

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 2017 年採択課題)

## 研究開発終了報告書

「個体ゲノム時代に向けた植物ゲノム情報解析基盤の構築」  
田畑 哲之 | かずさ DNA 研究所 所長

研究開発期間： 2017年4月～2022年3月



©2022 田畑 哲之 (かずさ DNA 研究所) Licensed under CC BY 4.0

## §1. 研究開発実施の概要

### 【研究開発の目的】

植物研究の大きな特徴は、モデル実験植物から実用植物までさまざまな植物種が研究開発の対象になっていることである。また、個々の植物を基礎研究から育種まで深く掘り下げるものから、植物種をまたいで有用遺伝子の探索や進化プロセスを推定する横断的なものまで、研究のスタイルも多様である。一方、ゲノム解析をめぐる状況は目まぐるしく変化し、2016 年度末の時点において、次世代型シーケンサ(NGS)が広く普及し、植物においても多くの種でゲノムの *de novo assembly* や多数の品種・系統のリシーケンスが実施されるようになった。そこで本研究開発では、今後一層の増加が見込まれるゲノム配列データを柔軟に取り入れ、ユーザーが必要な情報を抽出できる DB を新たに開発することとした。

### 【実施内容】

本事業の第二期まで開発を行っていた DNA マーカー情報等を格納するデータベース「PGDBj」を維持しつつ、PGDBjに格納していたデータに加え、ゲノム配列・遺伝子情報等を格納する「Plant GARDEN」の開発を新たに行った。また、オルソログ情報を活用した種を超えたゲノム情報の統合、カスタム型多型検出システムの構築、センテンスキュレーションを活用した遺伝子情報の収集、RDF による他 DB とのデータリンクなどを実施した。

### 【得られた成果の概要】

新規に開発した DB「Plant GARDEN」(<https://plantgarden.jp>)は本事業開始年度よりスキーマ設計を開始し、2018 年度に DB のデザインを設計・実装して、2019 年 3 月に β 版)を公開、2020 年 7 月に正規版を公開している。DB のデザインにあたっては、専門性にとらわれない幅広いユーザー層をターゲットとして「シンプルであること」「動作が直感に分かりやすいこと」に留意した。また、データ検索機能については、Elastic Search を利用した横断検索システム「pges」を開発し、Plant GARDEN に pges 実装するための API の整備と URL の Restful 化を行ってデータの高速検索を可能とした。安全性に関しては、β 版では http を用いていたが、正規版では https への変更を行い SSL へ対応するとともにサーバーの脆弱性評価をクラウド型脆弱性診断ツールである Vaddy を用いて実施した。

Plant GARDEN に格納するコンテンツはキュレーションによって順次増やしており、2019 年 3 月の β 版公開時の格納データは 9 種・9 ゲノム配列であったが、18 ヶ月後の 2021 年 9 月現在で 134 種 169 ゲノム配列、8,136,812 遺伝子配列の他、DNA マーカー、形質関連座、SNPs、連鎖地図情報を格納している。さらに Plant GARDEN に格納した全遺伝子配列を独自に開発したアノテーションツール「Hayai-Annotation」により再アノテーションし、OrthoDB より公開されているオルソログ遺伝子配列との対応づけを行った。この情報に基づいて、配列が類似した遺伝子の情報を種を超えて探索することができる。また、ゲノム配列多型情報については、公開されている SRA 配列と Plant GARDEN に格納しているゲノム配列を比較して検出した SNPs を独自コンテンツとして提供するとともに、ユーザー自らが有する配列情報をゲノム配列にマッピングし、SNPs を検出するツールの開発と提供も行った。

一方、本事業においては他プロジェクト等で開発された DB との連携も求められている。そこで、GlyCosmos Portal とのデータリンクを図るとともに、MicrobeDB.jp と MetaboBank 内に格納されているデータと RDF を介して連携検索を実施できるシステムを開発した。Plant GARDEN は正規版公開から 1 年 2 ヶ月であるが、アクセス数は順調に伸びている。また、コンテンツを拡充していないものの PGDBj のアクセス数も一定数あることから引き続き維持をする予定である。

## §2. 研究開発実施体制

### 1. 研究グループ

#### (1) 「かずさDNA研究所」グループ(研究代表者グループ)

- 担当項目

ゲノム横断的関連情報表示サイトの開発  
種を超えたゲノム情報統合のためのデータリンク基盤の構築  
カスタム型多型・ハプロタイプ検出システムの構築  
コンテンツの更新

#### (2) 「大阪大学」グループ(主たる共同研究者グループ(1)-1)

- 担当項目

種を超えたゲノム情報統合のためのデータリンク基盤の構築

#### (3) 「東京大学」グループ(主たる共同研究者グループ(1)-2, 主たる共同研究者の移動に伴い「大阪大学」グループから変更)

- 担当項目

種を超えたゲノム情報統合のためのデータリンク基盤の構築

#### (4) 「筑波大学」グループ(研究代表者グループ)

- 担当項目

ゲノム横断的関連情報表示サイトの開発

### 2. 有識者会議等

#### (1) 会議概要

名称	アドバイザー委員会
目的	本事業において開発するDBの開発方針等について、外部有識者と実施者の間で意見交換を行う。
委員数	10人

#### (2) 開催歴

年月日	場所	参加人数	主な議題
2017年8月3日	ステーションコンファレンス 東京	25人	プロジェクトの開始にあたって計画に対する意見交換
2018年2月8日	ステーションコンファレンス 東京	22人	新規データベースのスキーマや種を超えた遺伝子配列の比較に関する手法の検討について
2018年5月17日	ステーションコンファレンス	23人	新規データベースの開発状況や

	東京		種を超えた遺伝子配列の比較に関する手法の検討について
2019年2月18日	ステーションコンファレンス 東京	25人	Plant GARDENの開発状況と今後の方向性について
2019年8月7日	ステーションコンファレンス 東京	25人	Plant GARDENの開発状況と今後の方向性について
2020年10月1日	ステーションコンファレンス東京およびWeb開催	26人	Plant GARDENの開発状況と今後の方向性について
2021年3月10日	Web開催	27人	Plant GARDENの開発状況と今後の方向性について
2021年8月31日	Web開催	25人	Plant GARDENの開発状況と今後の方向性について

### §3. 研究開発の目的、実施内容及び成果

#### 1. 研究開発の背景

植物研究の大きな特徴は、モデル実験植物から実用植物までさまざまな植物種が研究開発の対象になっていることであろう。また、個々の植物を基礎研究から育種まで深く掘り下げるものから、植物種をまたいで有用遺伝子の探索や進化プロセスを推定する横断的なものまで、研究のスタイルもまちまちである。広範な植物種について多種多様なゲノム関連情報を提供する 2016 年度までに我々は植物ゲノムポータルサイト「PGDBj」を開発した。PGDBj は、このような特徴をもつ植物の研究開発に最適のものであることから、アカデミアから産業界まで幅広い学会や団体の支持を得ることができたと考えられる。さらに今回の提案では、複数の対象植物と複数のゲノム関連情報を組み合わせて表示するシステム、ユーザーが NGS で取得したデータを投入して DB 内の参照配列と対応させるツールの開発など、最新の要望にも対応するための開発が含まれている。PGDBj はアカデミアから種苗業界まで幅広いユーザーの利用を想定しており、その用途も千差万別であった。

一方、ゲノム解析をめぐる状況は目まぐるしく変化し、2016 年度末の時点において、次世代型シーケンサ(NGS)が広く普及し、植物においても多くの種でゲノムの *de novo assembly* や多数の品種・系統のリシーケンスが実施されるようになった。そこで本研究開発では、オルソログ検出手法を基軸に、今後一層の増加が見込まれるゲノム配列データを柔軟に取り入れ、ユーザーが必要な情報を抽出できる DB を新たに開発することとした。さらに個体ゲノムの時代が植物にも到来することを見据え、ユーザー自身が分析した NGS のショートリードデータを投入できるカスタム性の高い多型検出システムを開発・実装することとした。

#### 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。横断検索を用いることで植物に特化した多種類の DB へ効率的にアクセスできる。
Plant Genome And Resource Database ENtry	Plant GARDEN	植物ゲノム関連情報を統合化するハブとして本課題において新たに構築するポータルサイトである。多数の植物種のゲノム配列情報、遺伝子配列情報やマーカー情報および異なる植物種の遺伝子配列類似性情報を検索できる。また、ブラウザ等により異なる種類のデータの関係性を検出・閲覧できるシステムである。

• 上記以外のもの

正式名称	略称	概要
KusakiDB		植物の遺伝子アノテーションを高精度に実施するための遺伝子情報を格納した DB。遺伝子オルソログ情報を Ortho DB, Uniprot, NCBI nr の情報を一元化した。

