

2022年11月30日
第45回日本分子生物学会年会
フォーラム「生命科学のデータベース活用法」

マイクロバイオームの 国際的なデータハブの構築

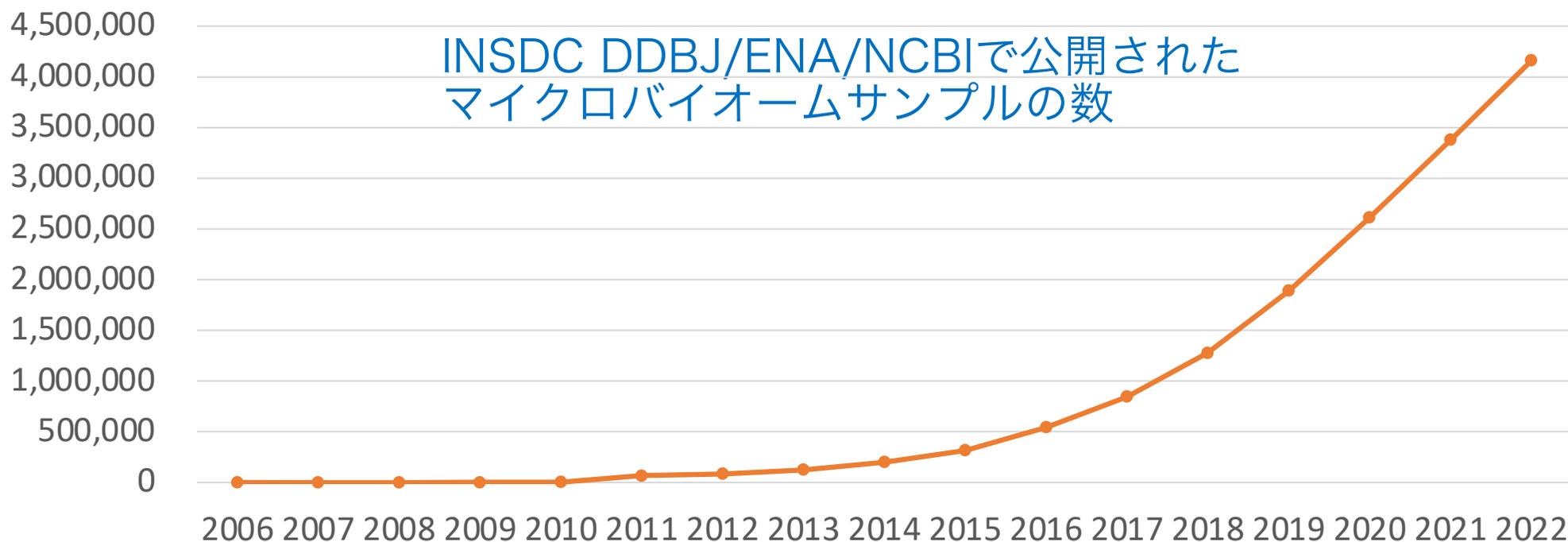
国立遺伝学研究所
情報研究系
森 宙史

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022森宙史(国立遺伝学研究所)



微生物研究の現状

- 微生物、特に原核生物は地球上の様々な場所に存在し環境と深く関係
- 微生物研究はバイオ分野に留まらず、他の多様な分野と連携可能
- 特に、微生物群集（マイクロバイオーム）の研究が急速な勢いで進展している

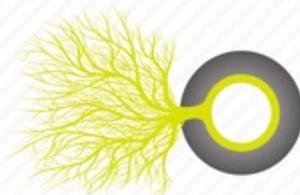




Microbe DB .JP

integrates lots of data related to microbes.

Especially, we integrate the microbial data that can be linked to **genomes**. since 2011



Microbe DB .JP

<https://microbedb.jp/>

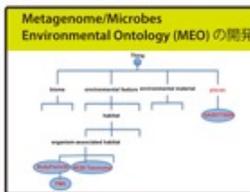
Microbe DB.jp

MicrobeDB.jp プロジェクトでは様々な微生物学上の知識を、ゲノム情報を核として遺伝子、系統、環境の3つの軸に沿ってセマンティックウェブの技術を使用して整理統合し、幅広い分野での微生物学の発展に資することの出来るデータベースの構築を目標としています。

