

DOI: 10.13376/j.cblls/2023176

文章编号: 1004-0374(2023)12-1608-05



刘勇勤, 兰州大学泛第三极环境中心教授、中国科学院青藏高原研究所研究员、国家杰出青年基金获得者、百千万人才工程国家级人选、享受政府特殊津贴专家。主要从事青藏高原冰川-湖泊微生物与气候环境关系的研究。近年来主持国家自然科学基金委重大研究计划重点支持项目、重点项目等课题 10 余项, 曾参与南极麦克默多干谷 LTER 计划 (MCM IV) 和北极 NASA 星外生命探索计划。以第一/通讯作者在 *Nature Biotechnology* 等期刊发表 SCI 论文 100 余篇, 研究成果揭示了冰川和湖泊微生物群落特征受气候环境影响的现代环境过程与冰芯中微生物的时间演化特征, 拓展了青藏高原冰芯气候环境变化的新指标。获第九届青藏高原青年科技奖、王宽诚西部学者突出贡献奖、中国科学院院长特别奖优秀指导老师奖等荣誉称号。

## 面向微生物与全球变化研究的泛第三极环境数据中心建设

刘勇勤<sup>1,2\*</sup>, 刘军志<sup>1</sup>, 原彬杰<sup>3</sup>, 张国庆<sup>4</sup>, 陈 蕾<sup>3</sup>,  
计慕侃<sup>1</sup>, 刘鹏飞<sup>1</sup>, 邢 鹏<sup>5</sup>, 田少艺<sup>1</sup>, 邓永翠<sup>3,6</sup>

(1 兰州大学泛第三极环境中心, 兰州 730000; 2 中国科学院青藏高原研究所, 青藏高原地球系统与资源环境全国重点实验室, 北京 100101; 3 南京师范大学地理科学学院, 南京 210023; 4 中国科学院上海营养与健康研究所, 上海 200031; 5 中国科学院南京地理与湖泊研究所, 南京 210008; 6 江苏省地理信息资源开发与利用协同创新中心, 南京 210023)

**摘 要:** 环境微生物学是生命科学与地球科学的交叉学科, 近年来发展迅速且得到学术界和社会的广泛关注, 但现有数据共享平台难以同时支持微生物数据和地学空间数据的高效组织和管理。针对该问题, 在第二次青藏高原综合科学考察研究任务五专题三支持下, 兰州大学建设了泛第三极环境数据中心, 实现了 Web 环境下 16S rRNA 基因、宏病毒组和宏基因组等环境微生物数据与相关环境因子数据的集成、管理和查询, 并以地图形式直观地展示样点空间分布, 为用户提供了便捷、易用的数据共享平台。中心未来将继续完善平台的在线数据可视化和分析功能, 以更好地服务于环境微生物和全球变化研究。

**关键词:** 微生物; 全球变化; 泛第三极; 数据中心

**中图分类号:** G250.74; P593; Q93 **文献标志码:** A

### Establishment of the Pan-Third Pole environmental data center for microbiology and global change research

LIU Yong-Qin<sup>1,2\*</sup>, LIU Jun-Zhi<sup>1</sup>, YUAN Bin-Jie<sup>3</sup>, ZHANG Guo-Qing<sup>4</sup>, CHEN Lei<sup>3</sup>, JI Mu-Kan<sup>1</sup>,  
LIU Peng-Fei<sup>1</sup>, XING Peng<sup>5</sup>, TIAN Shao-Yi<sup>1</sup>, DENG Yong-Cui<sup>3,6</sup>

(1 Center for the Pan-Third Pole Environment, Lanzhou University, Lanzhou 730000, China; 2 State Key Laboratory of Tibetan Plateau Earth System, Resources and Environment, Institute of Tibetan Plateau Research, Chinese Academy

收稿日期: 2023-03-13; 修回日期: 2023-08-28

基金项目: 国家自然科学基金重大研究计划(92251304, 91851207); 国家自然科学基金面上项目(42171132); 国家自然科学基金基础科学中心项目(42188102); 兰州大学中央高校基本科研业务费专项资金资助(lzujbky-2021-sp66, lzujbky-2022-ey08, lzujbky-2023-eyt01)

