

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

植物ゲノム情報統合ポータルサイト Plant GARDENの使い方

かずさDNA研究所・東京大学・筑波大学

発表者：かずさDNA研究所・磯部祥子



Licensed under a Creative Commons表示4.0国際ライセンス
(c)2019 かずさDNA研究所

時代に即したDB開発：PGDBjからPlant GARDENへ



2011～2016年度で開発

格納データ

マーカー・QTL・連鎖地図

オルソログ（遺伝子配列の配列類似性情報）

表示

Google様の検索

ユーザーの専門性により表示を変える



2017年度から開発開始

2018年度末にベータ版公開

2020年度に正規版公開

格納データ

ゲノム・遺伝子配列

マーカー・QTL・連鎖地図

SNP

表示

検索メニューから選択

専門性で区別せず、シンプルな表示



格納されているコンテンツ（2021年12月現在）

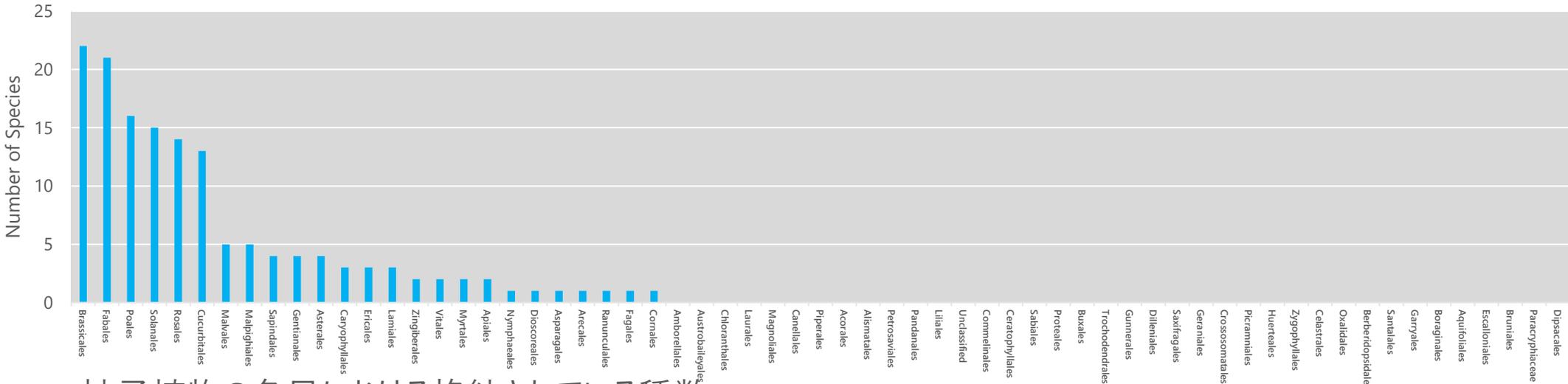
	植物種数	件数
ゲノム	134	169
遺伝子	126	8,136,812
DNAマーカー	34	313,115
形質と連関しているマーカー (QTL)	24	8,255
SNPs (SRAから検出した変異)	106	996

格納されている種の内訳

被子植物	147
緑藻植物	4
紅色植物	2
マゴケ植物	1
ゼニゴケ植物	1

被子植物では全64目（APGIV）のうち25目を格納

アブラナ > マメ > イネ > ナス > バラ



被子植物の各目における格納されている種数

Top PAGE ⇒ 各種のページへ

シンプルなデザイン
平易な表現
行いたいアクションを誘導

種ごとに検索したいケースが多いと想定して
設計

Lotus japonicus

この種について
科目: Fabaceae
マメ科
属名: Lotus
ミヤコグサ属

ゲノム配列をみる
すべて
▶ Lj2.5(Miyakojima MG-20)
▶ Lj3.0(Miyakojima MG-20)

遺伝子をさがす
すべて
▶ Lj2.5(Miyakojima MG-20)
▶ Lj3.0(Miyakojima MG-20)

マーカーをさがす
全て (1155)
SSR (1073)
CAPS (82)
SCAR (0)
その他 (0)