中身は基本的にRDFデータ

Ontology

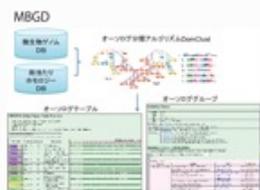
微生物エンサイクロペディアを目指した



Gene

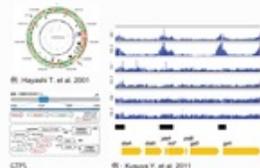
Taxon

Environment



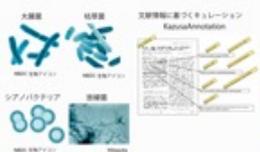
Ortholog: **MBGD**

オースログデータ



Genome: RefSeq

オミックスデータ



モデル微生物の高品質アノテーションデータ



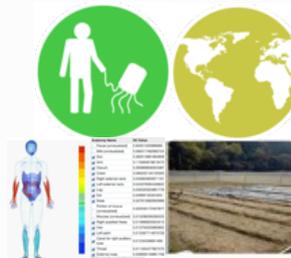
Taxonomy: NCBI Taxonomy

系統分類データ



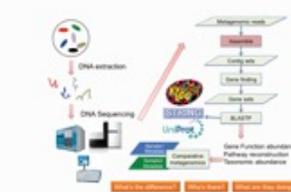
Culture Collection: **NBRC/JCM**

菌株データ



Metadata: **INSDC DRA**

環境のメタデータ



Metagenome: **INSDC DRA**

メタゲノムデータ

Togo picture gallery by DBCLS is licensed under a Creative Commons Attribution 2.1 Japan license (c)



Red color indicates our collaborators.

soil由来マイクロバイームサンプルの絞り込み検索

Home Document Analysis - Search Sign Up Sign in

MicrobeDB.jp

facet_metagenome_public 8531

hasMetagenomeAnalysis

taxonomy 8531

function 119

attribute name

attribute value

hasMEO (Text)

x

hasMEO: Component

Component for environment 8102

hasMEO: Env

Environment for microbes 7117

taxonomy (Text)

taxonomy

root 8531

hasHostTaxonomy (Text)

hasHostTaxonomy

root 1410

pH

0 14

Temperature

-100 150

HMADO (Text)

HMADO

Human microbiome associated 87 disease

HostEthnicity

Metagenomic samples 8531 results found in 155ms

hasMetagenomeAnalysis: taxonomy x hasMEO (Text): soil x Clear all filters

Previous **1** 2 3 4 ... Next

10 Select All Deselect All

Select	MDB SampleID	title	organism.name	organism.identifier	BioProjectID	SRAID	SRRID	BioSampleID	publishedDate
Remove	SAMD00003586	Urease gene-containing Archaea dominate autotrophic ammonia oxidation in two acid soils	soil metagenome	410658	PRJDB1924	DRS001577	DRR002212	SAMD00003586	2012-07-04T00:00:00.000
Remove	SAMD00009749	Active ammonia oxidizers in an acid soil are phylogenetically closely related to neutrophilic Nitrososphaera viennensis	soil metagenome	410658	PRJDB2274	DRS012638	DRR014314	SAMD00009749	2013-10-30T00:00:00.000
Add	SAMEA1559038	RmG_Sample_16S	soil metagenome	410658	PRJEB3363	ERS184934	ERR186224	SAMEA1559038	2012-11-01T00:00:00.000
Remove	SAMEA1559037	RMR_Sample_16S	soil metagenome	410658	PRJEB3363	ERS184936	ERR186222	SAMEA1559037	2012-11-01T00:00:00.000
Add	SAMEA1559036	RmGR_Sample_16S	soil metagenome	410658	PRJEB3363	ERS184935	ERR186225	SAMEA1559036	2012-11-01T00:00:00.000
Add	SAMEA1559035	RRR_Sample_16S	soil metagenome	410658	PRJEB3363	ERS184933	ERR186223	SAMEA1559035	2012-11-01T00:00:00.000
Remove	SAMD00018981	MIMARKS Survey related sample from rhizosphere metagenome	rhizosphere metagenome	939928	PRJDB2986		DRR021946 DRR021947 DRR021948	SAMD00018981	2015-01-16T00:00:00.000
Add	SAMN02054434	MIMARKS Survey related sample from Soil metagenome		410658	PRJNA198445	SRS416341	SRR835396 SRR835397 SRR835398	SAMN02054434	2013-04-23T00:00:00.000
Add	SAMN02054433	MIMARKS Survey related sample from Soil metagenome		410658	PRJNA198445	SRS416342	SRR835399 SRR835400 SRR835401	SAMN02054433	2013-04-23T00:00:00.000
Add	SAMN02054432	MIMARKS Survey related sample from Soil metagenome		410658	PRJNA198445	SRS416340	SRR835402 SRR835403 SRR835404	SAMN02054432	2013-04-23T00:00:00.000

