AJACS オンライン 必要な配列情報を入手する

植物ゲノムポータルサイト 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

Plant GARDENについて

統合化推進プログラムにおいて 2種類のDBを開発

https://plantgarden.jp

http://pgdbj.jp/



H23~28 年度に開発

格納データ マーカー オルソログ (遺伝子配列の配列類似性情報)

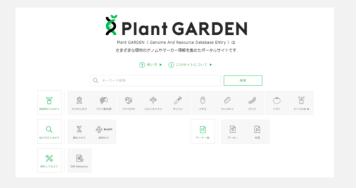




Ž Plant GARDEN

H29年度から開発開始

格納データ ゲノム・遺伝子配列 マーカー 遺伝子配列の類似性情報





GARDEN = Genome And Resource Database ENtry

覚えやすさを意識して命名

https://plantgarden.jp

B版公開:2019年3月

正規版公開:2020年7月

日本語版

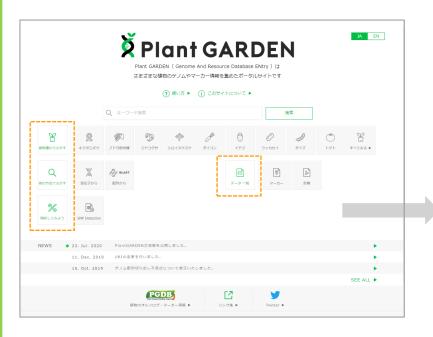


英語版





Top PAGE



- □ 「行うべきアクション」が分かりやすい項目表示 4つのメニュー表示 シンプルなデザイン・平易な言葉
- □ 想定するユーザー生物研究者/育種関係者 +教育機関
- 高精度なゲノム配列が公開されている植物種は可能な 限り網羅する方針

アクセス頻度の高い種をTopに表示(未実装) 2020年10月現在:119種の植物の全ゲノム・遺伝子情報を格納



格納されている主なコンテンツ 2020年10月現在

	植物種数	件数
ゲノム	119	133
遺伝子	119	5,890,957
DNAマーカー	34	287,703
形質と連関しているマーカー(QTL)	24	8,217
SNPs (SRAから検出した変異)	61	3,812

おことわり

現在、検索システムを調整中で、約2週間後に調整が終了する見込みです。

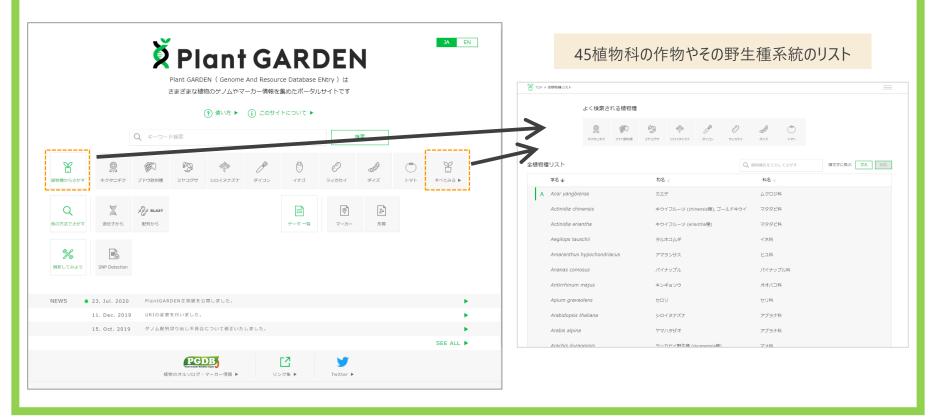
現在のバージョンでは意図した検索結果が出力されないケースもあります。

ご不便をおかけしますが、ご了承ください。

植物種からさがす

植物種からさがす

トップページの「植物種からさがす」または「すべてをみる」ボタンをクリック



全植物種リスト

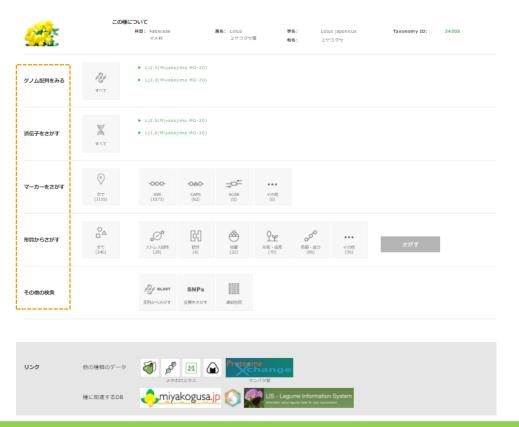
https://plantgarden.jp/ja/list/species



植物種別ページ

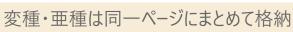
https://plantgarden.jp/ja/list/species





- □ ゲノム・遺伝子・SNPsは公開されている配列をキュレーション・解析してPlant GARDENに格納
- □ アセンブルされたゲノム配列はバージョン毎に表示
- □ DNAマーカー、形質と連関しているマーカー、連鎖地図はPGDBjでキュレーションしたコンテンツを移行中
- □ 各種に対応するメタボロミクス・タンパク質DBへのリンク
- □ 各種で整備されている代表的なDBの リンクを整備中

