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(平成 26 年度採択)
課題名「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」(研究代表者　黒川　顕)

微生物統合データベース「MicrobeDB.jp」の超高度化推進

黒川　顕
（東京工業大学地球生命研究所）

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。

我々は NBDC 第 1 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」を開発した。MicrobeDB.jp は、約 11 億トリプルで構成される Full RDF な DB であり、6 種類のオントロジー＆ボキャブラリおよび 110 種類の統合 DB 解析アプリケーション群（Stanza）を開発・実装した世界に類を見ない 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いる。

現在、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、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、単語の検索や単なる統計量の羅列ではなく、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 DB システムの構築を目指している。MicrobeDB.jp を利用することで、分野特異的のみならず分野横断的な新興研究に資する新たな知見を抽出することが容易となり、ゲノム情報を核とした我国の微生物研究の発展に寄与できると考える。