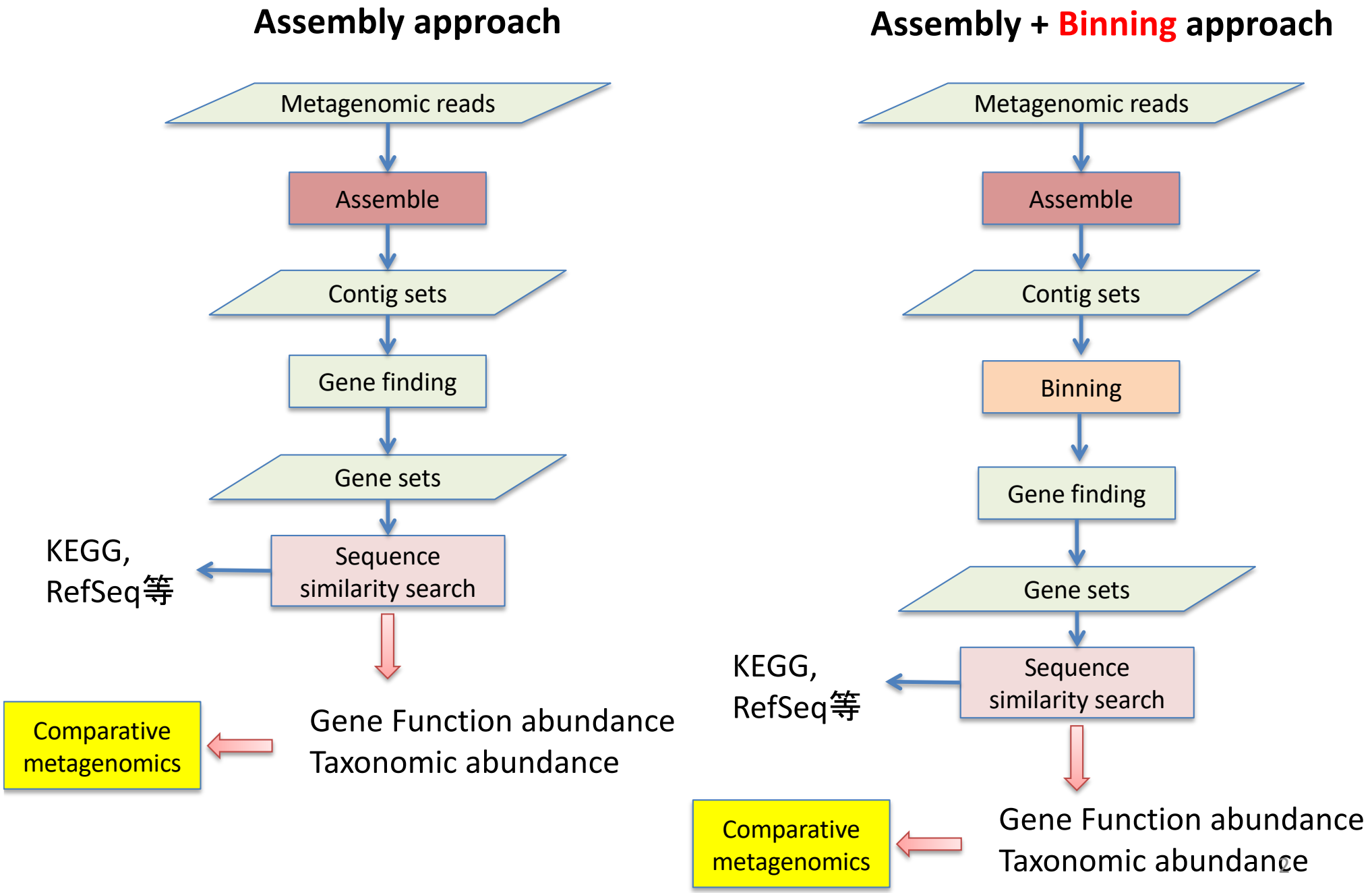
AJACS MAG (Metagenome-Assembled Genome) を
知って・学んで・使う」

2024年7月25日

メタゲノム・MAGデータを検索する

森　宙史, Ph.D.
(Hiroshi Mori)
国立遺伝学研究所
先端ゲノミクス推進センター
hmori@nig.ac.jp

遺伝子組成解析とMAG解析



遺伝子機能組成解析

利点

・群集全体でどの系統由来のどのような遺伝子機能が多いかがわかる
・群間での統計的仮設検定による比較解析を行いやすい
・同じ種類の機能遺伝子を持つ系統の多様性を解析可能

欠点

・存在量の多い遺伝子の大半は必須遺伝子等全系統が持つ機能遺伝子
・マイナー系統が担う機能については統計的に有意な差は出にくい
・数百万遺伝子の解析なので情報解析が複雑になりがち

MAG解析

利点

・単一種由来のドラフトゲノム配列を対象にするので群集中での頻度情報を考慮する必要が無い
・近縁種の既存のゲノム・MAGデータと比較ゲノム解析が可能
・数千遺伝子の解析なので情報解析が比較的単純

