

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」 (https://microbedb.jp/)
研究代表者氏名	黒川 颯
所属・役職	国立遺伝学研究所 情報研究系 教授 (2020年3月時点)

□目次

§1. 研究実施体制.....	1	① 概要.....	10
§2. 研究開発対象とするデータベース・ツール等.....	3	② 招待講演.....	11
(1) データベース一覧.....	3	③ 口頭講演.....	11
(2) ツール等一覧.....	3	④ ポスター発表.....	12
§3. 実施内容.....	4	(4) 知的財産権の出願.....	12
(1) 本年度の研究開発計画と達成目標.....	4	(5) 受賞・報道等.....	12
(2) 進捗状況.....	5	§5. 研究開発期間中に主催した活動 (ワークショップ等)	13
§4. 成果発表等.....	9	1. 進捗ミーティング.....	13
(1) 原著論文発表.....	9	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.....	13
① 論文数概要.....	9		
② 論文詳細情報.....	9		
(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. 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		https://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
5	Latent Environment Allocation	LEA	http://leamicrobe.jp

§3. 実施内容

(1) 本年度の研究開発計画と達成目標

MicrobeDB.jp の実用化を目指した研究開発計画 7 項目は以下の通りである。

① 徹底したユーザビリティの向上

MicrobeDB.jp ポータルサイトを公開し、また、ポータルサイトから **Stanza** およびアプリケーションを 10 種以上設置することによりユースケースを追加する。特にゲノム情報に関する **TogoStanza** の情報の提供ならびに比較解析機能を強化することでユーザビリティのさらなる向上を図る。ユーザ会、講習会等によって得られたユーザからのフィードバックを精査し、ポータルサイトの本公開後も継続的な改善を実施する。

② データ品質の向上

MEO の自動アノテーションについて、JCM および NBRC の菌株メタデータの MEO によるマニュアルアノテーション結果を教師データとして追加する。また、MEO の自動アノテーションツールが誤アノテーションするパターンの分析を行い、その結果を元にしてアノテーションルールの追加・修正を行うことで、MEO の自動アノテーションの精度向上を行う。ヒトマイクロバイームデータについては、メタデータのマニュアルキュレーション結果を、MEO の自動アノテーション結果とは区別した形で RDF 化すると共に、MEO の自動アノテーションの精度向上に利用する。

これまでにリスト化したヒトメタゲノム論文のうち、残りの約 700 報の論文に対して引き続きマニュアルキュレーションを行う。また新規に追加すべき論文を精査し論文リストを更新する。地理情報や疾患関連情報を論文横断的に収集することにより、公共データを用いた新たなデータ解析研究の素地となるデータベース構築を目指す。また、これまでに蓄積してきたマニュアルキュレーション手法を活用し、論文からの自動データ抽出を試みる。

③ キラーアプリケーションの開発

大量の真菌のメタ ITS 配列データから高速かつ正確に系統組成を推定する解析パイプラインを開発し公開する。また、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検討を続ける。

ゲノムアノテーションアプリケーションについて、前年度までに MMseqs2 および MAPLE を用いた解析パイプラインを作成したが、出力は従来の MAPLE 解析のままであり、MBGD との連携による付加価値の追加ができていない。そこで、対象ゲノムと近縁のゲノムとの間でモジュール構成を比較する機能、および同一モジュール内の遺伝子の系統プロファイルを比較する機能を追加し、対象ゲノムの機能的な特徴づけや、失われた遺伝子に対する代替経路の探索などをサポートするアプリケーションとして利用できるようにする。あわせて、この機能を効果的に使うための MyMBGD モードの改変、および MicrobeDB.jp との連携作業を行う。

系統プロファイル検索は、上記の代替経路の探索や、メタゲノムデータの解釈などにも使えることを念頭に、検索条件を指定するための使いやすいユーザインターフェースを作成する。また、MicrobeDB.jp 内のフェノタイプやメタゲノムデータなどに基づいて作成した系統プロファイルをあらかじめ用意し、これに基づく検索を行うことにより、その機能を手軽に試せるようにする。現状の実装では系統プロファイル間の類似性の評価が甘いので、種間の系統関係を反映させた系統プロファイルの評価方法に置き換える。

