

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）

「植物ゲノム情報活用のための
統合研究基盤の構築」

研究開発期間：平成26年4月～平成29年3月

研究開発終了報告書

研究代表者：田畑哲之

((公財)かずさDNA研究所 副理事長・所長)

§1. 研究開発実施の概要

平成 23 年度～平成 25 年度の統合化推進プログラム(以下、第一期)で構築された統合化のハブとして機能するポータルサイト(Plant Genome DataBase Japan;<http://pgdbj.jp>)によって、幅広い植物についてのゲノム関連諸情報への効率的なアクセスが可能となった。平成 26 年度～平成 28 年度の統合化推進プログラム(以下、第二期)では、こうして取得できるようになったゲノム関連情報を、さらに快適な環境で活用できる統合的な研究基盤の構築を目指し、以下に示す6つの開発項目を掲げた。この際、他課題との連携を深めることにより DB の統合範囲をより拡充することにも留意し、これまでに構築したコンテンツを生かすためのツールやシステムの開発を実施した。

(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 の構築に使用する NCBI の RefSeq 由来の配列データを更新した。また、各アミノ酸配列のゲノム配列中での位置情報(染色体番号と塩基位置)を RefSeq 内のデータに基づいて整理した。これによりゲノム配列中の位置を基準として構築された DB との統合が容易になった。さらに、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。PGDBj におけるオルソログコンテンツの表示システムの改善やウェブブラウザ経由で検索できるサイト(<http://pgdbj.jp/od2/>)も公開した。

(2) 植物リソース DB の拡充による植物ゲノム DB の統合: 第一期において開発した植物リソース横断検索システムに、新規のリソース情報として6植物種の遺伝子情報を整備、追加した。また、植物リソース DB エントリーである cDNA や EST クローンの塩基配列と、項目(1)のオルソログ DB をリンクさせ、遺伝子情報との関連付けを行なうことを目的とした、「PGDBj オルソログ DB-植物リソース連携 DB」を構築した。その結果、連携 DB から、オルソログ DB のリリース、クラスタ ID、植物種、情報の種類等を指定することで、オルソログクラスタ、植物リソースのエントリー情報、関係情報が取得可能となった。

(3) DNA マーカー情報の拡充による植物ゲノム DB の統合: 第一期から継続してキュレーションによって収集した DNA マーカーと QTL の情報を PGDBj から公開した。QTL および形質に連鎖する DNA マーカーについては、Trait Ontology (TO) 情報の付加と分類を実施した。また、文献から形質関連情報を持つ遺伝子をキュレーションし、TO 情報を介して DNA マーカー・QTL DB と統合した。これによって、形質に基づく DNA マーカーや遺伝子に関する情報の検索が可能となった。さらに、本 DB のユーザの拡大および種苗産業による利用促進を目的として、育種現場で必要とされる情報を選択・整理し、育種向け DNA マーカー検索ページを公開した。

(4) DB 間の連携による統合化: 第一期で構築された PGDBj の各 DB 間の連携をセマンティックウェブ技術によって強化することを目的として、PGDBj の各 DB エントリーとそれに付随する情報を RDF 化し、それらを管理・運用するための情報基盤を構築した。特に、オルソログ DB エントリーを介して、Uniprot 等の外部 DB エントリーとのリンクが容易になったことから、PGDBj で構築したデータベースの活用範囲が拡大した。

(5) オントロジーの整備による横断検索の効率化: 植物研究分野における実用的なオントロジー構築とその普及を目的として、オントロジー整備ワーキング委員会を設置した。植物関連学会で用語整備に取り組んでいることから、学会による用語の整備作業を支援する情報基盤を構築した。さらに、オントロジー整備に際し、植物研究コミュニティとの連携強化を目的として、オントロジーの有効な活用事例を示した。

(6) ゲノム情報のアノテーションの高度化: PGDBj のコンテンツを物理地図上に集約し、アノテーションの高度化を試みた。ゲノムブラウザとしては JBrowse (<https://jbrowse.org/>)を採用した(<http://pgdbj.jp/jbrowse/index.html?data=Brapa> 等)。PGDBj の JBrowse では、対象ゲノムの遺伝子モデルを、相同性に基づいて、機能アノテーションが充実しているシロイヌナズナやトマトなどの遺伝子モデルに関連付けしたデータが実装されており、その結果、対象ゲノム上に集約させた遺伝子の情報や QTL 情報、DNA マーカー情報について、生化学的機能や生物学的機能に関する知見を得ることが可能となった。

§2. 研究開発実施体制

1. 研究グループ

(1) 「研究代表者:田畑哲之」グループ (研究代表者グループ)

人員構成

氏名	所属機関	役職	研究開発項目	参加時期
田畑 哲之	(公財)かずさ DNA 研究所	副理事長・所長	総括	H26.4～ H29.3
平川 英樹	同上	グループ長	項目3～6	H26.4～ H29.3
中村 保一	同上	特別客員 研究員	項目2, 4	H26.4～ H29.3
駒井 多恵子	同上	特任研究員	項目3, 6	H27.10～ H29.3
白澤 沙知子	同上	特任研究員	項目3, 6	H26.12～ H28.2
浅水 恵理香	同上	特任研究員	項目3	H26.4～ H27.3
市原 寿子	同上	特任研究員	項目1～6	H26.4～ H29.3
小原 光代	同上	技術員	項目3～4	H26.4～ H29.3
藤代 維一	同上	技術員	項目3, 6	H26.4～ H27.3
石井 崇洋	同上	特任技術員	項目3～4	H26.4～ H27.4
阿久津 智子	同上	特任技術員	項目3	H26.4～ H29.3
佐藤 かおり	同上	補助員	項目3	H26.4～ H29.3
渡邊 裕佳	同上	補助員	項目3	H26.4～ H29.3
野口 佳代	同上	補助員	項目3	H26.5～ H29.3
小池 浩太郎	同上	補助員	項目3	H26.4～ H27.10
鹿島 尚悟	同上	補助員	項目3, 6	H28.1～ H29.3

担当項目

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
阪大グループが生成したオルソログのデータセットを PGDBj 横断検索システム上で効率的に検索するための機能を拡張した。また RDF 化し、RDF データの運用システムに実装した。
2. 植物リソース DB の拡充による植物ゲノム DB の統合
第一期において構築したリソース DB に対し、新たな植物種の配列情報とリソース情報を追加した。また、リソース DB エントリーとオルソログ DB エントリーを連携させた DB を構築するシス

ムを開発し、「PGDBj オルソログ DB-植物リソース連携 DB」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
第一期において未着手であったナス科、アブラナ科、マメ科等の作物に加え、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について、DNA マーカーと QTL 情報の文献キュレーションを実施した。また、検索インターフェースの改良と追加を実施した。
4. DB 間の連携による統合化
PGDBj の内部コンテンツを RDF 化し、RDF データの運用システムを構築した。統合化推進プログラムの他課題と RDF 構造を共有し、連携を強化した。また、第一期に開発した横断検索システムの機能を拡張した。BLAST 検索により、オルソログ DB エントリーと外部 DB エントリーとの対応付けと横断検索インデックスの更新を実施し、横断検索対象を拡充した。
5. オントロジーの整備による横断検索の効率化
PGDBj 内部コンテンツの RDF 化に使用する語彙を整備した。また、植物研究コミュニティからのオントロジー構築に関する要望を踏まえ、植物学用語を整備するための情報基盤を構築した。そして、オントロジー活用事例を植物研究コミュニティへ示した。
6. ゲノム情報のアノテーションの高度化
PGDBj の内部 DB に登録されている種々の配列情報を、BLAST 検索により物理地図上に関連付けて集約し、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実施した。

(2) 「研究分担者:中谷明弘」グループ (主たる共同研究者グループ(1))

人員構成

氏名	所属機関	役職	研究開発項目	参加時期
中谷 明弘	大阪大学 大学院医学系研究科	特任教授	項目1	H26.4～ H29.3

担当項目

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 で用いられている配列データの更新と対象植物種の追加を実施した。また、オルソログ DB コンテンツの表示システムを改良した。

(3) 「研究分担者:浅水恵理香」グループ (主たる共同研究者グループ(2))

人員構成

氏名	所属機関	役職	研究開発項目	参加時期
浅水 恵理香	龍谷大学 農学部	准教授	項目3	H27.4～ H29.3
岡 真規子	同上	補助員	項目3	H27.9～ H29.3

担当項目

3. DNA マーカー情報の拡充による植物ゲノム DB の統合
植物の重要形質に関わる遺伝子を、文献からキュレーション(以下、形質キュレーション)した。また、関連する遺伝子配列を収集し、他の PGDBj コンテンツとの配列ベースでの対応付けを実施した。さらに、形質キュレーションに基づいて Plant Ontology(PO), Trait Ontology (TO)を付与し、DNA マーカーDB と統合した。

2. 有識者会議等

○植物統合 DB アドバイザリー委員会

(1) 会議概要

名称	植物統合 DB アドバイザリー委員会
目的	植物関連 DB を運用・使用している研究者との意見交換を通じて、植物研究の現場において実用的な DB サービスを目指す。
委員数	8人

(2) 開催歴

年月日	場所	参加人数	主な議題
26年9月16日	ステーションコンファレンス東京 602号室 A+B	27人	研究開発課題(1)～(6)の進捗報告とそれに対する意見交換
27年3月4日	ステーションコンファレンス東京 605号室 B+C	27人	研究開発課題(1)～(6)の進捗報告とそれに対する意見交換
27年12月24日	ステーションコンファレンス東京 605号室 B+C	24人	研究開発課題(1)～(6)の進捗報告とそれに対する意見交換
28年8月18日	ステーションコンファレンス東京 605号室 B+C	27人	研究開発課題(1)～(6)の進捗報告とそれに対する意見交換

