

ライフサイエンスデータベース統合推進事業
(統合化推進プログラム)

研究開発中間報告書

「マイクロバイオー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」

森 宙史

国立遺伝学研究所 情報研究系 准教授



©2023 森 宙史(国立遺伝学研究所) Licensed under CC BY 4.0

目次

§1. 研究開発実施内容の概要	4
§2. 研究開発実施体制	5
1. 各グループの担当項目	5
(1) 森グループ（遺伝研）	5
(2) 内山グループ（基生研）	5
(3) 山田グループ（東工大）	5
(4) 松井グループ（東大から京大へ異動）	5
2. 有識者会議等.....	6
(1) 会議概要.....	6
(2) 開催歴	6
§3. 研究開発の実施計画、実施内容及び成果.....	7
1. 研究開発対象のデータベース・ツール.....	7
(1) データベース	7
(2) ツール等.....	7
2. 中間評価時までの達成目標及び達成状況.....	8
(1) 当初計画における中間評価時までの達成目標と達成状況	8
(2) 期間中に追加・削除・変更した実施計画及び達成目標.....	9
3. 実施内容	10
(1) 中間評価時までに実施した研究開発の内容	10
(2) 若手研究者の多様なキャリアパスの支援.....	14
§4. 主要なデータベースの利活用状況.....	15
1. アクセス数	15
(1) 実績	15
(2) 分析	15
2. データベースを利用して得られた研究成果・産業応用の例.....	15
3. その他.....	16
§5. 今後の研究実施計画	17
§6. 自己評価.....	18
§7. 外部発表等	19
1. 原著論文発表.....	19
(1) 論文数概要	19

(2) 論文詳細情報	19
2. その他の著作物（総説、書籍など）	20
3. 国際学会発表及び主要な国内学会発表.....	20
(1) 概要	20
(2) 招待講演.....	20
(3) 口頭講演.....	21
(4) ポスター発表	22
4. 知財出願.....	23
(1) 出願件数.....	23
(2) 一覧	24
5. 受賞・報道等.....	24
(1) 受賞	24
(2) メディア報道	24
(3) その他	24
§8. 研究開発期間中の活動.....	25
1. 進捗ミーティング	25
2. 主催したワークショップ、シンポジウム、アウトリーチ活動等	27

§1. 研究開発実施内容の概要

本研究開発では、爆発的な勢いでデータ蓄積が進み利用者からのニーズが多くかつ MicrobeDB.jp の強みでもあるマイクロバイオーームデータに研究開発の焦点を絞り、新たなデータの統合化の範囲をマイクロバイオーームに特化して拡大すると共に検索機能と解析機能の追加および洗練化を行うことで、微生物分野のエンサイクロペディア MicrobeDB.jp からマイクロバイオーーム研究の国際的なデータハブ Microbiome Datahub へと発展させる。具体的には、①ゲノム・メタゲノムの統合化、②リソース(モノ)との連結、③データ爆発への対応、④効率的な逐次 DB 更新・運用システムの"再"構築、⑤キラーデータセット・アプリケーションの開発、⑥マルチオミックス連携の 6 項目を中心に研究開発を実施する。中間評価までに、①については INSDC DDBJ/ENA/NCBI Assembly DB 由来の約 21 万の MAG データについて、アノテーションとクオリティチェックを行い、Microbiome Datahub で公開し検索・ダウンロード可能にした。②については、菌株保存機関の JCM と NBRC から菌株データを RDF ポータル経由で提供いただき、菌株データ中に存在する Phenotype データを抽出し、京大グループで開発している Phenotype 推定ツール Bac2Feature と連結可能にした。③については、爆発的な勢いで増加を続けるマイクロバイオーームデータに対応するために、リードデータの系統組成解析パイプラインで使用するツール群を見直し、大幅な高速化を実現した。大量の MAG に基生研グループで開発しているオーソログクラスタ DB である MBGD のオーソログクラスタ ID を高速にアサインする手法のテスト版を遺伝研グループと共同で開発した。④については、LLM 技術を応用した論文等のテキストデータからのメタデータ抽出・アノテーションツール EMBERS を東工大グループと遺伝研グループで共同で開発し、マイクロバイオーームのメタデータの高速アノテーションを実現した。⑤については、Microbiome Datahub のデータをユーザが自由に取得できるように、データ取得 API を開発し公開した。⑥については、Microbiome Datahub の MAG データをプロテオームグループや NEDO プロジェクト等で活用し連携を進めている。

§2. 研究開発実施体制

1. 各グループの担当項目

(1) 森グループ(遺伝研)

プロジェクト全体の研究統括を行うとともに、①ゲノム・メタゲノムの統合化(【研究開発実施項目 1-1】)、②リソース(モノ)との連結(【研究開発実施項目 1-2】)、③データ爆発への対応(【研究開発実施項目 1-3】)、④効率的な逐次 DB 更新・運用システムの"再"構築(【研究開発実施項目 1-4】)、⑤キラーデータセット・アプリケーションの開発(【研究開発実施項目 3-1】)、⑥マルチオミックス連携(【研究開発実施項目 1-6】)の6項目の研究計画について、①、③、⑤、⑥について主体となって研究開発を行い、②と④については、主体となる共同研究グループと連携して研究開発を行った。

(2) 内山グループ(基生研)

③のオーソログアサインメントパイプラインの開発と実行について主体となって研究開発を行った(【研究開発実施項目 2】)。

(3) 山田グループ(東工大)

④のヒトマイクロバイオーームに関する研究開発について、主体となって行った(【研究開発実施項目 3-1】)。また、①について Microbiome Datahub の論文化のために、遺伝研グループと共同で MAG データの解析や国際連携を行った(【研究開発実施項目 3-2】)。

(4) 松井グループ(東大から京大へ異動)

②の微生物の Phenotype データに関する研究開発について、主体となって行った(【研究開発実施項目 4】)。

2. 有識者会議等

(1) 会議概要

名称	Microbiome Datahub 利活用推進会議
目的	Microbiome Datahub のユーザビリティ向上および機能充実のための、DB ユーザインターフェースの改善案の提案、テスト開発した DB 機能やツールの改良すべき点等のアドバイス等の意見を出していただくための有識者会議
委員数	5 人

(2) 開催歴

年月日	場所	主な議題・指摘事項等
2024 年 3 月 14 日	日本ゲノム微生物学会年 会会場(千葉県)	Microbiome Datahub で扱うべき MAG と ゲノムデータの種類について。MGnify の MAG データについては Microbiome Datahub 内で検索できるようにすべきとの 提案があった。
2024 年 6 月 27 日	国内版 BioHackathon BH24.6 会場(群馬県)	Microbiome Datahub の検索インターフェ ースについて。さらに開発が必要な機能や 現状のインターフェースの挙動のバグ等につ いて指摘・提案があった。
2024 年 8 月 8 日	日本細菌学会総会会場 (北海道)	Microbiome Datahub の検索インターフェ ースとデータ取得 API について。現状のイ ンターフェース上で多数の MAG データを選 択しダウンロードすることが困難な点等につ いて改善すべきとの指摘があった。

§3. 研究開発の実施計画、実施内容及び成果

1. 研究開発対象のデータベース・ツール

(1) データベース

① 主要なもの

正式名称	略称	概要
Microbiome Datahub		https://mdatahub.org/

② 上記以外のもの

正式名称	略称	概要
Microbial Genome Database for Comparative Analysis	MBGD	https://mbgd.nibb.ac.jp/

