

第44回日本分子生物学会年会
フォーラム
生命科学のデータベース活用法

微生物統合データベース MicrobeDB.jpと関連ツール

国立遺伝学研究所
黒川 顕

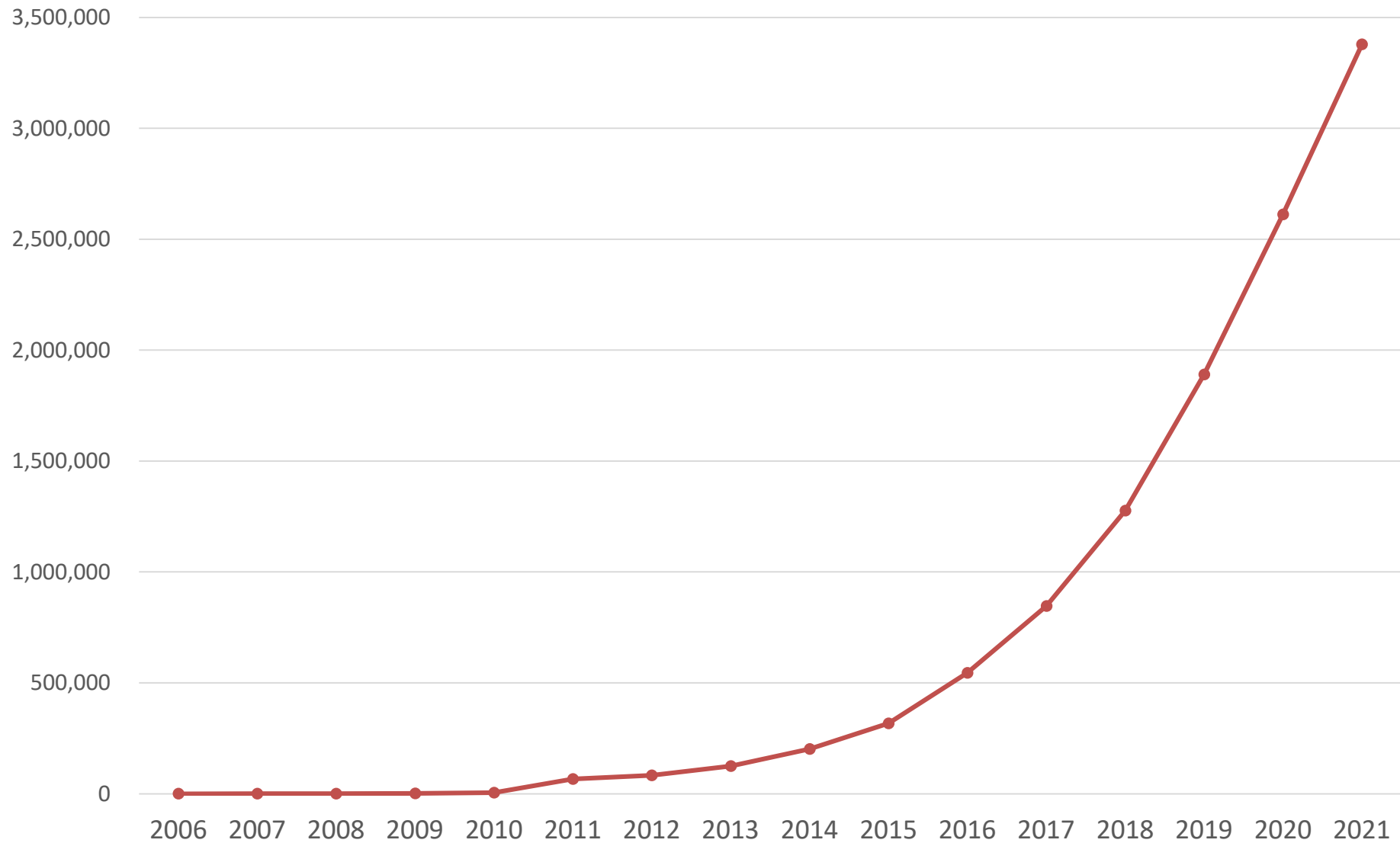


Licensed under a Creative Commons 表示4.0国際ライセンス ©2021
黒川 顕(国立遺伝学研究所)

微生物・微生物叢研究を取り巻く現状

- 微生物は地球上のいたる所に存在し、環境や他の生物と密接に関与している
- 微生物研究はバイオ分野のみならず、他の多くの分野と連携可能
- Metagenome、Microbiome等の研究が加速度的に進展している
- ゲノム、メタゲノム、環境を俯瞰できるデータベースがない

INSDC DRA/ERA/SRAに登録されている マイクロバイームデータのサンプル数 (積算：2021年10月時点)



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（2021年10月時点）

	サンプル数
マイクロバイオーーム	約 3,400,000
ヒト	約 1,050,000
マウス	約 160,000
土壌	約 460,000
水環境	約 280,000
人工環境	約 170,000
植物共生	約 220,000

340万サンプルはヒトや土壌等マイクロバイオーームサンプルの総数

MicrobeDB.jp ver.3

(<https://microbedb.jp/>)

The screenshot shows the MicrobeDB.jp website interface. At the top, there is a navigation bar with 'Home', 'Document', and 'Analysis' tabs. A search bar contains the text 'e.g. hot spring, Enterococcus faecalis, psb.' and a 'Search' button. The main content area features the MicrobeDB.jp logo and a description: 'Integrating and representing genome, metagenome, taxonomy resources and the analysis datasets with Semantic Web Technologies.' Below this is a 'Features' section with several sub-sections:

- Data sources of MicrobeDB.jp ver. 3**: A grid of colored boxes representing different data sources and their counts:
 - Metagenome and Microbes Environmental Ontology: 2401
 - Taxonomy: 129342
 - Ortholog Groups: 4203173
 - Microbial Phenotype Ontology: 277
 - Genome and Metagenome Sample: 1920339
 - Culture collections in Japan: 38414
 - Pathogenic Disease Ontology: 387
 - Human Microbiome Associated Disease Ontology: 305
 - KEGG Orthology: 22421

Last Modified date: 2020-02-16
- Keyword Search**: A section describing a keyword search function with a simple interface, providing free-text access to the literal fields of all RDF/OWL resources on MicrobeDB.jp. It includes a 'Text search' button.
- Representation and Visualization**: A section describing the TogoStanza framework, which enables the visualizing of reusable Web components that are embeddable into any Web applications. It includes a 'TogoStanza List' link.
- Comparative Analysis**: A section describing Comparative Analysis Tools between metagenome samples, environment terms, taxa, and sample metadata and taxonomic/functional analysis based on TogoStanza framework. It mentions that comparative analysis can be visualized by using a comparison tool.
- Upload Your Data**: A section describing how to upload data to MicrobeDB.jp for comparative analysis. It includes links for 'CREATE your account' and 'Sign in'.

MicrobeDB.jp version 3 data

Data category	Number of entry
Genome metadata (from RefSeq)	290,208 genomes
Ortholog cluster data (from MBGD)	375,228 clusters
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

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	BioSampleID	publishedDate
Add	SAMD00079966	Japanese469	human gut metagenome	408170	PRJDB4360	DRS057897	DRR092102	SAMD00079966	2018-01-19T00
Add	SAMD00024579	Yms34	human gut metagenome	408170	PRJDB3417	DRS020590	DRR028771	SAMD00024579	2015-08-24T00
Add	SAMD00058608	Microbiota of the fecal sample from subject 516	human gut metagenome	408170	PRJDB4998		DRR068403	SAMD00058608	2016-11-17T22
Add	SAMD00035679	NA40	human gut metagenome	408170	PRJDB4064		DRR041845	SAMD00035679	2016-06-30T22
Add	SAMD00043213	Japanese371	human gut metagenome	408170	PRJDB4360		DRR049363	SAMD00043213	2016-05-30T13
Add	SAMD00036348	TS-41	human gut metagenome	408170	PRJDB3601		DRR042663	SAMD00036348	2016-03-08T11
Add	SAMD00079965	Japanese468	human gut metagenome	408170	PRJDB4360	DRS057896	DRR092101	SAMD00079965	2018-01-19T00
Add	SAMD00079964	Japanese467	human gut metagenome	408170	PRJDB4360	DRS057895	DRR092100	SAMD00079964	2018-01-19T00
Add	SAMD00079963	Japanese466	human gut metagenome	408170	PRJDB4360	DRS057894	DRR092099	SAMD00079963	2018-01-19T00
Add	SAMD00079962	Japanese465	human gut metagenome	408170	PRJDB4360	DRS057893	DRR092098	SAMD00079962	2018-01-19T00

