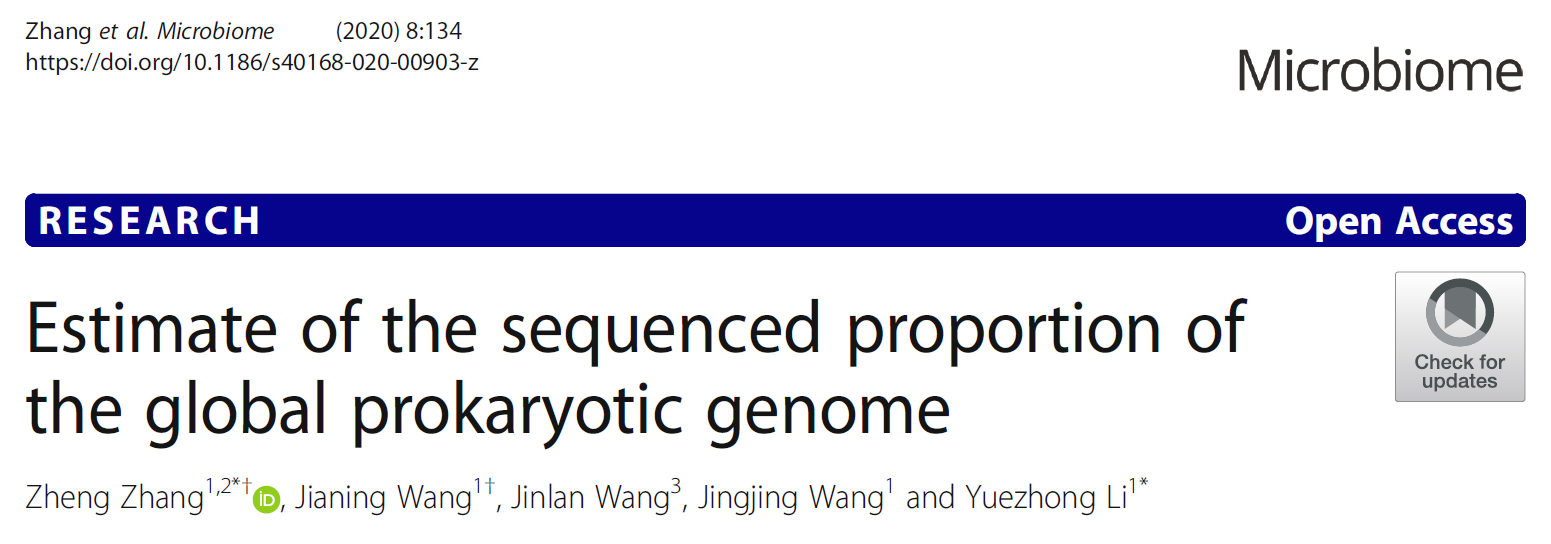
## •Microbiome：地球上有多大比例的原核生物已经被测序了基因组？

**全球原核生物基因组测序现状评估**

**Estimate of the sequenced proportion of the global prokaryotic genome**



* Article，2020-09-16
* Microbiome, [IF 11.607]
* DOI：https://doi.org/10.1186/s40168-020-00903-z
* 原文链接：<https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-020-00903-z>
* 视频摘要：<https://watch.wibbitz.com/preview/?id=b5cf250363a52433baa5e8783fa07cf03&next=none>
* 第一作者：Zheng Zhang（张正），Jianing Wang（王家宁）
* 通讯作者：Yuezhong Li（李越中，lilab@sdu.edu.cn），Zheng Zhang（张正，zhangzheng@sdu.edu.cn）
* 合作作者：Jinlan Wang（王金兰），Jingjing Wang（王晶晶）
* 主要单位：

山东大学微生物技术国家重点实验室（State Key Laboratory of Microbial Technology, Institute of Microbial Technology, Shandong University, Qingdao 266237, China）；

