

December 3, 2019

第42回日本分子生物学会年会
「生命科学のデータベース活用法2019」

MicrobeDB.jp version 3の活用法

森宙史

Hiroshi Mori

国立遺伝学研究所

情報研究系

<http://microbedb.jp/>



Text

Analysis

Statistics

Search

Environment: hot spring

Taxonomy: Enterococcus faecalis

Taxonomy: Streptomyces avermitilis

Gene: psbA

ID: 29

2011年から公開している
原核生物を主とした微生物の統合データベース (DB)

MicrobeDB.jp v.3 project members

National Institute of Genetics: (Genome, Metagenome, Ontology)

Ken Kurokawa, Yasukazu Nakamura, Hiroshi Mori,
Takatomo Fujisawa, Eli Kaminuma (TMDU), Koichi Higashi

National Institute of Basic Biology: (Ortholog)

Ikuo Uchiyama, Hirokazu Chiba (DBCLS), Hiroyo Nishide

Tokyo Institute of Technology: (Metagenome)

Takuji Yamada, Zenichi Nakagawa

Chiba University: (Fungal & Bacterial culture collection info.)

Hiroki Takahashi, Takashi Yaguchi

Technical adviser:

DBCLS (especially Shuichi Kawashima, Toshiaki Katayama)

微生物統合DB MicrobeDB.jpの基本的な開発方針

- 遺伝子と系統のリンク
 - その系統はどのような遺伝子機能が特徴的か？
- 遺伝子と表現型のリンク
 - ある表現型を担う遺伝子の候補は？
- 遺伝子と環境のリンク
 - その環境で生息するために必要な遺伝子機能は？
- 系統と環境のリンク
 - その環境で多い系統は？

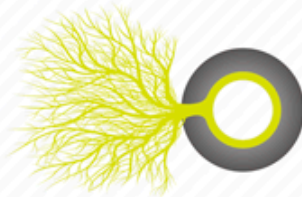
既存のデータにこれらのリンクを付与し、リンクから微生物に関する新たな知見を得られるDBを目指す



Microbe DB^{.JP}

integrates lots of data related to microbes.

