

研究開発実施報告

□概要

研究開発課題名	データサイエンスを加速させる微生物統合データベースの高度実用化開発
開発対象データベースの名称 (URL)	微生物統合データベース「MicrobeDB.jp」 (http://microbedb.jp/)
研究代表者氏名	黒川 顕
所属・役職	国立遺伝学研究所・生命情報研究センター・教授

□目次

§1. 研究実施体制	2	① 論文数概要	11
§2. 研究開発対象とするデータベース・ツール等	3	② 論文詳細情報	11
(1) データベース一覧	3	(2) その他の著作物(総説、書籍など)	11
(2) ツール等一覧	3	(3) 国際学会および国内学会発表	11
§3. 実施内容	4	① 概要	12
(1) 本年度の研究開発計画と達成目標	4	② 招待講演	12
(2) 進捗状況	6	③ 口頭講演	12
(3) 追加支援による実施結果	9	④ ポスター発表	13
① 実施タイプ	9	(4) 知的財産権の出願	14
② 実施内容と結果	9	(5) 受賞・報道等	14
③ 期待される将来的な効果	9	§5. 研究開発期間中に主催した活動(ワークショップ等)	
① 実施タイプ	10	14
② 実施内容と結果	10	1. 進捗ミーティング	14
③ 期待される将来的な効果	10	2. 主催したワークショップ、シンポジウム、アウトリー	
§4. 成果発表等	11	チ活動等	15
(1) 原著論文発表	11		

§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	黒川 顕	国立遺伝学研究所・教授	MicrobeDB.jpの実用化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・助教	オーソログデータの構築とそれに基づく比較ゲノム解析機能の開発
東工大グループ	山田 拓司	東京工業大学・准教授	ヒトマイクロバイオームデータの価値最大化を目指したメタデータ構築
千葉大グループ	高橋 弘喜	千葉大学・准教授	真菌類ゲノム・菌株・オミックス情報の収集と高度化

§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	MicrobeDB.jp		http://microbedb.jp

【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	http://mbgd.genome.ad.jp

(2) ツール等一覧

No.	名称	別称・略称	URL
1	Metagenome and Microbes Environmental Ontology	MEO	http://bioportal.bioontology.org/ontologies/MEO
2	Microbial Culture Collection Vocabulary	MCCV	http://bioportal.bioontology.org/ontologies/MCCV
3	Pathogenic Disease Ontology with Symptom	PDO-CSSO	http://bioportal.bioontology.org/ontologies/PDO , http://bioportal.bioontology.org/ontologies/CSSO
4	Metagenome Sample Vocabulary	MSV	http://bioportal.bioontology.org/ontologies/MSV

§3. 実施内容

(1) 本年度の研究開発計画と達成目標

1) 本年度の研究開発計画

本研究開発では、①徹底したユーザビリティの向上、②データ品質の向上、③キラーアプリケーションの開発、④さらなるデータの統合、⑤基盤データ解析技術の高度化、⑥効率的運用、⑦ホログenom対応、の7項目を中心に研究開発を実施し、MicrobeDB.jpの実用化を目指す。

本年度、①については、今までのMicrobeDB.jpのキーワード検索を想定したGoogle様インターフェースに加え、ユースケースに特化した幾つかのStanza・アプリケーションをリスト化した、ポータルサイトを構築し、ユーザ会等でのフィードバックを受けるためのテスト版として公開する。開発するポータルサイトにおいては、RDFストアに格納されている、遺伝子、環境、系統をはじめとする多様なデータセットの内容をポータルサイト上で可視化し、ユーザに対してすでにMicrobeDB.jpが提供している表示系、解析系Stanzaへのダイレクトな動線を提供する。また、MicrobeDB.jpのコアユーザからの意見や要望を収集するとともに、新規技術の開発や、解析Stanzaの開発などを技術共有しデータサイエンスの醸成を図ることを目的として、各関連学会会員、企業コンソーシアム、共同研究者等で構成された、MicrobeDB.jpユーザ会を設置し、ユースケースの発掘とDBのヘビーユーザの獲得等を行う。

②については、ゲノムデータの爆発的な増加にも対応し最新のオーソログデータを安定的にMicrobeDB.jpで提供するために、個々のゲノムデータの優先度評価と効果的な選定方法の開発や、差分的更新手順の確立を目指した研究開発を行う。また、既にINSDC DRA/ERA/SRA等で配列データが公開されているヒト共生細菌群集のメタ16S・メタゲノム解析のサンプルについて、マニュアルキュレーションまたはテキストマイニング技術を用いた論文からの、宿主についての疾患関連情報等の様々な情報の抽出を効率的に行う手法の開発を行う。また、人工物等を中心に解析される環境が多様化しているメタゲノムデータの、分離元環境の記述を適切な粒度で行うために、MEOやMSV等のオントロジーについて、新しい語句の追加や語句間の関係性の見直し等のオントロジーの更新作業を行う。

