Genome Editing-Guide RNA (sgRNA) Design

Selvaraju Kanagarajan
Associate Professor
Department of Plant Breeding

BREEDTECH Training
June 26, 2025

Where to start……

Method

Crop?

Tissue culture?

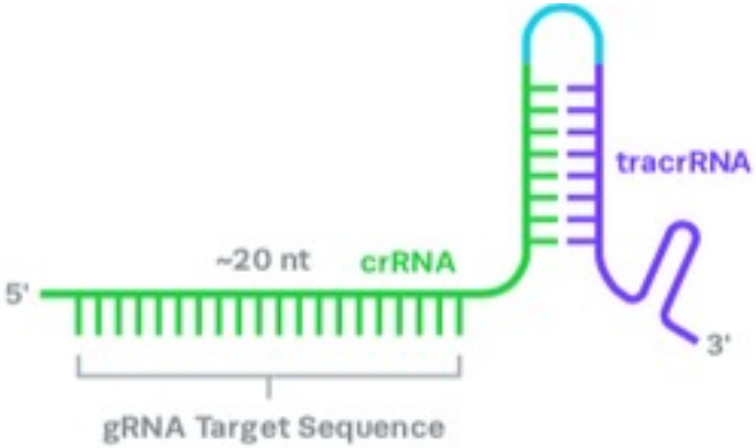
Gene Information and gRNA Design?

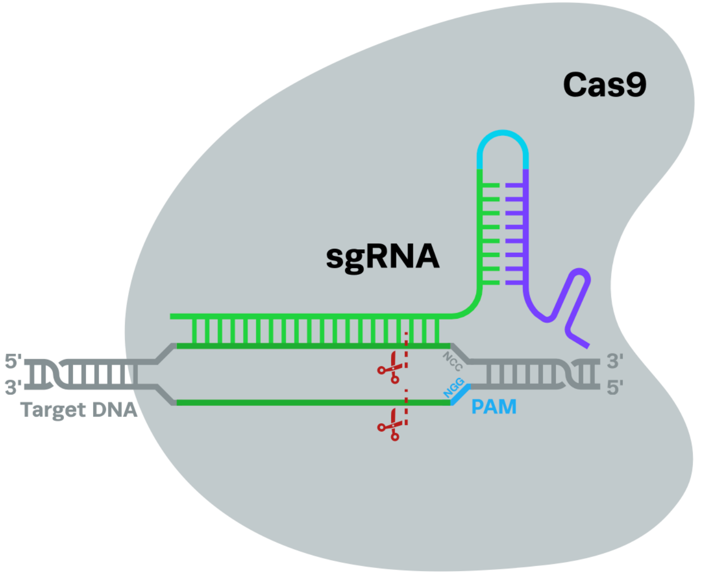
Mutant screening?

Regulation?

New Breeding Techniques (NBTs)-Targeted Gene Editing

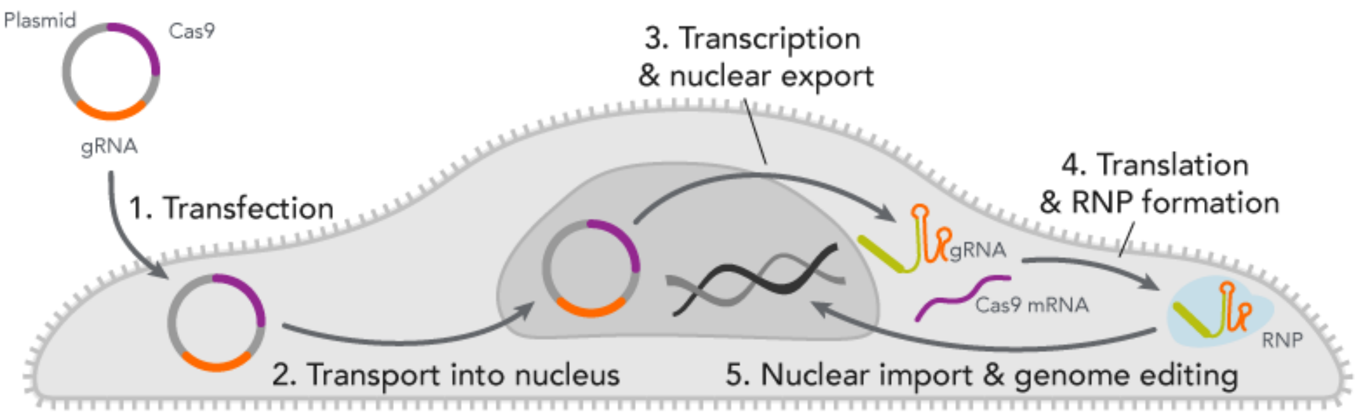
- ❖ Clustered regularly interspaced short palindromic repeats and CRISPR-associated protein-9 (CRISPR-Cas9)
- ❖ Directing a guide RNA (gRNA) complementary to a defined target, together with a CRISPR-associated endonuclease, Cas9, to create double-stranded DNA cleavage
- ❖ gRNA: crispr RNA (crRNA), a 17-20 nucleotide sequence complementary to the target DNA, and a tracrRNA, which serves as a binding scaffold for the Cas nuclease



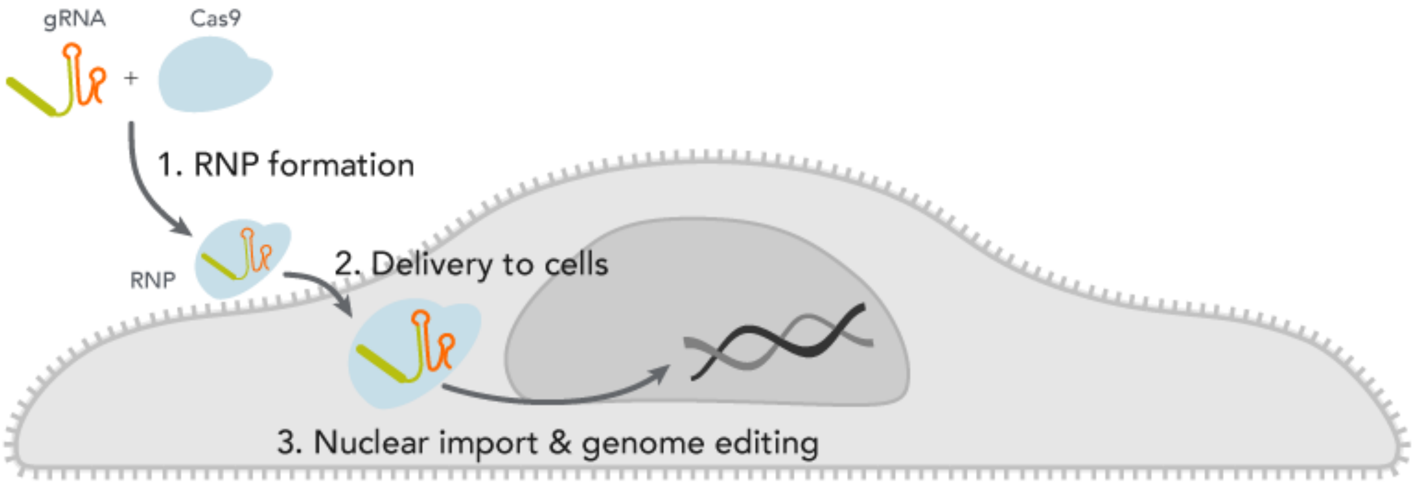


https://www.synthego.com/

Plasmid-based editing (Transgene-free)



