

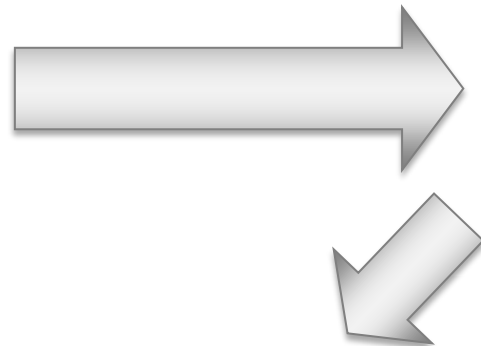
植物ゲノム情報活用のための統合研究基盤の構築

かずさDNA研究所植物ゲノム研究部
田畑哲之
大阪大学・大学院医学系研究科
中谷明弘



ゲノム・cDNA構造情報

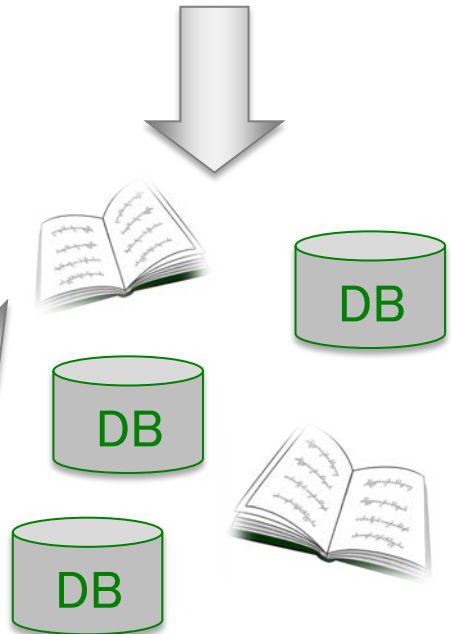
- クラミドモナス
- シアニディオシゾン
- ヒメツリガネゴケ
- ゼニゴケ
- イヌカタヒバ
- イネ
- コムギ
- オオムギ
- タバコ
- シロイヌナズナ
- ミヤコグサ
- ダイズ
- トマト
- イチゴ
- カーネーション
- ハクサイ
- キャッサバ
- ヤトロファ
- ポプラ
- ⋮



ゲノム機能情報

- 転写産物
- 代謝産物
- 変異体
- DNAマーカー
- 連鎖地図

遺伝子機能
ネットワーク
有用遺伝子
QTL解析
品種改良



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、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



JST 統合化推進プログラム H23～H25

「ゲノム情報に基づく植物データベースの統合」

- (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

#01



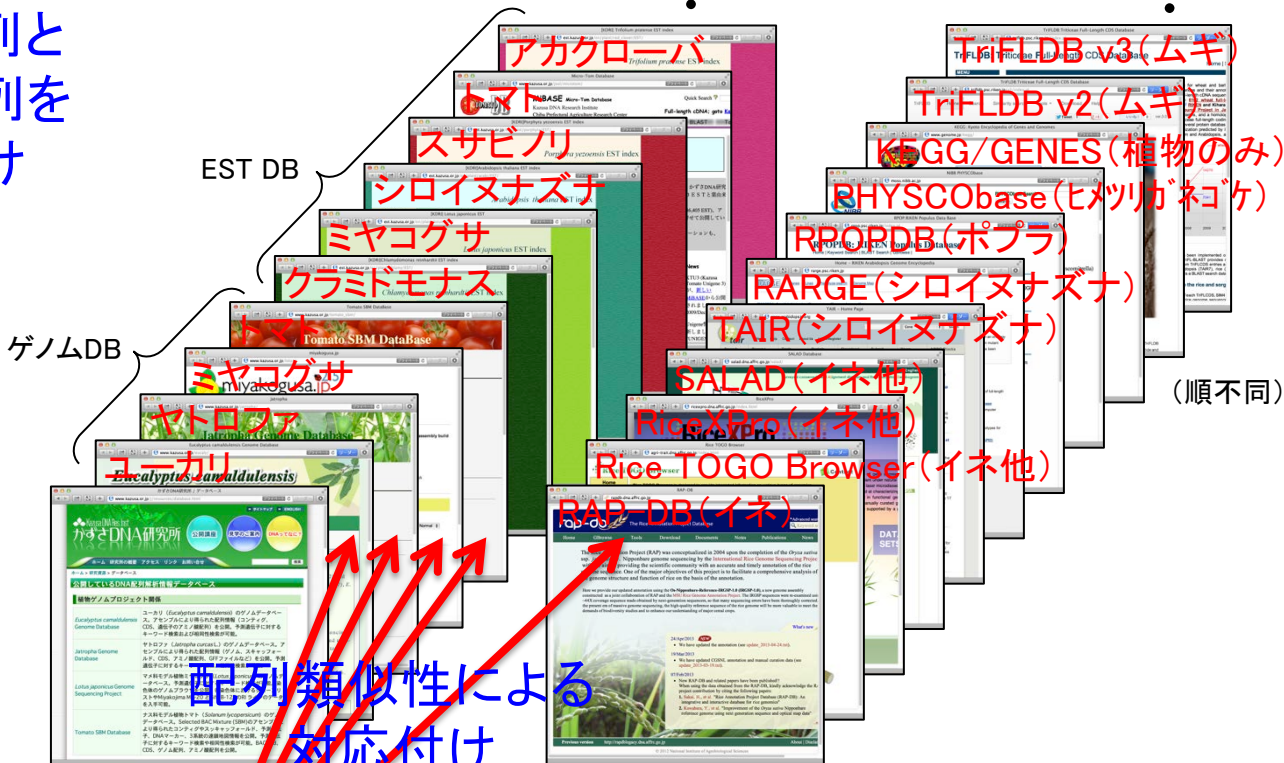
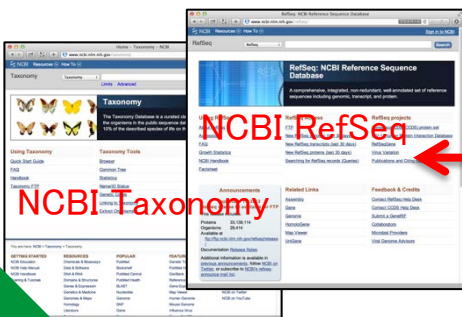
遺伝子オルソログDBの構築

