
万种原生生物基因组计划取得阶段性进展

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/25045.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

原生生物（Protist）是一大类单细胞真核生物的集合，包括单细胞真核藻类和原生动物等，组成了原生生物界。原生生物具有高度多样性，广泛分布于各类水环境中，在生态平衡、物质和能量循环、环境健康、动植物疾病等方面至关重要。它们是水生态系统的重要组成部分，是重要的初级生产力和氧气的制造者、碳循环的关键参与者，是水产动物的优良饵料、人类的营养品、生物能源，是水环境的“哨兵”、水华和赤潮的重要元凶，亦是人、畜、禽、鱼疾病的重要病原、互利共生的“好伙伴”。

NCBI分类系统已记录的原生生物种类超过6万种，未知数量难以估量。2019年12月，中国科学院水生生物研究所牵头发起万种原生生物基因组计划（Protist 10,000 Genomes Project, P10K），旨在建立一个大规模的原生生物遗传资源数据库，变革原生生物遗传资源数据极度缺乏的局面。

近日，水生所与中国科学院北京基因组研究所（国家生物信息中心）联合发布了P10K第一批数据。该数据通过万种原生生物基因组数据库（P10K database, <https://ngdc.cncb.ac.cn/p10k/>）共享发布释放。相关成果以The P10K database: a data portal for the protist 10000 genomes project为题，发表在《核酸研究》（Nucleic Acids Research）上

。P10K第一批数据收录了2959个原生生物数据集，包括1601个基因组和1358个转录组数据集，覆盖原生生物75%的纲和45%的目。其中，P10K团队在公共数据库整合了1858个数据集，新测序了1101个数据集，且以原生动物纤毛虫（Ciliate）为主。新测序数据将原生生物数据集规模整体提升了37%。新测序的样品由P10K团队在国内多种生境中采集、分离。对于不能实验室培养的大部分原生动物，该团队采用单细胞测序方法（约占新测序数据的98%）。同时，为了解决大规模单细胞组学数据的分析问题，P10K团队研发了一套针对原生生物单细胞测序数据的组装、去污染、物种鉴定、基因注释和评估的标准化分析流程。质量评估显示，该流程注释的基因组与公共数据库发布的基因组具有相似比例的中高等质量数据。

研究工作得到国家重点研发

计划、中国科学院战略性先导科技专项、中国科学院国际合作计划、

中国科学院青年创新促进会、国家自然科学基金和国际生物科学联盟开放生物多样性和健康大数据计划的资助，并获得国家水生生物种质资源库和中国科学院超级计算武汉分中心的支持。

作为“万种原生生物基因组计划”的重要组成部分，P10K数据库的建立和数据共享，将有助于推动对真核生物和多细胞生物起源、真核生物多样性、原生生物的极端环境适应以及微生物互作等重要基础科学问题的研究。同时，这一计划将促进对与生态环境保护、污染物降解和转化、营

养健康以及疾病防治相关的原生生物遗传资源的挖掘和潜在应用。同时，鉴于原生生物是浮游生物的关键组成部分，P10K数据库将为基于环境DNA的浮游生物鉴定提供支持，助力水生态健康评价。

P10K数据库建立了国家水生生物种质资源库/国家寄生虫资源库（活体种质资源）和国家基因组科学数据中心（遗传资源）之间的联系，对于促进国家科技资源共享服务平台的信息互联互通和数据共享具有重要意义。

[P10K计划](#)

[论文链接](#)

“万种原生生物基因组计划”的标志。设计思路：标志由不同原生生物组成。其中，钟虫（纤毛虫）代表字母“P”，裸藻（鞭毛虫）和团藻（绿藻）的组合代表数字“10”，阿米巴代表字母“K”。双螺旋既代表DNA又代表水波，即原生生物生存的水环境。“P”中喷发出诸多原生生物，代表1676年列文虎克发现钟虫之后，高度多样的各种原生生物被发现。

P10K数据库的特点和功能

研究团队单位：水生生物研究所

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发