目录

[**微生物组分析方法、应用与传播：现状与对策** 1](#_Toc26827268)

[**1. 方法优化和开发** 2](#_Toc26827269)

[**方法学的进展和存在问题** 2](#_Toc26827270)

[**已经发表的一些工作** 2](#_Toc26827271)

[**下一步工作重点** 3](#_Toc26827272)

[**2. 合作分析** 3](#_Toc26827273)

[**实验和计算生物学家的优缺点** 3](#_Toc26827274)

[**合作项目案例** 4](#_Toc26827275)

[**下一步工作计划** 4](#_Toc26827276)

[**3. 科学传播** 4](#_Toc26827277)

[**微生物组结合“互联网+”知识传播和教学** 5](#_Toc26827278)

[**综述撰写** 5](#_Toc26827279)

[**4. 现状和对策总结** 5](#_Toc26827280)

**微生物组分析方法、应用与传播：现状与对策**

刘永鑫1,2,3,4

1. 中国科学院遗传与发育生物学研究所，植物基因组学国家重点实验室
2. 中国科学院大学，生物互作卓越创新中心
3. 中国科学院-英国约翰英纳斯中心植物和微生物科学联合研究中心
4. 宏基因组公众号

总结自2019年12月7日，在中国科学院微生物研究所E201王军研究员主持的“中国微生物组研究现状与对策研讨会”上我的报告“**微生物组分析方法、应用与科学传播**”。

我叫刘永鑫，在东北农业大学获得微生物学学士和作物遗传育种硕士。14年加入中科院遗传发育所王秀杰组攻读生物信息学博士；毕业后在韩方普组从事博士后科研工作2年。于16年加入白洋组，任工程师，负责微生物组数据分析。目前我在Science[1], Nature Biotechnology[2] 等杂志发表论文18篇，其中第一作者(共同)论文9篇。应邀在遗传[3] 和Current Opinion in Microbiology[4] 等杂志发表综述文章。

我关注研究方向包括微生物组分析方法、应用与科学传播三个主要方向，我将从以下三个方向进行介绍.

**1. 方法优化和开发**

**方法学的进展和存在问题**

目前微生物组分析领域的处于高速发展阶段，方法和软件百花齐放、百家争鸣，很多软件形成三足鼎立(如扩增子分析引用万的流程有mothur[5]、QIIME[6]和USEARCH[7]三种)，并没有出现一家独大；对于绝大多数科研人员，开展分析中的主要难题不是没有软件可有，而是软件太多，不知如何选择，对方法缺少理解导致结果无法把握（关于微生物组研究中常用软件和点评，详见今年9月遗传发表的综述《微生物组数据分析方法与应用》[3]）。总结为易用性差、缺少国内的本地化支持、以及原创性工作太。

**已经发表的一些工作**

因此我们团队在这方面也做了一些工作，主要采用先学习，再跟跑的策略，希望将来能够成为领域者。

1. CAMI宏基因组基准数据集：用于评测宏基因组分析工具的金标准

宏基因组分析复杂，包括很多步骤而且每个步骤也有很多软件可选，而现实中一直缺少成份已知的宏基因组样本来评估目前软件的优缺点。我们团队参与构建了混合1300多种微生物的宏基因组标准样本，获得的可以对宏基因组分析结果评估的参考数据集(Critical Assessment of Metagenome Interpretation, CAMI)[8]，获得了本领域较全面的评估结果，也为本领域接下来的软件开发提供了指导方向，文章发表于Nature Methods杂志，两年获得两百多同行引用。

1. QIIME 2：可重复分析的微生物组工具

微生物组数据分析步骤多，也需要根据不同研究目标定制分析流程。在经验下足前，可以先系统学习主流分析方法，获得较优结果。我们也参与本领域最引用最多的分析流程QIIME系列最新版(QIIME 2)的开发和测试工作[9]，同时也系统掌握了最新的分析思路，而且为华人同行积累了数10万字的中文教程，使大家可以使用母语高效学习使用此软件。该软件于今年8月正式发表于Nature Biotechnology，仅上线3个月收获近百次引用。

1. 绝对定量扩增子方法：

面对扩增子无法绝对定量，结果假阳性较多的问题。我们在植物微生物组领域开发了一种可以确定宿主、细菌、真菌间的相对关系的方法，为扩增了分析提供更深层次的信息[10]。该成果作为创刊文章，发表于我们新创办的植物高水平期刊Plant Communications杂志。

