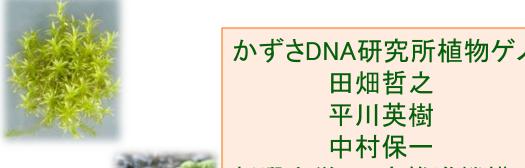


ゲノム情報に基づく植物データベースの統合



かずさDNA研究所植物ゲノム研究部 田畑哲之 平川英樹 中村保一 新潟大学・研究推進機構・超域学術院 中谷明弘



我が国の植物ゲノム研究の現状

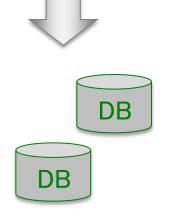
ゲノム・cDNA構造解析

```
単細胞藻類
  クラミドモナス
  シアニディオシゾン
コケ類
  ヒメツリガネゴケ
  ゼニゴケ
シダ類
  イヌカタヒバ
単子葉草本
  イネ
  コムギ
  オオムギ
双子葉草本
  タバコ
  シロイヌナズナ
  ミヤコグサ
  ダイズ
  トマト
  ハクサイ
  キャッサバ
樹木
  ヤトロファ
```

ポプラ

ゲノム機能解析
Transcriptomics
Proteomics
Metabolomics
変異体大規模解析
DNAマーカー
連鎖地図

有用遺伝子発見 遺伝子機能 ネットワーク QTL解析 品種改良



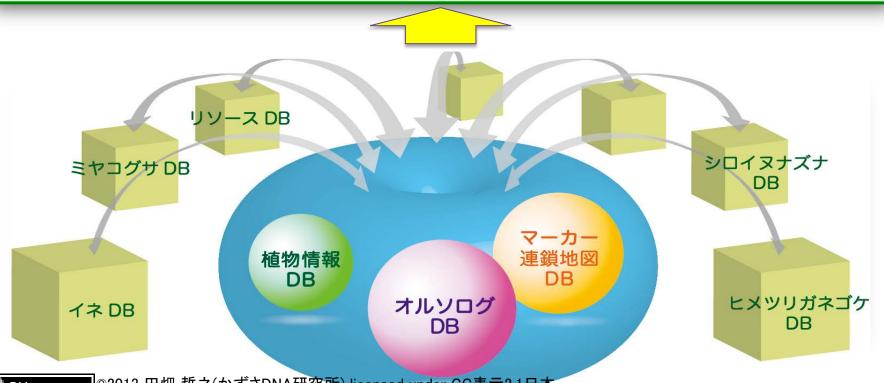


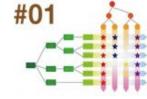


研究開発内容

日本の植物分子遺伝学、分子育種学研究

- ・モデル材料から農作物まで研究対象が多種多様である
- ・ゲノム構造、ゲノム機能、DNAマーカー、有用遺伝子連鎖解析、育種関連 情報など多彩な解析データが生産されDBおよび文献を通して公開されている
- 種々雑多な情報が多くのDBや文献に散在していて利便性が低い
- 情報が分散しているため、国際的なプレゼンスが十分示されない。





H24年度 計画 オルソログDBをハブにしたゲノムDBの統合

- 1. オルソログDBに対象ゲノムDBをリンクする。
 - ・DBエントリを遺伝子IDや配列ホモロジーで対応付け グループ内保有のモデル植物DB / 作物DBシロイヌナズナDB / イネDB / ヒメツリガネゴケDB / ゼニゴケDB・・・
- 2. 異生物種DBエントリ間の対応付けをする。
 - ・オルソログDBを経由してリンクを辿る。
- 3. 統合的検索GUI/API開発とポータルサイトとの連携を行う。
 - ・オルソログDB内の階層構造を反映させた検索の実現
- 4. セマンティックウェブ(RDF)に対応する。



H24年度までの成果 オルソログDBをハブにしたゲノムDBの統合

- 1. オルソログDBの構築と更新
 - ・緑色植物(20種)ラン藻(71種)
- 2. オルソログDBへの対象ゲノムDBのリンク
 - ・アミノ酸/EST配列を取得してblastp/blastxで対応付け



RAP-DB(イネ・農業生物資源研究所)

RARGE(シロイヌナズナ・理化学研究所)

SALAD database (ヒメツリガネゴケ他・農業生物資源研究所)

PHYSCObase(ヒメツリガネゴケ・基礎生物学研究所)

