏条形码技术分析其中的物种数目, 探讨是否存在入侵种的可能。Thomsen等(2016)使用6对引物成功检测到15种入侵鱼类, 同时检测到4种入侵鸟类。Emily等(2016)对加拿大海岸线和五大湖主要港口的浮游生物样本进行分析后鉴定到379个浮游动物物种, 有24个物种被确定为外来入侵种, 其中11种为首次在该地区检测到。非靶标生物的监测不仅能够定性监测到入侵生物的存在, 还可以定量评估生物入侵程度(李飞龙等, 2018), 有利于监管部门针对入侵程度制定相应的治理措施, 成为早期预警系统的主要监测方式。

### 4 环境DNA-宏条形码技术存在的问题及优化方案

尽管环境DNA-宏条形码技术在水生入侵物种监测方面具有显著的优势, 但该技术的应用层面目前还存在一系列问题(图1)。我们对影响该技术监测效果的原因进行了讨论, 并为以后的研究提出可能的解决方案。

#### 4.1 环境DNA-宏条形码技术的准确性

环境DNA监测中缺乏灵敏度高的通用引物、监测准确度有待标准化方法的支撑、低丰度生物群落重现性较差、环境DNA产生和降解的动力学过程还不甚明确(Zhan & Maclsaac, 2015; 马鸿娟等, 2016)等问题会使监测中产生假阴性和假阳性两种结果, 从而影响监测结果的准确性。

##### 4.1.1 环境DNA-宏条形码技术的假阴性

在实际监测过程中, 环境DNA-宏条形码技术存在I型和II型错误。I型错误是指拒绝实际上成立的

① 马竹欣 (2016) 利用环境DNA技术调查入侵种克氏原螯虾在元阳梯田的分布. 硕士学位论文, 云南大学, 昆明。



**表1 环境DNA-宏条形码技术在水生生态系统中的应用实例**  
**Table 1 Application examples of environmental DNA-metabarcoding technique in aquatic ecosystems**

监测类群 Taxon	监测地点 Location	遗传标记 Marker	引物序列 Primer sequences	实验目的 Main aims	参考文献 Reference
<b>淡水生态系统 Freshwater ecosystem</b>					
浮游动物 Zooplankton	加拿大安大略湖汉密尔顿港 Hamilton Harbour, Ontario, Canada	18S rRNA V4	Um18S: AGGGCAAKYCTGGTGCAGC; Um18SR: GRCGGTATCTRATCGYCTT	常见遗传标记PCR效率和入侵种监测能力比较 Comparison of the PCR efficiency of common genetic markers and their surveillance ability of invasive species	Zhan et al, 2014a
大型底栖无脊椎动物 Benthic macroinvertebrate	多伦多环境保护署 Toronto Region Conservation Authority (TRCA)	<i>COI</i>	LC01490: GGTCAACAAATCATAAAGATATTGG HC02198: TAAACTTCAGGGTGACCAAAAAATCA	评估淡水大型底栖无脊椎动物群落多样性及其入侵可能 Assessing biodiversity of freshwater benthic macroinvertebrate community and their invasive possibility	Hajibabaei et al, 2011
细菌/真菌 Bacteria/fungi	澳大利亚国王溪流域 Kings Creek Stream Network, Australia	16S rRNA V4 ITS2	515F: GTGNACGCMGCCGGGTAA 806R: GGACTACNVGGGTWCTAA ITS1F: CTTGGTCAITTAGACGAAGTAA ITS4: TCCTCCGCTTAT TGATATGC F1: ACTGGGATTAGATACCCC R1: TAGAACAGGCTCCTCTAG	木本植物入侵对水域细菌和真菌群落的影响 Effects of woody plant encroachment to bacterial and fungal communities	Veach et al, 2015
鱼类/两栖类 Fish/amphibians	丹麦沿海水域 Danish coastal waters	12S	F1: ACTGGGATTAGATACCCC R1: TAGAACAGGCTCCTCTAG	eDNA和高通量技术在入侵生物监测中的应用 Application of eDNA and high throughput technique in invasive biological monitoring	Valentini et al, 2016
节肢动物 Arthropod	德国Deilbach河 Deilbach River, Germany	16S	Ins16S_1short: TRRGACGAGAAAGACCCTATA ACGCTGTTATCCCTAAGGTA	引物设计及物种监测能力测试 Primers design and species monitoring capability test	Elbrecht et al, 2016
节肢动物 Arthropod	中国亚热带山区 A mountainous habitat in the subtropical region of China	<i>COI</i>	LepF1: ATTCAACCAATCATAAAGATAT LepR1: TAAACTTCTGGATGTCACAAAA	鸟枪法测序在生物多样性恢复中的准确性研究 Research on the accuracy of biodiversity recovery with Shotgun sequencing	Zhou et al, 2013
<b>海洋生态系统 Marine ecosystem</b>					
真核生物 Eukaryotes	波罗的海沿岸 Baltic Sea	<i>COI</i>	1380F: CCCTGCCHTTTGTACACAC 1510R: CCTTCYGCAGGTTACCTAC	压舱水入侵生物监测管理 Ballast water species surveillance and management	Zaiko et al, 2015b
细菌及浮游植物 Bacteria and phytoplankton	韩国, 日本海域 Japan Sea, Korea	23S cpDNA	A23SRV1: GGACARAAAGACCCTATG A23SRV1: AGATCAGCCTGTTATCC A23SRV2: CARAAAGACCCTATGMAGCT A23SRV2: TCAGCCTGTTATCCCTAG	引物评估及浮游植物入侵监测 Marker assessment and monitoring of the invasive species in 2016 phytoplankton community	Yoon et al, 2016
真核原生动物 Eukaryotic protists	美国切萨皮克湾 The Chesapeake Bay, USA	18S	3NDF: GGCAGTCTGGTGCCAG V4_euk_R2: ACGGTATCT(AG)ATC(AG)TCTTCCG 1380F: CCCTGCCHTTTGTACACAC 1510R: CCTTCYGCAGGTTACCTAC	压舱水寄生原生动物监测 Detecting parasitic protists in ballast water	Pagenkopp et al, 2016
鱼类 Fish	蒙特里湾水族馆 Monterey Bay Aquarium	12S	F1: ACTGGGATTAGATACCCC R1: TAGAACAGGCTCCTCTAG	环境监测中标记基因引物评估 Marker sequencing in environmental monitoring	Kelly et al, 2014
鱼类 Fish	格陵兰岛戴维斯海峡 Davis Strait, SW Greenland	12S	teleo_F: 50-ACACCCGCCCTCACTCT teleo_R: 50-CTTCCGGTACACTTACCATG	深海研究中高通量测序方法与传统方法的比较 Next-generation sequencing vs traditional surveys in deep ocean	Thomsen et al, 2016
<b>湿地生态系统 Wetlands ecosystem</b>					
线虫类 Nematoda	葡萄牙西南海岸Mira河口 Mira estuary, southwestern coast of Portugal	18S, <i>COI</i>	Seq18S_1: CATCCAAGGAAGGCAGCAGG Seq18S_2: GAGCTGGAATTTACCCGGG	河口底栖生物多样性监测 Estuary benthos biodiversity assessment	Avó et al, 2017
原生动物 Protozoa	挪威Finsevatn湖 Lake Finsevatn, Norway	18S rRNA V4	V4_euk_R1: GACTACGACGGTATCT(AG)ATC(AG)TCTTCCG V4_euk_R2: ACGGTATCT(AG)ATC(AG)TCTTCCG	原生群落多样性及地理分布监测 Investigate the diversity and distributional patterns of the protist	Bråte et al, 2010

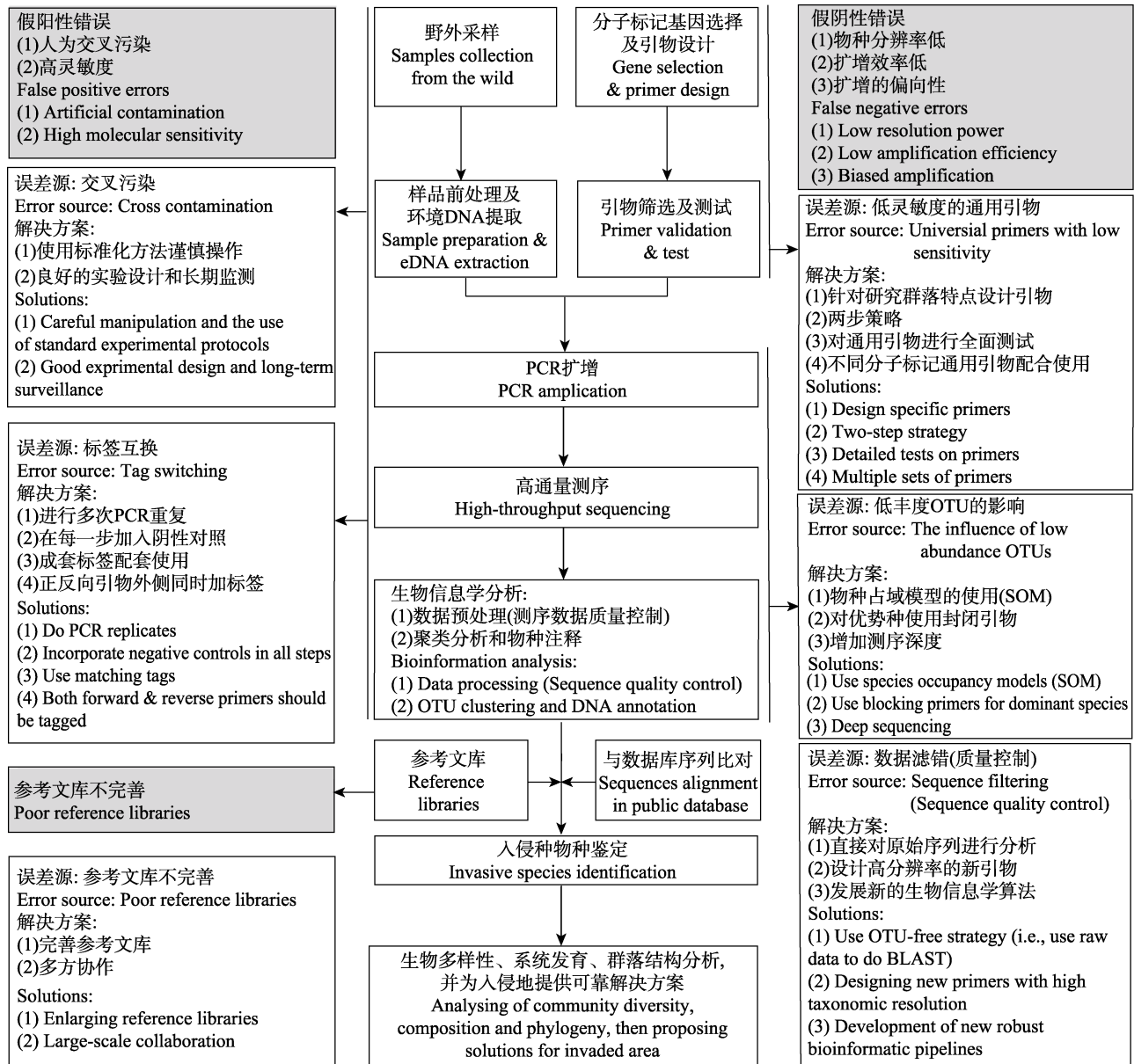


图1 基于环境DNA-宏条形码技术的水生入侵生物早期监测与预警技术现存的主要问题及可能的解决方案

Fig. 1 A summary of error sources and possible solutions for both false positive and false negative errors when using environmental DNA-metabarcoding methods for early detection and warning of aquatic invasive species

正确假设, 为“弃真”的错误。环境DNA监测技术中的I型错误是指未能监测到环境中真实存在的物种, 降低检出率, 带来假阴性结果。I型错误为部分入侵种的定殖爆发成灾提供了时间, 对外来入侵种早期监测造成了严重的影响。分子标记片段分辨率低、通用引物扩增效率低、数据滤错时严格度设置过高以及PCR的偏向性是导致I型错误的直接原因。

序列分辨率的高低在很大程度上是由分子标记片段的选

择造成的。低分辨率使得来源于不同物种的、具有高相似度的序列被分到同一个OTU, 可能使入侵物种被划为土著种。环境DNA-宏条形码鉴定物种的关键是寻找理想的分子标记基因。不同的基因片段在不同分类阶元系统进化研究中各有优缺点。已有研究表明, *COI*基因在鱼类不同属间的分辨率较低(Keskin & Atar, 2013; Deagle et al, 2014)。Zhan等(2014a)基于SSU基因研究了SSU V4-V9区域对入侵生物海鞘类群的分辨能力, 结果表明其分辨率在监测的类群间均可达到显著性水

平。例如,在玻璃海鞘属(*Ciona*)和二段海鞘属(*Didemnum*)的物种中种间差异大于5%,但不同类群间的分辨率存在一定差异,在拟菊海鞘属(*Botrylloides*)和菊海鞘属(*Botryllus*)等许多属中种间差异不到1% (Zhan et al, 2014a)。由于不同类群进化速度存在差异,而目前大部分水生类群缺乏亲缘关系程度的研究,故难以界定各物种间差异在何种水平会导致分子标记无法区分的情况,这就要求研究者在选择合适分子标记基因时,对水生生物各类群的进化速度和差异程度有深入的了解,并在通用性和分辨率两方面兼顾考虑。

当获取的环境DNA的量低于PCR能够检测到的阈值时,目标物种将难以被检出。首先,环境DNA在充满水和核酸酶的环境中极易随机降解成小片段,目标片段的缺失会造成扩增困难。其次,各类群在基因不同区域的多态性水平具有很大差异,可能导致各类群的PCR过程中在引物结合位点发生不同程度的错配,造成实际可以扩增的环境DNA总量低于检测阈值(Xiong et al, 2016)。高灵敏度的通用引物的使用可监测到生物总量占比仅 $2.3 \times 10^{-5}\%$ 的靶标物种,尽管如此,当入侵种生物量低于这个阈值范围时仍会出现假阴性结果(Zhan et al, 2014c)。

PCR的偏向性会导致低丰度和难扩增物种的数据丢失,这是导致条形码解析度和普适性较低的重要原因(陈炼等, 2016)。PCR偏向性是由不同模板的扩增效率的固有差异(Polz & Cavanaugh, 1998),或者是后期优势模板自我退火扩增的抑制作用(Suzuki & Giovannoni, 1996)造成的。同一体系中,模板DNA间扩增效率的固有差异,如GC含量高、富含二级结构使得不同物种对引物的亲和能力不同等均能导致PCR扩增效率的差异(Bellemain et al, 2010; Engelbrektson et al, 2010)。每轮PCR随机扩增都会带来偏向性,使我们无法准确判断物种的原始相对丰度,甚至发生漏检,成为入侵种检出的障碍。

PCR扩增的偏向性可通过不断优化扩增条件(包括降低引物量、适当增加模板量、减少循环次数、适当提高退火温度)来加以改善。此外,在明确自然条件下群落组成结构的情况下针对优势种设计封闭引物,也可减少不必要的DNA扩增(Boessenkool et al, 2012)。同时,可通过调整PCR程序,例如采用巢式PCR方法提高扩增特异性(Davey et al, 2014);

或者使用两步策略:在PCR第一轮扩增时,基于18S rRNA等分子标记基因设计使用较为通用的引物,对群落中的大多数类群进行扩增,在第二轮专门为稀有类群设计特异性强的引物实现对低丰度类群的检测,均可达到特异性富集的目的。

随着环境DNA-宏条形码技术的发展,基于不同分子标记基因保守区设计的多对通用引物配套使用的多条形码技术可对群落进行更全面的监测<sup>①</sup>。但多基因引物扩增得到的种类较为复杂,给序列对比和多套数据整合等带来困难,提高了假阳性率和监测成本。新一代的线粒体宏基因组技术,即PCR-free技术(如DNA捕获法、鸟枪法测序等)摒弃了PCR扩增,避免了PCR引入的一系列错误,使结果更加精确。除了评估物种组成外,通过推断每个物种中的线粒体DNA总量来评估各物种生物量乃至相对丰度,是一种全新的方法。虽然PCR-free对生物信息算法提出了更高要求,但该方法无疑为未来环境DNA-宏条形码技术提供了新的发展方向(Zhou et al, 2013)。

测序后一系列生物信息学分析方法的使用对检出率也存在极大影响。对原始数据进行数据滤错(质量控制)的过程中,低质量序列的去除在一定程度上可避免“人为物种”(artifacts, PCR或测序错误导致的假序列)的产生,但该过程又极易滤除环境中真实存在的稀有入侵种。真实条件下的稀有种和实验操作过程中产生的“人为种”的判断一直是困扰研究人员的问题。OTU聚类分析虽然能实现复杂水生生物群落的层级划分问题,但会产生低丰度序列检出生物量低以及稀有生物区系检出困难的问题。在具体操作中,对同一样本采用不同的聚类分析方法,甚至同一聚类方法采用不同分类水平检出的入侵生物数目都有很大差异。有研究表明数量庞大的低丰度序列多为交叉污染造成,对物种检出率影响很小(Kunin et al, 2010; Tedersoo et al, 2010)。但更多研究表明,序列丰富度低的OTU如单体(singletons)、二体(doubletons)和三体(tripletons)是群落中真实存在的稀有种(Kauserud et al, 2012; Zhan et al, 2013)。这些问题的存在都对生物信息方法的选择与后期的数据处理分析提出了更高要求。

① 张宛宛 (2017) 基于DNA宏条形码技术的浮游植物群落多样性监测研究. 硕士学位论文, 南京大学, 南京.



可以明确的是,在入侵早期,入侵物种处于低丰度水平,因此低丰度物种的完整监测是不可忽略的。在序列比对环节,将研究聚焦于比对平行样本间及各技术重复处理重现性高的序列和OTU可以提高低丰度物种的检出率。对于序列比对,最为理想的方法是使用OTU-free的方法直接比对原始序列,该方法在提出初期由于没有良好的算法导致工作量大、耗时长。随着环境DNA-宏条形码技术的不断发展,层出不穷的新算法已经可以满足对原始数据的直接分析,极大地降低了测序和算法导致的数据流失。一个有效的控制方法是在提高生物学重复的基础上使用物种占域模型(Species Occupancy Model, SOM; MacKenzie et al, 2002)评估监测结果。物种占域模型将空间占有率(occupancy rate)作为评估动物种群变化的一个度量标准(李勤等, 2013)。对入侵物种监测来说, SOM模型能估算出特定样点或区域被目标物种占有的概率,预测目标地点、目标物种在未来的分布变化、种群动态趋势以及干扰程度对物种生存的影响(万雅琼等, 2017)。增加生物学重复能够进一步提高低丰度物种的检出率, Ficetola等(2015)建议根据样本特征确定重复水平,对稀有分类群进行监测时采用8个以上的生物学重复,重复次数的增多为SOM模型的准确性提供了保障。理论生态学模型仅需少量调查便能够通过重复测量数据更好地对实验中可能存在的问题进行补充指导,对入侵物种的长期监测提供了很大的帮助。

#### 4.1.2 环境DNA-宏条形码技术的假阳性

在统计学上II型错误是指错误地接受无效假设,也就是犯了“存伪”的错误。在入侵种监测中, II型错误是指由于灵敏度高的高通量测序检测出不属于环境中的新序列并归为新群落,凭空创造出“人为入侵种”带来假阳性结果,造成人力物力的浪费。

交叉污染和分子标签互换是该现象出现的主要原因。有研究表明,在监测的202个宏基因组中145个可能已发生污染,占到总数的71.8%,而每个宏基因组中污染序列所占比例更是高达64% (Schmieder & Edwards, 2011)。造成交叉污染的原因很多,主要包括采样过程中采样器的污染以及各实验环节中样品间的相互污染等。PCR扩增时不可避免会形成异源双链(Kanagawa, 2003)、嵌合体(Odelberg et al, 1995)并产生随机性错误,这些PCR虚假产物的检出是出现假阳性结果的原因之一。此外,在分子

操作过程中对大量平行样本进行标记后, PCR产物中可能会出现标签互换(tag switching, 即由于合并样本带来的污染)现象也是交叉污染的间接来源。在早期的大量研究中,为节约成本同时消除测序批次带来的系统误差,在测序前处理时通常在引物外侧加入一段分子标签以达到对同一批次PCR扩增产物进行平行测序的目的,通过标签对不同来源的序列进行区分,进而分别进行序列拼接和数据处理及物种检出。这种操作存在一些潜在问题,例如操作不当会导致存在少部分未连接上标签的引物序列,保存不当会导致PCR产物稳定性下降使标签从引物上脱落或随机连接到其他序列上。这些情况都可能干扰扩增子测序,导致序列读取错误并造成交叉污染,发生II型错误(Carlsen et al, 2012)。

增加生物学样本重复数量是解决标签互换问题最为经济有效的方法。此外,可以使用成套的标签同时对正向和反向引物进行标记,避免相同批次样本对标签的重复使用(Carlsen et al, 2012)。每一步都设计相应的阴性对照,既可以减少系统误差,又可成为数据筛选时的重要参考。这种多因素导致的污染虽然难以避免,但是缩短PCR产物的保存时间、严格遵循标准化方法并执行规范操作,可以在一定程度上维持PCR产物的完整性并降低污染带来的影响。

#### 4.2 物种参考数据库的完整性

利用环境DNA-宏条形码技术进行物种鉴定,最终要回归到测序序列和参考数据库的比对上。因此,参考数据库的完整性和质量直接决定了运用环境DNA-宏条形码技术进行物种鉴定的可靠性(杨江华, 2017)。虽然近年来参考数据库中物种信息增长很快, GenBank、Barcode of Life (BOLD)等公共数据库收录了大量数据,但相对陆生生态系统,对水生生态系统收录的类群较少。数据库的不完善使得大量的环境DNA-宏条形码数据不能得到正确注释,从而大大降低了入侵种的检出率。Zhan等(2014a)的研究表明,从加拿大安大略省汉密尔顿港收集的浮游生物群落样本通过高通量测序后获得的353个OTU中,当以相似度99%拟合时,只有32个(9.1%)与公共数据库中收录的可用序列相匹配。另一方面,即使是收录在数据库中的物种,由于分类学的错综复杂性,也可能被归为不同的分类单元。此外,文库中收录的数据多为COI和16S rRNA等常见分子



标记基因的序列片段, 基于其他分子标记基因的数据较少, 在鉴定复杂群落和相似度较高的近缘种时不能满足需要。数据库收录的物种种类有限以及收录到的序列片段仅基于常见的分子标记基因, 使得文库数据在可用性上不尽人意。

随着环境DNA-宏条形码技术的高速发展, 世界各地在建的条形码数据库的开源与共享, 无疑为物种注释提供了丰富的资源。对于地方性入侵物种而言, 开展多方协作也是快速直接解决数据库问题的有效途径。

## 5 展望

生物入侵严重威胁着全球范围的水生生物多样性, 如果不及时采取有效的保护措施, 越来越多的本地物种将会以更快的速度面临濒危乃至灭绝。环境DNA-宏条形码技术具有高通量、高灵敏度、高分辨率的技术优势, 为水生生态系统入侵生物的早期监测预警系统的建立提供了条件。

随着测序技术的进一步发展, 第三代测序技术悄然而生, 单分子测序(如单分子实时测序技术、纳米孔单分子测序技术等)已逐渐成为新趋势。第三代测序技术在测序长度和准确性方面都有大幅度提高, 其极高的单分子分辨率及可以捕捉动态信息的优势是现有技术的补充。与第一代和第二代测序相比, 支持更高通量测序的第三代测序技术无需进行PCR扩增, 可以解决PCR扩增偏向性的问题, 使得序列比对分析将摆脱OTU聚类的方法, 这会极大程度上提高检出准确度。因此, 充分结合第三代测序技术优势, 优化环境DNA-宏条形码技术在水生入侵物种监测和预警过程中的应用, 是今后应该重点考虑的内容。

构建完善的物种DNA条形码的标准数据库及物种信息库、建立资源信息共享和应用平台是利用环境DNA-宏条形码技术进行入侵物种鉴定的基础。生物信息技术的发展将有助于环境DNA-宏条形码技术适应更加复杂多变的自然条件和物种类型, 提高监测灵敏度, 使得该项技术在更大的范围内得以推广应用(唐敏等, 2013)。环境DNA-宏条形码技术可以高效监测最易遭受生物入侵的区域, 在早期监测和预警中有着广阔的应用空间和发展潜能, 为水生入侵生物的检验检疫及立法工作提供数据基础和技术支持。该技术的发展将推动生物入侵

早期监测进入一个崭新的时代, 为全球生物多样性保护做出卓越的贡献。

## 参考文献

- Avó AP, Daniell TJ, Neilson R, Oliveira S, Branco J, Adão H (2017) DNA barcoding and morphological identification of benthic nematodes assemblages of estuarine intertidal sediments: Advances in molecular tools for biodiversity assessment. *Frontiers in Marine Science*, 4, 66.
- Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kausrud H (2010) ITS as an environmental DNA barcode for fungi: An *in silico* approach reveals potential PCR biases. *BMC Microbiology*, 10, 189.
- Brannock PM, Ortman AC, Moss AG, Halanych KM (2016) Metabarcoding reveals environmental factors influencing spatiotemporal variation in pelagic microeukaryotes. *Molecular Ecology*, 25, 3593–3604.
- Bråte J, Logares R, Berney C, Ree DK, Klaveness D, Jakobsen KS, Tabrizi KS (2010) Freshwater Perkinsea and marine-freshwater colonizations revealed by Pyrosequencing and phylogeny of environmental rDNA. *The ISME Journal*, 4, 1144–1153.
- Briski E, Cristescu ME, Bailey SA, MacIsaac HJ (2011) Use of DNA barcoding to detect invertebrate invasive species from diapausing eggs. *Biological Invasions*, 13, 1325–1340.
- Boessenkool S, Epp LS, Haile J, Bellemain E (2012) Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular Ecology*, 21, 1806–1815.
- Bourlat SJ, Borja A, Gilbert J, Taylor MI, Davies N, Weisberg SB, Griffith JF, Lettieri J, Field D, Benzie J, Glpckner FO, Rodríguez-Ezpeleta N, Faith DP, Bean TP, Obst M (2013) Genomics in marine monitoring: New opportunities for assessing marine health status. *Marine Pollution Bulletin*, 74, 19–31.
- Bucklin A, Ortman BD, Jennings RM, Nigro LM, Sweetman CJ, Copley NJ, Sutton T, Wiebec PH (2010) A “Rosetta Stone” for metazoan zooplankton: DNA barcode analysis of species diversity of the Sargasso Sea (Northwest Atlantic Ocean). *Deep-Sea Research Part II: Topical Studies in Oceanography*, 57, 2234–2247.
- Carlsen T, Aas AB, Lindner D, Vrålstad T, Chumacher T, Kausrud H (2012) Don’t make a mistake: Is tag switching an overlooked source of error in amplicon Pyrosequencing studies? *Fungal Ecology*, 5, 747–749.
- Chen L, Wu L, Liu Y, Xu HG (2016) Application of environmental DNA metabarcoding in ecology. *Acta Ecologica Sinica*, 36, 4573–4582. (in Chinese with English abstract) [陈炼, 吴琳, 刘燕, 徐海根 (2016) 环境DNA metabarcoding及其在生态学研究中的应用. *生态学报*, 36, 4573–4582.]
- Chen Y, Sun C, Zhan A (2017) Biological invasions in aquatic

- ecosystems in China. *Aquatic Ecosystem Health & Management*, 10, 402–412.
- Chen YY, Gao YC, Peng H, Xiong W, Li SG, Zhan AB (2018) Community structure of benthic macroinvertebrates and water quality assessment in the Songhua River. *Journal of Biosafety*, 27(2), 95–104. (in Chinese with English abstract) [陈义永, 高养春, 彭衡, 熊薇, 李世国, 战爱斌 (2018) 松花江流域大型底栖动物群落结构与水质评价. *生物安全学报*, 27(2), 95–104.]
- Cheung MK, Au CH, Chu KH (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 Pyrosequencing. *The ISME Journal*, 4, 1053–1059.
- Darling JA, Mahon AR (2011) From molecules to management: Adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environmental Research*, 111, 978–988.
- Davey ML, Kausserud H, Ohlson M (2014) Forestry impacts on the hidden fungal biodiversity associated with bryophytes. *FEMS Microbiology Ecology*, 90, 313–325.
- Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the cytochrome C oxidase subunit I marker: Not a perfect match. *Biology Letters*, 10, 20140562.
- Dudgeon D, Arthington AH, Gessner MO (2006) Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, 81, 163–182.
- Emily AB, Frederic JJ, Zhan A, Hugh JM, Melania EC (2016) Detection of aquatic invaders using metabarcoding reveals a high number of non-indigenous species in Canadian ports. *Diversity and Distributions*, 22, 1045–1059.
- Egan SP, Grey E, Olds B (2015) Rapid molecular detection of invasive species in ballast and harbor water by integrating environmental DNA and light transmission spectroscopy. *Environmental Science & Technology*, 49, 4113–4121.
- Elbrecht V, Taberlet P, Dejean T, Valentini A, Usseglio-Polatera P, Beisel JN, Coissac E, Boyer F, Leese F (2016) Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. *PeerJ*, 4, e1966.
- Engelbrekton A, Kunin V, Wrighton K, Zvenigorodsky N, Chen F, Ochman H, Hugenholtz P (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *The ISME Journal*, 4, 642–647.
- Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology Letters*, 4, 423–425.
- Ficetola GF, Pansu J, Bonin A, Coissac E, Giguët-Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Pompanon F, Rayé G, Taberlet P (2015) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, 15, 543–556.
- Galand PE, Casamayor EO, Kirchman DL (2009) Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing. *The ISME Journal*, 3, 860–869.
- Goldberg CS, Sepulveda A, Ray A, Baumgardt J, Waits LP (2013) Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science*, 32, 792–800.
- Hao YB, Zhang AJ, Liu JD, Gu ZM (2018) Application of environmental DNA technology in the study of fish resources. *Biotechnology Bulletin*, 34(12), 56–62. (in Chinese with English abstract) [郝雅宾, 张爱菊, 刘金殿, 顾志敏 (2018) 环境DNA技术在鱼类资源研究中的应用. *生物技术通报*, 34(12), 56–62.]
- Hajibabaei M, Shokralla S, Zhou X (2011) Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, 6, e17497.
- Hambler C, Henderson PA, Speight MR (2011) Extinction rates, extinction-prone habitats, and indicator groups in Britain and at larger scales. *Biological Conservation*, 144, 713–721.
- Holland RA, Darwall WRT, Smith KG (2012) Conservation priorities for freshwater biodiversity: The key biodiversity area approach refined and tested for continental Africa. *Biological Conservation*, 148, 167–179.
- Jerde CL, Mahon AR, Chadderton WL, Lodge DM (2011) “Sight-unseen” detection of rare aquatic species using environmental DNA. *Conservation Letters*, 4, 150–157.
- Jiang W, Wang QJ, Deng J (2016) Protocol optimization of eDNA analysis workflow for detecting *Hucho bleekeri*. *Chinese Journal of Applied Ecology*, 27, 2372–2378. (in Chinese with English abstract) [姜维, 王启军, 邓捷 (2016) 以川陕哲罗鲑为目标物种的水样环境DNA分析流程的优化. *应用生态学报*, 27, 2372–2378.]
- Ju RT, Li H, Shih CJ, Li B (2012) Progress of biological invasions research in China over the last decade. *Biodiversity Science*, 20, 581–611. (in Chinese with English abstract) [鞠瑞亭, 李慧, 石正人, 李博 (2012) 近十年中国生物入侵研究进展. *生物多样性*, 20, 581–611.]
- Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering*, 96, 317–323.
- Kausserud H, Kumar S, Brysting AK, Norden J, Carlsen T (2012) High consistency between replicate 454 Pyrosequencing analyses of ectomycorrhizal plant root samples. *Mycorrhiza*, 22, 309–315.
- Kelly RP, Port JA, Yamahara KM, Crowder LB (2014) Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE*, 9, e86175.
- Keskin E, Atar HH (2013) DNA barcoding commercially important fish species of Turkey. *Molecular Ecology Resources*, 13, 788–797.
- Kunin V, Engelbrekton A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: Pyrosequencing errors can

- lead to artificial inflation of diversity estimates. *Environmental Microbiology*, 12, 118–123.
- Lang DD, Tang M, Zhou X (2018) Qualitative and quantitative molecular construction of plant–pollinator network: Application and prospective. *Biodiversity Science*, 26, 445–456. (in Chinese with English abstract) [郎丹丹, 唐敏, 周欣 (2018) 传粉网络构建的定性定量分子研究: 应用与展望. *生物多样性*, 26, 445–456.]
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, 10, 34.
- Levy-Booth DJ, Campbell RG, Gulden RH, Hart MM, Powell JR (2007) Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry*, 39, 2977–2991.
- Li FL, Yang JH, Yang YN, Zhang XW (2018) Using environmental DNA metabarcoding to monitor the changes and health status of aquatic ecosystems. *Environmental Monitoring in China*, 34(6), 37–46. (in Chinese with English abstract) [李飞龙, 杨江华, 杨雅楠, 张效伟 (2018) 环境DNA宏条形码监测水生态系统变化与健康状态. *中国环境监测*, 34(6), 37–46.]
- Li Q, Wu JG, Kou XJ (2013) Applications of camera trap in wildlife population ecology. *Chinese Journal of Applied Ecology*, 24, 947–955. (in Chinese with English abstract) [李勤, 郭建国, 寇晓军 (2013) 相机陷阱在野生动物种群生态学中的应用. *应用生态学报*, 24, 947–955.]
- Lin Y, Gao Z, Zhan A (2015) Introduction and use of non-native species for aquaculture in China: Status, risks and management solutions. *Reviews Aquaculture*, 7, 28–58.
- Liu FM, Miu JL, Zheng Z, Wang YB (2007) Present situation, hazards and prevention and control measures of marine invasive alien species into China. *Coastal Engineering*, 26(4), 49–57. (in Chinese with English abstract) [刘芳明, 缪锦来, 郑洲, 王以斌 (2007) 中国外来海洋生物入侵的现状、危害及其防治对策. *海岸工程*, 26(4), 49–57.]
- Lockwood JL, Hoopes MF, Marchetti MP (2013) *Invasion Ecology*, 2nd edn. Blackwell Scientific Publications, Chichester, West Sussex.
- MacKenzie DIM, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255.
- Ma HJ, Stewart K, Ma LM, Ren WW, Zhao JF (2016) Environmental DNA and its application in protecting aquatic ecosystem. *Chinese Journal of Ecology*, 35, 516–523. (in Chinese with English abstract) [马鸿娟, Stewart K, 马利民, 任文伟, 赵建夫 (2016) 环境DNA及其在水生生态系统保护中的应用. *生态学杂志*, 35, 516–523.]
- McDonald LL (2004) Sampling rare populations. In: *Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters* (ed. Thompson WL), pp. 11–42. Island Press, New York.
- Odelberg SJ, Weiss RB, Hata A, White R (1995) Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase. *Nucleic Acids Research*, 23, 2049–2057.
- Pagenkopp LKM, Fleischer RC, Carney KJ, Holzer KK, Ruiz GM (2016) Amplicon-based Pyrosequencing reveals high diversity of protistan parasites in ships' ballast water: Implications for biogeography and infectious diseases. *Microbiology Ecology*, 71, 530–542.
- Pietramellara G (2009) Extracellular DNA in soil and sediment: Fate and ecological relevance. *Biology and Fertility of Soils*, 45, 219–235.
- Polz MF, Cavanaugh CM (1998) Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology*, 64, 3724–3730.
- Prosser JI (2010) Replicate or lie. *Environmental Microbiology*, 12, 1806–1810.
- Pysek P, Richardson DM (2010) Invasive species, environmental change and management, and health. *Annual Review of Environment and Resources*, 35, 25–55.
- Rees HC, Maddison BC, Middleditch DJ (2014) The detection of aquatic animal species using environmental DNA: A review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51, 1450–1459.
- Saunders GW (2005) Applying DNA barcoding to red macroalgae: A preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360, 1879–1888.
- Schmieder R, Edwards R (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS ONE*, 6, e17288.
- Shan XJ, Li M, Wang WJ (2018) Application of environmental DNA technology in aquatic ecosystem. *Progress in Fishery Sciences*, 39(3), 23–29. (in Chinese with English abstract) [单秀娟, 李苗, 王伟继 (2018) 环境DNA (eDNA)技术在水生生态系统中的应用研究进展. *渔业科学进展*, 39(3), 23–29.]
- Shi XL, Lepère C, Scanlan DJ, Vaultot D (2011) Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific Ocean. *PLoS ONE*, 6, e18979.
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proceedings of the National Academy of Sciences, USA*, 103, 12115–12120.
- Song JW, Small MJ, Casman EA (2017) Making sense of the noise: The effect of hydrology on silver carp eDNA detection in the Chicago area waterway system. *Science of the Total Environment*, 605, 713–720.
- Sun C, Zhao Y, Li H, Dong Y, MacIsaac HJ, Zhan A (2015) Unreliable quantitation of species abundance based on high-throughput sequencing data of zooplankton communities.



- Aquatic Biology, 24, 9–15.
- Suzuki MT, Giovannoni SJ (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology*, 62, 625–630.
- Taberlet P, Coissac E, Hajibabaei M (2012) Environmental DNA. *Molecular Ecology*, 21, 1789–1793.
- Takahara T, Minamoto T, Yamanaka H, Doi H, Kawabata Z (2012) Estimation of fish biomass using environmental DNA. *PLoS ONE*, 7, e35868.
- Tang CQ, Leasi F, Obertegger U, Kieneke A, Barraclough TG, Fontaneto D (2012) The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences, USA*, 109, 16208–16212.
- Tang M, Yi TS, Wang X, Tan MH, Zhou X (2013) The application of metabarcoding technology in identification of plant species diversity. *Plant Diversity and Resources*, 35, 769–773. (in Chinese with English abstract) [唐敏, 伊廷双, 王欣, 谭美华, 周欣 (2013) Metabarcoding技术在植物鉴定和多样性研究中的应用. *植物分类与资源学报*, 35, 769–773.]
- Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist*, 188, 291–301.
- Thomsen PF, Møller PR, Sigsgaard EE, Knudsen SW, Jørgensen OA, Willerslev E (2016) Environmental DNA from seawater samples correlate with trawl catches of Subarctic, deepwater fishes. *PLoS ONE*, 11, e0165252.
- Tréguier A, Paillisson J, Dejean T (2014) Environmental DNA surveillance for invertebrate species: Advantages and technical limitations to detect invasive crayfish (*Procambarus clarkii*) in freshwater ponds. *Journal of Applied Ecology*, 51, 871–879.
- Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin JM, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Møller PR, Willerslev E, Dejean T (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25, 929–942.
- Veach AM, Dodds WK, Jumpponen A (2015) Woody plant encroachment, and its removal, impact bacterial and fungal communities across stream and terrestrial habitats in a tallgrass prairie ecosystem. *FEMS Microbiology Ecology*, 91, fiv109.
- Wan FH, Guo JY, Zhang F (2009) *Research on Biological Invasions in China*. Science Press, Beijing. (in Chinese) [万方浩, 郭建英, 张峰 (2009) 中国生物入侵研究. 科学出版社, 北京.]
- Wan YQ, Guo WB, Wu J, Li L, Xu HG (2017) Site occupancy model: Principles and research progress. *Journal of Ecology and Rural Environment*, 33, 673–679. (in Chinese with English abstract) [万雅琼, 郭伟波, 吴军, 李莉, 徐海根 (2017) 物种占域模型的基本原理及其研究进展. *生态与农村环境学报*, 33, 673–679.]
- Wang ZH, Chen JF, Yang YF (2010) Control and management of harmful algal bloom species introduced by ballast water. *Marine Environmental Science*, 29, 920–922. (in Chinese with English abstract) [王朝晖, 陈菊芳, 杨宇峰 (2010) 船舶压舱水引起的有害赤潮藻类生态入侵及其控制管理. *海洋环境科学*, 29, 920–922.]
- Weber AAT, Pawlowski J (2013) Can abundance of protists be inferred from sequence data: A case study of foraminifera. *PLoS ONE*, 8, e56739.
- Wu B, Zhao Q, Liu HM, Liu ZH, Sun XJ, Sun C, Zhou LQ, Yang AG (2018) Comparative analysis of different DNA barcoding methods for Veneroida classification and identification. *Journal of Fishery Sciences of China*, 25, 880–890. (in Chinese with English abstract) [吴彪, 赵庆, 刘寒苗, 刘志鸿, 孙秀俊, 孙超, 周丽青, 杨爱国 (2018) 不同DNA条形码基因在帘蛤目贝类分类鉴定中的比较分析. *中国水产科学*, 25, 880–890.]
- Wu YS, Tang YK, Li JL, Liu K, Li HX, Wang Q, Yu JH, Xu P (2019) The application of environmental DNA in the monitoring of the Yangtze finless porpoise, *Neophocaena phocaenoides asaeorientalis*. *Journal of Fishery Sciences of China*, 26(1), 126–134. (in Chinese with English abstract) [吴昀晟, 唐永凯, 李建林, 刘凯, 李红霞, 王钦, 俞菊华, 徐跑 (2019) 环境DNA在长江江豚监测中的应用. *中国水产科学*, 26(1), 126–134.]
- Xiong W, Li H, Zhan A (2016) Early detection of invasive species in marine ecosystems using high-throughput sequencing: Technical challenges and possible solutions. *Marine Biology*, 163, 139.
- Xu N, Chang JB (2016) Preliminary study on fish species detection in the middle and lower Yangtze River using environmental DNA. *Journal of Hydroecology*, 37(5), 49–55. (in Chinese with English abstract) [徐念, 常剑波 (2016) 长江中下游干流环境DNA样本鱼类物种检测的初步研究. *水生态学杂志*, 37(5), 49–55.]
- Yang JH (2017) *Adaptive Strategies of Clonal Plants Growing in Heterogeneous Environments*. PhD dissertation, Nanjing University, Nanjing. (in Chinese with English abstract) [杨江华 (2017) 太湖流域浮游动物物种多样性与环境污染群落生态效应研究. 博士学位论文, 南京大学, 南京.]
- Yoon TH, Kang HE, Kang CK, Lee SH, Ahn DH, Park H, Kim HW (2016) Development of a cost-effective metabarcoding strategy for analysis of the marine phytoplankton community. *PeerJ*, 4, e2.
- Zaiko A, Martinez JL, Schmidt-petersen J (2015a) Metabarcoding approach for the ballast water surveillance — An advantageous solution or an awkward challenge? *Marine*



- Pollution Bulletin, 92, 25–34.
- Zaiko A, Samuiloviene A, Ardura A, Garcia-Vazquez E (2015b) Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. *Marine Pollution Bulletin*, 100, 53–59.
- Zhan A, Hulák M, Sylvester F, Huang X, Adebayo AA, Abbott CL, Adamowicz SJ, Heath DD, Cristescu ME, Maclsaac HJ (2013) High sensitivity of 454 Pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution*, 4, 558–565.
- Zhan A, Bailey SA, Heath DD, Maclsaac HJ (2014a) Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. *Molecular Ecology Resources*, 14, 1049–1059.
- Zhan A, Xiong W, He S, Maclsaac HJ (2014b) Influence of artifact removal on rare species recovery in natural complex communities using high-throughput sequencing. *PLoS ONE*, 9, e96928.
- Zhan A, He EA, Brown FJJ, Chain TW, Therriault CL, Abbott DD, Heath ME, Maclsaac Cristescu HJ (2014c) Reproducibility of Pyrosequencing data for biodiversity assessment in complex communities. *Methods in Ecology and Evolution*, 5, 881–890.
- Zhan A, Maclsaac HJ (2015) Rare biosphere exploration using high-throughput sequencing: Research progress and perspectives. *Conservation Genetics*, 16, 513–522.
- Zhan A, Ni P, Xiong W, Chen Y, Lin Y, Huang X (2017) Biological invasions in aquatic ecosystems in China. *Aquatic Ecosystem Health & Management*, 10, 402–412.
- Zhou J, Wu L, Deng Y, Zhi X, Jiang Y, Tu Q, Xie J, Van Nostrand JD, He Z, Yang Y (2011) Reproducibility and quantitation of amplicon sequencing-based detection. *The ISME Journal*, 5, 1303–1313.
- Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q (2013) Ultradeep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, 2(1), 4.

(特邀责任编辑: 周欣 责任编辑: 黄祥忠)



•综述•

# 模式动物在共生微生物研究中的作用

肖雅倩<sup>1</sup> 刘传<sup>2</sup> 肖亮<sup>2\*</sup>

1 (北京师范大学水科学研究院, 北京 100875)

2 (深圳华大生命科学研究院宏基因组研究中心, 广东深圳 518000)

**摘要:** 共生微生物是一类定植于宿主体表或体内, 可执行宿主本身无法完成的功能, 并依赖于宿主所提供的生长环境的微生物。众多研究表明, 人体肠道共生微生物与免疫、营养、代谢, 甚至精神健康等生理功能密切相关, 是重要的“微生物器官”。在早期的肠道微生物研究中, 模式动物就已经作为研究工具被使用。随着肠道微生物研究的不断深入, 模式动物作为不可替代的研究对象发挥了越来越重要的作用。本综述主要对几种重要的模式动物如斑马鱼(*Danio rerio*)、小鼠(*Mus musculus*)、猪(*Sus scrofa domestica*)和猕猴(*Macaca mulatta*)在肠道微生物研究中的应用进行了总结, 介绍了各种模式动物的发展过程及特点, 各自在应用于研究时的优缺点, 以及利用这些动物模型在共生微生物领域所取得的一些标志性的科研成果。同时, 也就近年来在共生微生物领域新兴的一些模式生物如蜜蜂(*Apis*)、果蝇(*Drosophila*)、秀丽隐杆线虫(*Caenorhabditis elegans*)等进行了一些探讨。旨在让该领域的研究者们了解模式动物与人体在共生微生物方面的异同, 为更好地利用这一研究工具提供参考。

**关键词:** 共生微生物; 宿主; 模式动物; 肠道微生物

## The role of model animals in the study of symbiotic microorganisms

Yaqian Xiao<sup>1</sup>, Chuan Liu<sup>2</sup>, Liang Xiao<sup>2\*</sup>

1 College of Water Sciences, Beijing Normal University, Beijing 100875

2 BGI-Shenzhen, Shenzhen, Guangdong 518000

**Abstract:** Symbiotic microorganisms colonize external or internal surfaces of a host depending on environmental factors, and may supply the host with special functions. More and more researchers have proven that symbiotic gut microorganisms are related to a diverse range of physiological functions of a host including immunity, nutrition, metabolism and even mental health. Thus, gut microorganisms comprise an important “microbial organ” in humans. Since the early days of microbiota research, animal models have been used frequently for their microbiota, contributing greatly to new research in this field. This review provides an overview of animals used as models in symbiotic microorganism studies, including zebrafish (*Danio rerio*), mice (*Mus musculus*), pigs (*Sus scrofa domestica*), and monkeys (*Macaca mulatta*). We provide insight into the development and characteristics of these model animals, highlighting the advantages and disadvantages of each model, as well as any outstanding scientific achievements based on their use. We also note that honey bee (*Apis*), fruit fly (*Drosophila*) and nematode (*Caenorhabditis elegans*) models are emerging as more prevalent in recent gut microbiota studies. This paper will contribute to better understanding the similarities and differences between the microbiota of model animals and humans, while providing useful information for effectively implementing these animal models in future research.

