

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

田畑 哲之

(公財)かずさ DNA 研究所

所長

「植物ゲノム情報活用のための統合研究基盤 の構築」



©2016 田畑 哲之(かずさDNA研究所) licensed under CC表示2.1 日本

§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

- ・遺伝子オルソログ DB の拡充による植物ゲノム DB の統合
- ・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・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

(3)「龍谷大」グループ

① 主たる共同研究者： 浅水 恵理香 (龍谷大学農学部 准教授)

② 研究項目

- ・DNA マーカー情報の拡充による植物ゲノム DB の統合

§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

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:現在、公開中のオルソログ DB で用いられている配列データの更新と新たにゲノム配列が解読された植物種の追加を行う。また、統合化のハブとして構築した PGDBj (Plant Genome DataBase Japan; <http://pgdbj.jp>) におけるオルソログテーブルや遺伝子情報に関するページを主な対象として表示システムの改善を行う。本年度は、前年度に作成したオルソログ DB のサブデータベースに対してウェブブラウザ経由で検索できるサイトを PGDBj 上で公開することを達成目標とした。

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 等)を明記し利便性の向上を図る。また、SABRE DB に収録された植物種リソースに国内の研究機関が保有する他の植物種のリソースを追加して統合化の範囲を拡大する。また、cDNA や EST クローンの塩基配列とオルソログ DB をリンクし、遺伝子情報との関連付けを行なう。さらに、リソース DB を永続的に運営するためのデータ更新や追加、相互リンク維持のための自動化を目的とした管理システムを構築する。本年度は、追加候補の植物リソースについてデータ構造の調査を実施し、またそれらを効率的に追加するための管理システム構築を達成目標とした。

3. DNA マーカー情報の拡充による植物ゲノム DB の統合:第一期研究開発で未着手のナス科、アブラナ科、マメ科等の主要作物の他、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ものを追加し、合計 80 種についての DNA マーカーと QTL 情報の文献キュレーションを進め公開する。また、形質に関わる機能遺伝子や重要アレルについてもキュレーションし(以下、形質キュレーション)、公開する。当初の計画では、形質キュレーションの対象として「病害抵抗性遺伝子」を中心に検討していたが、海外において充実した内容のデータベースが既存であったため、その他の形質に着目して進めることに変更した。本年度は、DNA マーカーと QTL 情報の文献キュレーション実施による情報の拡充に加え、形質キュレーションのためのシステム化を達成目標とした。

4. DB 間の連携による統合化:PGDBj のオルソログ DB、リソース DB、マーカー DB に対し、Resource Description Framework (RDF) 化を始めとしたセマンティックウェブ技術を適用し、別途 RDF 化が進められている統合化推進プログラムの他課題のデータベースと連携し、統合化を図る。かずさ DNA 研究所が保有する植物関連 DB に対しても RDF 化を実施し、情報の統合化を図る。本年度は、特に DNA マーカーが保有する情報の RDF 構造の決定を達成目標とした。

5. オントロジーの整備による横断検索の効率化:RDF によ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の整備を行い、共通の記述を介して PGDBj の内部および外部データベース間の連携を強化する。本年度は、前年度に選定した PGDBj コンテンツの RDF 化に必要であるが未定義となっている語彙 15 件を RDF 化に適用できるようにすることと、作物の形質語彙のオントロジー構築の方法決定を達成目標とした。

6. ゲノム情報のアノテーションの高度化:収集したオルソログ遺伝子や DNA マーカー付随配列、QTL 情報、変異体の原因遺伝子等の情報を物理地図に集約し、ゲノム上の様々な機能領域に対するアノテーションの高度化を図る。既にコンテンツが RDF 化・公開されている UniProt 等の外部データベースを活用し、効率的なリンク形成と情報集約を実施する。本年度は、前年度に引き続き BLAST 検索によるマーカー付随配列の物理地図座標の集積と、座標データと関連付けた