山东大学苏州研究院（Suzhou Research Institute, Shandong University, Suzhou 215123, China）；

## 摘要

**基因组测序提供了描绘原核生物进化和功能多样性的蓝图，深刻改变了我们对原核生物如何与其他原核生物、宿主以及周围环境相互作用的理解。然而，对于地球上有多大比例的细菌和古菌已经被测序了基因组这个基本问题，目前却仍不清楚。在本研究中，通过对地球微生物组计划数据和公共数据库中原核生物基因组信息的大规模比对，我们首次给出了答案。目前在地球上主要的原核生物群落中，已知基因组信息的细胞或类群占比的中位数已经分别达到38.1%（16.4%-86.3%）和18.8%（9.1%-52.6%）。原核生物群落基因组测序比例与其α多样性显著负相关，并且宿主依赖类环境的测序比例显著高于自由生活类环境。由于广布类群已知基因组信息的程度更高且仍存在大量低测序比例的稀有类群，全球原核生物类群整体上的基因组测序比例实际上仅有2.1%。少数高丰度的优势类群占据群落主体，其基因组测序比例也远高于数量庞大的稀有类群。这些结果首次全景展示了地球微生物组中原核生物的基因组测序状况，将有助于在未来更加合理高效的挖掘基因组资源。**

## 引言

原核生物是地球上最古老、最广泛的生命形式，生活在几乎所有的生态系统中，是地球元素循环的主要推动者。基因组测序提供了描绘原核生物进化和功能多样性的蓝图，深刻改变了我们对原核生物如何与其他原核生物、宿主以及周围环境相互作用的理解。那么，**目前地球上有多大比例的细菌和古菌细胞或类群已经测序了基因组？**这个看似简单和基础的问题却从未被解答。

自从1995年第一个完整的细菌基因组发布以来，原核生物的测序速度随着技术发展和成本降低而飞速上升，目前在公共数据库中具有完整或草图序列的细菌和古菌基因组已超过20万个。同时，测序通量和计算技术的进步已经可以利用分箱(binning)从宏基因组数据中获得不依赖培养的基因组(metagenome-assembled genomes, MAGs)，这进一步加速了对原核生物基因组信息的挖掘。而与基因组数据的指数级积累相反，最新估计的全球原核生物多样性仅为80-160万个OTU，远低于之前预测的上万亿个。因此，现在已有必要全面评估全球原核生物的基因组测序状况。

地球微生物组计划**（Earth Microbiome Project，EMP）**以前所未有的规模对地球微生物群落进行采样，用以评估原核生物在全球环境中的分布方式。在本研究中，**通过对EMP公布的首批数据和公共数据库中的原核生物基因组信息的大规模比对，我们第一次全景展示了地球微生物组中原核生物基因组的测序现状**。

## 结果

通过对EMP序列数据和近16万个RefSeq基因组信息的比对，我们评估了已知基因组信息的细胞或类群在特定原核生物群落中的比例（Bcell和BOTU）。结果显示，在16S rRNA序列完全一致的前提下，被研究的1万个样本的Bcell值中位数已达38.1%，上四分位数和下四分位数分别为16.4%和86.3%（图1）。这意味着**在超过一半的原核生物群落中，至少有38%的细胞的基因组信息已经被报告了**。类似的，对于BOTU值，其中位数也达到了18.8%（9.1%-52.6%）。并且，原核生物群落中已知基因组信息的比例与其所处环境密切相关，宿主依赖类环境的基因组测序程度显著高于自由生活类环境。原核生物群落的测序比例还与其α多样性呈现显著负相关性。具有低α多样性的原核生物群落往往具有较高的基因组测序程度，在细胞和类群水平上均是如此。



图1.原核生物群落中已知基因组信息的程度。(A) 群落中已测序细胞的比例。(B) 群落中已测序类群的比例。红色代表样品来自宿主依赖类环境，绿色代表来自自由生活类环境。

基于多个样本Meta分析得到的OTU的基因组测序比例被定义为POTU值。对1万个EMP样本的Meta分析共聚类到262011个OTU，以16S rRNA序列完全一致为标准，我们发现其中已知基因组信息的OTU仅占2.1%。**原核生物类群能够生存的样本或环境类型越多，其已知基因组信息的几率越高**（图2）。由于生活于多个群落的广布类群已知基因组信息的程度更高，而大量窄布/稀有类群的测序比例仍很低，导致了POTU值总是低于相应的BOTU值。