**Key words:** symbiotic microorganisms; host; animal models; gut microbiota

共生是两种生物彼此互利地生存在一起, 缺少彼都不能生存的一类种间关系, 是生物之间相互关系的高度发展。豆科植物和固氮细菌之间即是典型的植物和微生物之间的共生关系。动物和微生物

共生的案例也有很多, 如牛、羊等反刍动物与瘤胃微生物的共生就是其中之一。随着对人体内环境认识的发展, 人体和肠道菌群之间的共生关系也逐渐被了解。一般情况下, 人体肠道正常菌群的巨大数

收稿日期: 2019-02-21; 接受日期: 2019-04-29

\* 通讯作者 Author for correspondence. E-mail: xiaoliang@genomics.cn

量可以阻止和抑制外来微生物的入侵,从而抵御某些病原微生物的感染。此外,它们还能提供人体自身不能合成的维生素B1、B2、K及叶酸等营养物质(Ramakrishna, 2013; Rowland et al, 2018),而人体肠道则为这些微生物提供了良好的栖息场所和生长环境。因此,共生微生物可以被定义为一类定殖于宿主体表或体内,可执行宿主本身无法完成的功能,并依赖于宿主所提供的生长环境的微生物。

随着宏基因组测序技术的发展,我们已经可以从给定的环境样本中获取所存在的全部微生物的遗传信息。人体微生物特别是人体肠道微生物,其基因数目远超宿主本身,并且因为其在宿主代谢和免疫调节方面发挥的重要作用(Clemente et al, 2012; Sommer & Bäckhed, 2013; Marchesi et al, 2016)而备受研究者关注。越来越多的研究报道揭示了菌群和人类疾病之间的关联,表明肠道菌群的紊乱极有可能是多种疾病的关键风险因素(Xie et al, 2016; Berer et al, 2017; Jie et al, 2017; Liu et al, 2017)。随着研究领域的不断扩展,研究者的眼光也不仅仅局限于人体肠道微生物,皮肤、口腔、生殖道等环境中的微生物信息及其与宿主健康之间的关联也已经被报道(Al-Shehri et al, 2016; Kilian et al, 2016; Zákostelská et al, 2016; Haque et al, 2017)。越来越多的研究表明,人体各个环境的共生微生物对人体健康有着重要影响,与众多疾病的发生发展密切相关。

随着宏基因组研究的进展,发展出了一套较为成熟的研究方法,即宏基因组关联分析(metagenome-wide association) (Qin et al, 2012; Wang & Jia, 2016),并已应用于众多科研项目中。该方法可以解析出与疾病状态或各种相关指标存在关联的微生物基因、功能通路及特定微生物物种信息。随着宏基因组领域研究的不断深入,研究者们不再满足于“关联”研究,进而开始探索人体共生微生物与宿主健康状态之间的因果关系及互作机理,这将为人体共生微生物在医疗保健及临床干预方面的应用奠定理论基础。而这一深入探索首先在研究对象上遇到了瓶颈。在“关联”分析中,研究对象通常是人群,将来自不同人群的样本的宏基因组信息进行比对,找出差异及其与表型数据的关联。但由于不同环境因素(人种、居住地、饮食习惯、工作环境等)造成个体间共生微生物组成的巨大差异,研究者们通常需要较大量的样本才能在分析中排除数据“噪音”,

得出可靠的分析结果。而解析“因果”关系,不但同样需要较大的人群样本,还需要对人群进行组织和跟踪随访工作,在成本和研究周期方面要求很高,同时,对个体表型信息的跟踪、收集也有更高的要求。最为重要的一点,在正常的伦理限制下,“科赫法则”完全无法在以人为研究对象的科研中实现。因此,对于人体共生微生物的深入研究来说,人群本身并不是适合的研究对象。

动物模型在各个科学研究领域都扮演着重要角色。同样,在人体共生微生物领域,模式动物因其高度一致性的背景、可控的实验环境及可人工干预的特点,发挥着重要作用。特别是对于着眼于解析“因果”关系的研究者们来说,符合动物伦理的实验动物是深入研究共生微生物与宿主间相互作用的最佳研究对象和重要研究工具。在本综述中,我们将对宏基因组领域常见或不太常见的动物模型及相关的经典研究报道进行简析,展现动物模型在宏基因组研究中的关键作用,并讨论其优劣势,以便于研究者们更好地了解和运用动物模型辅助科学研究。

## 1 几种重要的模式动物在共生微生物研究中的应用

### 1.1 斑马鱼(*Danio rerio*)模型

斑马鱼是一种热带淡水鱼,因其体侧具有像斑马一样纵向的暗蓝色与银色相间的条纹而得名。2013年4月,英国桑格中心(Sanger Centre)协同全球最权威的14家斑马鱼研究机构,将斑马鱼全基因组和人类基因组进行比较分析发现,斑马鱼基因组含有的26,206个蛋白编码基因与人类基因高度同源,人类2万多个蛋白编码基因都至少可以在斑马鱼体内找到1个同源基因与之对应(Howe et al, 2013)。其独特的生物学、基因组学、遗传学优势及其高度保守的疾病信号传导路径,使其成为研究人类疾病信号传导途径及活体高通量药物筛选的最佳模式生物之一。

而早在斑马鱼自身的基因组被解析之前,研究者们已经开始意识到这一模式动物在肠道菌群研究方面的作用。Rawls等(2004)报道了无菌(Germ Free, GF)斑马鱼技术,首次将斑马鱼这一模式动物引入了菌群研究领域。研究者们将无菌斑马鱼消化道的基因表达情况跟常规斑马鱼进行了比较,发现

有超过200个基因受到肠道微生物调控; 同时将其与小鼠的消化道基因进行比较, 发现59个基因的表达情况在这两个物种中是保守的。这一研究成果为斑马鱼在宿主-肠道微生物互作研究中的应用奠定了科学基础。此外, Keller等(2008)建立的数字扫描激光光谱荧光显微技术, 可以实现对体内细胞的直接观察, 虽然该技术最初的应用是观察斑马鱼的胚胎发育过程, 但也为微生态研究提供了一种全新的技术手段。

菌群在肠道中的动态定殖过程一直是研究的难点, 在人体或普通实验动物上是难以实现的。研究者们将荧光标记的细菌菌株与无菌斑马鱼幼体一起培养, 然后利用荧光显微技术获取标记细菌在斑马鱼肠道中生长的三维图像, 并利用荧光强度实时测定细菌的多度。这项研究充分发挥了斑马鱼作为动物模型的优势, 在活的脊椎动物肠道内测定了微生物生长的动力学数据, 证实其符合对数生长模型, 并且在空间分布上并不均匀, 而是依赖于不同位置的微生物密度。这是首次巧妙地结合无菌斑马鱼模型和荧光显微技术对微生物在肠道内部的空间结构及生长动态进行研究(Jemielita et al, 2014)。随后, 研究者还利用斑马鱼模型开展了多项研究, 验证了不同肠道细菌物种对宿主免疫系统的影响(Rolig et al, 2015), 肠道细菌在宿主其他器官发育过程中所起的重要作用(Burns et al, 2016; Stephens et al, 2016)。更为重要的是, 越来越多的科研转化实验也开始在斑马鱼实验平台上开展, 研究者们利用这一动物模型证实了肠道菌群对于宿主肠道动力的影响(Wiles et al, 2016), 益生菌对于肥胖和糖尿病症状的改善(Falcinelli et al, 2017), 甚至是对抑郁和压力症状也有一定的干预作用(Davis et al, 2016)。可以说, 在宏基因组研究领域, 不论是基础研究还是应用转化, 斑马鱼模型都发挥了重要作用, 为研究者们提供了便捷、可靠的研究工具。

但值得注意的是, 斑马鱼模型在宏基因组方面的应用也存在一些缺点和局限。已有的数据表明, 有近70%的人体基因可在斑马鱼中找到至少一个直系同源基因(Howe et al, 2013), 证实了该种模式动物与人类在基因组层面的相似性。但对于宏基因组研究来说, 共生微生物的组成受环境因素的影响要大得多。首先, 斑马鱼生存于水生环境中, 这会显著影响其肠道或其他部位定殖的微生物种类, 造成

其与人体共生微生物之间产生了较大的差异; 其次, 饮食会对肠道微生物造成强烈的影响, 这一点是研究者们所公认的。而斑马鱼的食性与人类完全不同, 因此, 饲养过程中所使用的饲料成分也会显著影响其肠道微生物的组成。研究者们还特地对在不同实验室饲养的斑马鱼肠道微生物进行了比较, 结果发现不同实验室来源的斑马鱼, 其肠道菌群存在显著差异(Roeselers et al, 2011)。此外, 从实验技术层面来说, 斑马鱼体型细小, 基本上不可能进行个体操作, 并且在实验取材上也存在困难, 对于粪便样本, 往往是通过解剖获得肠道内容物, 这一做法的直接后果是粪便细菌DNA中混杂了大量其他来源的核酸物质, 对后续的测序及数据分析造成较大影响。

### 1.2 小鼠(*Mus musculus*)模型

不论是在健康还是疾病状况下, 要探索人体自身的生物学问题, 都需要特定的观察对象和大量的实验素材。但不论科学家们如何努力, 这一领域研究总是面临着诸多固有的困难: 人为引入致病或毒性物质, 在伦理上是严格禁止的; 很难从人体获得所需要的各种组织样本; 人体的某些特定时期, 例如胚胎期, 是极难被研究者所接触到的。除此之外, 从遗传学角度来看, 人类的代际相对较长, 很难开展长期的遗传学研究; 此外, 几乎不可能控制某一人群的配偶关系, 使其满足遗传学研究对其后代遗传背景的要求。尽管体外细胞培养系统可以解决部分问题, 但仍然难以再现有机体的整体性及生物体各系统之间的相互作用(Paigen, 2003)。因此, 研究者们需要一种合适的动物来替代人体作为研究对象。

代际时间短、繁殖快速、可形成近交系、饲养方便, 使得小鼠成为较为理想的哺乳动物研究对象。另外, 小鼠还有一个不容忽视的优点: 成本低廉。早在1981年, 第一只转基因小鼠在实验室的诞生, 新型的分子生物学技术与传统的小鼠遗传学系统相结合, 标志着哺乳动物生物学在分子水平的研究多了一项有力的工具。而小鼠模型与人体的各项相关性研究也已具备了较为完备的技术基础。发展至今, 数百种小鼠近交品系、多种突变型、众多的遗传图谱(Paigen, 2003), 已经构成了一个复杂的遗传体系, 为各领域的研究者提供了可靠的科研材料。

在漫长的进化过程中, 作为人类祖先的灵长类动物和作为小鼠祖先的啮齿类动物基因被证实早



在9,000万年前就已经“分道扬镳”(Lucas et al, 2018), 但二者的基因组仍有85%以上的相似度(Church et al, 2009)。而转录组分析也发现, 人和小鼠在转录因子及调控网络上也存在高度的保守性(Cheng et al, 2014)。比较而言, 在人体功能研究当中, 小鼠模型是更优于体外细胞实验的一种研究手段。肠道细菌在食物消化过程中发挥重要作用, 因此它们与宿主营养状况密切相关。营养学领域的研究者们很早就开始关注肠道菌群在宿主营养方面的作用, 早在20世纪三、四十年代, 已经有相关的研究探索磺胺类药物(Najjar, 1943; Hosp, 1945), 抗生素(Emerson et al, 1945; Smith et al, 1945), 以及特定饮食(Friedman, 1936)和肠道细菌之间的关系, 这些研究均在鼠类模型上开展。如前所述, 20世纪80年代小鼠遗传学的快速发展, 产生了各种近交系及人工突变品种, 也为小鼠模型的广泛应用奠定了基础。尽管测序技术还十分滞后, 当时的研究者们也开始利用16S技术探索小鼠肠道细菌的组成, 并对其中的特定物种进行分离、研究(Snel et al, 1995), 而基于小鼠和人在基因组及转录、调控方面的相似性, 宿主和肠道菌群之间的相互作用也成为了研究者的关注重点。通过无菌鼠和普通小鼠的对比, 研究者们发现小肠表皮细胞糖基化功能的实现, 需要有肠道细菌的存在(Bry et al, 1996), 肠道细菌和宿主本身之间存在着共同进化的关系, 二者相互影响, 为特定生理功能研究和新的治疗方法的开发提供了潜在可能(Hooper & Gordon, 2001)。而无菌小鼠(germ free mice)模型也在肠道微生物研究中发挥了重要作用, 为解析肠道菌群对宿主健康特别是代谢方面的影响提供了有力的证据(Rabot et al, 2010; Grover & Kashyap, 2014; Baothman et al, 2016; Kübeck et al, 2016)。

尽管小鼠模型在各领域科学研究中发挥着无可替代的作用, 但随着研究的广泛开展, 人们也越来越深入地了解到小鼠本身的一些局限性。如前所述, 人和小鼠在自身转录因子及调控网络上存在高度的保守性。但也有研究揭示, 在小鼠基因组中, 有将近一半的转录因子结合位点未在人类基因组中找到同源序列, 并且整体的基因表达情况在这两个物种间也存在较大差异(Cheng et al, 2014)。而在肠道微生物研究领域, 类似的问题也同样存在。在消化道的生理结构上, 小鼠和人体就存在一些差异,

其中最大的不同就是在小鼠的消化道中存在一个前胃(forestomach), 其角质化的鳞片粘膜区别于正常胃部组织, 并且覆盖了整个胃部的2/3, 其作用主要是食物的储存(Ghoshal & Bal, 1989)。这样的结构差异导致了其中微生物物种的不同, 有研究发现, 存在于小鼠消化道中的罗伊氏乳杆菌(*Lactobacillus reuteri*)和在人体消化道中发现的同种细菌在基因组上存在着较大差异(Frese et al, 2011)。

随着高通量测序技术的发展, 研究者们得以对小鼠和人的肠道细菌组成进行深入的比较。在门水平上, 二者呈现出较高的相似度, 拟杆菌门和厚壁菌门均为优势物种(Ley et al, 2006; Rawls et al, 2006)。但需要注意的是, 脱铁杆菌门在小鼠肠道中定殖较多, 而在人体肠道中极为少见, 只在胃部少量存在(Bik et al, 2006)。此外, 还有一类值得关注的物种: 分枝丝状杆菌(segmented filamentous bacteria, SFB)。此细菌被多次报道与固有免疫系统(innate immune system)的成熟密切相关(Suzuki et al, 2004; Gaboriau-Routhiau et al, 2009; Ivanov et al, 2009), 此物种在小鼠肠道中较为常见, 但几乎没有在成人的肠道中发现。

而针对大规模小鼠肠道样本的宏基因组研究也对小鼠肠道细菌的组成进行了深入研究(Xiao et al, 2015), 并与已报道的人肠道细菌分布(Qin et al, 2010)进行了比较, 在总共60个核心细菌属中, 在小鼠和人体肠道中同时存在的仅有25个, 且丰度各有不同。在微生物基因方面, 二者差异更为巨大, 仅有4%的微生物基因同时存在于二者的微生物基因集合当中。但值得注意的是, 尽管小鼠和人类的肠道菌群在物种和基因上差异甚大, 但在功能层面却具有较高的相似性, 二者共享了超过80%的可注释功能。目前为止, 已有超过1,500株分离自人体肠道的细菌菌株被用于深入研究(Rajilic-Stojanovic & de Vos, 2014; Lagier et al, 2016), 而只有100株左右来自小鼠肠道的细菌被分离出来(Lagkouvardos et al, 2016)。对于小鼠肠道微生物的研究, 还有很多工作要做。

小鼠作为研究工具的优势毋庸置疑, 这也是它成为最广泛使用的实验模型的原因。与此同时, 当我们将小鼠应用于肠道宏基因组领域时, 也有一些值得注意的限制。如前所述, 小鼠和人体肠道本身在肠道微生物组成方面存在先天的差异, 在将小鼠实

验的结果向人体外推时, 这些差异是需要关注的。无菌小鼠的出现, 部分缩小了小鼠和人之间在肠道微生态方面的差异, 研究者们将来源于人的肠道菌群移植进无菌小鼠的肠道中, 发现有88%的人源细菌属可以在无菌小鼠的肠道中定殖(Turnbaugh et al, 2009)。而这一技术最著名的应用在于以肥胖为代表的代谢性疾病研究(Ley et al, 2005; Geurts et al, 2011; Wang et al, 2017), 证实了肠道菌群与代谢性疾病之间的因果关系。但需要注意的是, 在高度流程化的小鼠实验过程中, 饲喂过程对于小鼠肠道微生态的影响极为巨大, 进而可能影响到最终的实验结果。在某些情况下, 鼠房的更新和改造, 会导致原本可观察的某些表型完全消失, 饲喂环境的影响甚至比遗传背景还要大(Friswell et al, 2010; Verbeke et al, 2015; Xiao et al, 2015)。这不得不引起人们的注意, 也是研究者在应用小鼠模型时所必须考虑的问题。

大鼠作为小鼠的“近亲”, 因为操作复杂度和成本等原因, 在微生态研究领域的应用广泛程度不及小鼠, 但因为二者的相似程度, 其特点和优缺点比较类似。也有研究者对SD (Sprague-Dawley)大鼠的肠道微生物特征进行了研究, 将人、小鼠、大鼠的肠道微生物数据集合进行了比较, 发现大鼠与人之间共享的肠道微生物基因数量和功能基因簇数量要略高于小鼠(Pan et al, 2018)。

### 1.3 猪(*Sus scrofa domestica*)模型和猕猴(*Macaca mulatta*)模型

除了上述的斑马鱼和小鼠, 猪作为一种大型哺乳动物也在很多科学研究中被作为模型使用。实际上, 作为人类重要的食物来源之一, 关于猪本身的各种营养和疾病相关问题早已经成为研究者们关注的领域。和小鼠类似, 营养学家们最早开始关注猪肠道菌群和宿主营养之间的联系。早在1936年, 就已经有了关于无菌猪技术的相关报道(Glimstedt, 1936), 随后的研究也将猪和人的肠道微生物进行了比较, 发现其在基因、物种和功能层面, 确实比小鼠更接近于人的肠道(Hildebrand et al, 2012; Xiao et al, 2016), 并且在消化系统及食性方面, 猪和人之间也更为相似。但和小型动物相比, 猪作为动物模型在研究应用中最大的问题在于维持成本较高。由于体型过大, 对饲喂场地要求较高; 指标测量、组织采样等操作也远比小型动物繁琐, 这些因素都限制了猪模型的使用范围。但随着遗传技术特别是

分子遗传技术的发展, 研究者们已经开始针对猪这个物种进行各种遗传操作, 以符合科研或产业应用的需求(Sachs & Galli, 2009), 这其中也包括小型化(Cyranoski, 2015), 这样的操作还存在一定的问题和争议, 但体型的缩小无疑解决了猪作为研究对象的一个关键性问题。

作为和人类最为接近的灵长类动物, 猴子无疑也是理想的模型动物。在医药研究中, 猩猩、恒河猴等非人灵长类(nonhuman primates, NHPs)动物为各类药物、疫苗的开发做出了巨大贡献。猴模型在肠道宏基因组中的应用比较晚, 但近年来也开始被广泛用于研究饮食、抗生素、遗传背景等对于肠道微生物的影响(Angelakis et al, 2016; Harris et al, 2016; Zevin et al, 2017)。通过人肠道微生态的比较, 也确实证实了白臀叶猴(*Pygathrix nemaeus*)、鬃毛吼猴(*Alouatta palliata*)、食蟹猴(*Macaca fascicularis*)在肠道微生物组成上较高的一致性(Clayton et al, 2016; Li et al, 2018)。与猪模型一样, 猴模型的应用也受到饲喂环境和成本的制约, 难以像小型动物那样大规模使用。并且也有研究报道, 笼养的NHPs肠道微生物与野生状态下存在较大差异(Clayton et al, 2016), 这其实也是所有模型动物在应用于宏基因组研究时所共同面临的问题: 同一类模型动物, 不同的饲喂场所, 环境因素往往会造成肠道微生态的差异, 可能造成实验结果的不稳定(Stappenbeck & Virgin, 2016), 这是在将动物模型应用于宏基因组研究时需要考虑的因素。而大型动物由于饲养环境复杂、控制度较低, 这一因素的影响会更为明显。

## 2 新兴动物模型及展望

传统的临床微生物研究主要关注某一种病原微生物对人类疾病的影响和作用, 随着高通量测序技术的突破, 人们已经可以针对特定环境中的微生物群落进行研究。现在已经了解到, 人体共生微生物由多种多样的细菌、古菌及真菌组成, 它们中的绝大多数是非病原菌, 但却通过各种机制、途径影响着宿主的健康, 包括: 营养及药物代谢(Wallace et al, 2010; de Clercq et al, 2016)、必需维生素的合成(Kau et al, 2011)、抵抗病原菌(Abt & Pamer, 2014)、宿主胆汁酸的次级代谢(Wahlström et al, 2016)、免疫调控(Round & Mazmanian, 2009; Belkaid & Hand, 2014)、感染的抵抗或敏感(Buffie et al, 2015), 甚至

宿主行为的调控(Dinan et al, 2015)。因为与人体健康之间的密切联系,使得以共生微生物为主要研究对象的宏基因组学成为了近年来的研究焦点。

在宏基因组学研究过程中,动物模型作为不可或缺的研究工具,自然也受到人们的重视。如前所述,研究者们不断尝试新技术、新方法,将动物模型应用到多种共生微生物(主要为肠道微生物)的研究中,并取得了众多的研究成果,其中一些结论甚至颠覆了人们之前的普遍认识。毫无疑问,模型动物为这一领域做出了重要贡献。但值得注意的是,模型动物的使用过程中也存在一些固有的问题和困难,以至于它们本身也成为了研究目的。研究者们同样进行了大量研究来探索如何更经济、更准确地使用模型动物。本文综述了一些常见的模型动物在宏基因组领域的研究情况,以及值得注意的问题和优缺点。实际上,模型动物的范围还在不断扩展,不断有新兴的模型被应用于特定领域的研究作为农业甚至整个生态环境的重要一员,蜜蜂(*Apis*)的肠道微生物生态早已引起了研究者的兴趣(Cox-Foster et al, 2007),并且蜜蜂作为一种较为新颖的模型昆虫也已经被应用在行为及认知功能方面的研究中(Menzel, 2012)。蜜蜂肠道菌群结构较为简单,繁殖和饲养相比大型动物来说更好操控,这些优点十分有利于将其应用于共生微生物的研究当中(Engel et al, 2016)。同样地,果蝇作为十分成熟的模型昆虫,被广泛应用在遗传学等多个领域的研究中(Adams et al, 2000; Reiter et al, 2001; Valente et al, 2004; Lloyd & Taylor, 2010),且研究者们基于该种模式生物构建了多种疾病模型(Pandey & Nichols, 2011),包括二型糖尿病模型(Álvarez-Rendón et al, 2018),越来越多的基于果蝇的微生态研究(Clark et al, 2015; Leftwich et al, 2018; Obadia et al, 2018)也预示着其作为该领域新型研究工具的重要性。秀丽隐杆线虫(*Caenorhabditis elegans*)作为一种经典的模式生物,其生命周期只有数周,在这期间完成成长、繁殖、衰老,最终死亡的整个过程,研究者们依据线虫的这一特点,研究了肠道微生物特定基因对线虫寿命的影响,取得了非常显著的进展(Han et al, 2017)。此外,线虫还被用于研究肠道微生物在癌症化疗中的作用(García-González et al, 2017; Scott et al, 2017),展现了其作为经典模式生物在微生态这一新兴领域所发挥的重要作用。与此同时,体外技

术也在不断发展,类器官、器官芯片等新技术的出现,让研究者们有了更多的选择。可以预见,共生微生物领域的研究将随着新模型、新技术的出现进入全新的阶段。

## 参考文献

- Abt MC, Pamer EG (2014) Commensal bacteria mediated defenses against pathogens. *Current Opinion in Immunology*, 29, 16–22.
- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Sidén-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC (2000) The genome sequence of *Drosophila melanogaster*. *Science*, 287, 2185–2195.
- Al-Shehri SS, Sweeney EL, Cowley DM, Liley HG, Rana-singhe PD, Charles BG, Shaw PN, Vagenas D, Duley JA,



- Knox CL (2016) Deep sequencing of the 16S ribosomal RNA of the neonatal oral microbiome: A comparison of breast-fed and formula-fed infants. *Scientific Reports*, 6, 38309.
- Álvarez-Rendón JP, Salceda R, Riesgo-Escovar JR (2018) *Drosophila melanogaster* as a model for diabetes type 2 progression. *BioMed Research International*, 8, 1417528.
- Angelakis E, Yasir M, Bachar D, Azhar EI, Lagier JC, Bibi F, Jiman-Fatani AA, Alawi M, Bakarman MA, Robert C, Raoult D (2016) Gut microbiome and dietary patterns in different Saudi populations and monkeys. *Scientific Reports*, 6, 32191.
- Baothman OA, Zamzami MA, Taher I, Abubaker J, Abu-Farha M (2016) The role of gut microbiota in the development of obesity and diabetes. *Lipids in Health and Disease*, 15, 108–116.
- Belkaid Y, Hand TW (2014) Role of the microbiota in immunity and inflammation. *Cell*, 157, 121–141.
- Berer K, Gerdes LA, Cekanaviciute E, Jia X, Xiao L, Xia Z, Liu C, Klotz L, Stauffer U, Baranzini SE, Kämpfel T, Hohlfeld R, Krishnamoorthy G, Wekerle H (2017) Gut microbiota from multiple sclerosis patients enables spontaneous autoimmune encephalomyelitis in mice. *Proceedings of the National Academy of Sciences, USA*, 114, 10719–10724.
- Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F, Perez-Perez G, Blaser MJ, Relman DA (2006) Molecular analysis of the bacterial microbiota in the human stomach. *Proceedings of the National Academy of Sciences, USA*, 103, 732–737.
- Bry L, Falk PG, Midtvedt T, Gordon JI (1996) A model of host–microbial interactions in an open mammalian ecosystem. *Science*, 273, 1380–1383.
- Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, Littmann E, van den Brink MRM, Jenq RR, Taur Y, Sander C, Cross JR, Toussaint NC, Xavier JB, Pamer EG (2015) Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature*, 517, 205–208.
- Burns AR, Stephens WZ, Stagaman K, Wong S, Rawls JF, Guillemin K, Bohannan BJ (2016) Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *The ISME Journal*, 10, 655–664.
- Cheng Y, Ma Z, Kim B-H, Wu W, Cayting P, Boyle AP, Sundaram V, Xing X, Dogan N, Li J, Euskirchen G, Lin S, Lin Y, Visel A, Kawli T, Yang X, Patacsil D, Keller CA, Giardine B, Consortium TME, Kundaje A, Wang T, Pennacchio LA, Weng Z, Hardison RC, Snyder MP (2014) Principles of regulatory information conservation between mouse and human. *Nature*, 515, 371–375.
- Church DM, Goodstadt L, Hillier LW, Zody MC, Goldstein S, She X, Bult CJ, Agarwala R, Cherry JL, DiCuccio M, Hlavina W, Kapustin Y, Meric P, Maglott D, Birtle Z, Marques AC, Graves T, Zhou S, Teague B, Potamousis K, Churas C, Place M, Herschleb J, Runnheim R, Forrest D, Amos-Landgraf J, Schwartz DC, Cheng Z, Lindblad-Toh K, Eichler EE, Ponting CP, Consortium TMGS (2009) Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biology*, 7, e1000112.
- Clark RI, Salazar A, Yamada R, Fitz-Gibbon S, Morselli M, Alcaraz J, Rana A, Rera M, Pellegrini M, Ja WW, Walker DW (2015) Distinct shifts in microbiota composition during *Drosophila* aging impair intestinal function and drive mortality. *Cell Reports*, 12, 1656–1667.
- Clayton JB, Vangay P, Huang H, Ward T, Hillmann BM, Al-Ghalith GA, Travis DA, Long HT, Tuan BV, Minh VV, Cabana F, Nadler T, Toddes B, Murphy T, Glander KE, Johnson TJ, Knights D (2016) Captivity humanizes the primate microbiome. *Proceedings of the National Academy of Sciences, USA*, 113, 10376–10381.
- Clemente JC, Ursell LK, Parfrey LW, Knight R (2012) The impact of the gut microbiota on human health: An integrative view. *Cell*, 148, 1258–1270.
- Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan PL, Briese T, Hornig M, Geiser DM, Martinson V, vanEngelsdorp D, Kalkstein AL, Drysdale A, Hui J, Zhai J, Cui L, Hutchison SK, Simons JF, Egholm M, Pettis JS, Lipkin WI (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*, 318, 283–287.
- Cyranoski D (2015) Gene-edited “micropigs” to be sold as pets at Chinese institute. *Nature News*, 526, 18.
- Davis DJ, Doerr HM, Grzelak AK, Busi SB, Jasarevic E, Ericsson AC, Bryda EC (2016) *Lactobacillus plantarum* attenuates anxiety-related behavior and protects against stress-induced dysbiosis in adult zebrafish. *Scientific Reports*, 6, 33726.
- de Clercq NC, Groen AK, Romijn JA, Nieuwdorp M (2016) Gut microbiota in obesity and undernutrition. *Advances in Nutrition*, 7, 1080–1089.
- Dinan TG, Stilling RM, Stanton C, Cryan JF (2015) Collective unconscious: How gut microbes shape human behavior. *Journal of Psychiatric Research*, 63, 1–9.
- Emerson GA, Smith DG, Wurtz E, Clark D (1945) Induction of nutritional deficiency by oral administration of streptomycin. *Journal of Pharmacology and Experimental Therapeutics*, 85, 336–342.
- Engel P, Kwong WK, McFrederick Q, Anderson KE, Barribeau SM, Chandler JA, Cornman RS, Dainat J, de Miranda JR, Doublet V, Emery O, Evans JD, Farinelli L, Flenniken ML, Granberg F, Grasis JA, Gauthier L, Hayer J, Koch H, Kocher S, Martinson VG, Moran N, Munoz-Torres M, Newton I, Paxton RJ, Powell E, Sadd BM, Schmid-Hempel P, Schmid-Hempel R, Song SJ, Schwarz RS, vanEngelsdorp D, Dainat B (2016) The bee microbiome: Impact on bee

- health and model for evolution and ecology of host-microbe interactions. *mBio*, 7, e02164-15.
- Falcinelli S, Rodiles A, Hatfe A, Picchiatti S, Cossignani L, Merrifield DL, Unniappan S, Carnevali O (2017) Dietary lipid content reorganizes gut microbiota and probiotic *L. rhamnosus* attenuates obesity and enhances catabolic hormonal milieu in zebrafish. *Scientific Reports*, 7, 5512.
- Frese SA, Benson AK, Tannock GW, Loach DM, Kim J, Zhang M, Oh PL, Heng NCK, Patil PB, Juge N, MacKenzie DA, Pearson BM, Lapidus A, Dalin E, Tice H, Goltsman E, Land M, Hauser L, Ivanova N, Kyrpidis NC, Walter J (2011) The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. *PLoS Genetics*, 7, e1001314.
- Friedman H (1936) A study of the fecal flora and the line test of normal rats, rachitic rats and heating rachitic rats. *Journal of Nutrition*, 12, 165–172.
- Friswell MK, Gika H, Stratford IJ, Theodoridis G, Telfer B, Wilson ID, McBain AJ (2010) Site and strain-specific variation in gut microbiota profiles and metabolism in experimental mice. *PLoS ONE*, 5, e8584.
- Gaboriau-Routhiau V, Rakotobe S, Lécuyer E, Mulder I, Lan A, Bridonneau C, Rochet V, Pisi A, De Paepe M, Brandi G, Eberl G, Snel J, Kelly D, Cerf-Bensussan N (2009) The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity*, 31, 677–689.
- García-González AP, Ritter AD, Shrestha S, Andersen EC, Yilmaz LS, Walhout AJM (2017) Bacterial metabolism affects the *C. elegans* response to cancer chemotherapeutics. *Cell*, 169, 431–441.
- Geurts L, Lazarevic V, Derrien M, Everard A, Van Roye M, Knauf C, Valet P, Girard M, Muccioli GG, Francois P, de Vos WM, Schrenzel J, Delzenne NM, Cani PD (2011) Altered gut microbiota and endocannabinoid system tone in obese and diabetic leptin-resistant mice: Impact on apelin regulation in adipose tissue. *Frontiers in Microbiology*, 2, <https://doi.org/10.3389/fmicb.2011.00149>.
- Ghoshal NG, Bal HS (1989) Comparative morphology of the stomach of some laboratory mammals. *Laboratory Animals*, 23, 21–29.
- Glimstedt G (1936) Der Stoffwechsel bakterienfreier Tiere I. *Skandinavisches Archiv Für Physiologie*, 73, 48–62. (in German)
- Grover M, Kashyap PC (2014) Germ-free mice as a model to study effect of gut microbiota on host physiology. *Neurogastroenterology & Motility*, 26, 745–748.
- Han B, Sivaramakrishnan P, Lin CJ, Neve IAA, He J, Tay LWR, Sowa JN, Sizovs A, Du G, Wang J, Herman C, Wang MC (2017) Microbial genetic composition tunes host longevity. *Cell*, 169, 1249–1262.
- Haque MM, Merchant M, Kumar PN, Dutta A, Mande SS (2017) First-trimester vaginal microbiome diversity: A potential indicator of preterm delivery risk. *Scientific Reports*, 7, 16145.
- Harris RA, Alcott CE, Sullivan EL, Takahashi D, McCurdy CE, Comstock S, Baquero K, Blundell P, Frias AE, Kahr M, Suter M, Wesolowski S, Friedman JE, Grove KL, Aagaard KM (2016) Genomic variants associated with resistance to high fat diet induced obesity in a primate model. *Scientific Reports*, 6, 36123.
- Hildebrand F, Ebersbach T, Nielsen HB, Li X, Sonne SB, Bertalan M, Dimitrov P, Madsen L, Qin J, Wang J, Raes J, Kristiansen K, Licht TR (2012) A comparative analysis of the intestinal metagenomes present in Guinea pigs (*Cavia porcellus*) and humans (*Homo sapiens*). *BMC Genomics*, 13, 514.
- Hooper LV, Gordon JI (2001) Commensal host-bacterial relationships in the gut. *Science*, 292, 1115–1118.
- Hosp H (1945) Sulfonamides and vitamin K deficiency. *Nutrition Reviews*, 3, 35–36.
- Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, Kilian B, Quintais LT, Guerra-Assunção JA, Zhou Y, Gu Y, Yen J, Vogel JH, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Cleve C, Oliver K, Clark R, Riddle C, Elliot D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthavadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper J, Tromans A, Grafham D, Sucke C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Urün Z, Eser C, Geiger H, Geisler M, Karotki L, Kirm A, Konantz J, Konantz M, Oberländer M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SM, Enright A, Geisler R, Plasterk RH, Lee C, Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nüsslein-Volhard C, Hubbard TJ, Roest Crollius H, Rogers J, Stemple DL (2013) The

- zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496, 498–503.
- Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, Wei D, Goldfarb KC, Santee CA, Lynch SV, Tanoue T, Imaoka A, Itoh K, Takeda K, Umesaki Y, Honda K, Littman DR (2009) Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell*, 139, 485–498.
- Jemielita M, Taormina MJ, Burns AR, Hampton JS, Rolig AS, Guillemin K, Parthasarathy R (2014) Spatial and temporal features of the growth of a bacterial species colonizing the zebrafish gut. *mBio*, 5, e01751-14.
- Jie Z, Xia H, Zhong S-L, Feng Q, Li S, Liang S, Zhong H, Liu Z, Gao Y, Zhao H, Zhang D, Su Z, Fang Z, Lan Z, Li J, Xiao L, Li J, Li R, Li X, Li F, Ren H, Huang Y, Peng Y, Li G, Wen B, Dong B, Chen JY, Geng QS, Zhang ZW, Yang H, Wang J, Wang J, Zhang X, Madsen L, Brix S, Ning G, Xu X, Liu X, Hou Y, Jia H, He K, Kristiansen K (2017) The gut microbiome in atherosclerotic cardiovascular disease. *Nature Communications*, 8, 845.
- Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI (2011) Human nutrition, the gut microbiome, and immune system: Envisioning the future. *Nature*, 474, 327–336.
- Keller PJ, Schmidt AD, Wittbrodt J, Stelzer EHK (2008) Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science*, 322, 1065–1069.
- Kelly S (2016) Mouse microbes may make scientific studies harder to replicate. *Science*, 353, 741–743.
- Kilian M, Chapple ILC, Hannig M, Marsh PD, Meuric V, Pedersen AML, Tonetti MS, Wade WG, Zaura E (2016) The oral microbiome—An update for oral healthcare professionals. *British Dental Journal*, 221, 657–666.
- Kübeck R, Bonet-Ripoll C, Hoffmann C, Walker A, Müller VM, Schüppel VL, Lagkouvardos I, Scholz B, Engel K-H, Daniel H, Schmitt-Kopplin P, Haller D, Clavel T, Klingenspor M (2016) Dietary fat and gut microbiota interactions determine diet-induced obesity in mice. *Molecular Metabolism*, 5, 1162–1174.
- Lagier J-C, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, Caputo A, Cadoret F, Traore SI, Seck EH, Dubourg G, Durand G, Mourembou G, Guilhot E, Togo A, Bellali S, Bachar D, Cassir N, Bittar F, Delerce J, Mailhe M, Ricaboni D, Bilen M, Dangui Niekou NPM, Dia Badiane NM, Valles C, Mouelhi D, Diop K, Million M, Musso D, Abrahão J, Azhar EI, Bibi F, Yasir M, Diallo A, Sokhna C, Djossou F, Vitton V, Robert C, Rolain JM, La Scola B, Fournier P-E, Levasseur A, Raoult D (2016) Culture of previously uncultured members of the human gut microbiota by culturomics. *Nature Microbiology*, 1, 16203.
- Lagkouvardos I, Pukall R, Abt B, Foessel BU, Meier-Kolthoff JP, Kumar N, Bresciani A, Martínez I, Just S, Ziegler C, Brugiroux S, Garzetti D, Wenning M, Bui TPN, Wang J, Hugenholtz F, Plugge CM, Peterson DA, Hornef MW, Baines JF, Smidt H, Walter J, Kristiansen K, Nielsen HB, Haller D, Overmann J, Stecher B, Clavel T (2016) The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nature Microbiology*, 1, 16131.
- Leftwich PT, Clarke NVE, Hutchings MI, Chapman T (2018) Reply to Rosenberg et al: Diet, gut bacteria, and assortative mating in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, USA*, 115, E2154–E2155.
- Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI (2005) Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences, USA*, 102, 11070–11075.
- Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006) Microbial ecology: Human gut microbes associated with obesity. *Nature*, 444, 1022–1023.
- Li X, Liang S, Xia Z, Qu J, Liu H, Liu C, Yang H, Wang J, Madsen L, Hou Y, Li J, Jia H, Kristiansen K, Xiao L (2018) Establishment of a *Macaca fascicularis* gut microbiome gene catalog and comparison with the human, pig, and mouse gut microbiomes. *GigaScience*, 7, 9.
- Liu R, Hong J, Xu X, Feng Q, Zhang D, Gu Y, Shi J, Zhao S, Liu W, Wang X, Xia H, Liu Z, Cui B, Liang P, Xi L, Jin J, Ying X, Wang X, Zhao X, Li W, Jia H, Lan Z, Li F, Wang R, Sun Y, Yang M, Shen Y, Jie Z, Li J, Chen X, Zhong H, Xie H, Zhang Y, Gu W, Deng X, Shen B, Xu X, Yang H, Xu G, Bi Y, Lai S, Wang J, Qi L, Madsen L, Wang J, Ning G, Kristiansen K, Wang W (2017) Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nature Medicine*, 23, 859–868.
- Lloyd TE, Taylor JP (2010) Flightless flies: *Drosophila* models of neuromuscular disease. *Annals of the New York Academy of Sciences*, 1184, E1–E20.
- Lucas BA, Lavi E, Shiue L, Cho H, Katzman S, Miyoshi K, Siomi MC, Carmel L, Ares M, Maquat LE (2018) Evidence for convergent evolution of SINE-directed Staufen-mediated mRNA decay. *Proceedings of the National Academy of Sciences, USA*, 115, 965–973.
- Marchesi JR, Adams DH, Fava F, Hermes GDA, Hirschfield GM, Hold G, Quraishi MN, Kinross J, Smidt H, Tuohy KM, Thomas LV, Zoetendal EG, Hart A (2016) The gut microbiota and host health: A new clinical frontier. *Gut*, 65, 330–339.
- Menzel R (2012) The honeybee as a model for understanding the basis of cognition. *Nature Reviews Neuroscience*, 13, 758–768.
- Najjar VA (1943) Vitamin synthesis by intestinal bacteria. *Nutrition Reviews*, 1, 175–176.
- Obadia B, Keebaugh ES, Yamada R, Ludington WB, Ja WW (2018) Diet influences host-microbiota associations in *Drosophila*. *Proceedings of the National Academy of Sciences, USA*, 115, E4547–E4548.
- Paigen K (2003) One hundred years of mouse genetics: An



- intellectual history. II. The molecular revolution (1981–2002). *Genetics*, 163, 1227–1235.
- Pan H, Guo R, Zhu J, Wang Q, Ju Y, Xie Y, Zheng Y, Wang Z, Li T, Liu Z, Lu L, Li F, Tong B, Xiao L, Xu X, Li R, Yuan Z, Yang H, Wang J, Kristiansen K, Jia H, Liu L (2018) A gene catalogue of the Sprague-Dawley rat gut metagenome. *GigaScience*, 7, 5.
- Pandey UB, Nichols CD (2011) Human disease models in *Drosophila melanogaster* and the role of the fly in therapeutic drug discovery. *Pharmacological Reviews*, 63, 411–436.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarnier F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, MetaHIT Consortium, Bork P, Ehrlich SD, Wang J (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464, 59–65.
- Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K, Wang J (2012) A metagenome-wide association study of gut microbiota in Type 2 diabetes. *Nature*, 490, 55–60.
- Rabot S, Membrez M, Bruneau A, Gérard P, Harach T, Moser M, Raymond F, Mansourian R, Chou CJ (2010) Germ-free C57BL/6J mice are resistant to high-fat-diet-induced insulin resistance and have altered cholesterol metabolism. *The FASEB Journal*, 24, 4948–4959.
- Rajilic-Stojanovic M, de Vos WM (2014) The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiology Reviews*, 35, 996–1047.
- Ramakrishna BS (2013) Role of the gut microbiota in human nutrition and metabolism. *Journal of Gastroenterology and Hepatology*, 28(Suppl. 4), 9–17.
- Rawls JF, Mahowald MA, Ley RE, Gordon JI (2006) Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell*, 127, 423–433.
- Rawls JF, Samuel BS, Gordon JI (2004) Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. *Proceedings of the National Academy of Sciences, USA*, 101, 4596–4601.
- Reiter LT, Potocki L, Chien S, Gribskov M, Bier E (2001) A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Research*, 11, 1114–1125.
- Roeselers G, Mittge EK, Stephens WZ, Parichy DM, Cavanaugh CM, Guillemin K, Rawls JF (2011) Evidence for a core gut microbiota in the zebrafish. *The ISME Journal*, 5, 1595–1608.
- Rolig AS, Parthasarathy R, Burns AR, Bohannon BJM, Guillemin K (2015) Individual members of the microbiota disproportionately modulate host innate immune responses. *Cell Host & Microbe*, 18, 613–620.
- Round JL, Mazmanian SK (2009) The gut microbiota shapes intestinal immune responses during health and disease. *Nature Reviews Immunology*, 9, 313–323.
- Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, Tuohy K (2018) Gut microbiota functions: Metabolism of nutrients and other food components. *European Journal of Nutrition*, 57, 1–24.
- Sachs DH, Galli C (2009) Genetic manipulation in pigs. *Current Opinion in Organ Transplantation*, 14, 148–153.
- Scott TA, Quintaneiro LM, Norvaisas P, Lui PP, Wilson MP, Leung K-Y, Herrera-Dominguez L, Sudiwala S, Pessia A, Clayton PT, Bryson K, Velagapudi V, Mills PB, Typas A, Greene NDE, Cabreiro F (2017) Host-microbe co-metabolism dictates cancer drug efficacy in *C. elegans*. *Cell*, 169, 442–456.
- Smith DG, Robinson HJ, Clark DM (1945) The influence of streptomycin and streptothricin on the intestinal flora of mice. *Journal of Bacteriology*, 50, 613–621.
- Snel J, Heinen PP, Blok HJ, Carman RJ, Duncan AJ, Allen PC, Collins MD (1995) Comparison of 16S rRNA sequences of segmented filamentous bacteria isolated from mice, rats, and chickens and proposal of “*Candidatus* Arthromitus”. *International Journal of Systematic and Evolutionary Microbiology*, 45, 780–782.
- Sommer F, Bäckhed F (2013) The gut microbiota—Masters of host development and physiology. *Nature Reviews Microbiology*, 11, 227–238.
- Stappenbeck TS, Virgin HW (2016) Accounting for reciprocal host–microbiome interactions in experimental science. *Nature*, 534, 191–199.
- Stephens WZ, Burns AR, Stagaman K, Wong S, Rawls JF, Guillemin K, Bohannon BJM (2016) The composition of the zebrafish intestinal microbial community varies across development. *The ISME Journal*, 10, 644–654.
- Suzuki K, Meek B, Doi Y, Muramatsu M, Chiba T, Honjo T, Fagarasan S (2004) Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proceedings of the National Academy of Sciences, USA*, 101, 1981–1986.
- Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI (2009) The effect of diet on the human gut microbiome: A metagenomic analysis in humanized gnotobiotic mice. *Science Translational Medicine*, 1, 6ra14.