**下一步工作重点**

根据目前研究中的现状和不足，以制定以下3方面研究重点，以进一步推动我国微生物组领域分析技术的全面发展和应用。

1. 提高易用性：分析流程主要的问题是易用化和灵活性是不可调和、相互矛盾的。简单好用的不灵活，太灵活的学习成本高不易上手。主要还是缺少团队系统的优化方法，国内缺少像QIIME团队这样专注一个软件10余年，持续精耕细作。计划开发兼顾灵活和易用的双版本(代码+网站)流程，本对中国人使用特点进行优化，实现较自主、有特色、易用的分析流程，适合各不同行业不同计算水平同行；
2. 微生物组数据库的国内备份站点：目前有微生物所、基因组所的大数据中心存原始数据，清华有R和conda镜像方便软件安装，但宏基因组相关数据库国内缺镜像，也更新较快，下载困难，需要提供国内地址；
3. 原创性：本人编程水平有限，希望和本领域更多计算生物学同行合作，开发能解决实际问题的先进算法，本质上推动本领域分析层面的新突破。

**2. 合作分析**

**实验和计算生物学家的优缺点**

**实验生物学家开展多组学研究的的痛点**

1. 数据分析无从下手：软件选择、安装和运行处处碰壁；
2. 分析中报错和优化无法把握：遇见错误很崩溃，认为分析就是软件运行一次(其实是多次反复分析、优化和筛选结果的探索过程)、无法整合现在工具服务于具体项目；
3. 结果解读：一般看不懂结果，无法选择合适的表达方式，结果与生物学意义很难结合；

**实验生物学家的优势**

1. 丰富的实验材料；
2. 精湛的实验设计；
3. 强大的背景知识。

每个人的时间和精力都是有限的，让实验生物学家精透数据分析，不是没有，但绝对是风毛菱角，成功的也绝对是人中龙凤。目前最好的方式是合作，而且根据统计文章作者数量是与影响因子最相关的因素，即合作更容易产出高水平成果。

**生物信息学家的优势**

1. 组学数据分析的完整思路；
2. 软件安装、错误解决和根据结果和经验反复优化的能力；
3. 结果解读和根据需求定制分析方案的能力；

与生物学家合作，挖掘数据背后的生物学意义，推动高水平成果产出。是目前最高效的方法。因为别人从零学习3个月的成果，并没有成熟经验的人3天优化的结果好和准确，所以分工和合作在科研领域需要进一步加强。

**合作项目案例**

1. 合作研究水稻根系微生物组时间序列分析，结合田间管理、采样和制备 + 数据分析挖掘，探索于水稻全生育期微生物组的动态变化规律，发现根系微生物组快速变化且在生长后期稳定，水稻生殖生长阶段特异富集氮代谢相关细菌，为农业益生菌的施用时间提供指导和理论基础。这一成果作为Science China Life Sciences(中国科学生命科学)2018年第6期封面文章发表[11]。该文章在发表后的一年多时间中，获得20次引用。
2. 合作研究水稻氮利用效率和微生物组的关系，进一步解析根系微生物组在氮利用中的作用。发现籼粳稻根系形成不同的微生物组，且籼稻比粳稻富集更多氮循环相关细菌，与氮肥利用效率有关；为农业益生菌在氮肥高效利用中的应用提供了理论和材料基础；该项成果作为今年6月Nature Biotechnology(自然生物技术)封面文章发表[2]。文章发表仅半年引用已知12次。
3. 通过多方合作研究拟南芥三萜和二半萜调控微生物组的机制，发现植物次生代谢物对微生物组的影响，植物二半萜和三萜类化合物对根系微生物组的调控规律，提供了人工调控微生物组的手段。文章今于年5月10日同天分别发表于Science(科学)[1]和Science China Life Sciences[12]，近半年来分别引用17次和5次。

**下一步工作计划**

合作共享。服务好工作单位植物基因组国家重点实验室，勇于攻关，在农业微生物组领域进一步取得原创性成果。

**3. 科学传播**

主要是经验和成果的宣传和共享，国内主要使用微信(Wechat)和微博(Weibo)等新媒体工具，而国外主要使用Twitter(推特)和Facebook(脸书)，这一种大区别在一定程度上阻碍国内和国际间的交流，但也是对各自文化特色尊重的具体体现。因为推动本领域的资源在本土化和国际化两方面的宣传和发展是值得重视的一个方向，将成为海内外重要成果的传播提供窗口。

**微生物组结合“互联网+”知识传播和教学**

3年前，微生物组领域的在百度几乎检索不到任何中文前沿资料。大量国内同行想要系统学习分析，英文资料上手困难，又缺少可读的中文学习资料。本团队成员创办了每日更新的宏基因组公众号，翻译、评测及注释大量国外最新发表文章的分析软件、方法和流程，方便华人圈同行高效使用中文资料学习微生物组数据分析，开展个性化的数据挖掘。目前经两年多发展，累计发表400多篇原创文章，共200多万字。得到了全球7万多同行的关注和好评，累计阅读量超1000万次。