\*通信作者: E-mail: yql@lzu.edu.cn; Tel: 0931-8912926

of Sciences, Beijing 100101, China; 3 School of Geography, Nanjing Normal University, Nanjing 210023, China; 4 Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai 200031, China; 5 Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China; 6 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China)

**Abstract:** Environmental microbiology is an interdisciplinary field that combines life sciences with earth sciences. In recent years, it has rapidly evolved, garnering extensive attention from both academia and society. However, existing data-sharing platforms struggle to efficiently organize and manage both microbial data and geospatial data. To address this problem, Lanzhou University has established the Pan-Third Pole Environmental Data Center under the support of Theme III of Task V of the Second Comprehensive Scientific Expedition to the Tibetan Plateau. This data center achieved the integration, management and querying of environmental microbiological data such as 16S rRNA genes, viromes, metagenomes, and related environmental factor data in a web environment. It visually showcases the spatial distribution of sampling points on maps, providing users with a convenient and user-friendly data-sharing platform. Moving forward, the center aims to further enhance the platform's online data visualization and analytical capabilities to better serve research in environmental microbiology and global change.

**Key words:** microorganisms; global change; Pan-Third Pole; data center

泛第三极地区指以青藏高原为主体的第三极向西扩展, 涵盖青藏高原、帕米尔高原、兴都库什、天山、伊朗高原、高加索、喀尔巴阡等山脉, 面积约 2 000 万平方公里, 人口多达 30 多亿, 与丝绸之路经济带高度重合<sup>[1]</sup>。该区域受气候变化和人类活动影响强烈, 生态环境较为脆弱, 其环境变化的重要性受到全球关注<sup>[2]</sup>。野外观测是研究泛第三极地区环境变化的基础, 对类型众多的野外观测数据进行系统集成和管理非常关键。

环境微生物研究在地学领域相对较新, 但意义重大。对于泛第三极地区而言, 冰川和冻土由于其低温特性, 封存了种类丰富的古老微生物及微生物演化的信息。温度、土地利用方式等历史气候变化及生态系统变迁过程都可以通过微生物的数量、物种组成与群落结构反映出来, 因此微生物的各种特征成为研究全球变化的良好生物代用指标<sup>[3-4]</sup>。近 50 年, 青藏高原增温速率约为全球同期升温速率的 2 倍<sup>[5]</sup>, 冰川和冻土快速消融, 影响了青藏高原及周边区域的生态系统, 例如冰川冻土中的微生物及大量有机碳氮养分在消融时被释放到下游, 影响下游生境中的微生物群落和碳氮循环过程; 冰川冻土中保存的微生物和病毒也会向下游释放, 对下游生态系统和人类生命健康造成影响等<sup>[6]</sup>。因此, 亟须加强对泛第三极区域环境微生物的研究。但是, 如何高效地管理环境微生物相关数据, 构建面向微生物与全球变化研究的数据平台, 是亟须解决的问题。

微生物数据包括 16S rRNA 基因、宏病毒组、宏基因组和宏转录组等多种形式, 要构建的数据平

台首先应全面、高效地支持这些类型的数据。同时, 进行地学研究时常需要将微生物数据与气候、地形、土壤和水体的理化性质等环境因子进行关联分析, 因此平台还需要支持微生物相关环境数据的高效组织和管理。地学领域现有的数据中心多侧重通用地学数据的集成和共享, 对微生物数据缺乏有效支持, 而生物领域的数据中心则对地学领域的空间数据支持不足, 难以满足环境微生物与全球变化研究的需要。

鉴于此, 本文在梳理和总结已有泛第三极数据中心特点的基础上, 设计和实现了面向微生物与全球变化研究的泛第三极环境数据中心。该中心将生物和地理大数据无缝整合起来, 实现了一个开放、合作、共赢的微生物及环境因子数据库共享平台, 从而为微生物与全球变化研究领域的科研工作者提供方便、快捷的数据服务。

