

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

「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」

黒川 顕 | 国立遺伝学研究所 情報研究系 教授

 ©2019 黒川 顕(情報・システム研究機構) Licensed under CC BY 4.0

§1. 研究開発実施の概要

【目的】

我々はこれまで、微生物データの統合化、超高度化を推進し、フル RDF による微生物統合データベース(以降 DB) MicrobeDB.jp を構築してきた。統合 DB においては、全データが連結された巨大なグラフとなっており、利用者は巨大グラフの全貌が不明であるため、「巨大グラフをどのように辿ればどのような答えが出てくるのか」を想定できず、これまでの DB の利用形態を適用する事が本質的に困難となる。これは統合 DB の利用そのものが、新しい科学的手法である事を意味しており、この新しい科学的手法こそが「データサイエンス」と言えよう。そこで本研究開発では、徹底的なデータ統合および高度化は継続しつつ、「統合化されたデータをどのように渡り歩き、どのような新規知見を得るか」という統合 DB の実用化に向けた、データサイエンスを加速させる統合 DB の利活用方法の開発に重点を置き、MicrobeDB.jp 活用による新たな科学的手法、すなわちデータサイエンス研究手法を提案するとともに、MicrobeDB.jp の実用化を目指す。

【方法】

本研究開発では、「統合化されたデータをどのように渡り歩き、どのような新規知見を得るか」、という統合 DB の実用化に向けた、データサイエンスを加速する統合 DB の利活用方法の開発に重点を置き、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ている。

【結果】

- ①徹底したユーザビリティの向上: キーワード検索に加えファセット検索可能な UI の開発、RDF モデル更新にともなう全データの更新。
- ②データ品質の向上: 微生物環境オントロジー MEO の更新、MEO 自動アノテーションの精度向上、ヒトマイクロバイオームのメタデータの徹底的な RDF 化。
- ③キラーアプリケーションの開発: 系統組成推定ツール VITCOMIC2 ローカル環境版、系統組成・遺伝子機能組成解析パイプライン MeGAP の Docker コンテナ版、微生物群集統合可視化ツール LEA、の開発。
- ④さらなるデータの統合: 真菌類ゲノム、ITS、RNA-Seq データの統合。
- ⑤基盤データ解析技術の高度化: メタゲノム解析パイプラインの高速化。オーソログ解析パイプラインの効率化。
- ⑥効率的運用: RDF の再設計、MEO 自動アノテーション技術開発。
- ⑦ホログenom対応: MicrobeDB.jp のサブセットである RhizoBase を介して、かずさ DNA 研究所 Plant Garden との密な連携を開始。

§2. 研究開発実施体制

1. 各グループの担当項目

(1) 「遺伝研」グループ(研究代表者グループ)

本研究開発グループでは、MicrobeDB.jp について、①徹底したユーザビリティの向上、②データクオリティの向上、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ホロゲノム対応のすべての研究開発項目において、中心となって研究開発を行う。

(2) 「基生研」グループ(主たる共同研究者グループ(1))

本研究開発グループでは、②データクオリティの向上、③キラーアプリケーションの開発、⑤基盤データ解析技術の高度化について、これまでに構築してきた微生物ゲノム間のオーソログ解析パイプラインをさらに効率化し、今後さらに大規模なゲノムデータが蓄積しても安定に更新が行えるような体制を構築する。また、RDF 化したオーソログデータと他のデータとを統合して、高度な比較ゲノム解析を行えるようなアプリケーションを開発する。必要に応じてオーソロジーデータを記述するためのオントロジーの拡張も行う。

(3) 「東工大」グループ(主たる共同研究者グループ(2))

本研究開発グループでは、②データクオリティの向上および⑦ホロゲノム対応について、医療や健康食品業界等の応用が見込まれる、ヒトマイクロバイオームデータについて、サンプルに付随する宿主側情報をメタデータとして整備し、MicrobeDB.jp 上で豊富かつ質の高いヒトマイクロバイオームデータを利用できるように研究開発を実施する。

(4) 「千葉大」グループ(主たる共同研究者グループ(3))

本研究開発グループでは、④さらなるデータの統合について、真菌類のゲノム情報と生息環境や生理活性等のメタデータの収集および分類情報の整理を核として、MicrobeDB.jp の真菌に関するデータの基盤整備を実施する。

2. 有識者会議等

該当なし

§3. 研究開発の目的、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

主要なもの

正式名称	略称	概要
MicrobeDB.jp		http://microbedb.jp

上記以外のもの

正式名称	略称	概要
Microbial Genome Database for Comparative Analysis	MBGD	http://mbgd.genome.ad.jp

(2) ツール等

正式名称	略称	概要
Metagenome and Microbes Environmental Ontology	MEO	http://bioportal.bioontology.org/ontologies/MEO
Microbial Culture Collection Vocabulary	MCCV	http://bioportal.bioontology.org/ontologies/MCCV
Pathogenic Disease Ontology with Symptom	PDO-CSSO	http://bioportal.bioontology.org/ontologies/PDO , http://bioportal.bioontology.org/ontologies/CSSO
Metagenome Sample Vocabulary	MSV	http://bioportal.bioontology.org/ontologies/MSV
Human Microbiome Associated Disease Ontology	HMADO	発症にヒトマイクロバイオームの関連が疑われるヒトの病気のオントロジー

2. 達成目標及び実施計画

(1) 当初の実施計画・達成目標 (開発提案書から抜粋)

1. 遺伝研グループ

研究開発実施項目(1): MicrobeDB.jp ポータルサイト構築

MicrobeDB.jp のユーザーインターフェース拡張によるユーザビリティの向上を図るため、これまでのシンプルなキーワード検索に加え、MicrobeDB.jp のフル RDF DB の利点を活用した、新しいポータルサイトを構築する。また、藻類、菌類等のデータセットや MBGD、MeGAP による解析データセットなどを中心にユースケースを積み上げて提示する。さらに、GenomeRefine を認証付きの API サービスとして改変し、新しいポータルサイトから呼び出すことで、ユーザのデータ登録環境もポータルサイト上に配置する。このポータルサイト構築にあたっては既存のユーザおよびユーザ会、また講習会におけるユーザからのフィードバックを受け、継続的に改善を図って行く。登録されたデータについては、公共 DB サービスとの間で、ユーザ認証および登録インターフェースを含めユーザ利便性向上のための連携を図る。

研究開発実施項目(2): MicrobeDB.jp ユーザデータアーカイブサービス拡張

GenomeRefine で開発した機能を、ユーザ・グループ認証サーバーを介して認証する機能を付与した API 化を実施する。メタデータと API 入力パラメータにセマンティクスを紐づけることによって、RDF 変換の汎用化を行う。本 API は構築予定のポータルサイトから呼び出すことにより、ユーザは MicrobeDB.jp 上で直接データ登録が可能になり、比較解析 Stanza との接続によるデータ解析機能への誘導がスムーズになる。また、この API を介することで、当グループが提供している微生物ゲノム解析パイプライン DFAST(<https://dfast.nig.ac.jp>)などの解析支援アプリケーションとの広範な連携を拡張していく。さらに、プライベートな API 等を開発し、協力体制を確立しながら NBDC ヒト DB のヒト資源由来のメタデータでの運用も目指していく。

研究開発実施項目(3):メタゲノムや RNA-Seq 等のマルチオミックスデータを用いた Stanza およびアプリケーションの開発

マルチオミックスデータを中心に、対応トピックモデル等の高度な統計解析手法やインタラクティブな視覚化手法を用いた様々な Stanza およびアプリケーションを開発し、ユースケースに応じた DB データの様々な解析結果を利用者に提供できるようにする。

