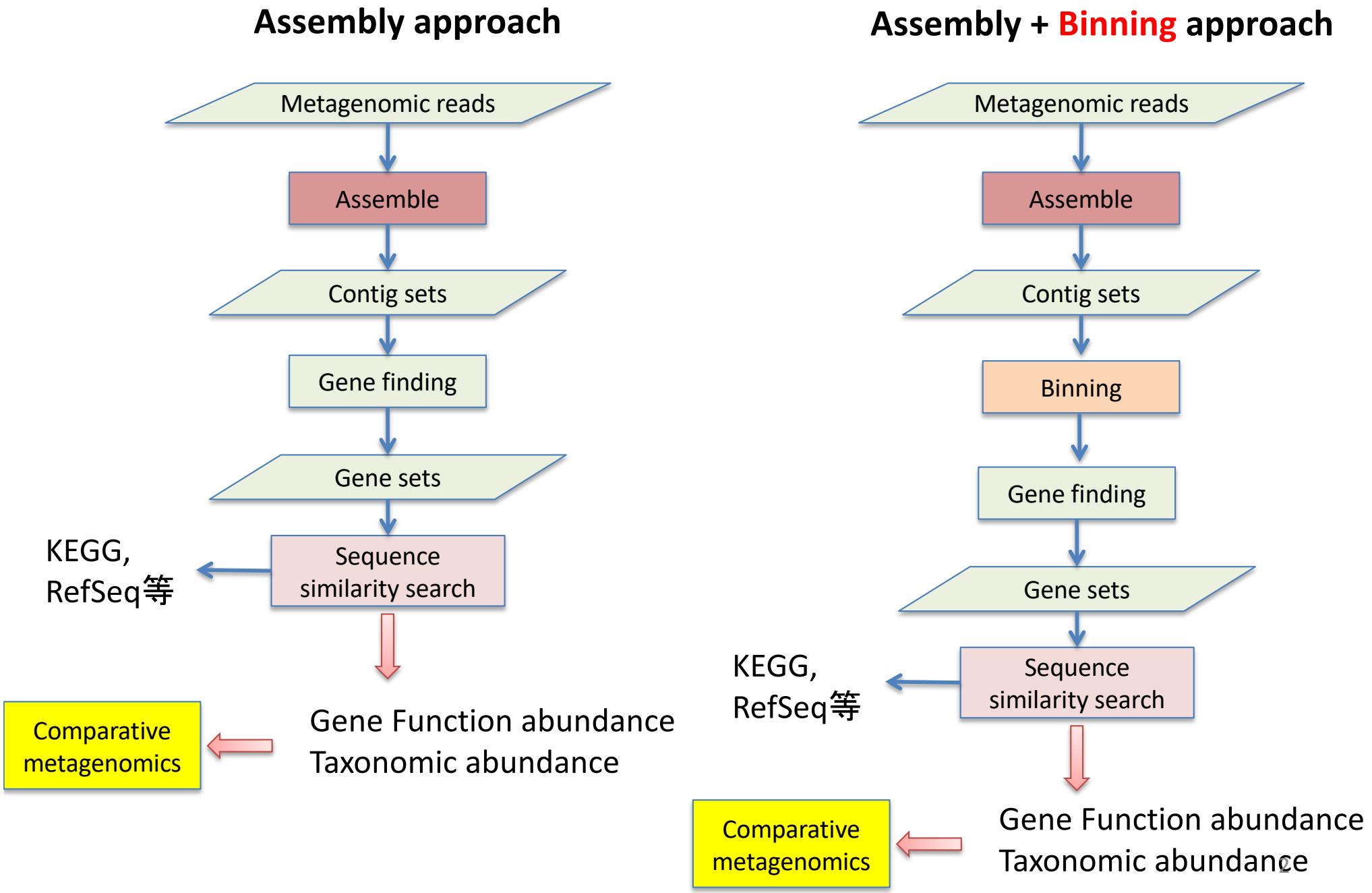
AJACS MAG (Metagenome-Assembled Genome) を
知って・学んで・使う」

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

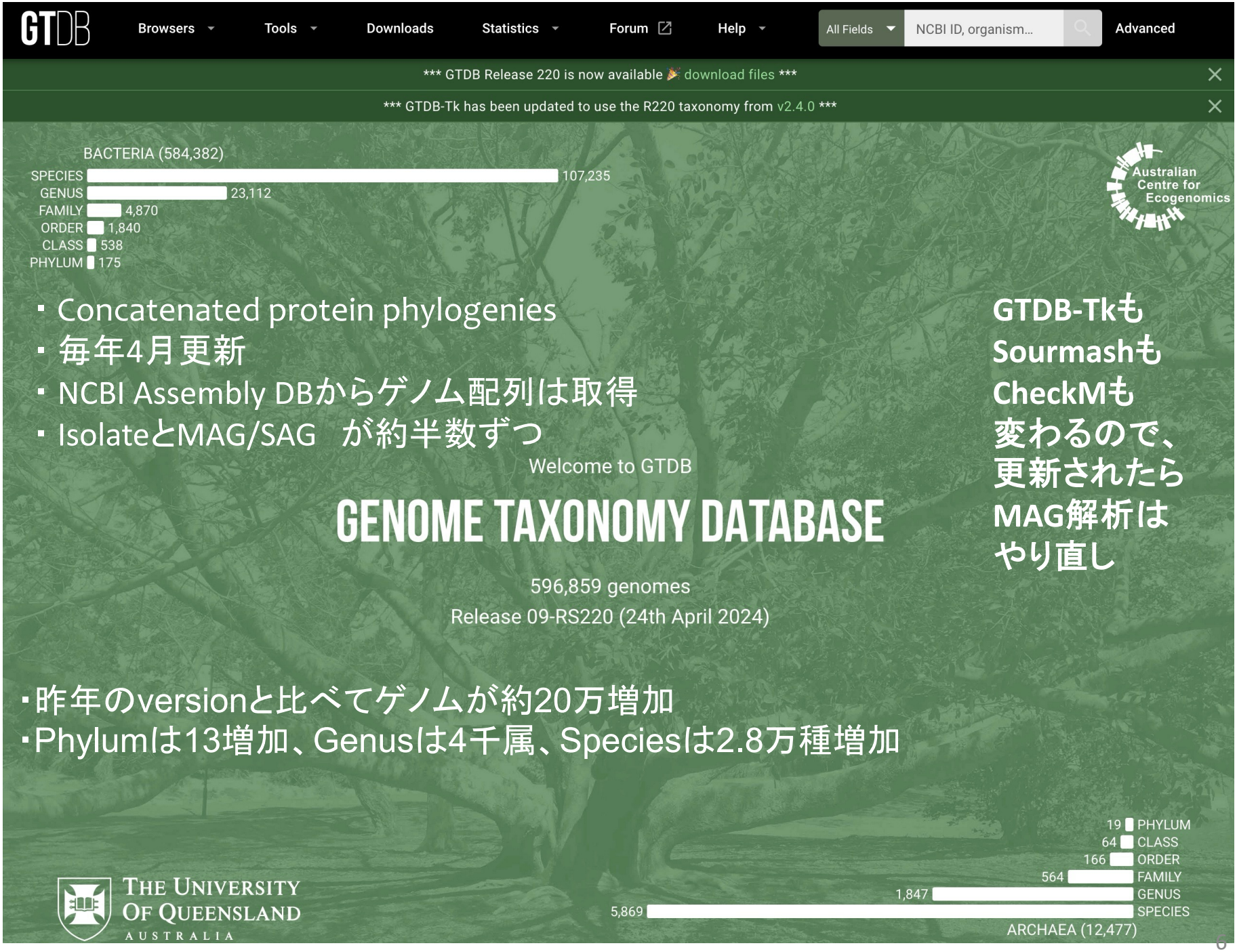
MAG関連のデータベース(DB)

様々なMAGをメタデータで検索し、配列データを取得する

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



GTDB-Tkも
Sourmashも
CheckMも
変わるので、
更新されたら
MAG解析は
やり直し

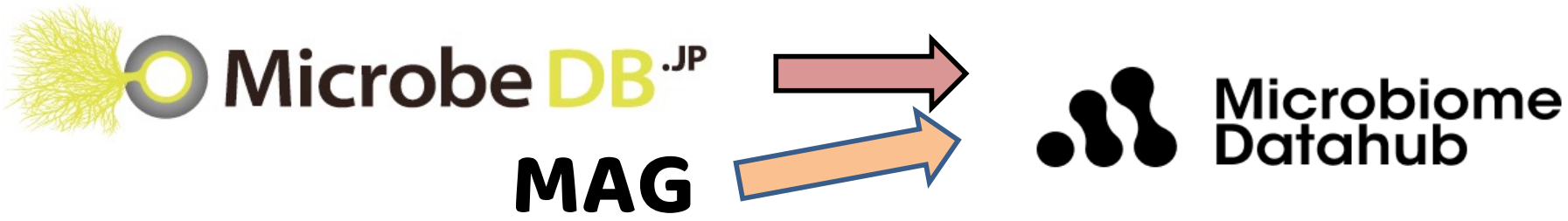
代表的なMAG DB間の比較

	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

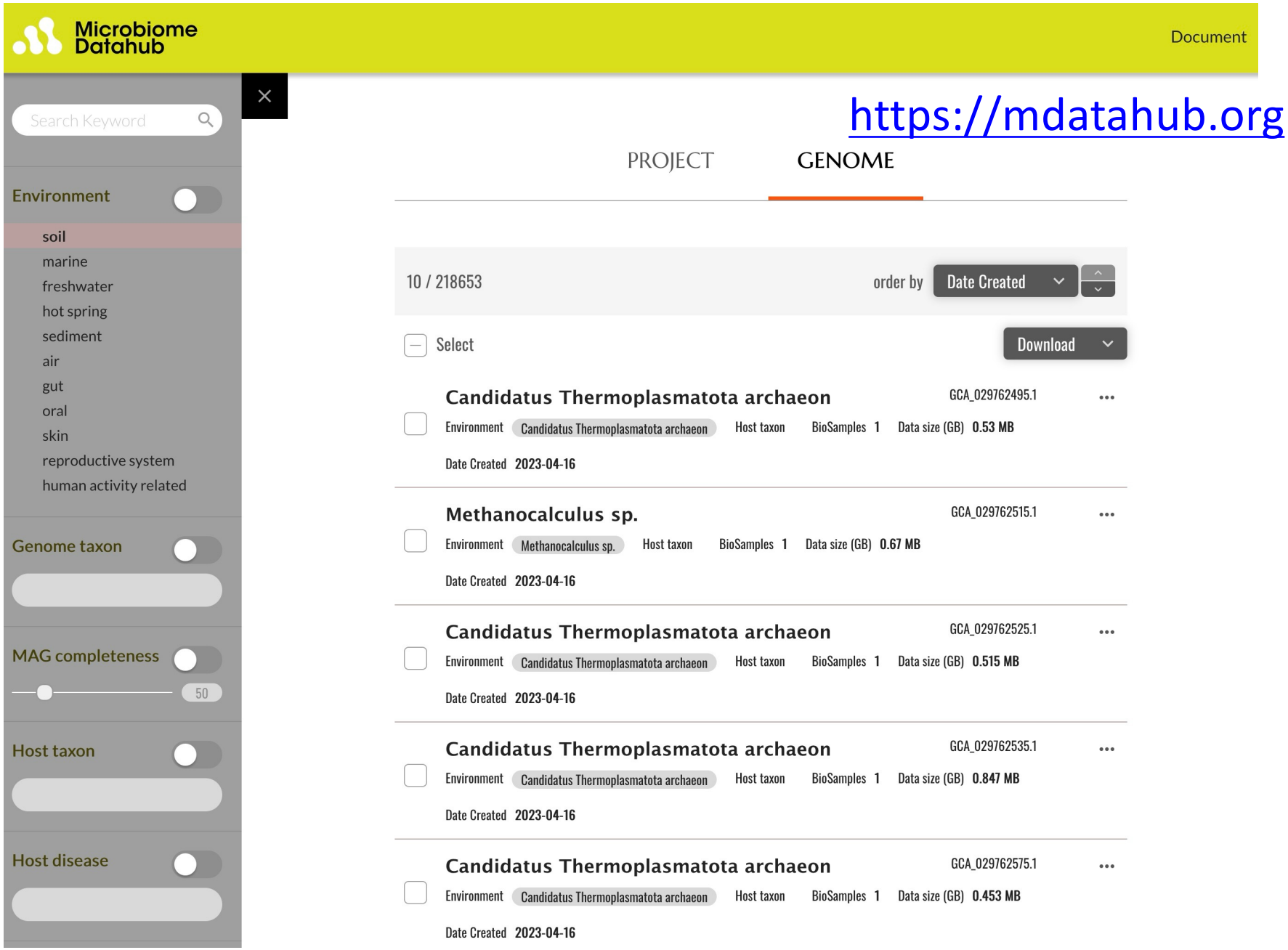
爆発的な勢いで増加するマイクロバイオームデータをいち早く収録し、検索・解析可能な統合DBとして、マイクロバイオーム研究の国際的なデータハブを構築し発展させることを目標とする。



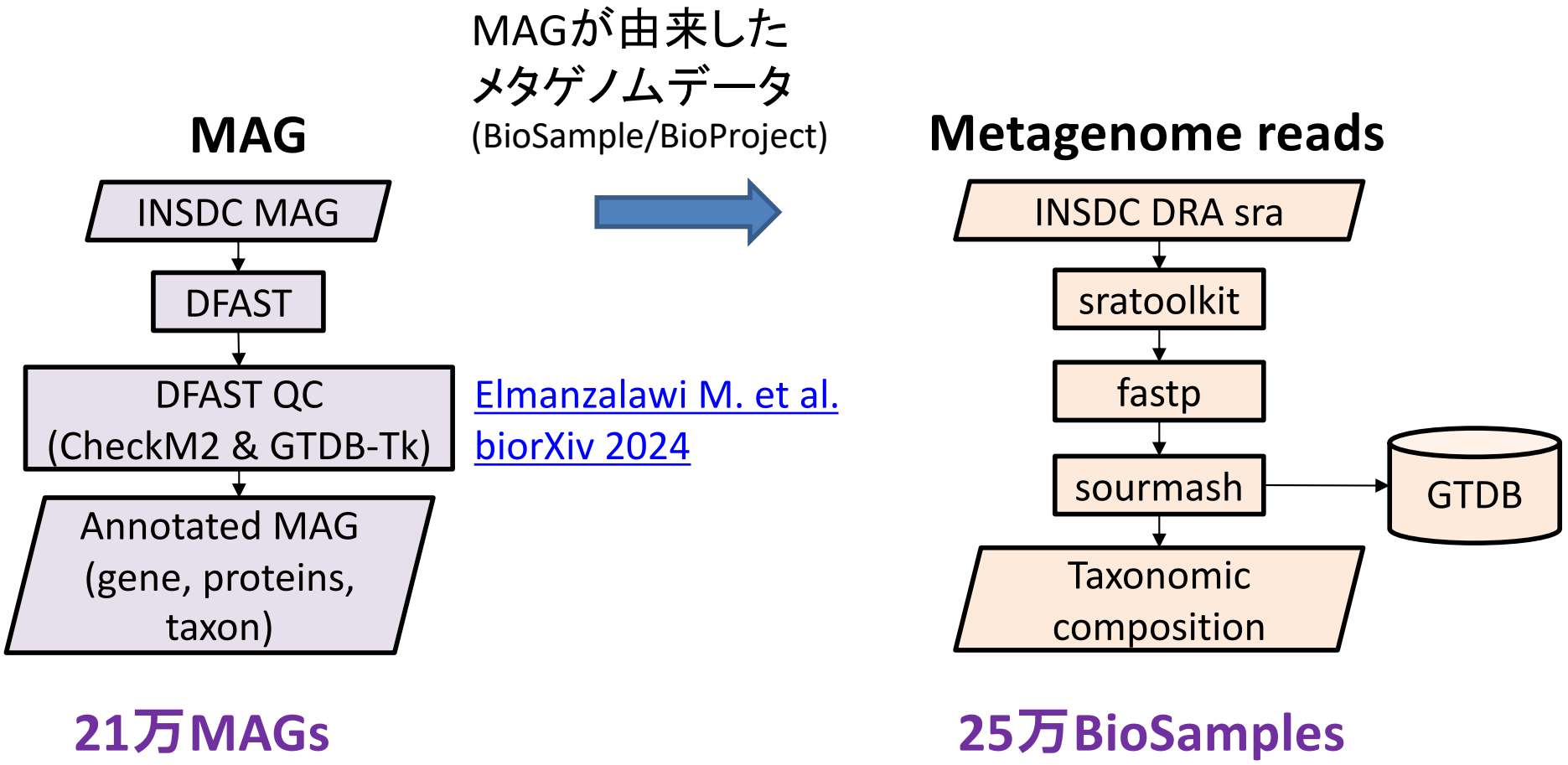
https://microbedb.jp → **https://mdatahub.org**

