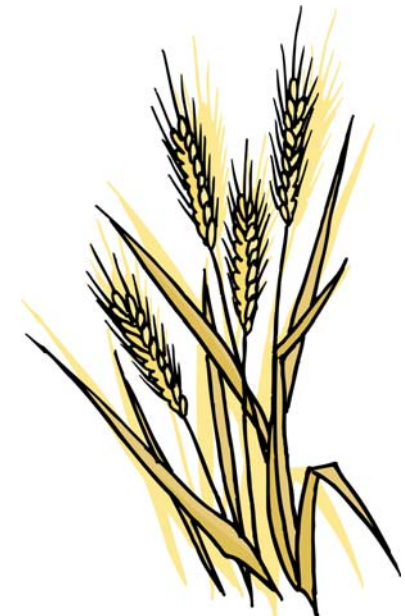


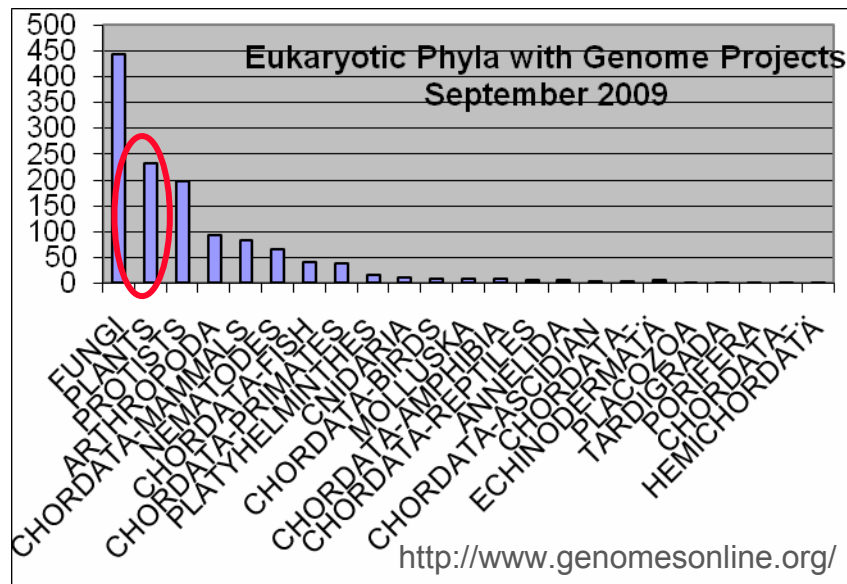
ゲノム情報に基づく植物データベースの統合



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



背景



Model plants

基礎研究

Crops

食料
工業原材料
創薬
エネルギー

品種改良

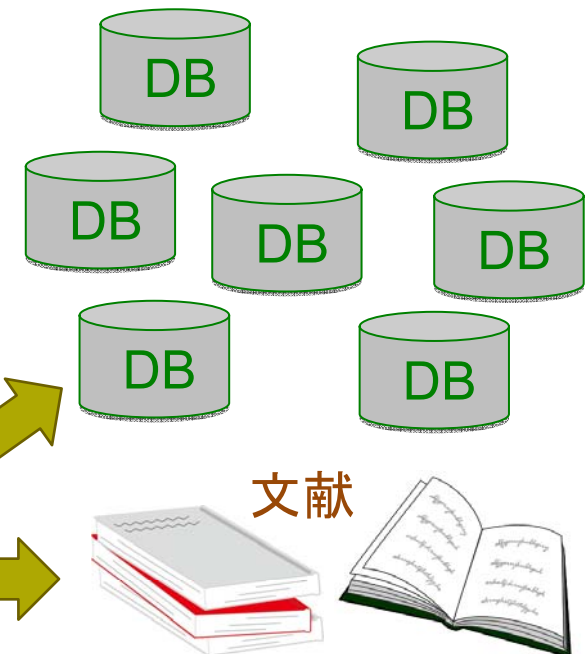
日本の植物ゲノム研究

タバコ、ゼニゴケ葉緑体 (1986年)
シロイヌナズナ
イネ
ミヤコグサ
ヒメツリガネゴケ
ヤトロファ
オオムギ
コムギ

ゲノム
cDNA
Omics

DNAマーカー
連鎖地図

QTL

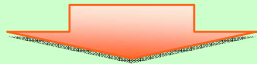


文献

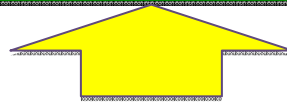
研究開発内容

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



(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

中谷明弘

(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

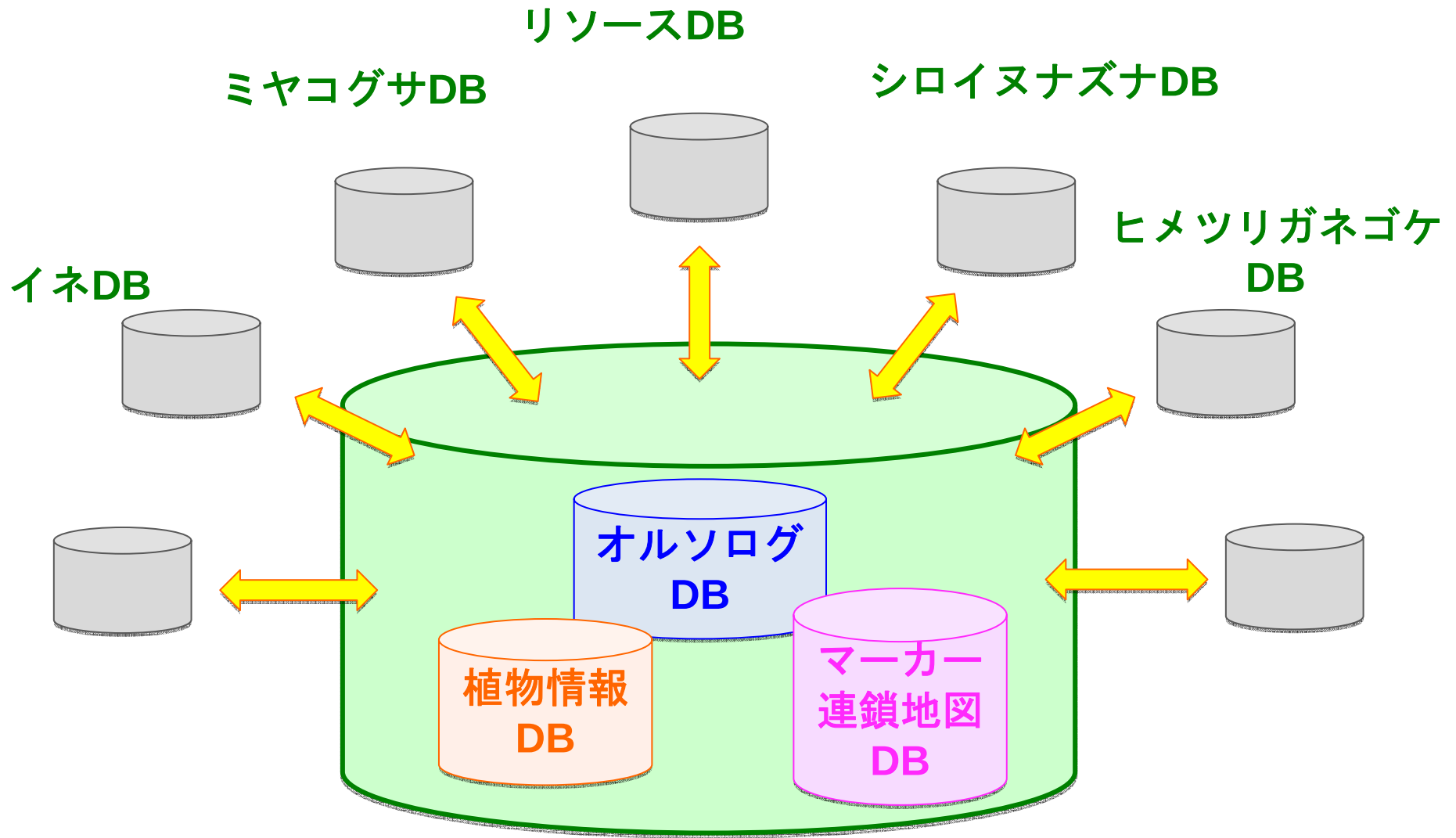
平川英樹

(3) 植物リソース情報DBの統合

中村保一

(4) 植物研究に関連する情報基盤の構築

平川英樹



目標：

- ・ DB化されていない生物情報、ゲノム関連情報をDB化する
- ・ できるだけ多くの項目について相互検索可能にする

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

H23年度 計画

植物とラン藻を対象にオルソログDBを構築する

1. アミノ酸配列DBを構築する。

遺伝子名とアミノ酸配列のDB化

2. 配列間ホモロジーDBを構築する。

全アミノ酸配列間のBLASTの結果をDB化

3. オルソログDBを構築する。

BLASTの結果を用いてアミノ酸配列をクラスタリング
生物種間の近縁関係に基づく階層構造を導入

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

H23年度 進捗と成果

階層的なオルソログDBを構築した。

1. 各DBのデータ形式の決定とDB生成を行なった。
アミノ酸配列DB / 配列間ホモロジーDB / オルソログDB
アミノ酸配列間の網羅的なBLASTとクラスタリングを
実行
2. 一連のDB生成処理をパイプライン化して自動化した。
アミノ酸配列と種間の系統関係を与えると自動実行。
3. ウェブGUIによるオルソログDBの閲覧方法を開発した
