

平成 27 年度 研究開発実施報告書

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）
研究代表者

黒川 顕

東京工業大学 地球生命研究所
教授

「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」



©2016 黒川 顕(東京工業大学) licensed under CC表示2.1 日本

§1. 研究実施体制

(1)「東工大」グループ

- ① 研究代表者: 黒川 顕 (東京工業大学地球生命研究所、教授)
- ② 研究項目:メタゲノムおよび微生物 DB システムの高度化と統合解析システムの開発

(2)「遺伝研」グループ

- ① 主たる共同研究者:中村 保一 (国立遺伝学研究所 生命情報研究センター、教授)
- ② 研究項目:微生物 DB の高度化と配列解析支援系の開発と提供

(3)「基生研」グループ

- ① 主たる共同研究者:内山 郁夫 (基礎生物学研究所 ゲノム情報研究室、助教)
- ② 研究項目:比較ゲノム解析に立脚した微生物ゲノム情報の統合化

§2. 研究実施内容

(1) 通期及び本年度の研究開発計画と達成目標

メタゲノムおよび微生物 DB システムの高度化と統合解析システムの開発 (東工大 G)

- 1) **メタゲノム DB の整備**
合計数百 TB 以上の配列データを高速に解析可能な解析手法の開発を行う。ユーザビリティの向上を目指し、検索システムの高度化を実現する。
- 2) **解析 Stanza の開発**
更新されたゲノムやメタゲノムデータを用いて比較ゲノム解析や比較メタゲノム解析等の高度な解析を行う Stanza を開発する。
- 3) **真菌類データの整備**
真菌類の RNA-seq データによる発現量を DB 化する。
- 4) **DB 更新自動化システムの開発**
オントロジーマッピングの半自動化システムを進展させ、人間の介入を最小化するシステム構築を目指す。

微生物 DB の高度化と配列解析支援系の開発と提供 (遺伝研 G)

- 1) **各種オミックスデータの整備**
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。
- 2) **藻類データの整備**
PGDBj (<http://pgdbi.jp>) と連携しつつ、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進める。
- 3) **MiGAP および MeGAP との連携強化**
MiGAP (<http://www.migap.org/>) 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する。また、MeGAP と連携することにより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。
- 4) **アクセスレベルの制限システムの構築**
OpenID にグループの概念を付与した認証システムを拡充し、MicrobeDB.jp 全域に適用する開発を実施する。

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 G)

- 1) **大量ゲノムデータへの対応**
大量に蓄積しつつある同種のゲノムデータを効率よく処理するため、種内比較と種間比較とに分けて実行し管理するような体制作りに取り組む。
- 2) **ドメイン単位のオーソログ分類の改良と活用**
ドメイン単位のオーソログ分類について、真核生物の遺伝子などドメイン構成が複雑なケースの分類精度を改善し、比較解析に効果的に活用できるようにする。
- 3) **オーソログデータの RDF 化に基づくデータ統合の推進**
オーソログデータを基盤として各種データを統合するため、オーソログ関連データの RDF 化と検索インターフェースの開発を行うとともに、他データベースとの連携やデータの標準化への対応を進める。

(2) 本年度の研究開発進捗状況

1. メタゲノムおよび微生物 DB システムの高度化と統合解析システムの開発 (東工大 G)

1) メタゲノム DB の整備

公共のメタゲノム塩基配列データの合計サイズが約 10 倍になったため、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う解析手法を開発した。公共の全メタゲノムサンプルの系統組成については計算完了し、RDF データを作成した。MG-RAST 及び IMG/M から取得したメタゲノムのメタデータを整理して RDF 化した。また、H26 年度に開発した半自動マッピングプログラムを用いて、MEO や PDO など開発したオントロジーに追加すべき語句をリスト化し、適宜追加した。これによって、公共の塩基配列 DB だけでなく、それらのデータベースのデータについてもオントロジーを介して統合することが可能になった。MicrobeDB.jp に収録されているデータの統計情報は以下の通りである。

- ・ 総トリプル数: 63 億トリプル
- ・ メタ 16S データ数: 87,620 サンプル
- ・ メタゲノムデータ数: 23,388 サンプル
- ・ ゲノムデータ数: 17,702 株 (76GB、12,553,423 遺伝子)
- ・ 菌株数: 14,961 株 (JCM)、17,367 株 (NBRC)