- NCBI RefSeq DBから取得したアミノ酸配列(緑色植物20種の約50万配列とラン藻111種の約50万配列)を用いてオルソログ情報の生成及び更新を行い、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と検索用のウェブサイトを構築
- ・かずさDNA研究所保有各種DB、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)内のアミノ酸配列データとリンク

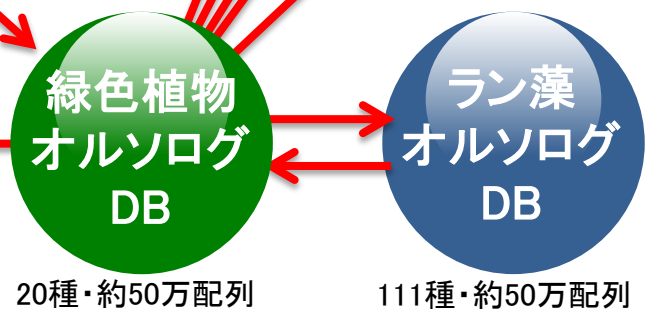
オルソログDBをハブとしたDBのリンク

各DBから取得した配列と
オルソログDB内の配列を
予めBLASTで対応付け

PGDBj 横断検索



配列類似性による
対応付け



研究開発の現状

#01



遺伝子オルソログDBの構築

- NCBI RefSeq DBから取得したアミノ酸配列(緑色植物20種の約50万配列とラン藻111種の約50万配列)を用いてオルソログ情報の生成及び更新を行い、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と検索用のウェブサイトを構築
- かずさDNA研究所保有各種DB、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)内のアミノ酸配列データとリンク

#02



DNAマーカーおよび連鎖地図情報

- 10種について75,975件のマーカー情報を収集、公開
- 20種について文献調査を行い、15,259件のマーカー情報1,767件のQTL情報を収集、公開
- 地図表示システム(遺伝地図、物理地図、地図間比較)を構築

地図表示システムの概要

2014年8月公開

トマト遺伝地図

解析集団の選択

The interface shows a genetic map with chromosomes labeled ch01 through ch08. A red circle highlights the ch01 chromosome, and a red arrow points to it from the text '連鎖群の選択' (Linkage group selection). Above the map, there are dropdown menus for 'Species: Solanum lycopersicum' and 'Population: AMF2'. A red circle highlights the 'AMF2' population selection, with a red arrow pointing to it from the text '解析集団の選択' (Analysis group selection).

連鎖群の選択

Species: *Solanum lycopersicum* Population: AMF2 Linkage Group: ch01

Linkage Group : ch01 : Compare to other maps **地図間の比較**

Page Format : 50 | SAVE ALL | Prev | Next | Page: 1 | Page | Data Count:93 TotalPages:2

Sort : 1 | Position | Asc. | 2 | Marker Name | Asc. | 3 | Not select | Asc. | Sort | Download EPS

