

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  
研究開発課題「ゲノム・メタゲノム情報を基盤とした  
微生物 DB の統合」

研究開発終了報告書

研究開発期間：平成23年4月～平成26年3月  
研究代表者：黒川 顕  
(東京工業大学地球生命研究所、教授)



## § 1 研究開発実施の概要

### (1) 実施概要

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「微生物エンサイクロペディア」の構築を目標とした。本研究開発では、セマンティックウェブ技術を徹底的に導入する事で、1:分類学的情報、2:菌株保存情報、3:モデル微生物における高品質データ、4:各種オミックスデータ、5:オーソログ遺伝子情報、6:メタゲノムデータ、などを統合したデータベース(DB)「MicrobeDB.jp」を開発した。具体的には、①微生物ゲノム情報の RDF 化、②MicrobeDB.jp 用の各種 Stanza の開発、③MBGD の RDF 化、④遺伝子やゲノムの各種 DB における ID 間の対応関係の RDF 化、⑤NBRC/JCM における菌株データの RDF 化、⑥メタゲノム・メタ 16S データの RDF 化、⑦各種オントロジーの開発、などセマンティックウェブ技術による DB の統合化に向け、研究グループが一丸となり各種データを RDF 化するとともに、各データを柔軟に統合するための多様なオントロジーを構築した。これら開発を経て、約 11 億トリプル数の Full RDF で記述した統合 DB「MicrobeDB.jp」を構築する事ができた。MicrobeDB.jp に実装した各 RDF とそのトリプル数は以下の通りである。

グラフ名	説明	作成元	トリプル数
refseq	RefSeq Prokaryoteゲノムデータ	DBCLS	550,273,744
mbgd	MBGD Orthologデータ	基生研	291,714,037
gtps	GTPSゲノムデータ	遺伝研	197,069,932
taxonomy	SPARQLthonで作成したNCBI Taxonomyオントロジー改良版	DBCLS, 遺伝研, 東工大	10,183,714
meta16S	各SRSメタ16Sの系統組成データ	東工大	9,831,600
gazetteer	地理オントロジー	外部機関	7,062,536
srs_metadata	SRSメタ16S・メタゲノムの様々なメタデータ	東工大	4,982,739
srs_ortholog	各SRSメタゲノムのMBGD Ortholog組成	東工大, 基生研	2,026,746
go	Geneオントロジー	外部機関	1,211,571
brc	JCM/NBRC菌株データ with NCBI Taxonomy ID	遺伝研, 東工大, DBCLS	903,319
gold	GOLDの個別ゲノムのMEO等へのオントロジーマッピングデータ	東工大, DBCLS	150,899
srs	SRSメタ16S・メタゲノムのMEO等へのオントロジーマッピングデータ	東工大	53,691
so	Sequenceオントロジー	外部機関	43,060
pdo	感染症オントロジー + 症状オントロジー + ゲノムへのオントロジーマッピングデータ	東工大	8,809
meo	微生物の生息環境オントロジー	東工大	4,975
msv	SRSメタ16S・メタゲノムのメタデータオントロジー	東工大	1,601
mcp	微生物フェノタイプオントロジー	DBCLS	734
mccv	菌株オントロジー	東工大, DBCLS	293
その他中間データ	いくつかのデータ集計系のSPARQLクエリは遅いため、MSSが集計結果のデータを作成		440,773
合計			1,075,964,773

Full RDF で記述した「MicrobeDB.jp」の利便性を向上するために、DB 利用のためのテンプレートである「Stanza」の開発を実施した。DBCLSグループとTogoStanzaの仕様を策定するとともに、本研究グループにより76種類のStanzaを開発した。DBCLSで開発された34種類のStanzaも併せてMicrobeDB.jpに実装する事で、MicrobeDB.jpを単に統計量を羅列するだけのデータベースではなく、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とする事に成功した。

また、MicrobeDB.jp内に収録しているデータの質を向上させるため、アノテーション情報の高度化を可能とするTogoAnnotationシステムの開発、遺伝子のオーソログ関係をより高精度に抽出するための新たなアルゴリズムの開発、メタゲノム断片配列を超高速にゲノム配列にマッピングする技術の開発、など、新規技術開発に基づくDBの高度化も実施した。さらに、研究者コミュニティのユーザビリティを向上させるため、さらにはデータ生産者から継続的にデータを受け付けるために、ドラフトゲノム解析を全自動で可能とする微生物ゲノム自動アノテーションシステム「MiGAP」と連携するためのゲートウェイ「GenomeRefine」を構築した。

## (2) 研究開発成果のデータベース等

- ・ 微生物統合データベース: MicrobeDB.jp
- ・ シアノバクテリアゲノムデータベース: CyanoBase
- ・ 根粒菌ゲノムデータベース: RhizoBase
- ・ 高度ゲノムアノテーションデータベース: TogoAnnotation
- ・ ゲノム再アノテーションデータベース: GTPS
- ・ 微生物オーソログ遺伝子データベース: MBDG
- ・ ゲノム・メタゲノム環境オントロジー: MEO

## § 2. 研究開発構想(および構想計画に対する達成状況)

### (1) 当初の研究開発構想

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「微生物エンサイクロペディア」の構築を目標とする。この目的のため、以下の観点から開発を行う: 1) 日本をはじめ世界中に散在している細菌(真正細菌、古細菌、バクテリオファージ)におけるゲノム、メタゲノム、トランスクリプトームなどの各種オミックス情報を広く収集する。2) これらのデータをホモロジー、オーソロジーに基づいて整理し、遺伝子、ゲノム(生物種)、環境の3つの軸に沿って整理統合するためのシステムを構築する。3) これら3つの軸に関わる、遺伝子機能、分類学的情報(16S 含む)、菌株保存情報、表現型情報(病原性、環境)などの知識を整理し、ゲノム情報を核として統合する。4) これらの微生物学上の知識について、広く研究者コミュニティからのフィードバックを得るためのシステムを開発する。5) 構築したシステムを幅広い分野の微生物学研究者に活用してもらうための使いやすいインターフェイスを作成する。特に微生物学分野のオミックス研究の発展に資することを目標とする。

この研究開発目標を達成するために、まず研究基盤となる以下の6項目の DB の整備に着手した。1: 分類学的情報、2: 菌株保存情報、3: モデル微生物(大腸菌、枯草菌、シアノバクテリア、放線菌)における高品質データ、4: 各種オミックスデータ、5: オーソログ遺伝子情報、6: メタゲノムデータ。

これらのデータの多くについては、本プロジェクトの参加グループがこれまでに開発してきたものであるため、それらをさらに充実させつつ相互に関連づけして統合していくという手順を踏む。その際、それぞれが独立に開発されてきた経緯に鑑みて、当面は ID レベルでの対応付けに基づく相互リンクによって連携をとることとするが、必要に応じてデータの内容に踏み込んだ、より深い連携をとることも検討し、全体を統一的に利用できるようにすることを目指す。

具体的に以下の手順で統合化を進める。①ゲノム決定済みのモデル微生物の研究結果や文献などの各種情報を本研究開発 DB に導入するとともに、枯草菌、大腸菌、シアノバクテリア、放線菌などのモデル微生物における研究者コミュニティとの強力な連携により、DB の高品質化を推進する。②モデル微生物ゲノムデータと、統一した基準で再アノテーションした公開微生物ゲノムデータとを併せてオーソログ遺伝子 DB を構築し、モデル微生物の高品質データを他の微生物ゲノムデータと連結させる。③これまで独立していた個別菌のゲノム情報を、メタゲノムデータを介して連結し、メタデータおよび分類学的情報、菌株保存情報と併せて、各ゲノムデータ上にマッピングする。④それぞれアップデートのタイミングが異なる DB を定期的に自動一斉アップデートし、常に最新のデータを提供可能なシステムを構築する。また各種 DB の整備および統合化に関しては、統合化推進プログラム研究総括および基盤技術開発プログラムの指導のもと、その手法を含めて検討する予定である。

### (2) 新たに追加・修正など変更した研究開発構想

研究計画を推進するにあたり、当初は ID レベルの対応付けにより独立した DB を統合しようとしていたが、基盤技術開発プログラムを実施している DBCLS の研究グループと頻りに意見交換を繰り返し、セマンティックウェブ技術を徹底的に導入し、これら多様なデータベースを統合していく事に決定した。この新たに決定した挑戦的な方針に従い、H23 年度には、まずは ID レベルの対

応付けにより微生物統合 DB「MicrobeDB.jp」のプロトタイプを立上げ、限定されたデータのみでの運用を開始する。また H24 年度には、①微生物ゲノム情報の RDF 化、②MicroboeDB.jp 用の各種 Stanza の開発、③MBGD の RDF 化、④遺伝子やゲノムの各種 DB における ID 間の対応関係の RDF 化、⑤NBRC/JCM における菌株データの RDF 化、⑥メタゲノム・メタ 16S データの RDF 化、⑦各種オントロジーの開発、などセマンティックウェブ技術による DB の統合化に向け、徹底的に各種データの RDF 化を実施するとともに、様々なオントロジーの構築を実施する。さらに H25 年度には、RDF/OWL ファイルを用いて構築した Full RDF なデータベースとなる「MicrobeDB.jp」の完成を目指す。また、MicrobeDB.jp 内に収録しているデータの質を向上させるため、アノテーション情報の高度化を可能とする TogoAnnotation システムの開発、遺伝子のオーソログ関係をより高精度に抽出するための新たなアルゴリズムの開発、メタゲノム断片配列を超高速にゲノム配列にマッピングする技術の開発、など、新規技術開発に基づく DB の高度化も実施する。さらに、研究者コミュニティのユーザビリティを向上させるため、さらにはデータ生産者から継続的にデータを受け付けるために、ドラフトゲノム解析を全自動で可能とする微生物ゲノム自動アノテーションシステム「MiGAP」と連携するためのゲートウェイ「GenomeRefine」を構築する。

(3) 達成状況

研究項目	H23 年度	H24 年度	H25 年度	変更点
1. メタゲノムデータベースの構築ならびにメタデータの整備、スパコンにおける解析システムの開発および実装 ・メタゲノムデータベースの構築 ・メタゲノムデータにおけるメタデータの整備 ・スパコンにおける解析システムの開発、実装 ・微生物 DB の統合化				・予定通り ・予定通り ・予定通り ・予定通り
2. 微生物ゲノム基盤情報資源ならびにアノテーションリファレンスの整備と共用化(遺伝研グループ) ・微生物ゲノム基盤情報資源の共用化 ・微生物ゲノムアノテーションリファレンスの整備と共用化 ・菌株保存情報の整備				・予定通り ・予定通り ・予定通り
3. 比較ゲノム解析に立脚した微生物ゲノム情報の統合化(基生研グループ) ・比較ゲノム解析に立脚した微生物ゲノム情報の統合化				・予定通り