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

HMADO(cancer)由来マイクロバイームサンプルの絞り込み検索

Home Document Analysis - Search Sign Up

Index

facet_sample 3587

hasMetagenomeAnalysis

taxonomy 41

attribute name

attribute value

hasMEO (Text)

hasMEO: Env

Environment for microbes 3587

taxonomy (Text)

taxonomy

root 3587

hasHostTaxonomy (Text)

hasHostTaxonomy

root 3542

HMADO (Text)

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	BioSampleID	publishedDate
<input type="button" value="Add"/>	SAMN07452485	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409846	SRR5903357 SRR5903369 SRR5903386 SRR5903748	SAMN07452485	2017-08-06T00:00:00.000
<input type="button" value="Add"/>	SAMN07452484	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409841	SRR5903342 SRR5903368 SRR5903385 SRR5903757	SAMN07452484	2017-08-06T00:00:00.000
<input type="button" value="Add"/>	SAMN07452483	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409840	SRR5903336 SRR5903367 SRR5903384 SRR5903756	SAMN07452483	2017-08-06T00:00:00.000
<input type="button" value="Add"/>	SAMN07452482	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409834	SRR5903329 SRR5903366 SRR5903383 SRR5903761	SAMN07452482	2017-08-06T00:00:00.000
<input type="button" value="Add"/>	SAMN07452481	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409835	SRR5903330 SRR5903365 SRR5903382 SRR5903760	SAMN07452481	2017-08-06T00:00:00.000
<input type="button" value="Add"/>	SAMN07452480	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409832	SRR5903327 SRR5903345 SRR5903381 SRR5903759	SAMN07452480	2017-08-06T00:00:00.000
<input type="button" value="Add"/>	SAMN07452479	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409833	SRR5903328 SRR5903346 SRR5903380 SRR5903758	SAMN07452479	2017-08-06T00:00:00.000
<input type="button" value="Add"/>	SAMN07452478	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409830	SRR5903325 SRR5903375 SRR5903379 SRR5903765	SAMN07452478	2017-08-06T00:00:00.000
<input type="button" value="Add"/>	SAMN07452477	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409831	SRR5903326 SRR5903337 SRR5903378 SRR5903764	SAMN07452477	2017-08-06T00:00:00.000
<input type="button" value="Add"/>	SAMN07452476	MIMS Environmental/Metagenome sample from human gut metagenome	human gut metagenome	408170	PRJNA397219	SRS2409828	SRR5903323 SRR5903338 SRR5903377 SRR5903763	SAMN07452476	2017-08-06T00:00:00.000

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

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 Home Document Analysis - e.g. hot spring, Enterococcus faecalis, psb. Search

Index
 facet_sample 247853

hasMetagenomeAnalysis
 taxonomy 8500
 function 246

Search id ...

attribute name
 Search attribute name ...

attribute value
 Search attribute value ...

hasMEO (Text)
 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	BioSampleID	publishedDate
<input type="button" value="Add"/>	SAMN02898402	MIMS Environmental/Metagenome sample from soil metagenome	soil metagenome	410658		SRS666697		SAMN02898402	2014-07-08T00:00:00
<input type="button" value="Add"/>	SAMN08408074	Metagenome or environmental sample from soil metagenome	soil metagenome	410658	PRJNA431674	SRS2872044	SRR6516229	SAMN08408074	2018-01-27T00:00:00
<input type="button" value="Add"/>	SAMN08408073	Metagenome or environmental sample from soil metagenome	soil metagenome	410658	PRJNA431673	SRS2872043	SRR6516228	SAMN08408073	2018-01-27T00:00:00
<input type="button" value="Add"/>	SAMN07281069	crop soil	soil metagenome	410658	PRJNA392052	SRS2317230	SRR5760457	SAMN07281069	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281068	crop soil	soil metagenome	410658	PRJNA392052	SRS2317229	SRR5760458	SAMN07281068	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281067	crop soil	soil metagenome	410658	PRJNA392052	SRS2317207	SRR5760480	SAMN07281067	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281066	crop soil	soil metagenome	410658	PRJNA392052	SRS2317228	SRR5760459	SAMN07281066	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281065	crop soil	soil metagenome	410658	PRJNA392052	SRS2317227	SRR5760460	SAMN07281065	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281064	crop soil	soil metagenome	410658	PRJNA392052	SRS2317226	SRR5760461	SAMN07281064	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281063	crop soil	soil metagenome	410658	PRJNA392052	SRS2317205	SRR5760482	SAMN07281063	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281062	crop soil	soil metagenome	410658	PRJNA392052	SRS2317206	SRR5760481	SAMN07281062	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281061	crop soil	soil metagenome	410658	PRJNA392052	SRS2317214	SRR5760473	SAMN07281061	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281060	crop soil	soil metagenome	410658	PRJNA392052	SRS2317215	SRR5760472	SAMN07281060	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281059	crop soil	soil metagenome	410658	PRJNA392052	SRS2317212	SRR5760475	SAMN07281059	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281058	crop soil	soil metagenome	410658	PRJNA392052	SRS2317213	SRR5760474	SAMN07281058	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281057	crop soil	soil metagenome	410658	PRJNA392052	SRS2317210	SRR5760477	SAMN07281057	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281056	crop soil	soil metagenome	410658	PRJNA392052	SRS2317211	SRR5760476	SAMN07281056	2017-06-27T00:00:00
<input type="button" value="Add"/>	SAMN07281055	crop soil	soil metagenome	410658	PRJNA392052	SRS2317208	SRR5760479	SAMN07281055	2017-06-27T00:00:00

dnaAの キーワード 検索

MicrobeDB.jp Home Document Analysis - e.g. hot spring, Enterococcus faecalis, psbA Search Sign Up Sign

Keyword search

DnaA x

Related keywords

Displaying related keywords . Please press for change to the new term instead of "DnaA"

Environment	0	
Taxonomy	0	
Gene	697	default:110688 default:1462 default:2063 default:2391 default:2642 default:279534 default:286151 default:3033 default:3161 default:3946 default:5118 default:5449 default:60088 default:647 default:6623 default:6795 default:777 default:821 default:8358 default:90
Phenotype	0	
Kegg	2	DnaA-homolog protein DnaA initiator-associating protein
HMADO	0	
PDO	0	
Strain	0	
Sample	0	

hot springの キーワード 検索

MicrobeDB.jp Home Document Analysis - e.g. hot spring, Enterococcus faecalis, psbA Search Sign Up Sign

Keyword search

hot spring x

Related keywords

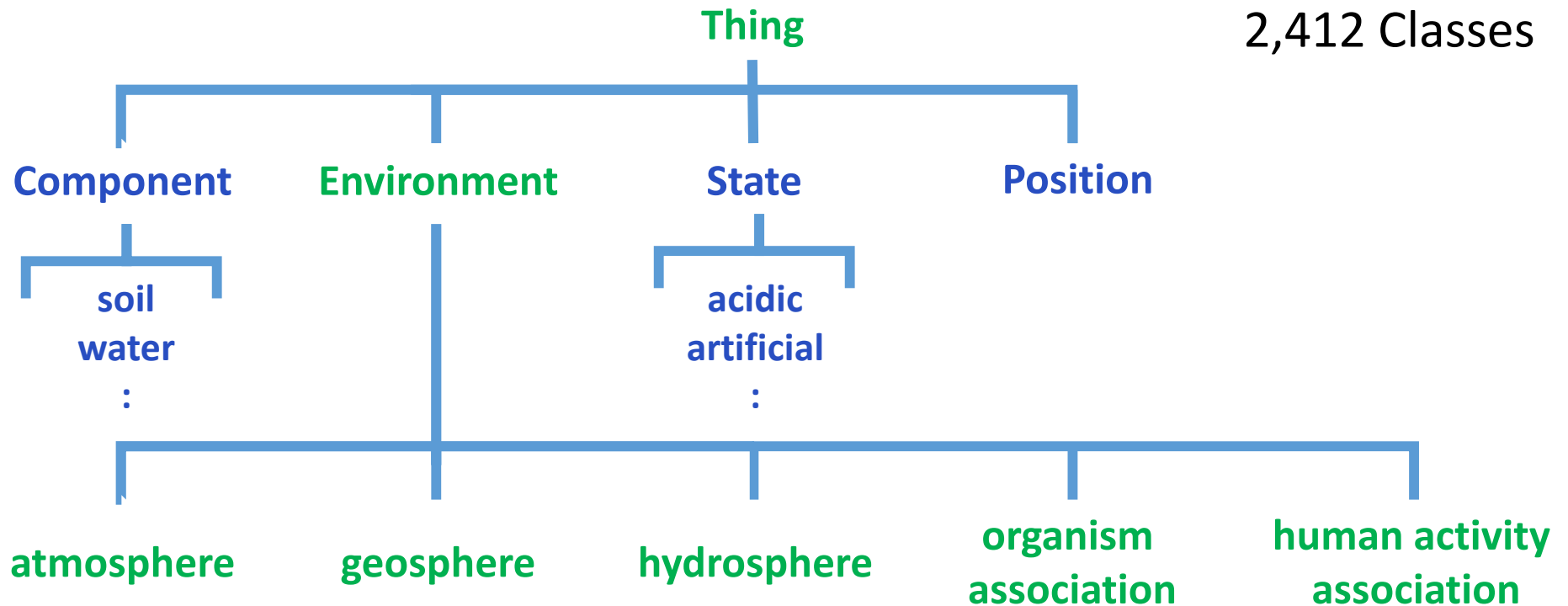
Displaying related keywords . Please press for change to the new term instead of "hot spring"