コアゲノム・パンゲノム解析機能については、ユーザが持つ多数の同種のゲノムを登録し、それらの中で効率的なオーソログ解析を行うのが本質的な部分であり、これを MyMBGD における系統特異的ゲノム比較モードの特殊なバージョンとして実装する。このため、大量ゲノムの一括登録、および MBGD で近縁ゲノム比較に用いているのと同等の高速ホモロジー検索を用いたオーソログ解析などの機能を追加する。

④ さらなるデータの統合

DBCLS と連携して開発した JSON2LD マッパーを利用して解析ツール出力結果の JSON 形式リソースを RDF に変換して微生物ゲノム情報とのデータ統合化を実施する。

公開されている大量の真菌のメタ ITS 配列データを解析し、サンプルごとに系統組成を推定してメタデータと共に RDF 化する。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ための解析手法のパイプラインの実装および既存の解析手法との比較検討を進める。

昨年度に引き続き、真菌類の完全長・ドラフトゲノムデータと、RNA-Seq データの収集・整理を進める。これらの RDF 化を進めるとともに、菌株リソースのメタデータの整理と RDF 化を行い、MicrobeDB.jp 上へデータを格納する。ITS 配列が登録済みの千葉大真菌センター保有菌株のメタ情報の整理に着手する。

⑤ 基盤データ解析技術の高度化

公共の大量のメタ 16S・メタゲノム配列データの一括解析および一括更新が既に困難になっているため、計算が終わった分だけ差分で更新することを可能にするための、解析パイプラインや参照データベースの改良を行う。

MBGD2018 年バージョンの公開が、2018 年の後半までずれこんだため、次期バージョンの更新作業の開始が予定より遅れているが、次年度内に完了することを目指して更新作業を行う。すでに対象となる完全ゲノムデータの列挙を行っており、ゲノム数は現状の約 2.5 倍の 15,421 にふくれあがるが、genus 数としては 1,198 で、これは現状の 20%弱増し程度であるため、前年度までに作成した階層的オーソログデータ構築プロトコルを用いれば更新可能と考えている。ただ、データサイズは大きくなるので、ホモロジーデータの差分更新や、大きなオーソロググループのサブグループ化などを行って、効率的な更新を行えるように更新プロトコルを工夫する。

ドラフトゲノムについては、これまで通り genus レベル以上でユニークなゲノムについては draft-plus データセットとして取り込むが、これは MMseqs2 によって既存のオーソロググループにアサインする形で行う。それ以外のドラフトゲノムについては、種内および近縁ゲノム間比較の材料として MyMBGD モードで使われることを念頭に、完成度などのクオリティに関するデータをつけて整理しておく。なお、当初予定ではここで各ゲノムデータに優先度づけを行うことを想定していたが、これを行う予定のポストドクが退職したこと、および現状で優先度づけの方針について目処が立っていないことから、当面の目標からは外すこととする。

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、各種データセットの追加、更新実施のために、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、API による追加、更新系の開発を完成する。MEQ の自動アノテーションや OpenRefine などを用いたキュレーションプロセスをワークフロー化し API を組み込むことで、キュレーション情報を反映し、継続的な精度向上を目指す。さらに、企業やユーザの要望に応じて MicrobeDB.jp ウェブアプリケーションのユーザ環境構築のため、データベース基盤システムやミドルウェア利用を見直し、コンテナ環境を導入することでシステムの効率化を図る。

⑦ ホロゲノム対応

かずさ DNA 研究所との共同研究として開発した、ホロゲノム解析支援ツールについて、ユースケースの一つとしてポータルサイトから公開する。また継続して国立環境研究所や藍藻ゲノムデータベース CyanoBase, 根粒菌ゲノムデータベース 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、リファレンスとなるホスト=共生微生物の対象の拡充を目指す。