形質からさがす
全て (246)
ストルシス耐性 (28)
粒性 (4)
収量 (22)
形態・成熟 (70)
品質・成分 (86)
その他 (36)

その他の検索
BLAST 配列からさがす
SNPs 変異をさがす
連続地図

Plant GARDEN
Plant GARDEN (Genome And Resource Database Entry) は
さまざまな植物のゲノムやマーカー情報を集めたポータルサイトです

? 使い方 ▶ i このサイトについて ▶

例: Solanum lycopersicum kinase 検索

植物種からさがす
他の方法でさがす
解析してみよう

キクタンニギク
ブドウ欧州種
ミヤコグサ
シロイヌナズナ
ダイコン
イチゴ
ラッカセイ
ダイズ
トマト
すべてみる ▶

遺伝子から
BLAST 配列から
データ一覧
マーカー
形質
SNP Detection

ゲノム配列の検索



この種について

科目: Fabaceae
マメ科

属名: *Lotus*
ミヤコ

ゲノム配列をみる



すべて

一覧

- ▶ Lj2.5(Miyakojima MG-20)
- ▶ Lj3.0(Miyakojima MG-20)

主な配列バージョン

アセンブル配列 + Variant Call (後述) を実施したSRA

アセンブルされた配列										
ゲノム配列名	Bio Sample	アセンブリレベル	アノテーション (gff / gff3)	遺伝子 (cds / gene)	Hayai Annotation ZEN	トランスクリプト (cDNA / mRNA)	タンパク (pep / aa)	ダウンロード	JBrowse	コメント
Lj2.5	Miyakojima MG-20	Chromosome	●	●			●			
Lj3.0	Miyakojima MG-20	Chromosome	●	●	●	●	●			

SRA配列							
配列の種類	Bio Sample	配列名	Accession	JBrowse			コメント
				0	1	2	
全ゲノムシーケンス (ショートリード)	Miyakojima MG-20	DRA001236	DRX013249				
全ゲノムシーケンス (ショートリード)	Miyakojima MG-20	DRA001236	DRX013250				
全ゲノムシーケンス (ショートリード)	nrsym1-1	DRA005940	DRX090745				

アセンブルされたゲノム配列・遺伝子情報

 *Brassica oleracea*

配列名 : Brassica_oleracea.BOL

意外と便利情報



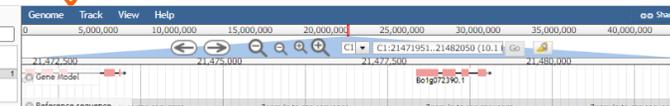
Index of /ja/download/t3712/t3712.G001

- Parent Directory
- Brassica_oleracea.BOL_42_gff3.gz
- Brassica_oleracea.BOL_cdna_all.fa.gz
- Brassica_oleracea.BOL_cds_all.fa.gz
- Brassica_oleracea.BOL_dna_toplevel.fa.gz
- Brassica_oleracea.BOL_dna_toplevel.fa.gz
- Brassica_oleracea.BOL_dna_rm_toplevel.fa.gz
- Brassica_oleracea.BOL_dna_sm_toplevel.fa.gz
- Brassica_oleracea.BOL_ncrna.fa.gz
- Brassica_oleracea.BOL_pep_all.fa.gz
- Brassica_oleracea_t3712.G001_zen_v1.0.gz

ゲノム配列の詳細			
配列名	Brassica_oleracea.BOL	系統名	TO1000DH3, kale-like type
配列数	32,928	染色体数	2n=18
配列長 (bp)	488,622,507	N50長 (bp)	48,366,697
シーケンシングの方法	Illumina, 454	取得した配列量	94x
アセンブリ方法	SOAPdenovo v. 1.05	推定ゲノムサイズ (Mb)	648
シーケンシングの方法のコメント		コメント	Kale-like type. A doubled haploid line TO1000DH3 derived from <i>B. oleracea</i> cultivar TO1424

