

# 微生物基因组数据挖掘和利用

吴林寰 wulh@im.ac.cn

- 中国科学院微生物研究所
- 国家微生物科学数据中心 (NMDC)
- 世界微生物数据中心 (WDCM)

# 汇报提纲

CONTENTS

01

微生物组学数据分析及流程

02

国家微生物科学数据中心的数据资源及平台

# 微生物资源的数字化转型成为学科发展的趋势

微生物资源是生命科学研究不可替代的基本材料，**对科技创新和产业发展**有重要的价值

## 传统微生物资源数据整合

- ◆ 以资源数据、样本数据为主
- ◆ 数据量小、复杂程度低，对整合度要求低
- ◆ 对数据的应用分析需求少



## 数据资源整合及应用的新需求

- ◆ **更加复杂的数据系统**：性状功能数据、组学数据、酶数据、代谢网络数据等
- ◆ 数据复杂度高、数据量大、对整合的要求高
- ◆ 数据分析和处理的要求

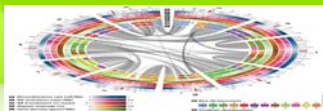
**综合性的微生物大数据**对微生物学研究和微生物资源利用有着至关重要的作用

传统分类学

分子生物学



基因组学



系统生物学



合成生物学



# 全球科技创新趋势

- 在实验观测、理论推演、计算仿真之后、形成数据驱动的科学范式（第四范式）
- 数据成为关键生产要素和战略性资源，科技创新和生产对数据的依赖程度将越来越高。

## 2018年：GTDB 分类系统



A complete domain-to-species taxonomy for Bacteria and Archaea

RESOURCE



A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life

Donovan H Parks, Maria Churochina, David W Waite, Christian Rinke, Adam Skarshewski, Pierre-Alain Chaumel & Philip Hugenholtz

2022年：  
基于序列数据描述的原核生物命名规则

nature microbiology

Explore content About the journal Publish with us

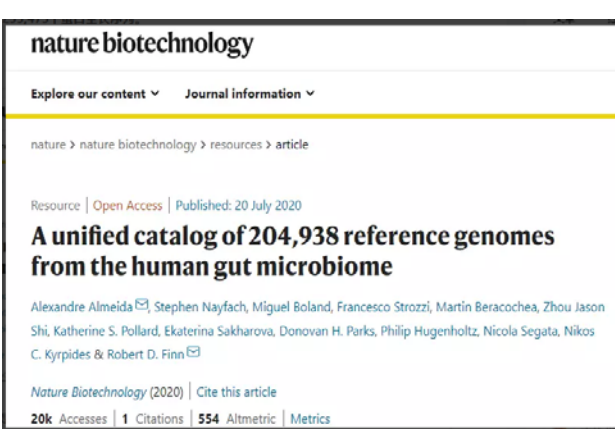
nature > nature microbiology > resources > article

Resource | Open Access | Published: 19 September 2022

**SeqCode: a nomenclatural code for prokaryotes described from sequence data**

Brian P. Hedlund, Maria Churochina, Philip Hugenholtz, Konstantinos T. Konstantinidis, Alison E. Murray, Marike Palmer, Donovan H. Parks, Alexander J. Probst, Anna-Louise Reysenbach, Luis M. Rodriguez-R, Ramon Rossello-Mora, Iain C. Sutcliffe, Stephanus N. Venter & William B. Whitman

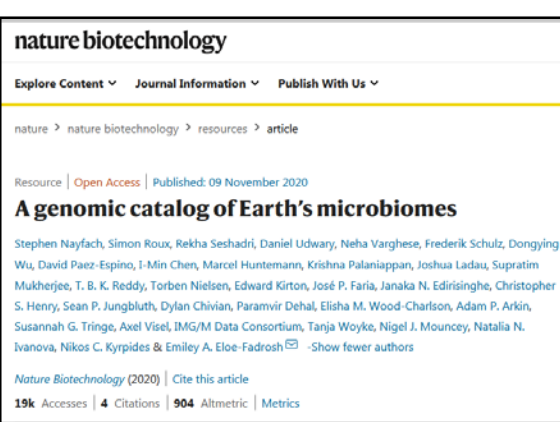
## JGI: 迄今为止最全面的人类肠道微生物组序列资源



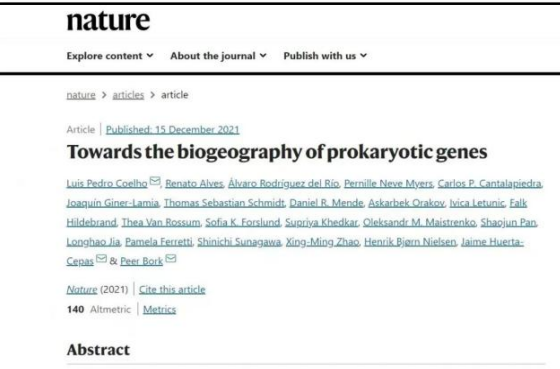
## EBI: 全球表层土壤中微生物组的结构和功能



## JGI & Argonne National Laboratory: 地球微生物组目录



## EBI: 全球微生物基因目录



## 微生物组数据如何改变人类健康?

nature portfolio

ADVERTISEMENT FEATURE Advertiser retains sole responsibility for the content of this article

**How microbiome multi-omics can bolster human health**

Sequencing technologies are enabling a deeper analysis of the gut's microbiome. Researchers can now explore what our microbial inhabitants are doing and how they contribute to, or protect from, disease.

Produced by

nature research | illumina®  
custom media

## 对于培养未被培养的大多数微生物的创新

nature reviews microbiology

Explore content Journal information Publish with us

nature > nature reviews microbiology > review articles > article

Review Article | Published: 22 October 2020

**Innovations to culturing the uncultured microbial majority**

William H. Lewis, Guillaume Tahon, Patricia Geesink, Diana Z. Sousa & Thijs J. G. Ettema

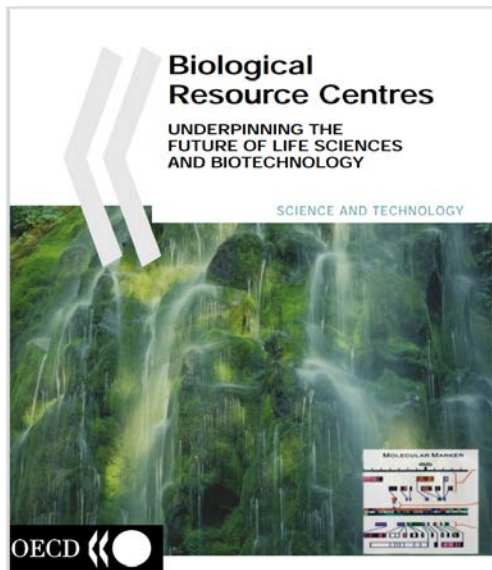
Nature Reviews Microbiology 19, 225–240 (2021) | Cite this article

10k Accesses | 17 Citations | 266 Altmetric | Metrics

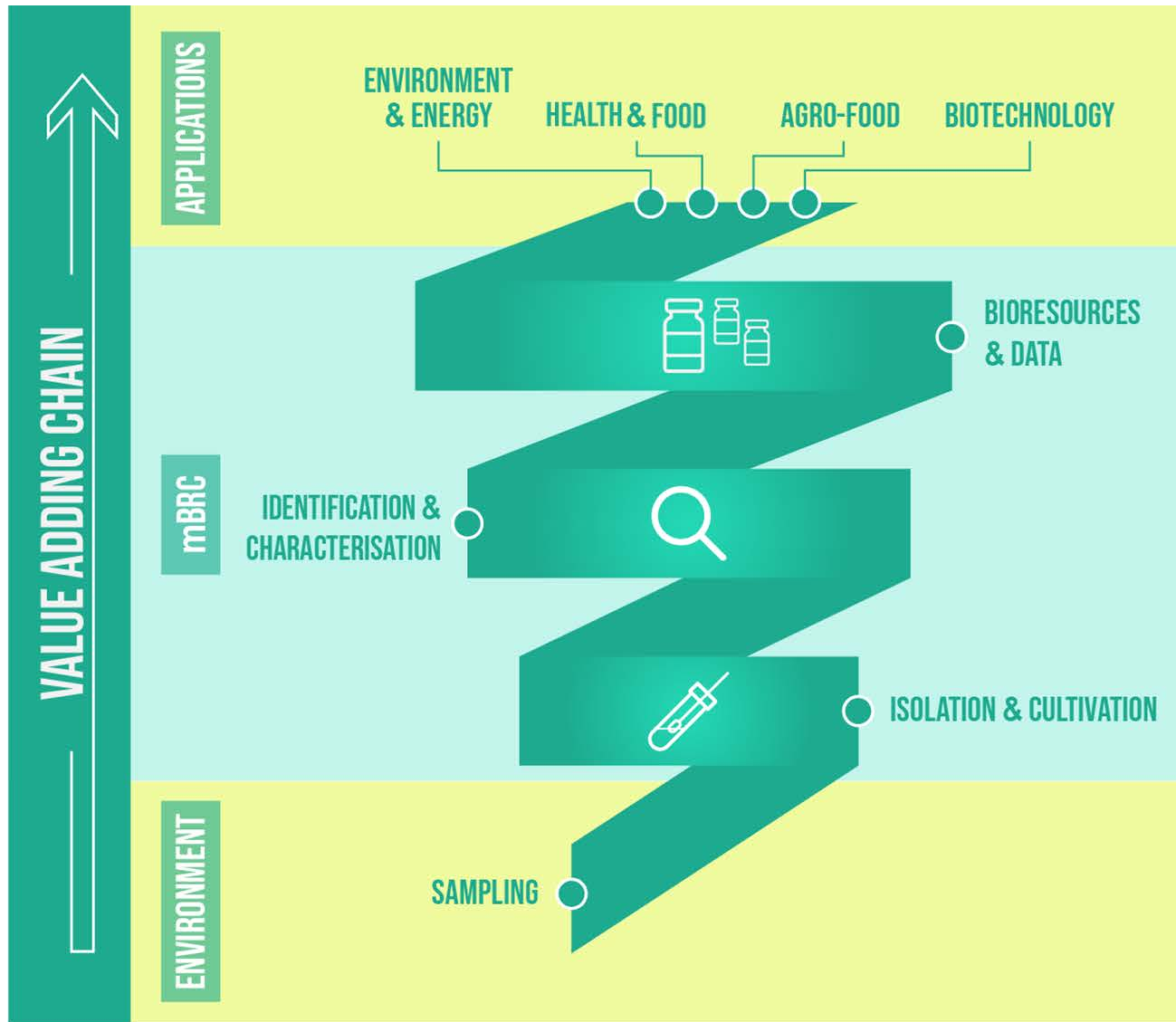
# 生命科学发展态势

- 生命科学研究的纵深发展，需要系统和深入地认识与解析生命
- 生命体和生物过程的数字化、模拟、推演成为一种重要的手段

# 保藏中心现代化发展及资源数字化



**Biological Resource Centres (BRCs):**  
 OECD in 2001 introduced a new concept of repositories and providers of high-quality biological materials and information



## CUTTING EDGE TECHNOLOGIES

- BIORESOURCE LONG-TERM PRESERVATION**
  - cryopreservation (robotically conducted)
  - freeze-drying
  - DNA encapsulation
- DATA**
  - data management
  - databases & network (FAIR)
  - statistical analysis & machine learning
- PCR TECHNOLOGY**
  - fingerprinting
  - gene expression/SNP analysis (qPCR)
- SEQUENCING**
  - long (Nanopore) reads
  - short (Illumina) reads
  - whole genomes
  - microbiomes
- MALDI-TOF MS**
- LC/uPLC MS/MS**
- HIGH-THROUGHPUT SCREENING**
  - multi-omics
  - enzyme/metabolite activities (robotically conducted)
  - multidrug-resistant & multi-toxins tests
- REAL TIME ISOLATION/DISSECTION**
  - micromanipulator
  - optical tweezers
  - fluorescent activated cell sorting
- CULTIVATION/CULTUROMICS**
  - high-throughput cultivation
  - dark matter/rebirth of culture techniques
- BSL-3 FACILITIES**
  - highly infectious strains
  - biohazard agents (toxins)
  - immunization assays

# 标准是资源数字化的基础

ISO 20387:2018 -- General requirements for biobanking

ISO/TR 22758:2020 -- Implementation guide for ISO 20387

ISO 21899:2020 -- General requirements for the validation and verification of processing methods for biological material in biobanks

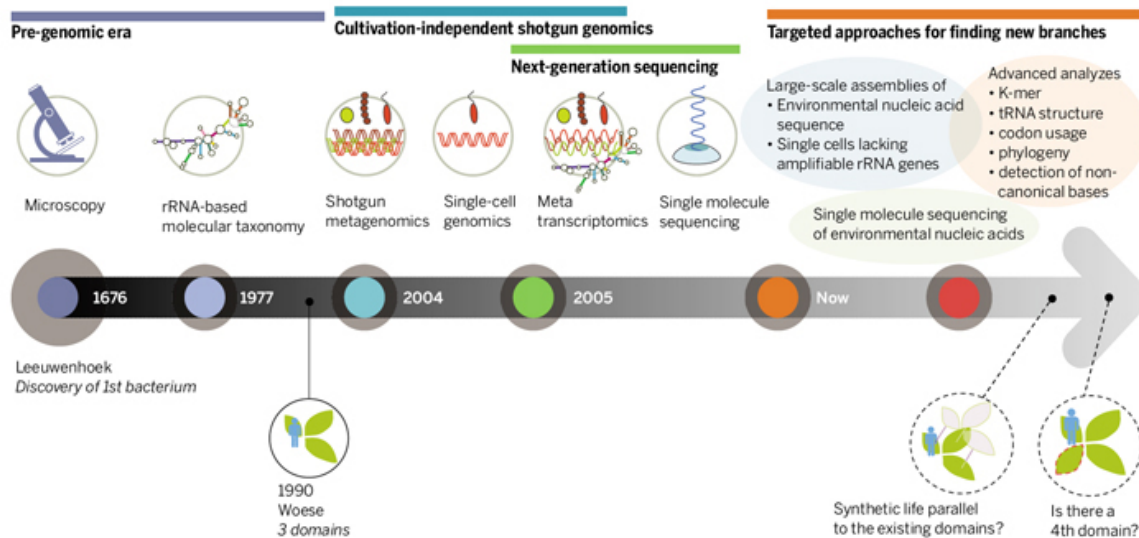
Human bio-material Biobank	Animal biobank	Plant biobank	Microbial biobank
<u>ISO 21709:2020/Amd 1:2021</u> <u>ISO/ TS 22859:2022</u> <u>ISO 24603:2022</u> <u>ISO 24651:2022</u> <u>ISO/DIS 18162</u> Requirements for human neural stem cells derived from pluripotent stem cells <u>ISO/DIS 18209-1</u> Biobanking of parasites — Part 1: Helminths <u>ISO/WD 20309</u> Requirements for deep-sea biological materials	<u>ISO/TS 20388:2021</u>	<u>ISO/TS 23105:2021</u>	<u>ISO 21710:2020</u> Specification on data management and publication in microbial resource centers <u>ISO 24088-1:2022</u> The collection, processing, storage and transportation criteria for microbial material — Part 1: Bacteria and archaea

# 微生物基因组数据将重构生命之树

基因组测序技术的发展，使得我们能够在表型、功能之外，从数据的角度，来认识和研究微生物

## 完善和重构生命系统发育树，建立基于组学数据的微生物分类“金标准”

随着微生物分类学进入基因组学时代，传统分类方法弊端凸显，建立利用基因组数据进行物种分类的“金标准”已迫在眉睫。



The search for new major branches on the tree of life. Cultivation-independent methods, novel sequencing technologies, and analytical approaches can be directed toward the detection of life outside currently established domains. [Woyke et al. 2014, Science/AAAS]

## INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

Official publication of the ICSP and the BAM Division of the IUMS

[About](#)

[Editorial Board](#)

[Latest Articles](#)

[Current Issue](#)

[Archive](#)

News : Genome sequencing data required with Taxonomic Descriptions.

08 December 2017

*International Journal of Systematic and Evolutionary Microbiology (IJSEM)* will soon be asking authors to provide genome sequencing data with descriptions of novel taxa in Taxonomic Descriptions.

Although not mandatory for publication, the inclusion of this data is highly recommended and will be expected to be included. If authors are unable to provide genome sequencing data for any reason this should be stated as such in their covering letter; exemptions will be considered on a case by case basis by the handling Editor.

Genome sequences are of great value to the systematics of prokaryotes. In addition to improving the general understanding of the biology of microorganisms, they improve the identification of prokaryotic species, identification of functional characteristics useful for resolving taxonomic groups and the resolution of the phylogeny of higher taxa.