Metagenome sample comparison analysis Compare 4

Taxonomic composition (bar)
Taxonomic composition (heatmap)
Diversity index
Hierarchical clustering
PCoA
Functional composition (bar)
Functional composition (heatmap)

Samples Taxonomic Composition

▲ 1/51 ▼

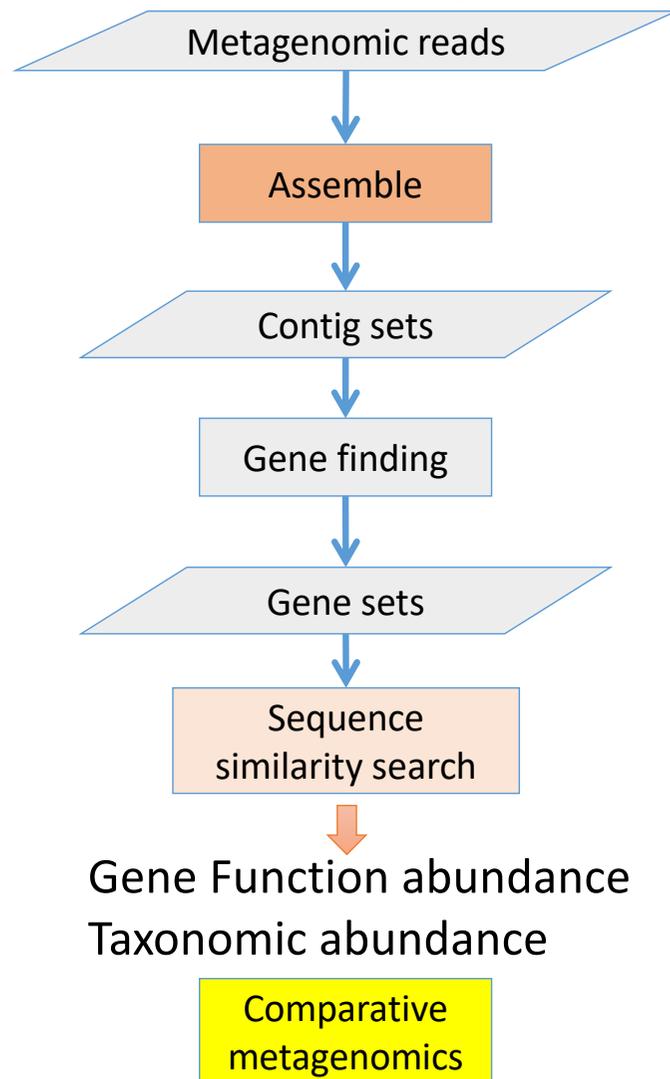
代表的なマイクロバイオームDB間の比較

	Admin	Database URL	Sequence data	Separate amplicon and shotgun?	Taxa	Function	Number of samples in Sept. 2022
NCBI Taxonomy + SRA	NCBI, USA	https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Undef&id=408169	○	×	△	×	>4,150,000
GOLD	JGI, USA	https://gold.jgi.doe.gov/	×	×	×	×	155,000
IMG/M	JGI, USA	https://img.jgi.doe.gov/cgi-bin/m/main.cgi	○	×	○	○	29,961
MG-RAST	Chicago U. USA	https://www.mg-rast.org/	○	○	○	○	80,144
MGNify	EBI, EU	https://www.ebi.ac.uk/metagenomics/	×	○	○	○	422,983
MicrobeDB.jp v.3	NIG, Japan	https://microbedb.jp/	×	○	○	○	1,533,607 (66,604)

