gcMeta Guidebook

User Manual for Global Catalogue of Metagenomics Database (gcMeta)

National Microbiology Data Center

Contents

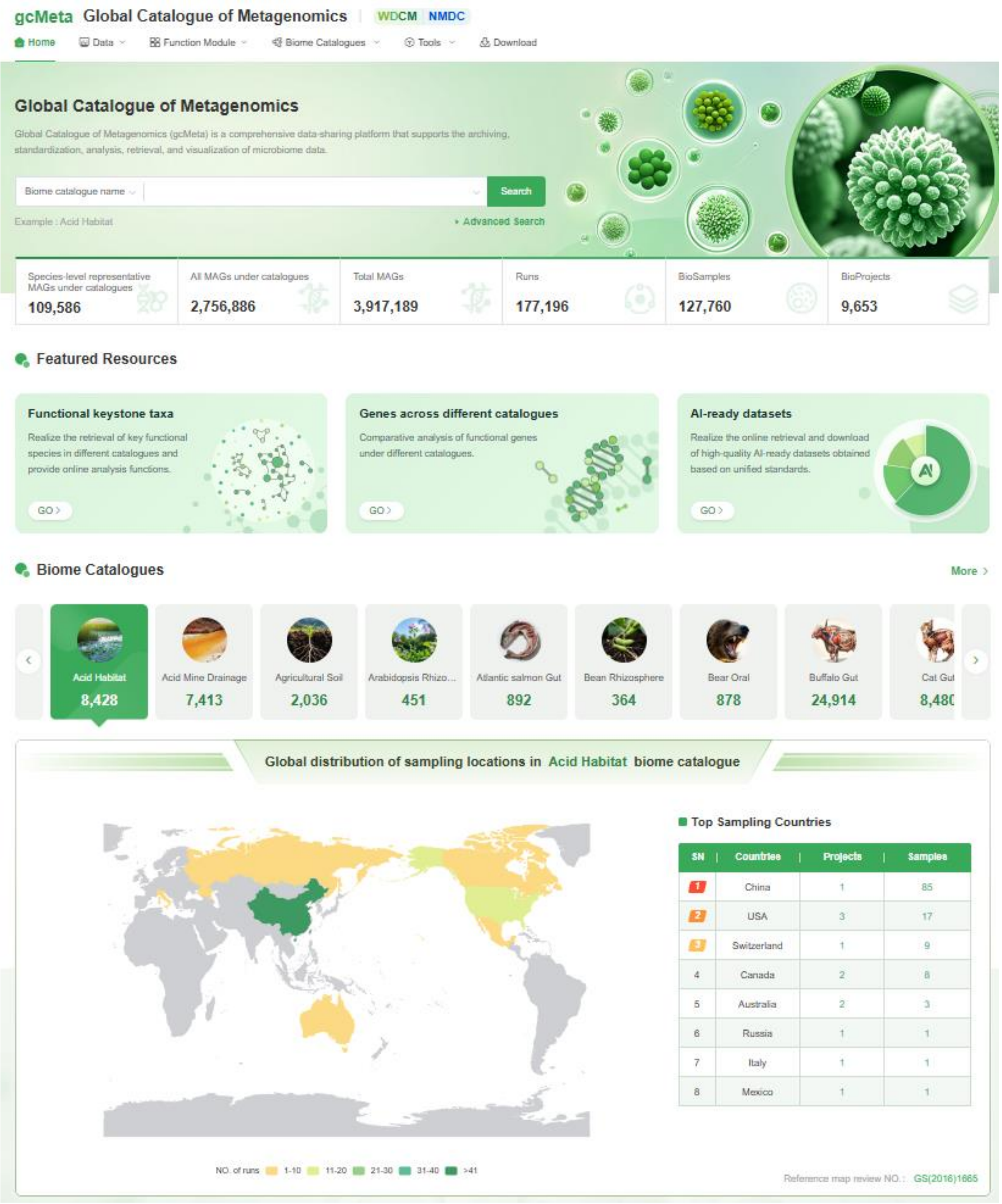
Introduction

The gcMeta database represents a comprehensive global resource for metagenome-assembled genomes (MAGs), integrating over 2.7 million MAGs from 104,266 samples covering more than 50 categories of biomes, covering human, animal, plant, marine, freshwater and extreme environments. The platform further enables cross-ecosystem functional comparisons, revealing niche-specific metabolic pathways and stress-response genes. Additionally, gcMeta provides AI-ready datasets with standardized annotations for microbial biogeochemical cycle metabolism enzymes, defense systems, and other functional modules, supporting machine learning applications. gcMeta also provides convenient tool for metagenome-assembled genome (MAG) analysis and offers customized analysis functions. This resource bridges microbial "sequence discovery" to "functional utilization", offering a foundation for ecological studies, industrial applications, and novel gene discovery.

gcMeta User Interface

Homepage

In the gcMeta homepage, we provide statistics and retrieval links of basic datasets, which mainly include number of total MAGs, species-level representative MAGs, as well as bioproject, biosample, and runs information associated with metagenomic samples. In the "Featured Resources" section, first two sections respectively enable the mining of functional keystone taxa and functional genes under different catalogues, as well as comparative analysis between catalogues. Meanwhile, various types of AI-ready datasets have been developed. Furthermore, "Biome Catalogues" provide access to the global distribution of sampling locations from each catalogue. Clicking on the corresponding catalogue allows access to the catalogue homepage.



Global Catalogue of Metagenomics

Global Catalogue of Metagenomics (gcMeta) is a comprehensive data-sharing platform that supports the archiving, standardization, analysis, retrieval, and visualization of microbiome data.

Species-level representative MAGs under catalogues	All MAGs under catalogues	Total MAGs	Runs	BioSamples	BioProjects
109,586	2,756,886	3,917,189	177,196	127,760	9,653

Functional keystone taxa

Realize the retrieval of key functional species in different catalogues and provide online analysis functions.

Genes across different catalogues

Comparative analysis of functional genes under different catalogues.

AI-ready datasets

Realize the online retrieval and download of high-quality AI-ready datasets obtained based on unified standards.

Global distribution of sampling locations in Acid Habitat biome catalogue

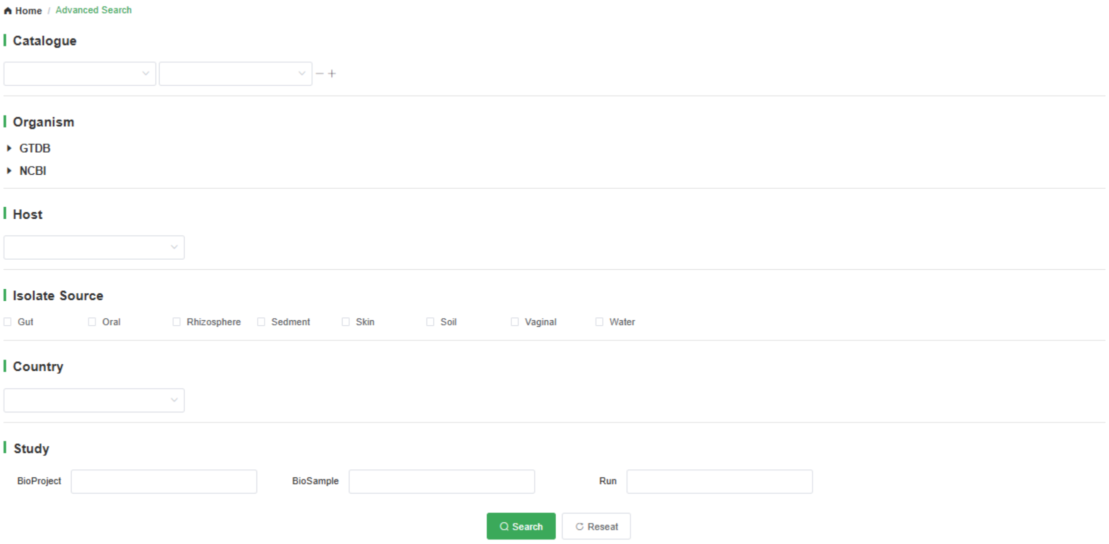
Top Sampling Countries

SN	Countries	Projects	Samples
1	China	1	85
2	USA	3	17
3	Switzerland	1	9
4	Canada	2	8
5	Australia	2	3
6	Russia	1	1
7	Italy	1	1
8	Mexico	1	1

Reference map review NO.: GS(2016)1665

Advanced Search

The ‘advanced search’ function allows users to conduct one or multiple criteria searches of manually curated metadata related to (i) catalogue, including catalogue group and catalogue name; (ii) Organism, including different species taxon levels of the GTDB or NCBI classification system and NCBI taxonomy ID; (iii) sample information, including host, isolate source, country, bioproject, biosample and run.

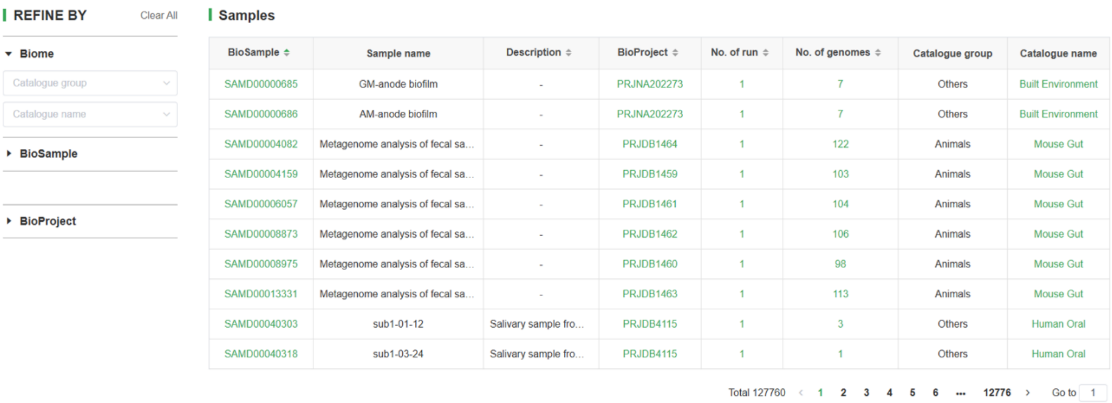


Data

Sample Information: Project, Samples, Runs

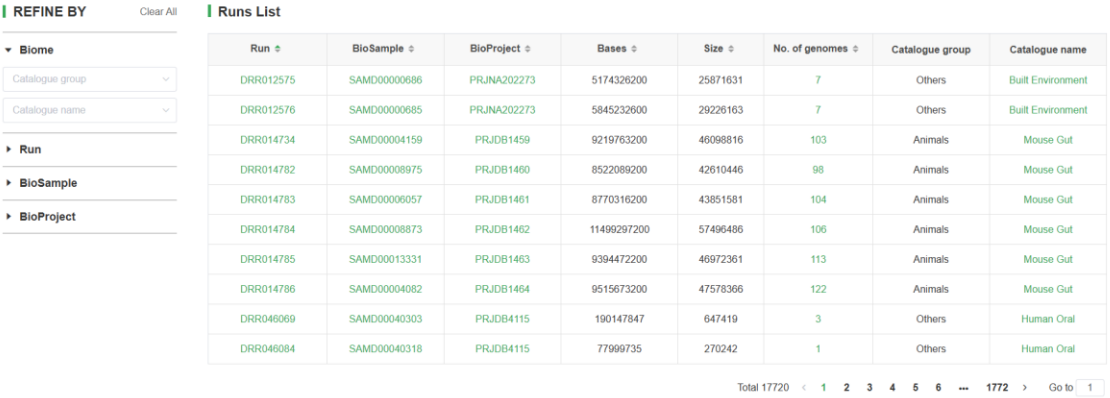
On the pages related to BioProject, BioSample, and Runs, genome information associated with them is provided. Filtering by Biome is supported, and mutual navigation between these pages is available.

BioProject	Project name	No. of BioSample	No. of run	No. of genomes	Catalogue group	Catalogue name
PRJDB10010	Metagenomic sequences of the Ofunato Bay in 2017	72	72	690	Marine	Marine Seawater
PRJDB10014	Metagenomic sequences of the Ofunato Bay in 2018	70	70	520	Marine	Marine Seawater
PRJDB10048	Metagenomic sequences of the Ofunato Bay in 2019	70	70	580	Marine	Marine Seawater
PRJDB10432	Shotgun metagenome sequencing of human saliva using HiSeq	1	1	9	Others	Human Oral
PRJDB11352	Taxonomic and functional specialization of the rumen microbiome of Japanese Black cattle revealed by next-generation sequencing	12	12	186	Large Livestock	Cattle Gut
PRJDB11444	Gut microbiome of prediabetes in Japan	290	290	14,012	Human	Human Gut
PRJDB11497	Shotgun metagenomic sequencing of the cryoconites collected from bipolar and Asian glaciers.	19	19	253	Others	Freshwater
PRJDB11511	Metagenome, metatranscriptome and 16S rRNA sequencing at multiple intestinal sites of common marmosets	6	36	25	Others	Host Associated
PRJDB12268	Arabidopsis thaliana metagenomic sequencing	36	36	225	Plant Rhizosphere	Arabidopsis Rhizosphere
PRJDB13503	Metagenome-assembled genomes from the rumen microbiome of cattle in Japan	9	9	66	Large Livestock	Cattle Gut



Samples

BioSample	Sample name	Description	BioProject	No. of run	No. of genomes	Catalogue group	Catalogue name
SAMD00000685	GM-anode biofilm	-	PRJNA202273	1	7	Others	Built Environment
SAMD00000686	AM-anode biofilm	-	PRJNA202273	1	7	Others	Built Environment
SAMD00004082	Metagenome analysis of fecal sa...	-	PRJDB1464	1	122	Animals	Mouse Gut
SAMD00004159	Metagenome analysis of fecal sa...	-	PRJDB1459	1	103	Animals	Mouse Gut
SAMD00006057	Metagenome analysis of fecal sa...	-	PRJDB1461	1	104	Animals	Mouse Gut
SAMD00008873	Metagenome analysis of fecal sa...	-	PRJDB1462	1	106	Animals	Mouse Gut
SAMD00008975	Metagenome analysis of fecal sa...	-	PRJDB1460	1	98	Animals	Mouse Gut
SAMD00013331	Metagenome analysis of fecal sa...	-	PRJDB1463	1	113	Animals	Mouse Gut
SAMD00040303	sub1-01-12	Salivary sample fro...	PRJDB4115	1	3	Others	Human Oral
SAMD00040318	sub1-03-24	Salivary sample fro...	PRJDB4115	1	1	Others	Human Oral