Species	Marker Name	Marker Type	Gene, trait	Population	LG or Chr	Position (cM)
Solanum lycopersicum	TES1447	EST-SSR		AMF2	ch01	0.0
Solanum lycopersicum	2758_112	EST-SNP		AMF2	ch01	0.691
Solanum lycopersicum	TES1019	EST-SSR		AMF2	ch01	1.355
Solanum lycopersicum	TGS1588	Genome-SSR		AMF2	ch01	1.66
Solanum lycopersicum	6172_1047	EST-SNP		AMF2	ch01	3.656
Solanum lycopersicum	8424_368	EST-SNP		AMF2	ch01	4.743
Solanum lycopersicum	TGS1009	Genome-SSR		AMF2	ch01	5.253
Solanum lycopersicum	2503_859	EST-SNP		AMF2	ch01	10.25
Solanum lycopersicum	1581_673	EST-SNP		AMF2	ch01	10.802
Solanum lycopersicum	TGS0604	Genome-SSR		AM		
Solanum lycopersicum	3850_896	EST-SNP		AM		
Solanum lycopersicum	10779_574	EST-SNP		AMF2	ch01	17.794
Solanum lycopersicum	7580_248	EST-SNP		AMF2	ch01	18.306
Solanum lycopersicum	2917_1793	EST-SNP		AMF2	ch01	18.808
Solanum lycopersicum	7553_763	EST-SNP		AMF2	ch01	18.808
Solanum lycopersicum	TES1738	EST-SSR		AMF2	ch01	18.808
Solanum lycopersicum	11051_1291	EST-SNP		AMF2	ch01	19.805
Solanum lycopersicum	TES0943	EST-SSR		AMF2	ch01	20.035
Solanum lycopersicum	3775_571	EST-SNP		AMF2	ch01	20.378
Solanum lycopersicum	2024_303	EST-SNP		AMF2	ch01	21.37
Solanum lycopersicum	21382_304	EST-SNP		AMF2	ch01	21.37

マーカーリスト

地図表示システムの概要

Compare

Species : Solanum lycopersicum Population or Map:AMF2
VS
Species : Solanum lycopersicum Population or Map: SL2.40

ch01

TES1447
TGS1019
TGS1009
TGS0604
TES1738
TES0943
TES0233
TGS0722
TGS0726
TEI0750
TGS0377
TGS2101
TGS0271
TGS0241
TGS0308
TGS0685
TES1150
TEI0950
TGS0564
TGS2456
TGS0486
TEI0549
TGS3365
TES0131
TEI0441
TGS0748
TEI0197
TGS0127

SL2.40ch01

TES1019
TES1447
TGS1009
TGS0604
TES1738
TES0943
TGS0722
TES0233
TGS0726
TEI0750
TGS2101
TGS0377
TGS0271
TGS0241
TGS0308
TES1150
TEI0950
TGS0564
TGS0486
TEI0549
TGS3365
TES0131
TEI0441
TGS0748
TEI0197
TGS0127

比較対象生物種の
選択

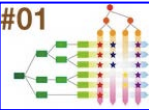
Search mapped marker

Species : Brachypodium distachyon
Genetic or Physical : Genetic
Population : Any
Linkage Group : Any
Marker Type : Indel
Map Position : from [] cM to [] cM
Marker Name : []
Comment : []
Sort : 1 Not select Asc. 2 Not select Asc. 3 Not select
Submit Reset

マップされたマーカーの検索

研究開発の現状

#01



遺伝子オルソログDBの構築

- NCBI RefSeq DBから取得したアミノ酸配列(緑色植物20種の約50万配列とラン藻111種の約50万配列)を用いてオルソログ情報の生成及び更新を行い、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と検索用のウェブサイトを構築
- かずさDNA研究所保有各種DB、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)内のアミノ酸配列データとリンク

#02



DNAマーカーおよび連鎖地図情報

- 10種について75,975件のマーカー情報を収集、公開
- 20種について文献調査を行い、15,259件のマーカー情報1,767件のQTL情報を収集、公開
- 地図表示システム(遺伝地図、物理地図、地図間比較)を構築

#03



植物リソース情報

- 理研BRCのSABREシステムを拡充、BRCとNBRPが保有する14種150万件のバイオリソース情報をPGDBj横断検索システムから統合検索可能に
- カンキツ類のリソース情報を近畿大、果樹研究所から収集

リソース情報の統合

NBRP
(遺伝研)

BRC
(理研)

カンキツ
リソースDB

PGDB Plant Genome DataBase Japan

atpA PGDBj 横断検索

新着情報 | ご意見・ご要望 | サイトマップ | ご要望 | サイトマップ

検索結果: 165hits
キーワード: "atpA" 同義語展開対象無し

1 2 3 > GoTo >>

[植物リソース] [Homologs of AT2G37040.1](#)
遺伝子モデル [AT2G37040.1](#) / 分子タイプ protein coding
遺伝子説明 PHE ammonia lyase 1
相同リソース