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



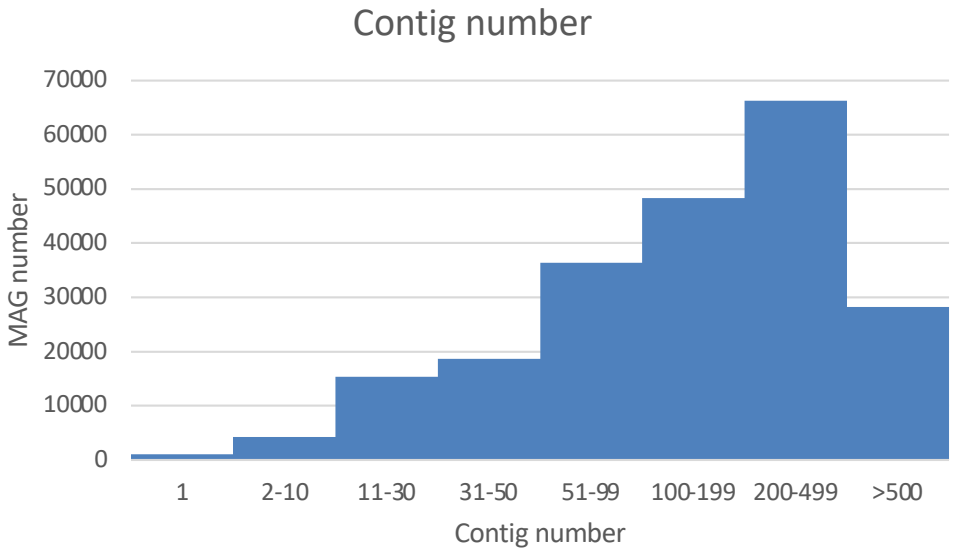
Microbiome DatahubにおけるMAGの収集とアノテーション

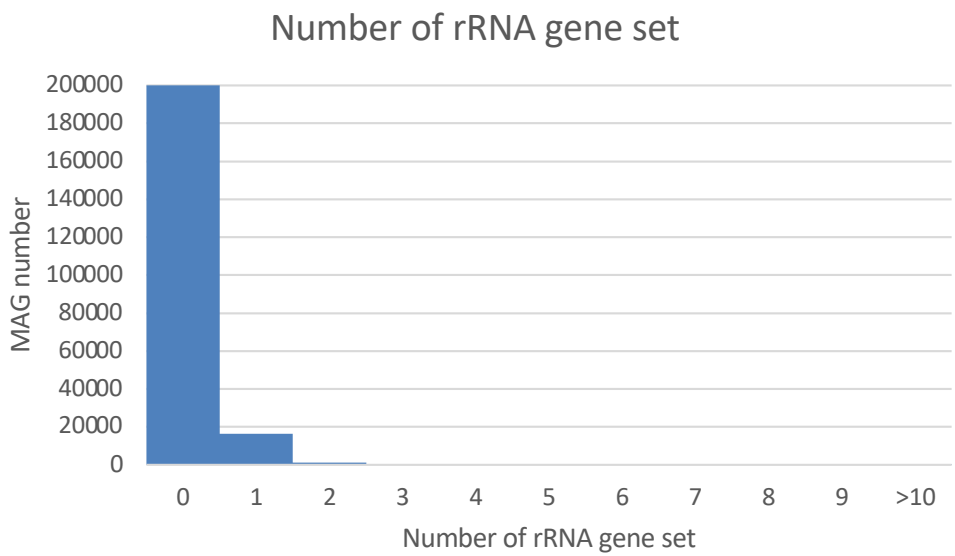


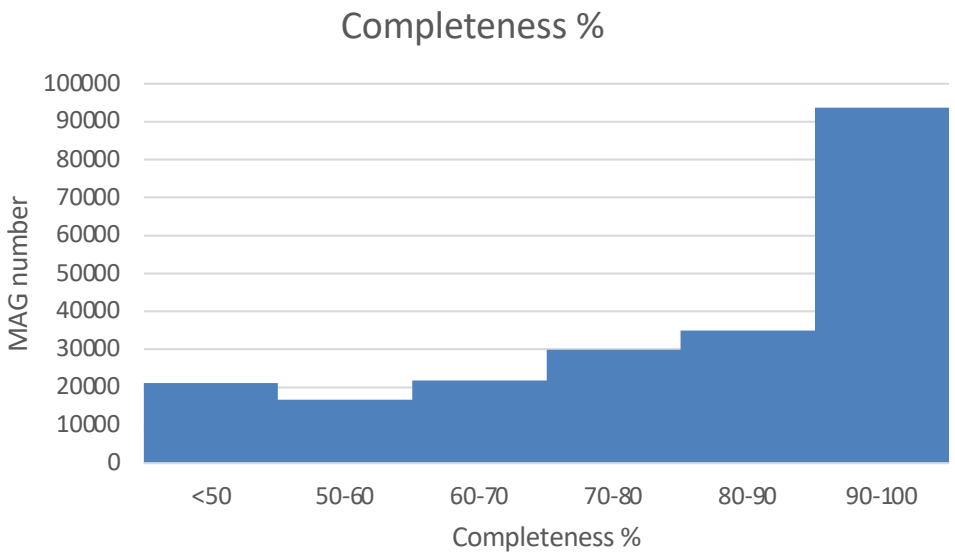
21万MAGs

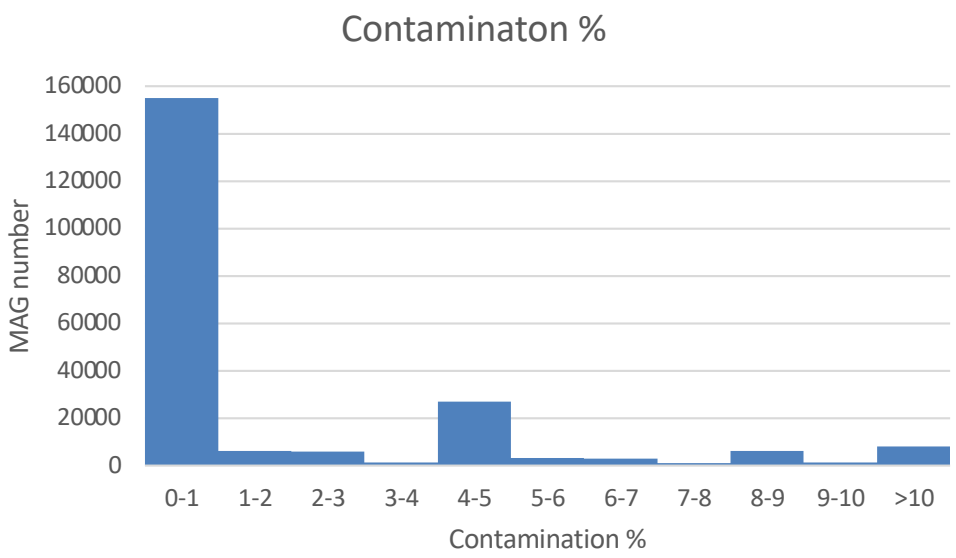
25万BioSamples

Microbiome Datahubにおける21万MAGの内訳

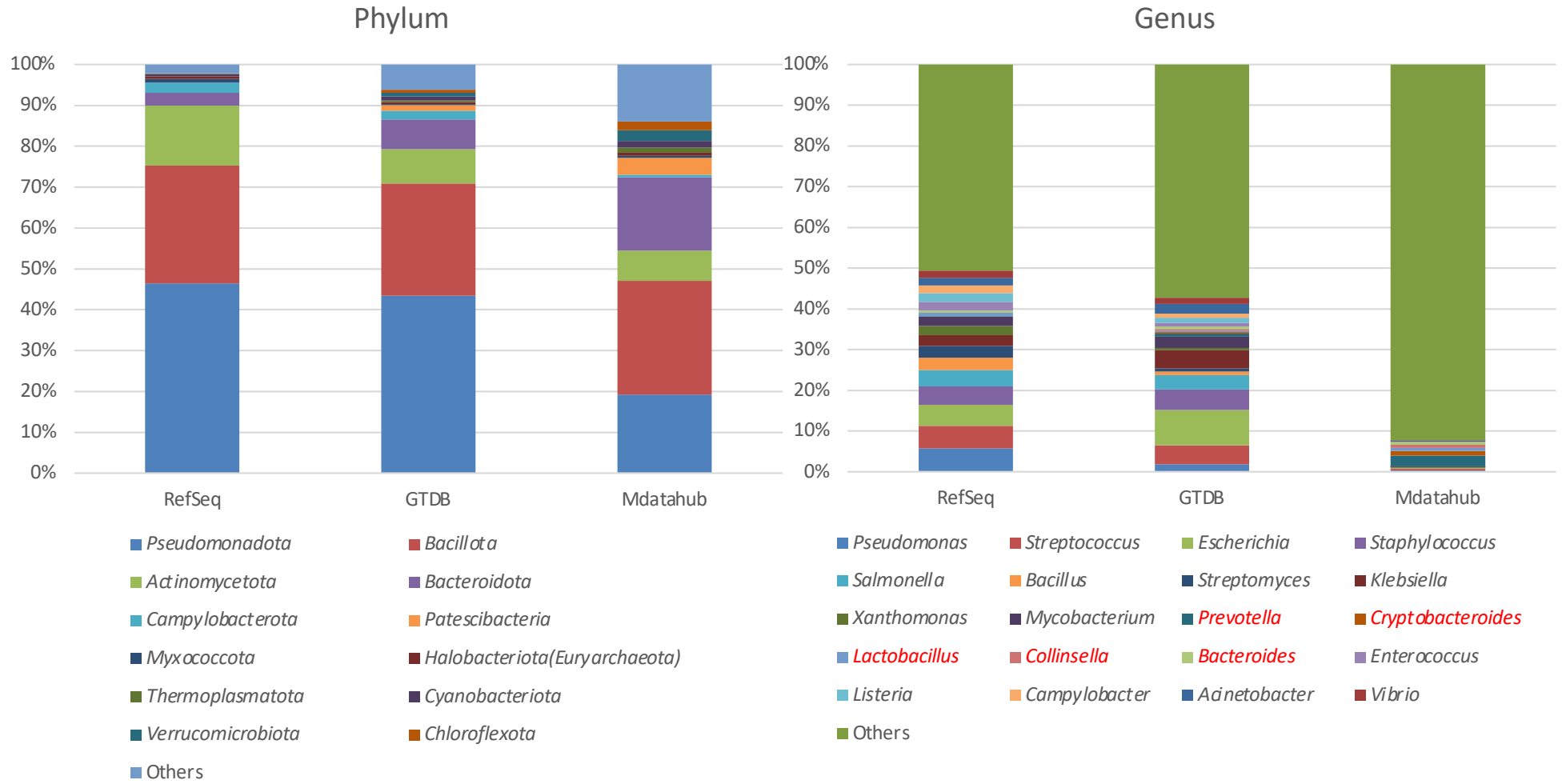








Microbiome Datahubにおける21万MAGの内訳と比較



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

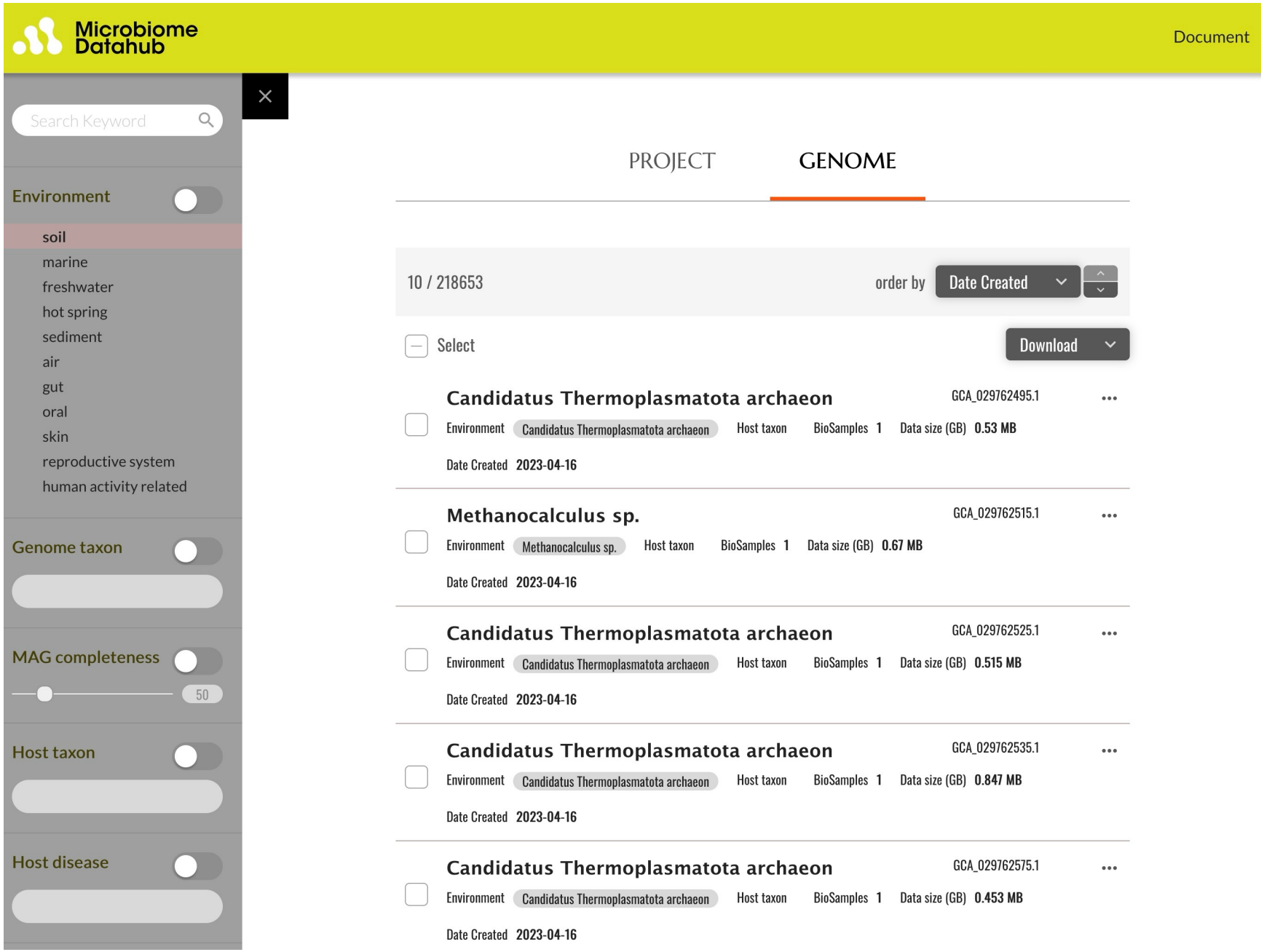
Microbiome Datahubでできること

- Projectのメタデータ、サンプルの系統組成の検索
  - 環境情報を用いてショットガンメタゲノムのINSDC BioProjectを検索、メタデータをTSV形式で取得
  - 環境情報を用いてショットガンメタゲノムのINSDC BioProjectを検索、サンプルの系統組成データをTSV形式で取得