RNP-based editing (Transgene & DNA-free)



https://eu.idtdna.com

Antinutrient factors (ANF) in rapeseed

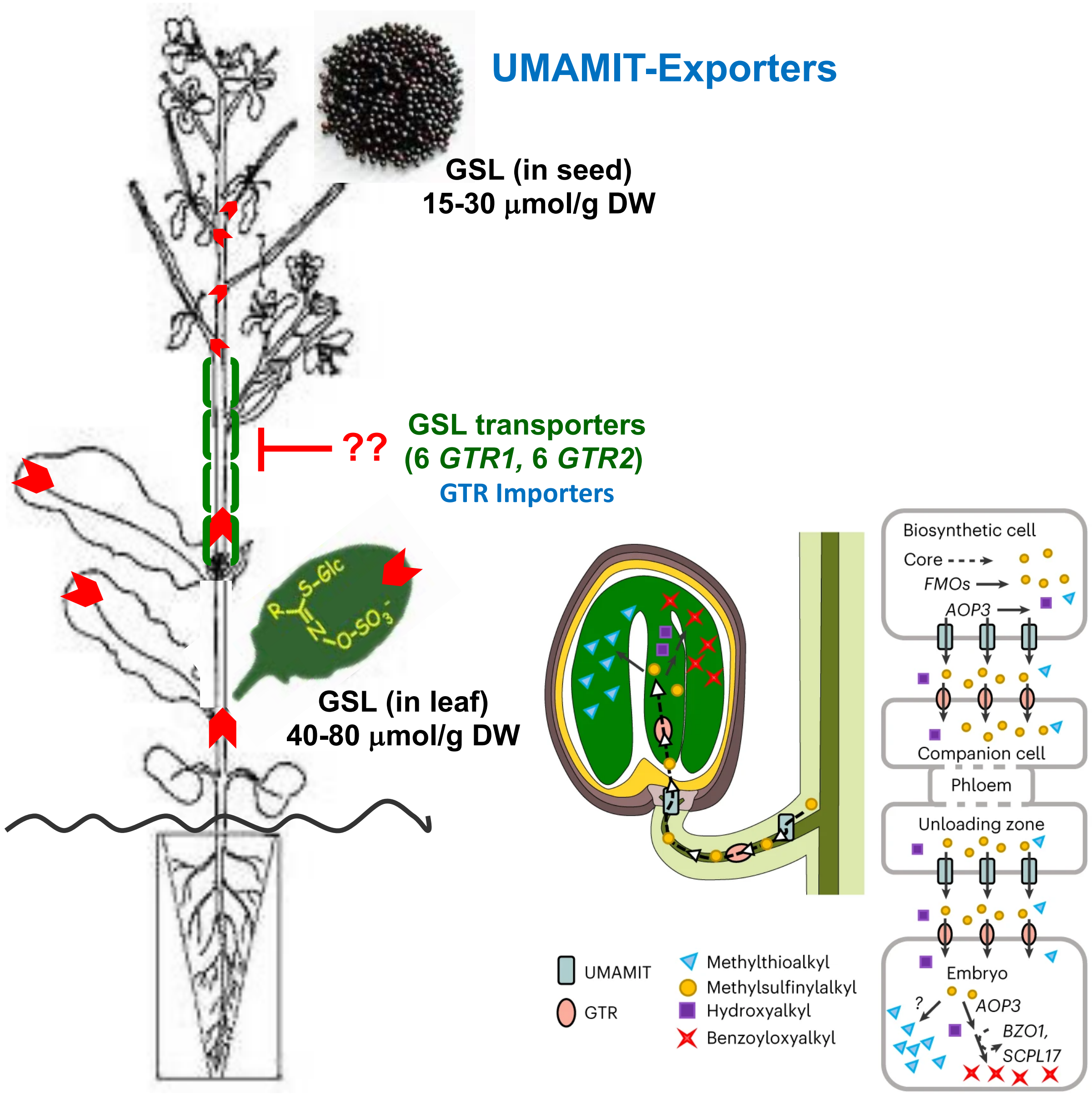

Photo: Anja Persson

Seedcake/meal
Protein: 35-45%
Rich in Tryp & Threonine
Good amounts of Lys, Met & Cys
Fiber: 10-12%
Fat: 3-5%

Anti-Nutritional Factors

Glucosinolates: Sulfur-containing compounds-bitter-interfere with thyroid function

- Synthesized in vegetative tissues and transported to seeds-regulated by glucosinolate transporter (*GTR*) genes (*GTR1* and *GTR2*) and *UMAMIT* genes
- Glucosinolates-sulfur- and/or nitrogenous secondary metabolites
- Glucosinolates-defense compounds against pests



Long-distance glucosinolate transport

Sanden et al. 2024

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 11 November 2024
DOI 10.3389/fpls.2024.1468675

Check for updates

OPEN ACCESS

EDITED BY
Kanwarpal S. Dhugga,
HiBioS Inc., United States

REVIEWED BY
Jasmeet Kaur,
Punjab Agricultural University, India
Zhengyu Wen,
KeyGene, United States

*CORRESPONDENCE
Ravikesavan Rajasekaran
✉ ravikesavan@tnau.ac.in
Selvaraju Kanagarajan
✉ selvaraju.kanagarajan@slu.se

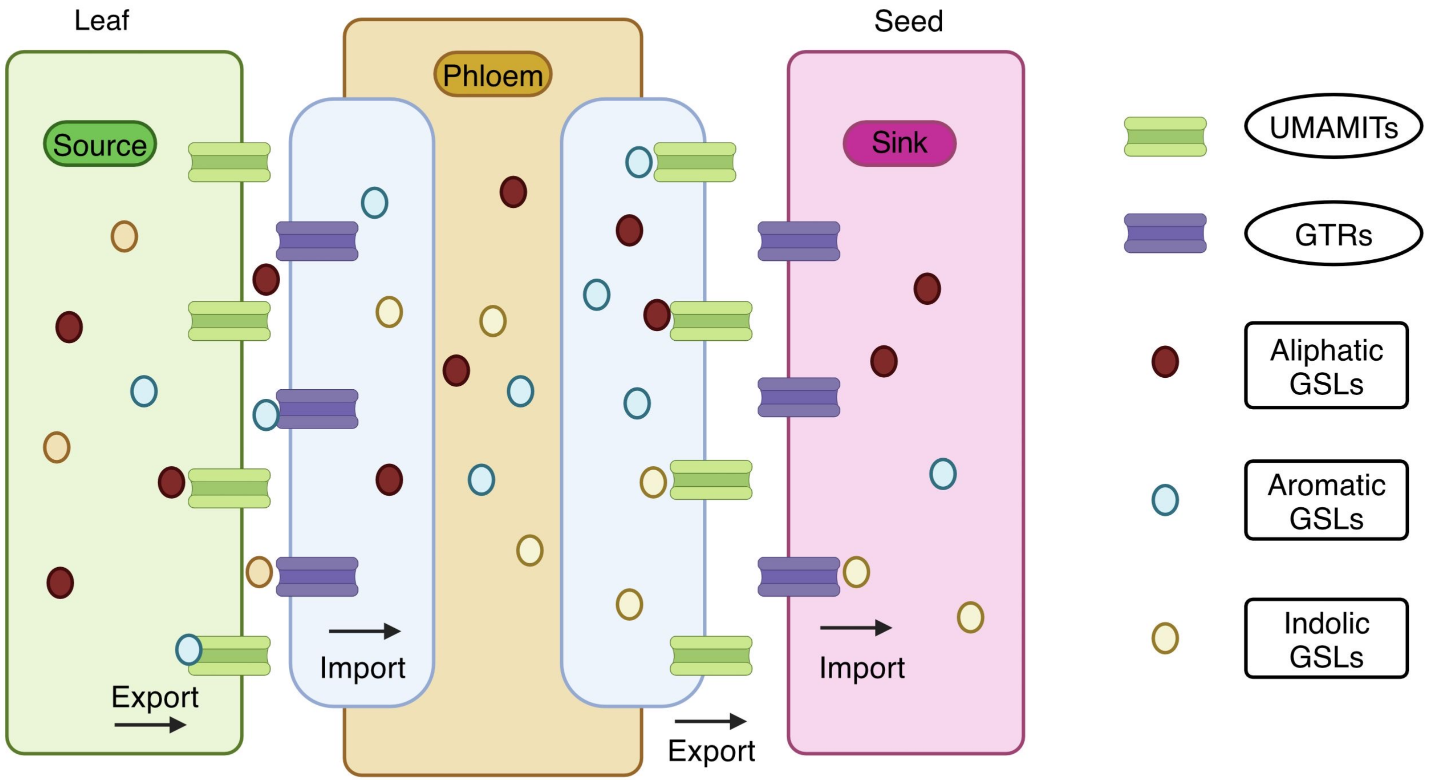
RECEIVED 22 July 2024
ACCEPTED 30 September 2024
PUBLISHED 11 November 2024