例) <http://gene.niigata-u.ac.jp/OD/>

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

階層的なオルソログDB

開発用ウェブDB

RefSeqからの16生物種の例

系統群・生物種の階層 (NCBI Taxonomy)

Streptophyta (phylum)

Poaceae (family)

Oryza sativa
Japonica Group
Zea mays
Sorghum bicolor

Viridiplantae (kingdom)

Chlorophyta (phylum)

Taxonomy TOC

http://192.168.10.130/OD/1.0.49/Part16/4479/TOC/4479.html

Hierarchical Orthologs/ Table of Clades

	BEP clade (TID:359160)>> [no rank] 25067 clusters (28555 proteins)	Viridiplantae (TID:147370)>> [kingdom] 25067 clusters (28555 proteins)	Oryza (TID:147380)>> [tribe] 25067 clusters (28555 proteins)	Oryza sativa (TID:4557)>> [genus] 25067 clusters (28555 proteins)	Oryza sativa Japonica Group (TID:4558)>> [species] 25067 clusters (28555 proteins)
Poaceae (TID:4479)>> [family] 47949 clusters (83706 proteins)	PACMAD clade (TID:147369)>> [no rank] 32497 clusters (55161 proteins)	Panicoidae (TID:147429)>> [tribe] 32497 clusters (55161 proteins)	Andropogoneae (TID:4557)>> [tribe] 32497 clusters (55161 proteins)	Zea (TID:4577)>> [genus] 17251 clusters (22159 proteins)	Sorghum (TID:4558)>> [genus] 17251 clusters (22159 proteins)
				Sorghum (TID:4557)>> [genus] 17251 clusters (22159 proteins)	Sorghum bicolor (TID:4558)>> [species] 17251 clusters (22159 proteins)