○オントロジー整備ワーキング委員会

(1) 会議概要

名称	オントロジー整備ワーキング委員会
目的	植物学およびオントロジーの専門家との意見交換を通じて、植物研究分野における実用的なオントロジー構築とその普及を目指す。
委員数	4人

(2) 開催歴

年月日	場所	参加人数	主な議題
26年12月11日	フクラシア東京ステーション 5階 会議室 I	13人	(1)植物研究分野におけるオントロジーの重要性、(2)RDF化
28年5月31日	メルパルク京都 6階 会議室 5	13人	進捗報告および最終年度としてのゴールの確認

§3. 研究開発の目的、実施内容及び成果

1. 研究開発の背景

平成23年度から実施した統合化推進プログラム(以下、第一期)では、植物ゲノム関連情報へのアクセシビリティを向上させるため、関連する国内外 DB の統合化を実施した。当時、国内において植物ゲノム DB が散在する一方で、ゲノム関連情報として重要である DNA マーカーの多くは DB 化されずに文献に留まる傾向にあった。また植物バイオソース情報もさらなる統合化が望まれていた。こうした状況を受け、(1)遺伝子オルソログ DB の構築とそれに基づく植物ゲノム DB の統合、(2)DNA マーカーおよび連鎖地図情報に基づく植物ゲノム DB の統合、(3)植物リソース情報 DB の統合、(4)植物研究に関連する情報基盤の構築、という4つの開発項目を掲げ、情報の収集、整備、相互リンクを通じて植物ゲノム関連データベースの統合を実施した。各種情報を統合するための基盤として横断検索システムを実装したポータルサイト PGDBj (Plant Genome DataBase Japan; <http://pgdbj.jp>) を構築・公開した。第一期の各開発項目と成果は以下のとおりである。

(1) 遺伝子オルソログ DB の構築とそれに基づく植物ゲノム DB の統合

オルソログを定義することは、植物種間のシンテニーに基づいた植物のゲノム進化や遺伝子機能の推定において重要な手掛かりとなる。モデル植物から農作物、藻類までの幅広い生物種を網羅したオルソログ DB は存在しなかったことから、生物種間の進化系統関係を反映させた階層的なオルソログ DB を構築した。対象生物は、20 種の緑色植物(約 50 万配列)と 111 種(約 50 万配列)のラン藻とした。また、ゲノムが決定されている植物種の 22 件の外部 DB に対して、オルソログ DB のエントリーとの間でリンク情報を整備した。本 DB 構築に際し、配列データと DB アクセス用の URL の情報で半自動的にデータ追加を可能とするシステムを構築した。また、RefSeq DB とのリンク情報を介して、各植物種ゲノム関連 DB への接続が可能となった。また URL の指定により、オルソログ DB のエントリーに直接アクセスできるように API を公開した。セマンティックウェブ技術の適用を想定した、DB の設計とタブ区切りデータの生成も実施した。

(2) DNA マーカーおよび連鎖地図情報に基づく植物ゲノム DB の統合

DNA マーカーや遺伝的連鎖地図は、遺伝学的研究や育種を行う上で重要なツールであり、モデル実験植物や実用植物において、様々な種類の DNA マーカーとそれに基づく連鎖地図が構築されてきた。近年の次世代型シーケンサーによる多種多様な植物種のゲノム配列分析の結果、大量の DNA マーカーを迅速かつ安価に開発できるようになると同時に、連鎖地図上の位置情報のみが既知であった DNA マーカーについても、その付随配列を用いてゲノム配列(物理地図)上にマッピングすることが可能になった。連鎖地図と物理地図との対応関係を明らかにすることによって、有用形質に関与する遺伝子の同定が加速することが期待できる。第一期では、全ゲノム配列が決定された植物種および DNA マーカーの情報が蓄積している 55 植物種を対象として、DNA マーカーと QTL に関する情報を文献や国内外の DB から収集した。これらの植物種のうち、主要な植物種(シロイヌナズナ、イネ、コムギ、トウモロコシ)については他研究機関において既にマーカー DB が整備されていたためリンクに留め、それ以外の植物種については、文献からマニュアルキュレーションを実施し、DB を構築した。本 DB 構築に際し、キュレーションの効率化を図るため、データ入力をウェブ上で行うシステムを開発した。また、ゲノム解読された植物種について、マーカーの位置を連鎖地図と物理地図上に集約させた。

(3) 植物リソース情報 DB の統合

国内に存在する種子や cDNA、ゲノム DNA クローンなど植物バイオリソースの情報を統合的に検索できるようにするため、PGDBj と理研 BRC の検索システム SABRE DB が連携し、外部利用が可能な API を整備した。この結果、PGDBj から理化学研究所バイオリソースセンター (<http://ja.brc.riken.jp/>、以下、理研 BRC) と遺伝学研究所のナショナルバイオリソースプロジェクト (<http://www.nbrp.jp/>、以下、NBRP) が保有する 15 植物種 154 万件のリソースに対する横断検索が可能となった。また、近畿大学と果樹研究所で維持されているカンキツ類リソース(cDNA クローンとゲノム塩基配列情報)として、約 900 種の在来種の保存株情報に関するインデックスサイトを構築した。また、公開可能な cDNA クローンリソースとゲノム塩基配列解析情報については、配列データベースを構築し、横断検索ができるようになった。

(4) 植物研究に関連する情報基盤の構築

モデル植物に限らず様々な作物についてゲノム配列が解読されつつあり、これらの研究論文には、ゲノム解析手法(アセンブルやマッピング、遺伝子抽出、アノテーション手法)や植物の諸性質の情報が記載されている。それらを整理し DB 化することで、研究の利便性が向上すると考えられた。そこで、全ゲノム配列が決定された植物および国内で DNA マーカーの研究が行われている 55 植物種を対象として、植物の基本情報(学術名、和名、科名、Taxonomy ID (NCBI へのリンク))や農学、園芸学、育種学的情報、そして、ゲノム配列が解読された植物種については、ゲノム解析手法(シーケンサーの種類やアセンブル手法、遺伝子情報抽出法、機能予測法など)を文献と各種 DB からマニュアルキュレーションし、ポータルサイトで公開した。また、収集した国内外植物関連 DB を分類し、DB リンク集として公開した。また、第一期における本プログラムの研究開発課題「メタボローム・データベースの開発」により構築された生物種・代謝物データベース DB (KN ApSAcK; <http://kanaya.naist.jp/KNApSAcK/>)と質量分析データ DB (MassBase; <http://web.s2.kazusa.or.jp/massbase/>)へのリンクを張り横断検索の対象とすることで連携の体制を整えた。

以上のように、第一期で構築された統合化のハブとして機能するポータルサイト(Plant Genom

e DataBase Japan; <http://pgdbj.jp>)によって、幅広い植物についてのゲノム関連諸情報への効率的なアクセスが可能となった。平成 26 年度～平成 28 年度の統合化推進プログラム(以下、第二期)では、こうして取得できるようになったゲノム関連情報をさらに快適な環境で活用できる統合的な研究基盤を構築すると共に、他課題との連携を深めることによって DB の統合範囲を拡充することを目的として、新たなツールやシステムの開発を実施した。

2. 研究開発対象のデータベース・ツール

(1) データベース

主要なもの

正式名称	略称	概要
Plant Genome Database Japan	PGDBj	植物ゲノム関連情報を統合化するハブとして構築したポータルサイトである。進化情報、リソース情報、ゲノム上の位置や構造情報を軸に遺伝子機能等を検索できる。植物に特化した多種類の DB へ効率的にアクセスできる。 http://pgdbj.jp/
PGDBj オルソログデータベース		緑色植物とラン藻のアミノ酸配列間のホモロジー情報に基づいて生成したオルソログ情報のデータベースである。サブデータベース(アミノ酸配列 DB、ホモロジーDB、オルソログ間の階層)を含めてウェブブラウザ経由で検索できる。 http://pgdbj.jp/od2/

上記以外のもの

正式名称	略称	概要
PGDBj DNA マーカー・QTL データベース		マニュアルキュレーションして構築した DNA マーカーと QTL の情報のデータベースである。名前、マーカーの種類、連鎖する形質の種類、物理地図上の位置情報等に基づき検索できる。PGDBj に内包されるデータベースである。 DNA マーカー: http://pgdbj.jp/markerdb/marker.html?sbj=m&ln=ja QTL: http://pgdbj.jp/markerdb/marker.html?sbj=q&ln=ja
PGDBj カンキツリソースデータベース		近畿大学と果樹研究所が保有する約 900 個体分の在来種、栽培品種のカンキツリソース情報についての情報カード型データベースである。INSDC に登録された Citrus 属の配列情報も含む。横断検索システムから検索できる。PGDBj に内包されるデータベースである。 http://pgdbj.jp/estui/citrus/CR.html?ln=ja&cr=cibase
ゲノム解読状況データベース		PGDBj 対象植物の中のゲノム解読に関する情報をまとめたデータベースである。中心となった解読機関サイトへ移動できる。PGDBj に内包されるデータベースである。 http://pgdbj.jp/plantdb/plantgenome.html?ln=ja