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



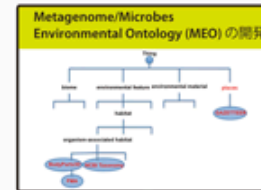
Microbe DB^{.JP}

<http://microbedb.jp/>

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

中身は基本的にRDFデータ

Ontology



Gene

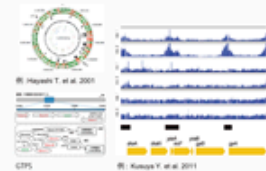
Taxon

Environment



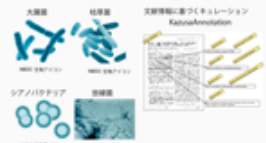
Ortholog: **MBGD**

オースログデータ



Genome: RefSeq

オミックスデータ



モデル微生物の高品質
アノテーションデータ



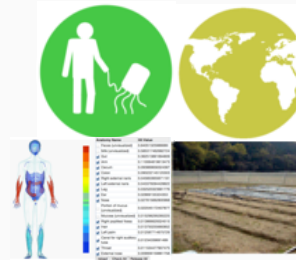
Taxonomy:
NCBI Taxonomy

系統分類データ



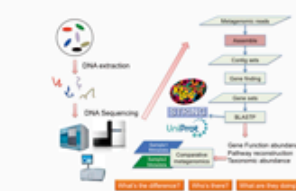
Culture Collection:
NBRC/JCM

菌株データ



Metadata:
INSDC DRA

環境のメタデータ



Metagenome:
INSDC DRA

メタゲノムデータ

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

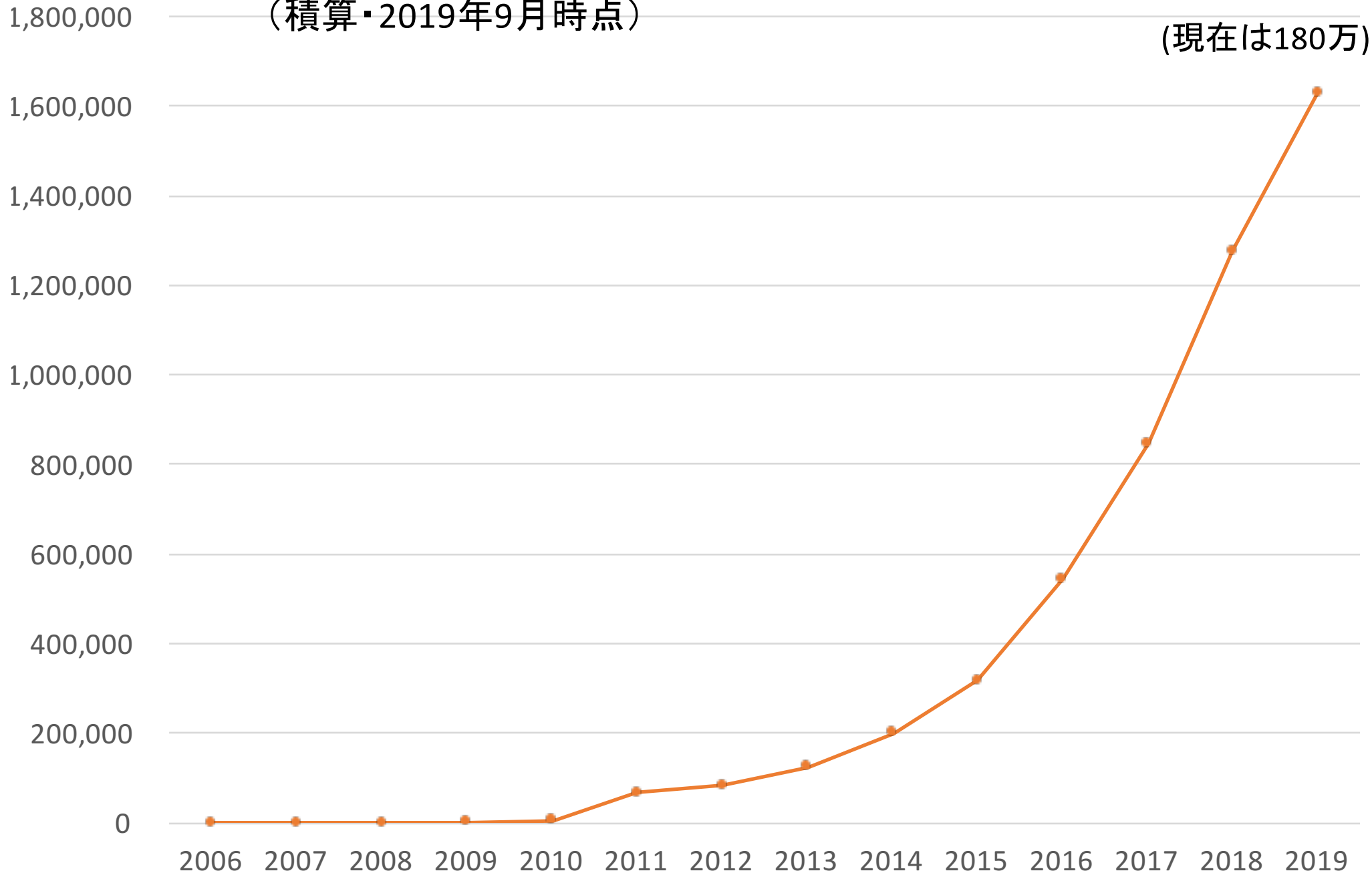


Red color indicates our collaborators.