(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

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、2020 年 2 月にリリースした。MicrobeDB.jp version 3 の公開にあわせて、MicrobeDB.jp のユーザ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、テスト版として公開・運用していたポータルサイトの正式版を、2020 年 2 月に公開し、ポータルサイトを MicrobeDB.jp のトップページとした。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。

ヒト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 2,300 報について確認した。年齢や血液データなどを含む関連メタデータは 399 項目であり、そのうち疾患数は 47 疾患、metadata が紐付けられている 18,282 サンプルを得ることができた。ここで抽出した疾患データを RDF 化するに際して、腸内環境関連疾患を表現するオントロジー Human Microbiome Associated Disease Ontology (HMADO) を新たに構築した。具体的には、論文マニュアルキュレーションによって得られた 47 項目の疾患名に HMADO ID を割り振り、その疾患を階層化、その後既存の疾患分類体系 (ICD、Mesh) と対応つけた。INSDC DRA/ERA/SRA などから大規模に抽出されたサンプルに付加する環境情報などのメタデータから、上記疾患名を抽出することで、100,349 サンプルに対して疾患名を紐つけることができた。上記情報はすでに MicrobesDB.jp に格納し公開済みであり、オンライン上で検索可能である。

新たなユースケースとして、これまでの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サンプルのファセット検索機能を実装し、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 UI を再構築することでシンプルな操作を実現した。ファセット検索機能には、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ームデータについて、宿主のヒトの病気、Ethnicity、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。

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、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、データの自動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。

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系統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 について、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、描画処理を効率化および 3D に拡張して公開した。また、LEA の遺伝機能組成版のアルゴリズムの開発を行った。

ゲノムアノテーションアプリケーションについては、複数のゲノムに対して Genomape (MAPLE から改称) を実行して、得られた結果を比較するインターフェイスを作成した。これは、利用者が指定した比較対象とするゲノムに対して、KEGG モジュールを縦、ゲノムを横に配置して、ゲノムごとに各 KEGG モジュールの有無を判定して比較できるようにした表と、個々の KEGG モジュールについて、その構成要素であるオーソロググループの有無をゲノムごとに表示した表、さらにこの情報をゲノム上の遺伝子の並びとして表して比較で

きるようにした図の3画面からなる。MBGD に登録されたゲノムについては、MBGD のオーソロググループに基づいてあらかじめ Genomaple 解析まで行った結果をストアしており、直接呼び出して比較することができる。これに、MyMBGD から登録した利用者のゲノムに対して行った解析結果も併せて比較できるようにした。

これに合わせて MyMBGD インターフェイスも一新し、新たに利用者のゲノムを既存のオーソログテーブル (standard ortholog table) と照合して、差分的に追加する機能をつけた。この解析に、Genomaple 解析をつなげることによって上述の解析が行えるようにした。MyMBGD インターフェイスは、この既存オーソログテーブルに追加するモードと、従来からある新規にクラスタリングを行うモードとに分けて、今後は後者の方でコアゲノム・パンゲノム解析までを行えるようにする予定である。

系統プロファイルアプリケーションは、系統群や表現系、生存環境などから興味のある条件を利用者が指定して、関連する系統プロファイル情報(各生物種に対するあるなしのパターン)を取得し、それに近いあるなしパターンを示すオーソロググループを検索できるようなインターフェイスを作成した。系統プロファイルの類似性は超幾何分布に基づいて統計評価を行うが、一般に系統関係の影響を受けて p 値が過小評価される傾向があるため、系統関係に基づいてパターンを圧縮して評価する手法を用いた。

④さらなるデータの統合

真菌メタ ITS 解析パイプラインについて、千葉大グループと連携して、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。

千葉大真菌センターが保有する真菌類約 19,000 株のリソースの RDF 化を進めている。これらのうち、まずは ITS 配列が紐づく菌株データ 851 株を対象として RDF 化を進めるとともに、新たに約 1,000 株の ITS 配列の決定およびアノテーションを実施した。これらの DDBJ への登録作業は継続して進めている。真菌類オミックスデータの整備に関しては、独自に解読した真菌類 10 株のドラフトゲノムデータについて、DDBJ アクセション番号の取得を完了した。引き続きドラフトゲノムデータの収集と整理を進めている。真菌類 RNA-seq データの整備も進めており、公共データ SRA (NCBI) に登録されている 19,432 件 (2017 年 7 月現在) のうち、約 4,000 件に関して、遺伝子発現量までの算出を完了し、前年度と合わせて 13,264 件のデータ整備が完了した。