Available Tracks

- Filter tracks
- Gene Model
- Reference sequence
- Reference sequence



論文 (DOIコード)	10.1186/gb-2014-15-6-r77
データソース名	EnsemblePlants

遺伝子をさがす

キーワードを入力して検索

検索 ▶ 詳細検索

Agri-Food
orsite Evr

遺伝子配列のキーワード検索

tml

詳細情報

ゲノム配列を切り出す

染色体 / Scaffolds: C1

開始位置: 1000

終了位置: 2000

相補鎖を出力する

リセット 配列を得る

配列の切り出し

検索機能の改良

検索システムの実装はユーザーの方が思うより、結構大変
Plant GARDENもまだまだ改良の余地があるのですが、地道な改善を続けています

遺伝子配列（遺伝子をさがす、で検索）
はキーワード、もしくは機能（GO, KEGGで検索）

Lotus japonicus

この種について
科目: Fabaceae
マメ科
属名: Lotus
ミヤコグサ属

ゲノム配列をみる
すべて

- Lj2.5(Miyakojima MG-20)
- Lj3.0(Miyakojima MG-20)

遺伝子をさがす
すべて

- Lj2.5(Miyakojima MG-20)
- Lj3.0(Miyakojima MG-20)

マーカーをさがす
全て (1155)

形質からさがす
全て (246)

その他の検索

キーワード検索

キーワードを入力して検索

リセット 検索

Plant GARDEN
Plant GARDEN (Genome And Resource Database ENtry) は
さまざまな植物のゲノムマーカー情報を集めたポータルサイトです

検索: Solanum tuberosum kinase

GO

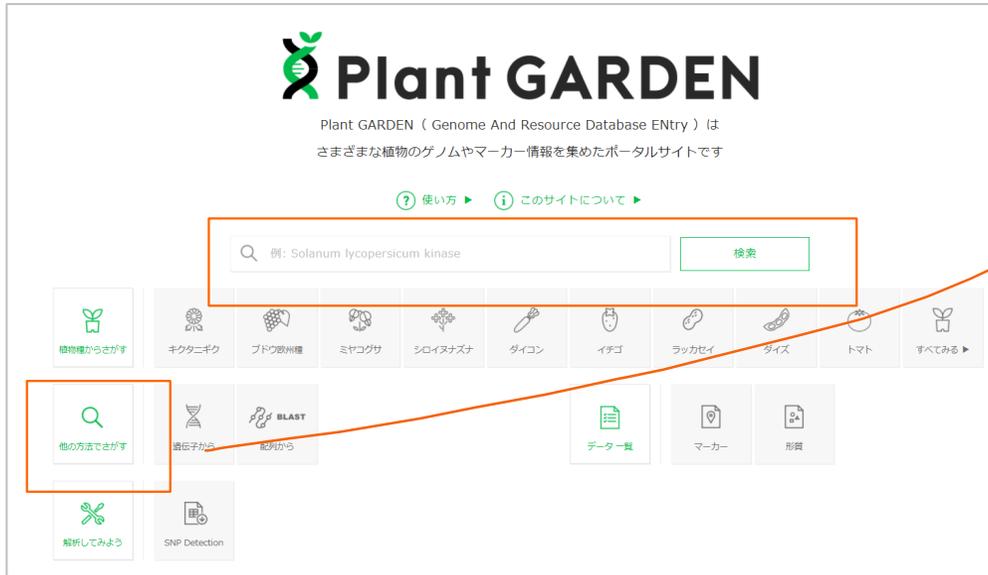
search ID / keyword

- alpha-1,6-mannosyltransferase activity | GO:0000009 (1)
- single strand break repair | GO:0000012 (2)
- phosphopyruvate hydratase complex | GO:0000015 (4)
- alpha-1,2-mannosyltransferase activity | GO:0000026 (4)
- ribosomal large subunit assembly | GO:0000027 (1)
- ribosomal small subunit assembly | GO:0000028 (3)