該当箇所 [agi_gene_model_id] AT2G37040.1 [molecular_type] protein_coding [sources] A. thaliana, T. halophkaryotes - 493 (source: NCBI BLINK). [gene_symbol] ATPAL1|PAL1 [gene_full_name] |PHE ammonia lyase 1 [prot

[植物リソース] [Homologs of AT3G53260.1](#)
遺伝子モデル [AT3G53260.1](#) / 分子タイプ protein coding
遺伝子説明 phenylalanine ammonia-lyase 2
相同リソース

該当箇所 [agi_gene_model_id] AT3G53260.1 [molecular_type] protein_coding [sources] A. thaliana, T. halophkaryotes - 489 (source: NCBI BLINK). [gene_symbol] ATPAL2|PAL2 [gene_full_name] |phenylalanine ammonia-lya

[植物リソース] [Homologs of AT4G34110.1](#)
遺伝子モデル [AT4G34110.1](#) / 分子タイプ protein coding
遺伝子説明 poly(A) binding protein 2
相同リソース

DB

すべて

オルソログDB (137)

遺伝マーカーDB (0)

植物DBリンク集 (0)

公開植物情報DB (0)

植物リソース (28)

Marker Type

すべて

QTL (0)

SSR (0)

SNP (0)

Transposable Element (0)

Uncategorized (0)

Unknown (0)

Others (0)

DB分類

すべて

情報提供 (0)

コンソーシアム (0)

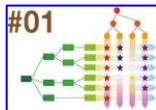
データベース (0)

情報統合 (0)

分類 (0)

研究開発の現状

#01



遺伝子オルソログDBの構築

- NCBI RefSeq DBから取得したアミノ酸配列(緑色植物20種の約50万配列とラン藻111種の約50万配列)を用いてオルソログ情報の生成及び更新を行い、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と検索用のウェブサイトを構築
- かずさDNA研究所保有各種DB、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)内のアミノ酸配列データとリンク

#02



DNAマーカーおよび連鎖地図情報

- 10種について75,975件のマーカー情報を収集、公開
- 20種について文献調査を行い、15,259件のマーカー情報1,767件のQTL情報を収集、公開
- 地図表示システム(遺伝地図、物理地図、地図間比較)を構築

#03



植物リソース情報

- 理研BRCのSABREシステムを拡充、BRCとNBRPが保有する14種150万件のバイオリソース情報をPGDBj横断検索システムから統合検索可能に
- カンキツ類のリソース情報を近畿大、果樹研究所から収集

#04



植物研究に関連する情報基盤

- 55種について基本情報と生物分類情報を公開
- 33種についてゲノム概要と解読手法を収集し各ゲノムDBへリンク
- 553件の関連DBを調査し、カテゴリー分類してリンク

研究開発の現状



PGDBj ポータルサイト
Plant Genome DataBase Japan
http://pgdbj.jp

横断検索機能
英語版PGDBj

リソース

コンテンツメニュー

- ▶ 植物リソース
NBRPの提供する植物リソース
(cDNAクローン)を検索(横断検索)
- ▶ 植物リソースについて

コンテンツメニュー

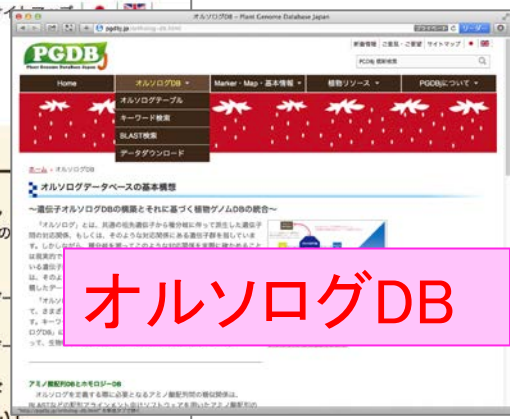
- ▶ オルソログテーブル
遺伝子のアミノ酸配列の
テーブル
- ▶ キーワード検索
オルソログDBの関連テ
- ▶ BLAST検索
オルソログDBの関連テ
相同性検索
- ▶ データダウンロード
- ▶ オルソログDBについて