## 1 泛第三极相关主要数据中心简介

目前, 与泛第三极地区科学研究相关的主要数据中心包括国家青藏高原科学数据中心、国家冰川冻土沙漠科学数据中心和国家地球系统科学数据中心, 其中前两个数据中心具有较为明显的区域特色, 而国家地球系统科学数据中心虽然不是专门面向青藏高原和泛第三极地区, 但数据涵盖范围较广, 也包括了与泛第三极研究相关的大量数据。

### 1.1 国家青藏高原科学数据中心

国家青藏高原科学数据中心(网址: <https://www.tpdc.ac.cn/home>)以中国科学院青藏高原研究

所为依托单位,是首批成立的20个国家科学数据中心之一,是针对青藏高原及周边地区科学数据门类最全、最权威的数据中心<sup>[7]</sup>。截至2023年2月16日,国家青藏高原科学数据中心拥有5643个数据集(其中开放数据集有3690个),数据文件量达到283TB,用户数达到6万余人。用户可以通过学科关键词、主题关键词、地点关键词进行检索,也可以通过主题关键词构成的数据知识图谱进行检索,并集成了一定的在线分析和模型模拟功能。

### 1.2 国家冰川冻土沙漠科学数据中心

国家冰川冻土沙漠科学数据中心(网址:<http://www.ncdc.ac.cn/portal/>)以中国科学院寒区旱区环境与工程研究所为依托单位,联合中国科学院新疆生态与地理研究所、中国科学院水利部成都山地灾害与环境研究所、中国科学院青海盐湖研究所、中国科学院青海高原生物研究所等单位共同组建,已形成七大研究室、三大研究系统的科研及支撑体系,突出了寒区旱区冰川、冻土、沙漠、大气、水土、生态、环境、资源、工程与可持续发展研究特色。目前,该数据中心数据总量超过50TB,站点数到达51个。用户可按主题、学科、时间、地点、站点、共享方式、贡献者、贡献单位和项目支持等进行检索,在冰川、冻土、沙漠数据的集成和共享方面具有优势。

### 1.3 国家地球系统科学数据中心

国家地球系统科学数据中心(网址:<http://www.geodata.cn/>)围绕地球系统科学与全球变化领域科技创新、国家重大需求与区域可持续发展,依托中国科学院地理科学与资源研究所共享共建,实现了我国地球系统科学数据共享从无到有,由国内走向国际的重大跨越。截至2020年,已开放共享数据集3.5万余个,数据资源量超过2.14PB。该中心包括多个数据分中心,在各分中心可按主题、学科、空间位置、数据产生方式、数据类型、空间分辨率、卫星/传感器类型、时间分辨率、空间尺度、比例尺和时间范围等进行检索,在国内外具有较高的影响力。

上述数据中心均主要面向地学工作者,采用的元数据规范在微生物方面考虑不足,对环境微生物相关数据格式和分析方法的支持较为薄弱,难以有效组织和集成环境微生物及相关环境因子数据。因此,亟须面向第三极地区微生物和全球变化研究的需求,构建具有领域特色、方便易用的数据中心。

## 2 面向微生物与全球变化研究的泛第三极环境数据中心建设

### 2.1 概述

在第二次青藏高原综合科学考察研究任务五“生物多样性保护与可持续利用”专题三“高原微生物多样性保护和可持续利用”(2019QZKK0503)支持下,兰州大学泛第三极环境中心建设了泛第三极环境数据中心(Pan-third Pole Environment Data Center, PTPE-DC, <http://tp.lzu.edu.cn/node>),实现了Web环境下16S rRNA基因、宏病毒组和宏基因组等环境微生物数据与相关环境因子数据的集成、管理和查询,并以地图形式直观地展示样点空间分布,为用户提供了便捷、易用的数据共享平台。

泛第三极环境数据中心依托兰州大学泛第三极环境中心建设,共建单位包括:中国科学院青藏高原研究所、上海营养与健康研究所生物医学大数据中心、南京地理与湖泊研究所、中国科学院城市环境研究所和厦门大学。中心与NODE(National Omics Data Encyclopedia database, <https://www.biosino.org/node/>)生物大数据共享平台达成合作协议,将作为分节点接入NODE平台,从而为更多用户提供方便、快捷的数据服务。

