

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)

「ゲノム・メタゲノム情報統合による
微生物 DB の超高度化推進」

研究開発期間：平成26年4月～平成29年3月

研究開発終了報告書

研究代表者：黒川 顕

(国立遺伝学研究所 生命情報研究センター 教授)

§1. 研究開発実施の概要

【目的】

ライフサイエンス研究において、知識の集積体であるデータベース(DB)は、知識の参照のみにとどまるものではなく、新たな研究分野を切り拓く上で欠く事のできない極めて重要な研究基盤である。ライフサイエンス研究の中でも、微生物研究は歴史も古く、これまで蓄積されたデータや知識は膨大かつ多岐にわたっている。さらに、昨今のゲノム科学の発展に伴い、ゲノムやメタゲノムなど圧倒的な量のデータが産出されており、これらを横断的にかつ簡便に利用できれば、新たな仮説や研究分野の創出がより容易になると期待される。そこで、申請者らは第一期統合化推進プログラムにて、国内外に散在する細菌の各種オミックス情報を広く収集し、遺伝子、ゲノム、環境の3つの軸に沿って遺伝子機能、分類学的情報、菌株保存情報、表現型情報などの知識を整理し、ゲノム情報を核としてセマンティックウェブ技術により統合した統合DB「MicrobeDB.jp」を開発してきた。第二期統合化推進プログラムにおける研究開発では、MicrobeDB.jp を細菌のみならず真菌類、藻類を対象として拡張するとともに、持続可能なシステムの構築、統合DB解析アプリケーション群(Stanza)の開発、さらには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、大規模データからの新規知識発見を容易に実現する事が可能なシステムの構築を目標とした。

【実施内容】

この目的のため、MicrobeDB.jp に対して以下の開発を実施してきた。1)各種オミックスデータへの対応、2)真菌類および藻類を対象とした拡張、3)各種オントロジー、ボキャブラリの開発、4)解析プロトコルを実装した各種 Stanza の開発、5)データの収集およびクオリティコントロール(QC)、更新の自動化など持続可能なシステムの構築、6)データ共有、公開におけるアクセスレベルの制御システムの構築、7)構築したシステムを幅広い分野の研究者に活用してもらうためのユーザビリティの向上。

【成果】

従来型のDBにおいては、単語の検索や統計量の羅列に留まっていたが、本研究開発により、多種多様なデータを横断的に利用した解析結果をユーザに提示する事が可能となった。主な研究開発成果を以下に記述する。

- 全データのRDF化に取り組み、MicrobeDB.jp ver.1(第一期)では約11億トリプルで構成されていたが、ver.2(第二期)では約90億トリプルのデータを整備、統合した。
- 公開されている173,359のメタ16S・メタゲノムサンプルに対して、開発した自動アノテーションツールにより、微生物生息環境オントロジー-MEOを付加した。
- 計算の高速化やユーザビリティの向上を目的に、SPARQLやJavascript、HTML/CSSのコードを抜本的に修正し、多様な解析Stanzaを195個開発した。
- MiGAPおよびMeGAPの解析出力をMicrobeDB.jpならびにINSDC公共DB受入システムへのスムーズな登録を支援するシステムGenomeRefineを整備した。
- 真核微生物のゲノムについては真菌類を中心に28種、菌株データについては16,671株の情報についてRDF化した。藻類に関しては、*Chlamydomonas*、*Klebsormidium*、*Nannochloropsis*を中心に、合計26種のゲノム情報から、藻類オーソログ遺伝子DBを整備した。DBCLSおよびPDBjグループとともに、単細胞藻類を対象にRDFデータモデルを構築した。
- MicrobeDB.jpに拡張OpenID認証システムを適用し、データのアクセスコントロールを可能とした。
- これまで手動でおこなっていたMEOアノテーションを、カーネルSVMを用いた教師あり学習による自動アノテーションツールを開発することで、DB自動更新において最も大きな障壁であった、MEOアノテーションが自動化され、DB更新の手間を大幅に削減する事に成功した。

§2. 研究開発実施体制

1. 研究グループ

(1) 「遺伝研」グループ 1 (研究代表者グループ)

人員構成

種別	氏名	所属機関	役職	研究開発項目	参加時期
○	黒川 顕	国立遺伝学研究所	教授	総括	H26.4～ H29.3
	山田 拓司	東京工業大学	准教授	代謝パスウェイ DBの構築	H26.4～ H29.3
	森 宙史	国立遺伝学研究所	助教	微生物統合 DBの構築	H26.4～ H29.3
*	山本 希	東京工業大学	特任助教	オントロジー開 発	H26.4～ H28.6

担当項目

1. メタゲノムDBの整備 (MG-RAST、IMG/M、JCHM)
2. 解析Stanzaの開発 (解析プロトコルの実装)
3. 真菌類データの整備 (オントロジー開発含む)
4. DB更新自動化システムの開発

(2) 「遺伝研」グループ 2 (主たる共同研究者グループ(1))

人員構成

種別	氏名	所属機関	役職	研究開発項目	参加時期
○	中村 保一	国立遺伝学研究所	教授	藻類データの 整備	H26.4～ H29.3
	神沼 英里	同上	助教	MeGAPとの連 携強化	H26.4～ H29.3
*	藤澤 貴智	同上	特任研 究員	アクセスレベ ルの制限シス テムの開発	H26.4～ H29.3

担当項目

1. 各種オミックスデータの整備
2. 藻類データの整備
3. MiGAPおよびMeGAPとの連携強化
4. アクセスレベルの制限システムの構築

(3) 基生研グループ (主たる共同研究者グループ(2))

人員構成

種別	氏名	所属機関	役職	研究開発項目	参加時期
○	内山 郁夫	基礎生物学研究所	助教	総括、システム 開発	H26.4～ H29.3
*	千葉 啓和	同上	専門研 究員	オーソログ分類 の精密化・オー ソログ情報の RDF化	H26.4～ H29.3

西出 浩世	同上	技術職 員	システム開発 補助	H26.4～ H29.3
-------	----	----------	--------------	-----------------

担当項目

- (1) 真核生物に対するオーソログ解析手法の開発
- (2) 真核生物を含むドラフトゲノムのオーソログ解析への活用
- (3) オーソログを基軸とした各種データの統合の推進

2. 有識者会議等

特定の外部有識者会議などは設置しておらず、学会および展示会等にて広く意見交換を行っている。

§3. 研究開発の目的、実施内容及び成果

1. 研究開発の背景

【本研究開発を実施するに至った経緯】

微生物研究の歴史は古く、これまでに蓄積された知識は、分類学的情報、菌株保存情報にはじまり、各種オミックス情報、メタゲノム情報など多様かつ膨大なものとなっている。しかし、もともと微生物研究においては、研究者の分野特異性が高く、自らの研究対象外に興味の範囲を拡大する機会は稀であったため、ゲノム情報の蓄積が進んだ後においても、通常の微生物学者がこれら膨大な情報を横断的に統合し活用する動きはあまり一般的ではなかった。しかし、新型シーケンサーに代表される実験機器類の飛躍的なハイスループット化により、ゲノムサイズが小さい微生物は比較的容易にゲノム解析が可能であることから、これまでゲノム研究と無関係であった微生物学研究者が自らゲノム解析を行うケースも増えてきており、状況が加速度的に変化している。すなわち、これまで1種類のゲノム解析にとどまっていた研究が、一度に複数種のゲノムを解読し、比較ゲノム解析やトランスクリプトーム解析、さらにはメタボローム解析などを複合的に活用し新たな知見を得るといふ、データ駆動型の大規模な研究が実現しつつある。さらに、環境中の微生物集団を丸ごとゲノム解析するメタゲノム解析も実現し、海洋、土壌などの自然環境だけでなく、ヒトの口腔内、腸内、皮膚に存在する微生物集団を対象としたヒトメタゲノム解析も急速な勢いで研究が進んでいる。さらに、USBメモリサイズのシーケンサーが登場するに至って、今後多様な産業分野における新たな微生物ゲノム情報の応用も加速度的に推進されると容易に想像できる。

しかしながら、これらのデータは、微生物学に革新的な進展をもたらす可能性を有しているにもかかわらず、この爆発的に生み出される情報を個々の微生物研究者やコミュニティが自らの研究のために活用することはますます困難になってきている。これはまさに DB システムの問題であり、効果的な統合 DB の構築によって、この状況を改善し大きな進展をもたらすことが期待できる。実際、様々なデータが横断的に統合され容易に利用可能な環境があれば、例えば、これまで病原菌にしか興味が無い研究者も、分野横断的な検索によって、研究対象の微生物が、他の環境中ではどのような挙動を示すのかといった幅広い情報を得る事ができ、その結果、従来の常識にとらわれない画期的な発見をもたらす可能性がある。

これら背景の元に、我々は第一期統合化推進プログラムにて、微生物統合 DB「MicrobeDB.jp」を開発してきた。MicrobeDB.jp では、ゲノム情報を核として様々な情報をセマンティックウェブ技術により統合した、約 11 億トリプルで構成される Full RDF な DB を開発するとともに、6 種類のオントロジー&ボキャブラリおよび 110 種類の Stanza を開発・実装し、世界に類を見ない DB となっている。環境データを主軸の一つとして DB を統合した事により、微生物研究者のみならず、臨床医学、地球惑星科学、建築など異分野の研究者からも利用されつつある。さらに、産業界においても、製薬、農業などバイオ産業を筆頭に、住宅、化学、分析、商社などの多様な業種からも関心を持たれており、様々な発展的要望も寄せられるようになってきている。しかしながら第一期では、様々なデータの統合は達成したものの、利用者の多様な要望に十分応えられるまでには至っていなかった。利用者の利便性を向上するためには、爆発的に生み出されるデータに対応し、データの充実を図るとともに、統合されたデータを利用者が効果的に活用するための手段を提供することが必要である。

【類似または関連する DB】

データ生産者のメタゲノム配列データを解析する機能を持った DB である MG-RAST や IMG/M などが、特に情報解析技術や計算機資源が無い研究者の間でよく使われるようになってきており、既に MG-RAST では 276,777 サンプル、129.69 Tbp のメタゲノムデータが登録されている(2017 年 1 月時点)。しかしながら、MG-RAST、IMG/M 共に、これらはあくまでメタゲノムの DB であり、RefSeq や INSDC 等に収録されているゲノムデータや系統データ、保存株リソースデータとは一切連携がない。また、両 DB 中の大部分のサンプルは、配列データが未だ公開されていない。

【背景状況の変化】

本研究開発開始当初より、メタゲノムデータの加速度的増大は予想していた。しかし、メタゲノム解析が世界中で一般化することで想定以上のペースでデータが産出されるようになった。第二期

においてMicrobeDB.jpのメジャーバージョンアップを予定していたが、既存の計算機群ではそれら膨大な情報を運用する事が困難であることがわかった。そこで第二期初年度にはストレージを、最終年度にサーバを増強する許可を頂き、MicrobeDB.jpの運用に支障が出ないよう対処した。

2. 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		微生物に関する様々な情報をゲノム情報を核として Semantic Web 技術を用いて統合した微生物統合 DB。
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 DB。

上記以外のもの

正式名称	略称	概要
CyanoBase		財団法人かずさ DNA 研究所で全ゲノム配列決定したシアノバクテリアを含む、シアノバクテリア、緑色硫黄細菌、紅色非硫黄細菌の全ゲノム配列解析結果を収録。ゲノム配列、染色体マップ、遺伝子アノテーション、遺伝子機能カテゴリ、文献、公共 DB のリンク等が閲覧可能。
RhizoBase		財団法人かずさ DNA 研究所で全ゲノム配列決定したミヤコグサ根粒菌 <i>Mesorhizobium loti</i> MAFF303099 を含む根粒菌の全ゲノム配列情報を収録。種毎に、ゲノム配列、染色体マップ、遺伝子アノテーション、遺伝子機能カテゴリ、文献、公共 DB のリンク等が閲覧可能。