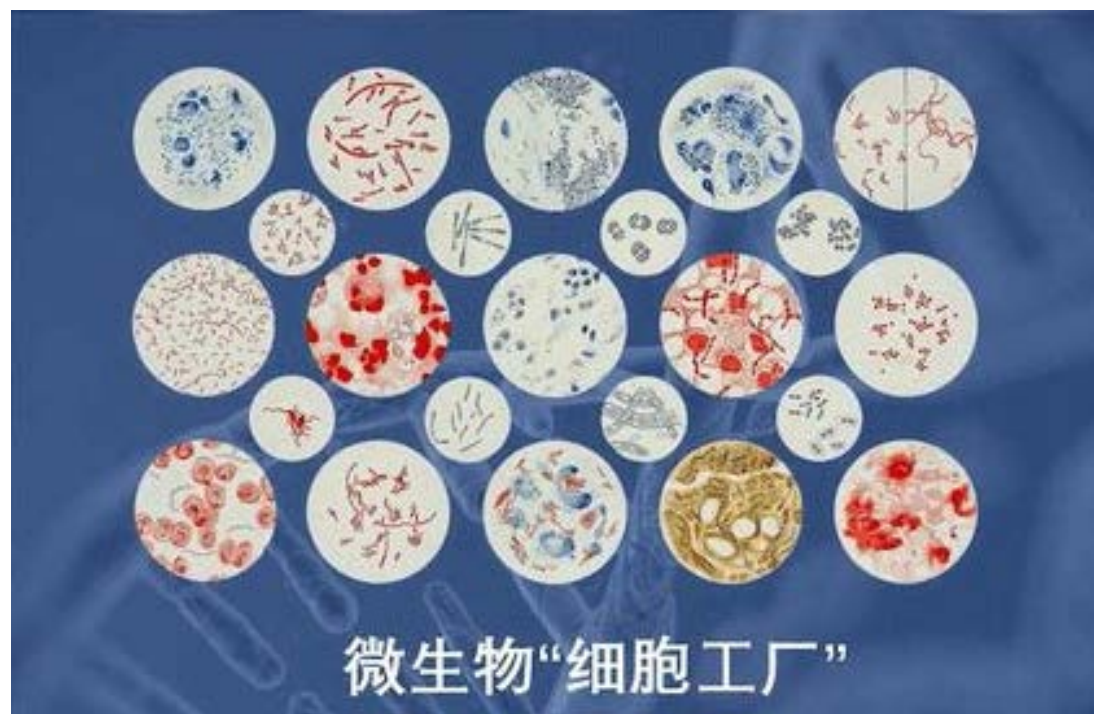
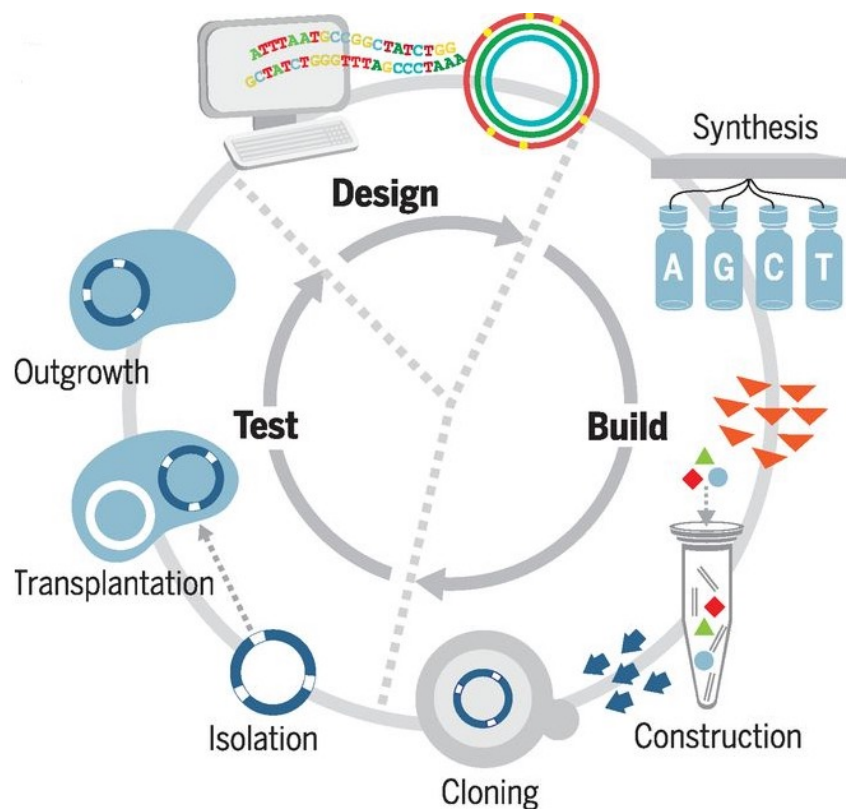
Science 2014: Searching for new branches on the tree of life

传统的微生物分类进入基因组学时代

# 微生物基因组学数据将有助认识并设计生命

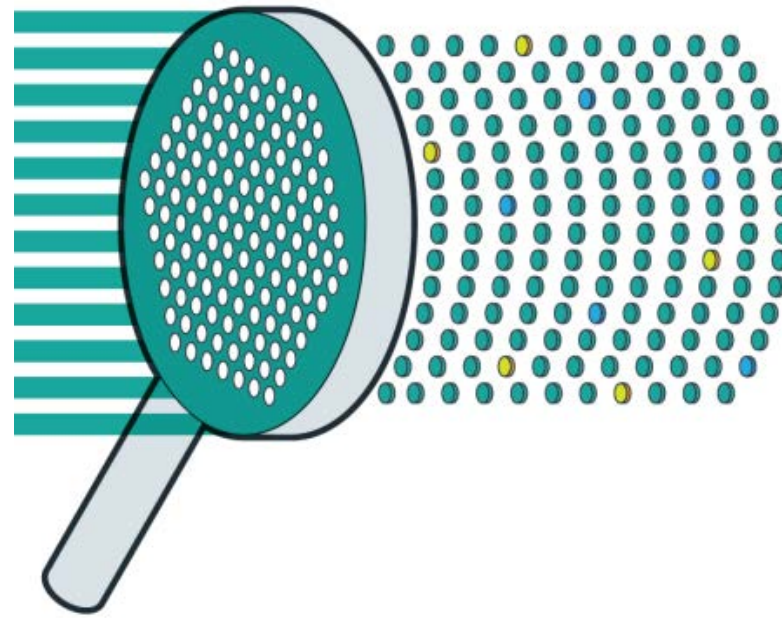
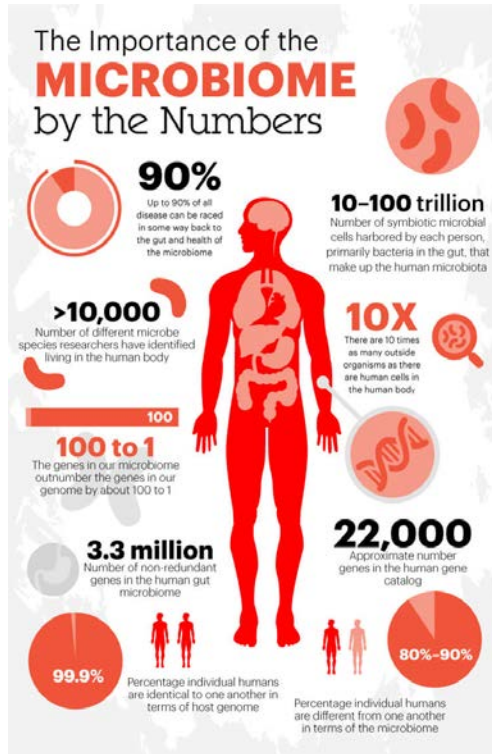
**为重要功能模块、代谢途径的挖掘提供必要基础数据，孵化产业链。**

微生物具有易改造、易合成、易发酵、富含天然产物等多种优势，基因组序列信息将为挖掘新合成生物元件、改造新功能菌株、诱导新天然产物的提供重要的线索

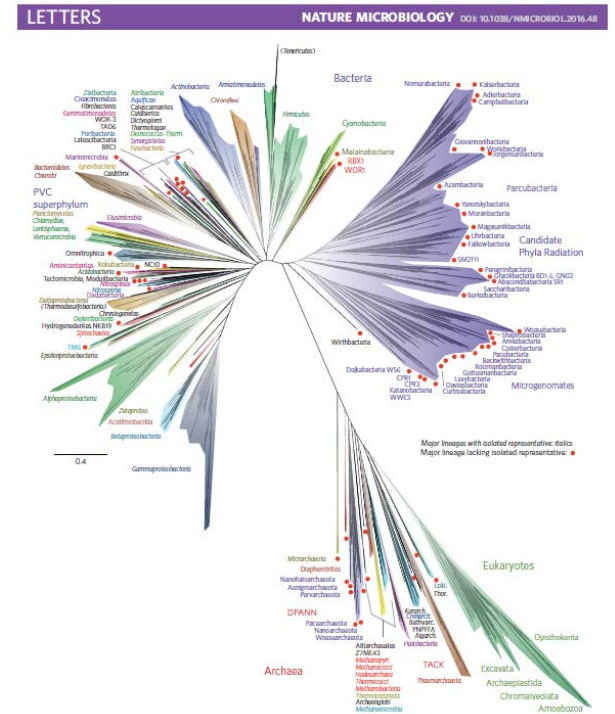




# 高质量的参考基因组是微生物组功能解析的关键

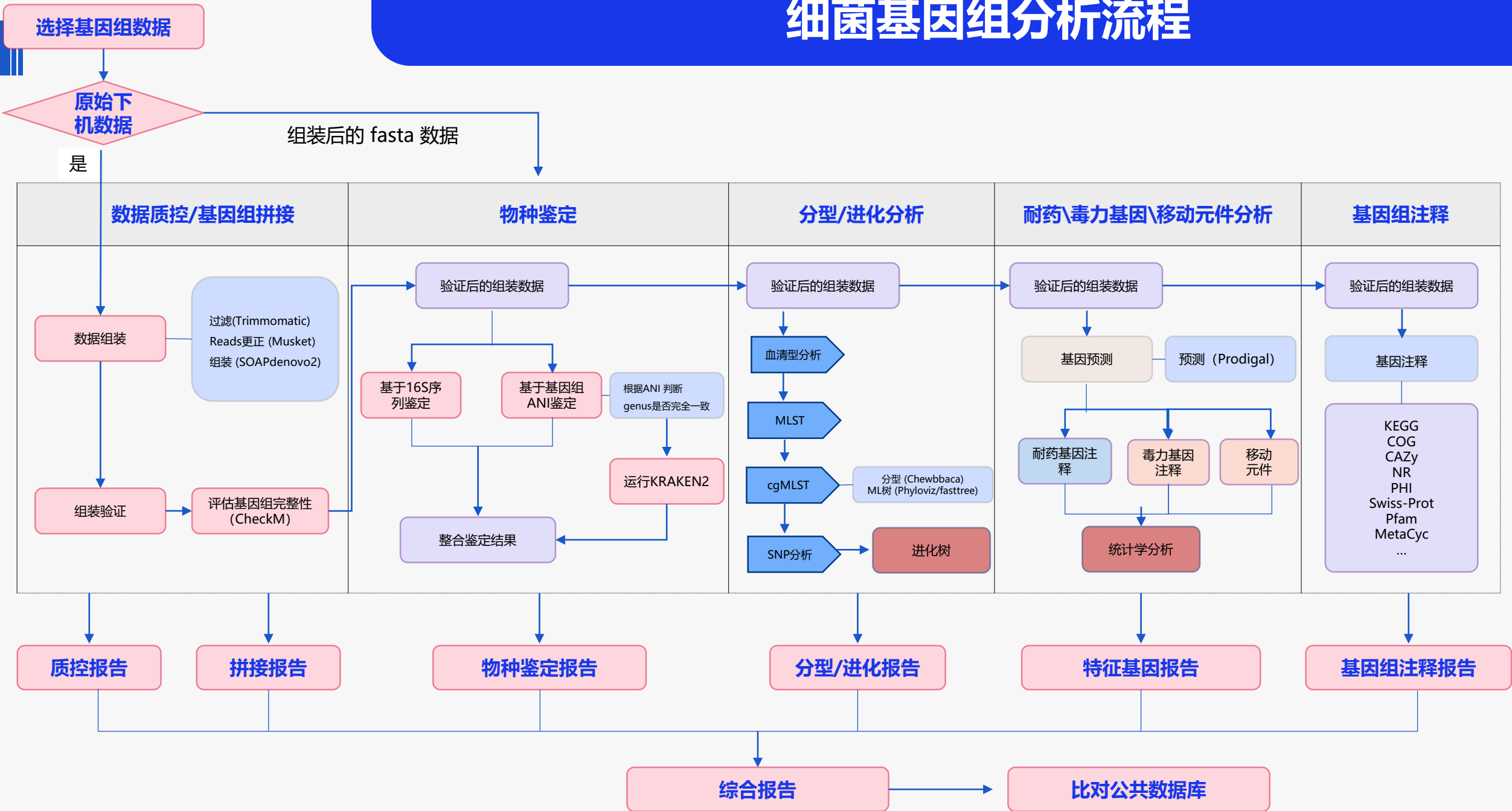


99%的微生物不可培养，数字表征将为必须....



**Figure 1 | A current view of the tree of life, encompassing the total diversity represented by sequenced genomes.** The tree includes 92 named bacterial phyla, 26 archaeal phyla and all five of the Eukaryotic supergroups. Major lineages are assigned arbitrary colours and named, with well-characterized lineage names, in *italics*. Lineages lacking an isolated representative are highlighted with non-italicized names and red dots. For details on taxon sampling and tree inference, see Methods. The names Thermoterricola and Thermodesulfobacteria are bracketed to indicate that these lineages branch within the Firmicutes and the Dehalospirillum, respectively. Eukaryotic supergroups are noted, but not otherwise delineated due to the low resolution of these lineages. The CPR phyla are assigned a single colour as they are composed entirely of organisms without isolated representatives, and are still in the process of definition at lower taxonomic levels. The complete ribosomal protein tree is available in rectangular format with full bootstrap values as Supplementary Fig. 1 and in Newick format in Supplementary Dataset 2.

# 细菌基因组分析流程



## 微生物组资源挖掘： 全面解析微生物组蛋白多样性和功能暗物质

nature

Explore content ▾ About the journal ▾ Publish with us ▾

nature > articles > article

Article | [Open access](#) | [Published: 11 October 2023](#)

### Unraveling the functional dark matter through global metagenomics

## 微生物组“生物银行”的建设面临巨大挑战

Trends in Microbiology

CellPress

Volume 29, Issue 2, February 2021, Pages 89-92

Scientific Life

### Development of Microbiome Biobanks – Challenges and Opportunities

## 从微生物组中挖掘新的抗生素

Microbiome

### Mining the microbiota for antibiotics

[Christine Beemelmans](#), [Andreas Keller](#) & [Rolf Müller](#)

[Nature Microbiology](#) **9**, 13–14 (2024) | [Cite this article](#)

## 微生物组中挖掘新肠菌酶

nature > nature microbiology > articles > article

Article | [Open access](#) | [Published: 03 January 2024](#)

### BilR is a gut microbial enzyme that reduces bilirubin to urobilinogen

[Brantley Hall](#) , [Sophia Levy](#), [Keith Dufault-Thompson](#), [Gabriela Arp](#), [Aoshu Zhong](#), [Glory Minabou Ndjite](#), [Ashley Weiss](#), [Domenick Braccia](#), [Conor Jenkins](#), [Maggie R. Grant](#), [Stephanie Abeysinghe](#), [Yiyan Yang](#), [Madison D. Jermain](#), [Chih Hao Wu](#), [Bing Ma](#) & [Xiaofang Jiang](#)

## 儿童免疫系统与菌群的互作

Science

Current Issue First release papers Archive About ▾ Submit

HOME > SCIENCE > VOL. 376, NO. 6596 > IMMUNE-MICROBE INTERACTIONS EARLY IN LIFE: A DETERMINANT OF HEALTH AND DISEASE LONG TERM

| SPECIAL ISSUE REVIEW

### Immune-microbe interactions early in life: A determinant of health and disease long term

## 微环境、系统和环境角度认识癌症

Cancer Cell

CellPress

Volume 39, Issue 10, 11 October 2021, Pages 1317-1341

Review

### Microbiome and cancer

[Nyssa Cullin](#)<sup>1,4</sup>, [Camila Azevedo Antunes](#)<sup>1,4</sup>, [Ravid Strausman](#)<sup>2</sup>, [Christoph K. Stein-Thoeringer](#)<sup>1</sup> , [Eran Elinav](#)<sup>1,3</sup>

## 微生物组与耐药传播

Microbiome

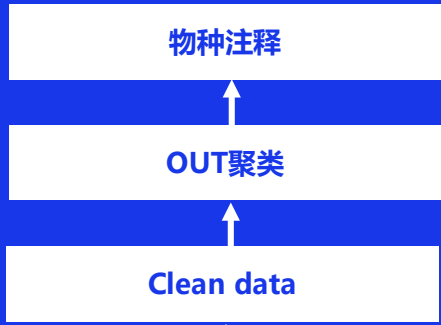
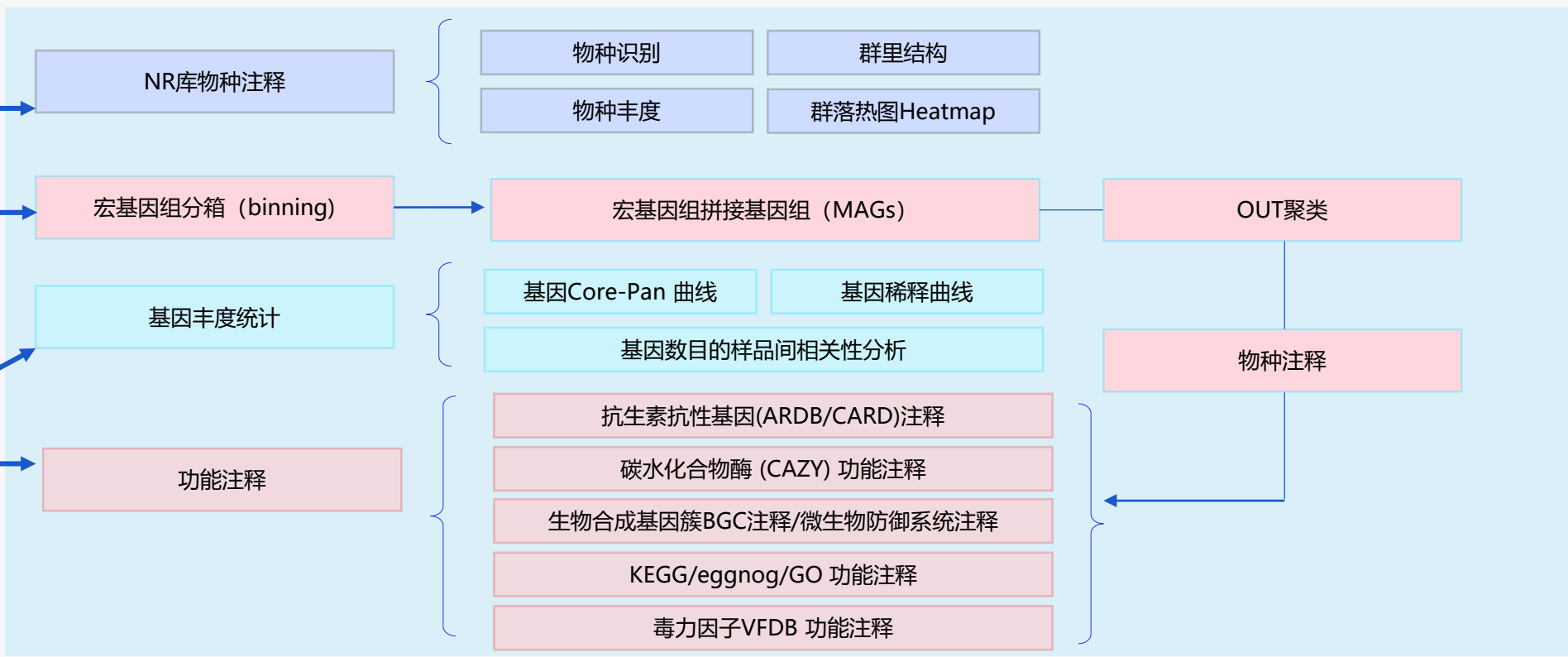
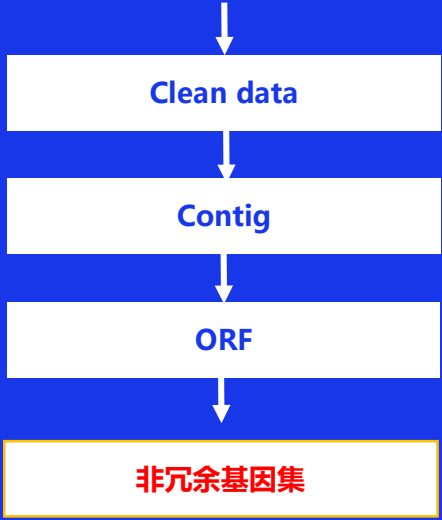
Home About Articles Sections Collections Submission Guidelines [Submit manuscript](#)