コンテンツメニュー

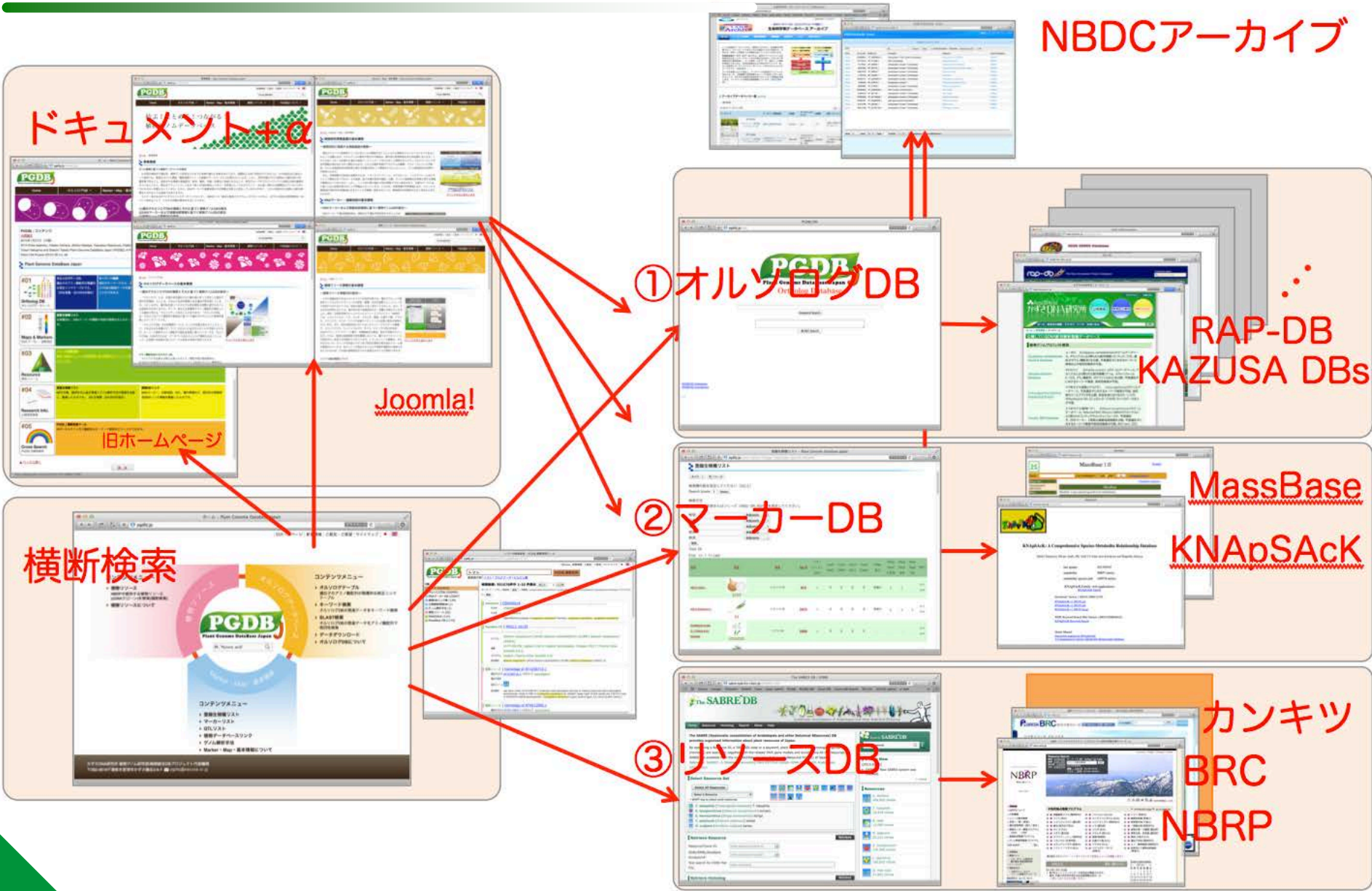
- ▶ 登録生物種リスト
- ▶ マーカーリスト
- ▶ QTLリスト
- ▶ 植物データベースリンク
- ▶ ゲノム解析手法
- ▶ Marker・Map・基本情報について