欠点

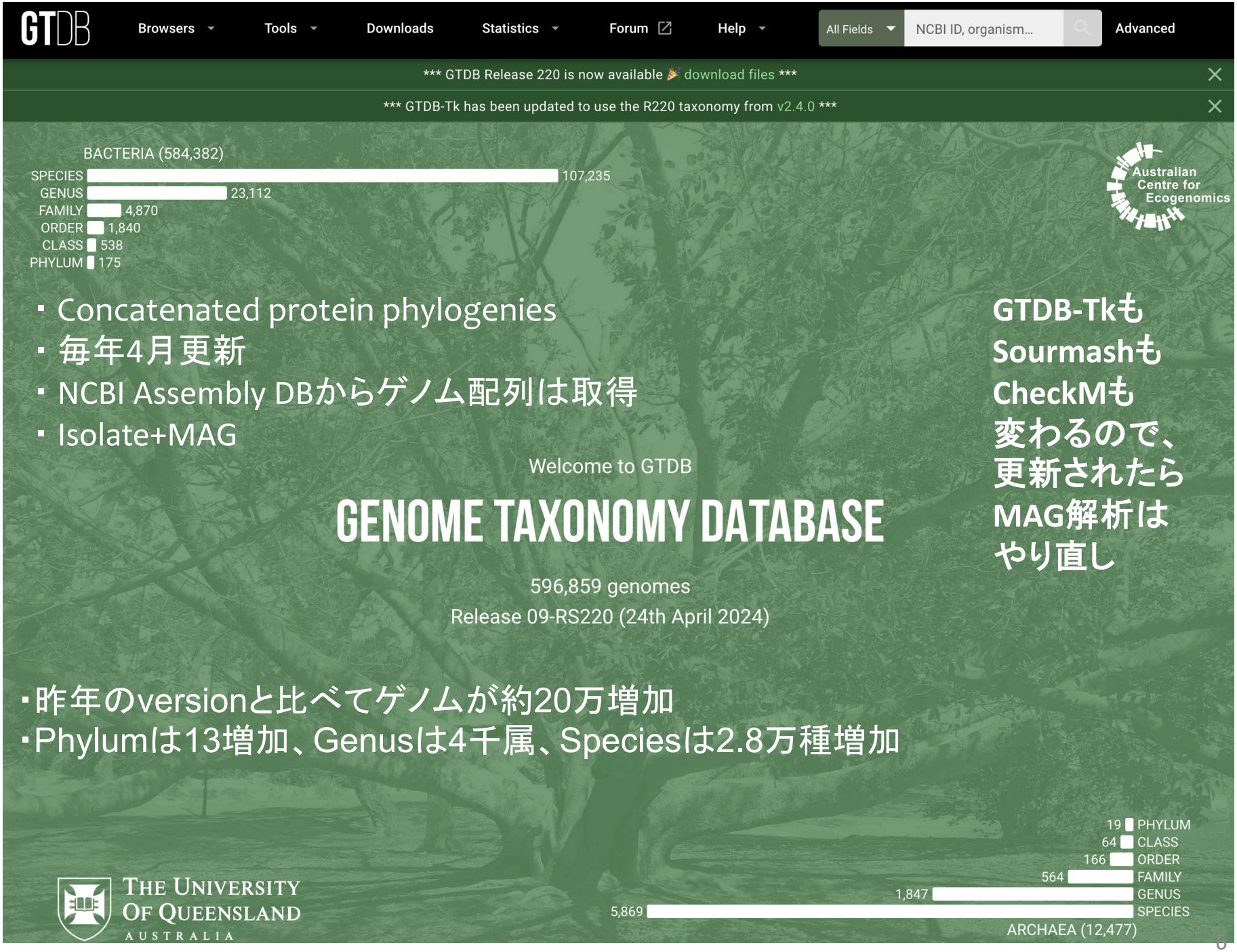
・rRNA遺伝子や水平伝播アイランド等MAGに含まれにくい遺伝子の存在
・キメラやコンタミの問題

MAG関連のデータベース(DB)

https://gtdb.ecogenomic.org/ 2023

- Concatenated protein phylogenies
- 毎年4月更新
- NCBI Assembly DBからゲノム配列は取得
- Isolate+MAG

https://gtdb.ecogenomic.org/ 2024



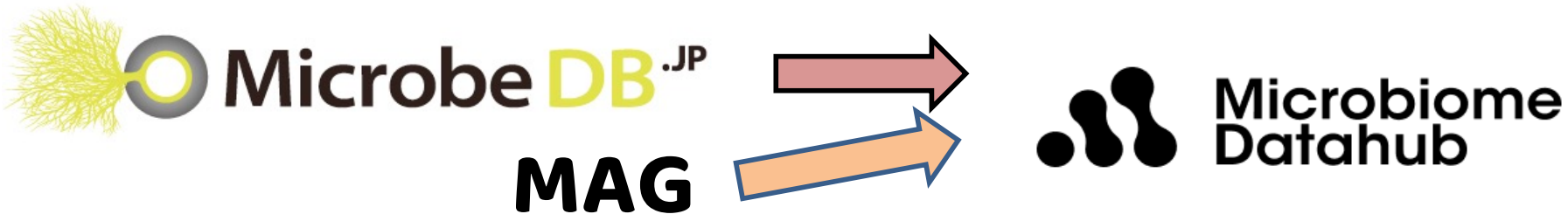
代表的なMAG DB間の比較

	Admin	Database URL	DNA sequence data	Protein sequence data	Number of MAGs in July 2024
NCBI Datasets	NCBI, USA	https://www.ncbi.nlm.nih.gov/datasets/genome/	○	○	443,763
IMG/M	JGI, USA	https://img.jgi.doe.gov/cgi-bin/m/main.cgi	○	○	25,507
SPIRE	EMBL, EU	http://spire.embl.de/	○	○	1,160,000
MGnify	EBI, EU	https://www.ebi.ac.uk/metagenomics	○	○	478,810
Microbiome Datahub	NIG, Japan	https://mdatahub.org/	○	○	218,653