Research | [Open access](#) | [Published: 28 August 2021](#)

### Longitudinal study on the effects of growth-promoting and therapeutic antibiotics on the dynamics of chicken cloacal and litter microbiomes and resistomes

[Chhedi Lal Gupta](#), [Shlomo E. Blum](#) , [Karuppasamy Kattusamy](#), [Tali Daniel](#), [Shelly Druyan](#), [Roni Shapira](#), [Oleg Krifucks](#), [Yong-Guan Zhu](#), [Xin-Yuan Zhou](#), [Jian-Qiang Su](#) & [Eddie Cytryn](#)

# 宏基因组测序数据



**Alpha多样性:** 某个群落内部的物种多样性。包括群落内物种的个数(species richness, 丰富度)以及每个物种的数量及分布(evenness, 均匀度)。

**Beta多样性:** 即在一个梯度上从一个生境到另一个生境所发生的种的多样性变化的速率和范围

# 汇报提纲

CONTENTS

01

微生物组学数据分析及流程

02

国家微生物科学数据中心的数据资源及平台

# 国家微生物科学数据中心 (NMDC)、世界微生物数据中心 (WDCM) 总体情况

- 科技部、财政部联合建立的二十家国家科学数据中心
- 我国生命科学领域**唯一**的世界数据中心



www.nmdc.cn

## 数据汇交与资源整合

建有**700**余个数据库，对31个专项的**496**个重点研发专项开展数据汇交，**全球60**个微生物领域数据库的**全部镜像**

## 国际领先的标准化体系

牵头发布**2**个国际ISO数据标准、**2**个国家标准和**5**个团体标准，形成微生物领域国际引领的标准化体系

## 全面的分中心体系

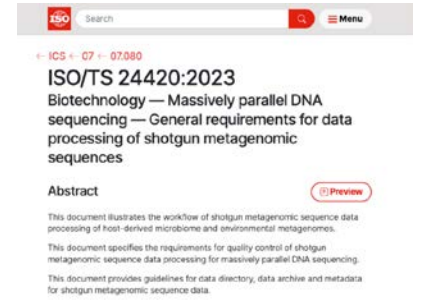
目前在国家疾控中心（**病原**）、**临床**、**军队**、**海洋**、**合成生物学领域**建有分中心

## 服务成效

有效服务于国家疾控、食品安全等国家战略需求  
通过**可用不可见**的技术手段，有效解决国家重要科学数据的有效利用

## 国际引领

牵头**全球51**个国家**152**个机构参加的微生物资源数据**国际合作**，形成**全球数据虹吸效应**，一系列数据产品具有重要国际影响力



## 两项国际ISO标准



## 三次世界互联网大会获奖



## 国家微生物科学数据中心 National Microbiology Data Center

## 重点专题数据库

### 重点数据库 / CHARACTERISTIC RESOURCES

<b>VarEPS</b> 新型冠状病毒变异点评...	<b>VarEPS-Influ</b> 流感病毒变异点风险评...	<b>gcPathogen</b> 全球病原微生物目录	<b>HIV</b> 中国艾滋病病毒基因序列...	<b>GCM</b> 全球微生物菌种目录
<b>gcType</b> 全球模式菌株测序计划	<b>Refs</b> 参考菌株数据库	<b>gcMeta</b> 微生物基因组平台	<b>ABC</b> 微生物资源利用知识库	<b>FungalNames</b> 真菌微生物名录

## 科学数据汇交

### 科技项目汇交专题 / SPECIAL TOPIC OF SCIENTIFIC PROJECTS

汇入口

17,2915 科学数据 | 194,921 图片及影像视频 | 147,945 文档及文字记录 | 305 项目 | 3,705 数据集 | 6,193 论文著作

<b>中国自然疫源性传染病流行病学数据库</b> 编号: 2013FY114600 负责人: 曹秀春 单位: 中国人民解放军军事医学科学院	<b>我国近海海洋生物DNA条形码资源库构建</b> 编号: 2014FY110500 负责人: 李新正 单位: 中国科学院海洋研究所
<b>中国孢子植物志的编研</b> 编号: 2013FY110400 负责人: 庄文娟 单位: 中国科学院微生物研究所	<b>西南喀斯特地区特色生物资源及多样性调查</b> 编号: 2014FY120100 负责人: 蔡磊 刘作勇 单位: 中国科学院微生物研究所

## 文章发表的数据提交

### 微生物数据资源 / MICROBIAL DATA RESOURCES

<b>生物项目数据</b> 685,144	<b>生物样本数据</b> 16,224,297	<b>核酸序列数据</b> 18,166	<b>原始组学数据</b> 863,026
<b>宏基因组数据</b> 106,807	<b>多元组学数据</b>	<b>蛋白质序列数据</b> 568,002	<b>晶体结构数据</b> 97,647

## 数据分析流程

### 分析平台 / ANALYSIS PLATFORM

查看更多 >

<b>Blast分析工具</b>	<b>宏基因组分析流程</b>	<b>基因组拼接工具</b>	<b>基因组结构分析</b>
<b>基因组注释分析</b>	<b>元基因组分析</b>	<b>比较基因组分析</b>	<b>便捷分析工具</b>





# 国际重要微生物数据全量备份 — 保障国家数据安全

## 已提供60个国际数据库的中国本地下载服务

(包括NCBI上26个与微生物有关的数据库, 20个EBI、JGI、MG-RAST、GOLD、KEGG等其他数据库)

### 核酸及蛋白质序列数据库

#### Genbank-核酸数据库

描述: GenBank-核酸数据库是美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)建立的一个全面的数据库, 包含近26万个物种的可公开的核苷酸序列。GenBank与欧洲的EMBL和日本的DDBJ之间每天进行数据交换确保了全球范围的数据覆盖。可通过FTP获取版本数据集和每日更新数据集。

中文关键词: 核苷酸序列

英文关键词: nucleotide sequence

文件大小: 1.1T

更新日期: 2023-12-09

下载链接: <ftp://download.nmdc.cn/genbank>

### 核酸及蛋白质序列数据库

#### Genbank-细菌与古菌核酸序列数据集

描述: GenBank-核酸数据库是美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)建立的一个全面的数据库, 包含近26万个物种的可公开的核苷酸序列。GenBank与欧洲的EMBL和日本的DDBJ之间每天进行数据交换确保了全球范围的数据覆盖。可通过FTP获取版本数据集和每日更新数据集。此数据集为Genbank细菌与古菌核酸序列数据集。

中文关键词: 核苷酸序列

英文关键词: nucleotide sequence

文件大小: 125G

更新日期: 2023-12-09

下载链接: [ftp://download.nmdc.cn/genbank\\_bct](ftp://download.nmdc.cn/genbank_bct)

### 核酸及蛋白质序列数据库

#### Genbank-病毒核酸序列数据集

描述: GenBank-核酸数据库是美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)建立的一个全面的数据库, 包含近26万个物种的可公开的核苷酸序列。GenBank与欧洲的EMBL和日本的DDBJ之间每天进行数据交换确保了全球范围的数据覆盖。可通过FTP获取版本数据集和每日更新数据集。此数据集为Genbank病毒核酸序列数据集。

中文关键词: 核苷酸序列

英文关键词: nucleotide sequence

文件大小: 143G

更新日期: 2023-12-09

下载链接: [ftp://download.nmdc.cn/genbank\\_vrl](ftp://download.nmdc.cn/genbank_vrl)

### 核酸及蛋白质序列数据库

#### DDBJ-核酸数据库

描述: DDBJ(DNA Data Bank of Japan), 为一级核酸数据库, 收集是世界三大核酸数据库之一, 与NCBI的genbank, EBI的EMBL数据库共同组成国际核酸数据库。三个库之间每日都交换更新数据和信, 每年更新四个版本。

中文关键词: 核苷酸序列

英文关键词: nucleotide sequence

## 处理形成参考数据库, 用于数据分析

BlastN

\* Sequence

```
>AAO40200002.1
GAGTTTGATCATGGCTCAGATTGAACGCTGGGGCAGGCCAATACACATGCAAGTGGAGCGGAAAGTACTTCGGTATGAGTAGA
GGGGCGGACGGGTGAGTAAACGCTAGGAATCTATCCAGTAGTGGGGGACAACTCGGGGAACTCGAGCTAATACCGCATACGT
CCTAAGGGAGAAAGCGGGGATCTCGGACCTCGCGCTATTGGAGGAGCTGCGTTGGATTAGCTAGTTGGTGGGTAAGGCC
TACCAAGGGGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACACCGGACTGAGACAGGCCCGGACTCTACGGGAGG
CAGCAGTGGGAAATTTGGCAATGGGCGAAAGCTGACGACGCAATGCCGGTGTGTGAAGAAGGCTTCGGTGTGAAGGC
ACTTTCATTTGGGAAGAAAGGTTAGTAGTATAAAGCTAGCTGTGACAT
```

example

\* Outfmt

6=Tabular

## 与自有数据进行整合

Home > Databases > BioProject

keywords

ID	NMDC Accession	Title	Organism Name	BioProject Type	Scope
1	PRJNA1	Comparison of the genome ...	Treponema denticola ATCC 35405	Genome sequencing	eMultiisolate
2	PRJNA8	Re-annotation and re-analys...	Campylobacter jejuni subsp. jejun...	Genome sequencing	eMultiisolate
3	PRJNA12	Genomic and genetic analys...	Pseudomonas fluorescens Pf0-1	Genome sequencing	eMultiisolate
4	PRJNA9	The complete genome sequ...	Francisella tularensis subsp. tulare...	Genome sequencing	eMultiisolate
5	PRJNA7	Genomic analysis of Campyl...	Campylobacter fetus subsp. vener...	Genome sequencing and ass...	eMultiisolate
6	PRJNA16	Genome sequence of Azoto...	Azotobacter vinelandii DJ	Genome sequencing	eMultiisolate
7	PRJNA23	Causes gonorrhoea	Neisseria gonorrhoeae FA 1090	Genome sequencing	eMultiisolate
8	PRJNA25	Comparative analysis of the ...	Bordetella parapertussis 12822	Genome sequencing	eMultiisolate
9	PRJNA24	Comparative analysis of the ...	Bordetella bronchiseptica RB50	Genome sequencing	eMultiisolate
10	PRJNA21	Genomic insights into metha...	Methylococcus capsulatus str. Bath	Genome sequencing	eMultiisolate

Total 68461 10/page < 1 2 3 4 5 6 ... 68466 > Go to 1

# 微生物计算环境及工具

## 8个流程88种在线数据分析工具

当前位置: 首页 > 分析工具

请输入关键词

搜索提示: SOAPdenovo2 / CANU / PfamScan / LEfSe /

- Blast分析工具 (5)
- 宏基因组分析流程 (3)
- 基因组拼接工具 (25)
- 基因组结构分析 (11)
- 基因组注释分析 (4)
- 元基因组分析 (20)
- 比较基因组分析 (13)
- 便捷分析工具 (7)

### BlastN

原始链接: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

工具介绍: 将给定的核酸序列与核酸数据库中的序列进行比对

关键字: local alignment

文章: Altschul, S.F., Gish, W., Miller, W., Myers, E.W., ...

[查看详情](#) [开始使用](#)

### BlastP

原始链接: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

工具介绍: 使用蛋白质序列与蛋白质数据库中的序列进行比对

关键字: local alignment

文章: Altschul, S.F., Gish, W., Miller, W., Myers, E.W., ...

[查看详情](#) [开始使用](#)

### BlastX

原始链接: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

工具介绍: 将给定的核酸序列与蛋白质数据库中的序列进行...

关键字: local alignment

文章: Altschul, S.F., Gish, W., Miller, W., Myers, E.W., ...

[查看详情](#) [开始使用](#)

### TblastN

原始链接: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

工具介绍: 将给定的蛋白序列与核酸数据库中的序列进行比对

关键字: local alignment

文章: Altschul, S.F., Gish, W., Miller, W., Myers, E.W., ...

[查看详情](#) [开始使用](#)

### TblastX

原始链接: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

工具介绍: 只在特殊情况下使用, 它将DNA被检索的序列和...

关键字: local alignment

文章: Altschul, S.F., Gish, W., Miller, W., Myers, E.W., ...

[查看详情](#) [开始使用](#)

### 基因组拼接工具

#### SOAPdenovo2

原始链接: <https://github.com/aquaskeyline/soapdenovo2>

中文描述: SOAPdenovo2是一款基于短读长拼接的软件, 针对人类基因组和illumina测序仪均有优化。

英文描述: SOAPdenovo2 is a novel short-read assembly method that can build a denovo draft assembly for the human-sized genomes. The program is specially designed to assemble Illumina GA short reads. It creates new opportunities for building reference sequences and carrying out accurate analyses of unexplored genomes in a cost effective way

关键字: large genome assembly sets

文章: Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience, 1, 18.

使用帮助: [SOAPdenovo2 教程](#) [SOAPdenovo2 使用须知](#)

- Blast分析工具 (5)
- 宏基因组分析流程 (3)
- 基因组拼接工具 (25)
- 基因组结构分析 (11)
- 基因组注释分析 (4)
- 元基因组分析 (20)
- 比较基因组分析 (13)
- 便捷分析工具 (7)

### SOAPdenovo2

原始链接: <http://github.com/aq...>

工具介绍: SOAPdenovo是一款基于短读长拼接的软件。针对...

关键字: large genome assem...

文章: Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan...

[查看详情](#) [开始使用](#)

### SPAdes

原始链接: <http://cab.spbu.ru/sof...>

工具介绍: SPAdes是一款整合型的基因组拼接软件, 适合...

关键字: genome, metagenom...

文章: Namiki, T., Hachiya, T., Tanaka, H. and...

[查看详情](#) [开始使用](#)

### MetaVelvet

原始链接: <http://www.ebi.ac.uk/...>

工具介绍: Velvet为一从头基因组组装工具, 专为短序列测...

关键字: metagenome assembly

文章: Namiki, T., Hachiya, T., Tanaka, H. and...

[查看详情](#) [开始使用](#)

### ALLPATH-LG

原始链接: <http://software.broad...>

工具介绍: 大型基因组拼接软件。

关键字: genome assembly

文章: Gnerre, S., Maccallum, I., Przybylski, D.,...

[查看详情](#) [开始使用](#)

### Meta-IDBA

原始链接: <http://i.cs.hku.hk/~als...>

工具介绍: 宏基因组拼接软件。

关键字: metagenome assembly

文章: Peng, Y., Leung, H.C., Yiu, S.M. and Chin, F.Y...

[查看详情](#) [开始使用](#)

### MEGAHIT

原始链接: <https://github.com/v...>

工具介绍: Megahit是一款针对大型复杂群落的宏基因组拼...

关键字: metagenome assembly

文章: Li, D., Liu, C.-M., Luo, R., Sadakane, K. and Lam...

[查看详情](#) [开始使用](#)

### RayMeta

原始链接: <http://denovoassemb...>

工具介绍: Ray为平行从头基因组组装工具, 工具内设计有...

关键字: metagenome assembly

文章: Boisvert, S., Raymond, F., Godzaridis, E.,...

[查看详情](#) [开始使用](#)

### CANU

原始链接: <https://github.com/m...>

工具介绍: Canu源自Celera组装工具, 主要用于高噪声单...

关键字: long reads assembly (...)

文章: Koren, S., Walenz, B.P., Berlin, K., Miller, J.R.,...

[查看详情](#) [开始使用](#)

### CAP3

原始链接: <http://doua.prabi.fr/s...>

工具介绍: 此程序可剪切序列5' 和3' 段低质量区。其以碱基...

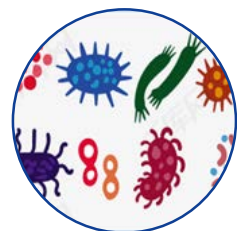
关键字: extension

文章: Huang, X. and Madan, A. (1999) CAP3: A DN...

[查看详情](#) [开始使用](#)

## 全球微生物资源目录

GCM

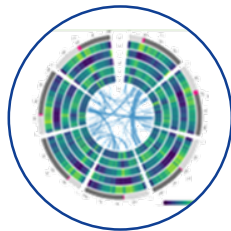


NAR database issue 2017

- 51个国家133个机构47万菌种资源
- 微生物资源相关物种、文献、生理生化等文本挖掘信息

## 全球模式微生物基因组

gcType

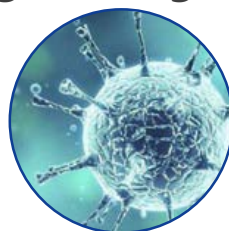


NAR database issue 2020

- 15个国家18个微生物机构, 16000个种模式微生物基因组

## 全球病原微生物基因组

gcPathogen



NAR database issue 2023

- 1400万基因组, 1100个种, 11类生态环境
- 4个知识库: 耐药基因、毒力基因、高致病性病毒等

## 全球真菌名录数据库

Fungal Names

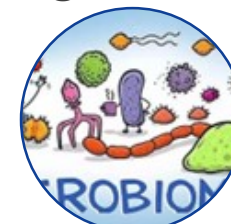


NAR database issue 2022

- 1567年至今数据
- 56.7万真菌名录信息
- 14.9万菌物信息
- 17023文章5153保藏中心

## 全球微生物宏基因组

gcMeta



NAR database issue 2019

- 120TB数据2000个样本
- 宏基因组拼接及注释信息
- 人体、冰川、土壤等专题资源

## 全球微生物参考菌株

gcRef



ISO标准正式发布 2020

- 国际ISO标准菌株信息
- 欧洲标准菌株信息
- 标准菌株基因组信息
- 标准菌株保藏和供应信息

## 全球新冠毒株及变异风险评估



NAR database issue 2021

- 1600万新冠病毒基因组
- 基于人工智能的风险评估
- 亲和力、免疫逃逸等

## 全球流感毒株及变异风险评估



NAR database issue 2023

- 100万流感基因组
- 传播力、跨种传播
- 流行病学数据及风险评估

# 全球微生物保藏中心注册系统CCINFO

## 世界菌种保藏联合会WFCC指定会员机构注册



WORLD FEDERATION FOR CULTURE COLLECTIONS

### WFCC Affiliate Membership application Additional Information Request Form

Culture Collections seeking admittance as WFCC Affiliates should in principle qualify to the standards laid down in the WFCC By-laws. The Executive Board of WFCC reviews all applications for new membership.