INSDC DRA/ERA/SRAで公開されたマイクロバイームサンプル数

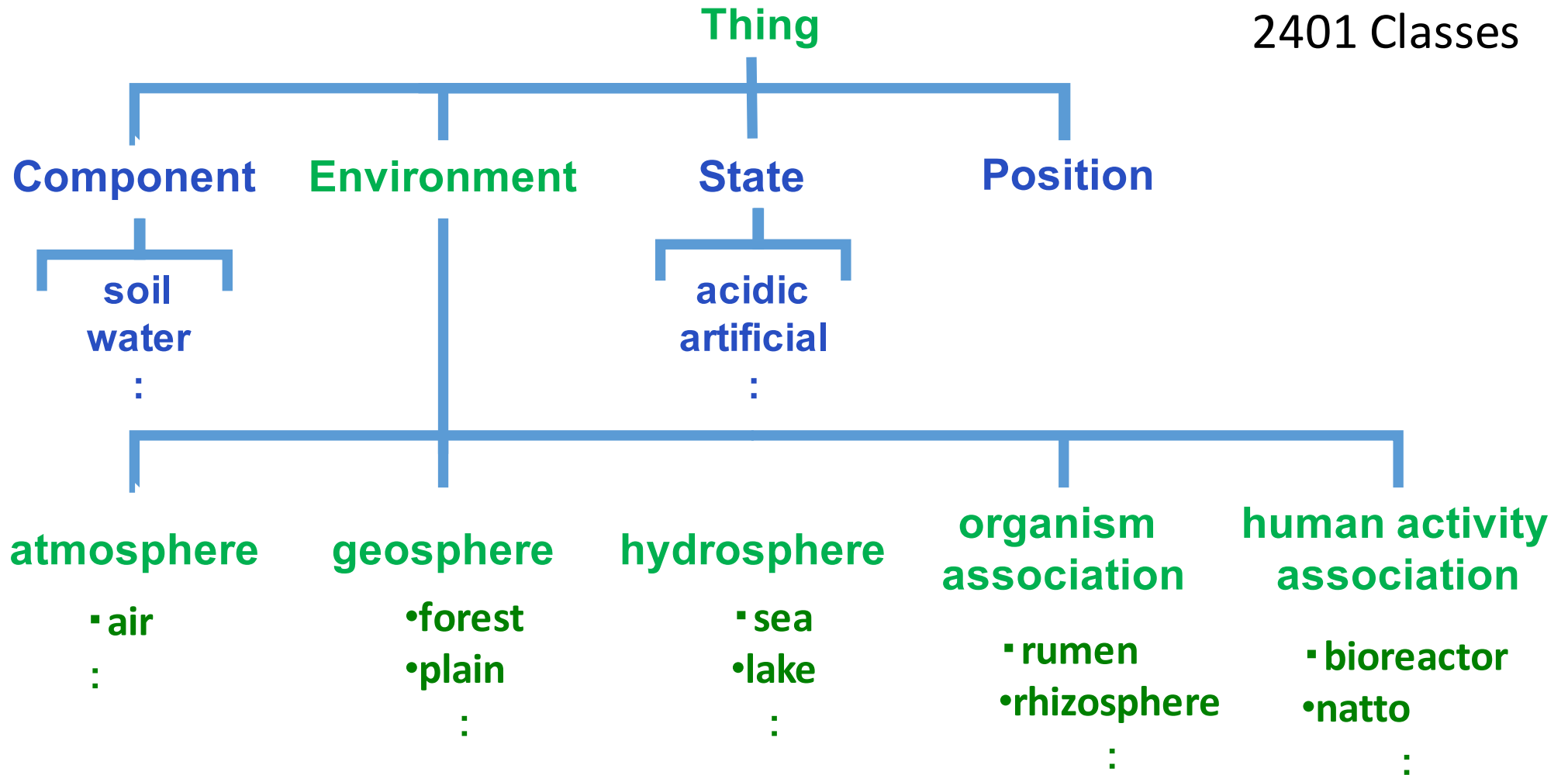
(積算・2019年9月時点)

(現在は180万)