研究開発実施項目(4):オミックスデータ RDF 拡張とアノテーション統合

配列 ID ならびに配列アライメント・マッピング情報を集積し、配列位置情報オントロジー FALDO に基づく変換系を構築することで、ゲノムアノテーションの統合と再利用をさらに広い範囲で可能とするゲノム RDF の変換系を実現する。また、RDF を介して統合されたアノテーションを新たなゲノムリファレンスとする GFF 形式等の、ゲノム解析の現場で汎用的に用いられているファイル形式に変換し、ユーザの利便性を図るとともに、RDF を起点とした新しい統合データの枠組みを提供することを目指す。

研究開発実施項目(5):メタゲノム解析パイプラインの速度・精度向上と真核メタ ITS 解析への対応

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。今後増加が予想される真核メタ ITS 解析データについて

では、高速かつ高精度な系統組成推定のための解析パイプラインを新たに開発し、第 3 年次末までに、公共の真核メタ ITS 解析データを網羅的に収集、再解析し、真核微生物についても環境中の分布を MicrobeDB.jp 上で検索、比較解析できるようにする。上記の開発した解析パイプラインを、ユーザのデータの解析にも利用できるようにするために、MeGAP を改良する。

研究開発実施項目(6):オントロジー自動アノテーションツールの精度向上

文献情報と紐づけられたサンプルについては、文献の要旨や本文等からテキストマイニングを行い、サンプルのオントロジーアノテーションに必要な情報を抽出した上で、自動アノテーションを行う機能をツールに付与することで、オントロジー自動アノテーションツールの精度向上を達成する。さらに、開発したオントロジー自動アノテーションツールを、MicrobeDB.jp における MEO 以外のオントロジーのアノテーションにも適用できるようにツールを改良する。

研究開発実施項目(7):藻類および植物ホログenom研究等を支援するサンプルメタデータおよび生物分類の再資源化

公共 DB に登録されている生物分類に関するメタデータについて、RDF 統計データを含む、データの RDF 化を実施する。さらに、RDF ガイドラインに従った既存の主要なオントロジーのマッピングによるリソース化を実施する。これらのリソースはすでに強力な連携を有する国立環境研究所や藍藻ゲノム DB CyanoBase, 根粒菌ゲノム DB RhizoBase のそれぞれの国内外のユーザコミュニティとの連携を図りながら、必要に応じて生物分類情報を拡張し、藻類および植物ホログenomの研究を支援できる RDF を開発する。また、RDF を起点とする微生物および植物ゲノムリファレンス情報をホログenom解析に利用可能とする。

2. 基生研グループ

研究開発実施項目(1):オーソログデータの構築

オーソログデータは、これまでに開発した階層的なオーソログ解析プロトコルに従いつつ、効率よく差分更新を行えるように、データ構築プロセスを修正しつつ更新漏れが起きないような頑健プロトコルを開発しデータ更新を行う。新規のゲノムデータを追加する際に、既存のゲノムデータとの重複度や重要度、クオリティなどに応じて優先度をつけ、優先度の低いものを更新対象から外すことによってデータ量を減らしつつ、必要十分なデータを含むオーソログデータを構築する仕組みについて検討する。その際、解析に必要なデータを後から差分的に付け加えられるような仕組み作りも行い、長期にわたって最善のオーソログデータの維持ができるようにすることを目指す。

研究開発実施項目(2):比較ゲノム解析アプリケーションの開発

これまでに作成したオーソログ DB やデータ表示用 Stanza を活用しつつ、利用者がより簡便に比較解析を実行できるようなアプリケーションを開発する。特に、利用者が指定したゲノム配列に対してオーソログデータを検索してアノテーションをつける「オーソログ検索・アノテーション」、指定した生物種群に存在し、それ以外の生物種に存在しないような遺伝子を探索する「系統プロファイル解析」、および同種・近縁種ゲノム間での遺伝子の保存性や多様性を解析する「コアゲノム・パンゲノム解析」などについて、利用者が柔軟に条件を指定して、その結果を様々な角度から検討できるようなインターフェースを作成する。また、

これらのアプリケーションを作成するに当たり、必要に応じて、これまでに開発したオーソログデータ用のオントロジー ORTH の拡張を行う。

3. 東工大グループ

研究開発実施項目:ヒトマイクロバイオーム関連のメタデータ整備

ヒトマイクロバイオームデータにおいて、公共 DB には登録されていない詳細な疾患関連情報に関するメタデータを論文本体から直接抽出し、MicrobeDB.jp の配列データと紐付けた RDF データを作成する。MicrobeDB.jp では、今までヒトの病気については感染症-症状オントロジーである PDO-CSSO に存在する感染症のみを扱ってきたが、本研究開発ではテキストマイニングの手法を用い、これまで感染症の病原菌としては報告されていないヒト常在菌がどのような疾患と関連するかを論文から抽出する。これらの結果をマニュアルキュレーションにより精査し、既存の標準的に用いられているヒトの疾患オントロジーを用いてアノテーションを行い、ヒトマイクロバイオームデータを解釈する上で重要になる、微生物-疾患関連性データの充実を行う。

4. 千葉大グループ

研究開発実施項目:真菌類ゲノム・菌株・オミックス情報の収集と高度化

真菌類のゲノム情報と生息環境や生理活性等のメタデータの収集および分類情報の整理を核として、MicrobeDB.jp の真菌に関するデータの基盤整備を実施する。具体的には、真菌類の完全長・ドラフトゲノムデータ、RNA-Seq データを公共塩基配列 DB から収集する。さらに、菌種同定に用いられる各種系統マーカー遺伝子 (ITS 配列など) の配列データの整備を行うことで、遺伝研グループで行う公共の配列 DB 由来のメタ ITS 解析データの菌種同定に用いることが可能なリファレンス配列データを作成する。菌株については、発酵や臨床面で最も重要なアスペルギルス属を中心に生理活性や培養条件等のメタデータ収集を行い、セマンティックウェブ技術を用いたデータの記述と他のデータとの統合化を行う。

第3年次末までの達成目標 (開発提案書に記載)

- MicrobeDB.jp トップページとして新たに開発するポータルサイトに、統合 DB 中のデータを駆使したユースケースを示すため、Stanza およびアプリケーションを 10 種以上設置する。
- 真核メタ ITS データに対する解析パイプラインを開発し、MicrobeDB.jp にて公共の真核メタ ITS データの検索および比較解析を可能にする。
- MicrobeDB.jp で行っている、MEO 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 60%低減する。
- オーソログデータを段階的に構築するプロトコルについて、差分更新によって効率的に更新が行える体制を構築する。これを用いて、公表されたゲノムデータ全体を取り込んだオーソログ DB の構築を行う。
- 更新対象とするゲノムに優先度を付けて、優先度の低い生物種を更新対象から外すことによって、データサイズを減らしつつ、必要十分なゲノムデータを取り込んだオーソログ DB を構築するための汎用的な仕組みを開発する。
- オーソログデータに基づいて比較ゲノム解析を行うアプリケーションとして「オーソログ検索に

よるアノテーション」、「系統プロファイル解析」、「コアゲノム・パンゲノム解析」などを実装する。

- これら開発を通して、MicrobeDB.jp の年間訪問者数を 13,000 人以上にする。

(2) 期間中に追加・削除・変更した実施計画・達成目標

【第 1(2017)年次追加支援 1】

高度実用化のユースケースの一つとして、ホログゲノム研究の基盤の提供およびそのためのホログゲノム研究支援ツール開発を実施する。将来的にヒトホログゲノム研究 DB 化や解析支援系にも活用可能な枠組みとして、まず、植物ホログゲノム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・川口正代司教授を生産者とするホログゲノムデータ連携の共同研究要請により、かずさ DNA 研究所で実施されている植物統合化推進プログラムと協力した開発として、①共生微生物ゲノム情報統合環境の構築と、②ホログゲノム解析支援ツールの開発を実施する。