#### (4) 研究開発の今後の展開について

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。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。

### § 3 研究開発実施体制

(1) 研究チームの体制について

①「研究代表者:黒川顕」グループ

研究参加者

氏名	所属	役職	研究開発項目	参加時期
○黒川 顕	東京工業大学地球生命研究所	教授	総括	H23.4～H26.3
小西 史一	東京工業大学大学院情報理工学研究科	特任准教授	スパコンにおける解析システムの開発	H23.4～H26.3
山田 拓司	東京工業大学大学院生命理工学研究科	講師	代謝パスウェイDBの構築	H24.4～H26.3
森 宙史	同上	助教	微生物統合DBの構築	H23.4～H26.3
*山本 希	東京工業大学地球生命研究所	産学官連携研究員	オントロジー開発およびメタゲノムDBの構築	H24.3～H26.3

②「研究分担者:中村保一」グループ

研究参加者

氏名	所属	役職	研究開発項目	参加時期
○中村 保一	国立遺伝学研究所・生命情報研究センター	教授	総括	H23.4～H26.3
菅原 秀明	同上	特任教授	システム開発	H23.4～H24.3
神沼 英里	同上	助教	データ解析	H23.4～H26.3
*藤澤 貴智	同上	特任研究員	システム開発	H23.4～H26.3
*飯田 直子	同上	特任研究員	システム開発	H25.12～H26.3

③「研究分担者:内山郁夫」グループ

研究参加者

氏名	所属	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	H23.4～H26.3
*千葉 啓和	同上	研究員	オーソログ分類の精密化、オーソログ情報のRDF化	H23.10～H26.3
西出 浩世	同上	技術職員	システム開発補助	H23.4～H26.3

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

1. 菌株保存データにおける連携  
NBRC や JCM などの菌株保存センターとデータの受け入れ、オントロジーの開発、遺伝子辞書の運用など、幅広く協力関係を築いてきた。
2. メタゲノムサンプルデータの標準化  
ゲノムやメタゲノムに付加する様々な情報についての国際標準を取り決める **Genomic Standards Consortium** と共同で、環境情報を記述する際のオントロジーやメタゲノムサンプルを記述するためのデータ形式等について共同声明を発表し、国際標準を意識した上での DB 開発に努めてきた。
3. 研究者コミュニティとの連携  
微生物なかでも細菌の研究に関連する各種学会や産業界と情報交換を促進し、学会員の要望を取り入れつつ開発を進めてきた。また、文部科学省科学研究費新学術領域研究(研究領域提案型)『生命科学系3分野支援活動』(呼称:ゲノム支援)において、多数の研究者らと細菌のゲノム・メタゲノム研究に関して共同研究を実施し、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。
4. ヒトメタゲノムデータを集積する仕組みの構築  
H25 年度に東工大内に **Japanese Consortium for Human Microbiome (JCHM)** を立ち上げ、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。
5. DBCLS との技術連携  
DBCLS が主催している **BioHackathon**、**SPARQLthon** や **STANZAthon** などのワークショップにも積極的に参加し、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た。
6. 配列アノテーションリソースのセマンティックウェブ技術連携  
DBCLS 主催の **BioHackathon** の場を中心に、DBCLS、DDBJ や UniProt などの塩基配列やタンパク質配列等の配列アノテーション情報の RDF 化やそれらを利用したアプリケーション開発に取り組んでいる研究者グループと密接に連携し、DDBJ エントリーおよびアノテーションのルールを記述するための **INSDC/DDBJ** オントロジーや配列の位置情報を記述するための **FALDO** オントロジー開発を実施した。
7. シアノバクテリア研究者との大規模ラン藻ゲノムの協調的解析のための連携  
国内のシアノバクテリア大規模ゲノム解析グループとならびに生物資源の収集、保存、配布機関である国立環境研究所(NIES)と連携し、新型シーケンサーを用いた大規模ゲノム解析を対象に協同的な解析ワークフロー構築を目指し連携している。
8. オーソログ標準形式の提案  
オーソログデータベース作成者が共通課題について話し合う国際コンソーシアムである **Quest for Orthologs** 会議へ参加し、RDF 形式による標準化を提案した。

## § 4 研究実施内容及び成果

4.1 研究課題名:メタゲノムデータベースの構築ならびにメタデータの整備、スパコンにおける解析システムの開発および実装(東京工業大学 黒川顕グループ)

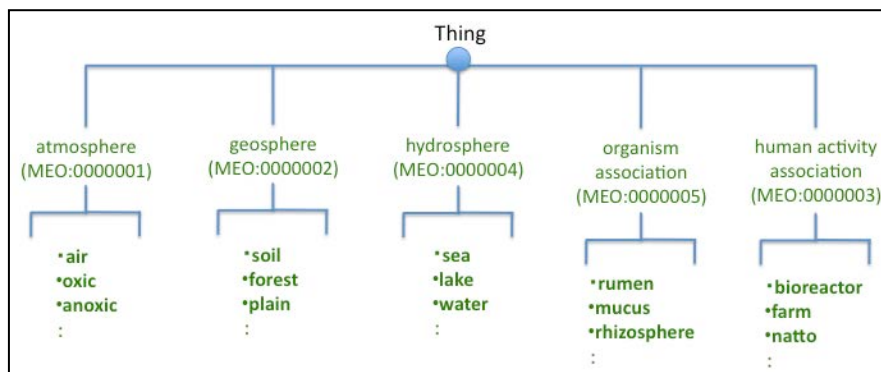
研究開発実施内容及び成果

### a. メタゲノムデータベースの構築

NCBI SRA塩基配列データベースに登録されている全てのメタ16Sデータおよびメタゲノムデータを統合DBに収録するために、それら全データを開発した解析パイプラインを用いて統一的に解析を行なった。得られた解析結果をRDF化しメタデータと統合する事により、DBの統合化が可能となる。H25年度までに、18,224のヒト共生細菌サンプル、5,827の環境細菌サンプルを取得した。SRAから取得したメタ16S配列を系統アサインメントした結果を、メタ16Sサンプルごとに系統名とその存在量を列挙した構造でRDF化した。系統の存在量は、サンプル間で系統組成を比較出来るように、サンプルごとにその系統分類階層にアサイン出来た全リード数で各系統の存在量を補正した。これらのサンプルごとの系統組成データをRDF化した結果9,831,600 トリプルとなった。また、SRAから取得したメタゲノム配列を遺伝子機能アサインメントした結果を、メタゲノムサンプルごとに遺伝子機能名とその存在量を列挙した構造でRDF化した。遺伝子機能の存在量は、各サンプル・各遺伝子機能分類における総和が1になるよう補正した。RDF化した結果3,102,678 トリプルとなった。

### b. メタデータの整備、オントロジー開発

メタデータとはメタゲノムデータに関する環境の記述であり、これが欠落しているメタゲノムデータは、単なる配列の集合体であり生物学的意味を全く持たない。メタデータのフォーマットは未だ統一されておらず、世界的に議論されている最中である。さらに、メタデータ記述中の語彙は登録者によって統一されておらず、語彙間の関係性の記述も存在しないため、語彙を単純に検索するだけでは目的の環境から採取されたメタゲノムデータを選択的に取得できない。この問題を解決するために、語彙間の階層性を定義し、構造化された語彙リストであるオントロジーを構築する必要がある。生物の生息環境についてのオントロジーは ENVO (Environmental Ontology)が既に存在する。微生物の生息環境に関する語彙は比較解析を行う上で非常に重要であるが、ENVO では生物全体の生息環境が定義されており、「anoxic」等の微生物特有の生息環境に関する語彙や、階層構造で表される生息環境の分類区分が不足している等の問題点があったため、H23 年度には、環境横断的なメタデータ定義を表現可能な新たなオントロジーMEO (Metagenome / Microbes Environmental Ontology)を開発し、H24 年度には国内外の開発者意見を取り入れた上で改良を施したMEO version0.3 をリリースした。またメタデータ名の同義語を整理したオントロジーMSV (Metagenome Sample Vocabulary)も構築した。



MEO (Metagenome / Microbes Environmental Ontology)の階層構造と語彙の例



メタデータは SRS 等の公共 DB に収録されているメタゲノムデータから取得したが、論文中にはメタデータの記述があるにもかかわらず、DB に登録された対応する配列エントリにはメタデータがほとんど記述されていないサンプルも多数存在した。そこで、メタデータをさらに充実させるために論文からメタデータを抽出した。環境細菌群集に関する PubMed ID を SRAs および SRA の XML ファイルから取得した結果、143 個の PubMed ID を取得できた。これらの PubMed ID と紐付けられたサンプル数は、全環境細菌群集サンプル 5,827 サンプル中 2,286 サンプル (39.2 %)、全ヒト共生細菌群集サンプル 18,224 サンプル中 12,365 サンプル (67.9 %) であった。これら論文から抽出したメタデータを RDF 化した結果、環境細菌に関するメタデータは 34,634 トリプル、ヒト共生細菌に関するメタデータは 40,577 トリプルとなった。

これらメタデータは、メタゲノムデータと密に連携しており、遺伝子配列データを介して他のオミックスデータおよび系統分類情報、菌株保存情報と統合し、MicrobeDB.jp において環境を表現する主要軸のひとつとなる。

#### c. メタゲノムデータのオントロジーマッピング

メタゲノムサンプルのメタデータ中の、微生物の生息環境に関する記述、地名に関する記述、宿主に関する記述をそれぞれ MEO, Gazetteer, NCBI Taxonomy に対してマッピングし、RDF 化した結果のトリプル数は 53,691 トリプルであった。また、メタゲノムサンプルのメタデータ名を MSV とマッピングし、メタデータ値と単位も RDF 化した結果、4,982,739 トリプルとなった。各オントロジーへマッピングを行なったことにより、メタデータ検索を行う際に、階層構造を用いた関連語の取得や、類義語を含めた結果取得が実現した。オントロジーマッピングを行なったことで、細菌群集の情報と環境情報を統合する基盤を構築できた。オントロジーマッピングを全て手動で行うことは困難であったため、SRS メタデータ中にオントロジー語彙と完全一致する記述があった場合は自動的にマッピングを行うツールを開発した。このツールを用いることで、ツールを用いたマッピングが不適切であると判断した部分のみ手動で修正を行い、オントロジーに不足している記述は追加する、という作業のみでマッピングを行えるようになった。

