

平成 26 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

[田畑 哲之]

[公益財団法人かずさ DNA 研究所・所長、副理事長]

[植物ゲノム情報活用のための統合研究基盤の構築]

§1. 研究実施体制

(1)「かずさ」グループ

① 研究代表者: 田畑 哲之 ((公財)かずさDNA研究所、所長)

② 研究項目

- ・植物リソースDBの拡充による植物ゲノムDBの統合
- ・DNAマーカー情報の拡充による植物ゲノムDBの統合
- ・DB間の連携による統合化
- ・オントロジーの整備による横断検索の効率化
- ・ゲノム情報のアノテーションの高度化

(2)「阪大」グループ

① 主たる共同研究者: 中谷 明弘 (大阪大学・大学院医学系研究科、特任教授)

② 研究項目

- ・遺伝子オルソログDBの拡充による植物ゲノムDBの統合

§ 2. 研究実施内容

(1) 遺伝子オルソログ DB の拡充による植物ゲノム DB の統合

各植物種の全遺伝子のアミノ酸配列間の網羅的 BLAST 検索に基づいたクラスタリングを行い、オルソログ DB の更新とエントリー追加を行った。最新版では RefSeq Database Release 66 に基づいている(緑色植物 40 種/ラン藻 213 種)。その際には、各アミノ酸配列のゲノム配列中での位置情報(染色体名と塩基位置)を RefSeq Database 内のデータに基づいて整理した。これまでも、配列の類似情報の他に、アミノ酸配列の ID や注釈情報中のキーワード等によって検索や DB 統合を行ってきたが、それらに加えて位置情報によってオルソログ DB の検索が可能になるため、ゲノム配列中の位置を基準として構築されているデータベースとの統合がより容易になった。個々の植物種の DB に登録された遺伝子配列については、配列類似情報に基づいてオルソログ DB との統合を進めている。また、上記のゲノム配列中の位置情報に基づくオルソログの表示手法を開発し、複数の種に跨がってシンテニー様の位置関係にあるオルソログの描画ツールを作成した。

(2) 植物リソース DB の拡充による植物ゲノム DB の統合

リソース DB を用いて検索できる各植物種のリソース情報(種子や cDNA クローン、EST など)を明記し利便性の向上を図ることを目標として開発を推進した。他の有用性の高い植物種として、園芸作物として価値の高いオランダイチゴの複数種に由来する 660,323 遺伝子ならびに我が国で生産量の高い花卉であるカーネーション由来の 43,266 遺伝子を PGDBj の統合検索システムに登録し、提供を開始した。また、これに並行して、理研 BRC が開発し、かずさ DNA 研究所が強化、これまでに 14 生物種、計 1,504,022 遺伝子情報を収めた、植物リソースに基づく遺伝子の串刺し検索データベースである SABRE DB に上記リソース情報を加えることで、SABRE ならびにその API による検索系を広く提供し、リソース情報の統合的な提供範囲の拡大を進めた。また、本課題のデータの RDF 化を実施し、統合的な利用を促進するためのデータの提供を開始した。

(3) DNA マーカー情報の拡充による植物ゲノム DB の統合

マーカー DB を拡充することで系統間あるいは植物間での比較解析基盤を作ることを目的として、新たにゲノム解読された植物種やユーザから要望のあった植物種を新たに 25 種加え、合計 80 種について基本情報、ゲノム解読情報、ゲノム関連 DB リンクの日本語版および英語版ページを作成し、公開した。DNA マーカー情報キュレーションでは、4 種のアブラナ科作物を含む計 11 種の DNA マーカーリストと QTL リストを公開し、これまでの公開植物種の数 41 とした。また、第一期研究開発で未着手のナス科、マメ科等の主要作物のキュレーションを行い、公開マーカー数を 101,341 件、QTL 数を 2,965 件とした。さらに、新たに遺伝地図と物理地図を追加し、それぞれ計 24 種および 11 種の情報を公開した。これらに加えて、病害抵抗性に関連する遺伝子やアルルなど機能情報の収集を進めており、マメ科ラッカセイとナス科トマトの約 70 の病害抵抗性遺伝子についての情報をまとめ、公開の準備を行った。今後は、マーカーや機能遺伝子と、各植物種のゲノム配列との対応付けを進める計画である。