Manikandan A, Muthusamy S, Wang ES, Ivarson E, Manickam S, Sivakami R, Narayanan MB, Zhu L-H, Rajasekaran R and Kanagarajan S (2024) Breeding and biotechnology approaches to enhance the nutritional quality of rapeseed byproducts for sustainable alternative protein

Breeding and biotechnology approaches to enhance the nutritional quality of rapeseed byproducts for sustainable alternative protein sources- a critical review

Anandhavalli Manikandan[1,2], Saraladevi Muthusamy[2], Eu Sheng Wang[2], Emelie Ivarson[2], Sudha Manickam[3], Rajeswari Sivakami[4], Manikanda Boopathi Narayanan[3], Li-Hua Zhu[2], Ravikesavan Rajasekaran[4*] and Selvaraju Kanagarajan[2*]

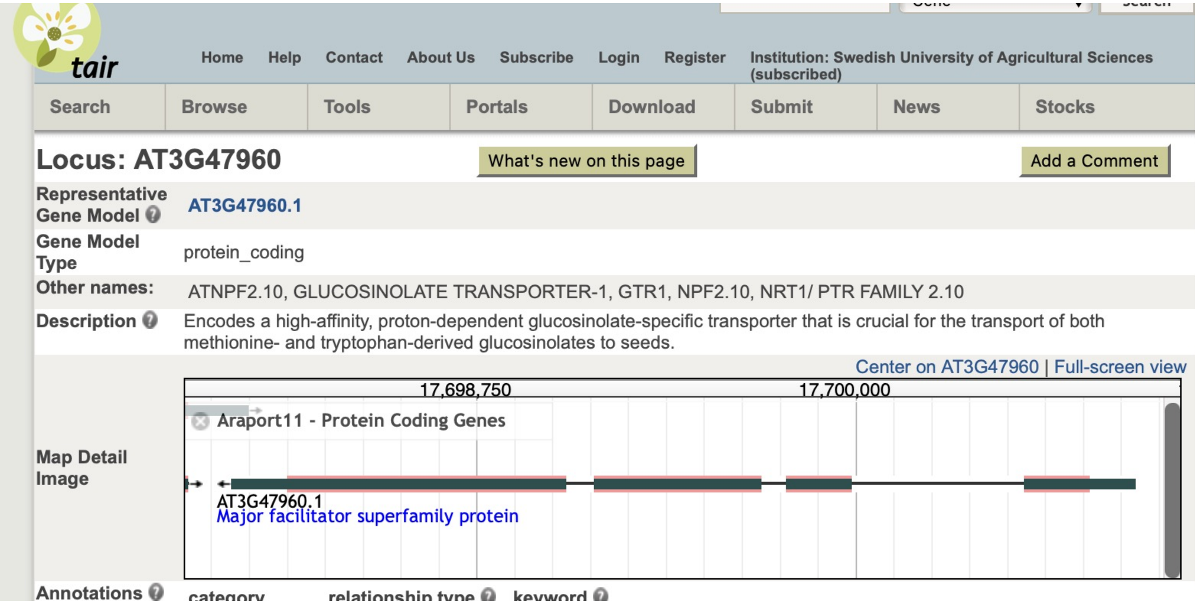
[1]Department of Genetics and Plant Breeding, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India, [2]Department of Plant Breeding, Swedish University of Agricultural Sciences, Lomma, Sweden, [3]Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India, [4]Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India



- Transport of GSL- Source to Sink
- Usually Multiple Amino Acids Move in And Out Transporter (UMAMIT)- Exporter- Source into apoplast
- Glucosinolate transporters (GTRs) – Importer-Apoplast into the phloem
- UMAMIT Exports Phloem into the apoplast
- GTR imports into the Sink.

Gene identification

Functionally characterized genes from Arabidopsis



Locus: AT3G47960

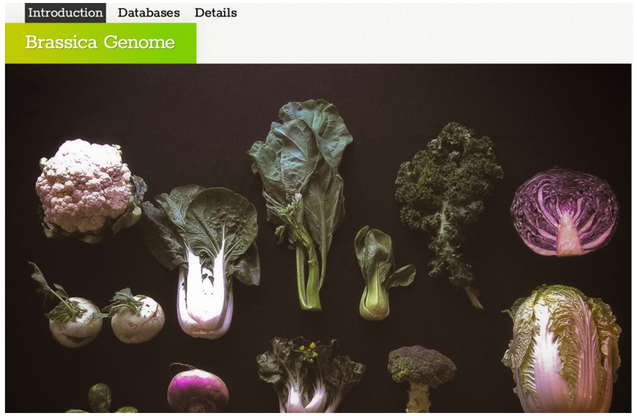
Representative Gene Model	AT3G47960.1
Gene Model Type	protein_coding
Other names:	ATNPF2.10, GLUCOSINOLATE TRANSPORTER-1, GTR1, NPF2.10, NRT1/ PTR FAMILY 2.10
Description	Encodes a high-affinity, proton-dependent glucosinolate-specific transporter that is crucial for the transport of both methionine- and tryptophan-derived glucosinolates to seeds.
Map Detail Image	
Annotations	category relationship type keyword