GO

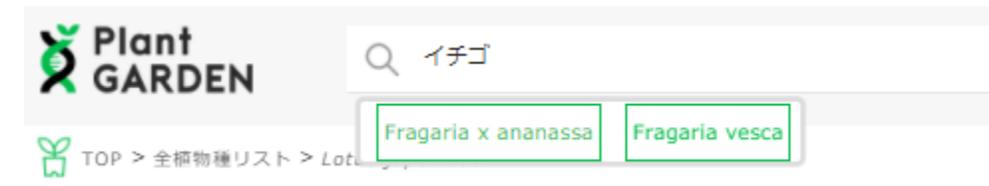
検索

検索機能の改良



Top Pageの検索窓は全データに対して検索

「他の方法でさがす」は「全データ」「遺伝子」「マーカー」のカテゴリ別に検索



全データ

検索結果: 全 310123 件中 50 件表示

Download

ヒットしたアイテム	種名	データのカテゴリ	詳細
CSS0036789.1	<i>Camellia sinensis</i>	遺伝子	Transcription factor, MADS-box
GWHPABKB002085	<i>Camellia sinensis</i>	遺伝子	Transcription factor, MADS-box
Solan12n033910.1	<i>Solanum pennellii</i>	遺伝子	Pectinesterase

遺伝子名	種名	ゲノム配列名	アノテーション (Ortho DB)
CSS0036789.1	<i>Camellia sinensis</i>	CSS_Ch Lev_2020050 6	Transcription factor, MADS-box
HPABKB002085	<i>Camellia sinensis</i>	DASZ	Transcription factor, MADS-box

遺伝子

マーカー

ヒットしたアイテム	種名	データのカテゴリ	詳細
SL20153_73	<i>Solanum lycopersicum</i>	マーカー	Others
TES4269	<i>Solanum lycopersicum</i>	マーカー	SSR
TES4238	<i>Solanum lycopersicum</i>	マーカー	SSR

ヘッダーの検索では候補wordの表示

遺伝子の再アノテーション

多数の種×異なるゲノム由来の遺伝子情報 → 比較したい

公開されている遺伝子配列のアノテーションの方法やアノテーションされる項目は実施者によってまちまち

→そのままでは比較が難しいので、格納している遺伝子配列全てを再アノテーション

→オリジナルのアノテーション情報はgffファイルをダウンロードサイトに格納

OrthoDB/Uniprot/Refseqを対象に配列類似性のSearch

遺伝子配列の詳細			
物理上の位置	遺伝子名	CDS	Lj4g3v1287420.2
		PEP	Lj4g3v1287420.2
	ゲノム配列上の位置	染色体	Lj3.0_chr4
		開始位置	19018818
		終了位置	19025821
		配列の向き	-
アノテーション情報 (Hayai Annotation ZEN)	KusakiDB	KusakiDB Accession	KUNI01299171
		KusakiDB バージョン	2
		参照DBにおける蛋白質/ 転写産物配列の有無	Yes
		DB間における配列の共通性	orthodb-refseq-uniprot