Eucalyptus camaldulensisGenome Database(ユーカリ・かずさDNA研究所) Jatropha Genome Database(ヤトロファ・かずさDNA研究所) Lotus japonicus Genome Sequencing Project(ミヤコグサ・かずさDNA研究

Tomato SBM Database(トマト・かずさDNA研究所)

CloverGarden(クローバー・かずさDNA研究所)

クラミドモナス EST Index (クラミドモナス・かずさDNA研究所)

ミヤコグサ EST Index (ミヤコグサ・かずさDNA研究所)

シロイヌナズナ EST Index (シロイヌナズナ・かずさDNA研究所)

スサビノリ EST Index (スサビノリ・かずさDNA研究所)

MiBASE Micro-Tom Database (トマト・かずさDNA研究所)

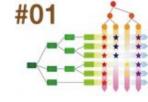
アカクローバー EST Index (アカクローバー・かずさDNA研究所)

シロクローバー EST Index (シロクローバー・かずさDNA研究所)



H24年度までの成果 オルソログDBをハブにしたゲノムDBの統合

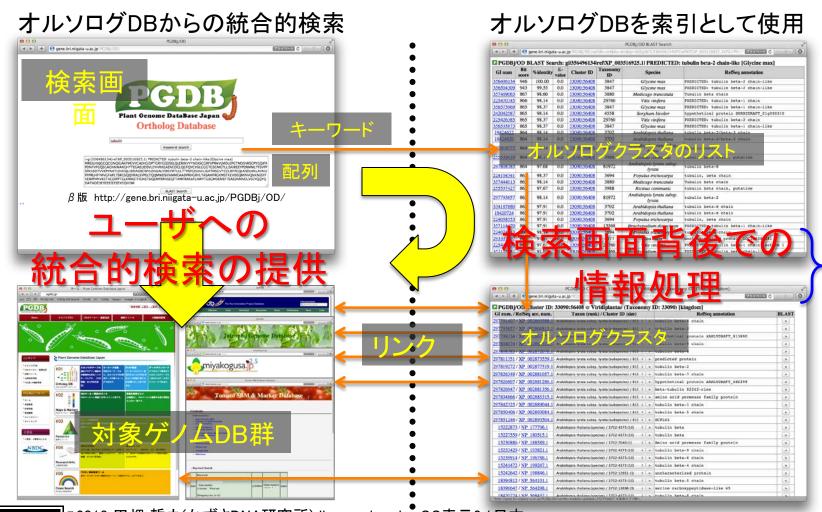
- 1. <u>オルソログDBの構築と更新</u>
 - 緑色植物(20種)ラン藻(71種)
- 2. オルソログDBへの対象ゲノムDBのリンク
 - ・アミノ酸/EST配列を取得してblastp/blastxで対応付け
- 3. <u>異生物種DBエントリ間の対応付け</u>
 - 上項のBLASTの結果をDB化してリンクを辿る
- 4. 統合的検索GUI/API開発
 - MySQL / PHP / JavaScript / Perl
- 5. セマンティックウェブ(RDF)対応
 - 現状では全てのデータをタブ区切りデータ化

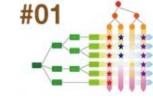


オルソログDBの検索

H24年度 成果内容

オルソログDBへの対象ゲノムDBのリンクと統合的検索





H24年度 成果内容

オルソログDBへの対象ゲノムDBのリンクと統合的検索

1. オルソログDBへ対象ゲノムDBをリンク

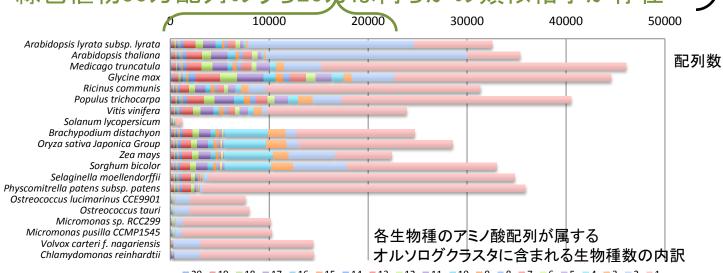
BLAST (E-value $\leq 10^{-5}$)