开展线下微生物组数据分析专题研讨会，带领同行针对性学习微生物组数据分析的实际操作和演示软件参数对结果的影响，解决零基础用户无从下手和个性化报错问题，让同行实现独立数据分析，自由开展数据挖掘和探索，极大的提高生物学规律的挖掘效率。两年来有500同行受益，他们自己优势与微生物组数据分析技术相结合，促进一批交叉学科成果的产出，有的博士顺利毕业，有的顺利职称进级，甚至一些学员直评为教授、优青和青千。

结合互联网+模式，开展微生物组学数据分析公开课教学，如在中国人民大学2019年R语言大会中报告《R语言在宏基因组数据分析和可视化中的应用》，线下60人的教室，线上直播有1826人同步收看。同时形成免费在视视频供同行随时观看学习。同时也尝试过千聊平台直播声音和幻灯片，同时2200多人在线观看学习，并可互动提问和即时解答。实现让大家足不出户也可以免费获得高质量的中科院前沿经验和技术的分享课程，并提供个性化问题交流的平台。受邀参加肠道演讲的录制，分析菌群分析入门指南20分钟视频，上线后累计观看达1.5万人次。

**综述撰写**

尤其是中文综述，为国内同行提供母语的入门学习资料。以为这种土鳖为例，我阅读中文的速度至少是英语的5倍以上，有时甚至10倍。这样可以极大提高我的学习效率，提高成果产出。今年了应邀遗传杂志撰写了《微生物组数据分析方法与应用》[3]，为中国同文快速入门微生物组数据分析提供基础学习材料和指明方向。应邀Current Opinion in Microbiology杂志撰写并发表《Reductionist synthetic community approaches in root microbiome research》[4]一文，为推动植物微生物组领域从描述向功能过滤提供参考。

**4. 现状和对策总结**

|  |  |  |
| --- | --- | --- |
| **中国微生物组研究** | **现状** | **对策** |
| 方法优化和开发 | 软件多  易生性差  国内原创少 | 评估+排名+点评  优化适合非专业人士  与数学、计算机联合开发 |
| 合作分析 | 缺少沟通  急功尽力  心态定位偏差 | 掌握必要的对方知识  实验和分析都是长期科研  换位思考、尊重他人贡献 |
| 科学传播 | 中文资料少  成果被埋没  报告讲座少 | 共享经验和笔记  解读、宣传和后续服务  视频直播+永久访问 |

1. Huang AC, Jiang T, Liu Y-X, Bai Y-C, Reed J, Qu B, et al. A specialized metabolic network selectively modulates Arabidopsis root microbiota. Science 2019; 364 (6440): eaau6389. doi: 10.1126/science.aau6389.

2. Zhang J, Liu Y-X, Zhang N, Hu B, Jin T, Xu H, et al. NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. Nature Biotechnology 2019; 37 (6): 676-684. doi: 10.1038/s41587-019-0104-4.

3. 刘永鑫, 秦媛, 郭晓璇, 白洋. 微生物组数据分析方法与应用. 遗传 2019; 41 (9): 845-826. doi: 10.16288/j.yczz.19-222.

4. Liu Y-X, Qin Y, Bai Y. Reductionist synthetic community approaches in root microbiome research. Current Opinion in Microbiology 2019; 49 97-102. doi: <https://doi.org/10.1016/j.mib.2019.10.010>.

5. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology 2009; 75 (23): 7537-7541. doi: 10.1128/aem.01541-09.

6. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 2010; 7 (5): 335-336. doi: 10.1038/nmeth.f.303.

7. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010; 26 (19): 2460-2461. doi: 10.1093/bioinformatics/btq461.

8. Sczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, et al. Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software. Nature Methods 2017; 14 1063. doi: 10.1038/nmeth.4458.

9. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 2019; 37 (8): 852-857. doi: 10.1038/s41587-019-0209-9.

10. Guo X, Zhang X, Qin Y, Liu Y-X, Zhang J, Zhang N, et al. Host-associated quantitative abundance profiling reveals the microbial load variation of root microbiome. Plant Communications 2019; doi: 10.1016/j.xplc.2019.100003.

11. Zhang J, Zhang N, Liu Y-X, Zhang X, Hu B, Qin Y, et al. Root microbiota shift in rice correlates with resident time in the field and developmental stage. Science China Life Sciences 2018; 61 (6): 613-621. doi: 10.1007/s11427-018-9284-4.

12. Chen Q, Jiang T, Liu Y-X, Liu H, Zhao T, Liu Z, et al. Recently duplicated sesterterpene (C25) gene clusters in Arabidopsis thaliana modulate root microbiota. Science China Life Sciences 2019; 62 (7): 947-958. doi: 10.1007/s11427-019-9521-2.