- ・系統群の階層に沿ってトーナメント式に「入れ子」でオルソログを生成
- ・系統群の上位/下位方向に「オルソログ」をリンク

Poaceaeレベルでのオルソログ

Poaceae (TID:4479) [family] [46]<< [0]>> [1]

47949 clusters (83706 proteins)
12812 MCOBP-multiple-organism multiple-protein clusters (42515 proteins)
2645 SJMAP-single-organism multiple-protein clusters (7691 proteins)
17494 SOBP-single-organism single-protein clusters (37494 proteins)

Poaceae 4479:0 Subclusters(s): 147370-147371 359160-170011

ATP synthase F0 subunit 1

Organism	Taxonomy ID	GenBank ID	RefSeq ID	Subcluster	Description
Oryza sativa Japonica Group	39947	194033257	YP_002000594.1	359160-170011	ATP synthase F0 subunit 1 [Oryza sativa Japonica Group] (7)
Sorghum bicolor	4558	115218526	XP_002447737.1	147370-147371	ATPase subunit 1 [Sorghum bicolor] (5)

Poaceae 4479:1 Subclusters(s): 147370-147371 359160-170011

ATPase subunit 4

Organism	Taxonomy ID	GenBank ID	RefSeq ID	Subcluster	Description
Oryza sativa Japonica Group	39947	194033257	YP_002000594.1	359160-170011	hypothetical protein OryzaM_03 [Oryza sativa Japonica Group] (1)
Sorghum bicolor	4558	115218526	XP_002447737.1	147370-147371	hypothetical protein SORBIRAFT_036602010 [Sorghum bicolor] (1)

Poaceae 4479:481 Subclusters(s): 147370-147371 359160-170011

Chl1g065990

Organism	Taxonomy ID	GenBank ID	RefSeq ID	Subcluster	Description
Oryza sativa Japonica Group	39947	115438919	NP_001043736.1	359160-170011	Chl1g065990 [Oryza sativa Japonica Group] (1)
Sorghum bicolor	4558	115218526	XP_002447737.1	147370-147371	hypothetical protein SORBIRAFT_036602010 [Sorghum bicolor] (1)
Zea mays	4577	167660166	XP_001108118.1	147370-147371	hypothetical protein LOC100384626 [Zea mays] (1)

Poaceae 4479:482 Subclusters(s): 147370-147371 359160-170011

phosphatidylinositol synthase 2

Organism	Taxonomy ID	GenBank ID	RefSeq ID	Subcluster	Description
Oryza sativa Japonica Group	39947	115438919	NP_001043736.1	359160-170011	Os02g0123600 [Oryza sativa Japonica Group] (1)
Sorghum bicolor	4558	115218526	XP_002447737.1	147370-147371	hypothetical protein SORBIRAFT_036602010 [Sorghum bicolor] (1)
Zea mays	4577	167660166	XP_001108118.1	147370-147371	phosphatidylinositol synthase 2 [Zea mays] (5)

Poaceae 4479:483 Subclusters(s): 147370-147371 359160-170011

phosphatidylinositol synthase 2

Organism	Taxonomy ID	GenBank ID	RefSeq ID	Subcluster	Description
Oryza sativa Japonica Group	39947	115438919	NP_001043736.1	359160-170011	Os02g0123600 [Oryza sativa Japonica Group] (1)
Sorghum bicolor	4558	115218526	XP_002447737.1	147370-147371	hypothetical protein SORBIRAFT_036602010 [Sorghum bicolor] (1)
Zea mays	4577	167660166	XP_001108118.1	147370-147371	phosphatidylinositol synthase 2 [Zea mays] (5)

(Ver. 1.0.49 (beta))