Environment	9	hot spring calcite hot spring acid hot spring alkaline hot spring hot spring water serpentine hot spring acidic hot spring water artificial hot spring neutral hot spring
Taxonomy	1	hot springs metagenome
Gene	0	
Phenotype	0	
Kegg	0	
HMADO	0	
PDO	0	
Strain	0	
Sample	1277	small RNA library in Yunohama Hot spring Soil metagenome isolated from a hot spring microbiota of hot spring water qSeq_replicate1 microbiota of hot spring water qSeq_replicate2 microbiota of hot spring water qSeq_replicate3 microbiota of hot spring Geothermal sediment in Onikobe hot spring area genus novel hyperthermophilic archaea from hot spring Trebouxioophyceae sp. KSI-1 is a green alga isolated from the seashore hot spring in Satsuma-Iojima, a volcanic island, in Kagoshima, Japan. This strain is highly tolerant to oxidative stress caused by methyl viologen and copper ion. hot spring water metagenome Kinyu sample hot spring water metagenome Yamanoshiro sample hot spring water metagenome Nonoykoya sample hot spring water metagenome Miyama sample hot spring water metagenome Kurinodake sample hot spring water metagenome Ioudani sample hot spring water metagenome Torigigoku sample hot spring water metagenome Ginyu sample hot spring water metagenome Nonoyufunki sample Marine metagenome isolated from Miyuki hot spring Hot spring microbial communities from Tufshi shyam1

10

Microbes/Metagenomes Environmental Ontology (MEO) ver. 0.9.3

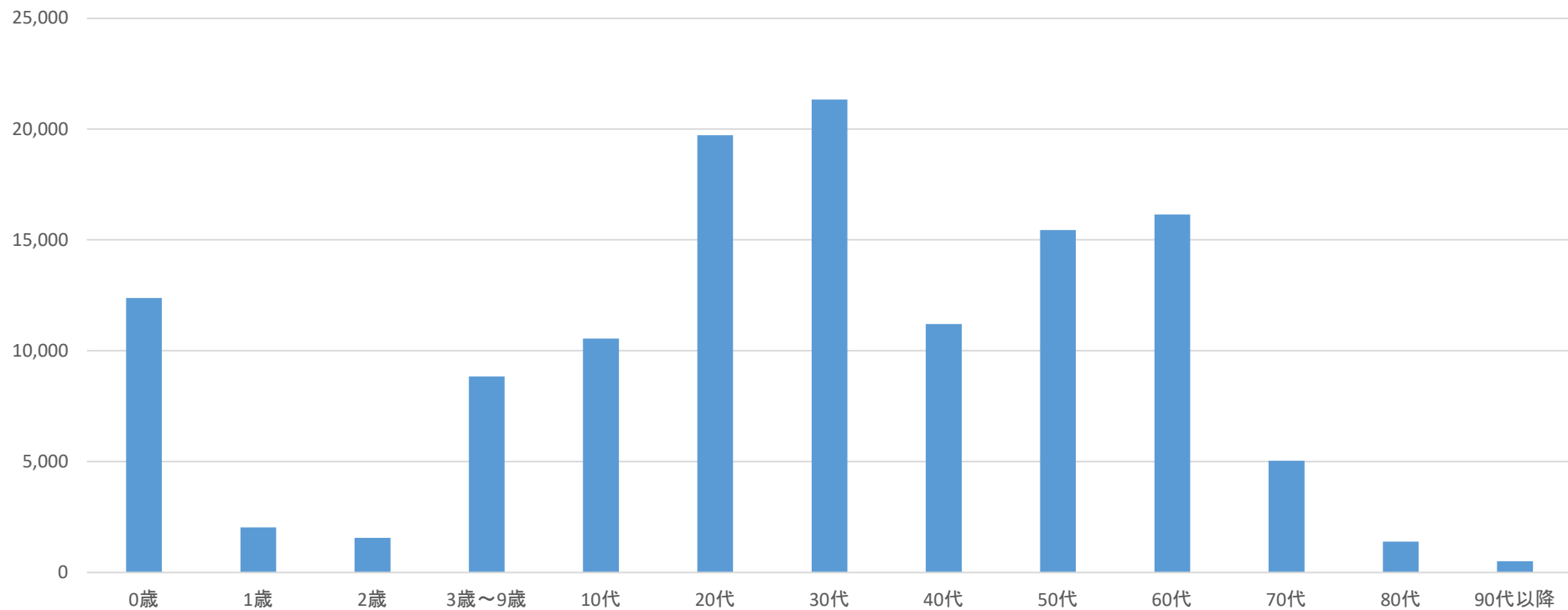
2,412 Classes



メタゲノムおよび菌株データの新規環境に対応するため、
クラスを約11追加（cryoconite等）

ヒトマイクロバイオームについて、特に詳細にメタデータをキュレーション(2021年10月時点)

	サンプル数
ヒト	約 1,050,000
ヒト (年齢情報あり)	約 122,000

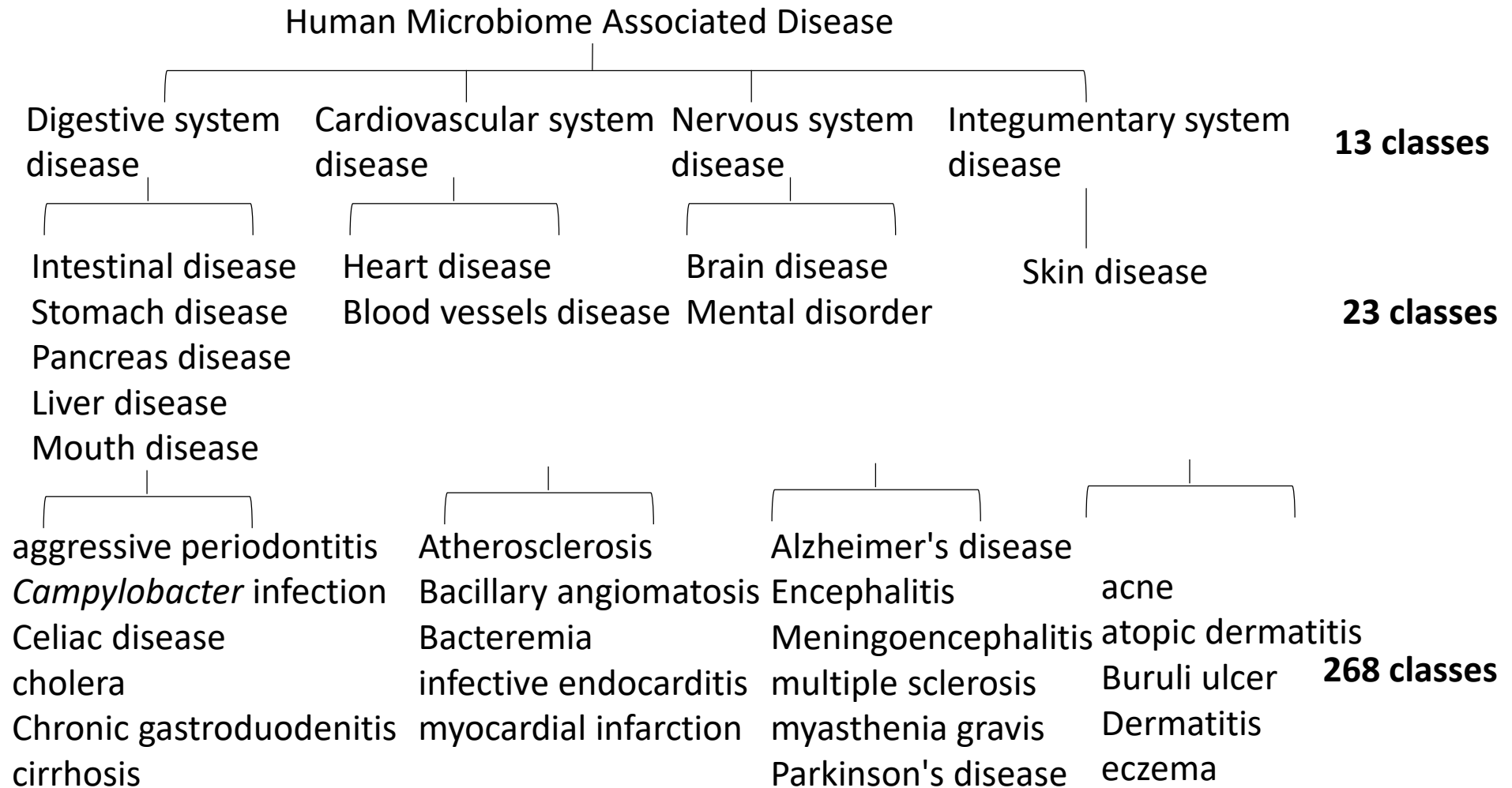


ヒトマイクロバイオームについて、特に詳細にメタデータをキュレーション(2021年10月時点)