(2) ツール等

正式名称	略称	概要
PGDBj 横断検索システム		PGDBj が内包するデータベースおよび植物関連の外部データベースの情報を横断的に検索できる。 http://pgdbj.jp/estui/pidb.html?ln=ja
PGDBj 育種向け DNA マーカーページ		育種に有用な DNA マーカー情報へ簡便にアクセスできるように、構築したインターフェースである。用途別に入り口を設置しており、必要とする情報を素早く取得できる。 http://pgdbj.jp/pages/index.html?dir=ag&page=menu&ln=ja
PGDBj JBrowse システム		ゲノムブラウザ JBrowse 上で PGDBj の DNA マーカーや QTL、オルソログ、リソース関連情報を表示および検索できる。 http://pgdbj.jp/jbrowse/index.html?data=Brapa

※主要なデータベース、ツールの詳細は別紙参照。

3. 達成目標及び実施計画

(1) 当初の実実施計画・達成目標

本研究開発では、第一期において統合化のハブとして構築されたポータルサイト PGDBj を、植物ゲノム情報をさらに快適な環境下で活用するための統合的な研究基盤とすることを目指し、以下の6項目 (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、(2)植物リソース DB の拡充による植物ゲノム DB の統合、(3)DNA マーカー情報の拡充による植物ゲノム DB の統合、(4)データベース間の連携による統合化、(5)オントロジーの整備による横断検索の効率化、(6)ゲノム情報のアノテーションの高度化、の研究開発を行った。そして、これらを通して、植物ゲノムに関する基礎研究の加速、さらには植物ゲノム研究の産業面への応用に繋がるツールや情報提供システムを開発した。

(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

- ・達成目標: 第一期で公開したオルソログ DB で使用さ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。次に、統合化のハブとして構築した 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を主な対象として、表示システムの改善を行う。NCBI の RefSeq DB に登録された遺伝子に加えて、個々の植物種で構築された DB に登録された遺伝子配列を合わせて用いることで、オルソログをより適切に定義する。また、現在のオルソログ DB の遺伝子機能は RefSeq DB の Product 名を引用しており、機能に関する情報が少ないため、ドメインや代謝経路関連の DB に対するリンクを張り、リソース DB やマーカーDB、さらには統合化推進プログラムの他課題「生物種メタボロームモデル・データベースの構築」、「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」や理化学研究所で公開されている SABRE DB と連携することでアノテーションの高度化を図る
- ・平成 26 年度は、ゲノム配列が決められた植物種の各個体についての配列情報を入手し、アミノ酸配列に対する総当たりの BLAST 検索を行いクラスタリングすることでオルソログを定義する。RefSeq DB の配列データの更新を行い、各個体により得られたオルソログ遺伝子間でリンクを張る。平成 27 年度は、SABRE DB をはじめ、統合化推進プログラムの他課題との連携、および、アノテーションの高度化のために各種 DB に対するリンクを整備する。平成 28 年度は、データおよびリンクの更新および追加の自動化スクリプトを作成する。

(2) 植物リソース DB の拡充による植物ゲノム DB の統合

- ・達成目標: リソース DB を通し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。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加えることで、統合化の範囲を拡大する。統合対象としては、オランダイチゴ、カーネーション、ダイコン、トマト、ナス、スサビノリを予定している。また、cDNA や EST クローンの塩基配列と(1)のオルソログ DB をリンクさせることで、遺伝子情報との関連付けを行なう。さらに、リソース DB を持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。
- ・平成 26 年度は、リソース DB から検索できる各植物種のリソースの種類(種子や cDNA クローン、EST など)を付記し、ユーザの利便性向上を図る。SABRE DB との横断検索機能を強化する。また、新たな植物種のリソースを追加する。平成 27 年度は、データ更新と維持を容易にするためにシステムの改善を行う。オルソログ DB との連携により種々の植物リソースから遺伝子との関連性を調べることができるツールの整備を行う。平成 28 年度は、持続的に運営するためのデータ更新や相互リンク維持のための自動化を行う管理システムを構築する。

(3) DNA マーカー情報の拡充による植物ゲノム DB の統合

- ・達成目標: 第一期では 31 種の植物について文献のキュレーションで収集した DNA マーカーと QTL に関する情報を公開した。第二期では第一期の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ズから要望のあった植物種についてキュレーションを実施する。これらの情報に次世代シーケンサーから得られた大量のゲノムやトランスクリプトームの配列データ、メタボローム等のオミックス情報、環境ストレス耐性や病害抵抗性などの機能情報を加えることによって、ビッグデータとし

て用いることができる植物ゲノム情報の集約を図る。具体的には、植物の病害やストレス耐性に関する情報を対象とする。

- ・平成 26 年度は、これまでの研究開発において、論文からのキュレーションが完了しなかった植物種について、DNA マーカーおよび QTL 関連情報の抽出を継続する。ゲノム配列が解読された植物種を追加する。平成 27 年度は、キュレーションおよびゲノム配列が解読された植物種の追加を引き続き行う。平成 28 年度は、データやリンクの更新や追加の自動化を目指したスクリプトの作成を行う。

(4) データベース間の連携による統合化

- ・達成目標：統合化推進プログラムの他課題「生物種メタボロームモデル・データベースの構築」、「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」との連携を深めるため、PGDBj のオルソログ DB、リソース DB、マーカー DB のデータに対し、Resource Description Framework (RDF) 化を始めとするセマンティックウェブ技術を適用する。また、かずさ DNA 研究所が保有するゲノム関連 DB (ミヤコグサ、ヤトロファ、ユウカリ、トマト) および EST 関連 DB (トマト、シロイヌナズナ、ミヤコグサ、クラミドモナス、アカクロバ、スサビノリ) についても同様に RDF 化および各データ間のリンク構築を行い、DB 間の連携による統合化を図る。
- ・平成 26 年度は、PGDBj 内の DB 間の連携を目的として、オルソログ DB、リソース DB、マーカー DB のコンテンツについて、タブ区切りのテキストデータを整備した後、RDF 化を行う。同時に、データ間をリンクするためのゲノム情報や生物種名、オルソログ遺伝子情報を整備する。この際に、統合化推進プログラムの他課題「生物種メタボロームモデル・データベースの構築」、「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」との連携にも対応可能な整備を行う。平成 27 年度は、統合化推進プログラムの他課題との連携を深め、さらに、国内の植物関連データベースで連携できうるものがあれば統合化を進める。平成 28 年度は、PGDBj 内の各種 DB および統合化推進プログラムの他課題、そして、他の研究機関のデータベース間と連携することで植物ゲノム DB の統合化を進める。

(5) オントロジーの整備による横断検索の効率化

- ・達成目標：RDF による横断検索の効率化を目的としてオントロジーの整備を行い、PGDBj の内部 DB だけではなく外部 DB 間についても共通の記述を介して連携を強化する。現在、Gene Ontology、Plant Ontology、Plant Trait Ontology、Environment Ontology が開発されているが、植物リソースや DNA マーカーに関するオントロジーは十分に整備されていない。このため、オルソログ、リソース、DNA マーカー関連に加え、表現型、ミュータント、原因遺伝子、アナトミー、発達ステージ、環境といったオントロジーを植物研究コミュニティと連携して作成し、Bioportal のオントロジー登録申請の仕組みを利用し登録することで、植物学のオントロジーの充実を図る。
- ・平成 26 年度は、PGDBj の内部 DB (オルソログ DB、リソース DB、マーカー DB) を統合するためのオントロジー整備を実施する。オントロジーの整備に向けて、専門家および植物学会の協力を得ながら、単語の定義や語彙の階層的な分類や定義について Plant Ontology、Plant Trait Ontology、Environment Ontology の内容を調査する。また、学会関係用語集や各種辞典 (生物学辞典、学術用語集、Oxford 植物学事典、カンキツ用語集、園芸学用語集、育種学用語集など) に掲載されていてすでに電子化されているデータについては、オントロジーエディタへ取り込み作業を進める。平成 27 年度は、オントロジーを調査し、登録すべき用語をまとめる。RDF にオントロジーを取り入れることで、他の統合化推進プログラムとの連携を深め、横断検索の効率化を図る。平成 28 年度は、オントロジーの調査の自動化を行うスクリプトを開発する。

(6) ゲノム情報のアノテーションの高度化

- ・達成目標：本研究課題で収集したオルソログ遺伝子や EST、完全長 cDNA 配列、DNA マーカー付随配列、QTL 情報、ミュータントの原因遺伝子、多型がみられる遺伝子といった情報を、連鎖地図と物理地図の表示システムと GBrowse 上に集約させる。また、オルソログ遺伝子やミュータントの原因遺伝子等の遺伝子配列に対して、タンパク質の機能情報のハブとして高精度にマニュアルキュレーションされた UniProtKB に対する BLAST 検索を行うことで、UniProtKB とのリンクを構築する。UniProt においては、データの RDF 化や SPARQ

L エンドポイント公開といった、セマンティックウェブ技術を活用した情報提供が行われているため、これらも利用する。さらに、モチーフやタンパク質立体構造、代謝経路など他の DB とリンクを構築することで、アノテーションの高度化を図る。京都大学の KEGG (<http://www.kegg.jp/kegg/>) などにおいても代謝経路や遺伝子の機能情報の RDF 化と SPARQL エンドポイントの公開準備が進められているので、これらを利用した効率的なリンク構築を実施する。