ゲノム配列情報 変種・亜種の取り扱い





アセンブルされたゲノム配列の情報

Brassica oleracea

DOWNLOAD



配列名:Brassica_oleracea.BOL

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 B. oleracea cultivar TO1434.		
論文 (DOIコード)	10.1186/gb-2014-15-6-r77	青任著者 系紛	Agri-Food Preside Evry Val d Essonile), Andrew G Sharpe (National Research Council Canada)		
データソース名	EnsemblePlants	データソースURL	http://plants.ensembl.org/index.html		

配列のダウンロード

Brassica oleracea

DOWNLOAD

配列名: Brassica_oleracea.BOL

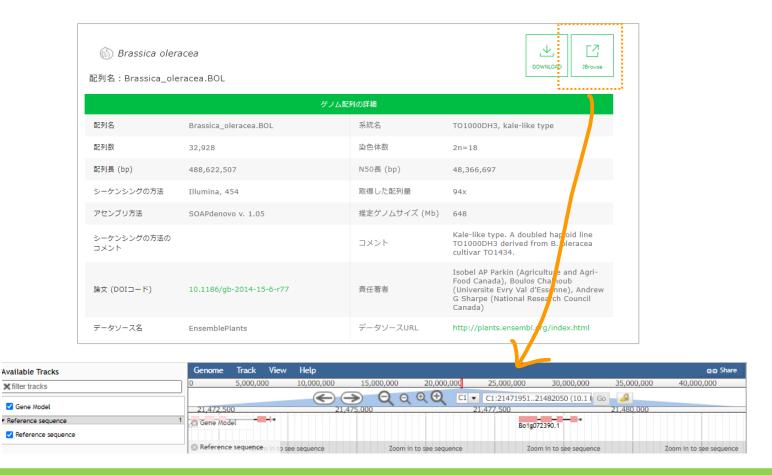
ゲノム配列の詳細						
配列名	Brassica_oleracea.BOL	系統名	TO1000DH3, kale-like type			
配列数	32,928	染色体数	2n=18			
配列長 (bp)	488,622,507	N50長 (bp)	48,366,697			
シーケンシングの方法	Illumina, 454	取得した配列量	94x			
アセンブリ方法	SOAPdenovo v. 1.05	推定ゲノムサイズ (Mb)	648			
シーケンシングの方法の コメント		コメント	Kale-like type. A doubled haploid line TO1000DH3 derived from B. oleracea cultivar TO1434.			
論文 (DOIコード)	10.1186/gb-2014-15-6-r77	責任著者	Isobel AP Parkin (Agriculture and Agri- Food Canada), Boulos Chalhoub (Universite Evry Val d'Essonne), Andrew G Sharpe (National Research Council Canada)			
データソース名	EnsemblePlants	データソースURL	http://plants.ensembl.org/index.html			



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
- 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.ncma.fa.qz
- Brassica oleracea.BOL.pep.all.fa.gz
- Brassica oleracea t3712.G001 zen v1.0.gz