マーカー
QTL
基本情報



オルソログDB

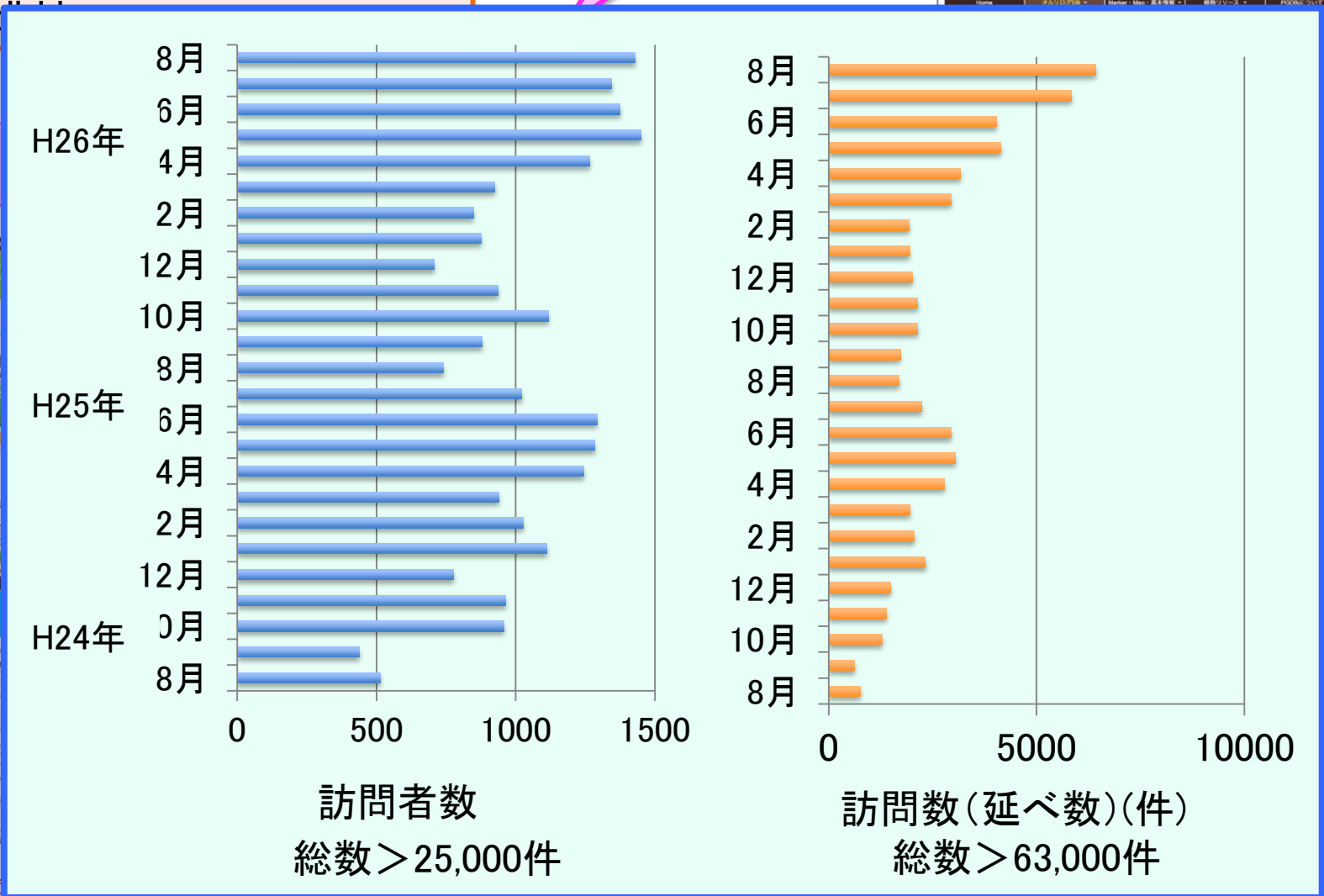
PGDBj内のDBの依存関係(～2013年度)



PGDBj ポータルサイト
Plant Genome DataBase Japan
<http://pgdb.jp>

横断検索機能
英語版PGDBjを公開

リソース



訪問者数
総数 > 25,000件

訪問数(延べ数)(件)
総数 > 63,000件

- ・平成23年度から25年度の統合化推進プログラムにおいて統合化のハブとして構築したポータルサイト**PGDBjのコンテンツを充実させる**
- ・統合化推進プログラムの**他課題との連携を深め、横断検索機能の拡充を図る**

植物ゲノム情報活用のための統合研究基盤の構築

- (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) DB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

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

(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

- ・現在のオルソログDBで用いられている**配列データの更新**
RefSeq Release 66 (July 2014)
- ・新たにゲノム配列が解読された**植物種の追加**
- ・オルソログテーブルや遺伝子情報ページ等について**表示システムの改善**
- ・RefSeqに加えて、**個々の植物関連DBで定義された遺伝子配列を利用**
- ・アノテーション高度化のため、**他のDBとのリンクを充実**

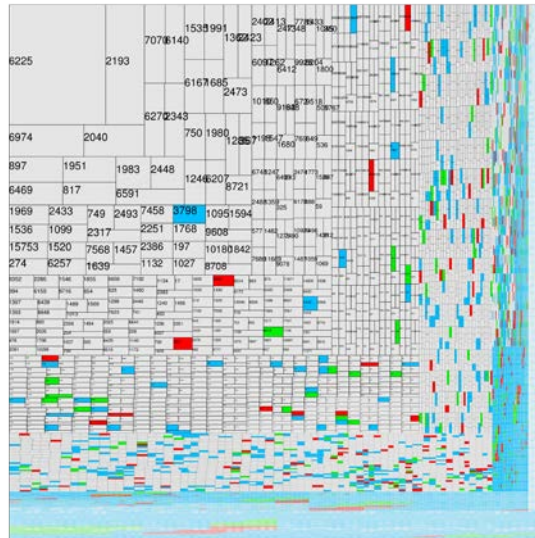
対象: ドメインや代謝経路関連DB、PGDBjのリソースやマーカーDB、
統合化推進プログラムその他課題、理研BRCのSABRE、等