#### d. 菌株データの RDF 化

NBRC (約 16,000 株) および JCM (約 14,000 株) の菌株保存データの RDF 化を実施した。これら2つの菌株データ群はそれぞれが 15 項目のデータから構成されているが、菌株データは配列データとは完全に独立しているため、ゲノムデータなど他のデータとの統合は不可能である。菌株データをゲノムデータ等と統合するために、新たなオントロジー MCCV (Microbial Culture Collection Vocabulary) を設計し、NBRC/JCM の菌株 ID と NCBI Taxonomy ID 間を名寄せした上で、90,3319 トリプルから構成される菌株 RDF を作成した。この事により、菌株データとゲノムデータ等が完全に統合され、MicrobeDB.jp により横断的な検索が可能となった。

#### e. スパコンにおける解析システムの開発および実装

メタゲノムデータはデータ量が膨大であり、ひとつのメタゲノムプロジェクトで産出されるデータが数 TB になる事もある。そこで、メタゲノムデータのアノテーションには、東工大スパコン TSUBAME2.0 や DDBJ スパコンなどの超高速大規模計算機を積極的に利用する必要がある。H23 年度は、マルチ GPU 対応の塩基配列相同性検索ソフトウェア CLAST を開発し、既存のゲノムレファレンス配列に対するメタゲノム配列マッピングを高速に実現することができた。CLAST は、BLAST と同様の検索精度を実現しつつ BWA などマッピングツールと同様の速度で検索可能となっている。H24 年度には、開発してきた CLAST を TSUBAME2.0 および DDBJ スパコンに実装し、メタゲノム解析パイプラインをそれぞれのシステムに特化するとともに、最新のシーケンサーにも対応させた上で、メタゲノムデータのゲノムマッピングを実施し、

メタゲノム DB と他の DB との統合化を図った。

#### f. 微生物統合 DB「MicrobeDB.jp」の構築

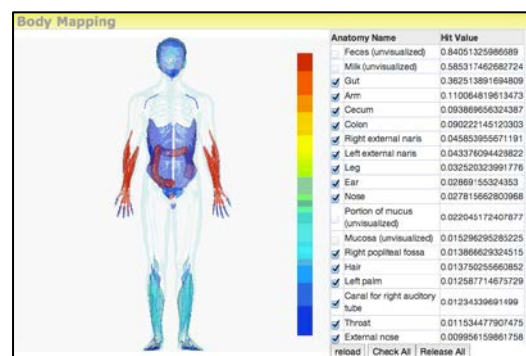
本研究開発において RDF 化した全データを統合し、微生物統合 DB「MicrobeDB.jp」を構築した。本研究では膨大な量の RDF データを格納すること、格納した膨大な RDF データに対する複雑な検索を実用時間内にレスポンスが得られることが必要だったため、これを実現できる Virtuoso を選択した。また、本 DB 利用の利便性を向上させるため、以下に挙げる各種オントロジーを構築した。

- ① FALDO (Feature Annotation Location Description Ontology): ゲノム中の各 feature の位置情報を記述するためのオントロジー (w/BioHackathon)
- ② INSDC Ontology: INSDC エントリの feature と qualifier のターム記述のためのオントロジー (w/DBCLS)
- ③ MCCV (Microbial Culture Collection Vocabulary): 菌株データを記述するためのオントロジー
- ④ MEO (Metagenome/Microbe Environmental Ontology): 細菌の生息環境を記述するためのオントロジー
- ⑤ MSV (Microbial Sample Vocabulary): メタデータの記述に用いられる語彙の表記揺れを整理したオントロジー
- ⑥ PDO/CSSO (Pathogenic Disease Ontology with Symptom): 細菌が引き起こす感染症の情報および感染症の症状を連結したオントロジー
- ⑦ GMO (Growth Media Ontology): 細菌の培地情報を記述するためのオントロジー (w/DBCLS)

さらに、MicrobeDB.jp における検索結果を表現するための Stanza の開発に着手した。Stanza は、再利用性、モジュール化、拡張が容易、という点で極めて優れたデータテンプレート(アプリケーション)であるため、統一した仕様となる TogoStanza を、MicrobeDB.jp ならびに Cyanobase をユースケースとして DBCLS と共に策定した。H24 年度中には、約 15 種類の Stanza を開発し MicrobeDB.jp に実装した。本報告書作成時点で、本研究グループにより 76 種類の Stanza を開発した。DBCLS で開発された 34 種類の Stanza も併せて 110 種類の Stanza を MicrobeDB.jp に実装する事で、MicrobeDB.jp を単に統計量を羅列するだけのデータベースではなく、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とする事に成功した。



MicrobeDB.jp トップ検索画面



HumanBodyMap Stanza

#### 全体的な成果の位置づけと類似研究との比較

MicrobeDB.jp は、微生物に関する情報をゲノム・系統・環境を軸としてセマンティックウェブ技術により統合化した DB である。このような DB は世界に類を見ないが、一方で、メタゲノムデータに特化して、研究者コミュニティからのデータを集積し、スパコンを利用した解析結果を提示する MG-RAST などの新たな DB システムも登場している。MicrobeDB.jp の

優位性は、全データを RDF にて記述した事により、今後も産出され続ける多様なデータとの統合が容易であるという点にある。この利点を生かし優位性をさらに高めるためには、MicrobeDB.jp にデータ集積のシステムを取り入れる必要がある。本研究開発においては、ゲノム自動アノテーションシステム「MiGAP」との連携を構築する事で、データ集積を促進するための基盤を開発した。本システムをさらに拡張し、メタゲノムデータに対応させるとともに、新たな Stanza 群の開発、データ更新技術開発など、DB システムの徹底した強化を実施する必要がある。

#### 4. 2 研究課題名:微生物ゲノム基盤情報資源ならびにアノテーションリファレンスの整備と共用化(国立遺伝学研究所 中村保一グループ)

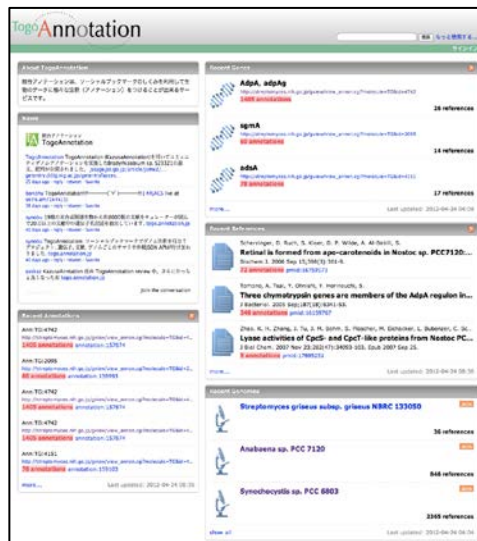
研究開発実施内容及び成果

##### a. 微生物ゲノム基盤情報資源の共用化

遺伝研・生命情報研究センターでは、我が国における大腸菌ゲノム解読の成果から始まって公開微生物ゲノム情報をGenome Information Broker (GIB)-Mとして公開してきたが、加えて、2010 年度までのバイオインフォマティクス推進事業におけるバイオ情報資源の高準化と共用化の課題において、微生物ゲノム情報の2次データベース群を構築提供してきた。すなわち、公開微生物ゲノム情報を定期的に再アノテーションし全タンパク質コード遺伝子の品質評価を加えたデータベースGene Trek in Prokaryote Space (GTPS)、公開ウイルスゲノム情報を集約したGenome Information Broker (GIB)-VならびにIS配列を集約したGIB-ISである。開発初年度には国際塩基配列データベース (INSDC) から公開されている微生物の完全ゲノム配列 3,250 件を対象に、Gene Trek in Prokaryote Space (GTPS) を再構築し、GTPS2001として微生物ゲノムの基盤情報を整備した。また、地球規模生物多様性情報機構日本ノードの協力を得て、GTPS2011のRDF化を試み、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実現した。GTPSのRDF化はライフサイエンス統合データベースセンターとの協力体制のもと、Sequence ontology とも整合する形式で作成し、197,070,005トリプルの規模で公開した。平成24年度は引き続きGTPSを更新し、そのRDF化による他のDBと他の要素データベースとの連携をすすめた。最終年度である平成25年度にはそれまでの開発に引き続き、関係するデータベース群と共用するためにRDFの記述形式を見直し、アノテーションレベルを統一した高品質なDBの共用化を促進し、利用価値を高めた。

##### b. 微生物ゲノムアノテーションリファレンスの基盤情報資源の共用化

新型シーケンサーの普及によってゲノムやトランスクリプトームの配列が従来に比較してさらに高速かつ大量に決定されるようになった現在、リファレンスとなるべきモデル微生物種群のゲノムアノテーションの信頼性を増すことは急務であると言える。とくに我が国のライフサイエンスの得意分野である、産業的に有用な微生物群のゲノムアノテーションの支援は重要な課題である。これまでかずさDNA研究所にて開発・維持されてきた、ソーシャルブックマークシステムによるデータ集積と整理を簡便に行う広域ゲノムアノテーション支援システムである「KazusaAnnotation」と、その基盤となる微生物ゲノムデータベースである「MicrobeBase」を国立遺伝学研究所内のサーバに移転するとともに拡張を実施し、本研究開発に於けるゲノム基盤として整備した。その際 KazusaAnnotation (<http://a.kazusa.or.jp>) はTogoAnnotation (<http://togo.annotation.jp>) へ名称ならびにURLを変更し、種々の最新更新情報を明示するようインターフェイスを改善した(次頁図)。