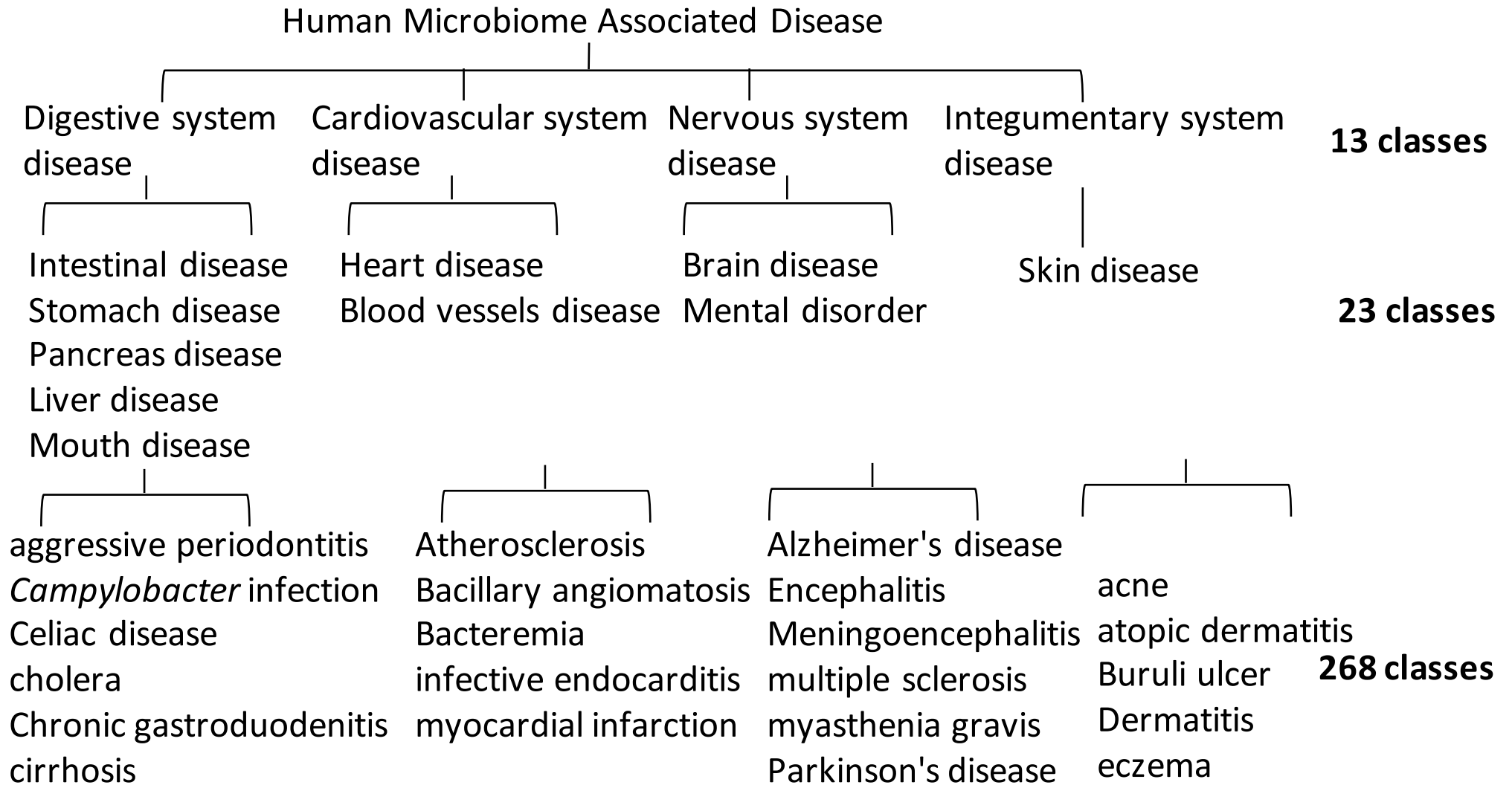
Metagenome/Microbes Environmental Ontology (MEO) Ver. 0.9.2

2401 Classes



意味・概念をDB内で統一してオントロジーとして語彙を定義

Human Microbiome Associated Disease Ontology (HMADO)