Runs List

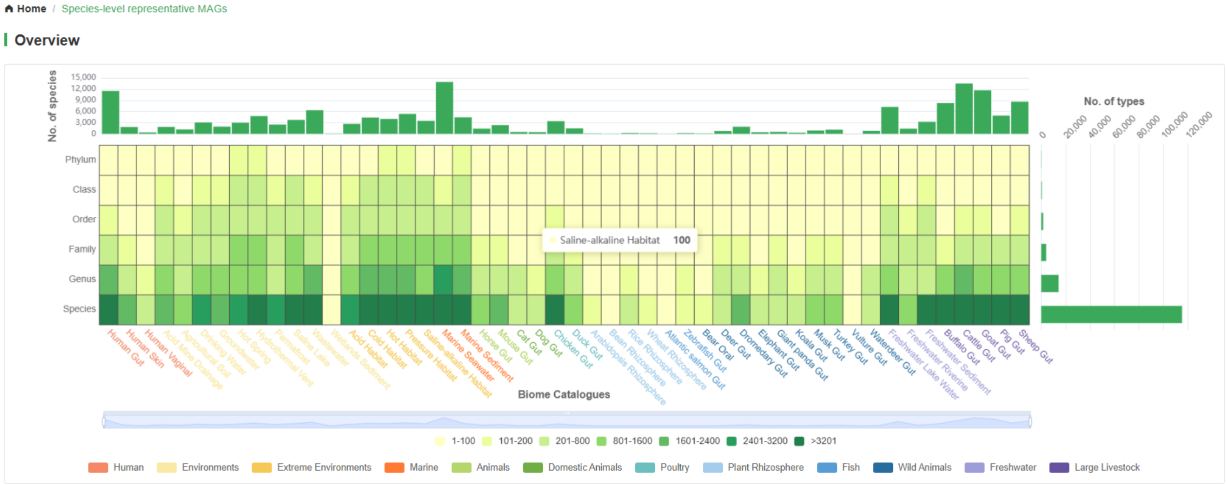
Run	BioSample	BioProject	Bases	Size	No. of genomes	Catalogue group	Catalogue name
DRR012575	SAMD00000686	PRJNA202273	5174326200	25871631	7	Others	Built Environment
DRR012576	SAMD00000685	PRJNA202273	5845232600	29226163	7	Others	Built Environment
DRR014734	SAMD00004159	PRJDB1459	9219763200	46098816	103	Animals	Mouse Gut
DRR014782	SAMD00008975	PRJDB1460	8522089200	42610446	98	Animals	Mouse Gut
DRR014783	SAMD00006057	PRJDB1461	8770316200	43851581	104	Animals	Mouse Gut
DRR014784	SAMD00008873	PRJDB1462	11499297200	57496486	106	Animals	Mouse Gut
DRR014785	SAMD00013331	PRJDB1463	9394472200	46972361	113	Animals	Mouse Gut
DRR014786	SAMD00004082	PRJDB1464	9515673200	47578366	122	Animals	Mouse Gut
DRR046069	SAMD00040303	PRJDB4115	190147847	647419	3	Others	Human Oral
DRR046084	SAMD00040318	PRJDB4115	77999735	270242	1	Others	Human Oral

Genome Information

Species-level representative MAGs under catalogues

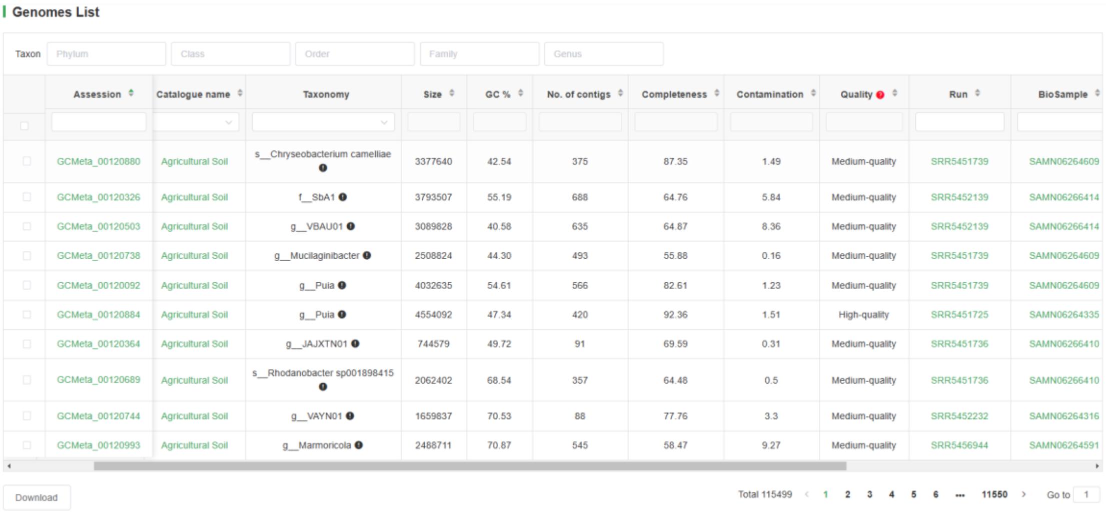
Overview

In the overview section, gcMeta provides statistics on species distribution across different taxonomy ranks for each catalogue. Hovering the mouse over the heatmap grid cells or bar chart bars allow viewing of specific values.



Genomes List

In the “Genomes List” section, queries for genomes of relevant species can be conducted by species taxonomic hierarchy or catalogue classification. Under Taxonomy, we also provide filtering options for Bacteria, Archaea, and novel species.

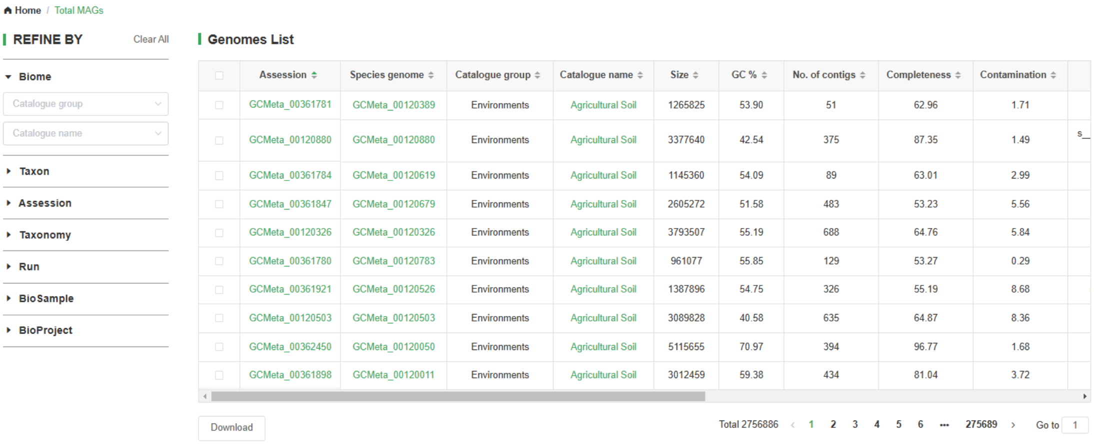


Genomes List

☐	Assession	Catalogue name	Taxonomy	Size	GC %	No. of contigs	Completeness	Contamination	Quality	Run	BioSample
☐	GCMeta_00120880	Agricultural Soil	s__Chryseobacterium camelliae	3377640	42.54	375	87.35	1.49	Medium-quality	SRR5451739	SAMN06264609
☐	GCMeta_00120326	Agricultural Soil	f__SbA1	3793507	55.19	688	64.76	5.84	Medium-quality	SRR5452139	SAMN06266414
☐	GCMeta_00120503	Agricultural Soil	g__VBAU01	3089828	40.58	635	64.87	8.36	Medium-quality	SRR5452139	SAMN06266414
☐	GCMeta_00120738	Agricultural Soil	g__Mucilaginibacter	2508824	44.30	493	55.88	0.16	Medium-quality	SRR5451739	SAMN06264609
☐	GCMeta_00120092	Agricultural Soil	g__Puia	4032635	54.61	566	82.61	1.23	Medium-quality	SRR5451739	SAMN06264609
☐	GCMeta_00120884	Agricultural Soil	g__Puia	4554092	47.34	420	92.36	1.51	High-quality	SRR5451725	SAMN06264335
☐	GCMeta_00120364	Agricultural Soil	g__JAJXTN01	744579	49.72	91	69.59	0.31	Medium-quality	SRR5451736	SAMN06266410
☐	GCMeta_00120689	Agricultural Soil	s__Rhodanobacter sp001898415	2062402	68.54	357	64.48	0.5	Medium-quality	SRR5451736	SAMN06266410
☐	GCMeta_00120744	Agricultural Soil	g__VAYN01	1659837	70.53	88	77.76	3.3	Medium-quality	SRR5452232	SAMN06264316
☐	GCMeta_00120993	Agricultural Soil	g__Marmoricola	2488711	70.87	545	58.47	9.27	Medium-quality	SRR5456944	SAMN06264591

All MAGs under catalogues and Total MAGs

In both sections, filtering is supported by Biome, Taxonomy, Accession, BioProject, BioSample, or Run. Users can view the basic information of all corresponding MAGs according to their needs, and the function of selecting and downloading the list is available.

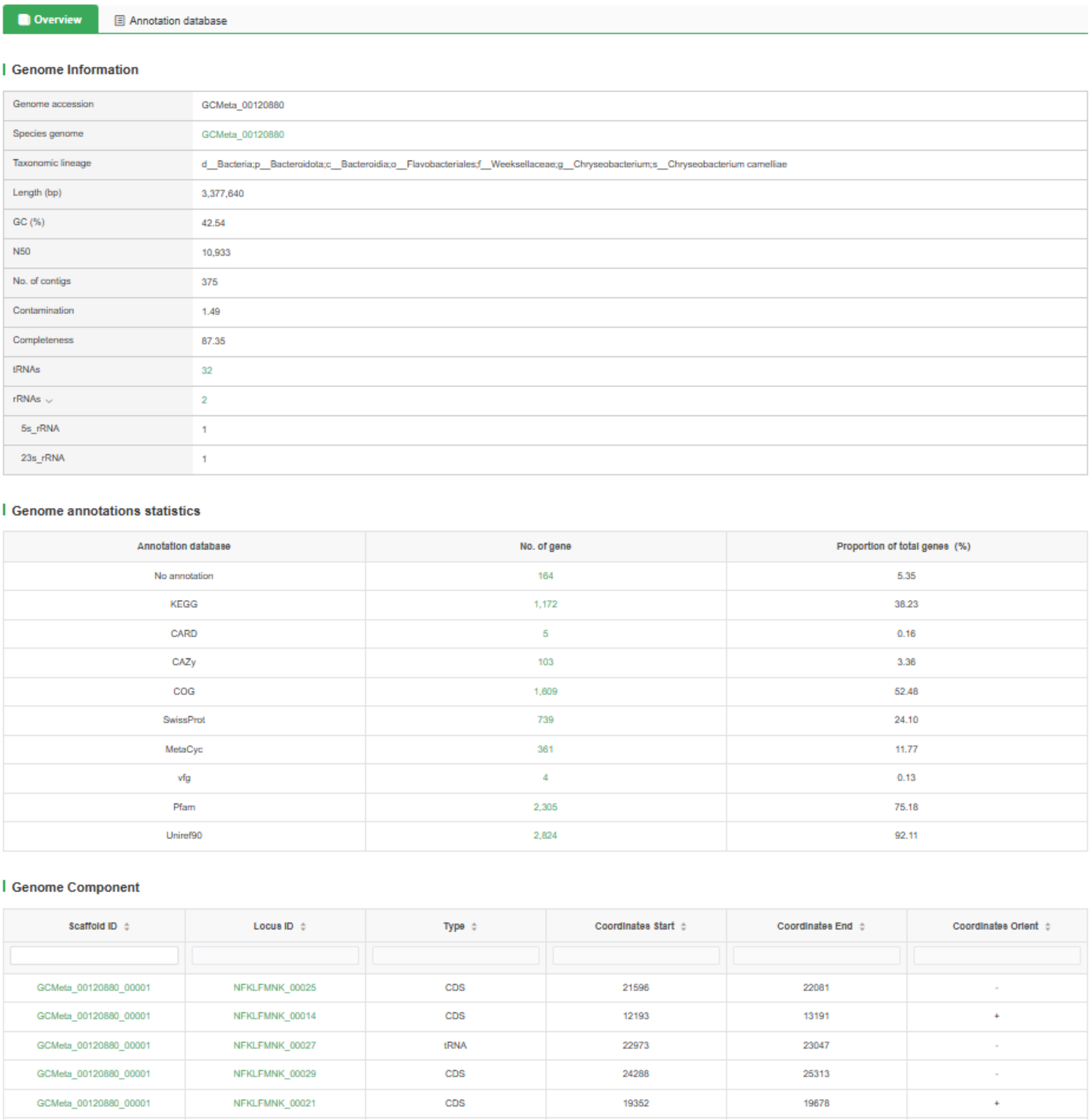


Genomes List

☐	Assession	Species genome	Catalogue group	Catalogue name	Size	GC %	No. of contigs	Completeness	Contamination	
☐	GCMeta_00361781	GCMeta_00120389	Environments	Agricultural Soil	1265825	53.90	51	62.96	1.71	
☐	GCMeta_00120880	GCMeta_00120880	Environments	Agricultural Soil	3377640	42.54	375	87.35	1.49	s__
☐	GCMeta_00361784	GCMeta_00120619	Environments	Agricultural Soil	1145360	54.09	89	63.01	2.99	
☐	GCMeta_00361847	GCMeta_00120679	Environments	Agricultural Soil	2605272	51.58	483	53.23	5.56	
☐	GCMeta_00120326	GCMeta_00120326	Environments	Agricultural Soil	3793507	55.19	688	64.76	5.84	
☐	GCMeta_00361780	GCMeta_00120783	Environments	Agricultural Soil	961077	55.85	129	53.27	0.29	
☐	GCMeta_00361921	GCMeta_00120526	Environments	Agricultural Soil	1387896	54.75	326	55.19	8.68	
☐	GCMeta_00120503	GCMeta_00120503	Environments	Agricultural Soil	3089828	40.58	635	64.87	8.36	
☐	GCMeta_00362450	GCMeta_00120050	Environments	Agricultural Soil	5115655	70.97	394	96.77	1.68	
☐	GCMeta_00361898	GCMeta_00120011	Environments	Agricultural Soil	3012459	59.38	434	81.04	3.72	

Genome details page

Regardless of which interface in gcMeta the genome id is clicked from, it will navigate to the detail page of that genome. This webpage is mainly divided into two sections. One is the “Overview”, which includes genome information, genome annotation statistics, and genome components; the other is the “Annotation database”, which contains the CDS distribution in the database and a list of annotation result categories, serving to display various functional annotation results of the genome. Each genome, including assembly information, the species cluster it belongs to, taxonomy, as well as the positional distribution of each gene on the genome and their functional annotation results, is recorded and displayed in this section.