©2012田畑 哲之(かずさDNA研究所) licensed under CC表示2.1日本

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

H24年度-25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。

H23年度 計画

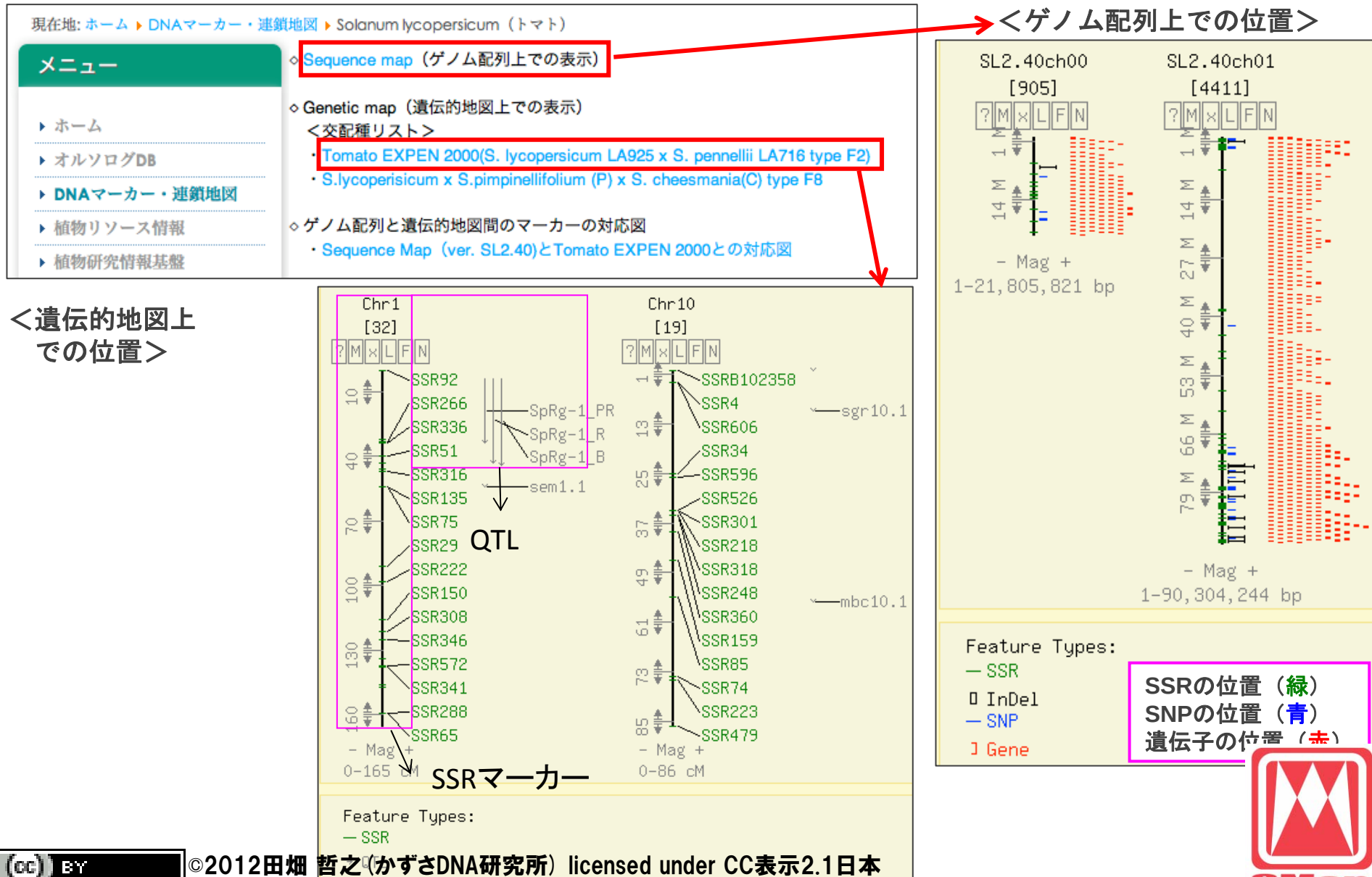
1. ゲノム配列決定の対象となっている植物をNCBIやGOLDなどのDBを用いて調べ、本研究開発の対象を決める。
2. 国内外のDNAマーカーやQTL、連鎖地図関連のDBを調査する。
3. DBや文献に記載されているDNAマーカーやQTL、連鎖地図に関する情報（多型の種類（SSRPやSNP）、プライマー配列など）を収集する。
4. ゲノム塩基配列にDNAマーカー情報を付与する。
5. 収集したDNAマーカーおよびQTL情報のゲノム配列上の位置を特定するシステムを構築する。

H23年度 進捗と成果

1. 55の植物種を本研究開発の対象とした。これらのうち、本年度は、ナス科（トマト、ジャガイモ、ナス、タバコ、トウガラシ）とイネ科を対象とした。
2. DBの調査は、43の植物種（約8割）について完了した。
3. マーカー情報の収集は、ナス科とイネ科（2012年）のDNAマーカー（SSRP, SNP, indelのプライマー）、連鎖地図、QTL情報をDBおよび文献から抽出した。
4. ナス科について、DNAマーカー付随配列を抽出しゲノム配列にマッピングした。
5. DNAマーカー付随配列のゲノム配列上および遺伝的地図上の位置の対応関係を表示するシステムを構築した。