(2) ツール等

正式名称	略称	概要
Hayai-Annotation Plants		植物を対象に遺伝子機能アノテーションを実施するツール。ローカル環境で動作し、実行速度が極めて速く、正確かつ包括的なアノテーションが可能である。
Hayai-annotation ZEN		Plant GARDEN 用の遺伝子機能アノテーションのソフトウェア。
Hayai-annotation v2		遺伝子機能アノテーションのソフトウェアで、Hayai-annotation Plants の次バージョン。KusakiDB をアノテーションに用いる。
SNP Detection		配列の精査、マッピングおよび変異検出を実施する解析パイプラインである。Plant GARDEN のユーザーを対象にβ版として公開中。FTP を通じてユーザーがリード配列とリファレンス配列をかずさDNA研究所のサーバーに Upload して解析を実行する。実行にはユーザー登録が必要である。
GS modeling		ユーザーが特定の個体集団に対して形質値と Genotype 情報を Upload し、Genotype から形質値を予測するための予測式を生成するツール
正式名称未定	My GARDEN	ローカル環境で動作するアプリケーションでユーザー自身が保有するデータと Plant GARDEN 内のゲノムデータ等を比較解析できるツール。

※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

① ゲノム横断的関連情報表示サイトの開発

統合化推進プログラム第二期までに、遺伝子情報とマニュアルキュレーションにより得た DNA マーカー、QTL 情報をゲノム配列上にマッピングすることで集約させ、JBrowse 上で閲覧できるようにした。対象は、Pseudomolecule として解読された 35 植物種であったが、そのうちエゾヘビイチゴ、リンゴ、ハクサイについては集約を完了した。本提案では、残りの 32 種について、キュレーションした情報を引き続き JBrowse 上に集約させる。さらに、かずさ DNA 研究所で解読した 16 植物種についても JBrowse 上への集約を行う。本研究開発を実施している間にも多くの新規植物のゲノム配列や変異体の全ゲノム配列が解読されるため、キュレーションした多型情報や原因遺伝子などを迅速に JBrowse 上に反映させるためのスクリプトを作成し自動化を試

みる。また、JBrowse には多数のプラグインが開発されており、その中でも本研究開発で有益だと思われる gwasviewer と multivariantviewer を PGDBj の JBrowse にインストールする。現在、さまざまな情報を JBrowse 上に掲載する試みがなされており、新たに開発されたプラグインは The GMOD Plugin Directory で公開されるため、有益なものがあれば採用する。これまでにキュレーションした DNA マーカーや QTL 情報、そして、本提案で新たにキュレーションする GWA S、RAD-Seq、QTL-Seq のデータを JBrowse 上で閲覧できるようにすることで、形質に関わる原因遺伝子の特定をサポートする。

さらに SRA データに登録されている内容(植物種名、品種・系統名、サンプリングされた器官、実験条件など)をキュレーションし、その内容を表にまとめることでユーザーに閲覧しやすい形で提供する。近年、ショートリードをゲノム配列に対してマッピングすることでゲノムワイドな SNP と InDel を検出する研究が行われているが、現在ではゲノムワイドな SSLP(Simple Sequence Length Polymorphism)も検出できるようになった。そこで、Pseudomolecule としてゲノム配列が決定されている植物種について、SRA からゲノム由来と転写産物由来のショートリードを入手し、精査したものをマッピングすることで SNP と InDel、SSLP 候補を検出する。その結果得られた多型候補について品種・系統間のジェノタイプ表を作成し公開する。植物種によっては倍数性が高いものやヘテロ性が高いものなどゲノム構造が多様であるため、検出された多型がヘテロかホモになっているかはユーザーが判断できるよう、マッピングで得られた VCF 形式をタブ区切り形式に変換した表を公開する。得られた多型候補については、最終的には JBrowse 上に集約させる。

## ② 種を超えたゲノム情報統合のためのデータリンク基盤の構築

複数の生物種に跨ったゲノム情報の統合を目的として、遺伝子の相同性(アミノ酸配列の類似性)の情報に基づいたデータリンク基盤の構築を行う。本研究開発では、以下の「関係」に基づいた情報を自動的に抽出してデータベースに蓄積するシステムの構築を行う。

### ①種内および種間での配列の類似関係

### ②類似関係にある配列のゲノム配列/染色体での位置関係

これらの「関係」に基づいて他の項目で整備する各種のゲノム情報データベースのエントリーを互いに関連付ける(データリンク)ための基盤を構築する。①に基づいて配列の類似情報によってエントリー間を直接関連づけることを可能にする。また、②に基づいて遺伝子群の近接関係が種を超えて保存されているゲノム配列や染色体の領域(シンテニー様領域)を抽出し、それに基づいて種間でのデータの関連づけを可能にする。これらによって、「名前情報」だけではなく、染色体名と塩基位置(bp 単位)で指定されたゲノム配列中の「位置情報」を別の生物種に対応づけて検索することが可能になる。

これまでの PGDBj オルソログデータベースでは、NCBI RefSeq database によるアミノ酸配列を使用していたが、これを生物種に固有の配列情報と識別子(ID)を用いるようにする。配列情報の取得から類似性情報の生成およびデータベースへの蓄積までの全てを自動化するために NCBI から一括してデータを取得していた部分を本研究開発で整備する配列情報を用いるように変更し、また、内部的には GI number ないし RefSeq ID をユニークな識別子(ID)としてシステム構築を行ってきた部分を生物種に固有の配列の識別子を使用するように変更する。全てが GI number で管理されているために NCBI のデータベースと暗黙のうちにリンクされているという特性があったが、必要に応じて NCBI による識別子との対応関係を生成して対応することとする。

従来の PGDBj オルソログデータベースでは、生物種の増加に対応するための部分的な追加や更新を想定していないという仕様であり(再構築が必要)、また、再構築の前後では新旧の情報の互換性が保証されていなかった。この点を改良して、既存の情報を可能な限り維持しつつ逐次的に生物種が追加できる仕組みを構築する。また、生物種を追加する順序に結果が依存しない性質(対称性)の実現を検討する。

これまでも生物種間の進化系統関係(NCBI Taxonomy databaseによる)に基づいて、近縁種から類似したアミノ酸配列をクラスタ化する手法を採用していたが、それに先立って全アミノ酸配列間の配列類似性の計算(BLAST)を実施していた。これは、 $O(N^2)$ の計算量とデータ容量が必要となって生物種の増加に伴ってデータベース構築の律速となっていたため、類似した配列群は代表間のみ絞る等の仕組みを構築して、より少数の配列類似性の計算のみで遺伝子の相同性の情報を蓄積できる仕組みを構築する。従来システムでは、予め計算した配列類似性の情報を蓄積してデータベース検索で配列類似性情報を取得可能にしていたが、必要となった際に動的に配列類似性情報を生成するように変更する。

### ③ カスタム型多型・ハプロタイプ検出システムの構築

ユーザーが自身のデータを専用のサーバーに投げ込み、ゲノム横断的関連情報表示サイトに格納されている配列からリファレンスとなる配列を選択して、GUI上でプログラムを操作して多型を検出するシステムを構築する。さらには複数の系統間の SNPs を比較して連鎖不平衡(Linkage Disequilibrium, LD)やハプロタイプブロックを検出する機能も追加する。

### ④ PGDBj コンテンツの更新

現在 PGDBj から提供されている植物リソース情報、登録植物種リスト、DNA マーカー関連情報、QTL 関連情報については、引き続き公開する。また、上記の情報のうち新たに文献が公表されればキュレーションを適宜行い、PGDBj に追加する。近年の NGS 技術の進歩により数多くの植物種についてのゲノム情報が得られているため、さらに植物種を拡大する。本研究開発期間中に新たに解読された植物ゲノムの情報や上記の関連情報については、外部 DB からの情報取得や文献のマニュアルキュレーションによって収集し、適宜追加し公開する。

また、キュレーションの自動化を目指して、NCBI の PubTator の他に、センテンスキュレーションをローカルで実施するためのツールを開発する。そのために、遺伝子名や病害抵抗性、ストレス耐性といった農業形質の語句をまとめた辞書を作成し、それらが含まれたセンテンスを自動的に抽出するプログラムを作成する。さらには、新たに追加されたコンテンツについても RDF 化を実施する。

第二期においてトップページに配置した育種家向けの検索システムについては、現在の形のまま引き続き PGDBj から公開する。今後、アドバイザー委員会や学会などを通じてユーザーから重要な意見があった場合、かずさ DNA 研グループと大阪大学グループとの間で協議し、研究開発計画を再検討することで対応する。また、オントロジーの整備についても、アドバイザー委員会での議論を参考に、植物学会と連携しながら引き続き検討を行う。

## (2) 期間中に追加・削除・変更した実施計画・達成目標

### ① ゲノム横断的関連情報表示サイトの開発

開発対象とする DB として、2016 年度末の計画立案時当初は第二期まで開発を行っていた PGDBj をベースとして新たな DB を構築することを予定していたが、PGDBj では対象としていな

