

## 2023年度 研究開発実施報告

### 概要

研究開発課題名	マイクロバイオ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
開発対象データベースの名称(URL)	Microbiome Datahub ( <a href="https://mdatahub.org/">https://mdatahub.org/</a> )
研究代表者氏名	森 宙史 (40610837)
所属・役職	情報・システム研究機構 国立遺伝学研究所 情報研究系 准教授



# 目次

概要	1
目次	2
§1. 研究実施体制	4
§2. 研究開発対象とするデータベース・ツール等	4
(1) データベース一覧	4
【主なデータベース】	4
【その他のデータベース】	4
(2) ツール等一覧	4
§3. 実施内容	5
(1) 本年度の研究開発計画と達成目標	5
① MAG データを用いたゲノム・メタゲノムの統合化(【研究開発実施項目 1-1】)、MAG データに関する国際連携(【研究開発実施項目 3-2】)	5
② 微生物 Phenotype 情報の RDF 化(【研究開発実施項目 1-2】)、微生物 Phenotype 情報のデータ構築(【研究開発実施項目 4】)	5
③ マイクロバイオーム解析パイプラインの大幅な速度向上(【研究開発実施項目 1-3】)、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の開発(【研究開発実施項目 2】)	5
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】)、ヒトマイクロバイオームのメタデータ整備とキラータセットの開発(【研究開発実施項目 3-1】)	5
⑤ Microbiome Datahub のデータを用いたキラータアプリケーションの開発(【研究開発実施項目 1-5】)	6
⑥ 植物やメタボローム・DDBJ 等の他の DB との連携(【研究開発実施項目 1-6】)	6
(2) 進捗状況	6
① MAG データを用いたゲノム・メタゲノムの統合化(【研究開発実施項目 1-1】)、MAG データに関する国際連携(【研究開発実施項目 3-2】)	6
② 微生物 Phenotype 情報の RDF 化(【研究開発実施項目 1-2】)、微生物 Phenotype 情報のデータ構築(【研究開発実施項目 4】)	7
③ マイクロバイオーム解析パイプラインの大幅な速度向上(【研究開発実施項目 1-3】)、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の開発(【研究開発実施項目 2】)	7
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】)、ヒトマイクロバイオームのメタデータ整備とキラータセットの開発(【研究開発実施項目 3-1】)	8
⑤ Microbiome Datahub のデータを用いたキラータアプリケーションの開発(【研究開発実施項目 1-5】)	8
⑥ 植物やメタボローム・DDBJ 等の他の DB との連携(【研究開発実施項目 1-6】)	8
§4. 成果発表等	10
(1) 原著論文発表	10
① 論文数概要	10
② 論文詳細情報	10
(2) その他の著作物(総説、書籍など)	10
(3) 国際学会および国内学会発表	10
① 概要	10
② 招待講演	11
③ 口頭講演	11
④ ポスター発表	12
(4) 知的財産権の出願(国内の出願件数のみ公開)	12
出願件数	12
(5) 受賞・報道等	12
① 受賞	12
② メディア報道	12
③ その他の成果発表	12
§5. 主要なデータベースの利活用状況	13
1. アクセス数	13

① 実績 .....	13
② 分析 .....	13
2. データベースの利用状況を示すアクセス数以外の指標 .....	13
3. データベースの利活用により得られた研究成果(生命科学研究への波及効果) .....	13
4. データベースの利活用によりもたらされた産業への波及効果や科学技術のイノベーション(産業や科学技術への波及効果) .....	14
§6. 研究開発期間中に主催した活動(ワークショップ等) .....	15
(1) 進捗ミーティング .....	15
(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 .....	15

## §1. 研究実施体制

グループ名	研究代表者または主たる共同研究者氏名	所属機関・役職名	研究題目
遺伝研グループ	森 宙史	情報・システム研究機構 国立遺伝学研究所	Microbiome Datahubのマイクロバイオーム研究の国際的なデータハブへの発展に向けた研究開発と運用
基生研グループ	内山 郁夫	基礎生物学研究所	微生物オーソログデータの構築と、そのゲノム機能推定への応用
東工大グループ	山田 拓司	東京工業大学	ヒトマイクロバイオームのメタデータ整備とキラードータセットの開発
東大グループ	西村 祐貴	東京大学	微生物Phenotype情報のデータ構築とRDF化