(2) ツール等

正式名称	略称	概要
Metagenome and Microbes Environmental Ontology	MEO	https://bioportal.bioontology.org/ontologies/MEO
Latent Environment Allocation	LEA	http://leamicrobe.jp
PZLAST		https://pzlast.nig.ac.jp

※データベース、ツールの詳細は別紙参照。

2. 中間評価時までの達成目標及び達成状況

(1) 当初計画における中間評価時までの達成目標と達成状況

達成目標	達成状況
・MAG データについて、Microbiome Datahub (仮)に 6 万 5 千種を目安にゲノムデータとして格納し、検索可能にする。	Microbiome Datahub に 218,653 MAGs のゲノムデータを格納し検索可能にした。
・Microbiome Datahub (仮)のマイクロバイオームの系統組成解析パイプラインについて、5 倍以上の速度向上を行う。	MicrobeDB.jp では 16S rRNA 遺伝子配列を用いた系統組成解析パイプラインを用いていたが、Microbiome Datahub では k-mer 組成を用いた系統組成解析パイプラインを新規に開発し、8 倍以上の速度向上を達成した。
・Microbiome Datahub (仮)のマイクロバイオームの遺伝子組成解析パイプラインについて、解析サンプル選択方法等の工夫も含め、トータルで 10 倍以上の速度向上を行う。	MicrobeDB.jp では KEGG アミノ酸配列データへの DIAMOND での配列類似性検索による遺伝子組成解析パイプラインを用いていた。深層学習ベースの DeepGO による遺伝子組成推定によって、10 倍以上の速度は達成できたが、アノテーションされた遺伝子機能の精度に問題があるため、Microbiome Datahub では高精度な KOfamScan を MAG データに対して実行することで MAG 単位で遺伝子機能組成を解析している。
・MAG に対する機能オーソロググループアノテーション機能を開発し、MAG の遺伝子に対して機能オーソログを高速にアサイン可能にする。	基研研グループの MBGD のデフォルトオーソログクラスタに対して遺伝研グループが開発した非常に高速な配列類似性検索ツール PZLAST を応用することで、大量の MAG データに対しても高速に MBGD オーソログ ID をアサインできるようにした。
・菌株に紐づく微生物 Phenotype 情報の RDF 化を行い、Microbiome Datahub (仮)から Phenotype 情報を検索可能にする。	日本を代表する菌株保存機関の JCM および NBRC の菌株データについて、NBDC RDF ポータルから菌株 RDF データを取得し、Phenotype 情報について微生物 Phenotype オントロジーをアノテーション付けし、Microbiome Datahub で検索するための基盤を整えた。MAG を中心とした Microbiome Datahub に菌株データを紐付けるために、今後京大グループが開発している Phenotype 推定ツール Bac2Feature を MAG に対して実行した結果を Microbiome Datahub で検索可能にする。
・Microbiome Datahub (仮)で行っている、MEO や HMADO 等のオントロジーを用いたメタデータのアノテーションとチェックにかかる時間を、オントロジー自動アノテーションツールの精度向上により 40%低減する。	LLM 技術を応用した論文等のテキストデータからのメタデータ抽出・アノテーションツール EMBERS を開発し、MEO や HMADO 等のオントロジーに自動でアノテーションできるようにした。EMBERS を用いることで、メタデータのアノテーションとチェックにかかる時間は

	MicrobeDB.jp の頃よりも 70%以上低減できた。
・Microbiome Datahub (仮)のデータを使ったキラーデータセット・アプリケーションを 2 つ以上開発し公開する。	Microbiome Datahub のデータをユーザが自由に取得できるように、データ取得APIを開発し公開した。現状メタゲノム BioProject のメタデータや系統組成データ、MAG のメタデータ、MAG の配列データがデータの ID を指定した URL にアクセスするだけで取得可能である。MicrobeDB.jp/Microbiome Datahub のメタゲノムアミノ酸配列データも活用して、ユーザが入力したメタゲノムアミノ酸配列データから微生物群集が生息する環境温度を推定するツール Metagenomic Thermometer(メタゲノム温度計)を開発し論文発表と web アプリケーションの公開を行なった。 マイクロバイオームの配列データに対する検索ツールである PZLAST について、約 21 万 MAG 由来の遺伝子のアミノ酸配列データに対する配列類似性検索を行える機能 (PZLAST-MAG)を開発し、現在論文執筆中である。2024 年度中に PZLAST-MAG を公開し論文投稿予定である。Python 言語と LLM の ChatGPT の API を用いた論文からのメタデータ抽出ツール EMBERS についても、論文から抽出し整理したメタデータと BioSample との対応関係を整理し、BioSample をクラスタリングすることでマイクロバイオームのメタ解析に使用できるサンプルメタデータ俯瞰ツールを開発した。EMBERS の論文を現在執筆中であり、論文公開と同時に公開予定である。
・これら開発を通して、Microbiome Datahub (仮)の年間訪問者数を 80,000 人以上にする。	MicrobeDB.jp は 2024 年度初めに閉鎖し、Microbiome Datahub は現在テスト版のみ公開し 2024 年中に正式版を公開予定であり、まだ訪問者数を評価できる段階ではない。

(2) 期間中に追加・削除・変更した実施計画及び達成目標

東大グループの研究分担者の異動に伴い、2024 年度から東大グループが無くなり、京大グループを新規に追加した。研究分担者の所属以外は研究開発参加者および研究開発内容に大きな変更はない。

3. 実施内容

(1) 中間評価時まで実施した研究開発の内容

1. 「ゲノム・メタゲノムの統合化」

メタゲノム由来の Metagenome Assembled Genome (MAG)について、INSDC DDBJ/ENA/NCBI Assembly DB および MGnify、GlobDB から MAG 配列およびメタデータを取得する処理の自動化を行い、データ更新を容易に行えるようにした。それらの MAG データのうち、INSDC DDBJ/ENA/NCBI Assembly DB 由来の約 21 万 MAG 配列データと、MGnify 由来約 31 万 MAG 配列データについて、ゲノムにコードされている遺伝子情報や系統情報等を得るために、遺伝研グループで開発した微生物ゲノムおよび MAG の自動アノテーションツールである DFAST および、MAG の系統推定と品質評価を行える DFASTQC [論文 1]を用いたワークフローで解析を行った。結果として得られたアノテーション済みの MAG 配列について、INSDC DDBJ/ENA/NCBI Assembly DB 由来の約 21 万 MAG データについて、先行して Microbiome Datahub で公開し検索・ダウンロード可能にした。加えて、単離菌のゲノムについても NCBI RefSeq DB から全種の代表ゲノム配列を取得し、DFAST および DFAST_QC で遺伝子アノテーションをした。それらのゲノム・MAG について、Average Nucleotide Identity (ANI)を高速に計算するツール skani を用いて配列間の ANI を計算し、ほぼ同一の配列群を一つにまとめる作業を現在行なっている。MAG データについては、EMBL で Peer Bork グループが中心になって進めている国際的な環境メタゲノムプロジェクト TREC(Traversing European Coastlines)に山田グループと森グループが日本側の担当者として参画し、土壌や海水メタゲノム・MAG データの記述方法や解析方法等を EMBL グループと話し合いを行った。また、ゲノム配列から原核生物の系統名を定義する動きの中心となっており MAG の系統 DB としてもスタンダードになっている、GTDB の Scientific Advisory Board に遺伝研グループの研究参加者の中村教授がなり、MAG データの INSDC の DB への登録推進や MAG のメタデータ項目、系統名の標準化等について議論し、Microbiome Datahub のメタデータ項目や系統情報等、TREC や GTDB 等と国際連携しつつ開発を進めている。