■ シロイヌナズナ

×

■ イネ

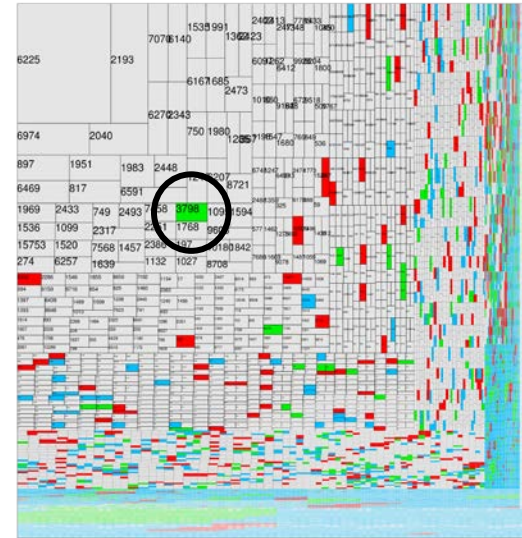
■ 両方で含有
■ 両方で欠損



■ シロイヌナズナ

×

■ ヒメツリガネゴケ

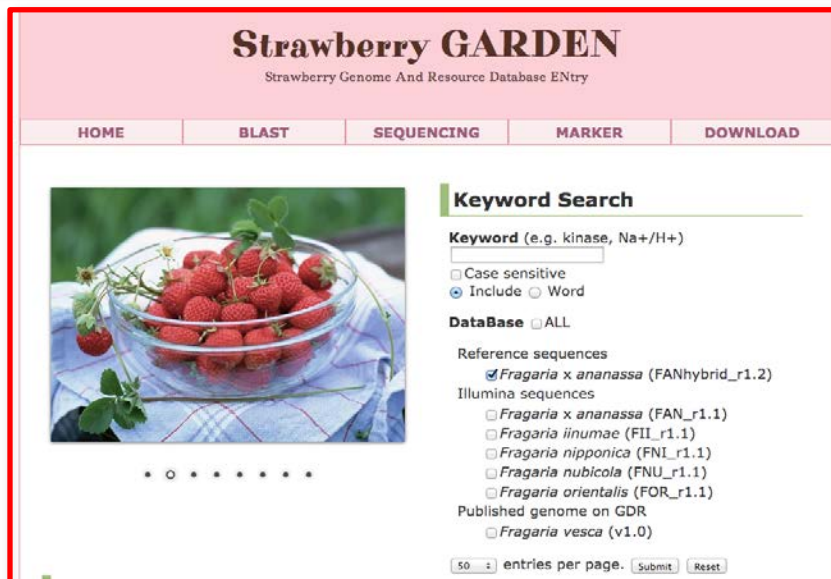


(2) 植物リソースDBの拡充による植物ゲノムDBの統合

- ・SABRE DB、NBRP DBに登録されている植物種に、**新たな植物リソース情報を追加**


対象: オランダイチゴ、カーネーション(ゲノム配列、DNAマーカー; かずさDNA研)、ダイコン(ゲノム配列、DNAマーカー; 東北大)、トマト(変異系統; 筑波大)、ナス(ゲノム配列、DNAマーカー; 農研機構)、スサビノリ(ゲノム・cDNA配列、DNAマーカー; 遺伝研、北海道大)、等

- ・リソースの配列情報に基づくオルソログDBとの連携を図り、**リソースと遺伝子情報の関連付けが行えるツールを整備、提供**



Strawberry GARDEN
Strawberry Genome And Resource Database ENtry

HOME BLAST SEQUENCING MARKER DOWNLOAD



Keyword Search

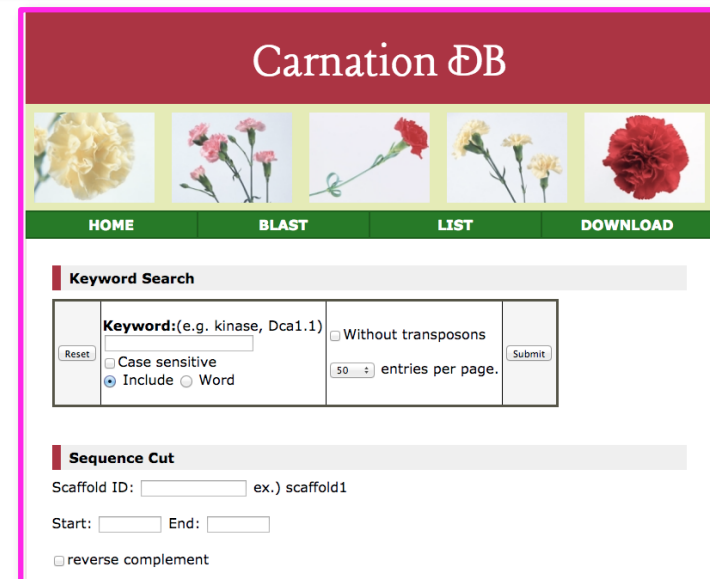
Keyword (e.g. kinase, Na+/H+)