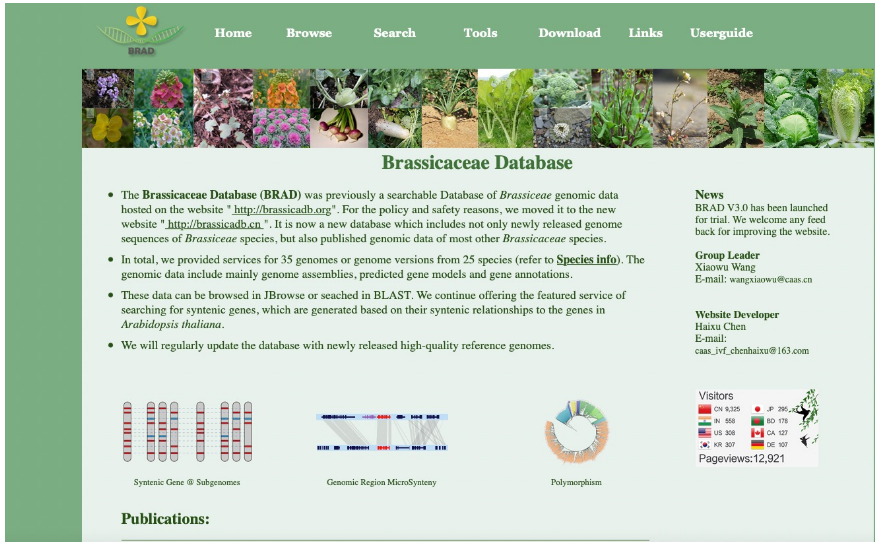
Arabidopsis database

GTR1 gene

```
         agtgtaaaaatgggcctacctaaatccagctaactcagcccacacattaaaaagaaacgttaaagggaaagtaataaatgaaaatgagtcactcagagac | A
17700900 agatgtctctatctttcacagccaaattcacaccattctttattttcatttttcccccttgcccagattccgacaaaatccactcgtttgattccacca | T
         ctatatataattgagcggagaagcttgacATGAAGAGCAGAGTCATTCTTAACCATAGAGACAGAAGAGATAAGAATCATAACAACAACAACACCAATCA | 3
17700700 TTACACACAGGTTGATACTATGGAGAGAAAGCCTCTTGAGGTTGAGCCATCTACTACGACGACTAACACCGATGTGGTCGATTCTTTCGAGGAAGAGCAG | G
         AGAAAAATCGTTTATAGAGGATGGAAAGTCATGCCTTTTATCATTGgtaacgctcctttttacttgtttgtgtctaattcagacatcattttttctttt | 4
17700500 ggggtgctttgtcggatttagggaagtgtaacattttgtccgtgatgtttttaaatttcgtttcttttctcttctctttgacttgactctgtcaaaaat | 7
         ctcgaatttggagttctttctctgagatttcgaagtgaaattggttttaaaactgagatttcgttgaaacaatgaaacccagataatcttacagatctgt | 9
17700300 ctaatggaaattatatttttttacaattgttttcattttattttcaaagagacttttgaaatctttaacacaaaaaacagagactttaaatcatgcagtta | 6
         agtttctgaaaaaagtagataagattttttgtgaatttctggataagacaaaattagtttatgggagttacaaaacagttataaagtttcatgtgaaccaa | 0
17700100 agtcagaacattggaagtccaaactttgatttgtgtgtttccttgtacattaaatgtaaaaaaaaaactgtaactgtattcttaatttcttatcattaat | .
         ttgctattgtagGTAATGAGACATTTGAGAAGCTTGGGATCATAGGGACATTATCAAACCTTCTTGTGTACCTAACTTCTGTATTCAACCTTAAGAGCTA | 1
17699900 CACAGCTGCAACTATCATCAATGCCTTTAGTGGTACAATCAATTTCGGGACTTTTATTGCTGCCTTTCTTTGCGACACTTACTTTGGTCGCTACAAGACT |
         CTCAGTGTCGCTGTCATCGCTTGTTTTCTGgtactacttttatatataatctttcatgactccggtttcgttttcgccaatgttttctaaaccattttttc |
17699700 atctctgtagGGATCGTTTGTGATACTACTTACTGCTGCAATTCCGTCGTTGCACCCCGTTGCTTGCGGAAACAAAATCTCGTGCGAAGGGCCAAGTGTG |
         GGACAGATACTGTTTCTGCTCATGGGTTTAGGCTTTCTTGTGGTAGGCGCTGGTGGTATCAGACCGTGTAATTTAGCGTTTGGTGCTGATCAGTTCAACC |
17699500 CCAAATCCGAATCAGGAAAGAAAGGGATCAACAGTTTCTTCAACTGGTATTTCTTCACTTTCACGTTCGCGCAGATCATCTCGCTCACAGCAGTTGTGTA |
         TATCCAGTCAAATGTGAGTTGGACAATCGGTTTGATCATTCCGGTGGCTCTTATGTTCTTGGCGTGTGTGATTTTCTTTGCTGGCGATAGATTGTATGTA |
17699300 AAAGTGAAAGCCTCGGGTAGTCCATTGGCTGGTATTGCTCGTGTTATAGCAGCCGCGATCAAGAAACGAGGTTTGAAGCCGGTTAAGCAGCCTTGGGTCA |
         ATCTTTACAACCATATCCCGTCTAACTATGCAAATACTACTCTAAAATACACCGACCAGTTCAGgtaaattgatatctacttattcttggttctcttgct |
17699100 ttatgtttctggcttttgttttgctgaaaaatgtttgtttattatgttgccagATTTCTAGACAAAGCAGCAATTATGACCCCCGAGGAGAAGTTGAACT |
         CCGATGGAACGGCTTCCGATCCATGGAAACTATGCACGTTGCAGCAAGTGGAAGAAGTGAAATGCATTGTAAGAGTGATTCCAATCTGGTTTGCATCCAC |
17698900 GATATACTACCTTGCAATAACTATACAAATGACTTATCCGGTCTTCCAAGCGCTCCAGAGCGACCGACGATTAGGTTCTGGTGGCTTCAGGATTCCTGCG |
         GCCACCTATGTAGTGTTCTTGATGACTGGCATGACAGTTTTCATCATATTCTACGACCGTGTACTTGTCCCGTCGCTCAGAAGAGTGACCGGGTTAGAAA |
17698700 CCGGTATATCACTCTTACAAAGAATCGGAGCTGGGTTTACATTTGCGATAATGAGTTTATTAGTTTCCGGGTTTATAGAGGAAAGGAGAAGAAACTTTGC |
         ATTGACCAAACCGACACTCGGGATGGCGCCTCGGACAGGAGAAATCTCATCAATGTCAGCGTTGTGGCTGATTCCGCAGCTTACACTTGCAGGCATAGCA |
17698500 GAAGCTTTTGCTGCCATTGGACAAATGGAGTTTTACTACAAGCAGTTTCCTGAAAACATGAAGAGCTTCGCGGGTTCTATCTTTTATGTCGGTGCAGGGG |
         TTTCGAGCTACCTCGCTAGCTTCTTGATATCAACTGTTCATCGAACAACCGCGCATTCACCTTCAGGGAATTGGTTAGCTGAGGATCTGAACAAAGCGAA |
17698300 ATTGGATTACTTCTACTTCATGCTCACAGGACTCATGGTTGTTAACATGGCTTACTTCTTGTTGATGGCTAGATGGTATAGATACAAAGGAGGTAATGAT |
         GAAGACATCACTGAGATTGAGACCAATGAAGAAGAAACCAAGCAACAGCAGCTACAAGACAAGAACTCTGTCTGAtctggaaacttatatatttttggta |
17698100 ttctacttcaatctccgtagtttttcttcacccgaaatcaaccttgttgtacttttatagattaattgaagaagaaaatcacttcatctttgtagacatg |
         tattgtaaaataagtttttttgttgctaatttcaacctccgtacgaagtttctgtttcttctttcatcaccacaacagaacagaacataacaaaccgctaa v
```

Brassica napus databases





Brassicaceae Database

- The **Brassicaceae Database (BRAD)** was previously a searchable Database of *Brassiceae* genomic data hosted on the website " http://brassicadb.org". For the policy and safety reasons, we moved it to the new website " http://brassicadb.cn ". It is now a new database which includes not only newly released genome sequences of *Brassiceae* species, but also published genomic data of most other *Brassicaceae* species.
- In total, we provided services for 35 genomes or genome versions from 25 species (refer to **Species info**). The genomic data include mainly genome assemblies, predicted gene models and gene annotations.
- These data can be browsed in JBrowse or seached in BLAST. We continue offering the featured service of searching for syntenic genes, which are generated based on their syntenic relationships to the genes in *Arabidopsis thaliana*.
- We will regularly update the database with newly released high-quality reference genomes.

News
BRAD V3.0 has been launched for trial. We welcome any feed back for improving the website.

