

平成 23 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「基盤技術開発プログラム」
平成 23 年度採択 研究代表者

米澤明憲

情報・システム研究機構・ライフサイエンス統合データベースセンター・センター長

データベース統合に関わる基盤技術開発

§1. 研究実施体制

(1)「DBCLS」グループ(研究機関別)

- ① 研究代表者: 米澤 明憲 (情報・システム研究機構・ライフサイエンス統合データベースセンター・センター長)
- ② 研究項目
 - ・ データベースの RDF による統合化
 - ・ 解析プラットフォームによる統合利用環境の整備
 - ・ インターネットを活用した高度検索技術の開発
 - ・ RDF 化に資するオントロジー、辞書、コーパスの整備、標準化
 - ・ 大規模ゲノム配列データの利用技術開発
 - ・ 情報統合化・知識発見のためのキュレーション支援
 - ・ 統合データベースに関わるコンテンツの作成、整備

(2)「CBRC」グループ(研究機関別)

- ① 主たる共同研究者: 浅井 潔 (産業技術総合研究所 生命情報工学研究センター、センター長)
- ② 研究項目
 - ・ 解析プラットフォームによる統合利用環境の整備

(3)「京都大学」グループ(研究機関別)

- ① 主たる共同研究者:五斗 進 (京都大学化学研究所、准教授)
- ② 研究項目
 - ・ DBGET/LinkDB システムの統合利用環境への応用
 - ・ メタゲノム・メタメタボローム等新規分野データ活用技術の開発
 - ・ 反応オントロジーの整備

§ 2. 研究実施内容

(1)目的

JST バイオサイエンスデータベースセンターの推進する統合データベースを実現するために必要な、基盤技術の確立を目指し、DBCLS、CBRC、京都大学の 3 機関が共同し、RDF を中心とするセマンティックウェブ、オントロジーや自然言語処理、解析ワークフロー、大規模データ活用等の技術をもとに、フェデレーション(分散)型の統合を実現するための研究開発を進める。

(2)進捗状況

計画初年度にあたる平成 23 年度は、データベースの RDF による統合を進めるための各種調査(有用 DB の優先度付調査、RDF 蓄積 DB 調査、オントロジー調査等)を実施、特定分野の DB の RDF 化やオントロジーの作成、解析ワークフローの RDF 化など当初計画に従って順調に研究開発を進めた。また、微生物ゲノムやプロテオーム分野を対象に統合検索のプロトタイプを開発し、情報提示法の検討を行った。8 月には国際会議 BioHackathon を開催、DDBJ、PDBj、LinkDB、UniProt や Bio2RDF、BioDBCCore など国内外の DB と共同で開発を行うことによって、国際連携、技術の標準化にも貢献した。大規模データの利用技術開発として急増する次世代シーケンスデータを対象にメタデータ検索技術の開発や評価、解析ツールの整備を行った。RDF を蓄積するトリプルストアの調査については、期間を延長して実施する。

(3)研究機関別の研究成果:

1) DBCLS グループ

- ・ DB の RDF による統合化を推進するため、NBDC とともに国際会議 BioHackathon2011 と統合 DB 技術情報交換ワークショップを主催、国内外の DB 開発者や研究者と連携し RDF 化に関するガイドラインを整備、会議で決定並びに開発された内容について複数の論文投稿を予定している。NBDC の主要サービスである DB アーカイブと連携し、寄託された DB のうち 6 件の RDF 化を行うとともに、アップロードしたデータを半自動的に RDF へ変換する機能をもつ TogoDB の後継となる、Semantic TogoDB を開発し、初期バージョンを公開した。
- ・ 解析プラットフォームによる統合利用環境の整備のため、DBCLS で開発を続けてきたゲノム解析プラットフォーム DBCLS Galaxy に対し、テキストマイニングツール群をレポジトリとして整備し、それらのツール群を Galaxy で利用可能とした。また、表形式の統合検索プロトタイプ TogoTable との連携機能を追加した。
- ・ 高度検索技術(統合検索技術)の開発を目指し、利用者のニーズ調査、有用分野でのプロトタイプピング、可視化技術の検討を実施した。具体的には NBDC や統合化推進プログラム、国内研究者の協力のもと、プロテオーム(TogoTable)、微生物ゲノム(RDF Genome)の統合検索プロトタイプを構築し試験公開した。既開発サービスの高機能化(DBカタログ、横断検索、DBアーカイブ)ではデータの RDF 化や配列検索機能の追加を実施した。
- ・ RDF 化に資するオントロジー、辞書、コーパスの整備、標準化を進めるために、NBDC と連携し、オントロジー構築のガイドラインを整備し、BioPortal を利用したオントロジー構築サポートツール OntoFinder/OntoFactory を開発し試験公開した。またコーパスアノテーションのためのツール PubCorpus の開発に着手した。さらに、統合化推進プログラム「ゲノム・メタゲノム情報を基盤とした微生物 DB の統合」グループと共同でメタゲノム環境オントロジー(MEO)を構築した。