- 平成 26 年度は、オルソログ DB、リソース DB、マーカー DB に登録されている遺伝子配列について、UniProtKB に対してリンクを構築する。また、遺伝子の機能情報として、モチーフやタンパク質立体構造、代謝経路など他の DB とリンクを構築する。平成 27 年度は、オルソログ DB、リソース DB、マーカー DB に登録されている遺伝子配列に加え、マーカー付随配列を物理地図上に集約させ、アノテーションの高度化を図る。平成 28 年度は、各種 DB とのリンクの更新・追加を行う。物理地図上に反映させる自動化スクリプトを作成する。

(2) 期間中に追加・削除・変更した実施計画・達成目標

- 項目(2)「植物リソース DB の拡充による植物ゲノム DB の統合」において、オルソログ DB と植物リソース DB のエントリーをリンクさせるシステム開発に際し、植物リソース DB のみにとどめず、連携対象となる外部 DB 等のエントリーにも適用可能とする設計に変更した(変更)。
- 項目(3)「DNA マーカー情報の拡充による植物ゲノム DB の統合」において、ユーザの拡大および種苗産業での利用促進を目的に、育種に有用な DNA マーカー情報に簡便にアクセスできる育種向け DNA マーカー検索ページを作成した(追加)。また、形質キュレーションの対象として「病害抵抗性遺伝子」から開始したが、海外において充実した内容の DB が存在するため、キュレーションの対象を重要形質全般に広げた(変更)。
- 項目(5)「オントロジーの整備による横断検索の効率化」において、植物関連学会がオントロジー整備をするための情報基盤を構築した(変更)。計画当初は、専門家および植物関連学会の協力を仰ぎながら、本プロジェクト内で用語の定義や語彙の階層分類の整備を進めることとしていたが、植物関連学会でも用語整備の取り組みが進んでいることから、そこで整備された情報の提供を前提として、学会による用語の整備作業を支援する情報基盤を構築することになった。また、オントロジー整備に際し、植物研究コミュニティとの連携強化を目的として、オントロジーの活用例を示すことを課題として追加した(追加)。
- 項目(6)「ゲノム情報のアノテーションの高度化」において、物理地図表示システムとして GBrowse の代わりに後継プログラムである JBrowse を採用した(変更)。また、PGDBj コンテンツの UniprotKB に対する BLAST 検索は中止した(削除)。オルソログ DB エントリーが UniprotKB エントリーと(基本的に)対応付けされている RefSeq ID を保有し、既に対応付けされていたためである。他の PGDBj コンテンツについては、オルソログ DB との対応付けが行われており、同様に本作業は不要であると判断した。

4. 実施内容

(1) 実施内容

開発項目ごとに記述する。

(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

平成 26 年度は、各植物種的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うことでオルソログ DB の更新とエントリー追加を行った。その際、各アミノ酸配列のゲノム配列中での位置情報(染色体番号と塩基位置)を RefSeq 内のデータに基づいて整理した。これまでに、配列の類似情報やアミノ酸配列の ID、注釈情報中のキーワードなどによって検索や DB 統合を行ったが、位置情報によるオルソログ DB の検索を可能にした。これによりゲノム配列中の位置を基準として構築されたデータベースとの統合が容易になった。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きオルソログ DB との統合を進めた。さらに、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(図 1)。

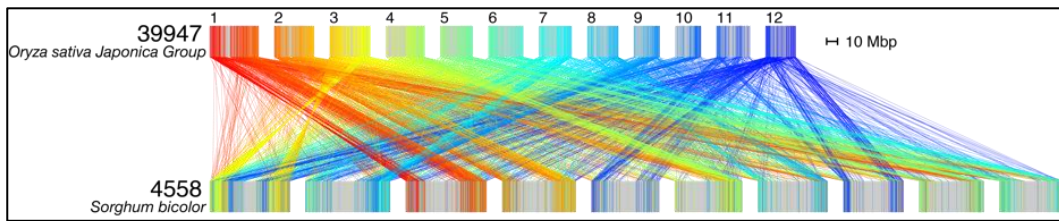


図1 オルソログ情報に基づくシンテニー様の染色体領域の抽出と描画

平成 27 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(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 PGDBj 上で公開した(<http://pgdbj.jp/od2/>)(図 2)。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、D B 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。



図 2 検索用ウェブページ

さらに、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ィールを主な対象として、表示システムの改善を行った(図 3)。



図 3 オルソログ情報の表示(左)とツリーマップによる系統プロフィール情報の表示(右)

平成 28 年度は、オルソログ DB で用いられている NCBI の RefSeq 由来の配列データ(緑色植物 40 種(約 114 万配列)とラン藻 213 種(約 80 万配列))を更新し、個々の植物種の DB において公開された遺伝子配列を合わせることでオルソログを新たに定義した(RefSeq に追従した更新は適宜実施を継続する)。その際、配列 ID の後方互換性について検討し、新旧エントリーのホモロジー情報で対応付ける方針とした。また、オルソログ DB の遺伝子機能は RefSeq の注釈(アノテーション)を引用しているが、クラスタ内のキーワードの頻度に基づいてより適切な情報を表示する機能を準備した。一方で、RDF データを介して PGDBj 内の他の DB および外部 DB のエントリーとの間で関連付けすることにより、オルソログ DB エントリーの注釈情報の充実化を目指した。これに先立ち、RDF 化作業に向けたデータ生成を行った。外部 DB は、ドメインや代謝経路関連の DB の他、統合化推進プログラムの他課題「生物種メタボロームモデル・データベースの構築」や「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」より提供される DB を対象とし、一連の作業の結果、構築したデータベースの活用範囲が拡大した。NBDC にてアーカイブ化(バージョン更新)すると共に、データベースの機能を記述した文書を出版した。

(2) 植物リソース DB の拡充による植物ゲノム DB の統合

第一期において開発した植物リソース横断検索システムに、新規のリソース情報としてかずさ DNA 研究所の DB に由来する 6 植物種の遺伝子情報を整備、追加した。具体的な追加数は、オランダイチゴとイチゴ野生種に由来する約 66 万件、カーネーションの約 4.3 万件、ダイコンの約 6.2 万件、トマトの約 3.5 万件、ナスの約 4.2 万件、スサビノリの約 1 万件、合計 85.2 万件である。

また、植物リソース DB エントリーである cDNA や EST クローンの塩基配列と、項目(1)のオルソログ DB をリンクさせ、遺伝子情報との関連付けを行なうことを目的とした、「PGDBj オルソログ DB-植物リソース連携 DB」を構築した。データベース構築に際し、オルソログ DB と植物リソースの連携データを生成するシステムを開発した。また、連携 DB に格納されたデータを取得・加工し、json および tsv 形式で提供するシステムを開発した。その結果、連携 DB から、オルソログ DB のリリース、クラスタ ID、植物種、情報の種類等を指定して、オルソログクラスタ、植物リソースのエントリー情報、関係情報が取得可能となった。さらに、本システムでは、オルソログ DB のクラスタ情報と連携させるデータセットとして、植物リソース DB だけでなく、項目(4)において連携対象とする各植物種の遺伝子モデルやその他の外部 DB の各エントリーへの適用も想定した設計を行った。これらの開発により、リソース DB およびオルソログ DB の更新や追加に対応した、リソース DB の持続的な運営を可能にする基盤を構築した。

(3) DNA マーカー情報の拡充による植物ゲノム DB の統合

第二期では第一期の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植物種を追加し、対象植物種を 80 種とした。これらの植物種については、ゲノム解読手法、DNA マーカー、QTL 情報、メタボローム、トランスクリプトーム情報へのリンク等、関連するゲノム情報を植物種ごとに集約した植物種基本情報を公開した(図 4 左)。80 種のうち、他研究機関においてすでに充実したデータベースが公開されている作物種を除いた 65 種を文献キュレーション対象とし、平成 29 年 1 月現在、第一期における取得分 31 種と合わせ DNA マーカーに関しては 65 種 264,733 件、QTL 情報に関しては 45 種 16,542 件を PGDBj より公開している。オントロジーによる統合化のため、QTL および形質に連鎖する DNA マーカーについて TO 分類を行い、植物種横断的な検索を可能にした。

ユーザの拡大および種苗産業における利用促進を目的に、育種向け DNA マーカー検索ページを作成、公開した。育種の選抜マーカーとしての利用が期待できる形質連鎖マーカー一覧については、連鎖形質、遺伝子との距離、遺伝子の候補領域など、実際の育種の現場で必要とされる情報を日本語で掲載した。また、染色体全域を大まかに調べるための基本 DNA マーカーセットも作成、公開した(図 4 右)。基本 DNA マーカーの設定には、設備の充実していない中小種苗会社や大学の研究室においても使えるよう、電気泳動で検出可能なマーカーを優先して選抜した。これらのマーカーの PGDBj マーカー詳細ページへのリンクにより、PGDBj の DNA マーカー詳細情報へのより簡便なアクセスが可能となった。

