



马占山学科组研发出两项重要宏基因组学技术

2020-04-29 来源：计算生物与医学生态学学科组



宏基因组技术系基因组技术的扩展：基因组技术通常是一次只测序（定）单一物种的基因组，而宏基因（组）技术则能够同时测序成千上万、甚至上百万微生物物种的基因组。故此，宏基因组学技术成为了肠道菌群、环境微生物、食品微生物，当然包括医学微生物学最重要的核心技术之一。事实上，如果没有宏基因组学技术，对肠道菌群的研究基本是无从下手。

然而，与基因组测序技术不同，宏基因组测序技术的质量管控通常是一笔糊涂账，特别是以测定标志性基因（例如16S-rRNA基因）为基础的“扩增子测序”(Amplicon)至今没有严格的科学标准确定其测序质量。这是因为，基因组测序数据在完成基因组组装后，通常可以与参考基因组相比较，就可以知道测序质量如何。而宏基因组测序技术之一的“全基因-宏基因测序”（whole genome or shotgun metagenome sequencing）所获得的基因数量通常是基因组测序所获得数量的上百倍，而人类对其中绝大多数的基因并没有参考基因组。又例如，宏基因测序另一技术、前面提到的“扩增子测序”使命是发现人类先前可能未知的大量微生物（细菌、病毒、噬菌体等），发现多少取决于测序质量，但往往“无从对质”。当然，可以尽可能加大测序“深度”，但测序“深度”与测序成本成正比。如何将钱花在刀刃上，更重要的是如何保证检测到所关心的微生物种类（例如机会性病原菌）或者所携带的基因则具有及其重要的现实意义。

马占山学科组此次发布的第一款宏基因组学技术则是解决了以上提及的问题之一：即用于“扩增子测序技术”质量控制方法。该技术的实际应用包括两个方面：（1）其一，可以对于某一项已经完成的扩增子测序项目的质量作出评估。例如：该研究发现，美国人肠道菌群计划（American gut microbiome project）项目质量（精度）大约是66%。这一发现也说明了宏基因测序技术的不确定性，或许AGM项目当初的设计目标应该高于这一仅仅是“合格”水平的质量。类似地、作为用户也可以对测序公司按合同所完成的测序项目质量（精度或误差率）作出评估，做到心中有数。（2）其二、测序公司可以根据以往项目积累的数据，建立测序精度、测序深度、以及测序成本之间关系的预估模型，从而在做市场营销时候对不同要求的客户提供不同价位、同时能够满足客户要求的测序服务。当然，预估肯定是有不确定成份，为此可以为用户设定不同的风险水平（类似于对小概率事件的控制）。

学科组此次发布的第二款技术则适用于宏基因组中“全基因-宏基因测序”技术。如前所提及，该技术最大挑战之一是要处理上百万基因信息（相比较人体自身基因只有两万多）。例如：要直接比较两人、或者两组（例如：“健康人群”与“肥胖人群”）宏基因样本实际上非常困难。首先是缺乏有效的“抓手”。马占山团队曾于2018年将生态多样性的概念和计算方法引入宏基因组学（Ma & Li 2018, Molecular Ecology）。但多样性度量有自己局限性；例如：多样性概念并没有考虑基因之间的关系。为此，此次发布的新技术则引入了“异质性”（Heterogeneity）的概念和度量方法。异质性与多样性之间既有联系，但更多的是区别。引用生态学家Aaron Ellison和哲学家Ayelet Shavit说法：“多样性适用于描述动物园”，而“异质性适应于描述生态系统”。简单讲，就是异质性强调系统内各个组成部分之间的差异，更重要的是它们之间是如何相互作用，从而完成整体功能。依此解释：东北原始森林（生态系统）中各种野生动物、林木、乃至土壤微生物之间存在及其复杂的扑食、寄生、共生等营养层关系，从而决定了森林生态系统的整体功能。而在城市动物园，虽然可能圈养着豺狼虎豹、大象熊猫、以及其它各种各样动物，但物种之间基本上“互不相识”。某一天，把一只老虎搬到其它动物园，或者老虎老死了，对动物园其它动物基本没什么影响，而“山中无老虎、猴子称霸王”所描述的景象则应该是对自然生态系统景象的描述。此次提出的“宏基因组异质性”概念和评估方法则是解决了宏基因多样性概念所无法涵盖的特征，从而为比较宏基因组样本提供了一项有效的分析技术。例如，该研究就发现，采用新提出的异质性研究方法，发现肥胖与正常体重、糖尿病人与健康对照之间均存在显著的异质性差异，而这类差别则可能就反映了疾病人群与健康对照之间宏基因功能的变化。当然，这些变化也可能用于疾病诊治方法的研究。另外，马占山研究员目前正在与 Aaron Ellison和

以上两项研究进展近日分别在线发表于 *Frontiers in Bioengineering and Biotechnology*以及*Frontiers in Microbiology*;
基于其研究的两项发明专利已进入实质审核阶段。

Ma ZS (2020) Estimating the optimum coverage and quality of amplicon sequencing with Taylor's power law extensions. *Front. Bioeng. Biotechnol.* | doi: 10.3389/fbioe.2020.00372

Ma ZS (2020) Assessing and Interpreting the Metagenome Heterogeneity With Power Law. *Front. Microbiol.* |11:648. doi: 10.3389/fmicb.2020.00648



Copyright © 2018-2020 中国科学院昆明动物研究所 .All Rights Reserved
地址: 云南省昆明市五华区教场东路32号 邮编: 650223
电子邮件: zhanggq@mail.kiz.ac.cn
滇ICP备05000723号  滇公网安备 53010202000920号