2. 「リソース(モノ)との連結」

菌株保存機関の JCM と NBRC から菌株データを RDF ポータル経由で提供いただき、菌株データ中に存在する細胞の形や酸素要求性、生育温度等の Phenotype データを抽出し、松井グループで開発している Phenotype 推定ツール Bac2Feature のデータ項目と対応付け、系統ごとの Phenotype リファレンスとして Bac2Feature で使用できるデータにした。Bac2Feature については、系統名や配列データから Phenotype を予測する際の予測精度の向上を行う上で、(i)最適な形質推定手法の選択、(ii)系統シグナルと進化距離が形質予測に及ぼす影響、が大きな問題になる。まず(i)については相同性ベース・系統群ベース・系統樹ベースの三つのアプローチに基づく推定手法を改めて設計・実装し、予測精度を比較した。その結果、連続値で表される形質については系統樹ベースの方法が、カテゴリ値で表される

形質については相同性ベースの方法が最も良い精度を示すことがわかった。(ii)については詳細な進化解析と組み合わせた性能検証を行った。その結果、生育温度や GC 含量、芽胞形成能、といった形質については進化的に非常によく保存されている一方で、細胞のサイズや至適 pH、細胞形状といった形質は進化的に非常に変化しやすいことがわかった。そしてこれらの保存度はそれぞれの形質の予測精度を非常によく説明できることが明らかになった。以上の成果は、Phenotype 予測パイプラインの改善に直接役立つのみならず、一般的な微生物表現型解析の基盤となる重要な知見であると考えている。今後 Bac2Feature の論文出版と並行して Microbiome Datahub の数十万の MAG データに対して Bac2Feature を実行し、各 MAG の Phenotype 推定を行い Microbiome Datahub に収録する。

菌株リソースと MAG とを連結する上で、ゲノム配列および系統名で個々の菌株と MAG とを対応付ける必要がある。JCM とは菌株のゲノム配列の相同性検索について連携し、JCM のゲノム解読済みの菌株に対してゲノム配列やアミノ酸配列で BLAST 検索を出来るよう JCM 側でシステム開発を行う上での技術アドバイスをを行い、2024 年 9 月に JCM のページで BLAST 検索機能が公開された。NBRC とは昨年度から NEDO グリーンイノベーション基金事業のプロジェクトを遺伝研の森も参加して共同で開始し、CO₂ 固定微生物を用いたバイオものづくりについて強固に連携しており、Microbiome Datahub 中の単離菌ゲノムおよび MAG と NBRC の菌株との配列や系統名による連結を進めている。

3. 「データ爆発への対応」

爆発的な勢いで増加を続けるマイクロバイームデータに対応するために、リードデータの系統組成解析パイプラインで使用するツール群を見直し、16S rRNA 遺伝子を用いた系統推定法から k-mer を用いた系統推定法に切り替えることで、MicrobeDB.jp の系統組成解析パイプラインと比較して 10 倍以上の高速化を実現した。また、系統名についても、MAG で標準的に用いられている GTDB の系統分類階層を用いることで、MAG の系統名と系統組成とで系統情報の齟齬が無いようにした。その上で、数万サンプル以上のメタゲノムデータについて系統組成を遺伝研スパコンで解析を行い、結果を Microbiome Datahub 上で閲覧可能にした。リードデータからの遺伝子機能組成推定についても、深層学習を用いて高速に Gene Ontology ID を付与する DeepGO や KEGG Orthology (KO) を付与する KOfamScan をテストし、DeepGO は速度は十分であるものの精度は低く、KOfamScan は完全長タンパク質配列については高精度であるものの速度面で問題がある等、数万以上にメタゲノムサンプルに対する遺伝子機能組成推定の解析ツールについては、未だ現実的な計算時間で実行可能で精度も良いツールは存在しない状況である。個別の MAG 単位では DFAST と KOfamScan をつなげた遺伝子機能組成解析パイプラインを開発し計算を実行済みであるため、MAG ベースでの遺伝子機能組成を Microbiome Datahub のコンテンツの一つにする予定である。

また、Microbiome Datahub の 21 万 MAG 由来の約 4.5 億タンパク質配列について、冗長性削減のためにスパコンを用いて配列相同性検索によるクラスタリングの計算を行い中であり、2024 年中には完了予定である。クラスタリングによって得られた代表配列について、

配列相同性を用いた MBGD オーソロググループへの紐付けを行う。MBGD オーソロググループ自体も、遺伝子系統樹と KO 分類とを重ねて適切なサブグループを切り出すプログラムを作成し、これを用いて MBGD と KO のクロスリファレンスを作成することで KO とも対応付けを行えるようにした。これを MyMBGD の Maple 解析パイプラインに組み込み、新規のゲノムや MAG に対する MBGD オーソログの検索から KEGG モジュールの有無の判定を行う精度を改善した。また、MAG に対するオーソログ対応付けの可能性を広げる目的で、現在の MBGD オーソロググループの作成手順に従って、種内、属内パンゲノムを構築してから全体のオーソロググループを作成するという、階層的な手順へと切り替えて MBGD オーソロググループの作成を高速化する開発を行った。並行して MBGD のデータ更新を進め、オーソログデータの構築を完了した。登録ゲノム数は 34079 件と倍増し、シングルトン含めたクラスタ数は 143 万クラスタと以前と比べて約 1.3 倍になった。

4. 「効率的な逐次 DB 更新・運用システムの"再"構築」

非常にニーズが多いヒトマイクロバイオームデータについて、優先度の高い論文からメタデータのマニュアルキュレーションを行い、ヒトマイクロバイオーム関連の約 4500 論文から約 41 000 サンプルのメタデータを抽出し、マイクロバイオーム配列データの個別の BioSample ID と紐付けた。それらのマニュアルキュレーション済みのヒトマイクロバイオーム BioSample メタデータを正解データとして、Python 言語と LLM の ChatGPT の API を用いた論文からのメタデータ抽出ツール EMBERS を開発した。EMBERS は PubMed Central から取得した論文データからのメタデータ抽出と (EMBERS-MINE)、メタデータの名寄せと MEO や HMDADO へのオントロジー自動アノテーション (EMBERS-FUSE) の二つのツールに分かれており、EMBERS を用いることで DB 更新の際に非常に手間がかかっていたメタデータ関連の作業を効率化できた。今後はマニュアルキュレーションと EMBERS を併用して高精度かつ高速にメタデータの抽出を行う。また、Microbiome Datahub のデータ更新を容易に行うために、BioProject および MAG データの取得、データ変換とアノテーション付け、DB へのデータ投入までの一連のワークフローの洗練化を行い、MicrobeDB.jp で用いていたデータ更新システムから一新した。DB における検索結果の表示においても、Python の Plotly ライブラリ等を用いた開発側にとって機能追加・修正が容易でかつユーザ側にとってよりわかりやすく動的にグラフでの結果表示を行えるようにした。