Group Leader
Xiaowu Wang
E-mail: wangxiaowu@caas.cn

Website Developer
Haixu Chen
E-mail:
caas_ivf_chenhaixu@163.com

Visitors

CN	9,325	JP	295
IN	558	BD	178
US	308	CA	127
KR	307	DE	107

Pageviews:12,921

Publications:

Brassica napus (AST_PRJEB5043_v1) ▾

Brassica napus Assembly and Gene Annotation

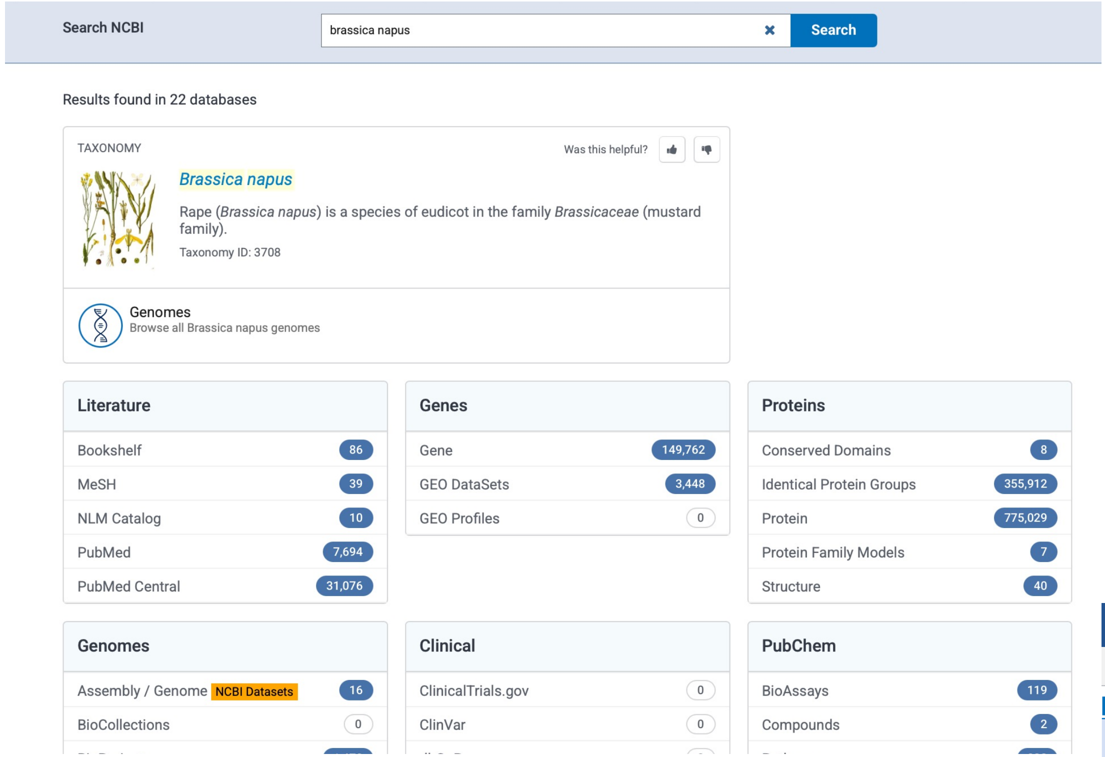
Brassica napus Genome Browser

Browse Data

Search using a sequence name, gene name, locus, or other landmark.
The wildcard character * is allowed.

Search:	**Display:**
chrA01:6,320,000..6,340,999	region of chromosome chrA01 from 6,320,000 to 6,340,999
chrC01:11,338,000..11,372,999	region of chromosome chrC01 from 11,338,000 to 11,372,999
BnaA01g00030D	Gene prediction BnaA01g00030D

NCBI databases for *B. napus*



Search NCBI: brassica napus — Search

Results found in 22 databases

TAXONOMY — Was this helpful?

Brassica napus

Rape (*Brassica napus*) is a species of eudicot in the family *Brassicaceae* (mustard family).

Taxonomy ID: 3708

Genomes
Browse all Brassica napus genomes

Literature	
Bookshelf	86
MeSH	39
NLM Catalog	10
PubMed	7,694
PubMed Central	31,076

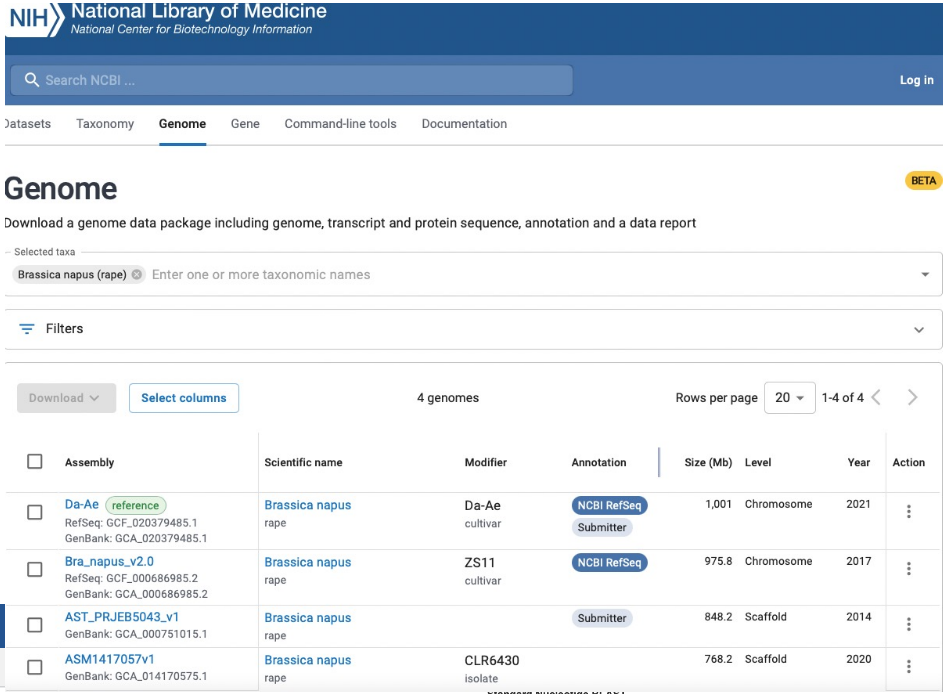
Genes	
Gene	149,762
GEO DataSets	3,448
GEO Profiles	0

Proteins	
Conserved Domains	8
Identical Protein Groups	355,912
Protein	775,029
Protein Family Models	7
Structure	40

Genomes	
Assembly / Genome NCBI Datasets	16
BioCollections	0

Clinical	
ClinicalTrials.gov	0
ClinVar	0

PubChem	
BioAssays	119
Compounds	2



NIH National Library of Medicine
National Center for Biotechnology Information

Genome

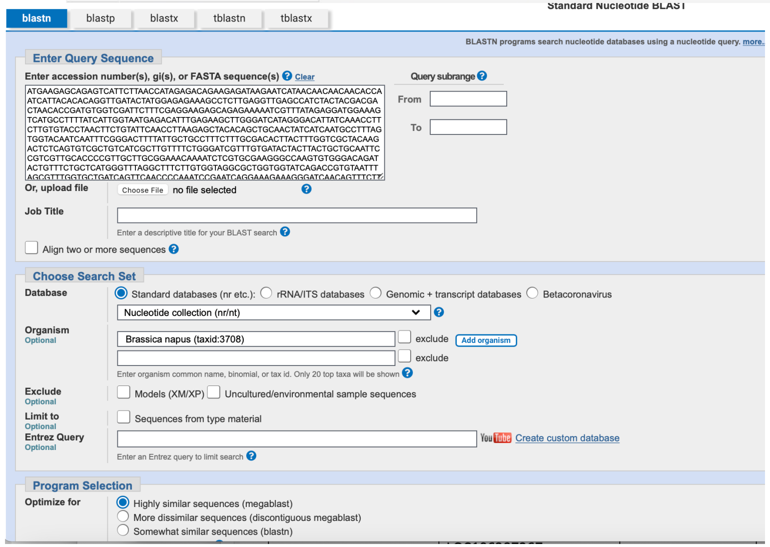
BETA

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa: Brassica napus (rape)

4 genomes — Rows per page 20 — 1-4 of 4

Assembly	Scientific name	Modifier	Annotation	Size (Mb)	Level	Year
Da-Ae reference RefSeq: GCF_020379485.1 GenBank: GCA_020379485.1	Brassica napus rape	Da-Ae cultivar	NCBI RefSeq Submitter	1,001	Chromosome	2021
Bra_napus_v2.0 RefSeq: GCF_000686985.2 GenBank: GCA_000686985.2	Brassica napus rape	ZS11 cultivar	NCBI RefSeq	975.8	Chromosome	2017
AST_PRJEB5043_v1 GenBank: GCA_000751015.1	Brassica napus rape		Submitter	848.2	Scaffold	2014
ASM1417057v1 GenBank: GCA_014170575.1	Brassica napus rape	CLR6430 isolate		768.2	Scaffold	2020



Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. more..