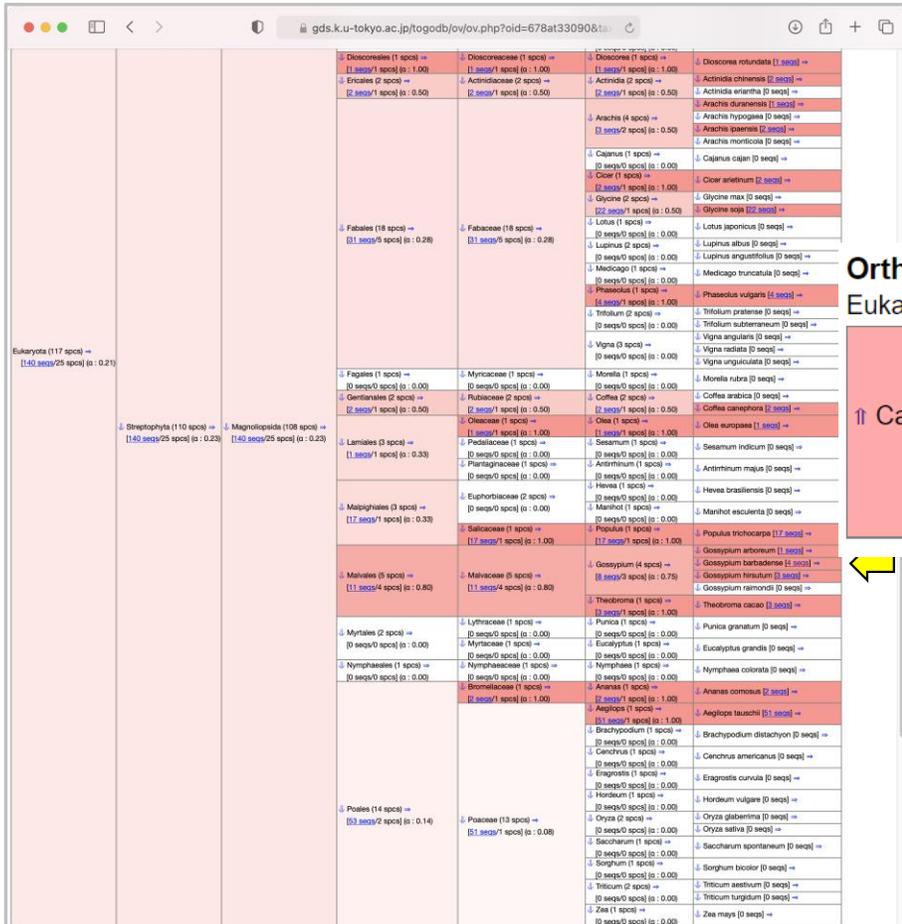
OrthoDB	DBのバージョン	10.1
	配列の説明	Tyrosyl-DNA phosphodiesterase I
	Gene ID	3880_0:00ba85
	Ortholog group unique ID	13760at33090
Functional annotation	出典元のDB	trembl
	蛋白質名	Putative phosphodiesterase I
	Accession	G7LGI2 詳細
	Uniprotアクセッション に対応するRefseqアクセ ッション	XP_003630434
	EC	3.1.4.1
	KEGG エントリー	NA
	KO ID	NA

Ortholog group を介して他種の遺伝子との類似性がわかる

開発中：Ortho group Unique IDを介した種間のオルソログ配列の検索

Ortho group ID毎にオルソログ配列を有している種のリストを作成
(Plant GARDENに格納されている配列が対象)

ドメイン 門 綱 目 科 属 種



Ortholog ID:678at33090 (Caryophyllales)

Eukaryota ← Streptophyta ← Magnoliopsida ← Caryophyllales

↑ Caryophyllales (3 spcs) ⇒ [3 seqs/2 spcs] (α : 0.67)	↓ Amaranthaceae (2 spcs) ⇒ [1 seqs/1 spcs] (α : 0.50)	↓ Amaranthus (1 spcs) ⇒ [0 seqs/0 spcs] (α : 0.00)	↓ Amaranthus hypochondriacus [0 seqs] ⇒
	↓ Polygonaceae (1 spcs) ⇒ [2 seqs/1 spcs] (α : 1.00)	↓ Beta (1 spcs) ⇒ [1 seqs/1 spcs] (α : 1.00)	↓ Beta vulgaris [1 seqs] ⇒



SNP情報

公開されているSRAからカバレッジが十分あるなどの指標で選抜した配列を格納しているゲノム配列に対して Mapping→Variant検出

 *Solanum lycopersicum*

ID	集団名	解析個体リスト	変異の種類	参照配列	配列の種類	SNP数	データ作成者	コメント	ダウンロード		TASUKE+	JBrowse
									表	Vcf file		
t4081.SNP000001	Solanum_lycopersicum_SRA_100_lines_20191118		SNPs, In/dels	SL3.0	WGS	2,191,117	Plant GARDEN	Bwa/bcftools 1.9, (QUAL≥200, 10≤DP≤200, max-missing=0.7)				