5. 「キラーデータセット・アプリケーションの開発」

Microbiome Datahub のユーザ数向上を目的として、Microbiome Datahub のデータをユーザが自由に取得できるように、データ取得 API を開発し公開した。現状メタゲノム BioProject のメタデータ (例: <https://mdatahub.org/api/dl/project/metadata/PRJNA13694>) や系統組成データ (例: <https://mdatahub.org/api/dl/project/composition/PRJNA464990>)、MAG のメタデータ (例: https://mdatahub.org/api/dl/genome/metadata/GCA_029762495.1)、MAG の配列データ (例: http://mdatahub.org/api/dl/sequence/genome/GCA_029762495.1) が取得可能である。大量データのダウンロードには別途ダウンロードサイトを

用意する予定であるが、個々のデータのダウンロードは API および Microbiome Datahub のサイトからマウスクリックのみでデータダウンロード API を呼び出せる機能を開発し、プログラム経由で API アクセスできない利用者にも API を利用できるようにしている。配列の検索については、マイクロバイオームの配列データに対する検索ツールである PZLAST について、約 21 万 MAG 由来の遺伝子のアミノ酸配列データに対する配列類似性検索を行える機能 (PZLAST-MAG)を開発し、現在論文執筆中である。2024 年度中に PZLAST-MAG を公開し論文投稿予定である。また、MicrobeDB.jp/Microbiome Datahub のメタゲノムアミノ酸配列データも活用して、ユーザが入力したメタゲノムアミノ酸配列データから微生物群集が生息する環境温度を推定するツール Metagenomic Thermometer (メタゲノム温度計)を開発し論文発表と web アプリケーションの公開を行なった (<http://palaeo.nig.ac.jp/metathermo/>) [論文 2]。4 で言及した Python 言語と LLM の ChatGPT の API を用いた論文からのメタデータ抽出ツール EMBERS についても、論文から抽出し整理したメタデータと BioSample との対応関係を整理し、BioSample をクラスタリングすることでマイクロバイオームのメタ解析に使用できるサンプルメタデータ俯瞰ツールを開発した。EMBERS の論文を現在執筆中であり、論文投稿し次第 web ツールとして公開予定である。

6. 「マルチオミックス連携」

MAG 配列データの遺伝子機能アノテーション等について、DDBJ で開発している DFAS T と連携して遺伝研スーパーコンピュータ上で高速に MAG のアノテーションをできるようにし、ゲノム・MAG のクオリティチェック行える DFAST_QC [論文 1]を公開した。また、DBCLS とは NITE NBRC と共に昨年度から NEDO グリーンイノベーション基金事業のプロジェクトを遺伝研グループと共同で進めており、DBCLS が開発しているゲノム DB TogoGenome や培地 DB TogoMedium と Microbiome Datahub の微生物ゲノムや MAG データ間で API を主とした連携を進めると共に、LLM を用いたメタデータの自動アノテーションについても DBCLS と毎月の Togothon で定期的に情報交換を行い、技術的な連携を行なっている。同じ統合化推進プログラムのプロテオームグループに Microbiome Datahub の MAG 由来のタンパク質配列データを渡し、メタプロテオームのリファレンスデータとして MAG データを使用するための連携を行なった。

7. その他

Microbiome Datahub の国際的な認知度を上げるために、Microbiome Datahub の論文を東工大グループと遺伝研グループが主となって執筆中であり、他の代表的な MAG DB である MGnify や GlobDB とのデータセットの重なり等の比較解析や新規の系統群についての解析結果を含めた論文として 2024 年度中に少なくとも bioRxiv で発表予定である。

(2) 若手研究者の多様なキャリアパスの支援

2023 年度前半に遺伝研グループで特任研究員として雇用していた村上匠博士について、Microbiome Datahub において大量のメタゲノムデータを扱った経験も評価され、2023 年 8 月から東京工業大学生命理工学院の助教としてキャリアアップした。2024 年中旬から遺伝研グループで一部本予算で人件費を支出し、MAG データの解析の仕事も行っていただいている特任研究員(飯塚朋代博士)についても、自身の科研費課題も同時に進めていただいております、解析技術面を含め独立した研究者としてキャリアアップするためのサポートを行っている。

§4. 主要なデータベースの利活用状況

1. アクセス数

(1) 実績

表 研究開発対象の主要なデータベースの利用状況(月間平均)

種別	2022 年度 (MicrobeDB.jp)	2023 年度 (MicrobeDB.jp)	2024 年度 (8 月末時点)
訪問数 (Visit)	3,392	1,408	1,579
訪問者数 (Unique)	1,983	945	935
閲覧ページ数 (Views)	239,538	208,614	7,747

(2) 分析

2022 年度と 2023 年度は、MicrobeDB.jp の最終更新を行ってから時間が経っており年月を経るほどデータが相対的に古くなったため、訪問数や訪問者数は減少傾向にあった。MicrobeDB.jp のコンテンツが古くなったことと、Microbiome Datahub への移行のために、2024 年 4 月に MicrobeDB.jp を閉鎖し、それ以降は Microbiome Datahub のテスト版のみ公開している状態であった。Microbiome Datahub では Web サイトのデザインとして検索結果を閲覧する際に MicrobeDB.jp とは異なりページ遷移をほぼしない形にしており、2024 年度は閲覧ページ数が激減した。

2. データベースを利用して得られた研究成果・産業応用の例

1. 遺伝研の黒川真臣研究員らと森らは、MicrobeDB.jp/Microbiome Datahub のメタゲノムアミノ酸配列データとメタデータをツールの評価に活用して、ユーザが入力したメタゲノムアミノ酸配列データから微生物群集が生息する環境温度を推定するツール Metagenomic Thermometer (メタゲノム温度計)を開発し論文発表と web アプリケーションの公開を行なった。

Masaomi Kurokawa, Koichi Higashi, Keisuke Yoshida, Tomohiko Sato, Shigenori Maruyama, Hiroshi Mori, Ken Kurokawa. “ Metagenomic Thermometer”, DNA Res., 30:dsad024, 2023 (DOI:10.1093/dnares/dsad024). プレ ス リ リ ー ス (https://www.nig.ac.jp/nig/ja/2023/11/research-highlights_ja/pr20231122.html) Metagenomic Thermometer URL: <http://palaeo.nig.ac.jp/metathermo/>

2. 慶應義塾大学の金井昭夫教授らの研究グループは、遺伝研の森らや日本原子力研究開発機構等多数の研究機関からなる共同研究によって、廃炉過程にある福島第一原子

力発電所原子炉建屋内に生息する微生物群集の系統組成を 16S rRNA 遺伝子のアンプリコン解析によって明らかにした。解析を行う上で、見つかった細菌群集が既知のどのような環境の群集と類似しているかを MicrobeDB.jp および LEA を用いて解析し、海水由来の細菌が優占していること等を明らかにし論文発表を行なった。

Tomoro Warashina, Asako Sato, Hiroshi Hinai, Nurislam Shaikhutdinov, Elena Shagimardanova, Hiroshi Mori, Satoshi Tamaki, Motofumi Saito, Yukihisa Sanada, Yoshito Sasaki, Kozue Shimada, Yuma Dotsuta, Toru Kitagaki, Shigenori Maruyama, Oleg Gusev, Issay Narumi, Ken Kurokawa, Teppei Morita, Toshikazu Ebisuzaki, Akihiko Nishimura, Yoshikazu Koma, Akio Kanai. “Microbiome analysis of the restricted bacteria in radioactive element-containing water at the Fukushima Daiichi Nuclear Power Station”, Appl Environ Microbiol., 90(4):e0211323 2024 (DOI:10.1128/aem.02113-23). プレスリリース(<https://www.iab.keio.ac.jp/news-events/2024/03191413.html>)