データベースによってMAGのクオリティはバラバラ
残りの時間で、各DBの特徴と使い方について解説する

JST-NBDC 統合化推進プログラム

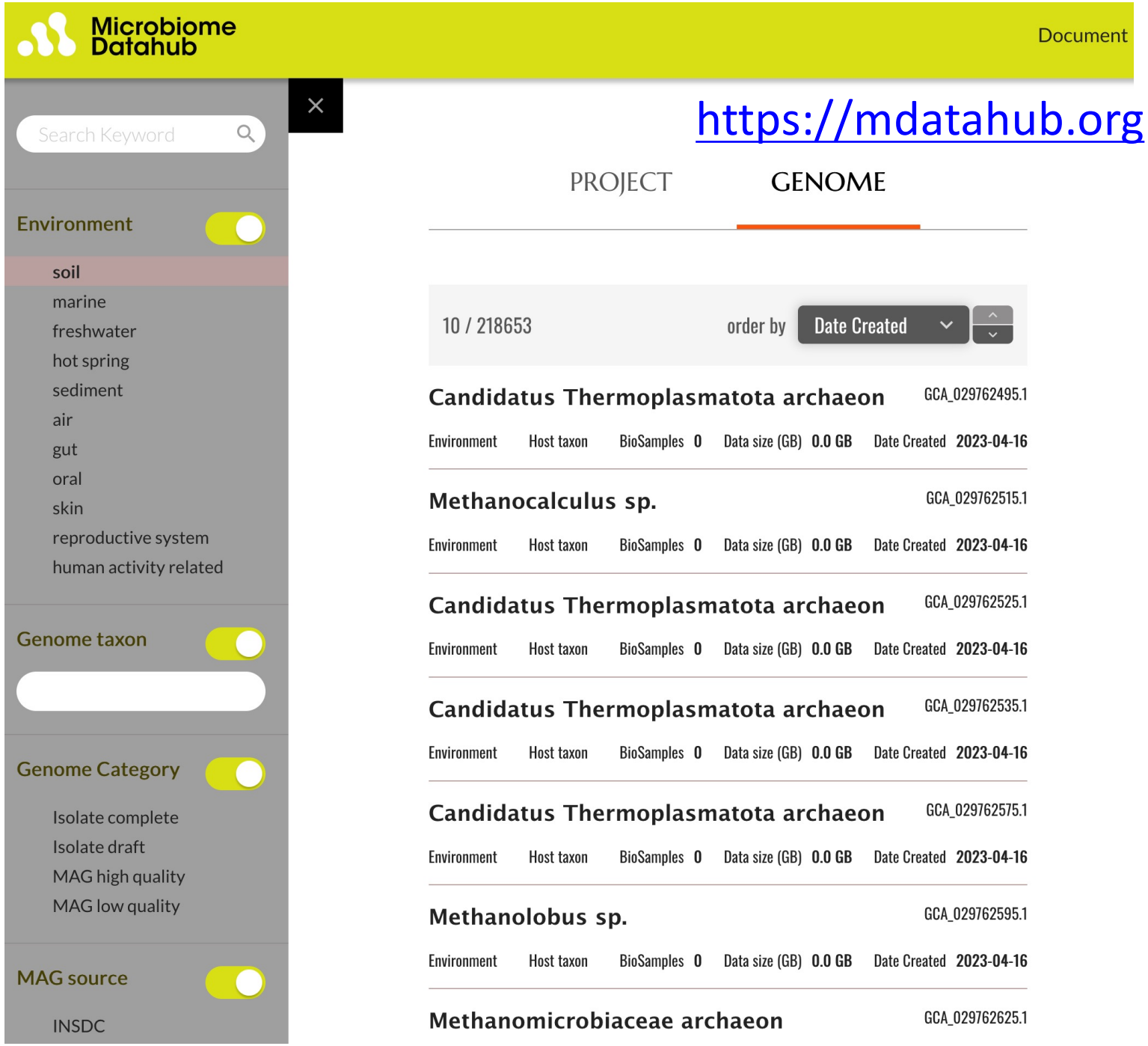
爆発的な勢いで増加するマイクロバイオームデータをいち早く収録し、検索・解析可能な統合DBとして、マイクロバイオーム研究の国際的なデータハブを構築し発展させることを目標とする。