③については、当グループで本プロジェクトとは独立に開発した、メタ16S・メタゲノム解析データからの系統組成推定ツールVITCOMIC2や、MicrobeDB.jpに存在するメタ16S・メタゲノム解析サンプルを微生物群集の組成と、対応する自然言語によるメタデータの記述の両者を統合して、世界中に存在する微生物群集の全体像を俯瞰可能な統合可視化ツールLEA等の、MicrobeDB.jpのデータとの親和性が高い解析ツールについて、それらの解析結果を閲覧するユーザのMicrobeDB.jpへの誘導を自然に行うために、機能の追加等を行い、MicrobeDB.jpのユーザ数増大に寄与するキラーアプリケーションにできるように開発を行う。また、昨年度までに出展した各種学会や展示会等で好評だった、何種類かの典型的な比較ゲノム解析をMicrobeDB.jp上でマウスクリックのみで行うことが可能な比較ゲノム解析Stanzaについて、MicrobeDB.jp中のゲノム以外の様々な情報をより活かした解析を行えるように改良を行う。さらなる比較ゲノムアプリケーション作成のため、MicrobeDB.jpとMBGDとの連携に加えて、MBGDとMAPLEを連携させた解析の実現に向けて、データの受け渡し方法などについて検討し、MBGDおよびMAPLE双方の修正を行う。それに基づいてMBGD/MicrobeDB.jpとMAPLEとを連携させた比較ゲノムアプリケーションの試作を行う。

④については、真核微生物であり、産業上有用な系統も多い真菌類について、完全長・ドラフトゲノムデータと、RNA-Seqデータの収集とRDF化を行うとともに、菌株リソースのメタデータの整理とRDF化を行う。また、それらの系統の環境中の分布を記述する上で有用な、メタITS解析について、公共DB中の大量サンプルを

統一した解析パイプラインで解析し、結果をユーザが **MicrobeDB.jp** で検索および比較解析できることを目的として、既存の解析手法や系統推定に用いる **Reference** 配列 DB の評価検討を行い、大量のメタ ITS 解析データから品質の高い系統組成データを得るために必要な研究開発項目の明確化を行う。

⑤については、爆発的な勢いで年々増加している公共のメタ 16S・メタゲノム解析データを網羅的に収集、再解析し、系統組成や遺伝子機能組成等の情報を高速かつ高精度に計算するために、**MicrobeDB.jp** で用いているメタ 16S・メタゲノム解析パイプラインを最新の解析手法を取り入れたものへとアップデートする。具体的には、**MicrobeDB.jp** で用いている、**VITCOMIC2** について、配列相同性検索に用いるリファレンス 16S rRNA 遺伝子配列 DB を作成するための配列クラスタリング手法を、当グループで開発した高速かつ高精度な配列クラスタリングツール **SaturnCluster** を用いることによって、より高精度な系統組成推定を実現する。また、メタゲノムデータからの遺伝子機能組成の推定については、現状のパイプラインで用いている配列相同性検索ツール **GHOSTX** を、同等の精度でより高速な **DIAMOND** 等のツールに置き換えることによって、より高速化を実現する。

⑥については、ゲノム・メタ 16S・メタゲノムデータのサンプル分離元の環境の記述に用いている **MEO** について、昨年度に開発した自動アノテーションツールの精度を向上するために、教師データとしている **MEO** のマニュアルアノテーションデータの追加等の作業を行う。

⑦については、ホログenom研究の基盤として **MicrobeDB.jp** を利用できるようにするためのユースケースとして、同様にセマンティックウェブ技術を用いて植物のデータの統合化を進めているかずさ DNA 研究所のグループと、**SPARQLthon** 等の機会を利用して、両 DB 間で統合したいデータおよび必要な作業の明確化を行うための話し合いを行い、連携を開始する。データベースを活用した共同研究として、まず「共生微生物ゲノム情報統合環境構築」を実施する。共生微生物ゲノムの情報インフラ整備として **MicrobeDB.jp** のサブセットである植物共生微生物ゲノムデータベース **RhizoBase** のゲノムもしくは共生アイランドの配列位置情報表示・情報連携のためのインタフェースとして **JBrowse** 環境を導入し、遺伝子情報など配列座標系にマップ可能な **RDF** トラックを追加することで、微生物共生体のゲノム上へのデータ統合を容易にするためのインタフェースを整備する。これと並行して、基礎生物学研究所共生システム研究部門・川口正代司教授、またかずさ DNA 研究所との共同研究として、要望に基づいた「ホログenom解析支援ツール開発」を実施する。これは、ヒト＝微生物叢ならびに植物＝微生物叢の全体像解明を目的としたホログenom解析支援系の構築である。ここでは植物・微生物共生系の全体像解明を目的として川口教授他が対象としているモデル研究システムであるミヤコグサ **Lotus japonicus** およびミヤコグサ根粒菌 **Mesorhizobium loti** を中心に、さまざまな動植物＝微生物叢のゲノムと相互作用の解析の支援ツールを構築する。具体的には、かずさ DNA 研究所による植物統合化推進プログラム・川口教授との共同研究により、**RhizoBase** が納める根粒菌ゲノムと、宿主である植物ゲノムを統合的に解析・閲覧可能な系を開発する。さらに、**RDF** を利用するメリットを生かし、各種メタデータ、宿主＝共生体生物横断的に遺伝子とその関連情報を提示するスタンプを開発することで、**Nod factor** 関連遺伝子や宿主側の菌体外多糖受容体等の共生関連遺伝子一覧等の解析支援情報の提供を可能にし、ホログenom研究者を支援するとともに、ヒト・植物・微生物それぞれの領域の研究者がホログenom研究を通じ未知の生物との関係性に関する知見を容易に得られることで、新たな健康・医療・食料・薬品等の開発・生産に役立つ統合されたデータベースを提供できるよう整備する。ホログenomデータ解析については、解析手法およびデータ公開等共同研究者の要求に応じて柔軟に対応するとともに、**MicrobeDB.jp** において利用しているメタゲノム解析パイプライン **MeGAP/16S** 系統解析データと変換 **RDF** を組み込む方法により開発する。