かったゲノム配列・遺伝子情報を取り扱うことから、データベース構造の抜本的な変革が必要であり、協議の結果全く新しい DB を新たに構築することとした。新しい DB の名称は「Plant GARDEN」とし、PGDBj の開発は停止して 2016 年度末の状態での維持・公開することとした。

#### ② 種を超えたゲノム情報統合のためのデータリンク基盤の構築

2017 年度～2019 年度までは 各植物種の遺伝子配列から新しい植物種と配列を随時追加してアミノ酸配列の相同性検索を実施し、類似度指標を算出、蓄積された類似度に基づく遺伝子クラスタリングを実施する方法により、種を超えたゲノム情報統合のためのデータリンク基盤を構築することで開発を進めていた。本開発は大阪大学が主体となって実施していたが、主たる共同研究者の大学異動と研究環境の変化のため、移動先の東京大学において同規模で開発を続けることが困難となった。そこで、本項目における開発方針を変更し、生物のオルソログ情報を格納している OrthoDB(<https://www.orthodb.org/>)の情報を利用し、Plant GARDEN に登録する遺伝子配列と OrthoDB が提供している gene ID と ortholog group unique ID を紐づけることにより、これらの ID を介して種横断的な遺伝子配列比較を行うこととした。

#### ③ カスタム型多型・ハプロタイプ検出システムの構築

当初は Plant GARDEN に格納されているゲノム配列に対し、ユーザー自らが有するリシーケンスデータを比較することにより、SNP 多型と検出された多型をもとにハプロタイプを検出するツール開発を予定していた。一方、研究開始後に、計算環境を丸ごと1つのパッケージに入れてコンテナ化する情報解析技術の革新が進んだ。このことから、Plant GARDEN 側がサーバーを設置してそこにユーザーがデータを投げ込んで計算をするのではなく、計算環境を整えたコンテナを作成、配布することがよりユーザーの利便性を高めると考えた。さらに、単に SNP 解析やハプロタイプ解析といった単体の解析パッケージを提供するのではなく、Plant GARDEN と同様の WUI でローカルデータと Plant GARDEN 上のデータを比較できるシステムを構築することが、将来的により良い情報活用環境をユーザーに提供できると考えた。そこで、ツール開発のうち、ハプロタイプ検出ツールの開発は実施しないこととし、かわりにローカル環境で動作するアプリケーションでユーザー自身が保有するデータと Plant GARDEN 内のゲノムデータ等を比較解析できるツール(開発名・My GARDEN)の開発を着手した。

#### ④ PGDBj コンテンツの更新

センテンスキュレーションについては、対象とする DB やキュレーション方法を再検討し、逐次最適な手法に変更した。また、RDF の活用として Plant GARDEN 内の情報と他 DB (MicrobeDB.jp、MetaboBank)をつなげ、RDF 間の連携検索を実施できるシステム開発を行うこととした。

## 4. 実施内容

### (1) 実施内容

#### ① ゲノム横断的関連情報表示サイトの開発

第二期までに構築した植物ゲノムポータルサイト PGDBj に格納されている DNA マーカーと QTL および連鎖地図情報に加え、多数の植物種のゲノム・遺伝子情報も格納した DB として、新たな DB「Plant GARDEN」の開発を行った。2017 年度は「PGDBj2」の仮称のもと、DB のスキーマを設計し、2018 年度に DB のデザインを設計・実装して、2019 年 3 月に β 版(日本語)を構築した。2019 年 6 月には英語の β 版を公開し、ユーザーからの意見を収集して DB を改良

した。2020年7月に正規版を公開(図1)、その後も検索システムの改良や微修正を実施して現在に至っている。



図1. Plant GARDEN のトップページ(<https://plantgarden.jp>)

DB のデザインにあたっては特に「シンプルであること」「動作が直感的に分かりやすいこと」に留意した。これは、ゲノム情報を必要とするユーザーがゲノム解析の専門研究者だけでなく、非専門の研究者、さらには産業界や教育現場など裾野が広がることを想定したためである。そのため、初心者であっても親しみをもって DB を操作できる WUI であることに留意した。

DB で最も重要な機能の1つである検索機能については、Plant GARDEN 用に Elasticsearch を利用した横断検索システム「pges」を開発した。また、Plant GARDEN のリクエストに従って pges を動作させるための API(pges-api)の整備とインデックスの調整を重ね、意図した検索が可能な限り速やかに実行できるようシステムの改良と実装を行った。さらに Elasticsearch による検索を可能とするために、URI を Restful 化した。安全性に関しては、β 版では http を用いていたが、正規版では https への変更を行い SSL へ対応するとともにサーバーの脆弱性評価をクラウド型脆弱性診断ツールである Vaddy を用いて実施し、安全性の強化を図った。

## ② 種を超えたゲノム情報統合のためのデータリンク基盤の構築

2017年度～2019年度は、各植物種の遺伝子配列に新しい植物種やその遺伝子配列を随時追加して翻訳アミノ酸配列の相同性検索を実施して類似度指標を算出し、蓄積された類似度に基づく遺伝子クラスタリングを実施する方法により、超えたゲノム情報統合のためのデータリンク基盤を構築することで開発を進めた。具体的には、新しい植物種と遺伝子配列を逐次追加してアミノ酸配列の相同性検索を実施し、類似度指標の情報を算出して蓄積するとともに、蓄積された類似度に基づく遺伝子クラスタリングの処理に際してバリエーション配列に対応できるように拡張を行った。さらに配列データの既存の類似配列セットへの逐次的な追加方法の開発を進め、対象とする生物種を近縁関係に基づいて複数の部分群に分割し、その部分群ごとに上

記の遺伝子クラスタリングを行って類似配列セットを生成した。生成した類似配列セットごとに配列プロファイルを作成し、配列プロファイルを部分群間で対応付けることにより、全ての生物種を網羅して類似配列をリンクすることが可能となった。

開発した手法は有望であったものの、データ生成にコストがかかり、かつ開発体制の変更が生じたため、協議の結果、2020年度から方針を変えて生物のオルソログ情報を格納している OrthoDB(<https://www.orthodb.org/>)の情報を利用し、Plant GARDEN に登録する遺伝子配列と OrthoDB が提供している gene ID と ortholog group unique ID を紐づけることにより、これらの ID を介して種横断的な遺伝子配列比較を行うこととした。

具体的には、OrthoDB に格納されているオルソログ遺伝子配列をベースとした植物遺伝子の機能アノテーションを実施する際の参照データセットとして KusakiDB を整備し、KusakiDB を用いた遺伝子機能アノテーションプログラム「Hayai-Annotation」を開発・改良した。

最新のアノテーションプログラム Hayai-Annotation ZEN は、検索データベースとして構築した KusakiDB を対象に、InterPro、PFAM、遺伝子オントロジー (分子機能, 生物学的プロセス, 細胞の構成要素) などの情報を検索するための新しい検索エンジンを実装している。一方、KusakiDB は3つの重要なデータベース OrthoDB、UniProt、RefSeq の情報が統合された、植物のタンパク質配列のオルソログデータベースである。種分化に伴い生じた Orthologous Group (OG) に着目し、OG を推定する機能 (OG Assessment) と、その推定のためのパラメータを調節する機能 (OG Management) を備えており、植物ゲノムの遺伝子機能の解析に用いることができる。本システムを用いて OG を比較すると、種や科に特有の OG を見つけたり、1つの OG が複数の機能を持っているかどうかを調べることができる。この機能を用いて、ユーザーが任意の遺伝子に対して類似した配列を有する他種、もしくは他ゲノムの遺伝子情報を抽出できるシステムを開発した。

Plant GARDEN に格納する遺伝子情報はすべて Hayai-Annotation ZEN による再アノテーションが実施されており、類似するオルソログ遺伝子配列が存在する場合はその ID が Plant GARDEN 上で公開されている。また、系統群を指定してオルソログ情報を表示したり、系統群間を移動して表示できるシステムの開発も行った (図2)。

**再帰的な木構造インデックスを実現するデータ構造**

- ↑/↓ で上位/下位の系統群への移動  
Superkingdom, phylum, class, order, family, genus, species
- 隣接する系統群への移動