例)トマト(ITAG2.3) 3.5万配列のうち3.0万がリンク

例)イネ(RAP-DB) 4.2万配列のうち3.7万がリンク

2. オルソログDB内での生物種を跨いだ類似関係

緑色植物50万配列のうち20万は何らかの類似相手が存在



全てをRDBに格納

©2013 田畑 哲之(かずさDNA研究所) licensed under CC表示2.1日本



最終年度 計画

ユーザビリティーの向上

- DBコンテンツの更新
 - ・オルソログDBの更新とリンク対象ゲノムDBの拡充
 - ・後方互換性の維持(再計算した場合のクラスタIDの管理他)
- 2. 統合的検索GUI/API開発とポータルサイトとの連携
 - 検索結果の効果的な表示方法の開発
 - -PGDBi内の他の項目との連携(横断検索)
- 3. 計算機による個別解析用のデータの整備
 - ・ダウンロード用データファイルの整備
 - ・セマンティックウェブ(RDF)への対応
- 4. ドキュメント(マニュアルやDB論文)の整備



H24年度 計画

- 55の植物種を対象として、DBや文献からDNAマーカー、 QTL、連鎖地図に関する情報を収集し、ポータルサイトで 公開する。
- 2. <u>塩基配列地図と連鎖地図がある植物種については、マーカー付随配列をマッピング</u>することで両地図間のマーカーの対応関係を調べる。これらのデータに対して収集したマーカー情報を加えることで、マーカー関連DBを整理する。
- 3. H23年度に構築した<u>地図表示システムの機能の充実</u>を図る。
- 4. 項目(1)~(4)に対して<u>横断検索</u>するためのデータ形式を - 決定する。



H24年度までの成果

- 1. <u>DNAマーカーとQTL</u>については、55の植物種のうち、<u>23種</u> (42%)の情報を収集し、そのうち11種(20%)をポータルサイトで 公開した。地図情報については、<u>16種(29%)の情報を生成・整</u> 備し、そのうち<u>7種(13%)を公開</u>した。
- 2. 地図表示システムに対して、プライマーで増幅されるマーカー を含んだ領域の類似度に基づき<u>複数の地図を関連づける機</u> <u>能を追加開発</u>した。
- 3. 項目(1)~(4)に対する横断検索に必要なデータ形式を決め、 収集した情報をその形式に変換する作業を開始した。



55植物種の マーカー・QTL・地図情報の整備状況リスト(抜粋)

•「科名」項をクリックしてソーティングした状態

. 14	ים זיא ני	. / / /				1127
<u>学名</u>	和名	科名▲	<u>Tax ID</u>	マーカーリスト	CMap 表示	アイコン
Arachis hypogaea	ラッカセイ	マメ科	3818	0	0	
<u>Trifolium</u> pratense	赤クローバー	マメ科	<u>57577</u>	0	0	-4-5
<u>Trifolium</u> <u>repens</u>	白クローバー	マメ科	3899	0	_	***
Glycine max	ダイズ	マメ科	3847	0	0	000
Lotus japonicus	<u> </u>	マメ科	<u>34305</u>	0	0	A DE
<u>Cajanus cajan</u>	<u> </u>	マメ科	3821	1 <u>22</u> 3	<u>122</u> ;	
Medicago truncatula	<u>タルウマゴヤシ</u>	マメ科	3880	_	-	
Citrus unshiu	ウンシュウミカン	ミカン科	55188	準備中	<u>125</u> ;	

マーカー・QTL情報がある 生物種件数

• 収集: 23 種/55 種(42%)

表中の「〇」と「準備中」

-公開:11 種/55 種(20%)

表中の「〇」

・地図情報がある生物種件数

•収集:16 種/55 種(29%)

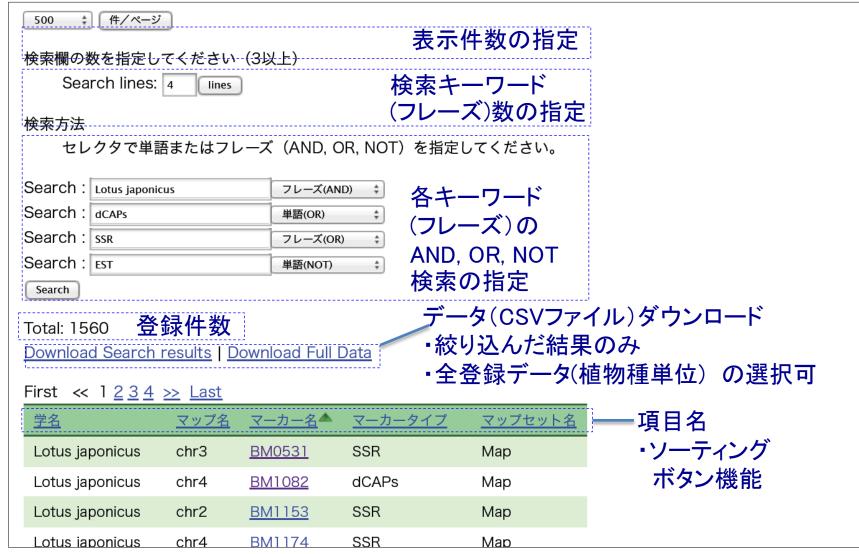
表中の「〇」と「準備中」

- 公開: 7種/55種(13%)