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

```
ATGAAGAGCAGAGTCATTCTTAACCATAGAGACAGAAGAGATAAGAATCATAACAACAACAACACCA
ATCATTACACACAGGTTGATACTATGGAGAGAAAGCCTCTTGAGGTTGAGCCATCTACTACGACGA
CTAACACCGATGTGGTCGATTCTTTCGAGGAAGAGCAGAGAAAAATCGTTTATAGAGGATGGAAAG
TCATGCCTTTTATCATTGGTAATGAGACATTTGAGAAGCTTGGGATCATAGGGACATTATCAAACCTT
CTTGTGTACCTAACTTCTGTATTCAACCTTAAGAGCTACACAGCTGCAACTATCATCAATGCCTTTAG
TGGTACAATCAATTTCGGGACTTTTATTGCTGCCTTTCTTTGCGACACTTACTTTGGTCGCTACAAG
ACTCTCAGTGTCGCTGTCATCGCTTGTTTTCTGGGATCGTTTGTGATACTACTTACTGCTGCAATTC
CGTCGTTGCACCCCGTTGCTTGCGGAAACAAAATCTCGTGCGAAGGGCCAAGTGTGGGACAGAT
ACTGTTTCTGCTCATGGGTTTAGGCTTTCTTGTGGTAGGCGCTGGTGGTATCAGACCGTGTAATTT
AGCGTTTGGTGCTGATCAGTTCAACCCCAAATCCGAATCAGGAAAGAAAGGGATCAACAGTTTCTT
```

Query subrange — From: — To:

Or, upload file: Choose File no file selected

Job Title: Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database: ● Standard databases (nr etc.): ○ rRNA/ITS databases ○ Genomic + transcript databases ○ Betacoronavirus

Nucleotide collection (nr/nt)

Organism Optional: Brassica napus (taxid:3708) ☐ exclude Add organism
☐ exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude Optional: ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to Optional: ☐ Sequences from type material

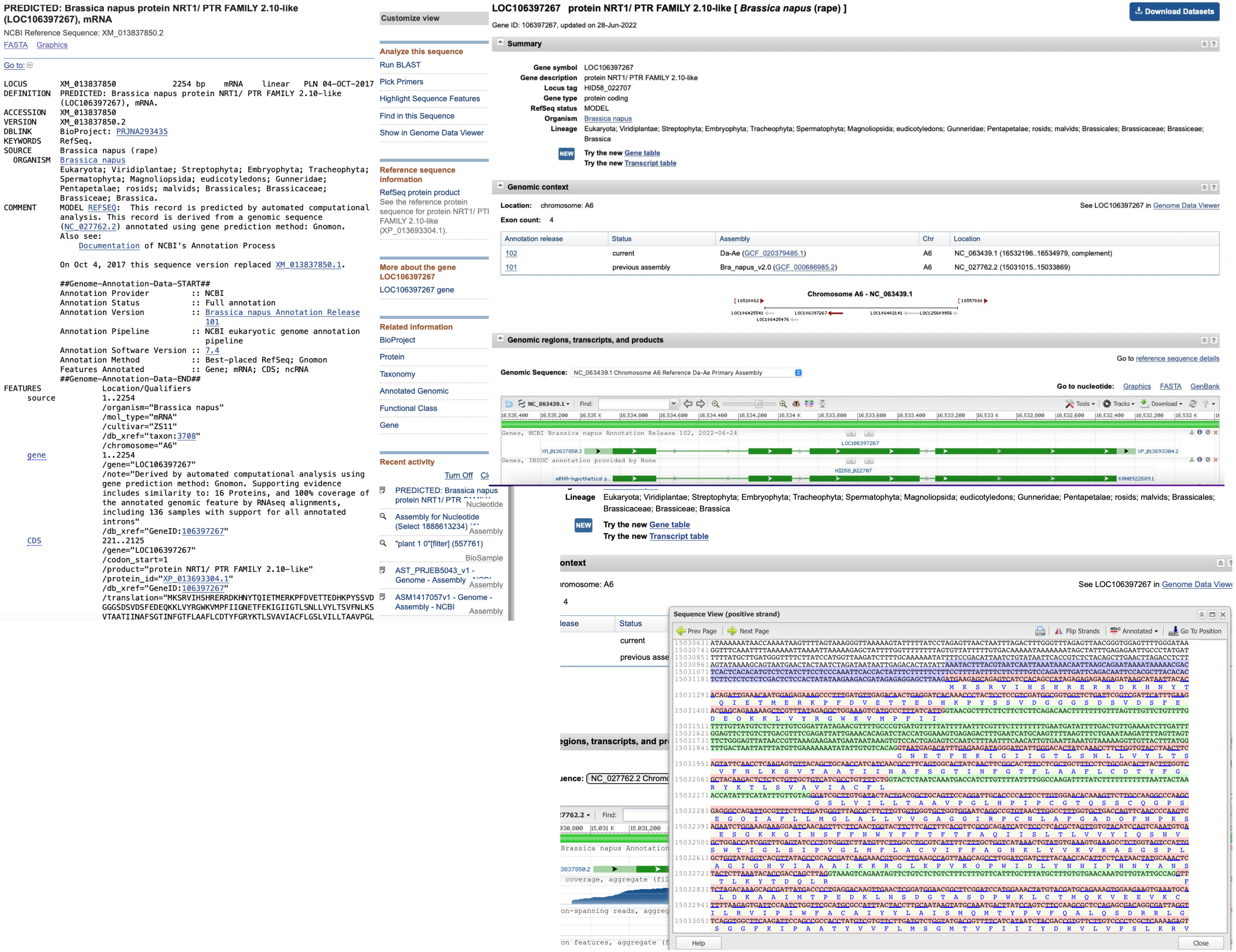
Entrez Query Optional: YouTube Create custom database

Enter an Entrez query to limit search

Program Selection

Optimize for:
● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)

NCBI genome database for *B. napus*



PREDICTED: Brassica napus protein NRT1/ PTR FAMILY 2.10-like (LOC106397267), mRNA