ヒトの体の部位	サンプル数	サンプル数 (日本人)
gut (feces)	507,479	17,566
oral cavity	118,779	7,477
respiratory system	55,807	7
skin	73,859	973
vagina	57,604	64

- 抗生物質投与 (約3-4万サンプル)
- Probiotics投与 (約8千サンプル)

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

INSDC DRA/ERA/SRAで公開された疾患患者のマイクロバイオームサンプルサイズ（2019年9月時点）

サンプルサイズが100以上ある疾患は約50疾患

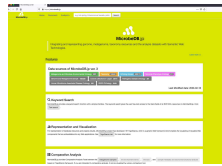
サンプルサイズが1,000以上ある疾患が右表の18疾患（未熟児と肥満は除外）

2021年10月時点

疾患患者由来のマイクロバイオームサンプル
約150,000サンプル

疾患名 (Disease Ontology)	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

微生物統合データベース「MicrobeDB.jp」の使い方



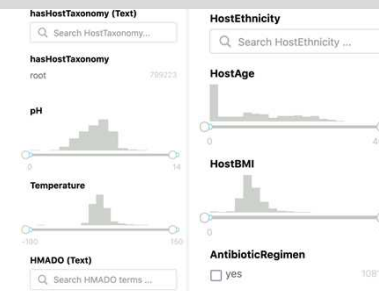
MicrobeDB.jpは、微生物のゲノム・遺伝子・系統・環境等の膨大かつ多様な情報を統合し、微生物が関与する幅広い分野において新たな知識を引き出す事ができる統合データベースです。世界中で産出されている微生物ゲノム、メタゲノム情報や系統情報などを収録し、遺伝子情報と環境情報を統合することにより、特定環境での遺伝子や種の動態、特異的遺伝子の環境分布、などを検索するだけで俯瞰することができます。

テキスト検索



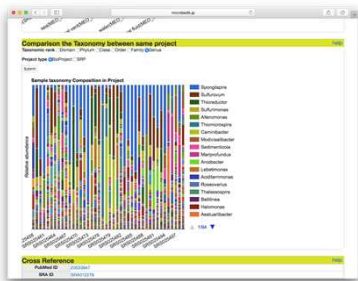
MicrobeDB.jp トップページからテキストによる検索が可能です。例えば *psbA* などの遺伝子名、*Enterococcus faecalis* などの細菌種名、hot spring や intestine など環境を表す単語などで検索する事が可能です。

ファセット検索



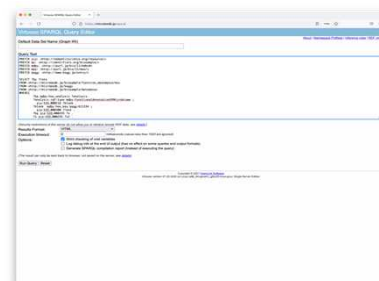
メタゲノムデータを遺伝子機能や環境要素により絞り込む事が可能です。テキストやサンプリング地点のpHや温度、被験者の地域、年齢やBMI、疾患や抗生物質使用履歴など、多様なメタデータでの絞り込みが可能です。

データ比較



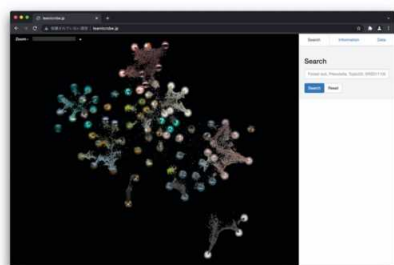
約210種類にもおよぶ解析ツールを実装しており、検索するだけで比較ゲノム解析、比較メタゲノム解析などが可能となっています。また自身で保有するデータをアップロードする事で、全データとの比較解析も可能としています。

SPARQL検索



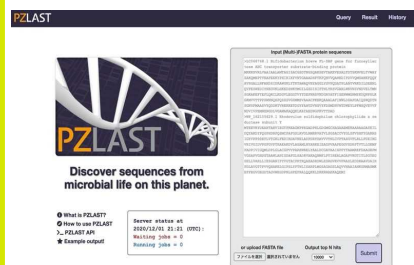
デフォルトで使う事ができる210種類のStanzaでは困難な複雑な検索を実現することができます。公開しているSPARQLエンドポイントにSPARQLによる命令を入力すれば高速に結果を得る事ができます。

群集構造によるサンプル&環境検索



微生物群集構造と環境との関係性を可視化したアプリケーションLEAを使って、微生物群集構造から環境を予測したり、類似のサンプルを検索したりすることができます。環境汚染の検出、疾患診断などに利用することが可能です。

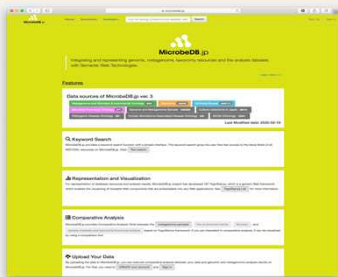
アミノ酸配列相同性検索



PZLASTは、MicrobeDB.jpに収録しているメタゲノムアミノ酸配列データ約2.5TBに対して、超高速にアミノ酸対アミノ酸の配列相同性検索を行うアプリケーションです。配列相同性を介して環境を検索することが可能です。

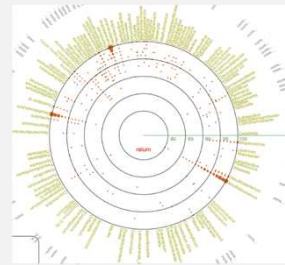
開発しているツール群

微生物統合DB「MicrobeDB.jp」



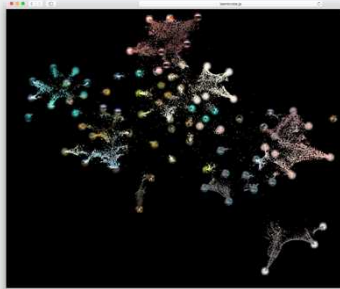
世界中から産出される微生物のゲノム・メタゲノムを中心とする各種オミックス情報を広く収集し、遺伝子機能、分類学的情報、菌株保存情報、表現型情報などの知識を整理し、ゲノム情報を核としてセマンティックウェブ技術によりデータ統合した世界最大の統合DBである。ゲノム・メタゲノム解析パイプラインも整備しており、解析した結果は速やかにDBに登録する事が可能。

群集構造解析ソフト「VITCOMIC2」



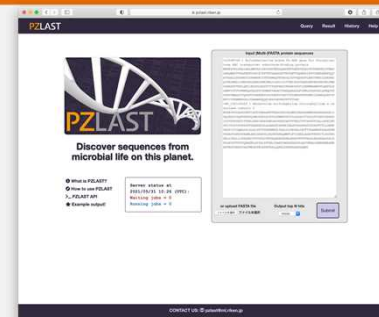
VITCOMIC2は、系統推定および系統組成描画ツールで、系統間の進化的な関係性を保持した形で可視化するソフトである。ゲノム解読されていない系統にも対応し、GPUを用いた高速な配列相同性検索、ゲノム内16S rRNA遺伝子コピー数の補正、メタゲノムからの完全長16S rRNA遺伝子配列の再構築等の様々な機能を搭載したwebサービスとなっている。

微生物群集/環境予測ソフト「LEA」



数万の微生物群集構造データを使用した機械学習により「環境」の概念を自動的に抽出し、「環境」概念と数万サンプルを同時に可視化、環境間のつながりや構造のパターンを明らかにした。過去数万サンプルすべての比較により、新規サンプルの「座標」を取得できる「微生物版GPS」。環境汚染の検出、疾患診断などに利用可能。