TogoAnnotation インターフェイスの刷新

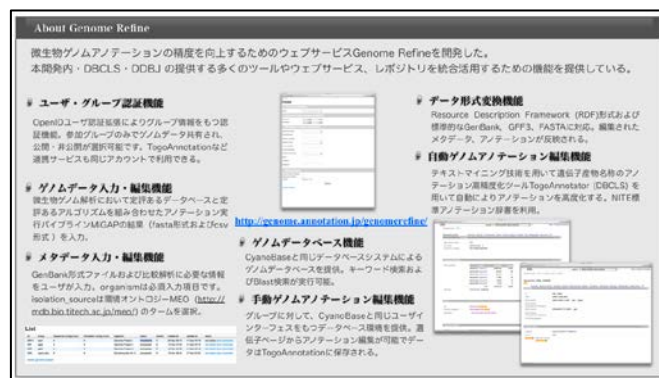
続いて、本システムを用いてリファレンスとして重要な菌株あるいは現象について、信頼性の高いマニュアルキュレーションに基づいたアノテーションの高度化を実施した。すでに大腸菌ならびに枯草菌では遺伝子に関する関連文献が集積されたデータベースが複数存在しているが、開発開始時点で文献情報の蓄積が不足しており整備が急務であると判断された放線菌 *Streptomyces coelicolor* A3(2)、*Streptomyces avermitilis* MA-4680、*Streptomyces griseus* IFO13350の3株を対象として、遺伝子および遺伝子セットについて言及した文／節を抽出 (sentence extraction)、遺伝子の名前をタグとしてさらに抽出 (named entity extraction) を手動で行うアノテーション・キュレーション方法を設計し、累計411報文から40,316注釈を登録し公開した。本システムには、本報告書作成時点で、24生物種の32,611遺伝子を対象とした、文献と遺伝子を関連付けたアノテーションとして253,749件を蓄積し公開している。

近年、微生物に於いても高速大量に集積されつつあるオミックスデータのゲノム情報への統合も課題であった。そのため、開発初年度に TogoAnnotation の API を整備し、平成24年度にはその API を活用することで、微生物オミックスデータを TogoAnnotation に格納し表現するための拡張開発を行った。具体的には、統合認証クライアントアプリケーション Tauth (<https://auth.annotation.jp>)を開発し、Excel ファイルによる TogoAnnotation へのバルク情報インポート機能を追加することで、文献データの一層の集約によるデータベース高度化を加速するとともに、MicrobeBase の遺伝子ならびに遺伝子群の表示システムである GeneView および GeneSetView ページへのオミックスデータベースエントリーの表示機能の実装と、検索インデックス作成データフローの集約整備を行った。同時に RDF リソースからの検索インデックス作成フローを追加することで、複雑な微生物オミックスデータの格納と表現を実現する環境の整備を実施し、放線菌について、上記のデータ統合整備環境を利用し、研究者コミュニティと協力してオミックスデータを取り扱う方法について検討し導入を行った。また、それ以外の我が国発のモデル微生物種群リファレンス情報として大腸菌、枯草菌をこの系に追加するための高速な論文収集モデルの開発をすすめた。

最終年度には、これまで本課題で拡張整備を実施した TogoAnnotation の安定運用を目標としたシステムのリファクタリングならびに追加開発を継続するとともに、引き続き大腸菌、枯草菌、シアノバクテリア他のモデル微生物に対してアノテーションの高品質化を実施し、同時にオミックス情報の蓄積をすすめた。データ蓄積と並行して、開発期間を通じて集積を続けたデータを情報統合のためより一層、共用化する目的で、セマンティックウェブ技術を

用いた提供環境を構築した。具体的には、CyanoBase、RhizoBaseおよび放線菌のゲノム情報を、DBCLS と連携して開発した生物の配列の位置情報記述のための FALDO オントロジーによるゲノム情報 RDF データモデルを用いて RDFに変換した。また、手動により TogoAnnotation を用いてキュレーションされた文献情報については、Open Annotation オントロジーを用いた RDF データモデルにより RDF に変換した。その他、オペロンやタンパク質複合体等の遺伝子セット情報も RDF 変換し、SPARQL エンドポイントに蓄積した(<http://genome.microbedb.jp/sparql>)。また、RDF の活用例を示すとともに実用性の高いデータ表示構造の提案を目標として、RDF をバックエンドにしてWWWページ上に自由に組み合わせて情報を表示するための情報パーツである「スタンザ」を、遺伝子セットを表示するスタンザと DNA 多型データ表示スタンザとして開発し利用に供した(<http://genome.microbedb.jp/stanza>)。

さらに、研究コミュニティからの要望により、自動ゲノム解析パイプラインMiGAPと連携し、ゲノムアノテーションを補完し、公開を支援するウェブサービスGenomeRefine(<http://genome.annotation.jp/genomerefine/>)を開発した。



GenomeRefine による認証、パイプライン、アノテーション支援技術の統合

新規に決定された複数の微生物ゲノムに関して、研究者コミュニティによるゲノムアノテーションの実施をTogoAnnotation システムと上記のグループ認証システムを活用することでサポートした。本DBの維持と精度向上に関しては研究者コミュニティとの連携が必須であり、期間を通じて連携への努力を継続しており、今後も引き続き有用な情報資源とツールを提供することで微生物研究者コミュニティとの良好な関係を維持し、ユーザーによる高品質化の永続性を目指すことが必要である。

### c. 菌株保存情報の整備

初年度に製品評価技術基盤機構と理化学研究所のそれぞれの菌株保存施設 (NITE Biological Resource Center ならびに Japan Collection of Microorganisms) の協力を得て微生物統合データベースの実現に適合した菌株データベースを構築した。具体的には、既存のリレーショナルデータベースから菌株のアクセッション番号をSubjectとして、個々の菌株の各項目のデータをできる限りURIで表現し、既存のオントロジーを活用し17,367株のデータをRDFへと変換した。その結果、菌株データ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も可能にした。次年度以降、ゲノム配列が決定された菌株情報を追加充実させると共に本研究開発で実施する情報基盤と連携させ、RDF化による他のDBとの共用化を促進した。また、日本産菌類集覧の構造化作業として、地球規模生物多様性情報機構日本ノードの協力を得て、日本菌学会関東支部が編集・出版した日本産菌類集覧を対象として、原稿ファイルから菌類データの機械抽出を試みた。機械抽出結果について菌類専門家からのアドバイスを得て、機械抽出とその後のマニュアル修正のルールを整備し、Abortiporus属からBotryobasidium属までについて属情報(属のシノニム情報、属のアナモルフ・テレオモルフ名、属のタイプ種情報)と属の下位分類の種情報(種のシノニ

ム情報、種のアナモルフ・テレオモルフ名)、種の国内採取記録文献ならびに種の宿主・基物を抽出することができた。

#### 全体的な成果の位置づけと類似研究との比較

我が国初の酸素発生型光合成細菌のシアノバクテリアの全ゲノム決定のデータを公開・共有するために1996年に作成したCyanoBaseの作成から20年弱が経過し、1998年にスタートした、ユーザーが作成したミュータント情報を自由に作成できる画期的なサービス提供をはじめとして、それに続くKazusaAnnotation (現TogoAnnotation) でのソーシャルブックマークによるゲノム上へのアノテーション付加インターフェイスの提供、またそれを用いた文献情報の大量収集によるアノテーション高度化など、本グループはゲノム上へのユーザー提供の情報を集約するためのサポートに努めて来た。マニュアルキュレーションを実施しているゲノムデータベースは多く存在するが、本課題研究ではそのコストをいかに下げて永続的な活動とするかを、一貫して追求してきた。NGS時代の到来とともに、情報収集や情報高度化の単位は、これまでの遺伝子単位からゲノム単位や大規模なオミクス実験情報単位へとシフトしようとしているが、本研究課題ではこれまで整備してきた TogoAnnotation と MicrobeBase を高度化することでそれに対応し、またその情報を MicrobeDB.jp あるいは DDBJ などの公共データベースへとさらに集約するためのフローを整備し、微生物ゲノムならびにメタゲノム情報の統合と再利用化の促進に寄与したものと考える。

### 4. 3 研究課題名:比較ゲノム解析に立脚した微生物ゲノム情報の統合化(基礎生物学研究所 内山郁夫グループ)

#### 研究開発実施内容及び成果

微生物比較ゲノムデータベース MBGD で構築しているオーソログテーブルに基づいて、種々の微生物情報を統合するための基盤を確立する。この目的のため、比較の基盤となる標準オーソログテーブルの高品質化を進めると共に、他グループと連携して、遺伝子の機能や微生物の表現型などの知見をオーソログ解析に基づいて相互に比較し、ゲノムの特徴付けなどに用いる方法などについて検討する。合わせて、基礎となるゲノムデータベースの整備を進めるとともに、微生物ゲノムデータベースのさらなる大規模化に備えたオーソログテーブルの更新体制を確立する。本研究開発の成果は、すでに公開している MBGD サービスの機能拡張として実装し提供するとともに、RDF 化したオーソログデータベースやそれを用いた表示アプリケーション(スタンザ)の提供を通じて MicrobeDB.jp データベースサービスの一部として機能させる。具体的には、以下の各項目について研究開発を実施した。

#### a. データソースの統合

今日のように爆発的にゲノムデータが蓄積する状況において、比較ゲノムデータベースにどのデータをどのように入れていくかということは、本質的に重要な戦略の一つであり、随時見直しを行っていく必要がある。本プロジェクト開始時点においては、MBGD には RefSeq の Complete genome セクションのデータのみを入れていたが、当時 RefSeq では同種別株のゲノムデータの蓄積が遅れる傾向にあったため、RefSeq に加えてオリジナルの GenBank および、遺伝研で独自にアノテーション付けして作成されている GTPS のデータを統合して用いることにした。このため、GTPS/GenBank/RefSeq の3つのデータベースの各エン트리について、染色体レベル、および遺伝子レベルでの対応付けを行い、統一プロトコルでアノテーション付けされた GTPS のアノテーション(信頼性が低い GradeX のものを除く)を中心として、不足する情報を RefSeq および GenBank から補う形で統合してデータ構築を行う体制を確立した。これにより RefSeq より充実した独自のデータセットを基盤とした比較ゲノム解析が可能になった。ただし、現在は RefSeq の更新の遅れが解消されたため、その違いは比較的小さいものとなっている。