- ・ 大規模ゲノム配列データの利用技術開発として、次世代シーケンスデータの公共リポジトリのダイジェスト版 SRAs (Survey of Read Archives)を開発するとともに、信頼性の高いデータの検索を可能にするウェブインターフェース Kusarinoko の開発に着手した。さらに、大規模配列 DB を高速に検索する統合遺伝子検索 GGRNA、遺伝子発現のリファレンスデータセットを表示提供する RefEx (Reference Expression dataset)の開発を実施、計 4 件のサービスを新規公開した。
- ・ 情報統合化・知識発見のためのキュレーション支援として、自然言語処理技術、テキストマイニング技術、コンピュータ支援による協働作業手法 (CSCW)を用い、多面的で実用的なキュレーション支援システムと協働作業運用技術の整備を実施、関係課題への環境構築支援を行うとともに、サービスとして Allie、inMeXes、TogoDoc、OReFil の運用並びに機能追加を行った。
- ・ 統合データベースに関わるコンテンツの作成、整備として、文科省統合 DB プロジェクトより継続してきたサービス、統合 TV、ライフサイエンス新着論文レビュー、BodyParts3D の開発運用を続けるとともに、新たに実験や生物画像のコンテンツ Togo Picture Gallery を公開した。

2) CBRC グループ

本グループでは、独自に開発したツールや有用な既存ツールを KNIME のプラットフォーム上で動作するように新規にノード化し、主として RNA 解析に関するワークフローを公開した。また DB の RDF 化と連携し、開発した高度な解析ツール群を広く利用可能とするために、セマンティック技術に対応した SADI (Semantic Automated Discovery and Integration) フレームワークを利用して、各ツールに RDF 入出力機能を追加し解析可能としたプロトタイプを公開している (<http://togo.cbrc.jp/semantic.html>)。加えて利用者に、よりワークフローを使い易くするためインストーラ及びユーザ・マニュアルを新規に作成し HP より公開した。

3) 京都大学グループ

本グループではデータベース検索のセマンティック化に必要とされる要素技術として LinkDB の開発を行うとともに、様々な解析ツールを整備し、メタゲノムなど新規大規模データへと応用することを目指している。H23 年度は DBCLS グループと協力し、LinkDB の RDF 化を検討、LinkDB の全データ (約 8 億 5000 万リンク) を RDF 化、ウェブ上の検索結果からダウンロードできるようにした。また、日次更新においても RDF 化に対応した。解析ツールの整備として、KAAS (KEGG Automatic Annotation Server) についてメタゲノムアノテーションで必要となる生物種情報の考慮を追加、アノテーション精度を向上させた。また、遺伝子ネットワーク予測ツール GENIES を整備し、サンプルデータを増やすなどして使い勝手を向上させた。これまで、遺伝子や化合物に比べて整備が遅れていた反応オントロジーについては、反応パターンに基づく階層分類を手作業で行い第 1 版を完成させた。また、その応用ツールである、類似パスウェイ検索ツール PathSearch を開発、半自動で分類した反応を用いたバージョンを H23 年 7 月に公開した。

(4) 今後の見通し

DB の RDF 化やオントロジー構築について、ガイドラインや方向性を確立した一方で、RDF を蓄積するトリプルストアの整備、DB の RDF 化やオントロジー構築にかかるコスト等、RDF を基盤とした開発を進めるための問題点も明らかになった。これらの課題は、昨今のウェブ技術の領域では、世界的に共通な問題を含んでおり、急速な技術的知見の蓄積や資産の発展も見込める。今年度実施した各種調査、プロトタイプ開発に引き続き、H24 年度には開発技術の有効性を示すフィジビ