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

マーカー付随配列のゲノム配列および遺伝的地図上の対応関係を表示するシステム



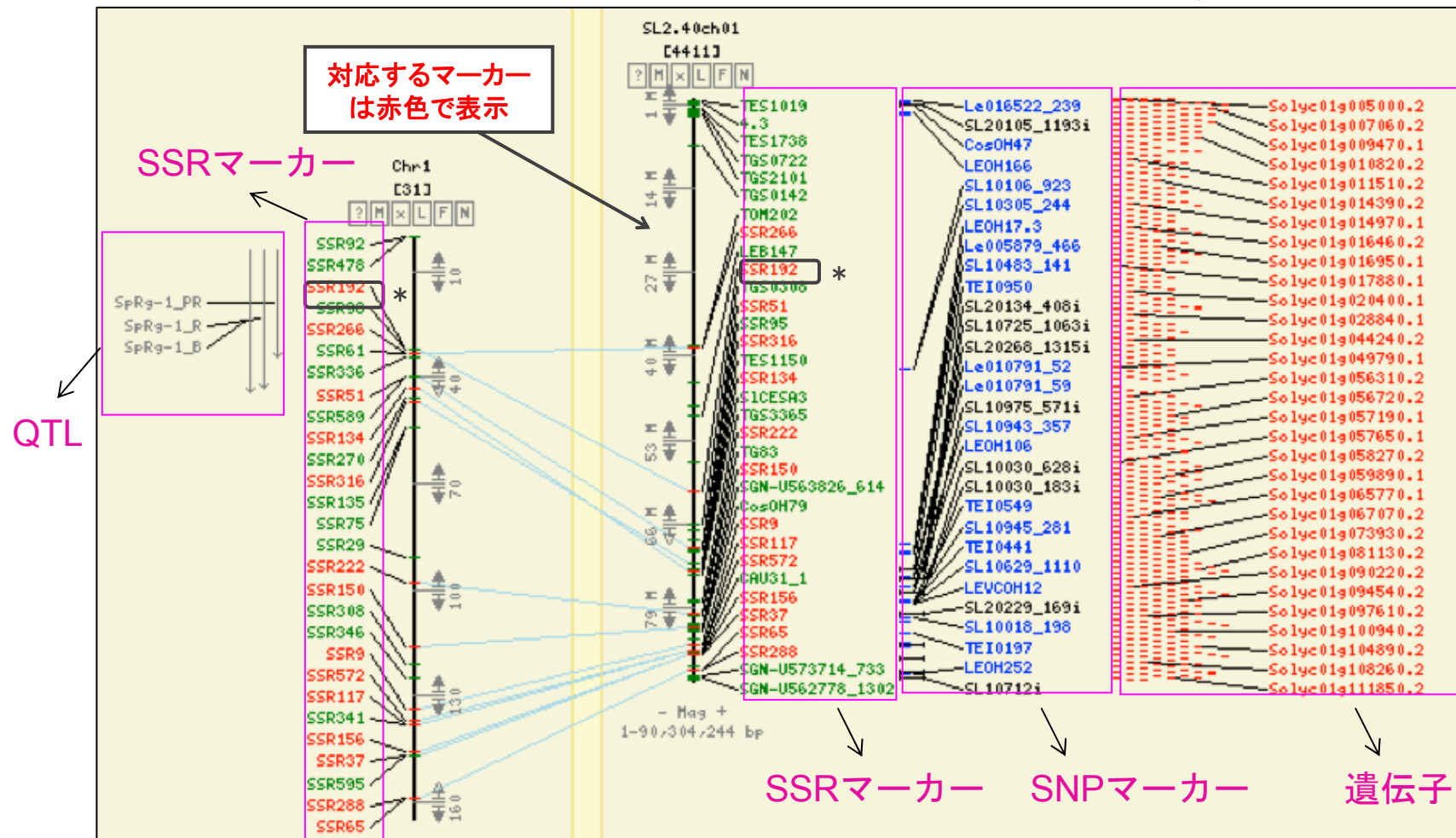
2. DNAマーカーおよび連鎖地図情報に基づく植物ゲノムDBの統合

ゲノム配列と遺伝的地図間のマーカーの対応図

(Genome Map (SL2.40)とEXPEN2000との対応図)

＜遺伝的地図上での位置＞

＜ゲノム配列上での位置＞



ゲノム配列と遺伝子配列上におけるマーカーの対応関係を表示することにより、表現型に影響を及ぼしうる遺伝子の推定に利用できる → ゲノム比較研究および育種プロセスの加速

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

<遺伝的地図上でのDNAマーカーの情報>

Feature Name: SSR192 (マーカー名)
Aliases: N/A
Accession ID: 132148846 (ID)
Feature Type: SSR
Species: Tomato (植物種)
Map: Map Set: Tomato EXPEN 2000(S. lycopersicum LA925 x S. pennellii LA716 F2)
Map Name: Chr1 (連鎖地図) (交雑)
Start: 32 cM (連鎖地図上の位置)
Stop: 32 cM

<ゲノム配列上でのDNAマーカーの情報>

Feature Name: SSR192 (マーカー名)
Aliases: N/A
Accession ID: 132262937 (ID)
Feature Type: SSR (マーカーの種類)
Species: Tomato (植物種)
Map: Map Set: Solanum lycopersicum Map (マップ名)
Map Name: SL2.40ch01 (染色体番号)
Start: 60,821,704 bp (プライマー開始位置)
Stop: 60,821,871 bp (プライマー終止位置)
Amplified Length: 168 (推定増幅長)
Forward: ACAACATGGGAAGCACTTGA (Fwプライマー配列)
Fw_align_length: 20 (Fwプライマー配列長)
Fw_mismatch: 0 (ミスマッチ数)
GBrowse: [[Link to GBrowse](#)] (Gbrowseへのリンク)
Reverse: ATTAAATTGGGCCATGGTGA (Rvプライマー配列)
Rev_align_length: 20 (Rvプライマー配列長)
Rev_mismatch: 0 (ミスマッチ数)
X Ref.(PubMed): [[21193580](#)] (PubMed)