- MAGのメタデータによる検索
  - 系統名やメタデータを用いてMAGデータを検索、メタデータや配列データを取得

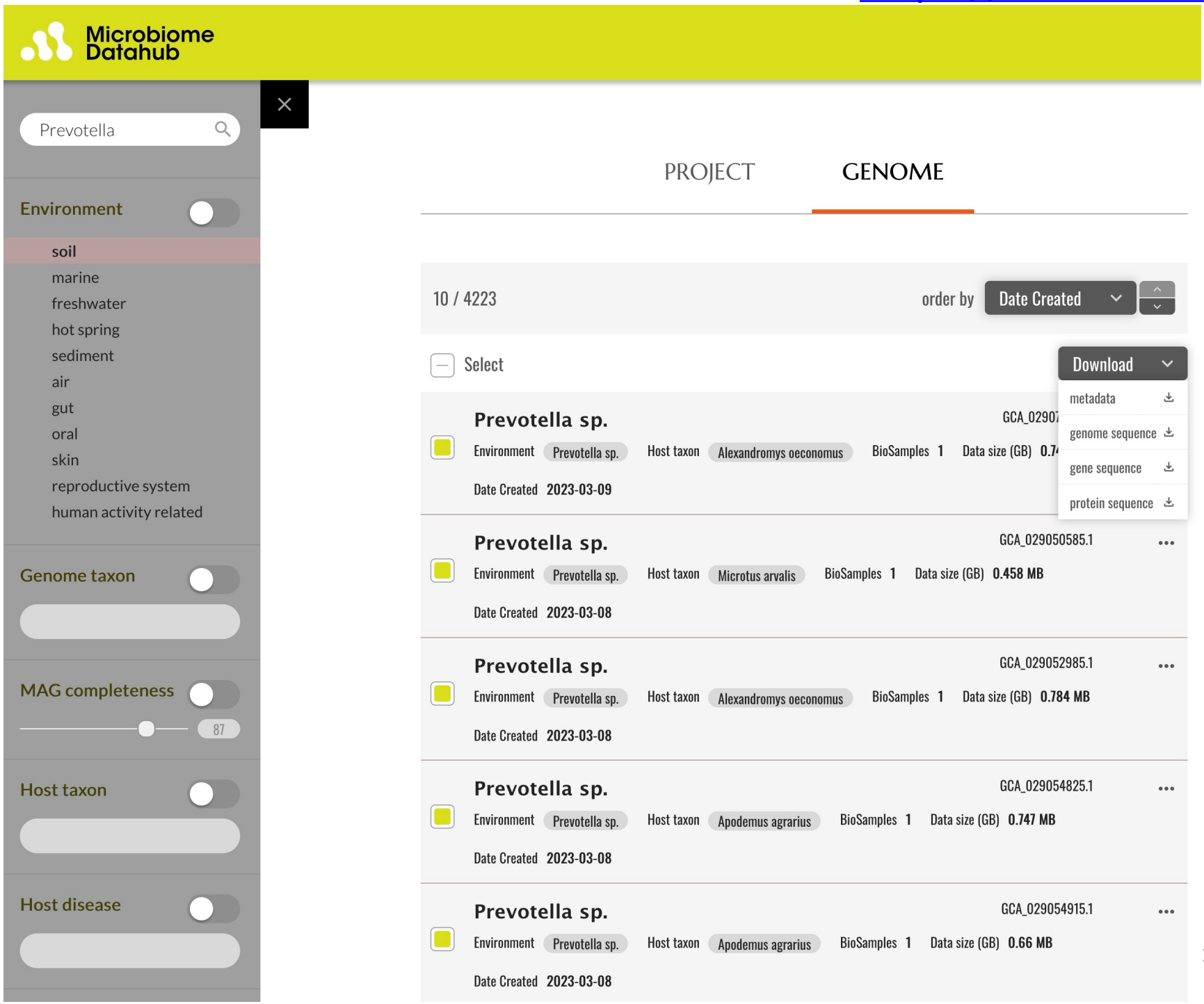
Microbiome DatahubのMAG検索

https://mdatahub.org



Microbiome DatahubのMAG検索

https://mdatahub.org



Microbiome DatahubのMAG検索のダウンロード結果例

MAG metadata例 (tsv形式)

DFAST結果

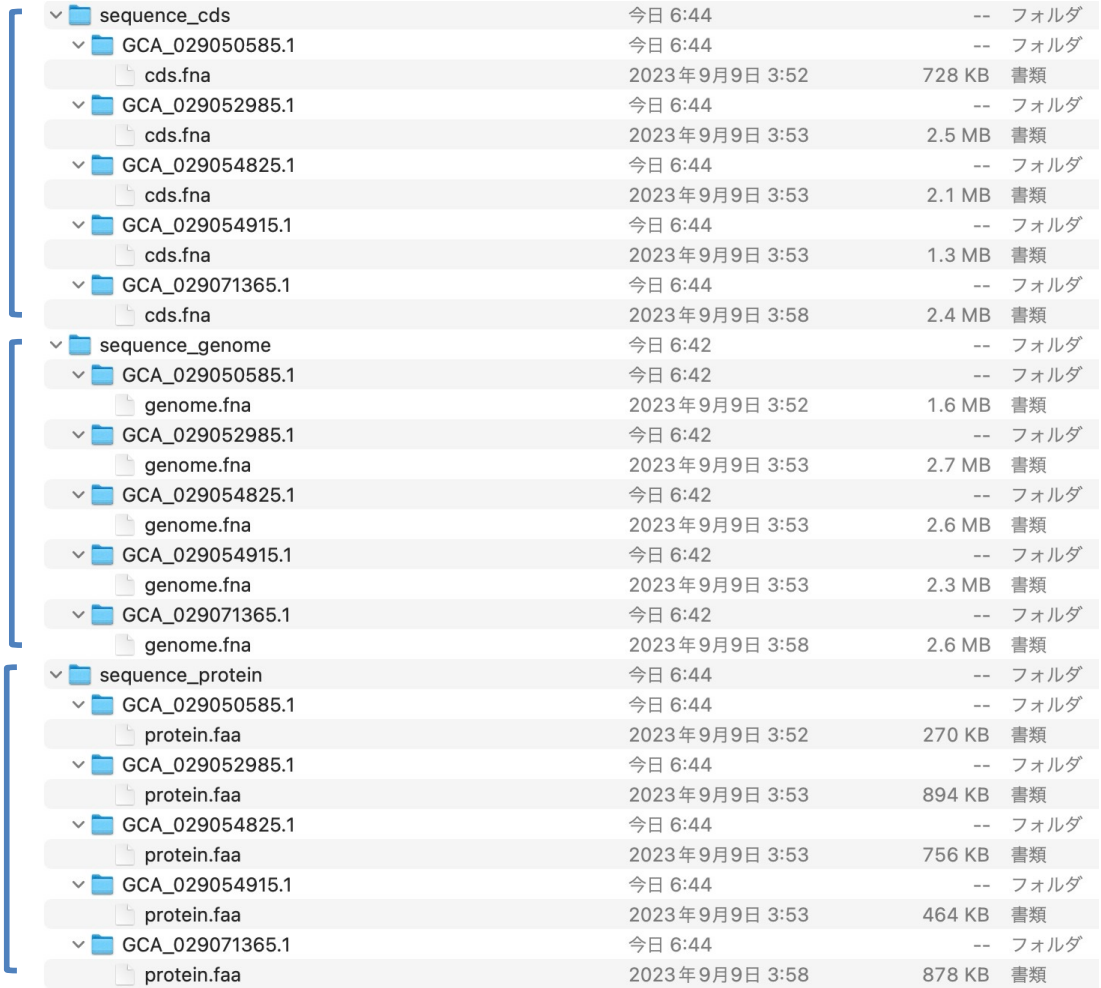
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	identifier	organism	quality	bioproject	biosample	species_taxid	san	san	sample_host_organism	sa	sa	sa	sample_host_location	sa	sa	Total Sequence Length (bp)	Number of Sequences	Longe	N50	GCcontent (%)	Number of CDSs
2	GCA_029050585.1	Prevotella sp.	2	PRJNA725899	SAMN19956079	59823	Pre	##	Microtus arvalis				Lithuania: Vilnius	{"	{"	1572930	703	####	###	51.3	813
3	GCA_029052985.1	Prevotella sp.	3	PRJNA725899	SAMN19955964	59823	Pre	##	Alexandromys oeconomus				Lithuania: Vilnius	{"	{"	2726451	158	####	###	46.8	2085
4	GCA_029054825.1	Prevotella sp.	3	PRJNA725899	SAMN19955874	59823	Pre	##	Apodemus agrarius				Lithuania: Vilnius	{"	{"	2610678	250	####	###	48.7	1875
5	GCA_029054915.1	Prevotella sp.	3	PRJNA725899	SAMN19955868	59823	Pre	##	Apodemus agrarius				Lithuania: Vilnius	{"	{"	2283067	724	####	###	52	1254
6	GCA_029071365.1	Prevotella sp.	3	PRJNA725899	SAMN19956450	59823	Pre	##	Alexandromys oeconomus				Lithuania: Vilnius	{"	{"	2595038	74	####	###	51	2081

Genome sequence, gene sequence, protein sequence例 (fasta形式)