Genome Information

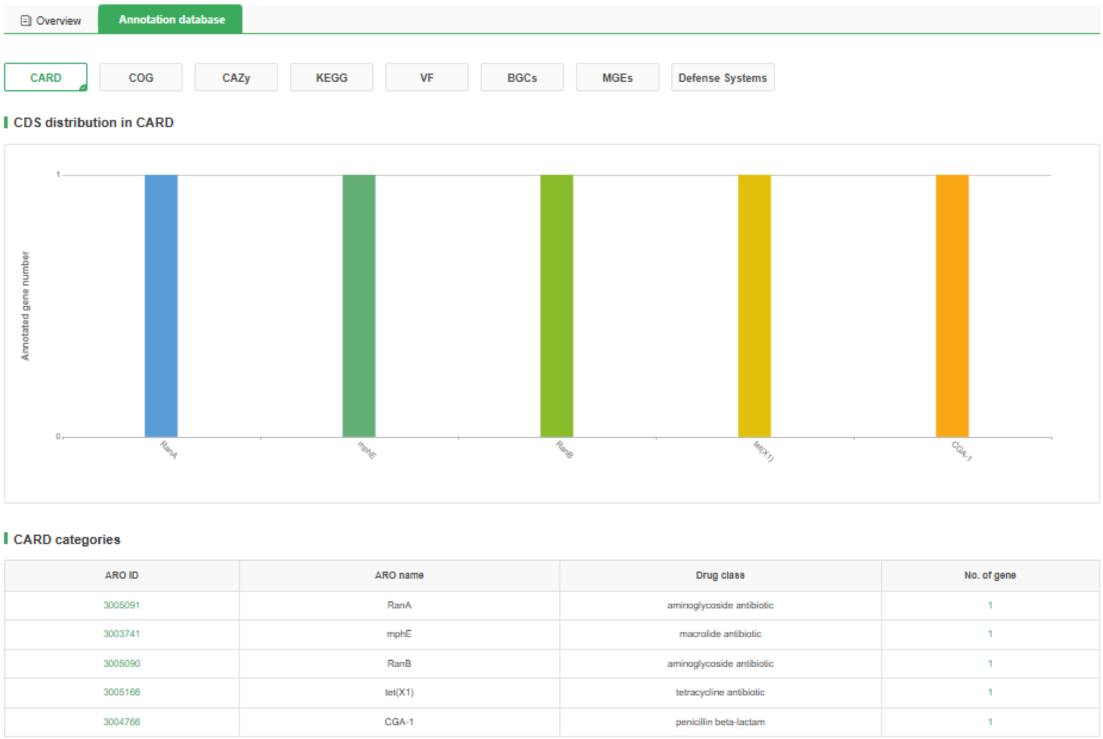
Genome accession	GCMeta_00120880
Species genome	GCMeta_00120880
Taxonomic lineage	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Weeksellaceae;g__Chryseobacterium;s__Chryseobacterium camelliae
Length (bp)	3,377,640
GC (%)	42.54
N50	10,933
No. of contigs	375
Contamination	1.49
Completeness	87.35
tRNAs	32
rRNAs	2
5s_rRNA	1
23s_rRNA	1

Genome annotations statistics

Annotation database	No. of gene	Proportion of total genes (%)
No annotation	164	5.35
KEGG	1,172	38.23
CARD	5	0.16
CAZy	103	3.36
COG	1,609	52.48
SwissProt	739	24.10
MetaCyc	361	11.77
vfg	4	0.13
Pfam	2,305	75.18
Uniref90	2,824	92.11

Genome Component

Scaffold ID	Locus ID	Type	Coordinates Start	Coordinates End	Coordinates Orient
GCMeta_00120880_00001	NFKLFMNK_00025	CDS	21596	22081	-
GCMeta_00120880_00001	NFKLFMNK_00014	CDS	12193	13191	+
GCMeta_00120880_00001	NFKLFMNK_00027	tRNA	22973	23047	-
GCMeta_00120880_00001	NFKLFMNK_00029	CDS	24288	25313	-
GCMeta_00120880_00001	NFKLFMNK_00021	CDS	19352	19678	+



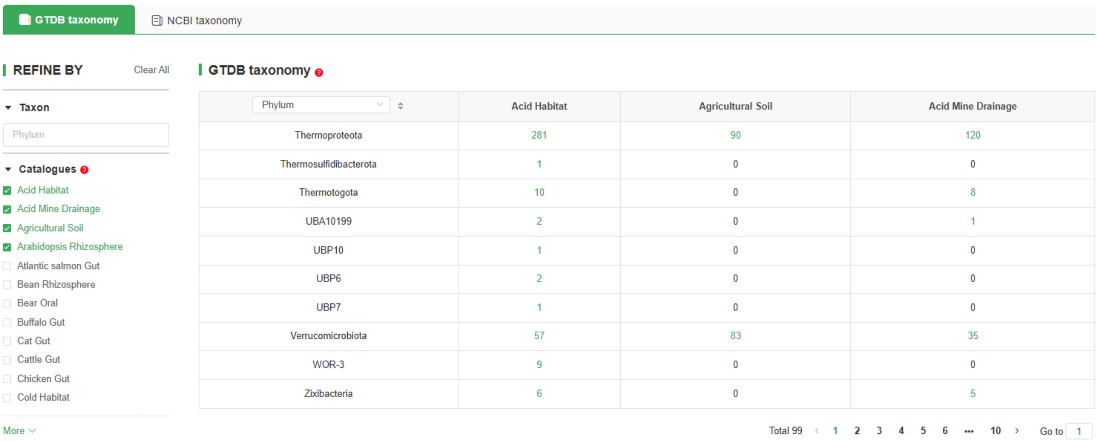
ARO ID	ARO name	Drug class	No. of gene
3005091	RanA	aminoglycoside antibiotic	1
3003741	mphE	macrolide antibiotic	1
3005090	RanB	aminoglycoside antibiotic	1
3005166	tet(X1)	tetracycline antibiotic	1
3004766	CGA-1	penicillin beta-lactam	1

Function Module

The "Function Module" provides users with a platform for free filtering and comparing species composition and functional diversity across different MAGs catalogues. Furthermore, users can also input terms to retrieve a specific taxon. In terms of functional comparison, this platform supports statistical comparison of 7 major functional categories, including ARGs, VFs, KEGG pathways, BGCs, CAZymes, MGEs, and defense systems. For each functional category, it provides comparisons of the distribution of functional genes based on different subgroups under that function, such as ARGs under different drug classes or metabolic genes under different pathway modules. Additionally, the platform enables taxonomic retrieval and statistical comparison of each specific functional gene or gene cluster.

Taxonomy Abundance

The "Taxonomy Abundance" supports retrieval based on either the GTDB or NCBI taxonomic system and allows filtering at various levels from phylum to genus.



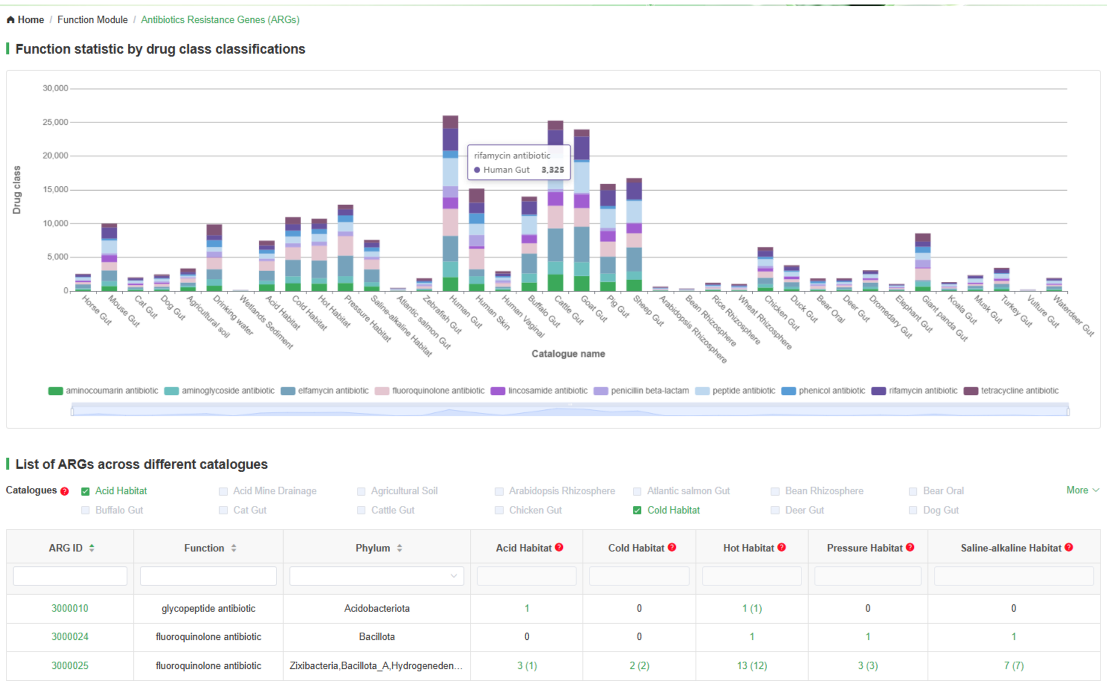
☑ Acid Habitat
☑ Acid Mine Drainage
☑ Agricultural Soil
☑ Arabidopsis Rhizosphere
☐ Atlantic salmon Gut
☐ Bean Rhizosphere
☐ Bear Oral
☐ Buffalo Gut
☐ Cat Gut
☐ Cattle Gut
☐ Chicken Gut
☐ Cold Habitat

GTDB taxonomy

Phylum	Acid Habitat	Agricultural Soil	Acid Mine Drainage
Thermoproteota	281	90	120
Thermosulfidibacterota	1	0	0
Thermotogota	10	0	8
UBA10199	2	0	1
UBP10	1	0	0
UBP6	2	0	0
UBP7	1	0	0
Verrucomicrobiota	57	83	35
WOR-3	9	0	0
Zixibacteria	6	0	5

Antibiotics Resistance Genes (ARGs)

Statistical comparison of the overall annotation results of antibiotic resistance genes (ARGs) across different catalogues. It supports the free selection of different catalogues to compare the distribution of species of each specific ARG across catalogues. By clicking on the numbers, users can obtain the corresponding list of species genomes. Detailed information can also be obtained by clicking on the "ARG ID".



Function statistic by drug class classifications

List of ARGs across different catalogues

Catalogues
☑ Acid Habitat ☐ Acid Mine Drainage ☐ Agricultural Soil ☐ Arabidopsis Rhizosphere ☐ Atlantic salmon Gut ☐ Bean Rhizosphere ☐ Bear Oral
☐ Buffalo Gut ☐ Cat Gut ☐ Cattle Gut ☐ Chicken Gut ☑ Cold Habitat ☐ Deer Gut ☐ Dog Gut

ARG ID	Function	Phylum	Acid Habitat	Cold Habitat	Hot Habitat	Pressure Habitat	Saline-alkaline Habitat
3000010	glycopeptide antibiotic	Acidobacteriota	1	0	1 (1)	0	0
3000024	fluoroquinolone antibiotic	Bacillota	0	0	1	1	1
3000025	fluoroquinolone antibiotic	Zixibacteria,Bacillota_A,Hydrogeneden...	3 (1)	2 (2)	13 (12)	3 (3)	7 (7)

KEGG Pathway

Statistical comparison of the overall annotation results of KO genes across different catalogues under KEGG orthology or modules. It supports the free selection of different catalogues to compare the distribution of species of each specific KO gene across catalogues. By clicking on the numbers, users can obtain the corresponding list of species genomes. Detailed information can also be obtained by clicking on the "KO catagory".

Home / Function Module / KEGG Pathway

Function statistic by KEGG classifications

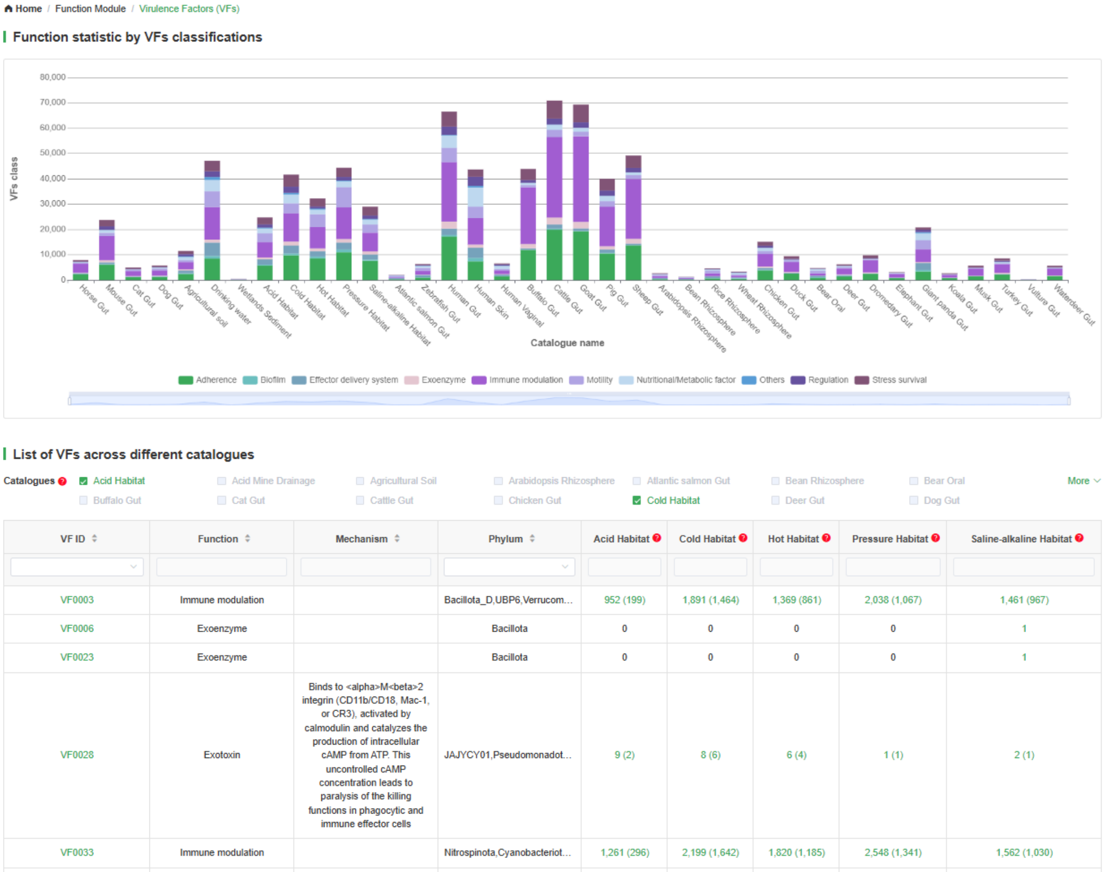
List of KO category across different catalogues

Catalogues ☑ Acid Habitat ☐ Acid Mine Drainage ☐ Agricultural Soil ☐ Arabidopsis Rhizosphere ☐ Atlantic salmon Gut ☐ Bean Rhizosphere ☐ Bear Oral More
☐ Buffalo Gut ☐ Cat Gut ☐ Cattle Gut ☐ Chicken Gut ☑ Cold Habitat ☐ Deer Gut ☐ Dog Gut

KO category	Description	Phylum	Acid Habitat	Cold Habitat	Hot Habitat	Pressure Habitat	Saline-alkaline Habitat
K00001	Glycolysis / Gluconeogenesis|Pyruvate metabolism|Fatty acid degradation|Tyrosine metabolism|Retinol metabolism|Chloroalkane and chloroalkene degradation|Naphthalene degradation|Metabolism of xenobiotics by cytochrome P450|Drug metabolism - cytochrome P450	Aquificota,Thermotogota,Methylomi...	149 (22)	277 (200)	205 (118)	195 (81)	293 (188)

Virulence Factors (VFs)

Statistical comparison of the overall annotation results of VF genes across different catalogues. It supports the free selection of different catalogues to compare the distribution of species of each specific VF gene across catalogues. By clicking on the numbers, users can obtain the corresponding list of species genomes. Detailed information can also be obtained by clicking on the "VF ID".