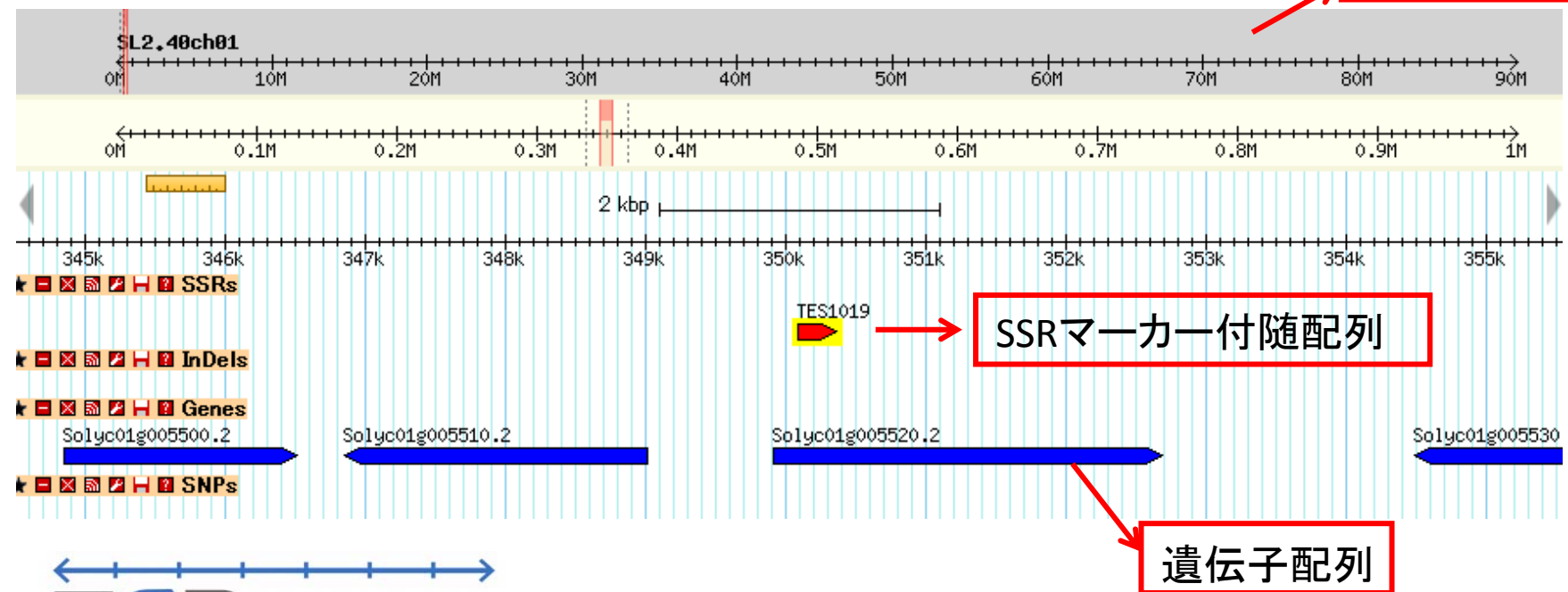
2. DNAマーカーおよび連鎖地図情報に基づく植物ゲノムDBの統合

マーカー付随配列のゲノム配列上の位置と遺伝子の対応関係の表示



SSR, SNP, Indelマーカーと遺伝子のゲノム配列上の位置をGbrowseにより表示

Solanum lycopersicum(SL2.40)



GBrowse

H24年度-25年度 計画

1. 55の植物種に対して国内外のDBや文献からのDNAマーカー、QTL、連鎖地図に関する情報の収集を引き続き行う。
2. イネ科やマメ科といった主要な作物については文献の量が膨大であり情報の収集に時間を要するため、基本的にマーカー関連DBへのリンクに留める。
3. ゲノム配列が解読された植物については、マーカー付随配列をゲノム配列上にマッピングすることで連鎖地図やQTLとの対応関係を明らかにし、DBの統合化を進める。
4. DNAマーカー付随配列のゲノム配列上および遺伝的地図上の位置の対応関係を表示するシステムの充実を図る。

3. 植物リソースの統合

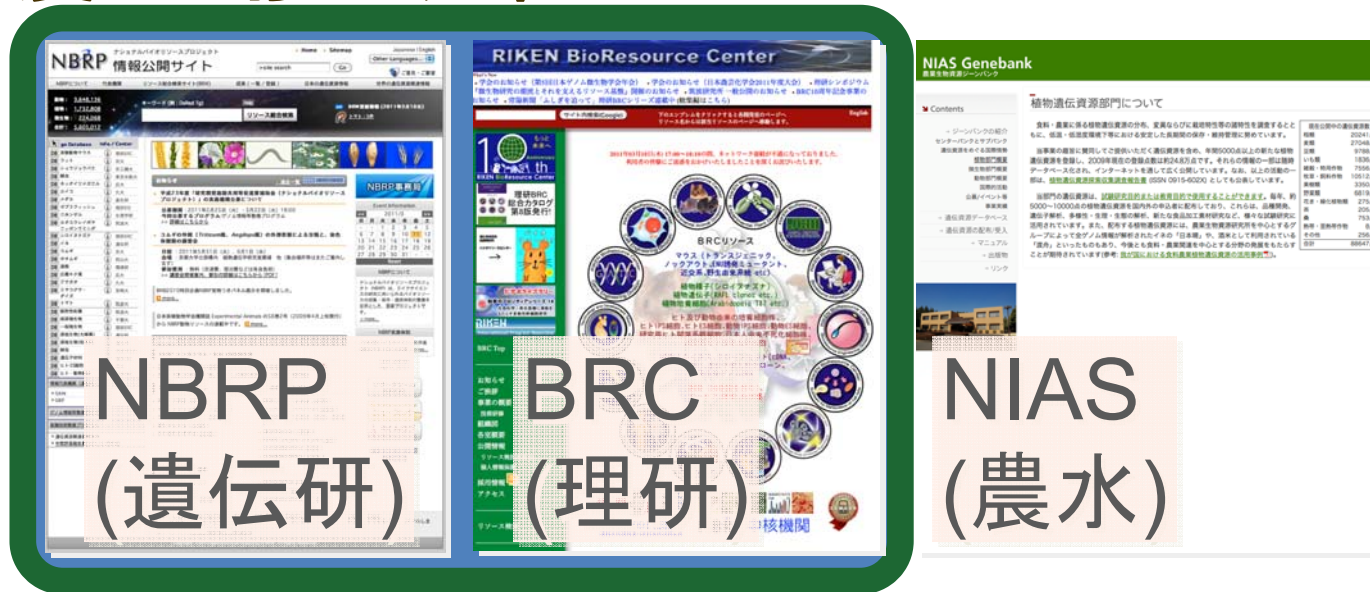
H23年度 計画



- ❖ 文部科学省のナショナルバイオリソースプロジェクト(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、ダイズ、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。
- ❖ 理化学研究所バイオリソースセンターの事業で、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ー、種子や培養細胞系の維持・配布が行われている。
- ❖ 農林水産省の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。
- ❖ 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整備し、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。

3. 植物リソースの統合

H23年度 進捗と成果



連携強化のための支援「植物実験材料の情報統合システムへのデータ加工・追加」を実施した。具体的には：

1. NBRP他のリソース情報として、オオムギ、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を整備する開発を支援した。
2. SABRE提供情報を遺伝研NBRPと統合検索可能するための検索システム設置し公開した。