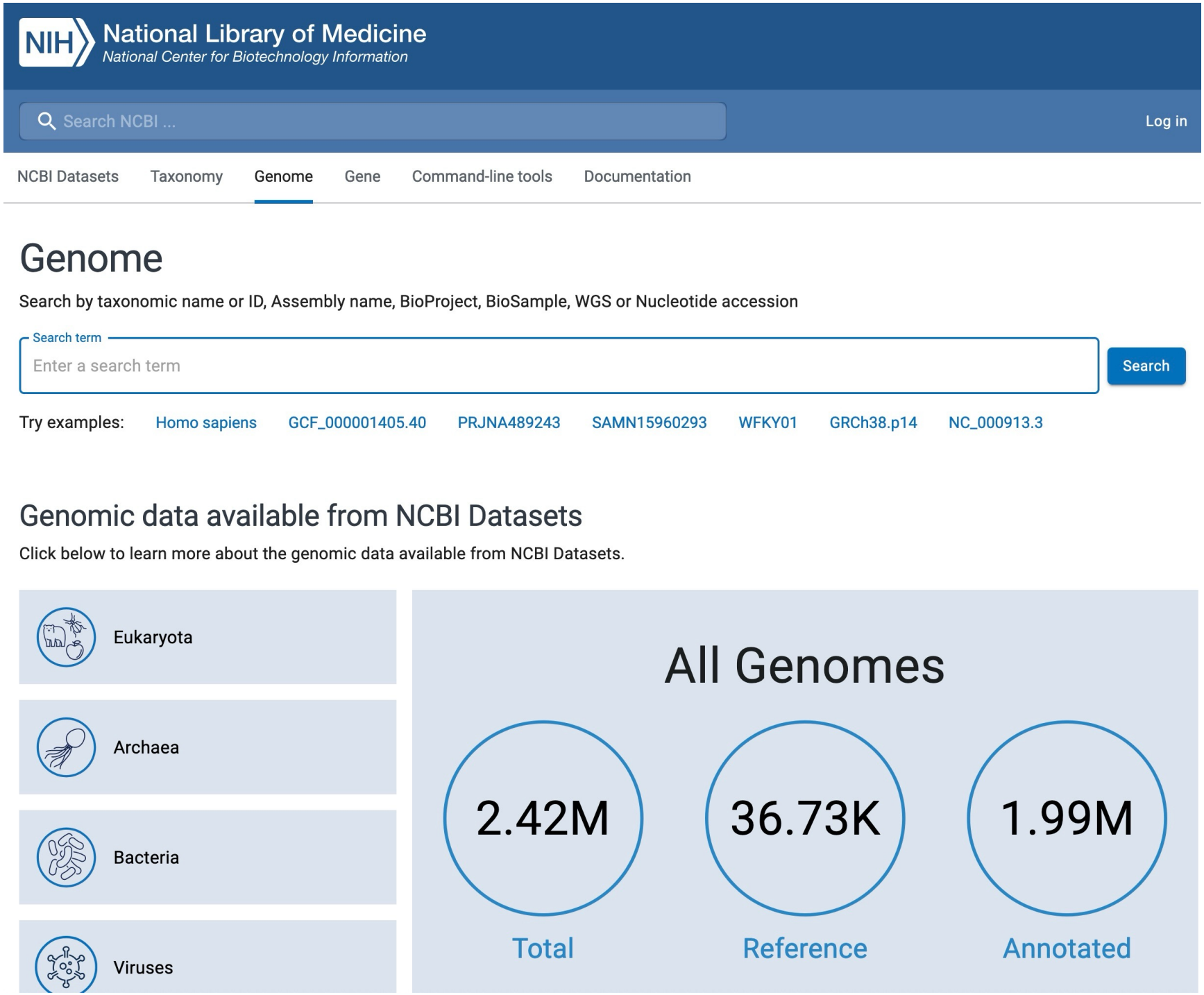
Gene sequence

Genome sequence

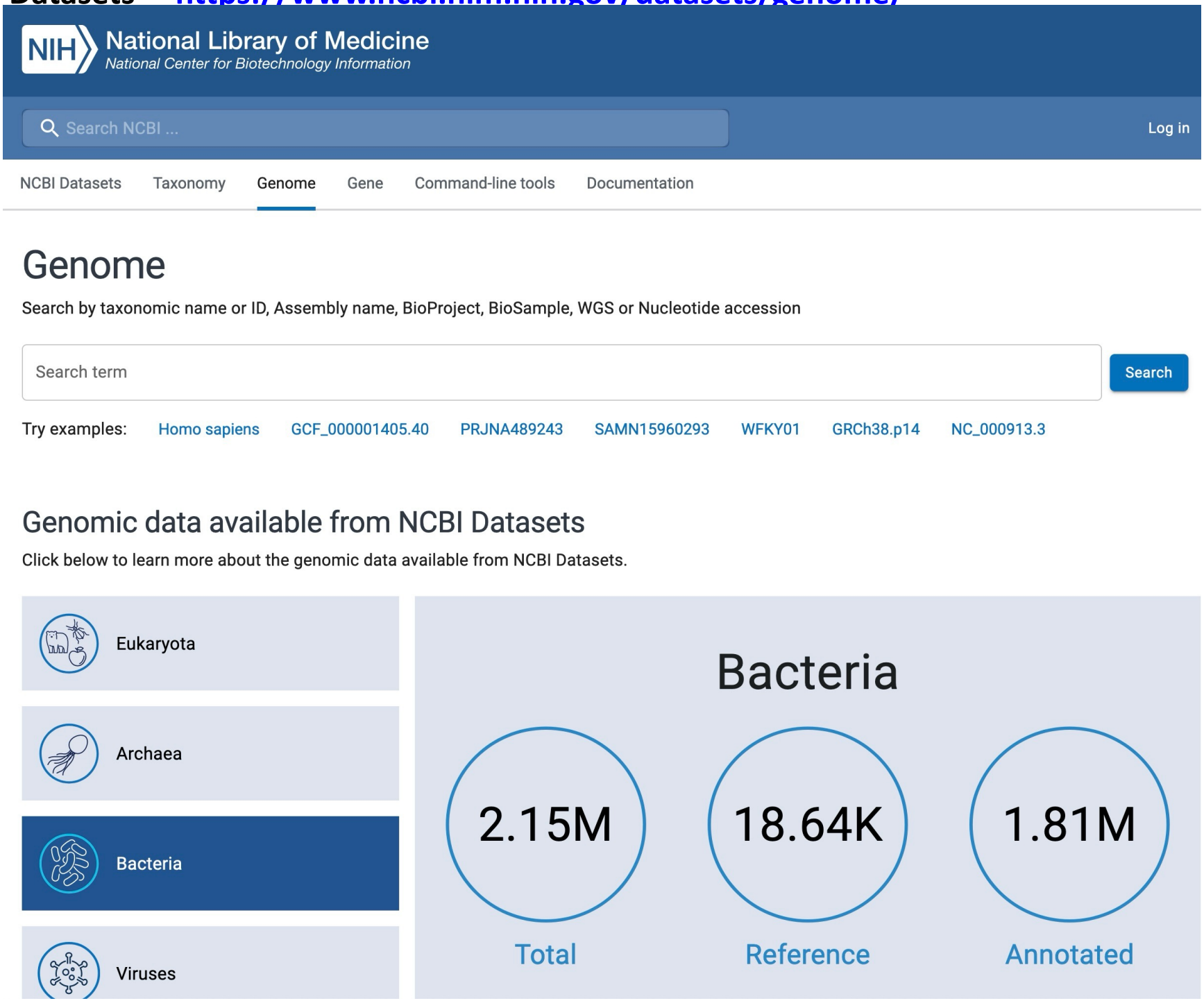
Protein sequence



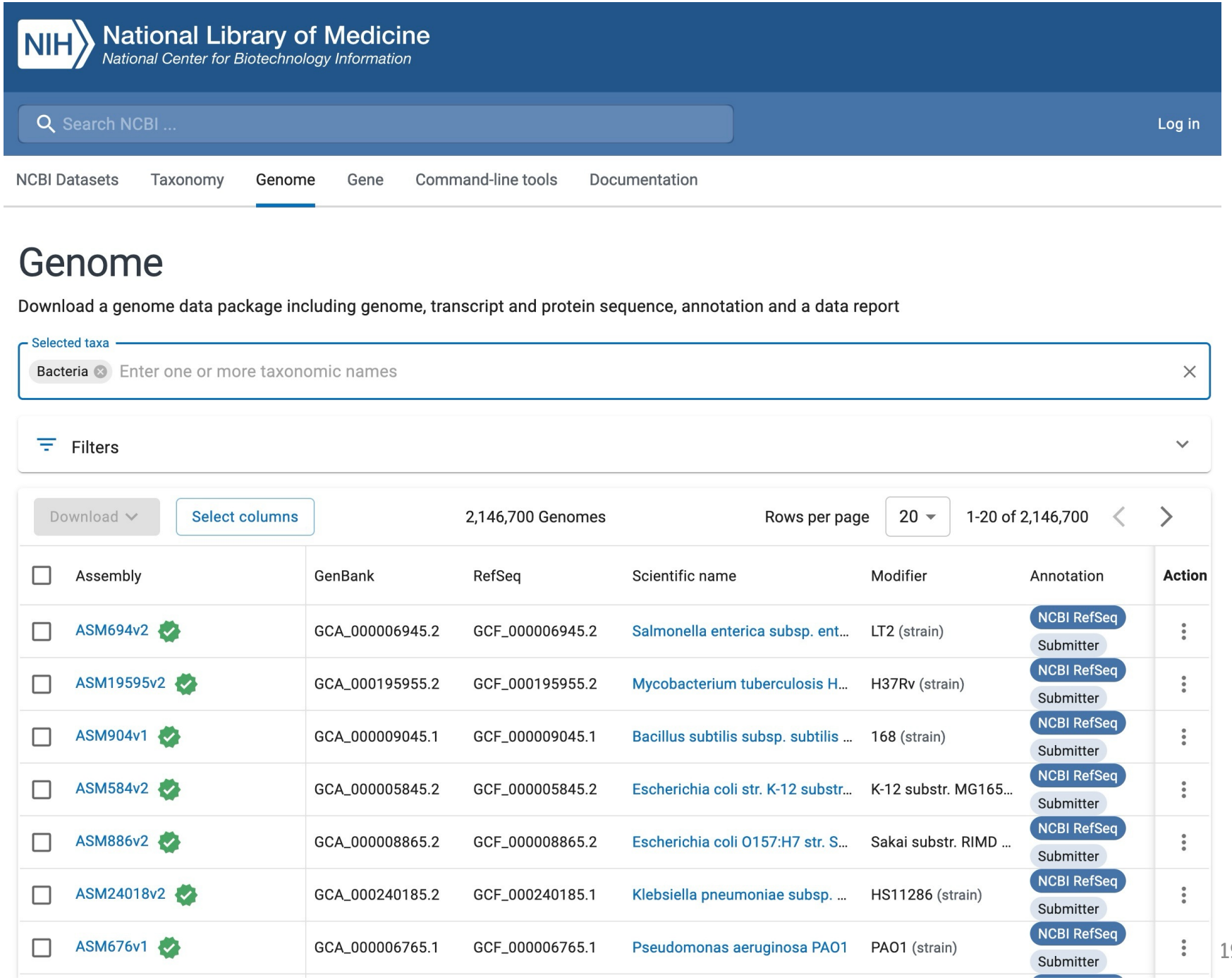
NCBI Datasets https://www.ncbi.nlm.nih.gov/datasets/genome/



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



NCBI Datasets https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=2



NCBI Datasets https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=2&mags_only=true

Genome

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa

Bacteria | Enter one or more taxonomic names

Filters MAGs

INCLUDE ONLY

- [ ] Reference genomes
- [ ] Annotated genomes
  - [ ] Annotated by NCBI RefSeq
  - [ ] Annotated by GenBank submitter
- [x] Metagenome-assembled genomes (MAGs)
- [ ] From type material
- [ ] From ICTV exemplar

EXCLUDE

- [ ] Atypical genomes
- [ ] Metagenome-assembled genomes (MAGs)
- [ ] Genomes from large multi-isolate projects

SEARCH WITHIN RESULTS

ASSEMBLY LEVEL

contig | scaffold | chromosome | complete

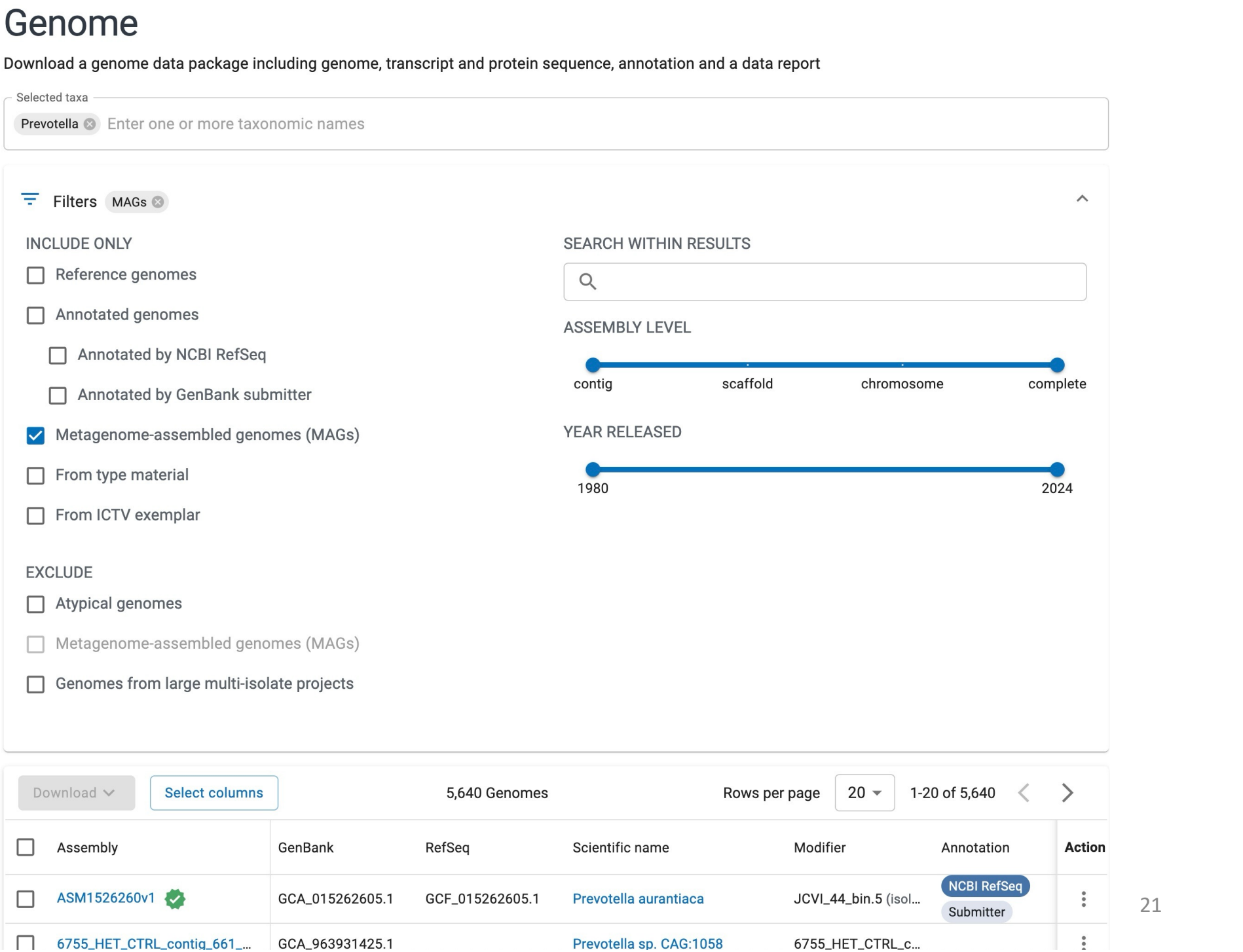
YEAR RELEASED

1980 — 2024

Download | Select columns | 424,066 Genomes | Rows per page 20 | 1-20 of 424,066

Assembly	GenBank	RefSeq	Scientific name	Modifier	Annotation	Action
piMetPara1.Trichococcus_floc…	GCA_963662535.1	GCF_963662535.1	Trichococcus flocculiformis	659e37ff-3402-402…	NCBI RefSeq	
ASM1318141v1	GCA_013181415.1	GCF_013181415.1	Ferrovum myxofaciens	S2.4 (isolate)	NCBI RefSeq	

NCBI Datasets https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=838&mags_only=true



NCBI Datasets https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=838&mags_only=true&assembly_level=2:3

Genome

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa

Prevotella | Enter one or more taxonomic names

Filters MAGs chromosome+

INCLUDE ONLY

- [ ] Reference genomes
- [ ] Annotated genomes
  - [ ] Annotated by NCBI RefSeq
  - [ ] Annotated by GenBank submitter
- [x] Metagenome-assembled genomes (MAGs)
- [ ] From type material
- [ ] From ICTV exemplar

EXCLUDE

- [ ] Atypical genomes
- [ ] Metagenome-assembled genomes (MAGs)
- [ ] Genomes from large multi-isolate projects

SEARCH WITHIN RESULTS

ASSEMBLY LEVEL

contig | scaffold | chromosome | complete

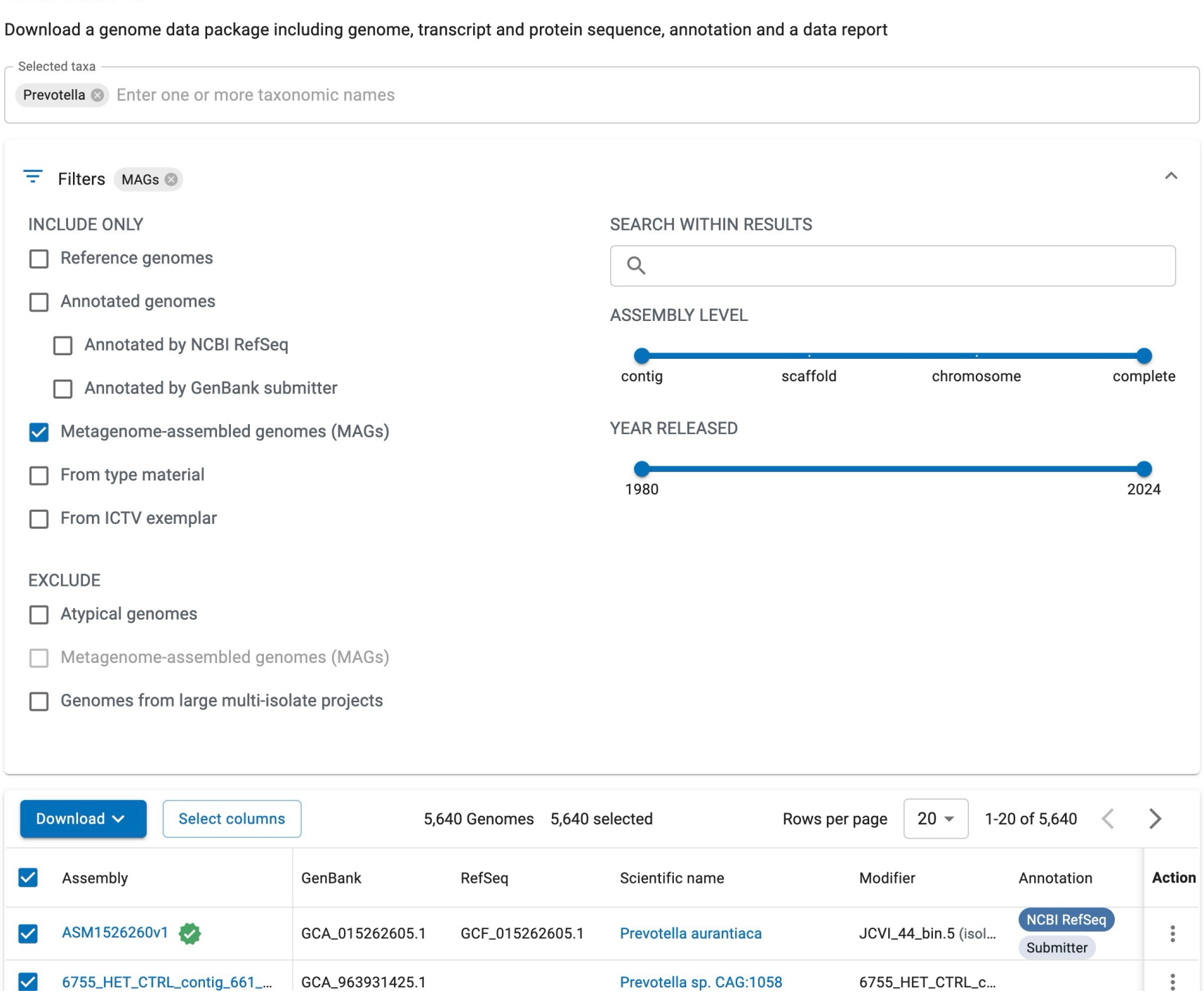
YEAR RELEASED

1980 — 2024

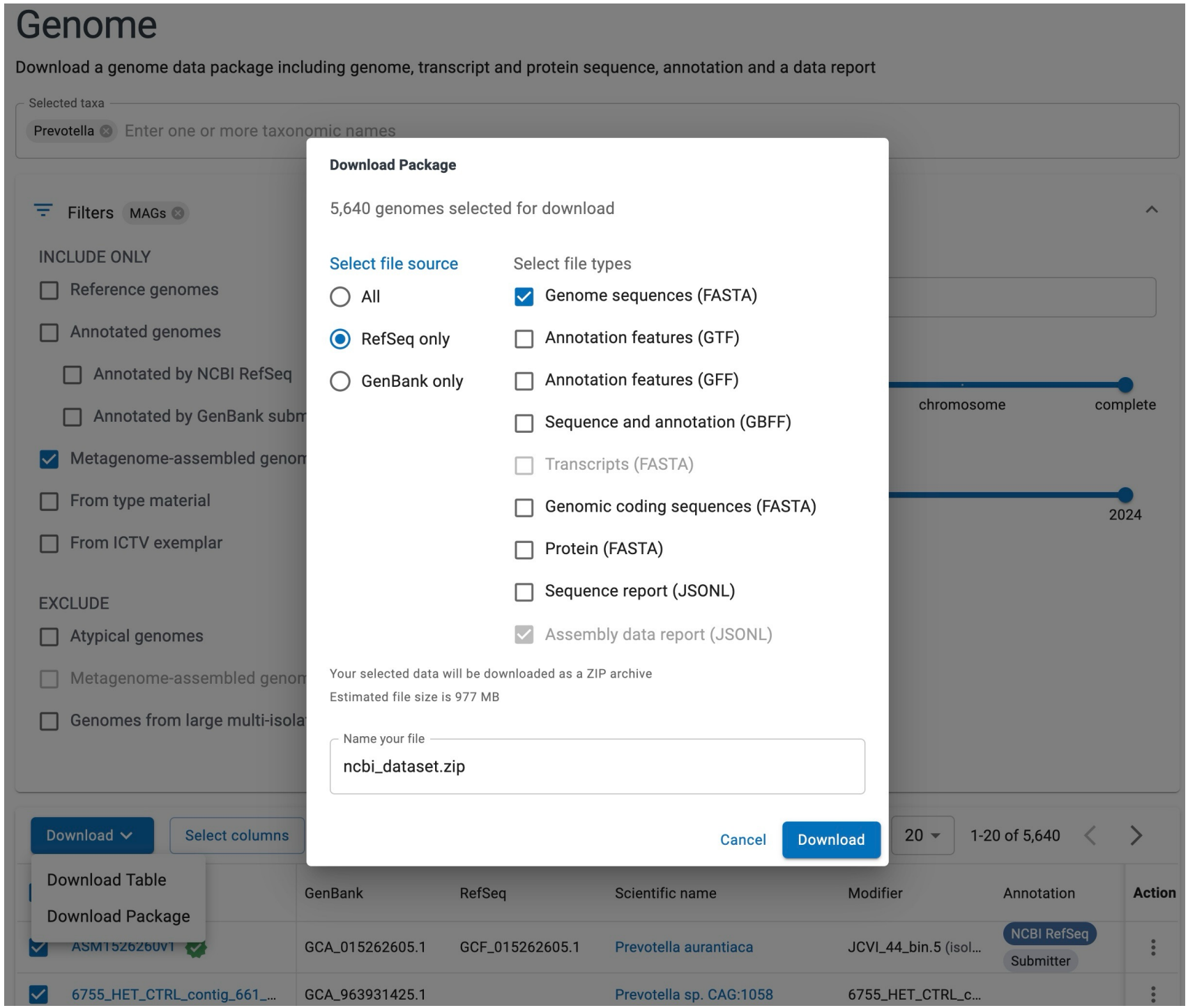
Download | Select columns | 3 Genomes | Rows per page 20 | 1-3 of 3

Assembly	GenBank	RefSeq	Scientific name	Modifier	Annotation	Action
6755_HET_CTRL_contig_661_...	GCA_963931425.1		Prevotella sp. CAG:1058	6755_HET_CTRL_c...		
metabat2_pooled_hifiasm.683_...	GCA_963931275.1		uncultured Prevotella sp.	metabat2_pooled_h...		