(4) DB 間の連携による統合化

PGDBj 内の DB 間の連携強化を目的として、オントロジーの調査と Resource Description

Framework (RDF) データ構造の検討、PGDBjコンテンツの RDF 化を実施した。コンテンツの内、オルソログ DB エントリー (RefSeq Database Release 57 に基づく緑色植物 20 種およびラン藻 111 種) とマメ科ラッカセイの病害キュレーションデータから RDF データを生成し、オルソログ DB に由来する 6,140,819 件、病害キュレーションデータに由来する 917 件について、SPARQL 検索ができるようにシステム整備を実施した。今後は、リソース DB やマーカー DB 等の残りの PGDBj コンテンツに加え、かずさ DNA 研究所が保有するゲノム関連 DB および EST 関連 DB についても RDF 化を実施する。また、SPARQL 検索に一層適した RDF データの構造改良を進める。更に、統合化推進プログラム全体で共有できるデータ構造について議論し、PGDBj コンテンツに適用することにより、統合化推進プログラムの他課題「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」等との連携を実施する。

(5) オントロジーの整備による横断検索の効率化

PGDBj の内部 DB 間および外部 DB 間にわたる横断検索の効率化を目的として、植物関連オントロジー整備のための調査を実施した。整備にあたり、国際標準となっている主要オントロジー Plant Ontology (PO) を始め、PO が推奨する 10 件の主要オントロジーを中心に調査を行い、PGDBj コンテンツの RDF 化に不足している語彙を抽出した。これまでに、PGDBj で取り扱っているマーカータイプに関係する語彙 5 件、作物品種などの集団に関係する語彙 4 件、植物病害に関係する語彙 6 件が主要オントロジーで未定義であり、優先的に整備される語彙として選定された。また、植物研究コミュニティとの連携として、オントロジー整備ワーキンググループ会議を開催し、植物研究の実用的なオントロジー整備に関する要望を聴取した。今後は、PGDBj コンテンツの RDF 化に向けて選定した語彙のオントロジー整備を実施する。また、研究コミュニティの要望を踏まえ、植物関連学会が進めている植物学用語整備を効率化するためのプラットフォームの作成と、農業形質および生育環境に関するオントロジー整備を実施する。

(6) ゲノム情報のアノテーションの高度化

本研究開発課題内で異なる観点に基づき収集されたオルソログ遺伝子や EST、DNA マーカー付随配列、変異体の原因遺伝子等の配列情報をマッピングし、集約するための基盤となる各生物種のゲノム情報 (配列バージョンや公開機関など) の調査を実施した。PGDBj で取り扱っている 32 科 80 生物種の内、25 科 53 種のゲノム配列が公開されていた。また、マーカー DB エントリーの 7 科 12 生物種のマーカー付随配列について BLAST 検索を実施し、物理地図上の位置情報 23,291 件を集積した。今後は、マーカー付随配列に加えて、遺伝子配列についても同様に物理地図上の位置情報を集積し、物理地図へのマッピングを実施する。また、高精度なマニュアルキュレーション DB である UniProtKB に対して BLAST 検索を実施し、遺伝子配列と機能情報のリンクを構築してゲノム情報のアノテーションの高度化を図る。この作業は、物理地図上への配列の集約やゲノム情報のアノテーションが不十分である植物種を対象として実施する。

§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) 0 件):

② 未発行論文数(“accepted”、“in press”等)(国内(和文) 0件、国際(欧文)0件)

③ 論文詳細情報

無し

(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)

平川英樹、市原寿子、浅水恵理香、中谷明弘、中村保一、田畑哲之、「PGDBj 植物ゲノム関連データベースの統合」実験医学増刊 vol. 32, No. 20, pp.74-75, 2014

(3-4) 国際学会および国内学会発表

① 招待講演 (国内 0件、国際 0件)

② 口頭講演 (国内 0件、国際 0件)

③ ポスター発表 (国内 1件、国際 1件)