#### b. ドラフトゲノムの拡充

シーケンスコストの低減により、爆発的にゲノムデータが生産されているが、その多くはコストを抑えてドラフト状態のまま公開されている。こうしたデータの中にも比較対象として有用なものが少なくない。そこで、RefSeq の Draft セクションのデータもデータベースに登録し、比較対象に加えられるようにした。ただし、オーソログデータを作成するに当たって、これら不完全なゲノムすべてを完全なものと同様にに入れていくことは、量の点で現実的でないだけでなく、質の点でも問題がある。そこで、中心となる標準オーソログテーブルは従来通り Complete のみで構築するが、質が比較的高く、属レベルで Complete に含まれないゲノムデータを追加したものを、拡張標準オーソログテーブルとして別途作成して提供することとした。合わせて、それ以外のドラフトデータもデータベースに格納しておき、利用者が MyMBGD 機能を通じてそれらのデータを利用できるようにした。MyMBGD は利用者が手持ちのデータをサブミットして、公開データと合わせてオンデマンドでデータを解析する機能であるが、ここに公開されたドラフトゲノムデータを含めることができる。これにより、ドラフトゲノムの利用について、更新時に必要となる処理量を減らしつつ、利用者の幅広い要求にも応えられるようにした。3 月末に公開を予定しているバージョンでは、完全ゲノム数 2749 に対してドラフトゲノム数が 5452、うち 358 種が拡張標準オーソログテーブルに追加される予定である。

#### c. 系統特異的オーソログ情報の整備

従来 MBGD では、属レベルでの重複を除いた微生物全体をカバーする生物種セットを用いた「標準オーソログテーブル」のみを事前に構築し、それ以外の生物種セットについては利用者が指定して動的にオーソログ解析を実行する方式としていた。MBGD の利用状況や利用者からの問い合わせ状況などから見て、実際に MBGD を効果的に活用している利用者の多くは、自分の興味がある生物種の近縁種を対象を絞って比較を行っていると考えられる。そこで、そうした解析をより効率的に行えるようにするため、NCBI のタクソミーデータベースを参照して、ゲノムデータがある程度蓄積した系統群(超界・門・綱・目・科・属・種)すべてについて、それらを対象としたオーソログテーブルを事前に作成して提供するようにした。これらのオーソログテーブルは、ウェブブラウザ上でタクソミーツリーに沿って利用者が興味のある系統群を選択することによって、切り替えて利用できるようにした。選択した際のトップページには、オーソロググループ全体のオーバービューとして、クラスターサイズの分布、出現パターンの分布、ゲノム間ペアワイズ比較のテーブルを切り替えて表示できるが、これらに加えて科以下の近縁系統群については、当グループで開発した CoreAligner プログラムを用いて、系統群内でシンテニー(遺伝子の並び)が保存された領域のコンセンサスな構造を構築して表示する機能も加えた。

#### d. オーソログの差分更新と MyMBGD の改良

ゲノムデータの蓄積が著しい状況において、オーソログ解析を効率的に行うことは重要な課題となっているが、有力な効率化のアプローチの一つは、既存のオーソログ分類に新しいデータを差分的に追加していくことである。我々が用いている DomClust プログラムは、ドメイン単位で階層的クラスタリングに基づく分類を行っているため、差分更新時にもドメイン分割と階層的クラスタリングツリーの再構築とを行う必要があり、これらの処理を行う MergeTree プログラムを開発した。これを用いて、標準オーソログテーブルに含まれない生物の遺伝子を、標準オーソログテーブル上に対応づけて、MBGD に登録されたすべての生物を含む拡張オーソログテーブルを作成し、提供する体制を構築した。このような差分的なデータの追加は、MyMBGD において利用者ゲノムの情報を既存のオーソログテーブルに載せて解析する際にも必要となる。そこで、MyMBGD のインターフェイスを改良して、利用者のゲノムを標準オーソログテーブル、または上述の系統特異的オーソログテーブル上に MergeTree を用いて追加し、閲覧できる機能を作成した。一方、利用者のゲノムを解析する機能は、MicrobeDB.jp プロジェクトにおいても重要な機能のひとつであるため、本プロジェクトの他グループと連携して利用者ゲノム情報を交換し、MyMBGD 機能を用いて利用者ゲノムにオー

ソロググループを割り当てたものを **MicrobeDB.jp** インターフェイスから利用できるような体制の構築も行った。

#### e. オーソログ分類の精密化

**MBGD**では、独自に開発した **DomClust** プログラムを用いてオーソログ分類を行っている。このプログラムは遺伝子融合や分裂が起きた場合に、ドメイン単位でオーソログの対応付けをするというユニークな機能を持つ。ただし、**DomClust** はホモロジー検索に基づくペアワイズ比較の結果を基にドメイン単位のクラスタリングを行っているため、アライメントの精度が低い場合には正しく分類できないケースが存在する。そこで、**DomClust** の分類結果からマルチプルアライメントを作成し、この情報を用いて分類精度を改善するプログラムの開発を行った。この目的のため、隣接ドメインを含むオーソロググループ対についてマルチプルアライメントを実行し、ドメインごとに **sum-of-pairs score (SP score)**によるアライメントスコアを計算し、その和によってドメイン単位の分類結果を評価するスコア体系(**DSP score**)を開発した。このスコア体系を用いて、**DomClust** で得られた分類結果から、隣接ドメインを含むオーソロググループ対について分類を再評価し、**DSP score** が最大になるように、マージもしくは境界を変更した再分割を行う。得られたグループについて、系統樹上の重複ノードを切断していく形でさらにグループ分割を行うことによってオーソロググループを得る。これら一連の処理をパイプラインとして実装したプログラム **DomRefine** を作成した。**DomRefine** によるオーソログドメインの再分類結果を、キュレーションにより作成された **COG** および **TIGRFAMs** データベースを用いて評価した結果、全体的には元の **DomClust** 分類と比べて期待通りに改善されていることを確認した。また、**DomClust/DomRefine** による分類が、**COG** を拡張して作成されている **eggNOG**より優れた分類になっているとの結果も得た。これらの成果は論文にまとめて現在投稿中である。本手続きは大規模なマルチプルアライメントや系統樹作成を含むため、計算時間は大きいが並列計算による高速化は可能であり、今後 **MBGD** の標準オーソログテーブルについて精密化を行い、マルチプルアライメントや系統樹の情報も含めて提供できるようにする予定である。

#### f. オーソロググループ情報の RDF 化

オーソログの情報は、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と統合することにより、それらの知識を相互に結びつけることが可能になる。この目的のため、**MBGD** に含まれるオーソロググループやその他の関連情報の **RDF** 化を行い、**MicrobeDB.jp** プロジェクトで収集している他のデータと統合していくことを試みた。オーソロググループの **RDF** 化に際しては、将来的に様々な他のオーソログデータベース間で概念を共有することも視野に入れて、汎用性を考慮した設計を行った。このため、オーソログデータを表現する共通形式として提唱されている **OrthoXML** の枠組みを参考にして、オーソログに関する基本的なオントロジーを定義し、それに含まれない概念を **MBGD** の独自オントロジーとして定義した。また、**RDF**化した **MBGD** データについては、独自に **SPARQL** エンドポイントを立ち上げて、**SPARQL**を用いて検索できるようにした。**RDF/SPARQL**に関する技術は必ずしも成熟していない部分もあるため、**DBCLS** で定期的開催された **SPARQLthon** に参加して先端的な技術情報を共有しつつ、**SPARQL** を使って他のデータベースと連携させた検索や推論を使ったタキソミーデータの検索などについて検討した。これらに基づき、**MicrobeDB.jp** のブラウザ上でオーソログデータを検索して表示するための基本的なアプリケーション(スタンザ)をいくつか作成した。

#### 全体的な成果の位置づけと類似研究との比較

**MBGD** は微生物に特化したオーソログデータベースで、微生物全体にまたがる比較から、個別菌研究者に向けた同種・同属間の比較までを満遍なくサポートしたものは他に類を見ない。この特徴は、今回の系統特異的オーソログ情報の整備や **MyMBGD** 機能の拡張によって、より一層強化された。遺伝子ではなくドメイン単位でオーソログ分類を行っていることも、



MBGD のユニークな特徴であるが、DomRefine の開発によってより高品質の分類が可能になり、リファレンスデータベースとしての価値が高まった。オーソログ情報の RDF 化については世界的に見ても緒についたばかりであるが、関連研究者との情報交換を含めて先導的な役割を果たせる位置にある。ゲノムデータベースの大規模化への対応についても一定の成果を上げたが、将来性の観点からはさらなる強化が必要である。微生物ゲノムデータ解析の統合的なインターフェイスについては IMG や MicrobesOnline など優れたものがいくつか存在するが、今後 MBGD を使った優れた解析インターフェイスを MicrobeDB.jp 上のスタンザとして提供していくことで対抗していけるものと考えている。

## § 5 成果発表等

(1) 原著論文発表 (国内(和文)誌 6 件、国際(欧文)誌 6 件)