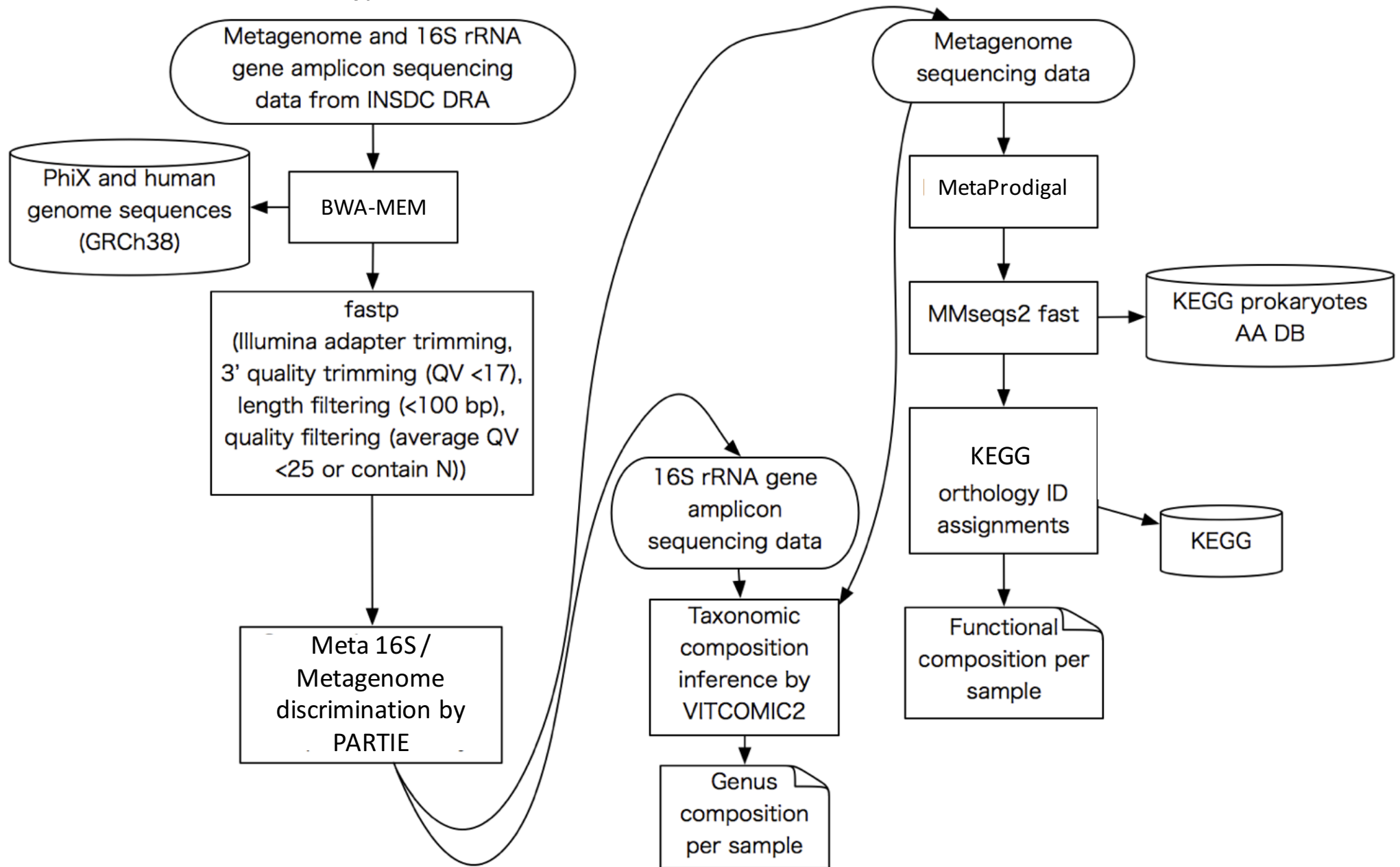
感染症なのか否かなど、微生物群集が関係するヒトの病気の分類

MicrobeDB.jp version 3 data

Data category	Number of entry
Genome metadata (from RefSeq)	290,208 genomes
Culture collection strain data from JCM/NBRC (RDF-Portal)	38,414 strains
Microbiome metadata (from INSDC DRA)	1,631,611 samples
Microbiome taxonomic composition data	96,766 samples
Microbiome functional composition data	4,784 samples

Ver. 2(2014)と比べてゲノムは約20倍、メタゲノムは約10倍の数に

MeGAP3 (MicrobeDB.jp ver. 3のパイプライン)



INSDC DRA/ERA/SRAで公開されたマイクロバイオーームサンプル数 (2019年9月時点)

	サンプル数
マイクロバイオーーム	約 1,600,000
ヒト	約 420,000
マウス	約 95,000
土壌	約 240,000
水環境	約 200,000
人工環境	約 140,000
植物共生	約 120,000

INSDC DRA/ERA/SRAで公開された
疾患患者のマイクロバイーム
サンプル数(2019年9月時点)

サンプル数が1000以上ある疾患が
右表の18疾患
(肥満と未熟児を除く)

大多数が16S rRNA遺伝子の
アンプリコン解析

疾患名 (HMADO)	サンプルサイズ
IBD	37,094
皮膚炎	13,219
にきび	7,488
喘息	6,178
乾癬	5,151
大腸炎	4,453
嚢胞性線維症	4,040
アレルギー	3,902
下痢	2,684
大腸がん	2,085
細菌性膣炎	2,030
アテローム性動脈硬化	1,968
赤痢	1,918
パーキンソン病	1,760
腺がん	1,742
統合失調症	1,712
歯周炎	1,555
虫歯	1,370



MicrobeDB.jp

Integrating and representing genome, metagenome, taxonomy resources and the analysis datasets with Semantic Web Technologies.

Database statistics

Total number of Metagenomic samples (SRA/SRS):	173,359 samples
- with taxonomic analysis results:	60,551 samples
- with functional analysis results:	4,048 samples
Total number of Assembled Genomes (RefSeq/Genbank):	16,983 taxa
Total number of Strains (JCM/NBRC):	16,671 strains
Total number of Environmental terms in ontology (MEO):	2,381 terms

Show graph

Search id or term...