**配列が含まれている生物種の割合を濃淡で表示**

**上位の系統群へ**  
**下位の系統群へ**

**System**  
**Genus**  
**Family**  
**Order**  
**Species**

**系統群に属する生物種の配列の表示**

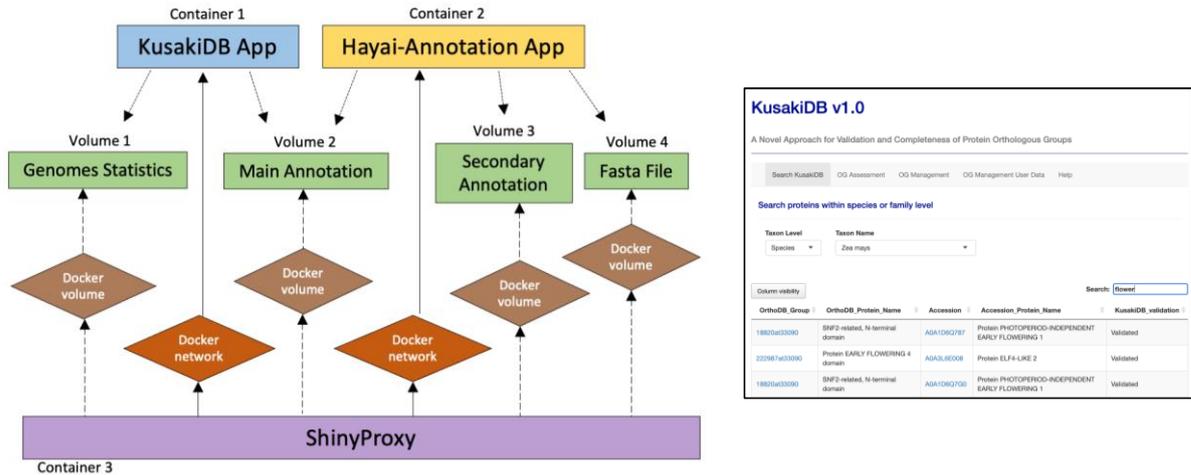
**上位の系統群を経由して別の系統群に移動可能**

図 2. オルソログ配列情報の表示法の開発

### ③ カスタム型多型解析システムの構築

SNP 検出用のサーバーを整備し、セキュリティ対策を実施したのち、解析用パイプラインを組み込み、ユーザーがデータをアップロードして配列データの品質評価、トリミング、リファレンスへのマッピング、SNP の検出、フィルタリングを実施できるシステムを開発して、Plant GARDE N から公開した。さらに、SNP 解析パイプラインのコンテナ化、ならびにユーザー自身が保有するデータと Plant GARDEN 内のゲノムデータ等を比較解析できるツール(開発名・My GARDE N)の開発を実施した。My GARDEN の開発では、My GARDEN 側のコンテナ整備だけでなく、Plant GARDEN が My GARDEN からのリクエストに応じてデータを送るための API の開発も同時に進めている。

また、遺伝子機能アノテーションを高速に実施するツール「Hayai-Annotation」と本ツールを実施する際に用いる KusakiDB に関しては、ShinyProxy を利用してツールのコンテナ化を実現し、ユーザー自身が解析を実施できる体制を整えた。これにより、ユーザーは個人の計算機環境内において、保有する配列データを用いた機能アノテーションを実行できる(図3)。



**KusakiDB v1.0**  
A Novel Approach for Validation and Completeness of Protein Orthologous Groups

Search KusakiDB | OIG Assessment | OIG Management | OIG Management User Data | Help

Search proteins within species or family level

Taxon Level: Species | Taxon Name: Zea mays

Search:

OrthoDB_Group	OrthoDB_Protein_Name	Accession	Accession_Protein_Name	KusakiDB_validation
1802ca33390	SNF2-related, N-terminal domain	ADA1D6Q787	Protein PHOTOPERIOD-INDEPENDENT EARLY FLOWERING 1	Validated
22097a62090	Protein EARLY FLOWERING 4 domain	ADA3L8D008	Protein ELFA-LIKE 2	Validated
1802ca33390	SNF2-related, N-terminal domain	ADA1D6Q755	Protein PHOTOPERIOD-INDEPENDENT EARLY FLOWERING 1	Validated

図3 開発した KusakiDB と Hayai-Annotation のコンテナ

### ④ PGDBj コンテンツの更新

Plant GARDEN 正規版の公開に伴って PGDBj のデータ更新を終了し、本年度は Plant GARDEN のコンテンツ拡充を図った。Plant GARDEN 内のコンテンツ拡充については、ゲノム配列を中心に染色体レベルにアセンブルされており、かつ遺伝子配列情報が付随しているデータを中心にキュレーションと格納を引き続き実施した。その結果、2019年3月のβ版公開当初に格納されたゲノム情報は9種、2020年2月では80種だったのに対し、2021年9月現在で134種169ゲノムの情報を格納できた(表1)。Plant GARDEN は Phytozome (<https://phytozome-next.jgi.doe.gov/>)など主要な植物ゲノムDBと同等の植物種数をカバーしている。また遺伝子ごとの詳細を閲覧できるページやDNAマーカー、SNPs情報など、ゲノムの多型性を検出するための情報も備えている。このことから、データの量・質ともに世界トップクラスの植物ゲノムポータルサイトを確立できたと考える。

表 1 Plant GARDEN に格納されている主なコンテンツ(2021年9月現在)

	植物種数	件数
ゲノム	134	169
遺伝子	126	8,136,812
DNA マーカー	34	313,115
形質と関連しているマーカー(QTL)	24	8,255
SNPs, In/dels (SRA から検出した変異)	105	636,047,952
SRA: Sequence Read Archive	105	996

センテンスキュレーションの検討では、QTL情報のキュレーションとして QTLTableMiner++(QTM, Singh et al, 2018, BMC Bioinformatics)を用いることを決定し、同プログラムをルーチンワークとして実施する体制を整えた。作業の並列化による解析の高速化もすすめ、Plant GARDEN および PGDBj 内に情報を格納した種に対するオートキュレーションを実施した。一方、QTLTableMiner++が対象とする論文 DB は限られていること、ならびにオートキュレーションの結果が必ずしも正確でないことから、マニュアルキュレーションの併用は不可欠であった。そこで、国内外のキュレーターと契約をし、オンラインでキュレーションを実施するシステムを整備した。

本事業においては他プロジェクト等で開発された DB との連携も求められている。そこで、GlyCosmos Portal とのデータリンクを図るとともに、MicrobeDB.jp と MetaboBank 内に格納されているデータと RDF を介して連携検索を実施できるシステムを開発した(図4)。



図 4. RDF を用いたデータ連携の概要(左)と MicrobeDB.jp と Metabobank との RDF の関係図

## (2) 実施内容のうちの特定項目の詳細

### ① 研究コミュニティを含むデータ提供者との連携・協業

本プロジェクト内にアドバイザー委員会を設置し、年2回のアドバイザー委員会を開催した。外部委員は、国内データベース維持管理者およびユーザーとなる植物ゲノム研究者や植物関連団体関係者を含み、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、日本育種学会の各学会から推薦を受けた研究者に外部委員を依頼した。さらに国内の種苗会

社の大部分が加盟する日本種苗協会や、理化学研究所環境資源科学研究センター長、農研機構次世代作物開発センター長にも相談をし、各組織より推薦を受けた有識者も外部アドバイザーに加えた。

② データベース利用者への周知、利用者との連携・協業

データベース利用者への周知として、学会発表を積極的に行った(計 47 件)。また、関係する学会での展示活動を積極的に行った(2021 年 9 月までに 24 件)。特に学会の展示活動では、ユーザーから Plant GARDEN に対する要望を直接聞き取り、開発に反映した。また、主に初心者に向けた DB 利用に関する講習会も 2 回実施し、参加者から好評を得た。さらに、Plant GARDEN 上においてもアンケートサイトを作成し、利用者の要望や不具合を日常的に聞き取れるようにした。

③ 利用者にとって有用なデータ基盤の構築

データベース間の連携に関しては、GlyCosmos Portal に糖鎖に関する遺伝子情報を提供し、相互リンクを作成した。また、Plant GARDEN で公開している植物種のゲノム配列、アノテーション、DNA マーカー、QTL 情報についてこれまで RDF 化を進めてきた。NCBI が提供している PubTator を用いて、各植物種の文献から種名 (Species) と Taxonomy ID、病気 (Disease)、遺伝子 ID (Gene) を含む文章を収集し、DBCLS で公開されている TogoDB に格納することで RDF 化を行った。MicrobeDB.jp と MataboBank に対し、RDF 間の連携検索を実施できるシステム開発を行っている。特に SPARQL 言語による検索の実施が困難なユーザーに対し、MicrobeDB.jp と MataboBank 側とも協働して、SPARQL 言語を使わずに検索できる使い易いウェブインターフェースを開発することで、統合化データベース連携検索システムを構築している。

さらにより多くのユーザーを取り込むための工夫として、特に若い年代は PC よりもタブレットを日常的に用いることを考慮し、タブレット版の開発も実施した。

④ 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

Plant GARDEN に格納するデータの収集においては、可能な限り自動化を進めるためにセンテンスキュレーションによる情報収集法の開発を行った。さらにマニュアルキュレーションについても体制を整備し、国内外のキュレーターと契約をしてオンラインでキュレーションを行うシステムを構築した。