上記研究開発と並行して、MicrobeDB.jp の基盤となるゲノム・メタゲノム・オーソログ等のデータを、最新のデータへと更新する作業を行う。

2) 本年度の達成目標

達成目標	タスク
ユースケースに特化した様々なアプリケーションをリスト化した、MicrobeDB.jp のポータルサイトを構築する。	・ポータルサイトの構築
MicrobeDB.jp ユーザ会を設置する。	・ユーザ会の設置
ヒト共生細菌群集のメタ 16S・メタゲノム解析のサンプルについて、公共 DB にはメタデータとして登録されていないが比較解析を行う上で重要な情報を収集するための手法開発	・論文から細菌群集の宿主についての疾患関連情報等の様々な情報を効率的に抽出する手法の開発
VITCOMIC2 等のキラアプリ候補に MicrobeDB.jp への誘導のための機能を付加する。	・個々のキラアプリ候補から、MicrobeDB.jp へユーザを誘導するための機能を開発
真菌類について、完全長・ドラフトゲノムデータと、RNA-Seq データ、菌株リソースのメタデータの収集・整理を行う。	・真菌類の完全長・ドラフトゲノムデータと、RNA-Seq データ、菌株リソースのメタデータの公共 DB 等からの収集・整理
MicrobeDB.jp で用いているメタ 16S・メタゲノム解析パイプラインを最新の解析手法を取り入れたものへとアップデートする。	・メタ 16S・メタゲノム解析パイプラインのアップデート
MEO の自動アノテーションツールの精度を向上するために、教師データとしている ME O のマニュアルアノテーションデータを追加する。	・ゲノム・メタゲノム等のメタデータの整理および環境関連語句の抽出と教師データ中に存在しない新規語句への MEO を用いたマニュアルアノテーション

(2)進捗状況

本研究開発では、MicrobeDB.jp の実用化を目指し、中心となる以下の 7 項目について研究開発を実施した。

①徹底したユーザビリティの向上

これまで MicrobeDB.jp において提供しているキーワード検索機能および表示系、解析系 Stanza 機能を見直し、ダイレクトな動線を設定するとともに、デザインを刷新しユーザインターフェースを改良することで、利便性の高いポータルサイトを構築した。これまで GenomeRefine を通じてユーザデータを登録していた機能を、ポータルサイト上で実施可能とするため、GenomeRefine のレポジトリ機能をポータルへ移植し、ウェブサービスの統合化を実施した。さらに、レポジトリ機能については、多様なメタデータの入力および DDBJ 登録形式に対応した簡便かつ高速な微生物ゲノムアノテーションパイプライン DFAST および MBGD オーソログ解析のポータルサイト上での実行と、DDBJ Read Annotation Pipeline に実装したメタゲノム解析パイプライン MeGAP 結果を利用可能とする連携を行った。さらに、MicrobeDB.jp 上での公開データとユーザデータとの比較解析において入力が必要なメタデータについては、DDBJ BioSample 登録でも再利用を可能とする連携を行なった。ユースケースとしては、これまでユーザデータのゲノムおよびメタゲノム比較解析をユースケースとしてリストし、MicrobeDB.jp ユーザ講習会での利用を目的としてテスト版 (<https://beta.microbedb.jp>) を公開した。

