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
研究開発課題
「ゲノム情報に基づく植物データベースの統合」

研究開発終了報告書

研究開発期間： 平成23年4月～平成26年3月

研究代表者：田畑哲之
（（公財）かずさDNA研究所、所長）

§1 研究開発実施の概要

(1)実施概要

植物ゲノムデータベース(以下DB)の統合に向けて、オルソログ遺伝子群の情報をアミノ酸配列の類似性に基づいて網羅的に整理したオルソログデータベース(OD)を構築した。次に、RAP-DB(イネ)を始めとする統合対象の植物ゲノムDB内の遺伝子ID や配列で検索できるDB項目(ウェブページ)をBLASTによる配列類似関係に基づいてこのODに対応づけ、同一のオルソログに対応づけられたDB項目を集計して表示することで、ODを経由したDB項目の相互リンクと横断的検索を実現した。上記 OD の構築に際しては、まず、アミノ酸配列のプールとしてNCBI RefSeq Databaseの緑色植物20生物種の約50万配列を取得し、これらの全配列間のBLASTによる類似情報を格納したDBを構築した。続いて、同DB内の類似情報に基づいて、生成した類似アミノ酸配列のクラスタをオルソログ情報としてDB化した。また、クラスタ生成の過程で使用した種間の系統関係やアミノ酸配列間の類似関係の階層構造もDB化して、その階層構造を上下することによって検索範囲の拡大と縮小を実現した。全情報は RDB(関係データベース)で管理し、SQL による柔軟な検索が可能な環境をWEB/DBサーバ上に構築した。

植物リソース情報DBの統合は、国内の植物バイオリソース情報を総合的に検索可能にし、我が国の植物研究をさらに推進させるための基盤として機能させることを目標としている。現在までに、理研BRCが開発したSABREシステム本体ならびにAPIの拡充を通じて、ゲノムに関連した情報をもつ文部科学省 NBRP(理研 BRC のもつ情報を含む)14 植物種・約150 万件のバイオリソース関連情報を、本研究開発において構築したポータルサイト Plant Genome DataBase Japan, PGDBjの横断検索システムから統合検索可能とした。また、新規な有用バイオリソース情報として、近畿大学、果樹研究所からカンキツ類のリソース情報を収集した。

近年DNAマーカーが大規模開発されている状況をうけ、本事業では国内外でゲノム解析やマーカー開発が行なわれている24科55種の植物を対象に、DNAマーカーの塩基配列情報をタグとして用いることによってマーカー情報の統合を進めてきた。これまでに、国内でDNAマーカー関連DBが公開されている10種について75,975件の情報を収集、公開した。うち7種(ラッカセイ、オランダイチゴ、ミヤコグサ、ダイコン、トマト、アカクローバ、シロクローバ)については、連鎖地図を閲覧することができる。さらに、20 種については文献からのキュレーションを完了し15,259件のマーカーおよび1,767件のQTL情報を収集、公開した。現在、異なる植物種間で配列情報に基づくリンク付けと地図表示を進めている。また、対象 55 植物種の基本情報やゲノム解析手法の整理、関連DBリンクの整備を行い、ユーザが必要な情報に迅速にたどり着けるよう利便性の向上に努めた。

上記の開発で構築された DB を有機的に結びつけ、ユーザの利便性を高めるため、ポータルサイトPlant Genome DataBase Japan, PGDBj(http://pgdbj.jp)を構築し、平成24年8月に公開を開始した。このサイトは、主に緑色植物20種に由来するアミノ酸配列間の関係を辿ることができる「オルソログテーブル」、キーワードやアミノ酸配列による検索機能、OD関連データダウンロードサービス、対象 55 植物種の基本情報、DNA マーカーおよび QTL の検索機能や地図表示、国内外の植物関連 DB リンク集を提供している。また、「PGDBj 横断検索」により、本事業で収集した全ての情報、リソース情報、オルソログ DB をハブとして関連づけた外部ゲノムDBの情報を横断的に検索できる。

(2)研究開発成果のデータベース等

- PGDBj Ortholog Database (OD)
- マーカーリスト (PGDBj Marker list)
- QTLリスト (PGDBj QTL list)
- 登録生物種リスト (Registered plant list)
- 植物データベースリンク (Plant DB link list)
- ゲノム解析手法 (Genome analysis methods)

- PGDBj カンキツリソースデータベース（仮）（年度内公開予定）

§2. 研究開発構想（および構想計画に対する達成状況）

（1）当初の研究開発構想

我が国の植物分子遺伝学、植物ゲノム研究の成果であるゲノム構造・機能情報やリソース情報は、論文および DB 上で公開される。しかし、研究対象がモデル材料から農作物まで多種多様であること、提供される情報が塩基配列、転写、翻訳、代謝、形質など多岐にわたること、研究グループやプロジェクトごとに個別の DB が構築されることから、異なるプラットフォームをもつ多くの DB が散在する状況にある。さらに、DNA マーカーや連鎖地図などの情報は文献上に留まっているものも多く、これらを総合的に検索して必要とする情報を入手することは容易ではない。そこで、我々はこれらの情報についての利便性を高めるため、(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、の4種類の研究開発を一体として進めることにした。

（2）新たに追加・修正など変更した研究開発構想

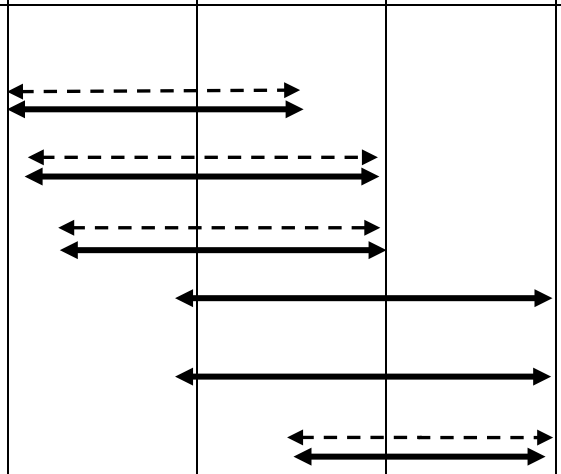
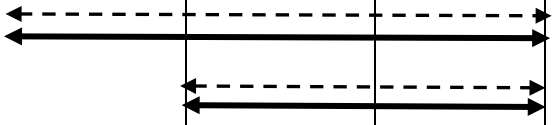
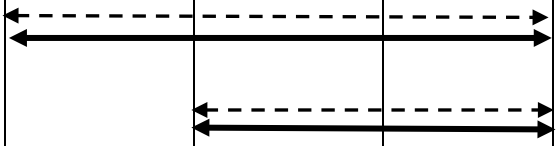
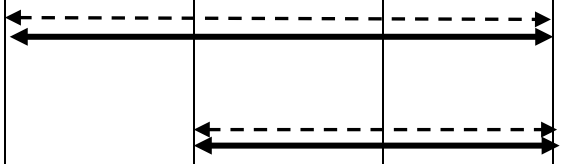
（新たな研究開発計画）