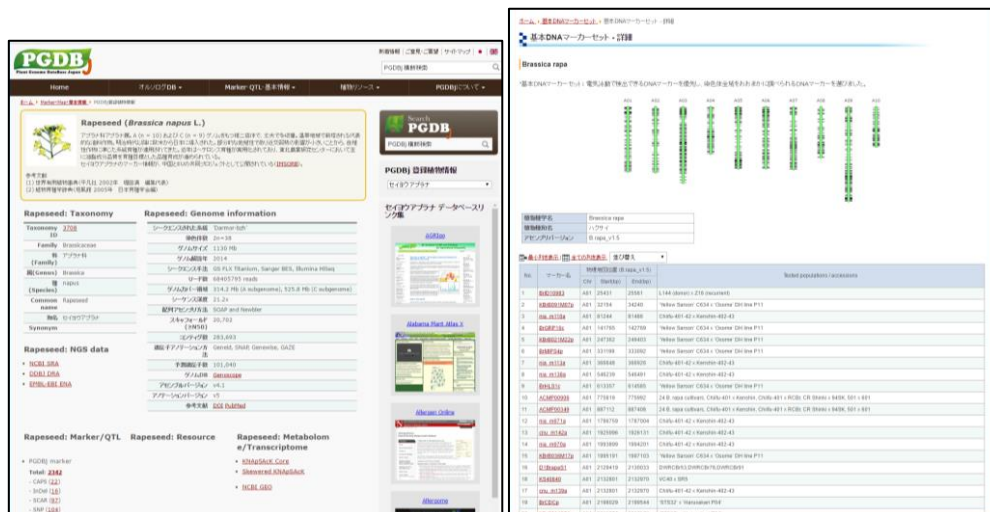


図 4 ゲノム情報を集約した植物種基本情報ページ例(左)と基本 DNA マーカーセット詳細ページ例(右)

形質キュレーションシステムの構築では、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。対象文献は 6,210 報であった。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 を使用した。313 のキーワードをもつ遺伝子(約 2,700)を 24 植物種から収集することに成功した。遺伝子の配列情報、形質キーワードの PO および TO カテゴリー分類により、DNA マーカー・QTL DB との統合を果たした。

データおよびリンクの更新、追加システムを構築、適用し、データベースの持続的な運営を容易にする基盤を整えた。

(4) DB 間の連携による統合化

第一期で構築された PGDBj の各 DB 間の連携を、セマンティックウェブ技術を用いて強化することを目的として、オントロジーの調査と RDF データ構造の検討、PGDBj の各 DB エントリーとそれに付随する情報の RDF 化を実施した。平成 29 年 1 月現在、RDF 化の対象とした情報数は、オルソログ DB の約 1001 万件、植物リソース DB の約 138 万件、DNA マーカー・QTL DB の約 149 万件である。また、RDF データを管理、運用するための情報基盤として Virtuoso(<https://virtuoso.openlinksw.com/>)を採用した。現在、SPARQL 検索ができるエンドポイント(<http://pgdbj.jp/sparql>)には、約 3424 万件の RDF データが格納されている。オルソログ DB エントリーについては、SPARQL エンドポイントを介して、立体構造 DB やモチーフ DB の情報のリンク付けが可能であることを確認した。

また、第一期で開発した PGDBj 横断検索システムを介した統合データ提供を目的として、横断検索システムの検索用インデックスの更新と機能拡張開発を実施した。特に、オルソログ DB の R66 版のデータ追加と共に、R57 版と R66 版から任意のリリースを選択できる機能やクラスタに含まれる配列数や生物種数に基づいた結果を絞り込みできる機能を追加した。さらに、項目(2)で構築した PGDBj オルソログ DB-植物リソース連携 DB に格納されたデータを横断検索システムに反映できるシステム開発と改変を実施した。この結果、理研 BRC や NBRP のエントリーについて、オルソログ DB エントリーとリンクした情報を外部 DB として追加した。これらの開発により、オルソログ DB やオルソログ DB と連携する DB の更新や追加に対応した、横断検索システムの持続的な運営を可能にする基盤を構築した。

(5) オントロジーの整備による横断検索の効率化

第一期、第二期を通じて構築・整備した PGDBj の内部 DB 間および外部 DB との間におけ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。調査では、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、その構造確認とともに、PGDBj コンテンツの RDF 化に不足している語彙 15 件を選定した。これらに、PGDBj 内部でのみ使用する仮版の位置

づけで ID を振り、内部コンテンツの RDF 化に適用した。

また、植物研究分野における実用的なオントロジー構築とその普及を目的として、オントロジー整備ワーキング委員会を開催した。植物関連学会が用語整備に取り組んでいるとの情報に基づき、そこで整備された情報が提供されることを前提として、学会による用語の整備作業を支援する情報基盤を構築した。基盤の構築には、MediaWiki (<https://www.mediawiki.org/wiki/MediaWiki>)を採用した。さらに、オントロジー整備に際し、植物研究コミュニティとの連携強化を目的として、オントロジーの活用例を示すことが課題として追加された。活用例として、藤井浩委員と共同で、オントロジーの整備により、作物の育種において系統情報や農業形質の情報を効率的に検索できるシステムを開発した。農研機構農業生物資源遺伝子バンクから公開されている植物遺伝資源の特性情報と農林水産省品種登録ホームページから公開されている品種登録情報の中から「交配情報」や「枝変わり情報」など由来に関する情報と、病害抵抗性などの「農業形質」の特性情報を収集し、オントロジー整備とデータの RDF 化を実施した。作物の中から情報の充実度の高さを期待して、全国で育成が推奨されている政令指定果樹を中心に作業を実施した。本システムにより、複数世代に遡った系統に関する農業形質の検索が可能となった。学会大会での発表や展示、共同研究者を通じた広報活動の結果、新しい共同研究の申し出も含め、好意的な意見を得た。

(6) ゲノム情報のアノテーションの高度化

第一期、第二期を通じて収集した PGDBj コンテンツの情報を物理地図上に集約し、植物ゲノム情報のアノテーションの高度化を試みた。情報基盤としては、ゲノムブラウザ JBrowse (<https://jbrowse.org/>)を採用した。本システムで情報を公開する植物種は、ゲノム情報が染色体に収束しており、かつ PGDBj において情報の充実度が高い 12 植物種とした。集約する情報に付随した塩基配列を問い合わせ配列として、データを集約させる生物種のゲノム配列(以下、対象ゲノム)に対して BLAST 検索を実施し、情報ごとの物理地図座標 71,554 件を算出した(平成 29 年 1 月現在)。これらの座標情報に基づき、収集した情報を対象ゲノム上にマッピングした。オルソログ DB エントリーについては、RefSeq データに記述されている座標情報を採用し、対象ゲノム上にマッピングした。また、それぞれの対象ゲノムに関して、精度の高い情報付加を目的として、生化学的機能および生物学的機能のアノテーションが充実しているシロイヌナズナの遺伝子モデルの情報を対象ゲノム上にマッピングした。まず、シロイヌナズナ等の遺伝子モデルを問い合わせ配列、対象ゲノムの遺伝子モデルを検索対象配列として BLAST 検索を実施した。この結果を用いて、シロイヌナズナ等の各遺伝子モデルを対象ゲノムの遺伝子モデルと関連付けし、最終的に対象ゲノムの遺伝子モデルの座標情報をシロイヌナズナ等の遺伝子モデルに関連付けした。その結果、本システムでは、対象ゲノム上にマッピングした遺伝子の情報や DNA マーカーあるいは QTL 情報を、異なる植物種の機能解析に由来する生化学的機能や生物学的機能の情報とともに表示あるいは検索が可能となった。平成 29 年 1 月現在、公開している植物種は、エゾヘビイチゴ、リンゴ、ハクサイである(<http://pgdbj.jp/jbrowse/index.html?data=Brapa> 等)。

(2) データベースの利便性に関する利用者ニーズと具体的な対応

利用者のニーズは、主にアドバイザー委員および植物関連学会大会での展示ブース来訪者から収集した。

- ・ **ニーズ 1:** 育種に有用な DNA マーカー情報のみを知りたい。

→ **対応 1:** 該当情報へ誘導する仕組みとして、育種向け DNA 検索ページを作成し、対応ツールを設置済み。育種現場で必要とされる情報をまとめたページの構築や、連鎖している形質情報等の分類を実施し、DNA マーカーの検索結果の表示を形質情報や検出方法などに基づいて切り替えられるようにした。また、形質の種類で表示させるマーカーの情報を設定できる。

- ・ **ニーズ 2:** オルソログ情報と植物リソース情報を組み合わせた検索をできるようにしてほしい。

→ **対応 2:** 対応機能を設置済み。オルソログ DB-植物リソース DB 連携 DB を構築することにより、実現した。

- ・ **ニーズ 3:** 特定形質に関わる遺伝子を効率的に推定するためのツールがほしい。

→ **対応 3:** 対応機能を設置済み。形質に関わる情報として QTL や形質キュレーションにより収

集した遺伝子情報を、機能アノテーションが充実しているシロイヌナズナの遺伝子モデルとともに検索、表示できるシステムを開発した。

(3) 持続的なデータベース運用体制の構築に向けた取り組み

リソース DB および横断検索システムの持続的な運営を可能にする基盤として、PGDBj オルソログ DB-植物リソース連携 DB を構築するためのシステムと、その連携 DB から指定した情報を出力させるシステムを構築した。DNA マーカー・QTL DB の持続的な運営を容易にする基盤として、DNA マーカーのキュレーションやデータ登録に際し、サブメントデータから機械的に処理して登録するシステムを構築した。RDF 化した PGDBj のコンテンツについて、Uniprot や KEGG から公開されている RDF データを利用した自動アサインに取り組み、データ収集の省力化を図った。植物関連オントロジーの構築に際し、植物関連学会による用語整備を支援するツールを構築し、その結果として整備されるデータの提供を受けることになった。整備される用語データが、本課題で開発している DB やポータルサイトへ容易に適用できることを目的とし、システムには MediaWiki を採用した。なお、オルソログ DB については、第一期の段階で、構築及び更新の処理は定型化され、クラスタリング処理を含めて独自開発した処理プログラムによって自動的に実行されるようになっている。