 Case sensitive
 Include Word


DataBase ALL

Reference sequences
 Fragaria x ananassa (FANhybrid_r1.2)
 Illumina sequences
 Fragaria x ananassa (FAN_r1.1)
 Fragaria iinumae (FIL_r1.1)
 Fragaria nipponica (FNI_r1.1)
 Fragaria nubicola (FNU_r1.1)
 Fragaria orientalis (FOR_r1.1)
 Published genome on GDR
 Fragaria vesca (v1.0)

50 entries per page.



Carnation DB



HOME BLAST LIST DOWNLOAD

Keyword Search

Keyword:(e.g. kinase, Dca1.1) Without transposons
 Case sensitive Include Word 50 entries per page.

Sequence Cut

Scaffold ID: ex.) scaffold1
 Start: End:
 reverse complement

(3) DNAマーカー情報の拡充による植物ゲノムDBの統合

- ・国内外でゲノム解読やDNAマーカー整備が行なわれた55植物種に、**新たにゲノムが解読された植物等25種を加えて**、マーカーやQTL情報のキュレーションを継続
- ・**重要な農業形質(病害抵抗性, 環境ストレス耐性, 収量など)に関する遺伝子、アليل、DNAマーカー情報を文献から収集し**、ゲノム配列上に位置づけることによって配列情報に基づく検索・比較ツールを構築し提供

25種を新たに追加

キウイフルーツ、テンサイ、アンボレラ、ギニアアブラヤシ、キク、ルベラナズナ、アサ、サツマイモ、スイカ、ヒヨコマメ、ソラマメ、クリ、マレーヤマバショウ、ドイツトウヒ、カナダトウヒ、テーダマツ、セイヨウナシ、ナシ、ウメ、スイートオレンジ、キヌワタ、モウソウチク、ゴマ、アワ、ハウレンソウ

植物病理論文のキュレーション

取得する情報	項目
文献	PMID, DOI, journal, author, year
宿主生物	scientific name, variety (cultivar, line), trait
病原生物	scientific name, classification, pathogen code, common name, disease name
宿主遺伝子	gene name, accession number, clone name, similar gene with known function, protein, enzyme (EC number), metabolite
導入遺伝子	(same as above)
遺伝子の効果	effect (defense, induction), biotic treatment, inoculated organ, evidence code
病害抵抗性の詳細	free comment

(4)DB間の連携による統合化

- ・統合化推進プログラムの他課題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のデータのセマンティックウェブ(RDF)化を実施**
- ・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DBとの**連携を模索**

□ RDF化に基づくDBの統合

- PGDBj内部DBのコンテンツからRDFデータを生成
- RDFデータを格納するDB(トリプルストア)の運用
- 検索言語SPARQLの検討
- 検索インターフェースの開発

□ 横断検索システムへの検索APIを外部サイトへ配布し、DB間の連携を拡大

連携するDB

□ PGDBjの内部DB同士

□ 統合化推進プログラムの他課題

- 「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」
- 「生物種メタボロームモデル・データベースの構築」
- 「生命と環境のフェノーム統合データベース」

□ かずさDNA研究所が保有するDB

- ゲノム関連DB:ミヤコグサ、ヤトロファ、ユーカリ、トマト
- EST関連DB:トマト、シロイヌナズナ、ミヤコグサ、スサビノリ、アカクローバ、クラミドモナス

(5) オントロジーの整備による横断検索の効率化

- ・PGDBjの内部DB（オルソログ、リソース、マーカー）と外部DBコンテンツの連携強化を図るため、**Gene Ontology (GO)**や**Plant Ontology (PO)**に基づく記述を充実
- ・辞典や書籍から**オントロジー**として整備する用語をまとめる
- ・プロジェクト内**オントロジー整備ワーキンググループ**と**植物学会**が協働し、様々な分野の専門家から意見聴取を行いながら、**オントロジーの整備**を進める

(6) ゲノム情報のアノテーションの高度化

- ・PGDBjの内部DB（オルソログ、リソース、マーカー）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る**
- ・**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とのリンクを張る**
- ・**モチーフやタンパク質立体構造、代謝経路**など他のDBとリンクを張ることで、**アノテーションの高度化**を図る

植物ゲノム情報と横断検索機能を拡 充させたPGDBjの構築

植物学会との連携

- (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



例: "Pyruvic acid"

- (4) DB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

- 統合化推進プログラムの他の採択課題との連携
- 他の研究機関のDBとの連携
- 関連学会での広報活動
- 海外の植物関連DBとの連携