【第 1(2017)年次追加支援 2】

ゲノム比較の結果をゲノムの代謝機能ポテンシャルの観点から評価するため、海洋研究開発機構の高見らによって開発された代謝ポテンシャル評価ツール MAPLE と連携させた解析機能を作成する。この目的のため、MBGD と MAPLE を連携させて、本システムのオーソログアノテーション結果から各ゲノムの KEGG Module の充足率を計算できるようにするとともに、その結果に基づいてゲノム間で代謝ポテンシャルの違いを評価するためのアプリケーションを開発する。このアプリケーションによって、未培養細菌から再構築したゲノムデータの機能上の特徴を、比較解析を通じてあぶりだすことを実現させて、本 DB の利用の呼び水となるような成果を導くことを目指す。

3. 実施内容

(1) 実施内容

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

MicrobeDB.jp のユーザーインターフェース拡張によるユーザビリティの向上を図るため、新しいポータルサイトを構築し、2018 年 3 月にポータルサイトをベータ版として公開した。ユーザビリティの見直しを行った結果、これまでのシンプルなキーワード検索に加え、フル RDF DB の利点を活用し、MEO などのオントロジーに基づいたファセット検索機能から可視化単位の TogoStanza を基盤とした比較解析結果の提示までのクリック数を可能な限り軽減し、UI を再構築することでシンプルな操作で実現した。また、ユーザデータの登録・更新機能については、計画当初 API によるウェブサービス間連携を計画していたが、定期的な DB 更新の実現にあたり、特にサンプルメタデータについて公開データおよびユーザデータの RDF データモデルおよび変換機能、自動アノテーション、データストア投入について共通基盤を再構築し、これまで GenomeRefine として独立していた機能をポータルサイトアプリケーションに統合することで、Web サイトを移動することなくそれらの機能を実現可能にした。ゲノム・メタゲノム対象データの拡張および全ての RDF を更新したデータセットを格納した MicrobeDB.jp version 3 を、2019 年 11 月にリリースする。本

リリースに合わせて、各種ドキュメントを整備し、ユーザ会、講習会、DBCLS の協力による TogoTV などポータルサイトの利用を促進するためのコンテンツ等を含む MicrobeDB.jp が提供するユースケースを提示した。本リリースには、ユーザ会等のフィードバックを精査、反映する形で、ヒトマイクロバイオームデータについて、宿主のヒトの病気、民族、抗生物質投与の有無、probiotics 摂取の有無を整理し、それらのメタデータによる絞り込み検索を実現した。サービス間がシームレスにつながる可能性を有するユーザ認証機能については、DDBJ へのデータ登録・公開の接続を目的にユーザ認証連携に着手した。

② データ品質の向上

最新のメタ 16S・メタゲノム・ゲノム・菌株のメタデータへの MEO 自動アノテーション結果を元に、cloud や floral nectar 等 MEO の語彙を追加し、より詳細に微生物の環境情報を記述可能にした。公開データおよびユーザによって入力されるサンプルのメタデータの品質向上を目的として、バリデーションを実施するために、DDBJ validation API の JSON 入力拡張開発を実施した。ヒトマイクロバイオームデータについては、ヒトのメタゲノム配列データが紐づいている論文約 4,500 報を文献情報 DB から収集し、約 1,200 報についてマニュアルキュレーションによってメタゲノムサンプルのメタデータ抽出を行った。それらのうち、DB に登録されているメタデータを相補する形で、実際に論文中に比較解析に有用な各サンプルのメタデータが記述されていた論文は 69 報であり、メタデータの項目数は表記揺れ等を除くと 176 項目であった。それらのメタデータを BioSample RDF と同様の構造で RDF 化した。

新しいオーソログデータ構築プロトコルを用いて構築したデータを、それに伴い必要となる修正作業を完了させた上で、MBGD の 2018 年バージョンとして公開した。登録ゲノム数は 6,318 で、2,547 種 1,019 属を含んでいる。新方式により作成したオーソログデータは階層的な構造を持っているが、その RDF 化に当たっては、オーソロジーデータ表現に用いている Orthology Ontology が階層的オーソログデータに対応しているので、それを用いてそのまま行うことができた。また、新バージョン公開にあたり、トップページのデザインを単純化し、遺伝子機能や生物種名での検索を容易に行えるような構成にした。これに伴い、Elasticsearch や jQueryUI ライブラリを使ったキーワード検索機能の強化を行った。

③ キラーアプリケーションの開発

大量のメタ 16S・メタゲノム配列データを持つユーザが、MicrobeDB.jp と同じ解析パイプラインを自らの解析環境で実行可能にするために、系統組成推定ツール VITCOMIC2 ローカル環境版、および系統組成・遺伝子機能組成解析パイプライン MeGAP の Docker コンテナ版を開発した。MicrobeDB.jp 中のメタゲノムサンプルの遺伝子機能組成データを用いて、微生物群集の全体像を俯瞰可能な統合可視化ツール LEA の、遺伝子機能組成版のアルゴリズムを開発した。LEA 遺伝子機能組成版については、ヒト以外のメタゲノムデータが増えた MicrobeDB.jp version 3 のデータを用いて学習を行い、系統組成版のアップデートとあわせて公開予定である。データ点が増えた状態でも、微生物群集構造と環境との関係性を直感的に理解できるようにするために、LEA を 3D に拡張して公開した。

ヒトマイクロバイオームと病気との関連性を整理し、病気横断的な検索や比較解析を可

能とするために、発症に微生物の関与が疑われる病気について、約 300 の Class で整理した Human Microbiome Associated Disease Ontology (HMADO) version 0.1 を構築し、HMADO でメタ 16S・メタゲノム・ゲノムのメタデータを自動アノテーションし、オントロジーを用いた統一的な検索および系統組成や遺伝子機能組成の比較解析を可能にした。

オーソログ検索・ゲノムアノテーションアプリケーションは、利用者のゲノム配列中の各遺伝子に対して、MBGD のオーソロググループの検索を行って対応付けする機能と、オーソロググループから KEGG Orthology へのクロスリファレンスを用いて、KEGG モジュール単位で機能を評価する MAPLE 解析とからなる。配列検索機能としては、HMMER を用いた HMM 検索に加えて、MMseqs2 を用いたプロフィール検索も実装し、特に後者はゲノム単位の解析においても十分な速度を持つことを確認した。この検索機能は、MBGD 2018 の機能としても公開した。この検索機能と、機能モジュール解析ツール MAPLE とを連動させて、ゲノムアノテーションツールの基礎的部分を作成した。このツールは、MBGD におけるユーザーゲノム解析機能である MyMBGD 機能の一環として動作する一方で、MicrobeDB.jp のアプリケーションとしても動作するように、MicrobeDB.jp の認証システムとも連携するように設計している。今後、さらに MBGD と連携させた比較解析機能の強化を行った上で、公開に向けての準備を進める。一方、系統プロフィール検索アプリケーションについては、利用者が指定した系統プロフィールに対して、類似性の高い系統プロフィールを持つオーソロググループを検索するアプリケーションのプロトタイプ作成を行った。特にメタゲノム解析との連携に向けて、メタゲノム解析から得られた系統プロフィールを入力として検索することを想定し、MicrobeDB.jp に登録されたデータと連携させることを試みた。こちらも今後使いやすいインターフェイスを作成して、公開に向けた準備を進める。