Function statistic by VFs classifications

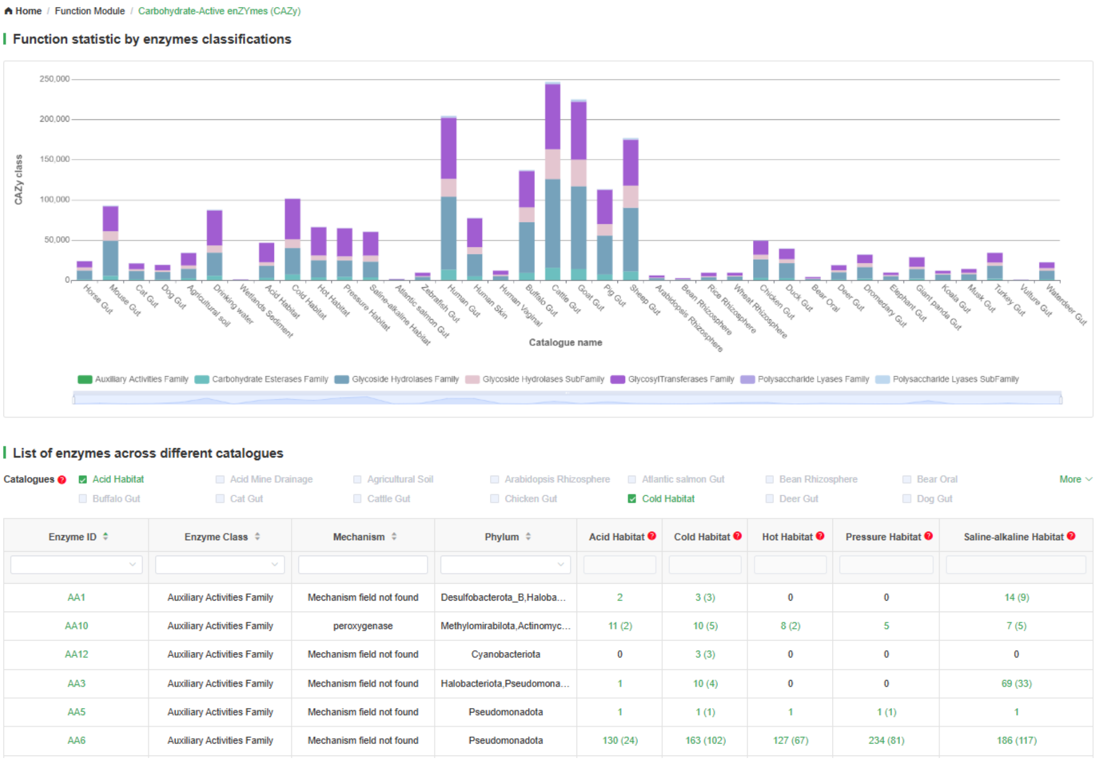
List of VFs across different catalogues

Catalogues
☑ Acid Habitat
☐ Acid Mine Drainage
☐ Agricultural Soil
☐ Arabidopsis Rhizosphere
☐ Atlantic salmon Gut
☐ Bean Rhizosphere
☐ Bear Oral
☐ Buffalo Gut
☐ Cat Gut
☐ Cattle Gut
☐ Chicken Gut
☑ Cold Habitat
☐ Deer Gut
☐ Dog Gut

VF ID	Function	Mechanism	Phylum	Acid Habitat	Cold Habitat	Hot Habitat	Pressure Habitat	Saline-alkaline Habitat
VF0003	Immune modulation		Bacillota_D,UBP6,Verrucom...	952 (199)	1,891 (1,464)	1,369 (861)	2,038 (1,067)	1,461 (967)
VF0006	Exoenzyme		Bacillota	0	0	0	0	1
VF0023	Exoenzyme		Bacillota	0	0	0	0	1
VF0028	Exotoxin	Binds to <alpha>M<beta>2 integrin (CD11b/CD18, Mac-1, or CR3), activated by calmodulin and catalyzes the production of intracellular cAMP from ATP. This uncontrolled cAMP concentration leads to paralysis of the killing functions in phagocytic and immune effector cells	JAJYCY01,Pseudomonadot...	9 (2)	8 (6)	6 (4)	1 (1)	2 (1)
VF0033	Immune modulation		Nitrospinota,Cyanobacteriot...	1,261 (296)	2,199 (1,642)	1,820 (1,185)	2,548 (1,341)	1,562 (1,030)

Carbohydrate-Active enZYmes (CAZy)

Statistical comparison of the overall annotation results of carbohydrate-active enzyme genes across different catalogues. It supports the free selection of different catalogues to compare the distribution of species of each specific enzyme gene across catalogues. By clicking on the numbers, users can obtain the corresponding list of species genomes. Detailed information can also be obtained by clicking on the "Enzyme ID".



Home / Function Module / Carbohydrate-Active enZYmes (CAZy)

Function statistic by enzymes classifications

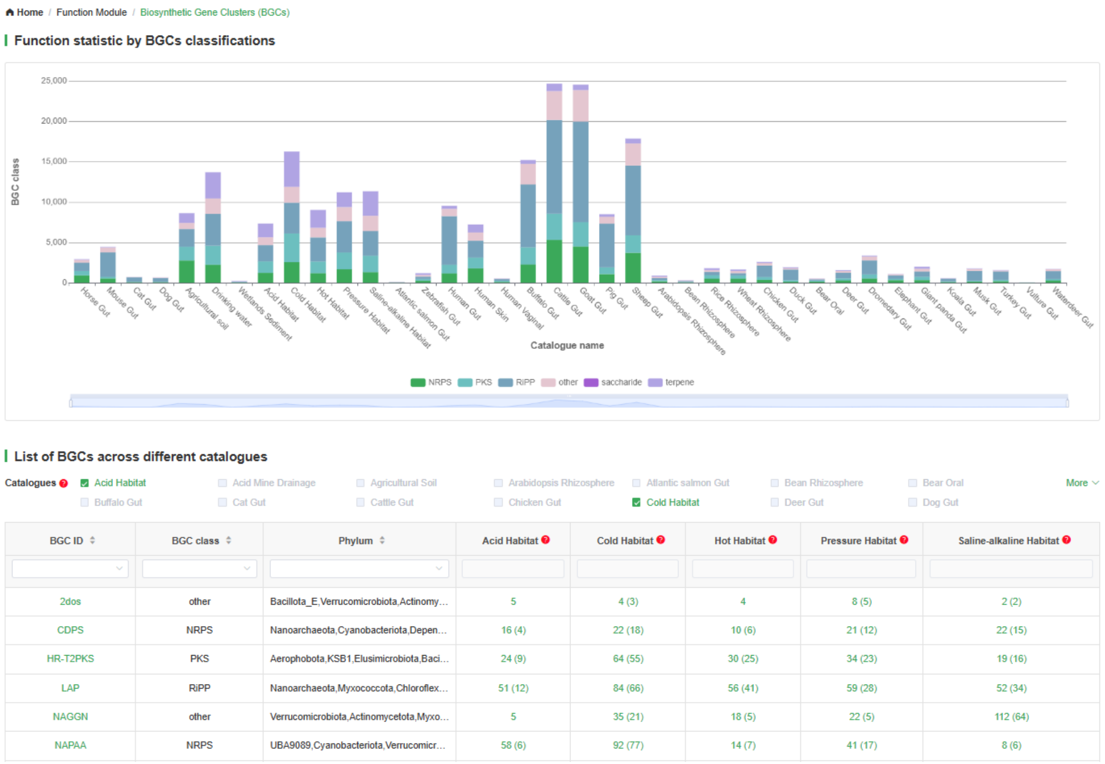
List of enzymes across different catalogues

Catalogues
☑ Acid Habitat ☐ Acid Mine Drainage ☐ Agricultural Soil ☐ Arabidopsis Rhizosphere ☐ Atlantic salmon Gut ☐ Bean Rhizosphere ☐ Bear Oral More
☐ Buffalo Gut ☐ Cat Gut ☐ Cattle Gut ☐ Chicken Gut ☑ Cold Habitat ☐ Deer Gut ☐ Dog Gut

Enzyme ID	Enzyme Class	Mechanism	Phylum	Acid Habitat	Cold Habitat	Hot Habitat	Pressure Habitat	Saline-alkaline Habitat
AA1	Auxiliary Activities Family	Mechanism field not found	Desulfobacterota_B,Haloba...	2	3 (3)	0	0	14 (9)
AA10	Auxiliary Activities Family	peroxygenase	Methylomirabilota,Actinomyc...	11 (2)	10 (5)	8 (2)	5	7 (5)
AA12	Auxiliary Activities Family	Mechanism field not found	Cyanobacteriota	0	3 (3)	0	0	0
AA3	Auxiliary Activities Family	Mechanism field not found	Halobacteriota,Pseudomona...	1	10 (4)	0	0	69 (33)
AA5	Auxiliary Activities Family	Mechanism field not found	Pseudomonadota	1	1 (1)	1	1 (1)	1
AA6	Auxiliary Activities Family	Mechanism field not found	Pseudomonadota	130 (24)	163 (102)	127 (67)	234 (81)	186 (117)

Biosynthetic Gene Clusters (BGCs)

Statistical comparison of the overall annotation results of BGCs across different catalogues. It supports the free selection of different catalogues to compare the distribution of species of each specific BGC across catalogues. By clicking on the numbers, users can obtain the corresponding list of species genomes. Detailed information can also be obtained by clicking on the "BGC ID".



Home / Function Module / Biosynthetic Gene Clusters (BGCs)

Function statistic by BGCs classifications

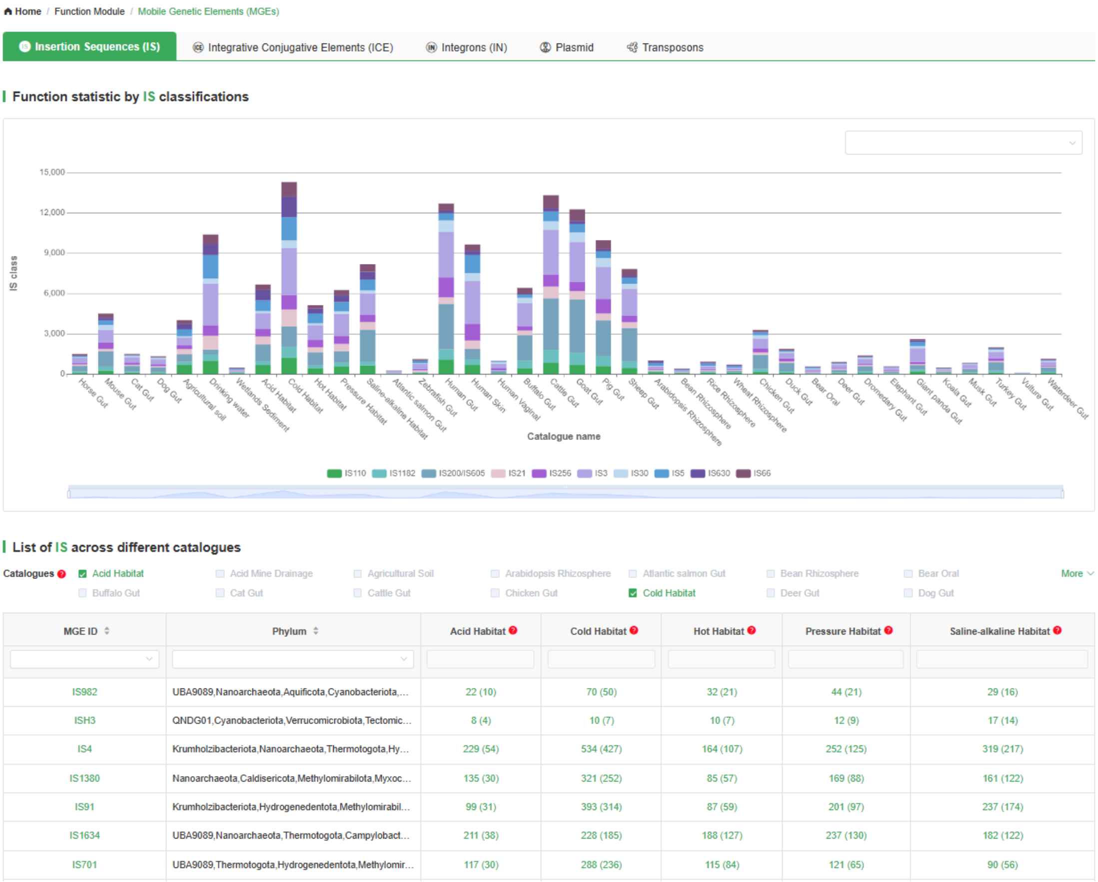
List of BGCs across different catalogues

Catalogues ☑ Acid Habitat ☐ Acid Mine Drainage ☐ Agricultural Soil ☐ Arabidopsis Rhizosphere ☐ Atlantic salmon Gut ☐ Bean Rhizosphere ☐ Bear Oral More
☐ Buffalo Gut ☐ Cat Gut ☐ Cattle Gut ☐ Chicken Gut ☑ Cold Habitat ☐ Deer Gut ☐ Dog Gut

BGC ID	BGC class	Phylum	Acid Habitat	Cold Habitat	Hot Habitat	Pressure Habitat	Saline-alkaline Habitat
2dos	other	Bacillota_E,Verrucomicrobiota,Actinomy…	5	4 (3)	4	8 (5)	2 (2)
CDPS	NRPS	Nanoarchaeota,Cyanobacteriota,Depen…	16 (4)	22 (18)	10 (6)	21 (12)	22 (15)
HR-T2PKS	PKS	Aerophobota,KSB1,Elusimicrobiota,Baci…	24 (9)	64 (55)	30 (25)	34 (23)	19 (16)
LAP	RiPP	Nanoarchaeota,Myxococcota,Chloroflex…	51 (12)	84 (66)	56 (41)	59 (28)	52 (34)
NAGGN	other	Verrucomicrobiota,Actinomycetota,Myxo…	5	35 (21)	18 (5)	22 (5)	112 (64)
NAPAA	NRPS	UBA9089,Cyanobacteriota,Verrucomicr…	58 (6)	92 (77)	14 (7)	41 (17)	8 (6)

Mobile Genetic Elements (MGEs)

Statistical comparison of the overall annotation results of MGEs (IS, ICE,IN, Plasmid and Transposons) across different catalogues. It supports the free selection of different catalogues to compare the distribution of species of each MGE across catalogues. By clicking on the numbers, users can obtain the corresponding list of species genomes. Detailed information can also be obtained by clicking on the "MGE ID".