 *Solanum lycopersicum*

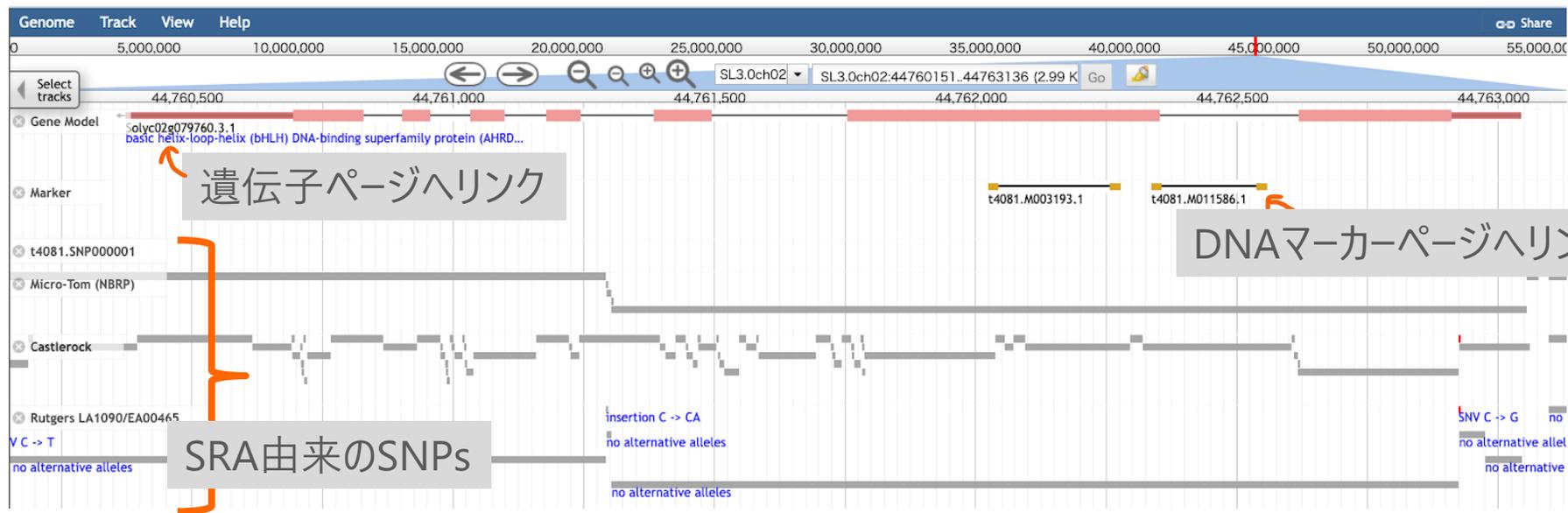
Bio Sample	配列名	Accession	JBrowse	gVCF
Prudens Purple	SRA861922	SRX5538730		
Silvery Fir Tree	SRA861922	SRX5538729		
Jerry's German Pink	SRA861922	SRX5538728		
M82	SRA861922	SRX5538727		
Glacier	SRA861922	SRX5538726		

各SRAのgvcfファイル

ブラウザを介した異なるカテゴリのデータ比較

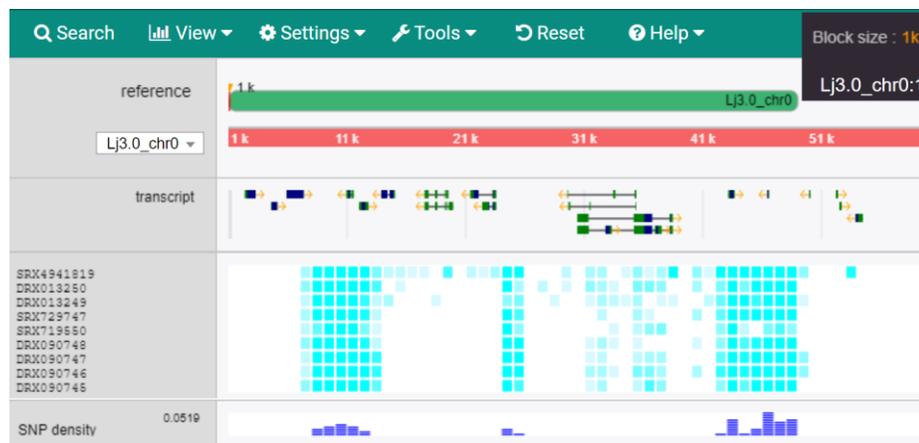


- 異なるカテゴリのデータ比較
- ・ゲノム配列
 - ・遺伝子
 - ・SNP
 - ・マーカー



TASUKE
(SNPページからのみアクセス)