表中の「〇」

ミヤコグサのマーカーリスト(抜粋)

•「マーカー名」項をクリックしてソーティングした状態



マーカー・QTLの掲載情報

Scientific Name		Solanum lycopersicum
Information Category		Genome-SSR 例:SSR —
Marker/QTL (trait)		TGS2757
Мар	population	AMF2(F2: S. lycopersicum 'Ailsa Craig' x S. lycopersicum 'Micro-Tom')
	Linkage group	ch11
	Position	57.131 - 57.131
	population	Expen2000 (F2: S. lycopersicum 'LA925' x S. pennellii 'LA716')
Мар	Linkage group	chll
	Position	67.665 - 67.665
Information		Forward: GTGGCCTAATTTCTTTACGAACG Reverse: CACACACACACACACACACACTCA SSR Fragment: 273 SSR Pattern: AT(mis1) SSR Repeat: 9 SSR PCR: SSR Detection: Enzyme: Related Marker: Related Species: PubMed: [20431859]

Scientific Name		Solanum lycopersicum			
Information Category		OTL 例:QTL —			
Marker/QTL (trait)		la5.1 [7] . Q I L			
	population	S.lycoperisicum x S.pimpinellifolium (P) x S. cheesmania(C) F8			
Мар	Linkage group	Chr05_Chr-P5a			
	Position	34.49 - 34.49			
Мар	population	S.lycoperisicum x S.pimpinellifolium (P) x S. cheesmania(C) F8			
	Linkage group	Chr05_Chr-C5			
	Position	11.01 - 11.01			
Information		Trait Affected: Total Leaf Area(LA)			
		Significant Marker: TG69_620			
		Coverage(cM):			
		PubMed: [18251001]			

改善すべき点:

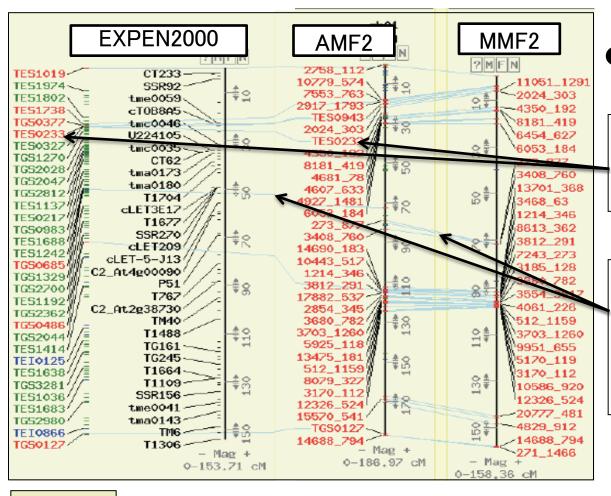
- ◆ データを公開している11種に おいても、収集した項目の一 部が未公開
 - ❖ マーカー情報:
 - 近傍の塩基配列等
 - ❖ QTL情報:

標本数, 年数, 季節, 自然環境 or 人工環境 等



横断検索用のファイルの変換作業時に未公開項目の登録作業を実施する

連鎖地図間(EXPEN2000-AMF2-MMF2)のマーカーの対応図



● 地図間で共通するマーカーの表示

同一色で表示→マーカーの種類が視覚的に捉え難い

マーカー間に線を 引くことで対応付け →線の色を変更して この表示のみにする 方が分かり易い

改善すべき点:

- ◆表示関連の改良
- ◆ 地図情報の登録作業

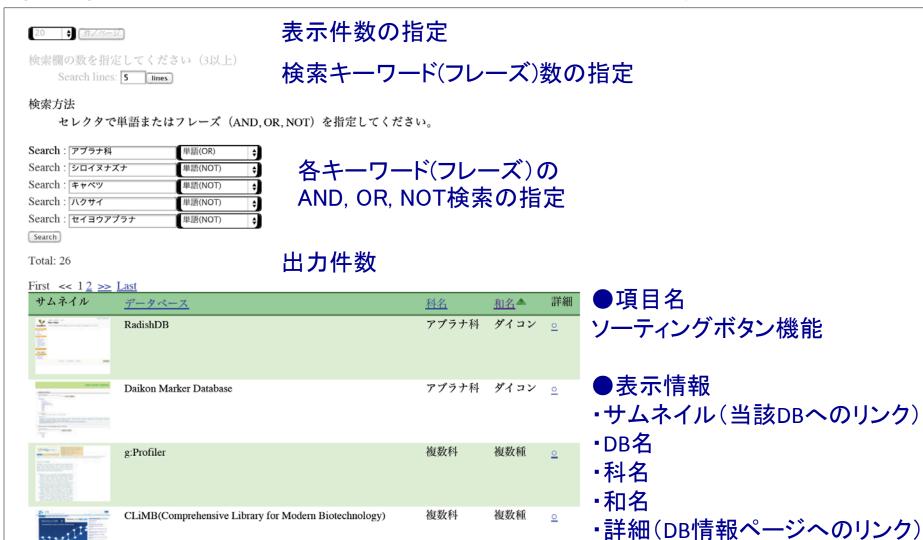
—SNP

-SSR



DBリンク集(抜粋)全73種

•検索機能等のインターフェイスはマーカーリスト同様



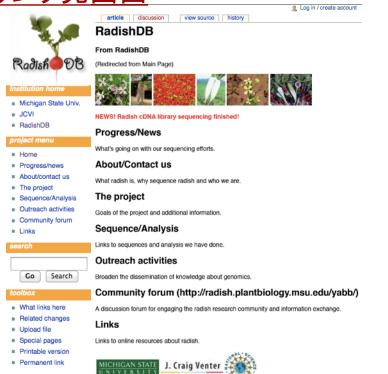
©2013 田畑 哲之(かずさDNA研究所) licensed under CC表示2.1日本



DBリンク登録データ(例: RadishDB)



サムネイル項をクリック



詳細項をクリック

登録データ

データ	D- disk DD		
ベース名	RadishDB		
記載植物	Raphanussativus		
科名	アブラナ科		
Family	Brassicaceae		
Name	Brassicaceae		
和名	ダイコン		
English	Radish		
Name	rausii		
围	USA		
機関	MiciganStateUniversity(ミシガン州立大学)、		
	J.CraigVenterInstitute(J.クレイグベンター研究所)		
URL	http://radish.plantbiology.msu.edu/index.php/Main_Page		
分類	総合サイト,primer配列		
説明	-		

改善すべき点:

©2013 田畑 哲之(かずさDNA研究所) licensed under CCまった。



最終年度計画

- 55の植物種におけるDNAマーカー、QTL、連鎖地図に関する情報の収集と公開を完了させる。
- 2. イネ科などの主要な作物についてはマーカー関連DBに リンクすることで完了させる。
- 3. 地図表示システムの問題点を解決する。
- 4. <u>塩基配列地図と連鎖地図がある植物種のマーカー付随配</u> <u>列をマッピングする</u>ことで生成される地図情報に 基づき植 物DBの統合を行い、完了させる。
- 5. 項目(1)~(4)に対する<u>横断検索</u>に必要なデータへの変換を 第二四半期までに完了する。
- 6. マニュアル等のドキュメントを作成・整備する。

3. 植物リソースの統合



H24年度 計画

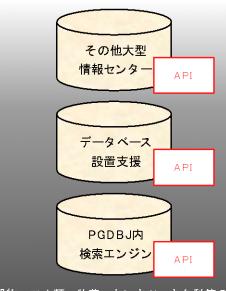
- ❖ 平成23年度までに、NBRP他の リソース情報としてムギ、ダイズ、ミヤコグサ、アサガオ、トマト、コムギ遺伝子を配列類似性によりシロイヌナズナ遺伝子と連結、理研BRCの検索システムSABREで検索可能にし、外部利用可能なAPIを整備する開発を行った。
- ❖ 平成24年度は、横断検索データの更新とともに、新規に有用なバイオリソース情報のデータ収集とID統合を追加し、同時に、既存のPlant Ontology (PO) 等の適用や拡張等を含め、共用化を容易にするためのオントロジーの策定と整備を計画した。





シロイヌナズナ、Thellungiella(環境研究植物)、ハクサイ、トマト、ミヤコグサ、ダイズ、キャッサバ、ストライガ(寄生植物)、アサガオ、コムギ、オオムギ、ポプラ、ヒメツリガネゴケのcDNAリソース情報検索システムの拡張