1. Éamonn Ó Tuama, John Deck, Gabriel Dröge, Markus Döring, Dawn Field, Renzo Kottmann, Juncai Ma, Hiroshi Mori, Norman Morrison, Peter Sterk, Hideaki Sugawara, John Wiczorek, Linhuan Wu, and Pelin Yilmaz, "Meeting Report: Hackathon-Workshop on Darwin Core and MIXS Standards Alignment (February 2012)", *Standards in Genomic Science*, 7, 1, 2012. (DOI: 10.4056/sigs.3166513).
2. Okubo T, Tsukui T, Maita H, Okamoto S, Oshima K, Fujisawa T, Saito A, Futamata H, Hattori R, Shimomura Y, Haruta S, Morimoto S, Wang Y, Sakai Y, Hattori M, Aizawa S, Nagashima KV, Masuda S, Hattori T, Yamashita A, Bao Z, Hayatsu M, Kajiya-Kanegae H, Yoshinaga I, Sakamoto K, Toyota K, Nakao M, Kohara M, Anda M, Niwa R, Jung-Hwan P, Sameshima-Saito R, Tokuda S, Yamamoto S, Yamamoto S, Yokoyama T, Akutsu T, Nakamura Y, Nakahira-Yanaka Y, Takada Hoshino Y, Hirakawa H, Mitsui H, Terasawa K, Itakura M, Sato S, Ikeda-Ohtsubo W, Sakakura N, Kaminuma E, Minamisawa K., "Complete genome sequence of Bradyrhizobium sp. S23321: insights into symbiosis evolution in soil oligotrophs", *Microbes Environ*, vol. 27, No. 3, pp 306-315, 2012. (DOI: 10.1264/jsme2.ME11321)
3. Ikuo Uchiyama, Motohiro Mihara, Hiroyo Nishide, Hirokazu Chiba. "MBGD update 2013: the microbial genome database for exploring the diversity of microbial world" *Nucleic Acids Res*, vol 41, pp. D631-D635, 2013. (doi:10.1093/nar/gks1006)
4. Toshiaki Katayama, Mark D Wilkinson, Gos Micklem, Shuichi Kawashima, Atsuko Yamaguchi, Mitsuteru Nakao, Yasunori Yamamoto, Shinobu Okamoto, Kenta Oouchida, Hong-Woo Chun, Jan Aerts, Hammad Afzal, Erick Antezana, Kazuharu Arakawa, Bruno Aranda, Francois Belleau, Jerven Bolleman, Raoul JP Bonnal, Brad Chapman, Peter JA Cock, Tore Eriksson, Paul MK Gordon, Naohisa Goto, Kazuhiro Hayashi, Heiko Horn, Ryosuke Ishiwata, Eli Kaminuma, Arek Kasprzyk, Hideya Kawaji, Nobuhiro Kido, Young Joo Kim, Akira R Kinjo, Fumikazu Konishi, Kyung-Hoon Kwon, Alberto Labarga, Anna-Lena Lamprecht, Yu Lin, Pierre Lindenbaum, Luke McCarthy, Hideyuki Morita, Katsuhiko Murakami, Koji Nagao, Kozo Nishida, Kunihiro Nishimura, Tatsuya Nishizawa, Soichi Ogishima, Keiichiro Ono, Kazuki Oshita, Keun-Joon Park, Pjotr Prins, Taro L Saito, Matthias Samwald, Venkata P Satagopam, Yasumasa Shigemoto, Richard Smith, Andrea Splendiani, Hideaki Sugawara, James Taylor, Rutger A Vos, David Withers, Chisato Yamasaki, Christian M Zmasek, Shoko Kawamoto, Kosaku Okubo, Kiyoshi Asai and Toshihisa Takagi The 3rd DBCLS BioHackathon: improving life science data integration with Semantic Web technologies. *Journal of Biomedical Semantics* 4:6. 2013. (doi:10.1186/2041-1480-4-6)
5. Fujisawa T, Okamoto S, Katayama T, Nakao M, Yoshimura H, Kajiya-Kanegae H, Yamamoto S, Yano C, Yanaka Y, Maita H, Kaneko T, Tabata S, Nakamura Y. CyanoBase and RhizoBase: databases of manually curated annotations for cyanobacterial and rhizobial genomes. *Nucleic Acids Res.*, 42(1): D666-670. 2014.

(doi: 10.1093/nar/gkt1145)

6. Toshiaki Katayama, Mark D Wilkinson, Kiyoko F Aoki-Kinoshita, Shuichi Kawashima, Yasunori Yamamoto, Atsuko Yamaguchi, Shinobu Okamoto, Shin Kawano, Jin-Dong Kim, Yue Wang, Hongyan Wu, Yoshinobu Kano, Hiromasa Ono, Hidemasa Bono, Simon Kocbek, Jan Aerts, Yukie Akune, Erick Antezana, Kazuharu Arakawa, Bruno Aranda, Joachim Baran, Jerven Bolleman, Raoul JP Bonnal, Pier Luigi Buttigieg, Matthew P Campbell, Yi-an Chen, Hirokazu Chiba, Peter JA Cock, Kevin B Cohen, Alexandru Constantin et al. BioHackathon series in 2011 and 2012: penetration of ontology and linked data in life science domains. *Journal of Biomedical Semantics*, 5:5, 2014. (doi:10.1186/2041-1480-5-5)

(2)その他の著作物(総説、書籍など)

1. 黒川顕, 微生物統合データベース MicrobeDB.jp の構築とその意義, 日本ゲノム微生物学会ニュースレター, No.4, 2011.
2. 黒川顕, 微生物統合データベース”MicrobeDB.JP”, 細胞工学, 30, 12, 2011.
3. 中尾光輝, 藤澤貴智, 解析プラットフォーム紹介, 実験医学増刊, Vol.29, No.15, 使えるデータベース・ウェブツール, 有田正規, 羊土社, 2011.
4. 森宙史, 山田拓司, 黒川顕, メタゲノム解析の現状と将来 知識データベースの開発, 情報管理, 55, 3, 2012.
5. Lynn Schriml, Ilene Mizrachi, Peter Sterl, Dawn Field, Lynette Hirschman, Tatiana Tatusova, Susanna Sansone, Jack Gilbert, David Schindel, Neil Davies, Chris Meyer, Folker Meyer, George Garrity, Lite Proctor, Marnix Medema, Yemin Lan, Anna Klindworth, Frank Oliver Glöckner, Tonia Korves, Antonia Gonzalez, Peter Dwayndt, Markus Göker, Anjette Johnston, Evangelos Pafilis, Susanne Schneider, Katherine Baker, Cynthia Parr, Granger Sutton, Heather Creasy, Nikos Kyrpides, K. Eric Wommack, Patricia L. Whetzel, Daniel Nasko, Hilmar Lapp, Takamoto Fujisawa, Adam M. Phillippy, Renzo Kottman, Judith A. Blake, Junhua Li, Judith A. Blake, Elizabeth M. Glass, Lynn Schriml, Ilene Mizrachi, Petra ten Hoopen, Jack Gilbert, Rob Knight, Susan Holmes, Curtis Huttenhower, Steven L. Salzberg, Bing Ma, Owen White. The 15th Genomic Standards Consortium meeting. *Stand. Genomic Sci.* 8:124-164, 2013. (doi:10.4056/sigs.3457)
6. Federhen S, Clark K, Barrett T, Parkinson H, Ostell J, Kodama Y, Mashima J, Nakamura Y, Cochrane G, Karsch-Mizrachi I. Toward richer metadata for microbial sequences: replacing strain-level NCBI taxonomy taxids with BioProject, BioSample and Assembly records. *Stand. Genomic Sci.*, 9(3), 2014. (doi:10.4056/sigs.4851102)

(3)国際学会発表及び主要な国内学会発表

① 招待講演(国内会議 15 件、国際会議 4 件)

〈国内〉

1. 黒川顕, 新型シーケンサーによるメタゲノム解析, 第 63 回日本生物工学会大会, 東京, 2011 年 9 月 27 日.
2. 黒川顕, ゲノムからメタゲノムへ, CBI/JSBi2011 合同大会, 神戸, 2011 年 11 月 9 日.
3. 内山郁夫, 原核・真核微生物の全ゲノム比較データベース MBGD, 微生物系統分類研究会、理化学研究所, 2011 年 11 月 25 日.
4. 黒川顕, 生態系の理解に向けたゲノム情報の解析と活用, 農業環境技術研究所公開セミナー, つくば, 2011 年 12 月 6 日.
5. 黒川顕, 新型シーケンサーによるメタゲノム解析, 第 12 回 iPAB(並列生物情報処理イニシアティブ)シンポジウム, 東京, 2011 年 12 月 9 日.
6. 森宙史, MicrobeDB.jp プロジェクトチーム, データ標準化の最新動向・微生物統合など, 第 5 回 NBRP データベース研究会, 国立遺伝学研究所・研究員宿泊施設2階セミナー室, 2012 年 3 月 22 日.

7. 黒川顕, 環境メタゲノミクス, 日本農芸化学会 2012 年度大会シンポジウム, 京都, 2012 年 3 月 25 日.
8. 内山郁夫, メタゲノム情報と系統分類, 日本進化学会第14回大会, 首都大学東京南大沢キャンパス, 2012 年 8 月 21-23 日.
9. 黒川顕, Development of an integrated analysis system for metagenomics and a global microbial database MicrobeDB.jp, 第 28 回日本微生物生態学会シンポジウム, 豊橋, 2012 年 9 月 20 日.
10. 黒川顕, メタゲノミクスの現状と未来, 情報・システム研究機構シンポジウム 2012「生命科学のビッグデータ革命」, 東京, 2012 年 11 月 9 日.
11. 内山郁夫, 原核・真核微生物の全ゲノム比較データベースMBGD, 微生物系統分類研究会, 理化学研究所, 2012 年 11 月 25 日.
12. 黒川顕, メタゲノム解析の現状と将来, 社団法人企業研究会 CAMM フォーラム, 東京, 2013 年 1 月 11 日.
13. 森宙史, メタゲノムインフォマティクスと微生物統合 DB, 第 41 回 GCOE 談話会, 東京大学情報生命科学実験棟 2 階講義室, 2013 年 7 月 12 日.
14. 中村保一, ソーシャルゲノムアノテーション, 明治大学科学技術研究所シンポジウム「大規模オミックス情報解析がもたらす生命科学の新たな展開」, 川崎, 2013 年 9 月 2 日.
15. 内山郁夫, CoreAligner: 系統群内で保存されたゲノムコア構造を構築するためのツール, 微生物系統分類研究会, 高山グリーンホテル, 2013 年 11 月 14-15 日.

〈国際〉

1. Ken Kurokawa, Development of an integrated analysis system for metagenomics and its applications for soil metagenome analysis, BiWO2011 (Bioinformatics Week in Odaiba 2011), 東京, 2012 年 1 月 26 日
2. Ken Kurokawa, Deciphering ancient ecosystems with environmental genomics and “Earth-Life DB”, The 1st ELSI International Symposium, 東京, 2013 年 3 月 29 日.
3. Yasukazu Nakamura, Cyanobase: an integrated database for cyanobacterial genomes, ESF conference on Molecular Bioenergetics of Cyanobacteria, Pultusk, Poland, 2013 年 4 月 17 日.
4. Yasukazu Nakamura, Integration of databases for microbes and plants from the viewpoint of (meta-)genomics, Joint Conference on Informatics in Biology, Medicine and Pharmacology (生命医薬情報学連合大会), 東京, 2013 年 10 月 15 日

② 口頭発表(国内会議 18 件、国際会議 7 件)

〈国内〉

1. 藤澤貴智, CyanoBaseおよびKazusa Annotation Suiteの今後の展開, ラン藻ゲノム研究交流会2011, 東京, 2011年7月2日.
2. 黒川顕, ゲノム・メタゲノム情報を基盤とした微生物DBの統合, 日本ゲノム微生物学会, 仙台, 2011年8月21日.
3. 黒川顕, 微生物エンサイクロペディアの構築に向けて, トーゴの日シンポジウム2011, 東京, 2011年10月5日.
4. Ikuo Uchiyama, A method for orthology and taxonomy classification in metagenomic data analysis, 第34回日本分子生物学会年会, パシフィコ横浜, 2011年12月13-16日
5. 内山郁夫, オーソログ解析に基づく新規ゲノム・メタゲノムデータのアノテーション, 第6回日本ゲノム微生物学会年会, 立教大学, 2012年3月10-12日.
6. 森宙史, MicrobeDB.jp プロジェクトチーム, データ標準化の最新動向・微生物統合など, 第5回NBRPデータベース研究会, 国立遺伝学研究所・研究員宿泊施設2階セミナー室, 2012年3月22日.
7. 藤澤貴智, 中村保一, TogoAnnotation を用いたゲノムアノテーションの利用とデータ共有,