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。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。

オーソログデータの構築に際して、RefSeq から完全ゲノムデータを抽出したところ、新規ゲノムが 9,088 件、トータルで 15,401 件となった。このうち複数株を持つ 991 種について種レベルのパンゲノムを作成し、さらに複数種を持つ 692 属について属レベルのパンゲノムを作成した。これらの近縁比較においては相同性検索のプログラムを ublast から diamond に変更し、またトップレベルの遠縁比較においては従来通り blast を用いているが、差分更新によって行うようにした。最終的に 1,444 属の(パン)ゲノムについて DomClust を用いたオーソログデータの作成を行った。現在、DomRefine を用いたマルチプルアライメントに基づくオーソロググループの修正を行なっているが、時間がかかっているため、ボトルネックの調査を行いつつ進めている。

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登録サンプルメタデータの RDF を再設計し、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。また、ユーザのニーズおよびデータ更新のための冗長な環境構築を効率的に実現するためにデータベース基盤システムのすべてのコンポーネントの Docker コンテナ化を実施した。

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ーフェイスとして 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ーフェイスを整備した。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 のデータとの変換 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 のホスト-共生微生物の対象を拡充し、植物ホストおよび 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。

§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)	9 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. 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 Yarimizu, 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. Rutger A Vos, Toshiaki Katayama, Hiroyuki Mishima, Shin Kawano, Shuichi Kawashima, Jin-Dong Kim, Yuki Moriya, Toshiaki Tokimatsu, Atsuko Yamaguchi, Yasunori Yamamoto, Hongyan Wu, Peter Amstutz, Erick Antezana, Nobuyuki P Aoki, Kazuharu Arakawa, Jerven T Bolleman, Evan Bolton, Raoul J P Bonnal, Hidemasa Bono, Kees Burger, Hirokazu Chiba, Kevin B Cohen, Eric W Deutsch, Jesualdo T Fernández-Breis, Gang Fu, Takatomo Fujisawa, Atsushi Fukushima, Alexander García, Naohisa Goto, Tudor Groza, Colin Hercus, Robert Hoehndorf, Kotone Itaya, Nick Juty, Takeshi Kawashima, Jee-Hyub Kim, Akira R Kinjo, Masaaki Kotera, Kouji Kozaki, Sadahiro Kumagai, Tatsuya Kushida, Thomas Lütteke, Masaaki Matsubara, Joe Miyamoto, Attayeb Mohsen, Hiroshi Mori, Yuki Naito, Takeru Nakazato, Jeremy Nguyen-Xuan, Kozo Nishida, Naoki Nishida, Hiroyo Nishide, Soichi Ogishima, Tazro Ohta, Shujiro Okuda, Benedict Paten, Jean-Luc Perret, Philip Prathipati, Pjotr Prins, Núria Queralt-Rosinach, Daisuke Shinmachi, Shinya Suzuki, Tsuyosi Tabata, Terue Takatsuki, Kieron Taylor, Mark Thompson, Ikuo Uchiyama, Bruno Vieira, Chih-Hsuan Wei, Mark Wilkinson, Issaku Yamada, Ryota Yamanaka, Kazutoshi Yoshitake, Akiyasu C Yoshizawa, Michel Dumontier, Kenjiro Kosaki, Toshihisa Takagi. *BioHackathon 2015: Semantics of data for life sciences and reproducible research*. *F1000Research* 2020, 9:136 (DOI: 10.12688/f1000research.18236.1).
3. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, Watanabe H, Masuda K, Nishimoto Y, Kubo M, Hosoda F, Rokutan H, Matsumoto M, Taka