3. NEDO グリーンイノベーション基金事業「バイオものづくり技術による CO₂ を直接原料としたカーボンリサイクルの推進」プロジェクトにおいて (https://www.nedo.go.jp/news/press/AA5_101619.html)、「CO₂ 固定微生物利活用プラットフォームの構築」として Microbiome Datahub の MAG DB やメタゲノム解析パイプラインを活用した研究開発を 2023 年度から DBCLS や NITE 等と共同で開始 (2023 年 6 月 1 日 プレスリリース <https://www.nig.ac.jp/nig/images/information/PR20230601.pdf>) している。この中で、日本国内のいくつかの企業と CO₂ 固定微生物利活用プラットフォームについて連携するために NITE が中心になってグリーンイノベーションフォーラム (GI フォーラム) を 2023 年度末に立ち上げた。プレスリリース (<https://www.nite.go.jp/data/000153394.pdf>)。今後、Microbiome Datahub の中の CO₂ 固定関連の MAG について NEDO プロジェクトで整理し再解析を行なった上で、GI フォーラムを介して国内のさまざまな企業に情報やデータを提供しデータベースおよびデータの産業利用を推進する予定である。
4. 遺伝研の MicrobeDB.jp/Microbiome Datahub とちとせグループ マイクロバイオームデータ利活用促進に向けたデータベース機能拡張について提携を開始 https://www.nig.ac.jp/nig/ja/2023/02/information_ja/pr20230220.html (2023 年 2 月プレスリリース)。現在、Microbiome Datahub の解析パイプラインやメタデータセットをちとせグループと共有し、特にヒトマイクロバイオームデータにおける産業応用を進めている。

3. その他

§5. 今後の研究実施計画

第一に Microbiome Datahub の機能拡張と Help 機能の充実等を行い、正式公開を 2024 年中に行う。また、Microbiome Datahub の論文完成を急ぎ、bioRxiv で論文公開(2024 年中)と国際誌へ論文投稿し出版を急ぐ。LLM 技術を用いた論文やテキストデータからのメタデータ抽出と名寄せツールである EMBERS の公開と論文文化も並行して行う。Microbiome Datahub の正式公開、論文出版、EMBERS の公開と論文出版等を経ることで Microbiome Datahub の国際的な認知度を上げ、多くのユーザ獲得と利活用事例の蓄積に努める。また、Microbiome Datahub の検索インターフェースについて、LLM を活用した文字列ベースでのよりわかりやすい検索インターフェースを開発し、従来型の絞り込み検索とユーザが使い分けられるようにする。個別のゲノム・MAG について微生物の Phenotype 推定ツール Bac2Feature の解析結果および MBGD オースログクラスタ情報を Microbiome Datahub 上に反映し、一つの DB 上で統一的に結果を表示できるようにする。

§6. 自己評価

今まで 10 年近く開発していた微生物エンサイクロペディア的な統合 DB である Microbe DB.jp から、マイクロバイオーーム、特に MAG を中心したマイクロバイオーーム特化の統合 DB へと開発方針を転換し、この 2 年半で DB の基幹システムやデータ構造、メタデータ抽出・整理方法等ほぼ全て新規に開発した。まだ Microbiome Datahub の正式公開や論文出版には至っていない点は大いに反省すべきであるが、申請時の段階では想定外なほど激変した生成 AI 時代に必要とされる DB を目指して、LLM 技術等をいち早く取り入れて DB の研究開発を行っており、ライフサイエンスに関わる国内外のデータを統合的に扱うための DB 開発という本プログラムの趣旨にそった研究開発を行なっていると自負している。本研究開発で大量のメタデータアノテーションを行う必要があつて開発した EMBERS 等の LLM 応用技術は他のライフサイエンス系 DB 開発でも広く応用可能であり、今後重要になる技術を開発できたと考えている。評価委員・アドバイザーの先生方からの意見を元に、今後 Microbiome Datahub を、マイクロバイオーームに興味があるユーザーに必要なデータ・機能を提供できる統合 DB として洗練させていきたい。

§7. 外部発表等

1. 原著論文発表

(1) 論文数概要

種別	国内外	件数
発行済論文	国内(和文)	0 件
	国際(欧文)	7 件
未発行論文 (accepted, in press 等)	国内(和文)	0 件
	国際(欧文)	0 件

(2) 論文詳細情報

1. Mohamed Elmanzalawi, Takatomo Fujisawa, Hiroshi Mori, Yasukazu Nakamura, Yasuhiro Tanizawa. “DFAST_QC: Quality Assessment and Taxonomic Identification Tool for Prokaryotic Genomes”, *bioRxiv*, 2024 (DOI: 10.1101/2024.07.22.604526).
概要: 原核生物ゲノムおよび MAG 配列について、Completeness (完成度) や Contamination 率、由来した系統等を高速に推定しゲノム配列のクオリティチェックを行うツール DFAST_QC を公開しその概要を紹介した論文である。
2. Masaomi Kurokawa, Koichi Higashi, Keisuke Yoshida, Tomohiko Sato, Shigenori Maruyama, Hiroshi Mori, Ken Kurokawa. “Metagenomic Thermometer”, *DNA Res.*, 30:dsad024, 2023 (DOI:10.1093/dnares/dsad024).
3. Susumu Morigasaki, Motomu Matsui, Iwao Ohtsu, Yuki Doi, Yusuke Kawano, Ryosuke Nakai, Wataru Iwasaki, Hisayoshi Hayashi, Naoki Takaya, “Temporal and fertilizer-dependent dynamics of soil bacterial communities in buckwheat fields under long-term management”, *Scientific Reports*, 14(1):9896, 2024 (DOI:10.1038/s41598-024-60655-w).
4. Hirotaka Matsumoto, Motomu Matsui, “NeTaGFT: A similarity network-based method for trait analysis”, *Methods in Ecology and Evolution*, 15:153–163, 2023 (DOI:10.1111/2041-210X.14251).
5. Hirotugu Shiroma, Youssef Darzi, Etsuko Terajima, Zenichi Nakagawa, Hirotaka Tsuchikura, Naoki Tsukuda, Yuki Moriya, Shujiro Okuda, Susumu Goto, Takuji Yamada, “Enteropathway: the metabolic pathway database for the human gut microbiota”, *bioRxiv*, 2023 (DOI:10.1101/2023.06.28.546710).
6. Tomoro Warashina, Asako Sato, Hiroshi Hinai, Nurislam Shaikhutdinov, Elena Shagimardanova, Hiroshi Mori, Satoshi Tamaki, Motofumi Saito, Yukihiisa Sanada, Yoshito Sasaki, Kozue Shimada, Yuma Dotsuta, Toru Kitagaki, Shigenori Maruyama, Oleg Gusev, Issay Narumi, Ken Kurokawa, Teppei Morita, Toshikazu Ebisuzaki, Akihiko Nishimura, Yoshikazu Koma, Akio Kanai. “Microbiome analysis of the restricted bacteria in radioactive element-containing

water at the Fukushima Daiichi Nuclear Power Station”, *Appl Environ Microbiol.*, 90(4):e0211323 2024 (DOI:10.1128/aem.02113-23).

7. Takao K Suzuki, Motomu Matsui, Sira Sriswasdi, and Wataru Iwasaki, “Lifestyle Evolution Analysis by Binary-State Speciation and Extinction (BiSSE) Model”, *Methods in Molecular Biology*, 2569:327–342, 2022, (DOI: https://doi.org/10.1007/978-1-0716-2691-7_16).