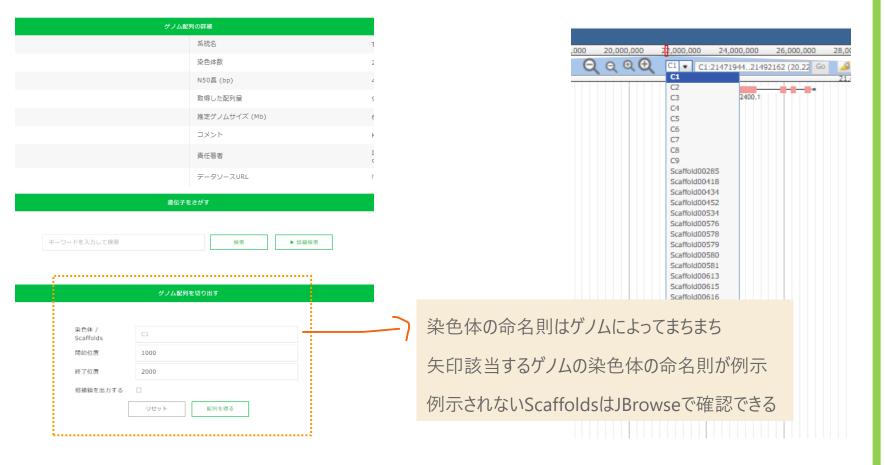
JBrowseへの表示



x filter tracks

Gene Model

配列の切り出し



ゲノム配列一覧





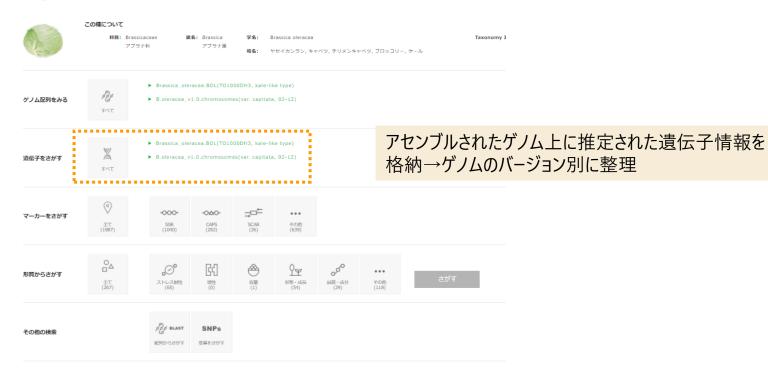
アセンブル配列された配列の異なる種類の配列一覧

SRA配列					
配列の種類	Bio Sample	配列名	Accession	JBrowse	コメント
HU/1021E3R	bio Sample	HG2943		B.oleracea_v1.0.chromosomes	3221
全ゲノムシーケンス (ショートリード)	sample_BKL_3712	ERA1462833	ERX2582715		
全年 ハベーケンフ パンコートリード)	camnle RKI 3712	FR41462833	FRY2582716	F2	



遺伝子配列の検索

Brassica oleracea



遺伝子キーワード検索





rassica_oleracea.BOL(TO1000DH3, kale-like type)

B.oleracea_v1.0.chromosomes(var. capitata, 02-12)

例えばゲノムのバージョンを選ぶと該当するゲノム 内で絞り込んだキーワード検索ができる

Q TOP > 検索 > 遺伝子 > Brassica oleracea > Brassica_oleracea.BOL

キーワード検索 ▶ 事前絞り込み条件 【植物種名 Brassica oleracea (ID:t3712) 【配列名 Brassica_oleracea.BOL (ID:t3712.G001) kinase リセット 検索 注意!現在、Safariでは検索が安定しません

遺伝子キーワード検索(ゲノムバージョン毎)

Brassica oleracea

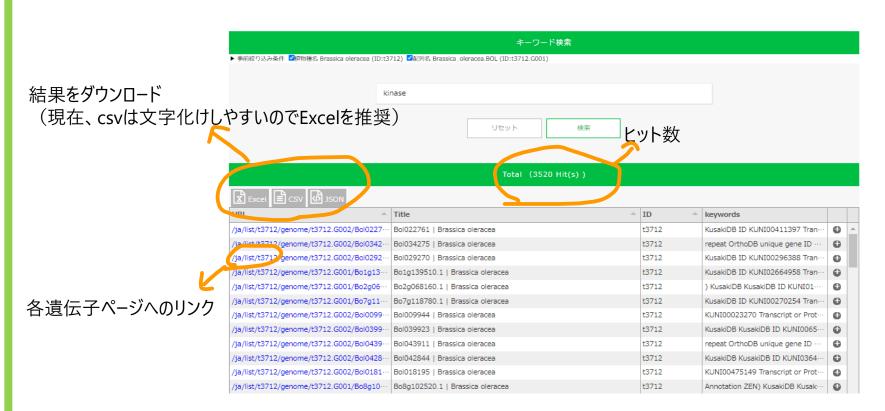
配列名: Brassica_oleracea.BOL





ゲノム配列を切り出す

遺伝子検索結果



検索結果の表示項目については、修正検討中です。

遺伝子のページ

Brassica oleracea

配列名:B.oleracea_v1.0.chromosomes



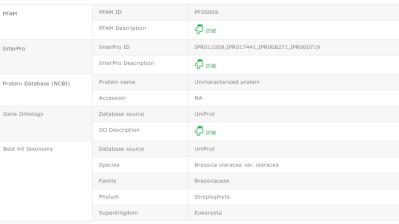
			遺伝子配列の詳細		
物理	物理上の位置	遺伝子名	CDS	Bol022761	
			PEP	Bol022761	
		ゲノム配列上の位置	Chromosome	C07	
			Start Position	7226852	
			End Position	7230434	
			Strand	+	
アノ	/テーション情報	KusakiDB	KusakiDB ID	KUNI00411397	
(Ha	yai Annotation ZEN)		Transcript or Protein existence	YES	
		OrthoDB	Protein name	Serine/threonine-protein kinase	
		OrthoDB unique gene ID	3711_0:0046d6		
			OG unique ID	95186at33090	PFAM
		UniProt	Protein name	Protein kinase domain-containing protein	PFAM
			Accession	A0A0D3CM56	
			Comment	NA	InterPro
			EC	NA	
			KO ID	NA	Protein Databa
			KO Description	NA	
					Cono Ontologo

つづき

遺伝子アノテーションはDB内の各遺伝子のアノテーション項目とアノテーション基準を統一させるために、オリジナルのアノテーションではなく、独自に開発したアノテーションツールにより再アノテーションしています。