一方、Plant GARDEN の独自コンテンツである SRA 由来の SNP に関しては、SNP 検出を高速に実施することができる専用のサーバー(イルミナ社 Dragen)を導入し、データ蓄積の加速化を図った。

⑤ 人材の育成

本課題では4名の特任研究員を雇用し、DBの開発を進めてきた。いずれも本課題を実施する過程で、データベース構築や遺伝子アノテーション解析など、新たな情報解析の技術を獲得し、プロジェクト期間中に2名は国立研究機関、1名は民間のゲノム情報解析会社に就職した。本課題は若手研究者のキャリアパスを支援する場として機能したと考える。

⑥ 国際連携・国際貢献

地球上の全生物のゲノム配列解析を目的とするプロジェクト Earth BioGenome Project と連携し、Affiliated Project Networksとして登録した。

(<https://www.earthbiogenome.org/affiliated-project-networks>)

- ⑦ その他  
特になし

## §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

表 1 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)

名称	種別	2017 年度	2018 年度	2019 年度	2020 年度	2021 年度 (9月時点)
Plant GARDE N	訪問者数		710	8,573	9,751	6,016
	訪問数		851	14,210	17,091	10,490
	ページ数		3,950	192,979	79,562	63,817
PGDBj	訪問者数	73,740	49,001	42,374	21,582	13,147
	訪問数	81,751	58,617	97,889	54,273	20,715
	ページ数	186,736	231,250	290,751	659,704	88,630

#### (2) 分析

Plant GARDEN は、訪問者数、ページ数ともに毎年度増加している。ベータ版を公開したのは2018年度末であり、正規版公開は2020年7月である。その後もコンテンツを拡充し、また関連学会への出展等でアピールに努めたことが、アクセス数の増加に影響したと考えられる。2021年度は6ヶ月の集計で訪問者数と訪問数は2020年度の60%であり、ページ数は80%となっていることから、今後も順調にアクセス数が増えることが期待できる。

PGDBjは今季においては開発をせず維持のみとなった。Plant GARDEN が開発されてこともあり、訪問数は毎年減少している。一方、コンテンツを更新しないにも関わらず一定のアクセス数があることから、今後も継続する必要があると考えている。

#### 2. データベースの利用状況を示すアクセス数以外の指標

特に該当はない

#### 3. データベースの利活用により得られた研究成果(生命科学研究への波及効果)

Plant GARDEN は本期において開発を開始したため、正規版の公開から1年と3ヶ月程度しか経過していない開発して間もないDBであり、現在のところDBの利活用により得られた研究成果は特に公開されていない。また、PGDBjについても、PGDBjのデータを用いて具体的な研究成果を出したことを明記した論文は公開されていなかった。PGDBjとPlant GARDENはすでに論文等とともに公開されているデータを集約したポータルサイトであり、学術的に新規な発見を行う場合は、データを生成した元の論文を引用することはあっても、Plant GARDENやPGDBjを引用したり謝辞として明記することは考えにくい。そのため、データベースの利活用により得られた研究成果を明示的に示すのは難しい状況にある。

一方、Plant GARDENのアクセスが増加しており、またPGDBjのアクセス数も一定数維持されていることから、植物のゲノム情報を取得するためのポータルサイトとして、生命科学研究に貢献できているものと考えている。

#### 4. データベースの利活用によりもたらされた産業への波及効果や科学技術のイノベーション(産

### 業や科学技術への波及効果)

Plant GARDEN はゲノムや遺伝子解析を行う専門家に対して質の高い情報を提供する一方で、これらの分野を専門としないユーザーもデータにアクセスしやすいよう、シンプルかつ操作が直観的に分かりやすいページデザインを心掛けた。また、出来るだけ多くの植物種の情報を格納することを目標として情報を整備している。その結果、様々な分野のユーザーから分かりやすいデータベースであるとの好評を受けている。Plant GARDEN が今後も植物ゲノム解析の裾野を広げ、アカデミアや産業界のゲノム解析の推進に貢献することが期待される。また、当初からの目標であるが、Plant GARDEN が高校や大学などの高等教育においてゲノムについて学習する際のプラットフォームとなることも目指している。

産業界においては、種苗産業などで利用者が多く学会等の展示会でもこれらの業界の方から利用している旨を伝えていただく機会も多い。一方、産業界においては具体的な製品の開発工程を公に出すことはないので、Plant GARDEN や PGDBj の情報を活用して製品化などに至った事例を把握するのは困難であった。

## §5. 今後の展開

今期で新たに開発した Plant GARDEN は、NGS の普及が進み多くの植物種で参照ゲノム配列が整備される中で、これらのゲノム・遺伝子配列の格納・提供に加えて、ユーザーが有するリード配列を Plant GARDEN に格納しているデータと比較するシステム開発を目指して整備を行った。

格納したデータの規模は Phytozome(<https://phytozome-next.jgi.doe.gov/>)など主要な植物ゲノム DB と同等の植物種数をカバーしているが、開発して間もない DB ということもあり、最新情報の Update が間に合っていない。今後は世界でもっとも植物のゲノム・遺伝子情報が充実した DB として世界に認知される DB となるよう、一層のデータ拡充に努めたい。生物のゲノム解読においては地球上のすべての生物のゲノム解読を目指した研究コンソーシアム Earth Biogenome Project (<https://www.earthbiogenome.org/>)が立ち上がり、世界の大型ゲノム解読プロジェクトがこのコンソーシアムに連なっている。Plant GARDEN も affiliated project として本コンソーシアムに参加している(Lewin et al, PNAS, 2021, accepted)。今後は世界の研究コミュニティとも連携して、Plant GARDEN の認知を広めたい。

一方、Plant GARDEN 内のデータとユーザーのデータを比較するシステム開発は、SNP 検出などのツール整備は行ったものの、理想とするスタイルの完成までには至らなかった。ユーザーのデータ比較については、計画時当初から開発を行うこととしていたが、利用できる技術やサービスが本事業の実施期間中に大きく変化した。すなわち、クラウドサービスの普及やコンテナ化技術の発展、WUI や API 開発に利用できるライブラリ等の充実である。そこで 2019 年度には追加予算もいただいて、これらを利用したユーザーのデータ比較システムを開始した(開発名: My GARDEN)。このシステムは、ユーザーがローカル PC 等に置く Web アプリケーションで、API により Plant GARDEN 内のデータを呼び出す。さらにこのアプリケーションとローカル PC、もしくは任意のクラウド上に置いたツールを連携させることで、ユーザー自身が意識せずに、異なるサーバー上に置かれたツールやデータを用いて、Plant GARDEN 内のデータとユーザーのデータを比較するシステムである。今期において My GARDEN 側のコンテナ整備や API の開発を実施しているところであるが、相当の開発資源を有する内容のためプロトタイプ開発までは終了の見込みであるが、今期中にコンテナをユーザーに配布できるまでに至らなかった。そこで、可能であれば My GARDEN の開発を継続し、これまででない新しいシステムをユーザーに提供したい。また、ゲノム情報だけでなく、画像情報など多くの大規模データをユーザーが保有する時代となり、ユーザーのデータを格納するプライベートな DB システムを開発する動きも広がりつつある。今後はこのような個人の DB とも連携できるシステムの開発を行っていきたい。

Plant GARDEN のもう 1 つの大きな特徴は、格納されている遺伝子配列が全て同一ツール(Hayai-Annotation)により再アノテーションされていることである。これにより、遺伝子アノテーションの品質とアノテーション項目が統一されたため、配列だけでなく遺伝子機能からも種を超えた遺伝子の比較を行うことが可能となった。各オルソログ配列に紐づく遺伝子の検索については今期中にプロトタイプの実装を終了する予定であるが、Plant GARDEN 上で実運用するには追加の修正作業等が必要になると考えている。オルソログ配列に基づく種を超えた遺伝子比較は、植物の研究において需要の高い解析である一方で、更新された最新のゲノム情報を用いて比較・検索できる DB の整備は進んでいない。本機能の充実が植物研究に大きく貢献すると期待しており、今後も機能の充実とより閲覧性の高い検索結果の表示に取り組むたい。

さらに今期においては、形質関連座として Plant GARDEN に掲載したのは QTL 解析などの遺伝解析で形質との関連が確認された座情報のみであった。一方、遺伝子に関しては、機能アノテ

ーションだけでなく、実験により機能が確かめられた遺伝子情報が多数存在する。今期はこれらの情報をキュレーションにより整備するまでにとどまったが、今後は機能が確認された遺伝子情報も形質座情報として Plant GARDEN に格納する。これは、正遺伝学と逆遺伝学のそれぞれの手法で機能との連関が確認された配列情報を同一 DB 上で比較できることを意味しており、生物研究への貢献は大きいと考える。また、研究開発開始当初にアドバイザー委員会では要望の高かった、ゲノム以外の配列や解析情報と Plant GARDEN 内の情報との比較、すなわち、RNA-Seq, Chip-Seq といった遺伝子発現解析や GWAS 解析といった遺伝解析と Plant GARDEN 内の遺伝子配列情報との比較については、今期は取り組むことができなかった。これらについては、技術的なアプローチを整理しつつ、実現の可能性について引き続き検討したい。