(2) ツール等

正式名称	略称	概要
MCCV		JCM や NBRC などの、菌株保存機関に蓄積されている菌株情報を記述するための統制語彙
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
PDO-CSSO		微生物がヒトに引き起こす感染症の病名を、ヒトの体の組織ごとにまとめて記述・整理した PDO と、感染症が引き起こす症状について整理した CSSO の二つのオントロジー
TogoAnnotator		ソーシャルブックマークによる文献情報集積プラットフォーム。
GenomeRefine		微生物ゲノムアノテーションの高品質化および解析支援のためのウェブサービス。ユーザ認証機能、メタデータ入力機能、解析パイプライン MiGAP、MeGAP の解析結果の登録機能、RDF 形式変換機能をもち、MicrobeDB.jp で参照可能。

※データベース、ツールの詳細は別紙参照。

3. 達成目標及び実施計画

(1) 当初の実実施計画・達成目標

本研究開発では、第一期統合化推進プログラムで開発してきた MicrobeDB.jp を基盤として、1) 各種オミックスデータへの対応、2) 真菌類および藻類を対象とした拡張、3) 各種オントロジー、ボキャブラリの開発、4) 解析プロトコルを実装した各種 Stanza の開発、5) データの収集および QC、更新の自動化など持続可能なシステムの構築、6) データ共有、公開におけるアクセスレベルの制御システムの構築、7) 構築したシステムを幅広い分野の研究者に活用してもらうためのユーザビリティの向上。以下、各実施項目の詳細を述べる。

1: 各種オミックスデータへの対応

近年の測定技術の劇的な高速化と低廉化により、微生物研究分野でも、環境の異なる条件下などでの大規模転写比較情報や翻訳産物インタラクション等、種々の観点からの各種オミックス情報が広く計測されるようになってきている。しかしながら、これまでの微生物ゲノム情報 DB では、そのような各種オミックスデータを統合的に整理し提供するための枠組みの整備が遅れている。本研究開発では、公開 DB および論文として発表されたが DB に収録されず、時間とともに失われる可能性の高いオミックス研究情報を集約し利用に供するために、オミックスデータを対象としたシステム整備と速やかな収集およびそのセマンティックリソース化を実施する(3年間継続)。

2: 真菌類および藻類を対象とした拡張

真菌類および藻類のデータに関しては、研究者コミュニティが充実している単細胞藻類および酵母、麹菌を対象として、我々がこれまで細菌のデータ整備で培ってきた技術と経験を生かし、データの整理、各種オントロジーの開発を経て、全データを RDF 形式で記述する事により MicrobeDB.jp 内に蓄積していく。これら生物種の遺伝子情報を統合するために、これら生物種を対象として新たにオーソログ遺伝子解析を実施する(3年間継続)。

3: 各種オントロジー、ボキャブラリの開発

上述した真菌類および藻類のデータを MicrobeDB.jp に蓄積し、比較解析やより有効な検索結果を得るためには、新たにオントロジーやボキャブラリを開発する必要がある。真菌類であれば、主要な研究課題である抗生物質に焦点をあてたオントロジーが必要であり、また藻類であれば、工業的利用や進化研究の観点から、脂質に焦点をあてたオントロジーの開発が必要となる。我々は、これまで多数のオントロジー&ボキャブラリを開発してきており、本研究開発においても引き続き、研究者コミュニティと連携しつつ開発を実施していく(3年間継続)。

4: 解析プロトコルを実装した各種 Stanza の開発

多種多様な情報が混在しているゲノムやメタゲノム等の複雑なデータから知識発見をするためには、データを要約したゲノムサイズや遺伝子数等の単純な統計量の提示だけでは不十分であり、それらのデータを用いた解析に習熟した研究者がその分野の最新の解析手法を用いて解析した様々な結果を、順序立って提供することが重要である。MicrobeDB.jp の能力を最大限生かす事が可能な、比較ゲノム解析やメタゲノム解析など様々な解析 Stanza を開発する(3年間継続)。

5: データの収集および QC、更新の自動化など持続可能なシステムの構築

DBは開発のみならず、その更新や維持管理する事が最重要課題である。更新や維持管理のコストを可能な限り低く抑制し、持続的な DB の維持を行うためには、データ収集や価値の付加、さらにはデータ更新を可能な限り自動化する必要がある。そこで本研究開発では、データ生産者から継続的にデータを受け付ける窓口のシステムとして、微生物ゲノム自動アンノテーションシステム「MiGAP」およびメタゲノム解析パイプライン「MeGAP」を利用し、

MicrobeDB.jp と一体運用を実現する。この事で、データ生産者に対して積極的なデータ登録を促す事ができる。また、これまで手作業で実施してきた DB の更新作業を可能な限り自動化し更新体制を強化する(3 年間継続)。

6: データ共有、公開におけるアクセスレベルの制御システムの構築

生命科学分野において特に配列情報等は、知識発見と論文発表が完了されるまでの間、解析者のコントロール下に置かれることが強く希望される傾向にある。また、大規模なゲノム・メタゲノム研究は単独の研究者ではなく広域に散在する研究者グループにより実施されることが多い。MicrobeDB.jp 上でアクセスコントロール環境を提供できれば、制限条件下での配列登録や解析が可能となり、利用と登録が促進され、論文公開後には速やかに配列ならびに解析情報を公開し利用に供することが期待できる。我々は、OpenID を拡張しグループの概念を付与した認証システムを開発しており、この認証システムをさらに拡充し、MicrobeDB.jp 全域に適用する開発を実施することで、グループ単位でのアクセスレベルコントロールを実現し、配列解析者の利便性を増すと同時に、配列公開の促進を図って行く(3 年間継続)。

7: 幅広い分野の研究者に活用してもらうためのユーザビリティの向上

MicrobeDB.jp では、潜在的なイノベータにも広く利用を促すために、これまでの生命系 DB とは全く異なる Google 様の簡潔なユーザインターフェースを採用している。しかし、専門用語を熟知している生命系研究者からは、検索語から想起される検索結果をピンポイントで提示できていないなどの意見も上がっている。そこで、検索語と各種 Stanza との対応関係を明確にし、検索語による Stanza の自動選択システムを開発する事で、検索システムを向上させる(3 年間継続)。

(2) 期間中に追加・削除・変更した実施計画・達成目標

【遺伝研グループ 1】

当初、MG-RAST や IMG/M に集積されているデータも取得し、MicrobeDB.jp に収録する計画にあった。MG-RAST および IMG/M で配列が公開されているサンプルの多くは、既に DRA/ERA/SRA でも配列が公開されているものも多く、逆に、メタデータのみ公開、またはメタデータも配列も公開されていないサンプルは、DRA/ERA/SRA でも配列が未公開の場合がほとんどであった。DRA/ERA/SRA からは配列が公開されていないがそれらの DB では配列が公開されているサンプルに絞って配列データの取得を試みたが、両 DB では多数のサンプルの配列データを一括で取得できる ftp サイト等が用意されておらず、サンプルごとにほぼ手作業で配列データを取得する必要があったため、手間と比較してデータの網羅性を上げる上での効果が少ないと判断したため、両 DB からの配列の取得は断念した。

一方、サンプルごとの環境情報の記述を含むメタデータについては、両 DB 共全サンプル一括で取得できたため、将来的にこれらの DB から配列データが取得できた場合に MicrobeDB.jp 上で DRA/ERA/SRA 由来のサンプルとスムーズな統合化をするために、微生物の生息環境に関する語彙をマニュアルで抽出し、MEO に適宜語彙を追加した。

【遺伝研グループ 2】

国内外のシアノバクテリア研究者コミュニティからの要望を受け、H28 年度において、CyanoBase への完全長およびドラフトゲノ情報のデータ拡充を計画し、337 生物種の追加を実施した。RDF および TogoStanza を利用する事でデータベースに呈示し、CyanoBase 20 周年に成果を報告した。

【基生研グループ】

当初、「真核生物を中心としてドメイン単位のオーソログ分類の精度に関する課題を検討して、さらなる改良を行う」計画にあったが、我々がを行っている「ドメイン単位のオーソログ分類」と

いう概念自体がユニークであって、オーソログ DB の国際的コミュニティの中においてさえ十分に浸透していないという事情が判明した(ほとんどの研究者はオーソログの対応付けに際してドメインを考慮しておらず、考慮している研究者は Pfam などの既存ドメイン DB に基づいて分類を行っている)ことから、まずは我々が定義したオーソロジドメインに対して、その特徴付けを行い、有用性を示していくことを優先すべきであるとの結論に達した。このため、上記の課題を代えて、「原核生物を中心としてオーソロジドメインの特徴付けを行う」という課題として、これを MBGD に基づくデータ解析のケーススタディの一環として行う方針に変更した。

また、真核微生物のゲノムデータの取り込み手順についても、NCBI のゲノムデータの公開方式が大きく変化したのに伴って大幅な修正を行ったが、その際に MBGD に取り込むデータのクオリティなどの基準についても再検討を行った。

4. 実施内容

(1) 実施内容

【遺伝研グループ 1】

研究開発実施項目(1):メタゲノム DB の整備

15 万サンプル以上の公共のメタ 16S・メタゲノムデータからメタゲノム配列データと付随するメタデータを取得し、メタデータは第一期で開発した MEO や MSV 等のオントロジーを用いてアノテーションし、メタ 16S・メタゲノム配列データは配列相同性検索によって系統や遺伝子機能をアノテーションした。メタ 16S・メタゲノムサンプル共に同じ基準で系統組成を推定する解析手法として、高精度な 16S rRNA 遺伝子配列 DB に対しての、GPU を用いた配列相同性検索による高速な系統組成推定手法を開発した。メタゲノムサンプルごとの遺伝子機能組成については、メタゲノム塩基配列データから遺伝子予測を行いアミノ酸配列データへと変換した後に、RefSeq の微生物ゲノム由来アミノ酸配列 DB に対して、BLAST の約 150 倍高速な配列相同性検索ツールである GHOSTX を用いて配列相同性検索を行い、その結果を用いて MBGD 及び KEGG に基づいたオーソログ組成を推定する解析手法を開発した。両解析手法を用いて遺伝研スーパーコンピュータで公共の全メタ 16S・メタゲノムサンプルの塩基配列データに対して計算を行い、計算結果から RDF データを作成した。クオリティフィルタリングを通過して系統組成を得ることができたメタ 16S・メタゲノムサンプル数は 60,551 サンプル、同様に遺伝子機能組成を得ることができたメタゲノムサンプル数は 4,048 サンプルであった。

公開メタ 16S・メタゲノムサンプルの分離源の環境情報を微生物生息環境オントロジーである MEO を用いて、後述する自動アノテーションツールを用いてアノテーションを行い、さらにその後マニュアルでアノテーション結果をチェックして、173,359 サンプルについて MEO アノテーションを付加した。また、メタゲノム解析関連のメタデータを記述するためのボキャブラリである MSV を用いて、温度や pH、溶存酸素濃度などの様々なメタデータについてマニュアルアノテーションを行い、153,879 サンプルについて MSV アノテーションを付加した。

研究開発実施項目(2):解析 Stanza の開発

RDF 化した微生物に関する様々なデータから、研究者が新たな仮説を発見しやすくすることを目的に、MEO による微生物の生息環境の分類を用いて、同一または近隣の MEO がアノテーションされているメタ 16S・メタゲノムサンプルの系統組成や遺伝子機能組成をヒートマップや階層的クラスタリング、多次元尺度構成法等の手法で比較解析できる Stanza や、特定の目的、手法でサンプリングされ、メタデータ記述もサンプル間で統一されているため、有意義な比較解析が可能な、同一プロジェクト内のサンプル間に絞って系統組成や遺伝子機能組成を比較解析できる Stanza など、目的に応じた様々な解析 Stanza を開発した。また、第一期で開発して既に存在する Stanza についても計算の高速化やユーザビリティの向上を目的に大規模に SPARQL や Javascript、HTML/CSS のコードを修正し、結果として本研究開発期間中に 195 個の Stanza を新たに開発または修正した。

研究開発実施項目(3):真菌類データの整備

原核・真核微生物の両方に共通したゲノムおよび RNA-Seq の RDF モデルを開発した。また、DBCLS と共同で開発している菌株保存機関の菌株関連ポキャブラリである MCCV について、菌株保存機関に存在する真核微生物の表現型情報や培養条件等の記述に対応するために、語彙の追加、語彙間の関係性の記述の修正等を行った。結果として、真菌類を中心に、真核微生物のゲノムについては 28 種、菌株データについては 16,671 株の情報について RDF 化することができた。