- Valente EM, Salvi S, Ialongo T, Marongiu R, Elia AE, Caputo V, Romito L, Albanese A, Dallapiccola B, Bentivoglio AR (2004) PINK1 mutations are associated with sporadic early-onset parkinsonism. *Annals of Neurology*, 56, 336–341.
- Verbeke KA, Boobis AR, Chiodini A, Edwards CA, Franck A, Kleerebezem M, Nauta A, Raes J, van Tol EA, Tuohy KM (2015) Towards microbial fermentation metabolites as markers for health benefits of prebiotics. *Nutrition Research Reviews*, 28, 42–66.
- Wahlström A, Sayin SI, Marschall H-U, Bäckhed F (2016) Intestinal crosstalk between bile acids and microbiota and its impact on host metabolism. *Cell Metabolism*, 24, 41–50.
- Wallace BD, Wang H, Lane KT, Scott JE, Orans J, Koo JS, Venkatesh M, Jobin C, Yeh L-A, Mani S, Redinbo MR (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science*, 330, 831–835.
- Wang J, Jia H (2016) Metagenome-wide association studies: Fine-mining the microbiome. *Nature Reviews Microbiology*, 14, 508–522.
- Wang S, Huang XF, Zhang P, Newell KA, Wang H, Zheng K, Yu Y (2017) Dietary teasaponin ameliorates alteration of gut microbiota and cognitive decline in diet-induced obese mice. *Scientific Reports*, 7, 12203.
- Wiles TJ, Jemielita M, Baker RP, Schlomann BH, Logan SL, Ganz J, Melancon E, Eisen JS, Guillemin K, Parthasarathy R (2016) Host gut motility promotes competitive exclusion within a model intestinal microbiota. *PLoS Biology*, 14, e1002517.
- Xiao L, Estellé J, Kiilerich P, Ramayo-Caldas Y, Xia Z, Feng Q, Liang S, Pedersen AØ, Kjeldsen NJ, Liu C, Maguin E, Doré J, Pons N, Le Chatelier E, Prifti E, Li J, Jia H, Liu X, Xu X, Ehrlich SD, Madsen L, Kristiansen K, Rogel-Gaillard C, Wang J (2016) A reference gene catalogue of the pig gut microbiome. *Nature Microbiology*, 1, 16161.
- Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, Li X, Long H, Zhang J, Zhang D, Liu C, Fang Z, Chou J, Glanville J, Hao Q, Kotowska D, Colding C, Licht TR, Wu D, Yu J, Sung JY, Liang Q, Li J, Jia H, Lan Z, Tremaroli V, Dworzynski P, Nielsen HB, Bäckhed F, Doré J, Le Chatelier E, Ehrlich SD, Lin JC, Arumugam M, Wang J, Madsen L, Kristiansen K (2015) A catalog of the mouse gut metagenome. *Nature Biotechnology*, 33, 1103–1108.
- Xie H, Guo R, Zhong H, Feng Q, Lan Z, Qin B, Ward KJ, Jackson MA, Xia Y, Chen X, Chen B, Xia H, Xu C, Li F, Xu X, Al-Aama JY, Yang H, Wang J, Kristiansen K, Wang J, Steves CJ, Bell JT, Li J, Spector TD, Jia H (2016) Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Systems*, 21, 572–584.
- Zákostelská Z, Málková J, Klimešová K, Rossmann P, Hornová M, Novosádová I, Stehlíková Z, Kostovčík M, Hudcovic T, Štěpánková R, Jůzlová K, Hercogová J, Tlaskalová-Hogenová H, Kverka M (2016) Intestinal microbiota promotes psoriasis-like skin inflammation by enhancing Th17 response. *PLoS ONE*, 11, e0159539.
- Zevin AS, Hensley-McBain T, Miller C, Smith E, Langevin S, Klatt NR (2017) Antibiotic treatment disrupts bacterial communities in the colon and rectum of simian immunodeficiency virus-infected macaques. *FEMS Microbiology Letters*, 15, 364.

(特邀责任编辑: 周欣 责任编辑: 时意专)



•综述•

# 洞察景观环境影响蜜蜂之新视角：肠道微生物

唐 敏<sup>1,2</sup> 邹 怡<sup>3</sup> 苏秦之<sup>2,4</sup> 周 欣<sup>1,2\*</sup>

1 (中国农业大学植物保护学院, 北京 100193)

2 (中国农业大学北京食品营养与人类健康高精尖创新中心, 北京 100193)

3 (西交利物浦大学健康与环境科学系, 江苏苏州 215123)

4 (中国农业大学食品科学与营养工程学院, 北京 100083)

**摘要:** 作为生态服务提供者的传粉蜜蜂与景观生态息息相关, 而以农田为主的景观组成显著降低了传粉蜜蜂的多样性。目前调查研究显示, 农田的扩张与蜜蜂多样性下降相关, 且农药残留对蜜蜂损害严重。景观中的开花植物决定了蜜蜂的食物(营养)组成, 其中花粉蛋白含量与蜜蜂的生长发育紧密相关。尽管研究已证实景观环境会显著影响蜜蜂蜂群的发展和个体的生长繁殖能力, 但未来还需要加强景观组成变化直接作用于蜜蜂的机制研究。另一方面, 大量研究表明蜜蜂肠道共生菌是影响宿主健康的重要因素: 可促进宿主吸收营养和抵抗病原菌。作为传粉者, 蜜蜂接触到的主要外部环境——花粉和花蜜都含有特殊的微生物, 很多研究暗示花源微生物是蜜蜂肠道菌来源之一。研究表明景观环境相关的食物(营养)、农药残留以及环境微生物都会显著影响肠道微生物。现有少量的研究证明不同景观的蜜蜂肠道微生物有差异, 景观环境可能通过作用于蜜蜂肠道微生物进而影响蜜蜂健康。然而不同景观环境中的微生物, 尤其是花源微生物和蜜蜂肠道菌之间的关联有待证明。景观对蜜蜂肠道微生物的影响值得研究, 希望可以从肠道菌的视角鉴别对蜜蜂友好的景观环境, 进而指导土地合理利用和蜜蜂保护。

**关键词:** 景观; 蜜蜂; 肠道微生物; 农药; 花粉营养; 环境微生物

## A new perspective on landscape impact in bee populations: Considering the bee gut microbiome

Min Tang<sup>1,2</sup>, Yi Zou<sup>3</sup>, Qinzhi Su<sup>2,4</sup>, Xin Zhou<sup>1,2\*</sup>

1 College of Plant Protection, China Agricultural University, Beijing 100193

2 Beijing Advanced Innovation Center for Food Nutrition and Human Health, China Agricultural University, Beijing 100193

3 Department of Health Environmental Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123

4 College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083

**Abstract:** Pollinator bees are providers of an important ecosystem service, and their survival relies completely on the landscape. Now with the landscape dominated by agriculture, bee diversity has been significantly reduced. Studies suggest that bee populations decline as agricultural land-use increases due to increased exposure to detrimental pesticides. Further, the protein content of pollen is highly important for the growth and development of a bee, and different landscapes provide distinct sources of nutrition. Although many studies have demonstrated the apparent impacts of landscape change on the population dynamics and individual survival of the bees, the underpinning mechanisms remain largely unknown. On the other hand, an increasing body of literature has shown that bee gut symbionts are of great importance to the health of the host bees in absorbing nutrients and resisting pathogens. When foraging, pollinator bees are exposed to particular microbes from pollen and nectar which have been suggested to be a source of some bee gut symbionts and could be either probiotics or pathogens. Together with landscape-related nutrition and pesticides, environmental microbes have been reported to affect bee microbiomes significantly. A number of pilot studies suggest that landscape change could affect bee microbiota, thereby influencing host health. An important linkage, however, is missing between environmental microbiota, especially those associated with

收稿日期: 2019-03-12; 接受日期: 2019-05-27

基金项目: 国家自然科学基金(31772493)和科技基础资源调查专项(2018FY100403)

\* 通讯作者 Author for correspondence. E-mail: xinzhoucaddis@icloud.com



the flowers, and that of the bee gut in a changing habitat. It is worth exploring how gut microbiomes respond to landscape changes. This will hopefully help us identify landscape types that are friendly to bees, so proper land-use can be implemented to protect the bees.

**Key words:** landscape; bees; gut microbiome; pesticide; pollen nutrition; environmental microbes

传粉昆虫在生态系统和农业生产中扮演着重要的角色(Klein et al, 2007; Gill et al, 2016), 80%以上的开花植物需要动物, 绝大多数是昆虫帮助授粉(Kremen et al, 2007; Ollerton et al, 2011)。近年来, 由于人类活动的影响, 自然景观生态发生了巨大的变化, 有研究发现土地利用变化是导致生物多样性降低的主要原因之一(Murphy & Romanuk, 2014)。其中农业用地的大面积扩展破坏了原有的自然、半自然生境, 大规模的集约化农业使得生境多样性降低(Ollerton et al, 2014; Ramankutty et al, 2018), 威胁着传粉昆虫及其生态服务功能(Biesmeijer et al, 2006; Burkle et al, 2013; Connelly et al, 2015)。栖息地碎片化直接威胁多种传粉昆虫, 破坏了生态系统中传粉网络的结构(Grass et al, 2018), 其中大部分蜜蜂群体受损明显(Winfrey et al, 2007)。本文讨论的蜜蜂指蜜蜂总科所有蜂种, 它们是主要的膜翅目传粉昆虫, 约有两万种, 其中社会性蜜蜂少于1,000种, 全球分布广泛。

景观中的开花植物种类直接决定了蜜蜂的食物和营养组成。Donkersley等(2014)发现景观生态组成直接影响蜂粮营养组成, 草地和阔叶林占比越高, 蜂粮的蛋白含量越高, 而蛋白质和氨基酸被认为是决定蜜蜂营养的关键因素(Di Pasquale et al, 2016; Glavinic et al, 2017)。也有研究认为景观组成对蜂粮没有决定性影响(Danner et al, 2017), 不过这些研究只关注了景观组成多样性的高低, 并未考虑不同景观生态系统中开花植物的种类和组成特点。景观生态常常被划分为不同类别, 比如农田、森林、草地等, 或者被归类于不同性质, 比如根据多样性的高低分为简单和复杂, 或者根据人为影响程度分为农田、自然和半自然。因此, 同一类性质的景观内植被组成可能非常不同。Colwell等(2017)调查了不同农田系统中西方蜜蜂(*Apis mellifera*)采集的蜂花粉的多样性和营养状况, 发现不同农田系统中蜜蜂的取食多样性和蜂粮蛋白含量均有显著差异。农业生产影响传粉蜜蜂的另一主要因素是农药残留(Potts et al, 2010; Gill et al, 2012; Botías et al, 2015; Rundlöf

et al, 2015; Crall et al, 2018), 它很可能是造成西方蜜蜂蜂群崩溃综合征(colony collapse disorder, CCD)的原因之一(Oldroyd, 2007; Goulson et al, 2015; Steinhauer et al, 2018)。

虽然与景观环境相关的因素明显影响着蜜蜂, 但其作用机制尚不明确。越来越多的研究显示肠道菌群与宿主健康紧密相关, 早在2007年就有研究表明患有崩溃综合征的蜂群其肠道菌群与正常蜂群不一样, 且与一种病毒有密切关系(Cox-Foster et al, 2007)。Ribièrè等(2018)发现健康蜂群和不健康蜂群的肠道微生物有显著差异。肠道有益菌可抵抗疾病, 促进宿主健康, 比如熊蜂肠道中的很多菌株都可以抵抗病原菌(Praet et al, 2018)。相反, 菌群紊乱会降低宿主对病原菌的抵抗力(Schwarz et al, 2016; Li et al, 2017a), 不利于蜜蜂健康(Anderson & Ricigliano, 2017)。因此, 通过肠道微生物的视角可以帮助我们理解景观环境影响蜜蜂健康的机制。

## 1 景观差异与蜜蜂肠道菌群差异相关

目前已有少量研究从景观水平上调查蜜蜂的肠道微生物。Jones等(2018)通过比较16S rRNA V4区序列发现, 景观生态对西方蜜蜂肠道微生物有显著影响, 虽然油菜花田这一景观因素对菌群差异的解释量只有6%, 低于取样点造成的差异(20%), 但如果综合考虑其他景观因素如开花植物的多样性组成、农田和森林等的组成比例、农药残留等, 可能会发现景观环境与肠道菌群演替有更显著的相关性。为了探究同一种东方蜜蜂(*Apis cerana*)在不同景观生态环境中的肠道菌群是否有差异, 我们分别从不同蜂场取工蜂并测序其肠道中16S rRNA V3区片段多样性(未发表数据)。4个蜂场分别位于四川关坝(SCGB)、四川小河沟自然保护区(SCXHG)、陕西杉树坪(SXSSP)和陕西杨家沟(SXYJG)(海拔分别为1,333 m、1,410 m、1,484 m和1,348 m)。关坝与小河沟地区的植被组成相似, 后者划入自然保护区时间长, 自然修复更好; 关坝和杨家沟蜂场附近有村庄和农田, 而且据当地调查, 农田有施农药的情况;

杉树坪和杨家沟同属秦岭自然保护区,前者自然修复更好,附近无人居住。所有样品工蜂属于同一东方蜜蜂群体(Chen et al, 2018),四川与陕西的工蜂肠道菌群并无显著差异(图1a),但来自同一省份的工蜂之间肠道菌群差异显著小于来自不同省份的(图1b),这很可能与植被组成不同有关,需要后期更多的数据验证。关坝与杨家沟采样点附近均有村庄,而小河沟与杉树坪采样点的自然修复程度更高,因此我们预期,关坝与杨家沟所采蜜蜂的肠道菌群组成差异应该更小,而小河沟与杨家沟所采蜜蜂的肠道菌群组成差异更大。但结果并不支持该假设,暗示了研究土地利用对肠道菌群影响的实验设计应该限定在更小的范围内,例如未来可以在同一省份植被组成相同或相似的地方设点采样。未来的工作中我们将验证是否是景观环境造成了这些差异,并研究其中哪些景观因素起决定性作用,其作用机制如何。

## 2 花粉和农药对蜜蜂肠道菌群的影响

花粉作为蜜蜂的主要食物来源,对蜜蜂健康非常重要(Frias et al, 2016),营养压力也被认为是导致CCD的原因之一(Naug, 2009; Potts et al, 2010; Goulson et al, 2015)。蜜蜂面临的另一主要生存压力是农药残留。花粉营养物质和农药残留对蜜蜂的影响

以及机制有待深入研究。

### 2.1 花粉营养显著影响蜜蜂及其肠道微生物

我们可以通过了解景观中蜜蜂获取的食物对其肠道菌及其代谢物的影响来理解不同景观对蜜蜂的影响。DeGrandi-Hoffman等(2016)研究发现相比人工配方喂养,天然饲养的西方蜜蜂更少染病,而且越冬存活率更高。Di Pasquale等(2016)的研究结果显示,食物中花粉比例越高,西方蜜蜂存活率越高。有研究探究了花粉的营养成分和多样性对蜜蜂健康的影响。Di Pasquale等(2013)给西方蜜蜂哺育蜂喂食单种或等量混合的4种不同花粉(对照组只喂食糖水),并监测蜜蜂生理特征及其对病原菌 *Nosema ceranae* 的抵抗力,发现不同的花粉对哺育蜂的发育和对病原菌抵抗力的影响存在显著差异:氨基酸含量高的花粉更有利于哺育蜂抵抗病原菌,而混合花粉未表现出优势。但在被 *N. ceranae* 感染后,喂食混合花粉或蛋白质含量最高的花粉的蜜蜂存活期更长。除蛋白质外,花粉还提供蜜蜂所需的脂类和微量元素,花蜜提供糖类物质,因此景观中的开花植物种类直接影响蜜蜂健康(Vaudo et al, 2015)。

大量研究表明食物(营养)是影响肠道微生物的重要因素(Muegge et al, 2011; Wu et al, 2011; David et al, 2014; Li et al, 2017b)。蜜蜂肠道菌群在宿主消

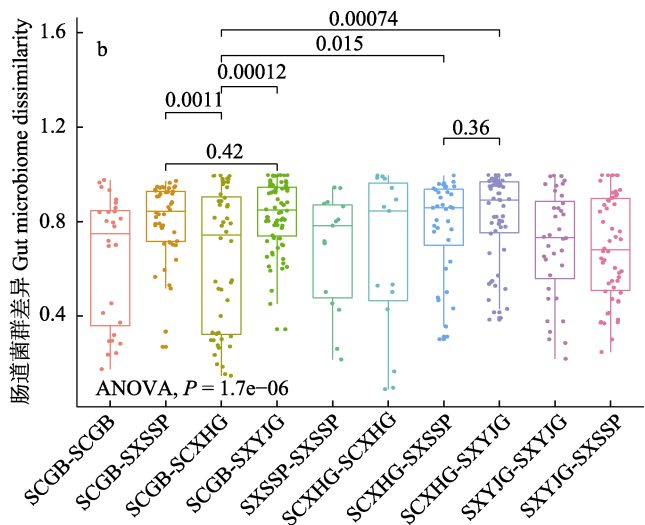
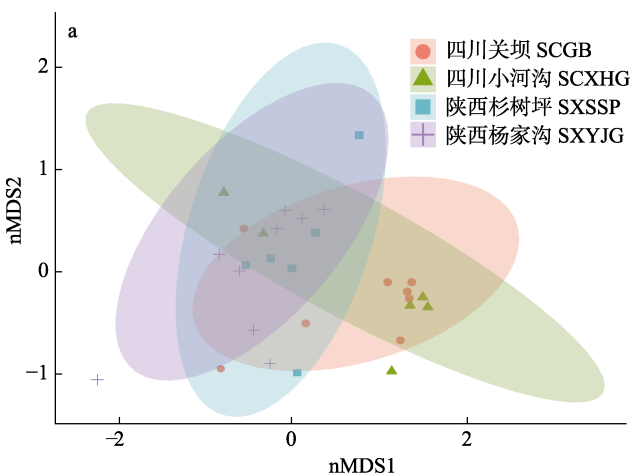


图1 来自4个蜂场的东方蜜蜂肠道菌群16S rRNA V3片段序列的非度量多维尺度分析(nMDS)图(a)及蜂场内和蜂场间的Bray-Curtis距离比较(ANOVA和t检验, b)

Fig. 1 Gut microbiome similarity on 16S rRNA V3 region of *Apis cerana* sampled from 4 apiaries showed by non-metric multidimensional scaling (nMDS) plot (a), and Bray-Curtis distance between gut microbiome compared among and between sampling sites and tested by using ANOVA and *t*-test (b). SCGB, Guanba, Sichuan; SCXHG, Xiaohogou, Sichuan; SXSSP, Shanshuping, Shaanxi; SXYJG, Yangjiagou, Shaanxi.

化食物、吸收营养的过程中扮演着关键的角色(Engel & Bonilla-Rosso, 2018)。例如西方蜜蜂肠道菌可以消化很多种类的粉源营养物质,更重要的是可以降解难以利用的植物次级代谢产物和花粉壁(Kešnerová et al, 2017),其肠道菌*Gilliamella apicola*的一些菌株还可以降解对蜜蜂有毒的糖类(Zheng et al, 2016)。食物可以通过影响肠道微生物进而作用于宿主。Maes等(2016)发现喂食陈旧的花粉可显著影响西方蜜蜂肠道微生物组成,并导致死亡率升高和生长发育缺陷,而且喂食花粉替代品会加重这些不良影响。

## 2.2 农药残留严重损害蜜蜂并显著改变蜜蜂肠道微生物

Pettis等(2013)从西方蜜蜂蜂粮中检测出35种残留农药,发现农药可削弱西方蜜蜂对病原菌*Nosema ceranae*的抵抗力,而且Kessler等(2015)研究发现西方蜜蜂不会趋避新烟碱类农药。大量研究表明:个体水平上,农药显著降低工蜂存活率(Henry et al, 2012; Tsvetkov et al, 2017)及蜂王的繁殖力和新蜂王产量(Whitehorn et al, 2012; Tsvetkov et al, 2017; Siviter et al, 2018);蜂群水平上,新烟碱农药被确证对西方蜜蜂、熊蜂和野生蜂群不利:蜂群增长量显著降低,后代明显减少,蜂王产出率也降低了85%(Whitehorn et al, 2012; Woodcock et al, 2017);还有一些研究发现农药在免疫(Tsvetkov et al, 2017)、发育(dos Santos et al, 2016)和行为(Henry et al, 2012)水平上对蜜蜂造成损伤。

另一方面,人们也意识到自然栖息地生态组成对传粉昆虫多样性的重要性。在农作系统中,人为增加生态多样性有利于缓解传粉生态功能缺失的状况(Blaauw & Isaacs, 2014a, b; Kovács-Hostyánszki et al, 2017)。在Colwell等(2017)研究的不同农作系统中,农药残留与开花植物的多样性呈负相关。生态强化过程中,可能是农药被“稀释”了(Park et al, 2015; Colwell et al, 2017),也可能是蜜蜂因为景观环境的改善变得更健康了。

已有少量研究从肠道微生物的角度解析农药影响蜜蜂的作用机制。Kakumanu等(2016)实验发现暴露于蝇毒磷、氟胺氰菊酯和百菌清的西方蜜蜂的肠道微生物发生显著改变。Yang等(2019)发现不同农药对东方蜜蜂和西方蜜蜂的中肠菌群的影响不同,科学家正致力于从机理层面解释这种影响。比

如,除草剂草甘膦的作用靶标物为烯醇式丙酮基莽草-3-磷酸合成酶,该酶存在于一些植物和微生物中,也存在于几乎所有已测序的西方蜜蜂肠道菌中。Motta等(2018)发现蜜蜂肠道菌编码的这种酶有两种类型,一种对草甘膦敏感,另一种不敏感,另外还有些菌株可以通过未知的方式抵抗草甘膦。这些基础研究可以帮助我们理解农药作用于蜜蜂和蜜蜂抵抗农药的机制。

## 3 景观花源微生物对蜜蜂肠道菌群的影响

除了食物和农药,景观环境中还有另外一个重要因素可扰动蜜蜂肠道微生物,即花源微生物。Lim等(2015)将景观看作一个整体变量,发现多种熊蜂的肠道微生物组成都与之相关,因此他们认为至少部分肠道菌是可以通过环境传播的。还有研究表明景观影响蜂巢和蜂粮微生物。Donkersley等(2018)发现土地利用会影响蜂粮菌群的组成,通过将多种景观进行比较,发现一种改善后的草地系统中的蜂粮菌群多样性最低,而蜂粮微生物可直接作用于蜜蜂肠道微生物(Powell et al, 2014)。景观生态的变化会显著改变花粉和花蜜资源(Baude et al, 2016),与之相应的花源微生物不可忽视。

### 3.1 花源微生物可能是肠道微生物的来源之一

荀二娜等(2017)总结了花蜜微生物的相关研究,发现每种花蜜所含细菌虽然平均只有1.4种,但发生率较高。这些细菌可以适应花蜜中的高糖环境,属于一类独特的微生物,一般为变形菌门,而且通常具有花蜜植物特异性(Fridman et al, 2012)。目前研究较多的西方蜜蜂肠道的共生菌种类不多,其中很多都属于变形菌(Martinson et al, 2011)。Anderson等(2013)培养和检测出传粉环境中的花蜜与西方蜜蜂肠道共有的多种细菌。目前蜜蜂肠道菌是否来源于环境还没有定论。研究较多的社会性蜜蜂——西方蜜蜂、东方蜜蜂和熊蜂拥有简单而稳定的肠道菌群(Martinson et al, 2011),而且与宿主健康息息相关(Koch & Schmid-Hempel, 2011; Anderson & Ricigliano, 2017; Raymann & Moran, 2018),其核心菌(Martinson et al, 2011; Kwong et al, 2017)未在其他环境中被发现,社会性行为很可能有助于这些蜜蜂肠道菌群的传播和稳定(Kwong et al, 2017)。不过,绝大多数研究都是采用16S rRNA扩增子片段作为分析手段,其中核心菌占比高达90%以上,非核心



菌容易被忽略。Saraiva等(2015)发现蜂花粉中微生物与西方蜜蜂肠道微生物组成存在显著差异,但依然共享了7%的细菌物种。从Anderson等(2014)研究西方蜜蜂蜂花粉的微生物结果可以看出,西方蜜蜂肠道的常见优势菌可以较长时间(至少3 d)存活于花粉环境,可满足菌群水平传播的必要条件。因此,这类拥有稳定菌群的社会性蜜蜂有可能获得并保留环境微生物,这些环境菌对蜜蜂及其肠道菌群的作用有待研究。

野生蜜蜂的肠道菌群与东方蜜蜂、西方蜜蜂和熊蜂显著不同,也与花粉明显相关,但共同点是它们都“招募”了对自身有益的细菌(Graystock et al, 2017)。McFrederick等(2012, 2014, 2017)发现多种野生蜂的肠道与花共享细菌,他们认为蜜蜂很可能从花粉和花蜜中获得了微生物。McFrederick和Rehan(2016)分析了一种木蜂巢内花粉和细菌的多样性,虽然总体上没有发现花粉多样性与细菌多样性的相关性,但其中某些花粉与对应的优势细菌显著相关,这提示花粉可能是细菌的来源。绝大多数野生传粉蜜蜂是独居蜂,没有社会性行为帮助它们保持稳定的肠道微生物,尽管对其肠道菌群的研究较少,但推测环境微生物可能对其有更显著的影响。

### 3.2 花源微生物对不同蜜蜂肠道菌的影响可能不同

Zou等(2017)研究了农田占比10%–70%的不同景观系统对传粉昆虫多样性的影响,发现农田比例越高,野生蜂的多样性越低,其中独居蜂受影响显著,而西方蜜蜂的丰度并未受到显著的影响。另外有研究表明,不同的野生蜂多样性与植被种类有关(Rollin et al, 2013; Wu et al, 2018),说明不同的蜜蜂受景观的影响不同,野生独居蜂比社会性蜜蜂更容易受到景观生态组成变化的影响(Neokosmidis et al, 2018)。尽管人工养殖的蜜蜂对农作物生产和自然生态系统传粉功能贡献显著(Hung et al, 2018; Saunders et al, 2018),CCD也让西方蜜蜂备受关注,但传粉者的多样性而非数量决定了传粉功能的水平(Winfrey et al, 2015; Blitzer et al, 2016; Kremen, 2018),且野生蜂对传粉网络和传粉服务的贡献不亚于饲养蜜蜂(Brittain et al, 2012; Holzschuh et al, 2012; Garibaldi et al, 2014)。过多地饲养蜜蜂可能会破坏自然的传粉网络结构并威胁野生传粉者(Torné-noguera et al, 2016, Magrach et al, 2017;

Geldmann & Gonzalez-Varo, 2018)。有研究表明野生熊蜂受制于养殖蜜蜂,但这种影响会因为景观生态的多样性升高而有所改善(Herbertsson et al, 2016)。由此可见景观环境对不同蜜蜂的发展十分重要。

鉴于各种蜜蜂肠道菌群的差异,研究景观环境对其的影响需要进行大量工作,蜜蜂肠道菌联盟(BeeBiome Consortium)也强调了这一研究方向的重要性(Engel et al, 2016)。景观生态的改变意味着环境微生物“菌池”的变化,这是一把双刃剑:既可能提供有益菌(Graystock et al, 2017; McFrederick et al, 2017),又可能隐藏致病菌(Graystock et al, 2015; Rothman et al, 2018)。最近发表的一项研究(Keller et al, 2018)表明,从巢穴环境分离培养出来的普遍存在于一种独居蜂体表、肠道和巢穴的一株细菌可有效抑制致病真菌和细菌。因此探明环境微生物影响肠道微生物进而作用于蜜蜂的机制和规律能为我们保护蜜蜂提供新的视角,改变蜜蜂肠道微生物可能是一种有效促进蜜蜂健康的干预方法(Crotti et al, 2013)。

综上所述,景观环境所决定的蜜蜂食物组成、花源微生物和农药都可能显著影响蜜蜂,其影响机制可以从肠道微生物的角度探究(图2)。在以上所述的景观环境差异影响蜜蜂的研究基础上,需要加入不同处理组与对照组的菌群比较分析,从差异显著的菌种、菌株、基因及其调控和代谢产物等方面验

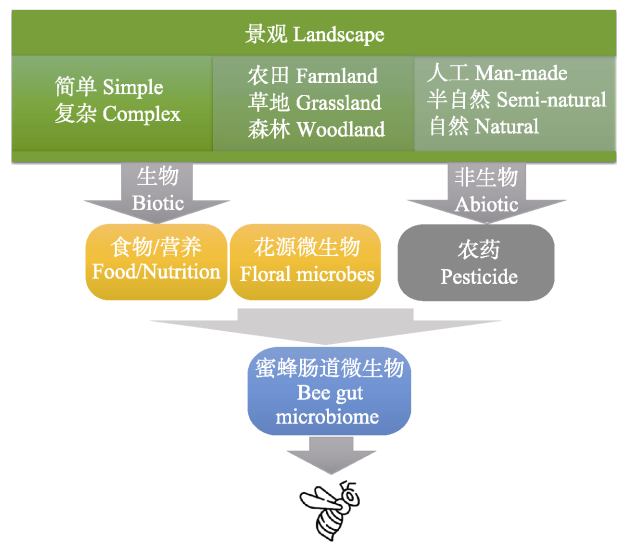


图2 从肠道微生物角度探究景观环境可能影响蜜蜂的主要途径示意图

Fig. 2 The main possible paths through which the landscape could affect bees by changing their microbiome

证景观变化扰动蜜蜂肠道微生物, 进而影响蜜蜂健康和蜂群发展的假设, 有望用蜜蜂肠道菌群指示景观组成是否生态友好。近期研究发现, 自然修复工作的联通栖息地和强化生态组成均有利于物种多样性的升高和传粉网络的稳固(Blaauw & Isaacs, 2014a, b; Aizen et al, 2016; Kovács-Hostyánszki et al, 2017; Kleijn et al, 2018), 通过确定其中蜜蜂肠道微生物的作用, 有望更合理高效地实施自然生态修复工作。

**致谢:** 感谢陕西省洋县长青林业局向定乾和四川省平武县关坝村李芯锐帮助收集工蜂样品, 感谢首都师范大学顾垒老师对植物多样性调查工作的指导。

### 参考文献

- Aizen MA, Gleiser G, Sabatino M, Gilarranz LJ, Bascompte J, Verdú M (2016) The phylogenetic structure of plant-pollinator networks increases with habitat size and isolation. *Ecology Letters*, 19, 29–36.
- Anderson KE, Riciigliano VA (2017) Honey bee gut dysbiosis: A novel context of disease ecology. *Current Opinion in Insect Science*, 22, 125–132.
- Anderson KE, Carroll MJ, Sheehan T, Mott BM, Maes P, Corby-Harris V (2014) Hive-stored pollen of honey bees: Many lines of evidence are consistent with pollen preservation, not nutrient conversion. *Molecular Ecology*, 23, 5904–5917.
- Anderson KE, Sheehan TH, Mott BM, Maes P, Snyder L, Schwan MR, Walton A, Jones BM, Corby-Harris V (2013) Microbial ecology of the hive and pollination landscape: Bacterial associates from floral nectar, the alimentary tract and stored food of honey bees (*Apis mellifera*). *PLoS ONE*, 8, e83125.
- Baude M, Kunin WE, Boatman ND, Conyers S, Davies N, Gillespie MAK, Morton RD, Smart SM, Memmott J (2016) Historical nectar assessment reveals the fall and rise of floral resources in Britain. *Nature*, 530, 85–88.
- Biesmeijer JC, Roberts SPM, Reemer M, Ohlemüller R, Edwards M, Peeters T, Schaffers AP, Potts SG, Kleukers R, Thomas CD, Settele J, Kunin WE (2006) Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science*, 313, 351–355.
- Blaauw BR, Isaacs R (2014a) Flower plantings increase wild bee abundance and the pollination services provided to a pollination-dependent crop. *Journal of Applied Ecology*, 51, 890–898.
- Blaauw BR, Isaacs R (2014b) Larger patches of diverse floral resources increase insect pollinator density, diversity, and their pollination of native wild flowers. *Basic and Applied Ecology*, 15, 701–711.
- Blitzer EJ, Gibbs J, Park MG, Danforth BN (2016) Pollination services for apple are dependent on diverse wild bee communities. *Agriculture, Ecosystems and Environment*, 221, 1–7.
- Botías C, David A, Horwood J, Abdul-Sada A, Nicholls E, Hill E, Goulson D (2015) Neonicotinoid residues in wildflowers, a potential route of chronic exposure for bees. *Environmental Science Technology*, 49, 12731–12740.
- Brittain C, Williams N, Kremen C, Klein AM (2012) Synergistic effects of non-*Apis* bees and honey bees for pollination services. *Proceedings of the Royal Society B: Biological Sciences*, 280, 20122767.
- Burkle LA, Marlin JC, Knight TM (2013) Plant-pollinator interactions over 120 years: Loss of species, co-occurrence, and function. *Science*, 339, 1611–1616.
- Chen C, Wang H, Liu Z, Chen X, Tang J, Meng F, Shi W (2018) Population genomics provide insights into the evolution and adaptation of the eastern honey bee (*Apis cerana*). *Molecular Biology and Evolution*, 35, 2260–2271.
- Colwell MJ, Williams GR, Evans RC, Shutler D (2017) Honey bee-collected pollen in agro-ecosystems reveals diet diversity, diet quality, and pesticide exposure. *Ecology and Evolution*, 7, 7243–7253.
- Connelly H, Poveda K, Loeb G (2015) Landscape simplification decreases wild bee pollination services to strawberry. *Agriculture, Ecosystems and Environment*, 211, 51–56.
- Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan PL, Briese T, Horning M, Geiser DM, Martinson V, vanEngelsdorp D, Kalkstein AL, Drysdale A, Hui J, Zhai J, Cui L, Hutchison SK, Simons JF, Egholm M, Pettis JS, Lipkin WI (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*, 318, 283–288.
- Crall JD, Switzer CM, Oppenheimer RL, Ford Versypt AN, Dey B, Brown A, Eyster M, Guerin C, Pierce NE, Combes SA, de Bivort BL (2018) Neonicotinoid exposure disrupts bumblebee nest behavior, social networks, and thermoregulation. *Science*, 362, 683–686.
- Crotti E, Sansonno L, Prosdociemi EM, Vacchini V, Hamdi C, Cherif A, Conella E, Marzorati M, Balloi A (2013) Microbial symbionts of honeybees: A promising tool to improve honeybee health. *New Biotechnology*, 30, 716–722.
- Danner N, Keller A, Härtel S, Steffan-Dewenter I (2017) Honey bee foraging ecology: Season but not landscape diversity shapes the amount and diversity of collected pollen. *PLoS ONE*, 12, e0183716.
- David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ (2014) Diet rapidly and reproducibly alters the human gut microbiome.

- Nature, 505, 559–563.
- DeGrandi-Hoffman G, Chen Y, Rivera R, Carroll M, Chambers M, Hidalgo G, de Jong EW (2016) Honey bee colonies provided with natural forage have lower pathogen loads and higher overwinter survival than those fed protein supplements. *Apidologie*, 47, 186–196.
- Di Pasquale G, Alaux C, Le Conte Y, Odoux JF, Pioz M, Vaissière BE, Belzunces LP, Decourtye A (2016) Variations in the availability of pollen resources affect honey bee health. *PLoS ONE*, 11, e0162818.
- Di Pasquale G, Salignon M, Le Conte Y, Belzunces LP, Decourtye A, Kretzschmar A, Suchail S, Brunet JL, Alaux C (2013) Influence of pollen nutrition on honey bee health: Do pollen quality and diversity matter? *PLoS ONE*, 8, e72016.
- Donkersley P, Rhodes G, Pickup RW, Jones KC (2018) Bacterial communities associated with honeybee food stores are correlated with land use. *Ecology and Evolution*, 8, 4743–4756.
- Donkersley P, Rhodes G, Pickup RW, Jones KC, Wilson K (2014) Honeybee nutrition is linked to landscape composition. *Ecology and Evolution*, 4, 4195–4206.
- dos Santos CF, Acosta AL, Dorneles AL, dos Santos PDS, Blochtein B (2016) Queens become workers: Pesticides alter caste differentiation in bees. *Scientific Reports*, 6, 31605.
- Engel P, Bonilla-Rosso G (2018) Functional roles and metabolic niches in the honey bee gut microbiota. *Current Opinion in Microbiology*, 43, 69–76.
- Engel P, Kwong WK, McFrederick Q, Anderson KE, Barribeau SM, Chandler JA, Cornman RS, Dainat J, de Miranda JR, Doublet V, Emery O, Evans JD, Farinelli L, Flenniken ML, Granberg F, Grasis JA, Gauthier L, Hayer J, Koch H, Kocher S, Martinson VG, Moran N, Munoz-Torres M, Newton I, Paxton RJ, Powell E, Sadd BM, Schmid-Hempel P, Schmid-Hempel R, Song SJ, Schwarz RS, vanEngelsdorp D, Dainat B (2016) The bee microbiome: Impact on bee health and model for evolution and ecology of host-microbe interactions. *mBio*, 7, e02164-15.
- Frias BED, Barbosa CD, Lourenço AP (2016) Pollen nutrition in honey bees (*Apis mellifera*): Impact on adult health. *Apidologie*, 47, 15–25.
- Fridman S, Izhaki I, Gerchman Y, Halpern M (2012) Bacterial communities in floral nectar. *Environmental Microbiology Reports*, 4, 97–104.
- Garibaldi LA, Steffan-dewenter I, Winfree R, Aizen MA, Bommarco R, Cunningham SA, Kremen C, Carvalheiro LG, Harder LD, Afik O, Bartomeus I, Benjamin F, Boreux V, Cariveau D, Chacoff NP, Dudenhöffer JH, Freitas BM, Ghazoul J, Greenleaf S, Hipólito J, Holzschuh A, Howlett B, Isaacs R, Javorek SK, Kennedy CM, Krewenka KM, Krishnan S, Mandelik Y, Mayfield MM, Motzke I, Munyuli T, Nault BA, Otieno M, Petersen J, Pisanty G, Potts SG, Rader R, Ricketts TH, Rundlöf M, Seymour CL, Schüepp C, Szentgyörgyi H, Taki H, Tscharntke T, Vergara CH, Viana BF, Wanger TC, Westphal C, Williams N, Klein AM (2014) Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *Science*, 339, 1608–1611.
- Geldmann J, Gonzalez-Varo JP (2018) Conserving honey bees does not help wildlife: High densities of managed honey bees can harm populations of wild pollinators. *Science*, 359, 392–393.
- Gill RJ, Baldock KCR, Brown MJF, Cresswell JE, Dicks LV, Fountain MT, Garratt MPD, Gough LA, Heard MS, Holland JM, Ollerton J, Stone GN, Tang CQ, Vanbergen AJ, Vogler A, Woodward G, Arce AN, Boatman ND, Brand-Hardy R, Breeze TD, Green M, Hartfield CM, O'Connor RS, Osborne JL, Phillips J, Sutton PB, Potts SG (2016) Protecting an ecosystem service: Approaches to understanding and mitigating threats to wild insect pollinators. *Advances in Ecological Research*, 54, 135–206.
- Gill RJ, Ramos-Rodriguez O, Raine NE (2012) Combined pesticide exposure severely affects individual- and colony-level traits in bees. *Nature*, 491, 105–108.
- Glavinic U, Stankovic B, Draskovic V, Stevanovic J, Petrovic T, Lakic N, Stanimirovic Z (2017) Dietary amino acid and vitamin complex protects honey bee from immunosuppression caused by *Nosema ceranae*. *PLoS ONE*, 12, e0187726.
- Goulson D, Nicholls E, Botías C, Rotheray EL (2015) Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science*, 347, 1255957.
- Grass I, Jauker B, Steffan-Dewenter I, Tscharnkte T, Jauker F (2018) Past and potential future effects of habitat fragmentation on structure and stability of plant-pollinator and host-parasitoid networks. *Nature Ecology & Evolution*, 2, 1408–1417.
- Graystock P, Goulson D, Hughes WOH (2015) Parasites in bloom: Flowers aid dispersal and transmission of pollinator parasites within and between bee species. *Proceedings of the Royal Society B: Biological Sciences*, 282, <https://doi.org/10.1098/rspb.2015.1371>.
- Graystock P, Rehan SM, McFrederick QS (2017) Hunting for healthy microbiomes: Determining the core microbiomes of *Ceratina*, *Megalopta*, and *Apis* bees and how they associate with microbes in bee collected pollen. *Conservation Genetics*, 18, 701–711.
- Henry M, Béguin M, Requier F, Rollin O, Odoux J, Aupinel P, Tchamitchian S, Decourtye A (2012) A common pesticide decreases foraging success and survival in honey bees. *Science*, 336, 348–350.
- Herbertsson L, Lindström SAM, Rundlöf M, Bommarco R, Smith HG (2016) Competition between managed honeybees and wild bumblebees depends on landscape context. *Basic and Applied Ecology*, 17, 609–616.
- Holzschuh A, Dudenhöffer J, Tscharnkte T (2012) Landscapes with wild bee habitats enhance pollination, fruit set and yield of sweet cherry. *Biological Conservation*, 153, 101–107.