ポータルサイト PGDBj において、オルソログ DB やリソース DB、マーカーDB といった内部 DB と理研 BRC の SABRE Database に対する横断検索システムを構築した結果、オルソログ DB における遺伝子機能情報（遺伝子名や遺伝子産物名）、リソース DB における遺伝子クローン情報（遺伝子機能情報）、マーカーDB における遺伝子の機能情報や形質などの用語が統一されていなかったため、正規表現などを取り入れた検索機能の強化やオントロジーの整備による横断検索の効率化を図る新たな研究開発計画が提案された。また、オントロジーの整備と共に、統合化推進プログラムにおいて導入が推奨されている Resource Description Framework（RDF）の概念を本研究開発に取り入れることにより、PGDBj の各種 DB のコンテンツの記述形式を統一化し DB 間の横断検索を強化することも新たな研究開発計画として提案された。

（修正点）

DNA マーカーおよび連鎖地図情報に基づく植物ゲノム DB の統合では、55植物種について文献からキュレーションを実施する予定であったが、イネやコムギ、トウモロコシといった主要作物については、Gramene（http://www.gramene.org）や WGGRC Wheat Genome Database（http://wggrc.plantpath.ksu.edu）、MaizeGDB（http://www.maizegdb.org）といった広く利用されているデータベースが存在するため、リンクに留めるべきと判断し、これら以外の作物についてのキュレーションを重点的に行った。また、本研究開発を開始した当初、日本語のコンテンツを作成していたが、複数の植物種を対象としたオルソログやリソース、マーカーといった幅広い内容での統合化 DB が国内外で見当たらなかったため、英語版のコンテンツを作成し海外での学会発表を通じて PGDBj の広報活動を行った。本研究開発では、4つの研究項目の内容を充実させそれらに対して横断検索機能を付与したポータルサイトを構築することを目標としたが、さらに、他研究課題との連携を深めるため、「メタボローム・データベースの開発」で開発された DB 間で相互リンクを張った。これに加え、外部サイトから PGDBj のコンテンツに対して検索できる API を開発することで、学会や協力機関のホームページを通じて PGDBj の利用者の拡大を図ることができる体制を整えた。

（3）達成状況

研究開発項目	H23年度	H24年度	H25年度	変更点
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 ・遺伝子名とアミノ酸配列のDB化（かずさグループ） ・配列間ホモロジーDBの開発（新潟大グループ） ・オルソログDBの開発（新潟大グループ） ・ゲノムDBへのリンク（新潟大グループ） ・統合的検索API開発（新潟大グループ） ・統合的検索GUI開発（新潟大グループ）				予定通り 予定通り 予定通り 追加 追加 予定通り
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、QTL関連のDBおよび文献の調査（かずさグループ） ・マーカー付随配列の連鎖地図上へのマッピングソフトの研究開発、DB管理システムの構築（かずさグループ）				予定通り 予定通り
3. 植物リソース情報DBの統合 ・リソース情報の統合DB開発とデータ集積（かずさグループ） ・リソース情報DBの共用化と永続化のための環境整備（かずさグループ）				予定通り 予定通り
4. 植物研究に関連する情報基盤の構築 ・植物のオミックス研究関連DBの調査、リンク情報の整理（かずさグループ） ・ポータルサイト、DB管理システムの構築（かずさグループ）				予定通り 予定通り

(4)研究開発の今後の展開について

本研究開発では、オルソログ DB、リソース DB、マーカーDB を基にした植物 DB の統合化を行ったが、今後はこれらの DB のコンテンツを自動的に更新・追加し、さらには、植物種の追加を容易に行うことで、PGDBj を永続的に運営するための技術開発が必要である。また、PGDBj の内部 DB の充実に加え、理研 BRC や農業生物資源研究所などで公開されている外部 DB や現統合化推進プログラムの他課題の成果や次期採択課題との連携を図ることで情報を共有し統合化をさらに進める必要がある。そのためには、オントロジーを整備し、データの記述を RDF といった形式で統一化し、それらを学会や論文発表を通じて研究者コミュニティに普及させることも重要な課題である。

今後、次世代シークエンサーの技術革新や計算機技術の向上がさらに進むことで、多種多様な植物種や品種、系統に関するゲノム構造情報や遺伝子発現関連情報が得られ、それにより、多数の DNA マーカーや連鎖地図情報が蓄積され、それらの情報から多様性の原因となる遺伝的背景が明らかになることが予想される。これらのデータに対して横断検索が可能な PGDBj を研究者コミュニティに提供することができれば、基礎研究や育種といった産業面への波及効果も得ることができる。こうした研究開発により植物ゲノム情報を統合化することは、新しい知識の獲得や革新的な研究の開発への貢献に繋がると考えている。

§3 研究開発実施体制

(1)研究チームの体制について

㊀「研究代表者:田畑哲之」グループ

研究参加者

氏名	所属	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	所長	統括	H23.4～H26.3
平川 英樹	同上	主任研究員	項目2、4統括	H23.4～H26.3
中村 保一	同上	特別客員研究員	項目3統括	H23.4～H26.3
＊浅水恵理香	同上	プロジェクト研究員	キュレーション統括	H25.4～H26.3
＊市原 寿子	同上	プロジェクト研究員	システム開発統括	H24.4～H26.3
＊眞板 寛子	同上	特任技術員	項目2、4キュレーション	H23.4～H25.3
小原 光代	同上	技術員	項目2、4キュレーション	H23.4～H26.3
山田 学	同上	技術員	項目2、4キュレーション	H23.4～H24.3
藤代 維一	同上	技術員	項目2、4キュレーション	H24.4～H26.3
＊実形 ゆりや	同上	プロジェクト技術員	項目2、4キュレーション	H23.4～H26.3
＊石井 崇洋	同上	プロジェクト技術員	項目2、4キュレーション	H23.4～H26.3
＊阿久津智子	同上	プロジェクト補助員	項目2、4キュレーション	H23.4～H26.3
＊佐藤 かおり	同上	プロジェクト補助員	項目2、4キュレーション	H23.4～H26.3
＊小池浩太郎	同上	プロジェクト補助員	項目2、4キュレーション	H23.4～H26.3