研究開発実施項目(4):DB 更新自動化システムの開発

MicrobeDB.jp の DB 更新自動化において、最も大きな障害は、遺伝子・系統・環境の 3 つの軸の一つである環境情報を記述する中心となる MEO を、メタゲノムやゲノム、菌株のメタデータを元に個々にアノテーションする作業である。第一期ではマニュアルで行っていたこの作業を自動化するために、第一期でマニュアルにて MEO をアノテーションした結果を用いて、カーネル SVM を用いた教師あり学習による MEO の自動アノテーションを行うツールを開発した。教師データに用いていない、菌株保存機関の菌株の分離源のメタデータと MEO とのマニュアルマッピングデータを用いた精度評価の結果からは、教師データに全く含まれていない新規なメタデータには弱い、そうでない場合は結果が十分実用に耐えうる精度があることを確認した。このツールを用いてゲノム・メタゲノム・菌株データについて MEO の自動アノテーションを行う事で、オントロジーアノテーションにかかる手間を大幅に削減することができた。

【遺伝研グループ 2】

研究開発実施項目(1):各種オミックスデータの整備

H26 年度は収集に関するシステム整備を実施し、コミュニティとの連携をもとに主要モデル微生物のデータ収集を開始した。また、収集したデータについては、これまでに開発したゲノム情報 RDF 等とのデータ統合し、SPARQL による検索性を高めるために RDF データモデルを設計し、RDF 化の開発に着手した。H27 年度は、ゲノムプロジェクトリストの拡張を実施し、RNA-Seq 解析の投入フロー整備ならびに対象となるデータ収集を継続推進した。H28 年度ではそれまでに作成した NCBI Genome Assembly (<http://www.ncbi.nlm.nih.gov/assembly>)由来のゲノムプロジェクトリストの更新系を整備するとともに、DDBJ が提供する各種リソースとの連携を実施、RNA-Seq の解析支援のための RDF データを作成した。また、遺伝研グループ 1 と協力し、MicrobeDB.jp のゲノム RDF 生成の自動化と、更新系の整備を実施した。この開発の実施により、RDF の活用によるゲノム以外の各種オミックスデータの取得と MicrobeDB.jp への取り込みが自動化された。

研究開発実施項目(2):藻類データの整備

H27 年度までにライフサイエンス統合データベースセンターと連携して進めてきた RDF によるゲノム基盤情報のデータモデルを単細胞藻類用途に拡張した。H28 年度はすでに紅藻類であるスサビノリや緑藻類のクラミドモナスの発現・リソース情報等を収集してきている植物統合課題の開発 DB である PGDBj (<http://pgdbj.jp>) と連携し、これまで相互に構築をすすめてきたオントロジー体系の共用化や、不足する語彙群の追加構築等を実施し、生物種群別の統合化推進プログラムの情報を繋ぐための基盤情報整備を実施することにより情報基盤の拡張を完了した。

研究開発実施項目(3):MiGAP および MeGAP との連携強化

本グループが開発した GenomeRefine (<http://genome.annotation.jp/genomerefine/>) は微生物ゲノムアノテーションの高品質化および解析支援のためのウェブサービスであり、これをセマンティックウェブ技術およびリソースを用いて機能強化し MiGAP との連携を実現す

ることで、MiGAP の解析出力を MicrobeDB.jp ならびに INSDC 公共 DB 受入システムへのスムーズな登録を支援するシステムとして整備した。それによって、配列決定され粗解析が終了しているものの DB に登録されていない微生物ゲノム情報の掘り起こしと公開の促進を図った。また、ゲノムのみならずメタゲノムの自動解析と公開支援のための拡張ならびにシステムの統合を実施しメタゲノムデータに対しても積極的なデータ公開・再利用の促進を図ることができた。H26 年度は GenomeRefine のセマンティックウェブ技術での強化に着手し、これまで実装した機能のセマンティックウェブ技術を利用した再実装をすすめると同時に DDBJ の提供する RDF 等、外部 DB との共用可能なリソースとの連動を整備した。H27 年度は引き続きメタデータ入力支援機能の強化と、INSDC/DDBJ リソースの OWL の共通化の推進、RDF データモデルの共通利用化を進めた。H28 年度は Genome Refine の MiGAP との API によるデータ送受信系の確立による DB 登録支援システムの拡張と、DDBJ スパコンで稼働している公共配列解析パイプラインである DDBJ Pipeline 上で MeGAP を実施可能とする開発を実施するとともに Genome Refine と連携させる開発を実施することで、MiGAP/MeGAP 出力の MicrobeDB.jp ならびに DDBJ への登録支援システム整備を完了した。

研究開発実施項目(4): アクセスレベルの制限システムの構築

H26 年度と H27 年度に MicrobeDB.jp 上の各種認証機能に対して拡張 OpenID 認証システムの適用を行うとともに、API 整備とドキュメント整備を実施し、当時利用されていた OpenID 認証機能を本拡張による認証に置き換え、解析系サービスをシームレスに連携させ得る認証システムの拡張を実施した。H28 年度は開発した認証システムの安定運用のための開発を実施するとともに、データのアクセスコントロール方法としてどのような内容が必要となるかを研究者コミュニティからの要望を聞き取り、開発を完了した。

【基生研グループ】

研究開発実施項目(1): 標準オーソログテーブルの精密化とその活用

MBGD でオーソログ解析に用いている DomClust は、遺伝子やドメインの融合や分裂が起きた場合に正しく対応づけができるように、ドメインの単位でオーソログ分類を行っているが、ペアワイズアライメントに基づいてドメイン分割を行っているために、アライメントの精度が低い場合に正しくドメイン分割されないケースが生じていた。そこで、より精度の高いマルチプルアライメントに基づいてドメイン分割を行うことにより、分類の精度を改善する DomRefine プログラムを開発した。DomRefine の基本的なアルゴリズムの開発と実装、およびテストデータを用いた評価は、前回の開発期間中にほぼ完了していたが、実際の大規模なデータに適用するには処理速度の点で問題があった。特に、処理の過程でマルチプルアライメントを繰り返し計算するのに膨大な時間がかかっていたため、中間結果のアライメントを保存しておいて、再利用が可能なケースでは計算を省くようにするなどの工夫をすることにより、処理を高速化した。その上で、これを MBGD の更新プロセスに組み込んで、MBGD の標準オーソログテーブルの精密化を行うようにした。DomRefine の副産物として、それぞれのオーソロググループについてマルチプルアライメントや系統樹の情報も計算されるので、これらを保存しておいて利用者が取得できるようにした。さらに、各オーソロググループのマルチプルアライメントに基づいてプロファイル隠れマルコフモデルを作成し、HMMER を用いて利用者のクエリ配列に対して対応するオーソロググループを検索できるようにした。

一方、このようにして構築したドメイン単位のオーソロググループは、今日広く用いられているドメイン DB である Pfam などとは違って、ホモロジーではなくオーソロジーに基づいてドメインを定義している。その意味で種分化と直結したドメイン構造の進化などを解析するのにより適していると考えられ、これまでにはないユニークなデータといえる。このデータを微生物ゲノム比較に効果的に活用するために、我々自身がこの DB を用いた解析を行って、ある程度その性質を明らかにする必要があると考えた。そこで、ドメイン融合遺伝子に基づいて、各オーソログドメインをノードとし(全部で 42,233 ノード)、それらの間の隣接関係をエッジとして定義

したオーソログドメインネットワークを構築し、各ノードについて遺伝子機能や生物系統などの観点から特徴付けを行って、次数の高さとの関連などを調べた。その結果、二次代謝関連遺伝子のオーソログドメインは比較的平均次数が高く、翻訳関連遺伝子では低いといった結果が得られ、遺伝子機能によって、ドメイン融合に対する進化的制約に違いがあることが示唆された。これらの成果は MBGD を用いたデータ解析のケーススタディの一つとして論文発表するべく準備を進めているが、合わせてオーソログドメインネットワークを対象とした検索など、今後 DB の検索機能の方にも反映させていきたいと考えている。

研究開発実施項目(2): データ取り込み手順の再構築

MBGD では従来、主に NCBI の RefSeq および GenBank において genomes セクションとして公開されていた完全ゲノムデータを取り込んできたが、NCBI の方で完全ゲノムとドラフトゲノム、原核生物と真核生物とを合わせて、すべて Assembly report を介して公開する体制への移行があり、それに合わせて我々のデータの取得・構築プロセスの全面的な修正を行った。Assembly report の中でゲノムの完成度に従った分類が行われているので、それを参考にしつつ、独自の基準も加えて、完成度の高さによって完全ゲノム、不完全ゲノム、ドラフトの3つに分けた。このうち、従来の MBGD では完成度の高い完全ゲノムのみを解析の対象としてきたが、近年完成度の低い不完全ゲノムや、染色体としてまとめられていないドラフトの状態で開催されるケースが増えており、それらも比較解析の対象として取り込むことができるようなデータ処理方式を構築した。

MBGD では、属(genus)当たり1つを代表にとった生物種セットで作成したオーソログテーブルをデフォルト(標準オーソログテーブル)として使っている。この標準オーソログテーブルに加えるのは完全ゲノムデータのみとするが、それ以外の不完全ゲノム全部と、ドラフトゲノムの中で一定のクオリティ基準を満たし、かつ標準オーソログテーブルに含まれない属に属するものを MBGD に追加し、標準オーソログテーブルを拡張した draft-plus テーブルとして利用できるようにした。それ以外のドラフトゲノムについては、DB には取り込むが事前のホモロジー計算の対象にはせず、利用者が MyMBGD 機能を通じて動的に解析に加えることができるようにした。ドラフトゲノムを取り込む方針については、前回の開発期間に仕様を作成して構築を進めていたが、利用者インターフェイスが整わず、公開までには至っていなかった。平成26年度に MyMBGD インターフェイスの全面的な改定を行い、これらの機能を利用できるようにして公開した。

なお、開発期間内に公開予定の最新版では、完全ゲノム数が 6,318、ドラフトゲノム数は 58,547 となっている。開発期間前は完全ゲノム数が 2,749(ドラフトはなし)であったので、倍以上に増加したことになる。

研究開発実施項目(3): 種内・属内オーソログテーブルと標準オーソログテーブルの統合

上述の通り、従来の MBGD では、ゲノム間の総当たりのホモロジー検索結果に基づいて、属当たり1つの代表生物種をとって比較した「標準オーソログテーブル」に加えて、種内や属内についてそれぞれオーソログ解析を行った「系統特異的オーソログテーブル」を作成して提供してきた。しかし、ゲノム数の急激な増加によって総当たりのホモロジー検索結果の維持が困難になってきており、また上記手順では種内や属内で保存されておらず、代表生物種に含まれない遺伝子が、全体比較の対照から漏れるという問題が残っていた。これらの問題を解決するため、種内・属内オーソログテーブルと、属間の比較である標準オーソログテーブルの作成プロセスを統合し、段階的に構成することによって、漏れない標準オーソログテーブルを効率的に作成する手続きを作成した。具体的にはまず種内オーソログ解析を行って、得られた遺伝子レパートリー(パンゲノム)を「種ゲノム」とし、これらを用いて属内オーソログ解析を行って得られたパンゲノムを「属ゲノム」とし、さらにこれらを用いて標準オーソログテーブルの作成を行う。種内・属内の総当たり比較には、超高速のホモロジー検索プログラムである UBLAST (USEARCH コマンドの-ublast オプション)を用いることによって高速化し、従来の

BLAST および Smith-Waterman 法を用いた検索は、データ数の伸びが比較的穏やかな属間比較に限ることとした。これにより、種内、属内のゲノム多様性の情報を維持しつつ、ホモロジー検索の対象となる遺伝子数を減らして更新処理を効率化するとともに、公開されたゲノム全体を使った漏れのない標準オーソログテーブルを作成することが可能となった。