- Hung K LJ, Kingston JM, Albretch M, Holway DA, Kohn JR (2018) The worldwide importance of honey bees as pollinators in natural habitats. *Proceedings of the Royal Society B: Biological Sciences*, 285, 20172140.
- Jones JC, Fruciano C, Hildebrand F, Al Toufalilia H, Balfour NJ, Bork P, Engel P, Ratnieks FL, Hughes WO (2018) Gut microbiota composition is associated with environmental landscape in honey bees. *Ecology and Evolution*, 8, 441–451.
- Kakumanu ML, Reeves AM, Anderson TD, Rodrigues RR, Williams MA (2016) Honey bee gut microbiome is altered by in-hive pesticide exposures. *Frontiers in Microbiology*, 7, 1–11.
- Keller A, Brandel A, Becker MC, Balles R, Abdelmohsen UR, Ankenbrand MJ, Sickel W (2018) Wild bees and their nests host *Paenibacillus* bacteria with functional potential of avail. *Microbiome*, 6, 1–10.
- Kešnerová L, Mars RAT, Ellegaard KM, Troilo M, Sauer U, Engel P (2017) Disentangling metabolic functions of bacteria in the honey bee gut. *PLoS Biology*, 15, e2003467.
- Kessler SC, Tiedeken EJ, Simcock KL, Derveau S, Mitchell J, Softley S, Radcliffe A, Stout JC, Wright GA (2015) Bees prefer foods containing neonicotinoid pesticides. *Nature*, 521, 74–76.
- Kleijn D, Bommarco R, Fijen TPM, Garibaldi LA, Potts SG, van der Putten WH (2018) Ecological intensification: Bridging the gap between science and practice. *Trends in Ecology and Evolution*, 34, 154–166.
- Klein A, Vaissie BE, Steffan-dewenter I, Cunningham SA, Kremen C, Cedex A (2007) Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences*, 274, 303–313.
- Koch H, Schmid-Hempel P (2011) Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proceedings of the National Academy of Sciences, USA*, 108, 19288–19292.
- Kovács-Hostyánszki A, Espíndola A, Vanbergen AJ, Settele J, Kremen C, Dicks LV (2017) Ecological intensification to mitigate impacts of conventional intensive land use on pollinators and pollination. *Ecology Letters*, 20, 673–689.
- Kremen C (2018) The value of pollinator species diversity. *Science*, 359, 741–742.
- Kremen C, Williams NM, Aizen MA, Gemmill-Herren B, Lebn G, Minckley R, Packer L, Potts SG, Roulston T, Steffan-Dewenter I, Vázquez DP, Winfree R, Adams L, Crone EE, Greenleaf SS, Keitt TH, Klein AM, Regetz J, Ricketts TH (2007) Pollination and other ecosystem services produced by mobile organisms: A conceptual framework for the effects of land-use change. *Ecology Letters*, 10, 299–314.
- Kwong WK, Medina LA, Koch H, Sing KW, Soh EJY, Ascher JS, Jaffe R, Moran NA (2017) Dynamic microbiome evolution in social bees. *Science Advances*, 3, e1600513.
- Li JH, Evans JD, Li WF, Zhao YZ, Degrandi-hoffman G, Huang SK, Li ZG, Hamilton M, Chen YP (2017a) New evidence showing that the destruction of gut bacteria by antibiotic treatment could increase the honey bee's vulnerability to *Nosema* infection. *PLoS ONE*, 12, e0187505.
- Li Q, Lauber CL, Czarnecki-Maulden G, Pan Y (2017b) Effects of the dietary protein and carbohydrate ratio on gut microbiomes in dogs of different body conditions. *mBio*, 8, e01703-16.
- Lim HC, Chu CC, Seufferheld MJ, Cameron SA (2015) Deep sequencing and ecological characterization of gut microbial communities of diverse bumble bee species. *PLoS ONE*, 10, e0118566.
- Maes PW, Rodrigues PAP, Oliver R, Mott BM, Anderson KE (2016) Diet-related gut bacterial dysbiosis correlates with impaired development, increased mortality and *Nosema* disease in the honeybee (*Apis mellifera*). *Molecular Ecology*, 25, 5439–5450.
- Magrach A, González-Varo JP, Boiffier M, Vilà M, Bartomeus I (2017) Honeybee spillover reshuffles pollinator diets and affects plant reproductive success. *Nature Ecology & Evolution*, 1, 1299–1307.
- Martinson VG, Danforth BN, Minckley RL, Rueppell O, Tingek S, Moran NA (2011) A simple and distinctive microbiota associated with honey bees and bumble bees. *Molecular Ecology*, 20, 619–628.
- McFrederick QS, Rehan SM (2016) Characterization of pollen and bacterial community composition in brood provisions of a small carpenter bee. *Molecular Ecology*, 25, 2302–2311.
- McFrederick QS, Thomas JM, Neff JL, Vuong HQ, Russell KA, Hale AR, Mueller UG (2017) Flowers and wild megachilid bees share microbes. *Microbial Ecology*, 73, 188–200.
- McFrederick QS, Wcislo WT, Taylor DR, Ishak HD, Dowd SE, Mueller UG (2012) Environment or kin: Whence do bees obtain acidophilic bacteria? *Molecular Ecology*, 21, 1754–1768.
- McFrederick QS, Wcislo WT, Hout MC, Mueller UG (2014) Host species and developmental stage, but not host social structure, affects bacterial community structure in socially polymorphic bees. *FEMS Microbiology Ecology*, 88, 398–406.
- Motta EVS, Raymann K, Moran NA (2018) Glyphosate perturbs the gut microbiota of honey bees. *Proceedings of the National Academy of Sciences, USA*, 115, 10305–10310.
- Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, Henrissat B, Knight R, Gordon JI (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science*, 332, 970–974.
- Murphy GEP, Romanuk TN (2014) A meta-analysis of declines in local species richness from human disturbances. *Ecology and Evolution*, 4, 91–103.
- Naug D (2009) Nutritional stress due to habitat loss may

- explain recent honeybee colony collapses. *Biological Conservation*, 142, 2369–2372.
- Neokosmidis L, Tscheulin T, Devallez J, Petanidou T (2018) Landscape spatial configuration is a key driver of wild bee demographics. *Insect Science*, 25, 172–182.
- Oldroyd BP (2007) What's killing American honey bees? *PLoS Biology*, 5, e168.
- Ollerton J, Erenler H, Edwards M, Crockett R (2014) Extinctions of aculeate pollinators in Britain and the role of large-scale agricultural changes. *Science*, 346, 1360–1362.
- Ollerton J, Winfree R, Tarrant S (2011) How many flowering plants are pollinated by animals? *Oikos*, 120, 321–326.
- Park MG, Blitzer EJ, Gibbs J, Losey JE, Danforth BN (2015) Negative effects of pesticides on wild bee communities can be buffered by landscape context. *Proceedings of the Royal Society B: Biological Sciences*, 282, 20150299.
- Pettis JS, Lichtenberg EM, Andree M, Stitzinger J, Rose R, vanEngelsdorp D (2013) Crop pollination exposes honey bees to pesticides which alters their susceptibility to the gut pathogen *Nosema ceranae*. *PLoS ONE*, 8, e70182.
- Potts SG, Biesmeijer JC, Kremen C, Neumann P, Schweiger O, Kunin WE (2010) Global pollinator declines: Trends, impacts and drivers. *Trends in Ecology and Evolution*, 25, 345–353.
- Powell JE, Martinson VG, Urban-Mead K, Moran NA (2014) Routes of acquisition of the gut microbiota of the honey bee *Apis mellifera*. *Applied and Environmental Microbiology*, 80, 7378–7387.
- Praet J, Parmentier A, Schmid-Hempel R, Meeus I, Smagghe G, Vandamme P (2018) Large-scale cultivation of the bumblebee gut microbiota reveals an underestimated bacterial species diversity capable of pathogen inhibition. *Environmental Microbiology*, 20, 214–227.
- Ramankutty N, Mehrabi Z, Waha K, Jarvis L, Kremen C, Herrero M, Rieseberg LH (2018) Trends in global agricultural land use: Implications for environmental health and food security. *Annual Review of Plant Biology*, 69, 789–815.
- Raymann K, Moran NA (2018) The role of the gut microbiome in health and disease of adult honey bee workers. *Current Opinion in Insect Science*, 26, 97–104.
- Rivière C, Hegarty C, Stephenson H, Whelan P, O'Toole PW (2018) Gut and whole-body microbiota of the honey bee separate thriving and non-thriving hives. *Microbial Ecology*, 78, 195–205.
- Rollin O, Bretagnolle V, Decourtye A, Aptel J, Michel N, Vaissière BE, Henry M (2013) Differences of floral resource use between honey bees and wild bees in an intensive farming system. *Agriculture, Ecosystems and Environment*, 179, 78–86.
- Rothman JA, Andrikopoulos C, Cox-Foster D, McFrederick QS (2018) Floral and foliar source affect the bee nest microbial community. *Microbial Ecology*, <https://doi.org/10.1007/s00248-018-1300-3>.
- Rundlöf M, Andersson GKS, Bommarco R, Fries I, Hederström V, Herbertsson L, Jonsson O, Klatt BK, Pedersen TR, Yourstone J, Smith HG (2015) Seed coating with a neonicotinoid insecticide negatively affects wild bees. *Nature*, 521, 77–80.
- Saraiva MA, Zemolin APP, Franco JL, Boldo JT, Stefenon VM, Triplett EW, de Oliveira Camargo FA, Roesch LFW (2015) Relationship between honeybee nutrition and their microbial communities. *Antonie van Leeuwenhoek*, 107, 921–933.
- Saunders ME, Smith TJ, Rader R (2018) Bee conservation: Key role of managed bees. *Science*, 360, 389.
- Schwarz RS, Moran NA, Evans JD (2016) Early gut colonizers shape parasite susceptibility and microbiota composition in honey bee workers. *Proceedings of the National Academy of Sciences, USA*, 113, 9345–9350.
- Siviter H, Brown MJF, Leadbeater E (2018) Sulfoxafloer exposure reduces bumblebee reproductive success. *Nature*, 561, 109–112.
- Steinhauer N, Kulhanek K, Antúnez K, Human H, Chantawanakul P, Chauzat MP, vanEngelsdorp D (2018) Drivers of colony losses. *Current Opinion in Insect Science*, 26, 142–148.
- Torné-noguera A, Rodrigo A, Osorio S, Bosch J (2016) Collateral effects of beekeeping: Impacts on pollen-nectar resources and wild bee communities. *Basic and Applied Ecology*, 17, 199–209.
- Tsvetkov N, Samson-Robert O, Sood K, Patel HS, Malena DA, Gajiwala PH, Maciukiewicz P, Fournier V, Zayed A (2017) Chronic exposure to neonicotinoids reduces honey bee health near corn crops. *Science*, 356, 1395–1397.
- Vaudo AD, Tooker JF, Grozinger CM, Patch HM (2015) Bee nutrition and floral resource restoration. *Current Opinion in Insect Science*, 10, 133–141.
- Whitehorn PR, O'Connor S, Wackers FL, Goulson D (2012) Neonicotinoid pesticide reduces bumble bee colony growth and queen production. *Science*, 336, 351–352.
- Winfree R, Fox JW, Williams NM, Reilly JR, Cariveau DP (2015) Abundance of common species, not species richness, drives delivery of a real-world ecosystem service. *Ecology Letters*, 18, 626–635.
- Winfree R, Griswold T, Kremen C (2007) Effect of human disturbance on bee communities in a forested ecosystem. *Conservation Biology*, 21, 213–223.
- Woodcock BA, Bullock JM, Shore RF, Heard MS, Pereira MG, Redhead J, Ridding L, Dean H, Sleep D, Henrys P, Peyton J, Hulmes S, Hulmes L, Sárospataki M, Saure C, Edwards M, Genersch E, Knäbe S, Pywell RF (2017) Country-specific effects of neonicotinoid pesticides on honey bees and wild bees. *Science*, 356, 1393–1395.
- Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman

- FD, Lewis JD (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science*, 334, 105–108.
- Wu P, Axmacher JC, Song X, Zhang X, Xu H, Chen C, Yu Z, Liu Y (2018) Effects of plant diversity, vegetation composition, and habitat type on different functional trait groups of wild bees in rural Beijing. *Journal of Insect Science*, 18, <https://doi.org/10.1093/jisesa/iey065>.
- Xun EN, Zhao JM, Guo JX, Zhang YW (2017) Nectar-dwelling microorganisms and their ecological functions. *Acta Ecologica Sinica*, 37, 1757–1768. (in Chinese with English abstract) [荀二娜, 赵骥民, 郭继勋, 张彦文 (2017) 花蜜微生物及其生态功能研究进展. *生态学报*, 37, 1757–1768.]
- Yang Y, Ma S, Yan Z, Liu F, Diao Q, Dai P (2019) Effects of three common pesticides on survival, food consumption and midgut bacterial communities of adult workers *Apis cerana* and *Apis mellifera*. *Environmental Pollution*, 249, 860–867.
- Zheng H, Nishida A, Kwong WK, Koch H, Engel P, Steele MI, Moran NA (2016) Metabolism of toxic sugars by strains of the bee gut symbiont *Gilliamella apicola*. *mBio*, 7, e01326-16.
- Zou Y, Bianchi FJJA, Jauker F, Xiao H, Chen J, Cresswell J, Luo A, Huang J, van der Werf W (2017) Landscape effects on pollinator communities and pollination services in small-holder agroecosystems. *Agriculture, Ecosystems and Environment*, 246, 109–116.

(责任编辑: 黄祥忠)





•综述•

# DNA条形码参考数据集构建和序列分析相关的新兴技术

刘山林\*

(中国农业大学植物保护学院, 食品营养与人类健康高精尖创新中心, 北京 100193)

**摘要:** 近年来DNA条形码技术迅速发展, 产生的条形码的数量及其应用范围都呈指数性增长, 现已广泛用于物种鉴定、食性分析、生物多样性评估等方面。本文重点总结并讨论了构建条形码参考数据库和序列聚类相关的信息分析的技术和方法, 包括: 基于高通量测序(high throughput sequencing, HTS)平台以高效并较低的成本获取条形码序列的方法; 同时还介绍了从原始测序序列到分类操作单元(operational taxonomic units, OTUs)过程中的一些计算逻辑以及被广泛采用的软件和技术。这是一个较新并快速发展的领域, 我们希望本文能为读者提供一个梗概, 了解DNA条形码技术在生物多样性研究应用中的方法和手段。

**关键词:** DNA条形码; 可操作物种单元; 聚类; 宏基因条形码; 高通量测序

## DNA barcoding and emerging reference construction and data analysis technologies

Shanlin Liu\*

Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Plant Protection, China Agricultural University, Beijing 100193

**Abstract:** DNA barcoding has been growing exponentially in terms of the number of barcode generated as well as its applications, e.g. as conservation tools in: species identification for damaged specimens, diet analysis from gut content and feces, biodiversity assessment from environmental DNA (eDNA), bulk arthropod samples or invertebrate-derived DNA (iDNA). These applications often require coupling with high throughput sequencing (HTS) technologies, and when done so are referred to as metabarcoding. Here, we discuss the methods used to generate reference barcodes using cost-efficient HTS platforms, and introduce several rules-of-thumb and some widely-used tools to conduct data quality control, denoising, and Operational Taxonomic Units (OTUs) clustering. We hope this review will help readers better understand how these emerging technologies can be implemented alongside existing technologies to accelerate biodiversity assessments in an accurate and efficient way.

**Key words:** DNA barcoding; OTUs; clustering; metabarcoding; high throughput sequencing

### 1 引言

DNA条形码是指用于鉴定物种的一个或多个标准化短基因片段(Hebert et al, 2003)。标准化DNA条形码参考数据库的构建需要遵循几个特征, 包括: (1)序列长度控制在目前测序技术可读取的长度范围内; (2)种间序列差异一般应大于种内差异; (3)具

有高度保守的侧翼序列以便于扩增引物的设计, 并保证其能覆盖足够多的代表物种; (4)最重要的是其应存在于绝大部分人们感兴趣的物种中(Kress & Erickson, 2012)。目前, 有多个符合这些标准的基因片段被广泛接受并作为DNA条形码应用。例如, 用于动物的有线粒体细胞色素氧化酶亚基(cytochrome c oxidase subunit 1, COI)的一段650 bp的序列

收稿日期: 2018-07-30; 接受日期: 2018-12-25

基金项目: 深圳市基础研究(自由探索)(JCYJ20170817150755701)

\* 通讯作者 Author for correspondence. E-mail: shanlin1115@gmail.com

(Hebert et al, 2003), 用于植物的有质体核酮糖1,5-二磷酸羧化酶基因(ribulose 1,5-bisphosphate carboxylase gene, *rbcL*)和成熟酶K基因(maturase K, *matK*) (Hollingsworth et al, 2009)以及用于真菌的转录间隔区(internal transcribed spacer, *ITS*) (Nilsson et al, 2009; Schoch et al, 2012) (表1)。

在Hebert等(2003)首次提出DNA条形码的概念后, DNA条形码技术在条形码数量及其应用方面都呈指数增长, 例如作为生物多样性保护的工用于: 受损标本的物种鉴定(Armstrong & Ball, 2005)、通过肠道内容物和粪便分析食性(Kunz & Whitaker Jr, 1983; Bohmann et al, 2014)、基于环境DNA (environmental DNA, eDNA) (Baird & Hajibabaei, 2012; Taberlet et al, 2012)和生物混合样品或无脊椎动物源DNA (invertebrate-derived DNA, iDNA)样品 (Yu et al, 2012; Bohmann et al, 2013; Liu et al, 2013)进行生物多样性评估。这些应用通常需要依赖于高通量测序(HTS)技术, 并被称为宏基因条形码技术 (metabarcoding) (Taberlet et al, 2012)。宏基因条形码技术或宏基因组学的方法最初主要应用于微生物学领域, 通过从各种环境样品中提取的DNA来分析表征微生物群落(Caporaso et al, 2011)。过去十多年的研究表明, 此种方法同样可以应用于动物和植物群落 (后文中将其称为大生物群落, macrobial community) (Hajibabaei et al, 2016; Deiner et al, 2017)。考虑到微生物宏基因组学研究已经非常成熟, 本文将重点总结和讨论与大生物群落相关的宏基因条形码的技术进展, 主要包括构建DNA条形码参考数据库和DNA序列聚类的一些技术和方法。

**表1 广泛用于DNA条形码技术的标记基因**

**Table 1 Marker genes widely used for barcoding**

标记基因 Marker gene	目标物种 Targeted group	数据库 Database
<i>16S</i>	细菌和古细菌 Bacteria and archaea (Sogin et al, 2006)	核糖体数据库项目 Ribosomal Database Project (RDP, Cole et al, 2008); Greengenes (DeSantis et al, 2006); SILVA (Pruesse et al, 2007)
<i>ITS</i>	真菌(Schoch et al, 2012)、植物(Group et al, 2011)、原生生物 (Pawlowski et al, 2012) Fungi (Schoch et al, 2012); plant (Group et al, 2011); protist (Pawlowski et al, 2012)	UNITE (Kõljalg et al, 2005); GenBank (Benson et al, 2012)
<i>18S</i>	原生生物 Protist (Pawlowski et al, 2012)	SILVA (Pruesse et al, 2007)
<i>matK + rbcL</i>	植物 Plant (Hollingsworth et al, 2009)	生命条形码数据库 Barcode of Life Data Systems (BOLD, Ratnasingham & Hebert, 2007); GenBank (Benson et al, 2012)
<i>COI</i>	动物群(Hebert et al, 2003)、原生生物(Pawlowski et al, 2012) Fauna (Hebert et al, 2003) and protist (Pawlowski et al, 2012)	核糖体数据库项目 Ribosomal Database Project (RDP, Cole et al, 2008)

## 2 构建DNA条形码参考数据库

通过过去十多年中全球范围内科学家的通力合作, DNA条形码参考序列数据库, 例如BOLD (Ratnasingham & Hebert, 2007), 已经初具规模。然而, 目前参考数据库存在的一个典型问题是其数据在地理空间和物种覆盖度方面均存在很大程度上的不平衡, 这主要是由于全球各地在条形码研究方面投入差异所致, 尤其是在物种多样性热点地区, 科研投入尤其是分子生物学相关方面的投入不足, 制约了这些地区物种信息的数字化。尽管HTS平台的单碱基的测序成本显著下降, 但由于测序长度相对较短, 并不适用于对长扩增子测序(例如, *COI*基因条形码包括引物序列长~719 bp, 而最新的Miseq系统最长可以完成双端300 bp的测序, 仍然无法测通标准条形码序列), 使得Sanger测序仍然是目前获取DNA条形码序列(表1)的主流技术。

不可否认, 随着集中式和工业化的实验室流程的普及, 用于标准DNA条形码技术的分析成本自2000年以来已显著降低(Hebert et al, 2016)。Meier等(2016)通过咨询条形码研究中心, 如安大略生物多样性研究所(Biodiversity Institute of Ontario, BIO)和加拿大DNA条形码中心(Canadian Centre for DNA Barcoding, CCDB), 真实地估算了分析成本。结果表明, 对于具有高质量DNA的样品, 如果不包含额外的服务(例如重新索要收据、二次抽样、返回剩余的组织或DNA等), CCDB的每个样品的商业成本约为20美元。如果序列和标本图片共享给国际生命条形码项目(International Barcode of Life, iBOL),

费用将降至10美元左右。由于Sanger测序技术即将接近其通量和相关化学成本的极限,因此基于此的条形码测序成本不太可能进一步大幅降低。据估计,在全球范围内对物种进行条形码登记(Hebert et al, 2016)需要测序1亿个样本。这意味着仅仅构建全球条形码参考序列的预算就需要约10亿美元。参考数据库的不足导致基于HTS的宏基因条形码研究的结论常常只能限于使用分类操作单元(operational taxonomic units, OTUs),无法确定其具体的物种信息(Linnaean species name),因而不能将现有的生物学和生态学的知识与用此方法得出的多样性分布规律结合起来。

科学家一直在尝试使用HTS平台以更低的成本和人工投入获取DNA条形码参考序列。为了解决读长的限制,不同的学者采用了不同的方法,包括:获取标准条形码序列的局部序列;用多轮PCR的方法扩增全长条形码的不同区域;又或者利用三代单分子高通量测序技术,如Pacific Biosciences的长读长单分子实时(Single Molecular Real Time, SMRT)测序系统;通过生物信息学算法来拼装填补由于读长限制而产生的序列中部的缺口(gap)。例如,早期的研究分别对不同物种单独进行PCR后再混合,通过Roche 454平台获取物种条形码序列(Shokralla et al, 2014),但是由于测序通量的限制,化学试剂成本高,454平台被迫退出市场,而且此方法同Meier等(2016)的方法(基于Illumina平台)一样无法获取标准全长的条形码序列。研究人员还试图应用两轮

PCR扩增,每一轮分别扩增全长条形码的部分序列,随后通过简单拼接获取全长序列(Shokralla et al, 2015; Cruaud et al, 2017)。另外,单分子测序平台SMRT技术可以获取环形一致序列(circular consensus sequences, CCSs),对一个分子多次测序,可以有效校正该平台固有的高测序错误率(10%–16%)(Eid et al, 2008)。因此,有研究测试了将SMRT技术应用于条形码数据库构建的可行性(Liu et al, 2017; Hebert et al, 2018),研究测试的混合样品中,最复杂的样品有来自多达10,000个不同样本的DNA中的COI扩增子。结果表明随着其测序成本的进一步下降,SMRT技术将会是构建DNA条形码数据库方向的强有力的方法之一。Liu等(2017)还提出了一套新的解决方案(HIFI-Barcode),这是一套准确高效的生物信息学替代方案。该方法可以使研究者基于目前最经济高效的Hiseq平台,以现有成本的1/10获取全长条形码序列,而且不需要额外的PCR步骤(表2)。与此同时,该研究组还利用BGISEQ-500平台最新的单端400 bp (SE 400)测序技术开发了一套简单有效的DNA条形码数据库的实验和分析流程(HIFI-SE)(Yang et al, 2018)。此方法利用测序读长的优势,可以通过简单的两端序列比对连接从而获得全长的DNA条形码序列。

基于高通量技术的条形码参考序列构建方法还体现了更高的灵敏度,进一步降低了操作成本。虽然目前有对Sanger测序峰图自动读取和识别的程序,但是Sanger测序往往要求对结果峰图文件进行

表2 利用高通量测序平台批量获取DNA条形码的方法

Table 2 High throughput methods to achieve barcode sequences

目标序列长度 Targeted region length (bp)	优势 Advantages	劣势 Disadvantages	参考文献 Reference
~300	–	无法处理较长的目标序列; Roche 454平台 Can not work on long fragments; Roche 454 platform	Shokralla et al, 2014
~180	简单, 易操作, 成本低 Straightforward, easy to operate, cost-efficient	目标序列偏短, 只能用于物种初筛 Short targeted region; can only be used for species pre-clustering	Meier et al, 2016
~650	标准DNA条形码全长 Standard full-length COI	普适性差; 需要多轮PCR过程 Poor universality; multiple rounds of PCR	Shokralla et al, 2015; Cruaud et al, 2017
~650	易操作, 标准DNA条形码全长 Easy to operate, standard full-length COI	相对较高的计算资源 Relatively high requirement for computational resources	Liu et al, 2017
~650	易操作, 标准DNA条形码全长 Easy to operate, standard full-length COI	SMRT平台成本高 High cost of SMRT platform	Hebert et al, 2018
~650	易操作, 标准DNA条形码全长 Easy to operate, standard full-length COI	测序平台暂时不够普及 Not a mass production	Yang et al, 2018



肉眼观察甄别以优化数据质量,因此很难开展高效的自动化分析流程。而基于HTS的方法就不存在这样的问题,使得高效自动化成为可能。此外,这些基于HTS的方法可以检测微量的PCR扩增子,因而可以获取那些“失败”的PCR扩增子(电泳凝胶上没有明显条带),进一步提高了总体条形码成功率(Liu et al, 2017)。虽然所有这些方法都需要在引物上添加样品特异的标签序列,这会导致在最开始订购引物时产生一次性的费用,然而一次性的引物合成可以进行数以千次的反应,每次反应中的引物成本可以少到忽略不计。

虽然上述所有的方法目前都只是在动物类群的COI相关基因中通过验证,但是根据其方法原理,SMRT技术和HIFI-barcode技术可以很容易转移应用于其他类型的标记基因,如用于植物的*rbcL*和*matK*基因,但是需要进一步的实验证明。总之,我们相信,如果这些新方法在分类学中得到迅速而广泛的应用,将为全球生物DNA条形码的生成记录开辟新纪元,最终完成全球生物物种条形码信息的数字化。

### 3 DNA序列聚类

对DNA分子测序时,HTS平台不同于Sanger测序,它对结合在测序芯片上的几乎所有DNA分子测序,分别产出单独的序列。因此,源自于PCR扩增的错误如单碱基替换和嵌合体,因为在扩增产物中比例较低,在Sanger测序中不会被测到,但是这些问题序列在HTS测序过程中将会被测到并输出为有效序列。这个问题在基于扩增子的生物多样性研究中尤为显著,导致OTU预测数目大大增加,产生的序列多样性远远高于抽样的生物群落的“真实”丰富度,进而高估多样性,产生不可信的生态发现。如何区分有意义的生物学变异(种内和种间变异)与PCR和测序过程中的错误,是目前相关分析工具包开发的主要问题,也是最核心的挑战之一。考虑到现有条形码参考数据库的完整性不足(在物种多样性及单一物种的空间覆盖度方面都有很大的偏向性),大部分的研究分析都是根据序列相似度采用从头聚类的方法,以便于真实评估所研究样点的物种多样性信息(alpha diversity),而非通过比较一个预先准备好的数据库进行多样性分析(Quast et al, 2012; Zhang et al, 2013)。本文将简要介绍几种广泛使用的

工具以及其背后的运算逻辑,以帮助读者更好地理解如何将大规模HTS序列聚类成有生态学意义的OUT(图1)。由于早期很多为Roche 454焦磷酸测序开发的软件已经不再被广泛使用,将不在本文中进行讨论。

首先,在进行聚类分析之前,通过Illumina HiSeq或Miseq测序产生的原始序列需要进行质量过滤,主要包括去除建库或测序过程中产生的错误序列,如:(1)含有任何测序接头的序列。在文库制备过程中,接头序列会添加到插入片段两端。如果模板长度小于测序长度,测序序列就会含有接头序列。接头也可能出现在序列中间,这是由于建库过程中引物结合到模板错误的位置,或者PCR反应中退火,延伸不足所致。对于前者,目前常见的可用工具有AdapterRemoval (Schubert et al, 2016)、Skewer (Jiang et al, 2014)和SOAPnuke (Chen et al, 2017)等,其中部分软件还可以去除PCR引物,如果提供标签序列,还可以进一步拆分样品。如果是后者,需要将这样的序列整条去除。大多数测序服务提供商都提供接头报告文件帮助用户去除接头序列。(2)含低质量(Q)碱基的序列。Q值平均值虽然是一个广泛采用的参数,但是在OTU聚类分析中被证明并不是一个可取的办(Edgar, 2013)。假设有2个相同长度为150 bp的序列,其质量值分别为 $140 \times Q35 + 10 \times Q2$ 和 $150 \times Q25$ ,那么它们的平均Q值是一样的,但是前一条序列的预期错误碱基数为6.4,而后者为0.5。因此,对于有低Q值的序列,正确的

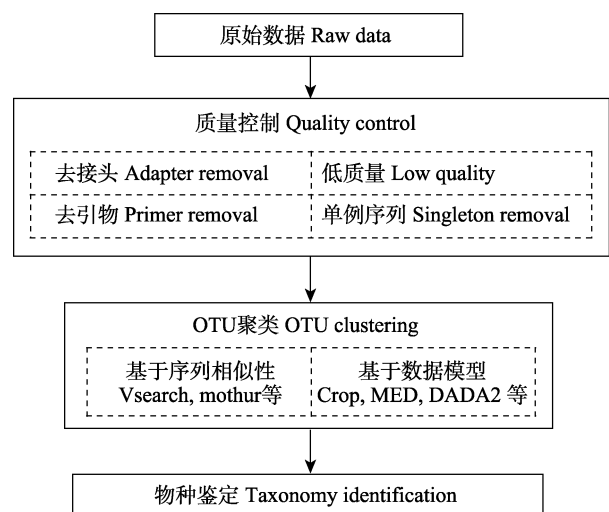


图1 条形码分析的数据处理流程图

Fig. 1 Diagram of DNA barcode data analysis

做法是删除整条的序列, 或删除其末端低质量的序列。

经过序列的预处理之后, 大多数OTU聚类流程, 如U/VSEARCH (Edgar, 2010; Rognes et al, 2016)、DADA2 (Callahan et al, 2016)、UPARSE (Edgar, 2013)等, 都会进行去除单例序列(singletons)的过程: 首先将相同序列合并为一条代表序列并记录其丰度, 随后去除单例序列(丰度为1的序列)。单例序列被普遍认为是错误序列。此外, 这一步降噪处理还会大大降低序列数量, 从而减轻后续分析的计算负荷。一些软件还包含额外的处理步骤, 如将序列删除为长度一致的序列(Edgar, 2013)或进行氨基酸翻译检查(Liu et al, 2013)。然而这些额外的降噪处理只适合某些特定的基因, 在使用时需要谨慎, 确认其使用范围。另外, 尽管不同的研究在嵌合体分析中, 有的分析流程倾向于在去除单例序列之后马上进行, 有的倾向于在OTU聚类之后进行, 但是其嵌合体鉴定的原理基本相似, 它们都试图找出双源嵌合体(bimera, two-parent chimera), 即嵌合体序列头尾两端分别来自于同一混合样品中高丰度的其他序列(Edgar et al, 2011; Callahan et al, 2016)。

现已发表的OTU聚类算法大致可以分为两类: (1)首先计算两两序列相似度, 然后以高丰度序列为根序列, 用一个预先设定的相似度(一般为97%)将这些序列进行分组。(2)用数学模型中心化和表征每个聚类单元。传统聚类方法需要计算比较所有的序列来计算距离矩阵, 已经很难处理现阶段研究的数据量(Matias Rodrigues & von Mering, 2013)。而目前广泛流行的USEARCH (Edgar, 2010), 亦或是开源的VSEARCH (Rognes et al, 2016), 或其特定版本的UPARSE (Edgar, 2013)都是基于快速启发式聚类算法。这个算法比传统方法能更快找到与目标序列相似的一个或几个代表序列, 极大降低了计算复杂度。因此目前主流的很多软件采取了类似的聚类方法, 包括: QIIME (UCLUST) (Caporaso et al, 2010)和mothur (Schloss et al, 2009)。CROP是一个基于高斯混合模型(Gaussian Mixture Model)的聚类方法(Hao et al, 2011)。它将高斯分布的平均值替换为一条中心序列, 以此来代表一个特定的组并利用高斯分布来处理测序错误和种间变异。DADA2 (Callahan et al, 2016)开发了一套基于测序质量值的模型, 用于估测Illumina扩增子测序中的错误。首先,

通过对错配碱基和其测序质量值之间进行加权LOESS局部建模(weighted loess fit), 然后通过对数据模型和观测数据的最佳拟合结果将序列聚类。MED (Minimum Entropy Decomposition) (Eren et al, 2015)采用了最小熵分解的算法, 利用序列间的信息不确定性来迭代分解数据集, 直到每个最终待解释的单元满足最大熵标准。

这些基于模型的方法不需要像前述基于序列相似性的聚类方法那样需要预先设定一个临界值(如97%), 而是通过数据本身特性对序列进行聚类。此外, Frøslev等(2017)提出了一个聚类后处理的方法(LULU), 结合序列相似度和共现(co-occurrence)模式从群落数据中去除错误的OTUs。此方法采用了一个类似于MED的算法, 但是其特点在于OTU聚类之后的数据处理, 将在涵盖多时空样本的研究中发挥重要作用。相比于传统基于序列相似度矩阵的方法, 这些基于数学模型的聚类方法可以降低OTU和 $\alpha$ 多样性高估的状况, 有望找到更多真实的变异, 同时减少错误序列, 从而获得更具有生态学意义的分类单元。得到OTU的参考序列之后, 大部分研究会进一步探讨其物种分类。本文不过多讨论与OTU物种鉴定相关的主题。简而言之, 物种鉴定可以通过使用BLAST或其他类似比对工具, 将序列比对到已建立的参考数据库, 如BOLD (Ratnasingham & Hebert, 2007)、Genbank (Benson et al, 2012)或其他用户定制的数据库。物种分类可以通过最佳匹配以及其序列相似度确定(Shi et al, 2018), 也可以采用基于系统发育树的方法, 根据序列在系统发育树中的位置来确定其物种分类(Zhang et al, 2013)。这两个方法都需要可靠的数据库。如果不能保证数据库的完整度和正确性, 目标序列和数据库数据比对的不确定性将导致模糊的, 甚至是错误的物种鉴定。

#### 4 总结和展望

我们正处于一个物种灭绝速度超过发现速度的时代, 很多物种在得到描述之前就灭绝了。完善正确地描述生物界所有物种的时空分布才能使全面了解生物多样性的现状及其驱动力(人为的、自然的, 或两者兼而有之), 进而才能给我们提供有价值的保护和管理策略, 缓解甚至遏止生物多样性持续降低的趋势。条形码技术旨在协助分类学家进

行物种鉴定并加速这一过程,而不是取代传统分类方法(Hebert et al, 2003)。条形码技术具有广泛的适用性(可以应用于生命之树上的几乎所有物种,无论个体的大小),而且所需的专业训练不及传统分类学繁重复杂,使得其不但成为分类学家的工具,也扩展繁衍出一系列为生态学家和公众服务的工具,应用于如受损样本和走私货物物种鉴定,结合 HTS 技术的生物多样性评估等。另外,条形码技术不能也不应该局限于基于扩增子的研究分析,随着测序成本的降低,条形码技术也在不断发展。略过 PCR 步骤(Zhou et al, 2013; Tang et al, 2015)和基于捕获芯片的技术(Liu et al, 2016)有望帮助科学家们更准确地理解生物多样性的组成及其驱动的生态过程。基于 eDNA 和 iDNA 的相关研究的顺利开展,将有助于揭示许多物种多样性的分布规律,尤其是那些小型的、隐秘的物种(Schnell et al, 2012; Mahon et al, 2013; Turner et al, 2014)。然而需要注意的是,条形码技术及其相关应用更多的是简化了对生态系统中物种鉴定和多样性评估的方法,但并不能取代合理的生态学设计,也不能减少样品采集的工作量。研究人员需要根据自身研究的特点设计样品采集方案,使其最终的研究具有统计学及生物学的意义,包括采集足够的样点分布和重复个数,以及涵盖相应的环境参数,如气候特征和 pH 梯度等。

## 参考文献

- Armstrong K, Ball S (2005) DNA barcodes for biosecurity: Invasive species identification. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360, 1813–1823.
- Baird DJ, Hajibabaei M (2012) Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, 21, 2039–2044.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2012) GenBank. *Nucleic Acids Research*, 41, D36–D42.
- Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Douglas WY, De Bruyn M (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29, 358–367.
- Bohmann K, Schnell IB, Gilbert MTP (2013) When bugs reveal biodiversity. *Molecular Ecology*, 22, 909–911.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581–583.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7, 335–336.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences, USA*, 108, 4516–4522.
- Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z (2017) SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Giga-Science*, 7, gix120.
- Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen A, McGarrell DM, Marsh T, Garrity GM (2008) The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, 37, D141–D145.
- Cruaud P, Rasplus J-Y, Rodriguez LJ, Cruaud A (2017) High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. *Scientific Reports*, 7, 41948.
- Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, Vere N (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26, 5872–5895.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* 72, 5069–5072.
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460–2461.
- Edgar RC (2013) UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10, 996–998.
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27, 2194–2200.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B (2008) Real-time DNA sequencing from single polymerase molecules. *Science*, 323, 133–138.
- Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML (2015) Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *The ISME Journal*, 9, 968–979.
- Frøslev TG, Kjølner R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, Hansen AJ (2017) Algorithm for post-clustering



- curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, 8, 1188.
- Group CPB, Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (*ITS*) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences, USA*, 108, 19641–19646.
- Hajibabaei M, Baird DJ, Fahner NA, Beiko R, Golding GB (2016) A new way to contemplate Darwin's tangled bank: How DNA barcodes are reconnecting biodiversity science and biomonitoring. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 371, 20150330.
- Hao X, Jiang R, Chen T (2011) Clustering 16S rRNA for OTU prediction: A method of unsupervised Bayesian clustering. *Bioinformatics*, 27, 611–618.
- Hebert PD, Braukmann TW, Prosser SW, Ratnasingham S, Ivanova NV, Janzen DH, Hallwachs W, Naik S, Sones JE, Zakharov EV (2018) A Sequel to Sanger: Amplicon sequencing that scales. *BMC Genomics*, 19, 219.
- Hebert PD, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, 270, 313–321.
- Hebert PD, Hollingsworth PM, Hajibabaei M (2016) From writing to reading the encyclopedia of life. *Proceedings of the Royal Society of London B: Biological Sciences*, 371, 20150321.
- Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, USA*, 106, 12794–12797.
- Jiang H, Lei R, Ding SW, Zhu S (2014) Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, 15, 182.
- Köljal U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjølner R, Larsson E (2005) UNITE: A database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist*, 166, 1063–1068.
- Kress WJ, Erickson DL (2012) DNA barcodes: Methods and protocols. In: *DNA Barcodes* (eds Kress WJ, Erickson DL), pp. 3–8. Humana Press, Totowa.
- Kunz TH, Whitaker JO Jr (1983) An evaluation of fecal analysis for determining food habits of insectivorous bats. *Canadian Journal of Zoology*, 61, 1317–1321.
- Liu S, Li Y, Lu J, Su X, Tang M, Zhang R, Zhou L, Zhou C, Yang Q, Ji Y (2013) SOAPBarcode: Revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons. *Methods in Ecology and Evolution*, 4, 1142–1150.
- Liu S, Wang X, Xie L, Tan M, Li Z, Su X, Zhang H, Misof B, Kjer KM, Tang M (2016) Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources*, 16, 470–479.
- Liu S, Yang C, Zhou C, Zhou X (2017) Filling reference gaps via assembling DNA barcodes using high-throughput sequencing—Moving toward barcoding the world. *GigaScience*, 6, 1–8.
- Mahon AR, Jerde CL, Galaska M, Bergner JL, Chadderton WL, Lodge DM, Hunter ME, Nico LG (2013) Validation of eDNA surveillance sensitivity for detection of Asian carps in controlled and field experiments. *PLoS ONE*, 8, e58316.
- Matias Rodrigues JF, von Mering C (2013) HPC-CLUST: Distributed hierarchical clustering for large sets of nucleotide sequences. *Bioinformatics*, 30, 287–288.
- Meier R, Wong W, Srivathsan A, Foo M (2016) \$1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. *Cladistics*, 32, 100–110.
- Nilsson RH, Ryberg M, Abarenkov K, Sjökvist E, Kristiansson E (2009) The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *FEMS Microbiology Letters*, 296, 97–101.
- Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Cepicka I, Decelle J, Dunthorn M (2012) CBOL protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biology*, 10, e1001419.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO (2007) SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35, 7188–7196.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2012) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41, D590–D596.
- Ratnasingham S, Hebert PD (2007) BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes*, 7, 355–364.
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75, 7537–7541.
- Schnell IB, Thomsen PF, Wilkinson N, Rasmussen M, Jensen LRD, Willerslev E, Bertelsen MF, Gilbert MTP (2012) Screening mammal biodiversity using DNA from leeches. *Current Biology*, 22, R262–R263.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL,

- Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW (2012) Nuclear ribosomal internal transcribed spacer (*ITS*) region as a universal DNA barcode marker for fungi. *Proceedings of the National Academy of Sciences, USA*, 109, 6241–6246.
- Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Research Notes*, 9, 88.
- Shi ZY, Yang CQ, Hao MD, Wang XY, Ward RD, Zhang AB (2018) FuzzyID2: A software package for large data set species identification via barcoding and metabarcoding using hidden Markov models and fuzzy set methods. *Molecular Ecology Resources*, 18, 666–675.
- Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M (2014) Next-generation DNA barcoding: Using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular Ecology Resources*, 14, 892–901.
- Shokralla S, Porter TM, Gibson JF, Dobosz R, Janzen DH, Hallwachs W, Golding GB, Hajibabaei M (2015) Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports*, 5, 9687.
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proceedings of the National Academy of Sciences, USA*, 103, 12115–12120.
- Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Environmental DNA. *Molecular Ecology*, 21, 1789–1793.
- Tang M, Hardman CJ, Ji Y, Meng G, Liu S, Tan M, Yang S, Moss ED, Wang J, Yang C (2015) High-throughput monitoring of wild bee diversity and abundance via metagenomics. *Methods in Ecology and Evolution*, 6, 1034–1043.
- Turner CR, Miller DJ, Coyne KJ, Corush J (2014) Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (*Hypophthalmichthys* spp.). *PLoS ONE*, 9, e114329.
- Yang C, Tan S, Meng G, Bourne DG, O’Brien PA, Xu J, Liao S, Chen A, Chen X, Liu S (2018) Access COI barcode efficiently using high throughput Single End 400 bp sequencing. *bioRxiv*, doi: 10.1101/498618.
- Yu DW, Ji YQ, Emerson BC, Wang XY, Ye CX, Yang CY, Ding ZL (2012) Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, 3, 613–623.
- Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29, 2869–2876.
- Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, 2, 4.

(特邀责任编辑: 周欣 责任编辑: 时意专)



•综述•

# 基于三代测序技术的微生物组学研究进展

许亚昆<sup>1,2</sup> 马越<sup>1,2</sup> 胡小茜<sup>1</sup> 王军<sup>1\*</sup>

1 (中国科学院微生物研究所, 北京 100101)

2 (中国科学院大学, 北京 100049)

**摘要:** 微生物在人类生活中无处不在, 过去人们对微生物的认识仅停留在单菌培养和定性研究上, 而测序技术的发展极大地促进了微生物组学的研究。越来越多的证据表明: 人体共生微生物、特别是肠道微生物与人类健康息息相关。二代测序技术凭借其高通量、高准确率和低成本的特点, 成为微生物组学研究中的主流测序技术。但是随着研究的深入, 二代测序技术的短读长(< 450 bp)增加了后续数据分析和基因组拼接难度, 也限制了该技术在未来研究中的应用。在此背景下, 第三代测序技术应运而生。第三代测序技术又称单分子测序, 能够直接对单个DNA分子进行实时测序, 而不需要经过PCR扩增。第三代测序技术的平均读长在2–10 kb左右, 最高可以达到2.2 Mb, 实现了长序列的高通量测序。凭借其超长的测序读长、无GC偏好性等优势, 三代测序技术为微生物基因组全长测序, 组装完整可靠的基因组提供了新的方法。本文在描述三代测序的技术特点和原理的基础上, 重点介绍了三代测序技术在微生物16S/18S rRNA基因测序、单菌的基因组组装以及宏基因组中的研究应用和进展。

**关键词:** 微生物; 三代测序; 16S/18S rRNA; 宏基因组

## Analysis of prospective microbiology research using third-generation sequencing technology

Yakun Xu<sup>1,2</sup>, Yue Ma<sup>1,2</sup>, Xiaoxi Hu<sup>1</sup>, Jun Wang<sup>1</sup>

1 Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101

2 University of Chinese Academy of Sciences, Beijing 100049

**Abstract:** Microbes are ubiquitous in human life. In years past, the study of microbes has only focused on single-bacteria cultures and qualitative analyses. The development of sequencing technology has greatly enhanced progress in microbiology research and more and more evidence shows that human symbiotic microbes, especially intestinal microbes, are closely related to human health. Second-generation sequencing technology is currently mainstream in microbiology research because of its high throughput, high accuracy and low cost. However, with the deepening complexity of research, the disadvantages of second-generation technology, i.e. short read length (< 450 bp), lead to subsequent challenges in data analysis and genome assembly, and limit the applicability to future research. In this context, the third-generation sequencing technology comes into being. The third-generation of sequencing (TGS) technology is also called single molecule sequencing. It directly carries out real-time sequencing of single DNA molecules without PCR amplifications. TGS technology significantly increases read length up to 2–10 kb or even 2.2 Mb. Because of its advantages of long read and no preference for GC, TGS provides a new method for full-length gene sequencing that facilitates the assembly of complete and reliable genome maps in microbes and that further reveals the diversity of microbial structures and functions. This review summarizes the technical characteristics and principles of TGS, and then mainly analyzes its applications and progress in 16S/18S rRNA gene sequencing, complete bacterial genome mapping and metagenomics research.