Hayai-Annotation

Ghelfi et al. (2019) Bioinformatics 35:4427-4429



アノテーション情報

△ Group/Domain

95186at33090

△ Name

Pkinase

Kinase-like_dom_sf

Ser/Thr kinase AS

Prot kinase dom

ATP binding

Protein_kinase_ATP_BS

Protein kinase domain-containing protein

protein serine/threonine kinase activity

Serine/threonine-protein kinase

KusakiDB	KusakiDB ID	KUNI00411397	
	Transcript or Protein existence	YES	
OrthoDB	Protein name	Serine/threonine-protein kinase	
	OrthoDB unique gene ID	3711_0:0046d6	
	OG unique ID	95186at33090	
UniProt	Protein name	Protein kinase domain-containing protein	
	Accession	A0A0D3CM56	
	Comment	NA	
	EC	NA	
	KO ID	NA	
	KO Description	NA	
PFAM	PFAM ID	PF00069	
	PFAM Description	₽詳細	Excel Scsv (b) 150
InterPro	InterPro ID	IPR011009,IPR017441,IPR008271,IPR000719	▼ UniProt Annotation (1 its
	InterPro Description	₹細	□ A0A0D3CM56 ▼ OrthoDB (1 item)
Protein Database (NCBI)	Protein name	Uncharacterized protein	□ 3711_0:0046d6 ▼ Pfam (1 item)
,	Accession	NA	□ PF00069 ▼ InterPro (4 items)
Gene Ontology	Database source	UniProt	☐ IPR011009 ☐ IPR017441
	GO Description	₩	☐ IPR008271 ☐ IPR000719
Best hit taxonomy	Database source	UniProt	▼ Gene ontology (2 items)
best me taxonomy	Species	Brassica oleracea var. oleracea	G0:0005524 G0:0004674
	Family	Brassicaceae	Search Checked ID
	Phylum	Streptophyta	
	Superkingdom	Eukaryota	

アノテーション用に開発したDB Kusaki DB (under preparation)

OrthoDB、UniProt, PFAM, InterPro, Protein DB, GOによるアノテーション

アノテーション時に参照したDB (UniProt もしくはProtain DB)でベストヒットだった種

_ Link

UniProt

Pfam

OrthoDB(SIB)

InterPro(EBI)

InterPro(EBI)

InterPro(EBI)

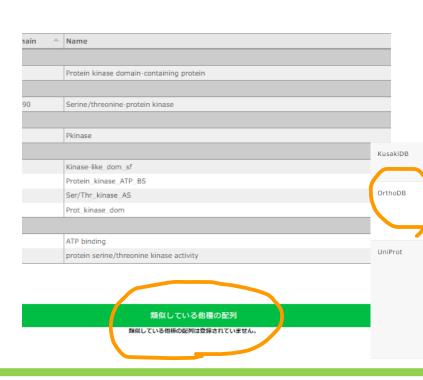
InterPro(EBI)

QuickGO(EBI)

OuickGO(EBI)

他種との配列比較 (予定)

アノテーションの詳細



今後の開発項目として、OrthoDB(オルソログ配列のDB)中のオルソログ情報を利用して、配列が類似する他種の遺伝子情報を表示することを計画している

KusakiDB ID	KUNI00411397
Transcript or Protein existence	YES
Protein name	Serine/threonine-protein kinase
OrthoDB unique gene ID	3711_0:0046d6
OG unique ID	95186at33090
Protein name	Protein kinase domain-containing protein
Accession	A0A0D3CM56
Comment	NA
EC	NA
KO ID	NA
KO Description	NA

遺伝子配列の取得



Brassica oleracea

osomes

	遺伝子配列の詳細	
	CDS	Bol022761
	PEP	Bol022761
の位置	Chromosome	C07
	Start Position	7226852
	End Position	7230434
	Strand	+
	KusakiDB ID	KUNI00411397
	Transcript or Protein existence	YES
	Protein name	Serine/threonine-protein kinase
	OrthoDB unique gene ID	3711_0:0046d6
	OG unique ID	95186at33090
	Protein name	Protein kinase domain-containing protein
	Accession	A0A0D3CM56
	Comment	NA
	EC	NA

	CDS 配列
遺伝子名	Bol022761
說明	[mRNA] locus=C07:7226852:7230434:+
R:FI	ATORATTACTOCKTOCKTTTATTCCKTGARAGGARAGGARAGCTCCCARGGATCAGAT ARCAGGACCOCCOGRAGGARAGCCAGGACGACGACGACCAGGACCAGGACAGACCACCC AGAGTTCCAGCTATAGCTCAGGGTTAGGAGGACAGACAGA

	アミノ酸配列
遺伝子名	Bol022761
記見明	[mRNA] locus=C07:7226852:7230434:+ [translate_table: standard]
R291	MRYFSCFYFHEKRATPKOSUNNSRABNGEVTGRONNKTHEELFITTEVKKHKUNDAEKET TINNTAAGTFTFELTATATHSTGEGLIVEGURSGLGONNETUSUULUSLIHKEHUNL TGKGAGGOGLIVETENISKGESLUSHLLDLTPGOFFELTATATATATATATATATATATATATATATATATATATA

	Transcript 配列
遺伝子名	NA
説明	NA
配列	NA

遺伝子配列のJBrowse表示 遺伝子配列 osomes Accessors and a second 遺伝子配列の詳細 CDS Bol022761 PEP Bol022761 上の位置 C07 Chromosome 10,000,000 15,000,000 20,000,000 25,000,000 30,000,000 35,000,000 40,000,000 C07 • C07:7226134..7231150 (5.02 Kb) Go 3 7,227,500 7,228,750 Bol022761 Protein name Serine/threonine-protein kinase OrthoDB unique gene ID 3711_0:0046d6 OG unique ID 95186at33090 Protein name Protein kinase domain-containing protein A0A0D3CM56 Accession Comment NA EC NA