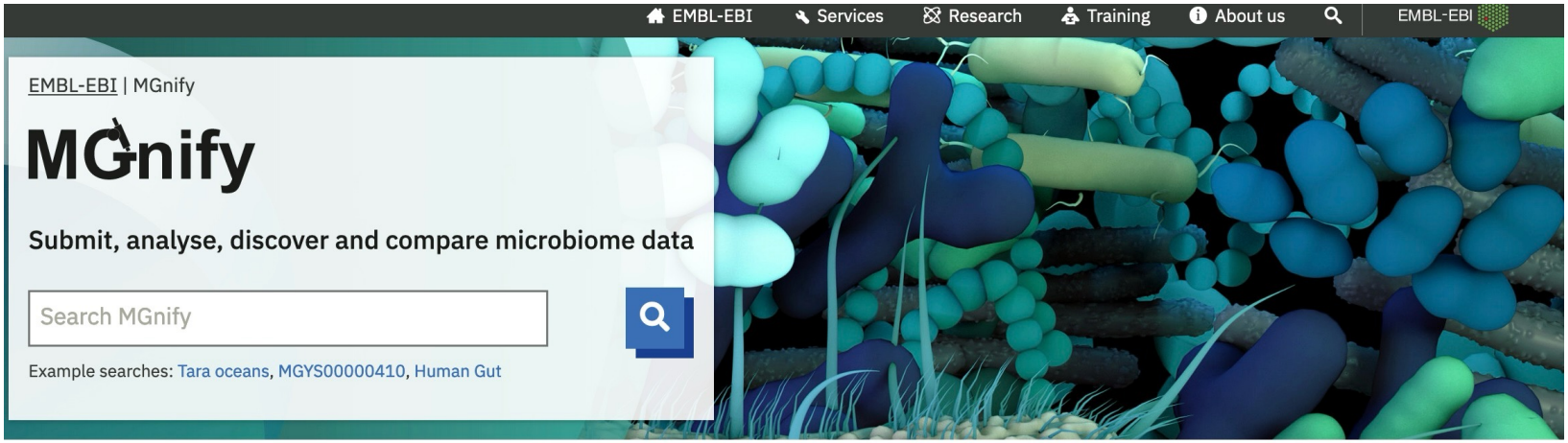
NCBI Datasets https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=838&mags_only=true



NCBI Datasets https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=838&mags_only=true



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



Overview | Submit data | Text search | Sequence search | Browse data | About | Help | Login

Search by

Text search → Name, biome, or keyword

Sequence search → Sequence search

Request analysis of

Submit and/or Request → Your data

Request → A public dataset

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

Latest studies

EMG produced TPA metagenomics assembly of the Gut microbial dysbiosis in young adults with obesity (...

The Gut microbial dysbiosis in young adults with obesity Third Party Annotation (TPA) assembly was derived from the primary whole genome shotgun (WGS) data set: PRJEB12123. This project includes samples from the following biomes: Host-associated, Hu...

Oil spill dispersant strategies and bioremediation efficientcy

After accidental oil spills to seawater from production or transport facilities the oil will be spread on the water surface and in the water column, and a number of weathering processes will appear. Typical operational methods to remove the oil from ...

nifH gene in KH-11-10 and KH-13-7 cruises

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

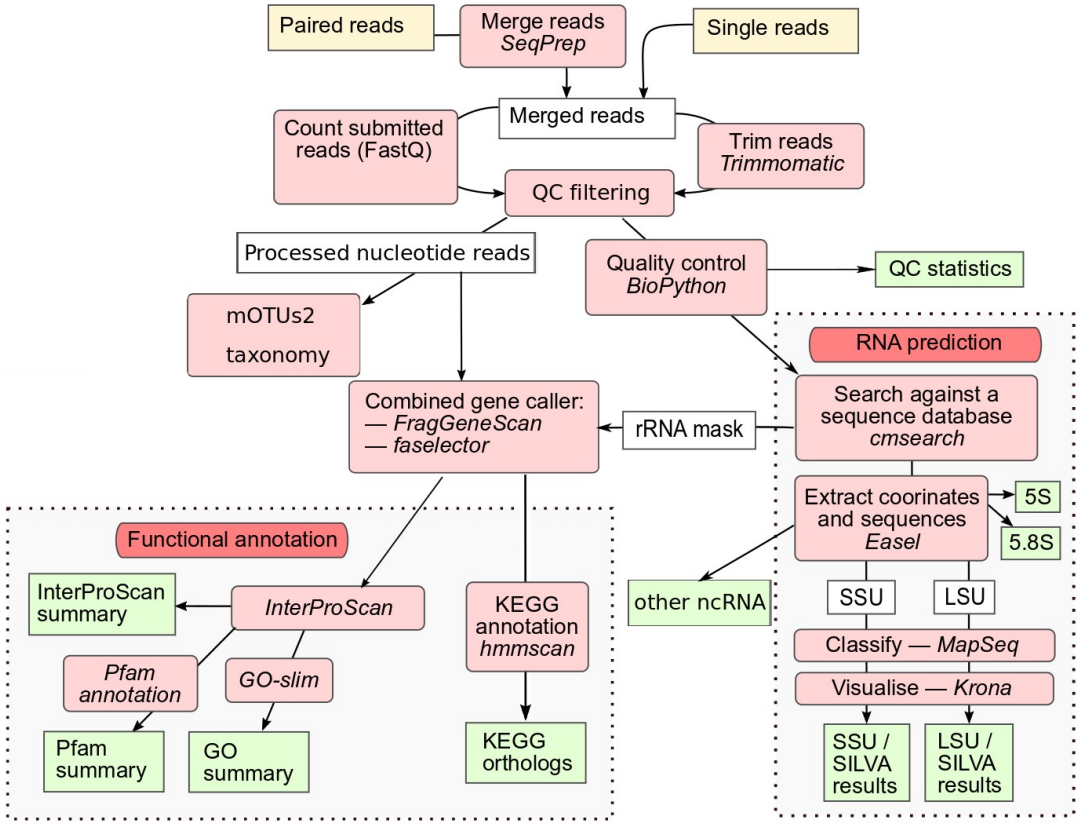
Public data

5004	studies
597736	analyses
478810	genomes in 11 MAG catalogues

Or by selected biomes

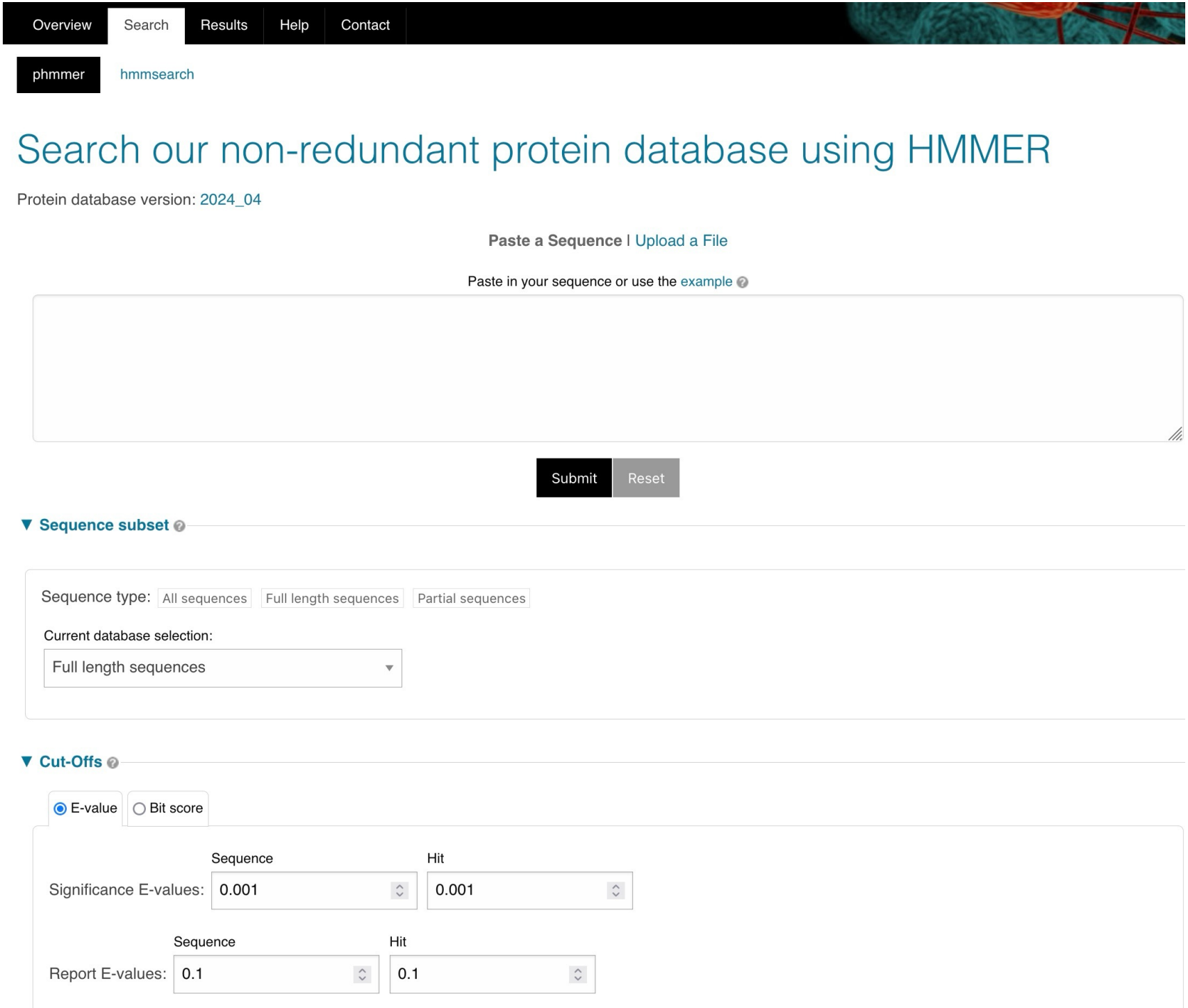
Human (213874)
Digestive system (111024)
Aquatic (51540)
Marine (38036)
Digestive system (35543)
Plants (28859)
Soil (25893)
Skin (11527)
Wastewater (4358)
Food production (2923)

View all biomes



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

MGnify https://www.ebi.ac.uk/metagenomics/sequence-search/search/phmmer



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

Home > Browse > Genomes

Browse MGnify

Super Studies | Studies | Samples | Publications | **Genomes** | Biomes

Genome catalogues are biome-specific collections of metagenomic-assembled and isolate genomes. The latest version of each catalogue is shown on this website. Data for current and previous versions are available on the FTP server.

If you use the MGnify Genomes resource, please cite:

MGnify Genomes: a resource for biome-specific microbial genome catalogues *Journal of Molecular Biology* (2023) doi:10.1016/j.jmb.2023.168016
Gurbich TA, Almeida A, Beracochea M, Burdett T, Burgin T, Cochrane G, Raj S, Richardson LJ, Rogers AB, Sakharova E, Salazar GA and Finn RD.

Catalogues list | All genomes | Gene search | MAG search

Select a catalogue in the table to browse or search its genomes.

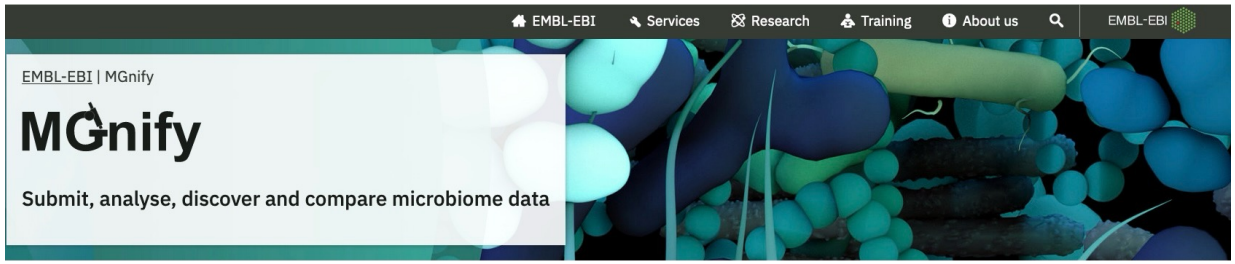
Filter biome

All root

Download

Biome	Catalogue ID	Catalogue name	Catalogue version	Species count	Total genomes count	Last Updated
	chicken-gut-v1-0-1	Chicken Gut v1.0.1	1.0.1	1322	13386	2024/1/23
	cow-rumen-v1-0-1	Cow Rumen v1.0.1	1.0.1	2729	5578	2024/1/23
	honeybee-gut-v1-0-1	Honeybee gut v1.0.1	1.0.1	154	627	2024/1/23
	human-gut-v2-0-2	Unified Human Gastrointestinal Genome (UHGG) v2.0.2	2.0.2	4744	289232	2024/1/23
	human-oral-v1-0-1	Human Oral v1.0.1	1.0.1	452	1225	2024/1/23
	human-vaginal-v1-0	Human Vaginal v1.0	1.0	280	618	2023/8/15
	marine-v2-0	Marine v2.0	2.0	13223	50866	2024/5/12
	mouse-gut-v1-0	Mouse Gut v1.0	1.0	2847	112951	2023/12/8
	non-model-fish-gut-v2-0	Non-model Fish Gut v2.0	2.0	178	254	2023/4/26
	pig-gut-v1-0	Pig Gut v1.0	1.0	1376	3972	2022/11/29
	zebrafish-fecal-v1-0	Zebrafish Fecal v1.0	1.0	79	101	2022/11/28

MGnify https://www.ebi.ac.uk/metagenomics/browse/genomes?browse-by=search-all&search=Prevotella



Overview | Submit data | Text search | Sequence search | Browse data | About | Help | Login

Home > Browse > Genomes

Browse MGnify

Super Studies | Studies | Samples | Publications | **Genomes** | Biomes

Genome catalogues are biome-specific collections of metagenomic-assembled and isolate genomes. The latest version of each catalogue is shown on this website. Data for current and previous versions are available on the FTP server.