Function statistic by IS classifications

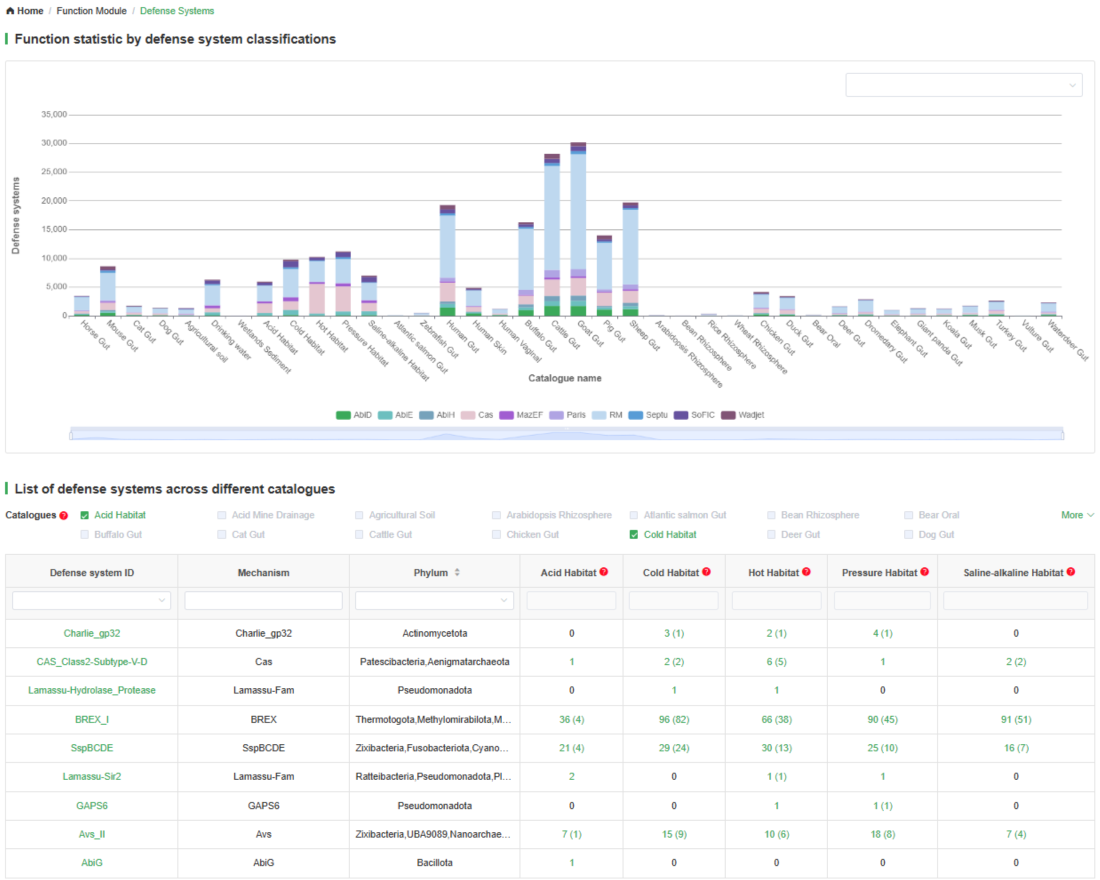
List of IS across different catalogues

Catalogues ☑ Acid Habitat ☐ Acid Mine Drainage ☐ Agricultural Soil ☐ Arabidopsis Rhizosphere ☐ Atlantic salmon Gut ☐ Bean Rhizosphere ☐ Bear Oral More
☐ Buffalo Gut ☐ Cat Gut ☐ Cattle Gut ☐ Chicken Gut ☑ Cold Habitat ☐ Deer Gut ☐ Dog Gut

MGE ID	Phylum	Acid Habitat	Cold Habitat	Hot Habitat	Pressure Habitat	Saline-alkaline Habitat
IS982	UBA9089,Nanoarchaeota,Aquificota,Cyanobacteriota,...	22 (10)	70 (50)	32 (21)	44 (21)	29 (16)
ISH3	QNDG01,Cyanobacteriota,Verrucomicrobiota,Tectomic...	8 (4)	10 (7)	10 (7)	12 (9)	17 (14)
IS4	Krumholzibacteriota,Nanoarchaeota,Thermotogota,Hy...	229 (54)	534 (427)	164 (107)	252 (125)	319 (217)
IS1380	Nanoarchaeota,Caldisericota,Methylomirabilota,Myxoc...	135 (30)	321 (252)	85 (57)	169 (88)	161 (122)
IS91	Krumholzibacteriota,Hydrogenedentota,Methylomirabil...	99 (31)	393 (314)	87 (59)	201 (97)	237 (174)
IS1634	UBA9089,Nanoarchaeota,Thermotogota,Campylobact...	211 (38)	228 (185)	188 (127)	237 (130)	182 (122)
IS701	UBA9089,Thermotogota,Hydrogenedentota,Methylomir...	117 (30)	288 (236)	115 (84)	121 (65)	90 (56)

Defense Systems

Statistical comparison of the overall annotation results of defense systems across different catalogues. It supports the free selection of different catalogues to compare the distribution of species of each specific defense system across catalogues. By clicking on the numbers, users can obtain the corresponding list of species genomes. Detailed information can also be obtained by clicking on the "Defense system ID".



Home / Function Module / Defense Systems

Function statistic by defense system classifications

List of defense systems across different catalogues

Catalogues
☑ Acid Habitat
☐ Acid Mine Drainage
☐ Agricultural Soil
☐ Arabidopsis Rhizosphere
☐ Atlantic salmon Gut
☐ Bean Rhizosphere
☐ Bear Oral
More
☐ Buffalo Gut
☐ Cat Gut
☐ Cattle Gut
☐ Chicken Gut
☑ Cold Habitat
☐ Deer Gut
☐ Dog Gut

Defense system ID	Mechanism	Phylum	Acid Habitat	Cold Habitat	Hot Habitat	Pressure Habitat	Saline-alkaline Habitat
Charlie_gp32	Charlie_gp32	Actinomycetota	0	3 (1)	2 (1)	4 (1)	0
CAS_Class2-Subtype-V-D	Cas	Patescibacteria,Aenigmatarchaeota	1	2 (2)	6 (5)	1	2 (2)
Lamassu-Hydrolase_Protease	Lamassu-Fam	Pseudomonadota	0	1	1	0	0
BREX_I	BREX	Thermotogota,Methylomirabilota,M…	36 (4)	96 (82)	66 (38)	90 (45)	91 (51)
SspBCDE	SspBCDE	Zixibacteria,Fusobacteriota,Cyano…	21 (4)	29 (24)	30 (13)	25 (10)	16 (7)
Lamassu-Sir2	Lamassu-Fam	Ratteibacteria,Pseudomonadota,Pl…	2	0	1 (1)	1	0
GAPS6	GAPS6	Pseudomonadota	0	0	1	1 (1)	0
Avs_II	Avs	Zixibacteria,UBA9089,Nanoarchae…	7 (1)	15 (9)	10 (6)	18 (8)	7 (4)
AbiG	AbiG	Bacillota	1	0	0	0	0

MAGs catalogues across different biomes

Each catalogue homepage is divided into four main sections, including an overview, genome list, taxonomy tree, and annotation.

Overview

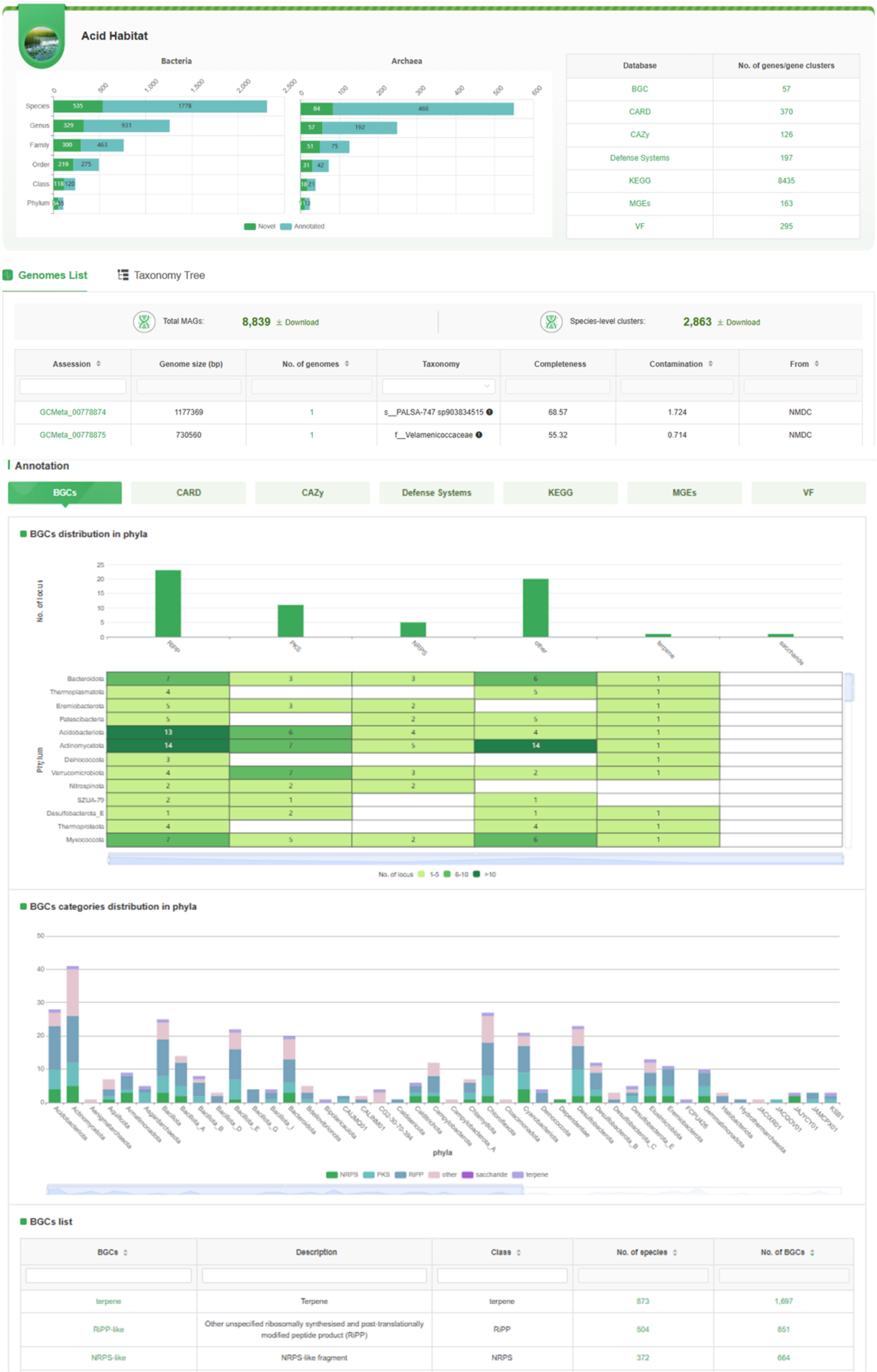
The overview section presents the distribution of annotated and novel species across different taxonomic ranks within the catalogue, as well as the overall annotation status of all species genomes in the catalogue. The numbers in the figures and tables link to the corresponding genome lists or detailed annotation result lists, respectively. Clicking on the database name in the table will navigate to the annotation section under the catalogue.

Genomes list

Genomes list section provides the distribution of genome counts within the catalogue, as well as basic information of species-level representative genomes, including assembly quality information, the number of genomes in the corresponding species cluster, taxonomy, and the source database of the genome. A GTDB-based taxonomy tree for each catalogue is available on the website enabling users to locate genomes from taxa of interest.

Annotation

The annotation section provides statistical displays of functional annotation results for all species-representative genomes in the catalogue, including the distribution of the number or different classes results at the phylum level, as well as detailed information for each annotation result. This includes descriptions of the annotation results, classifications, the species they originate from, and the specific gene loci involved. All numerical information herein can be clicked to navigate to the detailed list.



Acid Habitat

Database	No. of genes/gene clusters
BGC	57
CARD	370
CAZy	126
Defense Systems	197
KEGG	8435
MGEs	163
VF	295

Total MAGs: **8,839** ⤓ Download

Species-level clusters: **2,863** ⤓ Download

Assession ⇕	Genome size (bp)	No. of genomes ⇕	Taxonomy	Completeness	Contamination ⇕	From ⇕
GCMeta_00778874	1177369	1	s__PALSA-747 sp903834515 ❶	68.57	1.724	NMDC
GCMeta_00778875	730560	1	f__Velamenicoccaceae ❶	55.32	0.714	NMDC

Annotation

BGCs distribution in phyla

Phylum	RiPP	PKS	NRPS	other	terpene	saccharide
Bacteroidota	7	3	3	6	1	
Thermoplasmatota	4			5	1	
Eremiobacterota	5	3	2		1	
Patescibacteria	5		2	5	1	
Acidobacteriota	13	6	4	4	1	
Actinomycetota	14	7	5	14	1	
Deinococcota	3				1	
Verrucomicrobiota	4	7	3	2	1	
Nitrospinota	2	2	2			
SZUA-79	2	1		1		
Desulfobacterota_E	1	2		1	1	
Thermoproteota	4			4	1	
Myxococcota	7	5	2	6	1	