MicrobeDB.jp の RDF データ中の、特定の述語で記述されている目的語に対して、全文検索エンジンである Apache Solr を用いてキーワード検索を行う機能を実装し、大文字小文字の違いやスペルミス等を許容したキーワード検索を可能にした。また、ユーザが入力したキーワードが遺伝子・系統・環境・その他の内のどれであるかを、上記の Apache Solr による MicrobeDB.jp のデータの検索インデックス及び、環境であれば MEO、系統であれば NCBI Taxonomy などの対応するオントロジーを検索することで推定する機能を MicrobeDB.jp に実装した。それと同時に、各 Stanza が入力として求めるキーワードが、遺伝子・系統・環境の 3 軸のどれであるかを基に Stanza を 3 type に分類することで、ユーザが検索したキーワードの種類によって、表示する Stanza セットを変更する機能を実装した。これらの検索システムの改良によって、ユーザビリティを向上させた。

2) 解析 Stanza の開発

ゲノム間で各遺伝子機能の有無を Pathway レベルで比較解析できる Stanza#1、各メタゲノムサンプルの遺伝子機能組成を KEGG や MBGD 等の様々な機能カテゴリを用いて視覚化できる Stanza#2 等の解析 Stanza を新規に 16 個開発した。これにより、これまで開発した Stanza 合計は約 180 となる。

#1:http://fs2.bio.titech.ac.jp:9292/stanza/taxonomies_function_comparison_table?tax_id=519441,1001582&type=pathway

#2:http://fs2.bio.titech.ac.jp:9292/stanza/sample_function?srs_id=DRS000009&type=mbgd&chart=bar&submitButton=

また、統合化推進プログラムの榎屋グループと共同で、JCM の菌株の RDF データの更新を行い、ゲノム解読済みの株名を NCBI Taxonomy ID と紐付け、ゲノムデータとある程度自動で対応付けるプログラムを開発した。種名までしか NCBI Taxonomy ID が存在せず、かつゲ

ノムデータが存在する株については、手動で菌株 ID とゲノムの BioProject ID 間の関連付けを行った。

3) 真菌類および藻類データの整備

H26 年度に開発した RDF モデルを用いて、DBCLS と共同で、原核・真核問わず RefSeq の微生物ゲノムデータを RDF 化すると共に、MBGD のオーソログの RDF データと対応付けた。

4) DB 更新自動化システムの開発

現在の MicrobeDB.jp で用いられている、公共のメタゲノムデータのサンプリング環境のメタデータと MEO との 10,775 件のマニュアルマッピングデータを教師データとして用いて、入力されたサンプリング環境の自然言語文を、Natural language toolkit によって形態素解析し、教師データ中のそれらの出現頻度を元にして、ランダムフォレスト、線形カーネルサポートベクターマシン、Labeled LDA などの種々の統計的機械学習法を用いて自動的に MEO にマッピングするプログラムを開発した。教師データに用いていない、菌株保存機関の菌株の分離元のメタデータと 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ですが、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。

微生物 DB の高度化と配列解析支援系の開発と提供 (遺伝研 G)

1) 各種オミックスデータの整備

オミックスデータを対象とした整備として、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。本年度は、NCBI Assembly Reports など網羅的な解析に必要なソースの取得から RDF への変換までの仕組みを開発した。また、国際塩基配列データベース SRA に登録された微生物の RNA-Seq 解析で利用された 5,193 件を対象に、リファレンスとなるゲノム情報を NCBI Assembly Report のエントリーから 16S rRNA 遺伝子配列に基づきメタデータ経由で近縁ゲノム解析菌の検索を試行した。

2) 藻類データの整備

ゲノム情報は、これまで RefSeq からのゲノムアノテーションを利用しているが、単細胞性緑藻類ゲノムデータベースをはじめとする真核モデル生物のデータベース上では、主に GFF3 形式でゲノムアノテーション情報が公開されている。そこで、BioInterchange によって変換される RDF データモデルと GFVO の利用について調査を実施した。単細胞藻類のゲノム情報ならびにオミックスデータを MicrobeDB.jp 上に収集・整備のための環境として JBrowse 上へのマッピングを介して PGDBj (<http://pgdbj.jp>) と連携しつつ、単細胞性緑藻類ゲノムデータベースとの連携を開始した。

3) MiGAP および MeGAP との連携強化

前年に引き続き MiGAP ならびに MeGAP への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。具体的には、MicrobeDB.jp ユーザが自身のゲノム・メタゲノム解析データを利用するための登録サービス GenomeRefine (<http://genome.annotation.jp/genomerefine/>) において、MiGAP から解析された配列およびアノテーションファイルを受信するための API を開発した。