Plant GARDEN は専門の研究者にとって価値のある充実した情報を格納しつつ、専門性にとらわれない多くのユーザーにゲノム情報にアクセスしてもらうことを目指し、シンプルかつ直感的に分かりやすい操作性を意識してインターフェースを開発した。2021 年 9 月の時点で PC 版のみを公開しているが、タブレット版の開発も行っている。タブレット版の開発は、PC よりもタブレットを日常的に使い慣れている学生やポスドク研究者などの若年層の普及を狙っている。さらに Plant GARDEN の利用場面を研究開発だけでなく教育にも広げていきたいと考えており、タブレット版が新たな分野のユーザー層獲得につながることを期待している。今期においては、まずは Plant GARDEN を世に出すことを実現するために開発作業そのものに注力し、またコロナ感染拡大も伴ったことから、講習会を積極的に開催することはなかった。今後は Plant GARDEN の一層の充実に加えて、講習会を行ったり、使い方動画を充実するなどして、より多くの方に Plant GARDEN に触れてもらうための活動に取り組みたい。

## §6. 自己評価

本研究開発は、「個体ゲノム時代に向けた植物ゲノム情報解析基盤の構築」という課題名のもと、ゲノム解析を取り巻く技術の大きな変化に対応するために、第二期までに開発した DB「PGDBj」の開発を一旦終了し、PGDBj で格納していたコンテンツに加えてゲノム・遺伝子情報を格納する新しい DB を構築すると決断し、それを実行したものである。開発開始から 2 年間で β 版を構築し、3 年 4 ヶ月で正規版を公開、本年度は年間の訪問数が 2 万となる予定である。格納した植物種は 130 種を超え、また扱うデータの種類もゲノム、遺伝子、DNA マーカー、形質連関座、SNPs など多様である。このように多種多様なデータを格納する DB を短期間に開発、公開できたことは評価に値すると考えている。

一方で、今後の展開のところでも述べたように、目標の一つであった Plant GARDEN 内の情報とユーザーの情報を比較できるシステム開発については、目標もしくは理想とするシステムをユーザーに対して提示できるまでには至らなかった。このことは、担当者の努力が不十分であったというよりは Plant GARDEN 本体の開発に想定以上に労力を要したことが原因であったと考える。しかし、ユーザーのデータと Plant GARDEN 内の情報比較を行う需要はますます高まっていると感じており、本事業の中で構想を実現するための技術的なアプローチを検討し、プロトタイプを作成できたことは、Plant GARDEN の今後の展開を考える上で大きな価値があったと考えている。

統合化推進プログラムの趣旨のうち、「わが国の生命科学研究等によって産出された研究データを広く収集するデータベースを対象とし、より多くの多様な研究者にとってより価値のあるものへと発展させる」については、各研究者が整備したゲノム情報を収集・公開する基盤を構築できたと考えている。現在は人員不足で手が回らないため積極的な呼びかけはしていないが、アクセス制限付きの DB を Plant GARDEN で準備し、論文公開前から各研究者が Plant GARDEN にゲノム情報を格納できる体制の準備も整えている。論文公開後は速やかに Plant GARDEN の本体に情報を移行することで、個々では構築や管理が難しいゲノム情報を離散させることなく管理する基盤を整えた。

また、「データ提供者、データ利用者との緊密な連携・協業」については、学会等で積極的に発表・展示を行ったり、アドバイザー委員会を開催することで、その実現に努めた。「単独の研究からは得ることのできない関連分野の有用情報を発見し、また公開データを用いた大規模解析によって新たな知見を見いだす事が容易な情報基盤の確立」に関しては、種を超えた類似の遺伝子検索を、オルソログ情報を基軸として実行できる環境を整えたこと、また、公開されている SRA 情報を用いて多数の系統の多型情報を提供したことにより、実現できたと考えている。

## §7. 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)	1 件
未発行論文	国内(和文)	0 件
	国際(欧文)	0 件

#### (2) 論文詳細情報

1. Ghelfi A, Shirasawa K, Hirakawa H, Isobe S, Hayai-Annotation Plants: an ultra-fast and comprehensive gene annotation system in plants, *Bioinformatics* 35(21) 4427-4442, 2019 (DOI: 10.1093/bioinformatics/btz380)

### 2. その他の著作物(総説、書籍など)

1. 該当なし

### 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	8 件
	国際	2 件
口頭発表	国内	11 件
	国際	3 件
ポスター発表	国内	18 件
	国際	5 件

#### (2) 招待講演

〈国内〉

1. 磯部祥子. 大規模データ時代に育種が向かう先は? ~遺伝解析からデータ科学へ~. 千葉県バイオ・ライフサイエンスネットワーク会議. 2019 年 5 月
2. Isobe S. Plant GARDEN: a portal site for plant genome and DNA marker information. データ駆動型育種のためのデータ管理システム: 現状と展望. 2019 年 12 月
3. 磯部祥子. 植物ゲノムポータルサイト・PGDBj と Plant GARDEN の紹介第 9 回生命医薬情報学連合大会 (IIBMP2020) ワークショップ, Web 開催、2020 年 9 月
4. 磯部祥子. 植物ゲノム情報統合ポータルサイト Plant GARDEN の構築、第 43 回日本分子生物学会年会 生命科学のデータベース活用法 2020、Web 開催、2020 年 12 月
5. 磯部祥子. 植物ゲノムポータルサイト・Plant GARDEN の使い方、AJACS オンライン 2,

必要な配列情報を入手する、Web 開催、2020 年 10 月

6. 磯部祥子. ゲノム解析から植物の多様性を考えるバーチャル研究会 生物多様性の DNA 情報学~自然の計測と生命の理解のために. 2020 年
7. 磯部祥子、大規模データ時代の DB の役割は? ~Plant GARDEN の開発現場から~ 植物学会シンポジウム、Web 開催、2021 年 9 月 18 日
8. 市原寿子、平川英樹、Ghelfi, Andrea、小原光代、山田学、田村卓郎、中谷明弘、中村保一、白澤沙智子、杉原英志、田畑哲之、磯部祥子、植物ゲノムポータルサイト「Plant GARDEN」の使い方第 44 回日本分子生物学会年会 生命科学のデータベース活用法 2021、横浜、2021 年 12 月(予定)

〈国際〉

1. Isobe S. Kazusa DNA Research Institute Data Base construction. GOBii workshop, Tokyo University, 2019 年 6 月
2. Isobe S. "PlantGarden: a database for assisting plant breeding and genetics, Data management systems for data-driven breeding: the current status and prospect, TKP Ochanomizu Conference Center, 2019 年 12 月

### (3) 口頭講演

〈国内〉

1. Ghelfi A, Antezana E, Interaction between Genotype and Phenotype data, Functional Gene Annotation and Plant Breeding Ontology, 11th NBDC/DBCLS BioHackathon, Matsue, Shimane, 2018 年 12 月
2. 原田大士朗、市原寿子、中谷明弘、ジェルフィアンドレア、藤代継一、小原光代、平川英樹、田畑哲之、磯部祥子、植物ゲノム情報ポータルサイト・PlantGARDEN の開発と植物ゲノム解析の現状、日本育種学会第 134 回講演会、岡山大学(岡山)、2018 年 9 月
3. 原田大士朗、市原寿子、中谷明弘、ジェルフィアンドレア、藤代継一、小原光代、平川英樹、田畑哲之、磯部祥子、植物ゲノム情報ポータルサイト PlantGARDEN の開発、日本育種学会第 135 回講演会、千葉大学(千葉)、2019 年 3 月
4. 平川英樹、原田大士朗、Ghelfi Andrea、Fawcett Jeffrey、白澤沙知子、市原寿子、中谷明弘、磯部祥子、田畑哲之、植物ゲノム情報統合ポータルサイト Plant GARDEN の構築、第 41 回日本分子生物学会、横浜、2018 年 11 月
5. 平川英樹、品種間ゲノムワイド多型情報の収集および Plant GARDEN の構築、園芸学会平成 30 年度秋季大会イルミナランチョンセミナー、鹿児島、2018 年 9 月 22-24 日
6. 平川英樹、育種への利用を目指したゲノムデータベースの開発、園芸学会平成 31 年度春季大会小集会、東京、2019 年 3 月
7. 原田 大士朗, 市原 寿子, 中谷 明弘, ジェル フィ アンドレア, 藤代 継一, 小原 光代, 平川 英樹, 田畑 哲之, 磯部 祥子, 植物ゲノム統合ポータルサイト PlantGARDEN の開発と公開、第 37 回日本植物分子生物学会大会、京都、2019 年 9 月
8. 原田 大士朗, 市原 寿子, 中谷 明弘, ジェル フィ アンドレア, 藤代 継一, 小原 光代, 平川 英樹, 田畑 哲之, 磯部 祥子, 植物ゲノム情報ポータルサイト PlantGARDEN の 拡張、第 136 回育種学会秋季大会、奈良、2019 年 9 月
9. 原田 大士朗, 市原 寿子, 中谷 明弘, ジェル フィ アンドレア, 藤代 継一, 小原 光