ダウンロード		TASUKE+	JBrowse
表	Vcf file		



解析ツールの提供

解析ツールを選ぶ

-  お持ちのデータをアップロードしてブラウザ上で解析する
-  解析ツールをダウンロードしてローカル環境で解析する

アカウント登録をしてログインします

配列解析システム 1.03	
アカウント	sisobe
パスワード
<input type="button" value="ログイン"/>	
●アカウントの新規登録はここをクリック	

パスワードを忘れてしまった場合はplantgarden-at-kazusa.or.jpまでお問い合わせ下さい。(データはリセットされます)

SNP Detection	SNPやIn/Delを検出する解析パイプライン	 
Hayai-Annotation	遺伝子配列のアノテーション	 
Hayai-Annotation Plants v2.0	植物種に特化した機能アノテーションシステム	 
KusakiDB v1.0	タンパク質同族体群の検証と完全性のための新しいアプローチ	 

Launching Hayai_annotation_v2.0



ユーザーごとにDocker containerを自動で作成

Hayai-Annotation Plants v2.0 - Functional Annotation System Specialized in Plant Species

Database: KusakiDB v1.0
Reference: Hayai-Annotation Plants

Hayai Annotation Search Proteins

Perform Full Annotation

Query Sequence Type
 Protein
 DNA

Minimum Sequence Identity (%)
Slider: 20 to 100 (set to 50)

Minimum Query Coverage (%)
Slider: 20 to 100 (set to 50)

Maximum E-value
Slider: 1 to 100 (set to 1)

Minimum Target Coverage (%)
Slider: 20 to 100 (set to 50)

今後の開発

① スマホ版

The screenshot displays the Plant GARDEN mobile application interface. At the top, there is a search bar and the Plant GARDEN logo. Below this, a breadcrumb trail shows 'TOP > Cucumis sativus > ゲノム情報'. The main content area features a section for 'キュウリ, キュウリ半野生種 (Xishuangbannanensis変種), キュウリ野生種 (hardwickii変種) Schrenkiella parvula (formerly Thellungiella parvula)'. Below this, there is a section for 'アセンブリされた配列' (Assembled sequences) with a table showing genome assembly details. The table has columns for 'Genome Accession', 'Bio Sample', 'Assembly Level', 'DL', and 'JBrowse'. Two rows are visible, both for 'Lj2.5' with 'Miyakojima MG-20' as the bio sample and 'Chromosome' as the assembly level. Below this, there is a section for 'SRA配列' (SRA sequences) with a table showing SRA data. The table has columns for 'Sequence Type', 'Bio Sample', 'Accession', and 'JBrowse'. Five rows are visible, all for 'WGS' with 'Miyakojima MG-20' as the bio sample and 'DRA001236' as the accession. At the bottom, there is a 'リンク' (Links) section with various links and a 'さがす' (Search) section with filters for '植物種からさがす', '遺伝子からさがす', and '配列からさがす BLAST'. There is also a '解析' (Analysis) section with a link to '解析してみよう'.

② 解析ツール

Plant GARDEN上の配列を
ReferenceとしたSNP解析

現在はかずさDNA研に設置し
たサーバーにデータをUpload

→パイプラインをコンテナ化し、
ローカルPCで解析するツール



③ コンテンツの拡充

ゲノム・遺伝子情報
DNAマーカー/QTL情報

本年度末までの期間でキュレーターさん
随時募集中

論文/ゲノム情報ごとにお支払い
決まったノルマなどなし

対象は植物のみ

ご意見・ご要望をお待ちしています
plantgarden@Kazusa.or.jp