(4) 統合化推進プログラムの他のチームや DBCLS との連携

セマンティックウェブ技術に関連する作業については、SPARQLthon に定期的に参加し、情報の収集や共有に努めた。特に、RDF の設計、データの運用方法、関連ツールなどについて DBCLS や研究開発課題「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」(以下、微生物 DB) メンバーの千葉啓和氏(基礎生物学研究所)、藤沢貴智氏(遺伝学研究所)と協議、連携した。具体的には、項目(4)「データベース間の連携による統合化」において RDF 化したコンテンツのうち、オルソログ DB 関連情報について、千葉氏が作成したスクリプトの提供を受けて実施した。千葉氏は、微生物オルソログ DB コンテンツの RDF 化とデータの運用を担当し、RDF 構造について検討を重ねていたため、第二期の初期から連携している。この連携により、オルソログ情報を介して、微生物と植物の情報が繋がった。また RDF ポータルへの提供に際し、identifiers.org の情報とのリンクなど、他の研究開発課題とのリンクに有効な構造について、DBCLS の川島秀一氏、片山俊明氏、守屋勇樹氏も加わり協議、決定した。項目(6)「ゲノム情報のアノテーションの高度化」においては、微生物 DB のラン藻ゲノムを介した微生物と緑色植物の情報統合について、藤沢氏と協議した。

(5) データ産出を行う研究組織や研究室、プロジェクトとの連携

- ・ オントロジー整備ワーキング委員会での協議をきっかけに開始した、作物の育種における系統情報や農業形質の検索システム開発に際し、農研機構農業生物資源ジーンバンクと農林水産省品種登録ホームページのそれぞれから公開されている情報を使用している。用途と本開発により実現される機能について説明する場が設けられ、情報の使用に対する理解を得ることができた。
- ・ かずさ DNA 研究所が保有する作物の DNA マーカー DB (<http://marker.kazusa.or.jp/>) やゲノム DB (<http://strawberry-garden.kazusa.or.jp/>等)と連携し、エントリー情報やその構造情報の提供を受けている。公開する情報が補完的となるように連携している。
- ・ 農研機構野菜茶業研究所が保有する、作物の DNA マーカー DB である VegMarks (<http://www.naro.affrc.go.jp/project/results/laboratory/vegetea/2005/vegetea05-12.html>)と連携し、VegMarks のエントリー ID の提供を受けることが管理担当者との協議により決定した。
- ・ 理研 BRC および NBRP と連携し、SABRE に格納されている理研 BRC と NBRP が保有する 15 植物種 154 万件のリソースに関するリンク情報の提供を受けている。

(6) 人材の育成

キュレータに対し、ウェブブラウザを介した外部 DB ツールの使用方法やコマンドラインでの

BLAST による相同性検索の方法、結果の見方、UNIX 操作の基本から応用までを指導した。

(7) その他
該当なし

§4. 主要なデータベースの利活用状況

1. アクセス数

(1) 実績

名称	種別	平成 25(2013)年度	平成 26(2014)年度	平成 27(2015)年度	平成 28(2016)年度
PGDBj (全体, H28 年度は H28 年 12 月現在)	訪問者数	11,876	18,011	23,569	25,586
	訪問数	27,428	55,231	47,114	39,807
	ページ数	123,584	245,815	1,086,495	122,800

表 1-1 研究開発対象の主要なデータベースの利用状況(年度別)

名称	種別	平成 25(2013)年度	平成 26(2014)年度	平成 27(2015)年度	平成 28(2016)年度
PGDBj	訪問者数	990	1,501	1,964	2,843
	訪問数	2,286	4,603	3,926	4,423
	ページ数	10,299	20,485	90,541	13,644

表 1-2 研究開発対象の主要なデータベースの利用状況(月間平均)

(2) 分析

- PGDBj では、訪問者数が毎年度 30%以上ずつ増加した(平成 28 年度分は9ヶ月分で算出)。毎年度実施している植物関連学会大会での出展による広報活動が奏功している可能性がある。
- アクセスページ数については、平成 27 年度の値が極端に大きく、増加率は前年度比で+342%であった。この年度に閲覧されていたページの内容を確認すると、多くはオルソログ DB コンテンツであり、ページ総数は 791,119 ページ、全ページ数の中での割合は 73%を占めていた(表2)(増加率は前年度比で+556%)。DNA マーカーDB コンテンツについても、全ページ数の中での割合は 22%であったが、ページ総数は 237,711 ページと大きな値を示した(同+220%)。IP アドレスを調べた結果、海外の同じ地域の複数の IP アドレスから多数アクセスされていたことがわかった。平成 28 年度ではページ数の減少がみられた。これについては、前年度に多くアクセスしていたユーザの需要が満たされ、アクセスを終了した可能性もあるが、平成 27 年度の末から平成 28 年度の始めにかけて、ポータルサイトの運用システムにトラブルが発生し、機能が制限されたことが原因となった可能性もある。なお、運用システムは以前、Contents Management System の Joomla を採用していたが、トラブルの原因であることがわかったため、現在は Joomla の使用を中止し、Apache, MySQL, Perl による環境で管理できるように変更した。
- 4年間の閲覧ページ数の多いコンテンツを確認すると、オルソログ DB と DNA マーカーDB は常に 20%を超えていた(表2)。第一期の最終年度(平成 25 年度)の段階でも、アクセスの多いコンテンツであるとわかるが、「その他」の割合よりは小さい値を示していた。しかし、第二期に入ると「その他」の割合を上回るようになった。第二期は、PGDBj コンテンツの中でも、オルソログ DB と DNA マーカーDB について、インターフェースやデータの提示方法の改良に力を入れており、期待した通りの結果である。平成 29 年 1 月に入り、新たに公開したツールがあり、また3月に複数の植物関連学会での口頭発表と展示を計画しているため、今後さらなる増加が期待できる。

コンテンツ	平成 25 年度	平成 26 年度	平成 27 年度	平成 28 年度
トップページ	18,673 (15.1%)	18,771 (7.6%)	26,454 (2.4%)	24,431 (19.9%)

オルソログ DB	25,516 (20.6%)	118,917 (48.4%)	791,119 (72.8%)	25,291 (20.6%)
DNA マーカー-D B	36,416 (29.5%)	74,344 (30.2%)	237,711 (21.9%)	56,853 (46.3%)
その他	42,979 (34.7%)	33,783 (13.7%)	31,211 (2.9%)	16,225 (13.2%)
全体	123,584	245,815	1,086,495	122,800

表 2-1 PGDBj のコンテンツ毎のページ数(括弧内は全ページの中での割合) (年度別)

コンテンツ	平成 25 年度	平成 26 年度	平成 27 年度	平成 28 年度
トップページ	1,556 (15.1%)	1,564 (7.6%)	2,205 (2.4%)	2,715 (19.9%)
オルソログ DB	2,126 (20.6%)	9,910 (48.4%)	65,927 (72.8%)	2,810 (20.6%)
DNA マーカー-D B	3,085 (29.5%)	6,195 (30.2%)	19,809 (21.9%)	6,317 (46.3%)
その他	3,582 (34.7%)	2,815 (13.7%)	2,601 (2.9%)	1,803 (13.2%)
全体	10,299	20,485	90,541	13,644

表 2-2 PGDBj のコンテンツ毎のページ数(括弧内は全ページの中での割合) (月間平均)

2. データベースを利用して得られた研究成果事例

1. Takeshi Ara, et al., "Metabolonote: a wiki-based database for managing hierarchical metadata of metabolome analyses", *Fron. Bioeng. Biotechnol.*, 2015 (DOI: 10.3389/fbioe.2015.00038)
2. Kang, Y. Jae, et al., "Translational genomics for plant breeding with the genome sequence explosion", *Plant Biotechnology Journal*, vol. 14, pp.1075-1069, 2016 (DOI:10.1111/pbi. 12449).
3. Adeel Malik, "Genomics Resource for Plants", *Plant Omics: Trends and Applications*, pp 29 - 57, 2016 (DOI:10.1007/978-3-319-31703-8_2)
4. Martinez, Manuel. "Computational Tools for Genomic Studies in Plants." *Current Genomics* 17, 509-514, 2016 (DOI:10.2174/1389202917666160520103447)

3. その他

該当なし

§5. 研究開発期間中に得られた科学・技術や産業に対する波及効果

本研究開発では、第一期、第二期を通じて主に3つの DB(オルソログ DB、植物リソース DB、DNA マーカー・QTL DB)を構築、拡充した。

現在、植物オルソログ情報を扱う代表的な国際 DBとして、PLAZA (<http://bioinformatics.psb.ugent.be/plaza/>) や PlantOrDB (<http://bioinfolab.miamioh.edu/plantordb/interface/readme.php>) が挙げられる。これらの DB では、それぞれ 47 種、41 種の植物種を対象としている。これに対し、PGDBj のオルソログ DB では、構成配列として NCBI の RefSeqDB の情報を用いることによって、40 種の緑色植物と 213 種のラン藻を含み、さらに本 DB を HUB として 18 種の緑色植物のゲノム DB や完全長 cDNA DB を接続することによって、合計 58 種の緑色植物および 213 種のラン藻の配列情報を取得することができる。また、クラスタリングした遺伝子群に対して NCBI のタクソミー分類情報を活用しており、門・綱・目・科・属・種などの階層の単位で情報を表示、取