**WDCM number (required) :**.....

**Note:** WFCC can only assess your application if your collection is registered in the World Data Centre for Microorganisms (WDCM); please go to <http://www.wfcc.info/ccinfo/> where after registering in the CCINFO database (free of charge) you will receive a WDCM number.

Collection Full Name and  
Acronym:.....

Date and place:

Signature

Please send fully completed and signed form by e-mail to [m.groenewald@westerdijkinstituut.nl](mailto:m.groenewald@westerdijkinstituut.nl) or by regular postal mail to Marizeth Groenewald, WFCC Board, Westerdijk Institute, P.O. Box 85167, 3508AD, Utrecht, The Netherlands.

## 欧洲菌种保藏联盟指定会员机构注册



European Culture Collections' Organisation

HOME ABOUT COLLECTIONS ECCO MTA AND MDA MEETINGS NEWS NAGOYA PROTOCOL

### ECCO STATUTES

Article 5. Members are based in European countries that are affiliated to the Federation of European Microbiological Societies (FEMS). There are four categories:

1. Individual Members, open to scientists.
2. Corporate Members, open to public or private entities. Culture collections which apply for Corporate Membership must additionally:
  1. be registered with the World Data Center (of WFCC);
  2. produce a catalogue of holdings;
  3. supply cultures to the scientific community.

## International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) 国际藻类、真菌和植物的命名法，指定微生物保藏机构缩写

- Art. 7. Typification in general
- Art. 8. Specimens and gatherings
- Art. 9. Type

40.7. For the name of a new species or infraspecific taxon published on or after 1 January 1990 of which the type is a specimen or unpublished illustration, the single herbarium, collection, or institution in which the type is conserved must be specified (see also Rec. 40A.5 and 40A.6).

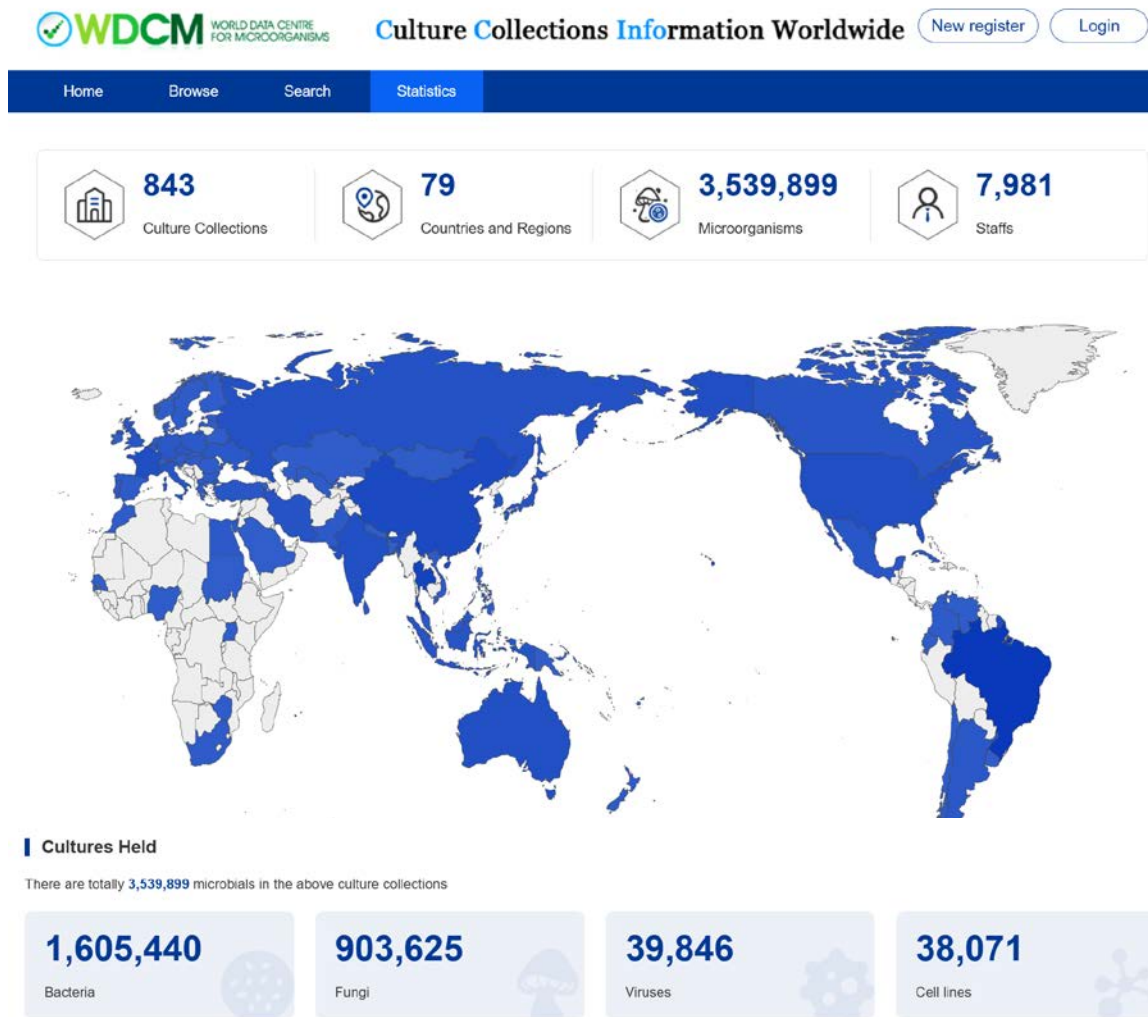
Ex. 8. In the protologue of *Setaria excurrens* var. *lesiflora* Kang ex S. L. Chen (in Bull. Nanjing Bot. Gard. 1988-1989: 3, 1990) the gathering Guangxi Team 4088 was indicated as “模式” [type] and the herbarium where the type is conserved was specified as “中国科学院植物研究所标本馆” [Herbarium, Institute of Botany, The Chinese Academy of Sciences], i.e. PE.

Note 4. Specification of the herbarium, collection, or institution may be made in an abbreviated form, e.g. as given in Index Herbariorum (<http://sweetgum.nybg.org/science/ih/>) or in the World Directory of Collections of Cultures of Microorganisms.

Ex. 9. When Y. Hart described “*Sesuvium eriocarpum* subsp. *spatinifolium*” (in Ot. Sist. Bot. Dergisi 2(2): 7, 1995) the name was not validly published because no herbarium, collection, or institution in which the holotype specimen was conserved was specified. Valid publication was effected when Y. Hart (in Strid & Tam, Fl. Hellen. 2: 325, 2002) wrote “Type ... Y. Hart HRT-27104 ... (U)” while providing a full and direct reference to his previously published Latin diagnosis (Art. 33.1).

确保全球微生物菌号体系的唯一性

# 全球保藏中心分布与保藏情况



Number of Holdings		
CCINFO	Collections	Strains
643	BCCM	216791
930	MCC	164652
825	NBRC	131062
550	CGMCC	107760
1194	BPUMS	100000
97	NRRL	96200
133	CBS	87500
1285	NIDB	86129
783	BCC	78202
59	BCRC	76339
1	ATCC	75079
1168	ECRC	75000
919	JMRC	50011
758	IBT	46500
597	KCTC	43829
32	CCUG	40500
1175	HGMCC	38050
637	MAFF	35804
274	DSMZ	33784
308	MUCL	32091

Number of Staffs	
Collection	Staffs
ATCC	400
DSMZ	155
CMRP	78
BR	67
BCRC	61
CICC	59
IBRC	56
EMCC	43
VKM	36
INACCS	35
INACCS	35
NBRC	32
KCTC	30
MDC	30
CBS	29
ESBL-Pak-121	29
TISTR	27
BACC	27
UCCAA	26
VURV	24

自2011年，全球新增299家，中国现有保藏中心48家

# 全球保藏中心服务

## 微生物分离、保藏及培养

- . Isolation and purification of strains
- . Freeze-drying
- . Optimisation of preservation conditions
- . Optimisation of cultivation/fermentation
- . Microbial counting/titer

## 分类工具及在线目录

- Bacdive/TYGS
- MycoBank
- YeastIP
- FungalDC
- Yeast-ID
- BIGSdb-Pasteur
- Klebsiella MALDI TypeR
- CLIMA

## 筛选、检测及鉴定

- Growth promoting / antimicrobial / antiviral bioassays
- . Microbial growth-promoting and antimicrobial tests
- . Antibiotic resistance assays
- . Biocontrol agents tests on plants
- . Biostimulating tests on plants
- . Virus resistance assays
- High-throughput screening
- . Metabolomic analyses
- . Analysis of the resistance/sensitivity of strains to physical and chemical stressors
- . Analysis of the strain performance for industrial application
- . Analysis of adhesive activity
- . Analysis of biosurfactant-producing activity
- . Detection of contaminants in raw materials and products
- . Material resistance testing

## 咨询、培训及合作研究

- . Consultancy (topics aligned with the MIRRI Clusters of Expertise)
- . Training courses
- . Contract Research

## 其他服务

- Other characterisation analyses
- . Mycovirus detection
- . Determination O<sub>2</sub> consumption / CO<sub>2</sub> production
- . Plasmid copy number quantification
- . Safety assessment of strains for food and feed
- Purification of cells/metabolites
- . Cell sorting applications (Flow cytometry)
- . Purification of metabolites
- Complementary services
- . DNA extraction
- . Construction and characterisation of intraspecific hybrids

# 物种及资源层面全球微生物菌种目录Global Catalogue of Microorganisms(GCM)

GCM Global Catalogue of Microorganisms | WDCM NMDC

Home Participants Statistics Join Us ACM Members ANRRC Members Help

Global Catalogue of Microorganisms

Please enter the keyword, Bacillus subtilis

Search Advanced Search Homology Search Species Info

ex: Bacillus subtilis | Escherichia coli | Halomicrobium mukohataei | Salmonella enterica

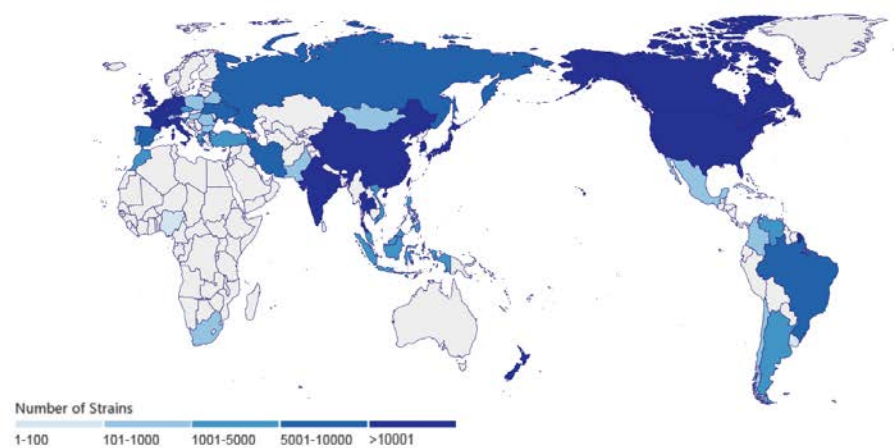
Culture Collections	Countries & regions	Species	Strains	Type Strains	Nucleotide	Literatures
151	51	51,630	530,386	24,130	8,596,028	29,908

<https://gcm.wdcm.org/>

- 2021年底: **133**个机构**47万**菌种资源
- 2022年底: **146**个机构**52万**菌株资源
- 2023年: **151**个机构**53万**菌株资源

	Country	Culture collection	Strains	Scale
1	Germany	3	59511	11.31%
2	Netherlands	1	58641	11.14%
3	Belarus	7	52825	10.04%
4	Japan	3	43055	8.18%
5	China	12	37760	7.17%
6	USA	5	36459	6.93%
7	Thailand	5	30999	5.89%
8	Korea	8	28953	5.50%
9	France	6	21776	4.14%
10	New Zealand	1	18340	3.48%
11	India	7	17960	3.41%
12	UK	4	17534	3.33%
13	Italy	3	17283	3.28%
14	Canada	1	10441	1.98%
15	Russia	6	9743	1.85%
16	Iran	6	6724	1.28%
17	Brazil	7	6618	1.26%
18	Ukraine	2	6395	1.22%
19	Spain	1	6057	1.15%
20	Czech	3	4961	0.94%
	Others	57	34264	6.51%
	<b>Total</b>	<b>148</b>	<b>526299</b>	

## Culture Collections Distribution



## Top update Collections

SN	Collection	Strains	Country
1	TBRC	21,346	Thailand
2	IRAN	4,560	Iran
3	BNCC	3,070	China
4	CMPUJ	735	Colombia
5	NBIMCC	350	Bulgaria
6	FRIM-MCC	294	Malaysia
7	MNCCM	203	Mongolia
8	mRB-UFRO	200	Chile
9	MAKU-MACC	90	Turkey

- 新增知识图谱、API关联服务等功能

- 服务于CBD/NP及DSI的实施

Scientific Name	<i>Bacillus subtilis</i> <span style="color: orange;">gcType</span>
Type Strains	NRRL B-4219 <sup>T</sup> ; NCIMB 3610 <sup>T</sup> ; VKM B-501 <sup>T</sup> ; CCUG 163 B <sup>T</sup> ; NBRC 13719 <sup>T</sup> (GCM10014672); NCFB 1769 <sup>T</sup> ; NRRL NRS-744 <sup>T</sup> ; NCDO 1769 <sup>T</sup> ; NCAIM B.01095 <sup>T</sup> ; CIP 52.65 <sup>T</sup> ; CCM 2216 <sup>T</sup> ; NBRC 16412 <sup>T</sup> ; JCM 1465 <sup>T</sup> ; NCCB 53016 <sup>T</sup> ; NRRL NRS-1315 <sup>T</sup> ; DSM 10 <sup>T</sup> (GCM10020737); CCRC 10255 <sup>T</sup> ; NCCB 70064 <sup>T</sup> ; CFBP 4228 <sup>T</sup> ; ATCC 6051-U <sup>T</sup> ; IMET 10758 <sup>T</sup> ; LMG 7135 <sup>T</sup> ; CCUG 163 <sup>T</sup> ; ATCC 6051 <sup>T</sup> ; IAM 12118 <sup>T</sup> ; NCCB 32009 <sup>T</sup> ; BCRC 10255 <sup>T</sup> ; NCTC 3610 <sup>T</sup> (GCM10015349); NCIB 3610 <sup>T</sup> (GCM10017364 GCM10018250 GCM10018249);

Species Related Information in GCM

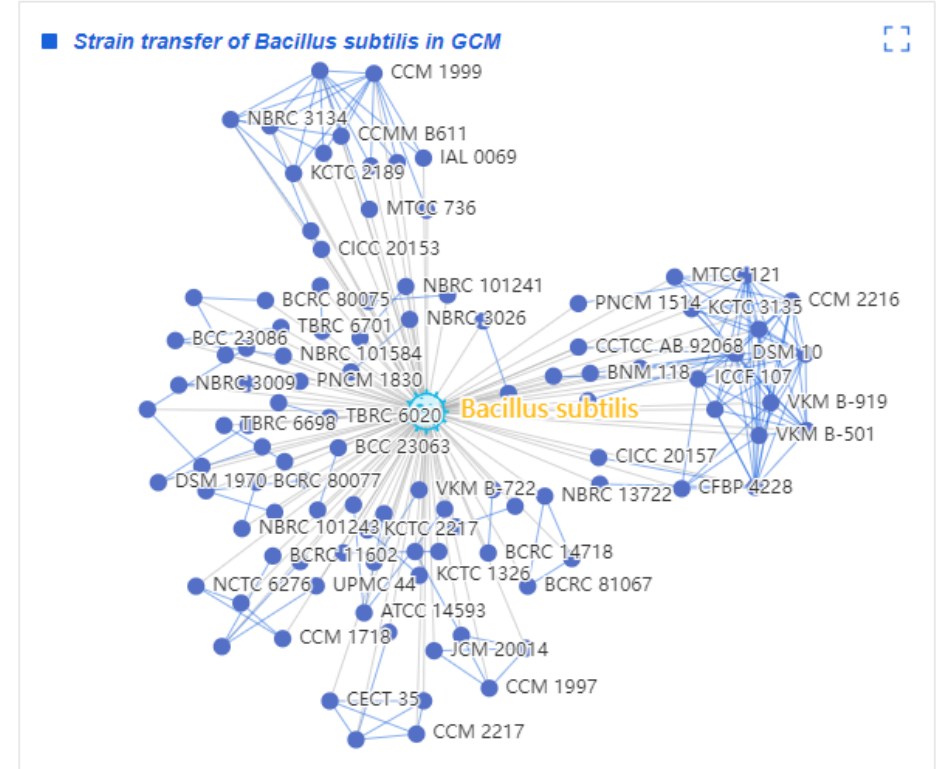
Strains	1,027	Culture Collections	60	Literatures	3,131
Genomes	367	Nucleotide	24,935	Patents	126

Bacillus subtilis cultured in Collections

Argentina (BNM)	Belarus (BIM)
Belgium (BCCM/LMG)	Brazil (IAL)
Bulgaria (NBIMCC)	Canada (UAMH)
China (ACCC,CCTCC,CICC,...)	Chinese Taipei (BCRC)
Colombia (CMPUJ)	Czech Republic (CCM)
France (CIP,CIRM-CFBP)	Germany (DSMZ,JMRC)
Greece (ACA-DC)	Hungary (DACT)
India (MCC,MTCC,NAIMCC,...)	Indonesia (FORDA-CC,I3LMCC,...)