研究開発実施項目(4):オーソログ情報の RDF 化

オーソログ情報は、様々な生物種において蓄積した遺伝子機能や表現型などの知識を相互に結びつけるための中枢として機能する。前回の開発期間からの継続課題として、セマンティックウェブ技術を用いて MicrobeDB.jp 内のデータやその他の DB と連携して MBGD のオーソログ情報を検索するためのオーソログ情報の RDF 化を進めた。この目的のため、まずオーソログの概念を定義した汎用のオントロジー Ortholog Ontology (OrthO)を開発し、これを用いて MBGD のオーソログ情報を RDF 化した³⁾。OrthO は汎用のオントロジーなので、eggNOG など他のオーソログ DB も OrthO を用いて RDF 化し、相互に比較することが可能である。作成したデータを検索するための SPARQL エンドポイントを立ち上げ、検索インターフェイスを備えたポータルサイト(MBGD SPARQL Search)を構築して、UniProt や Taxonomy DB などと統合的に検索できるようにした。SPARQL に不慣れな利用者でも検索できるように、典型的なクエリをあらかじめ多数用意し、パラメータを埋めて実行するだけで SPARQL クエリを構築して検索できるインターフェイスを用意した。

研究開発実施項目(5):オーソロジーオントロジーの統合

オーソロジーに対するオントロジーとしては、上述の我々が開発した OrthO の他に、Murcia 大の Fernández-Breis らが開発した OGO などが存在する。そこで、Fernández-Breis らと共同で、これらを統合しつつ、既存の利用可能な他のオントロジーとの関連も考慮して、より完成度の高い新たなオントロジー Orthology Ontology (ORTH; <https://github.com/qfo/OrthologyOntology>)を開発した。ORTH オントロジーのコアの部分は 21 クラスおよび 19 プロパティであり、その他にインポートしたオントロジーも含めると 4,613 クラスおよび 821 プロパティとなった。ORTH を用いたアプリケーションとして、オーソログ DB のコミュニティで用いられている、XML によるオーソログデータの標準形式 OrthoXML から、ORTH を用いて RDF データへ自動変換を行ったり、ORTH を使って複数のオーソログ DB に対する統合的な検索を行ったりする例について検討した。

研究開発実施項目(6):オーソログデータ検索のための Stanza の作成

オーソログデータを利用した解析結果の可視化を行う Stanza を作成し、Stanza ライブラリとして蓄積する体制を作った。基盤技術としては、DBCLS によって開発された TogoStanza フレームワークを用いた。DB への問い合わせ部分は SPARQL を用いて開発し、また必要に応じて JavaScript を用いて Stanza における可視化部分を開発した。作成した Stanza は MicrobeDB.jp および MBGD のサイトに埋め込み動作確認を行った。利用可能となった Stanza のリストを MBGD Stanza Catalog としてまとめた (<http://mbgd.genome.ad.jp/stanza/>)。このリストには、オーソロググループに関連づけられる Stanza (7 種)、生物または系統群に関連づけられる Stanza (4 種)、系統群を比較するための Stanza (7 種)が含まれている。オーソログデータには、利用目的に応じて複数のオーソログテーブルがあるため、これらのテーブルをアプリケーションによって使い分けられるように、対象データセットを Stanza 上でパラメータ指定できるようにした。

(2) データベースの利便性に関する利用者ニーズと具体的な対応

【利用者からの要望①:検索】

- 同一プロジェクト内のメタゲノムサンプルを比較解析したい。

- メタゲノムサンプルや系統の ID で検索できるようにしてほしい。
- キーワード検索で部分的に一致した各遺伝子機能名 (polymerase で検索した際の DNA polymerase subunit alpha や beta など) の一致件数も同時に表示してほしい。

【対応内容】

検索機能および該当する Stanza を実装することで対応完了した。

【利用者からの要望②:機能】

- 系統特異的なオーソログを調べたい。
- 各 MEO にアノテーションされているメタゲノムサンプルの数、さらには登録された年ごとの数を知りたい。

【対応内容】

統計情報解析 Stanza および該当する解析 Stanza を実装することで対応完了した。また、複雑な条件でのオーソログ検索に関しては、SPARQL インターフェイスを実装する事で可能とした。

【利用者からの要望③:その他】

- ガン患者由来の公開メタゲノムデータと、健常人の公開メタゲノムデータを比較したい。

【対応内容】

ガンは基本的に感染症ではないので、感染症オントロジーである PDO には収録されていない。かつ、現状のメタゲノム研究では宿主の詳細なメタデータは、DRA/ERA/SRA 等の公共の塩基配列 DB 中には、記述が必須ではないため、現状の集めたデータのみからでは、対応することが困難であり、今後の課題である。

【利用者からの要望④:その他】

- CyanoBase へのゲノム登録数を増やして欲しい。

【対応内容】

2016年10月に39株から376株への拡充を実施した。

(3) 持続的なデータベース運用体制の構築に向けた取り組み

- ゲノム・メタゲノムの解析パイプラインである MiGAP/MeGAP と、メタデータ入力・RDF 変換 Web アプリケーションである GenomeRefine を用いて、ユーザからのデータ収集・メタデータ記述の統一を自動化した。
- カーネル SVM を用いた教師あり学習による MEO の自動アノテーションツールを開発し、オントロジーアノテーションの手間を大幅に削減した。
- シアノバクテリア研究者コミュニティと連携し、本課題での開発物を活用した *Synechocystis* sp. PCC 6803 株のスクラッチからの再アノテーションを実施した。
- MBGD では、急速に増加するゲノムデータに対応するために、ホモロジー検索とオーソログ解析を、種内、属内、属間と段階的に行う手順を開発した。これによって今後の更新に必要なホモロジー計算の量を抑えることができ、当面の運用可能性を確保できたと考えている。

(4) 統合化推進プログラムの他のチームや DBCLS との連携

- ゲノム RDF を原核微生物と真核微生物両方に対応するために、DBCLS と連携して新たな RDF 構造を設計・開発した。
- 菌株保存機関の菌株関連ポキャブラリである MCCV について、菌株保存機関に存在する真核微生物の表現型情報や培養条件等の記述に対応するために、語彙の追加、語彙間

の関係性の記述の修正等を DBCLS、さらには他の統合化推進プログラムの理化学研究所榭谷グループと連携して行った。

- 菌株保存機関の一つである理研 JCM のデータの RDF 化については、第一期に我々のグループで開発した菌株の RDF モデルを基本として、理研内のバイオリソースデータの RDF 化を主に行う榭谷グループと連携して最新版のデータの RDF 化を行い、両グループで RDF データを共有した。
- DBCLS および植物統合グループと連携し、共通のゲノム RDF の利用およびシアノバクテリアおよび藻類について、FALDO によるゲノム座標系を記述した RDF データの jBrowse 上でのデータ統合について連携。
- ゲノムアノテーション支援ツール TogoAnnotator とそのサービス化について DBCLS と連携して開発を継続中。
- DBCLS 主催の毎月行われている SPARQLthon や、毎年行われている国内版・国際版 BioHackathon に毎回参加し、DB 開発の過程で生じた問題点を DBCLS や他の統合化推進プログラムのグループと共有して協議しつつ、開発を進めた。

(5) データ産出を行う研究組織や研究室、プロジェクトとの連携

- 新学術領域研究「ゲノム支援 (H22～27 年度)」および「先進ゲノム支援 (H28～33 年度)」: 多数の研究者らと微生物のゲノム・メタゲノム研究に関して共同研究を実施している。ここで得られたデータは DDBJ に登録するとともに、統一的なプロトコルで解析した 2 次データは MicrobeDB.jp 上に蓄積している。
- JST CREST「海洋生物多様性および生態系の保全・再生に資する基盤技術の創出 (H23 年度～)」では、海洋微生物のメタゲノム DB の構築に関してアドバイスしている。
- WPI 東工大地球生命研究所および新学術領域研究「冥王代生命学の創成」: 地球と生命の起源を研究する国内外の研究者とネットワークを作り、これら研究に活用可能な情報を提供している。海底熱水孔、蛇紋岩やカンラン岩などの地質学的専門用語で構成される微生物環境を表現するため MEO を改良している。
- ヒトマイクロバイオーム研究: H25 年度に東工大で立ち上げた Japanese Consortium for Human Microbiome (JCHM) において、大学法人として学界や産業界のみならず、広く一般からヒトメタゲノムデータを集積する仕組みを構築した。また、H29 年度からは 17 社の製薬企業を中心とする企業コンソーシアムが立ち上げ予定であり、日本人標準ヒトマイクロバイオームデータの整備に関して議論を続けている。
- 発酵食品 DB: H28 年 10 月に東工大で立ち上げた共同研究講座「ぐるなび食の価値創成共同研究講座」において、発酵食品に関わる麹菌ゲノムデータの整備を開始している。
- 藻類と陸上植物の中間的な存在である車軸藻植物門クレブソルミディウムのゲノム解析において、TogoAnnotation によるコミュニティゲノムアノテーションを支援した。また、シアノバクテリアゲノム解析のインフラ整備を目的として、豊橋技術科学大学および国立環境研究所と協力して、シアノバクテリアゲノム解析のインフラ整備の連携を実施した。
- オーソログ DB の提供者とその大口利用者 (UniProt, GeneOntology コンソーシアムなど) のコミュニティの間で、隔年で開催されている Quest for Orthologs 会議に出席している。前回は 2015 年にバルセロナで開催され、オーソログデータの評価法や表現法を含めて、様々な課題についての協議を行った。Murcia 大の Fernández-Breis とは前々回の本会議で知り合い、その後 BioHackathon での共同作業を通じて Ortholog Ontology の開発と OrthoXML から RDF の自動生成という共同研究につながっている。
- DBCLS が主催している BioHackathon、SPARQLthon や STANZAthon などのワークショップにも毎回欠かさずに参加し、DB の基盤技術などに関して、DB 研究者コミュニティとも連絡を密にしてきた。この事で、効果的な DB 開発や、重複の無い Stanza 開発などが実施

できている。

(6) 人材の育成

研究開発グループで雇用している研究員には、DBCLS 主催の BioHackathon や SPARQLthon、STANZAthon への積極的な出席を促し、専門家とのディスカッションを通して、高度なスキルを身につけてもらうよう心がけてきた。その結果、基生研グループの千葉研究員は、本プロジェクトで実施した研究成果などに基づいて博士号(論文博士)を取得している。また遺伝研グループ1に在籍した山本研究員は、H28 年度新設の東工大研究室にて主に発酵食品の DB 開発を担当する特任助教に着任した。

(7) その他

特になし。

§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

名称	種別	平成 25(2013) 年度	平成 26(2014) 年度	平成 27(2015) 年度	平成 28(2016) 年度
MicrobeDB.jp	訪問者数	5,712	6,673	7,489	5,196
	訪問数	15,867	24,957	16,741	7,625
	ページ数	313,272	561,522	729,093	122,093
MBGD	訪問者数	26,237	28,042	30,161	39,096
	訪問数	73,332	85,555	86,258	296,761
	ページ数	1,140,392	1,882,676	2,335,561	4,817,507

表 1-1 研究開発対象の主要なデータベースの利用状況（年度別）

名称	種別	平成 25(2013) 年度	平成 26(2014) 年度	平成 27(2015) 年度	平成 28(2016) 年度
MicrobeDB.jp	訪問者数	476	556	624	577
	訪問数	1,322	2,080	1,395	847
	ページ数	26,106	46,794	60,758	13,566
MBGD	訪問者数	2,186	2,337	2,513	4,344
	訪問数	6,111	7,130	7,188	32,973
	ページ数	95,033	156,890	194,630	535,279

表 1-2 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）

(2) 分析

- MicrobeDB.jp は、H28 年度の訪問者数、訪問数、ページ数がそれ以前よりも低い。これは、2016 年 11 月に MicrobeDB.jp ver.2 をリリースするまで、DB が公開されてから時間がかなり経っており収録していた情報が古くなっていたことと、2016 年 9 月中旬から 10 月末の間、研究代表者の異動に伴い、運用サーバを東工大から遺伝研に移設しさらには ver.2 へのアップデート作業のために、サービスを一時的に止めていたことが影響していると考えられる。ver.2 をリリース後は、2016 年 11 月のページ数が 44,143 ページ、12 月のページ数が 48,606 ページと順調に増えており、今後さらなる増加が期待できる。
- MBGD は、研究開発期間中は毎年増加傾向を示した。一時的な大量アクセスの影響も大きい一概には言えないが（特に H28 年度に訪問数とページ数が大幅に増えているのは、年度の前半に一時的な大量アクセスがあった影響が大きいと見られる）、すべての項目で一貫して増加傾向を示していることから、DB の認知度が順調に増加してアクセス数が増えていることを反映しているものと考えている。

2. データベースを利用して得られた研究成果事例

- Matsuki T, Yahagi K, Mori H, Matsumoto H, Hara T, Tajima S, Ogawa E, Kodama H, Yamamoto K, Yamada T, Matsumoto S, and Kurokawa K. A key genetic factor for fucosyllactose utilization affects infant gut microbiota development, *Nat. Commun.* 7:11939, 2016. (DOI: 10.1038/ncomms11939).
- Okai S, Usui F, Yokota S, Hori-i Y, Hasegawa M, Nakamura T, Kurosawa M, Okada S, Yamamoto K, Nishiyama E, Mori H, Yamada T, Kurokawa K, Matsumoto S, Nanno M, Naito T, Watanabe Y, Kato T, Miyauchi E, Ohno H. High-affinity monoclonal IgA regulates gut microbiota and prevents colitis in mice, *Nat. Microbiol.* 1:16103, 2016. (DOI: 10.1038/nmicrobiol.2016.103).