その他の バイオリソ*ー*ス



穀物、マメ類、牧草、カンキツ、キク科等のゲノム情報基盤を有するバイオリソースの情報の、大型センターとの相互検索系整備や散在データのデータベース化支援や収集を通じた、横断検索用の窓口機能(API)整備による横断検索の実現

各センターが持つこれらの情報を横断的に検索する機能を強化し、わ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。



H24年度 成果



- 理研BRCにて拡張したSABREの情報を検索するフォームをPGDBJ.jp 上に設置し、14植物種のリソース情報を、シロイヌナズナ遺伝子との類似により PGDBJ.jp にて串刺し検索可能とし、検索機能のAPI拡充とUI整備を実施し、検索系の共用化と再利用性を高めた。
- 2. 重要な新規リソース情報として、<u>カンキツ類リソース</u>として近畿大学ならびに佐賀大学より<u>のべ900個体の*Citrus*属在来種</u>と、農研機構・果樹研究所で整備されているカンキツ 類の<u>cDNAクローン+配列情報のデータベース</u>集約を開始した。
- 3. リソース情報を記述するための<u>オントロジ</u>として Plant Ontology, Trait Ontology, Gene Ontology が利用可能な植物種の対応表を作成し、それぞれの語彙のカヴァー程度とその不備等の調査を実施した。



最終年度 計画

- 1. 平成25年度は、平成24年度までに実施したゲノム情報に関連したバイオリソース横断検索システムの拡充と更新を引き続き行い、永続的に運営するためにデータ更新や相互リンクの維持の自動化を目的とした管理システムを構築する。
- 2. <u>新規な有用バイオリソースの情報</u>についてのデータ収集では、 平成24年度に開始した<u>カンキツ類バイオリソース情報のデータ</u> ベース化を推進し公開を完了するとともに、さらに<u>NBRPがサ</u> ポートしていない新規な有用リソース情報収集を行う。
- 3. 本事業のテーマ 1. <u>植物オルソログDB との連携</u>による統合環境の活用により、<u>種々の植物リソースから望みの遺伝子候補を選抜することを可能とするツール</u>の整備を行う。



H24年度 計画

- 1. 昨年度に構築したポータルサイトを整備し、一般公開する。
- 2. <u>55の植物種の学術情報をポータルサイトに掲載</u>する。
- 3. 項目(1)~(3)と連携し、基盤情報ページにリンクを設定する。
- 4. これまでに収集した<u>1,093件の植物ゲノム関連DB情報を分類,整</u> 備し、DBリンク集としてポータルサイトに掲載する。
- 5. 昨年度に続き、国内外DBを調査し、リンク情報を更新する。
- 6. DBのリンク切れと更新のための管理システムを構築する。
- 7. <u>項目(1)~(4)にわたる横断検索機能を実装するための準備</u>を開 始する。
- 8. 植物とゲノム解析手法に精通した<u>キュレーターとアノテータを育</u> 成する。



H24年度までの成果

- 1. <u>ポータルサイトを8月に一般公開</u>し、9月から本格的な 運用を 開始した。
- 2. <u>55の植物種の学術情報をポータルサイトに掲載</u>した。
- 3. 項目(2)の情報リンクを基盤情報トップページに設定した。
- 4. 外部DBのリンク切れと更新のための管理システムを構築した。
- 5. <u>収集した1,093件の植物ゲノム関連DB情報から</u>リンク切れや 重複などの除去作業を実施し、<u>709件についての情報をDBリ</u> ンク集としてポータルサイトに掲載した。
- 6. 項目(1)~(4)にわたる横断検索機能を実装するための準備を 開始した。
- 7. プロジェクト補助員(3名)にキュレータ、アノテータとしての教育を行った。



55植物の基盤情報トップページ(抜粋)

<u>学名</u>	和名	科名▲	<u>Tax ID</u>	マーカーリスト	CMap 表示	アイコン
Arachis hypogaea	<u>ラッカセイ</u>	マメ科	3818	0	0	
<u>Trifolium</u> <u>pratense</u>	赤クローバー	マメ科	<u>57577</u>	0	0	346
Trifolium repens	白クローバー	マメ科	3899	0	_	<u> </u>
Glycine max	<u> 41 x</u>	マメ科	3847	0	0	000
Lotus japonicus	ミヤコグサ	マメ科	<u>34305</u>	0	0	A DEL
<u>Cajanus cajan</u>	<u>+7</u> ×	マメ科	<u>3821</u>	10000 10 <u>44</u> 0	<u>:::</u> :	
Medicago truncatula	<u>タルウマゴヤシ</u>	マメ科	3880	-	-	3
Citrus unshiu	<u>ウンシュウミカン</u>	ミカン科	55188	準備中	<u> </u>	