②データ品質の向上

従来の MBGD では、属ごとに代表ゲノムを一つとって比較した標準オーソログテーブルに加えて、系統群ご

とに作成したオーソログテーブルを用意して、それらに用途に応じて使い分けるようにしていた。これを、種内比較、属内比較、属間比較と段階的に進め、各段階でパンゲノムをとって階層的につなげるプロトコルに変更した。これによって、対象となる 6,318 個の完全ゲノム(2,250 万遺伝子)に対して、全遺伝子レパートリーをカバーする代表遺伝子 460 万が抽出され、それらが 75 万あまりのオーソロググループに分類されることとなった。作成したオーソログテーブルは登録されたゲノム中の遺伝子全体をカバーしているが、それらの検索や表示を行うためには、種内、属内、属間というオーソログテーブルの階層的な構造を考慮して行う必要がある。そこで、MBGD の各情報表示画面や検索機能などについて必要な修正を行って、対応させた。年度終了時点で、修正が必要な点がまだ若干残っているために公開には至っていないが、次年度のなるべく早い時期に公開すべく準備を進めている。

メタゲノム配列が紐付いている論文約 4,500 報を文献情報データベースから収集し、そこからマニュアルキュレーションによってメタデータ抽出をおこない、これまでに 762 報について完了している。このうち、メタデータが利用可能な論文は 40 報あり、メタデータの項目数は 513 項目になる。これらのメタデータの中には、項目名が少し異なるだけでその意味がほぼ同一であるものが多く存在している。その意味を考慮し項目名をマニュアルキュレーションにより 176 項目へと大別した。また同時にメタデータのオントロジー整備やフォーマット策定作業も実施している。これら作業の完了後にはデータを RDF 形式に変換し、MicrobeDB.jp に格納する。

生物の環境記述の世界標準である ENVO の開発者と、BioHackathon の際に MEO と ENVO との対応付けの作業を行い、両オントロジー間での連携を強化した。

③キラーアプリケーションの開発

より広い範囲のユーザを MicrobeDB.jp へ誘導することを目的として、MicrobeDB.jp に存在するメタ 16S・メタゲノム解析サンプルを微生物群集の組成と、対応する自然言語によるメタデータの記述の両者を統合して、世界中に存在する微生物群集の全体像を俯瞰可能な統合可視化ツール LEA に、音声対話機能を付加した LEA/MDB Talk を開発した。

また、オーソログデータを活用したアプリケーションとして、利用者のクエリ配列に対してオーソログデータベースを検索して、オーソロググループへの分類を行うアプリケーションを作成した。検索の方式として、ホモロジー検索のトップヒットに基づく方法と、各グループ内配列のマルチプルアライメントを用いたプロファイル検索に基づく方法とを実装した。ホモロジー検索とプロファイル検索のプログラムとして様々なものを用いて比較した結果、このどちらの検索にも対応している MMseqs が検索速度、精度ともに優れていることが分かったのでこれを主な検索エンジンとして採用した。ただし、個別もしくは少数の遺伝子のアノテーションに向けた検索エンジンとして、速度は遅いがより精度の高い HMMER による実装も行った。一方、本アプリケーションは当面は MBGD において利用者ゲノムを取り込むための MyMBGD 機能の一部として作成しているが、最終的に MicrobeDB.jp サーバと連携させて、MicrobeDB.jp で登録されたゲノムデータの解析をも可能にする方針である。そのためデータ交換方式について検討するとともに、共通の認証系として TAUTH を用いたユーザ認証に対応させることについても検討した。

④さらなるデータの統合

真菌類の系統分類に用いる ITS 配列の整備を進めた。具体的には、千葉大真菌センターが保有する菌株の ITS 配列の収集・整理を行った。収集した 511 株の ITS 配列には、rRNA 領域、ITS 領域などのアノテーション情報が付与されてなかったため、まずは、既存の ITS 配列とのアライメントにより rRNA 領域と ITS 領域を同定した。その後、DDBJ への登録作業を進め、アクセッション番号の取得を完了した。一方で、真菌叢解析

手法の検討に着手した。細菌叢に比して、解析手法が確立されていないため、現状把握と問題点の洗い出しのために、論文調査を進め、国内雑誌へ寄稿した。

オミックスデータの整備については、RNA-seq データに関して整備を進めた。公共データ SRA (NCBI) に登録されているデータを調査したところ、19,432 件の真菌 RNA-seq データの登録を確認した(2017 年 7 月現在)。まず、これら全データを MicrobeDB.jp へ登録するために、遺伝子発現量算出のための解析パイプラインの検討を行った。検討したパイプラインを用いて、現在までに約 3,000 件(約 15%)について、各遺伝子のリード数算出まで完了した。

⑤ 基盤データ解析技術の高度化

爆発的な勢いで年々増加している公共のメタゲノム解析データを網羅的に収集、再解析し、遺伝子機能組成の情報を高速かつ高精度に計算するために、MicrobeDB.jp のメタゲノム解析パイプライン中で用いている配列相同性検索ツール GHOSTX を、同等の精度でより高速な DIAMOND に置き換えることで、さらなる高速化を実現した。