**Key words:** microbes; third-generation sequencing; 16S/18S rRNA; metagenomics

微生物通常指一切难以用肉眼观察到的微小生物, 包括细菌、病毒、古菌、真菌以及一些微小

收稿日期: 2018-07-30; 接受日期: 2018-12-25

基金项目: 科技部重点研发子课题(2018YFC2000504)和国家自然科学基金(31771481)

\* 通讯作者 Author for correspondence. E-mail: junwang@im.ac.cn



的原生物。微生物体积微小、结构简单,却又无处不在,在人类健康(Yang & Tarng, 2018)、工农业生产(Huang et al, 2018)、环境保护(Liu et al, 2018)和食品安全(Alori & Babalola, 2018)等方面都发挥着重要的作用。微生物传统的研究方法以分离纯菌培养为主,但由于培养技术的限制,超过99%的原核微生物无法在实验室培养得到,导致微生物结构和功能多样性的研究一直发展缓慢(Schloss & Handelsman, 2005)。

16S/18S rRNA基因存在于所有微生物的基因组中,是微生物分类鉴定最常用的标记分子。Sanger测序技术的兴起(Sanger et al, 1977),使得16S rRNA基因测序在细菌分类学中被广泛应用,由此发现了大量新的微生物门类,增加了微生物的多样性。1998年提出的宏基因组学(Hugenholtz et al, 1998)方法,是直接从环境样品中获取总的遗传物质来研究微生物群落的组成和结构,发现新的功能基因。宏基因组学的方法不仅解决了微生物分离培养的难题,而且可以全面分析微生物群落的多样性和丰度,研究微生物之间、微生物和环境或宿主之间的关系。而在过去的数年间,第三代测序技术的出现,以其超长的测序读长、无GC碱基偏好性和速度快等优势,在以16S/18S rRNA基因或全基因组为目标序列的微生物组学研究中取得了一系列进展,因此微生物的组成和群落结构、代谢特征、系统进化、微生物与环境的相互作用得到了更深层次的研究。

## 1 测序技术的发展

### 1.1 第一代测序技术

第一代测序技术(Sanger法)始于1977年,该测序方法完成了噬菌体X174全长5,375个碱基的基因测序,由此生命科学研究进入了基因组学时代(Sanger et al, 1977)。Sanger法即双脱氧终止法(Chain Termination Method),是指在含有四种脱氧核苷三磷酸(dNTP)的反应体系中加入一种不同的荧光标记的双脱氧核苷三磷酸(ddNTP)。由于ddNTP没有DNA延长所需要的3-OH基团, DNA合成随机地在G、A、T、C处终止,然后通过凝胶电泳确定待测DNA分子的长度。第一代测序技术的测序读长在1,000–1,500 bp,但一次只能测一条单独的序列,导致测序成本高、通量低,不能够应用于大规模的基

因测序。

### 1.2 第二代测序技术

二代测序技术又称“下一代”测序技术(Next-Generation Sequencing Technology, NGS)。该技术在保持测序准确度的前提下,主要解决了第一代测序通量低的问题。二代测序技术可以同时几十万到几百万条DNA分子进行测定,因此也被称为高通量测序技术(Metzker, 2010)。二代测序平台主要有Roche公司的454 (Sogin et al, 2006)、Illumina公司的Solexa、HiSeq (Gloor et al, 2010)和ABI公司的Solid等测序仪(Mardis, 2008),其原理分别为焦磷酸测序、边合成边测序、连接测序。二代测序技术既大大降低了测序成本又大幅提高了测序通量,同时保持了高准确性。比如使用Illumina公司的技术平台,在30×测序深度下,完成人类基因组测序仅仅需要1周(Niedringhaus et al, 2011)。但二代测序技术的主要问题在于测序读长短,只有250–300 bp,而且由于测序的系统偏好性,导致有些序列可能被测了很多次,但也有一些量少的序列无法被大量扩增,造成信息的丢失。

### 1.3 第三代测序技术

第三代测序技术主要有Pacific Biosciences公司的SMRT (Single Molecule, Real-Time Sequencing)技术和Oxford Nanopore公司的Nanopore技术(Eid et al, 2009)。三代测序技术的主要特点是单分子实时测序、长读长,但错误率偏高。下面从技术特点和测序原理两方面来介绍第三代测序技术。

#### 1.3.1 SMRT技术

SMRT技术的核心是能够实现单个DNA分子的测序,并且实时监控测序结果。2013年Pacific Biosciences公司成功推出商业化的三代测序仪PacBio RSII后,三代测序开始被广泛应用于微生物基因组研究中。经过不断的改良和升级,又在2015年10月推出全新的PacBio Sequel测序系统,测序技术的准确性和通量不断提高。SMRT技术测序系统主要有SMRT Cell、零模波导孔(Zero-Mode Waveguides, ZMW)和DNA聚合酶。

SMRT Cell是测序芯片, DNA样本完成建库后装载到Cell中才能上机测序。PacBio RSII系统配套的一个SMRT Cell的碱基产量为500 M到1 G,而最新升级的Sequel系统配套的一个SMRT Cell可产生5–10 G的数据,测序通量提升10倍左右。零模波导

孔ZMW是一种直径为50–100 nm的圆形纳米小孔。当DNA分子进入小孔后,因为从孔底发出的激发光不能穿透小孔进入上方的溶液区,被限制在底部一个足以覆盖被检测DNA部分的区域,而共聚焦显微镜只收集该区域的信号,从而将背景噪音降到最低。PacBio RSII使用的一个SMRT Cell中含有15万个ZMW孔,而最新的Sequel平台配套的升级版SMRT Cell中含有100万个ZMW孔。DNA聚合酶是实现SMRT技术长测序读长的关键之一,它在保持测序延续性的基础上,实现了DNA体内的合成速度。DNA聚合酶能够每秒钟读取10个碱基,因此测序速度是化学测序方法的2万倍。

### 1.3.2 Nanopore技术

与以往的测序技术均不同,Nanopore是采用电信号进行测序(Magi et al, 2017)。2014年,Oxford Nanopore公司开始对外提供Nanopore MinION试用项目计划,随后不断对早期版本仪器的高错误率和低通量问题进行改善。从2016年开始,Nanopore平台的通量得到较大提升,错误率也显著降低,在基因组中的应用也从微生物小基因组逐渐延伸到复杂动植物基因组。而更高通量测序平台GridION X5和PromethION的发布使Nanopore在复杂物种中的应用更为简单和便捷。新的测序仪在延续MinION核心测序技术的操作简单和文库制备快等优点的基础上,弥补了MinION测序仪通量低以及不适用于大批量样本或大基因组测序的不足。根据最新的文献记载,Nanopore MinION测序基于新开发的BulkVis工具首次生成了单个2.2 M序列(Payne et al, 2018)。

Nanopore测序系统主要包括纳米孔、薄膜和马达蛋白(motor protein)。纳米孔是一种跨膜蛋白形成的纳米孔道。不同版本的测序芯片使用不同的跨膜蛋白,目前R9版的芯片使用的是来自大肠杆菌的CsgG蛋白质经过基因工程改造后形成的通道。薄膜是人工合成的具有高电阻的膜,膜的两侧浸在含有离子的水溶液中,通过对膜上施加电势,离子通过薄膜上的蛋白纳米孔产生电流。马达蛋白是一种DNA解旋酶,在构建文库时,马达蛋白和引导接头(leading adaptor)一同加在DNA分子上,在测序过程中,马达蛋白会对双链DNA解螺旋使其变为单链,使得单链DNA以一定速度经过纳米孔。

Nanopore技术有3种不同的建库方式:(1)在1D

建库中,仅有引导接头,在测序过程中,马达蛋白对双链DNA解压和解链,引导接头通过纳米孔,随后模板链通过。(2)在2D建库中,既有引导接头,还有连接双链DNA分子的发夹接头(hairpin adaptor)。在测序过程中,首先是马达蛋白对双链DNA解螺旋使其变为单链,引导接头通过纳米孔,随后模板链通过,然后发卡接头和互补链通过。(3)在1D<sup>2</sup>建库中,DNA双链分别通过纳米孔,但并未如2D测序中通过发卡接头进行连接。当模板链完成测序后,纳米孔会捕获互补链的马达蛋白进行互补链测序(Clarke et al, 2009)。1D测序建库的优势在于文库构建更便捷,可低至10 min;相对1D测序,2D测序中可以得到更长的读长;1D<sup>2</sup>同时对模板链和互补链进行测序,可以得到高质量的一致性序列(Jain et al, 2018)。

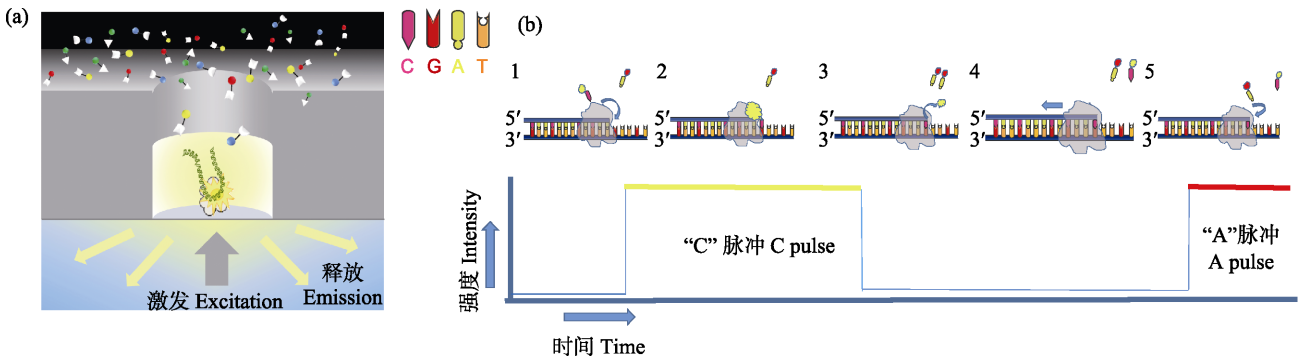
## 2 三代测序技术在微生物组学研究中的应用

### 2.1 16S/18S三代测序

16S rRNA基因存在于所有原核微生物的基因组中,序列长度约为1,500 bp (18S rRNA基因存在于所有真核微生物的基因组中,序列全长为1,500–2,000 bp)。16S rRNA基因有9个高变区(V1–V9),在一定程度上反映了微生物间的进化差异,可以利用高变区对不同菌属、菌种的细菌进行分类鉴定;而与高变区间隔分布的8个高度保守区域在细菌漫长的进化过程中则保持了结构和功能的稳定性。1977年,Woese等依据细菌16S rRNA基因序列上的区别,提出了三域系统,并分别命名为细菌域、古菌域和真核域(Woese & Fox, 1977)。目前16S/18S rRNA基因是研究微生物系统发育和分类鉴定最常用的分子标记(Woese, 1987)。

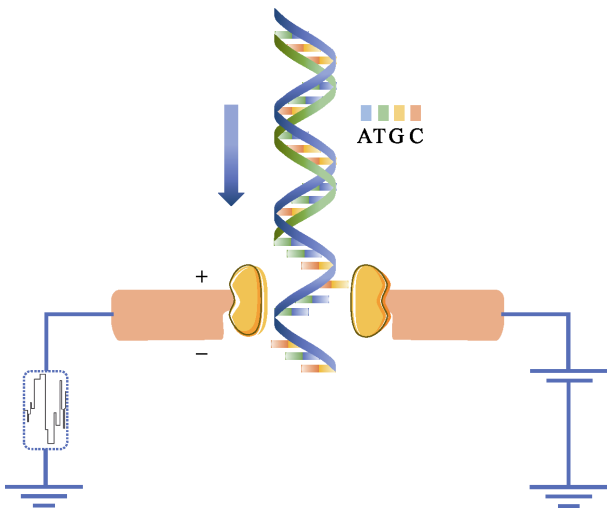
基于16S/18S的微生物基因测序即是利用16S/18S rRNA保守区域的基因序列设计引物,扩增研究所需的基因片段(Jonasson & Monstein, 2002)。包括基因组DNA的获取;设计标准引物对16S/18S基因的一个或多个区进行扩增;聚类,按照一定的相似性将扩增得到的序列进行分组,细菌中的分类操作单元(operational taxonomic unit, OTU)是基于序列97%相似性的分类;然后通过数据库比对为OTU的代表性序列提供分类学注释。

三代测序技术由于超长的测序读长,在微生物的分类鉴定和群落多样性的研究中更具有优势。二



**图1 PacBio SMRT测序原理。**(a)在零模波导孔(Zero-Mode Waveguides, ZMW)中, 单个DNA分子模板与引物和聚合酶结合后, 被固定到ZMW孔底部。DNA合成开始时, 新加入的荧光标记的dNTP由于碱基配对在ZMW底部停留较长时间, 激发后发出对应的荧光信号被共聚焦显微镜实时记录;(b) 1)荧光标记胞嘧啶脱氧核苷酸; 2)胞嘧啶脱氧核苷酸进入DNA链配对, 发射荧光信号; 3)荧光基团被DNA聚合酶切除, 荧光消失; 4)标记新的脱氧核苷酸; 5)继续新一轮合成。

Fig. 1 Schematic diagram of PacBio single molecule real-time sequencing. (a) In the ZMW hole, a single DNA molecule template combined with primers and polymerase is bind to the bottom of ZMW hole. At the beginning of DNA sequencing, the newly added fluorescent labeled dNTP remained at the bottom of ZMW for a long time due to base pairing, and the corresponding fluorescent signals were recorded by confocal microscopy in real time. (b) (1) Fluorescence labeling cytosine deoxynucleotides; (2) Cytosine deoxynucleotides entering DNA chain pairing, emitting fluorescent signals; (3) Fluorescent group is removed by DNA polymerase, fluorescence disappeared; (4) Label new deoxynucleotides; (5) Continue a new round.



**图2 Nanopore利用电信号检测出DNA的碱基序列。**纳米孔直径很小, 仅仅允许单个核苷酸通过。当DNA单链通过的时候, 就会对离子的流动造成阻碍, 从而使流过纳米孔的电流强度发生变化。由于ATCG四种碱基的带电性质不一样, 造成电流大小的波动也不一样, 因此可根据电流的变化鉴定所通过的碱基类型。

Fig. 2 Nanopore DNA sequencing using electronic signals as detection methods. The diameter of the nanoscale is very small that only a single DNA molecule is allowed to pass through. When a single strand of DNA passes through, it blocks the flow of ions and changes the current intensity across the nanopore. Because the charge properties of the four bases of ATCG are different, the type of base passed is identified according to the change of current.

代测序通常只能获得1-2段高变区域, 而三代测序的读长提升了2-3倍以上, 因而也提升了物种鉴定

的分辨率。研究人员对同一人的肠道菌群样本分别进行二代和三代的16S rRNA基因测序, 其中使用Illumina Miseq平台对16S rRNA基因的V3-V4区进行扩增测序, 使用PacBio SMRT技术直接读取和覆盖16S rRNA基因的V3-V7区, 实验结果表明有更多的OTUs基于SMRT测序数据被鉴定出来(Franzen et al, 2015)。Cusco等(2018)将Nanopore MinION应用于16S rRNA基因测序, 在犬类的皮肤微生物群体中发现了新的细菌门类。在另外一项针对已知菌种混合组成的模拟菌群的研究中, Nanopore MinION的测序数据使得一些物种能鉴定到种的水平(Benitez-Paez et al, 2016)。因此当三代测序技术的读长可以覆盖到更多基因片段时, 测序结果能够更准确地检测到微生物群体多样性组成。

### 2.2 细菌/真菌基因组三代测序

在微生物组学的研究中, 已有参考基因组的微生物数量远远低于自然界存在的微生物数量, 除了微生物难以纯化培养外, 还有一部分原因是二代测序读长短, 难以解决细菌/真菌基因组中的高重复区域和高GC区域的组装问题, 因此组装后的基因组中往往存在很多gap。而三代测序技术以其超长的测序读长和无GC偏好性克服了以上部分难题, 在单细菌/真菌的基因组组装中取得了很大突破。

2013年Chin等就使用SMRT技术组装出16个微生物基因组, 与参考基因组一致性达99.9999%



表1 三代测序技术的比较

Table 1 Comparison of three generation sequencing technologies

	技术平台 Technical platform	测序原理 Principle of sequencing	测序读长 Read length	优点 Advantages	缺点 Limitations
第一代 The first generation	Sanger	可中断测序 Chain-terminating sequencing	600–1,000 bp	读长长; 准确率高; 能很好地处理一些重复序列和多聚序列 Long reads; high accuracy; good ability to deal with repetitive and homopolymer regions.	通量低; 样品制备成本高, 难以做大量的平行测序 Low throughput; high cost of Sanger sample preparation; making massively parallel sequencing prohibitive.
第二代 The second generation	Roche/454	焦磷酸测序 Pyrosequencing	200–400 bp	在二代测序中读长最长; 高通量 Longest read lengths among the second-generation; high throughput.	样品制备较难; 难于处理重复和同种碱基多聚区域 Challenging sample preparation; hard to deal with repetitive/homopolymer regions.
	Illumina	边合成边测序 Sequencing by synthesis	2 × 150 bp	高通量 Very high throughput	读长短 Short reads
	ABI/Solid	连接测序 Sequencing by ligation	25–35 bp	高通量; 成本低 High throughput; low cost.	测序运行时间长; 读长短, 造成后续的数据分析困难和基因组拼接困难 Long sequencing runs (days); short reads, resulting in difficulties in subsequent data analysis and genome assembly.
第三代 The third generation	PacBio SMRT	边合成边测序/ DNA聚合酶 Sequencing by synthesis/DNA polymerase	~1,000 bp	高平均读长; 不需要扩增; 最长单个读长接近100 kb Long average read length; no amplification of sequencing fragments; longest individual reads approach 100 kb.	错误率高; 依赖DNA聚合酶的活性 Low accuracy; dependence on DNA polymerase activity.
	Nanopore	电信号测序/ 核酸外切酶 Electronic signals sequencing/exonuclease	最大记载2.2 M Maximum record 2.2 M	读长超长; 电学测序; 方便携带 Over-long read; electronic sequencing; portable.	错误率高 High sequencing error

(Chin et al, 2013)。Brown等(2014)分别使用Illumina和454这些二代测序平台和PacBio RS II平台技术对梭状芽孢杆菌(*Clostridium autoethanogenum*)进行基因组组装, 发现二代测序平台组装得到*Clostridium*的基因组仍然存在一些gap, 而使用PacBio SMRT测序直接组装得到一条完整的Contig; 并且之前由于技术的限制, 无法区分*C. autoethanogenum*和*C. ljungdahlii*两支菌株, 而基于PacBio SMRT的全基因组测序, 发现两者在CRISPR系统、氢化酶等方面具有显著差异。Ludden等(2017)用Nanopore MinION测序平台首次组装出了*Enterobacter kobei*的基因组完成图, 并且组装出编码blaOXA-48基因的质粒。Wick等(2017)对12个不同种的克雷伯氏肺炎菌进行测序, 然后与前期二代数据混合组装, 既保证了测序数据的完整性, 又保证了测序的准确性, 最终都组装成了基因组完成图。此外假单胞菌*Pseudomonas koreensis* P19E3的基因组含有长达70 kb的重

复序列, 使用二代平台的测序数据无法组装到完整的基因组, 而Schmid (2018)使用Nanopore MinION测序仪, 首次完成了对*Pseudomonas koreensis* P19E3的基因组组装。

真菌基因组大小在2.5–150 Mb之间, 介于细菌和大型动植物之间, 随着三代测序平台的改进和升级, 测序通量的不断提升, 三代测序技术在基因组研究中的应用也逐渐从细菌等小型基因组延伸到真菌等大型基因组中。*Orpinomyces* sp. strain C1A菌株是一种寄生在大型家畜肠胃中的厌氧真菌, 基因组超过100 Mb, GC含量只有17%, 是目前已知的GC含量最低的物种之一。Yousse等(2013)在对该菌株的基因组组装中发现, 对比Illumina测序数据得到的组装结果, PacBio RS的测序数据组装得到的contigs更长、数量大幅降低, 而且发现C1A菌株具有降解木质素的能力。此外Faino等(2015)基于PacBio RS II平台对黄萎病菌菌株JR2和VdLs17测



序并组装, 最终都得到了8条完整的染色体。2017年荷兰研究者Jansen等分离出了DDNA#1酵母菌株DNA, 使用Illumina和Nanopore MinION分别进行测序, 通过Illumina测序数据组装得到了14,764 contigs, 而使用Nanopore测序数据组装得到的contigs仅61条, 大幅度提高了酵母菌基因组数据的质量和完整度(Liem et al, 2017)。

### 2.3 宏基因组三代测序

宏基因组学是将环境中的微生物群落作为整体进行研究, 与16S/18S或全基因组分析相比, 宏基因组学能够对微生物群体基因组成及其功能、微生物群体的多样性、微生物与环境、微生物与宿主之间的关系进行全面解读。2016年Hug等就利用宏基因组技术发现了1,000多种未被培养或者了解甚少的微生物体, 进一步扩展了微生物的多样性, 而新的细菌门的出现, 也重新补充了进化树三域系统(Hug et al, 2016)。Fierer等(2007)通过构建牧场、沙漠、雨林土壤宏基因组文库, 对环境中的细菌、古生菌、真菌及病毒多样性进行了研究。结果发现土壤环境中包含着大量不可培养的新病毒种类, 其基因型特征与常规培养获得的病毒具有很大的差异。除此之外, 宏基因组学在海洋微生物多样性的探索中也取得了很大的进展, 促进了海洋微生物资源的开发。2015年Sunagawa领导的团队分析了来自世界各地的68个地点的远洋中上层水域的共243个样本, 构建了含有 $4 \times 10^7$ 个非冗余基因的海洋微生物宏基因组库, 并提出在垂直水平上海洋微生物群落的变化主要是由温度决定的; 同时对139个富含原核微生物的样本进行统计分析, 共发现了5,755个同源微生物组, 其中有40%的同源组与人类肠道菌群的同源组相同(Sunagawa et al, 2015)。

二代测序应用于宏基因组学时, 由于读长过短会导致一些基因信息的丢失。相比之下, 宏基因组学三代测序结果可以更加真实地反映菌群的组成情况, 较为准确地挖掘出新的功能基因。例如研究人员使用二代和三代测序平台分别对同一微生物发酵池样品进行了宏基因组测序分析, 并对结果做了比较。发现基于Illumina测序数据的组装, 由于读长较短, 结果存在大量不同物种间的同源序列错误组装, 无法真实呈现菌群组成; 而PacBio RS II得到的结果真实反映了两种优势菌株的组成情况(Frank et al, 2016)。Tsai等(2016)也采用了PacBio RS II和

Illumina HiSeq两种测序平台对人类手部和足部的菌群进行宏基因组分析, 结果表明三代测序技术显著减少了contigs的数量, 大大降低了序列拼接和基因组组装的难度, 而且从人的皮肤菌群样本中组装、注释、构建获得一例未知微生物的高质量基因组。在对Sakinaw Lake水体微生物的研究中, Singer等(2016)发现SMRT和Illumina的测序数据在门的水平上没有太大差别, 但是随着群落复杂度的增加, 二代和三代的测序数据在群落结构和系统发育分辨率上显示出显著差异。

新型三代测序设备的出现还扩展了宏基因组研究的应用场景。目前基因测序的工作主要局限于实验室, 而Nanopore MinION这种便携式测序仪为开展基因组实时测序提供了便利, 在鉴定新的病原体和细菌基因组测序方面得到了广泛应用。例如猪的肠道疾病通常是由不同的病原体引起的, 宏基因组学是一种常用的诊断方法, 而使用Nanopore技术在3小时内就可以完成样品测序和致病菌的诊断(Theuns et al, 2018)。MinION测序仪对于样本的预处理要求也较低, 对一些高致病性病毒如奇昆古尼亚病毒、埃博拉病毒和丙型肝炎病毒, MinION测序仪可以直接使用临床样本进行检测, 而不需要病毒的纯化培养(Greninger et al, 2015)。在埃博拉等暴发疾病的实时检测中, 研究人员就利用这种便携式测序仪在整个基因组库中鉴别出了新型的单核苷酸多态性(SNPs), 为流行病暴发的预警和病原体的进化提供了重要的方法(Quick et al, 2016)。

目前RNA病毒的基因组研究通常是先反转录为cDNA再进行测序。最近美国的研究者通过设计一种针对流感病毒基因组的Nanopore测序接头RTA, 实现了对流感病毒RNA的直接测序(Callaway, 2018)。而且传统方法对表观遗传修饰的分析, 需要对样本进行亚硫酸盐处理或者抗体沉降(pull-down)的预处理, 增加了检测的难度。通过Nanopore技术, Andrew Smith研究团队直接对大肠杆菌全长16S核糖体RNA进行测序, 从采样到出结果只需要2小时。通过数据分析, 该研究小组在16S rRNA的已知位点鉴定出了7-甲基鸟苷(m7G), 并且观察到了假尿嘧啶核苷等在内的表观遗传修饰的存在(Smith et al, 2017)。中国科学院水生生物研究所赵亮博士通过PacBio RS II测序, 获得3个微囊藻菌株的全基因组甲基化修饰图谱, 发现了18个未被鉴定的甲基化模

块,也发现不同微囊藻菌株的甲基化和甲基转移酶含量差异很大(Zhao, 2018)。

三代测序技术的兴起,极大地促进了微生物组学的发展,但是错误率偏高是三代测序技术面临的重大问题。15%–40%的错误率,大大地高于二代测序技术(< 1%),成为限制其商业应用开展的重要原因。不过三代测序的错误是随机发生的,可以靠覆盖度来纠错(但这又要增加测序成本)。为了提升准确率,PacBio公司对SMRT技术配套的技术不断优化,文献中经常看到的P4C2、P5C3以及P6C4代表的就是不断升级的试剂。同时,PacBio独有的环形一致性测序模式(circular-consensus sequence, CCS)极大地提高了单碱基测序的准确率。Nanopore公司也不断从Flow Cell、纳米孔、测序试剂和信号捕获及碱基识别软件等方面进行升级改进。其中,纳米孔最开始为R6,后来不断升级而出现了R7, R8, R9版本,到现在已经升级为R9.4,随之而来的是准确率和通量的提升。而且目前已经有比较成熟的软件Canu专门处理三代技术这类错误率较高的测序数据。Canu采用了先纠错、修整再组装的策略,通过序列与序列之间的重叠,进行单碱基的错误纠正,测序覆盖度越高,准确率就越高。

### 3 结语

近年来,微生物组学是生命科学研究的热门领域,而微生物组的深入研究和理解得益于测序技术的快速发展。第三代测序技术的出现,实现了很多技术上的突破:(1)凭借其超长读长的优势,解决了二代测序技术读长过短的问题,在微生物基因组学研究中减少了后续的基因组组装和注释的工作量,节省了大量的时间;(2) SMRT技术利用DNA聚合酶自身的合成速度,配上高分辨率的光学检测系统,实现了实时快速检测;(3)三代单分子测序技术的发展也让通过测序实时读取DNA碱基修饰成为可能,这对在基因组水平直接研究表观遗传现象有极大的帮助;(4) Nanopore技术还可以直接对RNA进行测序,省去了反转录过程。

但由于三代测序较高的错误率,目前三代测序技术在微生物组中的研究往往需要与二代测序数据结合使用,即利用二代测序数据对三代测序数据进行校正,也就意味着在实验中至少要制备2个不同的文库,造成了一定的人力和物力成本的增加。

随着三代技术的不断改进和新的组装方法的出现,测序的准确性和通量也在逐步上升。相信在不久的将来,基于三代测序技术的微生物组学研究能够从获得高质量的基因组组装结果,逐渐转向对物种生物学特征和进化历程的深入研究,研究策略也由单一组学测序逐渐延伸为基因组、转录组、代谢组和表观遗传组的多组学分析。

### 参考文献

- Alori ET, Babalola OO (2018) Microbial inoculants for improving crop quality and human health in Africa. *Frontiers in Microbiology*, 9, 2213.
- Benitez-Paez A, Portune KJ, Sanz Y (2016) Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION™ portable nanopore sequencer. *GigaScience*, 5, 4.
- Brown SD, Nagaraju S, Utturkar S, De Tissera S, Segovia S, Mitchell W, Land ML, Dassanayake A, Kopke M (2014) Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant Clostridia. *Biotechnology for Biofuels*, 7, 40.
- Callaway E (2018) Flu virus finally sequenced in its native form. *Nature*, 556, 420.
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, 10, 563–569.
- Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology*, 4, 265–270.
- Cusco A, Vines J, D'Andreano S, Riva F, Casellas J, Sánchez A, Francino O (2018) Using MinION to characterize dog skin microbiota through full-length 16S rRNA gene sequencing approach. *bioRxiv*, doi: <https://doi.org/10.1101/167015>.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong XX, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma CC, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, 323, 133–138.
- Faino L, Seidl MF, Datema E, van den Berg GC, Janssen A, Wittenberg AH, Thomma BP (2015) Single-molecule

- real-time sequencing combined with optical mapping yields completely finished fungal genome. *mBio*, 6, e00936-15.
- Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S, Knight R, Rohwer F, Jackson RB (2007) Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied and Environmental Microbiology*, 73, 7059–7066.
- Frank JA, Pan Y, Tooming-Klunderud A, Eijsink VGH, McHardy AC, Nederbragt AJ, Pope PB (2016) Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. *Scientific Reports*, 6, 25373
- Franzen O, Hu J, Bao X, Itzkowitz SH, Peter I, Bashir A (2015) Improved OTU-picking using long-read 16S rRNA gene amplicon sequencing and generic hierarchical clustering. *Microbiome*, 3, 43.
- Gloor GB, Hummelen R, Macklaim JM, Dickson RJ, Fernandes AD, MacPhee R, Reid G (2010) Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS ONE*, 5, e15406.
- Greninger AL, Naccache SN, Federman S, Yu GX, Mbala P, Bres V, Stryke D, Bouquet J, Somasekar S, Linnen JM, Dodd R, Mulembakani P, Schneider BS, Muyembe-Tamfum JJ, Stramer SL, Chiu CY (2015) Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Medicine*, 7, 99.
- Huang ZR, Hong JL, Xu JX, Li L, Guo WL, Pan YY, Chen SJ, Bai WD, Rao PF, Ni L, Zhao LN, Liu B, Lv XC (2018) Exploring core functional microbiota responsible for the production of volatile flavour during the traditional brewing of Wuyi Hong Qu glutinous rice wine. *Food Microbiology*, 76, 487–496.
- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hershberg AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF (2016) A new view of the tree of life. *Nature Microbiology*, 1, 16048.
- Hugenholtz P, Pitulle C, Hershberger KL, Pace NR (1998) Novel division level bacterial diversity in a Yellowstone hot spring. *Journal of Bacteriology*, 180, 366–376.
- Jain M, Koren S, Miga KH (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, 36, 338–345.
- Jonasson J, Monstein HJ (2002) Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16s rDNA fragments. *Apmis*, 110, 263–272.
- Liem M, Jansen HJ, Dirks RP, Henkel CV, van Heusden GPH, Lemmers R, Omer T, Shao S, Punt PJ, Spaik HP (2017) De novo whole-genome assembly of a wild type yeast isolate using nanopore sequencing. *F1000Research*, 6, 618.
- Liu H, Hu Z, Zhang Y, Zhang J, Xie H, Liang S (2018) Microbial nitrogen removal of ammonia wastewater in poly (butylenes succinate)-based constructed wetland: Effect of dissolved oxygen. *Applied Microbiology and Biotechnology*, 102, 9389–9398.
- Ludden C, Reuter S, Judge K, Gouliouris T, Blane B, Coll F, Naydenova P, Hunt M, Tracey A, Hopkins KL, Brown NM, Woodford N, Parkhill J, Peacock SJ (2017) Sharing of carbapenemase-encoding plasmids between Enterobacteriaceae in UK sewage uncovered by MinION sequencing. *Microbial Genomics*, 3, 1–12.
- Magi A, Semeraro R, Mingrino A, Giusti B, D'Aurizio R (2017) Nanopore sequencing data analysis: State of the art, applications and challenges. *Briefings in Bioinformatics*, 19, 1256–1272.
- Mardis ER (2008) Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, 9, 387–402.
- Metzker ML (2010) Sequencing technologies—The next generation. *Nature Reviews Genetics*, 11, 31–46.
- Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE (2011) Landscape of next-generation sequencing technologies. *Analytical Chemistry*, 83, 4327–4341.
- Payne A, Holmes N, Rakyan V, Loose M (2018) Whale watching with BulkVis: A graphical viewer for Oxford Nanopore bulk fast5 files. *bioRxiv*, doi: <https://doi.org/10.1101/312256>.
- Quick J, Loman NJ, Duraffour S, Simpson JT, Ettore S, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, Ouedraogo N, Afrough B, Bah A, Baum JHJ, Becker-Ziaja B, Boettcher JP, Cabeza-Cabrero M, Camino-Sanchez A, Carter LL, Doerrbecker J, Enkirch T, Garcia-Dorival I, Hetzelt N, Hinzmann J, Holm T, Kafetzopoulou LE, Koropogui M, Kosgey A, Kuisma E, Logue CH, Mazzarelli A, Meisel S, Mertens M, Michel J, Ngabo D, Nitzsche K, Pallasch E, Patrono LV, Portmann J, Repits JG, Rickett NY, Sachse A, Singethan K, Vitoriano I, Emanaberhan RLY, Zekeng EG, Racine T, Bello A, Sall AA, Faye O, Faye O, Magassouba N, Williams CV, Amburgey V, Winona L, Davis E, Gerlach J, Washington F, Monteil V, Jourdain M, Bererd M, Camara A, Somlare H, Camara A, Gerard M, Bado G, Baillet B, Delaune D, Nebie KY, Diarra A, Savane Y, Pallawo RB, Gutierrez GJ, Milhano N, Roger I, Williams CJ, Yattara F, Lewandowski K, Taylor J, Rachwal P, Turner DJ, Pollakis G, Hiscox JA, Matthews DA, O'Shea MK, Johnston AM, Wilson D, Hutley E, Smit E, DiCaro A, Wolfel R, Stoecker K, Fleischmann E, Gabriel M, Weller SA, Koivogui L, Diallo B, Keita S, Rambaut A, Formenty P, Gunther S, Carroll MW (2016) Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530, 228–232.
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, USA*, 74, 5463–5467.



- Schloss PD, Handelsman J (2005) Metagenomics for studying unculturable microorganisms: Cutting the Gordian knot. *Genome Biology*, 6, 229.
- Schmid M, Frei D, Patrignani A, Schlapbach R, Frey JE, Remus-Emsermann MNP, Ahrens CH (2018) Pushing the limits of de novo genome assembly for complex prokaryotic genomes harboring very long, near identical repeats. *Nucleic Acids Research*, 46, 8953–8965.
- Singer E, Bushnell B, Coleman-Derr D, Bowman B, Bowers RM, Levy A, Gies EA, Cheng JF, Copeland A, Klenk HP, Hallam SJ, Hugenholtz P, Tringe SG, Woyke T (2016) High-resolution phylogenetic microbial community profiling. *ISME Journal*, 10, 2020–2032.
- Smith AM, Jain M, Mulrone L, Garalde DR, Akeson M (2017) Reading canonical and modified nucleotides in 16S ribosomal RNA using nanopore direct RNA sequencing. *bioRxiv*, doi: <https://doi.org/10.1101/132274>.
- Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proceedings of the National Academy of Sciences, USA*, 103, 12115–12120.
- 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 C, 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 (2015) Structure and function of the global ocean microbiome. *Science*, 348, 1261359.
- Theuns S, Vanmechelen B, Bernaert Q, Deboutte W, Vandenhole M, Beller L, Matthijssens J, Maes P, Nauwynck HJ (2018) Nanopore sequencing as a revolutionary diagnostic tool for porcine viral enteric disease complexes identifies porcine kobuvirus as an important enteric virus. *Scientific Reports*, 8, 9830.
- Tsai YC, Conlan S, Deming C, Segre JA, Kong HH, Korch J, Oh J, Progra NCS (2016) Resolving the complexity of human skin metagenomes using single-molecule sequencing. *mBio*, 7, e01748-15.
- Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics*, 3, e000132.
- Woese CR (1987) Bacterial evolution. *Microbiological Reviews*, 51, 221–271.
- Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences, USA*, 74, 5088–5090.
- Yang CY, Targ DC (2018) Diet, gut microbiome and indoxyl sulphate in chronic kidney disease patients. *Nephrology*, 23, 16–20.
- Youssef NH, Couger MB, Struchtemeyer CG, Ligginstoffer AS, Prade RA, Najjar FZ, Atiyeh HK, Wilkins MR, Elshahed MS (2013) The genome of the anaerobic fungus *Orpinomyces* sp. strain C1A reveals the unique evolutionary history of a remarkable plant biomass degrader. *Applied and Environmental Microbiology*, 79, 4620–4634.
- Zhao L, Song Y, Li L, Gan N, Brand JJ, Song L (2018) The highly heterogeneous methylated genomes and diverse restriction-modification systems of bloom-forming *Microcystis*. *Harmful Algae*, 75, 87–93.

(特邀责任编辑: 周欣 责任编辑: 时意专)



•研究报告•

# 基于粪便DNA及宏条形码技术的 食肉动物快速调查及食性分析

邵昕宁<sup>1,2</sup> 宋大昭<sup>3</sup> 黄巧雯<sup>3</sup> 李 晟<sup>1,2</sup> 姚 蒙<sup>1,2\*</sup>

1 (北京大学生命科学学院, 北京 100871)

2 (北京大学生态研究中心, 北京 100871)

3 (中国猫科动物保护联盟, 北京 101011)

**摘要:** 在陆地生态系统中, 大型食肉动物对于稳定食物网结构和生态系统功能有重要作用。在世界范围内, 由于栖息地丧失和破碎化、猎杀、人类活动干扰以及病原体的传播, 大型食肉动物生存正面临严重威胁, 多种食肉动物地理分布范围及种群数量大幅度缩减。如何有效保护大型食肉动物物种多样性及种群已经成为世界关注的焦点问题和保护生物学的重要研究方向。川西高原地处我国西南山地与青藏高原东缘交界地带, 属于世界生物多样性热点地区, 是世界大型食肉动物物种最丰富的地区之一, 而日益增强的人类活动可能会加剧对当地动植物资源的破坏, 进而威胁野生食肉动物的生存。获得准确的物种多样性信息及食肉动物食性数据有助于深入了解该地区生态系统结构及食物网关系, 对研究物种共存机制及生物多样性保护有重要意义。本研究通过从四川甘孜藏族自治州新龙县和石渠县野外采集的食肉动物粪便样品中提取DNA, 利用DNA条形码进行物种鉴定, 快速获得该地区食肉动物物种构成信息。38份粪便样品经鉴定来自于7种食肉动物, 分别为5种大型食肉动物(狼*Canis lupus*、棕熊*Ursus arctos*、豹*Panthera pardus*、雪豹*P. unica*、狗*Canis lupus familiaris*)和2种中小型食肉动物(豹猫*Prionailurus bengalensis*、赤狐*Vulpes vulpes*)。进一步利用高通量测序和宏条形码技术对7种食肉动物粪便中的食物DNA进行精准食性分析, 得到包含19种哺乳类、8种鸟类和1种鱼类共计28个不同的食物分子可操作分类单元(molecular operational taxonomic unit, MOTU)。结果显示, 狼、狗、棕熊最主要的食物来源为偶蹄目动物, 其中取食频率最高的物种为家牦牛(*Bos grunniens*); 而豹猫和赤狐食物中小型哺乳动物如啮齿目和兔形目占重要比例, 其中高原松田鼠(*Neodon irene*)和高原鼠兔(*Ochotona curzoniae*)被取食频率最高。豹和雪豹的食物分别为偶蹄目的中华斑羚(*Naemorhedus griseus*)和岩羊(*Pseudois nayaur*)。本研究显示了粪便DNA及宏条形码技术在食肉动物多样性快速调查及高通量精确食性分析中的应用前景, 并为此类研究提供了技术路线的有力借鉴。

**关键词:** 生物多样性编目; 兽类调查; 非损伤性取样; DNA条形码; 食物网

## Fast surveys and molecular diet analysis of carnivores based on fecal DNA and metabarcoding

Xinning Shao<sup>1,2</sup>, Dazhao Song<sup>3</sup>, Qiaowen Huang<sup>3</sup>, Sheng Li<sup>1,2</sup>, Meng Yao<sup>1,2\*</sup>

1 School of Life Sciences, Peking University, Beijing 100871

2 Institute of Ecology, Peking University, Beijing 100871

3 Chinese Felid Conservation Alliance, Beijing 101011

**Abstract:** Large carnivores play an important role in the regulation of food-web structure and ecosystem functioning. However, large carnivores face serious threats that have caused declines in their populations and geographic ranges due to habitat loss and degradation, hunting, human disturbance and pathogen transmission. Conservation of large carnivore species richness and population size has become a pressing issue and an important research focus of conservation biology. The western Sichuan Plateau, located at the intersection of the mountains of southwest China and the eastern margin of the Tibetan Plateau, is a global biodiversity hotspot and has high carnivore species richness. However, increasing human activities may

exacerbate the destruction of local flora and fauna, thereby threatening the survival of wild carnivores. Information on species composition and dietary habits can improve our understanding of the structure and function of the ecosystem and food-web relationships in the study area. In addition, species composition and dietary habits are of great significance for understanding multi-species coexistence mechanisms and preserving biodiversity. This study collected carnivore fecal samples from Xinlong and Shiqu counties in the Ganzi Tibetan Autonomous Prefecture, Sichuan Province. DNA was then extracted from the samples and the species was identified based on DNA sequences and DNA barcoding techniques. Seven carnivores were identified, including five large carnivores (*Canis lupus*, *Ursus arctos*, *Panthera pardus*, *P. uncia* and *Canis lupus familiaris*) and two medium and small-sized carnivores (*Prionailurus bengalensis* and *Vulpes vulpes*). Using fecal DNA, high-throughput sequencing and metabarcoding, we conducted diet analysis for the seven carnivores and found 28 different food molecular operational taxonomic units (MOTUs), including 19 mammals, eight birds and one fish species. The predominant prey categories of wolves, dogs and brown bears were ungulates. The domestic yak (*Bos grunniens*) was the most frequently identified prey species. Small mammals such as rodents and lagomorphs accounted for a significant proportion in the diets of leopard cats and red foxes, The most frequent prey of this category of carnivore were the Chinese scrub vole (*Neodon irene*) and plateau pika (*Ochotona curzoniae*). In addition, leopards and snow leopards mainly fed on the Chinese goral (*Naemorhedus griseus*) and blue sheep (*Pseudois nayaur*), respectively. Our study highlights the utility of fecal DNA and metabarcoding technique in fast carnivore surveys and high-throughput diet analysis, and provides a technical reference and guidance for future biodiversity surveys and food-web studies.

**Key words:** biodiversity inventory; mammalian survey; non-invasive sampling; DNA barcoding; food-web

生物多样性是维持生态系统稳定性的基础,也是影响生态系统生产力、抗干扰能力及恢复能力等功能的重要因素(Hooper et al, 2005; Cardinale et al, 2012)。在衡量生物多样性的三个层次即遗传多样性、物种多样性和生态系统多样性中,物种多样性是对一个自然地理分布区域生物多样性情况最直观反映,物种的组成和结构对种群动态及生态系统功能有重要决定作用。在一个生态系统中共存的各种物种通过直接或间接的取食与被取食关系交织形成食物网,不同物种间的复杂相互作用促进生态系统功能的正常运转,进而维持着生态系统的稳定性和物种的多样性(Montoya et al, 2003)。对世界不同地区和多种生态系统类型的研究普遍显示,物种多样性的丧失是生态环境退化和生态系统失衡的主要因素(Hooper et al, 2012)。因此,对物种多样性及其相互作用的了解是认识生态系统过程的重要前提,也是保护生态环境和生物多样性的基础。

大型哺乳类食肉动物(以下简称大型食肉动物)处于陆地生态系统食物链的顶端,通过捕食、竞争、威慑和食物网级联效应调控大型食草动物、中型食肉动物及其他食物网成员的种群数量和动态,深刻影响食物网结构及生态系统功能(Wang & Macdonald, 2009; Wang et al, 2014)。多项研究均显示,大型食肉动物从原来的生物群落中消失后会引发一

系列下层营养级动物种群乃至植被和无机环境的重大变化,对生态系统产生不可逆转的影响(Ripple et al, 2014)。在世界范围内,由于一个多世纪以来人口的迅速增长和生产生活范围扩大,大型食肉动物原生栖息地丧失和破碎化,再加上捕猎以及细菌、病毒等病原体的传播,目前其种群生存普遍面临严重的威胁,多种食肉动物数量大幅度减少,分布的地理范围急剧收缩(Ripple et al, 2014)。如何有效保护大型食肉动物物种多样性及种群数量已经成为世界关注的焦点问题和保护生物学的重要研究方向。

我国西南山地北部与青藏高原东缘交界地带的川西高原属世界生物多样性热点地区之一(Myers et al, 2000),因其特殊的自然历史、地理条件及气候特征,为多种野生动物提供了适宜的栖息地环境,其哺乳类和鸟类数量分别占全国的20%和33%,有12种兽类和10种鸟类被列为国家一级重点保护野生动物,23种兽类和49种鸟类被列为国家二级重点保护野生动物。该区域也是世界范围内大型食肉动物物种最为丰富的地区之一(Ripple et al, 2014)。川西高原独特的自然地理环境和丰富的生物多样性使其具有极高的生态学研究和保护价值,而目前对该地区物种分布情况及生态系统过程了解甚少。获得准确的物种多样性信息及食肉动物食性数据有



助于深入了解该地区生态系统结构及食物网关系,对研究物种共存机制及生物多样性保护有重要意义。与低海拔的热带和亚热带森林生态系统相比,由于气候条件恶劣、水热资源缺乏、生物量和生产力低,高原生态系统更为脆弱(Cole & Landres, 1996);加之该地区人类活动不断加强,有可能加剧对当地动植物资源的破坏(Baker et al, 2008),因此对这一生态系统的研究和保护迫在眉睫。

物种和种群调查是有效开展濒危物种保护与管理工作的前提和基础(Bellemain et al, 2010)。由于食肉目动物活动范围广,多夜间行动且警觉性高,在野外很难对其进行直接观测,使对此类动物的调查和保护工作更加困难(Pelton et al, 2003)。传统的食肉动物物种识别方法主要有足迹识别法、毛发形态识别法、咬痕识别法等依据形态学和解剖学特征的鉴定方法(周用武和杨玉华, 2009; 张帅等, 2013)。由于近缘物种形态相似度高,而同种动物受生境条件、发育阶段及性别等因素的影响可能存在个体形态差异,此外还容易受到研究者经验等主观因素的干扰,因此形态学鉴定方法通常存在定种困难、错误率高、效率低、结果难以验证等问题(程希婷等, 2011)。

动物的捕食行为是重要的生态过程,对食性的分析及食物网络的研究可以直接体现出生态群落的功能和结构,揭示各级消费者之间的能量流动和物质利用转换,对捕食者在群落中的作用有更深理解,是生态系统研究的重要组成部分(王雪芹等, 2015),对动物食性的了解还可以为保护生态系统功能及物种多样性提供科学依据。对野生动物食性的研究方法传统上主要依赖于形态学分析,包括直接观察法、胃内容物分析法、粪便显微镜观察法等(尹华宝等, 2008)。但这些方法受限于食物类型和消化情况,并且可能存在多种形态相似难以区分的近缘种,需要分析人员有丰富的生物学分类经验,鉴定结果的效率和准确性较低(Pompanon et al, 2012)。

近几十年来分子生物学和生物信息学技术的飞速发展,推动DNA分析成为生态学和保护生物学研究的重要工具。尤其是DNA条形码(DNA barcoding)技术(即利用具有种内特异性和种间变异性的相对较短的标准DNA片段,建立一种新的物种身份识别系统),可以对物种进行快速、准确的识别和鉴定,成为物种多样性调查的有力手段(Hebert et al,

2003; Valentini et al, 2009; Kress et al, 2015)。二代测序技术的不断进步使得高通量、低成本地获取大量样品的DNA序列信息成为可能。DNA宏条形码(metabarcoding)技术就是利用高通量测序技术大量获得混合样本中多种条形码基因扩增子序列,继而通过生物信息学手段分析实现分类单元鉴定的方法(Creer et al, 2010; Porazinska et al, 2010; Voolstra et al, 2011)。DNA宏条形码以高通量、低成本、快速鉴定大量物种的优势,极大地拓展了DNA条形码应用的范围(Taberlet et al, 2012; 马兰, 2013)。在生态学领域, DNA宏条形码已广泛应用于生物多样性分析、食性分析、古生物研究、生态评估等多方面(Pompanon et al, 2012; Yoccoz et al, 2012; Murray et al, 2013; Kelly et al, 2014)。基于DNA宏条形码的分子食性分析方法通过从样品(如粪便)中提取DNA,设计针对某一种、某几种物种或某一类群的引物,利用聚合酶链式反应(PCR)扩增特定的条形码序列,并通过高通量测序和序列比对来识别食物种类(Valentini et al, 2009)。与基于显微观察的食性分析方法相比,这种方法的优势在于:(1)不依赖于研究者的形态学经验和主观判断;(2)对形态相似的近缘种和隐存种可以准确区分;(3)不受食物消化程度和硬质残留的局限,可以分析几乎所有食物类型(如肌肉、内脏等);(4)具有极高的灵敏度,微量的食物DNA也可以被检测到;(5)能对混合DNA样品进行测序且通量高,大大降低了分析成本和时间(Symondson, 2002; Valentini et al, 2009; Kircher & Kelso, 2010)。

本研究拟通过基于野外采集食肉动物粪便DNA的物种鉴定,获得位于我国西南山地的四川甘孜藏族自治州的食肉动物物种多样性初步信息,并基于高通量测序和DNA宏条形码技术,进行食肉动物精准食性分析。这些信息有助于深入了解研究地区生态系统食物网结构及功能,对研究物种共存机制及生物多样性保护有重要意义。同时本研究使用的粪便DNA及宏条形码技术为脊椎动物多样性快速调查及高通量精确食性分析提供了有力借鉴。

## 1 研究方法

### 1.1 研究区域

甘孜藏族自治州(以下简称甘孜州)位于青藏高原东南缘四川省西南部,介于97°22'–102°29' E、

27°58′–34°20′ N之间, 面积153,000 km<sup>2</sup>, 辖18个县。全州地貌大致分为: 山地、丘状高原、山原、高山峡谷等4种类型, 平均海拔3,500 m以上。该地区纬度位置属于亚热带气候区, 由于受到青藏高原复杂地形的影响, 境内气候呈现出青藏高原山地气候和大陆性气候特征, 属大陆性季风高原型气候。全州年平均气温8–10℃, 日照多、辐射强, 昼夜温差大, 干湿季明显, 年降水量300–600 mm, 水热同期。甘孜州总林地面积17,200 km<sup>2</sup>, 林草覆盖率达83%。

## 1.2 样品采集与保存

本研究粪便样品于2016年9–10月采集自甘孜州新龙县及石渠县境内(图1)。由调查人员沿野外样线进行搜索, 发现新鲜食肉动物粪便时, 首先用GPS手持机(Garmin 60CXS)记录采集时间及地理信息, 然后用一次性PE手套拾取单个粪便样品装入50 mL的无菌样品管中, 并在管盖和管身记录样品编号、详细地名及根据粪便形态鉴定的物种。同时填写“野外动物样品记录卡”, 包括编号、样品类型、生境类型等详细信息。在每次拾取样品时更换新的手套以避免交叉污染。当日将采集的样品浸没在无水乙醇中, 24 h后将乙醇倒掉, 倒入硅胶干燥保存。

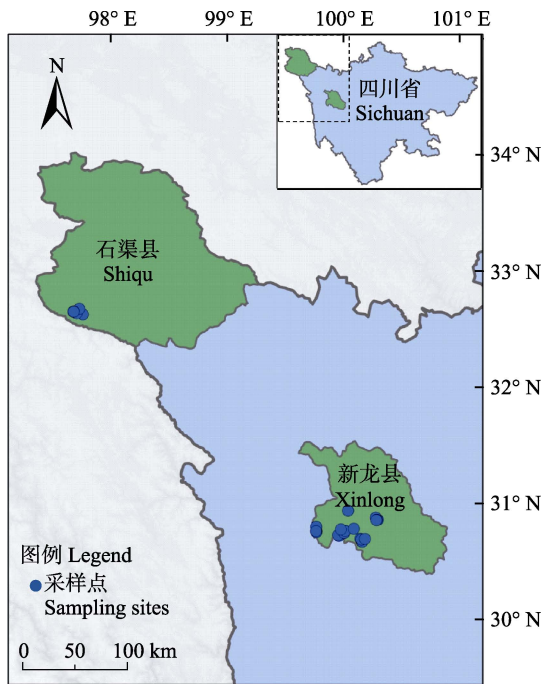


图1 四川省甘孜藏族自治州石渠县及新龙县采样地点示意图

Fig. 1 Map of sampling area in Shiqu and Xinlong counties of Ganzi Tibetan Autonomous Prefecture, Sichuan

## 1.3 粪便DNA提取、扩增及物种鉴定

本研究使用适用于食肉动物粪便的2CTAB-PCI法(Vallet et al, 2007)提取和纯化DNA, 并根据食肉动物粪便特点进行了若干改进(Xiong et al, 2017), 取样时从粪便内部多点夹取, 粪便总量不超过500 μL, 并捣碎均匀, 使其能充分反应, 提取过程中用CTAB及淀粉去除DNA中的PCR抑制物, 并用苯酚氯仿多次抽提DNA以达到提纯DNA的目的。