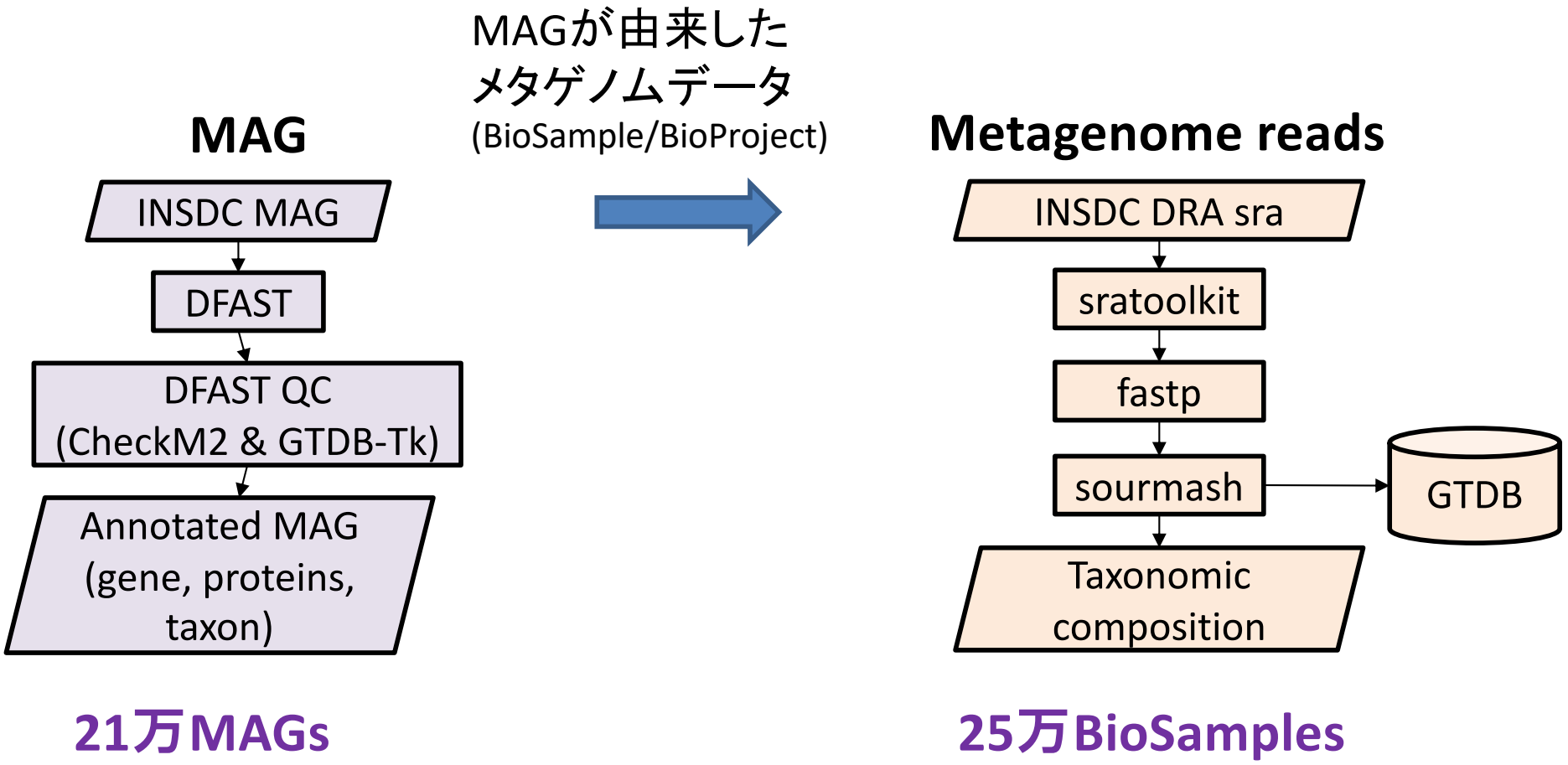
https://microbedb.jp → **https://mdatahub.org**