3. その他

特になし。

§5. 研究開発期間中に得られた科学・技術や産業に対する波及効果

新型シーケンサーの普及に伴い、微生物のゲノム・メタゲノムデータは猛烈な勢いで増大している。特にヒトマイクロバイオーム研究は、欧米を中心に加速度的に発展を続けており、潤沢な研究予算を背景に巨大なデータが産出され続けている。それらはアカデミアの研究分野にとどまらず、新興の企業を中心とした新たな創薬事業や健康産業へと成長しつつある。また、ヒトマイクロバイオームのみならず、農業をターゲットとした土壌マイクロバイオームや、都市環境や生活環境をターゲットとした環境マイクロバイオームなど、微生物が深く関与する多様な場において、微生物のゲノム・メタゲノム情報を基盤とした「ゲノム情報立脚型社会」が実現化しつつある。我々はこのような状況になる事を見越して、第一期より微生物 DB の統合化を推進し、論文のみならず学会や講演会、展示会などを通じて、ゲノム情報立脚型社会におけるデータ統合化の必要性および統合 DB を基盤とする新規融合領域の発展性を強く訴えてきた。

欧米に大きく遅れをとっていた我が国のマイクロバイオーム研究であるが、文科省から H28 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、我が国においても、ヒトマイクロバイオーム研究が強く推進されはじめており、我々が開発している MicrobeDB.jp に対して、これら研究にも大きく貢献する事が期待されている。

産業界においては、新たな産業のシーズとしてマイクロバイオーム研究への関心が強く、製薬、健康、化粧品、農業などバイオ産業を筆頭に、自動車、住宅、化学、分析、商社などの多様な業種も本分野への関心を持っており、DB に対する様々な発展的要望や、DB 利活用における共同研究など、MicrobeDB.jp による既産業の高度化、さらには新産業の開拓など、産業界への波及効果も高いと思われる。

マイクロバイオーム研究においては、細菌だけでなく藻類や菌類を含めた微生物叢研究、宿主であるヒトゲノムデータと、ヒトマイクロバイオームデータとの連携解析など、データ統合による新たな研究が展開され始めている。また、主に植物をターゲットとして、宿主および共生微生物すべてのゲノムを明らかにし、それら生物間の動的な相互作用や共進化などを研究するホログenom研究についても重点化されようとしている。このような状況において、我々の開発した微生物統合 DB 「MicrobeDB.jp」は、欠くことのできない研究基盤として、その重要性が一層高まると期待できる。

§6. 今後の展開

微生物は、海洋や河川、土壌や大気中、さらにはヒトや動物の腸内、皮膚、口腔内など地球上のあらゆる環境に存在し、その環境に特化した多様な微生物が群集を形成し棲息することで、地球環境における物質循環の根幹を形成しているといっても過言ではない。したがって、環境の根幹を形成する微生物群集の各個体および総体としての生命システムを明らかにするためには、環境中の微生物群集についての詳細な理解および基盤となる個別微生物種の高品質な情報が必須となる。

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」を開発し、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の基盤部分を構築した。今後、より多様な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を実装する事で、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。

これを実証するために、MicrobeDB.jp に収録されている約 3 万の 16S メタゲノムサンプルの微生物群集構造データに対して対応トピックモデルを適用し、それぞれのサンプルの潜在的環境因子をトピックとして抽出する事を試みた。その結果、微生物群集構造データを記述する「基底変数」

と見なしうる、微生物の部分群集構造と、その基底変数に対応する環境概念を表現する単語セットの集合が得られた。これらの基底変数の線形和として表現される各サンプルを低次元空間上で可視化することによって、どのような「環境」間で微生物群集構造の連続性が見られるのかを明らかにすることができた。これは、MicrobeDB.jp に収録されているメタゲノムデータが同一基準で解析されており、かつ全メタゲノムデータに標準化された MEO アノテーション情報が付されているからこそ初めて可能となる解析なのである。今後、新型シーケンサーに代表される実験機器の飛躍的なハイスループット化がますます加速され、ゲノム研究やメタゲノム研究に代表される先に述べたようなデータ駆動型研究がいよいよ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登場するに至って、今後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ると容易に想像できる。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り得えよう。

微生物は極めて多様であるため、微生物統合 DB の開発は3年間で完結するものではない。しかし、本課題で開発してきた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期待できるものである。実際、データ生産技術の発展を考えれば、今後動物や植物においても同様に多様性の大きなデータに直面する可能性が高く、また DNA 配列解析に依存しないプロテオームやメタボロームのデータについても、将来の技術革新により情報爆発を起こす可能性があるが、我々の開発しているシステムは、将来的にそのようなデータをも取り込んで発展させていく可能性を含んでいる。このような観点から将来を展望すれば、微生物に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生命分野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。日本が切り拓いてきたと言っても過言ではないゲノム科学は、現在欧米または中国に遅れをとっている感があるが、高品質な DB の統合化を基礎として新たな研究文化の創出を全世界に発信する事は、高度文化を築いてきた日本の責務であると考えられる。

§7. 自己評価

【研究代表者自身の評価】

本研究開発の成果は、参加した研究グループだけでは達成できず、DBCLS や他の研究開発グループとの相互依存的な連携により達成することができたと考えている。本研究開発では、第一期統合化推進プログラムで開発した **MicrobeDB.jp** を高度化することを目的として、大きく 7 つの開発項目を設けて研究開発を推進してきた。これらすべての開発目標は達成できたと考えている。特に、第一期開発終了時点で約 11 億トリプルであった DB を、第二期では約 90 億トリプルにまで発展させたことは、セマンティック Web 技術を駆使したトップクラスの統合 DB を開発した結果であると自負している。また、自動アノテーションツールの開発による統合 DB の自動更新技術、ゲノム・メタゲノム解析技術、オーソログ解析技術など、基盤技術も併せて向上させ、DB に収録するデータのハイレベルなクオリティの統一も達成している。さらに、それら統合データに容易にアクセス可能な 195 種類の **Stanza** 群の開発や検索インターフェイスの改良など、ユーザの利便性向上も図ってきた。これら成果をもって、最先端の統合 DB を開発できた事は、データ増大のペースを過小に予測していたものの、3年後を見通して立案した研究開発計画が適切であった事を示すものと考えている。

【本プログラムの趣旨に対する貢献】

微生物を対象とした DB は非常に多様であるが、それらの多くは基本的にはゲノムデータに紐付いている。我々はこの特徴に着目し、第一期、第二期を通じて、これら多様な DB を統合するために、DB に収録されている個々のデータを遺伝子、環境、系統に分類し、ゲノムデータを核として統合する戦略をとってきた。すなわち、単なる「DB の統合」ではなく、DB に収録されている個別の「データの統合」を目指してきた。しかし、データの統合は、統合するデータ間の関係性を解釈した上で、再解析を通じて新たな意味付けを徹底的に実施する必要があるため、無謀な研究開発戦略であると思われた。どうすれば作業を効率化できるか、どのような技術を開発する必要があるか、など、DBCLS と密に連携しつつ開発を続け、微生物統合 DB「**MicrobeDB.jp**」を構築するに至った。「データ統合」により全てのデータが巨大グラフで連結することで、問い合わせ方によって、個々の DB を意識することなく多種多様な結果をシームレスに得ることができる。また、環境をはじめとする各種オントロジーやボキャブラリを開発、整備したことで、それら自然言語を利用するあらゆる分野のデータとも親和性が高い。データ統合による **MicrobeDB.jp** は、統合化のメリットを最大化するものであり、ライフサイエンスのみならず、異分野からの参入をも容易に受け入れ可能な、新たな分野を切り拓く事ができるフロンティアとなり得るものと確信している。

§8. 外部発表等

1. 原著論文発表

(1) 論文数概要

種別	国内外	件数
発行済論文	国内(和文)	0 件
	国際(欧文)	36 件
未発行論文 (accepted, in press 等)	国内(和文)	0 件
	国際(欧文)	3 件

(2) 論文詳細情報

1. Yano M, Mori H, Akiyama Y, Yamada T, Kurokawa K. CLAST: CUDA implemented large-scale alignment search tool. *BMC Bioinformatics*, 15:406, 2014. (DOI:10.1186/s12859-014-0406-y).
2. Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, Sato S, Yamada T, Mori H, Tajima N, Moriyama T, Ikeuchi M, Watanabe M, Wada H, Kobayashi K, Saito M, Masuda T, Sasaki-Sekimoto Y, Mashiguchi K, Awai K, Shimojima M, Masuda S, Iwai M, Nobusawa T, Narise T, Kondo S, Saito H, Sato R, Murakawa M, Ihara Y, Oshima-Yamada Y, Ohtaka K, Satoh M, Sonobe K, Ishii M, Ohtani R, Kanamori-Sato M, Honoki R, Miyazaki D, Mochizuki H, Umetsu J, Higashi K, Shibata D, Kamiya Y, Sato N, Nakamura Y, Tabata S, Ida S, Kurokawa K, Ohta H. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* 5:3978, 2014. (DOI: 10.1038/ncomms4978).
3. Kanesaki Y, Masutani H, Sakanaka M, Shiwa Y, Fujisawa T, Nakamura Y, Yokota A, Fukiya S, Suzuki T, Yoshikawa H. Complete genome sequence of *Bifidobacterium longum* 105-A, a strain with high transformation efficiency. *Genome Announc.*, 2(6):e01311-14, 2014. (DOI:10.1128/genomeA.01311-14).
4. Tanizawa Y, Fujisawa T, Mochizuki T, Kaminuma E, Suzuki Y, Nakamura Y, Tohno M. Draft genome sequence of *Weissella oryzae* SG25T, isolated from fermented rice grains. *Genome Announc.* 2(4):e00667-14, 2014. (DOI:10.1128/genomeA.00667-14).
5. Tanizawa Y, Fujisawa T, Mochizuki T, Kaminuma E, Nakamura Y, Tohno M. Draft genome sequence of *Lactobacillus oryzae* strain SG293T. *Genome Announc.* 2(4):e00861-14, 2014. (DOI:10.1128/genomeA.00861-14).
6. Chiba H, Uchiyama I. “Improvement of domain-level ortholog clustering by optimizing domain-specific sum-of-pairs score”, *BMC Bioinformatics*, 15:148, 2014. (DOI:10.1186/1471-2105-15-148).
7. Wang T, Mori H, Zhang C, Kurokawa K, Xing X, Yamada T. DomSign: a top-down annotation pipeline to enlarge enzyme space in the protein universe. *BMC Bioinformatics*, 16:96, 2015. (DOI: 10.1186/s12859-015-0499-y).
8. Kojima H, Ogura Y, Yamamoto N, Togashi T, Mori H, Watanabe T, Nemoto F, Kurokawa K, Hayashi T, Fukui M. Ecophysiology of *Thioploca ingrica* as revealed by the complete genome sequence supplemented with proteomic evidence. *ISME J.*, 9:1166–1176, 2015. (DOI: 10.1038/ismej.2014.209).
9. Yamaichi Y, Chao C. M, Sasabe J, Clark L, Davis M. B, Yamamoto N, Mori H, Kurokawa K, Waldor M. K. High-resolution genetic analysis of the requirements for horizontal transmission of the ESBL plasmid from *Escherichia coli* O104:H4. *Nucleic Acids Res.* 43:348–360, 2015. (DOI: 10.1093/nar/gku1262).
10. Uchiyama T, Irie M, Mori H, Kurokawa K, Yamada T, FuncTree: Functional analysis and visualization for large-scale omics data. *PLoS One*, 10:e0126967, 2015. (DOI: 10.1371/journal.pone.0126967).
11. Kato H, Ogawa N, Ohtsubo Y, Oshima K, Toyoda A, Mori H, Nagata Y, Kurokawa K, Hattori M, Fujiyama A, Tsuda M, Complete Genome Sequence of a Phenanthrene