3. 植物リソースの統合

Bio Resource World beta @NBRP

ホームへ戻る Japanese | English

生物種 シロイヌナズナ / 植物細胞・遺伝子 カタログ Strain シロイヌナズナ・トランスポゾンタグライク から検索

キーワード At1g01490 検索 ""をワイルドカードとして検索できます Help

このページの情報は役に立ちましたか? ☐ はい ☐ いいえ 送信

ALL, DNA, BLAST, GO, Referenceのタブをクリックするとページが切り替わります。

ALL DNA BLAST Gene Ontology Reference

リソース数: 5,579,283

検索結果 8件 (キーワード: At1g01490)

系統 8件 系統の検索結果一覧へ DNA 0件

生物種	系統	生物種	DNA
シロイヌナズナ / 植物細胞・遺伝子	シロイヌナズナ・トランスポゾンタグライク 8件		リソースは見つかりませんでした。

系統一覧 8件 Page: 1 / 1 1 1-8 / 8

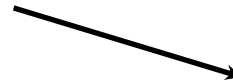
シロイヌナズナ / 植物細胞・遺伝子 pst14696 分譲可 理研 BRC

Category: シロイヌナズナ・トランスポゾンタグライク

- Line name: 12-5192-1
- MIPS code of insert position (5'): At1g01490
- Gene Ontology: GO:0046872, GO:0005575, GO:0030001

シロイヌナズナ / 植物細胞・遺伝子 nct13657 分譲可 理研 BRC

NBRP
(遺伝研)



At4g09000

- 名寄せ=ID統一
- 更新系作成支援

Resource Number	pst14696
Line name	12-5192-1

Insert position from 5'-flanking sequence	
Candidate position 1	182995
Candidate gene 1	1
MIPS code	At1g01490
5'1(1) Distance	-929
Title	unknown protein
MIPS code	At1g01500
5'1(2) Distance	-2265
Title	unknown protein
Insert position from 3'-flanking sequence	
Candidate position 1	

BRC
(理研)



H24年度-25年度 計画

1. 平成24年度は、平成23年度に実施した横断検索の更新とともに、有用なバイオリソース情報のデータ収集とID統合を追加する。
2. 既存のPlant Ontology (PO) 等の適用や拡張等を含め、共有化を容易にするためのオントロジーの策定と整備を実施する。
3. 平成25年度は、引き続きデータの追加を続けるとともに、植物オルソログDBとの連携による統合環境の活用により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が可能なツールの整備を行う。
4. ワンストップショップを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。

H23年度 計画

1. ゲノム配列の解読の対象となる植物の基本的特徴や農学、園芸学、育種学上の特徴、有用性、ゲノム配列の決定方法、解析手法について文献（論文や辞典、総説）から情報を収集する。
2. 植物のゲノム、トランスクリプトーム、メタボローム、インタラクトームといったオミックス研究のDBを調査し分類する。
3. 植物と解析手法に精通したキュレーターとアノテーターを育成する。
4. 本研究開発の4課題についての情報を閲覧し横断的検索が行えるポータルサイトを構築する。

4. 植物研究に関連する情報基盤の構築

H23年度 進捗と成果

1. 55の植物種についての学術的情報の収集が完了した。
1. 国内外における1,093件の植物のオミックス関連DBを収集した。
2. プロジェクト補助員3名をキュレーター、アノテーターとして育成した。
3. 55の植物種について、本年度に収集したマーカ情報やゲノム情報、オミックス関連DBを閲覧できるポータルサイトを構築した。

4. 植物研究に関連する情報基盤の構築

ポータルサイトのトップページ

 **植物統合データベース**

和名は全角カナで...  **検索**

[ホーム](#) [オルソログDB](#) [DNAマーカー・連鎖地図](#) [植物リソース情報](#) [植物研究情報基盤](#)

現在地: ホーム

メニュー

- ▶ [ホーム](#)
- ▶ [オルソログDB](#)
- ▶ [DNAマーカー・連鎖地図](#)
- ▶ [植物リソース情報](#)
- ▶ [植物研究情報基盤](#)
 - ▶ [植物百科事典](#)
 - ▶ [単子葉植物](#)
 - ▶ [双子葉植物](#)
 - ▶ [裸子植物](#)
 - ▶ [藻類・コケ類](#)
 - ▶ [植物データベースリンク](#)
 - ▶ [国内植物データベース](#)
 - ▶ [国外植物データベース](#)



<植物統合データベース 概要>

[オルソログDB](#) ~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~

植物ゲノム関連DBを統合する際、植物を対象とした計算機による客観的な条件設定に基づく遺伝子の分類が必要となるため、ゲノム構造が決定された植物を網羅するオルソログDBの構築を行っている。

[DNAマーカー・連鎖地図](#) ~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~

多様な植物種におけるDNAマーカーや遺伝的連鎖地図といった文献に留まっている情報を収集し、散在しているDBとの統合を行っている。

関連サイト

[独立行政法人科学技術振興機構 \(JST\)](#)
[バイオサイエンスデータベースセンター \(NBDC\)](#)
[ライフサイエンス統合データベースセンター \(DBCLS\)](#)

4. 植物研究に関連する情報基盤の構築

各植物に関するトップページ

NBDC 植物統合データベース

和名は全角カナで... 検索

ホーム オルソログDB DNAマーカー・連鎖地図 植物リソース情報 植物研究情報基盤

現在地: ホーム ▶ 植物研究情報基盤 ▶ 植物百科事典

メニュー

- ▶ ホーム
- ▶ オルソログDB
- ▶ DNAマーカー・連鎖地図
- ▶ 植物リソース情報
- ▶ 植物研究情報基盤
 - ▶ 植物百科事典
 - ▶ 単子葉植物
 - ▶ 双子葉植物
 - ▶ 裸子植物
 - ▶ 藻類・コケ類
 - ▶ 植物データベースリンク
 - ▶ 国内植物データベース
 - ▶ 国外植物データベース