得できるようになっている。オルソログを始めとして、分子進化情報は、遺伝子やタンパク質の機能や立体構造に関して重要な示唆を与えるものであり、基礎研究の要の一つである。本 DB が提供する情報を利用することにより、ユーザは既にまとめられた進化情報を、任意の階層範囲を選んで取得することができる。こうした機能は、解析を省力化・効率化し、基礎研究分野への発展に貢献すると期待できる。

植物リソースのデータベースについては、国内では理研 BRC (<http://ja.brc.riken.jp/>) の SABRE (<http://sabre.epd.brc.riken.jp/SABRE2.html>) や NBRP の情報公開サイト (<http://www.nbrp.jp/>)、農研機構の農業生物資源ジーンバンク (<https://www.gene.affrc.go.jp>) が存在する。SABRE は理研 BRC と NBRP (<http://www.nbrp.jp/>) が保有する 15 種の植物種のリソース情報を、農業生物資源ジーンバンクは農研機構が保有するリソース情報を登録したデータベースである。PGDBj では、SABRE と連携して検索 API を拡張開発し、また、データベース化されていなかった近畿大学と果樹研究所が保有するカンキツ類のリソース情報を整備して検索対象に追加した。この結果、SABRE に登録されているリソースと共に、カンキツ類のリソースを横断的に検索できるようにした。カンキツ類は、国内で育成・生産が推奨されている作物の一つであり、リソース情報検索を可能にしたことにより、国内の育種研究機関や種苗会社が求める情報提供に貢献している。

DNA マーカーに関する代表的な外部の DB としては、Gramene (<http://www.gramene.org>)、SOL Genomics Network (SGN; <https://solgenomics.net>)、GDR (<https://www.rosaceae.org/>) があり、それぞれイネ科、ナス科、バラ科などの各植物科に特化した DNA マーカーや QTL の情報がまとめられている。PGDBj の DNA マーカー・QTL DB では、DNA マーカーについては 27 科 65 種、QTL については 20 科 45 種を対象とし、多型や形質に関わる情報を文献からマニュアルキュレーションした。このように幅広い科や種について DNA マーカーを網羅した DB は他になく、広い科を跨って検索ができる点は本 DB の大きな特徴となっている。また、収集された情報に対し、多型の検出法や連鎖形質の種類等に基づいたフィルターを設置したことにより、ユーザのニーズに応じた情報の抽出が可能である。特に、電気泳動などの簡便な方法で検出できるマーカーに絞った情報抽出機能は、設備が充実していない中小種苗会社や大学の研究室から、評価する意見が多い。

本研究開発において情報整備の対象とした植物種の約 68% が実用植物であり、想定した活用事例は基礎研究から育種研究まで広範に及ぶ。現時点において、ユーザから好意的な意見が多いが、特に育種研究は成果が出るまでに時間がかかることもあり、本 DB を利用しての成果が出てくるには、もう少し時間が必要であると予想される。平成 29 年 1 月現在、PGDBj は OMICtools (<https://omictools.com>) に登録されている。OMICtools はバイオインフォマティクスのソフトウェアやデータベースなどの生物学的データの解析ツールを検索できるポータルサイトとして広く知られており、この中に登録されているということは、第三者からも有用性が認められていることを意味する。Global Engage のウェブページ (<http://www.global-engage.com/agricultural-biotechnology/best-plant-genomereadatabase/>) においては、PGDBj DNA マーカー・QTL DB が、有用な植物ゲノム関連 DB の中のトップ 10 の一つとして、紹介されている。Global Engage は学術界から産業界まで、薬学、農学、工学分野など幅広い分野の国際会議や展示会を運営し、各分野での有用情報の共有と発信を目指している企業である。このように PGDBj は海外においても徐々に認知されるようになってきていることから、今後、海外における活用事例の広がりも期待される。

§6. 今後の展開

本研究課題の第一期、第二期が進行する中で、整備対象とするデータについて、量・質ともに大きく変化した。たとえば、PGDBj オルソログ DB は NCBI RefSeq DB の配列エントリーを用いて構築されているが、第一期で採用されたリリース 57 に含まれる生物種と配列数は、緑色植物 20 種の約 50 万配列と、ラン藻 111 種の約 50 万配列であったのに対し、第二期で採用されたリリース 66 では、緑色植物 40 種の約 114 万配列と、ラン藻 213 種の約 80 万配列に増大した。今後も、リリースの更新に伴う対象植物種と対象配列の増加が見込まれ、計算機資源の強化を含めた対応策を検討する必要がある。また現在の PGDBj オルソログ DB は、外部の植物関連データベースのエントリー同士を結びつける HUB データベースとしての役割を担っている。現在は、NCBI の RefSeq DB の ID との対応 ID を保有しない外部 DB と、保有する外部 DB が混在し、前者については、BLAST による関連付けを実施した。一方、後者については、セマンティックウェブ技術の RDF による記述によって、関連付けされているものが存在した。第二期では、PGDBj コンテンツの RDF 化が主要な課題であり、そのための基盤整備を実施済みであったことから、こうしたデータに対しては、RDF 構造を利用した自動アサインに取り組んだ。今後、この機能を拡張することによって、BLAST によらない自動的な関連付けが可能であるかもしれない。

植物リソース DB に関しては、SABRE に登録された 15 の植物種とカンキツ類のリソースを検索できるように開発を実施したが、今後も、ユーザの視点から新たなリソース情報の登録を実施したい。例えば、近年、農業者の高齢化に伴い、生産される作物に変化が生じているとのことであり(例:ハクサイ農家→ブロッコリー農家)、種苗業界における育種対象も変化することが予想される。こうしたニーズを収集するための仕組みも作っていきたい。

DNA マーカー・QTL DB の構築においては文献からのキュレーションを実施したが、文献の報告年の進行に伴い、記述されていた解析手法と生成されるデータが大きく変化した。第一期の初期は、1 論文あたり、個別に開発・解析された 10 件未満程度の DNA マーカーのデータをキュレーションし、マニュアルでデータストアへ登録する作業が中心に実施されていた。これに対し、NGS を用いた大規模解析手法が主流となって以降、特に第二期の途中からは、数百から千件程度の DNA マーカーに関する情報をサブリメントデータから機械的に処理して登録する作業が多く見られるようになった。現在、完全な自動化は達成していないが、今後のデータ増加に耐えうる基盤は構築できていると考えられる。

また、データ数量のみならず、主流のマーカーの種類も、第一期では SSR マーカーであったものが、第二期では SNP マーカーに変化した。さらに、QTL に関連する情報も、実験交配に由来する分離集団を用いた従来の QTL 解析から、多様な品種・系統などを直接用いた Genome Wide Association Study (GWAS) が増えた。こうした変化に対し、ユーザが求めるツールの変化を想定し、情報のまとめ方や設置するフィルター、インターフェースの変更や追加を実施した。特に、PGDBj で収集、整備した DNA マーカーや QTL などのコンテンツを物理地図上に集約させるにあたり、ゲノムブラウザ JBrowse を使ったシステムを導入した。その結果、異なる性質や精度を示す複数の DNA マーカーや QTL 情報が一元化され、互いにゲノム上の近傍にある場合にも、同時に確認することが可能となった。PGDBj のコンテンツには、生化学的、生物学的機能アノテーションが充実したシロイヌナズナの遺伝子モデル情報も含まれ、QTL 情報と共に表示することにより、該当する QTL の形質の原因遺伝子の推定に有用であり、後の解析に必要な選抜マーカーの検索が可能である。今後 NGS データを取り扱うユーザの増加が見込まれることから、現在は PGDBj に実装済みのデータのみを表示する機能を提供しているが、将来的にはユーザが入力した NGS データから多型を検出し、その結果を PGDBj コンテンツと共に提示できるような新たな解析基盤を提供しうる。この仕組みによって、遺伝子型と形質との相関関係が一層明確となり、その結果として遺伝子単離や育種プロセスの加速への貢献が期待できる。

§7. 自己評価

国内外で公開されている多種多様な植物ゲノム情報を統合的に利用するための利便性が高い仕組みを構築することを目標として、今期は第一期で構築したポータルサイト PGDBjの機能を拡充するため本報告書に記載の 6 項目について研究開発を実施した。オルソログ DB をハブとすることによって種を超えた植物遺伝子の包括的な取り扱いが可能になったこと、文献に留まっていた多数の DNA マーカーが DB 化されゲノム物理地図と対応づけられたことなど、多様性を特徴とする植物ゲノム研究において新たな解析基盤を提供することに成功した。PGDBjへのアクセス数が増加を続けていることも、その有用性を示している。以上の理由から、本開発は本プログラムの趣旨に十分に合致していると判断できる。本研究開発は、マニュアルキュレーションによる文献情報の収集、データベース開発、RDF 化にともなうデータ処理等幅広い作業内容を含んでおり、それぞれの進捗度を厳密に評価することは困難であるが、全体としては開始当初の開発目標を十分に達成したと考えている。

一方で、改良すべき点も残されている。まず、ユーザ拡大の観点から、DB の利用に慣れていないユーザに向けて操作がより容易な UI の開発と丁寧な使用マニュアルの整備が必要であると考えられる。機能面では、近年加速度的に増加する新規植物や変異体の全ゲノム構造情報を迅速にデータベースに取り込んで表示させるシステムの実装が望ましい。また、大規模 SNP マーカーの開発を目的として、ユーザ自身が次世代型シーケンサーを用いて取得したショートリード配列データを PGDBj 内の参照配列にマップして多型を検出するツールの開発と実装も時宜を得たものであろう。