- Degrader, *Mycobacterium* sp. Strain EPa45 (NBRC 110737), Isolated from a Phenanthrene-Degrading Consortium. *Genome Announc.* 3:e00782-15, 2015. (DOI: 10.1128/genomeA.00782-15).
12. Kato H, Mori H, Maruyama F, Toyoda A, Oshima K, Endo R, Fuchu G, Miyakoshi M, Dozono A, Ohtsubo Y, Nagata Y, Hattori M, Fujiyama A, Kurokawa K, Tsuda M, Time-series metagenomic analysis reveals robustness of soil microbiome against chemical disturbance. *DNA Res.* 22:413–424, 2015. (DOI: 10.1093/dnares/dsv023).
 13. Higashi K, Tobe T, Kanai A, Uyar E, Ishikawa S, Suzuki Y, Ogasawara N, Kurokawa K, Oshima T. H-NS facilitates sequence diversification of horizontally transferred DNAs during their integration process. *PLoS Genet.* 12(1):e1005796, 2015. (DOI: 10.1371/journal.pgen.1005796).
 14. Kodama Y, Mashima J, Kosuge T, Katayama T, Fujisawa T, Kaminuma E, Ogasawara O, Okubo K, Takagi T, Nakamura Y. The DDBJ Japanese Genotype-phenotype Archive for genetic and phenotypic human data. *Nucleic Acids Res.* 43:D18-22, 2015. (DOI:10.1093/nar/gku1120).
 15. Hiraide Y, Oshima K, Fujisawa T, Uesaka K, Hirose Y, Tsujimoto R, Yamamoto H, Okamoto S, Nakamura Y, Terauchi K, Omata T, Ihara K, Hattori M, Fujita Y. Loss of cytochrome cM stimulates cyanobacterial heterotrophic growth in the dark. *Plant Cell Physiol.* 56(2):334-45, 2015. (DOI: 10.1093/pcp/pcu165).
 16. Hirose Y, Fujisawa T, Ohtsubo Y, Katayama M, Misawa N, Wakazuki S, Shimura Y, Nakamura Y, Kawachi M, Yoshikawa H, Eki T, Kanasaki Y. Complete genome sequence of cyanobacterium *Nostoc* sp. NIES-3756, a potentially useful strain for phytochrome-based bioengineering. *J. Biotechnol.* 218: 51-52, 2015. (DOI: 10.1016/j.jbiotec.2015.12.002).
 17. Mashima J, Kodama Y, Kosuge T, Fujisawa T, Katayama T, Nagasaki H, Okuda Y, Kaminuma E, Ogasawara O, Okubo K, Nakamura Y, Takagi T. DNA data bank of Japan (DDBJ) progress report. *Nucleic Acids Res.*, 44:D51–D57, 2015. (DOI: 10.1093/nar/gkv1105).
 18. Kodama Y, Mashima J, Kosuge T, Katayama T, Fujisawa T, Kaminuma E, Ogasawara O, Okubo K, Takagi T, Nakamura Y. The DDBJ Japanese Genotype-phenotype Archive for genetic and phenotypic human data. *Nucleic Acids Res.*, 43:D18-22, 2015. (DOI: 10.1093/nar/gku1120).
 19. Aoki-Kinoshita KF, Kinjo AR, Morita M, Igarashi Y, Chen Y, Shigemoto Y, Fujisawa T, Akune Y, Katoda T, Kokubu A, Mori T, Nakao M, Okamoto S, Katayama T, Ogishima S. Implementation of linked data in the life sciences at BioHackathon 2011. *J. Biomed. Semantics* 6:3, 2015. (DOI:10.1186/2041-1480-6-3).
 20. Uchiyama I, Mihara M, Nishide H, Chiba H. “MBGD update 2015: microbial genome database for flexible ortholog analysis utilizing a diverse set of genomic data”, *Nucleic Acids Res.*, 43:D270-276, 2015. (DOI: 10.1093/nar/gku1152).
 21. Chiba H, Nishide H, Uchiyama I. Construction of an ortholog database using the Semantic Web technology for integrative analysis of genomic data, *PLoS ONE*, 10, e0122802, 2015. (DOI:10.1371/journal.pone.0122802).
 22. Fernández-Breis, JT, Legaz-García, MC, Chiba H, Uchiyama I. Towards the semantic standardization of orthology content. *Proc. Semantic Web Applications and Tools for Life Sciences*, 74-83, 2015. (DOI:).
 23. Nakamura Y, Yamamoto N, Kino Y, Yamamoto N, Kamei S, Mori H, Kurokawa K, Nakashima N. Establishment of a multi-species biofilm model and metatranscriptomic analysis of biofilm and planktonic cell communities. *Appl. Microbiol. Biotechnol.* 100:7263–7279, 2016. (DOI: 10.1007/s00253-016-7532-6).
 24. Matsuki T, Yahagi K, Mori H, Matsumoto H, Hara T, Tajima S, Ogawa E, Kodama H, Yamamoto K, Yamada T, Matsumoto S, Kurokawa K. A key genetic factor for fucosyllactose utilization affects infant gut microbiota development. *Nat. Commun.*

- 7:11939, 2016. (DOI: 10.1038/ncomms11939).
25. Okai S, Usui F, Yokota S, Hori-i Y, Hasegawa M, Nakamura T, Kurosawa M, Okada S, Yamamoto K, Nishiyama E, Mori H, Yamada T, Kurokawa K, Matsumoto S, Nanno M, Naito T, Watanabe Y, Kato T, Miyauchi E, Ohno H, High-affinity monoclonal IgA regulates gut microbiota and prevents colitis in mice. *Nature Microbiol.* 1:16103, 2016. (DOI: 10.1038/nmicrobiol.2016.103).
 26. Bolleman JT, Mungall CJ, Strozzi F, Baran J, Dumontier M, Bonnal RPJ, Buels R, Hoehndorf R, Fujisawa T, Katayama T, Cock PJ. FALDO: a semantic standard for describing the location of nucleotide and protein feature annotation. *J. Biomed. Semantics*, 7:1–12, 2016. (DOI: 10.1186/s13326-016-0067-z).
 27. Fujisawa T, Narikawa R, Maeda S, Watanabe S, Kanesaki Y, Kobayashi K, Nomata J, Hanaoka M, Watanabe M, Ehira S, et al. CyanoBase: a large-scale update on its 20th anniversary. *Nucleic Acids Res.*, 45 (D1): D551-D554, 2016. (DOI: 10.1093/nar/gkw1131).
 28. Hirose Y, Fujisawa T, Ohtsubo Y, Katayama M, Misawa N, Wakazuki S, Shimura Y, Nakamura Y, Kawachi M, Yoshikawa H, Eki T, Kanesaki, Y. Complete genome sequence of cyanobacterium *Fischerella* sp. NIES-3754, providing thermoresistant optogenetic tools. *J. Biotechnol.*, 220:45–46, 2016. (DOI: 10.1016/j.jbiotec.2016.01.011).
 29. Hirose Y, Fujisawa T, Ohtsubo Y, Katayama M, Misawa N, Wakazuki S, Shimura Y, Nakamura Y, Kawachi M, Yoshikawa H, Eki T, Kanesaki, Y. Complete genome sequence of cyanobacterium *Nostoc* sp. NIES-3756, a potentially useful strain for phytochrome-based bioengineering. *J. Biotechnol.*, 218:51–52, 2016. (DOI: 10.1016/j.jbiotec.2015.12.002).
 30. Hirose Y, Fujisawa T, Ohtsubo Y, Katayama M, Misawa N, Wakazuki S, Shimura Y, Nakamura Y, Kawachi M, Yoshikawa H, Eki T, Kanesaki, Y. Complete Genome Sequence of Cyanobacterium *Leptolyngbya* sp. NIES-3755. *Genome Announc.*, 4, e00090-16, 2016. (DOI: 10.1128/genomeA.00090-16).
 31. Mashima J, Kodama Y, Kosuge T, Fujisawa T, Katayama T, Nagasaki H, Okuda Y, Kaminuma E, Ogasawara O, Okubo K, et al. DNA data bank of Japan (DDBJ) progress report. *Nucleic Acids Res.* 44:D51–D57, 2016. (DOI: 10.1093/nar/gkv1105).
 32. Nakai R, Fujisawa T, Nakamura Y, Baba T, Nishijima M, Karray F, Sayadi S, Isoda H, Naganuma T, Niki H. Genome sequence and overview of *Oligoflexus tunisiensis* Shr3 T in the eighth class Oligoflexia of the phylum Proteobacteria. *Stand. Genomic Sci.*, 11, 90, 2016. (DOI: 10.1186/s40793-016-0210-6).
 33. Nakai R, Fujisawa T, Nakamura Y, Nishide H, Uchiyama I, Baba T, Toyoda A, Fujiyama A, Naganuma T, Niki H. Complete genome sequence of *Aurantimicrobium minutum* type strain KNCT, a planktonic ultramicrobacterium isolated from river water. *Genome Announc.*, 4, e00616-16, 2016. (DOI: 10.1128/genomeA.00616-16).
 34. Tanizawa Y, Fujisawa T, Kaminuma E, Nakamura Y. DFAST and DAGA: web-based integrated genome annotation tools and resources. *Biosci. Microbiota, Food Heal.*, 35:173–184, 2016. (DOI: 10.12938/bmfh.16-003).
 35. Fernández-Breis JT, Chiba H, Legaz-García MC, Uchiyama I. The orthology ontology: development and applications. *J. Biomed. Semant.*, 7:34, 2016. (DOI: 10.1186/s13326-016-0077-x).
 36. Uchiyama I, Albritton J, Fukuyo M, Kojima K, Yahara K, Kobayashi I. A novel approach to *Helicobacter pylori* pan-genome analysis for identification of genomic islands, *PLoS One*, 11, e0159419, 2016. (DOI: 10.1371/journal.pone.0159419).
 37. Watanabe H, Goto S, Mori H, Higashi K, Hosomichi K, Aizawa N, Takahashi N, Tsuchida M, Suzuki Y, Yamada T, Horii A, Inoue I, Kurokawa K, Narita I. Comprehensive microbiome analysis of tonsillar crypts in IgA nephropathy.

Nephrol Dial Transplant. (in press).

38. Okai S, Usui F, Ohta M, Mori H, Kurokawa K, Matsumoto S, Kato T, Miyauchi E, Ohno H, Shinkura R. Intestinal IgA as a modulator of the gut microbiota. *Gut Microbes*, (in press).
39. Higashi K, Kawai Y, Baba T, Kurokawa K, Oshima T. Essential cellular modules for the proliferation of the primitive cell. *Geoscience Frontiers*, (in press).

2. その他の著作物(総説、書籍など)

1. 山本希, 森宙史, 山田拓司, 黒川顕, “メタゲノミクスの現状と未来”, 生命のビッグデータ利用の最前線, シーエムシー出版, pp.48-57, 2014.
2. 内山郁夫, 大量シーケンス時代の比較ゲノミクス基盤, 生命のビッグデータ利用の最前線, pp. 93-103, CMC 出版, 2014.
3. 森宙史, 黒川顕, “MicrobeDB.jp”, 今日から使えるデータベース・ウェブツール, 羊土社, pp.76-77, 32(20), 2014.
4. 森宙史, 東光一, 山田拓司, 黒川顕, “微生物統合データベースによる微生物と環境の研究への貢献”, 難培養微生物研究の最新技術 III—微生物の生き様に迫り課題解決へ—, シーエムシー出版, pp.10-19, 2015.
5. 内山郁夫, 大規模なゲノムデータの活用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ニュースレター No.11, 2015.
6. 工樂樹洋, 千葉啓和, ミーティングレポート 第4回 Quest for Orthologs ミーティング参加記, 日本進化学会ニュース Vol.16 No.2, 2015.
7. 内山郁夫, *Helicobacter pylori* の遺伝子レパートリー, *Helicobacter research*, 20, 30-36, 2016.
8. 東光一, 森宙史, 黒川顕, “メタゲノムデータの情報解析のオーバービュー”, 今すぐ始める! メタゲノム解析 実験プロトコール, 羊土社, pp.15-21, 2016.