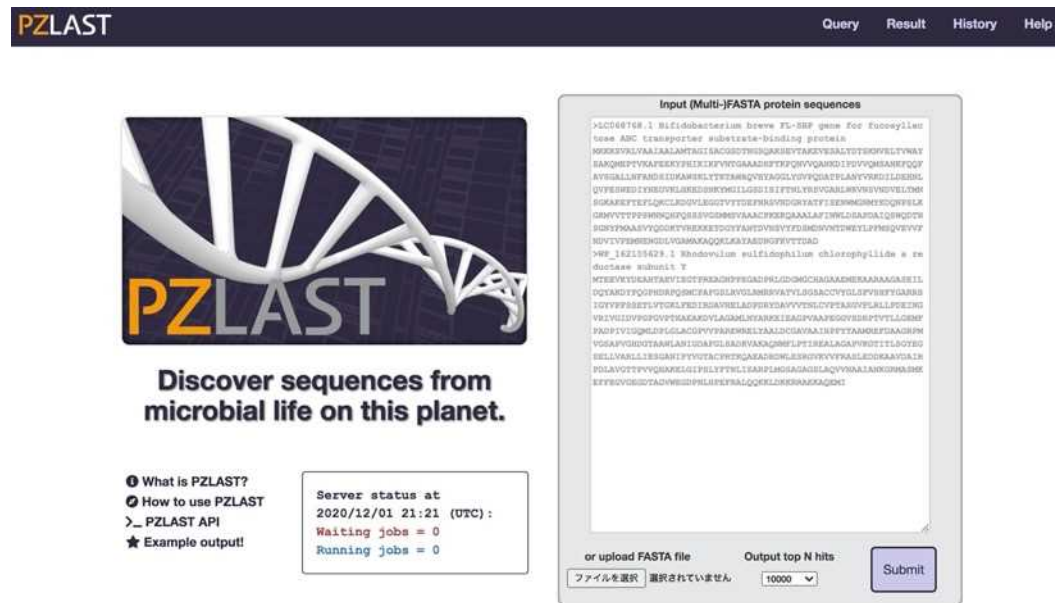
超高速メタゲノム検索ソフト「PZLAST」



PZLASTは、MicrobeDB.jpから取得した膨大なメタゲノム情報から予測したアミノ酸配列データをレファレンスとして(約2.5TB、ヒト遺伝子35万人分に相当)、極めて高速かつ高精度にアミノ酸配列の相同性検索を可能とするwebサービス。得られた結果は、アラインメント情報に加え、相同配列の存在した環境情報も可視化できる。

超高速メタゲノムデータ相同性検索システム PZLAST

<https://pzlast.riken.jp/meta>



The screenshot displays the PZLAST web interface. At the top, there is a navigation bar with 'PZLAST' on the left and 'Query', 'Result', 'History', and 'Help' on the right. The main content area is divided into two columns. The left column features a large image of a DNA double helix with the text 'PZLAST' overlaid in large, bold letters. Below the image, it says 'Discover sequences from microbial life on this planet.' Underneath this, there are four links: 'What is PZLAST?', 'How to use PZLAST', 'PZLAST API', and 'Example output!'. To the right of these links is a box showing 'Server status at 2020/12/01 21:21 (UTC):' with 'Waiting jobs = 0' and 'Running jobs = 0'. The right column is titled 'Input (Multi-FASTA protein sequences)' and contains a text area with several lines of protein sequence data. Below the text area, there are two buttons: 'or upload FASTA file' (with a file selection button) and 'Output top N hits' (with a dropdown menu set to '10000'). A 'Submit' button is located at the bottom right of the input area.

Amino acid vs amino acid sequence similarity search against protein sequences of public metagenome data (2.5TB) in MicrobeDB.jp

1 search needs approximately 5~10 minutes

PZLAST入力

- *Ideonella sakaiensis*のPET分解酵素 (PETase: poly(ethylene terephthalate) hydrolase)
- GAP38373

```
>ag:GAP38373 K21104 poly(ethylene terephthalate) hydrolase [EC:3.1.1.101] |  
(KEGG) poly(ethylene terephthalate) hydrolase (EC:3.1.1.101) (A)  
MNFPRASRLMQAAVLGGLMAVSAAATAQTNPYARGPNPTAASLEASAGPFTVRSFTVSRP  
SGYGAGTVYYPTNAGGTVGAIIVPGYTARQSSIKWWGPRLASHGFVVITIDTNSTLDQP  
SSRSSQQMAALRQVASLNGTSSSPIYGKVDTARMGVMGWSMGGGSLISAANNPSLKAAA  
PQAPWDSSTNFSSVTVP TLIFACENDSIAPVNSSALPIYDSMSRNAKQFLEINGGSHSCA  
NSGNSNQALIGKKGVAWMKRFMDNDTRYSTFACENPNSTRVSDFRTANCS
```