Bacillus subtilis geographic origin of strains in the world

China (36)	Japan (32)	Thailand (22)
Netherlands (15)	Romania (14)	Korea (12)
Nigeria (10)	Morocco (8)	Greece (8)
Ghana (8)	Spain (7)	India (7)
Italy (7)	Indonesia (6)	France (6)
Germany (6)	U.K. (3)	Chinese Taipei (3)
Turkey (2)	Malaysia (2)	Hungary (2)
USA (2)	Austria (1)	Denmark (1)



Air	Engineered	Environmental	Fermentation products	Food
5	2	16	12	10
Genetic engineering strain	Host-associated	Marine	Microbial-mat-biofilm	Plant-associated
26	28	2		70
Sediment	Sludge wastewater	Soil	Water	Others
8		160	13	672





## Strain Information

Strain Number	JCM 10629 <a href="#">Original Site</a>
GUID	<a href="https://doi.org/10.12210/JCM10629">https://doi.org/10.12210/JCM10629</a>
Original Catalogue URL	<a href="https://www.jcm.riken.jp/cgi-bin/jcm/jcm_number?JCM=10629">https://www.jcm.riken.jp/cgi-bin/jcm/jcm_number?JCM=10629</a>
WDCM number of Culture Collection	567 <a href="#">CCINFO</a>
Other Collection Numbers	ATCC 33234 ; DSM 402 ; KCTC 2217 ; NCIMB 10106
Organism Name	<i>Bacillus subtilis</i> subsp. <i>subtilis</i>
Organism Type	Bacterium
History of Deposit	H. Yoshikawa <-- S. Dusko Ehrich <-- C. Anagnostopoulos <-- J. Spizizen 168 (trpC2).
Optimum Temperature for Growth	37C

to create DOI for preserved resources

<https://doi.org/10.12210/jcm10696>



## Data from WDCM for any strain in GCM



## Strain Related Information

<a href="#">Literatures</a> 2	<a href="#">Genome</a> 0	<a href="#">Nucleotide Sequences</a> 1	<a href="#">Patents</a> 3
-------------------------------	--------------------------	--	---------------------------

<a href="#">Literatures(2)</a>		<a href="#">Genome(0)</a>	<a href="#">Nucleotide(1)</a>	<a href="#">Patents(3)</a>	
No.	Title			JournalTitle	Year
1	Enrichment and isolation of Flavobacterium strains with tolerance to high concentrations of cesium ion			SCIENTIFIC REPORTS	2016
2	Evaluation of the Bactericidal and Fungicidal Activities of Poly([2-(methacryloyloxy)ethyl]trimethyl Ammonium Chloride)(Poly (ME...			POLYMERS	2018

# GCM strains cover a wide range of isolation Source

Organism Type	Air	Engineered	Environmental	Fermentation products	Food	Host-associated	Marine	Microbial-mat-biofilm	Plant-associated	Sediment	Sludge wastewater	Soil	Water	Others
Algae	21	84	2,134	0	248	366	1,417	99	242	373	28	1,240	4,712	10,392
Antibody	0	0	0	0	0	0	0	0	0	0	0	0	0	48
Archaea	1	6	393	1	20	110	83	5	35	386	117	195	182	2,508
Bacterium	571	5,316	7,293	2,526	5,073	40,637	380	309	20,885	5,830	2,988	28,999	8,455	83,077
Chromista	0	0	0	0	0	0	0	0	0	0	0	3	0	27
Cyanobacteria	0	0	142	0	0	0	16	44	4	0	0	154	190	771
Fungi	1,814	28,816	7,357	967	4,539	15,595	25	17	37,073	270	287	26,097	858	100,215
Moss	0	0	5	0	0	0	0	0	0	0	0	2	0	0
Phage	1	9	0	0	1	6	0	0	5	1	40	17	17	364
Plasmid	0	0	0	0	0	0	0	0	0	0	0	0	0	2,133
Protozoa	0	0	18	0	0	481	63	0	17	0	0	30	154	673
Virus	0	1	0	0	0	2	0	0	0	0	0	2	4	349
Yeast	230	2,156	2,039	2,832	4,872	6,718	5	3	10,285	1,336	137	5,058	2,306	24,917
Others	0	0	5	0	0	0	0	0	2	0	0	0	0	6

4539 Fungi strains was isolated from Food

Number of strains



# 物种及资源层面真菌名录数据库FungalNames

国际菌物命名委员会 (NCF) 授权的3个菌物命名注册库之一

**About Fungal Names**

Fungal Names, a global data repository of fungal taxonomy, is established by the Institute of Microbiology, Chinese Academy of Sciences. The repository aims at providing integrated services on Fungi and fungus-like organisms involving fungal name registration, species identification, specimen preservation, taxonomists overview and related information query, statistics or data sharing for people worked with or interested in mycology.

Rank	Fungi s.l.	Fungi s.s.
Kingdom	3	1
Phylum	21	20
Class	67	63
Order	284	267
Family	1077	1036
Genus	10866	10697
Species	149512	147705

**Name Registration**  
Register a fungal name of new taxon, new rank, new combination, replaced name, etc. or a new typification

**Identification**  
Preliminary identification & Phylogenetic trees

**Datasets**  
Public Culture Collections & Herbaria for Microbes  
Journals & Publications on Fungal Taxonomy  
Taxonomists

**Tools & Resources**  
Name Standardization Tools  
Fungal Name List Download  
International Code of Nomenclature

**New Release**  
2 Dec./2022  
Today's new releases of fungal names: 0 records  
26 Sep./2022  
Annual review on nomenclature novelties of fungi in the world (2021)

**Related Websites**  
Index Fungorum  
Mycobank  
Fungarium (IRIAS), Institute of Microbiology, CAS  
Fungarium Union of China  
Macrofungi Outline (Basidiomycota)

- Fungal Names更新成为集菌物名称注册与菌物分类学知识为一体的综合信息服务平台。
- 已收录1596年以来出版的56.7万个菌物分类单元名称、模式信息，囊括当前全球已认知的14.9万个菌物物种，并定期同步更新3个菌物名称注册网站新释放名称信息（Fungal Names, Index Fungorum, MycoBank）。
- 提供已有菌物名称的综合信息（现用名，异名，分类信息，模式标本，保藏单位，发表信息等）。
- 提供全方位的菌物（包括狭义和广义真菌）注册、检索、鉴定、下载、数据标准化与可视化等服务。
- 多种实时统计数据展示（各级分类单元现用名，年度发表菌物名称最多的前10个属、作者、期刊，年度模式标本最多前10个的国家、保藏单位等）。

2021-2022年新注册物种182个

*Nucleic Acids Research*, 2022, 1  
<https://doi.org/10.1093/nar/gkac926>

## Fungal names: a comprehensive nomenclatural repository and knowledge base for fungal taxonomy

Fang Wang<sup>2,3,†</sup>, Ke Wang<sup>1,†</sup>, Lei Cai<sup>1</sup>, Mingjun Zhao<sup>1</sup>, Paul M. Kirk<sup>4</sup>, Guomei Fan<sup>2,3</sup>, Qinglan Sun<sup>2,3</sup>, Bo Li<sup>2,3</sup>, Shuai Wang<sup>2,3</sup>, Zhengfei Yu<sup>2,3</sup>, Dong Han<sup>2,3</sup>, Juncai Ma<sup>2,3</sup>, Linhuan Wu<sup>2,3,\*</sup> and Yijian Yao<sup>1,\*</sup>

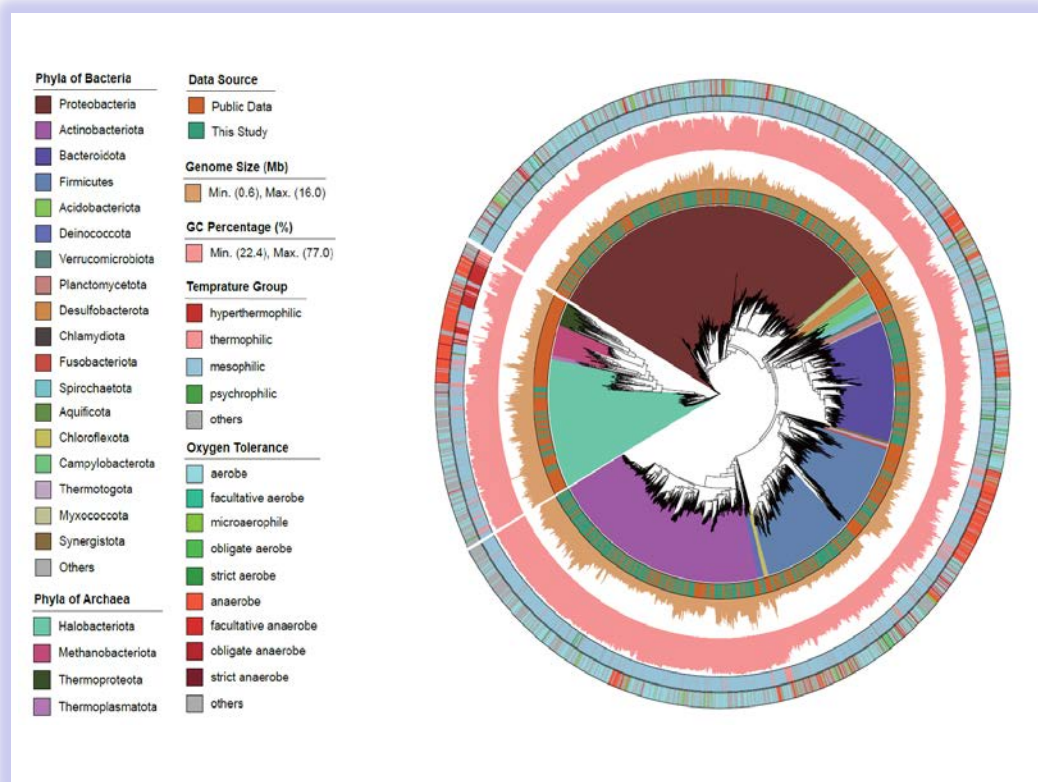
<sup>1</sup>State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China, <sup>2</sup>Microbial Resource and Big Data Center, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China, <sup>3</sup>Chinese National Microbiology Data Center (NMDC), Beijing 100101, China and <sup>4</sup>Jodrell Laboratory, Royal Botanical Gardens, Kew, Richmond, Surrey TW9 3DS, UK

Received August 08, 2022; Revised September 14, 2022; Editorial Decision September 30, 2022; Accepted October 07, 2022

# 全球万种微生物模式菌株测序计划及基因组数据库

40%

本计划填补了2018年之前全球已有物种缺失模式菌基因组测序约**40%**的空白



	测序中心	公布序列数
1	DOE Joint Genome Institute (JGI)	3066
2	<b>WDCM GCM 10K project</b>	<b>3911</b>
3	National Institute of Technology and Evaluation, Japan	386
4	University of Tokyo, Japan	272
5	Shanghai Majorbio Bio-pharm Technology Co.	183
6	J. Craig Venter Institute (JCVI)	147
7	Broad Institute	131
8	Washington University in St. Louis	125
9	Wellcome Trust Sanger Institute	97
10	Baylor College of Medicine	95
	OTHERS	7784

# 微生物模式基因组数据库

<http://gctype.wdcm.org/>

Global Catalogue  
of Microorganisms

## Type Strains Genome Database

Login

species/strain/genome/nucleotide/accession number  Search

Examples: *Streptomyces rubiginosus* / JCM 4416 / GCM60017754 [Advanced search](#)

**20,478** Species with validly published names

**85,086** Number of Type strains

**20,117** Number of Type strain Genomes

**4,805** Genomes sequenced by GCM

### Cite Us

- Shi W, Sun Q, Fan G, et al. gctype: a high-quality type strain genome database ... *Nucleic Acids Research*, 2020. [\[LINK\]](#) [\[PDF\]](#) [\[BibTeX\]](#) [\[EndNote\]](#)
- Wu L, McCluskey K, Desmeth P, et al. The global catalogue of microorganisms 10K ty... *GigaScience*, 2018, 7(6): gy026. [\[LINK\]](#) [\[PDF\]](#) [\[BibTeX\]](#) [\[EndNote\]](#)
- Wu L, Ma J. The Global Catalogue of Microorganisms (GCM) 10K type str... *Int J Syst Evol Microbiol*, 2019, 69(4): 895-898. [\[LINK\]](#) [\[PDF\]](#) [\[BibTeX\]](#) [\[EndNote\]](#)

### About the Project

Genomic information is essential to underpin taxonomic, phylogenetic, evolutionary and functional studies to comprehensively decipher the characteristics of microorganisms, to explore microbiomes through metagenomics and also answer fundamental questions of nature and human life. The research related to the taxonomy, evolution and functions of microorganisms have now entered into the genomic era. However, there is still a big gap of genomic sequencing information for the bacterial and archaeal species with validly published names by the International Committee on Systematics of Prokaryotes (ICSP). For fungal type strains, the remaining gap is even larger. Global Catalogue of Microorganisms (GCM)'s led type strain sequencing project is an international coordinated effort to close current gaps in the genomic maps of microbes and hence to promote research through deep mining of the genomic data. The project includes two core subprojects: 10K bacterial and archaeal type strain sequencing. [More](#)

### Services

**Genome Analysis**  
Visit the tool's main page here

**Species Identification**  
Visit the tool's main page here

### Sequenced species in phylum

Phylum	Percentage	Count
Pseudomonadota	87%	6207/7152
Actinomycetota	88%	4076/4606
Bacillota	84%	2963/3528
Bacteroidota	83%	2068/2502
Euryarchaeota	86%	511/593
Thermodesulfobacteriota	75%	223/296

## Features

### Data Services

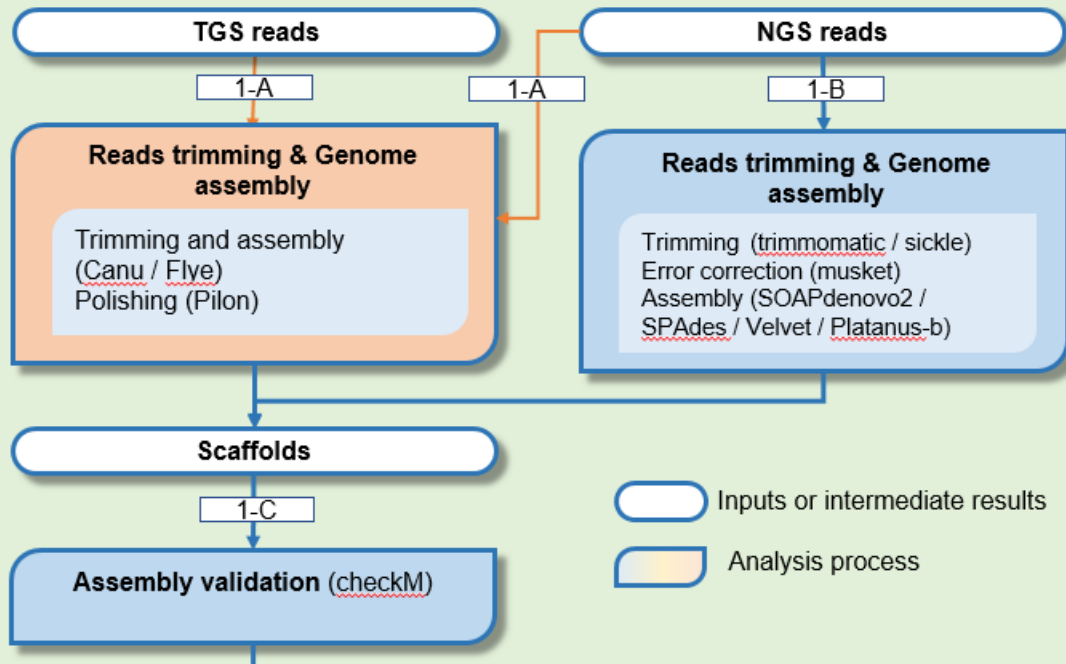
- Search the species, type strain, genome
- Browse the type strain by culture collections, check the sequencing status, sequences and annotation results

### Analysis Services

- Online genome assembly and annotation
- Novel species identification
- Phylogenetic analysis

# Providing updated reference data for global taxonomists

## 1. Processing of raw reads and assembly



## 2. Genomic component analysis

Gene prediction (Prodigal)  
 rRNA prediction (RNAmmer)  
 tRNA prediction (tRNAscan-SE)  
 CRISPR recognition (piller-cr)  
 Repeat detection (TRF)

## Gene

## 3. Gene annotation

KEGG, COG, Pfam, Swiss-Prot,  
 MetaCyc, PHI, CAZy, CARD, VFDB,  
 Anti-SMASH, ...

## Quality control and Assembly

NGS reads only  TGS reads only  NGS reads + TGS reads

inputReads\*

Add Grouped Multi File  
or drag & drop files here.

trimProgram\*

Sickle  Trimmomatic

assemblyProgram\*

SOAPdenovo2  SPAdes  Velvet  Platanus-b

sampleName\*

default

threads\*

12

## Genome structural analysis

structurePrograms\*

Prodigal  CheckM  RNAmmer  tRNAscan  TRF  PILER-CR

lineage\*

Bacteria  Archaea  Eukaryota

threads\*

12

## Genome annotation

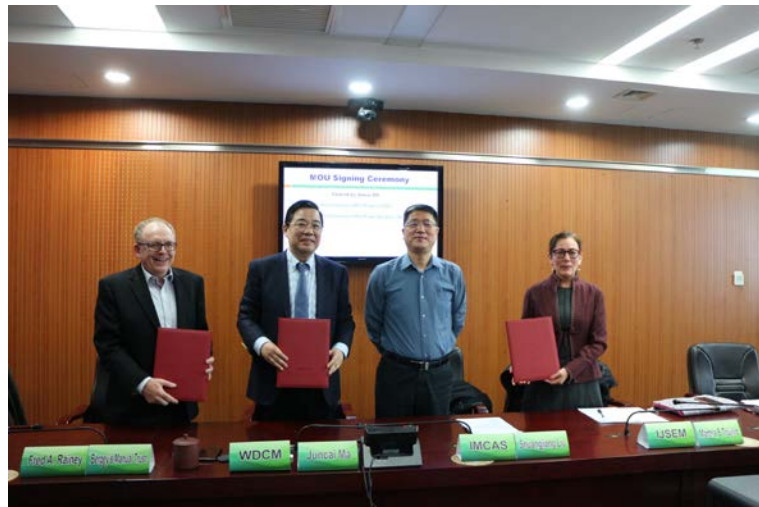
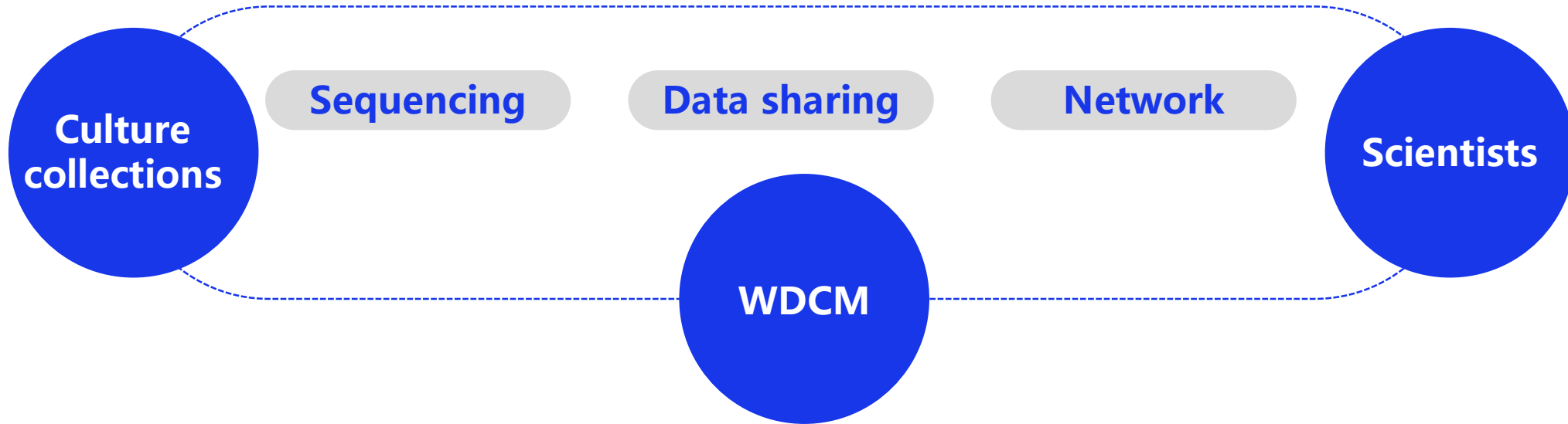
database\*

KEGG  COG  NR  CARD  CAZy  PHI  Swiss-Prot  VFDB  Pfam  MetaCyc  AntiSMASH

threads\*

12

# Cooperation with taxonomists



**INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY**  
Official publication of the ICSP and the BAM Division of the IUMS

About Editorial Board Latest Articles Current Issue Archive Col

**News : Genome sequencing data required with Taxonomic Descriptions.**  
08 December 2017  
*International Journal of Systematic and Evolutionary Microbiology (IJSEM)* will soon be asking authors to provide genome sequencing data with descriptions of novel taxa in Taxonomic Descriptions.