リテイスターディを意図したアウトプットを示せるよう、研究代表者のリーダーシップのもと研究計画を進める予定である。

§3. 成果発表等

(3-1) データベースおよびウェブツールの構築と公開

① 公開中のデータベース・ウェブツール等

サービス名: 統合 TV

概要: 生命科学分野の有用なデータベースやウェブツールの活用法を動画で紹介するウェブサイト

公開日: H19 年 7 月 19 日

URL: <http://togotv.dbcls.jp/>

アクセス数: H24 年 3 月の一意な訪問者数 5698 / 訪問回数 19591 / ページビュー: 67438 / 転送量 669.68 Gb

サービス名: OReFiL

概要: オンライン上に存在する多数の生命科学系の資源(データベースやソフトウェアなど)を効率的に見つけるための検索システム

公開日: H19 年 8 月

URL: <http://orefil.dbcls.jp/>

アクセス数: H24 年 3 月の一意な訪問者数 504 / 訪問回数 870 / ページビュー 2570

サービス名: BodyParts3D

概要: BodyParts3D(ボディパーツ 3D)は人体各部位の位置や形状を 3 次元モデルで記述したデータベース。Anatomography(アナトモグラフィ)を使って、BodyParts3D から解剖学用語を選択して自由に人体のモデル図を作成できる

公開日: H19 年 10 月 5 日

URL: <http://lifesciencedb.jp/bp3d/>

アクセス数: H24 年 3 月の一意な訪問者数 2,679 / 訪問回数 3,730 / ページビュー 121,912

サイト名: BioHackathon

概要: 最先端の研究開発者を招聘した国際的なソフトウェア開発会議の情報交換ならびに成果を公開するためのサイト。BioHackathon は H19 年度からウェブサービス標準化、統合利用環境構築、セマンティックウェブ、リンクトデータで 4 度開催。DB 統合利用に関する内外研究者の国際的ハブとしての地位も確立しつつある

公開日: H20 年 1 月

URL: <http://www.biohackathon.org/>

サイト名:MotDB

概要:講習会の開催情報、参加受付ならびに講習会で使用する資料等を配布するためのサイト。

平成 23 年度は 6 回の講習会を開催

公開日:H20 年 2 月

URL:<http://motdb.dbcls.jp/>

アクセス数:H24 年 3 月の一意な訪問者数 1167 / 訪問回数 1989 / ページビュー9186

サービス名:TogoDB

概要:エクセルなどのデータを簡単に読み込み、DB 化し、自動的に共通のウェブ検索インタフェースを生成するシステム

公開日:H20 年 4 月

URL:<http://togodb.dbcls.jp/>

アクセス数:H24 年 3 月の一意な訪問者数 2788 / 訪問回数 4914 / ページビュー33962

サービス名:TogoWS

概要:国内各ウェブサービスの稼働状況を継続的に監視し、サービスの連携に必要なデータ形式変換機能等を提供するウェブサービスを統合するシステム

公開日:H20 年 4 月

URL:<http://togows.dbcls.jp/>

アクセス数:H24 年 3 月の一意な訪問者数 446 / 訪問回数 1015 / ページビュー50299

サービス名:Allie

概要:MEDLINE を対象とし、出現する略字とその正規系のペアを検索するシステム。略字を入力することで、その使われ方を一覧表示する

公開日:H20 年 4 月

URL:<http://allie.dbcls.jp/>

アクセス数:H24 年 3 月の一意な訪問者数 7131 / 訪問回数 12522 / ページビュー27507

サービス名:DBCLS OpenID

概要:一つの ID で複数のサイトを認証できるシステム。各サイトで認証サービスを用意する必要がなく、サイト間のユーザー情報の集約が容易になる

公開日:H20 年 4 月

URL:<http://openid.dbcls.jp/>

アクセス数:H24 年 3 月の一意な訪問者数 400 / 訪問回数 757 / ページビュー5161

サービス名: Gendoo (Gene, Disease Features Ontology-based Overview System)

概要: 文献情報をもとに、遺伝子、疾患について、関連する疾患、薬剤、臓器、生命現象などの特徴をキーワードでリスト表示するツール

公開日: H20 年 12 月 12 日

URL: <http://gendoo.dbcls.jp/>

アクセス数: H24 年 3 月の一意な訪問者数 881 / 訪問回数 2112 / ページビュー 5883

サービス名: inMeXes

概要: MEDLINE を対象とし、利用者が入力した文字列とマッチする表現を逐次的(1文字入力毎)に検索するシステム。論文中の英語表現を容易に検索できる

公開日: H21 年 7 月

URL: <http://docman.dbcls.jp/im/>

アクセス数: H24 年 3 月の一意な訪問者数 875 / 訪問回数 1964 / ページビュー 28719 (TogoDoc との合算)