首先对每份样品的来源物种(即捕食者)进行分子物种鉴定。由于在粪便DNA中食物DNA降解较严重, 难以扩增长片段, 而捕食者自身DNA相对质量较好, 故首先使用脊椎动物通用引物16S-F/R(Xiong et al, 2016)(附录1)扩增长380 bp左右的捕食者线粒体16S片段。每个样品PCR总体积为30 μL, 包含DNA (10–100 ng/μL) 6 μL、2 × EasyTaq PCR SuperMix (全式金, AS111) 15 μL、正反向引物各0.2 μM及0.4 mg/mL牛血清白蛋白(BSA)。PCR反应程序为95℃预变性5 min; 然后95℃变性30 s, 55℃退火30 s, 72℃延伸30 s, 进行35个循环; 最后72℃延伸10 min。纯化后的PCR产物使用ABI PRISM3730xl测序仪(Applied Biosystems, Foster City, California)进行测序。测序结果通过软件Chromas v2.22 (Technelysium Pty Ltd., Australia)处理, 得到长度约320 bp的单一序列。然后用NCBI (National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov>)的BLAST程序检索测序片段在其数据库GenBank中的匹配序列。最佳匹配序列与待测序列的覆盖度(coverage)不低于95%, 一致度(identity)大于95%时认为检索结果有效。结合当地动物分布信息, 对序列进行筛选, 当覆盖度达到100%、一致度达到98%及以上, 且对应物种可能分布在研究地点时, 认为粪便样品来自匹配序列对应的物种。若存在不止一种物种的匹配序列, 则根据研究地点本底信息及物种的分布特征排除不符合的物种。当测序峰图正常、比对结果覆盖度达到100%而一致度小于98%时, 可能是GenBank数据库未收录该待测物种的序列, 则可根据当地物种分布情况, 推测可能的近缘物种或认为物种鉴定失败。

## 1.4 高通量测序及数据处理

确定粪便来源物种后, 进一步针对不同食肉动物进行基于DNA宏条形码的分子食性分析。粪便样

品中残存的被捕食者DNA较捕食者自身DNA量少且破碎化严重, 所以为了保证被捕食者DNA的成功扩增, 食性分析通用引物的扩增产物长度不宜超过150 bp。选用Riaz等(2011)针对脊椎动物设计的12SV5-F/R通用引物(附录1), 并在正反向引物5'端增加7个核苷酸的引物标签(Coissac, 2012)进行PCR扩增, 产物长度约140 bp。每个PCR采用具有不同标签序列的引物扩增, 后续高通量测序结果根据标签进行样品区分。该引物用ecoPCR软件通过数据库模拟PCR及多种食肉动物粪便样品进行验证, 通用性广且具有较高的脊椎动物物种分辨率, 对啮齿目、兔形目和鸡形目的物种鉴定表现佳(Shehzad et al, 2012a, b; Xiong et al, 2017)。粪便样品总DNA中同时包括捕食者与被捕食者DNA, 使用脊椎动物通用引物能够同时扩增二者的序列, 由于捕食者自身DNA占较多量, 扩增时会与被捕食者DNA竞争引物而影响食物DNA的扩增效率及成功率(Pompanon et al, 2012)。本文使用文献报道的豹猫(*Prionailurus bengalensis*)序列特异性阻抑引物(Shehzad et al, 2012a), 并自行设计了其他食肉动物物种的阻抑引物(附录1), 通过将各阻抑引物与当地主要物种的相应区段进行比较, 可能的被捕食者均与这段阻抑引物序列具有多于10个核苷酸位点的差异。阻抑引物浓度过高会影响PCR产物中被捕食者序列的扩增(Vestheim & Jarman, 2008)。经过前期预实验, 最终确定在反应体系中阻抑引物与通用引物的工作浓度比为10:1。对28份鉴种成功的食肉动物粪便每个样品的DNA进行3次PCR平行对照反应, 每个反应分别使用不同标签的通用引物加以区分。每个样品PCR总体积为30  $\mu$ L, 包含DNA (10–100 ng/ $\mu$ L) 6  $\mu$ L、2  $\times$  EasyTaq PCR SuperMix 15  $\mu$ L、正反向引物各0.2  $\mu$ M、阻抑引物2  $\mu$ M及0.4 mg/mL BSA。PCR反应程序为95 $^{\circ}$ C预变性5 min; 然后95 $^{\circ}$ C变性30 s, 60 $^{\circ}$ C退火30 s, 进行35个循环; 最后72 $^{\circ}$ C延伸10 min。为检验本底污染和交叉污染情况, 每批次扩增中包括DNA提取空白对照和不加模板的PCR空白对照反应, 共3个DNA提取空白对照和21个PCR空白对照。全部样品及对照反应共计108个PCR反应。

将所有PCR产物进行纯化并混合, 在北京诺和致源生物科技有限公司进行高通量测序。测序采用Paired-end法, 使用Illumina HiSeq2500测序平台(Illumina Inc., San Diego, CA, USA), 单端读段长度

为150 bp。高通量数据的处理参考网站<http://metabarcoding.org/obitools/doc/wolves.html>的说明, 用OBITools1.01.22程序包(<http://metabarcoding.org/obitools/doc>)经过拼接、标签识别、冗余序列清除、降噪、建立参考数据库、序列比对及分类等步骤, 进行序列的整理与分析。

当某一序列在本底数据库中最佳匹配序列多于1条时, OBITools中的程序会自动将物种鉴定结果记为能够涵盖所有匹配物种的最低分类单元, 但不能加入研究地物种分布信息的参考。结果中还可能不存在污染及PCR错误, 或者未能确定物种的序列, 所以需要在OBITools输出结果的基础上进行人工排查, 必要时用BLAST程序检索目标片段的匹配物种。

序列的物种鉴定原则如下:

(1)与单一物种比对的一致度  $\geq 99\%$ , 且符合物种分布记录, 记为该物种; 如发生与单一物种比对一致度  $\geq 99\%$ 但该物种非当地物种时, 则记为该物种在当地有分布的最近缘物种。

(2)与多个物种比对的一致度  $\geq 99\%$ 时, 首先排除当地无分布的物种, 如果仍不能确定为单一物种, 则记为能够涵盖所有一致度  $\geq 99\%$ 的当地物种的最低分类单元。

(3)比对结果中, 最高一致度  $< 99\%$ 且  $\geq 95\%$ 时, 一致度最高序列仅有1条, 且与得分次之的序列一致度差距  $\geq 2\%$ 时, 记为一致度最高序列的上一级分类单元。最高一致度对应有多个物种时, 则记为能够涵盖所有一致度最高的当地物种的最低分类单元。

(4)如果最高一致度  $\geq 95\%$ 范围内仅对应1种非当地物种, 则记为当地有记录的该物种最近缘种。

(5)比对结果中, 最高一致度  $< 95\%$ 时, 认为无法判定该序列的分类单元, 记为“未知”。

依据上述原则对每一序列进行物种鉴定与分类, 物种分布信息参考《中国兽类野外手册》(Smith和焱, 2009)。由于条形码分辨率及数据库的局限, 不是每条序列的识别都能达到物种水平, 所以我们使用分子可操作分类单元(molecular operational taxonomic unit, MOTU)的概念来对不同食物类群进行区分(Pompanon et al, 2012)。

### 1.5 食性分析

首先, 对食物MOTU进行筛选与整理。当单一PCR产物中某食物MOTU序列数少于同批PCR空白



对照或DNA提取空白对照,或单一PCR产物中某食物MOTU序列数少于该PCR总序列数的1%时,认为该序列可能来源于交叉污染,该MOTU不计入相应样品的食物结果。在每个样品中保留至少在2个PCR平行对照中出现的MOTU。序列处理中舍弃与GenBank比对后最高一致度低于95%的序列,并合并彼此相差小于2% (即相差1个核苷酸)的序列。物种鉴定时出现的食肉动物序列及可能来自实验污染的人的序列不计入食物MOTU。经过以上筛选和处理后,在每个样品中每一MOTU出现赋值为1,反之记为0,统计各食物MOTU的出现频次。

使用%FO和%RO两个指标估计各食肉动物的食性构成(Xiong et al, 2017)。对某一食肉物种而言,%FO是指含某食物MOTU的样品数占总样品数的百分比,计算公式如下:

$$\%FO_i = (N_i/N) \times 100\% \quad (1)$$

其中, $N_i$ 表示出现了第*i*类食物MOTU的样品数(次数), $N$ 是有效样品总数。

%RO是某食物MOTU的出现频次占所有食物MOTU出现总频次的百分比,计算公式如下:

$$\%RO_i = (N_i/\sum N_i) \times 100\% \quad (2)$$

其中, $N_i$ 含义同上, $\sum N_i$ 表示该物种所有食物MOTU出现次数之和。

单一物种全部粪便样品中所有食物MOTU的%RO之和为1,而当存在1个或多个样品中检出不止1种食物MOTU时,该物种的%FO之和大于1。%FO表示某个食物MOTU在该物种食性中出现的常见程度,%RO则侧重显示该食物MOTU在食性中所占份额和重要性(Wang et al, 2014; Gómez-Ortiz et al, 2015; Kasper et al, 2016)。

## 2 结果

### 2.1 样品采集及物种鉴定

于2016年秋季在石渠县(6份)和新龙县(32份)共采集38份食肉动物粪便样品,分别进行DNA提取,其中28份可以被通用引物16S-F/R扩增并成功判断出样品来源物种,物种鉴定成功率为73.7%。其余10份样品未能成功获得扩增产物进行测序及物种鉴定。

共鉴定出7种食肉动物(表1),包括狼(*Canis lupus*,  $N = 8$ )、狗(*C. lupus familiaris*,  $N = 6$ )、豹猫( $N = 6$ )、赤狐(*Vulpes vulpes*,  $N = 3$ )、棕熊(*Ursus*

*arctos*,  $N = 3$ )、豹(*Panthera pardus*,  $N = 1$ )和雪豹(*P. uncia*,  $N = 1$ )。

28份物种鉴定成功的样品中,有22份在样品采集时由野外工作人员根据粪便形态进行了物种预判,其中仅有8份样品预判物种与分子鉴定结果相同,且各食肉动物粪便的预判成功率各不相同,总体预判正确率仅为36.4%(表1)。

### 2.2 食性分析

高通量测序结果经过OBITools软件的过滤和降噪,并对这些序列进一步进行人工筛选和处理。序列分析结果显示所有28份样品均得到有效食性数据,共计28个不同的食物MOTU。其中17个MOTU鉴定到物种水平,占MOTU总数的60.7%,其他MOTU分别被鉴别至属、科及目水平(表2)。

28个食物MOTU包括19种哺乳类、8种鸟类和1种鱼类,分属于10个目(表3)。大型食肉动物狼、狗和棕熊的食物中偶蹄目均占最大比例,%RO分别为48%、47%和40%(图2);而狼和狗又体现了更多的食物多样性和相似性,家牦牛(*Bos grunniens*)是它

表1 物种分子鉴定结果及形态物种预判正确率

Table 1 Summary of species identification by DNA barcoding and fecal morphology

物种 Species	样品数 No. of sample	物种预判 样品数 No. of morphologically identified sample	预判正确 样品数 No. of correctly identified sample	预判正确率 Accuracy of morphological identification (%)
狼 <i>Canis lupus</i>	8	4	1	25.0
狗 <i>Canis lupus familiaris</i>	6	6	1	16.7
豹猫 <i>Prionailurus bengalensis</i>	6	4	4	100
赤狐 <i>Vulpes vulpes</i>	3	3	0	0
棕熊 <i>Ursus arctos</i>	3	3	0	0
豹 <i>Panthera pardus</i>	1	1	1	100
雪豹 <i>Panthera uncia</i>	1	1	1	100
合计 Total	28	22	8	36.4

表2 食物分子可操作分类单元(MOTU)中不同分类级别的数目及比例

Table 2 Number of prey MOTUs assigned to various taxonomic levels

分类等级 Taxonomic level	种 Species	属 Genus	科 Family	目 Order	总计 Total
MOTU数目 MOTU no.	17	8	2	1	28
比例 Percentage (%)	60.7	28.6	7.1	3.6	



们食物中出现频率最高的物种, %FO分别为75%和67%, 除偶蹄目之外, 它们都取食啮齿目、兔形目、鲤形目和鸡形目动物(图2, 3)。除偶蹄目外, 棕熊较多以食肉目的猪獾(*Arctonyx albogularis*)作为食物(表3)。中小型食肉动物豹猫和赤狐的食物中小型兽类如啮齿目和兔形目占主要比例, 其中出现频率最高的是啮齿目, 在豹猫和赤狐食性构成中%RO分

别为34%和38%, 兔形目其次, %RO分别为27%和25% (图2)。其中高原松田鼠(*Neodon irene*)和高原鼠兔(*Ochotona curzoniae*)是豹猫和赤狐主要取食的啮齿类和兔形目食物(表3)。与赤狐相比, 豹猫的食物多样性更高, 它还会较多取食如高原兔(*Lepus oiostolus*)、姬鼠属(*Apodemus*)、白腹鼠属(*Niviventer*)、鼯鼠属(*Petaurista*)等, 同时鼯鼠属(*Sorex*)只在豹猫

表3 7种食肉动物食物分子可操作分类单元(MOTU)分子鉴定结果。最佳匹配物种结果中斜线(/)用于分隔序列一致度相同的物种, 竖线(|)用于分隔一致度相差 ≥ 1%的物种。

Table 3 Summary of prey MOTUs of seven carnivores identified by molecular dietary analysis. Species with the identical sequence identity are separated by “/”, and species with ≥ 1% difference in sequence identity are separated by “|”.

食物 MOTU Food MOTU		在不同食肉动物样品中的出现频次 Occurrence frequency in different carnivore samples							GenBank 最佳匹配物种 Best match species in GenBank			
拉丁学名 Scientific name	MOTU 名称 MOTU name	狼 Wolf	狗 Dog	豹猫 Leopard cat	棕熊 Brown bear	赤狐 Red fox	豹 Leopard	雪豹 Snow leopard	中文名 Chinese name	拉丁名 Scientific name	最高一致度 Best identity (%)	序列号 Accession no.
		N = 8	N = 6	N = 6	N = 3	N = 3	N = 1	N = 1				
<b>偶蹄目 Artiodactyla</b>												
<i>Bos grunniens</i>	家牦牛	6	4	6	1	0	0	0	家牦牛	<i>Bos grunniens</i>	100	KX232527.1
<i>Elaphodus cephalophus</i>	毛冠鹿	2	2	3	0	0	0	0	毛冠鹿	<i>Elaphodus cephalophus</i>	99	DQ873526.1
<i>Sus scrofa</i>	野猪	2	2	3	0	0	0	0	野猪	<i>Sus scrofa</i>	100	KX886757.1
<i>Pseudois nayaur</i>	岩羊	2	0	0	0	1	0	1	岩羊	<i>Pseudois nayaur</i>	100	KP998469.1
<i>Moschus</i>	麝属	2	0	1	0	0	0	0	马麝/林麝	<i>Moschus chryso-gaster/ Moschus berezovskii</i>	100	KP684123.1/AY184425.1
<i>Przewalskium albirostris</i>	白唇鹿	1	0	0	1	0	0	0	白唇鹿	<i>Przewalskium albirostris</i>	100	JN632690.1
<i>Naemorhedus griseus</i>	中华斑羚	0	0	0	0	0	1	0	中华斑羚	<i>Naemorhedus griseus</i>	100	JN632664.1
<b>奇蹄目 Perissodactyla</b>												
<i>Equus caballus</i>	马	1	0	0	0	0	0	0	马	<i>Equus caballus</i>	100	KU575247.1
<b>兔形目 Lagomorpha</b>												
<i>Ochotona curzoniae</i>	高原鼠兔	2	3	5	0	2	0	0	高原鼠兔	<i>Ochotona curzoniae</i>	100	KM225732.1
<i>Lepus oiostolus</i>	高原兔	1	0	6	0	0	0	0	高原兔	<i>Lepus oiostolus</i>	99	AY745187.1
<b>啮齿目 Rodentia</b>												
<i>Neodon irene</i>	高原松田鼠	4	1	5	0	2	0	0	高原松田鼠	<i>Neodon irene</i>	100	HQ416908.1
<i>Apodemus</i>	姬鼠属	2	0	5	0	0	0	0	大耳姬鼠 高山姬鼠	<i>Apodemus latronum/ Apodemus chevrier</i>	100 99	HQ333256.1 HQ896683.1
<i>Marmota himalayana</i>	喜马拉雅旱獭	2	0	0	0	0	0	0	喜马拉雅旱獭	<i>Marmota himalayana</i>	100	JX069958.1
<i>Niviventer excelsior</i>	川西白腹鼠	0	1	1	0	0	0	0	川西白腹鼠	<i>Niviventer excelsior</i>	99	JQ927552.1
<i>Trogopterus xanthipes</i>	复齿鼯鼠	0	0	1	0	0	0	0	复齿鼯鼠	<i>Trogopterus xanthipes</i>	97	AY227546.1
<i>Petaurista</i>	鼯鼠属	0	0	1	0	0	0	0	红背鼯鼠 霜背大鼯鼠 红白鼯鼠	<i>Petaurista petaurista/ Petaurista philippensis/ Petaurista alborufus</i>	98 97	KP973556.1 KP973555.1 AY227541.1
<i>Eozapus setchuanus</i>	林跳鼠	0	0	0	0	1	0	0	林跳鼠	<i>Eozapus setchuanus</i>	98	KJ648495.1

表 3 (续) Table 3 (continued)

食物MOTU Food MOTU		在不同食肉动物样品中的出现频次 Occurrence frequency in different carnivore samples							GenBank最佳匹配物种 Best match species in GenBank			
拉丁学名 Scientific name	MOTU名称 MOTU name	狼 Wolf	狗 Dog	豹猫 Leopard cat	棕熊 Brown bear	赤狐 Red fox	豹 Leopard	雪豹 Snow leopard	中文名 Chinese name	拉丁名 Scientific name	最高一致度 Best identity (%)	序列号 Accession
		N = 8	N = 6	N = 6	N = 3	N = 3	N = 1	N = 1				
<b>食肉目 Carnivora</b>												
<i>Arctonyx albobularis</i>	猪獾	0	0	0	2	0	0	0	猪獾	<i>Arctonyx albobularis</i>	100	HM106329.1
<b>鼯形目 Soricomorpha</b>												
<i>Sorex</i>	鼯属	0	0	1	0	0	0	0	纹背鼯/小纹背鼯/长爪鼯/普通鼯	<i>Sorex cylindricauda</i> / <i>Sorex bedfordiae</i> / <i>Sorex unguiculatus</i> / <i>Sorex araneus</i>	99/ 98	KF696672.1/ GU981054.1/ KX754508.1/ KT210896.1
<b>鲤形目 Cypriniformes</b>												
Cyprinidae	鲤科	2	1	0	1	0	0	0	汪氏近红鲃/厚颌鲂	<i>Ancherythroculter wangi</i> / <i>Megalobrama pellegrini</i>	100	MG783573.1/ JX242529.1
<b>鸡形目 Galliformes</b>												
<i>Crossoptilon crossoptilon</i>	白马鸡	1	1	0	0	1	0	0	白马鸡	<i>Crossoptilon crossoptilon</i>	100	KP259808.1
<i>Gallus gallus</i>	家鸡	1	0	0	0	0	0	0	原鸡	<i>Gallus gallus</i>	100	KX987152.1
<i>Perdix</i>	山鹧属	0	0	0	0	1	0	0	斑翅山鹧/高原山鹧/灰山鹧	<i>Perdix dauurica</i> / <i>Perdix hodgsoniae</i> / <i>Perdix perdix</i>	100	KY411596.1/ KF027440.1/ KF781322.1
<b>雁形目 Anseriformes</b>												
<i>Anas</i>	鸭属	0	1	0	0	0	0	0	绿头鸭/琵嘴鸭/绿翅鸭/斑嘴鸭/针尾鸭	<i>Anas platyrhynchos</i> / <i>Anas clypeata</i> / <i>Anas crecca</i> / <i>Anas poecilorhyncha</i> / <i>Anas acuta</i>	100	KX592536.1/ KT345702.1/ KC771255.1/ KC466567.1/ KF312717.1
<b>雀形目 Passeriformes</b>												
<i>Phylloscopus</i> 1	柳莺属1	0	0	1	0	0	0	0	暗绿柳莺/淡眉柳莺	<i>Phylloscopus trochiloides</i> / <i>Phylloscopus humei</i>	99/ 98	KP267717.1/ KP267716.1
<i>Phylloscopus</i> 2	柳莺属2	0	1	0	0	0	0	0	褐柳莺/黄腰柳莺/乌嘴柳莺	<i>Phylloscopus fuscatus</i> / <i>Phylloscopus proregulus</i> / <i>Phylloscopus maculipennis</i>	100	JF505332.1/ AY635103.1/ /AY635100.1
Muscicapidae 1	鹟科	0	0	1	0	0	0	0	鹟/北红尾鹟	<i>Copsychus saularis</i> / <i>Phoenicurus aureoreus</i>	99	KU058637.1/ KF997864.1/ KF997863.1
Passeriformes	雀形目	0	0	1	0	0	0	0	家燕/红喉歌鸲	<i>Hirundo rustica</i> / <i>Luscinia calliope</i>	96	KX398931.1/ HQ690246.1

食物中出现(图3); 赤狐除捕食小型哺乳类外还取食鸡形目动物, 如白马鸡(*Crossoptilon crossoptilon*)和山鹧属(*Perdix*)。与狼和狗相比, 豹猫和赤狐食物中各目动物的出现比例更为平均。豹和雪豹的样品仅各1份, 各只鉴定出1种食物, 豹的食物为中华斑

羚(*Naemoredus griseus*), 雪豹的食物为岩羊(*Pseudois nayaur*)。食肉动物物种间食物重叠情况见图3。

此外, 家畜(家牦牛、马*Equus caballus*)在4种食肉动物中出现, 且狼、狗和棕熊取食家畜的%RO都在20%以上。

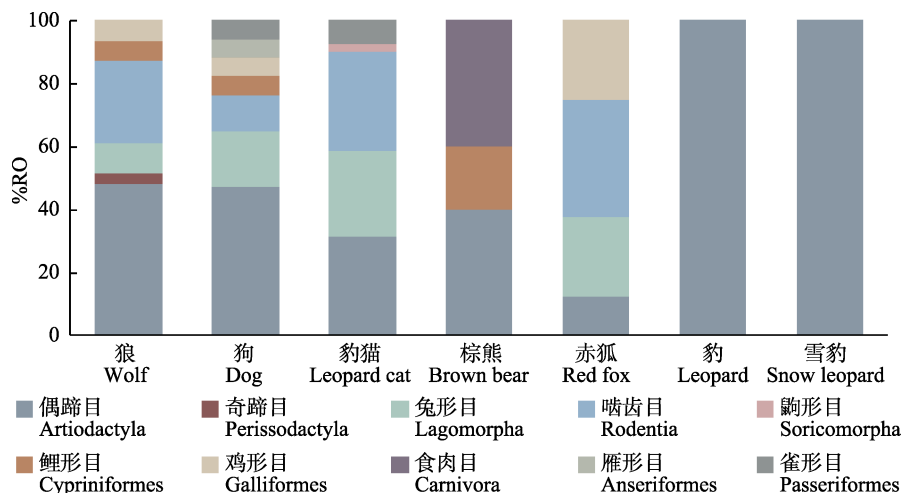


图2 目水平上不同食肉动物食物组成。%RO为不同食物分子可操作分类单元(MOTU)出现的相对频率。  
 Fig. 2 Dietary compositions by prey order identified in the diets of different carnivores. %RO indicates the relative frequency of occurrence of different food MOTUs.

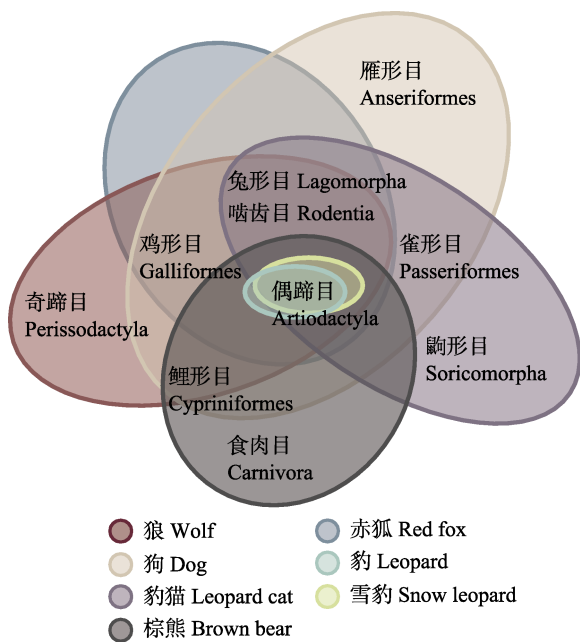


图3 食肉动物食性重叠情况  
 Fig. 3 Dietary overlap among carnivores

### 3 讨论

#### 3.1 食肉动物物种鉴定

本研究在甘孜藏族自治州新龙县和石渠县通过野外粪便DNA进行物种鉴定, 共发现食肉动物7种, 分别为狼、棕熊、豹、雪豹、豹猫、赤狐等6种野生食肉动物及1种家养食肉动物狗, 为这些物种在当地的分布提供了确凿的证据, 同时显示了基于粪便DNA进行物种快速调查的高效和便捷性。

形态学物种鉴定方法可通过粪便的外形、颜

色、气味、内容物等特征来判断来源物种, 此类方法受到多种物种自身因素、环境条件以及样品采集者的经验和主观判断的影响, 通常存在较高的错误率(Monterroso et al, 2013; Lonsinger et al, 2015)。本次样品采集时, 有22份样品由野外调查人员进行了物种预判, 但仅有8份与分子鉴定结果一致, 即传统物种鉴定方法的正确率为36.4%(表1)。由此可见, 基于DNA条形码的分子鉴定方法大大提高了由粪便进行物种识别的准确率。但就本次物种鉴定结果来看, 可能由于粪便样品质量或者PCR抑制物过多, 38份粪便样品中, 只有28份可以成功提取DNA并进行PCR扩增测序, 物种识别成功率为73.7%。这一问题可以通过优化DNA提取方法和后续采集样品时尽量采集新鲜粪便保证样品质量来提高成功率。

#### 3.2 食肉动物的食性组成

用DNA宏条形码和高通量测序相结合的方法, 从28份粪便样品中获得了28个食物MOTU, 快速展示了该地区7种食肉动物的部分食物物种组成。由于赤狐、棕熊、豹和雪豹的样品数均较少( $\leq 3$ ), 所以仅对样品数量相对较多的狼、狗和豹猫进行讨论。

北美及欧洲的大量研究表明, 有蹄类是狼的主要食物, 这与能量收益率和猎物的可利用度有关(Marfuard, 1998)。目前我们的结果显示, 狼取食最多的是偶蹄目动物, 其中家牦牛在食物组成中出现频率最高, 说明狼可能是对家畜危害最严重的食肉动物之一。原因可能是当地家牦牛数量远大于野生有蹄类, 使其成为狼的主要捕食目标, 但不排除部



分来源是食腐(取食死亡家畜残骸)或捡拾人类丢弃食物残余的可能。这一结果与我国其他很多地区相似,比如在内蒙古地区狼食性研究中,家畜在狼食谱中的频率(%RO)高达60%以上(颜文博等,2006)。在青海,狼食性中家畜的比例(%RO)从38%到78%不等(Liu & Jiang, 2003)。据国内外狼食性研究,狼在枯草期还会捕食啮齿类和兔类作为补充(张洪海等,2000;颜文博等,2006),是被捕食比例仅次于有蹄类的猎物类型,在我们的结果中啮齿目和兔形目物种也占狼食物的较大比例。而啮齿目比兔形目出现频率高,原因可能是在研究地区啮齿目动物数量更多且易捕食,或是啮齿目动物生物量远小于兔形目动物,需捕食更多个体才能满足营养需求。

甘孜州以农牧业生产方式为主,藏民多养狗来看家和放牧。由于宗教信仰,藏民不杀生,对伤人的狗和数量过剩的狗会进行驱逐而不是将其杀死,因此许多具有獒血统的逃逸和散养家狗汇集成群,穿梭于人类生活区与野生动物栖息地,很可能会与野生食肉动物直接竞争生存资源(Mitchell & Banks, 2005; Glen et al, 2006; Glen & Dickman, 2008)。本研究结果中,狗的食物中包括毛冠鹿(*Elaphodus cephalophus*)、野猪(*Sus scrofa*)、高原鼠兔及高原松田鼠等野生动物,这是狗捕食野生动物并与野生食肉动物竞争资源的直接证据。如果散养狗的种群数量不加以控制,可能会持续增长,通过捕食、竞争、排斥等方式对当地野生动物生存及生态系统稳定造成严重影响,需要管理者及早加以重视并进行有效管理。

以往食性研究显示豹猫主要取食啮齿目动物,其在豹猫粪便样品中的出现频率(%FO)可高达81%–96%(Austin et al, 2007; Rajaratnam et al, 2007; Shehzad et al, 2012a; Lee et al, 2014)。我们的分析结果中啮齿目也是豹猫食物中检出频率最高的类群(%FO = 100%),但具体物种构成上与以往调查有所不同。泰国、巴基斯坦、韩国等地豹猫取食的啮齿目食物主要为大鼠属(*Rattus*)和小鼠属(*Mus*)物种(Austin et al, 2007; Shehzad et al, 2012a; Lee et al, 2014),而我们在甘孜地区检测到豹猫食物中的啮齿类以姬鼠属、松田鼠属(*Neodon*)以及白腹鼠属为主,与在四川北部平武地区豹猫食性研究结果类似(Xiong et al, 2017)。此外,兔形目中高原鼠兔和高原兔出现频率最高(%FO = 83%–100%),而此前仅在

四川北部有豹猫取食兔形目物种的报道(Xiong et al, 2017)。豹猫的食物中还发现了家牦牛、毛冠鹿、野猪、麝(*Moschus spp.*)等大中型偶蹄类动物。由于DNA食性分析不能检测猎物的年龄,无法得知具体取食的情况。但考虑到豹猫与这些偶蹄类动物成年个体体型存在较大差异,直接捕食健康成体的可能性较小,推测豹猫有可能捕食其幼仔或取食动物尸体。

豹和雪豹的样品较少,均各1份,分别只检出了豹的猎物中华斑羚和雪豹的猎物岩羊。有蹄类常是豹和雪豹食物的最优选择(Seidensticker, 1976; Shehzad et al, 2012b)。由于大中型有蹄类体型较大,豹和雪豹一餐通常只摄入1种食物,因此每份粪便中的食物物种数少,符合其捕食行为特征(Shehzad et al, 2012b, 2015)。随着采样量的增加,预期豹和雪豹被检出的食物多样性也会有所增加。

不同食肉动物有不同的食物选择,可能体现其体型、生理、捕食策略等方面的差异。本研究食性分析的初步结果展示了当地7种食肉动物的部分主要食物,但由于每种食肉动物采样量不足,各食物物种在食肉动物食谱中所占比例可能不够准确,后续需要更广泛、长期地收集食肉动物粪便样品,以进行更加全面的食性研究,从而深入了解食肉动物的食物网结构、食性选择、竞争关系和共存机制。

### 3.3 问题与展望

采用DNA宏条形码技术并借助生物信息学方法的高通量食性分析有着传统方法无可比拟的优势。它采用较短的条形码设计通用引物,使降解的DNA更容易被扩增和检测,降低了对样品质量的要求。由于测序通量及深度远高于传统测序方法,高通量测序具有极高的灵敏度,微量食物也更容易被检测出来。同时一次测序大量样品产生大量数据,对于需要大规模样品分析的研究可以节省更多的时间和成本。由于形态学难以区分的物种通过DNA宏条形码技术也能被识别,研究者可以快速、大量、精确地获得物种信息,不需要预先长期积累物种分类学知识。

本研究结果表明,基于粪便DNA和DNA条形码技术的物种鉴定和食性分析,可以对研究地区的物种进行快速本底调查。对于本研究地区的食肉动物来说,高寒高海拔的温带森林及高山草甸生态系统生产力低,食物资源比较匮乏,对有限食物资源的利用差异可能比时间和空间资源的分化对于物

种共存更为重要。资源分配等生态数据的缺乏常常是高海拔地区食肉动物保护难以开展的重要原因(Wang et al, 2014)。从本次调查结果看, 川西高原地区分布有多种大型食肉动物, 精准分子食性分析可以从食物资源利用方面为该地区大型哺乳类食肉动物的共存机制提供科学数据, 从而更深入地了解当地生态系统结构和过程, 有助于更好地保护生态系统功能和物种多样性。

除了食性分析, DNA宏条形码的用途正迅速扩展到更多生态学研究领域。如用于植物多样性分析, 发现现有类群中的隐存分类单元等(Yoccoz et al, 2012)。Murray等(2013)通过DNA宏条形码技术调查了上千块来自澳洲西南的化石, 发现了一些当地物种, 还有灭绝或者是从未记录过的物种。Kelly等(2014)通过环境DNA评估海洋中的鱼类多样性及分布从而进行环境监测。在生物资源领域, DNA条形码可以帮助监控珍稀濒危物种的进出口贸易, 实现有害生物及入侵种的快速检验, 以及中药材市场管理和规范等(陈士林等, 2013)。此外, 日益更新的实验技术和数据分析方法更进一步推动了DNA宏条形码技术向更高效、量化和自动化的方向发展(Zhou et al, 2013; Shi et al, 2017; Zhang et al, 2017)。

尽管DNA条形码技术在近些年来飞速发展, 但仍存在缺陷与不足, 其应用潜能尚未完全开发。首先, 依赖DNA条形码的物种鉴定受到GenBank、EMBL、DDBJ等公共资源数据库收录情况的限制。在对一个地区进行分子物种鉴定时, 为避免由于各大公开数据库对地方特有种或隐存种序列收录不全面而导致无法分辨和准确鉴定物种, 需研究者先对该地区进行充分的本底物种调查并建立本地物种条形码数据库。其次, GenBank中收录的部分序列可能自身存在序列错误, 或序列对应物种信息错误, 因此仅依靠DNA序列比对进行物种鉴定也有误判风险, 需要研究者认真核实。再次, 单一DNA条形码对各类群物种的分辨率不同, 这可能会影响到物种鉴定的精确度。如本研究中, 12SV5条形码对雀形目物种分辨率较低, 同一条序列能与多个物种高度匹配, 只能将该MOTU归类为包含所有匹配物种的上一级分类单元。这类问题需要通过针对特定类群设计分辨率更高的条形码来解决。最后, 样品采集、DNA提取、PCR和高通量测序过程中均可能出现样品间的交叉污染问题, 导致假阳性结果。本研究采

取多项措施减少交叉污染造成的假阳性问题: (1)在PCR扩增和测序中同时进行DNA提取空白对照及不加DNA模板的PCR空白对照反应来进行假阳性检验; (2)数据处理时删除每个样品中序列数少于该样品总序列数1%的MOTU以减少交叉污染影响; (3)对每个样品设置3次PCR平行对照反应, 在 $\geq 2$ 次PCR中检测到某个食物MOTU时才认为结果有效。

总之, 随着DNA条形码和宏条形码技术被日益广泛地使用, 研究者需要在不断扩充共享数据库资源的同时加强对数据准确率的监管, 并在各类群条形码设计上不断优化和改善。同时需要强调的是, 对DNA条形码技术的推广不意味着可以忽视或取代传统分类学。对于分类尚不明确的物种, 需在传统分类学家的帮助下, 建立可靠的条形码数据库, 在此基础上, 生物条形码才能发挥越来越大的作用。

**致谢:** 野外样品采集得到四川省林业厅、山水自然保护中心、新龙县林业局、石渠县林业局及石渠洛须白唇鹿省级自然保护区的大力支持, 以及当地林业部门的向阳、周丽、罗布降泽、四郎仁青的协助, 在此表示感谢。

## 参考文献

- Austin SC, Tewes ME, Grassman LI, Silvy NJ (2007) Ecology and conservation of the leopard cat *Prionailurus bengalensis* and clouded leopard *Neofelis nebulosa* in Khao Yai National Park, Thailand. *Acta Zoologica Sinica*, 53, 1–14.
- Baker PJ, Boitani L, Harris S, Saunders G, White PCL (2008) Terrestrial carnivores and human food production: Impact and management. *Mammal Review*, 38, 123–166.
- Bellemain E, Swenson JE, Tallmon D, Brunberg S, Taberlet P (2010) Estimating population size of elusive animals with DNA from hunter-collected feces: Four methods for brown bears. *Conservation Biology*, 19, 150–161.
- Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, Kinzig AP, Daily GC, Loreau M, Grace JB, Larigauderie A, Srivastava DS, Naeem S (2012) Biodiversity loss and its impact on humanity. *Nature*, 486, 59–67.
- Chen SL, Pang XH, Luo K, Yao H, Han JP, Song JY (2013) DNA barcoding of biological resources. *Chinese Bulletin of Life Sciences*, 25, 458–466. (in Chinese with English abstract) [陈士林, 庞晓慧, 罗焜, 姚辉, 韩建萍, 宋经元 (2013) 生物资源的DNA条形码技术. *生命科学*, 25,

458–466.]

- Cheng XT, Wang AM, Gu ZF, Wang Y, Zhan X, Shi YH (2011) Current progress of DNA barcoding. *Genomics and Applied Biology*, 30, 748–758. (in Chinese with English abstract) [程希婷, 王爱民, 顾志峰, 王嫣, 战欣, 石耀华 (2011) DNA条形码研究进展. *基因组学与应用生物学*, 30, 748–758.]
- Coissac E (2012) OligoTag: A program for designing sets of tags for next-generation sequencing of multiplexed samples. *Methods in Molecular Biology*, 888, 13–31.
- Cole DN, Landres PB (1996) Threats to ecosystems: Impacts and research needs. *Ecological Applications*, 6, 168–184.
- Creer S, Fonseca VG, Porazinska DL, Giblin-Davis RM, Sung W, Power DM, Packer M, Carvalho GR, Blaxter ML, Lambshhead PJ, Thomas WK (2010) Ultrasequencing of the meiofaunal biosphere: Practice, pitfalls and promises. *Molecular Ecology*, 19, 4–20.
- Gómez-Ortiz Y, Monroy-Vilchis O, Mendoza-Martínez GD (2015) Feeding interactions in an assemblage of terrestrial carnivores in central Mexico. *Zoological Studies*, 54, 16.
- Glen AS, Fay AR, Dickman CR (2006) Diets of sympatric red foxes *Vulpes vulpes* and wild dogs *Canis lupus* in the Northern Rivers Region, New South Wales. *Australian Mammalogy*, 28, 101–104.
- Glen AS, Dickman CR (2008) Niche overlap between marsupial and eutherian carnivores: Does competition threaten the endangered spotted-tailed quoll? *Journal of Applied Ecology*, 45, 700–707.
- Hebert PDN, Cywinska A, Ball SL, Dewaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270, 313–321.
- Hooper DU, Adair EC, Cardinale BJ, Byrnes JE, Hungate BA, Matulich KL, Gonzalez A, Duffy JE, Gamfeldt L, O'Connor MI (2012) A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, 486, 105–108.
- Hooper DU, Chapin FS III, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S (2005) Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecological Monographs*, 75, 3–35.
- Kasper CB, Peters FB, Christoff AU, de Freitas TRO (2016) Trophic relationships of sympatric small carnivores in fragmented landscapes of southern Brazil: Niche overlap and potential for competition. *Mammalia*, 80, 143–152.
- Kelly RP, Port JA, Yamahara KM, Crowder LB (2014) Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE*, 9, e86175.
- Kircher M, Kelso J (2010) High-throughput DNA sequencing concepts and limitations. *Bioessays*, 32, 524–536.
- Kress WJ, Garcia-Robledo C, Uriarte M, Erickson DL (2015) DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology and Evolution*, 30, 25–35.
- Lee O, Lee S, Nam DH, Lee HY (2014) Food habits of the leopard cat (*Prionailurus bengalensis euptilurus*) in Korea. *Mammal Study*, 39, 43–46.
- Liu B, Jiang ZG (2003) Diet composition of wolves *Canis lupus* in the northeastern Qinghai-Tibet Plateau, China. *Acta Theriologica*, 48, 255–263.
- Lonsinger RC, Gese EM, Waits LP (2015) Evaluating the reliability of field identification and morphometric classifications for carnivore scats confirmed with genetic analysis. *Wildlife Society Bulletin*, 39, 593–602.
- Ma L (2013) Research progress in DNA minibarcoding and metabarcoding. *Journal of Agricultural Catastrophology*, 3(6), 58–60. (in Chinese with English abstract) [马兰 (2013) DNA微型条形码和复合条形码研究进展. *农业灾害研究*, 3(6), 58–60.]
- Marfuad P (1998) Food habits of Arctic wolves in Greenland. *Journal of Mammalogy*, 79, 236–244.
- Mitchell BD, Banks PB (2005) Do wild dogs exclude foxes? Evidence for competition from dietary and spatial overlaps. *Austral Ecology*, 30, 581–591.
- Monterroso P, Castro D, Silva TL, Ferreras P, Godinho R, Alves PC (2013) Factors affecting the (in)accuracy of mammalian mesocarnivore scat identification in Southwestern Europe. *Journal of Zoology*, 289, 243–250.
- Montoya JM, Rodríguez MA, Hawkins BA (2003) Food web complexity and higher-level ecosystem services. *Ecology Letters*, 6, 587–593.
- Murray DC, Haile J, Dortch J, White NE, Haouchar D, Bellgard MI, Allcock RJ, Prideaux GJ, Bunce M (2013) Scrapheap challenge: A novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Scientific Reports*, 3, 3371.
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, 403, 853–858.
- Pelton MR, Gittleman JL, Funk SM, Macdonald DW, Wayne RK (2003) Carnivore conservation. *Journal of Wildlife Management*, 67, 229–230.
- Pompanon F, Deagle BE, Symondson WO, Brown DS, Jarman SN, Taberlet P (2012) Who is eating what: Diet assessment using next generation sequencing. *Molecular Ecology*, 21, 1931–1950.
- Porazinska DL, Giblin-Davis RM, Esquivel A, Powers TO, Sung W, Thomas WK (2010) Ecometagenetics confirm high tropical rainforest nematode diversity. *Molecular Ecology*, 19, 5521–5530.
- Rajaratnam R, Sunquist M, Rajaratnam L, Ambu L (2007) Diet and habitat selection of the leopard cat (*Prionailurus bengalensis borneoensis*) in an agricultural landscape in



- Sabah, Malaysian Borneo. *Journal of Tropical Ecology*, 23, 209–217.
- Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E (2011) ecoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, 39, e145.
- Ripple WJ, Estes JA, Beschta RL, Wilmers CC, Ritchie EG, Hebblewhite M, Berger J, Elmhagen B, Letnic M, Nelson MP, Schmitz OJ, Smith DW, Wallach AD, Wirsing AJ (2014) Status and ecological effects of the world's largest carnivores. *Science*, 343, 1241484.
- Seidensticker J (1976) On the ecological separation between tigers and leopards. *Biotropica*, 8, 225–234.
- Shehzad W, Riaz T, Nawaz MA, Miquel C, Poillot C, Shah SA, Pompanon F, Coissac E, Taberlet P (2012a) Carnivore diet analysis based on next-generation sequencing: Application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, 21, 1951–1965.
- Shehzad W, McCarthy TM, Pompanon F, Purevjav L, Coissac E, Riaz T, Taberlet P (2012b) Prey preference of snow leopard (*Panthera uncia*) in South Gobi, Mongolia. *PLoS ONE*, 7, e32104.
- Shehzad W, Nawaz MA, Pompanon F, Coissac E, Riaz T, Shan SA, Taberlet P (2015) Forest without prey: Livestock sustain a leopard *Panthera pardus* population in Pakistan. *Oryx*, 49, 248–253.
- Shi ZY, Yang CQ, Hao MD, Wang XY, Ward RD, Zhang AB (2017) FuzzyID2: A software package for large dataset species identification via barcoding and metabarcoding using Hidden Markov Models and fuzzy set methods. *Molecular Ecology Resource*, 18, 666–675.
- Smith AT, Xie Y (2009) A Guide to the Mammals of China. Hunan Education Publishing House, Changsha. (in Chinese) [Smith AT, 解焱 (2009) 中国兽类野外手册. 湖南教育出版社, 长沙.]
- Symondson WO (2002) Molecular identification of prey in predator diets. *Molecular Ecology*, 11, 627–641.
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21, 2045–2050.
- Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Trends in Ecology and Evolution*, 24, 110–117.
- Vallet D, Petit EJ, Gatti S, Levréro F, Ménard N (2007) A new 2CTAB/PCI method improves DNA amplification success from faeces of Mediterranean (Barbary macaques) and tropical (lowland gorillas) primates. *Conservation Genetics*, 9, 677–680.
- Vestheim H, Jarman SN (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples—A case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*, 5, 12.
- Voolstra CR, Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, 6, e17497.
- Wang J, Laguardia A, Damerell PJ, Riordan P, Shi K (2014) Dietary overlap of snow leopard and other carnivores in the Pamirs of Northwestern China. *Chinese Science Bulletin*, 59, 3162–3168.
- Wang SW, Macdonald DW (2009) Feeding habits and niche partitioning in a predator guild composed of tigers, leopards and dholes in a temperate ecosystem in central Bhutan. *Journal of Zoology*, 277, 275–283.
- Wang XQ, Wang GH, Qiao F, Gao QK, Heong KL, Zhu ZR, Cheng JA (2015) Progress on high-throughput sequencing and its applications in food web analysis. *Acta Ecologica Sinica*, 37, 2530–2539. (in Chinese with English abstract) [王雪芹, 王光华, 乔飞, 高其康, Heong KL, 祝增荣, 程家安 (2015) 高通量测序及其在食物网解析中的应用进展. *生态学报*, 37, 2530–2539.]
- Xiong MY, Wang DJ, Bu HL, Shao XN, Zhang D, Li S, Wang RJ, Yao M (2017) Molecular dietary analysis of two sympatric felids in the Mountains of Southwest China biodiversity hotspot and conservation implications. *Scientific Reports*, 7, 41909.
- Xiong MY, Shao XN, Long Y, Bu HL, Zhang D, Wang DJ, Li S, Wang RJ, Yao M (2016) Molecular analysis of vertebrates and plants in scats of leopard cats (*Prionailurus bengalensis*) in southwest China. *Journal of Mammalogy*, 97, 1054–1064.
- Yan WB, Zhang HH, Yang HJ, Dou HS, Shen XQ (2016) Seasonal diet of wolves in the Dalaihu Nature Reserve, Inner Mongolia. *Chinese Journal of Zoology*, 41, 46–51. (in Chinese with English abstract) [颜文博, 张洪海, 杨红军, 窦华山, 沈秀清 (2006) 内蒙古达赉湖自然保护区狼食性的季节性变化. *动物学杂志*, 41, 46–51.]
- Yin HB, Yu GJ, Wang GL, Zhou YB, Wu MM (2008) Study methods of diets of carnivorous animals. *Journal of Anhui University (Natural Sciences Edition)*, 32(1), 90–94. (in Chinese) [尹华宝, 余冠军, 王贵林, 周友兵, 武梅梅 (2008) 食肉目动物食性研究方法. *安徽大学学报(自然科学版)*, 32(1), 90–94.]
- Yoccoz NG, Bråthen KA, Gielly L, Haile J, Edwards ME, Goslar T, Von SH, Brysting AK, Coissac E, Pompanon F (2012) DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, 21, 3647–3655.
- Zhang AB, Hao MD, Yang CQ, Shi ZY (2017) BarcodingR: An integrated R package for species identification using DNA barcodes. *Methods in Ecology and Evolution*, 8, 627–634.
- Zhang HH, Wang ZL, Ma WX, Sun YY (2000) Food habits of