⑥ 効率的運用

MicrobeDB.jp の基盤となるゲノム・メタゲノム・オーソログのデータおよび関連メタデータについて、最新データの追加・更新を実施するために、一次リソースから RDF 変換およびデータベース投入までのデータフローを整理し、更新系についてメジャー更新および差分更新系整備のための仕様設計に着手した。

ゲノム・メタ 16S・メタゲノムデータのサンプル分離元の環境の記述に用いている MEO について、自動アノテーションツールの精度を向上するために、教師データとしている MEO のマニュアルアノテーションデータに、菌株保存機関である JCM 由来の菌株の分離元の環境の MEO を用いたマニュアルアノテーション結果を追加した。

⑦ ホログenom対応

ホログenom研究の基盤として MicrobeDB.jp を利用できるようにするためのユースケースとして、同様にセマンティックウェブ技術を用いて植物のデータの統合化を進めているかずさ DNA 研究所のグループとの連携を開始した。具体的には、本年度追加支援を受けて、共生微生物ゲノムの情報インフラ整備として MicrobeDB.jp のサブセットである植物共生微生物ゲノムデータベース RhizoBase のゲノム全体もしくは共生アイランドの配列位置情報表示・情報について、連携のためのインタフェイスとして JBrowse 環境を導入し、遺伝子情報など配列座標系にマップ可能な RDF トラックを追加することで、微生物共生体のゲノム上へのデータ統合を容易にするためのインタフェイスを整備することで「共生微生物ゲノム情報統合環境構築」を実施した。また、ヒト＝微生物叢ならびに植物＝微生物叢の全体像解明を目的としたホログenom解析支援系の構築を目的として、基礎生物学研究所共生システム研究部門・川口正代司教授、かずさ DNA 研究所との共同研究として、「ホログenom解析支援ツール開発」を実施した。具体的には、植物・微生物共生系モデル研究システムであるミヤコグザ *Lotus japonicus* およびミヤコグサ根粒菌 *Mesorhizobium loti* について、ゲノム情報および共生関係の RDF 化を実施し、MicrobeDB.jp において利用しているメタゲノム解析パイプライン MeGAP の 16S rRNA 遺伝子を用いた系統組成解析データの変換 RDF を組み込む方法により MicrobeDB.jp 上で解析を可能とする機能開発を行った。

(3) 追加支援による実施結果

(追加支援 1)

① 実施タイプ

タイプ A: データベースを活用した共同研究の実施

② 実施内容と結果

ホログenom研究に関するツールの開発にあたっては、ヒト=微生物叢ならびに植物=微生物叢研究の最新動向に対応するために、実際のデータ生産者・あるいは利用者としてのユーザのニーズの収集と、リアルタイムの開発へのフィードバックを得るために、共同研究ベースで基盤ならびに支援ツールの整備をすすめることが不可欠となる。我々は、長年にわたって植物共生微生物である根粒菌ゲノムアノテーションデータベース RhizoBase (<http://genome.microbedb.jp/rhizobase>) を運用しており、植物・微生物相互作用研究コミュニティとの連携や活用の実績があり、また、共生関連遺伝子や相互作用機作にかかわる文献由来アノテーションを集積済みである。そこで、本追加支援においては、将来的にヒトホログenom研究データベース化や解析支援系にも活用可能な枠組みとして、まず、植物ホログenom解析支援のための開発を共同研究として実施する。具体的には、基礎生物学研究所共生システム研究部門・川口正代司教授とかずさ DNA 研究所での植物統合化推進プログラムとの共同研究にて ①共生微生物ゲノム情報統合環境の構築と、②ホログenom解析支援ツールの開発を実施した。

① 共生微生物ゲノム情報統合環境構築

共生微生物ゲノムの情報インフラ整備として MicrobeDB.jp のサブセットである植物共生微生物ゲノムデータベース RhizoBase のゲノムもしくは共生アイランドの配列位置情報表示・情報連携のためのインタフェースとして JBrowse 環境を導入し、遺伝子情報など配列座標系にマップ可能な RDFトラックを追加し、微生物共生体のゲノム上へのデータ統合を容易にするためのインタフェースを整備した。