I. sakaiensis のPETaseは40°Cで高いPET分解活性を示す。

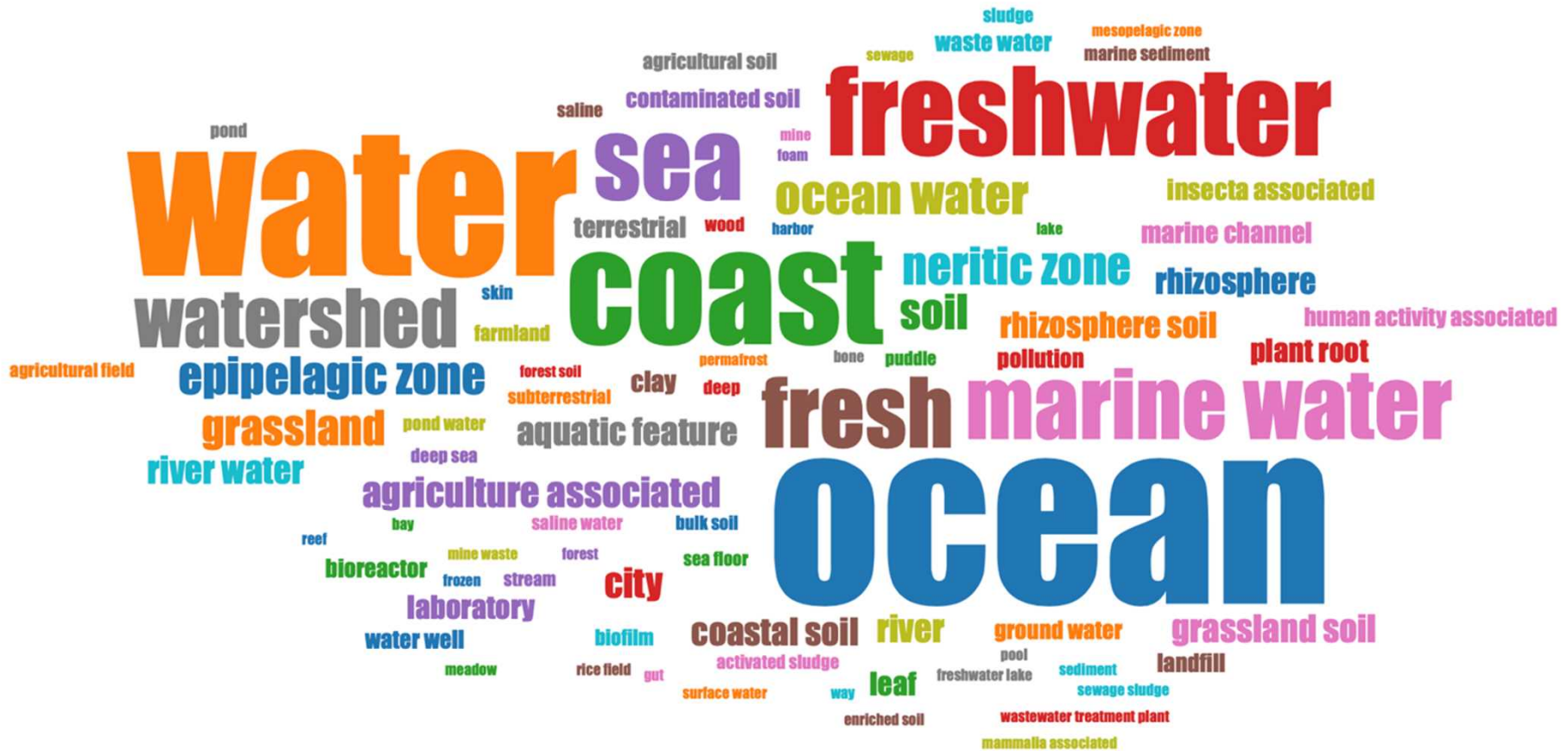
PZLAST結果

Search

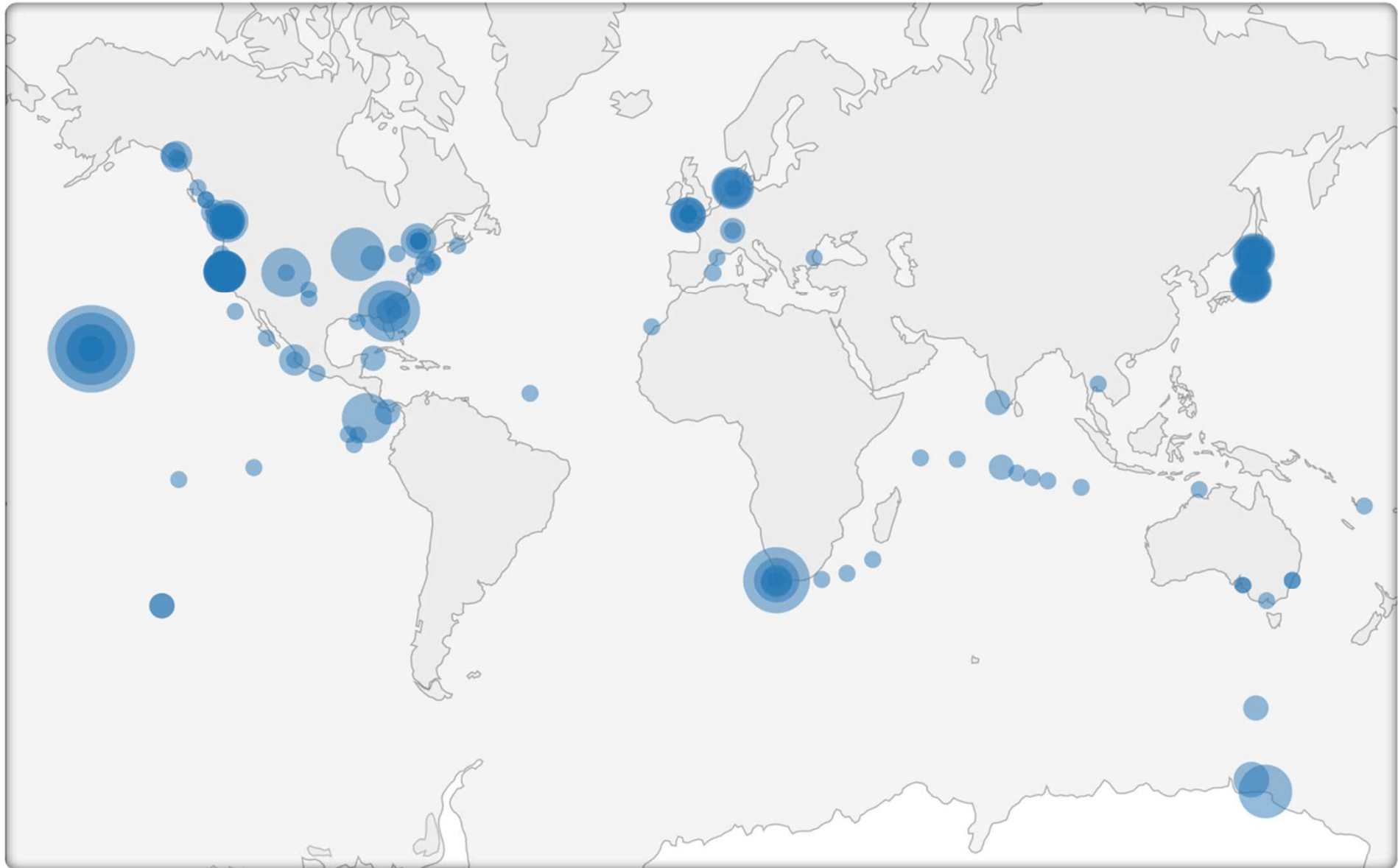
Query ID	SRR ID	BioSample ID	Query start	Query end	#Match	E-value	Alignment
test	ERR986603	SAMEA3512143	61	222	90	3.92e-44	Click to view
test	SRR1653561	SAMN03196583	22	200	95	1.30e-39	Click to view
test	ERR833297	SAMEA3313092	86	235	84	1.76e-36	Click to view
test	ERR164415	SAMEA2620066	28	167	77	1.76e-36	Click to view
test	ERR833614	SAMEA3313342	55	203	82	5.13e-36	Click to view
test	SRR1653561	SAMN03196583	9	188	94	1.14e-35	Click to view
test	SRR332522	SAMN00710591	162	289	78	1.49e-35	Click to view
test	SRR1653561	SAMN03196583	109	289	93	9.68e-35	Click to view
test	ERR833275	SAMEA3313071	42	289	90	6.93e-33	Click to view
test	ERR833280	SAMEA3313076	69	250	79	6.93e-33	Click to view

Showing 1 to 10 of 656 rows 10 rows per page

PZLAST word cloud



PZLAST global map



PZLAST結果

test	ERR833289	SAMEA3313084	45	203	50	3.13e-9	Click to view
------	---------------------------	------------------------------	----	-----	----	---------	-------------------------------

Query : test
Reference start : 56
Reference end : 212
Score : 188
Bit score : 7.70e+1
Reference sequence :
MKKIFLYNFILLSSFLYSECSLDLDYDLCMQYSQYCEWDETTGSCFEVGGGGGGGGSGSGPYEFATITESDGMNRNGPDYRD
GVLYYPLNGNPPYKNNVLSPGFGGNSASMSSWGAFYASHGFVAMTVGPNDIINDSHYQRGEGLIDGTQTIIEENSRIIGSP
VYGLIDVDSFTVSGYSMGGGASHDAALIAQANELDHVVAVISLNPTVIFEDCNLCPANS

Alignment:
Ref SGSGPYEFATITESDGMNRNGPDYRDGVLYYPLNGNPPYKNNVLSPGFGGNSASMSSWGAFYASHGFVAMTVGPNDIINDS
.:|:***:.*| **| |**|.:.||*|..*****.:*|:*....:|.***|:|:|:|.:.:***|.*:|. .
Qry ASAGPFTVRSFTVS--RPSG--YGAGTVYYPTNAGGTVGAIIVPGYTARQSSIKWWGPRLASHGFVVITIDTNSTL-DQ

Ref HYQRGEGLIDGTQTIIEENSRIIGSPVYGLIDVDSFTVSGYSMGGG-----ASHDAALIAQANEL--DHVVAVISLN-PTV
|.:.*:.*:*.:.**|.||:|*|.***.*|:|:|:| |:::.*|*|*:* |**.***|:. ||:
Qry PSSRSSQMAALRQVASLNGTSSSPIYGKVDTARMGVMGWSMGGGSLISAANNPSLKAAPQAPWDSSTNFSSVTVPTL

Ref IFEDC
|*|
Qry IFA-C

MicrobeDB.jp (SAMEA3313064)

MicrobeDB.jp Home Document Analysis e.g. hot spring, Enterococcus faecalis, psbA Search Sign Up Sign in

Result

This search term has exact match.
Now displaying stanzas in the category: **Sample** . Parameters are sample_id: **SAMEA3313084**

Metadata [Help](#)

Sample description	Water sample taken on 2004-01-09T18:47:00+00:00 from site GS017 (lat:20.52833333, long:-85.427), 0.1µm filter fraction. Samples were collected by filtering water pumped through a 20µm prefilter, then size-fractionating by serial filtration through 3µm, 0.8µm, and 0.1µm membrane filters, then finally, in some samples, through tangential flow filtration fitted with a 50kDa pore size filter to concentrate a viral fraction. This genetic information may be considered to be part of the genetic patrimony of Mexico, the country from which the sample was obtained. Users of this information agree to: 1) acknowledge Mexico as the country of origin in any country where the genetic information is presented and 2) contact the CBD focal point identified on the CBD website: http://www.cbd.int/countries/ if they intend to use the genetic information for commercial purposes.
Package	Generic.1.0
Submission date	2015-09-06T09:06:43.643
Last update	2019-04-17T15:35:22.547
Publication date	2015-09-05T00:00:00.000
Organism	marine metagenome
geo_loc_name	Caribbean Sea:Western Caribbean Marine Ecoregion:Yucatan Channel, Mexico
depth	1.6

Annotation [Help](#)

Environment	water ocean
Environment temperature	27.0
Environment PH	
Host taxonomy	
Host age	
Host BMI	
Host disease	
Host disease status	

<https://microbedb.jp> を新規タブで開く

BLASTP検索結果

(入力 : SAMEA3313064=PZLASTで見つかった配列)