BGCs categories distribution in phyla

BGCs list

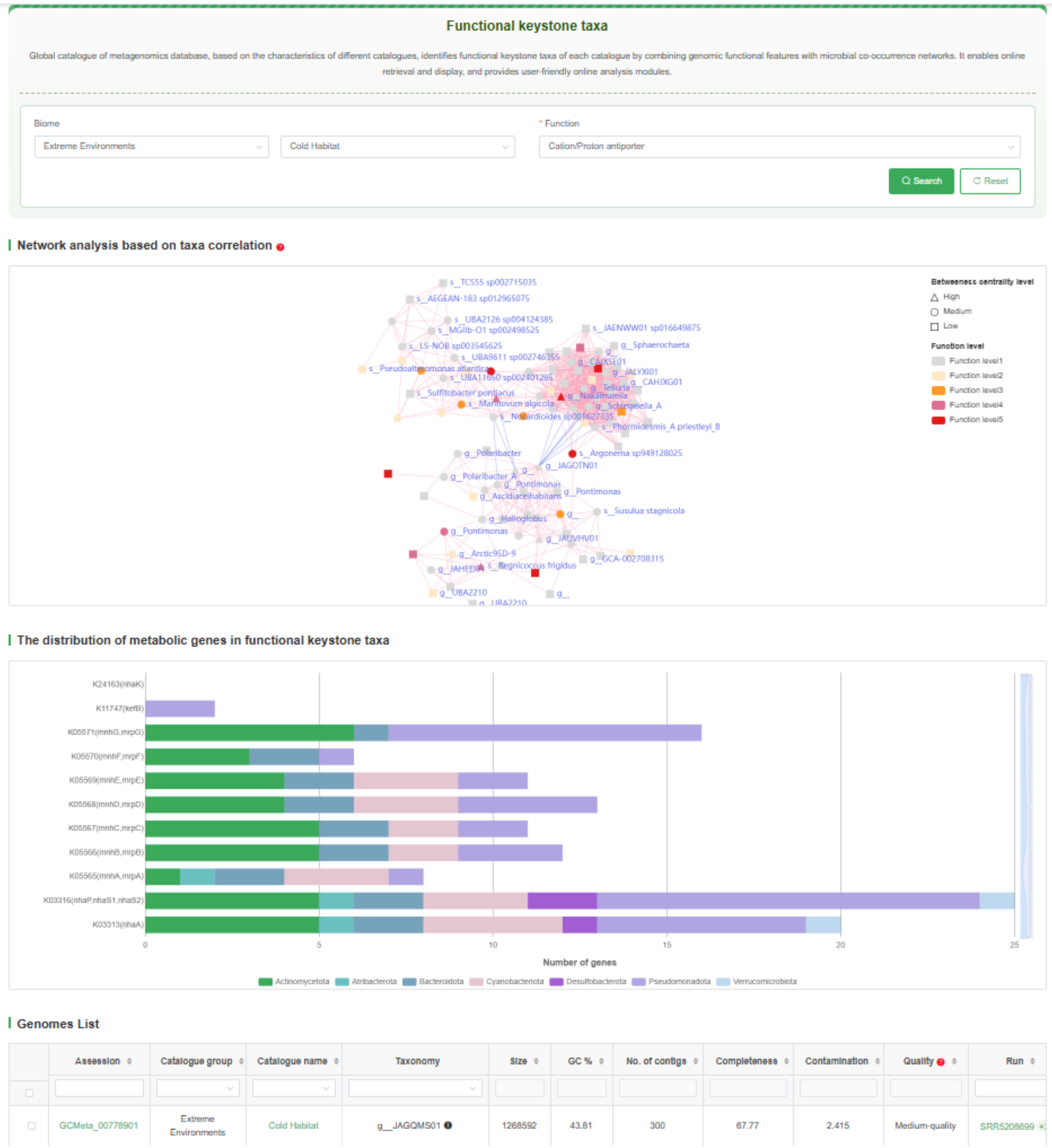
BGCs ⇕	Description	Class ⇕	No. of species ⇕	No. of BGCs ⇕
terpene	Terpene	terpene	873	1,697
RiPP-like	Other unspecified ribosomally synthesised and post-translationally modified peptide product (RiPP)	RiPP	504	851
NRPS-like	NRPS-like fragment	NRPS	372	664

Feature resources

Featured Resources comprises three main modules: functional keystone taxa, genes across different catalogues, and AI-ready datasets.

Functional keystone taxa

By integrating genomic features and co-occurrence networks, “Functional keystone taxa” aims to identify ecosystem-critical microorganisms and to reveal their roles in biogeochemical cycles (e.g., N/S metabolism) and environmental adaptation (e.g., salinity/heavy metal resistance), thereby enabling the transition from sequence discovery to functional utilization. This section includes interactive network analysis, which synchronously presents the distribution of key functional genes within taxa and the distribution of key species across different catalogues.



Functional keystone taxa

Global catalogue of metagenomics database, based on the characteristics of different catalogues, identifies functional keystone taxa of each catalogue by combining genomic functional features with microbial co-occurrence networks. It enables online retrieval and display, and provides user-friendly online analysis modules.

Biome: Extreme Environments, Cold Habitat

* Function: Cation/Proton antiporter

Network analysis based on taxa correlation

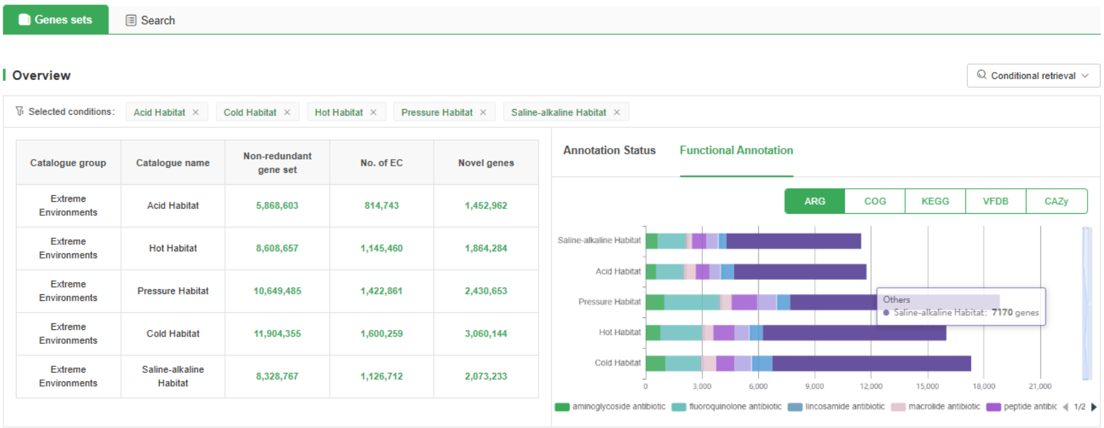
The distribution of metabolic genes in functional keystone taxa

Genomes List

☐	Assession	Catalogue group	Catalogue name	Taxonomy	Size	GC %	No. of contigs	Completeness	Contamination	Quality	Run
☐	GCMeta_00778901	Extreme Environments	Cold Habitat	g__JAGQMS01	1268592	43.81	300	67.77	2.415	Medium-quality	SRR5208699

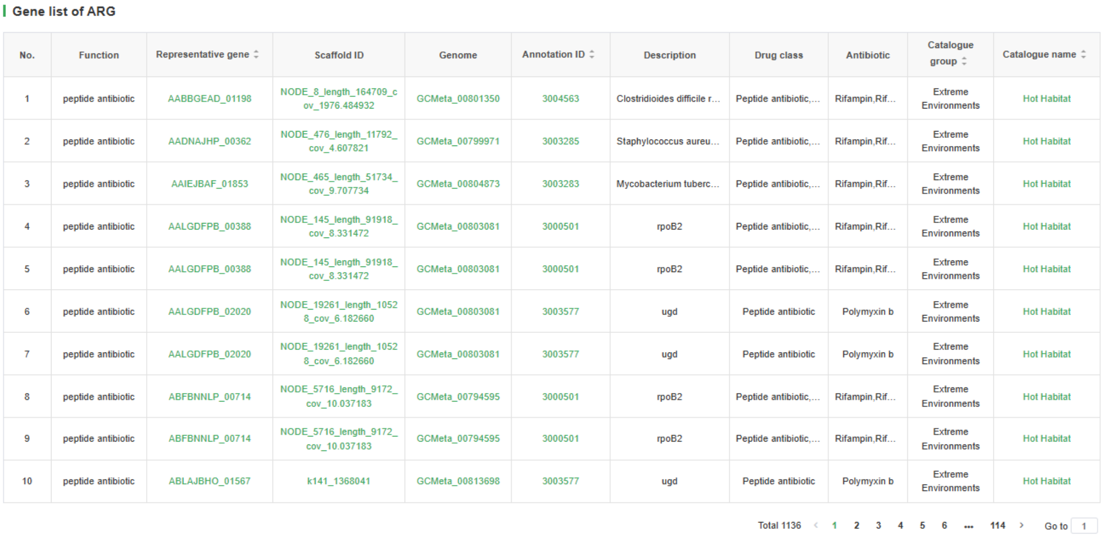
Genes across different catalogues

In the “Genes across different catalogues” section, we have enabled analysis of the distribution of functional genes in non-redundant gene sets. Users can perform statistical comparisons of non-redundant gene sets across different catalogues via the "conditional retrieval" function under the overview of "Gene sets".



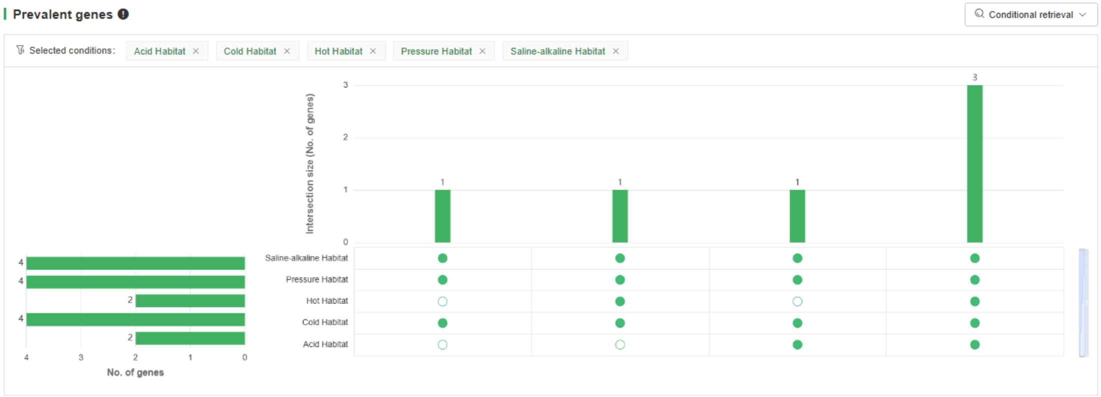
Catalogue group	Catalogue name	Non-redundant gene set	No. of EC	Novel genes
Extreme Environments	Acid Habitat	5,868,603	814,743	1,452,962
Extreme Environments	Hot Habitat	8,608,657	1,145,460	1,864,284
Extreme Environments	Pressure Habitat	10,649,485	1,422,861	2,430,653
Extreme Environments	Cold Habitat	11,904,355	1,600,259	3,060,144
Extreme Environments	Saline-alkaline Habitat	8,328,767	1,126,712	2,073,233

These comparisons include the distribution of functional enzymes and novel genes, as well as their annotation results in various functional databases. With a single click, users can obtain a detailed list of genes.


Gene list of ARG

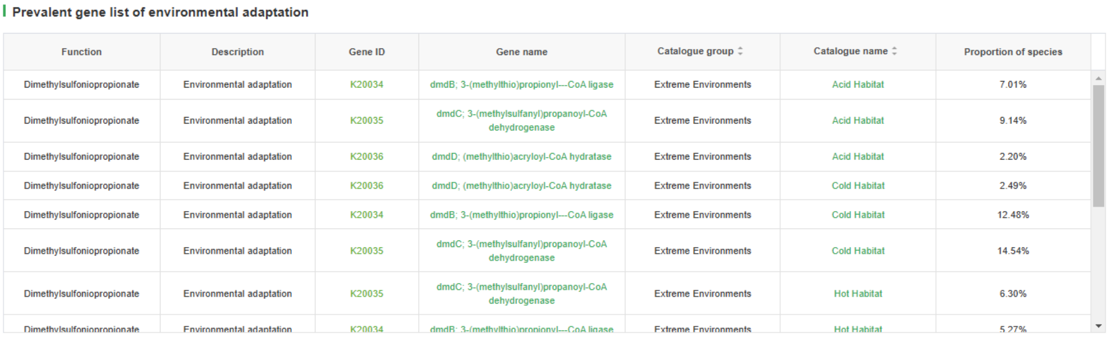
No.	Function	Representative gene	Scaffold ID	Genome	Annotation ID	Description	Drug class	Antibiotic	Catalogue group	Catalogue name
1	peptide antibiotic	AABBGEAD_01198	NODE_8_length_164709_cov_1976.484932	GCMeta_00801350	3004563	Clostridioides difficile r...	Peptide antibiotic,...	Rifampin,Rif...	Extreme Environments	Hot Habitat
2	peptide antibiotic	AADNAJHP_00362	NODE_476_length_11792_cov_4.607821	GCMeta_00799971	3003285	Staphylococcus aureu...	Peptide antibiotic,...	Rifampin,Rif...	Extreme Environments	Hot Habitat
3	peptide antibiotic	AAIEJBAF_01853	NODE_465_length_51734_cov_9.707734	GCMeta_00804873	3003283	Mycobacterium tuberc...	Peptide antibiotic,...	Rifampin,Rif...	Extreme Environments	Hot Habitat
4	peptide antibiotic	AALGDFPB_00388	NODE_145_length_91918_cov_8.331472	GCMeta_00803081	3000501	rpoB2	Peptide antibiotic,...	Rifampin,Rif...	Extreme Environments	Hot Habitat
5	peptide antibiotic	AALGDFPB_00388	NODE_145_length_91918_cov_8.331472	GCMeta_00803081	3000501	rpoB2	Peptide antibiotic,...	Rifampin,Rif...	Extreme Environments	Hot Habitat
6	peptide antibiotic	AALGDFPB_02020	NODE_19261_length_10528_cov_6.182660	GCMeta_00803081	3003577	ugd	Peptide antibiotic	Polymyxin b	Extreme Environments	Hot Habitat
7	peptide antibiotic	AALGDFPB_02020	NODE_19261_length_10528_cov_6.182660	GCMeta_00803081	3003577	ugd	Peptide antibiotic	Polymyxin b	Extreme Environments	Hot Habitat
8	peptide antibiotic	ABFBNNLP_00714	NODE_5716_length_9172_cov_10.037183	GCMeta_00794595	3000501	rpoB2	Peptide antibiotic,...	Rifampin,Rif...	Extreme Environments	Hot Habitat
9	peptide antibiotic	ABFBNNLP_00714	NODE_5716_length_9172_cov_10.037183	GCMeta_00794595	3000501	rpoB2	Peptide antibiotic,...	Rifampin,Rif...	Extreme Environments	Hot Habitat
10	peptide antibiotic	ABLAJBHO_01567	k141_1368041	GCMeta_00813698	3003577	ugd	Peptide antibiotic	Polymyxin b	Extreme Environments	Hot Habitat

Meanwhile, based on core species, we have achieved cross-ecosystem functional gene mining and comparative analysis from multiple perspectives. Users are allowed to freely select interested catalogues, metabolic pathways or functional genes under the "conditional retrieval" of "Prevalent genes", realizing the comparison of shared and unique genes among the core species under these catalogues, and can obtain the detailed list of genes by clicking. This provides a systematic reference for revealing the diversity of microbial ecological adaptive functions, as well as for the discovery of novel biological functional elements and the application of unique functional genes.