〈国内〉

市原寿子、浅水恵理香、平川英樹、中谷明弘、中村保一、田畑哲之、植物ゲノム情報活用のための統合研究基盤の構築、第37回日本分子生物学会年会、パシフィコ横浜、横浜、2014年11月25~27日(特別企画:バイオデータベースコーナー)

〈国際〉

Ichihara H, Asamizu E, Nakaya A, Nakamura Y, Shirasawa S., Hirakawa H, Tabata S. Plant Genome DataBase Japan (PGDBj) in 2015: a comprehensive database covering information of plant genome-related databases in Japan, PAG XXIII, San Diego, USA, 2015年1月12日

(3-5) 知的財産権の出願

無し

(3-6) 受賞・報道等

無し

(3-7) その他(学会年会等における展示ブースでの広報活動)

〈国内〉

1. 植物ゲノム情報活用のための統合研究基盤の構築(<http://pgdbj.jp/>), 第32回日本植物細胞分子生物学会, アイーナ(いわて県民情報交流センター), 盛岡, 2014年8月21~22日(企業展示)
2. 植物ゲノム情報活用のための統合研究基盤の構築(<http://pgdbj.jp/>), 日本植物学会第78回大会, 明治大学, 川崎, 2014年9月12~14日, (企業展示)
3. 植物ゲノム情報活用のための統合研究基盤の構築, 126回日本育種学会秋季大会,

南九州大学, 都城, 2014年9月26~27日(企業展示)

4. 公益財団法人かずさDNA研究所, 日本園芸学会平成26年度秋季大会, 佐賀大学, 佐賀, 2014年9月27~28日(企業展示)
5. 植物ゲノム情報活用のための統合研究基盤の構築(<http://pgdbj.jp/>), 第56回日本植物生理学会年会, 東京農業大学, 東京, 2015年3月16~18日(企業展示)
6. Plant Genome DataBase Japan (PGDBj), 127回日本育種学会春季大会, 玉川大学, 町田, 2015年3月21~22日(企業展示)
7. 植物ゲノム情報活用のための統合研究基盤の構築(<http://pgdbj.jp/>), 日本植物病理学会100周年記念大会, 明治大学, 東京, 2015年3月27~29日(企業展示)

〈国際〉

1. PAG XXIII Plant Genomics Bioinformatics Outreach Consortium, PAG XXIII, San Diego, USA, 2015年1月11~13日(企業展示)

§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
H26年6月30日	グループ合同ミーティング(非公開)	かずさDNA研究所	6人	研究計画確認、進捗報告のためのミーティング
H26年9月16日	平成26年度 第一回アドバイザー委員会(非公開)	ステーションコンファレンス東京	27人	本研究開発課題全体に対し、外部有識者から意見聴取するためのミーティング
H26年12月11日	平成26年度 植物統合DBオントロジー整備ワーキンググループ委員会(非公開)	フクラシア東京ステーション	13人	植物関連オントロジー整備に向けた外部有識者からの意見聴取のためのミーティング
H27年2月18日	グループ合同ミーティング(非公開)	大阪大学	6人	研究進捗報告のためのミーティング
H27年3月4日	平成26年度 第二回アドバイザー委員会(非公開)	ステーションコンファレンス東京	27人	本研究開発課題全体に対し、外部有識者から意見聴取するためのミーティング
H27年3月30日	果樹関連オントロジー構築ミーティング(非公開)	明治大学	3人	果樹関連オントロジー整備に向けた調査報告と作業方針決定のための外部有識者を含むミーティング

以上

別紙1 既公開のデータベース・ウェブツール等

No.	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)	備考
01	植物ゲノム情報活用のための統合研究基盤の構築	植物ゲノム統合データベース (PGDBJ, Plant Genome Database Japan)	国内で提供される植物ゲノム関連情報(塩基配列、遺伝子、DNAマーカー、連鎖地図等)や植物リソース情報の総合検索をめざしたデータベースです。ゲノム解析とマーカー開発の対象となった植物の基本情報も提供しています。	http://pgdbj.jp/			オルソログDBはエントリ数が膨大であり、結果表示に長時間要する為、本機能の対象外とした