If you use the MGnify Genomes resource, please cite:

MGnify Genomes: a resource for biome-specific microbial genome catalogues *Journal of Molecular Biology* (2023) doi:10.1016/j.jmb.2023.168016
Gurbich TA, Almeida A, Beracochea M, Burdett T, Burgin T, Cochrane G, Raj S, Richardson LJ, Rogers AB, Sakharova E, Salazar GA and Finn RD.

Catalogues list | **All genomes** | Gene search | MAG search

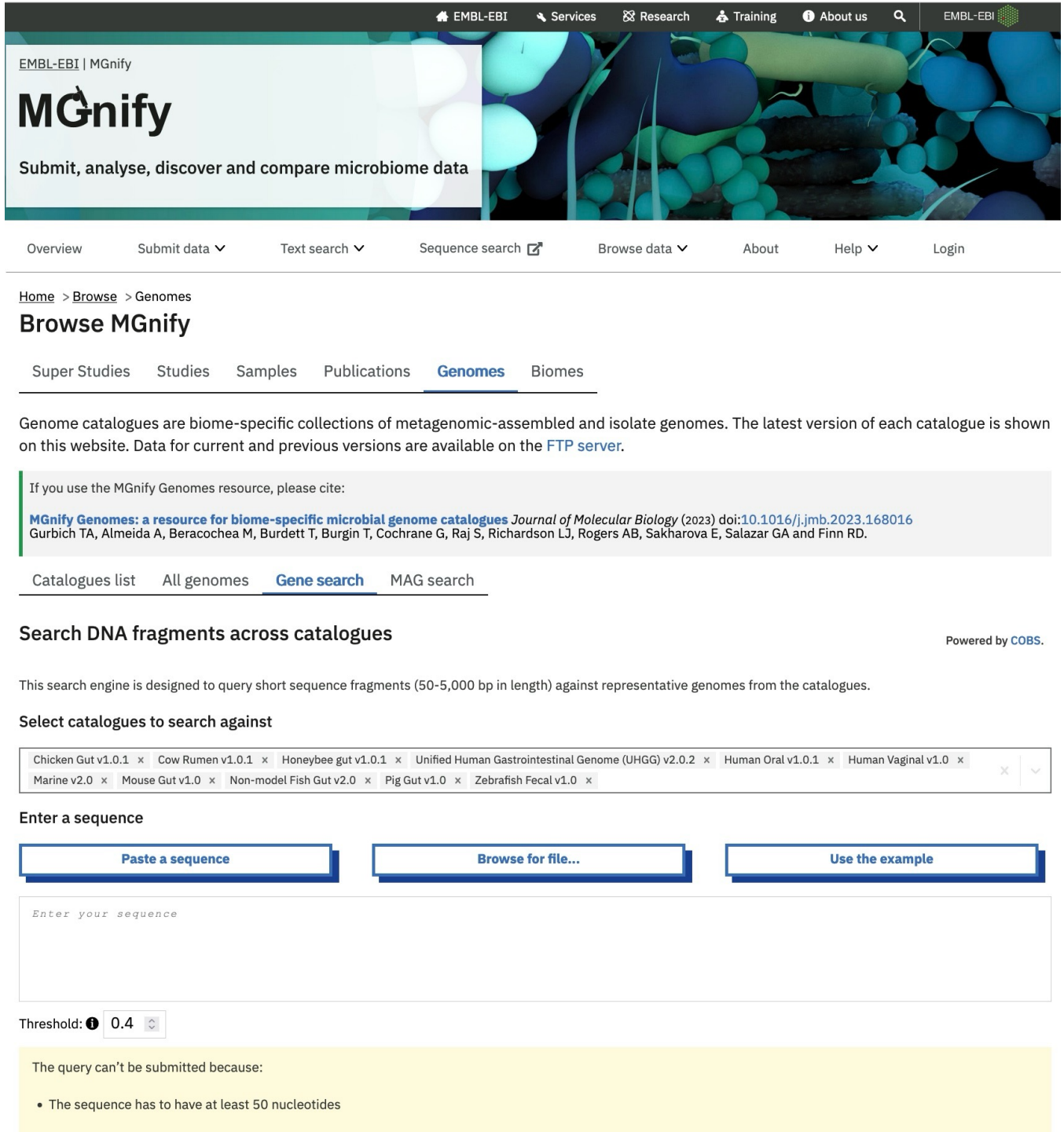
Search for genomes across all catalogues.

Genomes (552)

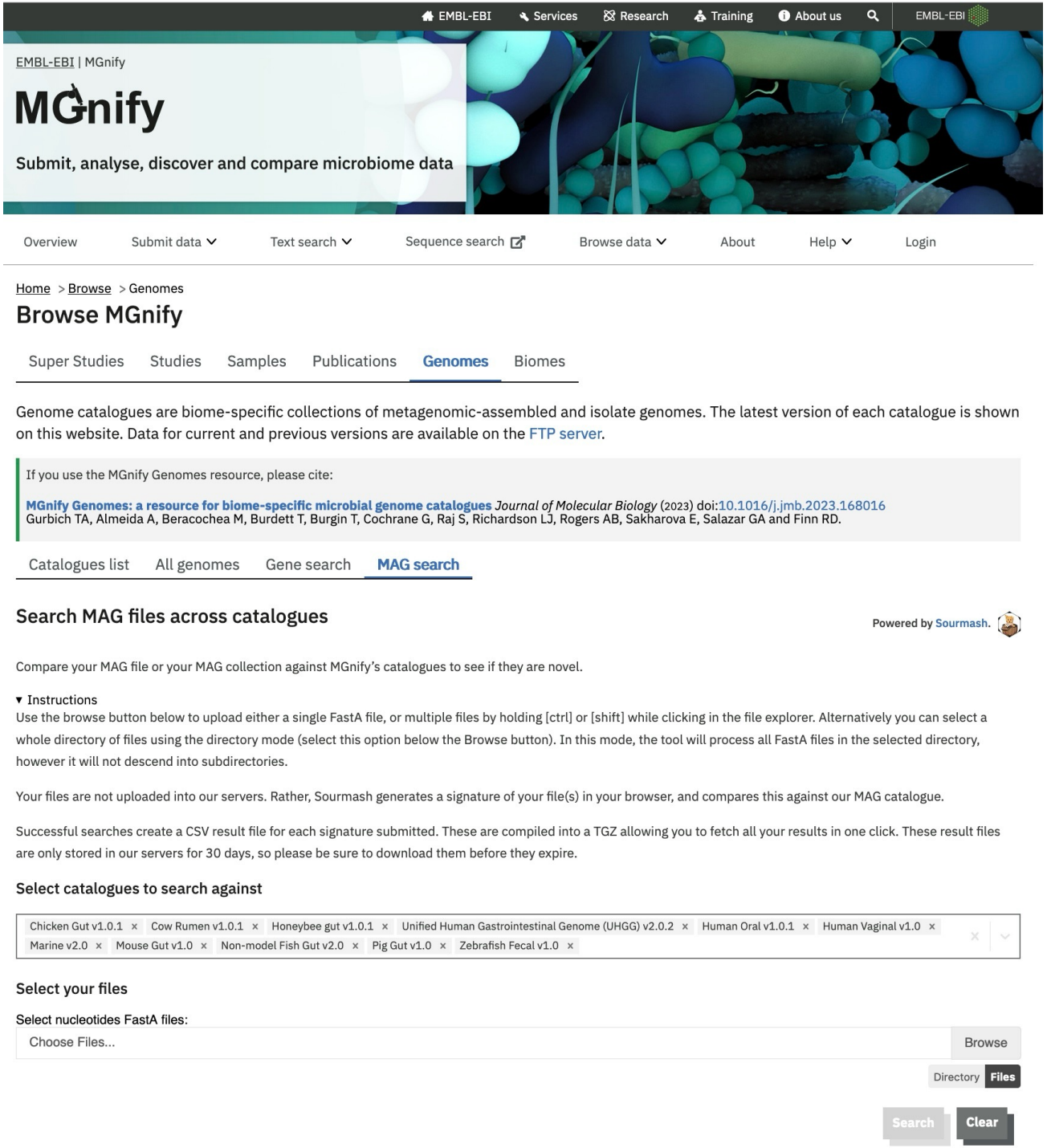
Filter: Prevotella — Download

Biome	Accession	Catalogue	Type	Taxonomy
	MGYG000411028	mouse-gut-v1-0	MAG	Prevotella rodentium
	MGYG000410844	mouse-gut-v1-0	MAG	Paraprevotella sp910586505
	MGYG000409046	mouse-gut-v1-0	MAG	Paraprevotella sp910584955
	MGYG000348300	mouse-gut-v1-0	MAG	Prevotella sp002933775
	MGYG000331390	mouse-gut-v1-0	MAG	Prevotella sp002265625
	MGYG000331388	mouse-gut-v1-0	MAG	Alloprevotella sp900542795
	MGYG000331371	mouse-gut-v1-0	MAG	Prevotella sp000436035
	MGYG000331314	mouse-gut-v1-0	MAG	Prevotella sp900553595
	MGYG000331239	mouse-gut-v1-0	MAG	Prevotella pectinovora
	MGYG000330961	mouse-gut-v1-0	MAG	Prevotella sp900552675

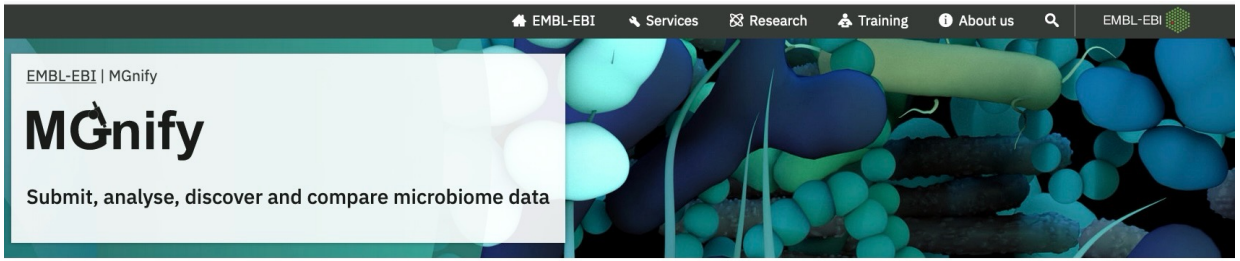
MGnify https://www.ebi.ac.uk/metagenomics/browse/genomes?browse-by=gene-search



MGnify https://www.ebi.ac.uk/metagenomics/browse/genomes?browse-by=mag-search



MGnify https://www.ebi.ac.uk/metagenomics/genome-catalogues/human-gut-v2-0-2



Overview | Submit data | Text search | Sequence search | Browse data | About | Help | Login

Home > Genomes > Unified Human Gastrointestinal Genome (UHGG) v2.0.2

Unified Human Gastrointestinal Genome (UHGG) v2.0.2

289232	4744	FTP Site	Pipeline v1.2.1
Total genomes	Species-level clusters	Download full catalogue	View workflow & tools

This is an update to the UHGG1.0 catalogue described by Almeida et al., Nature Biotechnol (2021).

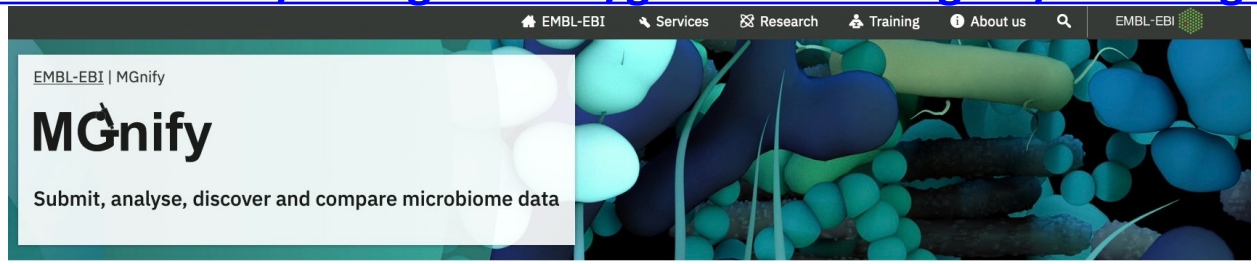
Genome list | Taxonomy tree | Protein catalogue | Search by Gene | Search by MAG

Species-level cluster representatives

Filter: Enter your search terms

Biome	Accession	Length	Num. of genomes	Completeness	Contamination	Type	Taxonomy	Last Updated
	MGYG000000001	3221717	4	98.59	0.7	Isolate	GCA-900066495 sp902362365	2024/1/23
	MGYG000000002	4441003	359	99.37	0	Isolate	Blautia_A faecis	2024/1/23
	MGYG000000003	3234232	1181	100	0	Isolate	Alistipes shahii	2024/1/23
	MGYG000000004	3707250	25	98.66	0.22	Isolate	Anaerotruncus colihominis	2024/1/23
	MGYG000000005	3932530	2	99.3	0	Isolate	Terrisporobacter glycolicus_A	2024/1/23
	MGYG000000006	2823639	1	99.26	1.39	Isolate	Staphylococcus xylosus	2024/1/23
	MGYG000000007	2017471	1	98.94	0	Isolate	Lactobacillus intestinalis	2024/1/23
	MGYG000000008	1972012	24	99.22	0.78	Isolate	Lactobacillus johnsonii	2024/1/23
	MGYG000000009	2221226	2	99.21	0	Isolate	Ligilactobacillus murinus	2024/1/23

MGnify https://www.ebi.ac.uk/metagenomics/genome-catalogues/human-gut-v2-0-2



Overview | Submit data | Text search | Sequence search | Browse data | About | Help | Login

Home > Genomes > Unified Human Gastrointestinal Genome (UHGG) v2.0.2

Unified Human Gastrointestinal Genome (UHGG) v2.0.2

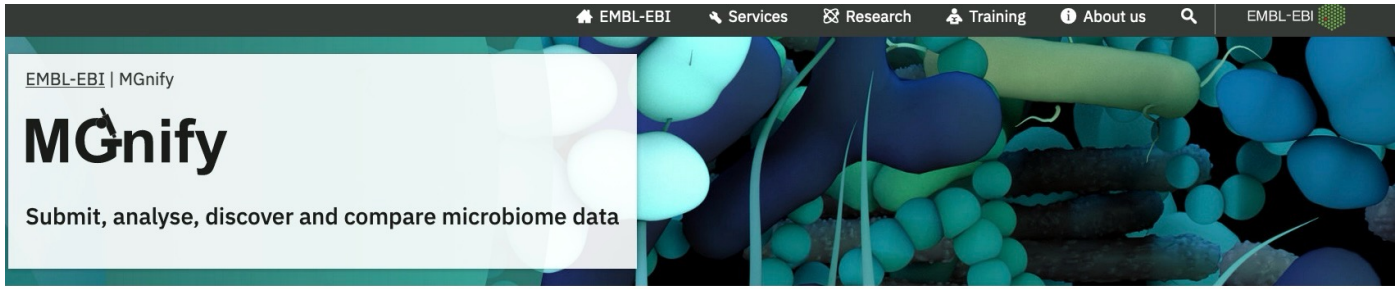
289232	4744	FTP Site	Pipeline v1.2.1
Total genomes	Species-level clusters	Download full catalogue	View workflow & tools

This is an update to the UHGG1.0 catalogue described by Almeida et al., Nature Biotechnol (2021).