Prevalent genes

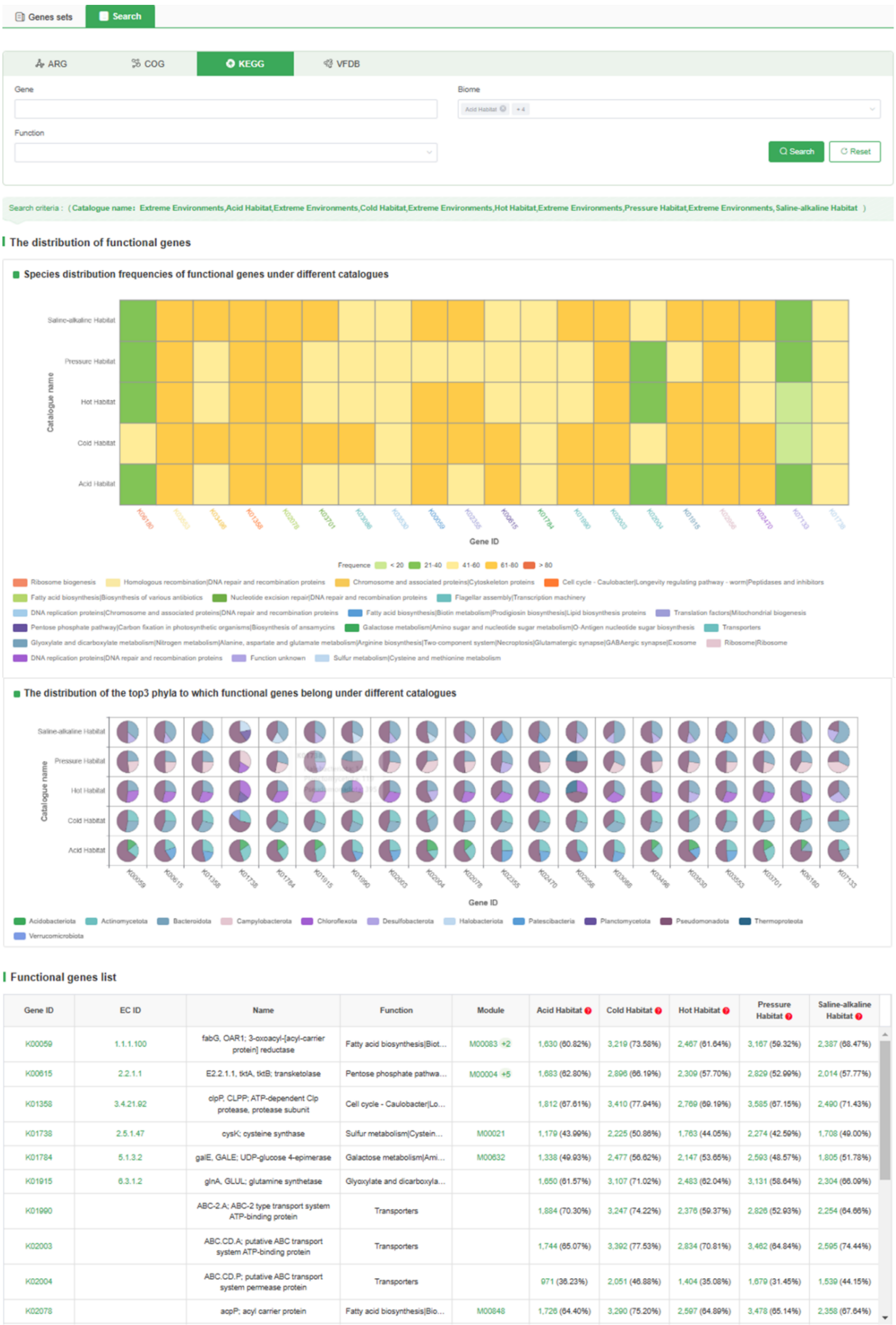
Selected conditions: Acid Habitat × Cold Habitat × Hot Habitat × Pressure Habitat × Saline-alkaline Habitat ×



Prevalent gene list of environmental adaptation

Function	Description	Gene ID	Gene name	Catalogue group	Catalogue name	Proportion of species
Dimethylsulfoniopropionate	Environmental adaptation	K20034	dmdB; 3-(methylthio)propionyl---CoA ligase	Extreme Environments	Acid Habitat	7.01%
Dimethylsulfoniopropionate	Environmental adaptation	K20035	dmdC; 3-(methylsulfanyl)propanoyl-CoA dehydrogenase	Extreme Environments	Acid Habitat	9.14%
Dimethylsulfoniopropionate	Environmental adaptation	K20036	dmdD; (methylthio)acryloyl-CoA hydratase	Extreme Environments	Acid Habitat	2.20%
Dimethylsulfoniopropionate	Environmental adaptation	K20036	dmdD; (methylthio)acryloyl-CoA hydratase	Extreme Environments	Cold Habitat	2.49%
Dimethylsulfoniopropionate	Environmental adaptation	K20034	dmdB; 3-(methylthio)propionyl---CoA ligase	Extreme Environments	Cold Habitat	12.48%
Dimethylsulfoniopropionate	Environmental adaptation	K20035	dmdC; 3-(methylsulfanyl)propanoyl-CoA dehydrogenase	Extreme Environments	Cold Habitat	14.54%
Dimethylsulfoniopropionate	Environmental adaptation	K20035	dmdC; 3-(methylsulfanyl)propanoyl-CoA dehydrogenase	Extreme Environments	Hot Habitat	6.30%
Dimethylsulfoniopropionate	Environmental adaptation	K20034	dmdB; 3-(methylthio)propionyl---CoA ligase	Extreme Environments	Hot Habitat	5.27%

In addition, the "Search" function provides personalized retrieval of different catalogues and different functional genes, enabling the comparison of species distribution of functional genes across catalogues.



Search criteria : （Catalogue name： Extreme Environments,Acid Habitat,Extreme Environments,Cold Habitat,Extreme Environments,Hot Habitat,Extreme Environments,Pressure Habitat,Extreme Environments,Saline-alkaline Habitat ）

The distribution of functional genes

Species distribution frequencies of functional genes under different catalogues

The distribution of the top3 phyla to which functional genes belong under different catalogues

Functional genes list

Gene ID	EC ID	Name	Function	Module	Acid Habitat	Cold Habitat	Hot Habitat	Pressure Habitat	Saline-alkaline Habitat
K00059	1.1.1.100	fabG, OAR1; 3-oxoacyl-[acyl-carrier protein] reductase	Fatty acid biosynthesis|Biot...	M00083 +2	1,630 (60.82%)	3,219 (73.58%)	2,467 (61.64%)	3,167 (59.32%)	2,387 (68.47%)
K00615	2.2.1.1	E2.2.1.1, tktA, tktB; transketolase	Pentose phosphate pathwa...	M00004 +5	1,683 (62.80%)	2,896 (66.19%)	2,309 (57.70%)	2,829 (52.99%)	2,014 (57.77%)
K01358	3.4.21.92	clpP, CLPP; ATP-dependent Clp protease, protease subunit	Cell cycle - Caulobacter|Lo...		1,812 (67.61%)	3,410 (77.94%)	2,769 (69.19%)	3,585 (67.15%)	2,490 (71.43%)
K01738	2.5.1.47	cysK; cysteine synthase	Sulfur metabolism|Cystein...	M00021	1,179 (43.99%)	2,225 (50.86%)	1,763 (44.05%)	2,274 (42.59%)	1,708 (49.00%)
K01784	5.1.3.2	galE, GALE; UDP-glucose 4-epimerase	Galactose metabolism|Ami...	M00632	1,338 (49.93%)	2,477 (56.62%)	2,147 (53.65%)	2,593 (48.57%)	1,805 (51.78%)
K01915	6.3.1.2	glnA, GLUL; glutamine synthetase	Glyoxylate and dicarboxyla...		1,650 (61.57%)	3,107 (71.02%)	2,483 (62.04%)	3,131 (58.64%)	2,304 (66.09%)
K01990		ABC-2.A; ABC-2 type transport system ATP-binding protein	Transporters		1,884 (70.30%)	3,247 (74.22%)	2,376 (59.37%)	2,826 (52.93%)	2,254 (64.66%)
K02003		ABC.CD.A; putative ABC transport system ATP-binding protein	Transporters		1,744 (65.07%)	3,392 (77.53%)	2,834 (70.81%)	3,462 (64.84%)	2,595 (74.44%)
K02004		ABC.CD.P; putative ABC transport system permease protein	Transporters		971 (36.23%)	2,051 (46.88%)	1,404 (35.08%)	1,679 (31.45%)	1,539 (44.15%)
K02078		acpP; acyl carrier protein	Fatty acid biosynthesis|Bio...	M00848	1,726 (64.40%)	3,290 (75.20%)	2,597 (64.89%)	3,478 (65.14%)	2,358 (67.64%)

AI-ready datasets

In “AI-ready datasets” section, we offer standardized, quality-controlled datasets, pre-annotated functional categories (biogeochemical enzymes, defensive system elements, etc.), rich metadata for machine learning applications, which is facilitating the development of predictive models in microbial ecology and biotechnology. For each functional enzyme, gcMeta provides five datasets under different thresholds, facilitating users downloading and using them according to their actual needs.

AI-ready datasets

Diverse and model-compatible AI-ready datasets, which have undergone unified data quality control and normalization, are available for online retrieval and download on this page. The gcMeta team will continue to monitor the list for updates and changes.

The distribution proportion of data entries within different datasets

The source distribution of AI-reay datasets

Microbial biogeochemical cycle metabolism datasets

Dataset	Enzyme	Size(M)	No. of sequences	No. of catalogues	No. of genomes	Identity	Qcover	Scover	E-value	Model	Download
gcMeta Nitrogen cycling 1	nitrogenase	3.2	8,937	36	4,302	60	60	60	0.00001	ESM ProtTrans protGPT2	gcMeta_Nitrogen_cycling_1_606060.fasta.gz gcMeta_Nitrogen_cycling_1_606060.fasta.gz.md5 gcMeta_Nitrogen_cycling_1_606060_meta.txt
gcMeta Nitrogen cycling 1	nitrogenase	3.2	8,786	36	4,248	60	70	70	0.00001	ESM ProtTrans protGPT2	gcMeta_Nitrogen_cycling_1_607070.fasta.gz gcMeta_Nitrogen_cycling_1_607070.fasta.gz.md5 gcMeta_Nitrogen_cycling_1_607070_meta.txt
gcMeta Nitrogen cycling 1	nitrogenase	1.8	4,964	36	2,605	80	60	60	0.00001	ESM ProtTrans protGPT2	gcMeta_Nitrogen_cycling_1_806060.fasta.gz gcMeta_Nitrogen_cycling_1_806060.fasta.gz.md5 gcMeta_Nitrogen_cycling_1_806060_meta.txt
gcMeta Nitrogen cycling 1	nitrogenase	1.8	4,787	36	2,527	80	80	80	0.00001	ESM ProtTrans protGPT2	gcMeta_Nitrogen_cycling_1_808080.fasta.gz gcMeta_Nitrogen_cycling_1_808080.fasta.gz.md5 gcMeta_Nitrogen_cycling_1_808080_meta.txt

Microbial functional enzyme datasets

Dataset	Enzyme	Size(M)	No. of sequences	No. of catalogues	No. of genomes	Identity	Qcover	Scover	E-value	Model	Download
gcMeta α-amylase	α-amylase	8.9	16,913	38	13,937	60	60	60	0.00001	ESM ProtTrans protGPT2	gcMeta_α-amylase_606060.fasta.gz gcMeta_α-amylase_606060.fasta.gz.md5 gcMeta_α-amylase_606060_meta.txt
gcMeta α-amylase	α-amylase	8.8	16,604	38	13,703	60	70	70	0.00001	ESM ProtTrans protGPT2	gcMeta_α-amylase_607070.fasta.gz gcMeta_α-amylase_607070.fasta.gz.md5 gcMeta_α-amylase_607070_meta.txt
gcMeta α-amylase	α-amylase	3.5	6,467	38	5,331	80	60	60	0.00001	ESM ProtTrans protGPT2	gcMeta_α-amylase_806060.fasta.gz gcMeta_α-amylase_806060.fasta.gz.md5 gcMeta_α-amylase_806060_meta.txt
gcMeta α-amylase	α-amylase	3.4	6,269	38	5,178	80	80	80	0.00001	ESM ProtTrans protGPT2	gcMeta_α-amylase_808080.fasta.gz gcMeta_α-amylase_808080.fasta.gz.md5 gcMeta_α-amylase_808080_meta.txt
gcMeta α-amylase	α-amylase	19	34,207	38	23,238	40	40	40	0.00001	ESM ProtTrans protGPT2	gcMeta_α-amylase_404040.fasta.gz gcMeta_α-amylase_404040.fasta.gz.md5 gcMeta_α-amylase_404040_meta.txt

Defense system datasets

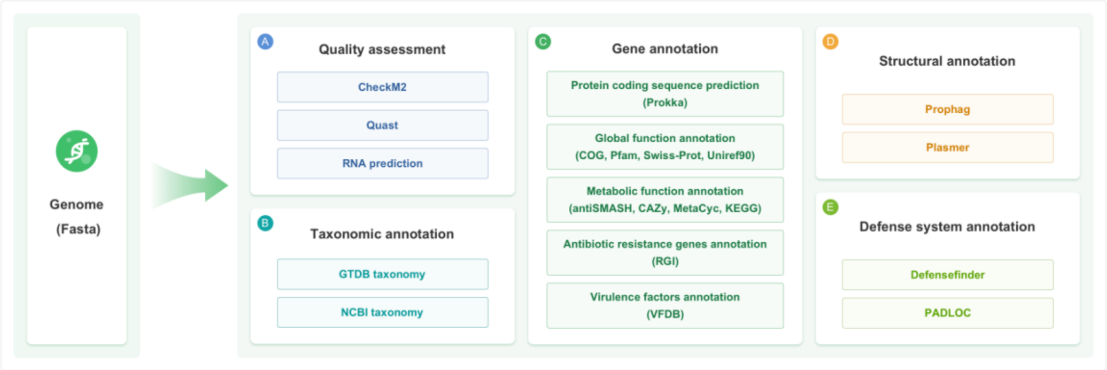
Dataset	Description	Size(M)	No. of sequences	No. of catalogues	No. of genomes	Model	Download
gcMeta AbiD	Antiphage defense system AbiD	3	9,455	36	7,824	ESM ProtTrans protGPT7	AbiD.fasta.gz AbiD.fasta.gz.md5 AbiD_meta.txt
gcMeta AbiE	Antiphage defense system AbiE	4	14,465	38	6,017	ESM ProtTrans protGPT6	AbiE.fasta.gz AbiE.fasta.gz.md5 AbiE_meta.txt
gcMeta BREX	Antiphage defense system BREX	9.2	12,776	35	2,593	ESM ProtTrans protGPT11	BREX.fasta.gz BREX.fasta.gz.md5 BREX_meta.txt
gcMeta Cas	CRISPR-Cas system	54	137,126	38	19,887	ESM ProtTrans protGPT3	Cas.fasta.gz Cas.fasta.gz.md5 Cas_meta.txt
gcMeta Hachiman	Antiphage defense system Hachiman	1.2	1,953	37	947	ESM ProtTrans protGPT13	Hachiman.fasta.gz Hachiman.fasta.gz.md5 Hachiman_meta.txt
gcMeta MazEF	MazEF toxin–antitoxin system	1.1	8,628	38	3,853	ESM ProtTrans protGPT8	MazEFfasta.gz MazEF.fasta.gz.md5 MazEF_meta.txt