Public/Private

metagenome_public 3729

hasMetagenomeAnalysis

taxonomy 3729

function 609

hasMEO (Text)

gut x

hasMEO: Component

Component for environment 389

hasMEO: Env

Environment for microbes 3729

hasMEO: Position

Position toward environment 5

hasMEO: State

Metagenomic samples 3729 results found in 112ms

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

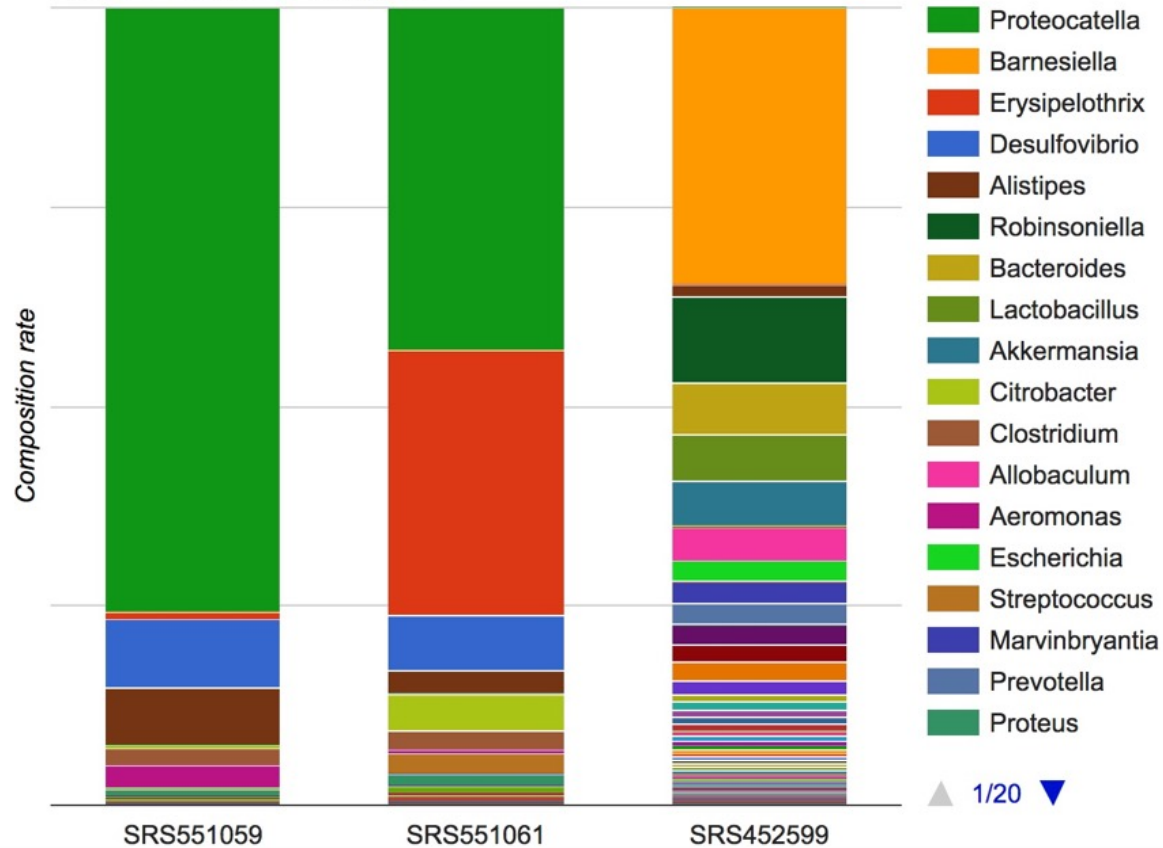
Previous 1 2 3 4 ... Next

10 Select All Deselect All

Select	MDB SampleID	msv:sampleTitle	msv:scientificName	msv:hasTaxonID	msv:hasBioProjectID	msv:hasBioSar
Remove	SRS551059	Content of the intestinum from animals fed one meal of heparinized sheep blood	gut metagenome	749906	PRJNA237098	SAMN026145
Remove	SRS551061	Content of the intestinum of animals fed two meals of heparinized sheep blood four weeks apart	gut metagenome	749906	PRJNA237098	SAMN026145
Remove	SRS452599	Environmental/Metagenome sample for mouse gut metagenome	mouse gut metagenome	410661	PRJNA209582	SAMN022138
Add	SRS367344	Pooled 16S rRNA gene sequences	Bacteria	2	PRJNA209582	SAMN017587
Add	SRS369251	Pooled 16S rRNA gene sequences		2	PRJNA209582	SAMN017656



Samples Taxonomic Composition



**ここからversion 3の新機能
(12月中に切り替わります)**

MEO:soil 由来マイクロバイームサンプルの絞り込み検索



MicrobeDB.jp

Index

facet_sample 247853

hasMetagenomeAnalysis

taxonomy 8500

function 246

Search id ...

attribute name

Search attribute name ...

attribute value

Search attribute value ...

hasMEO (Text)

soil x

hasMEO: Component

Component for environment 237292

hasMEO: Env

Environment for microbes 171749

taxonomy (Text)

Search taxonomy terms ..

taxonomy

root 247853

hasHostTaxonomy (Text)

Search HostTaxonomy...