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	24 件
	国際	3 件
口頭発表	国内	8 件
	国際	2 件
ポスター発表	国内	23 件
	国際	2 件

(2) 招待講演

(国内)

1. 森宙史, 古代ゲノムと微生物統合データベース, 北海道大学低温科学研究所研究集会「低温環境への適応:ゲノム進化と環境変動指標」, 北海道大学, 2014年7月22-23日.
2. 黒川顕, 統合データベースを活用した冥王代類似環境微生物のゲノム情報解析, 大阪, 日本進化学会第16回大阪大会, 2014年8月23日.
3. 森宙史, 藤澤貴智, 大山彰, 菅原秀明, 中村保一, 黒川顕, MiGAP を用いた微生物ゲノムオートアノテーションとMicrobeDB.jp 連携, 第8回日本ゲノム微生物学会若手の会, ろうきん研修所富士センター, 2014年9月28-29日.
4. 内山郁夫, 微生物比較ゲノムデータベースMBGD, 第3回生命医薬情報学連合大会・オープンサイエンスアワード2014, 仙台国際センター, 2014年10月2-4日.
5. 黒川顕, 微生物統合データベース「MicrobeDB.jp」, 環境微生物系学会合同大会

- 2014, 浜松, 2014年10月22日.
6. 森宙史, 細菌群集解析のためのメタ 16S・メタゲノム解析パイプラインの開発, Illumina BaseSpace Worldwide Developers Conference in Japan, 東京大学柏の葉キャンパス駅前サテライト, 2014年10月29-30日.
 7. 森宙史, メタゲノム解析と微生物統合データベース, 第30回DDBJing講習会 in 東京, 独立行政法人科学技術振興機構東京本部, 2014年12月18日.
 8. 黒川顕, メタゲノミクスの現状と未来, 如水会講演会, 東京, 2015年3月3日.
 9. 黒川顕, 微生物統合データベース「MicrobeDB.jp」, 第88回日本細菌学会総会, 長良川国際会議場, 2015年3月26日.
 10. 内山郁夫, 属、科など上位分類群に保存されたコア遺伝子の定義と分類への応用, 第88回日本細菌学会総会, 長良川国際会議場, 2015年3月26-28日.
 11. 黒川顕, メタゲノミクスの未来, 蔵前工業会, 神戸, 2015年4月4日.
 12. 黒川顕, 微生物統合データベース「MicrobeDB.jp」, マリンバイオテクノロジー学会, 東京, 2015年5月30日.
 13. 黒川顕, メタゲノミクスの未来, リバネス「細菌叢研究が拓く未来」, 東京, 2015年7月23日.
 14. 森宙史, 微生物統合データベース MicrobeDB.jp のメタゲノム情報を用いた超高度化, ラン藻ゲノム交流会 2015, 東京大学駒場キャンパス 16号館 119・129号室, 2015年7月25日.
 15. 中村保一, CyanoBase 愛されて 20年, 平成 27年度 ラン藻ゲノム交流会, 東京, 2015年7月25日.
 16. 黒川顕, メタゲノミクスと未来社会, ATIS(技術情報サービス協会)講演会, 東京, 2015年9月16日.
 17. 千葉啓和, SPANG:コマンドライン環境での SPARQL 利用促進, 人工知能学会 第37回 SWO 研究会, 慶応大学日吉キャンパス, 2015年11月13日.
 18. 森宙史, 微生物ゲノムアノテーションの現状と問題点, 2015年度国立遺伝学研究所研究会 Annotathon 2015, 国立遺伝学研究所, 2015年11月12日.
 19. 黒川顕, メタゲノミクスのその先, イルミナゲノムサミット 2016, 東京, 2016年6月1日.
 20. 黒川顕, 微生物統合データベース「MicrobeDB.jp」, 電子情報通信学会・第9回アクセラレーション技術発表討論会「IT 農業」, 静岡, 2016年9月1日.
 21. 黒川顕, 微生物統合データベース「MicrobeDB.jp」, イルミナマイクロバイオーームワークショップ, 東京, 2016年10月20日.
 22. 黒川顕, メタゲノム解析パイプライン, マイクロバイオーーム公開セミナー, 大阪, 2016年11月4日.
 23. 黒川顕, 微生物統合データベース「MicrobeDB.jp」, 東北大学微生物進化機能開発寄附講座開設記念シンポジウム, 仙台, 2016年11月24日.
 24. 黒川顕, ショットガンメタゲノム解析手法, 第1回細菌叢ワークショップ, 大阪, 2016年12月22日.

〈国際〉

1. Ken Kurokawa, “Small Bugs, Big Data”: Developing an integrated database for microbes with semantic web technologies, ICSTI2014 General Assembly & Annual Conference in Tokyo, Tokyo, 2014年10月20日.
2. Hirokazu Chiba, Hiroyo Nishide, Ikuo Uchiyama, Construction of ortholog database using the Semantic Web technology BioHackathon2014 Symposium, 東北大学, 2014年11月9日.
3. Yasukazu Nakamura, Towards better genome annotation, The 13th JCK Bioinformatics Symposium, OIST Sea Side House, Okinawa, 2015年8月11日.

(3) 口頭講演

〈国内〉

1. 藤澤貴智, 中村保一, ゲノムアノテーション支援サービス「Genome Refine」の開発と現状, ラン藻ゲノム研究交流会 2014, 東京, 2014年7月19日.
2. 森宙史, 内山郁夫, 菅原秀明, 中村保一, 黒川顕, MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp の超高度化, 第9回日本ゲノム微生物学会年会, 神戸大学, 2015年3月6-8日.
3. 中村保一, CyanoBase 愛されて 20 年. ところで今後もご入り用ですか?, 平成 27 年度 ラン藻ゲノム交流会, 東京大学駒場キャンパス, 2015年7月25日.
4. 藤澤貴智, 栗井光一郎, CyanoBase 再アノテーションの進捗について, 平成 27 年度 ラン藻ゲノム交流会, 東京大学駒場キャンパス, 2015年7月25日.
5. 内山郁夫, 三原基広, 西出浩世, 千葉啓和, 大規模ゲノムデータの活用に向けた微生物比較ゲノムデータベース MBGD の改良, ゲノム微生物学会, 神戸大学, 2015年3月6-8日.
6. 森宙史, 環境情報を用いた微生物統合データベースの超高度化, 生命医薬情報学連合大会 2015 年大会, 京都大学宇治キャンパスおうばくプラザ, 2015年10月31日.
7. 森宙史, 藤澤貴智, 千葉啓和, 山本希, 内山郁夫, 菅原秀明, 中村保一, 黒川顕, MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp の検索システムの高度化と新解析パイプライン, 第 10 回日本ゲノム微生物学会年会, 東京工業大学, 2016年3月4-5日.
8. 森宙史, 微生物統合データベース MicrobeDB.jp 2.0 のデータ統合の実際, 第 5 回生命医薬情報学連合大会, 東京国際交流館プラザ平成, 2016年9月29日.

〈国際〉

1. Ikuo Uchiyama, New developments in MBGD: incorporation of draft genome data and refinement of domain-level ortholog identification, Quest for Orthologs, Barcelona, May 25-27, 2015.
2. Hirokazu Chiba, Jesualdo Tomás Fernández-Breis, María del Carmen Legaz-García, Ikuo Uchiyama, Applying RDF/SPARQL to ortholog databases, BioHackathon 2015 Symposium, Nagasaki, September 13, 2015.

(4) ポスター発表

〈国内〉

1. Hirokazu Chiba, Hiroyo Nishide, Ikuo Uchiyama, Construction of ortholog database using the Semantic Web technology, JSBi2014・第3回生命医薬情報学連合大会, 仙台国際センター, 2014年10月2-4日.
2. 中村保一, DDBJ, トーゴの日シンポジウム 2014, 東京, 2014年10月5-6日.
3. 藤澤貴智, Genome Refine: ゲノムアノテーション支援ウェブサービスにおけるセマンティックウェブ技術の利用, トーゴの日シンポジウム 2014, 東京, 2014年10月5-6日.
4. 内山郁夫, 三原基広, 西出浩世, 千葉啓和, 大規模比較ゲノム基盤としての微生物ゲノムデータベース MBGD, トーゴの日シンポジウム 2014, 時事通信ホール, 2014年10月5日.
5. 鈴木真也, 山本希, 森宙史, 黒川顕, 微生物統合データベースの超高度化推進のための微生物の生息環境オントロジーと解析アプリケーションの開発, 第9回日本ゲノム微生物学会年会, 神戸大学, 2015年3月6-8日.
6. 藤澤貴智, Genome Refine: 自動ゲノムアノテーション統合環境を提供するウェブサービス, 第9回ゲノム微生物学会年会, 神戸, 2015年3月6-8日.
7. 森宙史, 藤澤貴智, 千葉啓和, 内山郁夫, 菅原秀明, 中村保一, 黒川顕,

- MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp の超高度化, トーゴの日シンポジウム 2015, 東京大学弥生講堂一条ホール, 2015 年 10 月 5-6 日.
8. 鈴木真也, 山本希, 森宙史, 黒川顕, 微生物統合データベースのための微生物採集環境オントロジーと解析アプリケーションの開発, トーゴの日シンポジウム 2015, 東京大学弥生講堂一条ホール, 2015 年 10 月 5-6 日.
 9. 山本希, 鈴木真也, 森宙史, 黒川顕, MicrobeDB.jp プロジェクトチーム, 疾病関連語句オントロジーを利用したゲノム、メタゲノムデータの RDF 化と利用, トーゴの日シンポジウム 2015, 東京大学弥生講堂一条ホール, 2015 年 10 月 5-6 日.
 10. 中村保一, DDBJ, トーゴの日シンポジウム 2015, 東京, 2015 年 10 月 5-6 日.
 11. 藤澤貴智, GenomeRefine: ゲノム・メタゲノム解析データに対してアノテーションするウェブサービス, トーゴの日シンポジウム 2015, 東京, 2015 年 10 月 5-6 日.
 12. 内山郁夫, 三原基広, 西出浩世, 千葉啓和, 大規模比較解析に向けた微生物ゲノムデータベース MBGD の拡張, トーゴの日シンポジウム 2015, 東京大学弥生講堂, 2015 年 10 月 5 日.
 13. 千葉啓和, 西出浩世, 内山郁夫, セマンティック Web 技術を用いたオーソログデータベースの統合化, トーゴの日シンポジウム 2015, 東京大学弥生講堂, 2015 年 10 月 5 日.
 14. 鈴木真也, 山本希, 森宙史, 黒川顕, 微生物統合データベースを高度化する微生物環境オントロジーと解析アプリケーションの開発, 第 10 回日本ゲノム微生物学会年会, 東京工業大学, 2016 年 3 月 4-5 日.
 15. 藤澤貴智, シアノバクテリアに特化した CyanoBase ゲノムアノテーションパイプラインの構築, 第 10 回日本ゲノム微生物学会年会, 東京, 2016 年 3 月 4 日.
 16. 千葉啓和, 内山郁夫, ドメイン単位のオーソログ分類に基づく融合遺伝子の解析, 第 10 回日本ゲノム微生物学会年会, 東京工業大学大岡山キャンパス, 2016 年 3 月 4 日.
 17. 鈴木真也, 山本希, 森宙史, 黒川顕, 微生物統合データベースを高度化する微生物環境オントロジーと解析アプリケーションの開発, 日本農芸化学会 2016 年度大会, 札幌コンベンションセンター, 2016 年 3 月 28-30 日.
 18. 森宙史, 藤澤貴智, 千葉啓和, 鈴木真也, 山本希, 内山郁夫, 菅原秀明, 中村保一, 黒川顕, MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp 2.0, トーゴの日シンポジウム 2016, 東京大学弥生講堂一条ホール, 2016 年 10 月 5 日.
 19. 鈴木真也, 山本希, 森宙史, 黒川顕, 微生物由来環境情報の RDF 化による統合化, トーゴの日シンポジウム 2016, 東京大学弥生講堂一条ホール, 2016 年 10 月 5 日.
 20. 中村保一, DDBJ, トーゴの日シンポジウム 2016, 東京, 2016 年 10 月 5-6 日.
 21. 藤澤貴智, セマンティック・ウェブ技術を用いた CyanoBase 更新系の整備, トーゴの日シンポジウム 2016, 東京, 2016 年 10 月 5-6 日.
 22. 内山郁夫, 三原基広, 西出浩世, 千葉啓和, 大規模な微生物ゲノムに対応したオーソログ解析手順の改良, トーゴの日シンポジウム 2016, 東京大学弥生講堂, 2016 年 10 月 5 日.
 23. 千葉啓和, 内山郁夫, セマンティック SPANG による RDF データベースの統合的利用, トーゴの日シンポジウム 2016, 東京大学弥生講堂, 2016 年 10 月 5 日.