＊江頭　博子	同上	プロジェクト補助員	項目2、4キュレーション	H23.4～H24.3

㊀「研究分担者：中谷明弘」グループ
研究参加者

氏名	所属	役職	研究開発項目	参加時期
○中谷　明弘	新潟大学研究推進機構	准教授	サブ課題1 システム開発	H23.4～H26.3

(2)国内外の研究者や産業界等との連携によるネットワーク形成の状況について

本プロジェクトは、日本植物学会、日本植物生理学会、日本植物細胞分子生物学会の会長の支持のもとに実施されており、年会において展示、発表等によって会員への情報提供が行われた。さらに、年2回開催のアドバイザリー委員会においては、外部委員として上記学会幹部、国内植物関連DB管理者を招聘し、情報収集、情報提供を行った。また、本プログラムの「メタボローム・データベースの開発」の研究代表者である金谷重彦博士と連携し、収集した遺伝子情報と代謝産物情報をリンクした。

§4 研究実施内容及び成果

4．1 研究課題名：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、植物リソース情報 DB の統合、および植物研究に関連する情報基盤の構築（かずさ DNA 研 田畑哲之グループ）

研究開発実施内容及び成果

国内外で DNA マーカーが大規模に開発されている状況を受け、本課題では代表的な 24 科 55 種の植物を対象に、DNA マーカーの塩基配列情報をタグとして用いることによってマーカー情報の統合を進めた。まず国内で DNA マーカー関連 DB が公開されている 10 種について計 75,975 件の情報を収集、公開した。更に文献からのキュレーションを進め、20 種について計 15,259 件のマーカーおよび 1,767 件の QTL 情報を収集、公開した（表 1)。

連鎖地図について、各解析集団に基づいて DNA マーカーを遺伝地図上に位置付け、マーカー名による地図間比較機能を提供する地図表示システムを構築し、公開した（図1)。マーカーの塩基配列情報を利用してゲノム物理地図上に位置付け、ゲノムアセンブリが連鎖群に収束している 7 種について、物理地図を公開した。

対象 55 植物種について個別の情報ページを設け、基本情報、分類情報、ゲノムの特徴と解読手法、DNA マーカーや地図情報へのリンク、外部メタボローム DB へのリンク、関連サイトへのリンクを提供した。特に関連サイトについては、リンク先をトップページから当該植物のページに掘り下げることで、必要な情報に迅速にたどり着けるよう利便性の向上に努めた。

表 1. 文献キュレーションによるマーカーおよび QTL 情報（公開済み）

植物種名	論文数	マーカー数	QTL 数
アサガオ	1	75	0
ウンシュウミカン	8	58	0
エゾヘビイチゴ			
オランダイチゴ	15	341	74
カカオ	13	200	39
キマメ	7	487	13
キャッサバ	14	7,993	185
ジャガイモ	80	613	368
セイヨウヤマカモジ	7	214	3
タバコ	10	912	16
チャノキ	12	855	0
トウゴマ	5	223	0
ナツメヤシ	4	42	0
パパイア	9	53	14
モモ	22	405	217
ヤトロファ	10	856	18
ユーカリ	13	68	39
ヨーロッパブドウ	53	494	161
リンゴ	72	929	450
レタス	25	441	170
計	380	15,259	1,767

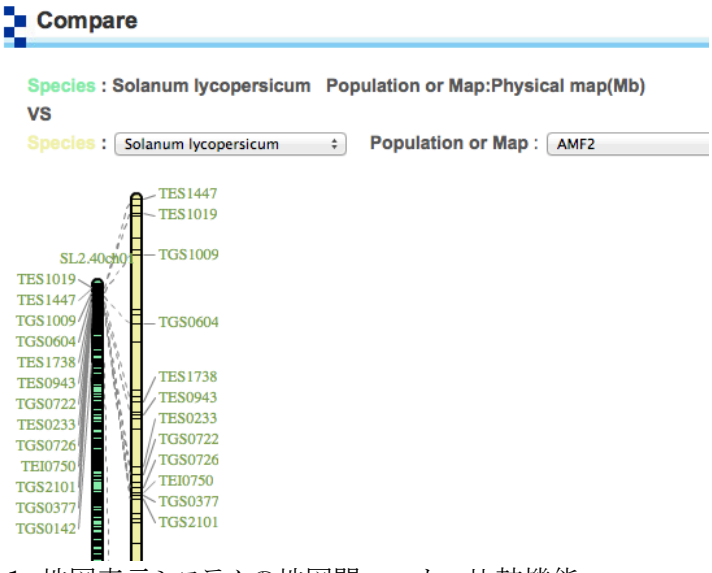


図 1. 地図表示システムの地図間マーカー比較機能

モデル実験材料であるシロイヌナズナの研究分野では、遺伝子クローンや変異体ラインなどのバイオリソースの開発、蓄積とそれらの共有が国際コミュニティ内で活発に行われてきたことが、研究の進歩に大きく寄与してきた。また、作物においては、さまざまな germplasms が種子等の形で長期保存され、育種に利用されてきた。現在、文部科学省のナショナルバイオリソースプロジェクト(以下 NBRP)では、シロイヌナズナ、イネ、コムギ、オオムギ、藻類、広義キク属、アサガオ、ミヤコグサ、タイズ、トマトの各種バイオリソースの収集と配布が実施されている。また、理化学研究所においてもバイオリソースセンターの事業として、シロイヌナズナ、ヒメツリガネゴケ、ポプラ、キャッサバ等の完全長 cDNA クローンライブラリ、種

子や培養細胞系の維持・配布が行われている。農林水産省では、農業生物資源ジーンバンクで穀物、豆類、牧草を中心に数万点以上のバイオリソースが保存され、配布する体制をとっている。こうしたリソース情報は、それぞれのセンター個々に工夫をこらしてユーザが利用しやすい形で特性解説や配布受付フォーム等が整備されている。各センターがもつこれらの情報を横断的に検索できるようになれば、わが国の植物研究をさらに推進させるための基盤となる。

本開発では、国内の主要リソースセンターが収集・配布している植物リソースのワンストップショップを構築することを目標として実施した。各リソースセンターが持つリソース情報を横断的に検索可能にするとともに、各種植物 DB や本研究開発の他の課題で整備した情報と、本事業で開発したポータルサイト PGDBj 上で統合検索を可能にした。ゲノム情報を基盤とした植物リソースの統合プラットフォームである PGDBj への組み込みによって、植物研究者が研究目的に適ったリソースに辿り着きやすくなるよう整備した。それぞれのリソースの周辺関連情報の入手を容易にし、DB の情報からその情報源であるリソースへのアクセスを容易にする機能を実現し、植物研究の利便性の向上を図った。