④ さらなるデータの統合

真菌類の系統分類に用いる参照 ITS 配列の整備を行うために、千葉大真菌センターが保有する真菌 1,358 株の ITS 配列を DDBJ に登録した。また、真菌類 10 株のドラフトゲノム配列データに遺伝子アノテーション情報の付与を完了した。真菌類のゲノム配列の充実については、東工大グループでも別プロジェクトで行なっており、それらをあわせると、100 株以上の真菌ゲノム配列データが近日中に DDBJ から公開予定である。真菌類の RNA-Seq データについては、DRA に登録されている約 4.6 万サンプルのうち、約 1.3 万サンプルについて、遺伝子発現量の算出を完了した。真菌メタ ITS 解析パイプラインについては、参照配列データとしてよく利用されている UNITE を用いて VITCOMIC2 と同様の解析手法で系統組成を推定する解析パイプラインのテスト版を開発した。今後、千葉大の真菌 ITS データとあわせて参照配列 DB を構築し、真菌メタ ITS 解析パイプラインの性能向上を行う予定である。

⑤ 基盤データ解析技術の高度化

メタ 16S・メタゲノムデータの解析パイプラインで用いる系統アサインメント用の参照 16S rRNA 遺伝子配列 DB を、Ribosomal Database Project の分類体系および配列データを基に構築し、VITCOMIC2 および MeGAP から利用できるようにした。さらに、MicrobeDB.jp で系統分類の基盤としている NCBI Taxonomy の分類体系と整合性がとれるように、Ribosomal Database Project と NCBI Taxonomy 間の分類体系のマッピング RDF を構

築した。

爆発的な勢いで増加するメタ 16S・メタゲノムデータの解析を高速化するために、解析パイプライン MeGAP の各解析に用いるプログラムを見直した。具体的には、配列のクオリティフィルタリングを行うツールを Python で書かれた cutadapt から C++ で書かれた fastp に変更し、アミノ酸の配列相同性検索のためのツールを GHOSTX からより高速な MMseqs2 に変更等の改良により、前 version (MeGAP2) と同等の解析精度を保った上で、10 倍近い高速化を実現した。

⑥ 効率的運用

ゲノム、メタゲノムのサンプル情報 RDF に関して、公開 BioSample データおよびユーザー登録サンプルメタデータの RDF 再設計し、データモデルの共通化を実施した。MicrobeDB.jp データ更新のための情報基盤に関して、一次リソースから RDF 変換および DB 投入、検索系のデータフローを整理し、メジャー更新および差分更新系を実施する API の開発に着手した。ユーザのゲノムデータと公開ゲノムデータとの比較解析を実現しユーザ利便性を向上するため、自動ゲノムアノテーションパイプライン DFAST に MBGD ID のアサイン機能を実装した。また、MEO の自動アノテーションのアルゴリズムを見直し、教師データの追加および改良を行い、MEO 自動アノテーションの精度向上を行なった。MEO 自動アノテーションの精度向上によって、自動アノテーション後のマニュアルでの修正に必要な時間と手間が激減し、MEO アノテーションに必要な時間を MicrobeDB.jp version 2 の 1/10 以下に減らすことができた。MEO 自動アノテーション機能については、DBCLS や理研 JCM、NITE BRC など他のグループからも使用したい旨要望があったため、MEO 自動アノテーション API としてテスト版を公開した。MEO 自動アノテーションの精度向上については、次年度も引き続き取り組む予定である。

⑦ ホロゲノム対応

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ゲノム解析支援系の構築を目標に、ホロゲノム研究の基盤として 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連携を開始した。DB 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」を実施した。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 DB 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した。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「ホロゲノム解析支援ツール開発」を実施した。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ザ *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した。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した。

ホロゲノム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中である。

Lotus japonicus および *Mesorhizobium loti* MAFF303099 の宿主-共生微生物の対象を拡充し、植物宿主および RhizoBase の根粒菌 5 株について RDF 化を実施し、ホロゲノム解析支援ツールにおいて利用するメタデータを整備した。また、生物分類情報アノテーションを記載するためのサンプル RDF および解析データ RDF データモデルを設計した。2019 年度において、計画を前倒し、かずさ DNA 研究所 Plant Garden との密な連携の実施を開始した。ゲノム情報およびマーカー情報についてミヤコグサに続いて、SIP における農業領域との連携を視野に対象生物にイチゴも含める拡張を実施した。さらに ACCEL 解析データ (VITCOMIC2、LEA) およびそれらのメタデータの解析結果の提示とさらなる解析支援を目的とした開発に着手するとともに植物および微生物ゲノム情報の統合化における課題を整理した。

- (2) 「対象とするデータベース条件」のうち採択時に未達成であった項目の対応状況
該当なし。

(3) 統合化推進プログラムの他のチームや DBCLS との連携

DBCLS とは、毎月行われている SPARQLthon 等で現状の課題の共有を頻繁に行なっており、その結果として、本グループで開発した MEO 自動アノテーション API は、DBCLS でも菌株の分離環境のアノテーション等で用いられている。また、植物統合グループとは、ホロゲノムデータの統合化においてゲノムの RDF 化やマイクロバイオーームデータ等について連携して開発を進めている。

§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	2016年度	2017年度	2018年度	2019年度 (9月末時点)
ユニーク IP 数	837	6,764	5,927	4,683
ページ数	20,908	28,203	29,444	33,932

(2) 分析

2016 年末に MicrobeDB.jp version 2 を公開したことで、2016 年度と 2017 年度以降ではユニーク IP 数、ページ数ともに大きな差があり、特にユニーク IP 数は 4 倍以上と飛躍的に増加している。ページ数については、version 2 へのアップデートの際に、できるだけページ遷移を減らすよう工夫したため、2016 年度と 2017 年度以降ではユニーク IP 数ほど大幅な増加はしていない。version 2 へ更新した後約 3 年が経過しているため、ユニーク IP 数が減少してきているが、大規模に情報更新した version 3 へのアップデートによりこの減少傾向は回復すると期待できる。

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

本グループで開発した MEO 自動アノテーション API は、DBCLS および NITE でも菌株の分離環境のアノテーション等で用いられている。企業とは MicrobeDB.jp を核とした共同研究を進めている(計 7 社)。また、受託分析会社に VITCOMIC2 を導入し、顧客に対して MicrobeDB.jp からの知識提供サービスを展開しようとしている。さらに、食品会社とは、マイクロバイオーーム情報を徹底的に活用する事を目的とした共同研究を開始しており、MicrobeDB.jp システム全体を企業内サーバに導入する予定である(2019 年度)。

革新的先端研究開発支援事業(AMED-CREST:2017 年度～)および戦略的イノベーションプログラム(SIP:2018 年度～)では、遺伝研が領域から得られるマイクロバイオーームデータを集約する研究開発拠点を担っており、そのデータ共有コアとして MicrobeDB.jp が活用されている。

