

研究開発課題別事後評価結果

1. 研究開発課題名

ゲノム・メタゲノム情報を基盤とした微生物 DB の統合

2. 代表研究者名

東京工業大学地球生命研究所 教授 黒川 顕

3. 研究実施概要

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情報の統合的データベースの構築を目指し、本課題では、セマンティックウェブ技術を徹底的に導入することで、分類学的情報、菌株保存情報、などを統合したデータベース「MicrobeDB.jp」を開発した。さらに、利便性の向上を目的としたデータベース利用のためのテンプレートである Stanza の開発や、アノテーション情報の高度化を可能とする「TogoAnnotation」システム、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。

- 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及び実装

メタゲノムデータベースの構築のため、National Center of Biotechnology

Information(NCBI) のSequence Read Archive(SRA)塩基配列データベースに登録されているデータを解析し、その結果をRDF化した。さらに、メタデータの整備のため、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)の開発を行い、メタゲノムデータを「MEO」、「Gazetteer」、「NCBI Taxonomy」の各オントロジーへマッピングした。また、菌株保存データをゲノムデータとマッピングさせるための新たなオントロジーの構築も行った。RDF化した全データを統合して、微生物統合データベース「MicrobeDB.jp」を構築しスパコンに実装、さらにゲノムマッピング等の解析システムも開発した。

- 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
平成22年度までのバイオインフォマティクス推進事業において開発した、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Insertion sequence (IS) 配列を集約した「GIB-IS」について、データの追加及びデータベースの再構築を行い、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データベースの共用化を促進し、利用価値を高めた。有用な微生物群のゲノムアノテーションの支援のため、「TogoAnnotation」（旧「KazusaAnnotation」）を構築し、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、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進めた。データ蓄積と並行して、情報統合のため、開発期間を通じて集積を続けたデータをより一層、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。

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

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について検討を行い、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した。本研究開発の成果は、すでに公開している **Microbial Genome Database for Comparative Analysis (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 (Stanza) 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た。

4. 事後評価結果

4-1. 当初計画の達成度

本課題は、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及び実装、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、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、という当初の研究計画を達成した。微生物ゲノム情報の統合化を進め、データのRDF化も実施した。NBDCの生命科学系アーカイブサービスへの預け入れと公開も実現しており、統合化は進んだといえる。

4-2. 研究開発成果の公開および利用の状況等

開発したデータベースは統合化データベース「**MicrobeDB.jp**」として公開されている、すべてのデータセットのダウンロードも可能となっている。しかし、統合化データベース「**MicrobeDB.jp**」については、月間のユニークIPアクセス数は600件程度にとどまっており、利用の状況は少ない。

4-3. 研究開発成果によるライフサイエンス分野のデータ活用への波及効果

データベース開発の段階において、オントロジー整備の面で国際的な連携がなされており、この点において波及効果があるとみることができる。なお、「MicrobeDB.jp」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たデータベースであり、幅広い分野での微生物学の発展に寄与可能であると考えられるが、データベースそのものの波及効果については今後の普及によるところもあり、現時点では正しく評価することが難しい。

4-4. 広報・アウトリーチ活動等

論文発表や国内外の学会にて講演・発表を通じて、研究成果を積極的に周知し、学会等でのブース出展を実施しており評価できる。

5. 総合評価

本課題は、環境情報を中心とした各種オントロジーの開発とセマンティックウェブ技術の活用により、微生物情報と環境情報が融合し、多様な情報解析が可能な優れた微生物関連情報が統合されたデータベースを構築、提供している。現時点における利用者は少ないが、今後、本データベースを真菌類等、より上位の微生物にも拡張し、広く微生物データの統合を進めて、さらにメタボロームや植物ゲノム等の他の統合化推進課題との連携も継続して実施することで、今後、データベースの利用価値を高め、様々な研究分野で活用されることを期待する。