图2.原核生物广布类群已知基因组信息的程度更高，导致POTU值远小于BOTU值。

因为一个OTU有可能出现在多个样本中，所以我们通过随机抽样评估了样本数量对POTU值的影响（图3）。结果显示，随着样本数量增加，POTU值呈现指数衰减趋势，并最终稳定在2.13%±0.03%。因此，基于1万个EMP样本估计的POTU值已经非常接近全球所有原核生物类群中已知基因组信息的比例了。基于全部宿主依赖类环境样本计算的POTU值为4.6%，而全部自由生活类环境样本的POTU值则仅有2.1%。总之，**尽管原核生物基因组信息在快速积累，但目前全球原核生物类群整体上的基因组测序比例仍然很低**。



图3.全球原核生物类群的基因组测序比例仍很低。

丰度最高的前1%的原核生物类群占了全球原核生物群落总丰度的超过70%，并且在各种环境类型中均是如此（图4）。我们发现，丰度最高的前1%的原核生物类群中已经有高达38.0%被报告了基因组序列，而在占比59.8%的低丰度类群（总序列数不超过10条）中则仅有0.6%已知基因组信息。因此，**具有高基因组测序程度的少数优势类群占据了全球原核群落主体**。



图4.原核生物高丰度类群已知基因组信息的程度更高。

我们估计了原核生物在门纲目科属各个分类水平的基因组测序程度（补充数据）。POTU值在不同类群间差异巨大，而同一类群的POTU值在环境间也有明显差别。此外，RefSeq基因组全部来自已培养菌株，而数据库中也已有大量基于宏基因组binning获得的未培养细菌和古菌的基因组。利用这些MAG，我们评估了菌株可培养性对目前基因组测序偏好的影响。结果显示，可培养性影响物种间的基因组测序偏好但不影响环境间差异。

## 讨论

基因组是了解原核生物生理、生态和进化的基本资源。在这个工作中，我们首次全面评估了全球原核生物基因组测序的现状。目前地球上多数原核生物群落的基因组测序程度已经比较高了，在细胞水平这个比例的中位数已经达到38.1%（16.4%-86.3%），在类群水平也有18.8%（9.1%-52.6%）。因此，当我们关注一个具体的原核生物群落时，这个群落的遗传信息可能已经在相当高的程度上被展示了。然而，与群落的情况不同，全球原核生物类群在整体上的基因组测序比例仍然很低。我们估计，目前所有原核生物类群中仅有2.1%已被报告了基因组信息。由于原核生物基因组测序目前仍集中于少数广布类群和高丰度类群，对于大量稀有类群基因组信息的了解尚非常有限，对这个巨大遗传资源的挖掘工作其实才刚刚开始。

原核生物仅有1%可培养的说法对微生物生态学产生了深远影响，但在最近产生争论（https://mp.weixin.qq.com/s/ouwxRrUMPS1i9lIxxXOhFA）。因为RefSeq基因组数据基本全部来自已培养类群，而相当比例的可培养类群尚未测序基因组，所以我们估计全球原核生物类群的可培养率应高于2.1%的基因组测序比例。而与高丰度优势类群具有更高的基因组测序比例相似，优势类群也应具有比稀有类群高的多的可培养率，所以原核生物细胞水平的可培养率也会大大高于类群水平。因此，我们认为**不论对于类群还是细胞，原核生物仅有1%可培养的说法都已经过时了**。

## 作者简介



**张正**，山东大学微生物技术研究院副研究员，微生物技术国家重点实验室“药源微生物基础和应用创新团队”成员。从事微生物学和生物信息学研究多年，作为第一作者或通讯作者在Molecular Biology and Evolution、Microbiome、Nucleic Acids Research等杂志上发表SCI论文15篇。



**王家宁**，山东大学微生物技术研究院博士研究生，微生物技术国家重点实验室“药源微生物基础和应用创新团队”成员。目前主要研究方向是环境微生物组。作为第一作者或共同第一作者在Microbiome、Journal of Hazardous Materials、mSystems、International Biodeterioration and Biodegradation等杂志上发表研究论文5篇。



**李越中**，山东大学微生物技术研究院教授，博士生导师，国家杰出青年基金获得者，微生物技术国家重点实验室“药源微生物基础和应用创新团队”PI。长期致力于重要药源微生物相关的基础和应用基础研究，作为通讯作者在ISME、Microbiome、mSystems等杂志上发表研究论文100余篇，获得国家发明专利10余项、国际专利2项。

作者主页：http://www.mbtechinst.qd.sdu.edu.cn/info/1040/1707.htm

博士后招聘：https://q.eqxiu.com/s/VVq1Mybr