## §2. 研究開発対象とするデータベース・ツール等

### (1) データベース一覧

#### 【主なデータベース】

No.	名称	別称・略称	URL
1	Microbiome Datahub		<a href="https://mdatahub.org/">https://mdatahub.org/</a>

#### 【その他のデータベース】

No.	名称	別称・略称	URL
1	Microbial Genome Database for Comparative Analysis	MBGD	<a href="https://mbgd.nibb.ac.jp/">https://mbgd.nibb.ac.jp/</a>

### (2) ツール等一覧

No.	名称	別称・略称	URL
1	Metagenome and Microbes Environmental Ontology	MEO	<a href="https://bioportal.bioontology.org/ontologies/MEO">https://bioportal.bioontology.org/ontologies/MEO</a>
2	Latent Environment Allocation	LEA	<a href="http://leamicrobe.jp">http://leamicrobe.jp</a>
3	PZLAST		<a href="https://pzlast.nig.ac.jp">https://pzlast.nig.ac.jp</a>

### §3. 実施内容

#### (1) 本年度の研究開発計画と達成目標

新たなデータの統合化の範囲をマイクロバイオ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キラードータセット・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。

#### ① MAG データを用いたゲノム・メタゲノムの統合化(【研究開発実施項目 1-1】)、MAG データに関する国際連携(【研究開発実施項目 3-2】)

メタゲノム由来の Metagenome Assembled Genome (MAG)について、INSDC DDBJ/ENA/NCBI を中心に GTDB や、IMG/M、MGnify などの主要なメタゲノムの DB から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する。MAG の系統推定と品質評価、DFAST を用いた自動アノテーションについて、数十万ゲノムを超える MAG データに対して高速に計算を実行可能なように昨年度開発したワークフローの改良を行う。また、単離菌のゲノムデータと MAG データについて、Microbiome Datahub で検索および結果表示するコンテンツの追加と洗練化を行い、MAG データを扱うデータベースとして論文化および MAG データを利用した海外の研究者と国際連携が出来るようにする。

#### ② 微生物 Phenotype 情報の RDF 化(【研究開発実施項目 1-2】)、微生物 Phenotype 情報のデータ構築(【研究開発実施項目 4】)

昨年度 Phenotype 情報付きの菌株データを提供いただいた菌株保存機関の JCM と NBRC から更新版の菌株データを提供いただく。提供いただいた菌株データ中に存在する Phenotype データについて、東大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た上で、ゲノム解読済みの株については Phenotype データを Microbiome Datahub で閲覧可能にする。さらに、単離菌のみを対象としていた Bac2Feature について、MAG に対しても Phenotype 推定が可能か検証する。

#### ③ マイクロバイオーム解析パイプラインの大幅な速度向上(【研究開発実施項目 1-3】)、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の開発(【研究開発実施項目 2】)

爆発的な勢いで増加を続けるマイクロバイオームデータに対応するために、昨年度大幅な改良を行った系統組成解析パイプラインについて、大量のプロジェクトのマイクロバイオームデータに対して計算を実行し、Microbiome Datahub で BioProject 単位で系統組成を表示可能にする。また、大量のゲノムや MAG データに対する高速なオーソログアサインメントパイプラインを構築する。

#### 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】)、ヒトマイクロバイオームのメタデータ整備とキラードータセットの開発(【研究開発実施項目 3-1】)

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うと共に、散らばっている MAG データについて大規模プロジェクトから優先して論文からの MAG データ置き場の情報の収集を行い、Microbiome Datahub でそれらのプロジェクト由来の MAG データも検索可能にする。昨年度開発した MEO 等のオントロジー自動アノテーションツール MDatahubAnnotator については、例外処理の追加などのプログラムの改良を行う。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う。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て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い結果表示を行えるようにする。

## ⑤ Microbiome Datahub のデータを用いたキラアアプリケーションの開発(【研究開発実施項目 1-5】)

Microbiome Datahub のユーザ数向上を目的として、LEA や PZLAST 等の、Microbiome Datahub のマイクロバイオームデータを用いた既存のツールの機能強化と論文化を行う。特に、メタゲノムのアミノ酸配列データに対する高速な配列類似性検索ツール PZLAST については、MAG が持つアミノ酸配列データに対する配列類似性検索機能も追加し、PZLAST2 として公開して論文化を行う。また、Microbiome Datahub の MAG データに対してユーザが自分の MAG データを入力して DNA 対 DNA で配列類似性検索ができる機能を開発する。利用者視点に立った DB 開発を進めるために、Microbiome Datahub についてアドバイザリーボードによるアドバイザリー委員会を今年度開催し、DB についての幅広いユーザの意見を取り入れて開発を進める。