- ◆ 表示情報
- 学名(各植物情報ページへのリンク)
- •和名(同上)
- •科名
- Taxonomy ID(NCBIへのリンク)
- •Cmap表示

「QTL, 地図関連情報

積み残し作業:

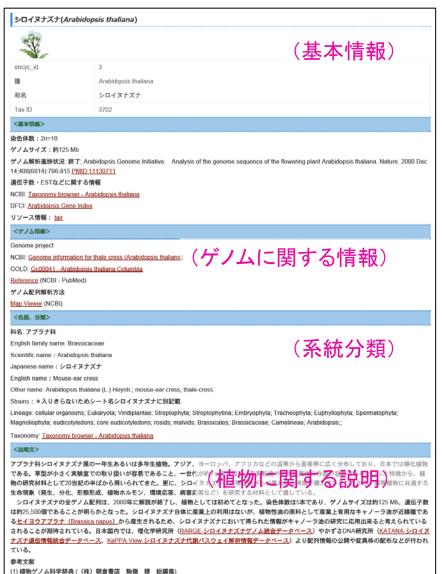
項目(1)オルソログDB関連情報, 項目(3)リソース関連情報が未掲載



整備が済んでいる植物種 から順次リンクを設定



各植物におけるページ(例: Arabidopsis thaliana)



積み残し作業:

- ◆ 項目整備が不十分
- ❖ 項目の整理
- ❖ マーカーに関する情報が未掲載整備が済んでいるものから順次リンクを設定
- ❖ 関連DBに関する情報が未掲載

整備は完了順次リンクを設定



最終年度 計画

- 1. ポータルサイトについて、<u>項目(1)~(4)にわたる横断検索機能を</u> 実装し、植物統合DBを完成させる。
- 2. ユーザーからの意見を参考にして内容の充実を図る。
- 1. <u>DBリンク集で公開している709件の各DBについて分類作業を完</u> <u>了</u>させ、リンク集に反映する。
- 2. DBの調査を継続しリンク情報を更新する。

 引き続き、プロジェクト補助員に<u>キュレーター、アノテータとしての</u> 教育を行う。

http://pgdbj.jp/



新着情報 ご意見・ご要望



Home



Home



PGDB

新着情報 ご意見・ご要望 サイトマップ

植物リソース





PGDB

Home

|新着情報 | ご意見・ご要望 | サイトマップ

公開植物情報

コンテンツ

- ▶オルソログDB
 - ▶ オルソログテーブ/
 - ▶ キーワード検索
 - ▶ BLAST検索
 - ▶ データダウンロー
- ▶ DNAマーカー・連鎖
- ▶植物リソース

▶公開植物情報

▶ PGDBi 内横断検索

PGDBjについて

コンテンツ

- ▶オルソログDB
- ▶ DNAマーカー・連鎖
 - ▶マーカー関連DBリ
 - ▶ 登録生物種リスト
- ▶植物リソース

PGDBjについて

▶新着情報

▶ 関連機関

▶ login

▶ご意見・ご要望

▶サイトポリシー

NBDC

▶サイトマップ

- トホーム
- ▶事業概要
- ▶新着情報
- ▶ご意見・ご要望
- ▶ 関連機関
- ▶ サイトポリシー
- ▶サイトマップ
- ▶管理メニュー
- ▶ login

NBDC

- ▶公開植物情報
- ▶ PGDBj 内横断検索

▶ホーム

- ▶事業概要
 - - ▶事業概要
- ▶管理メニュー



コンテンツ

トオルソログDB

▶植物リソース

▶ 公開植物情報

スト)