3. その他

特になし。

§5. 今後の展開

本研究開発では、ゲノム情報を核として様々な微生物学上の知識を統合し、幅広い分野での微生物学の発展に資することのできる統合 DB「MicrobeDB.jp」を、持続可能なシステムに昇華させるとともに、ユーザビリティの向上を徹底する事で、単なる統計量の検索や羅列ではなく、大規模データから新規知識発見を容易に引き出す事が可能なシステムを構築する事を目的としている。微生物データの統合化、さらには統合 DB に対する解析プロトコルの実装は、微生物の体系的な理解を促進し、これまでの仮説検証型の研究のみならず、膨大なデータの中から新たな仮説を導くデータ駆動型の研究を強力に推進する事を可能とする。新型シーケンサー等の飛躍的なハイスループット化は今後ますます加速され、ゲノム研究やメタゲノム研究に代表されるデータ駆動型研究が本格化すると考えられる。さらに、USB メモリサイズのシーケンサーが一般化し、多様な産業分野における微生物ゲノム情報の応用も加速度的に推進されようとしている。このような状況下では、これまでブラックボックス化していた微生物の生命システムを、統合化 DB を利用したデータ駆動型研究により明らかにすることは、工業、農業、畜産業や医薬、健康など、微生物が関与する広範な産業にとって最重要研究課題のひとつとなる。この潮流はすでに始まっており、我々は MicrobeDB.jp を武器として、工業(廃水处理)、農業(農作物品質管理、高機能作物)、畜産業(養豚)、医薬(創薬、疾患マーカー探索)、健康(食品開発)、新規産業(ヒト環境)に関する企業との共同研究を進めている。

微生物関連情報は極めて多様であるため、微生物統合 DB の開発は与えられた研究期間内で完結するものではない。しかし、我々が開発している、基盤的な技術やアプリケーション群、コンセプトは汎用性の高いものであり、さらなる多様かつ大量のデータの統合化に向けて、将来にわたって発展させていくことが可能である。微生物にとって宿主となる動物や植物に関するホログノミックなデータや、プロテオームやメタボロームのデータについても今後情報爆発を起こす可能性があるが、MicrobeDB.jp ではそれらデータも統合化し発展していけるよう、すでに準備を開始している。

このような観点から将来を展望すれば、微生物データのみには固執するのではなく、植物や動物、タンパクやメタボロームなど他の基盤 DB とのさらなる統合化は、科学や産業の分け隔てなく社会の発展にとって極めて重要である事は容易に理解できる。今後発展が期待されているゲノム合成分野においても欧米や中国に遅れをとっており追いつくのは容易ではないが、基盤となる高品質な統合 DB は基礎から応用に至るまで必須であり、常に先を見据えつつ腰を落ち着けて開発を継続する事が重要と考えている。

§6. 自己評価

【目標の達成状況】

第3年次末までの目標はすべて達成できている、もしくは今年度中に達成予定である。

【実施内容に対する研究代表者自身の評価】

我々は、統合化推進プログラムの中では先陣を切ってセマンティックウェブ技術によるデータ統合化を推進し、DBCLSの強力なバックアップを得ながら、多様なRDFモデル、オントロジー、ボキャブラリ、Stanzaを開発して、フルRDFによるMicrobeDB.jpを構築してきた。一方で、DBCLSによる技術の高度化に追随する必要性が生じたため、RDFモデルを一新するなど、今期開発のDBシステムは基盤から再構築する必要があった。この大規模な再開発にも関わらず、第3年次までの目標をすべて達成し、当初の目標を大幅に上回るDBアクセス数を獲得したことは高く評価されるべきと考えている。また、当初の予想通り、国内においてもマイクロバイーム分野が急速に発展しており、先を見通した開発が実施できている事を大いに評価している。

【本プログラムの趣旨に対する、本研究開発課題への貢献の程度】

本DBが対象とするのは、主に微生物および微生物叢に関する研究分野である。本分野では、2016年3月にJST CRDS戦略プロポーザル「微生物叢(マイクロバイーム)研究の統合的推進～生命、健康・医療の新展開～」が発表後、文科省から2016年度研究開発目標として「宿主と微生物叢間のクロストーク・共生の解明と健康・医療への応用」が発表され、ヒトマイクロバイーム研究が強く推進されはじめた。これを受けて、2016年度以降AMEDでは革新的先端研究開発支援事業や次世代がん医療創生研究事業、難治性疾患実用化研究事業などで、ヒトマイクロバイーム関連研究が開始された。ヒトマイクロバイーム研究以外においては、2014年度からJST ACCEL「共生ネットワークの分子基盤とその応用展開」において、菌根菌と宿主植物との関係性を明らかにする研究が開始されるとともに、JST CRDS「研究開発の俯瞰報告書 研究開発の新しい動向(2016年)」にて、土壌・環境マイクロバイームの評価・制御基盤技術が、今後大きく注目すべきテーマとして挙げられ、宿主とマイクロバイームの相互作用、動態、共進化などをゲノムレベルで解析するホログenom研究も重点化されようとしている。さらに、2018年度には、戦略的イノベーションプログラムSIP第2期が開始され、「スマートバイオ産業・農業基盤技術」課題においては、産官学連携でのオープンイノベーション型の取り組みによる、農業からヒト健康に至るまでの情報統合を目指している。これらいずれのプロジェクトにおいてもMicrobeDB.jpが少なからず活用されており、統合化推進プログラムを構成する研究グループとしてその役割を果たしていると考えている。

§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)	1 件
	国際(欧文)	19 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. Yohei Shimura, Yuu Hirose, Naomi Misawa, Sachiko Wakazuki, Takatomo Fujisawa, Yasukazu Nakamura, Yu Kanesaki, Haruyo Yamaguchi, Masanobu Kawach, "Complete Genome Sequence of a Coastal Cyanobacterium, *Synechococcus* sp. S train NIES-970", *Genome Announc.*, vol. 5 No. 14, e00139-17, 2017 (DOI: 10.1128/genomeA.00021-18).
2. Yasuhiro Tanizawa, Takatomo Fujisawa, Yasukazu Nakamura, "DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication", *Bioinformatics*, vol. 34, Issue 6, pp. 1037-1039, 2018 (DOI: 10.1093/bioinformatics/btx713).
3. Yu Kanesaki, 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).
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).
5. 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. Youssef Darzi, Ivica Letunic, Peer Bork, Takuji Yamada, "iPath3. 0: interactive pathways explorer v3", *Nucleic acids research*, 46, W1, W510-W513, 2018 (DOI: 10.1093/nar/gky299).
7. Hikaru Watanabe, Issei Nakamura, Sayaka Mizutani, Yumiko Kurokawa, Hiroshi Mori, Ken Kurokawa, Takuji Yamada, "Minor taxa in human skin microbiome contribute to the personal identification", *PLoS one*, 13, 7, e0199947, 2018 (DOI:10.1371/journal.pone.0199947).
8. Atsushi Ueda, Atsushi Kobayashi, Sayaka Tsuchida, Takuji Yamada, Koichi Murata, Hiroshi Nakamura, Kazunari Ushida, "Cecal microbiome analyses on wild Japanese rock ptarmigans (*Lagopus muta japonica*) reveals high level of coexisten