遺伝子機能からの詮索



- Brassica_oleracea.BOL(TO1000DH3, kale-like type)
- B.oleracea_v1.0.chromosomes(var. capitata, 02-12)

遺伝子機能(GO, KEGG)からの検索も可能 (Hayai-annotationによるアノテーション結果に基づきます)



DNAマーカー情報の検索

Brassica oleracea

該当するマーカー数:1956

Brassica oleracea



この種について

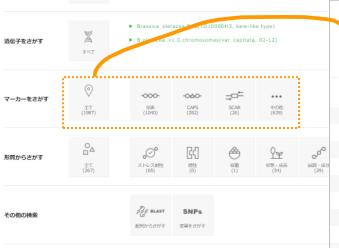
料目: Brassicaceae アプラナ科 具名: Brassica アブラナ風 学名: Brassica oleracea

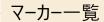
1名: ヤセイカンラン, キャベツ, チリメンキャベツ, ブロッコリー, ケール

Taxonomy I

一覧のダウンロード

PCRベース(プライマー情報あり)のマーカーを格納





1 2 3 4 5 … ▶

1-50 件表示 ▼

マーカーID 🔟	マーカー名 ↓	タイプ ↓	Fw プライマー配列 開始位置		
		912 (配列名 ↓	染色体 🔱	開始位置 🔟
t3712.M000001.1	BoAP1-B	Others	NA	NA	0
t3712.M000002.1	bo2	Others	NA	NA	0
t3712.M000003.1	BoCAL-a	CAPS	BOL_v1.0	C03	40353865
t3712.M000004.1	BoAP1-a	Others	NA	NA	0
t3712.M000005.1	BoCAL-a_SSR	SSR	BOL_v1.0	C03	40355283
t3712.M000006.1	bo4	Others	NA	NA	0
t3712.M000007.1	bo5	Others	BOL_v1.0	C06	40488687
t3712.M000008.1	bo10	Others	NA	NA	0
t3712.M000009.1	bo12	Others	BOL_v1.0	C05	3915816
+2712 M000010 1	ho12	Othoro	NIA.	NIA.	^

各マーカーのページへ

DNAマーカーのページ

Brassica oleracea



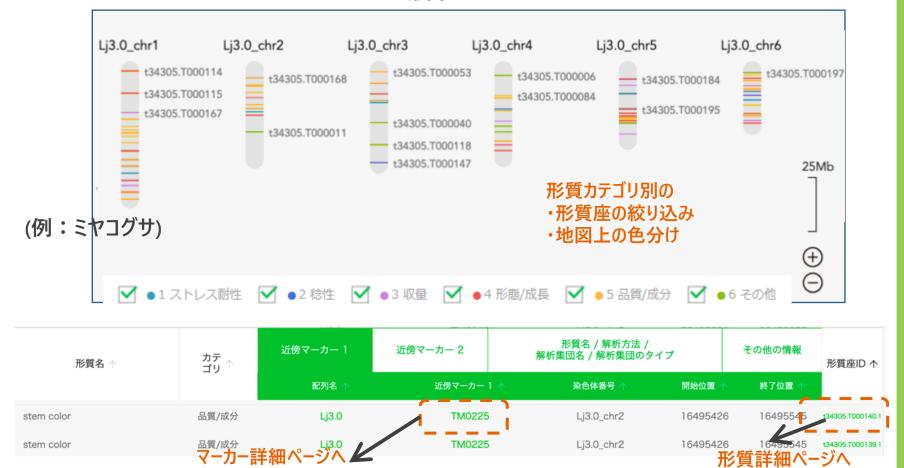
マーカーID		t3712.M000003.1
マーカー名		BoCAL-a
タイプ		CAPS
物理上の位置	配列名	BOL_v1.0
	染色体	C03
	Fw プライマー配列	CATACAAACGCAAACATCTC
	開始位置	40353865
	終了位置	40353884
	Rv プライマー配列	GGACAAAGCCCACCTACATG
	開始位置	40354608
	終了位置	40354589
ターゲット配列		NA
アレル [検出された系統]		digested [var. botrytis (cauliflower)]; digested and not digested bands [var. botrytis (cauliflower) x var. italica (broccoli)]; not digested [var. italica (broccoli)]
制限酵素		SpeI
コメント		
文献 (DOI)		10.1023/A:1011370525688
文献 (PubMed)		NA NA

形質と連関するマーカー情報 の検索

Brassica oleracea



形質ページ



形質一覧ページからの マーカー詳細ページ表示



形質詳細ページ

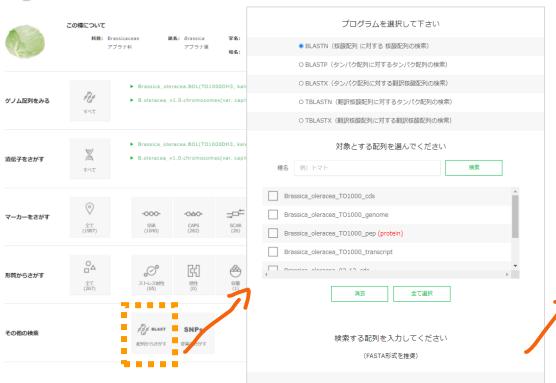
- 現在はPGDBj(Plant Genome DataBase japan, http://pgdbj.jp/)のデータを再整理・移行中
- 今後はキュレーションによりコンテンツを拡充