- wolf (*Canis lupus*) in Xing'anling Mountains. *Journal of Qufu Normal University*, 26(1), 80–82. (in Chinese) [张洪海, 王振龙, 马文祥, 孙玉英 (2000) 大、小兴安岭地区狼的食性. *曲阜师范大学学报*, 26(1), 80–82.]
- Zhang S, Bao QQ, Tang PZ, Zhang SL, Yang YX, Li GL, Bao WD (2013) Description on foot tracts and fecal morphology of several carnivore species. *Sichuan Journal of Zoology*, 32, 857–861. (in Chinese with English abstract) [张帅, 鲍清泉, 汤鹏展, 张书理, 杨永昕, 李桂林, 鲍伟东 (2013) 几种食肉目动物的足迹与粪便形态描述. *四川动物*, 32, 857–861.]
- Zhou YW, Yang YH (2009) Review of non-DNA methods for identification of animal species. *Journal of Tonghua Teachers College*, 30(10), 58–61. (in Chinese) [周用武, 杨玉华 (2009) 动物物种鉴定的非DNA方法评述. *通化师范学院学报*, 30(10), 58–61.]
- Zhou X, Li YY, Liu SL, Yang Q, Su X, Zhou LL, Tang M, Fu RB, Li JG, Huang QF (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, 2, 4.
- (特邀责任编辑: 周欣 责任编辑: 闫文杰)

## 附录 Supplementary Material

### 附录1 本研究使用的引物信息

Appendix 1 Summary of primer information in this study

<http://www.biodiversity-science.net/fileup/PDF/2018214-1.pdf>



•研究报告•

# 用于蜜蜂和熊蜂肠道微生物分类的 细菌16S rRNA数据库优化

张雪<sup>1</sup> 李兴安<sup>2</sup> 苏秦之<sup>1</sup> 曹棋钠<sup>1</sup> 李晨伊<sup>1</sup> 牛庆生<sup>2\*</sup> 郑浩<sup>1\*</sup>

1 (北京市食品营养与人类健康高精尖创新中心, 中国农业大学食品科学与营养工程学院, 北京 100083)

2 (吉林省养蜂科学研究所蜜蜂遗传育种省重点实验室, 吉林省吉林市 132108)

**摘要:** 蜜蜂和熊蜂是重要的传粉昆虫, 对农业生产及生态平衡的维持具有重要作用。近年来, 研究发现蜜蜂及熊蜂肠道内含有大量微生物, 其组成简单、特异。正常的肠道微生物群落对蜜蜂的生长、激素调节、致病菌抵抗等具有重要作用。随着高通量测序的发展, 研究者们也可快速获得传粉蜂肠道微生物组成, 这给生物多样性和物种保护及蜂类健康等的研究带来了便捷。但是由于蜜蜂和熊蜂肠道微生物群落均由特殊菌种组成, 目前的细菌16S rRNA数据库无法对其进行准确的分类, 并且部分东方蜜蜂(*Apis cerana*)特有的肠道微生物菌种缺乏16S rRNA序列信息。本文从来源于5个不同省份的东方蜜蜂肠道中分离得到在东方蜜蜂中普遍含有的*Apibacter*菌属纯菌, 获取其全长16S rRNA序列, 并对目前蜜蜂和熊蜂肠道的5个核心菌种的分类进行了综述, 对其分类和命名进行了修正。根据蜜蜂肠道微生物的明确分类, 在目前常用的SILVA细菌分类数据库基础之上对其进行了命名及分类优化, 并加入东方蜜蜂中普遍含有的*Apibacter*序列, 从而获得了优化数据库Bee Gut Microbiota-Database (BGM-Db)。通过1组东方蜜蜂及1组西方蜜蜂(*Apis mellifera*)的肠道菌群高通量测序结果, 分析不同数据库的表现, 我们发现相比于SILVA和Ribosomal Database Project (RDP), BGM-Db对蜜蜂肠道16S rRNA高通量测序短序列实现了菌种级别的分类, 分辨率更高。

**关键词:** 蜜蜂; 肠道微生物; 高通量测序; 数据库; 16S rRNA

## A curated 16S rRNA reference database for the classification of honeybee and bumblebee gut microbiota

Xue Zhang<sup>1</sup>, Xing'an Li<sup>2</sup>, Qinzhi Su<sup>1</sup>, Qina Cao<sup>1</sup>, Chenyi Li<sup>1</sup>, Qingsheng Niu<sup>2\*</sup>, Hao Zheng<sup>1\*</sup>

1 Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083

2 Provincial Key Laboratory of Bee Genetics and Breeding, Apiculture Science Institute of Jilin Province, Jilin City, Jilin Province 132108

**Abstract:** Honey and bumble bees are import pollinators, playing significant roles in the agricultural industry and maintaining the bio-ecosystem balance. Recently, it was found that the bees harbor a simple, yet specific gut microbiota. The normal bee gut microbiota makes essential contributions to host growth, endocrine signaling, and pathogen resistance. With the development of high through-put sequencing technology, researchers can now quickly identify the gut community structure for a low cost. This is helpful for biodiversity, conservation and bee health studies. However, the currently-used 16S rRNA databases are not specific enough to classify the bee gut microbiota properly. Many of the specific bacteria that enrich the gut of *Apis cerana* are in the genus *Apibacter*. Here, we isolated *Apibacter* species from *A. cerana* collected in five provinces of China, and added them to the current SILVA database. We also curated the nomenclature of some existing sequences and re-classified them in the updated database. Based on the analysis of the 16S rRNA sequencing data from one *A. cerana* and one *Apis mellifera* sample, our Bee Gut Microbiota-Database (BGM-Db) offers a more accurate classification of bee gut microbiota at a higher resolution than either the

收稿日期: 2019-01-25; 接受日期: 2019-05-07

基金项目: 国家自然科学基金(31870472; 31470123)

\* 共同通讯作者 Co-authors for correspondence. E-mail: 1463199779@qq.com; hao.zheng@cau.edu.cn



### SILVA or Ribosomal Database Project (RDP) database.

**Key words:** honey bee; gut microbiota; high-throughput sequencing; database; 16S rRNA

蜜蜂是重要的农业昆虫,除生产蜂蜜、蜂王浆、蜂胶等蜂产品外,其更重要的功能是作为媒介为农作物授粉,保障农业产量(Klein et al, 2007)。近年来,由于生态环境改变、农药的过量施用等原因,蜜蜂种群数量锐减。自2006年首次报道以来,全球范围内多次暴发蜂群崩溃综合征(colony collapse disorder, CCD),使得美国23%的蜂农蜜蜂损失达45%以上(Cox-Foster et al, 2007)。导致CCD的原因是多方面的(vanEngelsdorp et al, 2009)。近年来对于蜜蜂肠道微生物群落的研究发现,正常的肠道微生物对于蜜蜂的生长、健康、食物代谢、激素调节、提高免疫力、抵抗外源致病菌入侵等方面都有重要作用(Raymann et al, 2017; Zheng et al, 2017)。通过对西方蜜蜂(*Apis mellifera*)肠道菌的16S rRNA测序发现,其肠道菌群组成简单并且较为保守,受地理和环境等因素的影响较小。蜜蜂肠道菌群主要由8种特异核心菌种组成,占其肠道微生物群落的95%以上(Engel et al, 2014; Martinson et al, 2011),包括3种革兰氏阳性菌及5种革兰氏阴性菌。而其他蜂种,例如东方蜜蜂(*Apis cerana*)、熊蜂(*Bombus*)及其他无刺蜂也具有和西方蜜蜂相似的肠道菌群结构(Martinson et al, 2011)。蜜蜂肠道菌具有高度特异性,目前发现的来源于蜜蜂肠道的菌种均为蜜蜂特有菌种,在其他环境中并不存在,并且不同蜂种也具有各自特异的菌种(Kwong et al, 2014a, 2017)。因此,研究不同蜂种、蜂群肠道菌群多样性及其与蜂群健康的关系等需要在较高的分辨率上准确分析。

蜜蜂及熊蜂的主要肠道菌种都可以进行体外培养。近年来随着对蜜蜂及熊蜂的深入研究,分离得到大量核心肠道菌并且获得了大量纯菌基因组信息,测得的全长或接近全长的16S rRNA序列储存在SILVA (Pruesse et al, 2007)和核糖体数据库项目(Ribosomal Database Project, RDP; Cole et al, 2014)等通用数据库中,用于高通量测序结果的比对。但是通用数据库并没有针对来源于蜜蜂肠道菌的序列进行优化,其中的蜜蜂肠道菌存在命名模糊、菌种水平分类不清等问题。另外,虽然东方蜜蜂和熊蜂具有很多和西方蜜蜂相近的肠道菌属,但

也含有其特殊肠道菌。Kwong等(2017)发现,东方蜜蜂、大蜜蜂(*Apis dorsata*)和熊蜂中含有不同于西方蜜蜂的一种特殊肠道菌属——*Apibacter*,而目前测得的*Apibacter*的16S rRNA全长序列较少,分离得到的纯菌也仅有来自于熊蜂的1株*Apibacter men-salis* (Praet et al, 2016)和来自于东方蜜蜂和大蜜蜂的3株*Apibacter adventoris* (Kwong & Moran, 2016)。因此目前数据库中*Apibacter*属的参考序列多样性较低,这会影响到对东方蜜蜂、熊蜂和大蜜蜂中*Apibacter*菌种的鉴定(Kwong & Moran, 2016; Kwong et al, 2017)。

随着测序技术的发展,通过对16S rRNA基因进行高通量测序来鉴定蜜蜂肠道菌群,对不同蜂种及不同地域、不同健康状况的蜂群进行比较分析等具有重要作用。而目前基于Illumina等平台的高通量测序得到的16S rRNA序列主要为部分的高变区域,基于片段序列的菌种鉴定则需要利用RDP的朴素贝叶斯分类器(Ribosomal Database Project's Naïve Bayesian Classifier, RDP-NBC)对微生物群落的分类学组成及结构进行预测(Wang et al, 2007),而分类的准确性很大程度上依赖作为训练集(training set)的接近16S rRNA全长(细菌约1,500 bp)的参考数据库分类的深度和解析度(Mikaelyan et al, 2015)。目前最常用的16S rRNA短序列分类数据库包括SILVA (Yilmaz et al, 2014)和RDP (Cole et al, 2014)。而对于某些特殊环境如昆虫肠道中的样品,由于缺乏长片段的16S rRNA克隆文库序列,无法构建用于分类鉴定的训练集,影响了分类的准确性(Newton & Roeselers, 2012; Mikaelyan et al, 2015)。特别是对于蜜蜂肠道,其细菌群落均为特异的分类类群,而在其他环境中基本没有发现。另外,虽然蜜蜂肠道菌群结构相对简单,在菌种水平上只有8种核心菌,但是在16S rRNA序列相似性大于97%的同一菌种内存在较高的菌株间多样性。之前对于蜜蜂肠道菌株间多样性的研究发现,某些菌株即使其16S rRNA序列相似性达到99%,它们的基因组结构及功能也存在很大差异(Engel et al, 2014; Zheng et al, 2016)。另外目前对于蜜蜂肠道的研究主要集中于西

方蜜蜂, 对于东方蜜蜂等其他蜂种肠道菌群的分离鉴定还存在不足, 造成数据库的分辨率和分类准确性不高。

因为目前并没有针对蜜蜂等蜂类肠道菌优化的16S rRNA数据库, 本文基于最为常用且更新频率较高的SILVA数据库, 将其中蜜蜂肠道特有细菌的分类和命名进行系统校正。另外, 来自吉林、四川、青海、海南和江西的东方蜜蜂中分离出96株东方蜜蜂特有菌属*Apibacter*菌株, 利用Sanger测序获得接近全长的16S rRNA序列, 并将其补充到优化的数据库中。通过新序列的补充及对原有数据库分类和名称的优化, 我们构建了针对蜜蜂16S rRNA高通量测序分析的数据库Bee Gut Microbiota Database (BGM-Db)。并且通过对1组东方蜜蜂和1组西方蜜蜂的肠道菌16S rRNA的高通量测序结果进行分类, 对SILVA、RDP和BGM-Db数据库的表现进行了比较。

## 1 材料与方法

### 1.1 蜜蜂样品及菌株分离

东方蜜蜂样品分别采自吉林省敦化市大石头林业局沟口林场、青海省海南藏族自治州贵德县河阴镇、海南省三亚市临春岭公园、江西省永修县燕山镇七里长坑、四川省平武县木皮藏族乡关坝村蜂场和马尔康市松岗镇阿坝东方蜜蜂保种场。将活蜜蜂冻于冰上5–10 min后用尖头镊子解剖肠道并在100 uL的25%甘油溶液中(v/v) (PBS pH = 7.4)充分匀浆。用接种环挑取少量匀浆, 分别在添加有5%脱纤维无菌绵羊血的脑心浸液培养基(HIA, OXOID)、5%脱纤维无菌绵羊血的哥伦比亚固体培养基(OXOID)及添加1 g/L莫匹罗星(大连美仑生物技术有限公司)的TPY琼脂培养基(青岛海博生物技术有限公司)进行划线。35℃下5% CO<sub>2</sub>培养箱中培养48 h或者在厌氧环境操作台内培养144 h后挑取单菌落, 进行16S rRNA菌落PCR扩增(引物分别是27F和1492R; Lane, 1991), 将测序结果与NCBI 16S rRNA数据库序列进行Blast比对, 鉴定是否为*Apibacter*。

### 1.2 *Apibacter*全长16S rRNA序列分析

将获得的全长*Apibacter* 16S rRNA序列在MEGA 7软件中进行序列比对, 并将获得的比对序列进行人工修正, 然后在MEGA 7软件中利用最大似然法构建系统发育树。获得的*Apibacter*的16S

rRNA序列提交到GenBank数据库, 序列号为MH478207–MH478300。

### 1.3 数据库分类优化

BGM-Db的分类框架优化基于SILVA数据库的原有系统发育分类。对SILVA数据库中含有大部分蜜蜂肠道菌来源序列的细菌进化分支进行系统发育分析, 从而重新定义或进一步解析基于节点的分类, 其命名基于前期对蜜蜂肠道菌的分类命名。经过质控的96株*Apibacter*菌株的接近全长16S rRNA序列, 通过与Mothur网站上的SILVA数据库参考序列进行比对, 利用ARB软件包上传到SILVA数据库。利用UCLUST, 以98%相似性为阈值去除冗余的16S rRNA基因序列。

### 1.4 16S rRNA基因的高通量测序和分类比较

选取东方蜜蜂和西方蜜蜂各1只, 将抽取的肠道在含有20 uL蛋白酶K溶液(20 mg/mL)的十六烷基三甲基溴化铵(CTAB)缓冲液中匀浆, 参考已有的微球破碎方法提取肠道DNA, 进行16S rRNA基因的V3+V4区PCR扩增。对扩增子进行Illumina高通量测序, 将获得结果作为检测不同数据库表现的原始数据。原始数据进行质量控制后分别利用BGM-Db及原始的SILVA数据库和常用的RDP数据库进行细菌分类。采用Mothur软件包的RDP朴素贝叶斯分类法, 基于80%的置信区间, 将排名前10的分类结果进行比较。

## 2 结果

### 2.1 对数据库中蜜蜂核心菌分类名称的校正

由于目前对于蜜蜂肠道细菌的分类比较模糊, 命名也比较繁杂、多样, 因此首先对蜜蜂肠道核心菌的发现历史进行简要总结, 进而对其分类及命名进行规范。2010年Martinson等(2011)通过16S rRNA克隆建库的方法对西方蜜蜂、大蜜蜂、黑小蜜蜂(*Apis andreniformis*)和熊蜂肠道共生菌进行了系统的研究, 发现社会性蜜蜂核心肠道共生菌8种, 将相应的分类类群命名为Alpha-1, Alpha-2.1, Alpha-2.2, Beta, Firm-4, Firm-5, Gamma-1, Gamma-2(表1)。但是当时并没有得到这几种核心菌的纯菌, 因此没有具体的菌种命名。在接下来的研究当中, 陆续分离得到了相应纯菌, 逐步构建了蜜蜂肠道菌的命名体系。

(1)  $\beta$ -变形菌*Snodgrassella*菌属。高通量测序分

表1 蜜蜂肠道核心菌命名校正

Table 1 List of the nomenclature of the curated bacterial species from bee gut

属名 Genus	种名 Species	其他名 Other names	BGM-Db数据库聚类名 BGM-Db names	参考文献 References
<i>Snodgrassella</i>	<i>Snodgrassella alvi</i>	Beta	<i>Snodgrassella</i>	Martinson et al, 2012; Kwong & Moran, 2013
<i>Gilliamella</i>	<i>Gilliamella apicola</i>	Gamma-1	<i>Gilliamella</i>	Martinson et al, 2012; Kwong & Moran, 2013
<i>Frischella</i>	<i>Frischella perrara</i>	Gamma-2	<i>Frischella</i>	Engel et al, 2013
<i>Schmidhempelia</i>	' <i>Candidatus Schmidhempelia bombi</i> '		<i>Schmidhempelia</i>	Martinson et al, 2014
<i>Bartonella</i>	<i>Bartonella apis</i>	Alpha-1	<i>Bartonella apis</i>	Moran, 2015
<i>Commensalibacter</i>	<i>Commensalibacter intestini</i>	Alpha-2.1	<i>Commensalibacter</i> Alpha2.1	Kwong et al, 2014b
<i>Bombella</i>	<i>Bombella apis</i>	Alpha-2.2	<i>Bombella</i> Alpha2.2	Yun et al, 2017
	<i>Bombella intestini</i>			Yun et al, 2017
<i>Parasaccharibacter</i>	<i>Parasaccharibacter apium</i>			Moran, 2015
<i>Lactobacillus</i>	<i>Lactobacillus mellis</i>	Firm-4	<i>Lactobacillus</i> Firm-4	Olofsson et al, 2014; Moran, 2015
	<i>Lactobacillus mellifer</i>			
	<i>Lactobacillus apis</i>	Firm-5	<i>Lactobacillus</i> Firm-5	Kwong et al, 2014b
	<i>Lactobacillus helsingborgensis</i>			Olofsson et al, 2014; Moran, 2015
	<i>Lactobacillus melliventris</i>			
	<i>Lactobacillus kimbladii</i>			
	<i>Lactobacillus kullabergensis</i>			
	<i>Lactobacillus apinorum</i>			
	<i>Lactobacillus kunkeei</i>	/	<i>Lactobacillus kunkeei</i>	Moran, 2015
<i>Bifidobacterium</i>	<i>Bifidobacterium asteroides</i>	/	<i>Bifidobacterium asteroides</i>	Bottacini et al, 2012
	<i>Bifidobacterium coryneforme</i>	/	<i>Bifidobacterium coryneforme/indicum</i>	Ellegaard et al, 2015
	<i>Bifidobacterium indicum</i>	/		Ellegaard et al, 2015
	<i>Bifidobacterium bombi</i>	/	<i>Bifidobacterium bombi/commune/bohemicum</i>	Killer et al, 2009
	<i>Bifidobacterium commune</i>	/		Praet et al, 2015
	<i>Bifidobacterium bohemicum</i>	/		Killer et al, 2011
<i>Bombiscardovia</i>	<i>Bombiscardovia coagulans</i>	/	<i>Bombiscardovia</i>	Killer et al, 2010
<i>Apibacter</i>	<i>Apibacter adventoris</i>	/	<i>Apibacter</i>	Kwong & Moran, 2016
	<i>Apibacter mensalis</i>	/		Praet et al, 2016

析发现,  $\beta$ -变形菌(Beta)这个分类在蜜蜂肠道中丰度较高, 在西方蜜蜂和东方蜜蜂中丰度约为12%, 在熊蜂中丰度更高, 为14%–40% (Kwong et al, 2017), 主要定殖于回肠部位, 贴肠道壁生长。16S rRNA序列分析表明, 这些“Beta”菌属于 *Neisseriaceae* 科, 与 *Simonsiella*、*Neisseria* 和 *Alysiella* 菌属相近, 序列差异性大于5%, 并形成单独的分类群(Martinson et al, 2012)。Kwong等(2014a)从西方蜜蜂和熊蜂中分离到了多株Beta单菌wkB2、wkB12和wkB29, 其间的相似度在99%以上, 与之关系最近的是来源于白蚁肠道的 *Stenoxybacter acetivorans* (93.8%)。由于形成新的分类簇, 作者将其命名为 *Snodgrassella* 菌属, 菌

种命名为 *Snodgrassella alvi*。Beta菌在肠道中主要分布于回肠, 通过FISH方法对西方蜜蜂肠道菌的可视化研究发现, *Snodgrassella*和肠道壁接触生长(Martinson et al, 2012)。基因组分析发现, 这2种菌富含与细胞运动和分泌功能相关的基因, 这些功能与细胞膜的生成相关, 而细胞膜的形成可以抵抗外源致病菌对肠道的入侵。并且这2种菌具有编码RTX蛋白的基因, RTX蛋白可以发挥细菌素的功能从而抵抗外源致病菌。Steele等(2017)发现 *Snodgrassella alvi*含有多种T6SS作用因子, 并且在蜜蜂肠道内其表达增强, 说明了Beta菌在肠道中与宿主、其他共生菌和致病菌等相互作用。不同 *Snodgrassella alvi*



单菌中T6SS作用因子有所不同,说明在不同蜜蜂当中不同分类群的*Snodgrassella*功能或者基因组具有不同特点。

(2)  $\gamma$ -变形菌 *Gilliamella*, *Frischella*, ‘*Candidatus Schmidhempelia bombi*’。蜜蜂中的 $\gamma$ -变形菌主要来自 *Orbaceae* 科,包括Gamma-1和Gamma-2两个不同菌属(Martinson et al, 2011)。16S rRNA分析表明Gamma-1广泛存在于各种社会性蜜蜂肠道中,且丰度较高,为15%–40% (Kwong et al, 2017)。Gamma-1与 *Pasteurellaceae* 和 *Enterobacteriaceae* 相近,但是形成不同的聚类,其差异超过10%,该菌属被命名为 *Gilliamella* (Martinson et al, 2012), 主要定殖于回肠部位,占比约为47%。Kwong和Moran (2013)从西方蜜蜂和熊蜂中分离到3株 *Gilliamella* 纯菌,并将菌种命名为 *Gilliamella apicola*。宏基因组分析发现Gamma-1基因组中富集碳水化合物代谢途径,并且含有能够降解花粉主要成分果胶的裂解酶(Engel et al, 2012)。花粉降解产物中部分单糖对蜜蜂具有毒性,如甘露糖。Zheng等(2016)对42株 *Gilliamella* 基因组测序发现,部分 *Gilliamella* 菌株编码甘露糖降解酶,并通过体外实验验证了其功能,说明其在蜜蜂解毒中发挥作用。

Gamma-2与其近源物种 *Orbus* 和 *Gilliamella* 有大概95%的相似度,进化上形成独立分支。Engel等(2013)从西方蜜蜂中分离到Gamma-2纯菌,命名为 *Frischella perrara*。该菌特殊定殖在西方蜜蜂中肠与后肠之间的幽门区域,但不存在于熊蜂中, *Frischella* 的定殖引起蜜蜂的类似免疫黑化反应,形成痂状表型,该菌与西方蜜蜂的相互作用及其在免疫和代谢中发挥的作用尚待研究(Engel et al, 2015)。

Martinson等(2014)从熊蜂(*Bombus impatiens*)的宏基因组中拼接得到熊蜂特有的1株Gamma变形菌的基因组信息,与 *Gilliamella* 和 *Frischella* 同属于 *Orbales* 目,但形成不同的单独分类簇。用野生和实验室饲养的 *B. impatiens* 对该菌属进行特异性PCR验证,发现该菌存在于90%的 *B. impatiens* 中。基因组分析发现该菌中缺乏大部分参与三羧酸循环的基因和NADH脱氢酶,但保留和低氧环境代谢相关的基因,具有共生菌特有的简化的基因组。表明了该菌与熊蜂的共同进化过程。Martinson等(2014)将其命名为“*Candidatus Schmidhempelia bombi*”,在本文的BGM-Db中也将相关序列群命名为“*Schmid-*

*hempelia*”。

(3)  $\alpha$ -变形菌。蜜蜂肠道的 $\alpha$ -变形菌是主要的肠道共生菌,约占肠道微生物总量的20%左右(Engel et al, 2012)。蜜蜂中Alpha菌群主要有3类: Alpha-1、Alpha-2.1和Alpha-2.2,分别对应 *Bartonella*、*Commensalibacter* 和 *Bombella* 属。*Bartonella* 为 *Apis* 属蜜蜂的特有菌(Kwong et al, 2017),实验表明 *Bartonella* 能降解果胶中的多聚半乳糖醛酸,帮助蜜蜂代谢花粉(Engel et al, 2012),并且编码大量氨基酸及辅助因子生物合成、氮代谢相关的基因(Segers et al, 2017)。但其在肠道中的定殖位置还未有报道。*Commensalibacter* 主要存在于熊蜂中(Praet et al, 2018),也存在于果蝇肠道中(Kim et al, 2012)。Alpha-2.2主要存在于蜜蜂的食物、蜜囊和幼虫中。聚类为3种,分别命名为 *Bombella intestini*, *Bombella apis* 和 *Parasaccharibacter apium*。*Bombella intestine* 是由Li等(2015)从熊蜂蜜囊中分离到的乙酸菌,能以氨作为唯一碳源。*Bombella apis* 从西方蜜蜂中肠中分离得到,为严格厌氧菌。*Parasaccharibacter apium* 能帮助提高幼虫存活率和抵抗螨虫的感染(Corby-Harris et al, 2014, 2016)。

(4) 乳杆菌 *Lactobacillus Firm-4*, *Firm-5*。蜜蜂肠道中的 *Lactobacillus* 与人肠道的相同菌属属于不同的分类分支。*Lactobacillus* 主要有2种系统型,分别是属于厚壁菌门的 *Firm-4* 和 *Firm-5*, 主要分布在蜜蜂的后肠(回肠和直肠),在蜜蜂微生物中的丰度为5%–50%,具有丰富的属内及种内多样性(Ellegaard et al, 2015)。含有大量编码细胞膜蛋白的基因和涉及糖摄取的磷酸转移酶系统(Kwong et al, 2014b),可能与花粉等植物组分的吸附和降解相关(Ellegaard et al, 2015)。*Firm-4* 包括 *Lactobacillus mellis* 和 *Lactobacillus mellifer*, *Firm-5* 在蜜蜂肠道中丰度更高,包括多个菌种,主要有 *Lactobacillus apis*、*Lactobacillus helsingborgensis*、*Lactobacillus melliventris*、*Lactobacillus kimbladii*、*Lactobacillus kullabergensis* 和 *Lactobacillus apinorum* (Kwong et al, 2014b; Olofsson et al, 2014)。不属于 *Firm-4* 和 *Firm-5* 的另一种 *Lactobacillus kunkeei* 主要存在于蜂蜜、花粉、蜜蜂蜜囊中,可以诱导产生IgA抗体,提高蜜蜂的免疫响应(Asama et al, 2015)。

(5) 双歧杆菌。*Bifidobacterium* 是另一蜜蜂核心肠道菌,具有丰富的属内多样性,主要分布在蜜蜂



的直肠,丰度较高,在*Apis*属的蜜蜂肠道中丰度范围在10%–30%,在不同熊蜂种群中丰度差异较大(Kwong et al, 2017)。目前从蜜蜂中分离到的*Bifidobacterium*可分为多个菌种(表1),其中*Bifidobacterium asteroides*被分离到的概率最高,说明其分布广、丰度高。1株从西方蜜蜂后肠分离到的该菌种单菌PRL2011于2012年被Bottacini等(2012)测序,证实其富集碳水化合物代谢相关基因,在呼吸代谢上发挥重要的作用。*Bifidobacterium coryneforme*和*Bifidobacterium indicum*由于物种相似度较高,与*Bifidobacterium asteroides*进化距离较近,但系统发育形成一个单独聚类,在本文建立的BGM-Db中被命名为*Bifidobacterium coryneforme/ indicum*组。而从熊蜂中分离到的*Bombiscardovia coagulans*,其16S rRNA序列与这2种菌的相似度较高,在92%左右,但分类到Bifidobacteriaceae目的另一菌属(Killer et al, 2010)。与以上菌种进化距离较远的*Bifidobacterium bombi*、*Bifidobacterium commune*和*Bifidobacterium bohemicum*均从熊蜂中分离得到(Killer et al, 2009, 2011; Praet et al, 2015), 3个菌种通过200多个核糖体基因被聚为一类,在本文建立的BGM-Db中命名为*Bifidobacterium bombi/commune/bohemicum*。

虽然蜜蜂的核心肠道菌具有保守性,但其组成又随分工的不同而有所差别。在西方蜜蜂中,工蜂和蜂王、雄蜂的肠道菌也不同。蜂王肠道菌以 $\alpha$ -变形菌为主,这与工蜂蜜囊中的菌群组成相似。雄蜂的肠道菌群组成与工蜂相似,但是具有更多的*Lactobacillus* (Tarpy et al, 2015)。不同种类的蜜蜂其肠道菌群结构有所不同,不同蜜蜂含有某些特异的核细菌,如东方蜜蜂、大蜜蜂和部分熊蜂中含有Flavobacteriales目的特殊肠道菌,命名为*Apibacter*。该菌目前已知只有4株培养纯菌,分别是从小蜜蜂和大蜜蜂中分离到的*Apibacter adventoris* (Kwong & Moran 2016; Kwong et al, 2017)和从熊蜂中分离到的*Apibacter mensalis* (Praet et al, 2016)。由于已知菌株较少,会影响数据库中*Apibacter*参考序列的解析度,从而影响对16S rRNA测序结果的比对。

**(6) 东方蜜蜂肠道菌*Apibacter*。**为增加数据库中*Apibacter*参考序列的深度,我们从5个省份的东方蜜蜂个体中分离了96株*Apibacter*菌,对其进行接近全长的16S rRNA测序。将测得的序列与数据库中已有的*Apibacter* 16S rRNA序列进行比对并构建进

化树(图1)。系统发育分析结果表明,东方蜜蜂与熊蜂中的*Apibacter*在进化树上处于并行分支,而与西方蜜蜂和大蜜蜂中的序列进化距离较远;我们分离到的东方蜜蜂*Apibacter*与Kwong和Moran (2016)从马来西亚东方蜜蜂中分离得到的菌株按照地理位置的不同而聚类在一起,表明存在宿主与共生*Apibacter*的共进化现象。而不同省份之间的*Apibacter*聚类又有交叉,说明除地理分化的影响外,*Apibacter*属内又可能进化出多种不同的菌种。我们将测得的东方蜜蜂肠道*Apibacter*接近全长的序列添加到对蜜蜂其他肠道核心菌的命名校正过的SILVA数据库中,构建了新的用于蜜蜂肠道菌16S rRNA测序分类的新的数据库,并命名为BGM-Db。

## 2.2 不同数据库的分类结果比较

为比较不同16S rRNA数据库在对蜜蜂肠道菌测序结果分类上的差别,我们分别选取1只西方蜜蜂和1只东方蜜蜂样品,利用Illumina对其肠道菌16S rRNA扩增结果进行高通量测序,将测序结果分别利用BGM-Db与已有的SILVA和RDP数据库进行分类(图2)。

提取西方蜜蜂和东方蜜蜂中丰度前10的分类簇,计算其百分比。其中西方蜜蜂中丰度最高的为 $\alpha$ -变形菌,44%以上的菌种在不同数据库其分类结果有所不同(图2)。BGM-Db将其鉴定为*Brucella*菌属,而其他2个数据库则将其分类为*Ensifer*菌属。*Brucella*和*Ensifer*属于Rhizobiales目不同科的菌属,与西方蜜蜂的核心肠道菌*Bartonella*属于同一个目。而3个数据库的分类结果均未将其分类到*Bartonella*属,甚至*Bartonellaceae*科,可能是由于样品测序污染所致,或者是该西方蜜蜂主要肠道菌与已知的*Bartonella*进化距离较远,分化为新的物种,使其分类到其他科的菌属。另一种 $\alpha$ -变形菌,通过BGM-Db可以将其分类到菌种水平,为蜜蜂肠道菌*Commensalibacter* Alpha-2.1;而SILVA数据库只能将其分类到*Commensalibacter*属,RDP数据库则只能将其分类到Rhodospirillales科。

对于蜜蜂肠道核心菌*Lactobacillus*,利用BGM-Db数据库可以将其准确分类到Firm-4和Firm-5这2个组,其中Firm-5中的菌种丰度明显高于Firm-4。而利用SILVA数据库则只能将其分类到*Lactobacillus*属,另外有0.12%的测序读长只能将其分类到*Lactobacillales*目,0.04%分类到Firmicutes门。RDP

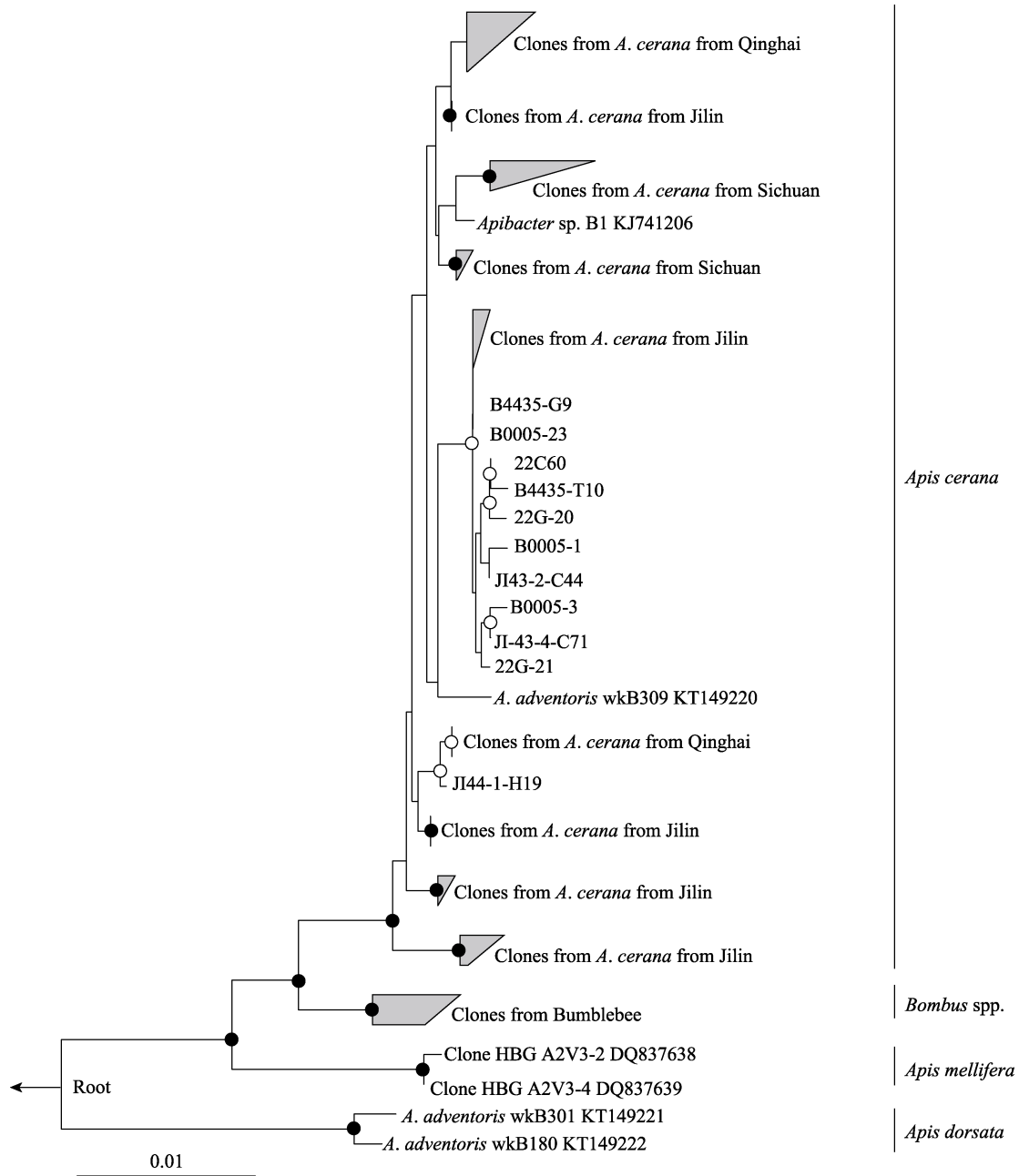


图1 基于接近全长16S rRNA序列(1,168 bp)的东方蜜蜂*Apibacter*菌属系统发育树。节点处白圈表示自展支持率大于75%，黑圈表示大于95%。

Fig. 1 A maximum-likelihood tree of the *Apibacter* genus from *A. cerana* based on the near full-length 16S rRNA sequences (1,168 bp). Bootstrap values are indicated at the branching nodes (● > 95%, ○ > 75%).

数据库的分类结果类似，将其分类到*Lactobacillus*属和*Lactobacillales*目。

对于蜜蜂肠道核心菌*Gilliamella*，BGM-Db和RDP的分类结果比SILVA分别多出了0.03%和0.51%的序列只分类到*Orbaceae*科，这可能是由于SILVA首先对其数据库中所有的序列进行建模聚类，再从每个聚类中挑取代表性的序列用于查询序列的比

对，而BGM-Db和RDP则用其数据库中所有的序列对查询序列进行比对。在比对到*Orbaceae*科的下一分类等级时，可能有多多个属的参考序列与查询序列相似，因此BGM-Db和RDP数据库无法将其准确分类到*Gilliamella*属。这也说明蜜蜂肠道菌存在属间多样性。

与西方蜜蜂相比，东方蜜蜂肠道菌的

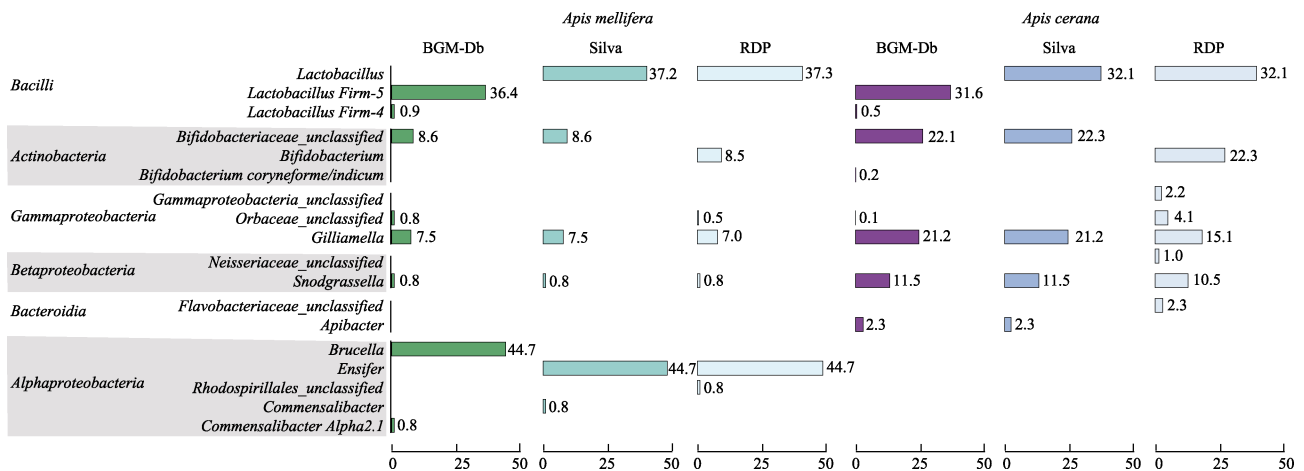


图2 利用BGM-Db、SILVA及Ribosomal Database Project (RDP)数据库对东方蜜蜂和西方蜜蜂16S rRNA测序结果分类比较  
Fig. 2 Comparison of the behaviors of the BGM-Db, SILVA, and Ribosomal Database Project (RDP) databases when they are used in the classification of the gut microbiota of *Apis cerana* and *Apis mellifera*

*Lactobacillus*丰度更高,而 $\alpha$ -变形菌则比较稀少。同样,对于*Lactobacillus*,与其他2个数据库相比,BGM-Db数据库可以将其准确鉴定到菌种水平,说明蜜蜂肠道的*Lactobacillus*属内主要菌种为已知的蜜蜂核心菌Firm-4和Firm-5。而对于*Bifidobacteriaceae*科,相比于SILVA数据库,BGM-Db数据库可以从中分出占比0.19%的*Bifidobacterium*属的*coryneforme*和*indicum*种,说明对数据库进行优化后,可以增加对16S rRNA基因序列的分辨率。对大部分的*Bifidobacteriaceae*,BGM-Db和SILVA都只将其分类到科的级别,而RDP数据库将其全部分类为*Bifidobacterium*属,这可能是由于RDP的参考序列较其他2个数据库少,因此在比对时,排除了其近源菌属的干扰。但较少的参考序列在与其他一些菌属比对时,如对 $\beta$ -变形菌分类时,对于一部分的序列只能将它们分类到科和目的级别,而BGM-Db和SILVA数据库则将绝大部分序列准确地分类到了菌属水平。又如*Flavobacteriaceae*科下的*Apibacter*菌属,作为新分离的蜜蜂肠道菌(Praet et al, 2016; Kwong & Moran, 2016; Kwong et al, 2017)在RDP数据库中未得到及时更新,因此只能将东方蜜蜂中的*Apibacter*序列分类到上一级的*Flavobacteriaceae*科,BGM-Db和SILVA数据库则将其准确分类到*Apibacter*菌属。

### 3 讨论

SILVA和RDP作为全细菌16S rRNA基因序列数据库,对于某些特殊环境菌,如昆虫肠道共生菌缺

乏系统的校正,分辨率较低。我们以SILVA数据库为基础,根据蜜蜂肠道菌的命名规范,对SILVA数据库中的蜜蜂肠道菌分支进行了系统校正和重新命名,并且对于*Flavobacteriaceae*科下新的蜜蜂肠道菌属*Apibacter*,从东方蜜蜂肠道中分离了96株新的菌株,将其16S rRNA全长序列上传到SILVA数据库,更正后的新的蜜蜂肠道特殊分类数据库命名为BGM-Db。相比于SILVA和RDP数据库,BGM-Db对蜜蜂肠道16S rRNA高通量测序短序列实现了菌种级别的分类,分辨率更高;并且系统命名后也更方便将新测得的序列与原有的已知蜜蜂肠道菌进行比较,对于蜜蜂肠道菌的分类鉴定具有重要意义。BGM-Db包括一个16S rRNA序列数据库的FASTA文件及一个物种分类信息的taxonomy文件(<https://www.zhenghaolab.com/bgm-db>)。BGM-Db文件适用于Mothur分析软件,可直接替换Mothur中classify\_seqs步骤的默认文件,相关使用方法可参照Mothur手册([https://www.mothur.org/wiki/Classify\\_seqs](https://www.mothur.org/wiki/Classify_seqs))。

### 参考文献

- Asama T, Arima TH, Gomi T, Keishi T, Tani H, Kimura Y, Tatefuji T, Hashimoto K (2015) *Lactobacillus kunkeei* YB38 from honeybee products enhances IgA production in healthy adults. *Journal of Applied Microbiology*, 119, 818–826.
- Bottacini F, Milani C, Turrone F, Sánchez B, Foroni E, Duranti S, Serafini F, Viappiani A, Strati F, Ferrarini A, Delledonne M, Henrissat B, Coutinho P, Fitzgerald GF, Margolles A, van Sinderen D, Ventura M (2012) *Bifidobacterium asteroides*