まず、文部科学省の植物バイオリソースが集約されている二大センターとして、NBRP ならびに理研バイオリソースセンター（以下 BRC）の提供する情報の調査と横断検索のための開発を実施した。両者においては既に NBRP サイトにて BRC リソースの検索が可能となっていたが、調査の結果、リファレンスゲノムに遺伝子 ID のバージョン情報が明示されていない場合があること、リファレンスゲノムの指定が適切でなく消失している場合がある等の問題点があった。これらの問題を解消することでリソース情報とゲノム情報の対応状況を改善し、また BRC からも NBRP 由来情報を統合的に検索するための情報整備の支援を実施した。BRC が構築、提供している、シロイヌナズナの遺伝子を基準として、配列類似性によって植物の遺伝子クローンを横断的に検索する機能をもつ Systematic consolidation of Arabidopsis and other Botanical REsource（SABRE）データベース上に、遺伝研 NBRP モデル植物 6 種（コムギ、オオムギ、アサガオ、ミヤコグサ、ダイズ、トマト）の遺伝子クローンの情報を追加し、合計 14 万種の遺伝子クローンを統合的に検索できる拡張を実施した。また、他のサイトからも直接活用できるよう、REST API による検索機能を追加した。このことによって、本開発による植物統合検索が実現した際に SABRE の API を活用することによって幅広い植物リソース情報にゲノム塩基配列の類似性を基盤として到達できる機能を実現した。

NBRP が集約した植物リソースは膨大であり極めて有用であるが、プロジェクトの集約には限界がある。そのため、これまで集約できていないが有用性の高いと思われる植物バイオリソース情報を調査したところ、カンキツ類の原種や過去の栽培種などのバイオリソースが近畿大学と農研機構果樹研究所に存在することがわかった。果樹研究所では cDNA クローンと塩基配列が公開されており、またウンシュウミカンのゲノム塩基配列決定も実施中であるため、本課題のゲノムを基盤としたリソース統合の目的に合致することから、上記のデータの散逸を防ぎ、また統合データベース上での集約を目標として、聞き取り調査とデータの収集、データベースの作成を実施した（図2）。

最終的に本開発によって作成された PGDBj 上から、遺伝研 NBRP と理研 BRC が有する 14 生物種計 1,504,022 件のリソース情報に SABRE の検索 API を用いることで NBRP、BRC それぞれのサイトとの横断検索を実現し、NBRP 以外の重要なリソースであるカンキツリソース情報として 900 個体分の在来種、栽培品種情報、公開可能な cDNA 配列情報を統合した。また、今後、ウンシュウミカンのゲノム塩基配列がリリースされた際にはその配列情報も統合できるよう準備を整えた。

最終年度には、本開発で実現したワンストップショップを永続的に運営することを目標としてデータの収集系と検索インデックスの自動化を実施し、作業の軽減を追求した。しかしながら、現実にはゲノム塩基配列のバージョンが変わることによる配列や ID の不整合を完全な形で自動トレースする事には困難が多く、統合データベースとしていかに永続性をもたせることができるかは、今後も引き続き取り組むべき課題である。

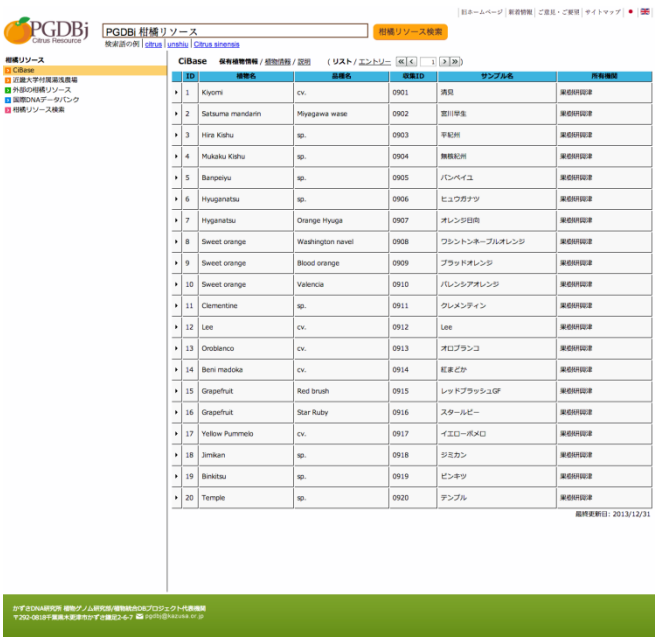


図2:PGDBj カンキツリソース DB

本研究開発の成果は、ポータルサイト PGDBj を通じて公開した(日本語版:http://pgdbj.jp, H24年8月公開;英語版:http://pgdbj.jp/?ln=en, H25年8月公開;図3)。公開後、ポータルサイトに設置した要望フォームやメール、各学会年会での講習会、公開展示を通じてユーザからの意見を収集し、コンテンツの充実やユーザーインターフェースの改良を実施した。また、各研究開発項目で構築した DB コンテンツを横断検索可能にするための「検索インデックス作成システム」及び「検索システム」を構築し、一般公開した(http://pgdbj.jp/estui/search/pidb.html, H25年7月公開;図4)。横断検索システムでは、本課題の4研究開発項目のコンテンツの他に、他の研究開発課題「メタボローム・データベースの開発」で構築された「生物種・代謝物 DB(KNApSAcK;http://kanaya.naist.jp/KNApSAcK/)」と「質量分析データ DB(MassBase;http://webs2.kazusa.or.jp/massbase/)」を検索対象に加えることにより、ユーザが植物のゲノム関連情報とメタボローム関連情報を取得できるように整備した。さらに、横断検索システムについては、外部ウェブサイトから PGDBj に対する横断検索を可能とする検索窓(検索 API)を開発した。PGDBj へのアクセス数は、公開を開始してからの約1年半で 34,570 件を超え、年毎の月平均を比較すると、平成 24 年では 1,100 件、平成 25 年では 2,200 件とアクセス数が1年間で倍増した。