RDF データを格納・表示するためのシステム構築を達成目標とした。

(2) 本年度の研究開発進捗状況

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:オルソログ DB のサブデータベースである配列類似 DB、配列 DB、系統群 DB に対して、ウェブブラウザ経由で検索できるサイトを公開した (<http://pgdbj.jp/od2/>)。リンク先の DB 群が保持する情報に基づく注釈情報の整備や、アミノ酸配列単位になっているオルソログ情報の遺伝子単位への集約などを行った。また、DB 間のリンク時に個別に生成している DB エントリー間の対応情報を蓄積することにより、ID 変換データベースの構築を行った。更に、PGDBj におけるオルソログテーブルや遺伝子情報に関するページ、ツリーマップに基づく系統プロファイルを主な対象として表示システムの改善を行った。

2. 植物リソース DB の拡充による植物ゲノム DB の統合:PGDBj の横断検索システムに新規に追加する植物リソースとして、ダイコン、トマト、ナス、スサビノリについて内包するデータ構造を調査し、それぞれ、約 62,000 件、約 35,000 件、約 42,000 件、約 10,000 件の予測遺伝子が追加候補として挙げられた。これらの構造に基づき、横断検索システム用の検索インデックスを効率的に生成し、横断検索へ反映させるための管理システムを構築した。

3. DNA マーカー情報の拡充による植物ゲノム DB の統合: DNA マーカー情報に関しては 96,824 件(総数:62 植物種、198,165 件)、QTL 情報に関しては 2,560 件(総数:35 植物種、5,525 件)の情報をキュレーションし PGDBj から公開した。形質キュレーションについては、形質を表すキーワードを NIAS Genebank「植物遺伝資源の特性情報」から選択し、文献要旨から遺伝子名と形質に関するセンテンスを収集した。システム化に際してキュレーションツールを検討した結果、NCBI で作成された PubTator を使用することにした。562 件のキーワードを用いて検索を行ったところ、64,283 件の文献が得られ、この内、6,210 件の文献に由来する生物種と遺伝子にアノテーションが付けられた。この手法はキュレーションの自動化につながる事が期待された。

4. DB 間の連携による統合化:PGDBj 内コンテンツの RDF 化を拡充すると共に、セマンティックウェブ技術による横断検索(SPARQL 検索)に適した RDF データ構造の改良を進めた。特に、DNA マーカー情報に関する RDF 化について、トマト1番染色体に座乗する約 1900 件の SSR マーカーのデータをテストケースとして RDF 化を実施し、関連付ける情報と RDF 構造を決定した。

5. オントロジーの整備による横断検索の効率化:前年度に未定義であった語彙 15 件に PGDBj の仮オントロジーID を定義して PGDBj コンテンツの RDF 化に適用した。作物形質のオントロジー構築については、育成が推奨されている植物 18 種について、農林水産省品種登録ホームページから公開されている「特性表」を用いて調査した。植物毎に語彙の数や定義が異なるため(例:バラ科果樹の語彙数の差-西洋スモモ 88 件、ウメ 43 件)、植物毎のオントロジー構築を決定した。また、尺度を考慮したオントロジー構造のデザインを行った。

6. ゲノム情報のアノテーションの高度化:ゲノム配列が染色体数に収束している 12 植物種に由来する DNA マーカーの付随配列について BLAST 検索を実施し、物理地図上の座標情報 67,001 件を集積した。また、JBrowse のシステム整備を実施し、トマトをテストケースとして、DNA マーカーDB については集積した座標情報を、オルソログ DB については Refseq から抽出した座標情報を用いて物理地図上への集約を試みた。また、RDF ストアを介して検索した DNA マーカーおよびオルソログ DB エントリーを JBrowse 上に表示させるシステムを構築した。

§3. 成果発表等

(3-1) 原著論文発表

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

① 公開中のデータベース・ウェブツール等

別紙1参照。

② 未公開のデータベース・ウェブツール等 (非公開)

別紙2参照。

(3-3) データベース・ウェブツール等の活用事例 (非公開)

無し

(3-4) その他の著作物(総説、書籍など)

1. Akihiro Nakaya, Hisako Ichihara, Erika Asamizu, Sachiko Shirasawa, Yasukazu Nakamura, Satoshi Tabata and Hideki Hirakawa, “Plant Genome DataBase Japan (PGDBj)”, *Methods in Molecular Biology: Plant Genomics Databases*. (accepted)