[Download](#) [GenPept](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

hypothetical protein [Candidatus Marinimicrobia bacterium]

Sequence ID: [MBS29659.1](#) Length: 390 Number of Matches: 1

Range 1: 1 to 220 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
367 bits(941)	4e-124	Compositional matrix adjust.	186/220(85%)	197/220(89%)	1/220(0%)
Query 1	MKKIFLYNFILLSSFLYSECSLDLDYDL	CMQYSQYCEWDETTGSCFEVGGGGGGGSGS-	59		
Sbjct 1	MKKIFL NFIL +FL SECSLDYDLC+ YS	QYCEW ET C E+GGGG GGG	60		
Query 60	PYEFATITESDGMRNGPDYRDGVLYYPL	NGNPPYKNVVLSPGFGGNSASMSSWGA	FYASH 119		
Sbjct 61	PYEFATITESDG+RNGPDYRDGVLYYPL+GN	PYKN+VLSPGFGGNSASMSSW AFYASH	120		
Query 120	GFVAMTVGPNDEINDSHYQRGEGLIDGT	QTIIIEENSRIQSPVYGLIDVDSFTVSG	YSMGG 179		
Sbjct 121	GFVAMTVGPNDEINDSHYQRGEGLIDG	QTI+EEN+RIGSPVYGLID +SFTVSG	YSMGG 180		
Query 180	GASHDAALIAQANELDHVVAVISLNPTV	IFEDCNLCPANS 219			
Sbjct 181	GASHDAALIAQAN D+V AVISLNPTV	IFEDCNLCPAN+ 220			

[Download](#) [GenPept](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

hypothetical protein [Candidatus Marinimicrobia bacterium]

Sequence ID: [MAN35465.1](#) Length: 390 Number of Matches: 1

Range 1: 1 to 220 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
367 bits(941)	4e-124	Compositional matrix adjust.	185/220(84%)	198/220(90%)	1/220(0%)
Query 1	MKKIFLYNFILLSSFLYSECSLDLDYDL	CMQYSQYCEWDETTGSCFEVGGGGGGGSGS-	59		
Sbjct 1	MKKIFL NFIL ++FL+SECSLDYDLC+ YS	QYCEW ET C E+GGGG GGG	60		
Query 60	PYEFATITESDGMRNGPDYRDGVLYYPL	NGNPPYKNVVLSPGFGGNSASMSSWGA	FYASH 119		
Sbjct 61	PYEFATITESDG+RNGPDYRDGVLYYPL+GN	PYKN+VLSPGFGGNSASMSSW AFYASH	120		



Marinimicrobia bacterium



ARTICLE

DOI: [10.1038/s41467-017-01376-9](https://doi.org/10.1038/s41467-017-01376-9)

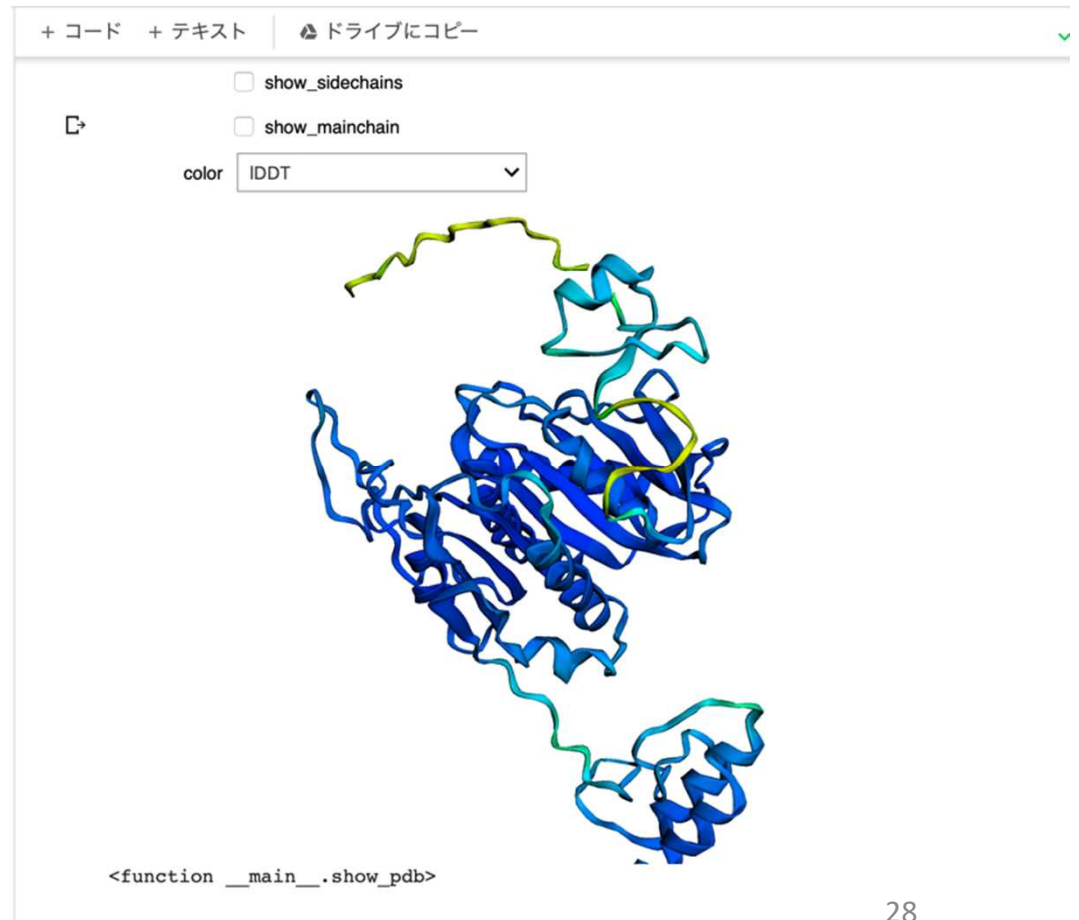
OPEN

Diverse *Marinimicrobia* bacteria may mediate coupled biogeochemical cycles along eco-thermodynamic gradients

Alyse K. Hawley ¹, Masaru K. Nobu^{2,3}, Jody J. Wright¹, W. Evan Durno⁴, Connor Morgan-Lang⁴, Brent Sage⁴, Patrick Schwientek⁵, Brandon K. Swan^{6,11}, Christian Rinke ⁷, Monica Torres-Beltrán¹, Keith Mewis⁸, Wen-Tso Liu², Ramunas Stepanauskas ⁶, Tanja Woyke ⁵ & Steven J. Hallam ^{1,4,9,10}


AlphaFold2簡易版 (MBS29659.1=BLASTで見つかった配列)

```
>MBS29659.1 hypothetical protein  
CMG39_00660 [Candidatus  
Marinimicrobia bacterium]  
MKKIFLINFILSITFLXSECSDDL DYDLDCIYYSQYCEW  
SETASLCTEIGGGGTGGGGSGSGPYEFATITESDGLR  
NGPDYRDGVLYYPLDGNAPYKNIVLSPGFGGNSASMS  
SWAAFYASHGFVAMTVGPNDEINDSHYQRGEGLIDGV  
QTILEENNRIGSPVYGLIDENSFTVSGYSMGGGASHD  
AALIAQANGYDYVKAVISLNPTVIFEDCNLCPANTYE  
GETYCICLVPELIAHSLPSLIFAGQFELNELTAYDGL  
LGQDIYANLPAETDKILFEGANSGHGFAELSNPDVEE  
KVLAFMNYFILNDDNYCETLQEIPASASQYLTFDCE  
QVMVGDLNADLIINVQDVVLTVNLVLTAGDYNISADL  
NSDNIVNVQDIILLINIILN
```



PETase (PDB: 5XH3)

Biological Assembly 1 ?



3D View: Structure | Electron Density | Ligand Interaction

Global Symmetry: Asymmetric - C1 ?
Global Stoichiometry: Monomer - A1 ?

5XH3

Crystal structure of a novel PET hydrolase R103G/S131A mutant in complex with HEMT from *Ideonella sakaiensis* 201-F6

DOI: [10.2210/pdb5XH3/pdb](https://doi.org/10.2210/pdb5XH3/pdb)

Classification: **HYDROLASE**

Organism(s): *Ideonella sakaiensis*

Expression System: *Escherichia coli* BL21(DE3)

Mutation(s): Yes ?