単離菌ゲノムとMAGを基盤とした
マイクロバイオームの統合データベース

Microbiome Datahubの今のUI



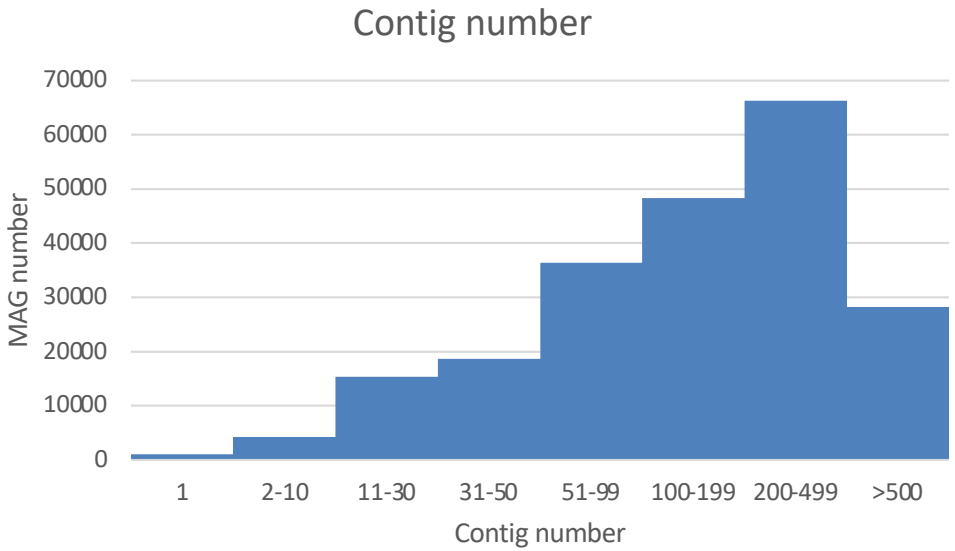
Microbiome DatahubにおけるMAGの収集とアノテーション

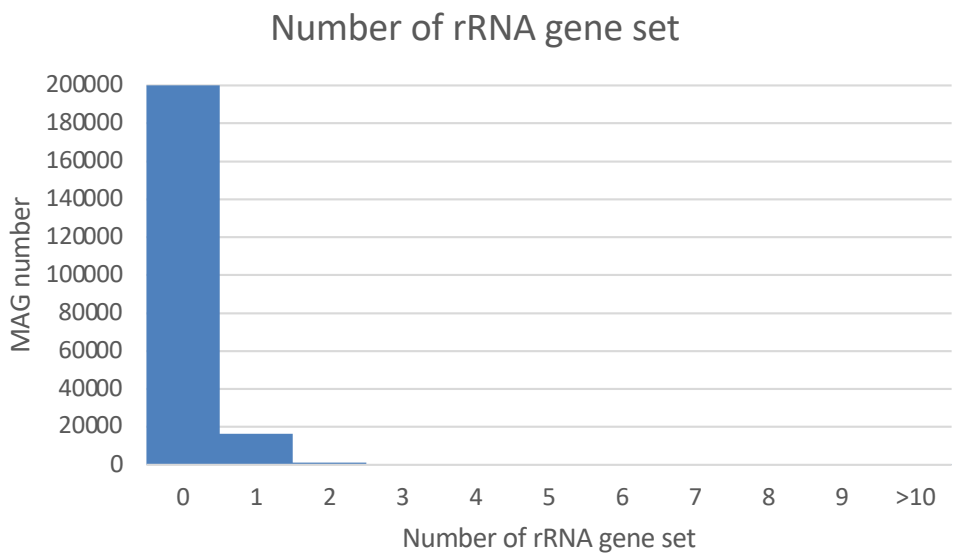


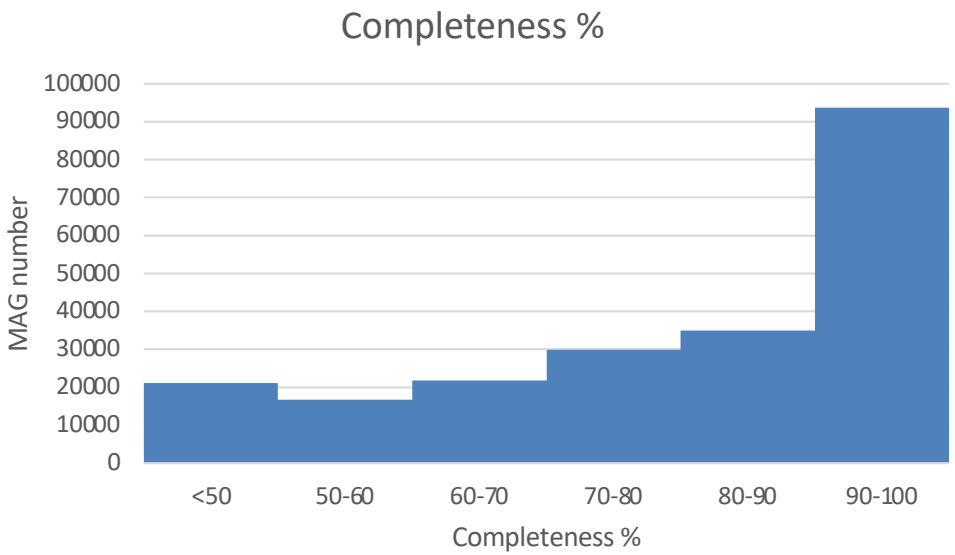
21万MAGs

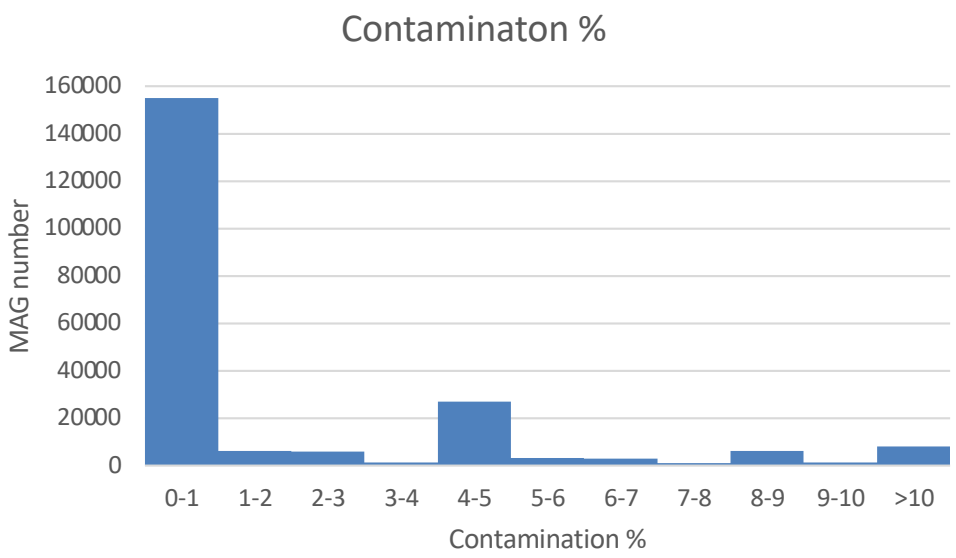
25万BioSamples

Microbiome Datahubにおける21万MAGの内訳

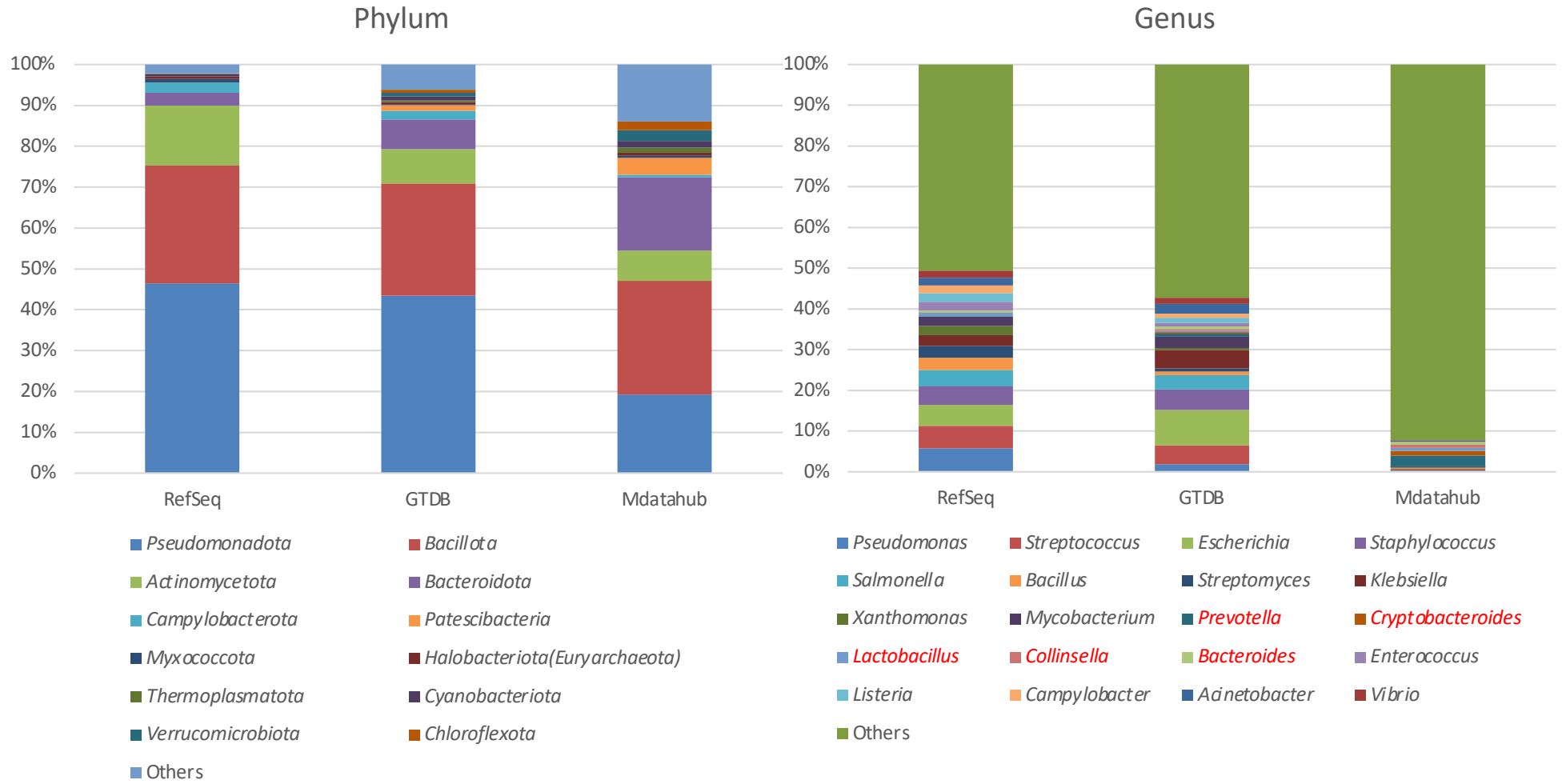








Microbiome Datahubにおける21万MAGの内訳と比較



RefSeq単離菌31万ゲノムデータの内訳

Escherichia: 3.5万、*Klebsiella*: 2万、*Staphylococcus*: 2万、
Streptococcus: 1.8万、*Pseudomonas*: 1.5万、*Salmonella*: 1.2万、
Acinetobacter: 1万　RefSeqは病原菌が全体の半数以上を占める

MGnify https://www.ebi.ac.uk/metagenomics/

Search by

Text search

Name, biome, or keyword

Sequence search

Sequence search

Or by data type

Analysis types

480962	amplicon
57629	assemblies
2050	metabarcoding
39920	metagenomes
2581	metatranscriptomics
2	long reads assemblies