2. その他の著作物(総説、書籍など)

1. 森宙史、東光一、黒川顕、微生物統合情報とリソース、微生物資源の整備と利活用の戦略、エヌ・ティー・エス出版、pp325-335、2023 年.
2. 森宙史、Microbiome Datahub と PZLAST、実験医学増刊 バイオデータベース・ウェブツール、羊土社、pp101-103、2022 年.

3. 国際学会発表及び主要な国内学会発表

(1) 概要

種別	国内外	件数
招待講演	国内	11 件
	国際	0 件
口頭発表	国内	12 件
	国際	3 件
ポスター発表	国内	11 件
	国際	2 件

(2) 招待講演

〈国内〉

1. 松井求、系統樹推定の常識を問い直す、国立遺伝学研究所研究集会、国立遺伝学研究所、2022 年 8 月 3 日.
2. 松井求、ポストコッホ機能生態学が目指すもの、ポストコッホ生態研究集会、長崎大学、2022 年 8 月 28 日.
3. 森宙史、マイクロバイオーームからの MAG/SAG データの取得と活用の展望、NITE スポンサーセミナー、BioJapan 2022、オンライン、2022 年 10 月 12-14 日.
4. 松井求、微生物ダークマターは今何色か?、第 35 回日本 Archaea 研究会、東京、2023 年 6 月 30 日.

5. 松井求、環境微生物研究におけるバイオインフォマティクス活用法、アグリバイオ、東京、2023 年 7 月 24 日.
6. 松井求、Tree of Life は完成できるか?、IIBMP2023 シンポジウム、千葉県柏市、2023 年 9 月 7 日.
7. 森宙史、マイクロバイオーーム研究を先導するハブとなる統合データベース **Microbiome Datahub** の開発、第 75 回日本生物工学会大会 シンポジウム「データ駆動型研究が導く未培養微生物 (微生物ダークマター) の開拓と展望」、名古屋大学東山キャンパス、名古屋、2023 年 9 月 5 日.
8. 森宙史、マイクロバイオーーム研究を先導するハブを目指した **Microbiome Datahub** の開発、信州大学理学部生物学コース教室セミナー、信州大学理学部 8 番講義室、2023 年 9 月 14 日.
9. 松井求、バイオインフォマティクスが駆動する微生物学、MPIP セミナー、広島県東広島市、2023 年 10 月 31 日.
10. 松井求、分子系統学の進化可能性を考える、Intelligence and Cultural Evolution Theory ワークショップ、東京、2023 年 12 月 22 日.
11. 松井求、バイオインフォマティクスが駆動する微生物科学、微生物科学イノベーション特論 II、東京、2023 年 12 月 26 日.

〈国際〉

(3) 口頭講演

〈国内〉

1. 森宙史、マイクロバイオーーム研究を先導するハブを目指した **Microbiome Datahub** の開発、第 1 回統合微生物学研究会、ラフォーレ修善寺、2022 年 9 月 21 日.
2. 森宙史、マイクロバイオーーム研究を先導するハブを目指した **Microbiome Datahub** の開発、トーゴの日シンポジウム 2022、オンライン、2022 年 10 月 5 日.
3. 鈴木誉保、松井求、森ヶ崎進、大津厳生、土肥裕希、河野祐介、林久喜、高谷直樹、岩崎渉、**Trait-based** アプローチでせまる 細菌の群集構造とその安定性、第 35 回日本微生物生態学会、2022 年 10 月 31 日.
4. 森宙史、マイクロバイオーームの国際的なデータハブの構築、第 45 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」、2022 年 11 月 30 日.
5. 内山郁夫、高柳正彦、CGB: 大規模ゲノム比較のための比較ゲノムブラウザ、第 17 回日本ゲノム微生物学会年会、かずさアカデミアホール、2023 年 3 月 8-10 日.
6. 森宙史、マイクロバイオーーム研究を先導するハブを目指した **Microbiome Datahub** の開発、第 96 回日本細菌学会総会 シンポジウム「微生物バイオインフォマティクスの最前線」、アクリエ姫路、2023 年 3 月 17 日.
7. 山田拓司、ゲノム・メタゲノムデータの可視化手法の開発、第 96 回日本細菌学会総会 シンポジウム「微生物バイオインフォマティクスの最前線」、アクリエ姫路、2023 年 3 月 17 日.

8. 森宙史、微生物のゲノム・MAG データを統合したデータベース Microbiome Datahub の開発、雪氷生物研究会 2023、JAXA、つくば、2023 年 4 月 4 日。
9. 森宙史、MAG を基盤とした Microbiome Datahub の開発、第 15 回日本ゲノム微生物学会若手の会研究会、湯河原温泉 おんやど恵、湯河原、2023 年 9 月 19 日。
10. 森宙史、メタゲノムのアノテーション、Annotathon 2023、国立遺伝学研究所 W403-405 会議室、三島、2023 年 11 月 16 日。
11. 森宙史、細菌の多様性を解明するための大規模 MAG 解析、第 97 回日本細菌学会総会、札幌コンベンションセンター、札幌、2024 年 8 月 7 日。
12. 森宙史、大規模メタゲノムデータから俯瞰するバクテリアの複製・転写・翻訳装置、第 96 回日本遺伝学会大会、高知工科大学、高知、2024 年 9 月 5 日。

〈国際〉

1. Ikuro Uchiyama, Recent developments in MBGD and its application to genomic functional inference. Novel challenges in the quest for orthologs (7th Quest for Orthologs meeting), Sitges, Spain, Sep. 17-18, 2022.
2. Motomu Matsui, Reconstructing Phylogenetic Reconstruction, DATA SCIENCE FOR GENOME DYNAMICS, Online, 2024/01/09
3. Hiroshi Mori, Microbiome Datahub: an integrated database to accelerate microbiome research, HUPO-PSI Kyoto Symposium 2024, Fujita Commemorative Lecture Hall, Kyoto University, Kyoto 2024/03/21

(4) ポスター発表

〈国内〉

1. 藤吉真生、鈴木誉保、松井求、岩崎渉、Bac2Feature：系統情報を利用した微生物の形質推定パイプライン、第 24 回日本進化学会、静岡県沼津市、2022 年 8 月 5 日。
2. 松井求、岩崎渉、シン・近隣結合法：PANJEP 法が問い直す系統解析の常識、第 24 回日本進化学会、静岡県沼津市、2022 年 8 月 5 日。
3. 鈴木誉保、松井求、森ヶ崎進、大津厳生、土肥裕希、林久喜、高谷直樹、岩崎渉、微生物群集は普遍的に極めて安定な形質組成を持つ、第 60 回日本生物物理学会、2022 年 9 月 28 日。
4. 内山郁夫、三原基広、西出浩世、千葉啓和、高柳正彦、高見英人、MBGD 2022: 大規模な微生物オーソログデータベースの効果的な活用に向けて、トーゴの日シンポジウム 2022、2022 年 10 月 5 日。
5. 松井求、明石基洋、鈴木誉保、藤吉真生、土肥裕希、森ヶ崎進、林久喜、高谷直樹、岩崎渉、ポストコッホ生態学が解き明かす輪作に呼応した微生物動態、第 35 回日本微生物生態学会、2022 年 10 月 31 日。
6. 藤吉真生、鈴木誉保、松井求、岩崎渉、Bac2Feature による微生物群集の表現型解析、第 35 回日本微生物生態学会、2022 年 10 月 31 日。
7. 森宙史、藤澤貴智、東光一、谷澤靖洋、中川善一、西出浩世、中村保一、山田拓司、