- ラン藻ゲノム交流会 2012, 駒場, 2012年7月14日.
- 千葉啓和, 内山郁夫, 微生物ゲノムデータベース MBGD におけるオーソログ分類の改良, 日本進化学会第14回大会, 首都大学東京南大沢キャンパス, 2012年8月21-23日.
  - 内山郁夫, シンテニーの保存性に基づく微生物系統群のコアゲノム構造, 第35回日本分子生物学会年会, 福岡国際会議場, 2012年12月11-14日.
  - 千葉啓和, 内山郁夫, マルチプルアライメントを用いたドメイン単位のオーソログ分類の改良, 生命情報科学若手の会, 岡崎コンファレンスセンター, 2013年3月1-3日.
  - 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, 第7回日本ゲノム微生物学会年会, 長浜バイオ大学, 2013年3月8-10日.
  - 内山郁夫, 西出浩世, 三原基広, 千葉啓和, シンテニーに基づくコアゲノムデータベース, 第7回日本ゲノム微生物学会, 長浜バイオ大学, 2013年3月8-10日.
  - 藤澤貴智, 照井敬子, 桧原直子, 山田佐知子, 石井英治, 加藤香奈, 手塚武揚, 大西康夫, 神沼英里, 岡本忍, 黒川顕, 中村保一, 放線菌ゲノムアノテーションプラットフォームの構築とリファレンス情報の統合, 日本農芸化学会 2013年度大会, 仙台, 2013年3月27日.
  - 山本希, 森宙史, 中村保一, 内山郁夫, 菅原秀明, 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 の開発, 日本農芸化学会 2013年度大会, 東北大学川内北キャンパス, 2013年3月24-27日.
  - 千葉啓和, 西出浩世, 内山郁夫, オーソログ情報の RDF 化とその検索, 第17回オープンバイオ研究会, 北陸先端大, 2013年3月29日~30日
  - 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, 第7回日本細菌学会若手コロッセウム, 広島フォレストヒルズガーデン, 2013年8月7-9日.
  - 藤澤貴智, 中村保一, ゲノム基盤情報のためのオントロジー開発と CyanoBase2013, ラン藻ゲノム研究交流会 2013, 東京, 2013年7月13日.
  - 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 の開発, 第7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, 2013年9月19-20日.

〈国際〉

- Ikuo Uchiyama, A rapid incremental update procedure for ortholog grouping and its application to metagenomic data, Quest for Orthologs Meeting 2011, Wellcome Trust Conference Centre, Hinxton, UK, 17-19, June, 2011.
- Ikuo Uchiyama, A rapid procedure for large-scale ortholog assignment and its application to metagenomic data. SMCB 2011, Kyoto Univ, 27-30, July, 2011.
- Hiroshi Mori, Environmental contextualisation in MicrobeDB.jp, The 13th Workshop of the Genomic Standards Consortium, Kingkey Palace Hotel, Shenzhen, China, March 5-7, 2012.
- Ikuo Uchiyama, Enhancement of the Microbial Genome Database for Comparative Analysis (MBGD). Quest for Orthologs Meeting 2013, University of Lausanne, Jul 24-26, 2013.
- Hirokazu Chiba, Ikuo Uchiyama, Construction of ortholog database using the semantic web technology. Quest for Orthologs Meeting 2013, University of Lausanne, Jul 24-26, 2013.
- Yasukazu Nakamura, TogoAnnotation / CyanoBase: Social genome curation for microbial reference genome annotation improvement, CSH Asia meeting: Frontiers in Bioinformatics and Computational Biology, 中国蘇州, 2013年9月25日.
- Hiroshi Mori, Data integration in the MicrobeDB.jp using Semantic Web technology, WDCM and CODATA Joint Workshop, Beijing Friendship Hotel, Beijing, China, Sept. 26, 2013.

③ ポスター発表(国内会議 28 件、国際会議 3 件)

〈国内〉

1. 藤澤貴智, 中尾光輝, 岡本忍, 山本泰智, 中村保一, 大規模なゲノムアノテーション情報の手動による精度と再利用性の向上, 日本ゲノム微生物学会ワークショップ, 仙台, 2011 年 8 月 20 日
2. 黒川顕, 中村保一, 内山郁夫, 菅原秀明, ゲノム・メタゲノム情報を基盤とした微生物 DB の統合, トーゴの日シンポジウム 2011, 日本科学未来館, 2011 年 10 月 5 日
3. 藤澤貴智, 岡本忍, 神沼英里, 菅原秀明, 中村保一, 微生物ゲノムアノテーションリファレンスの整備と共用化, トーゴの日シンポジウム 2011, 東京, 2011 年 10 月 5 日
4. 藤澤貴智, 岡本忍, 神沼英里, 菅原秀明, 内山郁夫, 黒川顕, 中村保一, 「微生物エンサイクロペディア」と CyanoBase における微生物ゲノムアノテーションリファレンス整備にむけた取り組み, ラン藻の分子生物学 2011, 東京, 2011 年 12 月 3 日
5. 東京工業大学, 国立遺伝学研究所, 基礎生物学研究所, 微生物統合データベース: MicrobeDB.jp, 第 34 回日本分子生物学会年会, パシフィコ横浜, 2011 年 12 月 13-16 日
6. 藤澤貴智, 岡本忍, 神沼英里, 菅原秀明, 内山郁夫, 黒川顕, 中村保一, 統合微生物ゲノムアノテーションリファレンスの整備と共用化に向けた取り組み, 第 34 回日本分子生物学会年会, 横浜, 2011 年 12 月 14 日
7. MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp, 第 6 回日本ゲノム微生物学会年会, 立教大学池袋キャンパス, 2012 年 3 月 10-12 日
8. 森宙史, 岡本忍, 川島秀一, 竹原潤一, 吉野弘二, MicrobeDB.jp プロジェクトチーム, 微生物の棲息環境データの記述・整理のためのオントロジー構築, 第 6 回日本ゲノム微生物学会年会, 立教大学池袋キャンパス, 2012 年 3 月 10-12 日.
9. 藤澤貴智, 岡本忍, 照井敬子, 桧原直子, 加藤香奈, 石井英治, 神沼英里, 菅原秀明, 内山郁夫, 黒川顕, 中村保一, TogoAnnotation: ゲノムアノテーションリファレンス情報集積システム整備と放線菌アノテーションの試み, 第6回ゲノム微生物学会, 東京, 2012 年 3 月 10 日
10. 千葉啓和, 内山郁夫, 微生物比較ゲノムデータベース MBGD におけるオーソログ分類の改良:ドメイン境界の精緻化を中心に, 第6回ゲノム微生物学会年会, 立教大学池袋キャンパス, 2012 年 3 月 10-12 日
11. 森宙史, 岡本忍, 川島秀一, 竹原潤一, 吉野弘二, 山本希, MicrobeDB.jp プロジェクトチーム, 微生物の生息環境データの記述および高度な検索のためのオントロジー構築, トーゴの日シンポジウム 2012, 時事通信ホール, 2012 年 10 月 5 日
12. 黒川顕, 中村保一, 内山郁夫, 菅原秀明, MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp, トーゴの日シンポジウム 2012, 時事通信ホール, 2012 年 10 月 5 日
13. 黒川顕, 中村保一, 内山郁夫, 菅原秀明, MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp, BioJapan2012, パシフィコ横浜, 2012 年 10 月 10-12 日
14. 黒川顕, 中村保一, 内山郁夫, 菅原秀明, MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp, 第 35 回日本分子生物学会年会, 福岡国際会議場・マリンメッセ福岡, 2012 年 12 月 11-14 日
15. 藤澤貴智, 岡本忍, 長崎英樹, 神沼英里, 菅原秀明, 内山郁夫, 黒川顕, 中村保一, TogoAnnotation: 研究コミュニティのためゲノムアノテーションプラットフォーム構築, 第 35 回日本分子生物学会年会, 福岡, 2012 年 12 月 14 日
16. 千葉啓和, 内山郁夫, 微生物ゲノムデータベース MBGD におけるオーソログ分類の改良, 第 35 回日本分子生物学会年会, 福岡国際会議場, 2012 年 12 月 11-14 日
17. 森宙史, 岡本忍, 川島秀一, 竹原潤一, 吉野弘二, 山本希, MicrobeDB.jp プロジェクトチーム, 微生物の生息環境データの記述および高度な検索のためのオントロジー構築, トーゴの日シンポジウム 2012, 時事通信ホール, 2012 年 10 月 5 日.

18. 黒川顕, 中村保一, 内山郁夫, 菅原秀明, MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp, トーゴーの日シンポジウム 2012, 時事通信ホール, 2012 年 10 月 5 日.
19. 黒川顕, 中村保一, 内山郁夫, 菅原秀明, MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp, 第 7 回日本ゲノム微生物学会年会, 長浜, 2013 年 3 月 8-10 日.
20. 藤澤貴智, 神沼英里, 岡本忍, 菅原秀明, 内山郁夫, 黒川顕, 中村保一, TogoAnnotation: 研究コミュニティ支援ゲノムアノテーションプラットフォームの整備とセマンティックウェブ技術利用の取り組み, 第 7 回日本ゲノム微生物学会, 長浜, 2013 年 3 月 8-10 日
21. 千葉啓和, 内山郁夫, マルチプルアラインメントを用いたドメイン単位のオーソログ分類の改良, 第 7 回日本ゲノム微生物学会, 長浜, 2013 年 3 月 8-10 日
22. 黒川顕, 中村保一, 内山郁夫, 菅原秀明, MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp, 日本農芸化学会 2013 年度大会, 東北大学川内北キャンパス, 2013 年 3 月 24-27 日.
23. 黒川顕, 中村保一, 内山郁夫, 菅原秀明, MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp, トーゴーの日シンポジウム 2013, 時事通信ホール, 2013 年 10 月 4-5 日.
24. 山本希, 岡本忍, 川島秀一, 竹原潤一, 吉野弘二, 森宙史, MicrobeDB.jp プロジェクトチーム, 微生物の生息環境データの記述および高度な検索のためのオントロジー構築, トーゴーの日シンポジウム 2013, 時事通信ホール, 2013 年 10 月 4-5 日.
25. 藤澤貴智, Genome Refine: 微生物ゲノム基盤情報ならびにアノテーションリファレンスの RDF を活用したゲノムアノテーションサービス開発, トーゴーの日シンポジウム 2013, 東京, 2013 年 10 月 5-6 日.
26. Chiba Hirokazu, Ikuo Uchiyama, A method to improve domain-based ortholog clustering. 日本バイオインフォマティクス学会 2013 年年会, タワーホール船堀, 2013 年 10 月 29-31 日
27. 藤澤貴智, 森宙史, 岡本忍, 山本泰智, 片山俊明, 川島 秀一, 谷澤靖洋, 神沼英里, 大山彰, 菅原秀明, 内山郁夫, 黒川顕, 中村保一, ゲノムアノテーションの精緻化と解析支援のためのサービス, 第 35 回日本分子生物学会, 神戸, 2013 年 12 月 3 日.
28. 千葉啓和, 内山郁夫, ドメイン単位のオーソログ分類の改良法, 第 36 回日本分子生物学会年会, 神戸国際会議場, 2013 年 12 月 3-6 日.