▶ 植物DBリンク

▶ PGDBi 内横断検索

PGDBj について

▶ホーム

▶事業概要

▶新着情報

▶関連機関 ▶サイトポリシー

▶ご意見・ご要望

▶サイトマップ



▶ DNAマーカー・連鎖地図

▶ 百科事典(登録生物種リ

- ▶オルソログDB
- ▶ DNAマーカー・連鎖^は
- ▶ 植物リソース

コンテンツ

- ▶リソース横断検索
- ▶ 公開植物情報
- ▶ PGDBi 内横断検索

PGDB_iについて

- ▶ ホーム
- ▶新着情報
- ▶ ご意見・ご要望
- ▶関連機関
- ▶ サイトポリシー
- ▶サイトマップ
- ▶管理メニュー

NBDC



オルソログDB



DNAマーカー・連鎖地図



japanese name ネギ



1/3ページ Display 20 * of 55

base info img

Allium cepa

タマネギ



4679

35875



4

Arabidopsis lyrata

Arabidopsis thaliana

ミヤマハダサオ

シロイヌナズナ

59689



Arachis hypogaea

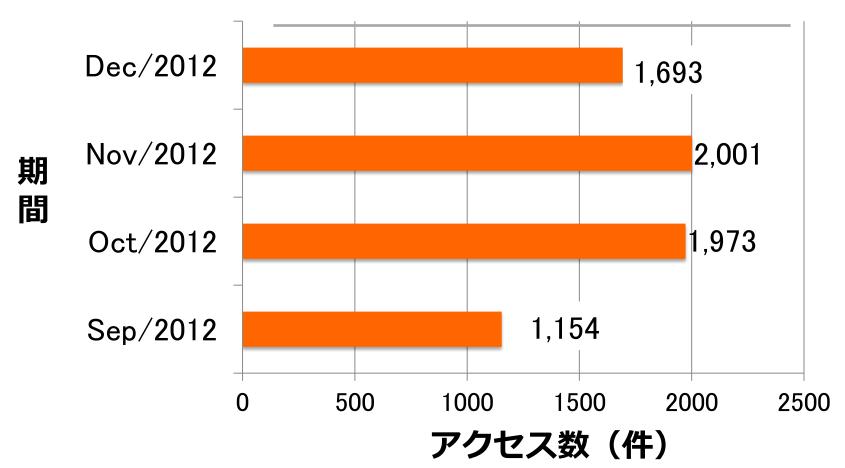
ラッカセイ

3818

●ポータルサイトへのアクセス実績(9/1~12/31)



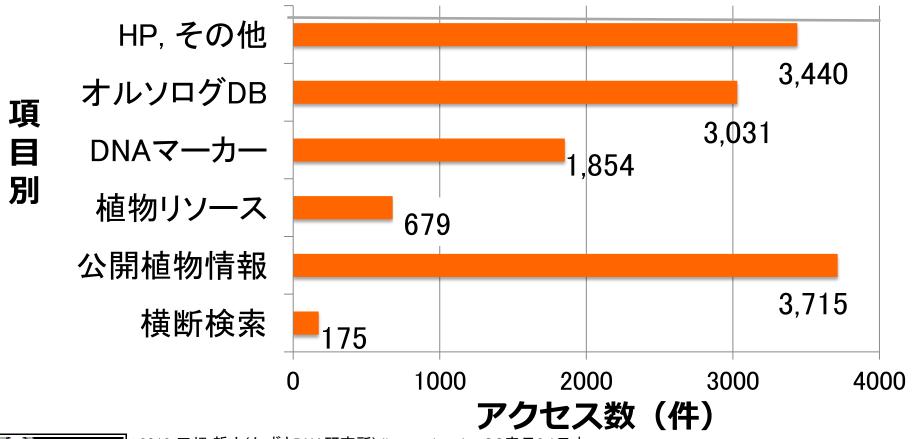
- ●サイトへの訪問者数:6,821件/4ヶ月間
- 同日 同IPからの複数アクセスは1としてカウント



●ポータルサイトへのアクセス実績(9/1~12/31)



- ●30回以上アクセスのあったページ:109件/4ヶ月間
- ●上記ページへのアクセス総数:12,894件/4ヶ月間
- •項目(1)~(4)ではコンテンツが多いものほど閲覧回数が増える傾向
- 未整備の横断検索へのアクセスが少なからずある



広報・コミュニティーとの連携

パンフレット作成

学会発表、学会展示

第30回日本植物細胞分子生物学会大会・シンポジウム(奈良

·一般口頭発表、展示 (2012年8月3日~5日)

日本植物学会第76回大会(姫路)

・展示 (2012年9月15日~17日)

第35回日本分子生物学会年会(福岡)

・展示 (2012年12月11日~14日)

第54回日本植物生理学会年会(岡山)

・DBシンポジウム、展示 (2013年3月21日~23日)

第54回日本植物生理学会年会(仙台)

・展示 (2013年3月24日~28日)

アドバイザリー委員会設置

- ・目的:コミュニティーへの周知、意見収集
- ・年2回(第一回は2012年11月22日、東京)
- ・外部委員9名(植物DB関係者、植物関連学会幹事等)