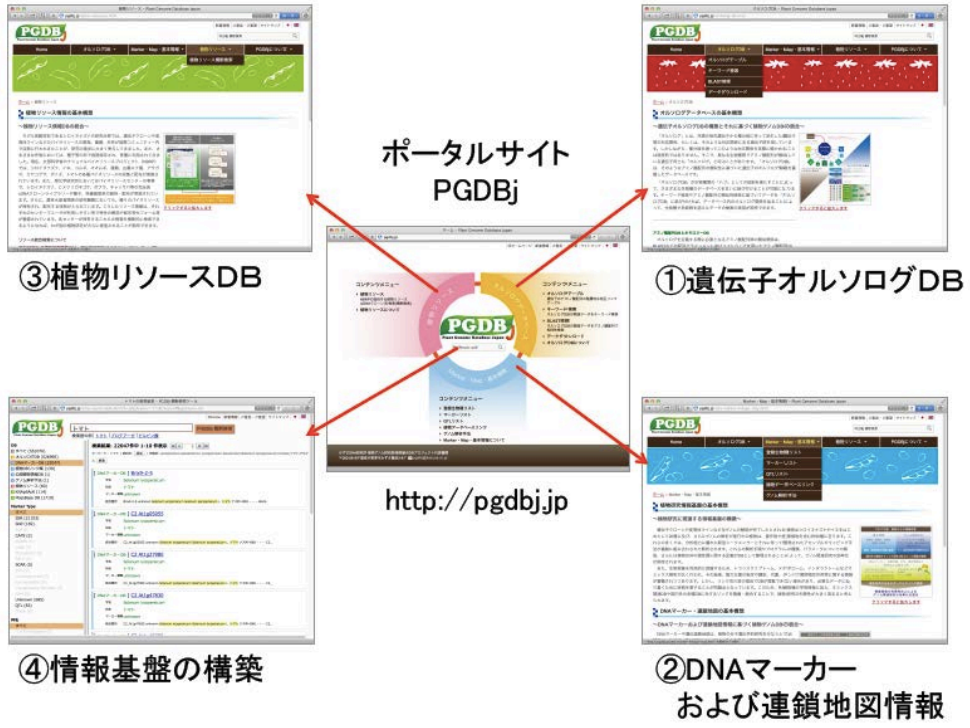


図3. ポータルサイト PGDBj（Plant Genome DataBase Japan）

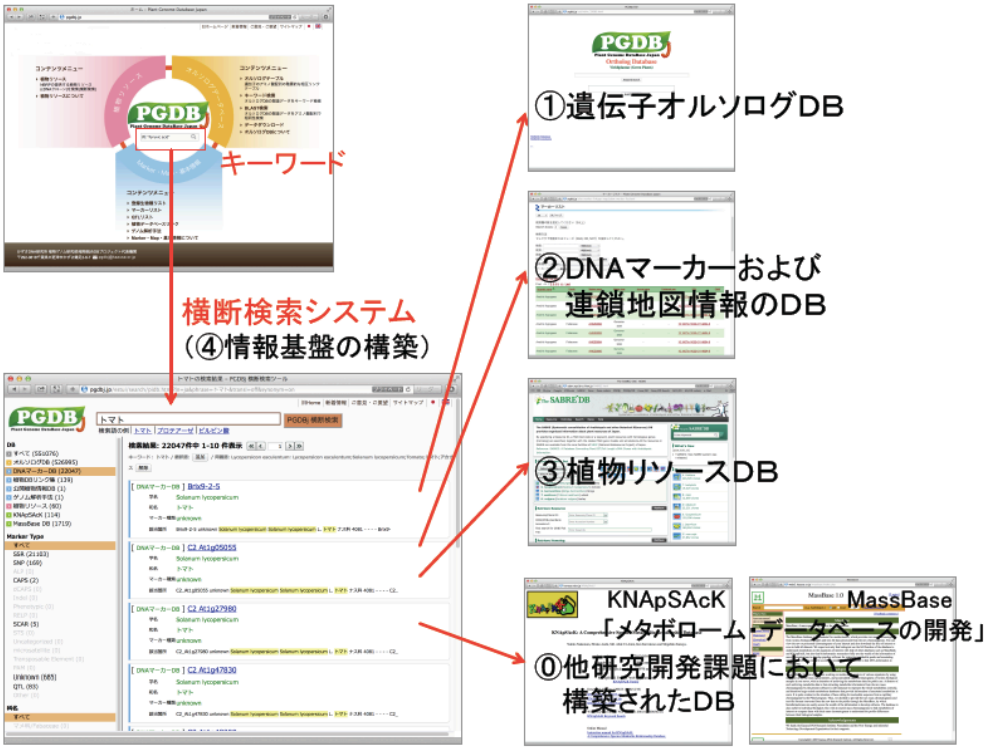


図4. PGDBj 横断検索システム

４．２ 研究課題名：遺伝子オルソログDBの構築と植物ゲノムDBの統合（新潟大学　中谷明弘グループ）

研究開発実施内容及び成果

モデル植物や作物についてのゲノム塩基配列の解読が精力的に進められており、それに伴い遺伝子配列も蓄積しているが、植物を主な対象としたオルソログ関係の整理は十分になされていない状況にある。そこで、植物についてのゲノム関連DB（以下、植物ゲノムDB）を遺伝子レベルで統合するために、生物種間の進化系統関係を反映させた階層的な遺伝子オルソログDBを作成し、遺伝子間のオルソログ関係によって植物ゲノムDB群のエントリを相互リンクした。これによって、項目の細分化が進んでいるDBの全体像を、遺伝子IDをタグとして統一的かつ俯瞰的に探索可能にした。

そもそも、植物ゲノムDBに含まれる遺伝子IDは必ずしも共通化されておらず、さまざまなスプライシングバリアントや断片配列の情報を含むことや、遺伝子IDもDBやバージョン毎に異なることもあるため、DBエントリ間を遺伝子IDの文字列比較のみで対応付けることは簡単ではない。ユーザが保有する遺配列情報と植物ゲノムDBとの対応づけも同様である。このため、計算機処理のみによって取得可能である遺伝子配列の類似性の情報は、DBエントリにタグ付けられた遺伝子ID間の対応付けを遺伝子シンボルや機能注釈等の既存知識に依存しないで行うための必須の情報であるといえる。

しかしながら、ユーザが検索対象DBの全ての遺伝子配列に対して類似性検索を実行して得られた結果を整理し、そこから様々なデータの抽出や解釈をすることは、現実的には難しい。そこで、本システムでは、一定規模数の生物種を網羅するアミノ酸配列のプールを準備し、そこに含まれる全アミノ酸配列を予め計算した網羅的な配列類似性情報に基づいてカテゴリ化（機能分類）し、統合対象DBのエントリやユーザ保有の遺伝子をこのカテゴリに対応付けることによって、それらを相互リンクした。これらのカテゴリ情報とカテゴリへの対応付け情報が遺伝子オルソログDBに格納されている。従って、検索クエリは、ひとたびオルソログDBの何れかのエントリにヒット出来てしまえば、以降はDB内の配列類似度情報に基づいて全ての関連データを辿ることが可能になる。