Genome list | **Taxonomy tree** | Protein catalogue | Search by Gene | Search by MAG

- Domain
  - Archaea 28
  - Bacteria 4716
    - Firmicutes_A 1906
    - Bacteroidota 619
      - Bacteroidia 619
        - Bacteroidales 609
          - Rikenellaceae 61
          - Bacteroidaceae 260
            - Bacteroides 53
            - Phocaeicola 36
            - Prevotella 113
              - Prevotella stercorea 1
              - Prevotella hominis 1
              - Prevotella buccae 1
              - Prevotella sp900770395 1
              - Prevotella sp900543975 1
              - Prevotella sp900548535 1
              - Prevotella sp002299635 1
              - Prevotella sp900546535 1
              - Prevotella sp000436035 1
              - Prevotella sp900552675 1
              - Prevotella sp900546575 1
              - Unknown 15
              - Prevotella sp000431975 1
              - Prevotella sp900545525 1
              - Prevotella sp900549175 1
              - Prevotella sp900540375 1
              - Prevotella sp900554695 1
              - Prevotella sp900554045 1
              - Prevotella sp000436595 1
              - Prevotella sp900556395 1
              - Prevotella sp002251365 1
              - Prevotella pectinovora 1
              - Prevotella sp900543585 1
              - Prevotella lascolaii 1
              - Prevotella sp900557405 1
              - Prevotella salivae 1
              - Prevotella oralis 1

MGnify https://www.ebi.ac.uk/metagenomics/genomes/MGYG000000526#overview



Overview | Submit data | Text search | Sequence search | Browse data | About | Help | Login

Home > Genomes > human-gut-v2-0-2 > MGYG000000526

Genome MGYG000000526

Type: MAG

Taxonomic lineage: Bacteria > Bacteroidota > Bacteroidia > Bacteroidales > Bacteroidaceae > Prevotella > Prevotella sp900770395

Overview | Browse genome | COG analysis | KEGG class analysis | KEGG module analysis | Downloads

Genome statistics

Type:	MAG
Length (bp):	3510035
Contamination:	0.19%
Completeness:	97.78%
Num. of contigs:	38
Total number of genomes in species:	4
Number of proteins:	2880
GC content:	47.78%
Taxonomic lineage:	Bacteria > Bacteroidota > Bacteroidia > Bacteroidales > Bacteroidaceae > Prevotella > Prevotella sp900770395
N50:	182536

Genome annotations

InterPro coverage:	87.6%
EggNog coverage:	91.11%

Pan-genome statistics

Pan-genome size:	3119
Pan-genome core size:	1696
Pan-genome accessory size:	1423

Genome RNA coverage

rRNA 5s total gene length coverage:	0%
rRNA 16s total gene length coverage:	0%
rRNA 23s total gene length coverage:	0%
tRNAs:	19
ncRNA:	17

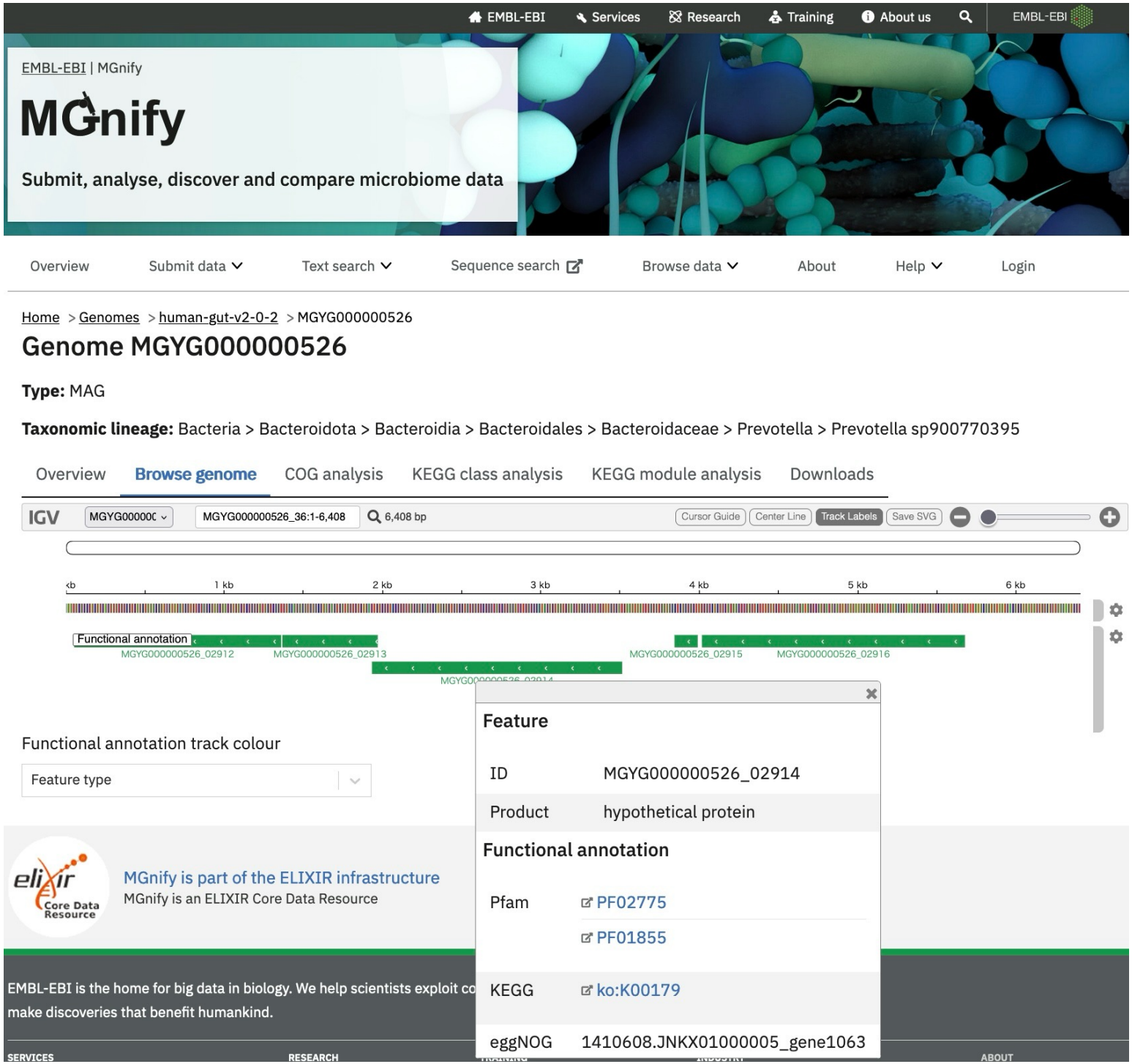
Geographic metadata

Origin of representative genome:	Oceania
Geographic range of pan-genome:	Europe, Oceania

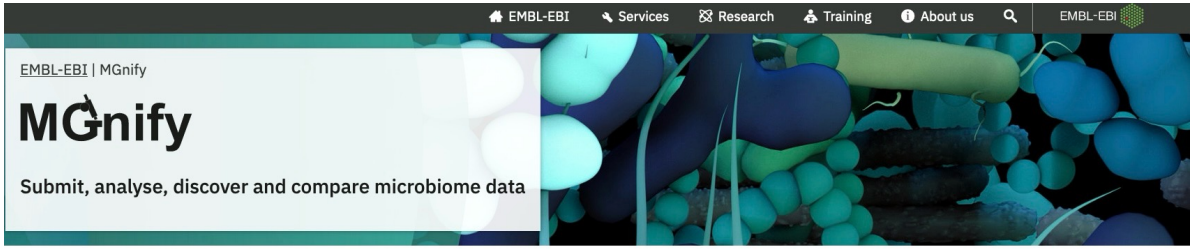
External links

ENA sample accession:	SRS476306
ENA study accession:	SRP029441

MGnify https://www.ebi.ac.uk/metagenomics/genomes/MGYG000000526



MGnify https://www.ebi.ac.uk/metagenomics/genomes/MGYG000000526#kegg-module-analysis



Overview | Submit data | Text search | Sequence search | Browse data | About | Help | Login

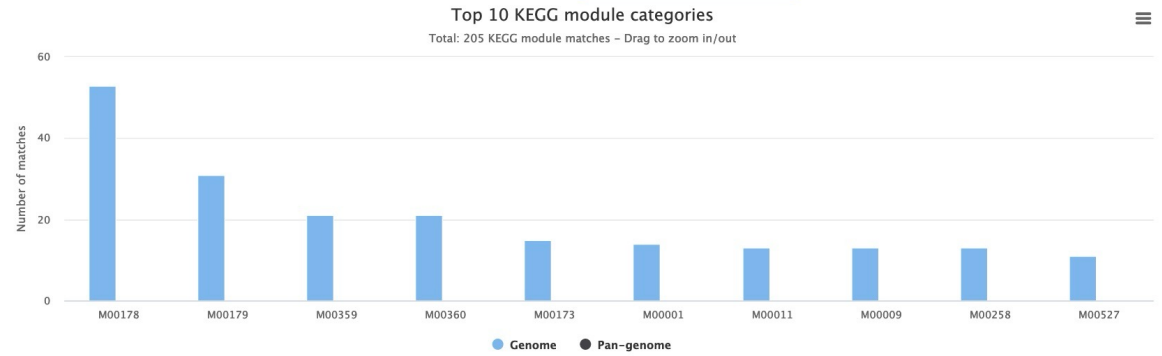
Home > Genomes > human-gut-v2-0-2 > MGYG000000526

Genome MGYG000000526

Type: MAG

Taxonomic lineage: Bacteria > Bacteroidota > Bacteroidia > Bacteroidales > Bacteroidaceae > Prevotella > Prevotella sp900770395

Overview | Browse genome | COG analysis | KEGG class analysis | **KEGG module analysis** | Downloads

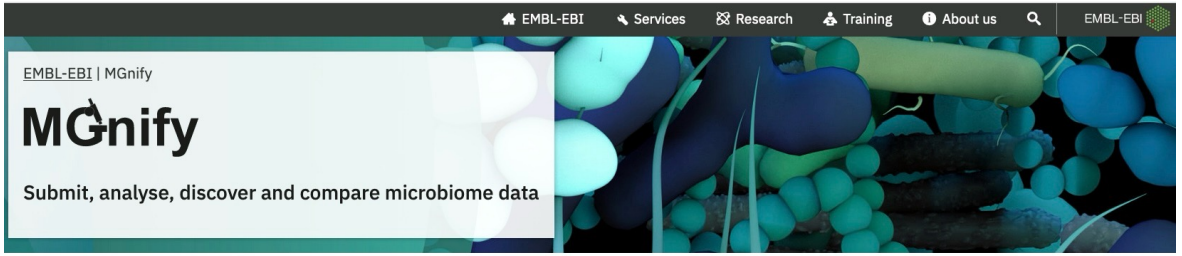


All 190 KEGG modules

Module ID	Description	Genome Count	Pan-genome count
M00178	Ribosome, bacteria	53	
M00179	Ribosome, archaea	31	
M00359		21	
M00360		21	
M00173	Reductive citrate cycle (Arnon-Buchanan cycle)	15	
M00001	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	14	
M00011	Citrate cycle, second carbon oxidation, 2-oxoglutarate => oxaloacetate	13	
M00009	Citrate cycle (TCA cycle, Krebs cycle)	13	
M00258		13	
M00527	Lysine biosynthesis, DAP aminotransferase pathway, aspartate => lysine	11	

Page Size: Show 10 | Previous 1 2 3 4 5 6 7 ... 18 19 Next

MGnify https://www.ebi.ac.uk/metagenomics/genomes/MGYG000000526



Overview | Submit data | Text search | Sequence search | Browse data | About | Help | Login

Home > Genomes > human-gut-v2-0-2 > MGYG000000526

Genome MGYG000000526

Type: MAG

Taxonomic lineage: Bacteria > Bacteroidota > Bacteroidia > Bacteroidales > Bacteroidaceae > Prevotella > Prevotella sp900770395

Overview | Browse genome | COG analysis | KEGG class analysis | KEGG module analysis | Downloads

Pan-genome analysis

Name	Compression	Format	Action
List of core genes in the entire pangenome	-	TAB	Download
Presence/absence binary matrix of the pan-genome across all conspecific genomes	-	TSV	Download
Tree generated from the pairwise Mash distances of conspecific genomes	-	Newick format	Download
DNA sequence FASTA file of the pangenome	-	FASTA	Download

Genome analysis

Name	Compression	Format	Action
All predicted CDS	-	FASTA	Download
DNA sequence FASTA file of the genome assembly of the species representative	-	FASTA	Download
DNA sequence FASTA file index of the genome assembly of the species representative	-	FAI	Download
Genome GFF file with various sequence annotations	-	GFF	Download
eggNOG annotations of the protein coding sequences	-	TSV	Download
InterProScan annotation of the protein coding sequences	-	TSV	Download

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).

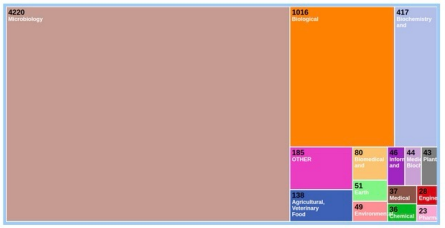
If you use IMG web resources or data to assist in research publications or proposals. Please cite: IMG - Chen et al., 2022 (Nucleic Acids Research, gkac976).
GOLD v9 - Mukherjee et al., 2022 (Nucleic Acids Research, gkac974).

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

IMG/M https://img.jgi.doe.gov/cgi-bin/m/main.cgi?section=TaxonList&page=isolateList&domain=Bacteria%20MAGs

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 Search History

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

Home > Find Genomes

14457 Loaded

IMG Survey

Bacteria - Metagenome-Assembled Genome

Data downloadには
アカウント登録が必要

Table Configuration

hint: ***Data Statistics with** [assembled, unassembled, both] means metagenomes counts or percentages only use assembled data or unassembled data or both (assembled data and unassembled data) for its calculations. This does not apply to isolates. The [assembled, unassembled, both] pick list is avaiable under the Table Configuration Data Statistics. **The default is assembled data.***

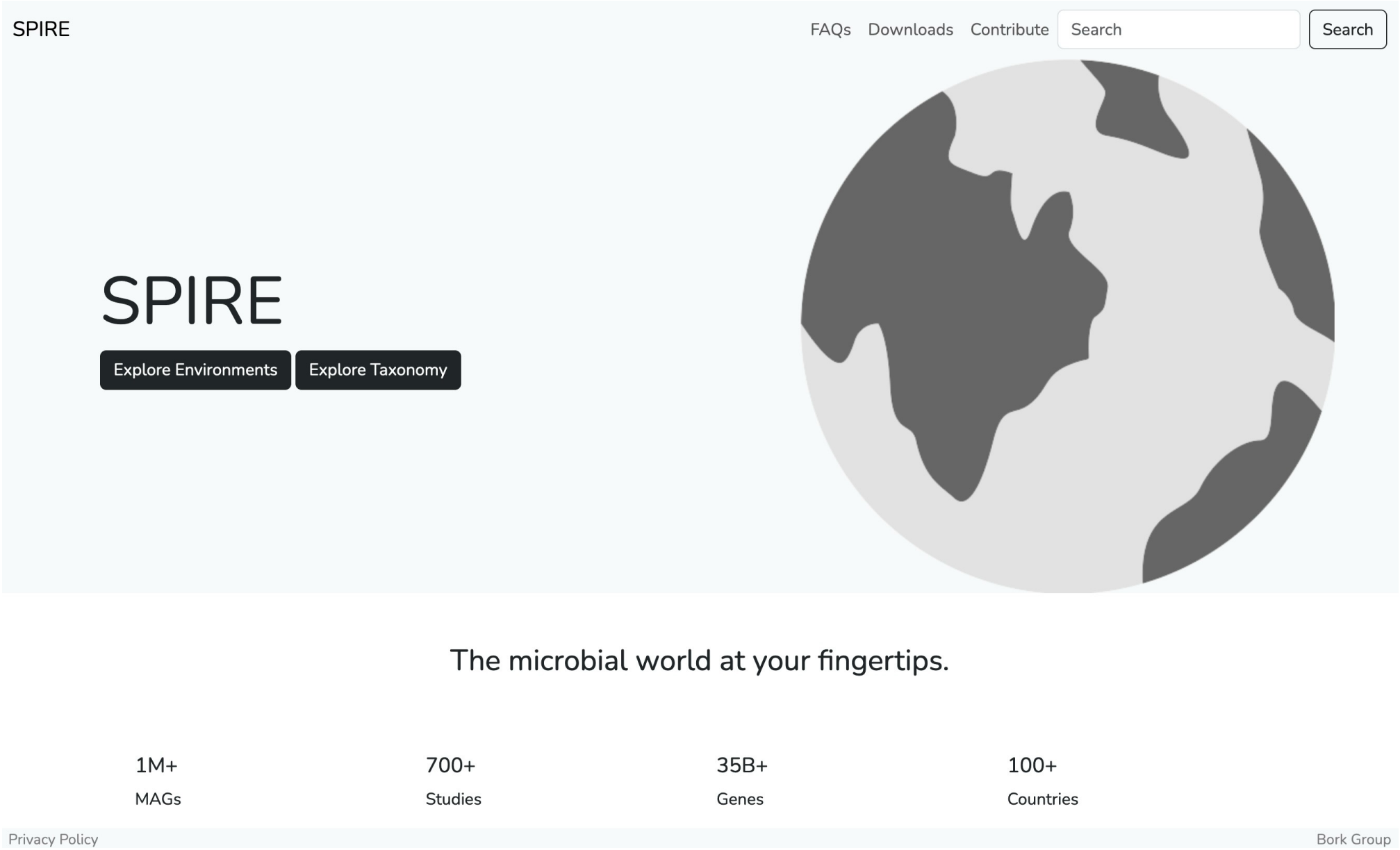
Add Selected to Genome Cart | **Display Selected on Map** | **View Phylogenetically**