hasHostTaxonomy

root 30348

pH

Metagenomic samples 247853 results found in 527ms

hasMEO (Text): soil x Clear all filters

Previous 1 2 3 4 ... Next

50 Select All Deselect All

Select	MDB SampleID	title	organism.name	organism.identifier	BioProjectID	SRAID	SRRID
Add	SAMN02898402	MIMS Environmental/Metagenome sample from soil metagenome	soil metagenome	410658		SRS666697	
Add	SAMN08408074	Metagenome or environmental sample from soil metagenome	soil metagenome	410658	PRJNA431674	SRS2872044	SRR666697
Add	SAMN08408073	Metagenome or environmental sample from soil metagenome	soil metagenome	410658	PRJNA431673	SRS2872043	SRR666697
Add	SAMN07281069	crop soil	soil metagenome	410658	PRJNA392052	SRS2317230	SRR555555
Add	SAMN07281068	crop soil	soil metagenome	410658	PRJNA392052	SRS2317229	SRR555555
Add	SAMN07281067	crop soil	soil metagenome	410658	PRJNA392052	SRS2317207	SRR555555
Add	SAMN07281066	crop soil	soil metagenome	410658	PRJNA392052	SRS2317228	SRR555555
Add	SAMN07281065	crop soil	soil metagenome	410658	PRJNA392052	SRS2317227	SRR555555
Add	SAMN07281064	crop soil	soil metagenome	410658	PRJNA392052	SRS2317226	SRR555555
Add	SAMN07281063	crop soil	soil metagenome	410658	PRJNA392052	SRS2317205	SRR555555
Add	SAMN07281062	crop soil	soil metagenome	410658	PRJNA392052	SRS2317206	SRR555555
Add	SAMN07281061	crop soil	soil metagenome	410658	PRJNA392052	SRS2317214	SRR555555
Add	SAMN07281060	crop soil	soil metagenome	410658	PRJNA392052	SRS2317215	SRR555555
Add	SAMN07281059	crop soil	soil metagenome	410658	PRJNA392052	SRS2317212	SRR555555
Add	SAMN07281058	crop soil	soil metagenome	410658	PRJNA392052	SRS2317213	SRR555555
Add	SAMN07281057	crop soil	soil metagenome	410658	PRJNA392052	SRS2317210	SRR555555
Add	SAMN07281056	crop soil	soil metagenome	410658	PRJNA392052	SRS2317211	SRR555555
Add	SAMN07281055	crop soil	soil metagenome	410658	PRJNA392052	SRS2317208	SRR555555

HMADO:cancer 由来マイクロバイオームサンプルの絞り込み検索

Index

facet_sample 3587

hasMetagenomeAnalysis

taxonomy 41

Search id ...

attribute name

Search attribute name ...

attribute value

Search attribute value ...

hasMEO (Text)

Search MEO terms ...

hasMEO: Env

Environment for microbes 3587

taxonomy (Text)

Search taxonomy terms ..

taxonomy

root 3587

hasHostTaxonomy (Text)

Search HostTaxonomy...

hasHostTaxonomy

root 3542

HMADO (Text)

Search cancer x

HMADO

Metagenomic samples 3587 results found in 43ms

HMADO (Text): cancer x Clear all filters

Previous 1 2 3 4 ... Next

10 Select All Deselect All

Select	MDB SampleID	title	organism.name	organism.identifier	BioProjectID	SRAID	SRRID
Add	SAMN07452485	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409846	SRR5903357 SRR5903369 SRR5903386
Add	SAMN07452484	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409841	SRR5903342 SRR5903368 SRR5903385
Add	SAMN07452483	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409840	SRR5903336 SRR5903367 SRR5903384
Add	SAMN07452482	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409834	SRR5903329 SRR5903366 SRR5903383
Add	SAMN07452481	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409835	SRR5903330 SRR5903365 SRR5903382
Add	SAMN07452480	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409832	SRR5903327 SRR5903345 SRR5903381
Add	SAMN07452479	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409833	SRR5903328 SRR5903346 SRR5903380
Add	SAMN07452478	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409830	SRR5903325 SRR5903375 SRR5903379
Add	SAMN07452477	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409831	SRR5903326 SRR5903337 SRR5903378
Add	SAMN07452476	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409828	SRR5903323 SRR5903338 SRR5903377

Metagenome sample comparison analysis Compare 0 Please select items within range 2 - 100

Japanese由来マイクロバイームサンプルの絞り込み検索



MicrobeDB.jp

Home Document Analysis -

e.g. hot spring, Enterococcus faecalis, psb.