Although not mandatory for publication, the inclusion of this data is highly recommended and will be expected to be included. If authors are unable to provide genome sequencing data for any reason this should be stated as such in their covering letter; exemptions will be considered on a case by case basis by the handling Editor.

Genome sequences are of great value to the systematics of prokaryotes. In addition to improving the general understanding of the biology of microorganisms, they improve the identification of prokaryotic species, identification of functional characteristics useful for resolving taxonomic groups and the resolution of the phylogeny of higher taxa.

This requirement will come into effect from January, 2018 and we ask that authors deposit the data in an established, freely available public database that does not require viewer registration (i.e. GenBank, ENA or DDBJ).

For any queries regarding these requirements, please email [ijsem@microbiologysociety.org](mailto:ijsem@microbiologysociety.org).

**INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY**

**Bergey's Manual Trust**

# Cooperation with taxonomists

## *Clostridium prolinivorans* sp. nov., a thermophilic bacterium isolated from an anaerobic reactor degrading propionate

Yan Huang,<sup>1,2</sup> Zhixian Wei,<sup>1,2</sup> Lulu Cong,<sup>3</sup> Zhongwei Guo,<sup>1,2</sup> Rui Chen,<sup>1,2</sup> Yi Shichun Ma<sup>1,2,\*</sup>

**Author affiliations:** <sup>1</sup>Biogas Institute of Ministry of Agriculture, Section 4-13, Chengdu, Si and Application of Rural Renewable Energy of Ministry of Agriculture, Section 4-13, CI Geography and Limnology, Chinese Academy of Sciences, Nanjing 210006, PR China.  
**\*Correspondence:** Shichun Ma, mashichun@caas.cn  
**Keywords:** *Clostridium*; thermophilic; *Clostridium prolinivorans*; saccharolytic.  
**Abbreviation:** PYG, peptidic yeast glucose.  
The GenBank accession number for the 16S rRNA gene is MH020668. The DDBJ/ENA/Gen for the draft genome sequence are RJW00000030 and GCM60049815, respectively. Four supplementary figures are available with the online version of this article.

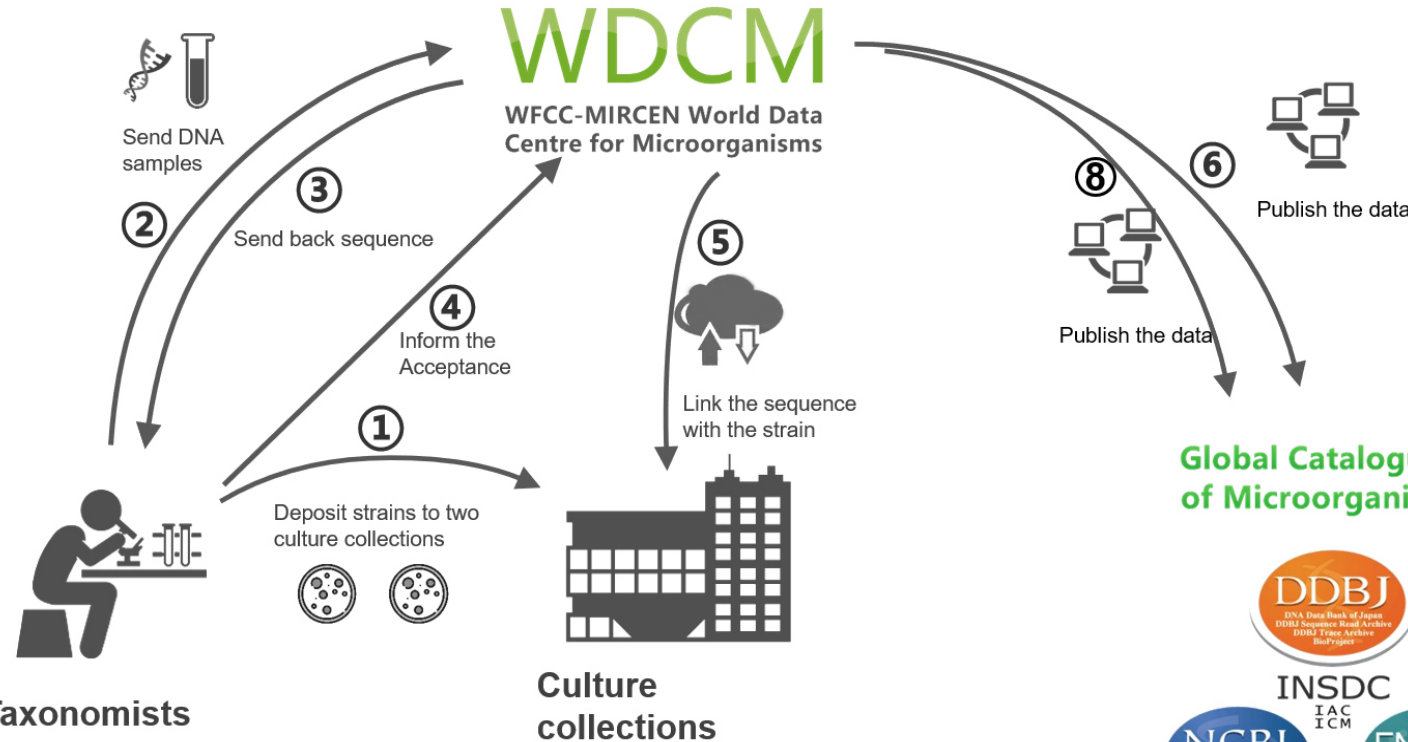
03523 © 2019 IJMS

## *Halostella limicola* sp. nov., isolated from saline soil sampled at the Tarim Basin

Dong Han, Li-Guo Hong, Qin Xu and Heng-Lin Cui\*

### Abstract

A halophilic archaeon, strain LT12<sup>T</sup>, was isolated from saline soil sampled at the Tarim Basin, PR China. The novel strain



## INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

Official publication of the ICSP and the BAM Division of the IUMS

390 species sequenced

Acronym	Full Name/Institution	Country
1 ATCC®	American Type Culture Collection	USA
2 BCRC	"Bioresource Collection and Research Center/Food Industry Research and Development Institute"	Chinese Taipei
3 CAIM	Collection of Aquatic Important Microorganisms/CIAD/Mazatlan Unit for Aquaculture and Environmental Management	Mexico
4 CCM	Czech Collection of Microorganisms/ Masaryk University	Czech
5 CCGU	Culture Collection University of Gothenburg	Sweden
6 CECT	Spanish Type Culture Collection / University of Valencia	Spain
7 CGMCC	China General Microbiological Culture Collection Center	China
8 CICC	China Center of Industrial Culture Collection	China
9 CIP	Collection de l'Institut Pasteur	France
10 DSMZ	Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH	Germany
11 ICMP	International Collection of Microorganisms from Plants	New Zealand
12 JCM	Japan Collection of Microorganisms/RIKEN BioResource Center	Japan
13 KACC	Korean Agricultural Culture Collection	Korea
14 KCTC	Korean Collection for Type Cultures, Korea Research Institute of Bioscience and Biotechnology	Korea
15 KMM	G.B. Elyakov Pacific Institute of Bioorganic Chemistry, Far-Eastern Branch, Russian Academy of Sciences	Russian
16 NBRC	Biological Resource Center/National Institute of Technology and Evaluation	Japan
17 NCAIM	National Collection of Agricultural and Industrial Microorganisms	Hungary
18 NCTC	National Collection of Type Cultures	United Kingdom
19 PCU	Pharmaceutical Sciences Chulalongkorn University Culture Collection/ Chulalongkorn University	Thailand
20 TBRC	Thailand Bioresource Research Center/National Center for Genetic Engineering and Biotechnology	Thailand
21 TISTR	TISTR Culture Collection/Bangkok MIRCEN	Thailand
22 VKM	All-Russian Collection of Microorganisms	Russian Federation



# 启动全球“未培养微生物培养组”计划，带动20个国家参与

目标：将全球可培养微生物的比例从1%大幅提高，建立新型资源库体系

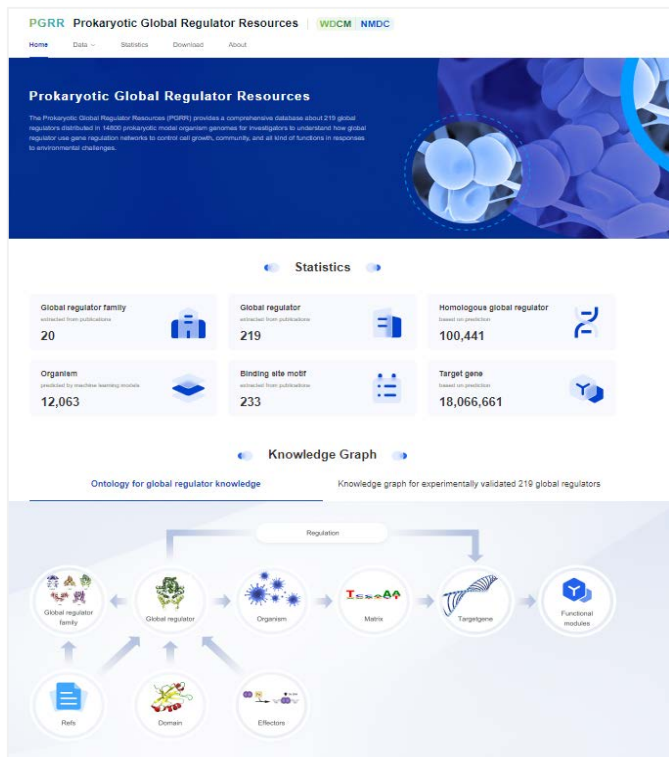


目标：

1. 建立一系列微生物采样、建库、测序及数据分析的**国际规范与标准**
2. 建立一套人工智能**大数据**指导下的、高通量定向培养组的**技术方案体系**
3. 共建一个覆盖全球广泛生境的微生物组**大数据平台**
4. **显著提升微生物新物种的发现和保藏**，挖掘大量新的功能元件，为合成生物学发展提供重要基础

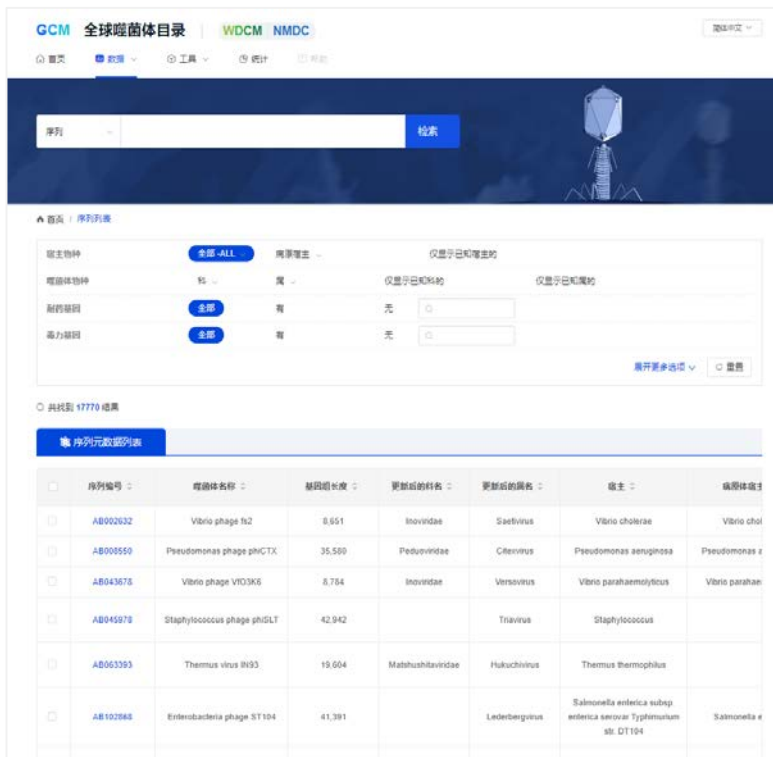
## 利用人工智能挖掘形成一系列重要的数字化元件库

### 全局调控因子数据库



12063个物种的100 441个调控元件，作用于18066661个靶基因，包括重要的代谢通路

### 噬菌体与微生物防御系统数据库



100万病原菌超过400万噬菌体及防御系统数据，并实现噬菌体与宿主的虚拟配型预测

### 移动元件与病原微生物耐药基因数据库



506种病原菌110万基因组7884种移动元件家族，对耐药基因的转移具有重要意义

# 宏基因组分析流程及数据库

## 1. Global sample collection

NCBI ~ 23.08 million BIOS amples (2022.01)

**Metadata collation and filtering**  
Obtain metagenomic data of more than 80,000 runs

Metadata of NCBI sample is extracted :  
Habitat coverage: air, human, animals, plants, soil, salt lakes, extreme environment, and other hundreds of habitats.

## 2. Analysis based on reads

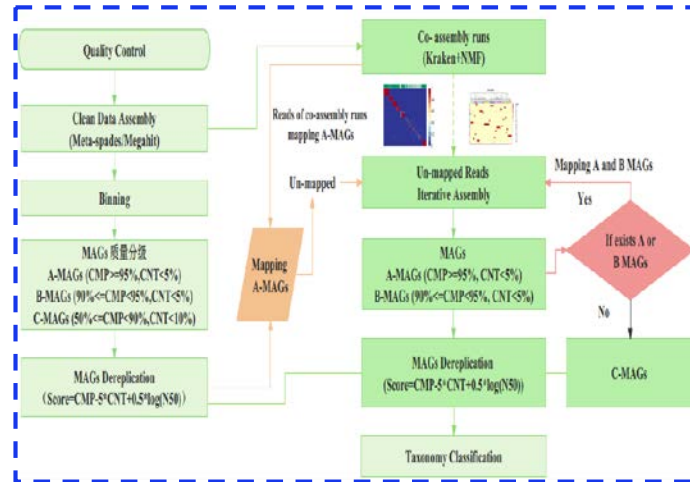
**Annotation analysis of raw data**  
(More than 50,000 runs have been completed)

Quality Control

Kraken Annotation

CARD Annotation

## 3. Metagenome assembly and binning



## 4. Genome annotation

System annotation based on structure and function  
Combining with the special research needs, the annotation base database is built.

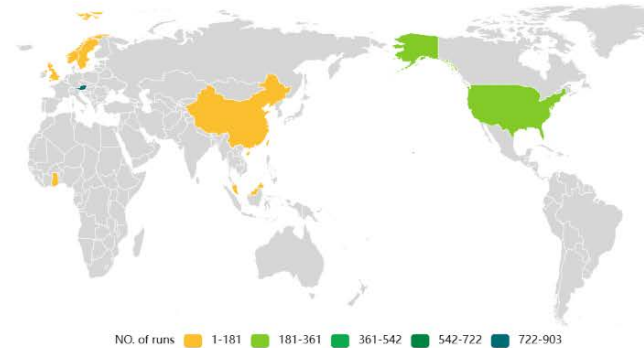
Structural Annotation  
(Prodigal, RNAmmer, tRNAscan...)

Functional Annotation  
(COG, KEGG, CARD, VFDB, CAZy, Pfam, Swissprot, Uniref100, Antismash, Rfam, MetaCyc...)

Functional Annotation  
(Self-database: (PcbC/EryCIII/GenB3/OxyB))

The screenshot shows the homepage of the Global Catalogue of Metagenomics. It features a search bar with a dropdown menu for 'Project' and a search button. Below the search bar, there is a list of statistics: Projects (6,575), Samples (63,803), Experiments (74,393), Runs (74,835), and MAGs (58,829). The website also includes a navigation menu with options like Home, Data, Function module, MAGs catalogues, Download, and Contact Us.