Deposited: 2017-04-19 Released: 2017-12-20

Deposition Author(s): Han, X., Liu, W.D., Zheng, Y.Y., Chen, C.C., Guo, R.T.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.30 Å

R-Value Free: 0.145

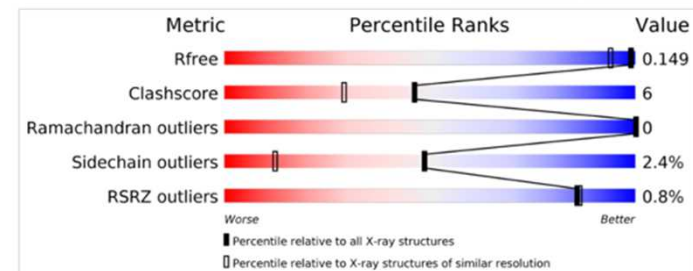
R-Value Work: 0.118

R-Value Observed: 0.119

wwPDB Validation ?

3D Report

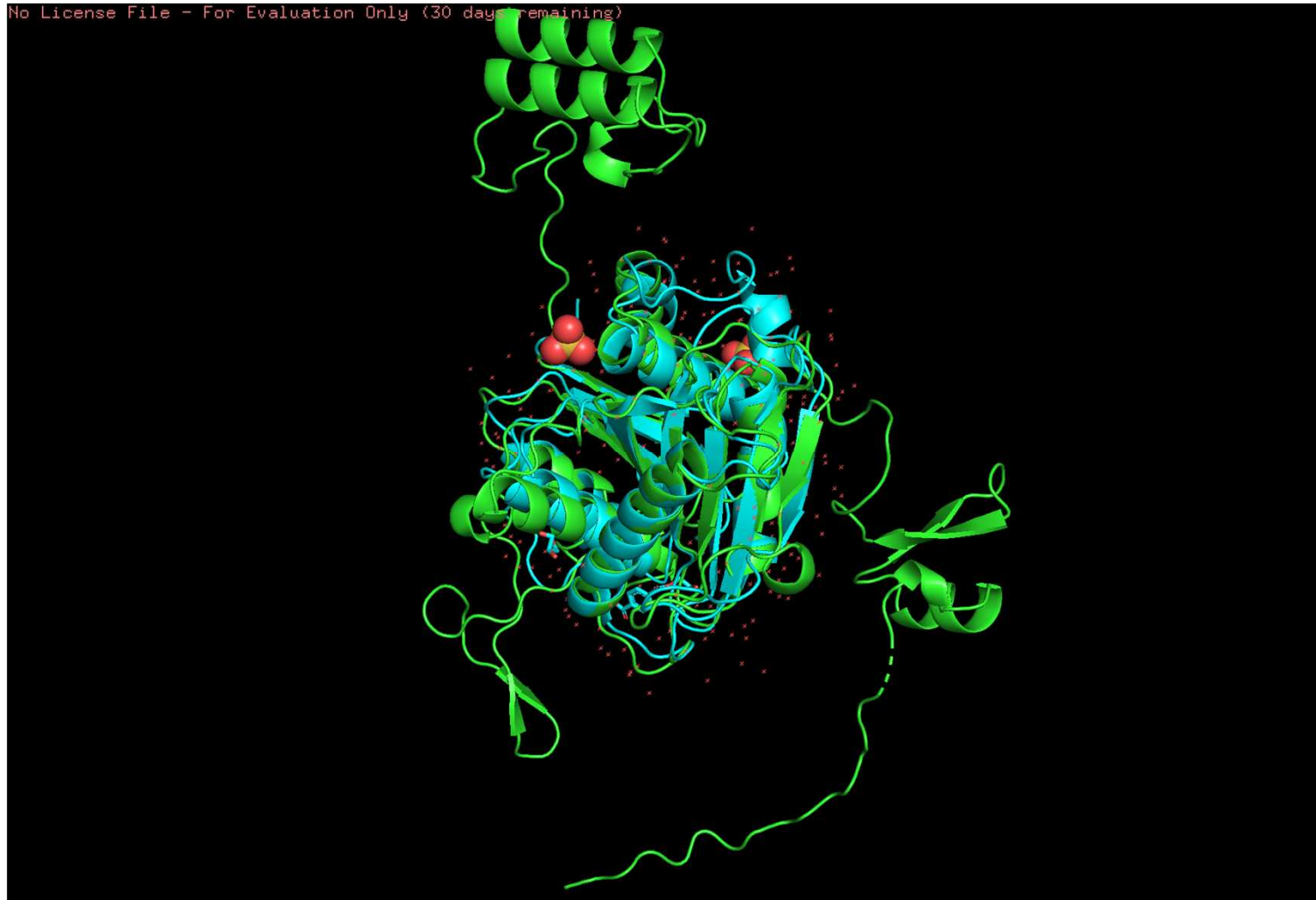
Full Report



Display Files

Download Files

PyMOLでのalign(PETaseとBLAST配列) (align MBS29659.1, 5XH3)



AlphaFold2は5XH3を学習しているはずなので、当たり前の結果
(単なるhomology modeling)

PyMOLでのalign(PZLAST配列とBLAST配列) (align SAMEA3313064, MBS29659.1)



PyMOLでのalign(PETaseとPZLAST配列) (align SAMEA3313064, 5XH3)



PZLASTの応用例

- PETase (GAP38373)をクエリとしてblastpをしても、*M. bacterium*のMBS29659.1を検索する事は困難。
- PZLASTではPETaseをクエリとしてSAMEA3313064を検索できた。
- SAMEA3313064をblastpするとCMG39_00660にヒットした。MBS29659.1のアノテーションは“hypothetical protein”であった。
- AlphaFold2の3D予測では、SAMEA3313064、MBS29659.1ともに3D構造はPETaseとほぼ一致した。
- PETaseと3D構造が類似のアミノ酸配列を得る事ができた。
- SAMEA3313064の採取地点の温度は27℃。この酵素は27℃で活性を持つと考えられる。

主たる共同研究者

国立遺伝学研究所

黒川 顕：微生物DBにおける研究統括

中村保一：DDBJ&植物統合DBの連携

森 宙史：統合DB構築、オントロジー&Stanza開発

藤澤貴智：セキュリティの高度化

東 光一：機械学習による統合データ解析技術開発

基礎生物学研究所

内山郁夫：オーソログデータの構築

千葉啓和：オーソログ情報のRDF化、Stanza開発 (DBCLS)

西出浩世：オーソログデータの構築補助

東京工業大学

山田拓司：ヒトマイクロバイオームデータのメタデータ構築

千葉大学

高橋弘喜：真菌類ゲノム・菌株・オミックス情報の収集と高度化

矢口貴志：真菌類分類情報の整理

MicrobeDB.jp ver.3

(<https://microbedb.jp/>)

The screenshot shows the MicrobeDB.jp website in a browser window. The browser's address bar displays "microbedb.jp". The website has a yellow header with the MicrobeDB.jp logo and navigation links: Home, Document, Analysis, and a search bar containing "e.g. hot spring, Enterococcus faecalis, psb." with a "Search" button. On the right of the header are "Sign Up" and "Sign in" links. The main content area features the MicrobeDB.jp logo and a tagline: "Integrating and representing genome, metagenome, taxonomy resources and the analysis datasets with Semantic Web Technologies." Below this is a "Features" section with several sub-sections:

- Data sources of MicrobeDB.jp ver. 3**: A grid of colored boxes representing different data sources and their counts:
 - Metagenome and Microbes Environmental Ontology: 2401
 - Taxonomy: 129342
 - Ortholog Groups: 4203173
 - Microbial Phenotype Ontology: 277
 - Genome and Metagenome Sample: 1920339
 - Culture collections in Japan: 38414
 - Pathogenic Disease Ontology: 387
 - Human Microbiome Associated Disease Ontology: 308
 - KEGG Orthology: 22421

Last Modified date: 2020-02-16
- Keyword Search**: A section with a magnifying glass icon and text: "MicrobeDB.jp provides a keyword search function with a simple interface. The keyword search gives the user free-text access to the literal fields of all RDF/OWL resources on MicrobeDB.jp. Click [Text search](#)."
- Representation and Visualization**: A section with a bar chart icon and text: "For representation of database resources and analysis results, MicrobeDB.jp project has developed 197 TogoStanza, which is a generic Web framework which enables the visualizing of reusable Web components that are embeddable into any Web applications. See [TogoStanza List](#) for more information."
- Comparative Analysis**: A section with a list icon and text: "MicrobeDB.jp provides Comparative Analysis Tools between the [metagenome samples](#), the environment terms, the taxa, and [sample metadata and taxonomic/functional analysis](#) based on TogoStanza framework. If you are interested in comparative analysis, it can be visualized by using a comparison tool."
- Upload Your Data**: A section with an upload icon and text: "By uploading the data to MicrobeDB.jp, you can execute comparative analysis between your data and genomic and metagenomic analysis results on MicrobeDB.jp. For that, you need to [CREATE your account](#) and [Sign in](#)."