- ce of lactic acid bacteria and lactate-utilizing bacteria", *Microorganisms*, 6, 3, 7, 2018 (DOI:10.3390/microorganisms6030077).
9. Naofumi Yoshida, Takuo Emoto, Tomoya Yamashita, Hikaru Watanabe, Tomohiro Hayashi, Tokiko Tabata, Namiko Hoshi, Naoya Hatano, Genki Ozawa, Naoto Sasaki, Taiji Mizoguchi, Hilman Zulkifli Amin, Yushi Hirota, Wataru Ogawa, Takuji Yamada, Ken-ichi Hirata, "Bacteroides vulgatus and Bacteroides dorei reduce gut microbial lipopolysaccharide production and inhibit atherosclerosis", *Circulation*, 138, 22, 2486-2498, 2018 (DOI:10.1161/CIRCULATIONAHA.118.033714).
 10. Tomohiro Hayashi, Tomoya Yamashita, Hikaru Watanabe, Kenjiro Kami, Naofumi Yoshida, Tokiko Tabata, Takuo Emoto, Naoto Sasaki, Taiji Mizoguchi, Yasuhiro Irino, Ryuji Toh, Masakazu Shinohara, Yuko Okada, Wataru Ogawa, Takuji Yamada, Ken-ichi Hirata, "Gut microbiome and plasma microbiome-related metabolites in patients with decompensated and compensated heart failure", *Circulation Journal*, 83, 1, 182-192, 2018 (DOI:10.1253/circj.CJ-18-0468).
 11. Sachiko Naito, Noriko Takeuchi, Misako Ohkusu, Azusa Takahashi-Nakaguchi, Hiroki Takahashi, Naoko Imuta, Junichiro Nishi, Keigo Shibayama, Mayumi Matsuoka, Yuko Sasaki, Naruhiko Ishiwada, "Clinical and bacteriologic analysis of nontypeable *Haemophilus influenzae* strains isolated from children with invasive diseases in Japan from 2008 to 2015", *J Clin Microbiol*, 56, 7, e00141-18, 2018 (DOI:10.1128/JCM.00141-18).
 12. 高橋 (中口) 梓, 萩原大祐, 高橋弘喜, 酒井香奈江, 豊留孝仁, 渡辺哲, 亀井克彦, 五ノ井透, "次世代シーケンサーを活用したオミクス解析による *Aspergillus fumigatus* 病原因子探索", *Med Mycol J*, 59, 2, J35-J40, 2018 (DOI:10.3314/mmj.18.005).
 13. Kenzo Matsuo, Akira Haku, Beibei Bi, Hiroki Takahashi, Nobuhiko Kamada, Takashi Yaguchi, Shinobu Saijo, Mitsutoshi Yoneyama, Yoshiyuki Goto, "Fecal microbiota transplantation prevents *Candida albicans* from colonizing the gastrointestinal tract", *Microbiol Immunol*, 63(5):155-163, 2019 (DOI: 10.1111/1348-0421).
 14. Masaaki Shimizu, Yoko Kusuya, Yikelamu Alimu, Cai Bian, Hiroki Takahashi, Takashi Yaguchi, "Draft Genome Sequence of *Aspergillus awamori* IFM 58123NT", *Microbiol Resour Announc*, 8(4): e01453-18, 2019 (DOI: 10.1128/MRA.01453-18).
 15. Takahito Toyotome, Saho Hamada, Satoe Yamaguchi, Hiroki Takahashi, Daisuke Kondoh, Masahiko Takino, Yu Kanesaki, Katsuhiko Kamei, "Comparative genome analysis of *Aspergillus flavus* clinically isolated in Japan", *DNA Res*, 6, 1, 95-103, 2019 (DOI:10.1093/dnares/dsy041).
 16. Mohammad Vahed, Jun-ichi Ishihara, Hiroki Takahashi, "Dipartite: A tool for detecting bipartite motifs by considering base interdependencies", *PLoS One*, 14(8):e0220207, 2019 (DOI: 10.1371/journal.pone.0220207).
 17. Laure Nicolas Annick Ries, Jacob L. Steenwyk, Patrícia Alves de Castro, Pollyne Borborema Almeida de Lima, Fausto Almeida, Leandro José de Assis, Adriana Oliveira Manfiolli, Azusa Takahashi-Nakaguchi, Yoko Kusuya, Daisuke Hagiwara, Hiroki Takahashi, Xi Wang, Joshua J. Obar, Antonis Rokas, Gustavo H. Goldman, "Nutritional Heterogeneity Among *Aspergillus fumigatus* Strains Has Consequences for Virulence in a Strain- and Host-Dependent Manner", *Front Microbiol*, 1

- 0:854, 2019 (DOI: 10.3389/fmicb.2019.00854).
18. Toshiaki Katayama, Shuichi Kawashima, Shinobu Okamoto, Yuki Moriya, Hirokazu Chiba, Yuki Naito, Takatomo Fujisawa, Hiroshi Mori and Toshihisa Takagi, "TogoGenome/TogoStanza: modularized Semantic Web genome database", *Databases*, vol. 2019, 2019 (DOI: 10.1093/database/bay132).
 19. Ikuo Uchiyama, Motohiro Mihara, Hiroyo Nishide, Hirokazu Chiba, Masaki Kato. MGD 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).
 20. Toshiaki Katayama, Shuichi Kawashima, Gos Micklem, Shin Kawano, Jin-Dong Kim, Simon Kocbek, Shinobu Okamoto, Yue Wang, Hongyan Wu, Atsuko Yamaguchi, Yasunori Yamamoto, Erick Antezana, Kiyoko F. Aoki-Kinoshita, Kazuharu Arakawa, Masaki Banno, Joachim Baran, Jerven T. Bolleman, Raoul J.P. Bonnal, Hidemasa Bono, Jesualdo T. Fernández-Breis, Robert Buels, Matthew P. Campbell, Hirokazu Chiba, Peter J. A. Cock, Kevin B. Cohen, Michel Dumontier, Takatomo Fujisawa, Toyofumi Fujiwara, Leyla Garcia, Pascale Gaudet, Emi Hattori, Robert Hoehndorf, Kotone Itaya, Maori Ito, Daniel Jamieson, Simon Jupp, Nick Juty, Alex Kalderimis, Fumihiko Kato, Hideya Kawaji, Takeshi Kawashima, Akira R. Kinjo, Yusuke Komiyama, Masaaki Kotera, Tatsuya Kushida, James Malone, Masaaki Matsubara, Satoshi Mizuno, Sayaka Mizutani, Hiroshi Mori, Yuki Moriya, Katsuhiko Murakami, Takeru Nakazato, Hiroyo Nishide, Yosuke Nishimura, Soichi Ogishima, Tazro Ohta, Shujiro Okuda, Hiromasa Ono, Yasset Perez-Riverol, Daisuke Shinmachi, Andrea Splendiani, Francesco Strozzi, Shinya Suzuki, Junichi Takehara, Mark Thompson, Toshiaki Tokimatsu, Ikuo Uchiyama, Karin Verspoor, Mark D. Wilkinson, Sarala Wimalaratne, Issaku Yamada, Nozomi Yamamoto, Masayuki Yurimizu, Shoko Kawamoto, Toshihisa Takagi. BioHackathon series in 2013 and 2014: improvements of semantic interoperability in life science data and services. *F1000Research* 2019, 8:1677, 2019. (DOI: 10.12688/f1000research.18238.1).

2. その他の著作物(総説、書籍など)

1. 高橋弘喜、組成解析(真菌)、化学療法の領域、Vol.33、No.7、2017年
2. Ikuo Uchiyama, Ortholog identification and comparative analysis of microbial genomes using MGD and RECOG. In "Protein function prediction: Methods and protocols", (D. Kihara ed.), pp.147-168, Humana Press, 2017.
3. 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
4. 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
5. 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
6. 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
7. 黒川顕、森宙史、東光一、9章 メタゲノム解析、よくわかるバイオインフォマティクス入門、講談社、pp123-pp136、2018年

- 村上匠、森宙史、黒川顕、メタゲノムデータの情報解析とデータベース、腸内細菌叢 健康と疾患を制御するエコシステム、実験医学増刊、羊土社、37、pp23-28、2019 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	27 件
	国際	8 件
口頭発表	国内	13 件
	国際	7 件
ポスター発表	国内	14 件
	国際	7 件

(2) 招待講演

〈国内〉

- 千葉啓和、ドメインレベルのオーソログ解析でみる融合遺伝子の進化、ラン藻ゲノム交流会、東京大学駒場キャンパス、2017 年 6 月 24 日
- 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 年 7 月 28 日
- 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 年 10 月 9 日
- 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 年 10 月 12 日
- 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 年 11 月 10 日
- 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 年 12 月 22 日
- 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 年 1 月 25 日
- 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 年 3 月 27 日
- 高橋弘喜、進歩が著しいマイクロバイオームおよび微生物ゲノム研究法、第 117 回日本皮膚科学会総会、広島市、2018 年 6 月 2 日
- 中村保一、塩基配列データの爆発的増加と DDBJ をとりまく現状、遺伝研研究会・生命情報データの拡大と分子進化研究の進展、三島、2018 年 8 月 4 日
- 内山郁夫、非コア遺伝子に着目した水平移動遺伝子の検出、遺伝研研究集会「DNA 水