泛第三极环境数据中心以构建开放、合作、共赢的泛第三极地区微生物和环境因子数据共享平台为目标,经过2年的建设,在软件平台建设、数据汇交、科学数据共享与服务等方面取得了显著进展,有力支撑了第二次青藏高原综合科学考察研究等国家重大科研项目。

### 2.2 软件平台建设

泛第三极环境数据中心在中国科学院上海营养与健康研究所生物医学大数据中心开发的NODE平台基础上进行定制化开发,基于PostGIS、Leaflet等开源软件,开发了空间数据存储、查询与地图展示等功能,形成了集微生物与地理空间数据于一体的综合数据共享平台(图1)。

在微生物数据方面,平台支持宏基因组、宏转录组、甲基化、单细胞、微生物组、病毒组等测序数据,蛋白质组、代谢组等质谱数据,以及影像、流式细胞等其他数据。微生物相关的元数据包含6个模块:Project、Sample、Experiment、Run、Data和Analysis。其中,Project是对一个研究项目的总体描述,可以关联多个Experiment;Experiment用于描述样本测序实验,一个Experiment只对应一个



图1 泛第三极环境数据中心网站门户

Project ; Sample 对数据来源的样本信息进行描述, 包括理化等环境数据, 用户可以在线直接填写, 也可以下载表格填写后上传; Run 是高通量测序的上机流程信息; Analysis 是对原始数据进行分析得到的结果信息。

在地理空间数据方面, 平台支持 Shapefile、KML、GeoTiff 等常用矢量和栅格数据的上传、存储、查询和可视化。同时, 为方便不同来源数据的集成与整合分析, 对生物地球化学循环相关的常用地理环境因子制定了统一的分类和命名标准, 用户上传数据时, 平台根据用户指定的映射关系为上传数据赋予语义信息, 为多源数据的无缝整合奠定了基础。

与通用数据平台相比, 泛第三极环境数据中心在生物地球化学循环领域的数据集成和共享方面更为便捷和高效。

### 2.3 已有数据情况

目前平台已集成青藏高原 5 个综合区冰川、湖泊、湿地、河流、土壤等多圈层微生物数据和环境因子数据。微生物数据包括青藏高原 5 061 个样品的 16S rRNA 基因数据 (图 2), 涵盖冰雪、冰尘、湖水和沉积物等生境类型, 数据量 258 GB; 1 763 个样品的三极宏基因组数据, 数据总量达 20 TB; 3 830 个样点的高原大型真菌、地衣种质资源和高原典型土著哺乳动物等肠道微生物的数据 (7 842 个