② ホログenom解析支援ツール開発

ヒト=微生物叢ならびに植物=微生物叢の全体像解明を目的としたホログenom解析支援系の構築を行う。ここでは植物・微生物共生系の全体像解明を目的として川口教授他が対象としているモデル研究システムであるミヤコグサ *Lotus japonicus* およびミヤコグサ根粒菌 *Mesorhizobium loti* を中心に、さまざまな動植物=微生物叢のゲノムと相互作用の解析の支援ツールを開発し、MicrobeDB.jp ポータルの機能を拡張する形で実装した。既存ホスト-微生物共生菌のメタデータについては、ホストの拡充を想定して、Relation Ontology の語彙を利用した RDF を作成し、ホログenom解析支援ツールから参照可能にした。ゲノム情報については、かずさ DNA 研究所 miyakogusa.jp および RhizoBase 由来のデータをデータ統合のためのゲノム、配列、遺伝子の URI を整理しながら RDF 化を実施し、TogoStanza の仕組みを利用して、PGDBj から公開予定の JBrowse および RhizoBase の JBrowse と合わせて遺伝子情報を提示することで、ホスト-微生物共生菌のゲノム情報を含めた既存の知識を参照可能にした。また、ホログenom解析データの公開および MicrobeDB.jp のデータ更新により、ファセット検索機能によりホスト情報による絞り込みを可能とした。ファセット検索機能については、ポータルサイトの汎用的な機能として開発することで、ホログenomを構成する微生物の選択インターフェースを合わせてファセット検索を可能にした。さらに、ホログenom解析において着目する任意の微生物ゲノムを選択するためのインターフェースを用意した。

③ 期待される将来的な効果

近年、ヒトマイクロバイオームや植物微生物相互作用の解明や農業分野への応用を目的としたホロゲノム解析プロジェクトが進行し、解析系の確立や大量のデータ公開の必要性が予想されているが、ホロゲノム解析を支援し、ホスト生物と微生物群集の統合的なデータベースは存在しない。本年度においては、植物ホロゲノム研究支援のために植物共生微生物叢メタゲノム解析のプライベートデータの解析支援とデータ公開に備えた開発を実施し、MicrobeDB.jp - PGDBj のデータベース間連携のもと、他に例を見ないセマンティック・ウェブ技術をフルに活用したホロゲノム解析支援ツールの提供が可能となる。MicrobeDB.jp の主要なユーザとしては微生物研究者をとっているが、本申請による植物統合データベースとの連携による基盤整備を通じて、ヒトおよび植物ホロゲノム研究者のニーズを実現し、微生物ゲノム情報を介した、ヒト・植物研究者への利用ユーザの拡大と活用事例の増大が期待される。

(追加支援 2)

① 実施タイプ

タイプ A: データベースを活用した共同研究の実施

② 実施内容と結果

本追加支援においては、本体研究計画の「③キラーアプリケーションの開発」における利用者ゲノムに対するオーソログ対応付けの次のステップとして、遺伝子単位ではなく、一連の機能を果たす上で必要な遺伝子集合としての「モジュール」の単位でアノテーションを行えるようにすることを目的として行った。そのような機能モジュールを収集したデータベースとして KEGG MODULE があり、これを利用してモジュールの有無を評価するプログラムとして MAPLE があるため、本支援課題では MBGD と MAPLE を連携させることによってこの機能を実現させることとした。MAPLE はもともと KEGG を直接検索してモジュールの有無を評価する処理までを含んでいるが、その中から KO のリストから KEGG モジュールの充足率を計算する機能を切り出して、これを MBGD オーソログと KO の対応付けに基づいて作成した KO リストに適用することによって連携させた。これにより、利用者のゲノムに対して MBGD オーソログに基づくアノテーションを行い、それに基づいて各 KEGG モジュールの有無の評価を行うことが可能となった。さらに、得られた結果を KEGG パスウェイ図上に表示し、他のゲノムと比較する機能も実装した。

③ 期待される将来的な効果

今後、本機能の完成度を上げて MicrobeDB.jp のゲノム解析アプリケーションの一つとして公開し、利用可能とする。これにより、利用者のゲノムを機能モジュール単位で評価し、他のゲノムと比較する機能を本データベースのアプリケーションの一つとして提供できるようになる。併せて、共同研究者で MAPLE の開発者でもある高見博士と共同で、本開発により作成したツールを利用して、環境中や生体内のメタゲノムサンプルから構築された未培養菌のゲノムデータから、その微生物を機能ポテンシャルの点から特徴づけることを試みる。こうした研究を通じて、本データベースを、環境微生物学や医学などの分野において重要なはたらきを持つ新規未培養菌の機能的特徴づけに活用する道が開かれることが期待される。

§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)	4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. Yasuhiro Tanizawa, Takatomo Fujisawa, Yasukazu Nakamura, "DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication", *Bioinformatics*, vol. 34, Issue 6, , p p. 1037-1039, 2018 (DOI: 10.1093/bioinformatics/btx713)
2. Yu Kanasaki, Masaki Hirose, Yuu Hirse, Takatomo Fujisawa, Yasukazu Nakamura, Satoru Watanabe, Shigeru Matsunaga, Hiroko Uchida, Akio Murakami, "Draft Genome Sequence of the Nitrogen-Fixing and Hormogonia-Inducing Cyanobacterium *Nostoc cycadae* Strain WK-1, Isolated from the Coralloid Roots of *Cycas revoluta*", *Genome Announc.*, vol. 6, No. 7, e00021-18, 2018 (DOI: 10.1128/genomeA.00021-18)
3. Yohei Shimura, Yuu Hirose, Naomi Misawa, Sachiko Wakazuki, Takatomo Fujisawa, Yasukazu Nakamura, Yu Kanasaki, Haruyo Yamaguchi, Masanobu Kawach, "Complete Genome Sequence of a Coastal Cyanobacterium, *Synechococcus* sp. Strain NIES-970", *Genome Announc.*, vol. 5 No. 14, e00139-17, 2017 (DOI: 10.1128/genomeA.00139-17)
4. Hiroshi Mori, Takayuki Maruyama, Masahiro Yano, Takuji Yamada, Ken Kurokawa, VITCOMIC2: visualization tool for the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing, *BMC Syst Biol.*, vol. 12, 30, 2018 (DOI: 10.1186/s12918-018-0545-2)