サービス名: TogoWiki

概要: 国内版バイオハッカソンの情報交換ならびに成果を公開するためのサービス

公開日: H21 年 8 月

URL: <http://wiki.lifesciencedb.jp/mw/>

アクセス数: H24 年 3 月の一意な訪問者数 485 / 訪問回数 1038 / ページビュー 5129

サービス名: DBCLS galaxy

概要: 生命科学データに特化したウェブベースの対話的ツール組み合わせインタフェース DBCLS で開発されたツール群も組み込んだ。

公開日: H21 年 10 月

URL: <http://galaxy.dbcls.jp/>

アクセス数: H24 年 3 月の一意な訪問者数 170 / 訪問回数 295 / ページビュー 1177

サービス名: TogoDoc

概要: 文献情報及び論文 PDF を管理し、また、特定の文献情報群に関連する論文情報を提示するシステム。TogoDoc Client と連携して文献を管理することも可能なほか、スマートフォンにも対応。

公開日: H21 年 12 月

URL: https://docman.dbcls.jp/pubmed_recom/

アクセス数: H24 年 3 月の一意な訪問者数 875 / 訪問回数 1964 / ページビュー 28719 (InMeXes との合算)。TogoDoc Client からのアクセスは 49672 ヒット(H24 年 4 月)、サーバへの登録者は 826 人(H24 年 4 月現在)

サービス名:LifescienceQA

概要:ライフサイエンス分野の研究者の情報交換を促すための Q&A サービス。現在までに 116 の質問が投稿され、224 の回答が得られている。

公開日:H22 年 11 月

URL:<http://qa.lifesciencedb.jp/>

アクセス数:H24 年 3 月の一意な訪問者数 1263 / 訪問回数 4350 / ページビュー13080

サービス名:ライフサイエンス 新着論文レビュー

概要: Nature、Science、Cell などのトップジャーナルに掲載された日本人を著者とする生命科学分野の論文について、論文の著者自身の執筆による日本語によるレビューを、だれでも自由に閲覧・利用できるよう、いち早く公開するオンラインジャーナルサービス。本年度末までの 19 か月間に 300 本のレビューを公開した

公開日:H22 年 9 月 1 日

URL:<http://first.lifesciencedb.jp/>

アクセス数:H24 年 3 月の一意な訪問者数 12,453 / 訪問回数 43,644 / ページビュー254,600

サービス名:SRAs (Survey of Read Archives)

概要:公共データベース(SRA [NCBI]、ENA [EBI]、DRA[DDBJ])に登録された「次世代シーケンサ」データについて、目的別、機器別、生物種別等、さまざまな統計情報から閲覧、比較、データのダウンロードができる目次サイト。論文からのデータの検索も可能である

公開日:H23 年 1 月 5 日

URL:<http://sra.dbcls.jp/>

アクセス数:H24 年 3 月の一意な訪問者数 326 / 訪問回数 429 / ページビュー 1635

サービス名:togo picture gallery

概要:ライフサイエンス分野のイラストをだれでも自由に閲覧・利用できるよう Web 上にて無料で公開しているウェブサイト。

公開日:H23 年 4 月 1 日

URL:<http://g86.dbcls.jp/togopic/>

アクセス数: H24 年 3 月の一意な訪問者数 309 / 訪問回数 586 / ページビュー 4,648、公開から H24 年 4 月 11 日までの総アクセス数 53075

サービス名:統合遺伝子検索 GGRNA

概要: 遺伝子や転写産物をさまざまなキーワードから素早く検索し、その結果をわかりやすく提示することができる遺伝子検索エンジン。遺伝子名や各種 ID、タンパクの機能や特徴などのキーワードだけでなく、短い塩基配列やアミノ酸配列から遺伝子を高速に検索することも可能である

公開日:H23年5月18日

URL: <http://GGRNA.dbcls.jp/>

アクセス数: H24年3月の一意な訪問者数 873 / 訪問回数 1666 / ページビュー 9901、公開から H24年4月11日までの総アクセス数 32000 ページ / 44000 ヒット

サービス名:Kusarinoko

概要: 公共の次世代シーケンスデータレポジトリである SRA に登録されたデータにおいて、登録者により記述されたメタデータを整理し、登録されたデータを元に発表された論文の情報や、予め計算したシーケンスのクオリティ情報を併せて提示することで、より信頼性の高いデータの検索・利用を補助するウェブインターフェース。