- 平伝搬」、遺伝研、2018年8月20-21日
12. 内山郁夫、オーソログを探求する: Quest for Orthologs 会議について、日本進化学会第20回大会、東京大学駒場キャンパス、2018年8月22-25日
 13. 千葉啓和、オーソログデータベースに基づくドメイン融合解析、日本進化学会第20回大会、東京大学駒場キャンパス、2018年8月22-25日
 14. 中村保一、ゲノム解析手法のトレンドと DDBJ・遺伝研スーパーコンピュータ、第二回木村資生記念進化学セミナー、八王子、2018年9月3日
 15. 黒川顕、微生物統合データベース MicrobeDB.jp、JBA 発酵と代謝研究会、東京、2018年9月3日
 16. 黒川顕、メニーコアプロセッサが可能とするリアルタイムメタゲノム解析、アクセラレーション技術発表討論会、油津、2018年9月7日
 17. 森宙史、Metagenome informatics to understand microbial communities, イルミナマイクロバイオロジーワークショップ 2018『メタバーコーディングからショットガンメタゲノムによる全ゲノムへの展望』, 東京コンファレンスセンター品川, 2018年9月18日
 18. 高橋弘喜、病原真菌 *Aspergillus fumigatus* の環境応答能の解析, 第101回日本細菌学会関東支部総会, 東京都, 2018年11月1日
 19. 黒川顕、マイクロバイオームと未来社会、第2回経団連・産業連携セミナー、東京、2019年2月25日
 20. 中村保一、INSDC の一員としての DDBJ 事業の紹介と塩基配列決定の国際動向と INSDC への塩基配列データの登録と管理の実際、製品評価技術基盤機構 NITE バイオテクノロジーセンターセミナー、木更津、2019年2月22日
 21. 黒川顕、微生物統合データベース MicrobeDB.jp、第11回日本ゲノム微生物学会、東京、2019年3月8日
 22. 高橋弘喜、病原真菌アスペルギルスフミガタスの表現型解析および遺伝型解析、感染症診断と治療におけるゲノム解析、伊勢原市、2019年3月15日
 23. 黒川顕、微生物統合データベース MicrobeDB.jp、日本農芸化学会、東京、2019年3月26日
 24. 高橋弘喜、病原真菌 *Aspergillus fumigatus* の環境応答能の理解に向けて、第92回日本細菌学会総会、札幌市、2019年4月24日
 25. 黒川顕、マイクロバイオームデータの可能性、第32期 CAMM フォーラム本例会、東京都、2019年5月10日
 26. 黒川顕、マイクロバイオームデータの可能性、トヨタ自動車、豊田市、2019年6月7日
 27. 高橋弘喜、次世代シーケンサー (NGS) の進展と微生物研究への応用、第56回小児アレルギー学会、千葉市、2019年11月2日(予定)

〈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
2. 中村保一、DDBJ/INSDC/DFAST、GCM 2.0 workshop、北京、2018/6/26
3. Ken Kurokawa, Integrated database for microbe "MicrobeDB.jp", JSME2018&10th

ASME, Okinawa, 2018/7/13

4. 中村保一、Genome analyses using NGS for next-generation model species, The 16th CJK (日中韓) Bioinformatics Symposium、葉山、2018/8/30
5. 中村保一、DDBJ / the INSDC is ready for GCM 2.0, 8th WDCM / GCM 2.0 meeting、北京、2018/11/21
6. 中村保一、Introduction of the INSDC, data submission & genome annotation, GCM 2.0 training course、北京、2018/11/26
7. Ken Kurokawa, Hiroshi Mori, Integrated Database for Microbes, "MicrobeDB.jp", Laser Solutions for Space and the Earth 2019 (LSSE2019), Yokohama, April 23-26, 2019.
8. Uchiyama, I. Ortholog relationships among pan-genomes: the Microbial Genome Database (MBGD) for exploring microbial diversity, The 67th NIBB Conference "Quest for Orthologs", National Institute for Basic Biology, Okazaki, Jul. 31-Aug.2, 2019.

(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 年 5 月 23 日
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 年 6 月 2 日
3. 森宙史、東光一、黒川頭、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 年 6 月 30 日
4. 黒川頭、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 年 8 月 31 日
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 年 9 月 11 日
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 回日本医用マススペクトル学会、一橋講堂、2017 年 9 月 14 日
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 年 9 月 28 日
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 年 12 月 9 日
9. 山田拓司、大腸がん発病に伴うヒト腸内環境のマルチオミクス解析、第45回日本毒性学会学術年会、大阪国際会議場、2018 年 7 月 19 日
10. 森宙史、黒川頭、微生物統合データベース MicrobeDB.jp の活用法、ワークショップ「生命科学のデータベース活用法 2018」、第 41 回日本分子生物学会年会、パシフィコ横浜、2018 年 11 月 28 日
11. 森宙史、メタゲノムインフォマティクスとヒトマイクロバイオーーム統合データベース、第 92 回日本生化学会大会シンポジウム「共生微生物叢研究の最前線」、パシフィコ横浜、2019

年 9 月 20 日

12. 森宙史、東光一、黒川顕、微生物統合データベース MicrobeDB.jp と関連するツール、ワークショップ「微生物—植物相互作用を解明して育種に繋げるには?」、トーゴの日シンポジウム 2019、日本科学未来館、2019 年 10 月 5 日
13. Cai Bian, Yikelamu Alimu, Hiroki Takahashi, Takashi Yaguchi、Characterization and proposal of two new species in *Aspergillus* section *Nigri*、第 63 回日本医真菌学会総会・学術集会、千葉市、2019 年 10 月 11 日

〈国際〉

1. Ikuo Uchiyama, A new protocol for constructing the ortholog table in Microbial Genome Database for Comparative Analysis, NIG International Symposium 2017, Mishima, May 27-29, 2017.
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, Jun. 8-10, 2017.
3. Hirokazu Chiba, Fernández-Breis, Jesualdo 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, Jun. 8-10, 2017.
4. Hiroshi Mori, MicrobeDB.jp: integrating microbial genome and metagenome data using semantic web technology, BioHackathon 2017 symposium, JST Science Plaza, Japan, Sep. 10, 2017.
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, Nov. 1, 2017.
6. Hiroshi Mori, Ken Kurokawa, VITCOMIC2: Visualization and analysis tool for the phylogenetic composition of microbial communities based on 16S rRNA gene sequences, ASM Microbe 2018, Georgia World Congress Center, Atlanta, USA, June 7-12, 2018.
7. Hirokazu Chiba, Ikuo Uchiyama, Unraveling the microbial gene repertoire by sub-gene level orthologous clustering. The 67th NIBB Conference “Quest for Orthologs”, National Institute for Basic Biology, Okazaki, Jul. 31-Aug.2, 2019.

