

マイクロバイーム研究を先導する 微生物統合データベース



Microbiome Datahub

<https://mdatahub.org/>



Microbiome Datahubは、INSDC由来の約21万件のMAGデータとRefSeq由来の約2.6万件のリファレンス単離菌ゲノムおよび関連するBioProject情報を収録し、系統や品質、宿主の系統等でMAGを絞り込み検索して、配列データやアノテーション情報等をWebページやAPI経由でダウンロードできるようにしています。

MAGは培養が困難な微生物のゲノムをも明らかにできることから、微生物の多様性を記述する上で、近年、注目されています。Microbiome DatahubにはINSDCで公開されているMAGを収集し、品質管理を行った上で、高品質なMAGデータ 21万件について、微生物ゲノム自動アノテーションツールDFASTで解析した遺伝子アノテーション情報とその配列情報、(2) 由来した系統やサンプル等のメタデータ、(3) 微生物ゲノム比較解析システムMBGDで同定されたオーソログ情報、(4) Bac2Featureで予測された細胞形態、グラム染色性、胞子形成能や運動性の有無、ゲノム情報、生育条件等の環境特性など、その微生物のさまざまな特徴や性質の情報を掲載しています。またAPIを用いれば、これらの情報をまとめて取得することもできます。

Microbiome Datahubの一番の特徴は、その収載データの多様さです。MAGデータを収載するデータベースはほかにもいくつか存在しますが、その多くは主にヒトの腸内や皮膚などに生息するマイクロバイーム由来のMAGデータを収集しています。一方、Microbiome Datahubは、ヒトの腸内や皮膚などに限らず、土壌、海水、淡水、温泉、堆積物、空中などさまざまな環境サンプルを対象としており、多種多様な微生物系統種のMAGデータが収載されています。

Microbiome Datahubとあわせて公開されているPZLAST-MAGを用いると、Microbiome Datahub に収載されているMAGに由来する約4.5億個のタンパク質配列に対し、高速な相同性検索をすることができます。

< Microbiome Datahubの収載データ数 > (2025年6月5日 現在)

単離菌由来ゲノム	26,076件
Metagenome-Assembled Genome (MAG)	218,653件

Metagenome-Assembled Genome (MAG) : 土壌や環境水、糞便や皮膚組織などから微生物を単離・培養することなく網羅的に解析したゲノム配列を「メタゲノム(Metagenome)」と呼びます。メタゲノムには実験室では培養できない多様な微生物種に由来する断片化したゲノム配列が含まれています。これらの断片配列をコンピュータ上でつなぎ合わせ、元の微生物ゲノムを再現した仮想配列を「Metagenome-Assembled Genome (MAG)」と呼びます。MAGからは、これまで知られていなかった微生物が明らかになるとともに、その微生物が有する新規な酵素遺伝子など、産業応用上においても有用な情報が含まれていることが期待されています。

Microbiome Datahubは、統合化推進プログラムの研究開発課題「マイクロバイーム研究を先導するハブを目指した微生物統合データベースの特化型開発」(研究代表者 情報・システム研究機構 国立遺伝学研究所 森 宙史 准教授)の一環で開発・提供しています。

参考文献

Mori H. *et al.*, Microbiome Datahub: an open-access platform integrating environmental metadata, taxonomy, and functional annotation for comprehensive metagenome-assembled genome datasets. *Microbiome* (2026)
DOI: 10.1186/s40168-026-02385-x

Bac2Featureで予測された表現型

MBGD	tf	count	tf	description
1.76	1	K02074		ABC transporter-ATP-binding protein
1.77	15			ABC transporter-ATP-binding protein
2.17	1	K02477		Response regulator protein domain DNA-binding protein
2.42	1			Response regulator protein domain DNA-binding protein
3.151	6			PAS domain-containing sensor histidine kinase
3.152	1			PAS domain-containing sensor histidine kinase
3.180	1			PAS domain-containing sensor histidine kinase
6.10	1	K01897		CoA synthetase, long-chain-fatty acid:CoA ligase
6.2	3	K01895		CoA synthetase, long-chain-fatty acid:CoA ligase
6.32	1	K00666		CoA synthetase, long-chain-fatty acid:CoA ligase

絞り込み検索タブ **MBGDで同定されたオーソログ情報**



国立研究開発法人科学技術振興機構 (JST)
情報事業部 NBDC事業推進室

Tel: 03-5214-8491 E-mail: nbdc-dicp@jst.go.jp



Creative Commons 表示4.0国際ライセンス

(2026年4月)

Microbiome Datahubの使い方



Microbiome Datahub

<https://mdatahub.org/>



Microbiome Datahub で取得可能なデータ

- ・ MAG のメタデータのTSV
- ・ MAG の塩基配列のfasta
- ・ MAG がコードする遺伝子配列のfasta
- ・ MAG がコードする遺伝子のアミノ酸配列のfasta
- ・ MAG がコードする遺伝子にアサインされたMBGD Ortholog / KEGG Orthology IDリストのTSV

データの取得方法

- 1) API を使い URL で MAG の ID を指定してダウンロード（詳細はAPI マニュアル (<https://mdatahub.org/apimanual>)を参照)
- 2) Web ブラウザ上からダウンロード（下記参照）。

使い方の解説はDocument をクリック
APIの使い方はAPI Manual をクリック

The screenshot shows the Microbiome Datahub web interface. On the left, there is a sidebar with search filters: 'Search Keyword', 'SUBMIT', 'Genome taxon', 'MAG completeness' (set to 50), 'Host taxon', 'Quality' (with star ratings from 1 to 5), and 'Data Source' (with options for INSDC MAG and Isolate Genome). The main area displays a list of MAGs under the 'GENOME' tab. The list includes columns for 'Select', 'MAG ID', 'Name', 'Environment', 'Host taxon', 'BioSamples', and 'Data size (GB)'. The first entry is 'Bombilactobacillus bombi' with ID GCF_964340835.1 and a size of 0.533 MB. Other entries include 'Pedobacter helvus', 'Pedobacter urelyticus', 'Xanthomonas vasicola', 'Kwoniella pini CBS 10737', and 'Methylocapsa polymorpha'. A 'Download' button is visible at the top right of the list.

↑ クリックで各MAGの詳細ページに移動

MAGの詳細ページで閲覧できる情報

- ①MAGメタデータ：INSDCでのBioProject・BioSample・ゲノムID・系統名等のメタデータ
- ②MAGのMBGD Orthology / KEGG Orthology 組成：MAGが持つ遺伝子のMBGD Orthology IDとKO IDを独自の配列相同性検索 (PZLAST-MAG) によってアサインしたもの (※)
- ③MAGのDFAST：ゲノムアノテーションツールDFASTの統計量
- ④DFAST_QC：クオリティチェックツールDFAST_QC (CheckM)の結果
- ⑤Bac2Feature 結果：Phenotype推定ツールBac2Featureの結果

※ 個別のMAGのMBGD Orthology ID / KO ID組成のTSVは、ダウンロードAPIを用いて取得できる。
例えば、MAG ID: GCA_029762515.1であれば、下記URLから取得可能
https://mdatahub.org/api/genome/mbgd/GCA_029762515.1