Tool

Metagenome-assembled Genome Analysis

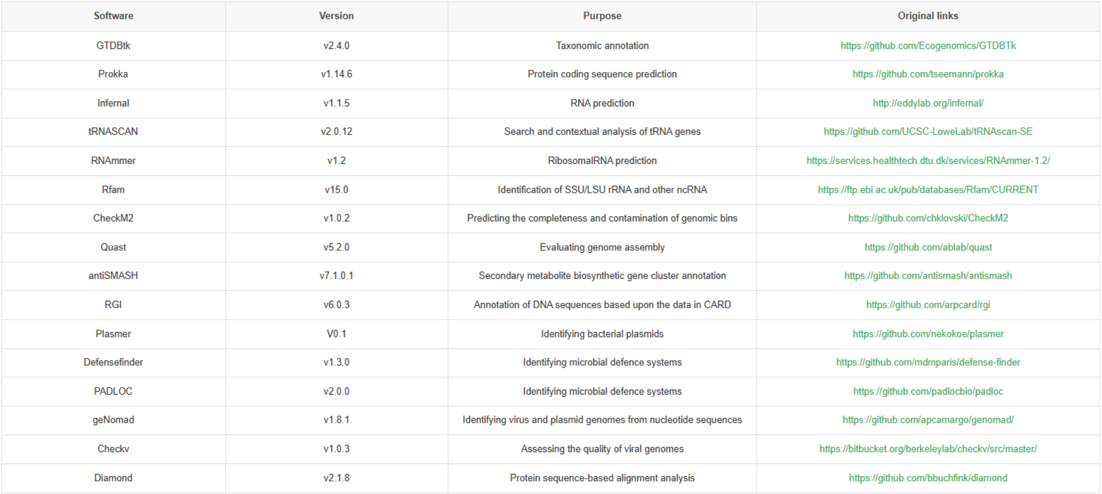
The genome analysis tool can perform detailed annotation analysis of metagenome-assembled genome (MAG), primarily including five components: genome quality assessment, taxonomic annotation, gene annotation, structural annotation, and defense system annotation. We have provided detailed introductions to the analysis workflow, related software, and analysis results. Additionally, a personalized screening function is available in the “program selection” module to facilitate users in making choices based on their actual needs.

Analysis Process



1. Original link

We provide detailed software introductions and original links on the Tools page.



Software	Version	Purpose	Original links
GTDBtk	v2.4.0	Taxonomic annotation	https://github.com/Ecogenomics/GTDBTk
Prokka	v1.14.6	Protein coding sequence prediction	https://github.com/tseemann/prokka
Infernal	v1.1.5	RNA prediction	http://eddylab.org/infernal/
tRNASCAN	v2.0.12	Search and contextual analysis of tRNA genes	https://github.com/UCSC-LoweLab/tRNAscan-SE
RNAmmer	v1.2	RibosomalRNA prediction	https://services.healthtech.dtu.dk/services/RNAmmer-1.2/
Rfam	v15.0	Identification of SSU/LSU rRNA and other ncRNA	https://ftp.ebi.ac.uk/pub/databases/Rfam/CURRENT
CheckM2	v1.0.2	Predicting the completeness and contamination of genomic bins	https://github.com/chklovski/CheckM2
Quast	v5.2.0	Evaluating genome assembly	https://github.com/ablab/quast
antiSMASH	v7.1.0.1	Secondary metabolite biosynthetic gene cluster annotation	https://github.com/antismash/antismash
RGI	v6.0.3	Annotation of DNA sequences based upon the data in CARD	https://github.com/arpcard/rgi
Plasmer	V0.1	Identifying bacterial plasmids	https://github.com/nekokoe/plasmer
Defensefinder	v1.3.0	Identifying microbial defence systems	https://github.com/mdmparis/defense-finder
PADLOC	v2.0.0	Identifying microbial defence systems	https://github.com/padlocbio/padloc
geNomad	v1.8.1	Identifying virus and plasmid genomes from nucleotide sequences	https://github.com/apcamargo/genomad/
Checkv	v1.0.3	Assessing the quality of viral genomes	https://bitbucket.org/berkeleylab/checkv/src/master/
Diamond	v2.1.8	Protein sequence-based alignment analysis	https://github.com/bbuchfink/diamond

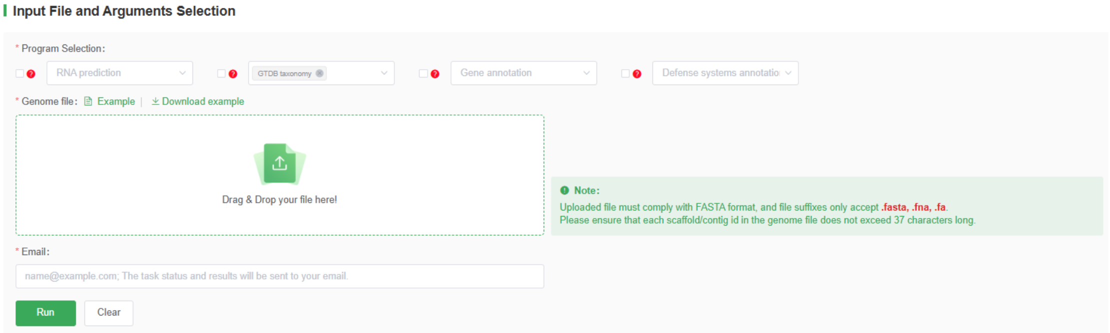
2. Introduction for use

1) Please select the analysis process you need in the analysis interface below. Select the corresponding analysis modules as prompted by the process.

2) In the "Genome files" window, upload your sequence file.

3) Please leave your email address, and all analysis results will be returned by email.

* Required fields are marked with an asterisk.



Input File and Arguments Selection

* Program Selection:

☐ RNA prediction ☐ GTDB taxonomy ☐ Gene annotation ☐ Defense systems annotatio

* Genome file: Example | Download example

Note:
Uploaded file must comply with FASTA format, and file suffixes only accept **.fasta, .fna, .fa**.
Please ensure that each scaffold/contig id in the genome file does not exceed 37 characters long.

* Email:

name@example.com; The task status and results will be sent to your email.

3. Input files and parameter selection

1) Upload the input file in fasta format.

a) The uploaded file must be in FASTA format. Accepted file extensions are: .fna , .fa or .fasta.

b) File names cannot contain special characters, such as: #;;,,""""`$&()-=+[]{}|^:!?%".

c) Please ensure that each scaffold/contig id in the genome file does not exceed 37 characters long.

2) Program Section

a) RNA prediction: Select the desired software to perform non-coding RNA prediction.

b) Taxonomy: GTDB annotation is selected by default here, and NCBI annotation can be added as needed.

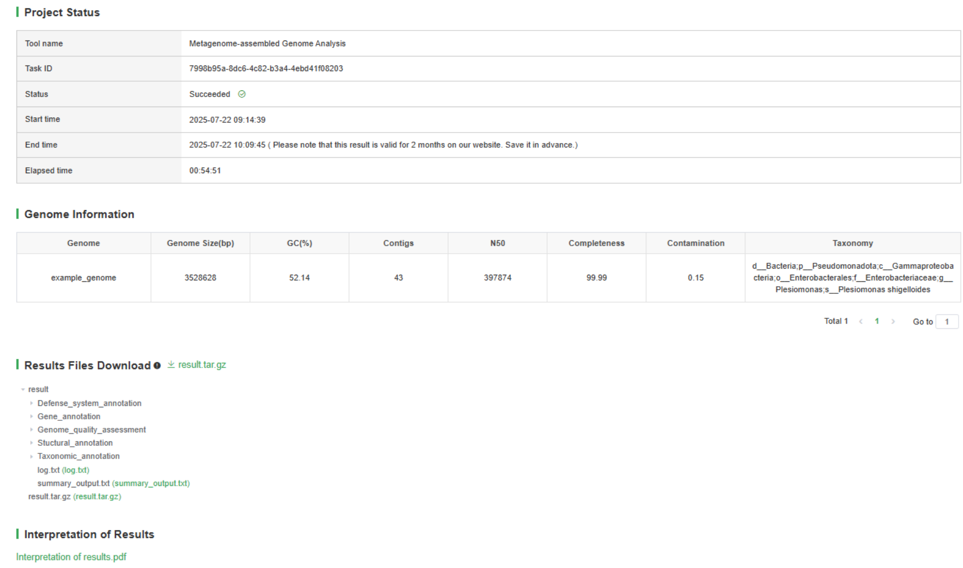
c) Gene annotation: Select the desired annotation content, and free combination is allowed.

d) Defense system annotation: Select the desired software.

4. Leave your email address and click “Run”.

5. Viewing analysis results

gcMeta provides task running details and basic statistical displays of genome. Results of all selected analysis content can be viewed and downloaded under "Results Files Download". In addition, the "Interpretation of results.pdf" file, which contains detailed explanations of the analysis results, can be downloaded as needed.



Project Status

Tool name	Metagenome-assembled Genome Analysis
Task ID	7998b95a-8dc6-4c82-b3a4-4ebd41f08203
Status	Succeeded ✓
Start time	2025-07-22 09:14:39
End time	2025-07-22 10:09:45 (Please note that this result is valid for 2 months on our website. Save it in advance.)
Elapsed time	00:54:51

Genome Information

Genome	Genome Size(bp)	GC(%)	Contigs	N50	Completeness	Contamination	Taxonomy
example_genome	3528628	52.14	43	397874	99.99	0.15	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Plesiomonas;s__Plesiomonas shigelloides

Results Files Download ⤓ result.tar.gz

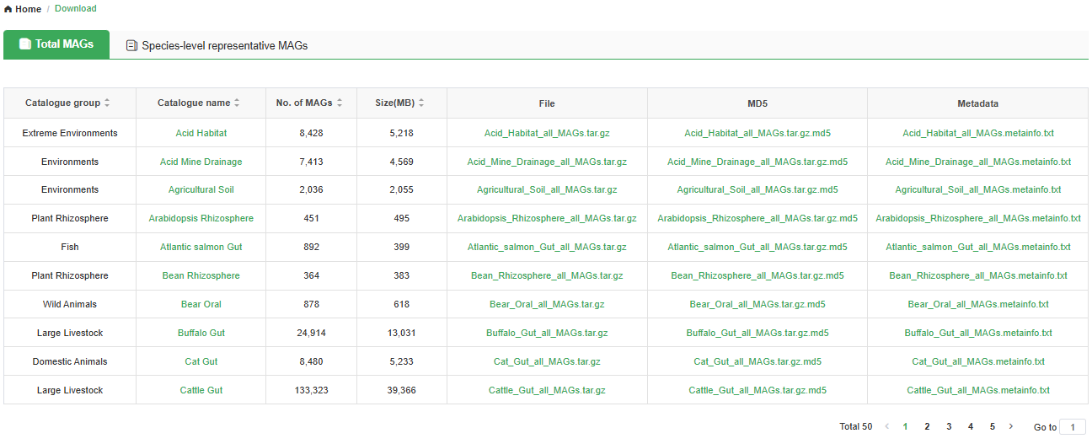
- result
  - Defense_system_annotation
  - Gene_annotation
  - Genome_quality_assessment
  - Stuctural_annotation
  - Taxonomic_annotation
  - log.txt (log.txt)
  - summary_output.txt (summary_output.txt)
- result.tar.gz (result.tar.gz)

Interpretation of Results

Interpretation of results.pdf

Download

gcMeta provides sequence file downloads for total MAGs and species-level representative MAGs grouped by catalogue, along with corresponding MD5 files and metadata files for the compressed packages. The metainfo includes the following items: "MAG, Catalogue_group, Catalogue_name, Biosample, Catalogue_OTU_id, Catalogue_OTU_status, Total_OTU_id, Total_OTU_status, GTDB_taxonomy, Completeness, Contamination, Genome_size, N50, GC_content, and No_of_scaffolds". "Catalogue_OTU_id" indicates the species clustering affiliation within a specific catalogue. "Total_OTU_id" indicates the species clustering affiliation across all catalogues.



Catalogue group	Catalogue name	No. of MAGs	Size(MB)	File	MD5	Metadata
Extreme Environments	Acid Habitat	8,428	5,218	Acid_Habitat_all_MAGs.tar.gz	Acid_Habitat_all_MAGs.tar.gz.md5	Acid_Habitat_all_MAGs.metainfo.txt
Environments	Acid Mine Drainage	7,413	4,569	Acid_Mine_Drainage_all_MAGs.tar.gz	Acid_Mine_Drainage_all_MAGs.tar.gz.md5	Acid_Mine_Drainage_all_MAGs.metainfo.txt
Environments	Agricultural Soil	2,036	2,055	Agricultural_Soil_all_MAGs.tar.gz	Agricultural_Soil_all_MAGs.tar.gz.md5	Agricultural_Soil_all_MAGs.metainfo.txt
Plant Rhizosphere	Arabidopsis Rhizosphere	451	495	Arabidopsis_Rhizosphere_all_MAGs.tar.gz	Arabidopsis_Rhizosphere_all_MAGs.tar.gz.md5	Arabidopsis_Rhizosphere_all_MAGs.metainfo.txt
Fish	Atlantic salmon Gut	892	399	Atlantic_salmon_Gut_all_MAGs.tar.gz	Atlantic_salmon_Gut_all_MAGs.tar.gz.md5	Atlantic_salmon_Gut_all_MAGs.metainfo.txt
Plant Rhizosphere	Bean Rhizosphere	364	383	Bean_Rhizosphere_all_MAGs.tar.gz	Bean_Rhizosphere_all_MAGs.tar.gz.md5	Bean_Rhizosphere_all_MAGs.metainfo.txt
Wild Animals	Bear Oral	878	618	Bear_Oral_all_MAGs.tar.gz	Bear_Oral_all_MAGs.tar.gz.md5	Bear_Oral_all_MAGs.metainfo.txt
Large Livestock	Buffalo Gut	24,914	13,031	Buffalo_Gut_all_MAGs.tar.gz	Buffalo_Gut_all_MAGs.tar.gz.md5	Buffalo_Gut_all_MAGs.metainfo.txt
Domestic Animals	Cat Gut	8,480	5,233	Cat_Gut_all_MAGs.tar.gz	Cat_Gut_all_MAGs.tar.gz.md5	Cat_Gut_all_MAGs.metainfo.txt
Large Livestock	Cattle Gut	133,323	39,366	Cattle_Gut_all_MAGs.tar.gz	Cattle_Gut_all_MAGs.tar.gz.md5	Cattle_Gut_all_MAGs.metainfo.txt