(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 年 5 月 22 日-24 日
2. 内山郁夫、三原基広、西出浩世、千葉啓和、MBGD 2018:大規模な微生物ゲノムデータにより効果的な比較に向けた改良、トーゴの日シンポジウム 2017、東京大学弥生講堂 一条ホール、2017 年 10 月 4-5 日
3. 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、黒川顕、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 年 10 月 4-5 日
4. 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 回ゲノム微生物学会年会、京都大学桂キャンパス、2018 年 3 月 5-7 日
 5. 藤澤貴智、森宙史、谷澤靖洋、神沼英里、内山郁夫、山田拓司、高橋弘喜、中村保一、黒川顕、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、京都市、2018 年 3 月 5 日-7 日
 6. 森宙史、黒川顕、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 年 3 月 5-7 日
 7. 豊留孝仁、高橋弘喜、亀井克彦、日本臨床分離 *Aspergillus flavus* 株と *A. flavus* NRRL 3357 株および *A. oryzae* RIB40 株とのゲノムワイド SNP 比較解析、日本マイコトキシソ学会第 82 回学術講演会、帯広市、2018 年 8 月 24 日
 8. 藤澤貴智、森宙史、谷澤靖洋、神沼英里、内山郁夫、山田拓司、高橋弘喜、中村保一、黒川顕、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、トーゴの日シンポジウム 2018、2018 年 10 月 5 日
 9. 森宙史、黒川顕、MicrobeDB.jp のメタゲノム解析パイプライン、トーゴの日シンポジウム 2018、日本科学未来館、2018 年 10 月 5 日
 10. 内山郁夫、非コア遺伝子の並び順の保存性に基づくゲノミックアイランドの抽出、第 13 回日本ゲノム微生物学会年会、首都大学東京南大沢キャンパス、2019 年 3 月 6-8 日
 11. 森宙史、藤澤貴智、西出浩世、矢口貴志、高橋弘喜、山田拓司、内山郁夫、中村保一、黒川顕、微生物統合データベース MicrobeDB.jp の更新とヒトマイクロバイオームデータの統合、第 13 回日本ゲノム微生物学会年会、八王子市、2019 年 3 月 6-8 日
 12. 内山郁夫、三原基広、西出浩世、千葉啓和、高柳正彦、高見英人、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 2019、日本科学未来館、2019 年 10 月 5 日
 13. 藤澤貴智、森宙史、谷澤靖洋、児玉悠一、内山郁夫、中川善一、山田拓司、高橋弘喜、中村保一、黒川顕、統合微生物データベース MicrobeDB.jp ポータルサイト拡張、トーゴの日シンポジウム 2019、2019 年 10 月 5 日
 14. 森宙史、藤澤貴智、中川善一、山田拓司、黒川顕、MicrobeDB.jp ver. 3 のマイクロバイオームデータの概要、トーゴの日シンポジウム 2019、日本科学未来館、2019 年 10 月 5 日

〈国際〉

1. Yuumi Nakamura, Hiroki Takahashi H, Akiko Takaya, Yuzaburo Inoue, Yuki Katayama, Yoko Kusuya, Rena Oguma, Fumiya Yamade, Naoki Shimojo, Gabriel Nunez, Hiroyuki Matsue, Cutaneous acquisition of *Staphylococcus*-sensing agr mutations protects against atopic dermatitis, International Eczema Council at SID meeting/76th SID Annual Meeting, Portland, Apr. 26-29, 2017
2. Hirokazu Chiba, Ikuo Uchiyama, Applying the semantic web technology to dive

rsified biological resources for comparative omics、NIG International Symposium 2017 Commemorating the 30th Anniversary of DDBJ、Mishima Citizens Cultural Hall、May. 27-29、2017

3. 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、May. 27-29、2017
4. 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、Sep. 19-22、2017
5. Hiroshi Mori、Ken Kurokawa、VITCOMIC2 and MicrobeDB.jp: Analyzing taxonomic composition of microbial communities based on 16S rRNA gene sequences without sequence clustering、ISME17、Congress Center Leipzig、Leipzig、Germany、Aug. 13-17、2018
6. Hiroshi Mori、Ken Kurokawa、VITCOMIC2 and MicrobeDB.jp: Analyzing and comparing microbial community sequencing data、Frontiers of Genome Science 2019、Ito International Research Center、University of Tokyo、Japan、Jan. 9-10、2019
7. Hiroki Takahashi、Maico Umemura、Akihiro Ninomiya、Masaaki Shimizu、Yoko Kusuya、Shun-ichi Urayama、Akira Watanabe、Katsuhiko Kamei、Takashi Yaguchi、Daisuke Hagiwara、Diversified transcriptional regulation of secondary metabolic gene clusters in closely related *Aspergillus* species、30th Fungal Genetics Conference、Pacific Grove、Mar. 13、2019

4. 知財出願

(1) 出願件数

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

(2) 一覧

①国内出願

該当なし。

②海外出願

1. 情報処理システム、情報処理方法、及びプログラム、黒川顕、東光一、森宙史、2018/1/

19、PCT/JP2018/001594

2. INFORMATION PROCESSING SYSTEM, INFORMATION PROCESSING METHOD, AND PROGRAM, Ken Kurokawa, Koichi Higashi, Hiroshi Mori, 2018/1/19, 米国 16/480,539

③その他の知的財産権

該当なし。

5. 受賞・報道等

(1) 受賞

1. Linked Open Data チャレンジ 2017 アイデア部門優秀賞、藤澤貴智・山本泰智、2018年3月10日
2. Linked Open Data チャレンジ Japan 2018 プロGRESS賞、山本 泰智、藤澤 貴智、小林 巖生、2018年12月8日

(2) メディア報道

該当なし。

(3) その他

特になし。

§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
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	同上
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
2018年 4月25日	チーム内ミーティング (非公開)	遺伝研	4人	進捗報告のためのミーティング
2018年 5月16日	チーム内ミーティング (非公開)	東工大	3人	同上
2018年 6月20日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2018年 7月5日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2018年 9月6日	チーム内ミーティング (非公開)	遺伝研	4人	同上
2018年 11月5日	チーム内ミーティング (非公開)	遺伝研	4人	同上
2018年 11月21日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2018年 12月25日	チーム内ミーティング (非公開)	遺伝研	7人	産業応用のためのミーティング
2018年 12月26日	チーム内ミーティング (非公開)	遺伝研	4人	進捗報告のためのミーティング
2019年 1月30日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2019年 2月19日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2019年 4月11日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2019年 5月10日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2019年	チーム内ミーティング	東工大	3人	同上

年月日	名称	場所	参加人数	目的・概要
5月24日	(非公開)			
2019年 6月3日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2019年 6月20日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2019年 7月1日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2019年 7月22日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2019年 7月26日	チーム内ミーティング (非公開)	東工大	3人	同上
2019年 8月13日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2019年 8月26日	チーム内ミーティング (非公開)	遺伝研	3人	同上
2019年 9月17日	チーム内ミーティング (非公開)	遺伝研	5人	ワークショップ準備のための話し合い
2019年 9月30日	チーム内ミーティング (非公開)	東工大	3人	進捗報告のためのミーティング
2019年 10月17日	チーム内ミーティング (非公開)	遺伝研	7人	進捗報告のためのミーティング
2019年 10月18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

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
2019年7 月31日～8 月2日	Quest for Orthologs	基礎生物学 研究所・岡崎 コンファレン	50人	オーソログ DB や推定手法の開発者と利用者が集まって隔年で開催される国際コンソー

年月日	名称	場所	参加人数	目的・概要
		スセンター		シム会議。内山がオーガナイザーとして、基生研コンファレンスの一環として開催した。
2019年10月5日	ワークショップ「微生物—植物相互作用を解明して育種に繋げるには？」	日本科学未来館コンファレンスルーム土星	40人	MicrobeDB.jp のホログenom連携を推し進めるために、ホログenom研究を行なっている研究者および植物統合DB開発者と共に、ホログenom研究に有用な統合DBとはどのようなDBなのか等について話し合った。

以上