- maru H, Yamada M, Matsuda T, Iwasaki M, Yamaji T, Yachida T, Soga T, Kurokawa K, Toyoda A, Ogura Y, Hayashi T, Hatakeyama M, Nakagama H, Saito Y, Fukuda S, Shibata T, Yamada T. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine*. 2019, 25, 968–976, doi: 10.1136/gutjnl-2019-319188.
4. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, Beghini F, Manara S, Karcher N, Pozzi C, Gandini S, Serrano D, Tarallo S, Francavilla A, Gallo G, Trompetto M, Ferrero G, Mizutani S, Shiroma H, Shiba S, Shibata T, Yachida S, Yamada T, Wirbel J, Schrotz-King P, Ulrich CM, Brenner H, Arumugam M, Bork P, Zeller G, Cordero F, Dias-Neto E, Setubal JC, Tett A, Pardini B, Rescigno M, Waldron L, Naccarati A, Segata N. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nature Medicine*. 2019, 25, 667-678. doi: 10.1038/s41591-019-0663-4.
 5. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, Fleck JS, Voigt AY, Palleja A, Ponnudurai R, Sunagawa S, Coelho LP, Schrotz-King P, Vogtmann E, Habermann N, Niméus E, Thomas AM, Manghi P, Gandini S, Serrano D, Mizutani S, Shiroma H, Shiba S, Shibata T, Yachida S, Yamada T, Waldron L, Naccarati A, Segata N, Sinha R, Ulrich CM, Brenner H, Arumugam M, Bork P, Zeller G. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medicine*. 2019, 25, 679-689. doi:10.1038/s41591-019-0406-6.
 6. Erawijantari PP, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, Saito Y, Fukuda S, Yachida S, Yamada T. Influence of gastrectomy for gastric cancer treatment on faecal microbiome and metabolome profiles. *Gut*. 2020, January 16 online, doi:10.1136/gutjnl-2019-319188
 7. 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*, 10, 854, 2019 (DOI:10.3389/fmicb.2019.00854).
 8. 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).
 9. 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:10.1111/1348-0421.12680).

(2) その他の著作物(総説、書籍など)

特になし

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

② 招待講演

〈国内〉

1. 森宙史, メタゲノムインフォマティクスと微生物統合データベース, 産総研セミナー, 産業総合技術研究所つくばセンター, 2019 年 11 月 1 日.
2. 高橋弘喜, 病原真菌 *Aspergillus fumigatus* の環境応答能の数理モデル化による理解の試み, 第 93 回日本細菌学会総会、ウインクあいち, 2020 年 2 月 19 日
3. 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、幕張メッセ, 2019 年 11 月 2 日
4. 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、札幌コンベンションセンター, 2019 年 4 月 24 日

〈国際〉

1. Ikuo Uchiyama, Motohiro Mihara, Hiroyo Nishide, Orthology Relationships among Pan-Genomes: the Microbial Genome Database (MBGD) for Exploring Microbial Diversity, The 67th NIBB Conference "Quest for Orthologs", Okazaki, Jul. 31 - Aug. 2
2. Metagenomic and metabolomic analyses reveal dynamic shifts in gut microbiota along the adenoma-carcinoma sequence in colorectal cancer, 山田拓司, OIST Workshop "A World of Microbiota", 2019/7/4,
3. Metagenomic and metabolomic analyses in colorectal cancer, 山田拓司, Microbiome Movement - Drug Development, Singapore, 2019/9/17

③ 口頭講演

〈国内〉

1. 内山郁夫, 微生物比較ゲノムデータベースを用いた水平移動遺伝子動態の網羅的解析、遺伝学研究集会「環境中の DNA 循環」、国立遺伝学研究所, 2019 年 8 月 19-20 日
2. 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.
3. 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?」, トーゴーの日シンポジウム 2019, 日本科学未来館, 2019 年 10 月 5 日.
4. 森宙史, 中川善一, 山田拓司, Human Microbiome Associated Disease Ontology の構築と比較メタゲノム解析, 第 13 回日本ゲノム微生物学会若手の会研究会, 山喜旅館, 2019 年 10 月 27 日.
5. 森宙史, 微生物の生息環境のアノテーションとヒトコホートのメタゲノム研究, Annotathon 2019, 東京大学柏の葉キャンパス駅前サテライト, 2019 年 11 月 13 日.