## ⑥ 植物やメタボローム・DDBJ 等の他の DB との連携(【研究開発実施項目 1-6】)

マイクロバイオームの配列データと同時に測定されることが増えつつあるメタボロームデータやヒトや植物等の宿主のオミクスデータ、および微生物が持つ糖鎖情報について、主に統合化推進プログラムの他の DB とのデータ連携について他のグループと協議しつつ連携を進める。また、同じ遺伝研にある DDBJ と特に MAG データについて連携を強め、急速な勢いで増加し INSDC でも毎年のようにメタデータ項目や格納するデータベース等が変わる MAG データに対して迅速に対応できるようにする。

これらの研究開発を、遺伝研グループを中核とし、基生研グループ、東工大グループ、東大グループの計 4 グループで分担して実施する。

## (2) 進捗状況

### ① MAG データを用いたゲノム・メタゲノムの統合化(【研究開発実施項目 1-1】)、MAG データに関する国際連携(【研究開発実施項目 3-2】)

INSDC から取得した約 21 万個の MAG 配列データおよび対応する BioProject のメタデータを Microbiome Datahub 上で Elasticsearch によって環境情報等で BioProject と MAG を絞り込み検索できるようにしており、ユーザインターフェースや絞り込み条件について精査し改良を行なった。BioProject と MAG データについては、メタデータを JSON 形式でダウンロード可能な API を開発し Microbiome Datahub から公開した。データの拡充についても、EMBL-EBI で開発し公開している MAG データベースである MGnify から MAG 配列データ約 31 万個を取得し、DFAST および DFAST QC を用いて MAG ごとに遺伝子アノテーションとクオリティチェックを行い、現在 Microbiome Datahub に掲載準備中である。

MAG データに関する国際連携については、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)に東工大グループと遺伝研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 グループと話し合った。

## ② 微生物 Phenotype 情報の RDF 化(【研究開発実施項目 1-2】)、微生物 Phenotype 情報のデータ構築(【研究開発実施項目 4】)

菌株保存機関の NBRC から Phenotype 情報付きの菌株データの RDF ファイルを 2023 年に新たに RDF ポータル経由で最新版として提供いただいた。提供いただいた菌株データ中に存在する Phenotype データについて、昨年度と同様に項目と値をリスト化し、OMP や MPO 等の既存の微生物のフェノタイプオントロジーでのカバー率を調べた。これまで開発してきた Phenotype アノテーションツールである Bac2Feature について、性能向上のためのさまざまな検証やその結果に基づくパイプラインの見直しを行なった。本パイプラインを使った Phenotype アノテーションにあたっては、特に(1)最適な形質推定手法の選択、(2)系統シグナルと進化距離が形質予測に及ぼす影響、(3)16S rRNA 遺伝子のアンプリコン解析のためのプライマーセットが形質予測に及ぼす影響、が大きな問題になる。まず(1)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。(2)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分かった。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。(3)については各種プライマーセットを想定したシミュレーションを行うことで、プライマーセットの違いによる影響を検討した。結果、V1-V2 領域以上の長い領域を増幅しているのであれば、16S rRNA 遺伝子全長を使った場合とほとんど遜色のない形質予測精度が得られることが分かった。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。これらの新たな知見を踏まえ、次年度以降は、他グループと連携して MAG の情報などを積極的に取り込みつつ、実データへの適用例を積み上げていくことで、本パイプラインの有用性を示したい。

## ③ マイクロバイーム解析パイプラインの大幅な速度向上(【研究開発実施項目 1-3】)、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の開発(【研究開発実施項目 2】)

大量に存在する MAG 配列間の高速度な距離計算を行えるツール skani を 21 万 MAG データに対して試し、現実的な時間で MAG 間の距離を計算可能であることを確認した。21 万 MAG 由来の約 4.5 億タンパク質配列について、スーパーコンピュータ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試し、計算時間は数十日レベルでかかるものの、実行可能であることを確認した。MAG 配列データと MAG 由来タンパク質配列の冗長性削減の計算を現在実施中であり、よりユーザにとって使いやすい MAG データセットを構築すると共に 2024 年度に行うデータ更新の際に解析パイプラインで計算が必要な MAG 配列の削減を行う。MBGD オーソロググループについては、遺伝子系統樹と KEGG Orthology(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対応づけを改善した。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、MBGD のオーソロググループをドラフトゲノムに拡張する方法の改良を行っている。これは、従来はドラフトゲノムを既存のオーソロググループと対応付けることのみ