また、複数のメタデータおよび MeGAP 解析データを一括登録可能にするバルク登録機能を整備するとともに、DDBJ BioSample Package 毎に必須項目が変動する 334 項目の属性データ項目に対応した入力メタデータの入力を可能にした。また、OWL/RDF を用いた DDBJ BioSample 定義情報との共用化について、DDBJ における BioSample バリデーターの開発を通じて連携を開始した。今年度において、GenomeRefine は、14 グループによって、MiGAP 12、MeGAP 87 の解析データが利用された。

4) アクセスレベルの制限システムの構築

OpenID にグループの概念を付与した認証システムを拡充し、MicrobeDB.jp 全域でのグループ単位でのアクセスレベルコントロールを強化した。今年度は、特に解析系サービスをシームレスに連携させるための認証システムの拡張を実施した。これまで、認証サーバーは、OpenID プロトコルを用いて認証を行っていたが、OpenID Connect 用のインターフェースを追加した。ウェブアプリケーションだけでなく、モバイルアプリケーションやシングルページアプリケーションからも OpenID Connect プロトコルを用いた認証と所属グループ情報を含む、ユーザー情報の取得が可能となった。

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 G)

1) データベースの更新

微生物比較ゲノムデータベース Microbial Genome Database for Comparative Analysis (MBGD; <http://mbgd.genome.ad.jp>)のデータ更新を行った。新たに 1283 の完全ゲノムを追加し、登録された完全ゲノム数は全体で 4,742 となった (前年比 1.37 倍)。この他に、My MBGD 機能によって動的に比較に加えられるドラフトゲノムも 15,767 件追加登録しており、全体で 29,181 件となった (前年比 2.05 倍)。各属から 1 種のゲノムをとって作成した標準オーソログテーブルに含まれるゲノム数は 868 (増加分 81)、ドラフトも含めた標準オーソログテーブルのゲノム数は 1,801 (増加分 217) となった。

2) ホモロジー検索の管理方式の変更

同種のゲノムが大量に蓄積していることを受けて、ホモロジー検索を種内比較と種間比較を分けて行い、効率的に管理するよう修正を行った。具体的には、まず高速ホモロジー検索プログラム UBLAST を用いて種内での総当たりの比較を行い、それを用いて種内オーソログ解析を行って代表配列を抽出した種ごとの pan-genome を作成する。次にそれらを用いて種間の総当たり比較を行う。従来、属ごとに代表生物種をとって作成していた標準オーソログテーブルについても、まず属内のオーソログ解析に基づいて pan-genome を作成した上で、それらの間のオーソログ解析を行って作成する。これら一連のプロトコルを実装した。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せるようになった。次回更新時にこのプロトコルを用いたデータ作成を行うとともに、このデータを利用するための利用者インターフェースの修正を行う予定である。

3) ドメイン単位のオーソログ分類

MBGD におけるドメイン単位のオーソロググループの情報を効果的に活用することを目的として、ドメイン融合遺伝子に基づいて各オーソログドメイン間の隣接関係を定義したオーソログドメインネットワークを構築した (35,080 ノード)。このネットワーク

の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い、次数の高さとの関連などを調べた。例えば、二次代謝関連遺伝子のオーソログドメインは平均次数が 9.8 と高い一方、翻訳関連遺伝子では平均次数が 2.0 と少ない等の結果が得られ、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。

4) オーソロジーオントロジーの統合

Murcia 大の Fernández-Breis らと共同で、オーソログ関連用語のオントロジー OrthO と OGO を統合して発展させた新たなオントロジー Orthology Ontology (ORTH; <https://github.com/qfo/OrthologyOntology>) を開発した(文献 1)。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 812 プロパティとなった。すでに一部で利用されている XML による共通形式 OrthoXML からこのオントロジーを用いた RDF データへの自動変換や、このオントロジーを使って統合的な検索を行う応用について検討した。

5) オーソログデータの RDF 化

完全ゲノムおよびデフォルトオーソロググループに加えて、新たにドラフトゲノムデータと、系統特異的オーソロググループのデータを RDF 化して格納した(合計 2,137,264,913 トリプル)。MBGD RDF を利用したオーソログデータを検索する基本的なクエリについてのテンプレートを用意し、それらを MBGD SPARQL 検索ポータルサイト(<http://mbgd.genome.ad.jp/sparql>) 上で参照できるようにした。また、前項 ORTH オントロジーを使った検索ができるよう修正した。

6) オーソログデータ検索のためのスタンプの作成

オーソログデータを利用した解析結果の可視化を行うスタンプを作成し、スタンプライブラリとして蓄積する体制を作った。オーソロググループに関連づけられるスタンプ(5 種)、遺伝子に関連づけられるスタンプ(1 種)、系統群に関連づけられるスタンプ(3 種)を作成した。これらは今後 MicrobeDB.jp および MBGD から利用できるようにする予定である。