これまでに、NCBI RefSeq Databaseから取得したアミノ酸配列を用いてオルソログ情報の生成及び更新を行い（RefSeq Release57及び62まで追随）、アミノ酸配列、配列間類似性、種間系統関係、オルソログ情報を一体化させたオルソログDBを構築し、検索用のウェブサイトを構築した（http://pgdbj.jp/OD/）。DB検索用の基本的なAPI（SQLとPHPで実現した検索用URLの記述方法）を定義し、この基本的なAPIを組み合わせることによってより複雑な検索を実現できるようにした。現在、オルソログDBには緑色植物20生物種の約50万のアミノ酸配列とラン藻111生物種の約50万のアミノ酸配列が含まれている（合計約100万配列）。構築及び更新の処理は定型化されており、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、生成された情報は全て関係データベースに格納されるようになっている。この関係データベースから作成した検索用インデクスを介し、上記のAPIを用いてデータベース検索を行うことによって、PGDBjの横断検索中のオルソログDB部分が実現されている。

当初の統合対象DBは、整備が進んでいるモデル植物・作物のゲノムDB（研究代表者グループ保有）や、イネ、シロイヌナズナ、ヒメツリガネゴケ等のゲノムDB（各関連機関保有）とし、状況に応じて順次拡大した。これまでに、上記のオルソログDBに、研究代表者グループ保有のゲノムDB（ユーカリ、ヤトロファ、ミヤコグサ、トマト他）とEST DB（クラミドモナス、ミヤコグサ、シロイヌナズナ、スサビノリ、トマト、アカクローバ他）に加えて、RAP-DB、Rice TOGO、RiceXPRO、SALAD（農業生物資源研究所）、TAIR（米国）、RARGE、RPOPDB、TriFLDB（理化学研究所）、PHYSCObase（基礎生物学研究所）、KEGG/GENES（京都大学）他をリンクしている（図5）。これらのDB内のアミノ酸配列は、BLASTによってオルソログDBのアミノ酸配列と対応付けを行っており、DBのエントリ間を遺伝子レベルでリンクしている。配列情報とリンク先のDBアクセス用URLの情報があれば、ほぼ機械的にリンクを追加することが可能になっている。

オルソログ DB 内に登録されたデータはタブ区切りデータとして整理されており、ウェブブラウザを用いた GUI による検索の他、オフラインでの解析や RDF 化等への対応が可能なように管理している。オルソログ情報はダウンロードページから取得できるようにすると共に、アーカイブ化に向けたデータ提供を継続する。現状では、RDF 化や SPARQL 検索環境の構築は実施していないが、「ゲノム・メタゲノム情報を基盤とした微生物 DB の統合」をはじめとする他課題と、国際的にも使用されつつあるオルソログに関するオントロジーを共有することを確認しており、それに基づいた RDF 化を行うことによって、より広範な系統群に渡る一体化した分散検索環境の構築が期待される。

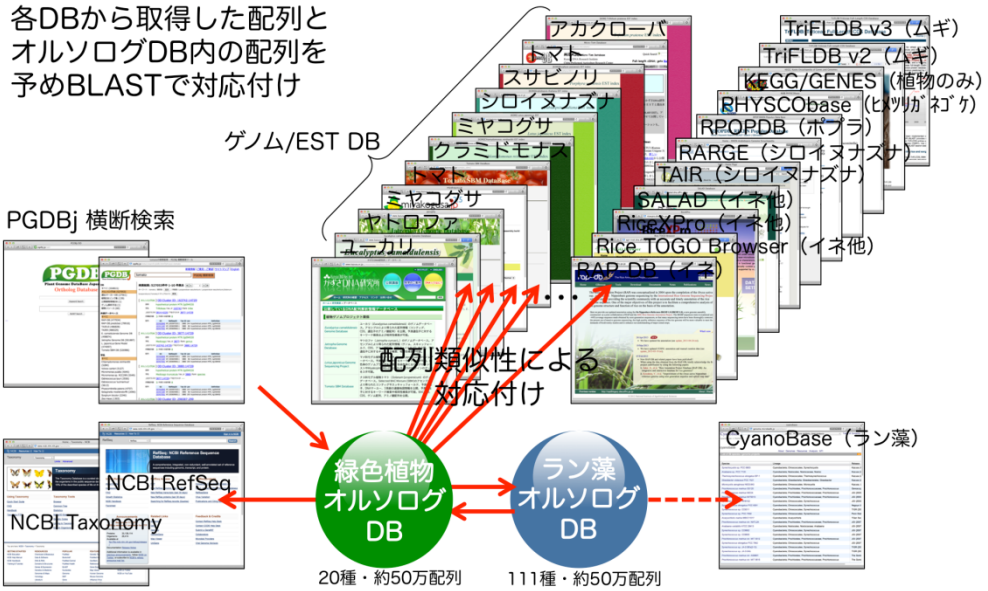


図5. オルソログ DB をハブとした植物ゲノム/EST DB の統合

§5 成果発表等

(1)原著論文発表　(国内(和文)誌　0 件、国際(欧文)誌　2 件)

1. Asamizu E, Ichihara H, Nakaya A, Nakamura Y, Hirakawa H, Ishii T, Tamura T, Fukami-Kobayashi K, Nakajima Y, Tabata S. (2014) Plant Genome DataBase Japan (PGDBj): a portal website for the integration of plant genome-related databases, *Plant Cell Physiol.*, 55(1):e8. doi: 10.1093/pcp/pct189.

2. Fukami-Kobayashi K, Nakamura Y, Tamura T, Kobayashi M. (2014) SABRE2: A Database Connecting Plant EST/Full-Length cDNA Clones with Arabidopsis Information. *Plant Cell Physiol.*, 55(1): e5. doi: 10.1093/pcp/pct177.

(2)その他の著作物(総説、書籍など)

1. 田畑哲之、「植物ゲノムデータベースの統合」 細胞工学 vol. 30, No. 12, pp.1314-1317, 2011

(3)国際学会発表及び主要な国内学会発表

① 招待講演(国内会議　3 件、国際会議　1 件)

1. 市原寿子、統合化推進プログラムーゲノム情報に基づく植物データベースの統合ー、第54回日本植物生理学会大会、岡山大学、H24 年 3 月 22 日

2 中村保一、Integration of databases for microbes and plants from the viewpoint of

(meta-)genomics 、Joint Conference on Informatics in Biology, Medicine and Pharmacology (生命医薬情報学連合大会)、タワーホール船堀、H24 年 10 月 15 日

3. 市原寿子、−ゲノム情報に基づく植物データベースの統合−、第31回日本植物細胞生物学会(札幌)大会、北海道大学、H25 年 9 月 12 日

4. 市原寿子、シンポジウム「データベース講習会」:Plant Genome DataBase Japan (PGDBj)-植物統合 DB2013 年度版の紹介-、第55回日本植物生理学会大会、富山大学、H26 年 3 月 18 日