Showing 1 to 10 of 14,457 entries

First Previous 1 2 3 4 5 6 7 8 9 10 ... 1,446 Next Last | Export | Select All | Clear All | Select - page | Deselect - page | Toggle | Column Selector | Show 10

Domain	Sequencing Status	Study Name	Genome Name / Sample Name	Sequencing Center	IMG Genome ID	Genome Size * assembled	Gene Count * assembled
Search Domain	Search Sequencing Status	Search Study Name	Search Genome Name / Si	Search Sequencing Center	Search IMG Genome ID	Search Genome Size * as:	Search Gene Count * asse
Bacteria	Permanent Draft	Candidatus Caldatribacterium californiense OP9-cSCG	Candidatus Caldatribacterium californiense OP9-cSCG	Unknown	2527291510	2080528	2534
Bacteria	Permanent Draft	Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage	SAR86 cluster bacterium SAR86A	Unknown	2597489919	1245342	1339
Bacteria	Permanent Draft	23 microbial genomes reconstructed from anaerobic bioreactors	Sulfurovum sp. G1	Beijing Genomics Institute (BGI)	2600254903	1781553	1864
Bacteria	Permanent Draft	Candidatus Cloacamonas acidaminovorans	Candidatus Cloacamonas acidaminovorans Evry	Unknown	642555115	2246820	1868
Bacteria	Permanent Draft	Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage	SAR86 cluster bacterium SAR86B	Unknown	2597489920	1679540	1890
Bacteria	Permanent Draft	CSP_777133	Candidatus Endomicrobia bacterium trichonymphae	Unknown	2014613004	3375854	4350
Bacteria	Permanent Draft	Uncultured BPS1	Uncultured Chlorobi bacterium Chlorobi5	Unknown	2061766010	343942	738
Bacteria	Permanent Draft	Uncultured Spirochaete Spiro3	Uncultured Spirochaete Spiro3	Unknown	2061766011	512975	1181
Bacteria	Permanent Draft	Red algae associated microbial communities from Porphyra blades in Schoodic Point, Maine, United States	Planctomycetes bacterium P1	DOE Joint Genome Institute (JGI)	2509276000	8468922	7084

SPIRE http://spire.embl.de/



SPIRE http://spire.embl.de/faqs

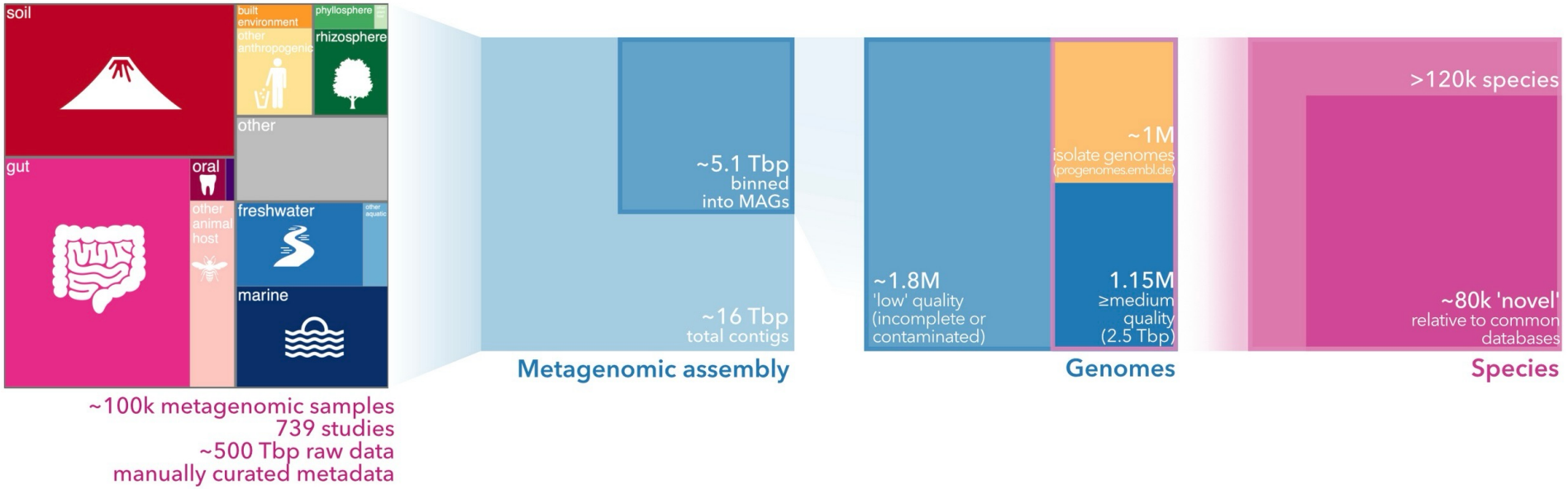
FAQs

What is SPIRE?

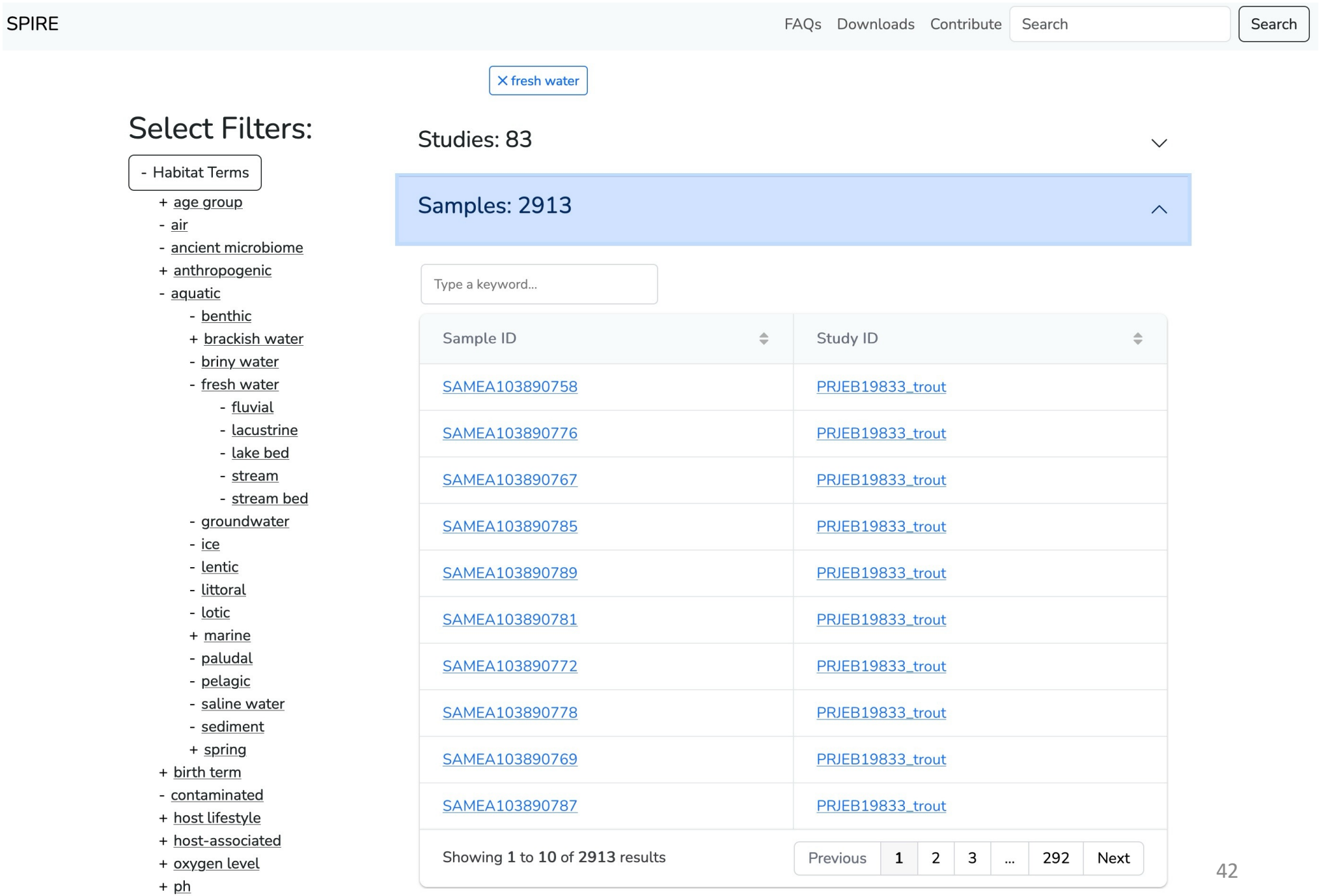
SPIRE is short for **S**earchable **P**lanetary-scale m**I**crobiome **RE**source: a one stop shop for microbial data, integrated and consistently processed across habitats and phylogeny, at global scales. SPIRE holds five main types of data, derived from ~100k metagenomic samples in 739 studies encompassing ~500Tbp of publicly available raw metagenomes:

- manually curated contextual data, including annotations against a custom microntology (See FAQ: What is microntology?)
- ~16Tbp worth of metagenomic assemblies (contigs)
- ~35 billion metagenomic open reading frames (ORFs) called from these contigs, functionally annotated in various ways 1.16 million metagenome-assembled genomes (MAGs) of medium or high quality, co-clustered with isolate genomes into ~120 thousand species-level clusters
- a custom mOTU database mOTUs to allow taxonomic profiling against these species-level clusters, as well as pre-computed taxonomic profiles (coming soon...)

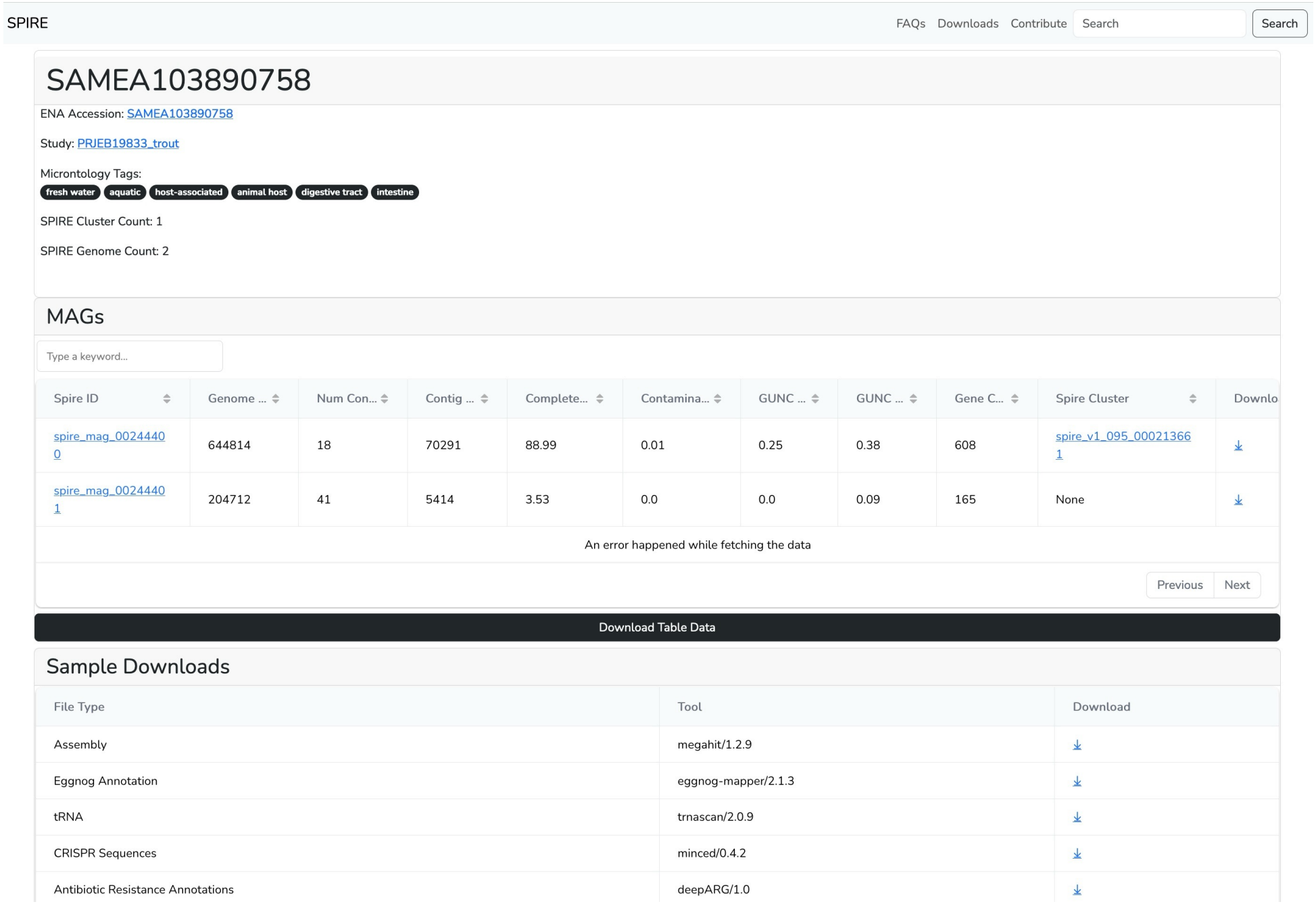
The aim of SPIRE is to provide an integrated view of microbial diversity at various taxonomic granularities, at genome and at gene level, across known microbial habitats.



SPIRE http://spire.embl.de/environment?microntology=&add=fresh+water&page=1



SPIRE http://spire.embl.de/sample/SAMEA103890758



SPIRE

FAQs Downloads Contribute Search Search

SAMEA103890758

ENA Accession: SAMEA103890758

Study: PRJEB19833_trout

Microntology Tags:

fresh water | aquatic | host-associated | animal host | digestive tract | intestine

SPIRE Cluster Count: 1

SPIRE Genome Count: 2

MAGs

Type a keyword...

Spire ID	Genome ...	Num Con...	Contig ...	Complete...	Contamina...	GUNC ...	GUNC ...	Gene C...	Spire Cluster	Downlo
spire_mag_00244400	644814	18	70291	88.99	0.01	0.25	0.38	608	spire_v1_095_000213661	
spire_mag_00244401	204712	41	5414	3.53	0.0	0.0	0.09	165	None	

An error happened while fetching the data

Previous Next

Download Table Data

Sample Downloads

File Type	Tool	Download
Assembly	megahit/1.2.9	
Eggnog Annotation	eggnog-mapper/2.1.3	
tRNA	trnascan/2.0.9	
CRISPR Sequences	minced/0.4.2	
Antibiotic Resistance Annotations	deepARG/1.0	

代表的なMAG DB間の比較

	Admin	Database URL	環境の多さ	プロジェクトの多さ	Number of MAGs in July 2024
NCBI Datasets	NCBI, USA	https://www.ncbi.nlm.nih.gov/datasets/genome/	未整理	大	443,763
IMG/M	JGI, USA	https://img.jgi.doe.gov/cgi-bin/m/main.cgi	小	小	25,507
SPIRE	EMBL, EU	http://spire.embl.de/	中	中	1,160,000
MGnify	EBI, EU	https://www.ebi.ac.uk/metagenomics	小	大	478,810
Microbiome Datahub	NIG, Japan	https://mdatahub.org/	大（整理中）	大	218,653

環境情報をいかに整理するかが各MAG DBの今後の課題