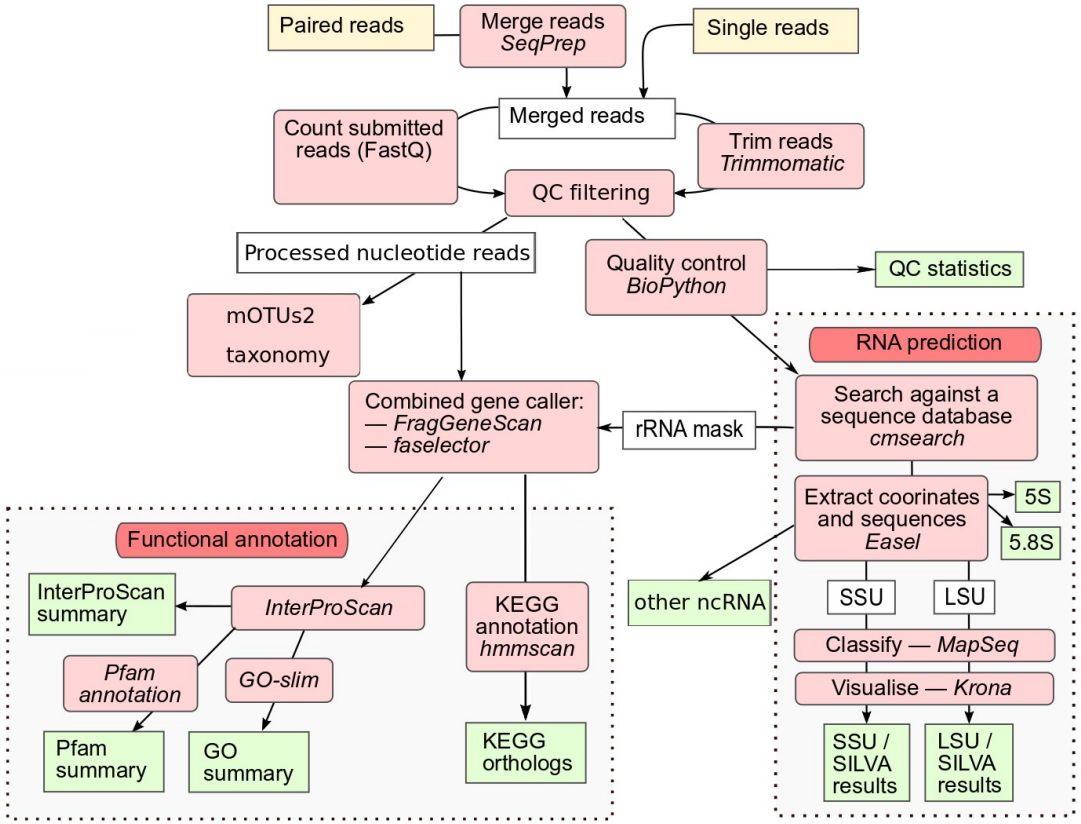
Public data

5004	studies
597736	analyses
478810	genomes in 11 MAG catalogues

Or by selected biomes

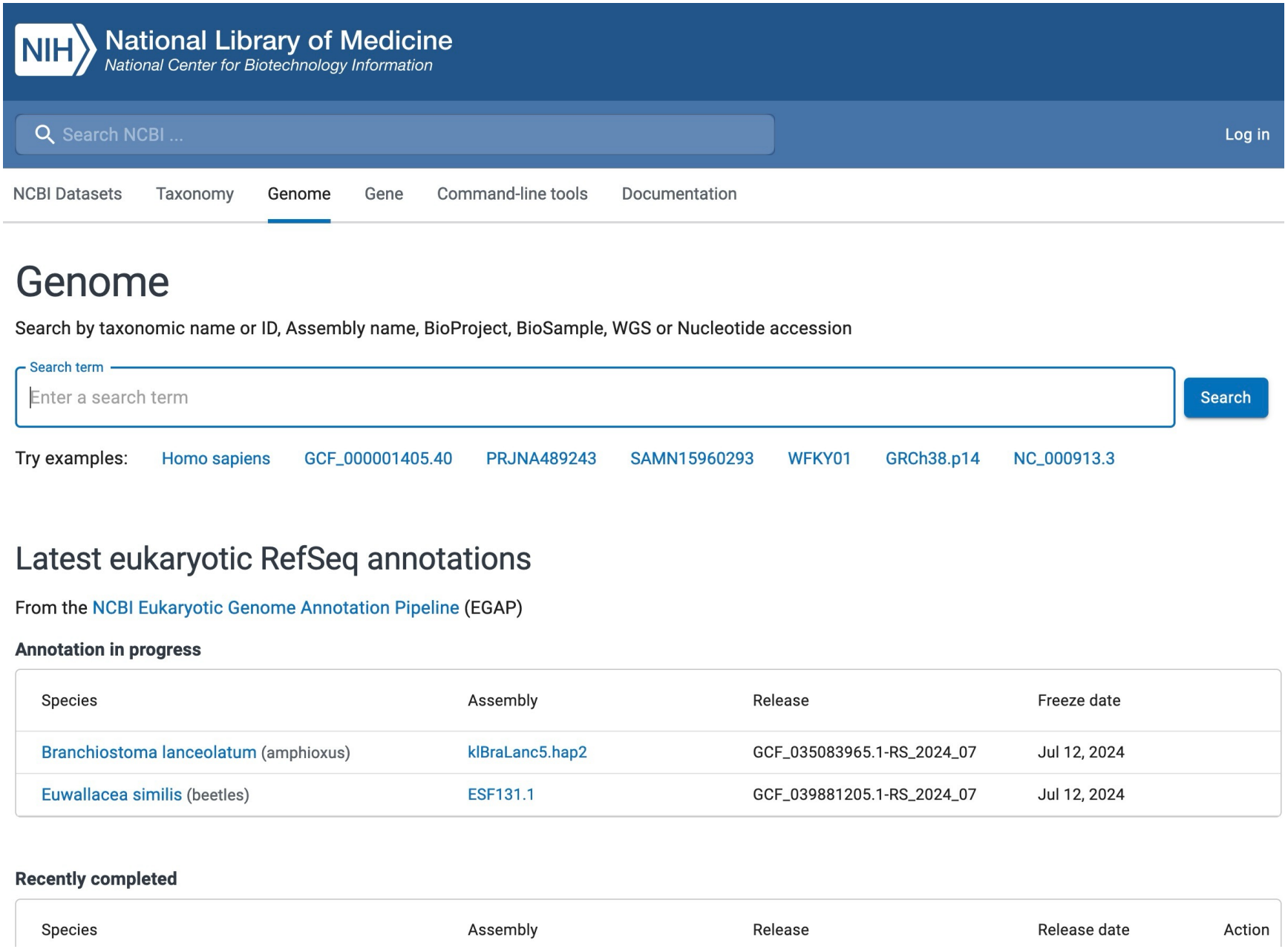
Human (213874)
Digestive system (111024)
Aquatic (51540)
Marine (38036)
Digestive system (35543)
Plants (28859)
Soil (25893)
Skin (11527)
Wastewater (4358)
Food production (2923)