形質詳細ページ

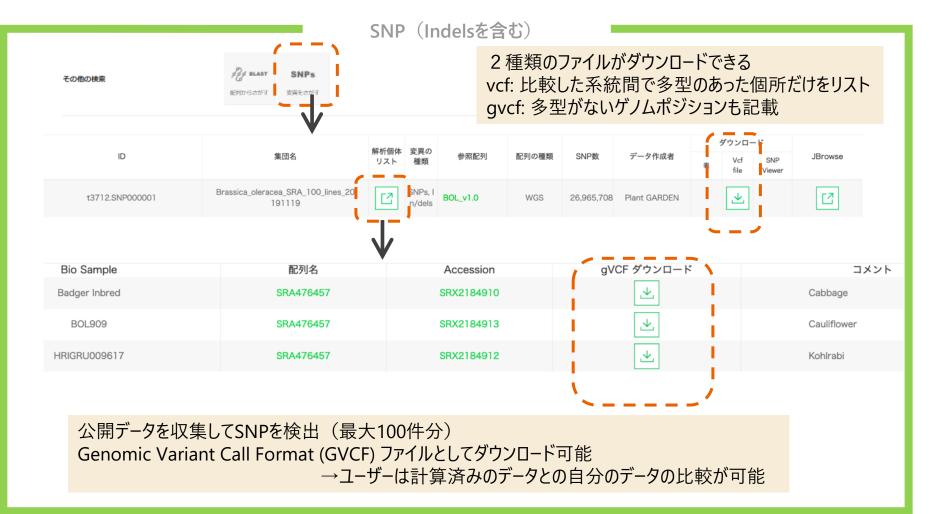


BLAST検索

Brassica oleracea







Topページから

その他の検索





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



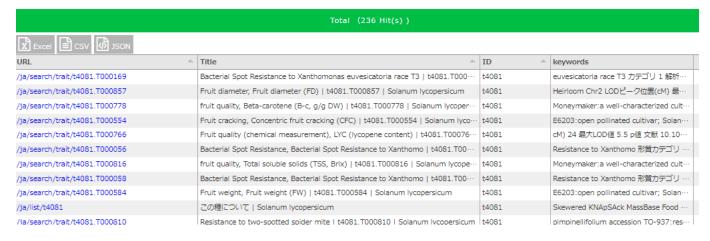
Top Pageからのキーワード検索

Ž Plant GARDEN

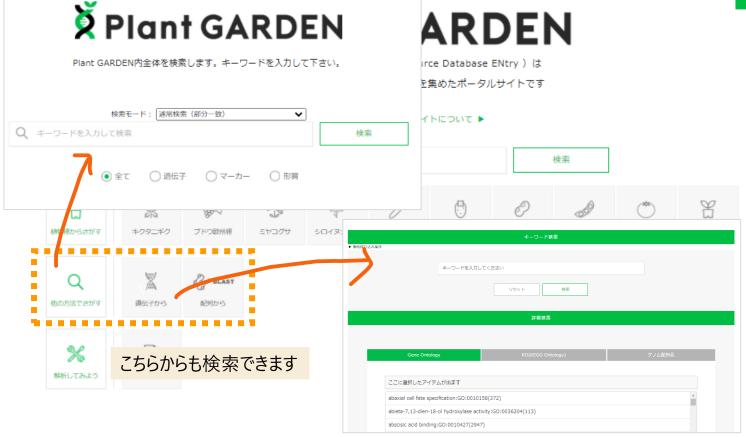
Plant GARDEN内全体を検索します。キーワードを入力して下さい。



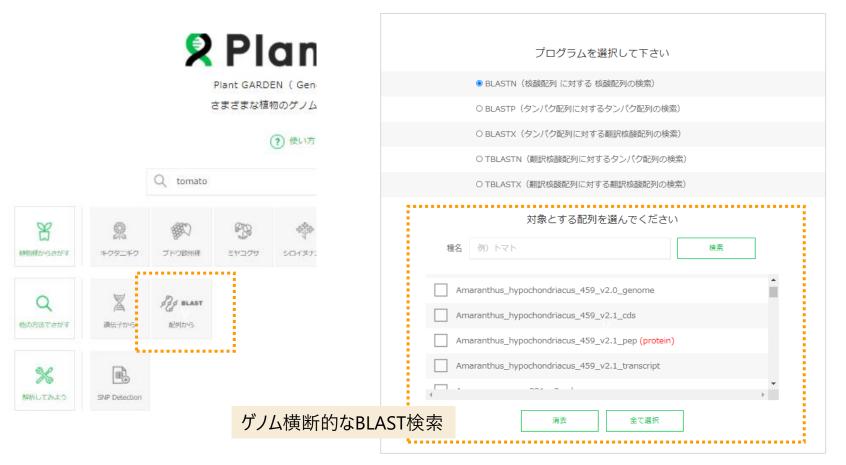
カテゴリーで絞り込みできる







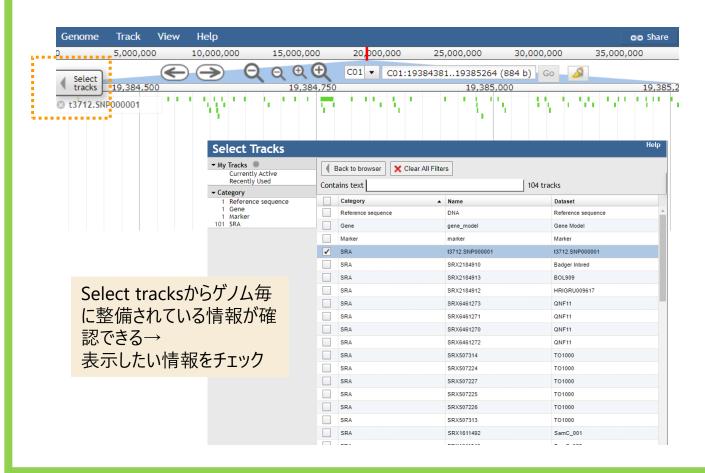
TopページからのBLAST検索

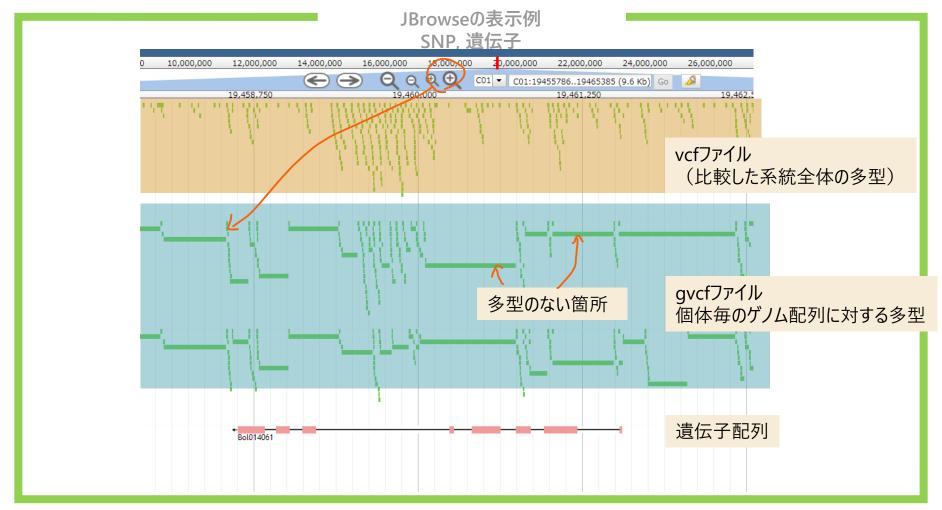






JBrowseで表示するデータを選ぶ



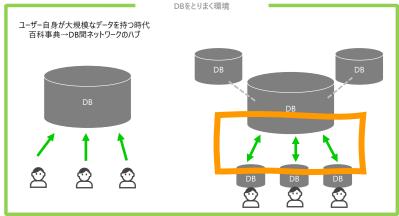


Jbrowse上のSNP表示 16,000,000 18,000,000 20,000,000 22,000,000 26,000,000 10,000,000 12,000,000 14,000,000 24,000,000 C01 C01:19455786..19465385 (9.6 Kb) Go 19,458,750 19,461,250 19,462,5 SNV SNPの詳細情報 Primary Data SNV Type 65 Score 拡大表示 Description SNV T -> G **Position** C01:19460012..19460012 SNV T -> A Length 1 bp no alternative alleles no alternative alleles Attributes no alternative alleles AC AD SNV G -> A 42 14 SNPEffによるアノテーション AN ANN G | downstream_gene_variant | MODIFIER | Bol014060.gene | Bol014060.gene | transcript | Bol014060 | protein_coding||c.*2243T>G|||||2243| G | intron variant | MODIFIER | Bol014061, gene | Bol014061, gene | transcript | Bol014061 | protein cod ing|5/7|c.645+227A>C||||| BQB 0.855738 Bol014061 DP 56 DP4 31 11 0 14 HOB 0.5 ICB MQ MOOF 0.0357143

解析ツール

つないで使う

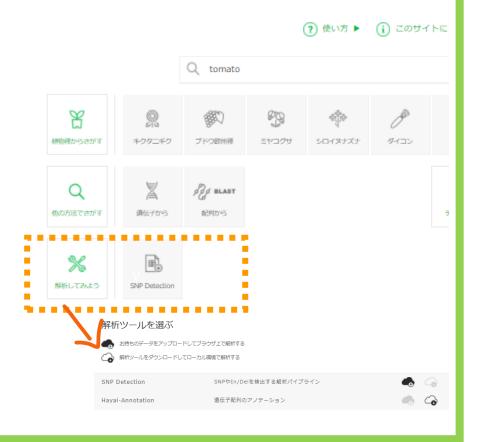
Ž Plant GARDEN





カスタム型データ解析システム

Plant GARDEN内のデータを参照データとし、 ユーザーのデータとできるだけシームレスに比較 できるツールの開発



SNP検出パイプライン





かずさ内の解析サーバ ーに実装



配列解析シ	ステム 1.03	
アカウント	sisobe	
パスワード		
	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)