NCBI Reference Sequence: XM_013837850.2

```
LOCUS       XM_013837850            2254 bp    mRNA    linear   PLN 04-OCT-2017
DEFINITION  PREDICTED: Brassica napus protein NRT1/ PTR FAMILY 2.10-like
            (LOC106397267), mRNA.
ACCESSION   XM_013837850
VERSION     XM_013837850.2
DBLINK      BioProject: PRJNA293435
KEYWORDS    RefSeq.
SOURCE      Brassica napus (rape)
  ORGANISM  Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae;
            Pentapetalae; rosids; malvids; Brassicales; Brassicaceae;
            Brassiceae; Brassica.
COMMENT     MODEL REFSEQ:  This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NC_027762.2) annotated using gene prediction method: Gnomon.
            Also see:
                Documentation of NCBI's Annotation Process

            On Oct 4, 2017 this sequence version replaced XM_013837850.1.

            ##Genome-Annotation-Data-START##
            Annotation Provider         :: NCBI
            Annotation Status           :: Full annotation
            Annotation Version          :: Brassica napus Annotation Release
                                           101
            Annotation Pipeline         :: NCBI eukaryotic genome annotation
                                           pipeline
            Annotation Software Version :: 7.4
            Annotation Method           :: Best-placed RefSeq; Gnomon
            Features Annotated          :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..2254
                     /organism="Brassica napus"
                     /mol_type="mRNA"
                     /cultivar="ZS11"
                     /db_xref="taxon:3708"
                     /chromosome="A6"
     gene            1..2254
                     /gene="LOC106397267"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Gnomon. Supporting evidence
                     includes similarity to: 16 Proteins, and 100% coverage of
                     the annotated genomic feature by RNAseq alignments,
                     including 136 samples with support for all annotated
                     introns"
                     /db_xref="GeneID:106397267"
     CDS             221..2125
                     /gene="LOC106397267"
                     /codon_start=1
                     /product="protein NRT1/ PTR FAMILY 2.10-like"
                     /protein_id="XP_013693304.1"
                     /db_xref="GeneID:106397267"
                     /translation="MKSRVIHSHRERRDKHNYTQIETMERKPFDVETTEDHKPYSSVD
                     GGGSDSVDSFEDEQKKLVYRGWKVMPFIIGNETFEKIGIIGTLSNLLVYLTSVFNLKS
                     VTAATIINAFSGTINFGTFLAAFLCDTYFGRYKTLSVAVIACFLGSLVILLTAAVPGL
```

See the reference protein sequence for protein NRT1/ PTI FAMILY 2.10-like (XP_013693304.1).

LOC106397267 protein NRT1/ PTR FAMILY 2.10-like [*Brassica napus* (rape)]

Gene ID: 106397267, updated on 28-Jun-2022

Summary

Gene symbol LOC106397267
Gene description protein NRT1/ PTR FAMILY 2.10-like
Locus tag HID58_022707
Gene type protein coding
RefSeq status MODEL
Organism Brassica napus
Lineage Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Brassiceae; Brassica

NEW Try the new Gene table
Try the new Transcript table

Genomic context

Location: chromosome: A6

See LOC106397267 in Genome Data Viewer

Exon count: 4

Annotation release	Status	Assembly	Chr	Location
102	current	Da-Ae (GCF_020379485.1)	A6	NC_063439.1 (16532196..16534979, complement)
101	previous assembly	Bra_napus_v2.0 (GCF_000686985.2)	A6	NC_027762.2 (15031015..15033869)

Chromosome A6 - NC_063439.1

Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence: NC_063439.1 Chromosome A6 Reference Da-Ae Primary Assembly

Go to nucleotide: Graphics FASTA GenBank

Lineage Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Brassiceae; Brassica

NEW Try the new Gene table
Try the new Transcript table

ontext

romosome: A6

See LOC106397267 in Genome Data View

4

Sequence View (positive strand)

```
15030631 ATAAAAAATAACCAAAATAAGTTTTAGTAAAGGGTTAAAAAGTATTTTTATCCTAGAGTTAACTAATTTAGACTTTGGGTTTAGAGTTAACGGGTGGAGTTTTGGGATAA
15030741 GGTTTCAAATTTTAAAAAATTAAAATTAAAAAGAGCTATTTTGGTTTTTTTTAGTGTTATTTTTGTGACAAAAATAAAAAAATAGCTATTTGAGAGAATTGCCCTATGAT
15030851 TTTTATGCTTGATGGGTTTTCTTATCCATGGTTAAGATCTTTTGCAAAAAATATTTTCCGACATTAATCTGTATAATTCACCGTCTCTACAGCTTGAACTTAGACCTCTT
15030961 AGTATAAAAGCAGTAATGAACTACTAATCTAGATAATAATTGAGACACTATATTAAATACTTTACGTAATCAATTAAATAAACAATTAAGCAGAATAAAATAAAAACGAC
15031071 TCACTCACACATGTCTCTATCTTCCTCCCAAATTCACCACTATTTCTTTTTCTTTCCTTTTATTTTCTTCTTTGTCCAGATTTGATTCAGACAATTCCACGCTTACACAC
15031181 TCTTCTCTCTCTCGACTCTCCACTATATAAGAAGACGATAGAGAGGAGCTTAAGATGAAGAGCAGAGTCATCCACAGCCATAGAGAGAGAAGAGATAAGCATAATTACAC
                                                                M  K  S  R  V  I  H  S  H  R  E  R  R  D  K  H  N  Y  T
15031291 ACAGATTGAAACAATGGAGAGAAAGCCCTTTGATGTTGAGACAACTGAGGATCACAAACCCTACTCCTCCGTCGATGGCGGTGGTTCTGATTCGGTCGATTCATTTGAAG
          Q  I  E  T  M  E  R  K  P  F  D  V  E  T  T  E  D  H  K  P  Y  S  S  V  D  G  G  G  S  D  S  V  D  S  F  E
15031401 ACGAGCAGAAAAAGCTCGTTTATAGAGGCTGGAAAGTCATGCCCTTTATCATTGGTAACGCTTTCTTCTTCTCTTCAGACAACTTTTTTGTTTAGTTTGTTCTGTTTTG
         D  E  Q  K  K  L  V  Y  R  G  W  K  V  M  P  F  I  I
15031511 TTTGTTATGTCTCTTTTGTCGGATTATAGAACGTTTTGCCCGTGATGTTTTTATTTTAATTTCGTTTCTTTTTTTTGAATGATATTTTGACTGTTGAAAATCTTGATTT
15031621 GGAGTTCTTGTCTTGACGTTTCGAGATTATTGAAACACAGATCTACCATGGAAAGTGAGAGACTTTGAATCATGCAAGTTTTAAGTTTCTGAAATAAGATTTTAGTTAGT
15031731 TTCTGGGAGTTATAACCGTTAAAGAAGAATGAATAATAAAGTGTCCACTGAGAGTCCAATCTTTAATTTCAACATTGTGAATTAAATGTAAAAAGGTTGTTACTTTATGG
15031841 TTTGACTAATTATTTATGTTGAAAAAAATATATTGTGTCACAGGTAATGAGACATTTGAGAAGATAGGGATCATTGGGACACTATCAAACCTTCTGGTGTACCTAACTTC
                                                    G  N  E  T  F  E  K  I  G  I  I  G  T  L  S  N  L  L  V  Y  L  T  S
15031951 AGTATTCAACCTCAAGAGTGTTACAGCTGCAACCATCATCAACGCCTTCAGTGGCACTATCAACTTCGGCACTTTCCTCGCTGCTTTCCTCTGCGACACTTACTTTGGTC
           V  F  N  L  K  S  V  T  A  A  T  I  I  N  A  F  S  G  T  I  N  F  G  T  F  L  A  A  F  L  C  D  T  Y  F  G
15032061 GCTACAAGACTCTCTCTGTTGCTGTCATCGCCTGTTTCTGGTACTCTAATCAAATGACCATCTTGTTTTATTTGGCCAAGATTTTATCTTTTTTTTTTTAATTACTAA
         R  Y  K  T  L  S  V  A  V  I  A  C  F  L
15032171 ACCATATTTCATATTTGTTGTAGGGATCGCTTGTGATACTACTGACGGCTGCAGTTCCAGGATTGCACCCCATTCCTTGTGGAACACAAAGTTCTTGCCAAGGCCCAAGC
                                 G  S  L  V  I  L  L  T  A  A  V  P  G  L  H  P  I  P  C  G  T  Q  S  S  C  Q  G  P  S
15032281 GAGGGCCAGATTGCGTTTCTTCTGATGGGTTTAGCGCTTCTTGTGGTGGGTGCTGGTGGAATCAGGCCGTGTAACTTGGCCTTTGGTGCTGACCAGTTCAACCCCAAGTC
          E  G  Q  I  A  F  L  L  M  G  L  A  L  L  V  V  G  A  G  G  I  R  P  C  N  L  A  F  G  A  D  Q  F  N  P  K  S
15032391 AGAATCTGGAAAGAAAGGAATCAACAGTTTCTTCAACTGGTACTTCTTCACTTTCACGTTCGCGCAGATCATCTCCCTCACGCTAGTTGTGTACATCCAGTCAAATGTGA
          E  S  G  K  K  G  I  N  S  F  F  N  W  Y  F  F  T  F  T  F  A  Q  I  I  S  L  T  L  V  V  Y  I  Q  S  N  V
15032501 GCTGGACCATCGGTTTGAGTATCCCTGTGGGTCTTATGTTCTTGGCCTGCGTCATTTTCTTTGCTGGTCATAAACTGTATGTGAAAGTGAAAGCCTCTGGTAGTCCATTG
         S  W  T  I  G  L  S  I  P  V  G  L  M  F  L  A  C  V  I  F  F  A  G  H  K  L  Y  V  K  V  K  A  S  G  S  P  L
15032611 GCTGGTATAGGTCACGTTATAGCCGCAGCGATCAAGAAACGTGGCTTGAAGCCAGTTAAGCAGCCTTGGATCGATCTTTACAACCACATTCCTCATAACTATGCAAACTC
          A  G  I  G  H  V  I  A  A  A  I  K  K  R  G  L  K  P  V  K  Q  P  W  I  D  L  Y  N  H  I  P  H  N  Y  A  N  S
15032721 TACTCTTAAATACACCGACCAGCTTAGGTAAAGTCAGAATAGTTCTGTCTCTGTCTTTCTTTGTTCATTTGCTTTATGCTTTGTGTGAACAAATGTTGTATTGCCAGGTT
          T  L  K  Y  T  D  Q  L  R                                                                                  F
15032831 TCTAGACAAAGCAGCGATTATGACCCCTGAGGACAAGTTGAACTCGGATGGAACGGCTTCGGATCCATGGAAACTATGTACGATGCAGAAAGTGGAAGAAGTGAAATGCA
          L  D  K  A  A  I  M  T  P  E  D  K  L  N  S  D  G  T  A  S  D  P  W  K  L  C  T  M  Q  K  V  E  E  V  K  C
15032941 TTTTAAGAGTGATTCCAATCTGGTTCGCATGCGCCATTTACTACCTTGCAATAAGTATGCAAATGACTTATCCAGTCTTCCAAGCGCTCCAGAGCGACAGGCGATTAGGT
         I  L  R  V  I  P  I  W  F  A  C  A  I  Y  Y  L  A  I  S  M  Q  M  T  Y  P  V  F  Q  A  L  Q  S  D  R  R  L  G
15033051 TCAGGTGGCTTCAAGATTCCAGCCGCCACCTATGTCGTGTTCTTGATGTCTGGTATGACGGTTTTCATCATAATCTACGACCGTGTTCTTGTCCCCTCGCTCAAAAGAGT
          S  G  G  F  K  I  P  A  A  T  Y  V  V  F  L  M  S  G  M  T  V  F  I  I  I  Y  D  R  V  L  V  P  S  L  K  R  V
```