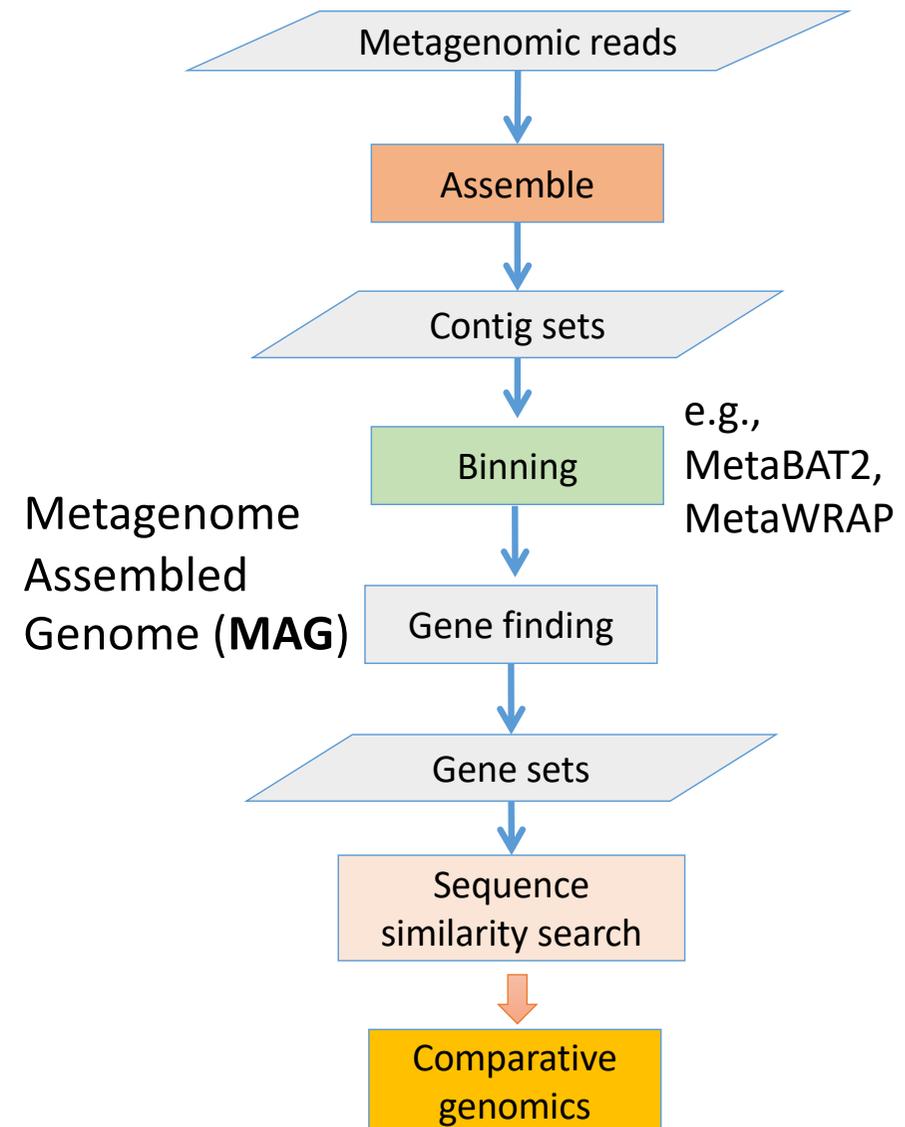
Metagenome Assembled Genome (MAG)

Metagenome analysis strategy

Assembly approach



Assembly + **Binning** approach



メタゲノムの全体像を議論するのではなく、優占系統のゲノム配列を抽出して各ゲノムが持つ機能について議論する

培養を経ずにメタゲノム配列データから 機能や進化的な類縁関係が推定された系統が多数存在

In Sept. 2022

Taxonomic rank	Cultured and validly described taxa	Genome-based putative taxa including uncultured taxa
Phylum	42	166
Class	147	477
Order	276	1,571
Family	656	4,070
Genus	3,865	16,686
Species	22,601	65,703

List of prokaryotic names with standing in nomenclature

(<https://lpsn.dsmz.de/text/numbers>)

Genome Taxonomy Database (<https://gtdb.ecogenomic.org/stats>)

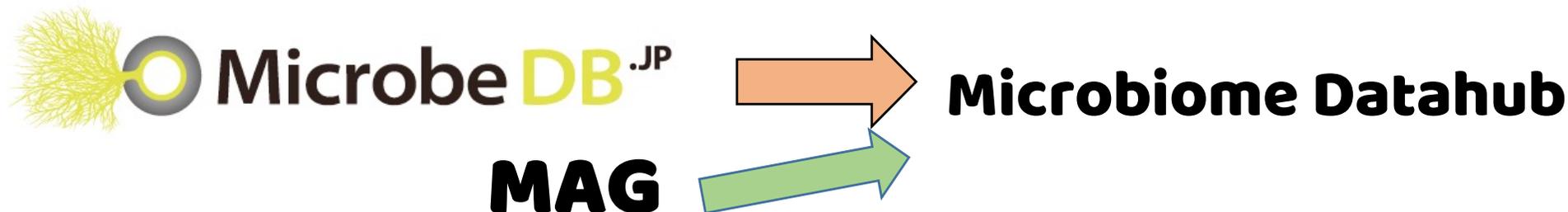
本研究開発の構想・目標

MicrobeDB.jpを

マイクロバイオームに特化した統合DBに作り変える

- ・ 爆発的なデータの増加に対応しいち早くデータを収録
- ・ マイクロバイオームデータに特化した検索機能やUI
- ・ 様々な解析ワークフローに対応したデータ解析機能の実装