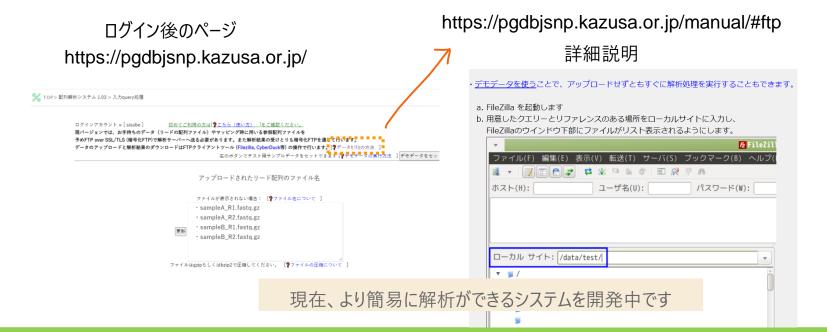
> アカウント登録をして ログインします

ログイン後の画面

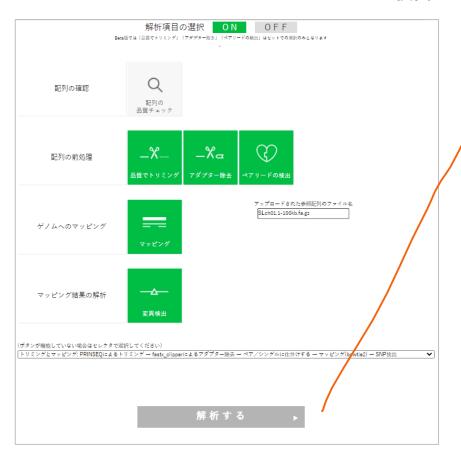
	dogay coxtillation and a feet compile	デモデータの実行方法] デモデータ:
	アップロードされたリード配列のファイル名	
	ファイルが表示されない場合: [?ファイル名について]	
	· sampleA_R1.fastq.gz	
	· sampleA_R2.fastq.gz	
軍新	· sampleB_R1.fastq.gz	
~~	· sampleB_R2.fastq.gz	
更新		

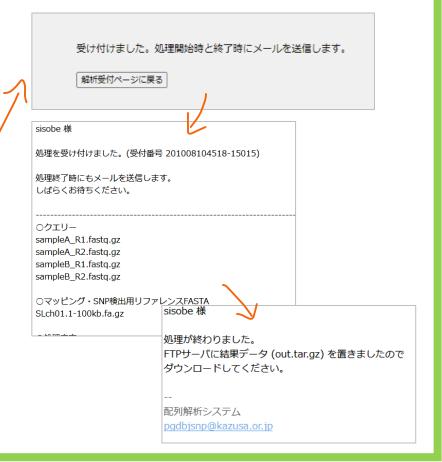
SNP検出パイプライン

データのUploard、Downloadはftpサーバー経由で実施する
(事前にftpクライアントツールのインストールが必要です)
現在はPlant GARDENに格納されたデータとのリンクもされていないので、参照配列も一旦Plant GARDENからダウンロードした後にこちらのサイトにUp loadする必要があります。



SNP検出パイプライン







事前アンケート

植物ゲノムのDB関連で具体的に知りたいこと

- □ DBを活用して新たな価値を生んだ事例があれば
 - →DB開発側にも直接レポートが来るわけでないので、分からないのですが、研究開発者の<調べる>行為のサポート役と考えています
- □ ストレス関連遺伝子
- □ フェノミクスとゲノミクスを含む統合データベース
 - →現在、DNAマーカーについては一部形質との連関情報を掲載しています。遺伝子に関しても同様の情報収集を始めており、今後掲載する予定です
- □ 植物種からアミノ酸配列 C 末側100AA程度の 配列をあつめたい。
 - →残念ながらDB機能としては対応していません。pep配列ファイルをダウンロードしていただき、そこからの切り出しになると思います。
- □ どのような手順(研究スケジュール的な視点)で利用するか
 - →目的によって異なりますので一概にはいえませんが、まずは対象とする植物種を決めて格納されているデータを確認されるのが良いと思います。
- □ マイクロアレイ の結果/eFP browserにおける検索ID
- □ 共発現データベース
- □ 選択的スプライシング 動物と植物のゲノム間での大きな違い。動物におけるゲノム情報の多様性と、植物におけるゲノム情報の多様 性の差異
 - →こちらはPlant GARDEN外の内容のなりますので、今回対応できず申し訳ありません。NBDCで次回の講習会に取り上げて下さればと思います。

ユーザーの皆様へ

https://plantgarden.jp

Ž Plant GARDEN

- □ Plant GARDENは、発展途上のDBです。不具合、もしくは機能不足が生じてご迷惑をおかけすることもあると思いますが、お使いになってみて問題点があれば、ぜひお知らせください。
- 格納するデータについては、できるだけユーザーの意見を取り入れたいと思います。格納をご希望する データがございましたらお知らせください。優先的に格納してまいります。

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