### Global distribution of sampling locations in certain biome



### Top Sampling Countries

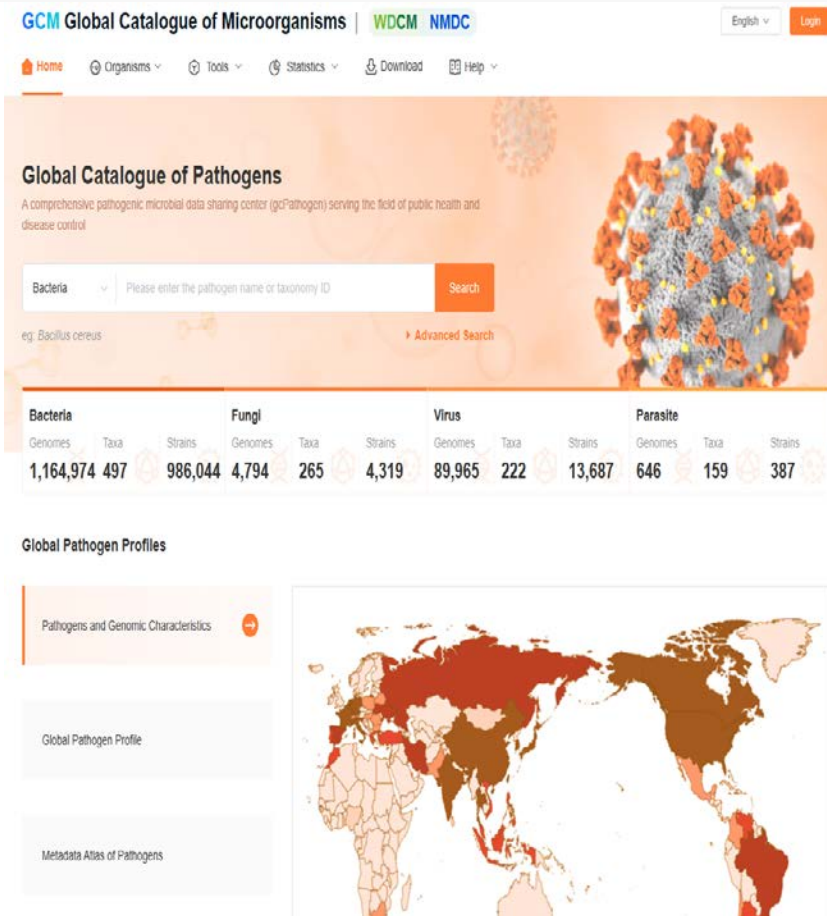
SN	Countries	Samples	Runs
1	Singapore	903	903
2	Austria	1	777
3	USA	233	233
4	Norway	126	126
5	China	100	100
6	U.K.	41	41
7	Sweden	32	32
8	Ghana	7	7
9	Malaysia	6	6
10	Puerto Rico	2	2

### Biomes

Human Gut	Host Associated	Human Associated	Water	Soil	Built Environment	Misc Environment	Human Oral
143,028	81,959	55,729	27,046	17,983	10,161	8,598	7,105
Wastewater Sludge	Human Skin	Plant Associated	Air	Sediment	Human Vaginal	Microbial Mat Biofilm	More >
6,886	5,329	4,962	2,647	1,920	1,251	59	

# 全球病原微生物数据库

- **病原名录**-中国卫健委、美国疾控中心、美国传染病学会和微生物学会的指导意见
- **病原全覆盖**-收录细菌497种、病毒222种、真菌265种，寄生虫159种
- **全球暴发史**-通过已发表的Review类文章、地方爆发报告和研究性文章，确认流行暴发史



## Meta库

### 流行病学数据

### 多组学数据库 测序信息 质控信息

### 时空数据库 气象、交通 环境、地理

### 耐药表型数据 药敏实验数据 临床症状数据 用药追踪数据

### 科学文献数据

## 参考序列库

- Reference序列
- 预警因子序列
- MLST管家基因
- cgMLST Schema

## 耐药基因库

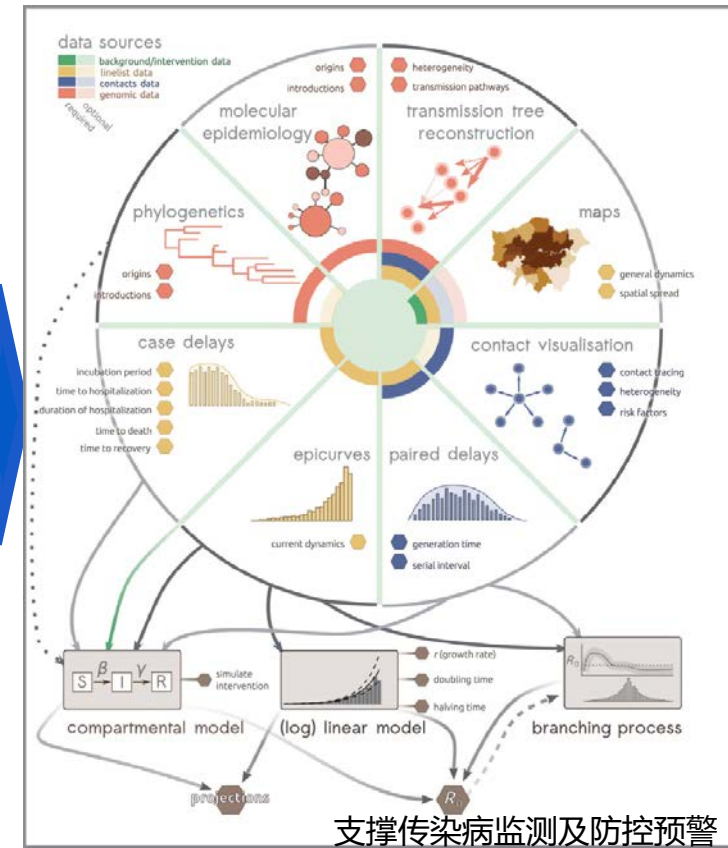
- 耐药基因序列
- 基因功能
- 耐药机制

## 致病因子库

- 致病因子序列
- 基因功能
- 致病机制

## 抗生素分子库

- 抗生素分类
- 抗生素功能机制





## 病原名录

中国卫健委、美国疾控中心、美国传染病学会和微生物学会的指导意见

## 病原全覆盖

收录细菌499种、病毒233种、真菌260种，寄生虫188种

## 全球暴发史

通过已发表的Review类文章、地方爆发报告和研究性文章，确认流行暴发史。

### Meta库

### 流行病学数据

### 多组学数据库

测序信息  
质控信息

### 时空数据库

气象、交通  
环境、地理

### 耐药表型数据

药敏实验数据  
临床症状数据  
用药追踪数据

### 科学文献数据

### 参考序列库

Reference序列  
预警因子序列  
MLST管家基因  
cgMLST Schema

### 耐药基因库

耐药基因序列  
基因功能  
耐药机制

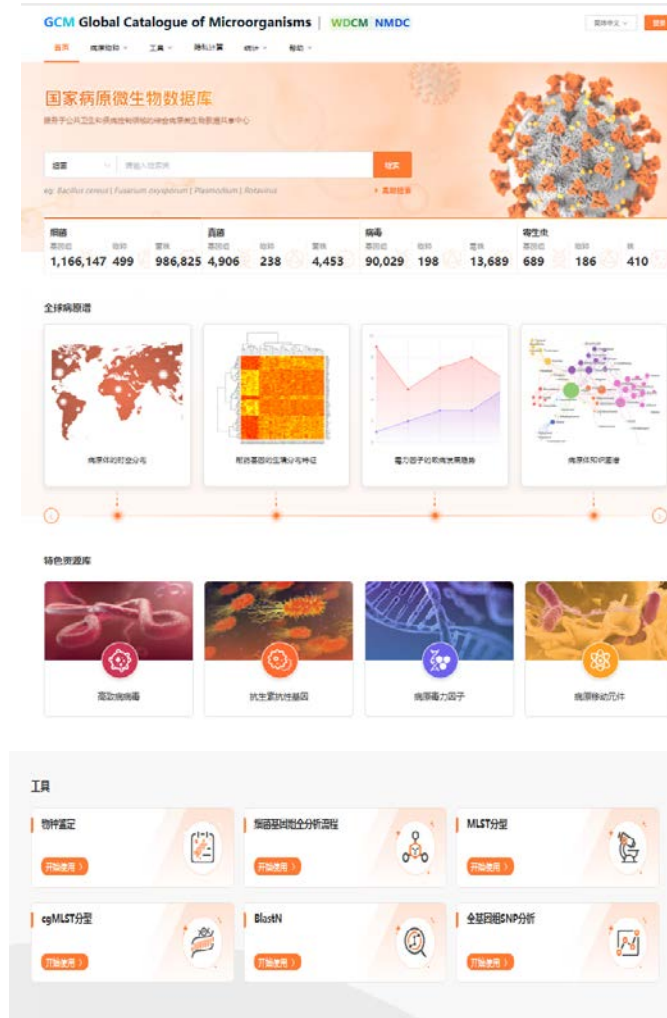
### 致病因子库

致病因子序列  
基因功能  
致病机制

### 抗生素分子库

抗生素分类  
抗生素功能机制

<https://nmdc.cn/gcpathogen/>





# 病原的全球数据整合与监测

首页总结了各类病原体的物种数和数据量

GCM Global Catalogue of Microorganisms | WDCM NMDC

Home Organisms Tools Publication Statistics Download Copyright Statement Cite us Help

## Global Catalogue of Pathogens

A comprehensive pathogenic microbial data sharing center (gcPathogen) serving the field of public health and disease control

Bacteria Please enter the pathogen name or taxonomy ID Search

ex: Bacillus cereus Advanced Search

细菌

真菌

寄生虫

病毒

Bact			Fungi			Virus			Parasite		
Genomes	Taxa	Strains	Genomes	Taxa	Strains	Genomes	Taxa	Strains	Genomes	Taxa	Strains
1,164,974	497	986,044	4,794	265	4,319	89,965	222	13,687	646	159	387

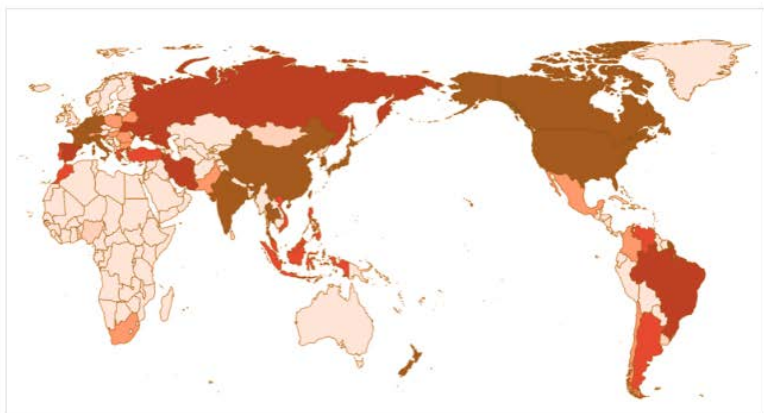
### Global Pathogen Profiles

Pathogens and Genomic Characteristics

Global Pathogen Profile

Metadata Atlas of Pathogens

Gene Co-occurrence Network



## E.coli Overview: 40年的样本、基因组>21万条、6340个STs

### Escherichia coli

Taxonomy ID: 562

Rank: species

Genome Count: 169581

GC (%): 36.7 ~ 63.0

Reference Paper: 7727633

Reference Genome: GCA\_000005845.2

Strain Count: 140849

Completed Genome Count: 2174

Collection Date: 1884 ~ 2022

Lineage: Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales; Enterobacteriaceae; Escherichia

提交单位: 87个

国家: 103个

分离源: 10种

宿主: 10类

宿主疾病: 9类

on (CDC) (43958) / National Center for Biotechnology Information (36487) / Food and Drug Administration (FDA) (25273) / IRN-CVM-FDA (2110) / US Food and Drug Administration (1867) / Hospital Henri Mondor APHP (1843) / Walter Reed Army Institute of Medicine (1289) / PASTEUR (1289) / Microbiology, Graduate School of Medical Sciences, Kyushu University (1288)

USA, China (5872), Australia (5744), Canada (4911), France (3856), Japan (3231), Germany (3147), Denmark (2035), Netherlands (1891)

Human, Animal, Human, Food, Urine, Blood/Body fluid, Environment, Abdominal drainage, Culture, Environment (hospital)

Companion animal, Environment, Food, Human, Laboratory, Livestock, Other, Plants, Poultry, Wild animal

Respiratory system, Urinary system, Reproductive system, Central nervous system, Bone/Joint, Eye/Ear/Oral/Nasal, Skin/Soft tissue, Digestive system

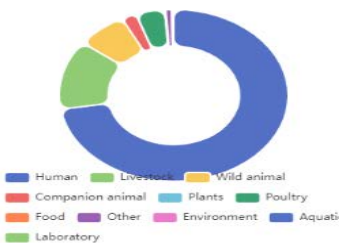
Drug Class: Fluoroquinolone antibiotic, Sulfonamide antibiotic, Antibiotic without defined classification, Beta-lactam antibiotic, Peptide antibiotic, Aminoglycoside antibiotic, Diaminopyrimidine antibiotic, Rifamycin antibiotic, Aminocoumarin antibiotic, Tetracycline antibiotic

Antibiotics: Fosfomycin, Norfloxacin, Nalidixic acid, Ciprofloxacin, Enoxacin, Streptomycin, Tetracycline, Sulfadoxine, Sulfadimidine, Sulfamethoxazole

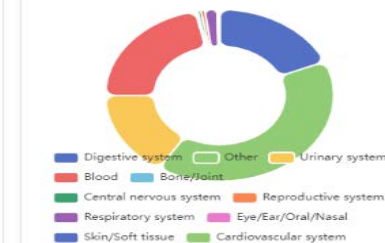
Antibiotic Resistance Genes Name: AcrB, AcrD, AcrF, EmrB, TolC, BaeR, MdtM, ATP-binding cassette (ABC) antibiotic efflux pump, Major facilitator superfamily (MFS) antibiotic efflux pump, Beta-lactamase

Virulence Factor Name: Peritrichous flagella, LOS, Capsule, TTSS secreted effectors, Type 1 fimbriae, ETT2, TTSS, Heme biosynthesis, LPS, Lateral flagella

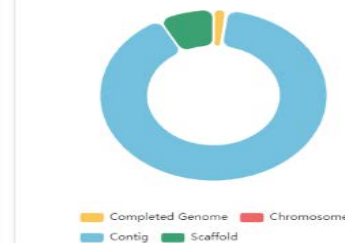
### 宿主分类占比



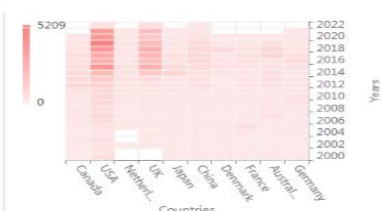
### 疾病分类占比



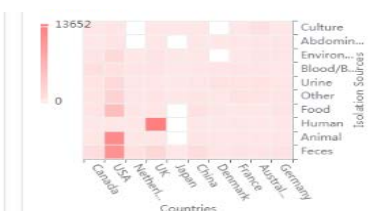
### 基因组类型



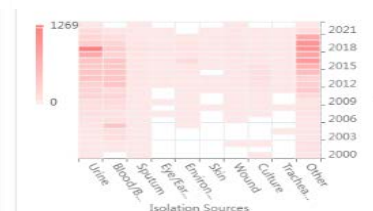
### E.coli的全球各国时空分布



### E.coli的全球各国宿主分布



### E.coli的分离源时间分布



## 物种鉴定

通过使用KRAKEN2、RNAmmer、BLASTn、Mash和FastANI对目标序列与参考库序列进行相似度比较，并给出最优比较结果，对病原序列进行物种鉴定。



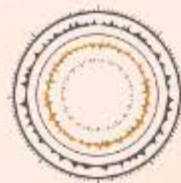
## 移动元件及相关抗性基因和毒力因子分析

用于病原菌中可移动元件的注释，及与其有交互关系的耐药基因和毒力因子的检测。



## 有参拼接组装

用于完成细菌全基因组的序列拼接和组装，而且还对拼接后序列进行了完整...



## BLAST-pathogen

通过将上传序列与我们自构建的病原菌和病毒全基因组序列进行比对，得到...



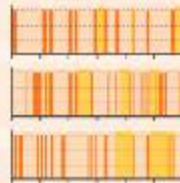
## MLST分型

MLST工具使用传统的PubMLST分型方案对目标序列进行测序分型。



## 无参拼接组装和基因组注释

用于基于以下几个数据库对查询基因...



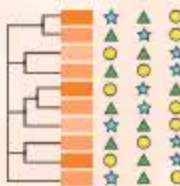
## SNP 分析

用于对细菌全基因组组装序列与数据库中的参考基因组进行比对，获得待测...



## cgMLST分型


用于病原体菌株的序列分型、遗传进...



# 建立了微生物领域完善的数据体系和国际权威数据库系统

## 全球产业用参考菌株数据库 Global Catalogue of Reference Strain


### Reference Strain Catalogue








#### Reference Strain Catalogue

This catalogue was produced to enable broader and easier access to the reference strains listed by the ISO TC 34 SC 9 Joint Working Group 5 and by the Working Party on Culture Media of the International Committee on Food Microbiology and Hygiene (ICFMH-WPCM) in their publication Handbook of Culture Media for Food and Water Microbiology. It fulfils a need expressed by these bodies for a unique system of identifiers for strains recommended for use in quality assurance.