§3. 成果発表等

(3-1) 原著論文発表

- ① 発行済論文数 (国内 (和文) 0 件、国際 (欧文) 10 件)
 - ② 未発行論文数 (国内 (和文) 0 件、国際 (欧文) 3 件)
 - ③ 論文詳細情報
1. Yamaichi Y, Chao C. M, Sasabe J, Clark L, Davis M. B, Yamamoto N, Mori H, Kurokawa K, Waldor M K (2015) 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. (DOI: 10.1093/nar/gku1262)
 2. Uchiyama T, Irie M, Mori H, Kurokawa K, Yamada T (2015) FuncTree: Functional analysis and visualization for large-scale omics data. *PLoS One*, 10:e0126967. (DOI: 10.1371/journal.pone.0126967)
 3. Kato H, Ogawa N, Ohtsubo Y, Oshima K, Toyoda A, Mori H, Nagata Y, Kurokawa K, Hattori M, Fujiyama A, Tsuda M (2015) Complete Genome Sequence of a

- Phenanthrene Degradar, *Mycobacterium* sp. Strain EPa45 (NBRC 110737), Isolated from a Phenanthrene-Degrading Consortium. *Genome Announc.* 3:e00782-15. (DOI: 10.1128/genomeA.00782-15)
4. 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 (2015) Time-series metagenomic analysis reveals robustness of soil microbiome against chemical disturbance. *DNA Res.* 22:413–424. (DOI: 10.1093/dnares/dsv023).
 5. Hirose Y, Fujisawa T, Ohtsubo Y, Katayama M, Misawa N, Wakazuki S, Shimura Y, Nakamura Y, Kawachi M, Yoshikawa H, Eki T, Kanasaki Y. (2015) Complete genome sequence of cyanobacterium *Nostoc* sp. NIES-3756, a potentially useful strain for phytochrome-based bioengineering. *J Biotechnol.*, 218: 51-52. (DOI: 10.1016/j.jbiotec.2015.12.002)
 6. Mashima J, Kodama Y, Kosuge T, Fujisawa T, Katayama T, Nagasaki H, Okuda Y, Kaminuma E, Ogasawara O, Okubo K, Nakamura Y, Takagi T. (2015) DNA data bank of Japan (DDBJ) progress report. *Nucleic Acids Res.*, pii: gkv1105.
 7. 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. (2015) Loss of Cytochrome cM Stimulates Cyanobacterial Heterotrophic Growth in the Dark. *Plant Cell Physiol.*, **56**, 334-345. (DOI: 10.1093/pcp/pcu165)
 8. Kodama Y, Mashima J, Kosuge T, Katayama T, Fujisawa T, Kaminuma E, Ogasawara O, Okubo K, Takagi T, Nakamura Y. (2015) The DDBJ Japanese Genotype-phenotype Archive for genetic and phenotypic human data. *Nucleic Acids Res.*, **43**, D18-22. (DOI: 10.1093/nar/gku1120)
 9. Kanasaki Y, Masutani H, Sakanaka M, Shiwa Y, Fujisawa T, Nakamura Y, Yokota A, Fukiya S, Suzuki T, Yoshikawa H. (2015) Complete Genome Sequence of *Bifidobacterium longum* 105-A, a Strain with High Transformation Efficiency. *Genome Announc.*, **2**, pii: e01311-14. (DOI: 10.1128/genomeA.01311-14)
 10. Fernández-Breis, J.T., Legaz-García, M.C., Chiba, H., and Uchiyama, I. (2015) Towards the semantic standardization of orthology content. *Proc. Semantic Web Applications and Tools for Life Sciences*, 74-83.
 11. 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.* (in press).
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.*, (in press). *本論文は昨年度投稿し H28 年度出版予定である。

12. 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, Shinkura R, High-affinity monoclonal IgA regulates gut microbiota and prevents colitis in mice. *Nat. Microbiol.*, (in press).

(3-2) データベースおよびウェブツール等の構築と公開

別紙1参照。

(3-3) その他の著作物(総説、書籍など)

1. 森宙史, 東光一, 山田拓司, 黒川頭, 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ー, シーエムシー出版, pp10-19 (2015)
2. 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)
3. 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)

(3-4) 国際学会および国内学会発表

① 招待講演 (国内 3件、国際 1件)

〈国内〉

1. 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
2. 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
3. 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

〈国際〉

1. Yasukazu Nakamura, Towards better genome annotation, The 13th JCK Bioinformatics Symposium, OIST Sea Side House, Okinawa, Japan, October 11, 2015.