6. 森宙史, MicrobeDB.jp version 3 の活用法, 第 42 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 2019」, 福岡国際会議場, 2019 年 12 月 3 日.
7. 内山郁夫、微生物比較ゲノムデータベース MBGD に基づく水平移動遺伝子動態の網羅的解析、日本ゲノム微生物学会年会、ウインクあいち、2020 年 3 月 6-8 日
8. 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、オークラ千葉ホテル、2019 年 10 月 11 日

〈国際〉

1. Hirokazu Chiba, Ikuo Uchiyama, Unraveling the microbial gene repertoire by sub-gene level orthologous clustering, The 67th NIBB Conference “Quest for Orthologs”, Okazaki, Aug. 1, 2019.

④ ポスター発表

〈国内〉

1. 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, トーゴーの日シンポジウム 2019, 日本科学未来館, 2019 年 10 月 5 日.
2. 内山郁夫、微生物比較ゲノムデータベース MBGD、トーゴーの日シンポジウム 2019、日本未来科学館、2019 年 10 月 5 日
3. 千葉啓和、オーソログ情報の統合化と利活用、トーゴーの日シンポジウム 2019、日本科学未来館、2019 年 10 月 5 日
4. 内山郁夫、微生物比較ゲノムデータベース MBGD を用いた水平移動遺伝子動態の網羅的解析、日本分子生物学会年会、マリンメッセ福岡、2019 年 12 月 3-6 日
5. 森宙史、藤澤貴智、西出浩世、矢口貴志、高橋弘喜、中川善一、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 version 3、第 14 回日本ゲノム微生物学会年会、ウインクあいち、2020 年 3 月 6 日-8 日

〈国際〉

1. Hiroshi Mori, Development of the semantic-web based integrated database of human microbiome, The 8th Global Network Forum on Infection and Immunity: Microbiome, Chiba University Hospital, January 10–11, 2020.

(4) 知的財産権の出願

特になし

(5) 受賞・報道等

特になし

§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
2019年 4月11日	チーム内ミーティング (非公開)	遺伝研	3人	進捗報告のためのミーティング
2019年 5月10日	チーム内ミーティング (非公開)	遺伝研	3人	進捗報告のためのミーティング
2019年 5月24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 6月3日	チーム内ミーティング (非公開)	遺伝研	3人	進捗報告のためのミーティング
2019年 6月20日	チーム内ミーティング (非公開)	遺伝研	3人	進捗報告のためのミーティング
2019年 7月1日	チーム内ミーティング (非公開)	遺伝研	3人	進捗報告のためのミーティング
2019年 7月22日	チーム内ミーティング (非公開)	遺伝研	3人	進捗報告のためのミーティング
2019年 7月26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 8月13日	チーム内ミーティング (非公開)	遺伝研	3人	進捗報告のためのミーティング
2019年 8月26日	チーム内ミーティング (非公開)	遺伝研	3人	進捗報告のためのミーティング
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
2019年 10月31日	チーム内ミーティング (非公開)	遺伝研	3人	ホログenom連携のためのミーティング
2020年 1月23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
2020年 1月29日	チーム内ミーティング (非公開)	遺伝研	3人	進捗報告のためのミーティング
2020年 2月5日	チーム内ミーティング (非公開)	遺伝研	3人	進捗報告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
2019年 7月31日	The 67th NIBB Conference “Quest for Orthol	岡崎コンファレンスセ	50人	オーソログ推定手法やデータベースの開発者と利用者が2年に1度集まって開

年月日	名称	場所	参加人数	目的・概要
～8月2日	ogs”	ンター		かれる国際会議。今回は、内山がオーガナイザーとして、基生研主催の NIBB conference の一環として行った。
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

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。	https://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	
3	Latent Environment Allocation	LEA	微生物環境推定ツール	http://leamicrobe.jp	2018/6/6	維持・発展	ツール等	Higashi K, Suzuki S, Kurosawa S, Mori H, Kurokawa K. Latent environment allocation of