松井求、内山郁夫、マイクロバイオーーム研究を先導するハブを目指した **Microbiome Datahub** の開発、第 17 回日本ゲノム微生物学会年会、かずさアカデミアホール、2023 年 3 月 8-10 日。

8. 内山郁夫、河合幹彦、西出浩世、千葉啓和、三原基広、微生物比較ゲノムデータベース MBGD、トーゴの日シンポジウム、2023 年 10 月 5 日。
9. 内山 郁夫、三原 基広、西出 浩世、千葉 啓和、高柳正彦、河合 幹彦、MAG のアノテーションに向けた微生物比較ゲノムデータベース MBGD の改良、第 18 回日本ゲノム微生物学会年会、かずさアカデミアホール、2024 年 3 月 12 日。
10. Mikihiro Kawai, Motohiro Mihara, Hiroyo Nishide, Hirokazu Chiba, Ikuo Uchiyama, Phylogenetic profiling re-visited: exploration of alternate orthologs in functional modules, 第 18 回日本ゲノム微生物学会年会、かずさアカデミアホール、2024 年 3 月 12 日。
11. Jiang Zhenbo, Uchiyama Ikuo, Extraction of the shared co-expression patterns based on extended pan-network and core-network concepts, 第 18 回日本ゲノム微生物学会年会、かずさアカデミアホール、2024 年 3 月 12 日。

〈国際〉

1. Hiroshi Mori, Takatomo Fujisawa, Koichi Higashi, Hiroyo Nishide, Zenichi Nakagawa, Yasuhiro Tanizawa, Yasukazu Nakamura, Motomu Matsui, Ikuo Uchiyama, Takuji Yamada, Microbiome Datahub: an integrated database of microbiome samples and MAG data, IHMC Kobe 2022, Kobe Portopia Hotel, Kobe, November 8-19, 2022.
2. Hiroshi Mori, Takatomo Fujisawa, Koichi Higashi, Hiroyo Nishide, Zenichi Nakagawa, Yasuhiro Tanizawa, Yasukazu Nakamura, Motomu Matsui, Ikuo Uchiyama, Takuji Yamada, Microbiome Datahub: an integrated database of microbiome projects and MAG data, The 14th International Workshop on Advanced Genomics, Hitotsubashi Hall, Tokyo, October 4-6, 2023.

4. 知財出願

(1) 出願件数

種別		件数
特許出願	国内	0 件
	国外	0 件
その他の知的財産出願		0 件

(2) 一覧

① 国内出願

該当なし。

② 海外出願

該当なし。

③ その他の知的財産権

該当なし。

5. 受賞・報道等

(1) 受賞

該当なし。

(2) メディア報道

1. (JST-NBDC からのプレスリリース(2023 年 10 月 4 日))” 単行本「微生物資源の整備と利活用の戦略」の「微生物統合情報とリソース」の章で、MicrobeDB.jp と Microbiome Datahub の機能を紹介”, <https://biosciencedbc.jp/news/20231004-02.html/>)
2. 遺伝研の MicrobeDB.jp/Microbiome Datahub とちとせグループ マイクロバイオームデータ活用促進に向けたデータベース機能拡張について提携を開始
https://www.nig.ac.jp/nig/ja/2023/02/information_ja/pr20230220.html
(2023 年 2 月プレスリリース)
3. NEDO グリーンイノベーション基金事業「バイオものづくり技術による CO2 を直接原料としたカーボンリサイクルの推進」プロジェクトにおいて (https://www.nedo.go.jp/news/press/AA5_101619.html)、「CO2 固定微生物利活用プラットフォームの構築」として Microbiome Datahub の MAG DB やメタゲノム解析パイプラインを活用した研究開発を 2023 年度から DBCLS や NITE 等と共同で開始 (2023 年 6 月 1 日プレスリリース <https://www.nig.ac.jp/nig/images/information/PR20230601.pdf>)

(3) その他

該当なし。

§8. 研究開発期間中の活動

1. 進捗ミーティング

年月日	名称	場所	参加人数	目的・概要
2022年 4月20日	チーム内キックオフ 会議（非公開）	オンライン	9人	プロジェクト内のキック オフ会議
2022年 4月25日	チーム内開発会議（非 公開）	オンライン	5人	プロジェクト内の開発 会議
2022年 5月9日	チーム内開発会議（非 公開）	オンライン	4人	プロジェクト内の開発 会議
2022年 5月23日	チーム内開発会議（非 公開）	オンライン	4人	プロジェクト内の開発 会議
2022年 6月11日	チーム内開発会議（非 公開）	オンライン	4人	プロジェクト内の開発 会議
2022年 7月4日	チーム内開発会議（非 公開）	オンライン	4人	プロジェクト内の開発 会議
2022年 7月19日	チーム内開発会議（非 公開）	オンライン	4人	プロジェクト内の開発 会議
2022年 7月28日	企業との連携に関す る会議（非公開）	オンライン	5人	企業との連携に関する 会議
2022年 8月1日	チーム内開発会議（非 公開）	オンライン	4人	プロジェクト内の開発 会議
2022年 8月18日	統合化推進プログラ ムの他のグループと の連携についての会 議（非公開）	オンライン	3人	統合化推進プログラ ムの他のグループとの連 携についての会議
2022年 8月22日	チーム内開発会議（非 公開）	オンライン	4人	プロジェクト内の開発 会議
2022年 9月11日	チーム内開発会議（非 公開）	オンライン	3人	プロジェクト内の開発 会議
2022年 9月26日	チーム内開発会議（非 公開）	オンライン	4人	プロジェクト内の開発 会議
2022年 10月11日	チーム内開発会議（非 公開）	オンライン	6人	プロジェクト内の開発 会議
2022年 10月17日	チーム内開発会議（非 公開）	オンライン	5人	プロジェクト内の開発 会議
2022年 10月18日	チーム内開発会議（非 公開）	オンライン	4人	プロジェクト内の開発 会議
2022年 10月27日	チーム内開発会議（非 公開）	オンライン	6人	プロジェクト内の開発 会議
2022年 11月2日	チーム内開発会議（非 公開）	オンライン	6人	プロジェクト内の開発 会議
2022年 11月7日	チーム内開発会議（非 公開）	オンライン	4人	プロジェクト内の開発 会議
2022年	チーム内開発会議（非 公開）	オンライン	4人	プロジェクト内の開発