(2) その他の著作物(総説、書籍など)

1. 森宙史、細菌のゲノム進化と群集ダイナミクス、*生体の科学*、pp155-159、Vol. 68、No. 2、2017年
2. 高橋弘喜、組成解析(真菌)、*化学療法の領域*、pp105-109、Vol.33、No.7、2017年
3. 森宙史、黒川顕、メタゲノム解析・ホログゲノム解析、*化学療法の領域*、pp111-117、Vol. 33、No. 7、2017年
4. 東光一、黒川顕、メタゲノム解析と食品開発への応用について、pp22-27、*明日の食品産業* 7・8月号、2017年
5. Ikuo Uchiyama, Ortholog identification and comparative analysis of microbial genomes using MBGD and RECOG. In "Protein function prediction: Methods and protocols", (D. Kihara ed.), p p.147-168, Humana Press, 2017.

(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	7 件
	国際	1 件
口頭発表	国内	8 件
	国際	5 件
ポスター発表	国内	5 件
	国際	3 件

② 招待講演

〈国内〉

1. 山田拓司、ヒト腸内細菌と大腸がん発病モデル、日本マイクロバイオームコンソーシアム(JMBC)設立記念シンポジウム、THE GRAND HALL 品川、2017/7/28
2. 黒川顕、メタゲノム研究の技術的進展と情報の取扱いについて、ヒトゲノム研究倫理を考える会、グランフロント大阪、2017/10/9
3. 黒川顕、微生物ビッグデータチャレンジ: データサイエンスを加速させる微生物統合データベース "MicrobeDB.jp"、BioJapan2017 次世代微生物革命セミナー、パシフィコ横浜、2017/10/12
4. 黒川顕、データサイエンスを加速させる微生物データの統合化、CBI 学会講演会「創薬・ヘルスケア応用に向けた腸内細菌の最新動向」、キャンパスイノベーションセンター東京、2017/11/10
5. 森宙史、微生物研究開発における微生物統合データベース MicrobeDB.jp の利用法、千葉県バイオ・ライフサイエンス・ネットワーク会議 平成 29 年度講演会「バイオ産業の活性化にむけた公的データベースの活用法、ペリエホール、2017/12/22
6. Hiroshi Mori、Developing tools and database for microbial community analysis、マイクロバイオーム研究開発の最前線、日本橋ライフサイエンスハブ、2018/01/25
7. 黒川顕、微生物統合データベース「MicrobeDB.jp」、第 91 回日本細菌学会総会シンポジウム、福岡国際会議場、2018/3/27

〈国際〉

1. 中村保一、How to make bio-DB's and services sustainable?、日中韓バイオインフォマティクスシンポジウム、ソウル、2017/06/23

③ 口頭講演

〈国内〉

1. 森宙史、微生物群集の系統組成推定・描画ツール VITCOMIC2 の開発と応用、NGS 現場の会第五回研究会、仙台国際センター展示棟、2017/5/23
2. 山田拓司、大腸がん発病に伴うヒト腸内細菌のマルチオミクス解析、第 17 回日本抗加齢医学会総会、東京国際フォーラム、2017/6/2
3. 森宙史、東光一、黒川顕、微生物群集ビッグデータの統合と機械学習を利用した群集予測・診断システム、BIOTech2017 アカデミックフォーラム、東京ビッグサイト、2017/6/30
4. 黒川顕、データサイエンスを加速させる微生物データの統合化、環境微生物系学会合同大会 2017、東北大学川内キャンパス、2017/8/31

5. 山田拓司、ヒト腸内細菌データベースとその利用、第 69 回日本生物工学会大会。早稲田大学西早稲田キャンパス、2017/9/11
6. 山田拓司、大腸がん発病に伴うヒト腸内細菌のマルチオミクス解析、第 42 回日本医用マスペクトル学会、一橋講堂、2017/9/14
7. 山田拓司、Human gut microbiome associated with colorectal cancer、第 76 回日本癌学会学術総会、パシフィコ横浜、2017/9/28
8. 中村保一、MicrobeDB.jp、日本分子生物学会年会 NBDC ConBio フォーラム「生命科学のデータベース活用法」、神戸、2017/12/09