Search

Index

facet_sample 1539

hasMetagenomeAnalysis

taxonomy 1385

Search id ...

attribute name

Search attribute name ...

attribute value

Search attribute value ...

hasMEO (Text)

Search MEO terms ...

hasMEO: Component

Component for environment 4

hasMEO: Env

Environment for microbes 1539

taxonomy (Text)

Search taxonomy terms ..

taxonomy

root 1539

hasHostTaxonomy (Text)

Search HostTaxonomy...

hasHostTaxonomy

root 1539

HMADO (Text)

Metagenomic samples

1539 results found in 147ms

HostEthnicity: japan x Clear all filters

Previous

1

2

3

4

...

Next

10

Select All

Deselect All

Select	MDB SampleID	title	organism.name	organism.identifier	BioProjectID	SRAID	SRRID
Add	SAMD00079966	Japanese469	human gut metagenome	408170	PRJDB4360	DRS057897	DRR092102
Add	SAMD00024579	Yms34	human gut metagenome	408170	PRJDB3417	DRS020590	DRR028771
Add	SAMD00058608	Microbiota of the fecal sample from subject 516	human gut metagenome	408170	PRJDB4998		DRR068403
Add	SAMD00035679	NA40	human gut metagenome	408170	PRJDB4064		DRR041845
Add	SAMD00043213	Japanese371	human gut metagenome	408170	PRJDB4360		DRR049363
Add	SAMD00036348	TS-41	human gut metagenome	408170	PRJDB3601		DRR042663
Add	SAMD00079965	Japanese468	human gut metagenome	408170	PRJDB4360	DRS057896	DRR092101
Add	SAMD00079964	Japanese467	human gut metagenome	408170	PRJDB4360	DRS057895	DRR092100
Add	SAMD00079963	Japanese466	human gut metagenome	408170	PRJDB4360	DRS057894	DRR092099
Add	SAMD00079962	Japanese465	human gut metagenome	408170	PRJDB4360	DRS057893	DRR092098

Metagenome sample comparison analysis



Compare 0

Please select items within range 2 - 100

Ethnicity: Japanese

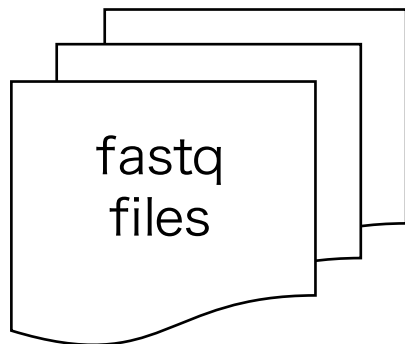
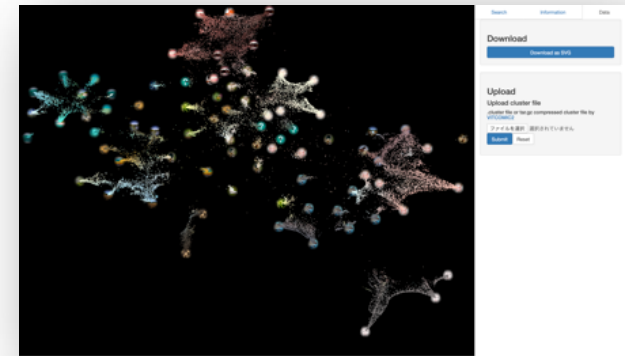
MicrobeDB.jp ver. 3の特徴

- かなりマイクロバイオームに特化した統合DB
- 環境や病気のオントロジーはメタデータから自動的にアノテーション
- 160万サンプル以上のマイクロバイオームデータをメタデータ等で検索可能
- MeGAP3による計算が終わったサンプルから、数ヶ月に一度、MicrobeDB.jpで検索できるように
- 現在主にセキュリティ面等をテスト中であり、12月中には<http://microbedb.jp/>から公開

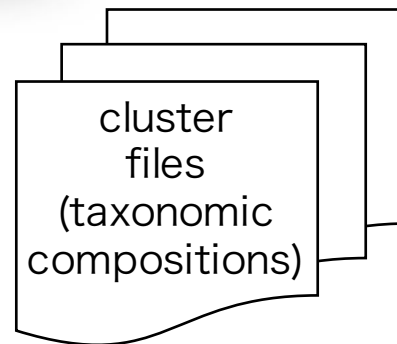
How to map your metagenome data on LEA

VITCOMIC2
(<http://vitcomic.org/>)

LEA
(<http://leamicrobe.jp/>)



~ 6 min. /
a sample
(100,000
sequences)



~ 0.5 sec. /
a sample