を行っていたところを、対応がつかなかった配列を対象に新たなクラスタリングを行うことで、新規のオーソロググループを作成するよう改良するものである。これを効率的に行うため、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、階層的な手順として実装しているが、対象とするドラフトゲノム数が巨大になるため、さらなる効率化の可能性についても検討している。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 34,079 件と倍増し、シングルトン含めたクラスター数は 1,433,670 で約 1.3 倍になった。最後にドラフトゲノムへの拡張処理が残っているが、先述の拡張プロトコルはまだ完成していないため、さしあたりは従来通りの差別的な拡張のみを行って次期バージョンとして公開する方向で準備している。その後、新規の拡張プロトコルを用いてオーソログレパートリーの拡張を試みる予定である。

#### 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】)、ヒトマイクロバイオームのメタデータ整備とキラデータセットの開発(【研究開発実施項目 3-1】)

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、ChatGPT 4.0 の API を用いて論文中からサンプルが由来した国や病名、DNA 抽出キットの情報等のメタデータを自動で抽出するプログラムを作成し、性能評価を行なった。評価の結果、東工大グループによって論文からマニュアルで抽出したメタデータと自動抽出の結果がよく一致することが確認でき、さらに BioSample からのメタデータ抽出では欠落していたメタデータが得られていることを確認した。2024 年度に ChatGPT 4.0 の API を利用したメタデータ抽出プログラムをヒトマイクロバイオーム論文数万論文に適用してメタデータ抽出を行う。また、BioSample に記述されている環境情報について環境オントロジー MEO を用いたアノテーション付けについても、ChatGPT 4.0 API を利用して自動で行う。一方でマニュアルキュレーションによる論文からのメタデータ収録は継続して行っている。マニュアルキュレーションデータは LLM を活用したデータ取得の正当性を検証するための検証データとしても有用であり、ChatGPT 4.0 の API を利用したメタデータ抽出プログラムでサポートしつつ今後も継続的に行う予定である。

#### ⑤ Microbiome Datahub のデータを用いたキラアプリケーションの開発(【研究開発実施項目 1-5】)

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 のメタデータや MAG のメタデータが取得可能であるが、2024 年度に BioProject ごとのサンプルの系統組成や MAG の配列データについても API 経由でダウンロード可能にする。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現在開発中である。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)を実現する上で必要な配列類似性検索のアルゴリズム部分を開発した。配列をクラスタリングした結果をリファレンスに用いるか否かは現在検討中であるが、各 BioSample の環境情報とあわせて、2024 年度に PZLAST-MAG を公開予定である。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。

#### ⑥ 植物やメタボローム・DDBJ 等の他の DB との連携(【研究開発実施項目 1-6】)

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 DFAST 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た。また、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 Togethon 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 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。

## §4. 成果発表等

### (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)	5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. 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).
2. 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).
3. Hirotsugu Shiroma, Youssef Darzi, Etsuko Terajima, Zenichi Nakagawa, Hirotaka Tsukahira, 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).
4. 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).
5. 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).

#### (2) その他の著作物(総説、書籍など)

1. 森宙史, 東光一, 黒川頭, 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・ティイー・エス出版, pp325–335, 2023年.

### (3) 国際学会および国内学会発表

#### ① 概要

種別	国内外	件数
招待講演	国内	8件

種別	国内外	件数
	国際	0 件
口頭発表	国内	3 件
	国際	2 件
ポスター発表	国内	4 件
	国際	1 件

## ② 招待講演

〈国内〉

1. 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, 2023/6/30
2. 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, 2023/07/24
3. 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, 2023/09/07
4. 森宙史, マイクロバイオ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, 2023/09/05
5. 森宙史, マイクロバイオ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, 2023/09/14
6. 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, 2023/10/31
7. 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, 2023/12/22
8. 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, 2023/12/26

〈国際〉

該当なし。

## ③ 口頭講演

〈国内〉

1. 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.
2. 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.
3. 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.