〈国際〉

1. Ikuo Uchiyama、A new protocol for constructing the ortholog table in Microbial Genome Database for Comparative Analysis、NIG International Symposium 2017 Commemorating the 30th Anniversary of DDBJ、Mishima Citizens Cultural Hall、2017/05/27-29
2. Ikuo Uchiyama、A new protocol for constructing ortholog database for large-scale microbial genome comparison、Quest for Orthologs 5、University of Southern California、Los Angeles、2017/06/08-10
3. Hirokazu Chiba、Jesusaldo Tomás Fernández-Breis、Ramón García Martínez、Yuki Moriya、Susumu Goto、Ikuo Uchiyama、Development of orthology ontology and its application to orthology meta-search、Quest for Orthologs 5、University of Southern California、Los Angeles、2017/06/08-10
4. Hiroshi Mori、MicrobeDB.jp: integrating microbial genome and metagenome data using semantic web technology、BioHackathon 2017 symposium、JST Science Plaza、Japan、2017/09/10
5. Hiroshi Mori、Takayuki Maruyama、Masahiro Yano、Takuji Yamada、Ken Kurokawa、VITCOMIC2: Visualization tool for the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing、GIW 2017、Hotel Samjung、Seoul、2017/11/01

④ ポスター発表

〈国内〉

1. 内山郁夫、西出浩世、比較ゲノム解析ツール RECOG を用いた NGS データの比較解析、第 5 回 NGS 現場の会、仙台国際センター、2017/05/22-24
2. 森宙史、藤澤貴智、鈴木真也、千葉啓和、東光一、神沼英里、西出浩世、矢口貴志、高橋弘喜、山田拓司、内山郁夫、中村保一、黒川颯、微生物統合データベース MicrobeDB.jp の高度実用化開発、トーゴの日シンポジウム 2017、2017/10/04-05
3. 藤澤貴智、MicrobeDB.jp ポータル: 統合微生物データベースのポータルサイト構築、第 12 回日本ゲノム微生物学会年会、京都大学桂キャンパス、2018/03/05-07
4. 森宙史、黒川颯、微生物群集の系統組成推定ツール VITCOMIC2 の MicrobeDB.jp への応用、第 12 回日本ゲノム微生物学会年会、京都大学桂キャンパス、2018/03/05-07

5. 内山郁夫、三原基広、西出浩世、千葉啓和、公開ゲノムの全遺伝子レパートリーをカバーする微生物オーソログテーブルの構築、第12回日本ゲノム微生物学会年会、京都大学桂キャンパス、2018/03/05-07

〈国際〉

1. Hiroshi Mori, Takayuki Maruyama, Masahiro Yano, Takuji Yamada, Ken Kurokawa, VITCOMIC2: visualization tool for the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing data, NIG International Symposium 2017 Commemorating the 30th Anniversary of DDBJ, Mishima Citizens Cultural Hall, 2017/05/27-29
2. Hirokazu Chiba, Ikuo Uchiyama, Applying the semantic web technology to diversified biological resources for comparative omics, NIG International Symposium 2017 Commemorating the 30th Anniversary of DDBJ, Mishima Citizens Cultural Hall, 2017/05/27-29
3. Hiroshi Mori, Takayuki Maruyama, Masahiro Yano, Takuji Yamada, Ken Kurokawa, VITCOMIC2: visualization of the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing, ProkaGENOMICS 2017, Georg-August-University Gottingen, Germany, 2017/09/19-22

(4) 知的財産権の出願

種別	件数	
特許出願	国内	0件
	国外	1件
その他の知的財産出願	0件	

(5) 受賞・報道等

該当無し

§5. 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
2017年 5月12日	チーム内ミーティング(非公開)	遺伝研	4人	産業応用のためのミーティング
2017年 6月26日	チーム内ミーティング(非公開)	遺伝研	8人	進捗報告のためのミーティング
2018年 1月19日	チーム内ミーティング(非公開)	遺伝研	9人	進捗報告のためのミーティング
2018年 2月26日	チーム内ミーティング(非公開)	遺伝研	4人	講習会準備のためのミーティング

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
2018年 3月7日	MicrobeDB.jp 講習会 (第12回 日本ゲノム微生物学会年会内)	京都大学桂 キャンパス	20人	MicrobeDB.jp のポータルサイトの紹介と使い方の講習会、およびユーザ会としてのDBへのニーズの収集
2018年 3月29日	MicrobeDB.jp 講習会 (第91回 日本細菌学会総会内)	福岡国際会 議場	50人	MicrobeDB.jp のポータルサイトの紹介と使い方の講習会、およびユーザ会としてのDBへのニーズの収集

以上

別紙 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。	http://microbedb.jp	2011/12/12	維持・発展	データベース等	
2	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。	http://mbgd.genome.ad.jp	1997/7/1	維持・発展	データベース等	