Cooperative unit



#### WDCM Reference Strains cover

 Strains 202	 Species 134	 Culture Collections 49	 ISO and other standards 70	 Culture media 98
---	---	--	--	--

- 微生物领域ISO标准、FDA等国际标准中的参考菌株
- 整合了来自70个国际标准的**134个物种**的202个参考菌株信息
- 为每一个参考菌株提供**WDCM唯一编号**

几乎所有的与**微生物检测**相关的**ISO标准**均使用美国**ATCC菌株**，并且大量的ISO标准仅使用一株ATCC菌株。为了解决这种ISO标准的**“孤儿菌现象”**，ISO与世界微生物数据中心（WDCM）开展了正式合作，以确保ISO标准在全球的可持续运行。

我们按照ISO和WDCM联合制定的操作规范，将中国普通微生物菌种保藏中心、食品发酵研究院在的**50株菌**正式列为**食品和饲料、肉及肉制品、乳制品、水质等领域的56个ISO标准的等效用菌**，实现了我国ISO标准菌的**零的突破**，避免了卡脖子。

进入WDCM参考菌株目录后，受到我国产业用户的广泛关注，菌种销售数量稳步提升，2021-2023年期间，面向社会共享**2402支**，销售范围涉及国内**1386余家**单位，其中**企业910家**，涉及食品、制药、日化、健康和环保等多个行业。

**对我国生物产业的高质量可持续发展和国际标准的有效应用，提供了有效支撑**



**毒蘑菇表征及基因组数据库**  
CHARACTERIZATION AND GENOME DATABASE OF POISONOUS MUSHROOMS

BRP M

首页 物种浏览 数据库 物种筛选鉴定平台 技术服务 资源 登录与注册

### 毒蘑菇表征及基因组数据库 / About Database

毒蘑菇类型和基因组数据库覆盖全球的毒蘑菇物种名录。为了弥补我国毒蘑菇数据库空白，加强国内外毒蘑菇研究学术交流，有力推动我国毒蘑菇研究的系统性及前沿性，服务大众需求，构建了毒蘑菇表征与基因组服务平台，数据库主要包括凭证样本库、基因组序列库、表征性状库。

种(Species) e.g Bulgaria inquinans

e.g. Bulgaria inquinans / Gyromitra esculenta / Gyromitra infusa / Halvella pseudoreflexa

物种 43 凭证样本 131 序列 97 图片 44

数据库 Database

凭证样本库

Reference database for voucher specimens

多基因序列库

Reference database for multi-locus sequences

表观性状库

Reference database for morphological characteristics

凭证样本库: Reference database for voucher specimens

收集世界毒蘑菇名录记录的物种，依据分类信息并结合标准及凭证样本的形态鉴定，收集模式标本或凭证样本，经分类学专家逐一检验后，进行统一、标准化管理样本库，整理物种名称、样本编号、分类地位、产地、生境类型等相关信息名称，整理成库。

**植物检疫性菌物数据库**  
Reference Databases for Plant Quarantine Fungi

BRP M

首页 物种浏览 数据库 物种筛选鉴定平台 技术服务 资源 关于

### 植物检疫性菌物数据库

植物检疫性菌物数据库平台搭建了植物检疫性菌物标准库及检测平台，建立了形态、分子检测标准化的参考库(模式、凭证样本实物库及信标库、参比物文库、形态特征库以及多基因基准序列数据库)，同时，通过整合多个信标库资源开发从检测样品到物种精准鉴定的平台模块，实现跨、检一体的“一站式”检测服务。更多>>

种

例如: Colletotrichum kahawae | Colletotrichum asianum | Dimorphoseta oshua | Aflaria putrefacta

物种 749 凭证样本 1,649 序列 4,876 标准分子 1,313 图片组 378

数据库

凭证样本库

GO >

参比物文库

GO >

多基因序列基准数据库

GO >

表观性状库

GO >

凭证样本库

以我国检疫性菌物为主要对象，同时选取口岸的检验检疫数据、国家外来有害生物信息数据、作物保护大全(Crop Protection Compendium)及国内外文献报道的检测对象及高风险植物病原菌物为核心对象，依据分类信息并结合标准及凭证样本的形态鉴定，收集模式标本或凭证样本，经分类学专家逐一检验后，进行统一、标准化管理样本库，整理物种名称、样本编号、寄主、分类地位、产地、生境类型等相关信息名称，整理成库。

# 大型真菌生物多样性数据库

MTSEM Macrofungal taxonomic system and economic mushrooms



Chinese ▾

Home Dataset Expert Team About Contact Us

## Macrofungal taxonomic system and economic mushrooms

Macrofungal taxonomic system and economic mushrooms, a global data repository of macrofungi, aims to provide the latest taxonomic information for each macrofungal genus and aggregates relevant research progress of important species (edible, medicinal, and poisonous).

Genus

eg: Agaricus | Ganoderma | Amanita

Species with Current Names	Strains	Taxa Relevant Literatures	Collection & Herbarium	Country & region of Collection & Herbarium
56,383	15,446	7,494	455	81



Rank	Total	Macrofungi	Edible Fungi	Medicinal Fungi	Poisonous Fungi
Kingdom	1	1	1	1	1
Phylum	2	2	2	2	2
Class	26	10	9	8	5
Order	82	40	29	21	17
Family	286	182	104	86	54
Genus	2,301	1,809	356	202	135
Species	56,383	44,664	1,822	546	574

All Macrofungi

Species 44,664 Strains 13,316

Edible Fungi Medicinal Fungi Poisonous Fungi

Edible Fungi

Species 1,822 Strains 2,716

Medicinal Fungi

Species 546 Strains 3,408

Poisonous Fungi

Species 574 Strains 650

# 冰川微生物组

## Global Glacier Genome and Gene database (4Gdb)

Home Project data Function module Blast Statistics Download About

**4Gdb** — A specialised genomic and genetic database of glacial microbiomes of global glaciers.

Project Search by taxonomy, CAZs, BGCs, VFs, AMRs... Search

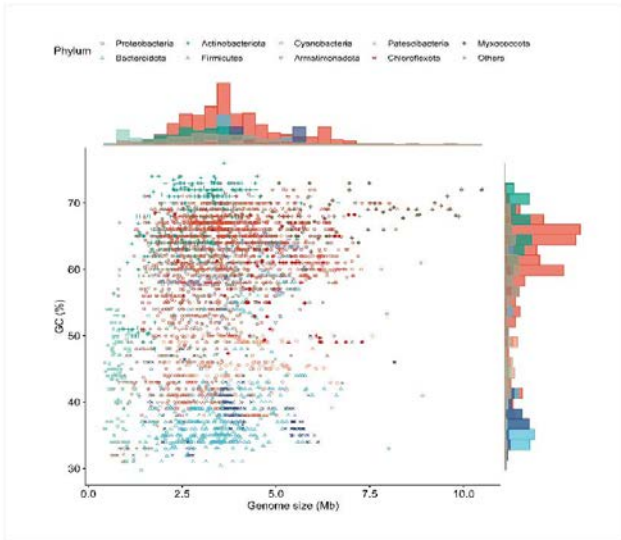
ex: Bacteria | China | Yuzhu

### Project data

Studies: 248	Publications: 23	Samples: 1,213
Isolates: 1,005	Metagenomes: 208	MAGs: 3,375

### Function module

CAZs: 3,269,431	BGCs: 28,034	VFs: 589,029
AMRs: 1,731	Tax Abundance: 477,360	



### News and Updates

**18** 2022.06 4Gdb v1.0 is updated with bug fixed and new features added.

**1** 2022.05 4Gdb v1.0 is publicly accessible.

### Please Cite

- A genome and gene catalogue of glacier microbiomes, Nature Biotechnology

### Global visitor map



# 人肠道微生物组

## hGMB human Gut Microbial Biobank (hGMB)

Home Data Search Download Metabolites Lachnospiraceae

**About the Project**

An open access human Gut Microbial Biobank (hGMB) were constructed by bacterial isolation and cultivation from feces of health volunteers. In total, 1,170 isolates representing 400 different species from 159 genera, 53 families and 6 phyla were included in hGMB and deposited in International Depository Authority (IDA) - the China General Microbiological Culture Collection Center (CGMCC) for long-term preservation and public use worldwide. The strain identities, genomic accessions and basic taxonomic information of known species were available in the webpage. Moreover, by genome sequencing and polyphysically characterization, 102 new species including 29 novel genera and 3 novel families were recognized by the hGMB work, and denominated following ICNP rules. Detailed genomic, taxonomic and biologic descriptions of the 102 new species were also provided in the webpage.

**400**

Number of species

**159**

Number of Genera

**53**

Number of Families

**6**

Number of phyla

Search

Candidatus Wujia chipingensis/NSJ-41 Search

Examples: Candidatus Wujia chipingensis/ NSJ-41 / 4P15T from human feces / Hydrogenisotidium Advanced search

序号	Species Name	Type/Representative strains	Family	Deposit Accessions
1	<i>Wujia hominis</i>	NSJ-40	Tianchiaceae	CGMCC 1.32813
2	<i>Luoshenia tenuis</i>	NSJ-44	Christensenellaceae	CGMCC 1.32817
3	<i>Feltania hominis</i>	BX7	Feltaniaceae	CGMCC 1.32862
4	<i>Bianquea renqueensis</i>	NSJ-32	Bianqueaceae	CGMCC 1.32805
5	<i>Gehongia tenuis</i>	NSJ-53	Christensenellaceae	CGMCC 1.32829

# 鸡肠道微生物组

## ICRGGC Integrated Chicken Reference Genomes and Gene Catalog (ICRGGC)

Home Data Download

**About the Project**

An open access the integrated chicken reference genomes and gene catalog (ICRGGC) were constructed by integrating the metagenomic sequencing data of chicken gut microbiome from four studies in China and European countries. In total, 12,339 metagenomic-assembled genomes on strain level from 1,978 species were generated. According to the genomes, there were 893 novel species and 58 novel genera. In addition, we constructed a gene catalog consist of 10.6 million genes. The draft genomes, gene catalog and its relative abundance were available in this webpage.

**1,978**

Number of Species

**893**

Number of New Species

**12,339**

Number of Strain

**9,845**

Number of New Strain

**16,565,684**

Number of Genes

**68.2%**

the Proportion of ICRGGC Annotated Genes

Search

Genome ID Search

Examples: MAGs\_cg\_1

NO.	Genome ID	Phylum	Class	Order	Family	Genus	Species
1	MAGs_cg_1	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Ligilactobacillus	
2	MAGs_cg_2	Bacteroidota	Bacteroidia	Bacteroidales	Bacteroidaceae	Prevotella	Prevotella sp000540415
3	MAGs_cg_3	Firmicutes A	Clostridia	Lachnospirales	Lachnospiraceae	Eisenbergiella	
4	MAGs_cg_4	Firmicutes A	Clostridia	Papillistrepococ...	Papillistrepococ...	Ramboutasia	
5	MAGs_cg_5	Actinobacteriota	Actinomycetia	Actinomycetales	Dermabacteraceae	Brachyobacterium	Brachyobacterium sp003...
6	MAGs_cg_6	Firmicutes A	Clostridia	Lachnospirales	Lachnospiraceae	952a	
7	MAGs_cg_7	Synergistota	Synergistia	Synergistales	Synergistaceae	An23	An23 sp000544636
8	MAGs_cg_8	Firmicutes A	Clostridia	Lachnospirales	Lachnospiraceae	952a	
9	MAGs_cg_9	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Facklamia A	Facklamia A tabacinaeade
10	MAGs_cg_10	Firmicutes A	Clostridia	Oxallospirales	Oxallospiraceae	Flavovibractor	Flavovibractor sp002161...

# 全局调控因子数据库

PGRR Prokaryotic Global Regulator Resources | WDCM NMDC

Home Data Statistics Download About

## Prokaryotic Global Regulator Resources

The Prokaryotic Global Regulator Resources (PGRR) provides a comprehensive database about 219 global regulators distributed in 14800 prokaryotic model organism genomes for investigators to understand how global regulator use gene regulation networks to control cell growth, community, and all kind of functions in responses to environmental challenges.

Statistics

Global regulator family  
extracted from publications

20



Global regulator  
extracted from publications

219



Homologous global regulator  
based on prediction

100,441



Organism  
predicted by machine learning models

12,063



Binding site motif  
extracted from publications

233



Target gene  
based on prediction

18,066,661



# 噬菌体数据库

GCM 全球噬菌体目录 | WDCM NMDC

全球噬菌体目录

数据统计

噬菌体基因组	67	噬菌体蛋白	1,113	噬菌体基因	11,783
--------	----	-------	-------	-------	--------

噬菌体序列数据库

噬菌体一站式分析流程

# 中国沙门基因组数据库

The Chinese Local Salmonella Genome DataBase version 2 (CLSGDB v2)

CLSGDB v2

The Chinese Local Salmonella Genome DataBase version 2 (CLSGDB v2)

About the database

An open access Chinese local Salmonella genome database version 2 (CLSGDB v2) was constructed by integrating 1982 newly generated Salmonella whole-genome sequencing (WGS) data and 4187 publicly available genomes from 35 Chinese provinces, covering 1959-2022. In total, 1767 high-quality genomes including 144 accessory and 285 resistance genes (RGs) were obtained. Using the CLSGDB v2, we present a comprehensive picture of antimicrobial resistance genes, virulence and mobile elements in Salmonella from China. We also uncovered the geographic distribution of enteric fever, head, typhoid, and enteric fever genes in China, all of them associated to the clinically important antimicrobials including nalidixic acid, fluoroquinolones, and the 3rd-generation cephalosporins. We further observed that economic, climatic, and social factors can drive the rise of AMR. We hope that the CLSGDB v2 and correlation analysis throughout the case for data collection and sharing for the public good, especially WGS, which provides data on the early identification of emerging resistance genes and outbreaks. The Salmonella genome sequences, statistical analysis results and a list of accession number for host genomes are available on this web page.

Overall, this study represents the largest Salmonella genome database from China (CLSGDB v2) and presents the most comprehensive genome analysis of Salmonella to date, with around 8000 high-quality genomes originating from 35 Chinese provinces between 1959 and 2022. This database will assist future genomic surveillance studies and will help inform interventions for foodborne diseases, AMR, food safety, and public health.

# Bacterial Cyclic-di-GMP Database

Home Search Data Tools References About Help

Bacterial Cyclic-di-GMP Database

Search by taxonomy, CAZy's, BGCs, VFs, AMRs...

Search

ex: PA0169 | PA0290 | siaD

## Promoter-gfp reporter library

MORE >

The promoter regions of 41 genes that are (or are predicted to be) involved in c-di-GMP metabolism in *Pseudomonas aeruginosa* PAO1 were cloned into the vector pPROBE-AT<sup>+</sup> respectively, resulting in a library of promoter-gfp transcriptional fusion reporter plasmids, which can monitor the expression of the corresponding genes in *Pseudomonas aeruginosa*.

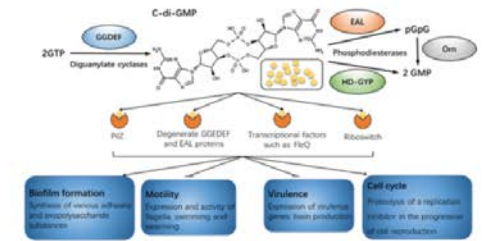
GGDEF only						
NO.	Plasmid	Promoter region	FIG	OD <sub>600</sub>	PA#	Encoded enzyme
1	pPPA0169 <sup>+</sup>		13505.55	1.65	PA0169	DGC
2	pPPA0290		3510.07	1.65	PA0290	DGC <sup>+</sup>
3	pPPA0338 <sup>+</sup>		17886.19	1.61	PA0338	DGC <sup>+</sup>
4	pPPA0847		957.78	1.67	PA0847	DGC
5	pPPA1107		9917.39	1.63	PA1107	DGC

GGDEF-EAL						
NO.	Plasmid	Promoter region	FIG	OD <sub>600</sub>	PA#	Encoded enzyme
21	pPPA0285		832.27	1.62	PA0285	PDE
22	pPPA0575		3821.01	1.67	PA0575	PDE
23	pPPA0861		12296.51	1.66	PA0861	PDE,DGC
24	pPPA1181		4381.87	1.71	PA1181	DGC
25	pPPA1433 <sup>+</sup>		5413.46	1.70	PA1433	EP

## Cyclic-di-GMP

C-di-GMP is synthesized from two GTP molecules by diguanylate cyclases (DGC) that usually harbor a conserved GGDEF domain and is hydrolyzed by c-di-GMP specific phosphodiesterases (PDE) with conserved EAL or HD-GYP domains, which degrade c-di-GMP to pGpG or GMP. Deletion or overexpression of DGC will affect cell motility, psl and biofilm.


MORE >






# 自有知识产权的数据管理软件

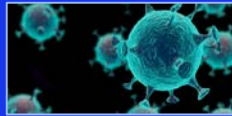
### 微生物类型



细菌




真菌




病毒

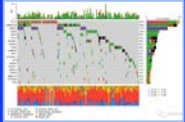
### 数据类型




实物资源



基因组



变异组



图谱

高效识别与实时处理技术

数据质量控制模型

生物信息分析

可视化决策支持

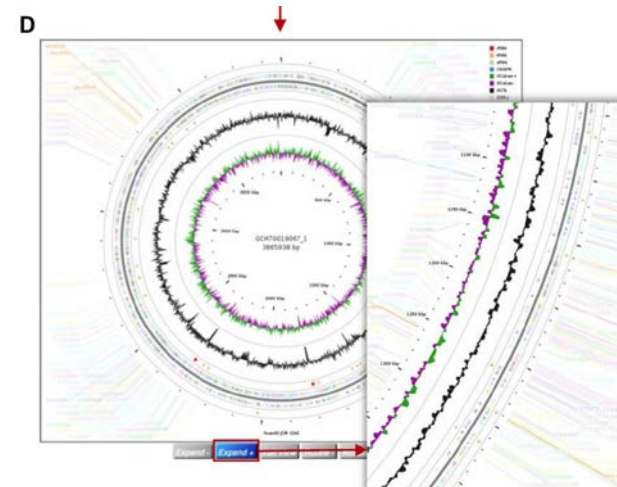
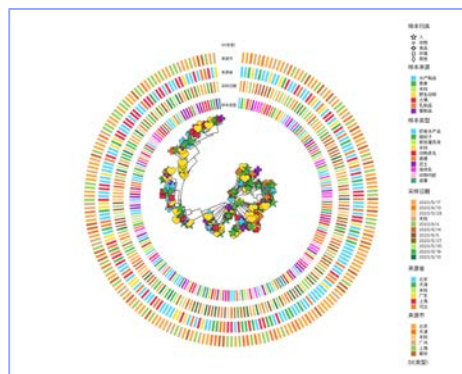
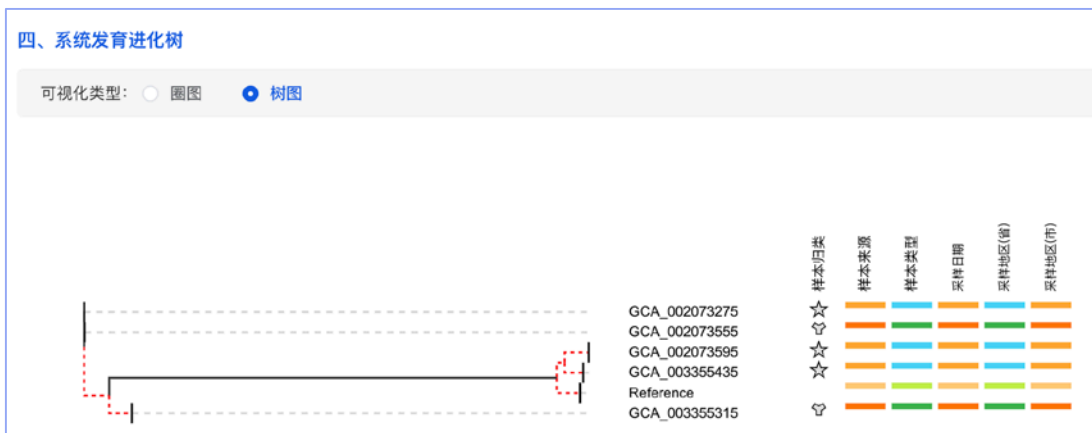
智能化管理与分析软件



Microbench



对标国际商用软件  
自有知识产权  
数据进 报告出



# 请您批评指正

Please criticize and correct

---



中国科学院微生物研究所微生物资源与大数据中心



国家微生物科学数据中心

National Microbiology Data Center