- 代, 平川 英樹, 田畑 哲之, 磯部 祥子, 植物ゲノム情報ポータルサイト Plant GARDEN の拡張. 令和元年園芸学会秋季大会. 2019 年 9 月
10. 市原 寿子, 原田 大士朗, ジェルフィ アンドレア, 小原 光代, 山田 学, 白澤 沙知子, フォーセット ジェフリー, 田村 卓郎, 杉原 英志, 中谷 明弘, 中村 保一, 平川 英樹, 田畑 哲之, 磯部 祥子. 植物ゲノム情報ポータルサイト Plant GARDEN の正規版公開. 第 138 回育種学会講演会. Web 開催, 2020 年 10 月
  11. 市原 寿子, 小原 光代, 山田 学, ジェルフィ アンドレア, 平川 英樹, 白澤 沙知子, 田村 卓郎, 杉原 英志, 中村 保一, 中谷 明弘, 田畑 哲之, 磯部 祥子. 植物ゲノム情報ポータルサイト Plant GARDEN の改訂(2020 年度・第 4 四半期版). 第 139 回育種学会講演会. Web 開催, 2021 年 3 月

〈国際〉

1. Ghelfi A, Hayai-Annotation: An Ultra-Fast and Comprehensive Gene Annotation System in Plants', PAG ASIA 2018, Seoul, 2018 年 5 月
2. Ghelfi A, Hayai-Annotation: An Ultra-Fast and Comprehensive Gene Annotation System in Plants, Argonne National Laboratory, Chicago, 2018 年 7 月
3. Ichihara, H, Hirakawa, H, Ghelfi A, Kohara M, Yamada M, Tamura T, Nakaya A, Nakamura Y, Shirasawa S, Sugihara E, Tabata S, and Isobe S. Plant GARDEN: a portal web site for accessing plant genome, DNA marker and SNP information. Biotech 2021, Web 開催. 2021 年 7 月

**(4) ポスター発表**

〈国内〉

1. 市原 寿子, 菊地 正隆, 長谷川 舞衣, 小原 光代, 平川 英樹, 磯部 祥子, 田畑 哲之, 中谷 明弘、種を超えたゲノム情報のためのデータリンク基盤の構築、トーゴの日シンポジウム 2017、東京大学弥生講堂、10 月
2. 市原 寿子, 菊地 正隆, 長谷川 舞衣, 小原 光代, 平川 英樹, 磯部 祥子, 田畑 哲之, 中谷 明弘、Construction of a data link system for integrating genomic information across species、第 40 回日本分子生物学会年会、神戸国際展示場、12 月
3. 市原 寿子, 原田大士朗, Jeffrey Fawcett, 白澤沙知子, 小原 光代, 菊地 正隆, 長谷川 舞衣, 平川 英樹, 磯部 祥子, 田畑 哲之, 中谷 明弘、種を超えたゲノム情報統合のためのデータリンク基盤の構築、トーゴの日シンポジウム 2018、日本科学未来館、10 月
4. 市原 寿子, 原田大士朗, Jeffrey Fawcett, 白澤沙知子, 小原 光代, 菊地 正隆, 長谷川 舞衣, 平川 英樹, 磯部 祥子, 田畑 哲之, 中谷 明弘、種を超えたゲノム情報統合のためのデータリンク基盤の構築、第 41 回日本分子生物学会年会、パシフィコ横浜、11 月
5. Ghelfi A, Shirasawa K, Hirakawa H, Isobe S, Hayai-Annotation: five levels of high-accurate ultra-fast gene annotation in plants, 第 134 回講演会(岡山), 2018 年 9 月
6. 原田大士朗、市原寿子、中谷明弘、ジェルフィアンドレア、藤代継一、小原光代、平川英樹、田畑哲之、磯部祥子、世界における植物ゲノム解析の現状と課題、トーゴの日シン

ポジウム 2018、日本科学未来館(東京)、10 月

7. 原田大士朗、市原寿子、中谷明弘、ジェルフィアンドレア、藤代継一、小原光代、平川英樹、田畑哲之、磯部祥子、植物ゲノム情報ポータルサイト・PlantGARDEN の開発にむけて、第 41 回日本分子生物学科年会、パシフィコ横浜(神奈川)、2018 年 11 月
8. 平川英樹、原田大士朗、Andrea Ghelfi、Jeffrey Fawcett、白澤沙知子、市原寿子、中谷明弘、磯部祥子、田畑哲之、植物ゲノム統合ポータルサイト Plant GARDEN の構築、トーゴの日シンポジウム 2018、東京、2018 年 10 月
9. 市原寿子、磯部祥子、平川英樹、原田大士朗、Ghelfi Andrea、小原光代、山田学、白澤沙知子、中村保一、田村卓郎、杉原英志、田畑哲之、中谷明弘、種を超えたゲノム情報統合のためのデータリンク基盤の構築、トーゴの日シンポジウム 2019、日本科学未来館、10 月
10. 平川 英樹、藤沢 貴智、森宙史、ジェルフィ アンドレア、市原 寿子、中村 保一、磯部 祥子、田畑 哲之、黒川 颯. 生物間相互作用の解明に向けた植物・微生物統合データベース間の連携. トーゴの日シンポジウム 2020. Web 開催、2020 年 10 月
11. ジェルフィ アンドレア、中村 保一、磯部 祥子. KusakiDB v1.0: New protein database of orthologous genes in plant species. トーゴの日シンポジウム 2020. Web 開催、2020 年 10 月
12. 市原 寿子、原田 大士朗、ジェルフィ アンドレア、小原 光代、山田 学、白澤 沙知子、田村 卓郎、杉原 英志、中谷 明弘、平川 英樹、田畑 哲之、磯部 祥子. 植物ゲノム情報ポータルサイト「Plant GARDEN」の改訂(2021年度・第2四半期版). トーゴの日シンポジウム 2020. Web 開催、2020 年 10 月
13. Ichihara Hisako, Harada Daijiro, Ghelfi Andrea, Kohara Mitsuyo, Yamada Manabu, Shirasawa Sachiko, Fawcett Jeffrey, Tamura Takuro, Sugihara Eiji, Nakaya Akihiko, Nakamura Yasukazu, Hirakawa Hideki, Tabata Satoshi, Isobe Sachiko. Release of the official version of Plant GARDEN, a plant genome information portal site. 第 43 回日本分子生物学会. Web 開催、2020 年 12 月
14. Ghelfi Andrea, Nakamura Yasukazu, Isobe Sachiko. KUSAKIDB: a database of orthologous proteins specialized in plant species. バーチャル研究会 生物多様性の DNA 情報学~自然の計測と生命の理解のために. Web 開催、2020 年 12 月
15. 市原 寿子、小原 光代、山田 学、ジェルフィ アンドレア、平川英樹、白澤 沙知子、田村 卓郎、杉原 英志、中村 保一、中谷 明弘、田畑 哲之、磯部 祥子. 植物ゲノム情報ポータルサイト Plant GARDEN 正規版の現状. 超分野植物科学研究会第1回研究会 2020. Web 開催、2021 年 6 月
16. 平川 英樹、藤沢 貴智、森宙史、ジェルフィ アンドレア、市原 寿子、中村 保一、磯部 祥子、田畑 哲之、黒川 颯. 生物間相互作用の解明に向けた植物・微生物統合データベース間の連携. 超分野植物科学研究会第1回研究会 2021. Web 開催、2021 年 6 月
17. 市原 寿子、原田 大士朗、ジェルフィ アンドレア、小原 光代、山田 学、白澤 沙知子、田村 卓郎、杉原 英志、中村 保一、中谷 明弘、平川英樹、田畑 哲之、磯部 祥子. 植物ゲノム情報ポータルサイト「Plant GARDEN」の改訂(2021年度・第2四半期版). トーゴの日シンポジウム 2020. Web 開催、2021 年 10 月(予定)
18. 平川 英樹、藤沢 貴智、長崎英樹、森宙史、福島篤史、ジェルフィ アンドレア、市原 寿子、中村 保一、金谷重彦、有田正規、黒川 颯、磯部 祥子、田畑 哲之. 植物ゲノム統

合データベース Plant GARDEN における微生物、メタボローム統合データベースとの連携. トーゴの日シンポジウム 2021. Web 開催、2021 年 10 月 (予定)