§8. 外部発表等

1. 原著論文発表

(1) 論文数概要

種別	国内外	件数
発行済論文	国内 (和文)	0 件
	国際 (欧文)	0 件
未発行論文 (accepted, in press 等)	国内 (和文)	0 件
	国際 (欧文)	0 件

(2) 論文詳細情報

該当なし

2. その他の著作物（総説、書籍など）

1. Nakaya A, Ichihara H, Asamizu E, Shirasawa S, Nakamura Y, Tabata S, and Hirakawa H: Plant Genome DataBase Japan (PGDBj). Methods Mol Biol., 15 33:45-77, 2017 (DOI:10.1007/978-1-4939-6658-5_3)
2. 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 32:74-75、2014 (ISBN 978-4-7581-0343-5)

3. 国際学会発表及び主要な国内学会発表

(1) 概要

種別	国内外	件数
招待講演	国内	1 件
	国際	0 件
口頭発表	国内	1 件
	国際	0 件
ポスター発表	国内	9 件
	国際	3 件

(2) 招待講演

〈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、園芸学会平成 28 年度春季大会小集会、厚木、2016 年 3 月 25 日

〈国際〉

該当なし

(3) 口頭講演

〈国内〉

1. 柴谷(駒井)多恵子、PGDBj より新たに公開した「育種向け DNA マーカー検索ページ」について、園芸学会平成 28 年度秋季大会、名古屋、2016 年 9 月 11 日

〈国際〉

該当なし

(4) ポスター発表

〈国内〉

1. 白澤沙知子、PGDBj～植物 80 種のゲノム情報の集約～、日本育種学会大128回講演会、新潟、2015年9月12日
2. 市原寿子、植物ゲノム情報へのセマンティックウェブ技術の適用、日本育種学会大128回講演会、新潟、2015年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、園芸学会平成27年度秋季大会、徳島、2015年9月27日
4. 中谷明弘、遺伝子オルソログ DB による植物ゲノム DB の統合、JST バイオサイエンスデータベースセンター「トーゴーの日シンポジウム 2016」、東京、2015年10月5日
5. 柴谷(駒井)多恵子、植物ゲノム情報活用のための統合研究基盤の構築、JST バイオサイエンスデータベースセンター「トーゴーの日シンポジウム 2016」、東京、2015年10月6日
6. 市原寿子、作物品種情報に対するセマンティックウェブ技術の適用、JST バイオサイエンスデータベースセンター「トーゴーの日シンポジウム 2016」、東京、2015年10月6日
7. 柴谷(駒井)多恵子、PGDBj における文献より収集した DNA マーカー・QTL 情報の公開および育種向け DNA マーカー検索、平成28年度果樹バイオテク研究会、宇都宮、2016年10月19日
8. 市原寿子、セマンティックウェブ技術を用いた果樹形質検索システムの構築、平成28年度果樹バイオテク研究会、宇都宮、2016年10月19日
9. 市原寿子、作物品種の形質情報に対するセマンティックウェブ技術の適用、第39回日本分子生物学会年会、横浜、2016年12月2日

〈国際〉

1. Hisako Ichihara, Plant Genome Database Japan (PGDBj) in 2015: A Comprehensive Database Covering Information of Plant Genome-Related Database in Japan, Plant & Animal Genome Conference XXIII, San Diego, CA, 2015年1月12日
2. Hisako Ichihara, Plant Genome Database Japan (PGDBj) in 2016: A Comprehensive Database Covering Information of Plant Genome-Related Database in Japan, Plant & Animal Genome Conference XXIV, San Diego, CA, 2016年1月11日
3. Akihiro Nakaya, Plant Genome Database Japan (PGDBj) in 2017: a research infrastructure for genomics and genetics of plants, Plant & Animal Genome Conference XXV, San Diego, CA, 2017年1月16日

4. 知財出願

該当なし

5. 受賞・報道等

(1) 受賞

該当なし

(2) メディア報道

該当なし

(3) その他

植物関連学会大会での出展による広報活動

(国内)

1. 第32回 日本植物細胞分子生物学会(盛岡), 2014/08/21-22

2. 日本育種学会 平成 26 年度秋季大会(東京), 2014/09/11-14
3. 日本育種学会 平成 26 年度秋季大会(宮崎), 2014/09/25-28
4. 日本園芸学会 平成 26 年度秋季大会(佐賀), 2014/09/27-29
5. 第 37 回 日本分子生物学会年会(横浜), 2014/11/25-27
6. 第 56 回 日本植物生理学会年会(東京), 2015/03/16-18
7. 日本植物病理学会大会創立 100 周年記念大会(東京), 2015/03/29-31
8. 第 33 回 日本植物細胞分子生物学会(東京), 2015/08/10-12
9. 日本植物学会 第 79 回大会(新潟), 2015/09/06-08
10. 第 128 回 日本育種学会 平成 27 年秋季大会(新潟), 2015/09/11-12
11. 日本園芸学会 平成 27 年秋季大会(徳島), 2015/09/26-27
12. BioJapan 2015 World Business Forum(横浜), 2015/10/14-16
13. 第 38 回 日本分子生物学会年会・第 88 回 日本生化学会大会(神戸), 2015/12/01-04
14. 第 57 回 日本植物生理学会年会(盛岡), 2016/03/18-20
15. 第 129 回 日本育種学会 平成 28 年春季大会(横浜), 2016/03/21-22
16. 日本園芸学会 平成 28 年春季大会(東京), 2016/03/26-27
17. 第 34 回 日本植物細胞分子生物学会(上田), 2016/09/1-3
18. 第 130 回 日本育種学会 平成 28 年秋季大会(鳥取), 2016/09/24-25
19. 第 39 回 日本分子生物学会年会(横浜), 2016/11/30-12/2
20. 第 58 回 日本植物生理学会年会(鹿児島), 2016/03/16-18
21. 日本園芸学会 平成 29 年春季大会(藤沢), 2016/03/19-20
22. 第 131 回 日本育種学会 平成 29 年春季大会(名古屋), 2016/03/29-30

(国際)

1. PAG XXIII (San Diego, CA), Booth “Plant Genomics Outreach Consortium”, 2015/01/10-14
2. PAG XXIV (San Diego, CA), Booth “Agricultural Biological Database Outreach Consortium”, 2016/01/09-13
3. PAG XXV (San Diego, CA), Booth “Agricultural Biological Database Outreach Consortium”, 2017/01/14-18

§9. 研究開発期間中の活動

1. 進捗ミーティング

年月日	名称	場所	参加人数	目的・概要
平成 26 年 4 月 7 日より 毎週月曜日	植物統合 DB ラボミーティ ング(第 1 回～第 128 回) (非公開)	かずさ DN A 研究所	10 人	研究進捗報告のためのミ ーティング
平成 26 年 6 月 30 日(月)	H26 年度第 1 回植物統合 DB 内部検討会(非公開)	かずさ DN A 研究所	4 人	かずさ研究所担当分の 開発項目について打ち 合わせ
平成 27 年 2 月 18 日(水)	H26 年度第 2 回植物統合 内部検討会(非公開)	大阪大学	7 人	H26 年度の進捗報告
平成 27 年 3 月 30 日(月)	作物関連オントロジー整備 会議(非公開)	明治大学	3 人	作物関連オントロジー整 備の為の打ち合わせ
平成 27 年 6 月 26 日(金)	H27 年度内部検討会 (非公開)	龍谷大学	8 人	進捗報告と H27 年度の 実施内容確認
平成 28 年 8 月 5 日(金)	H28 年度内部検討会 (非公開)	ステーショ ンコンファレ ンス東京	6 人	進捗報告と、最終年度と して目指すゴールの確認

2. 主催したワークショップ、シンポジウム、アウトリーチ活動等 該当なし

以上

別紙1 研究開発対象のデータベース等

No.	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系データベースアーカイブ	NBDCヒトデータベース	NBDC RDFポータル	関連文献 (論文リストに記載があれば、その番号でも可)
1	Plant Genome DataBase Japan (PGDBj)	Plant Genome DataBase Japan (植物ゲノム統合データベース)	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	継続・発展	データベース等	公開済	対象外		Erika Asamizu et al., "Plant Genome DataBase Japan (PGDBj): a portal website for the integration of plant genome-related databases.", <i>Plant Cell Physiol</i> , vol. 55, pp. e8, 2014 (DOI: 10.1093/pcp/pct189) Akihiro Nakaya, Hisako Ichihara, Erika Asamizu, Sachiko Shirasawa, Yasukazu Nakamura, Satoshi Tabata and Hideki Hirakawa, "Plant Genome DataBase Japan (PGDBj)", <i>Methods Mol Biol</i> . vol. 1533 (<i>Plant Genomics Databases</i>), pp. 45-77, 2017 (DOI: 10.1007/978-1-4939-6658-5_3)
2	PGDBj Ortholog Database	PGDBj Ortholog DB	PGDBj オルソログDBの基本の検索ページです。オルソログDBのサブデータベース(アミノ酸配列DB、ホモロジーDB、オルソログ間の階層)に対してウェブブラウザ経由で検索ができます。	http://pgdbj.jp/od2/	2015/2/28	新規	データベース等	公開済	対象外	調整中	Akihiro Nakaya, Hisako Ichihara, Erika Asamizu, Sachiko Shirasawa, Yasukazu Nakamura, Satoshi Tabata and Hideki Hirakawa, "Plant Genome DataBase Japan (PGDBj)", <i>Methods Mol Biol</i> . vol. 1533 (<i>Plant Genomics Databases</i>), pp. 45-77, 2017 (DOI: 10.1007/978-1-4939-6658-5_3)