View all biomes



https://emg-docs.readthedocs.io/en/latest/analysis.html

NCBI Datasets https://www.ncbi.nlm.nih.gov/datasets/genome/



IMG/M https://img.jgi.doe.gov/cgi-bin/m/main.cgi

JGI IMG/M INTEGRATED MICROBIAL GENOMES & MICROBIOMES

Quick Genome Search: Go

Login into IMG/MER

My Analysis Carts: 0 Genomes | 0 Scaffolds | 0 Functions | 0 Genes | 0 Genome Search History | 0 Gene Search History | 0 Scaffold Search History | 0 Bin S

Home | IMG/M | Find Genomes | Find Genes | Find Functions | Compare Genomes | OMICS | My IMG | Collaborations | Help

Home

IMG Survey

IMG Content

	Datasets	JGI	All
Bacteria		21131	137535
Archaea		786	2968
Eukarya		370	591
Viruses		13	19529
Metagenome		19982	34779
Cell Enrichments		2799	2849
Single Particle Sorts		7368	7758
Metatranscriptome		6219	8722
Combined Assembly		275	694
Total Datasets			217387

Last Datasets Added On:

Genome	2024-06-13
Metagenome	2024-07-14

Project Maps: Genomes Metagenomes

IMG Webinar YouTube Playlist

Download Isolate Genomes
Download Metagenome BINs

The **Integrated Microbial Genomes (IMG)** system serves as a community resource for analysis and annotation of genome and metagenome datasets in a comprehensive comparative context. The **IMG data warehouse** integrates genome and metagenome datasets provided by IMG users with a comprehensive set of publicly available genome and metagenome datasets.

IMG provides users with tools for analyzing publicly available genome datasets and metagenome datasets (Nucleic Acids Research, January 2019).

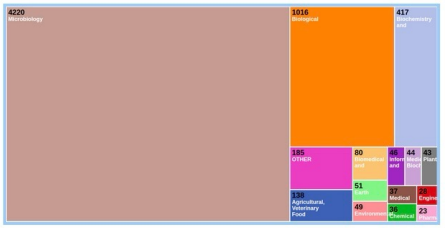
If you use IMG web resources or data to assist in research publications or proposals. Please cite: IMG - Chen et al., 2022 (Nucleic Acids Research, gkac976).
GOLD v9 - Mukherjee et al., 2022 (Nucleic Acids Research, gkac974).

IMG Statistics | Data Usage Policy

View Citations

	Isolates		SAGs		MAGs	
Sequenced at:	JGI	All	JGI	All	JGI	All
Bacteria	13458	113595	2282	9447	5357	14457
Archaea	220	1365	318	601	248	1002
Eukarya	370	591	0	0	0	0
Viruses	13	16510	0	75	0	2918

(Only data sets with GOLD metadata were counted.)



Combined assembly data sets were excluded in the following metagenome and metatranscriptome table statistics.

Metagenome | Metatranscriptome

Engineered	JGI	ALL	Environmental	JGI	ALL	Host-associated	JGI	ALL
Artificial ecosystem	173	203	Air	60	116	Algae	53	168
Bioreactor	902	1074	Aquatic	16111	20696	Annelida	138	152
Bioremediation	25	61	Terrestrial	9184	12514	Arthropoda	0	1
Biotransformation	3	8				Arthropoda: Crustaceans	0	7
Built environment	297	1428				Arthropoda: Insects	123	324
Food production	0	61				Birds	17	48
Industrial production	20	43				Cephalochordata	0	4
Lab culture	4	4				Fish	0	29
Lab enrichment	5	69				Fungi	145	146
Modeled	11	91				Human	2	1001

SPIRE http://spire.embl.de/