② 口頭発表(国内会議 2 件、国際会議 0 件)

1. 田畑哲之、市原寿子、中谷明弘、中村保一、平川英樹、統合化推進プログラム−ゲノム情報に基づく植物データベースの統合−、植物細胞分子生物学会、奈良先端科学技術大学院大学、H24 年 8 月 4 日

2. 浅水恵理香、市原寿子、平川英樹、中谷明弘、中村保一、田畑哲之、ゲノム情報に基づく植物データベースの統合−Plant Genome DataBase Japan (PGDBj)−、第 31 回植物細胞分子生物学会年会、札幌、H25 年 9 月 11 日

③ ポスター発表(国内会議 2 件、国際会議 1 件)

1. 平川英樹、中村保一、中谷明弘、田畑哲之、ゲノム情報に基づく植物データベースの統合、第 34 回日本分子生物学会年会、パシフィコ横浜、H23 年 12 月 13 日～16 日

2. Asamizu E, Ichihara H, Nakaya A, Nakamura Y, Hirakawa H, Tabata S. Plant Genome DataBase Japan (PGDBj), a comprehensive database covering information of plant genome-related databases in Japan, PAG XXII, San Diego, USA, H26 年 1 月 13 日

3. 浅水恵理香、市原寿子、平川英樹、中谷明弘、中村保一、田畑哲之、Plant Genome DataBase Japan (PGDBj), a comprehensive database covering information of plant genome-related databases in Japan、第 55 回植物生理学会年会、富山、H26 年 3 月 18 日

(4)知財出願

無し

(5)受賞・報道等

無し

§6 研究開発期間中に主催した会議等

主なワークショップ、シンポジウム、アウトリーチ等の活動

年月日	名称	場所	参加人数	概要
H23 年 4 月 27 日	グループ合同ミーティング(非公開)	東京八重洲ビジネスセンター	4 人	研究進捗報告のためのミーティング
H23 年 5 月 21 日	グループ合同ミーティング(非公開)	東京八重洲ビジネスセンター	4 人	研究進捗報告のためのミーティング

H23 年 7 月 22 日	グループ合同ミーティング（非公開）	かずさ DNA 研究所	4 人	研究進捗報告のためのミーティング
H23 年 10 月 28 日	グループ合同ミーティング（非公開）	かずさ DNA 研究所	10 人	研究進捗報告のためのミーティング
H24 年 1 月 6 日	グループ合同ミーティング（非公開）	かずさ DNA 研究所	10 人	研究進捗報告のためのミーティング
H24 年 2 月 11 日	グループ合同ミーティング（非公開）	かずさ DNA 研究所	4 人	研究進捗報告のためのミーティング
H24 年 4 月 20 日	グループ合同ミーティング（非公開）	かずさ DNA 研究所	5 名	DB 開発計画の確認と進捗報告
H24 年 7 月 20 日	グループ合同ミーティング（非公開）	かずさ DNA 研究所	5 名	各項目の進捗報告
H24 年 8 月 31 日	グループ合同ミーティング（非公開）	かずさ DNA 研究所	4 名	各項目の進捗報告
H24 年 10 月 5 日	グループ合同ミーティング（非公開）	時事通信ホール	5 名	各項目の進捗報告
H24 年 11 月 22 日	アドバイザリー委員会（非公開）	東京ステーションコンファレンス	22 名	外部有識者からの PGDBj に対する意見聴取
H24 年 12 月 12 日	グループ合同ミーティング（非公開）	マリンメッセ福岡	5 名	各項目の進捗報告
H24 年 12 月 13 日	「データベースから始まる分子生物学～研究の新しいスタイルを模索する～」第 35 回日本分子生物学会年会 ワークショップ	マリンメッセ福岡		ワークショップ企画・オーガナイズ
H25 年 3 月 26 日	グループ合同ミーティング（非公開）	東北大学川内北キャンパス	4 名	各項目の進捗報告
H25 年 4 月 26 日	グループ合同ミーティング（非公開）	かずさ DNA 研究所	6 名	DB 開発計画の確認と進捗報告
H25 年 7 月 1 日	植物統合 DB・画期的ゲノム情報 DB 情報交換会（非公開）	農業生物資源研究所	9 名	農水省 DB プロジェクトと統合推進事業との情報交換
H25 年 7 月 10 日	グループ合同ミーティング（非公開）	新潟大学駅南キャンパス	8 名	目標達成に向けた年度内スケジュールの確認、課題の検討
H25 年 7 月 30 日	アドバイザリー委員会（非公開）	東京ステーションコンファレンス	24 名	外部有識者からの PGDBj に対する意見聴取
H25 年 10 月 17 日	オントロジー整備 WG（非公開）	東京ステーションコンファレンス	8 名	植物オントロジー整備に向けた意見交換
H25 年 12 月 6 日	データベースを使い倒した新しい研究スタイルによる分子生物学」第 36 回日本分子生物学会年会 ワークショップ	パシフィコ横浜		ワークショップ企画・オーガナイズ
H26 年 3 月 4 日	アドバイザリー委員会（非公開）	東京ステーションコンファレンス	25 名（予定）	外部有識者からの PGDBj に対する意見聴取

§7 ユーザー評価結果への対応

≪平成 25 年 7 月に実施した「NBDC における事業活動のユーザー評価」(http://biosciencedbc.jp/user-hyouka-2013/result-summary)で得られたユーザーの意見、提案等 (詳細は別紙 2 を参照)に対し、実施済み若しくは実施予定の対応策等を具体的に記載してください。)

㊀実施済み
㊁実施予定

- コメント1:便利だと思います。さらにデータ数が増える事を期待します。

  ㊀実施済みの対応策:いずれの研究開発項目においても、掲載の対象とした植物種について文献および DB から取得可能なデータの整備と追加を実施している。

- コメント2:具体的な活用事例をのせてほしい。

  ㊁実施予定の対応策:活用事例を含むユーザーマニュアルを作成し、ポータルサイトにて公開する。

- コメント3

  コメント(1):Cluster ID等が数字だけなので、その値の意味するところがなかなか理解できないです。

  ㊁実施予定の対応策:Cluster IDは単なる通し番号であるため、値それ自体には意味はないことを強調するなど、データ自体に関する情報を充実させる。横断検索システムを介して提示される情報では、Cluster IDと共に、それが包括する植物種や分子名などを表示しているように、Cluster IDを介して集約される情報によって意味付けするという方針となっている。検索結果の見方等について説明したユーザーマニュアルを作成し、ポータルサイトにて公開すると共に、Cluster IDを意識させないUIの構築も行う。