〈国際〉

1. Hiroshi Mori, Nozomi Yamamoto, Takuji Yamada, Ken Kurokawa, MicrobeDB.jp: Integration of a microorganism database based on genomic and metagenomic information, the 3rd ELSI International Symposium, Kuramae Hall, Tokyo, Japan, January 13-15, 2015.

2. Takatomo Fujisawa, Toshiaki Katayama, Mitsuteru Nakao, Shinobu Okamoto, Yasukazu Nakamura, Semantic data integration in CyanoBase and RhizoBase, Biocuration 2015, Beijing, China, April 25, 2015.

4. 知財出願

なし

5. 受賞・報道等

(1) 受賞

1. 手島精一記念研究賞、森宙史、黒川顕、2015年1月.
2. 第10回日本ゲノム微生物学会研究若手賞、森宙史、2016年3月.

(2) メディア報道

なし

(3) その他

なし

§9. 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
2014年 7月5日	チーム内第1回全体会議（非公開）	東工大田町キャンパス	10人	研究進捗報告のための会議
2015年 2月21日	チーム内第2回全体会議（非公開）	東工大大岡山キャンパス	8人	同上
2015年 4月6日	チーム内第3回全体会議（非公開）	東工大大岡山キャンパス	8人	同上
2015年 7月22日	菌株会議（非公開）	東工大田町キャンパス	5人	海外の菌株保存機関のデータベース開発者と連携を模索する会議
2015年 8月23日	菌株会議（非公開）	中国科学院	8人	同上
2016年 9月15日	チーム内第4回全体会議（非公開）	国立遺伝学研究所	7人	研究進捗報告のための会議
2016年 12月21日	チーム内第5回全体会議（非公開）	国立遺伝学研究所	6人	同上

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
2016年 10月8日	基礎生物学研究所一般公開	基礎生物学研究所	4716人 (全体)	本プロジェクトで作成したスタンザを改変したインターフェイスを使って簡単な比較ゲノム解析のデモを行った。

以上

別紙 研究開発対象のデータベース等

No.	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系データベースアーカイブ	NBDCヒトデータベース	NBDC RDFポータル	関連文献 (論文リストに記載があれば、その番号でも可)
1	MicrobeDB.jp		本DBは、微生物に関する多種多様な情報を遺伝子・系統・環境の3つの軸に沿って整理統合し、セマンティックWeb技術を利用して単一の検索ウィンドウからそれらの情報を検索可能な統合DBです。	http://microbedb.jp/	2011年12月	継続・発展	データベース等	調整中	対象外	調整中	
2	MCCV	Microbial Culture Collection Vocabulary	JCMやNBRCなどの、菌株保存機関に蓄積されている菌株情報を記述するための統制語彙MCCV (Microbial Culture Collection Vocabulary)をOWLファイルで提供するDBです。	http://bioportal.bioontology.org/ontologies/MCCV	2011年12月	継続・発展	ツール等	調整中	対象外	調整中	
3	CyanoBase	Genome database for Cyanobacteria	財団法人かずさDNA研究所で全ゲノム配列決定したシアノバクテリアを含む、シアノバクテリア、緑色硫黄細菌、紅色非硫黄細菌の全ゲノム配列解析結果を収録しています。ゲノム配列、染色体マップ、遺伝子アノテーション、遺伝子機能カテゴリ、文献、公共データベースのリンク等が閲覧できます。	http://genome.microbedb.jp/cyanobase	1995年12月	継続・発展	データベース等	調整中	対象外	調整中	Takatomo Fujisawa, Shinobu Okamoto, Toshiaki Katayama, Mitsuteru Nakao, Hidehisa Yoshimura, Hiromi Kajiya-Kanegae, Sumiko Yamamoto, Chiyoko Yano, Yuka Yanaka, Hiroko Maita, Takakazu Kaneko, Satoshi Tabata, Yasukazu Nakamura, "Nucleic Acids Res.", 42:D666-70,2014 (DOI:10.1093/nar/gkt1145)
4	RhizoBase		財団法人かずさDNA研究所で全ゲノム配列決定したミヤコグサ根粒菌Mesorhizobium loti MAFF303099を含む根粒菌の全ゲノム配列情報を収録しています。種毎に、ゲノム配列、染色体マップ、遺伝子アノテーション、遺伝子機能カテゴリ、文献、公共データベースのリンク等が閲覧できます。	http://genome.microbedb.jp/rhizobase	2002年2月	継続・発展	データベース等	調整中	対象外	調整中	Takatomo Fujisawa, Shinobu Okamoto, Toshiaki Katayama, Mitsuteru Nakao, Hidehisa Yoshimura, Hiromi Kajiya-Kanegae, Sumiko Yamamoto, Chiyoko Yano, Yuka Yanaka, Hiroko Maita, Takakazu Kaneko, Satoshi Tabata, Yasukazu Nakamura, "Nucleic Acids Res.", 42:D666-70,2014 (DOI:10.1093/nar/gkt1145)
5	MBGD	Microbial Genome Database for Comparative Analysis	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です。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できます。	http://mbgd.genome.ad.jp/	1997年7月	継続・発展	データベース等	調整中	対象外	公開済	Uchiyama, I., Mihara, M., Nishide, H. and Chiba, H. Nucleic Acids Res., 43, D270-276, 2015

No.	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系データベースアーカイブ	NBDCヒトデータベース	NBDC RDFポータル	関連文献 (論文リストに記載があれば、その番号でも可)
6	MBGD SPARQL Search		MBGDのRDF版をSPARQLで検索するためのウェブサイトです。	http://mbgd.genome.ad.jp/sparql/	2014年10月	継続・発展	ツール等	対象外	対象外	対象外	Hirokazu Chiba, Hiroyo Nishide, and Ikuo Uchiyama. PLOS ONE, 10(4), e0122802, 2015.
7	TogoAnnotation		ソーシャルブックマークによる文献情報集積プラットフォームです。	http://togo.annotation.jp/	2006年12月	継続・発展	ツール等	調整中	対象外	調整中	
8	Genome Refine		微生物ゲノムアノテーションの高品質化および解析支援のためのウェブサービスです。ユーザ認証機能、メタデータ入力機能、解析パイプラインMiGAP、MeGAPの解析結果の登録機能、RDF形式変換機能をもち、MicrobeDB.jpで参照可能です。	http://genome.annotation.jp/genomerefine/	2014年7月	継続・発展	ツール等	対象外	対象外	対象外	
9	GTPS	Gene Trek in Prokaryote Space	<個別の公開ページは閉鎖。開発成果はMicrobeDB.jpへ統合されている。> ゲノム配列が決定された原核生物を対象に各種解析を行い再アノテーションを行ったデータベースです。GIB (Genome Information Broker) に登録されている真正細菌と古細菌のORFを共通プロトコルを用いて再アノテーションし、グラフィカルユーザインタフェースやフラットファイルとしてデータ提供をしています。 セマンティックWeb技術によって、GTPSと微生物統合データベースにおける他の要素データベースとの連携を可能にするGTPS2011のRDF化版が http://gtps.ddbj.nig.ac.jp/rdf/ よりダウンロード可能です。		2009年3月	休止・閉鎖	データベース等	対象外	対象外	対象外	Takehide Kosuge, Takashi Abe, Toshihisa Okido, Naoto Tanaka, Masaki Hirahata, Yutaka Maruyama, Jun Mashima, Aki Tomiki, Motoyoshi Kurokawa, Ryutaro Himeno, Satoshi Fukuchi, Satoru Miyazaki, Takashi Gojobori, Yoshio Tateno and Hideaki Sugawara. "Exploration and Grading of Possible Genes from 183 Bacterial Strains by a Common Protocol to Identification of New Genes: Gene Trek in Prokaryote Space (GTPS)", DNA Res, 13(6):245-254, 2006 (DOI:10.1093/dnares/dsl014)
10	GTPS/RDF		<個別の公開ページは閉鎖。開発成果はMicrobeDB.jpへ統合されている。> セマンティックWeb技術によって、GTPSと微生物統合データベースにおける他の要素データベースとの連携を可能としたGTPS2011のRDF化版です。		2012年4月	休止・閉鎖	データベース等	対象外	対象外	対象外	
11	MEO	Metagenome/Microbes Environmental Ontology	<個別の公開ページは閉鎖。開発成果はMicrobeDB.jpへ統合されている。> 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であるMEO (Metagenome/Microbes Environmental Ontology)のOWLファイルと、公共のメタゲノムデータのメタデータをMEO相手にマッピングしてサンプルごとに整理した結果のRDFファイルの2つから構成されたDBです。	http://bioportal.bioontology.org/ontologies/MEO	2012年1月	休止・閉鎖	ツール等	調整中	対象外	調整中	

No.	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系データベースアーカイブ	NBDCヒトデータベース	NBDC RDFポータル	関連文献 (論文リストに記載があれば、その番号でも可)
12	Metagenome.jp (Human Meta BodyMap)		〈個別の公開ページは閉鎖。開発成果はMicrobeDB.jpへ統合されている。〉 世界中で公開されているヒトメタゲノム解析で得られた配列情報を、独自に注意深く再アノテーションしたメタデータに基づき参照できるようにしたDBです。メタデータ検索機能や配列相同性検索ツールBody-BLASTを用いたメタデータ検索機能も提供しています。		2012年3月	休止・閉鎖	データベース等	対象外	対象外	対象外	
13	PDO-CSSO	Pathogenic Disease Ontology and Clinical Signs and Symptoms Ontology	〈個別の公開ページは閉鎖。開発成果はMicrobeDB.jpへ統合されている。〉 微生物がヒトに引き起こす感染症の病名を、ヒトの体の組織ごとにまとめて記述・整理したオントロジーであるPDO (Pathogenic Disease Ontology)のOWLファイルと、感染症が引き起こす症状についてのオントロジーであるCSSO (Clinical Signs and Symptoms Ontology)のOWLファイル、さらに公共のゲノムデータのメタデータをPDO-CSSO相手にマッピングし、宿主となる生物の系統名と、その感染症に関連する細菌名とを関連づけてゲノムごとに整理した結果のRDFファイルから構成されたDBです。	http://bioportal.bioontology.org/ontologies/PDO , http://bioportal.bioontology.org/ontologies/CSSO	2013年5月	休止・閉鎖	ツール等	調整中	対象外	調整中	
14	Orthology Ontology	ORTH	オーソログ情報を記述するために一般的に利用可能なオントロジーです。	https://github.com/qfo/OrthologyOntology	2015年11月	新規	ツール等	対象外	対象外	対象外	Jesualdo Tomás Fernández-Breis, Hirokazu Chiba, Maria del Carmen Legaz-Garcia, Ikuo Uchiyama. Journal of Biomedical Semantics, 7, 34, 2016.
15	SPANG		分散型のRDFデータベースを統合的に利用するための汎用SPARQLクライアントです。SPANGから呼び出し可能なSPARQLテンプレートのライブラリも含まれています。	http://purl.org/net/spang	2015年3月	継続・発展	ツール等	対象外	対象外	対象外	