年月日	名称	場所	参加人数	目的・概要
12月5日	公開)			会議
2022年 12月7日	チーム内開発会議 (非公開)	オンライン	3人	プロジェクト内の開発会議
2022年 12月21日	チーム内開発会議 (非公開)	オンライン	4人	プロジェクト内の開発会議
2023年 1月24日	チーム内開発会議 (非公開)	オンライン	3人	プロジェクト内の開発会議
2023年 1月25日	チーム内開発会議 (非公開)	オンライン	7人	プロジェクト内の開発会議
2023年 2月9日	チーム内開発会議 (非公開)	オンライン	4人	プロジェクト内の開発会議
2023年 2月17日	チーム内開発会議 (非公開)	オンライン	6人	プロジェクト内の開発会議
2023年 3月2日	チーム内開発会議 (非公開)	オンライン	6人	プロジェクト内の開発会議
2023年 3月22日	チーム内開発会議 (非公開)	オンライン	6人	プロジェクト内の開発会議
2023年 4月5日	チーム内開発会議 (非公開)	オンライン	8人	プロジェクト内の開発会議
2023年 4月14日	チーム内開発会議 (非公開)	オンライン	3人	プロジェクト内の開発会議
2023年 5月9日	チーム内開発会議 (非公開)	オンライン	4人	プロジェクト内の開発会議
2023年 7月7日	チーム内開発会議 (非公開)	オンライン	7人	プロジェクト内の開発会議
2023年 8月23日	チーム内開発会議 (非公開)	オンライン	7人	プロジェクト内の開発会議
2023年 9月20日	チーム内開発会議 (非公開)	オンライン	7人	プロジェクト内の開発会議
2023年 10月6日	プロテオーム連携会議 (非公開)	京都大学東京オフィス	6人	Microbiome Datahub とプロテオームデータベースの連携のための会議
2023年 10月10日	チーム内開発会議 (非公開)	オンライン	7人	プロジェクト内の開発会議
2023年 10月24日	チーム内開発会議 (非公開)	東京工業大学緑ヶ丘キャンパス	4人	プロジェクト内の開発会議
2023年 10月26日	チーム内開発会議 (非公開)	オンライン	7人	プロジェクト内の開発会議
2023年 11月17日	EMBL-EBI の TREC グループとの連携会議 (非公開)	東京工業大学緑ヶ丘キャンパス	5人	EMBL-EBI の TREC プロジェクト代表者 (Peer Bork 先生) との連携会議
2023年 11月29日	チーム内開発会議 (非公開)	東京工業大学緑ヶ丘キャンパス	4人	プロジェクト内の開発会議
2023年 12月6日	チーム内開発会議 (非公開)	オンライン	7人	プロジェクト内の開発会議

年月日	名称	場所	参加人数	目的・概要
2023 年 12 月 27 日	チーム内開発会議（非公開）	オンライン	4 人	プロジェクト内の開発会議
2024 年 1 月 10 日	チーム内開発会議（非公開）	東京工業大学緑ヶ丘キャンパス	4 人	プロジェクト内の開発会議
2024 年 3 月 6 日	チーム内開発会議（非公開）	オンライン	7 人	プロジェクト内の開発会議
2024 年 3 月 19 日	チーム内開発会議（非公開）	オンライン	7 人	プロジェクト内の開発会議
2024 年 4 月 26 日	チーム内開発会議（非公開）	オンライン	7 人	プロジェクト内の開発会議
2024 年 5 月 24 日	チーム内開発会議（非公開）	オンライン	7 人	プロジェクト内の開発会議
2024 年 6 月 12 日	チーム内開発会議（非公開）	オンライン	4 人	プロジェクト内の開発会議
2024 年 6 月 19 日	チーム内開発会議（非公開）	オンライン	4 人	プロジェクト内の開発会議
2024 年 7 月 24 日	チーム内開発会議（非公開）	オンライン	4 人	プロジェクト内の開発会議
2024 年 8 月 28 日	チーム内開発会議（非公開）	オンライン	4 人	プロジェクト内の開発会議

2. 主催したワークショップ、シンポジウム、アウトリーチ活動等

年月日	名称	場所	参加人数	目的・概要
2023 年 2 月 5-7 日	チーム内 Hackathon（非公開）	ホテルアソシア 静岡	7 人	プロジェクト内の DB 開発合宿
2023 年 3 月 12-14 日	チーム内 Hackathon（非公開）	ホテルアソシア 静岡	6 人	プロジェクト内の DB 開発合宿
2023 年 4 月 18-19 日	チーム内 Hackathon（非公開）	三井ガーデンホテル 柏の葉	5 人	プロジェクト内の DB 開発合宿
2023 年 6 月 12-14 日	チーム内 Hackathon（非公開）	三井ガーデンホテル 柏の葉	5 人	プロジェクト内の DB 開発合宿
2024 年 1 月 22-24 日	チーム内 Hackathon（非公開）	ホテルアソシア 静岡	7 人	プロジェクト内の DB 開発合宿

以上

別紙 研究開発対象のデータベース等

No.	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系 データベース アーカイブ	NBDCヒト データ ベース	NBDC RDFポ ータル	関連文献 (論文リストに記載があれば、その番号で も可)
1	Microbiome Datahub		メタゲノム由来のゲノムであるMAGを中心に微生物のゲノムとメタゲノムデータを統合したマイクロバイオームの統合データベース。	https://mdatabub.org/	2023年5月31日 (webサイトはテスト版として公開。正式アナウンスはまだ)	新規	データベ ース等	提供前	対象外	提供前	
2	MicrobeDB.jp		ゲノムを軸にして微生物学の様々なデータをセマンティックウェブ技術を用いて統合した微生物統合データベース。	https://microbedb.jp	2011年12月	休止・閉鎖	データベ ース等	対象外	対象外	提供前	
3	Microbial Genome Database for Comparative Analysis	MBGD	オーソログ解析に基づいて微生物ゲノムの比較解析を行うためのデータベース。公開されたゲノム全体を含む標準オーソログテーブルに基づいて、各オーソロググループの系統プロファイルの比較などを行えるほか、動的なオーソログ解析機能によって、利用者自身が持つゲノム配列も含めて、興味のある生物種セットを対象を絞った比較を行うことも可能。	https://mbgd.niib.ac.jp	1997年7月	継続・発展	データベ ース等	提供前	対象外	公開済	Ikuko Uchiyama, Motohiro Mihara, Hiroyo Nishide, Hirokazu Chiba, Masaki Kato, "MBGD update 2018: microbial genome database based on hierarchical orthology relations covering closely related and distantly related comparisons", Nucleic Acids Res., 47, D382–D389, 2019. (DOI: 10.1093/nar/gky1054).
4	Metagenome and Microbes Environmental Ontology	MEO	微生物の生息環境に関するメタデータを記述し整理するためのオントロジー。	https://bioportal.bioontology.org/ontologies/MEO	2012年2月	継続・発展	ツール等	対象外	対象外	対象外	
5	Latent Environment Allocation	LEA	Microbiome Datahubに存在する大量のマイクロバイオームサンプルの系統組成と環境を対応させて視覚化するWebアプリケーション。ユーザが自分の群集組成データをアップロードしてLEA上にマッピングすることで、系統組成から環境を大まかに推定することも可能。	http://leamicro	2018年6月	継続・発展	ツール等	対象外	対象外	対象外	Koichi Higashi, Shinya Suzuki, Susumu Kurosawa, Hiroshi Mori, Ken Kurokawa, "Latent environment allocation of microbial community data", PLoS Comput Biol, 14, e1006143, 2018 (DOI:10.1371/journal.pcbi.1006143).
6	PZLAST		Microbiome Datahubに存在する大量のメタゲノムサンプルのアミノ酸配列データに対して高速な配列相同性検索を行うことが可能なWebアプリケーション。	https://pzlast.n	2020年4月	継続・発展	ツール等	対象外	対象外	対象外	Hiroshi Mori, Hitoshi Ishikawa, Koichi Higashi, Yoshiaki Kato, Toshikazu Ebisuzaki, Ken Kurokawa, "PZLAST: an ultra-fast amino acid sequence similarity search server against public metagenomes", Bioinformatics., 37, 3944–3946, 2021. (DOI: 10.1093/bioinformatics/btab492).