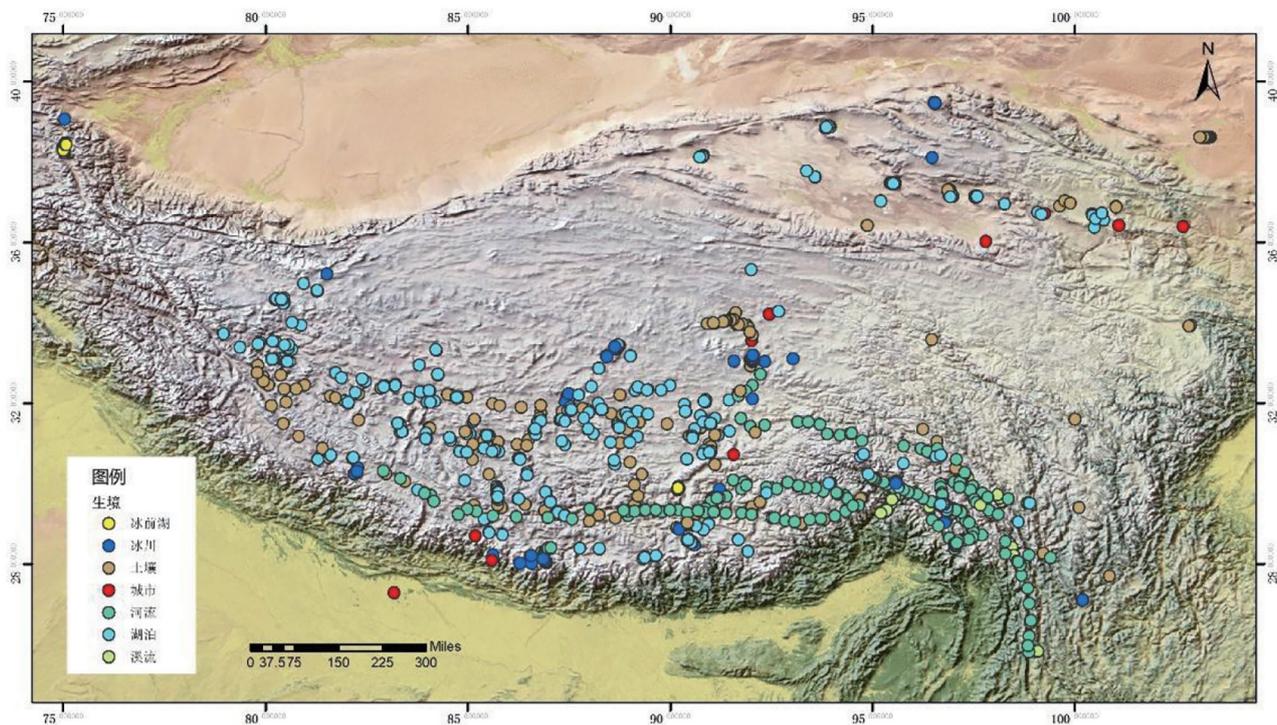


图2 平台已集成的青藏高原16S rRNA基因扩增子数据空间分布图(5 061个样品)

测序数据, 数据量超过 10 TB), 弥补了第一次青藏高原综合科学考察研究中微生物数据的空白。环境因子数据包括青藏高原冰川碳氮数据集、青藏高原湖泊流域环境因子数据集等, 数据发布后已被德国波茨坦大学、日本京都大学等知名学术机构的同行广泛使用。

#### 2.4 未来展望

泛第三极环境数据中心在环境微生物及环境因子的集成和管理方面已初步显示出易用性和便捷性, 但目前还停留在数据的查询和下载方面, 对数据在线可视化和在线分析支持不足。未来将基于 Web 开发领域的最新技术, 在微生物数据和地理空间数据的在线展示方面开展更加深入的工作; 同时, 增加空间查询和分析等地理信息系统的功能, 以及宏基因组、宏转录组等微生物组学分析功能, 让用户不必把数据下载到本地即可方便、快捷地实现定制化的分析。

### 3 小结

环境微生物和全球变化是当前学术界的研究前沿。在第二次青藏高原综合科学考察研究任务五专

题三支持下, 兰州大学建设了泛第三极环境数据中心, 实现了网络环境下环境微生物数据与相关环境因子数据的高效集成、管理和查询, 为用户提供了便捷、易用的数据共享平台。中心未来将继续完善平台的在线数据可视化和分析功能, 以更好地服务于环境微生物和全球变化研究。

#### [参 考 文 献]

- [1] 姚檀栋, 陈发虎, 崔鹏, 等. 从青藏高原到第三极和泛第三极. 中国科学院院刊, 2017, 32: 924-31
- [2] 孙鸿烈, 郑度, 姚檀栋, 等. 青藏高原国家生态安全屏障保护与建设. 地理学报, 2012, 67: 3-12
- [3] 褚海燕. 高寒生态系统微生物群落研究进展. 微生物学通报, 2013, 40: 123-36
- [4] 姚檀栋, 秦大河, 王宁练, 等. 冰芯气候环境记录研究: 从科学到政策. 中国科学院院刊, 2020, 35: 466-74
- [5] Mu CC, Abbott BW, Norris AJ, et al. The status and stability of permafrost carbon on the Tibetan Plateau. *Earth-Sci Rev*, 2020, 211: 103433
- [6] 刘勇勤, 计慕侃, 刘军志, 等. 青藏高原冰川冻土微生物如何记录和影响气候变化? 地球科学, 2022, 47: 3825-6
- [7] 潘小多, 李新, 冉有华, 等. 开放科学背景下的科学数据开放共享: 国家青藏高原科学数据中心的实践. 大数据, 2022, 8: 113-20