〈国際〉

1. Developing a Plant Genome Database PGDBj in the NGS Era, Fawcett JA, Hirakawa H, Kohara M, Ghelfi A, Fujishiro T, Ichihara H, Nakaya A, Isobe S, Tabata S, Animal Genome XXVI Conference, P0022, San Diego, California, 2018 年 1 月
2. Ghelfi A, Shirasawa K, Isobe S, Hosokawa, Development of a New Pipeline for Haplotype-Specific Expression: case-study in F1 reciprocal cross in pepper, International Society for Computational Biology, Chicago, 2018 年 7 月
3. Harada D, Ichihara H, Nakaya A, Ghelfi A, Fujishiro T, Kohara M, Hirakawa H, Tabata S, Isobe S, Development of plant a genome portal site, Plant GARDEN, PAG Asia 2019, 中国・深圳、2019 年 6 月
4. Harada D, Ichihara H, Nakaya A, Ghelfi A, Hirakawa H, Tabata S, Sachiko I. Development of Plant Genome Portal Site PlantGARDEN. Plant and Animal Genome XXVIII, PE0190, San Diego, USA, 2020 年 1 月
5. Nakamura Y, Shirasawa S, Sugihara E, Tabata S, and Isobe S. Plant GARDEN: a portal web site for accessing plant genome, DNA marker and SNP information. Biodiversity Genomics 21, Web 開催. 2021 年 9 月

4. 知財出願

(1) 出願件数

種別	件数
特許出願	国内 0 件
	国外 0 件
その他の知的財産出願	0 件

(2) 一覧

① 国内出願

1. 該当なし

② 海外出願

1. 該当なし

③ その他の知的財産権

1. 該当なし

5. 受賞・報道等

(1) 受賞

1. 該当なし

(2) メディア報道

1. 該当なし

### (3) その他

#### 1. 学会広報活動

〈国内〉

- 1) 第 35 回 日本植物細胞分子生物学会(埼玉)、2017 年 8 月 29～31 日
- 2) 日本育種学会 第 132 回講演会(盛岡)、2017 年 10 月 7～8 日
- 3) 2017 年生命科学系学会合同年次大会(ConBio2017)(神戸)、2017 年 12 月 6～9 日
- 4) 日本育種学会 第 133 回講演会(福岡)、2018 年 3 月 25～26 日
- 5) 第 59 回 日本植物生理学会年会(北海道)、2018 年 3 月 28～30 日
- 6) 第 36 回 日本植物細胞分子生物学会(金沢)、2018 年 8 月 26～28 日
- 7) 日本育種学会秋季大会 第 134 回講演会(岡山)、2018 年 9 月 22～23 日
- 8) 園芸学会平成 30 年度秋季大会(鹿児島)、2018 年 9 月 22～23 日
- 9) 第 41 回 日本分子生物学会年会(横浜)、2018 年 11 月 28～30 日
- 10) 第 60 回 日本植物生理学会年会(名古屋)、2019 年 3 月 13～15 日
- 11) 日本育種学会春季大会 第 135 回講演会(千葉)、2019 年 3 月 16～17 日
- 12) 園芸学会平成 31 年度春季大会(川崎)、2019 年 3 月 23～24 日
- 13) 第 138 回日本育種学会(オンライン) 2020 年 10 月 10 日～11 日)
- 14) 第 43 回日本分子生物学会年会(オンライン) 2020 年 12 月 2 日～4 日
- 15) 第 61 回日本植物生理学会年会(オンライン) 2020 年 3 月 14 日～16 日
- 16) 令和 3 年園芸学会春季大会(オンライン) 2021 年 3 月 24 日～30 日
- 17) 第 139 回日本育種学会(オンライン) 2021 年 3 月 19 日～21 日
- 18) 超分野植物科学研究会第1回研究会 2021. Web 開催 2021 年 6 月 3 日～4 日
- 19) 第 38 回日本植物バイオテクノロジー学会(つくば) 大会 2021 年 9 月 9 日～11 日
- 20) 令和 3 年園芸学会秋季大会(オンライン) 2021 年 9 月 8 日～14 日
- 21) 第 140 回日本育種学会(オンライン) 2021 年 9 月 23 日～25 日

〈国際〉

- 1) Plant & Animal Genome XXVI (米国、サンディエゴ)、2018 年 1 月 13 日～17 日
- 2) PAG ASIA 2019 (中国、深圳)、2019 年 6 月 6 日～8 日
- 3) Plant & Animal Genome XXVII (米国、サンディエゴ)、2019 年 1 月 11 日～15 日

#### 2. イルミナ DRAGEN 広告 You Tube 2021 年 7 月

<https://www.youtube.com/watch?v=OMT1IcVNGqs>

公開から 1 ヶ月で約 1 万 5000 回の視聴

## §8. 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
2017年 10月4日	担当者ミーティング	東京大学	7人	プロジェクトの進捗報告と協議を行う担当者会議。トーゴの日にあわせて実施した。
2018年 1月26日	担当者ミーティング	東京ステーションコンファレンス	7人	プロジェクトの進捗報告と協議を行う担当者会議
2018年 5月17日	担当者ミーティング(非公開)	TKP 品川カンファレンスセンター	10人	研究進捗報告と今後の進め方の協議のためのミーティング
2019年 4月15日	担当者ミーティング(非公開)	TKP 品川カンファレンスセンター	9人	研究進捗報告と今後の進め方の協議のためのミーティング
2020年度 より随時	担当者ミーティング(非公開)	随時		関係者が一同に会してのミーティングは開催しなかったが、検索システム、オルソログ情報表示など解決問題に応じて、随時 web ミーティングを実施した。

### 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

年月日	名称	場所	参加人数	目的・概要
2018年 3月9日	植物ゲノム情報活用に向けたデータベース利用法	フクラシア東京ステーション会議室	85人	PGDBj ユーザーの獲得増加を目指し、PGDBj、DDBJ の利用法とゲノム配列・マーカー解析法の講習会を実施した。
2019年12 月	Data management systems for data-driven breeding: the current status and prospect	TKP 御茶ノ水カンファレンスセンター	80人	データ駆動型システムの開発に向けた世界的な動向を把握し、かつ、開発に向けた世界的な協働の道を探ることを目的として、こうしたシステム開発を先導している研究者に現況と展望について話題提供をいただいた。

以上

## 別紙 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 RDP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)		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。横断検索を用いることで植物に特化した多種類のDBへ効率的にアクセスできる。	<a href="http://pgdbj.jp">http://pgdbj.jp</a>	2012/8/20	継続・発展	データベース等	対象外	対象外	公開済	
2	Plant GARDEN		植物ゲノム関連情報を格納したポータルサイトである。PGDBjでは格納されていなかった全ゲノム配列情報を基盤に、植物種毎に情報を閲覧できるページを提供している。他に遺伝子配列、アノテーション、PCRベースのDNAマーカー、SNPs、形質関連マーカー等の情報を格納し、ゲノムブラウザ (Jbrowse) 上で横断的に情報を検索することも可能である。また、異なる植物種間で類似する遺伝子配列を検索することも可能である。	<a href="https://plantgarden.jp">https://plantgarden.jp</a>	2020/7/23	新規	データベース等	対象外	対象外	提供前	
3	Hayai-Annotation Plants		植物を対象に遺伝子機能アノテーションを実施するツール。ローカル環境で動作し、実行速度が極めて速く、正確かつ包括的なアノテーションが可能である。	<a href="https://github.com/kdri-genomics/Hayai-Annotation-Plants">https://github.com/kdri-genomics/Hayai-Annotation-Plants</a>	2018/11/20	新規	ツール等	対象外	対象外	対象外	Ghelfi A, Shriasawa K, Hirakawa H, Isobe S. "Hayai-Annotation Plants: an ultra-fast and comprehensive functional gene annotation system in plants", Bioinformatics, Vol135, No21, pp4427-4429, 2019 (DOI 10.1093/bioinformatics/btz380)
4	SNP Detection		配列の精査、マッピングおよび変異検出を実施する解析パイプラインである。Plant GARDENのユーザーを対象に公開中。FTPを通じてユーザーがリード配列とリファレンス配列をかすさDNA研究所のサーバーにUploadして解析を実行する。実行にはユーザー登録が必要である。	<a href="https://pgdbjsnp.kazusa.or.jp/">https://pgdbjsnp.kazusa.or.jp/</a>	2019/3/11	新規	ツール等	対象外	対象外	対象外	
5	Kusaki DB		植物の遺伝子アノテーションを高精度に実施するための遺伝子情報を格納したDB。遺伝子オルソログ情報をOrtho DB, Uniprot, NCBI nrの情報を一元化した。	<a href="http://pgdbjsnp.kazusa.or.jp/app/kusakidb">http://pgdbjsnp.kazusa.or.jp/app/kusakidb</a>	2020/12/1	新規	データベース等	対象外	対象外	対象外	