② 口頭講演 (国内 4件、国際 2件)

〈国内〉

1. 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
2. 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, 黒川頭, 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

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. 藤沢貴智、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

〈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.

③ ポスター発表 (国内 10 件、国際 3 件)

〈国内〉

1. 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
2. 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
3. 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
4. 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
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
6. 藤澤貴智、森宙史、神沼英里、大山彰、菅原秀明、内山郁夫、黒川顕、中村保一 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、東京大学弥生講堂、2015年10月6日
7. 藤澤 貴智、広瀬 侑、兼崎 友、志村 遥平、谷澤 靖洋、山本 泰智、李 慶範、河地 正伸、中村 保一、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、東京工業大学大岡山キャンパス、2016年3月4日
8. 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6日

9. 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、トーゴーの日シンポジウム 2015、東京大学弥生講堂、2015年10月6日
10. 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

〈国際〉

1. Takatomo Fujisawa, Toshiaki Katayama, Mitsuteru Nakao, Shinobu Okamoto, Yasukazu Nakamura, Semantic data integration in CyanoBase and RhizoBase, Biocuration 2015, Beijing, China, April 25, 2015
2. Hirokazu Chiba, Hiroyo Nishide, Jesualdo Tomás Fernández-Breis, Ikuo Uchiyama, Applying RDF/SPARQL to ortholog databases, Quest for Orthologs 4, Barcelona, May 25, 2015.
3. Hirokazu Chiba, Ikuo Uchiyama, SPANG: A command-line client supporting query generation for distributed SPARQL endpoints, Semantic Web Applications and Tools for Life Sciences 2015, Cambridge, December 9, 2015.

(3-5) 知的財産権の出願

該当無し

(3-6) 受賞・報道等

該当無し

§4. 研究開発期間中に主催した活動(主催したワークショップ等)

年月日	名称	場所	参加人数	概要
2015年4月6日	チーム内第1回全体会議(非公開)	東工大大岡山キャンパス	8人	研究進捗報告のための会議
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	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

以上

別紙1 既公開のデータベース・ウェブツール等

No.	状態	研究開発課題名	データベース、ウェブツール等の名称	概要 (150字程度)	URL	公開開始日	関連論文 (論文リストに記載があれば、その番号でも可)
1	継続	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	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/		
2	継続	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	CyanoBase (MicrobeBase)	シアノバクテリアに代表される酸素発生型光合成細菌とその関連バクテリアのゲノム情報を集積したデータベースです。	http://genome.microbedb.jp/cyanobase	H7年12月	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, "CyanoBase and RhizoBase: databases of
3	継続	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	GTPS	本DBでは微生物ゲノムの最新情報を基にした再アノテーション情報を提供しています。	http://gtps.ddbj.nig.ac.jp/gtps_data/	H21年3月	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
4	継続	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	GTPS/RDF	本DBはセマンティック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。	http://gtps.ddbj.nig.ac.jp/rdf/	H24年4月26日	
5	継続	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	Metagenome.jp (Human Meta BodyMap)	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。	http://www.metagenomics.jp/mg/		
6	継続	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	MBGD	オーソログ解析に基づいて微生物ゲノムの比較解析を行うためのデータベースです。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/		
7	継続	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	MEO	本DBは、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://mdb.bio.titech.ac.jp/meo		
8	継続	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	RhizoBase (MicrobeBase)	窒素固定植物共生細菌である根粒菌(共生関連領域のみの部分配列を含む)ゲノム情報を集積したデータベースです。	http://genome.microbedb.jp/cyanobase	H13年1月	Takatomo Fujisawa, Shinobu Okamoto, Toshiaki Katayama, Mitsuteru Nakao, Hidehisa Yoshimura, Hiromi Kajiya-Kanegae, Sumiko Yamamoto, Chiyoko Yano, Yuka Yanaka, Hiroko Maita, Takakazu Kaneko,
9	継続	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	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/	H19年6月1日	

10	継続	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	PDO-CSSO	本DBは、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://mdb.bio.titech.ac.jp/pdo http://mdb.bio.titech.ac.jp/csso		
11	継続	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	MCCV	本DBは、JCMやNBRC等の、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であるMCCV (Microbial Culture Collection Vocabulary)のOWLファイルを提供するDBです。DBCLSと共同で開発しております。	http://bioportal.bioontology.org/ontologies/MCCV		
12	継続	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	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/	H26年7月	
13	継続	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	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。SPARQLによる検索例や、RDFアーカイブ、Ortholog Ontology (OrthO)のドキュメントなども含んでいます。	http://mbgd.genome.ad.jp/sparql		