公開日:H23年6月9日

URL: <http://g86.dbcls.jp/kusarinoko/>

アクセス数:H24年3月の一意な訪問者数 93 / 訪問回数 120 / ページビュー567、公開から H24年4月までの総アクセス数 2038

サービス名:RefEx (Reference Expression dataset)

概要: EST、GeneChip、CAGE、RNA-seq の 4 種類の異なる手法 によって得られたヒトおよびマウス、ラットにおける遺伝子発現データを並列に表示し、遺伝子発現解析を行う上で基準となるリファレンス(参照)データベースとして利用することを目的とした遺伝子発現データベース

公開日:H23年10月1日

URL: <http://refex.dbcls.jp/>

アクセス数:H24年3月の一意な訪問者数 304 / 訪問回数 586 / ページビュー 2127

サービス名:Allie RDF Data

概要:Allie の SPARQL エンドポイント

公開日:H23年

URL: <http://data.allie.dbcls.jp/>

アクセス数:H24年4月の一意な訪問者数 197 / 訪問回数 261 / ページビュー737

サービス名:Semantic TogoDB

概要:TogoDB の機能に加え、アップロードしたデータを半自動的に RDF へ変換する機能をもつシステム

公開日:H23年12月

URL: <http://semantic.togodb.dbcls.jp/>

サービス名: 統合 DB 情報基盤サイト

概要: CBRC が独自に開発したソフトウェアや解析ツール及び有用な既存ツールのワークフローを提供するサイト。利用者は、KNIME のプラットフォーム上で解析ツールを結合し、組み合わせることで容易にワークフローを構築でき、データの読み込み、計算、解析、可視化を行うことが可能である

公開日: H22 年 7 月 12 日 (ログ解析開始日)

URL: <http://togo.cbrc.jp/>

アクセス数: 公開日から H24 年 3 月 31 日現在 訪問者数 3435 回

データベース名: LinkDB

概要: 世界各国で開発されている分子生物学関連のデータベースをエントリー間の関係で結んだデータベース。データベースエントリー間の関係とその属性(original, reverse, equivalent)の 3 項関係で表現している。KEGG を中心に164 のデータベースを約 8 億 5000 万のリンクで結んでいる

公開日: H6 年 11 月 26 日

URL: <http://www.genome.jp/linkdb/>

アクセス数:

年月	一意な訪問者数	ページ数
2011 年 4 月	7,651	101,558
2011 年 5 月	9,293	138,052
2011 年 6 月	8,944	260,031
2011 年 7 月	8,228	157,856
2011 年 8 月	8,178	190,290
2011 年 9 月	9,790	193,192
2011 年 10 月	10,754	94,630
2011 年 11 月	11,546	620,762
2011 年 12 月	8,706	188,662
2012 年 1 月	9,625	135,678
2012 年 2 月	10,653	73,715
2012 年 3 月	12,257	108,094

ウェブツール名: ゲノムネット ケミカル情報解析ツール

概要: 代謝化合物や反応を中心としたケミカル情報を解析するためのツール群。類似化学構造検索ツール SIMCOMP、部分構造検索ツール SUBCOMP、反応経路計算・予測ツール PathComp、PathPred、PathSearch、EC 番号割り当てツール E-zyme からなる

公開日: H16 年 10 月 1 日

URL: http://www.genome.jp/ja/gn_tools_ja.html#chemical

アクセス数:

年月	一意な訪問者数	ページ数
2011年4月	1,904	112,398
2011年5月	2,090	69,869
2011年6月	2,144	45,067
2011年7月	1,824	33,067
2011年8月	1,843	40,131
2011年9月	2,098	36,230
2011年10月	2,186	34,278
2011年11月	2,307	40,634
2011年12月	1,841	20,004
2012年1月	1,873	23,555
2012年2月	1,886	40,112
2012年3月	1,999	48,453

ウェブツール名: ゲノムネット ゲノム情報解析ツール

概要: ゲノムから遺伝子の機能アノテーションを中心としたゲノム情報を解析するためのツール群。遺伝子の自動アノテーションとパスウェイマッピングのための KAAS、EST からコンセンサスコンテイングを作成する EGassembler、オミックスデータ統合による遺伝子ネットワーク予測ツール GENIES からなる。