〈国際〉

1. Motomu Matsui, Reconstructing Phylogenetic Reconstruction, DATA SCIENCE FOR GENOME DYNAMICS, Online, 2024/01/09
2. 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. 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。
2. 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、日本ゲノム微生物学会、かずさアカデミアホール、2024年3月12日。
3. Mikihiko Kawai, Motohiro Mihara, Hiroyo Nishide, Hirokazu Chiba, Ikuo Uchiyama, Phylogenetic profiling re-visited: exploration of alternate orthologs in functional modules, 日本ゲノム微生物学会、かずさアカデミアホール、2024年3月12日。
4. Jiang Zhenbo, Uchiyama Ikuo, Extraction of the shared co-expression patterns based on extended pan-network and core-network concepts, 日本ゲノム微生物学会、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 projects and MAG data, The 14th International Workshop on Advanced Genomics, Hitotsubashi Hall, Tokyo, October 4-6, 2023.

#### (4) 知的財産権の出願（国内の出願件数のみ公開）

出願件数

種別	件数
特許出願	0件

#### (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/>

##### ③ その他の成果発表

該当なし。

## §5. 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 1 研究開発対象の主要なデータベースの利用状況(月平均)

DB名	種別	2023年度
MicrobeDB.jp (2024年度4月に公開停止)	訪問者数	895
	訪問数	1,303
	閲覧ページ数	274,273
Microbiome Datahub (2024年度6月に正式公開予定)	訪問者数	N.D.
	訪問数	N.D.
	閲覧ページ数	N.D.

#### ② 分析

旧データベースから新データベースへの移行期であるため評価は難しいが、MicrobeDB.jp については最終更新から年月から経っており、Metagenomic Thermometer や福島第一原発のマイクロバイオーーム解析の論文等でMicrobeDB.jp について言及はしているものの、2023年度はアクセス数は減少傾向にあった。MicrobeDB.jp と Microbiome Datahub は同じサーバで運用しており、2024年度6月にMicrobeDB.jp から Microbiome Datahub へと完全に切り替えるため、今後は Microbiome Datahub のアクセス数を元に判断していきたい。

### 2. データベースの利用状況を示すアクセス数以外の指標

特に無し

### 3. データベースの利活用により得られた研究成果(生命科学研究への波及効果)

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](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>)

#### 4. データベースの利活用によりもたらされた産業への波及効果や科学技術のイノベーション(産業や科学技術への波及効果)

1. 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 CO<sub>2</sub> 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([https://www.nedo.go.jp/news/press/AA5\\_101619.html](https://www.nedo.go.jp/news/press/AA5_101619.html))において、「CO<sub>2</sub> 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 等と共同で開始している(2023 年 6 月 1 日プレスリリース <https://www.nig.ac.jp/nig/images/information/PR20230601.pdf>)。この中で、日本国内のいくつかの企業と CO<sub>2</sub> 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 CO<sub>2</sub> 固定関連の MAG について NEDO プロジェクトで整理し再解析等を行なった上で、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。

## §6. 研究開発期間中に主催した活動(ワークショップ等)

### (1) 進捗ミーティング

年月日	名称	場所	参加人数	目的・概要
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
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

### (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 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1 既公開のデータベース・ウェブツール等

No.	正式名称	別称・略称	概要	URL	公開日	状態	分類	関連論文
1	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。	<a href="https://microbedb.jp">https://microbedb.jp</a>	2011/12/12	休止・閉鎖	データベース等	
2	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。	<a href="https://mdatahub.org/">https://mdatahub.org/</a>	2023/5/31	維持・発展	データベース等	
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。	<a href="https://mbgd.nibb.ac.jp">https://mbgd.nibb.ac.jp</a>	1997/7/1	維持・発展	データベース等	Ikuo 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", <i>Nucleic Acids Res.</i> , 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。	<a href="https://bioportal.bioontology.org/ontologies/MEO">https://bioportal.bioontology.org/ontologies/MEO</a>	2012/2/21	維持・発展	ツール等	
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。ユーザー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。	<a href="http://leamicrobe.jp">http://leamicrobe.jp</a>	2018/6/6	維持・発展	ツール等	Koichi Higashi, Shinya Suzuki, Susumu Kurosawa, Hiroshi Mori, Ken Kurokawa, "Latent environment allocation of microbial community data", <i>PLoS Comput Biol</i> , 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。	<a href="https://pzlast.nig.ac.jp">https://pzlast.nig.ac.jp</a>	2020/4/1	維持・発展	ツール等	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", <i>Bioinformatics.</i> , 37, 3944-3946, 2021. (DOI: 10.1093/bioinformatics/btab492).