  コメント(2):検索結果などが膨大であると、表示に時間がかかったり時には止まってしまうため、結果表示数を変えるなどユーザビリティの向上を期待します。

  ㊀実施済みの対応策:横断検索システムについては、検索と表示の仕組みに対して根本的な改修を実施し、公開当初よりも速く結果が表示出来るようにした。

  コメント(3):植物研究のポータルサイトとして、認知度を上げる必要があると感じました。

  ㊀実施済みの対応策:各種学会年会でのブース出展や講習会、植物研究者を対象としたメーリングリストでの定期的なアナウンス、日本植物生理学会から刊行されている国際誌「Plant & Cell Physiology」のデータベース特集号への掲載等を通じた広報活動を実施した。
  ㊁実施予定の対応策:外部ウェブサイトからPGDBjに対する横断検索システムを可能とする検索APIを開発済みであり、今後協力を得られる機関や組織のウェブサイトに設置を依頼する予定である。

- コメント4

コメント(1):オーソログの定義の説明がほしい。現在の記述は不十分だし、その手法がオーソログの定義として確立しているのかもよくわからない。(ゲノムデータでのオーソログ定義が怪しいことは多々あるので、このサイトだけの問題ではないかもしれません)

⊖実施済みの対応策:「Plant & Cell Physiology」のデータベース特集号にてもオーソログ情報の抽出手順を記述した。指摘の通り、オーソログの定義は難しいため、手続き的な定義(作成の手順とパラメータ値による定義)に基づいている。

⊖実施予定の対応策:パラメータ違いのDB群の準備や配列類似度の閾値等を指定した検索の実現などによって、オーソログの範囲を動的に設定できるシステムを検討中。また、オーソログのオントロジーを共有してRDF化して、より広範な系統群のオーソログを含めたメタ検索を実現することにより、緑色植物やラン藻のみでは意味付けが難しいオーソログへ他の系統群からの情報の転送を行うことも検討中。

コメント(2):検索結果の一括ダウンロードなどの機能があるとありがたい。

⊖実施予定の対応策:横断検索システムでのオルソログ配列データと機能データについては、検出されたCluster ID単位でダウンロードできるように改修している段階である。

コメント(3):DNAマーカー・連鎖地図はまだ完成していないのか?

⊖実施済みの対応策:情報収集が完了した植物種については、DBにて公開した。また、DNAマーカー・連鎖地図情報のDBに移動するリンクボタンの中に、正しくアクセスできないものが含まれていたため、これらを修正した。

コメント(4):リンク集というところも使えなかった。

⊖実施済みの対応策:植物関連DBについて情報収集と整備を実施し、リンク集として公開した。また、リンク集に移動するリンクボタンの中に、正しくアクセスできないものが含まれていたため、これらを修正した。

- コメント5:遺伝子によってオーソロググループが大きすぎたり小さすぎたりすることがあるので、状況に応じて類似性の閾値を変化させることができると使いやすいと思います。

  ⊖実施予定の対応策:上掲の通り、オーソロググループの範囲を動的に設定できるようなシステムを検討中。現在は、デフォルト値設定での情報をそのまま検索結果としているが、過度に複雑にならない範囲で検索パラメータの変更、もしくは、検索結果の絞り込みを行えるようにする予定である。

- コメント6:植物関連の学会HPトップにPGDBjのリンク・バナーを置いてもらってはどうでしょうか?

  ⊖実施予定の対応策:本研究開発は、日本植物学会会長、日本植物生理学会会長、日本植物分子細胞生物学会会長の支持と協力のもと実施されており、今後、各学会HPの運営方針に沿ってPGDBjのリンク・バナーおよび開発したPGDBj検索APIを設置してもらえる様にお願いする予定である。

- コメント7：最初のページだけ力を入れていて、検索結果が全くデザインされていないのが残念。

  ㊀実施済みの対応策：検索結果の表示内容および機能を強化し、デザインについても改良した。具体的には、検索対象となる各種DBの分類項目に基づいて表示枠を色分けし、検出された情報を視覚的に捉えやすくする工夫をした。

§８ その他

（1）研究代表者として、研究開発、プロジェクト運営等について、上記以外に報告したいことがあれば、自由に記載してください。

本研究期間中に以下のような学会展示で広報活動を行った。

1. 市原寿子、平川英樹、中谷明弘、中村保一、田畑哲之、統合化推進プログラムーゲノム情報に基づく植物データベースの統合ー、第 30 回日本植物細胞分子生物学会大会、奈良先端科学技術大学院大学、H24 年 8 月3日～5 日
2. 市原寿子、平川英樹、中谷明弘、中村保一、田畑哲之、統合化推進プログラムーゲノム情報に基づく植物データベースの統合ー、日本植物学会第76回大会、兵庫県立大学、H24 年9月15日～17日
3. 市原寿子、平川英樹、中谷明弘、中村保一、田畑哲之、統合化推進プログラムーゲノム情報に基づく植物データベースの統合ー、第35回日本分子生物学会大会、マリンメッセ福岡、H24 年12月11日～14日
4. 市原寿子、平川英樹、中谷明弘、中村保一、田畑哲之、統合化推進プログラムーゲノム情報に基づく植物データベースの統合ー、第54回日本植物生理学会岡山大会、岡山大学、 H25年3月21日～23日
5. 市原寿子、平川英樹、中谷明弘、中村保一、田畑哲之、統合化推進プログラムーゲノム情報に基づく植物データベースの統合ー、日本農芸化学会 2013 年度大会、東北大学、H25年3月24日～27日
6. 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、ゲノム情報に基づく植物データベースの統合 (http://pgdbj.jp/)、第31回日本植物細胞生物学会（札幌）大会・シンポジウム、北海道大学、H25年9月10日～12日
7. 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、ゲノム情報に基づく植物データベースの統合 (http://pgdbj.jp/)、日本植物学会第77回大会（札幌）、北海道大学、H25年9月13日～15日
8. 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、（特別企画展「使ってみようバイオデータベースーつながるデータ、広がる世界」）ゲノム情報に基づく植物データベースの統合 (http://pgdbj.jp/)、第36回日本分子生物学会年会、神戸国際展示場、H25年12月3日～5日
9. 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、ゲノム情報に基づく植物データベースの統合（http://pgdbj.jp/）、第55回日本植物生理学会年会、富山大学、H26年3月18日～20日
10. 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、（附設展示会）ゲノム情報に基づく植物データベースの統合（http://pgdbj.jp/）、日本農芸化学会 2014 年度（東京）大会、明治大学、H26年3月27日～30日

以上