公開日: H17年7月1日

URL: http://www.genome.jp/ja/gn_tools_ja.html#genome

アクセス数:

年月	一意な訪問者数	ページ数
2011年4月	1,581	30,916
2011年5月	1,570	51,831
2011年6月	1,518	80,515
2011年7月	1,336	71,808
2011年8月	1,536	69,007
2011年9月	1,528	82,314
2011年10月	1,856	81,501
2011年11月	2,002	91,585
2011年12月	1,912	48,435
2012年1月	1,659	27,852
2012年2月	1,900	27,756
2012年3月	2,091	30,745

(3-2) 原著論文発表

- ① 発行済論文数(国内(和文) 0件、国際(欧文) 7件):
- ② 未発行論文数(“accepted”、“in press”等)(国内(和文) 0件、国際(欧文) 2件)

③ 論文詳細情報

1. Hiromasa Ono, Yoshinao Oki, Hidemasa Bono and Koichiro Kano, Gene expression profiling in multipotent DFAT cells derived from mature adipocytes., *Biochem Biophys Res Commun.*, Apr 15; 407(3), pp.562-567, 2011 (PMID:21419102) (DOI: 10.1016/j.bbrc.2011.03.063)
2. Yamamoto Y, Yamaguchi A, Bono H, Takagi T., Allie: a database and a search service of abbreviations and long forms., *Database*, Apr 15;2011:bar013, 2011 (PMID:21498548)(DOI: 10.1093/database/bar013)
3. Jin-Dong Kim, Yue Wang, Toshihisa Takagi, Akinori Yonezawa, Overview of Genia Event Task in BioNLP Shared Task 2011, *Proceedings of BioNLP Shared Task 2011 Workshop*, pp.7-15, 2011 (ISBN: 9781937284091)
4. Katayama T, Wilkinson MD, Vos R, Kawashima T, Kawashima S, Nakao M, Yamamoto Y, Chun HW, Yamaguchi A, Kawano S, Aerts J, Aoki-Kinoshita KF, Arakawa K, Aranda B, Bonnal RJ, Fernández JM, Fujisawa T, Gordon PM, Goto N, Haider S, Harris T, Hatakeyama T, Ho I, Itoh M, Kasprzyk A, Kido N, Kim YJ, Kinjo AR, Konishi F, Kovarskaya Y, von Kuster G, Labarga A, Limviphuvadh V, McCarthy L, Nakamura Y, Nam Y, Nishida K, Nishimura K, Nishizawa T, Ogishima S, Oinn T, Okamoto S, Okuda S, Ono K, Oshita K, Park KJ, Putnam N, Senger M, Severin J, Shigemoto Y, Sugawara H, Taylor J, Trelles O, Yamasaki C, Yamashita R, Satoh N, Takagi T., The 2nd DBCLS BioHackathon: interoperable bioinformatics Web services for integrated applications. *J. Biomed. Semantics*, Aug 2;2:42011 (PMID: 21806842)(DOI: 10.1186/2041-1480-2-4)
5. Shin Kawano, Hiromasa Ono, Toshihisa Takagi, and Hidemasa Bono, Tutorial videos of bioinformatics resources: online distribution trial in Japan named TogoTV., *Brief Bioinform.*, 13 (2), pp.258-268, 2012 (PMID:21803786) (DOI: 10.1093/bib/bbr039)
6. 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 SI, 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 SI, 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 and Environments*, Advance Publication, 2012 (DOI:10.1264/jsme2.ME11321)
7. Masaaki Kotera, Toshiaki Tokimatsu, Minoru Kanehisa and Susumu Goto, “MUCHA: multiple chemical alignment algorithm to obtain building block substructures of orphan metabolites”, *BMC Bioinformatics*, vol. 12, S1, 2011. (DOI: 10.1186/1471-2105-12-S14-S1)

8. Atsuko Yamaguchi, Yasunori Yamamoto, Jin-Dong Kim, Toshihisa Takagi and Akinori Yonezawa, Discriminative Application of String Similarity Methods to Chemical and Non-chemical Names for Biomedical Abbreviation Clustering.,BMC Genomics, 2012 (to appear)
9. Jin-Dong Kim, Ngan Nguyen, Yue Wang, Jun'ichi Tsujii, Toshihisa Takagi and Akinori Yonezawa, The Genia Event and Protein Coreference tasks of the BioNLP Shared Task 2011,BMC Bioinformatics, 2012 (in press)