〈国際〉

1. Takatomo Fujisawa, Hideki Nagasaki, Yasukazu Nakamura, Introduction of a genome annotation platform for the Marchantia research community, Marchantia Workshop 2012, Aso, Kumamoto, Japan, November 15-17, 2012.
2. Toshiaki Katayama, Shinobu Okamoto, Shuichi Kawashima, Hiroshi Mori, Takamoto Fujisawa, Semantic Integration of the Genome Annotations, Semantic Web Applications and Tools for Life Sciences 2012, Paris, France, November 28-30, 2012.
3. Takatomo Fujisawa, Toshiaki Katayama, Shinobu Okamoto, Hiroshi Mori, Yasunori Yamamoto, Yasukazu Nakamura, Development of an ontology for the INSDC Feature Table Definition, The 15th Genomic Standards Consortium meeting, Bethesda, US, April 22-24, 2013.

(4)知財出願

①国内出願  
なし

②海外出願  
なし

③その他の知的財産権  
なし

(5)受賞・報道等

①受賞(顕著な受賞の前に\*を付記してください)

\*Linked Open Data チャレンジ Japan 2011 審査員特別賞「ライフサイエンス賞」,  
MicrobeDB.jp プロジェクトチーム, 応募作品 Metagenome/Microbes Environmental  
Ontology, 2012年3月8日.

②マスコミ(新聞・TV等)報道(プレス発表をした場合にはその概要もお書き下さい。)

③その他

## § 6 研究開発期間中に主催した会議等

主なワークショップ、シンポジウム、アウトリーチ等の活動  
アウトリーチ(学会等ブース出展)

年月日	名称	場所	参加人数	概要
2011年12月13-16日	第34回日本分子生物学学会年会	パシフィコ横浜		微生物統合データベース: MicrobeDB.jp
2012年10月10-12日	BioJapan2012	パシフィコ横浜		微生物統合データベース: MicrobeDB.jp
2012年12月11-14日	第35回日本分子生物学学会年会	福岡国際会議場・マリンメッセ福岡		微生物統合データベース: MicrobeDB.jp
2013年3月8-10日	第7回日本ゲノム微生物学会年会	長浜バイオ大学		微生物統合データベース: MicrobeDB.jp
2013年3月24-27日	日本農芸化学会 2013年度大会	東北大学川内北キャンパス		微生物統合データベース: MicrobeDB.jp
2013年12月3-5日	第36回日本分子生物学学会年会	神戸ポートアイランド	1	微生物統合データベース: MicrobeDB.jp
2013年10月9-11日	BioJapan 2013	パシフィコ横浜		微生物統合データベース: MicrobeDB.jp

チーム内ミーティング

年月日	名称	場所	参加人数	概要
2011年4月19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	9人	研究進捗報告のための会議
2011年7月20日	チーム内第2回全体会議(非公開)	DBCLS	9人	研究進捗報告のための会議
2011年8月11日	チーム内 RDF 技術会議(非公開)	DBCLS	7人	RDF DB 開発のための会議

2011年9月16日	チーム内 RDF 技術会議(非公開)	DBCLS	7人	RDF DB 開発のための会議
2011年11月5日	チーム内 DB 開発会議(非公開)	DBCLS	4人	DB 開発のための会議
2011年11月11日	チーム内オントロジー開発会議(非公開)	DBCLS	10人	オントロジー開発のための会議
2011年12月23日	チーム内 DB 開発会議(非公開)	DBCLS	6人	DB 開発のための会議
2012年1月11日	チーム内第3回全体会議(非公開)	DBCLS	22人	研究進捗報告のための会議
2012年1月18日	チーム内オントロジー開発会議(非公開)	DBCLS	5人	オントロジー開発のための会議
2012年2月11日	チーム内オントロジー開発会議(非公開)	DBCLS	4人	オントロジー開発のための会議
2012年3月21日	チーム内第4回全体会議(非公開)	DBCLS	15人	研究進捗報告のための会議
2012年4月19日	チーム内オントロジー開発会議(非公開)	DBCLS	6人	オントロジー開発のための会議
2012年4月24日	チーム内オントロジー開発会議(非公開)	DBCLS	6人	オントロジー開発のための会議
2012年4月27日	チーム内 DB 開発会議(非公開)	東工大すずかけ台	4人	DB 開発のための会議
2012年5月10日	チーム内オントロジー開発会議(非公開)	DBCLS	9人	オントロジー開発のための会議
2012年5月30日	チーム内第5回全体会議(非公開)	DBCLS	19人	研究進捗報告のための会議
2012年7月5日	チーム内 DB 開発会議(非公開)	東工大すずかけ台	4人	DB 開発のための会議
2012年7月13日	チーム内第6回全体会議(非公開)	DBCLS	19人	研究進捗報告のための会議
2012年8月2日	チーム内 DB 開発会議(非公開)	DBCLS	6人	DB 開発のための会議
2012年8月16日	チーム内 DB 開発会議(非公開)	DBCLS	5人	DB 開発のための会議
2012年8月30日	チーム内オントロジー開発会議(非公開)	DBCLS	4人	オントロジー開発のための会議
2012年9月14日	チーム内 DB 開発会議(非公開)	東工大すずかけ台	4人	DB 開発のための会議
2012年10月3日	チーム内 DB 開発会議(非公開)	DBCLS	6人	オントロジー開発のための会議
2012年10月19日	チーム内第7回全体会議(非公開)	DBCLS	17人	研究進捗報告のための会議
2012年10月25日	チーム内 DB 開発会議(非公開)	DBCLS	9人	DB 開発のための会議
2012年11月21日	チーム内 DB 開発会議(非公開)	東工大すずかけ台	4人	DB 開発のための会議



2013年1月14日	チーム内 DB 開発会議 (非公開)	東工大田町キャンパス	6人	DB 開発のための会議
2013年3月15日	チーム内 DB 開発会議 (非公開)	DBCLS	8人	DB 開発のための会議
2013年5月15日	チーム内第8回全体会議 (非公開)	DBCLS	13人	研究進捗報告のための会議
2013年6月6日	チーム内 DB 開発会議 (非公開)	DBCLS	8人	DB 開発のための会議
2013年8月1日	チーム内 DB 開発会議 (非公開)	DBCLS	9人	DB 開発のための会議
2012年10月3日	チーム内 DB 開発会議 (非公開)	DBCLS	6人	オントロジー開発のための会議
2013年9月21日	チーム内第9回全体会議 (非公開)	東工大田町キャンパス	7人	研究進捗報告のための会議
2013年11月9日	チーム内第10回全体会議 (非公開)	DBCLS	10人	研究進捗報告のための会議
2014年1月6日	チーム内第10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

## §7 ユーザー評価結果への対応

≪平成25年7月に実施した「NBDCにおける事業活動のユーザー評価」(<http://biosciencedbc.jp/user-hyouka-2013/result-summary>)で得られたユーザーの意見、提案等(詳細は別紙2を参照)に対し、実施済み若しくは実施予定の対応策等を具体的に記載してください。)

### ①実施済み

- ・ 使い方や、何ができるのかについて説明がほしい。
- ・ 各情報がどのようなソースから構築されているかの情報が欲しいです。
- ・ ドキュメントやヘルプが多いと助かります。
- ・ ヘルプ機能があると完璧になります。
- ・ 使い方に関して、もう少し説明があった方が良い。

検索結果のページを構成する各 Stanza に対して Help ページを用意し、その Stanza がどのようなデータから構成されているのか、およびその Stanza がどのような情報を示しているのかがユーザーにわかるようにした。Help の例:

[http://mdb.bio.titech.ac.jp/microbedb.jp\\_help\\_page/gene/ortholog-stanza#content](http://mdb.bio.titech.ac.jp/microbedb.jp_help_page/gene/ortholog-stanza#content)

- ・ 活用事例の紹介などしていただけると助かります。

研究開発終了報告書と共に、MicrobeDB.jp の活用事例を紹介した資料を作成した。

### ②実施予定

- ・ 検索結果で blank 部分が非常に多い。先にでてきた TogoStanza というのと重複している部分が多いと思われるが、こちらは UI/デザイン面で劣る。

MicrobeDB.jp は TogoStanza と異なり、微生物の遺伝子・系統・環境に関する多様なデータを統合した DB であるため、ユーザーが検索に用いる語句も非常にバラエティに富んだものとなっている。そこで MicrobeDB.jp では、検索語に対応する検索結果を漏れなく表示できるように、様々な Stanza にて独立に SPARQL 検索を実行している。そのため、検索結果が得られた Stanza からは表示が返ってくるが、検索しても全く結果が得られなかった場合にはその Stanza は空白を表示

していた。現在検索システムの洗練および検索しても全く結果が得られ無かった **Stanza** の表示形式の変更を行うことを予定しており、この問題は今後改善していく予定である。

## § 8 その他

(1) 研究代表者として、研究開発、プロジェクト運営等について、上記以外に報告したいことがあれば、自由に記載してください。

DB統合化を推進するにあたり、**DBCLS**の関係者には強く連携を取って頂き技術面において大きくサポート頂いた。心から感謝申し上げます。また、所有する全データを積極的に提供して頂いた**NBRC** および **JCM** の関係者には心から感謝申し上げます。

以上