爆発的な勢いで増加するマイクロバイオームデータをいち早く収録し、検索・解析可能な統合DBとして、MicrobeDB.jpをマイクロバイオーム研究の国際的な**データハブ**へ発展させる。



Microbiome Datahub

MicrobeDB.jpの重要なデータや機能は引き継ぎつつ、
<https://mdatahub.org> で新DBを構築して今年度末公開予定

研究開発実施体制

国立遺伝学研究所

森宙史：プロジェクト全体の研究統括

中村保一：DDBJ & 他の統合DBとの連携

藤澤貴智：効率的な逐次DB更新・運用システムの”再”構築

東光一、谷澤靖洋：MAGデータの統合・キラアアプリケーションの開発

来年度雇用予定研究員：マイクロバイオームデータに関する開発

基礎生物学研究所

内山郁夫：オーソログデータのマイクロバイオームデータとの統合

千葉啓和：オーソログ情報のRDF化と解析手法の効率化（DBCLS）

西出浩世：オーソログデータの開発補助

東京工業大学

山田拓司：ヒトマイクロバイオームデータのキラデータセットの開発

中川善一：ヒトマイクロバイオームデータのメタデータ整備

東京大学

松井求：微生物PhenotypeデータのRDF化とキラアプリケーション開発

鈴木誉保、西村祐貴、雇用予定研究員：微生物Phenotypeデータの構築

Microbiome Datahubの 主な研究開発項目

1. ゲノム・メタゲノム(MAG)の統合化
2. メタゲノムとリソース（菌株）との連結
3. データ爆発への対応
4. 効率的な逐次DB更新・運用システムの"再"構築
5. キラーデータセット・アプリケーションの開発
6. マルチオミックス連携

ゲノムとMAGの統合化、MAGのクオリティチェック、
メタゲノムと菌株の連結等、
MAGを中心にマイクロバイームデータを整理したDBとして
Microbiome Datahubの開発を進めております

Keyword

Search

Environment

soil	2500
marine	2000
freshwater	3232
hot spring	250
sediment	500
air	100
gut	5000
oral	200
skin	100
reproductive system	50
human activity related	520

Host taxon

Host disease

Host location

2202 Hits

Project ID	Environment	Host taxon	BioSamples	Data size (GB)	Name	...
PRJNA826698	gut	Homo sapiens	120	800	Reference human gut microbiome project	
PRJNA826700	gut	Mus musculus	20	55	Wild mouse gut microbiome project	

Project単位での検索のイメージ

Keyword

Search

Genome taxon

2400 Hits

Genome category

Isolate complete 6000
 Isolate draft 12000
 MAG high quality 30000
 MAG low quality 92000

MAG source

INSDC 100000
 MGnify 30000
 Other DBs 250000
 Original 2000

MAG completeness



Environment

soil 3500
 marine 4000

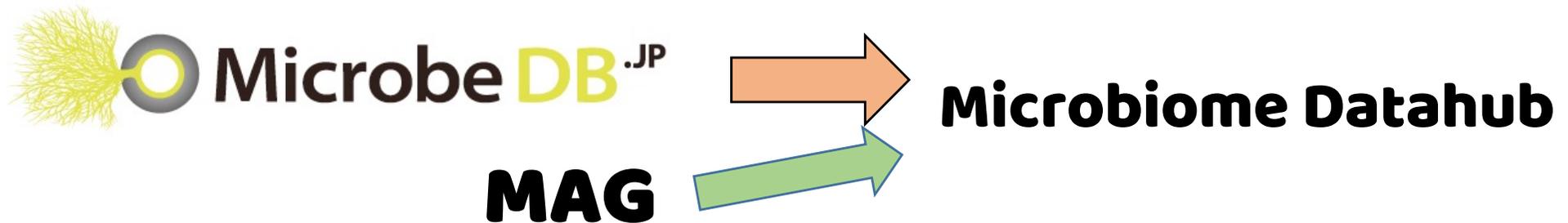
Genome ID	Environment	Host taxon	Project ID	Genome category	Project Name	...
PRJNA826698#bin1	gut	Homo sapiens	PRJNA826698	MAG high quality	Reference human gut microbiome project	
PRJNA826700#bin100	gut	Mus musculus	PRJNA826700	MAG low quality	Wild mouse gut microbiome project	

MAG単位での検索のイメージ

まとめ

MicrobeDB.jpを
マイクロバイオームに特化した統合DBに作り変える

爆発的な勢いで増加するマイクロバイオームデータをいち早く収録し、検索・解析可能な統合DBとして、MicrobeDB.jpをマイクロバイオーム研究の国際的なデータハブへ発展させる。



<https://microbedb.jp> → <https://mdatahub.org>