- PRL2011 genome analysis reveals clues for colonization of the insect gut. *PLoS ONE*, 7, e44229.
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM (2014) Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, 42, D633–D642.
- Corby-Harris V, Snyder L, Meador CAD, Naldo R, Mott B, Anderson KE (2016) *Parasaccharibacter apium*, gen. nov., sp. nov., improves honey bee (Hymenoptera: Apidae) resistance to *Nosema*. *Journal of Economic Entomology*, 109, 537–543.
- Corby-Harris V, Snyder LA, Schwan MR, Maes P, McFrederick QS, Anderson KE (2014) Origin and effect of Alpha 2.2 Acetobacteraceae in honey bee larvae and description of *Parasaccharibacter apium* gen. nov., sp. nov. *Applied and Environmental Microbiology*, 80, 7460–7472.
- Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan PL, Briese T, Hornig M, Geiser DM, Martinson V, vanEngelsdorp D, Kalkstein AL, Drysdale A, Hui J, Zhai J, Cui L, Hutchison SK, Simons JF, Egholm M, Pettis JS, Lipkin WI (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*, 318, 283–287.
- Ellegaard KM, Tamarit D, Javelind E, Olofsson TC, Andersson SGE, Vásquez A (2015) Extensive intra-phylotype diversity in *Lactobacilli* and *Bifidobacteria* from the honeybee gut. *BMC Genomics*, 16, 284.
- Engel P, Bartlett KD, Moran NA (2015) The bacterium *Frischella perrara* causes scab formation in the gut of its honeybee host. *mBio*, 6, e00193-15.
- Engel P, Kwong WK, Moran NA (2013) *Frischella perrara* gen. nov., sp. nov., a Gammaproteobacterium isolated from the gut of the honeybee, *Apis mellifera*. *International Journal of Systematic and Evolutionary Microbiology*, 63, 3646–3651.
- Engel P, Martinson VG, Moran NA (2012) Functional diversity within the simple gut microbiota of the honey bee. *Proceedings of the National Academy of Sciences, USA*, 109, 11002–11007.
- Engel P, Stepanauskas R, Moran NA (2014) Hidden diversity in honey bee gut symbionts detected by single-cell genomics. *PLoS Genetics*, 10, e1004596.
- Killer J, Kopečný J, Mrázek J, Havlík J, Koppová I, Benada O, Rada V, Kofroňová O (2010) *Bombiscardovia coagulans* gen. nov., sp. nov., a new member of the family Bifidobacteriaceae isolated from the digestive tract of bumblebees. *Systematic and Applied Microbiology*, 33, 359–366.
- Killer J, Kopečný J, Mrázek J, Koppová I, Havlík J, Benada O, Kott T (2011) *Bifidobacterium actinocoloniiforme* sp. nov. and *Bifidobacterium bohemicum* sp. nov., from the bumblebee digestive tract. *International Journal of Systematic and Evolutionary Microbiology*, 61, 1315–1321.
- Killer J, Kopečný J, Mrázek J, Rada V, Benada O, Koppová I, Havlík J, Straka J (2009) *Bifidobacterium bombi* sp. nov., from the bumblebee digestive tract. *International Journal of Systematic and Evolutionary Microbiology*, 59, 2020–2024.
- Kim EK, Kim SH, Nam HJ, Choi MK, Lee KA, Choi SH, Seo YY, You H, Kim B, Lee WJ (2012) Draft genome sequence of *Commensalibacter intestini* A911T, a symbiotic bacterium isolated from *Drosophila melanogaster* intestine. *Journal of Bacteriology*, 194, 1246.
- Klein AM, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, Tscharntke T (2007) Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences*, 274, 303–313.
- Kwong WK, Engel P, Koch H, Moran NA (2014a) Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proceedings of the National Academy of Sciences, USA*, 111, 11509–11514.
- Kwong WK, Mancenido AL, Moran NA (2014b) Genome sequences of *Lactobacillus* sp. strains wkB8 and wkB10, members of the Firm-5 clade, from honey bee guts. *Genome Announcements*, 2, e01176-14.
- Kwong WK, Medina LA, Koch H, Sing KW, Soh EJY, Ascher JS, Jaffé R, Moran NA (2017) Dynamic microbiome evolution in social bees. *Science Advances*, 3, 1–16.
- Kwong WK, Moran NA (2013) Cultivation and characterization of the gut symbionts of honey bees and bumble bees: Description of *Snodgrassella alvi* gen. nov., sp. nov., a member of the family Neisseriaceae of the betaproteobacteria, and *Gilliamella apicola* gen. nov., sp. nov., a member of Orbaceae fam. nov., *Orbales* ord. nov., a sister taxon to the order 'Enterobacteriales' of the Gammaproteobacteria. *International Journal of Systematic and Evolutionary Microbiology*, 63, 2008–2018.
- Kwong WK, Moran NA (2016) *Apibacter adventoris* gen. nov., sp. nov., a member of the phylum *Bacteroidetes* isolated from honey bees. *International Journal of Systematic and Evolutionary Microbiology*, 66, 1323–1329.
- Lane DJ (1991) 16S/23S rRNA Sequencing. In: *Nucleic Acid Techniques in Bacterial Systematics* (eds Stackebrandt E, Goodfellow M), pp. 115–175. John Wiley and Sons, New York.
- Li L, Praet J, Borremans W, Nunes OC, Manaia CM, Cleenwerck I, Meeus I, Smagghe G, De Vuyst L, Vandamme P (2015) *Bombella intestini* gen. nov., sp. nov., an acetic acid bacterium isolated from bumble bee crop. *International Journal of Systematic and Evolutionary Microbiology*, 65, 267–273.
- Martinson VG, Danforth BN, Minckley RL, Rueppell O, Tingek S, Moran NA (2011) A simple and distinctive microbiota associated with honey bees and bumble bees. *Molecular Ecology*, 20, 619–628.
- Martinson VG, Mago T, Koch H, Salzberg SL, Moran NA

- (2014) Genomic features of a bumble bee symbiont reflect its host environment. *Applied and Environmental Microbiology*, 80, 3793–3803.
- Martinson VG, Moy J, Moran NA (2012) Establishment of characteristic gut bacteria during development of the honeybee worker. *Applied and Environmental Microbiology*, 78, 2830–2840.
- Mikaelyan A, Köhler T, Lampert N, Rohland J, Boga H, Meuser K, Brune A (2015) Classifying the bacterial gut microbiota of termites and cockroaches: A curated phylogenetic reference database (DictDb). *Systematic and Applied Microbiology*, 38, 472–482.
- Moran NA (2015) Genomics of the honey bee microbiome. *Current Opinion in Insect Science*, 10, 22–28.
- Newton IL, Roeselers G (2012) The effect of training set on the classification of honey bee gut microbiota using the Naïve Bayesian Classifier. *BMC Microbiology*, 12, 221.
- Olofsson TC, Alsterfjord M, Nilson B, Butler E, Vásquez A (2014) *Lactobacillus apinorum* sp. nov., *Lactobacillus mellifer* sp. nov., *Lactobacillus mellis* sp. nov., *Lactobacillus melliventris* sp. nov., *Lactobacillus kimbladii* sp. nov., *Lactobacillus helsingborgensis* sp. nov. and *Lactobacillus kullabergensis* sp. nov., isolated from the honey stomach of the honeybee *Apis mellifera*. *International Journal of Systematic and Evolutionary Microbiology*, 64, 3109–3119.
- Praet J, Aerts M, Brandt ED, Meeus I, Smagghe G, Vandamme P (2016) *Apibacter mensalis* sp. nov.: A rare member of the bumblebee gut microbiota. *International Journal of Systematic and Evolutionary Microbiology*, 66, 1645–1651.
- Praet J, Meeus I, Cnockaert M, Aerts M, Smagghe G, Vandamme P (2015) *Bifidobacterium commune* sp. nov. isolated from the bumble bee gut. *Antonie van Leeuwenhoek*, 107, 1307–1313.
- Praet J, Parmentier A, Schmid-Hempel R, Meeus I, Smagghe G, Vandamme P (2018) Large-scale cultivation of the bumblebee gut microbiota reveals an underestimated bacterial species diversity capable of pathogen inhibition. *Environmental Microbiology*, 20, 214–227.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO (2007) SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35, 7188–7196.
- Raymann K, Shaffer Z, Moran NA (2017) Antibiotic exposure perturbs the gut microbiota and elevates mortality in honeybees. *PLoS Biology*, 15, e2001861.
- Segers FH, Kešnerová L, Kosoy M, Engel P (2017) Genomic changes associated with the evolutionary transition of an insect gut symbiont into a blood-borne pathogen. *The ISME Journal*, 11, 1232–1244.
- Steele MI, Kwong WK, Whiteley M, Moran NA (2017) Diversification of type VI secretion system toxins reveals ancient antagonism among bee gut microbes. *mBio*, 8, e01630-17.
- Tarpy DR, Mattila HR, Newton ILG (2015) Development of the honey bee gut microbiome throughout the Queen-Rearing Process. *Applied and Environmental Microbiology*, 81, 3182–3191.
- vanEngelsdorp D, Evans JD, Saegerman C, Mullin C, Haugbruge E, Nguyen BK, Frazier M, Frazier J, Cox-Foster D, Chen Y, Underwood R, Tarpy DR, Pettis JS (2009) Colony collapse disorder: A descriptive study. *PLoS ONE*, 4, e6481.
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian Classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73, 5261–5267.
- Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO (2014) The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Research*, 42, D643–D648.
- Yun JH, Lee JY, Hyun DW, Jung MJ, Bae JW (2017) *Bombella apis* sp. nov., an acetic acid bacterium isolated from the midgut of a honey bee. *International Journal of Systematic and Evolutionary Microbiology*, 67, 2184–2188.
- Zheng H, Nishida A, Kwong WK, Koch H, Engel P, Steele MI, Moran NA (2016) Metabolism of toxic sugars by strains of the bee gut symbiont *Gilliamella apicola*. *mBio*, 7, e01326-16.
- Zheng H, Powell JE, Steele MI, Dietrich C, Moran NA (2017) Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. *Proceedings of the National Academy of Sciences, USA*, 114, 4775–4780.

(特邀责任编辑: 周欣 责任编辑: 黄祥忠)



•方法•

# 基于Nextflow构建的宏条形码 自动化分析流程EPPS

李诣远\* David C. Molik Michael E. Pfrender

(Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46554, USA)

**摘要:** 基于宏条形码技术的物种快速检测有助于生物多样性的评估、预测和保护。本文介绍了常用宏条形码分析的步骤和参数设定方法。我们利用Nextflow搭建了一款宏条形码分析流程EPPS, 可以自动化地运行从原始数据的质量控制到环境多样性的比较。Nextflow软件还拥有流程监控的功能, 可视化输出每个进程所消耗的时间与内存。本文还使用测试数据和已发表数据证明该平台能够有效地分析宏条形码数据并可靠地分析环境生物多样性的相似性。

**关键词:** 环境DNA; USEARCH; Trimmomatic; 主成分分析

## EPPS, a metabarcoding bioinformatics pipeline using Nextflow

Yiyuan Li\*, David C. Molik, Michael E. Pfrender

*Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46554, USA*

**Abstract:** Metabarcoding helps to quickly assess biodiversity. In this study, we discuss popular metabarcoding analytical tools and parameter settings. We also develop a metabarcoding bioinformatics pipeline, EPPS, to process data from quality control of raw reads to biodiversity comparisons between samples using a pipeline building program, Nextflow. The EPPS pipeline can summarize the time and memory cost of each process in the pipeline. We also apply the pipeline on a test dataset and a public dataset from a previous study. The result suggests that this pipeline can reliably analyze metabarcoding data and facilitate pipeline sharing of metabarcoding studies.

**Key words:** environmental DNA; USEARCH; Trimmomatic; principal component analysis

生物多样性为人类提供了重要的生态系统服务, 包括洁净的水源和空气、食物、气候调节、碳循环、航运、休闲娱乐等等(Millennium Ecosystem Assessment, 2005)。近年来由于人类活动和环境变化, 生物多样性正受到严重的影响(Worm et al, 2006; Collen et al, 2014; Pimm et al, 2014; Newbold et al, 2015)。对生物多样性全面详细地了解有助于评价其状况, 并预测其发展趋势, 促进保护工作的开展。

随着基因组学技术的发展, 第二代测序(next-generation sequencing)技术越来越多地被应用于生物多样性的调查(Pfrender et al, 2010; Lodge et

al, 2012; Bohmann et al, 2014; Thomsen & Willerslev, 2015; Deiner et al, 2017a; Simon & Evans, 2017), 包括: 生物多样性的监测、入侵物种的监测、食性分析等等。生物多样性调查时, 往往通过检测基因组的一段或多段DNA序列(如: DNA条形码)与已知物种的DNA序列的匹配来实现物种的快速分类(Taberlet et al, 2012)。常见的技术包括: 宏条形码(metabarcoding)技术、线粒体基因组捕获技术(Dowle et al, 2016; Liu et al, 2016; Wilcox et al, 2018)和全线粒体基因组测序技术(Zhou et al, 2013; Crampton-Platt et al, 2015, 2016; Tang et al, 2015; Deiner et al, 2017b; Bista et al, 2018)。由于宏条形码

技术成本低, 实验操作简便, 所需的DNA起始量低, 因此应用最为普遍。宏条形码的一般流程是提取混合物种的DNA, 并通过PCR扩增目标DNA片段, 利用高通量测序获取PCR扩增子的DNA序列, 并通过生物信息学方法进行多样性分析(Thomsen et al, 2012; Ji et al, 2013; Liu et al, 2013; Evans et al, 2016, 2017; Olds et al, 2016; Li et al, 2018)。

基于PCR的宏条形码生物信息分析流程的基本步骤包括: 测序质量控制, PCR引物的删除, 双向序列的拼接, 分子可操作分类单元(molecular operational taxonomic unit, MOTU)的聚类分析, 生物多样性分析以及物种分类(图1)。目前宏条形码的常用分析软件有很多, 如QIIME (Caporaso et al, 2010)、DADA2 (Callahan et al, 2016)、Mothur (Schloss et al, 2009)、USEARCH (Edgar, 2010, 2013)、VSEARCH (Rognes et al, 2016)、obitools (Boyer et al, 2016)等。其中, 有很多软件开发时主要是针对微生物宏基因组16S rRNA基因的生物信息分析, 如QIIME、DADA2、Mothur和USEARCH。在这些软件中, QIIME主要基于python并且整合了各种宏基因组分析的软件。DADA2、Mothur和USEARCH则提供了一个一体化的分析流程, 自身包含了分析所需的所有命令。obitools是一款面向宏条形码生物信息

分析的软件, 与上述几个软件不同的是, obitools包含了很多为宏条形码设计的功能, 例如, 提供引物设计的帮助和针对真核生物的物种鉴定。除了通用的宏条形码分析软件, Sato等(2018)还设计了针对鱼类的宏条形码分析软件MiFish。MiFish提供了在线鱼类宏条形码序列分析的解决方案, 通过USEARCH和BLAST软件将序列与鱼类宏条形码数据库MitoFish比对获得多样性信息。虽然以上软件包含了大部分宏条形码的分析步骤, 但分析的步骤依旧需要调用众多的软件和命令, 针对不同的研究宏条形码的分析参数也有差异。因此快捷地调用和调整不同的程序和命令可以极大地提高宏条形码分析的效率。

近年来在微生物领域已经有多个研究致力于设计宏基因组(Piro et al, 2017; Urtskiy et al, 2018; Visconti et al, 2018)和16S rRNA数据(<https://github.com/h3abionet/h3abionet16S>)的分析流程。但在宏条形码研究领域, 还没有一套可以快速建立分析大量样品的分析流程。面对越来越多的第二代测序数据和宏条形码样本, 一套能够在不同平台上快速建立并根据样品情况调整参数的信息分析流程将有助于快速分析大量样品, 并可以针对不同的样品调整流程。本文利用Nextflow (Di Tommaso et al, 2017)搭建了一款宏条形码分析流程EPPS。Nextflow是一款基于Groovy语言的流程管理系统(workflow management system), 可以方便快捷地调度程序和中间文件, 降低了创建流程的复杂性。同时Nextflow支持Linux、MacOS和云计算平台, 便于将流程扩展到不同的计算平台。本文将详细介绍宏条形码分析中不同生物信息分析步骤的推荐参数并介绍如何修改参数, 方便其他宏条形码研究使用。借助于Nextflow的优势, EPPS可以让研究者快速地建立宏条形码的分析流程并调整参数, 同时支持并行计算, 能够分析和监控大量宏条形码样品的分析。EPPS的分析流程可以从GitHub下载([https://github.com/lyy005/epps\\_nf](https://github.com/lyy005/epps_nf))。

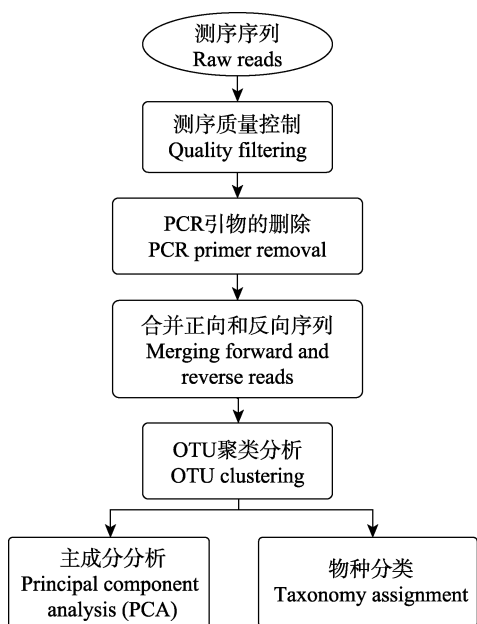


图1 EPPS的主要分析步骤。OTU聚类分析还包括去除重复序列、OTU聚类和嵌合体的检测。

Fig. 1 The workflow of EPPS. OTU clustering step includes, dereplication, OTU clustering and chimera detection.

## 1 EPPS流程设计

EPPS的分析流程包括: 原始测序结果的质量控制, 引物序列的删除, 正向和反向序列的合并, OTU聚类分析和群落多样性分析(主成分分析)(图1)。EPPS使用Nextflow工作流程语言, 简化了用



户在分析过程中的重复操作,使其可以一键完成大量宏条形码样品的分析,并提供了进程所占的内存和时间消耗,便于用户对进程的监控。EPPS流程提供常用的OTU表格输出格式,便于用户将结果输入其他的后期分析软件中,如QIIME和Phinch (Bik & Interactive Pitch Inc., 2014)。

(1)软件安装。EPPS流程基于Linux和MacOS操作系统。用户在运行EPPS之前需要安装Nextflow、Java 7、R (R Core Team, 2016)和USEARCH软件。R软件可以从<https://www.r-project.org>下载安装。USEARCH软件可以从<http://drive5.com/usearch/download.html>下载安装。另外, EPPS流程还包含了VSEARCH的可执行文件,用户可以使用命令:

```
git clone git@github.com:lyy005/epps_nf.git  
或者前往https://github.com/lyy005/epps\_nf下载流程。
```

(2)测序文件的输入。EPPS分析流程的输入文件分为两部分:第一部分是fastq格式的Illumina双向测序结果文件;第二部分是一个名为primer.fas的fasta格式文件,用来指定宏条形码使用的引物序列。EPPS下载完成后,用户使用Linux或者MacOS终端命令进入EPPS文件夹: `cd epps_nf`。文件夹里包含3个文件夹: `./bin`、`./example`和`./input`。`./bin`文件夹包含运行脚本所需的所有程序,下载的USEARCH可执行文件也需要添加到bin文件夹中。`example`文件夹包含有可以直接运行的测试文件,其中包括8个fastq文件,分别是4个测试样品的正向和反向测序文件,并且文件需要以“样品名.1.fastq”和“样品名.2.fastq”的方式命名。`example`文件夹中还包括1个名为primer.fas的文件,该文件包含有宏条形码测序的引物序列。宏条形码测序时,通常是将不同样品的标签序列(index)添加到Illumina平台的测序接头(adaptor)上,这样Illumina测序平台可以直接将不同样品分流(demultiplexing),生成每个样品的fastq文件。具体实验方法可以参见: <http://www.earthmicrobiome.org/protocols-and-standards/16s/> (Caporaso et al, 2011)。默认的测序文件输入格式后缀为fq。如果输入文件格式为fq.gz或者fastq.gz,用户可以将epps\_v190209.nf第4行的命令改成相应的格式。例如,如果是fastq.gz格式,可以将命令改为:  
`params.reads = "$PWD/input/*{1,2}.fastq.gz"`

分析前先将输入文件(fastq文件和primer.fas文件)复制到input文件夹中,再使用命令

```
./nextflow run epps_v190209.nf -with-timeline
```

运行分析流程并获得分析结果。EPPS流程分析测试样品的时间和内存消耗参见图2。本文将通过流程自带的4个测试样品test1到test4具体介绍工作流程和分析结果。使用“-with-timeline”命令会产生一个时间消耗文件(timeline.html),如图3。该文件包括了EPPS流程中每个进程消耗的时间以及虚拟存储的峰值。

(3)测序质量控制。首先,输入的测序文件(fastq格式)会使用Trimmomatic (version 0.38)软件进行严格的质量控制。Trimmomatic的输入参数如下:“ILLUMINACLIP:combined.Illumina.fasta:3:30:6:1:true SLIDINGWINDOW:10:20 MINLEN:50”。Trimmomatic首先使用16 bp的种子序列(seed)与Illumina的测序接头进行匹配。如果有≤ 3 bp的错配,种子序列将继续延长匹配。接着,Trimmomatic会计算匹配得分,如果双向测序序列与测序引物接头的匹配达到30分(约50 bp的匹配)或者单向测序序列与测序引物接头的匹配达到10分(大约17 bp的匹配),序列的匹配部分将会被删除。Trimmomatic还会使用一个10 bp长的滑动窗口(sliding window)为测序序列计算平均序列质量值。如果平均序列质量值小于20,该序列将不会被使用。最后,任何< 50 bp的经过测序接头筛查和质量控制的序列均不会被后续分析使用。只有正向和反向序列同时通过了Trimmomatic的质量控制,该序列才会被用于后续分析。用户可以在epps\_v190209.nf文件的25行对Trimmomatic的参数或者测序接头文件进行修改。

Trimmomatic通常包含4个fastq格式的输出文件,其中有2个配对的fastq文件,包含正向和反向都已通过质量控制的序列。还有2个不配对的fastq文件,包含了只有正向或者只有反向通过质量控制的序列。EPPS中只保留了正反双向序列都通过Trimmomatic质量控制的fastq文件。该结果包含在output文件夹中,文件名分别是“样品名称.pe.1.fq”和“样品名称.pe.2.fq”。

(4) PCR引物的删除。PCR引物通常加在PCR扩增子的两端。去除PCR引物有多种方法。由于引物序列通常是由合成的引物连接在扩增子的末端并进行PCR扩增, EPPS流程使用PCR引物完全匹配的方法进行引物序列的查找与删除。同时,由于引物通常位于测序序列的开端,且有很高的测序质量。

## Processes execution timeline

Launch time: 10 Feb 2019 00:05

Elapsed time: 22.7 s

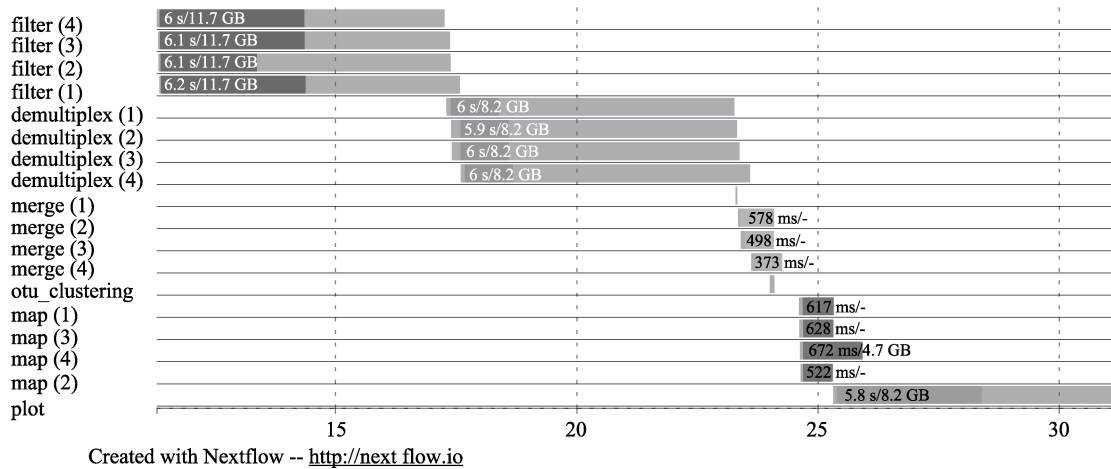


图2 EPPS流程每一个进程的时间消耗。横坐标代表时间, 单位是秒。最左列的名称分别对应了宏条形码分析的流程。filter: 测序质量控制; demultiplex: PCR引物的删除, 如果有多个引物则将各个引物分开; merge: 合并正向和反向序列; otu\_clustering/map: OTU聚类分析; plot: 主成分分析。由于测试数据有4个样品, 因此每个进程的右侧括号里有1-4的序号。浅色进度条代表进程所消耗的系统时间。深色进度条代表的是每个进程的CPU时间。每个进度条包含有两个数字, 第1个数字代表每个进度的系统时间, 第2个数字代表虚拟内存的峰值。

Fig. 2 The timeline chart of EPPS. The x-axis is the amount of time for each process in seconds. Each row indicates the name of different stages of the analysis. filter, Data filtering; demultiplex, Primer removal and demultiplex if there are multiple primers; merge, Merging of forward and reverse reads; otu\_clustering and map, OTU clustering and mapping of reads; plot, Plotting PCA plot. As there are four samples in the testing data set, there are four processes (1 to 4) for filter, demultiplex, merge, and map steps. Each bar indicates the time for each process. The dark area in each bar represents the real execution time. Each bar displays two numbers: the task duration time and the virtual memory size peak.

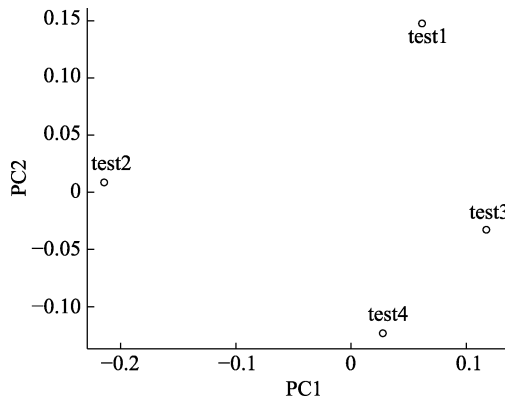


图3 基于测试数据的主成分分析结果。图中每一个点代表一个测试数据的样品。点与点之间距离越近代表样品之间的物种组成相似度越高。例如, test3和test4的相似度大于test3和test1的相似度。

Fig. 3 PCA plot based on testing data. Each dot in the figure represents a test sample. The distance between dots indicates the dissimilarity between samples. For example, the similarity between test3 and test4 is higher than test1 and test2.

如果在序列开端出现引物的错配, 说明该序列是由于PCR错误或者测序错误而导致序列的不一致。因此, EPPS通过完全匹配引物可以进一步筛选序列的

测序质量。其他的分析软件中, QIIME通过局部比对 (truncate\_reverse\_primer.py命令)匹配并删除引物序列, 默认值是容许引物序列与测序结果有2个错配 (mismatch)。DADA2通过在序列的开端截取指定长度的碱基删除引物序列。除此之外, 用户也可以使用cutadapt (Martin, 2011)或Trimmomatic (Bolger et al, 2014)进行引物序列的匹配。

EPPS流程中使用了Perl语言脚本对PCR引物序列进行匹配和删除。Perl语言脚本搜索正向和反向测序序列中完全匹配的引物序列。通过匹配的PCR引物序列, 所有的扩增子均被调整为同一个方向, 以方便下一步OTU聚类分析。测序的结果同样可以在output文件夹中找到。命名的格式为“样品名称.demul.F.fq”和“样品名称.demul.R.fq”。“样品名称.demul.F.fq”包含所有匹配正向PCR引物的序列, “样品名称.demul.R.fq”包含了所有匹配反向PCR引物的序列。

(5)合并正向和反向序列。合并正反向序列是宏条形码分析的重要一步。当PCR扩增子小于两端测

序读长时,宏条形码分析流程通常会利用正向和反向测序序列的重叠部分将它们合并为1个序列。这样做的优点有两个:(1)合并测序序列可以保证PCR扩增子的完整,不会出现由于扩增子部分的缺失影响OTU聚类;(2)合并测序序列可以利用正反向序列重叠的部分纠正测序序列末端的错误,从而提高测序序列的质量值(Zhou et al, 2011; Masella et al, 2012; Edgar, 2013)。EPPS流程使用USEARCH (version 10.0.240) `fastq_mergepairs`命令合并正向和反向的测序结果,并利用USEARCH针对合并的正反向序列检测测序质量。如果测序结果的期望测序错误率高于0.5,该序列将被删除。相比于USEARCH默认的期望错误率(1.0),EPPS流程选用了相对严格的错误率筛选以提高最终物种多样性分析的可靠度。合并之后的结果存放在output文件夹中,命名的方式为“样品名称.merged.rename.fasta”。在不同分析流程中都有合并这一步,例如:QIIME的`vsearch join-pairs`命令(Rognes et al, 2016),DADA2的`mergePairs`命令,Mothur的`make.contigs`命令,USEARCH的`fastq_mergepairs`命令,obitools的`obijoin-pairedend`命令。这些程序通常会重新计算合并后序列的错误率(Edgar, 2013; Rognes et al, 2016)。

(6)去除重复序列和OTU聚类分析。在合并正反向序列之后,EPPS会把相同的序列合并以提升聚类分析的效率。合并的序列的丰度(read number)会被保留,序列将会按照丰度从高到低排序。EPPS使用USEARCH (version 10.0.240) `usearch_global`命令或者VSEARCH (version 2.10.4)的`cluster_size`命令对前一步获得的序列进行聚类分析。聚类分析使用97%的全局序列相似度。如果需要使用不同的聚类相似度,可以修改`epps_v190209.nf` 88行`-id 0.97`。1个OTU里丰度最高的序列会被作为OTU的代表序列,用于后续的比较分析和物种分类。OTU聚类的结果存放在`step1_otu_clustering/otus.fasta`文件中。每个样品的测序序列再通过USEARCH或者VSEARCH的`usearch_global`命令利用全局比对方法和OTU的代表序列进行比对,以获取OTU在每个样品中的丰度。每个样品的比对结果存放在`step2_mapping`文件夹中,命名的方式为“样品名称.uc”。最后,EPPS使用perl脚本将不同的比对结果合并,结果存放在“`combined.uc.table`”文件中。该文件的每一列代表1个样品,每一行是1个OTU的丰度。

通常在合并正反向序列之后,可以选择直接进行序列比对或者OTU的聚类分析。在有相对完整的参考序列数据库(reference database)时,通过将序列与数据库进行比对可以直接获得样品中的物种多样性信息。常见的比对软件包括BLAST, UCLUST和QIIME。当数据库不完备时,为了获得相对完整的物种信息,通常信息分析流程使用OTU聚类进行物种多样性分析。常用的聚类软件包括QIIME、USEARCH、CROP、SWARM。根据样品中物种亲缘关系的差异,聚类的相似度一般设为97%–99%。

(7)检测嵌合体。宏条形码的PCR过程中会导致PCR嵌合体(chimera)的产生。检测嵌合体通常有两种方法:基于参考序列;从头检测。EPPS流程里USEARCH `cluster_otus`命令自带从头检测嵌合体的方法。而VSEARCH不能自动检测嵌合体。如果用户使用VSEARCH,需要额外运行从头检测嵌合体`uchime_denovo`命令(`epps_v190209.nf` 92行)。

(8)主成分分析。用户获得OTU的序列和OTU在样品中的分布之后,就可以进行样品内部的多样性( $\alpha$ 多样性)、样品之间的多样性( $\beta$ 多样性)或者整体样品的多样性( $\gamma$ 多样性)分析。其中, $\beta$ 多样性分析可以通过主成分分析完成。通过比较样品之间物种多样性组成的差异将样品之间的相似性可视化。计算样品与样品间的距离( $\beta$ 多样性)有多种方法(Cardoso et al, 2009),包括Jaccard相似性指数、Shannon相似性指数、Sørensen相似性指数等等。由于PCR的影响,基于PCR的宏条形码的序列数量往往不被作为物种丰度和生物量的参考(Tang et al, 2015; Deiner et al, 2017a; Bista et al, 2018),因此EPPS流程使用Jaccard相似性指数计算样品间的相似性。

EPPS流程使用R语言和R分析包ggplot2 version 3.0.0.9 (Wickham, 2016),vegan version 2.4.6 (Oksanen et al, 2013)和ggrepel version 0.8.0 (Slowikowski, 2018)计算样品间的差异并作图。主成分分析的结果为PDF格式的文件,存放在`step3_profiling_table/plot.pdf`中。该文件中不同的点代表不同的宏条形码样品,样品间的距离越近代表相似度越高。

(9)物种分类。由于不同的物种类群常常有不同的参考序列集和分类方法,EPPS流程本身不包含物种分类的软件。用户可以基于EPPS流程产生的OTU代表序列,并根据研究的需要采用不同的方法进行分类,常用的有三种:基于相似度分类、基于基因



序列特征分类、基于系统发育树分类(Bazinet & Cummings, 2012)。(1)基于相似度分类。通过将测序序列与已知物种的序列比对来获得物种分类结果。常见的分类软件包括BLAST (Camacho et al, 2009)、MEGAN (Huson et al, 2007)、MetaPhyler (Liu et al, 2010)和CARMA (Gerlach & Stoye, 2011)。该方法的优点在于在参考序列集完整的时候准确性非常高,所以参考序列集的完整性对物种分类的影响非常显著。(2)基于基因序列特征的分类。通常是利用DNA序列的特征或者kmer频率进行物种分类。常见的分类软件包括: SINTAX (Edgar, 2016)、RDP (Wang et al, 2007)、NBC (Rosen et al, 2010)、PhyloPythiaS (Patil et al, 2012)、Phymm和PhymmBL (Brady & Salzberg, 2009)。这一类方法的优势在于一旦模型建立,对测序序列分类的速度非常快,缺点在于如果基因序列长度较短会限制有效的序列特征的数量,从而影响分类的准确性。(3)基于系统发育树的分类。该方法是三种方法中最准确的,并且由于分类的方法建立在系统发育树的基础上,可以精确到不同分类阶元。该方法的限制也很明显,为了构建系统发育树,测序序列与参考序列集需要进行多序列联配和系统发育树的构建,这两个步骤往往需要消耗较多的计算资源和时间。基于系统发育树的软件包括pplacer (Matsen et al, 2010)、EPA (Berger et al, 2011)和FastTree (Price et al, 2009)。同时,也有软件同时利用多种方法进行分类,例如Statistical Assignment Package (SAP) (Munch et al, 2008)。不同研究可以根据参考序列集的完整情况和分析数据量的大小进行选择。

如果用户没有事先准备高质量参考序列集进行分类,可以使用在线SAP软件(<https://services.birc.au.dk/sap/server>)。SAP首先通过NetBlast将每个OTU的代表序列比对到NCBI NR数据库。从NCBI比对获得的同源序列会进行多序列联配,并通过对代表序列和同源序列构建系统发育树计算代表序列属于特定物种的贝叶斯后验概率。

## 2 测试数据以及EPPS的输出结果

为了方便测试, EPPS自带了模拟数据。该数据模拟了12,500条Illumina测序结果,每条序列的读长为300 bp。另外,我们还使用了已发表的公共数据(Li et al, 2018)来评估该流程分析较大样本量的表现。

### (1)模拟数据

模拟数据存放于./example文件夹中。开始分析之前,用户首先需要将example文件夹里所有fastq文件和primer.fas文件复制到input文件夹中:

```
cp ./example/* ./input/
```

运行命令:

```
nextflow run epps_v180726.nf -with-timeline
```

重要的结果文件存放在文件夹: ./output/step3\_profiling\_table/。combined.uc.table文件是OTU表格文件。plot.pdf文件是主成分分析文件,该文件主要是基于OTU表格进行的主成分分析。从图3可以看出, test3和test4的物种组成相似度较高, test1和test2的物种组成相似度较低。在资源消耗方面, EPPS流程需要大约15s (图2),消耗54 M的硬盘空间,大约是原始测序数据所占空间(63 M)的0.86倍。

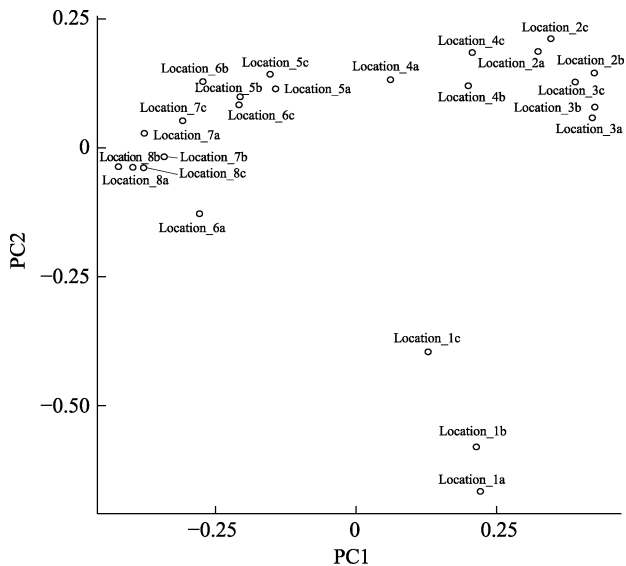
### (2)公共数据

为了进一步证明流程的可靠性,我们分析了已经发表的宏条形码数据(Li et al, 2018)。Li等(2018)沿Kalamazoo河支流Eagle溪的上游至下游选择了8个样品采集点(分别以Location 1–8命名)。每个采集点采集3个水体样品(分别以a, b, c命名),用来研究水体中鱼类组成的多样性。8个采集点中,采集点1是河流的最上游,水流量非常小。采集点2–6的水流量逐渐增加。采集点7–8位于Kalamazoo河的干流中,水流量最大。该研究发现从上游到下游,物种多样性的相似度逐渐下降。采集点内的样品的相似性显著大于采集点间的相似性。8个样品的数据可以分别从NCBI SRA SRS2894037–SRS2894043, SRS2894045中下载。

将下载的fastq文件存放在./input文件夹后,运行EPPS并获得结果(图4)。从图4可以看出,最上游的采样点有独特的鱼类多样性。位于河流中游的采样点2–6有相似的鱼类多样性组成。采样点7–8也有较为类似的鱼类多样性,与文中的结论一致(Li et al, 2018)。在资源消耗方面,由于EPPS全部为自动化,省去了人工的时间,流程仅需要12 min即可完成24个样品的分析,消耗18 G的硬盘空间,大约是原始测序数据(6.8 G)的2.6倍。

EPPS有助于研究结果的重复分析及不同研究分析方法的共享。通过分析测试数据和已发表的公共数据,我们获得了可靠的多样性分析结果,证明了EPPS是一款快速可靠的宏条形码分析流程。





**图4** 公共数据的分析结果。样品的名称编号1–8分别代表从最上游到下游的8个采样点。编号的后缀a, b, c分别代表同一个采样地点的3次独立的重复取样。基于图中的结果, 最上游的样品Location 1有独特的鱼类多样性组成。Location 3–6有类似的鱼类多样性组成。最下游的样品Location 7–8有类似的鱼类多样性组成。

Fig. 4 EPPS result based on public data set. Samples are named from 1 to 8 from upstream to downstream. The suffix “a”, “b” and “c” indicate three replicates of the same sampling location. Based on the PCA, the most upstream sample (Location 1) has unique fish composition. Location 3–6 have similar fish composition. The downstream samples (Location 7–8) share similar fish composition.

## 参考文献

- Bazinet AL, Cummings MP (2012) A comparative evaluation of sequence classification programs. *BMC Bioinformatics*, 13, 92.
- Berger SA, Krompass D, Stamatakis A (2011) Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Systematic Biology*, 60, 291–302.
- Bik HM, Interactive Pitch Inc. (2014) Phinch: An interactive, exploratory data visualization framework for–Omic datasets. *bioRxiv*, 009944.
- Bista I, Carvalho GR, Tang M, Walsh K, Zhou X, Hajibabaei M, Shokralla S, Seymour M, Bradley D, Liu S, Christmas M (2018) Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources*, 18, 1020–1034.
- Bohmann K, Evans A, Gilbert MT, Carvalho GR, Creer S, Knapp M, Douglas WY, De Bruyn M (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29, 358–367.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120.
- Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E (2016) obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, 16, 176–182.
- Brady A, Salzberg SL (2009) Phymm and PhymmBL: Metagenomic phylogenetic classification with interpolated Markov models. *Nature Methods*, 6, 673.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 421.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences, USA*, 108, 4516–4522.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7, 335.
- Cardoso P, Borges PA, Veech JA (2009) Testing the performance of beta diversity measures based on incidence data: The robustness to undersampling. *Diversity and Distributions*, 15, 1081–1090.
- Collen B, Whitton F, Dyer EE, Baillie JE, Cumberlidge N, Darwall WR, Pollock C, Richman NI, Soulsby AM, Böhm M (2014) Global patterns of freshwater species diversity, threat and endemism. *Global Ecology and Biogeography*, 23, 40–51.
- Crampton-Platt A, Timmermans MJ, Gimmel ML, Kutty SN, Cockerill TD, Vun Khen C, Vogler AP (2015) Soup to tree: The phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample. *Molecular Biology and Evolution*, 32, 2302–2316.
- Crampton-Platt A, Douglas WY, Zhou X, Vogler AP (2016) Mitochondrial metagenomics: Letting the genes out of the bottle. *GigaScience*, 5, 15.
- Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, de Vere N, Pfrender ME (2017a) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26, 5872–5895.
- Deiner K, Renshaw MA, Li Y, Olds BP, Lodge DM, Pfrender ME (2017b) Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods in Ecology and Evolution*, 8, 1888–1898.
- Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C (2017) Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35, 316.
- Dowle EJ, Pochon X, Banks JC, Shearer K, Wood SA (2016) Targeted gene enrichment and high-throughput sequencing

- for environmental biomonitoring: A case study using freshwater macroinvertebrates. *Molecular Ecology Resources*, 16, 1240–1254.
- Edgar RC (2016) SINTAX: A simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *bioRxiv*, 074161.
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460–2461.
- Edgar RC (2013) UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10, 996.
- Evans NT, Li Y, Renshaw MA, Olds BP, Deiner K, Turner CR, Jerde CL, Lodge DM, Lamberti GA, Pfrender ME (2017) Fish community assessment with eDNA metabarcoding: Effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, 74, 1362–1374.
- Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, Mahon AR, Pfrender ME, Lamberti GA, Lodge DM (2016) Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*, 16, 29–41.
- Gerlach W, Stoye J (2011) Taxonomic classification of metagenomic shotgun sequences with CARMA3. *Nucleic Acids Research*, 39, e91.
- Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Research*, 17, 377–386.
- Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, Nakamura A, Kitching R, Dolman PM, Woodcock P, Edwards FA, Larsen TH (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, 16, 1245–1257.
- Li Y, Evans NT, Renshaw MA, Jerde CL, Olds BP, Shogren AJ, Deiner K, Lodge DM, Lamberti GA, Pfrender ME (2018) Estimating fish alpha- and beta-diversity along a small stream with environmental DNA metabarcoding. *Metabarcoding and Metagenomics*, 2, e24262.
- Liu B, Gibbons T, Ghodsi M, Pop M (2010) MetaPhyler: Taxonomic profiling for metagenomic sequences. In: *Bioinformatics and Biomedicine (BIBM)*, 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 95–100.
- Liu S, Wang X, Xie L, Tan M, Li Z, Su X, Zhang H, Misof B, Kjer KM, Tang M, Niehuis O (2016) Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources*, 16, 470–479.
- Liu S, Li Y, Lu J, Su X, Tang M, Zhang R, Zhou L, Zhou C, Yang Q, Ji Y, Yu DW (2013) SOAPBarcode: Revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons. *Methods in Ecology and Evolution*, 4, 1142–1150.
- Lodge DM, Turner CR, Jerde CL, Barnes MA, Chadderton L, Egan SP, Feder JL, Mahon AR, Pfrender ME (2012) Conservation in a cup of water: Estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, 21, 2555–2558.
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17, 10–12.
- Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD (2012) PANDAsq: Paired-end assembler for Illumina sequences. *BMC Bioinformatics*, 13, 31.
- Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: Linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11, 538.
- Millennium Ecosystem Assessment (2005) *Ecosystem and Human Well-being: Biodiversity Synthesis*. World Resources Institute, Washington, DC.
- Munch K, Boomsma W, Huelsenbeck JP, Willerslev E, Nielsen R (2008) Statistical assignment of DNA sequences using Bayesian phylogenetics. *Systematic Biology*, 57, 750–757.
- Newbold T, Hudson LN, Hill SL, Contu S, Lysenko I, Senior RA, Börger L, Bennett DJ, Choimes A, Collen B, Day J (2015) Global effects of land use on local terrestrial biodiversity. *Nature*, 520, 45.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, Simpson GL, Solymos P, Stevens MH, Wagner H (2013) Package 'vegan'. *Community Ecology Package*, version. 2. (accessed on 2018-08-01)
- Olds BP, Jerde CL, Renshaw MA, Li Y, Evans NT, Turner CR, Deiner K, Mahon AR, Brueseke MA, Shirey PD, Pfrender ME (2016) Estimating species richness using environmental DNA. *Ecology and Evolution*, 6, 4214–4226.
- Patil KR, Rouné L, McHardy AC (2012) The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. *PLoS ONE*, 7, e38581.
- Pfrender M, Hawkins C, Bagley M, Courtney G, Creutzburg B, Epler J, Fend S, Ferrington L Jr, Hartzell P, Jackson S, Larsen D (2010) Assessing macroinvertebrate biodiversity in freshwater ecosystems: Advances and challenges in DNA-based approaches. *The Quarterly Review of Biology*, 85, 319–340.
- Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO (2014) The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, 344, 1246752.
- Piro VC, Matschkowski M, Renard BY (2017) MetaMeta: Integrating metagenome analysis tools to improve taxonomic profiling. *Microbiome*, 5, 101.
- Price MN, Dehal PS, Arkin AP (2009) FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, 26, 1641–1650.
- R Core Team (2016) *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/>. (accessed

- on 2018-08-01)
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584.
- Rosen GL, Reichenberger ER, Rosenfeld AM (2010) NBC: The Naive Bayes Classification tool webserver for taxonomic classification of metagenomic reads. *Bioinformatics*, 27, 127–129.
- Sato Y, Miya M, Fukunaga T, Sado T, Iwasaki W (2018) MitoFish and MiFish pipeline: A mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding. *Molecular Biology and Evolution*, 35, 1553–1555.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW (2009) Introducing Mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75, 7537–7541.
- Simon TP, Evans NT (2017) Environmental quality assessment using stream fishes. In: *Methods in Stream Ecology*, 3rd edn. (eds Hauer FR, Lamberti G), pp. 319–334. Elsevier, London.
- Slowikowski K (2018) ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. <https://CRAN.R-project.org/package=ggrepel>. (accessed on 2018-08-01)
- Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Environmental DNA. *Molecular Ecology*, 21, 1789–1793.
- Tang M, Hardman CJ, Ji Y, Meng G, Liu S, Tan M, Yang S, Moss ED, Wang J, Yang C, Bruce C (2015) High-throughput monitoring of wild bee diversity and abundance via metagenomics. *Methods in Ecology and Evolution*, 6, 1034–1043.
- Thomsen PF, Kielgast JO, Iversen LL, Wiuf C, Rasmussen M, Gilbert MT, Orlando L, Willerslev E (2012) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21, 2565–2573.
- Thomsen PF, Willerslev E (2015) Environmental DNA—An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183, 4–18.
- Uritskiy GV, DiRuggiero J, Taylor J (2018) MetaWRAP—A flexible pipeline for genome-resolved metagenomic data analysis. *bioRxiv*, 277442.
- Visconti A, Martin TC, Falchi M (2018) YAMP: A containerised workflow enabling reproducibility in metagenomics research. *GigaScience*, 7, giy072.
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73, 5261–5267.
- Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York. <http://ggplot2.org>. (accessed on 2018-08-01)
- Wilcox TM, Zarn KE, Piggott MP, Young MK, McKelvey KS, Schwartz MK (2018) Capture enrichment of aquatic environmental DNA: A first proof of concept. *Molecular Ecology Resources*, 18, 1392–1401.
- Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS, Jackson JB, Lotze HK, Micheli F, Palumbi SR, Sala E (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science*, 314, 787–790.
- Zhou HW, Li DF, Tam NF, Jiang XT, Zhang H, Sheng HF, Qin J, Liu X, Zou F (2011) BIPES, a cost-effective high-throughput method for assessing microbial diversity. *The ISME Journal*, 5, 741.
- Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, 2, 4.

(特邀责任编辑: 周欣 责任编辑: 闫文杰)