(3-5) 国際学会および国内学会発表

① 招待講演 (国内 1 件、国際 0 件)

〈国内〉

1. 平川英樹・市原寿子、植物統合データベース PGDBj (Plant Genome DataBase Japan) の構築、日本園芸学会平成 28 年春季大会小集会、東京農業大学農学部厚木キャンパス、3 月 25 日

② 口頭講演 (国内 0 件、国際 0 件)

③ ポスター発表 (国内 4 件、国際 1 件)

〈国内〉

1. 白澤沙知子、PGDBj～植物 80 種のゲノム情報の集約～、第 128 回日本育種学会平成 27 年秋季大会、新潟、9 月 12 日
2. 市原寿子、植物ゲノム関連情報へのセマンティックウェブ技術の適用、第 128 回日本育種学会平成 27 年秋季大会、新潟、9 月 12 日
3. 市原寿子、植物ゲノム情報活用のための統合研究基盤の構築、日本園芸学会平成 27 年秋季大会、徳島、9 月 27 日
4. 市原寿子、植物ゲノム関連情報を統合するためのポータルサイト PGDBj のコンテンツに対するセマンティックウェブ技術の適用、第 129 回日本育種学会平成 28 年春季大会、横浜、3 月 22 日

〈国際〉

1. Hisako Ichihara, Plant Genome Database Japan (PGDBj) in 2016: A Comprehensive Database Covering Information of Plant Genome-Related

(3-6) 知的財産権の出願

無し

(3-7) 受賞・報道等

無し

(3-8) その他(学会年会等における企業展示ブースでの広報活動) (非公開)

〈国内〉

1. 第33回日本植物細胞分子生物学会、東京大学農学部弥生キャンパス、東京、2015年8月10～12日
2. 日本植物学会第79回大会、朱雀メッセ:新潟コンベンションセンター、新潟、2015年9月6～8日
3. 第128回日本育種学会平成27年秋季大会、新潟大学五十嵐キャンパス、新潟、2015年9月11～12日
4. 日本園芸学会平成27年秋季大会、徳島大学常三島キャンパス、徳島、2015年9月26～27日
5. BioJapan 2015 World Business Forum、パシフィコ横浜、横浜、2015年10月14～16日
6. BMB2015、神戸ポートアイランド、神戸、2015年12月1～4日
7. 第57回日本植物生理学会年会、岩手大学上田キャンパス、盛岡、2016年3月18～20日
8. 第129回日本育種学会平成28年春季大会、横浜市立大学、横浜、2016年3月21～22日
9. 日本園芸学会平成28年春季大会、東京農業大学厚木キャンパス、厚木、2016年3月26～27日

〈国際〉

1. Agricultural Biological Database Outreach Consortium, PAG XXIV, San Diego, USA, 2016年1月9～13日

§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
H27年6月26日	グループ合同ミーティング(非公開)	龍谷大学瀬田キャンパス	8人	研究計画確認、進捗報告のためのミーティング
H27年12月24日	平成27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	25人	外部有識者からの意見聴取のためのミーティング

以上

別紙1 既公開のデータベース・ウェブツール等

No.	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	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/	2012/8/8	Erika Asamizu et al., "Plant Genome DataBase Japan (PGDBj): a portal website for the integration of plant genome-related databases.", Plant Cell Physiol, vol. 55, pp. e8, 2014 (DOI: 10.1093/pcp/pct189)
2	新規	植物ゲノム情報活用のための統合研究基盤の構築	PGDBj Ortholog Database	オルソログDBのサブデータベース(配列類似DB、配列DB、系統群DB)に対してウェブブラウザ経由で検索できるように公開したサイト	http://pgdbj.jp/od2/		Akihiro Nakaya, Hisako Ichihara, Erika Asamizu, Sachiko Shirasawa, Yasukazu Nakamura, Satoshi Tabata and Hideki Hirakawa, "Plant Genome DataBase Japan (PGDBj)", Methods in Molecular Biology: Plant Genomics Databases. (accepted)