関連サイト

- 独立行政法人科学技術振興機構 (JST)
- バイオサイエンスデータベースセンター (NBDC)
- ライフサイエンス統合データベースセンター (DBCLS)

Allium cepa (タマネギ)	Allium fistulosum (ネギ)	Arabidopsis lyrata (ミヤマハタザオ)	Arabidopsis thaliana (シロイヌナズナ)
		No Image	
Arachis hypogaea (ラッカセイ)	Brachypodium distachyon (セイヨウヤマカモジ)	Brassica napus (セイヨウアブラナ)	Brassica oleracea (キャベツ)
	No Image		
Brassica rapa (ハクサイ)	Cajanus cajan (キマメ)	Camellia sinensis (チャ)	Capsicum annuum (トウガラシ)
	No Image		

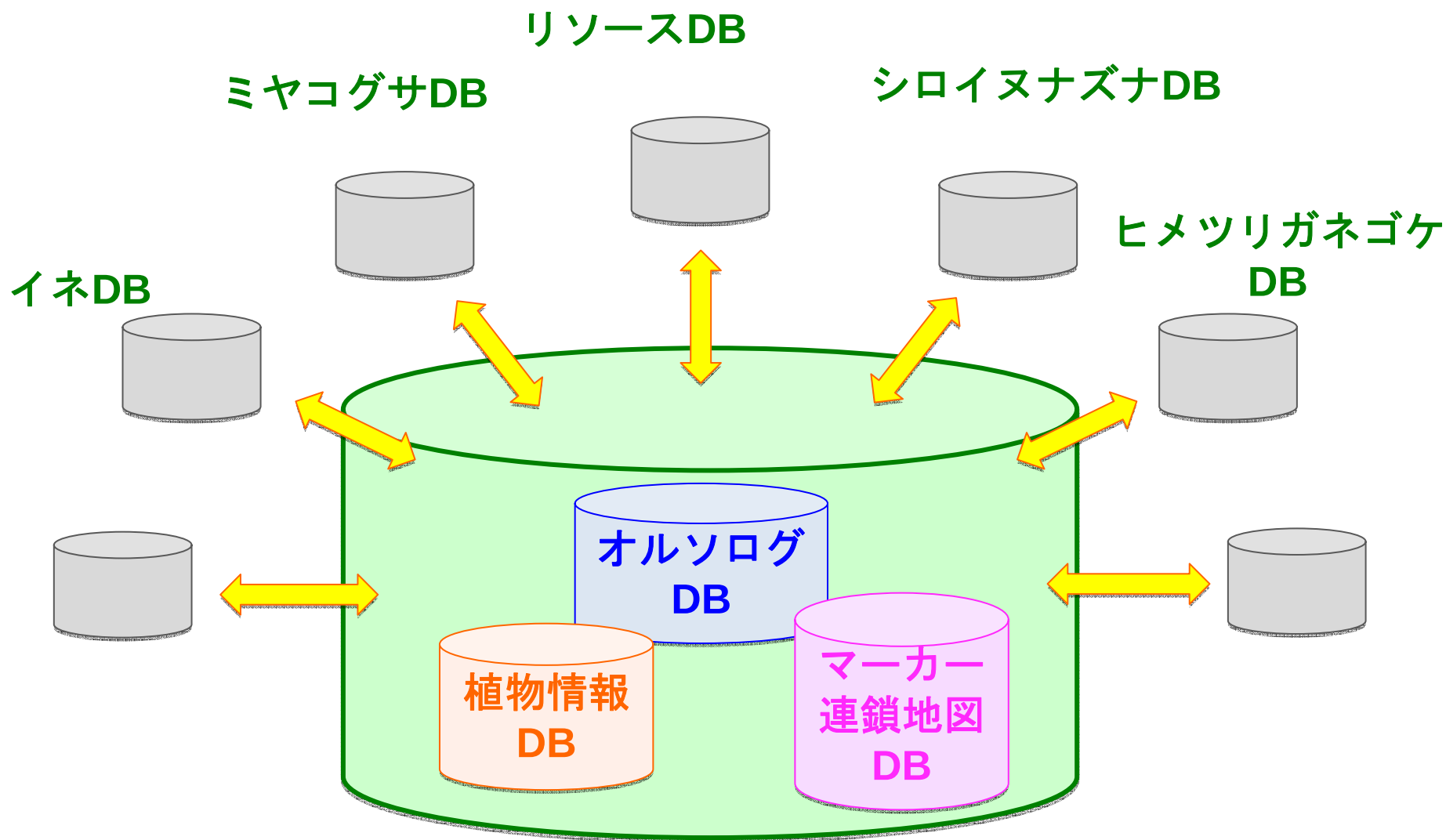
各植物におけるページ（例：*Solanumlycopersicum*）

 CC BY

4. 植物研究に関連する情報基盤の構築

H24年度-25年度 計画

1. 対象とした55の植物種についての学術的な情報の収集を終了させる。（平成24年度）
2. 収集した1,093件の国内外のゲノムやトランスクリプトーム、メタボローム、インタラクトームといった各オミックス研究のDBを分類しポータルサイトに掲載する。（平成24年度）
3. DBの調査を継続しリンク情報を更新する。
4. キュレーターとアノテーターを育成する。
5. ポータルサイトに本研究開発の4課題を掲載することで情報の連携を図り、横断的検索の機能を付与する。
6. 情報がある程度揃った時点でポータルサイトを一般に公開する。（平成24年度）
7. ポータルサイトに登録したDBのリンク切れなどを調べ更新することができるDB管理システムを構築する。



H24年度計画：

- ・ DBの公開（6月）
- ・ 広報、学会、コミュニティへの周知、意見収集