General workflow for cloning of a target genes

Primer design based on untranslated regions/genes based on reference genome

↓

DNA extraction and PCR

↓

Agarose gel electrophoresis

↓

Gel extraction of PCR product

↓

Cloning into vector and *E. coli* transformation

↓

Plasmid purification

↓

Sanger Sequencing with vector specific and gene specific primers

General workflow for cloning of a target gene from cDNA

RNA extraction

↓

cDNA synthesis

↓

PCR using gene specific primers

↓

Agarose gel electrophoresis

↓

Gel extraction of PCR product

↓

Cloning into vector and transformation into *E.coli*

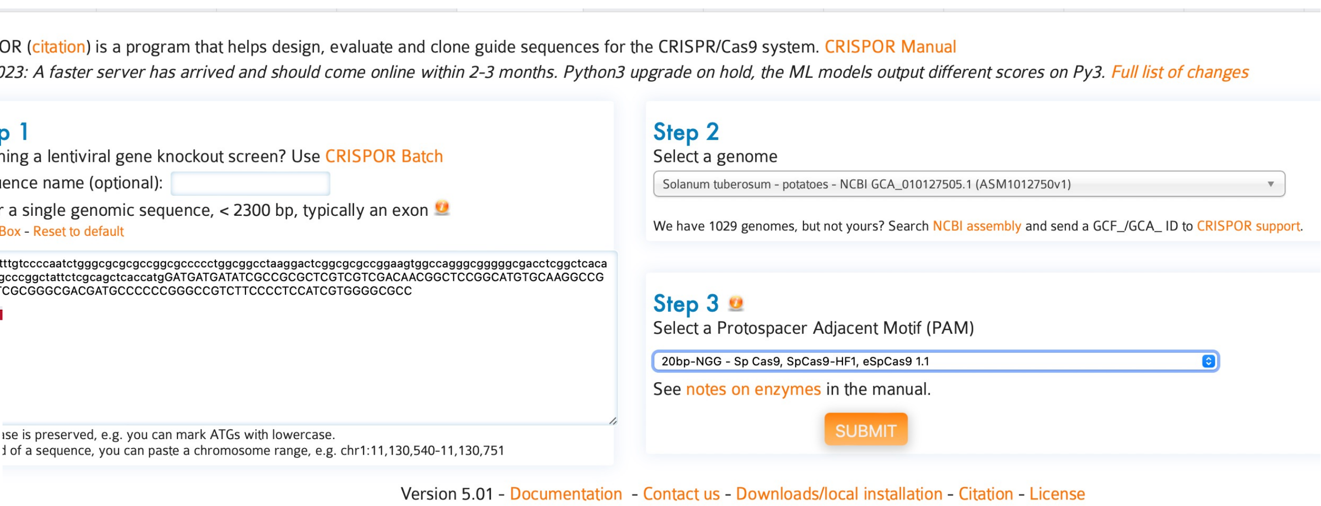
↓

Plasmid purification

↓

Sanger sequencing with vector specific
and gene specific primers

gRNA Design


CRISPOR (citation) is a program that helps design, evaluate and clone guide sequences for the CRISPR/Cas9 system. CRISPOR Manual

July 2023: A faster server has arrived and should come online within 2-3 months. Python3 upgrade on hold, the ML models output different scores on Py3. Full list of changes

Step 1

Planning a lentiviral gene knockout screen? Use CRISPOR Batch

Sequence name (optional):

Enter a single genomic sequence, < 2300 bp, typically an exon

Clear Box - Reset to default

```
cttcctttgtccccaatctgggcgcgcgccggcgccccctggcggcctaaggactcggcgcgccggaagtggccagggcgggggcgacctcggctcaca
gcgcgcccggctattctcgcagctcaccatgGATGATGATATCGCCGCGCTCGTCGTCGACAACGGCTCCGGCATGTGCAAGGCCG
GCTTCGCGGGCGACGATGCCCCCCGGGCCGTCTTCCCCTCCATCGTGGGGCGCC
```

se is preserved, e.g. you can mark ATGs with lowercase.

of a sequence, you can paste a chromosome range, e.g. chr1:11,130,540-11,130,751

Step 2

Select a genome

Solanum tuberosum - potatoes - NCBI GCA_010127505.1 (ASM1012750v1)

We have 1029 genomes, but not yours? Search NCBI assembly and send a GCF_/GCA_ ID to CRISPOR support.

Step 3

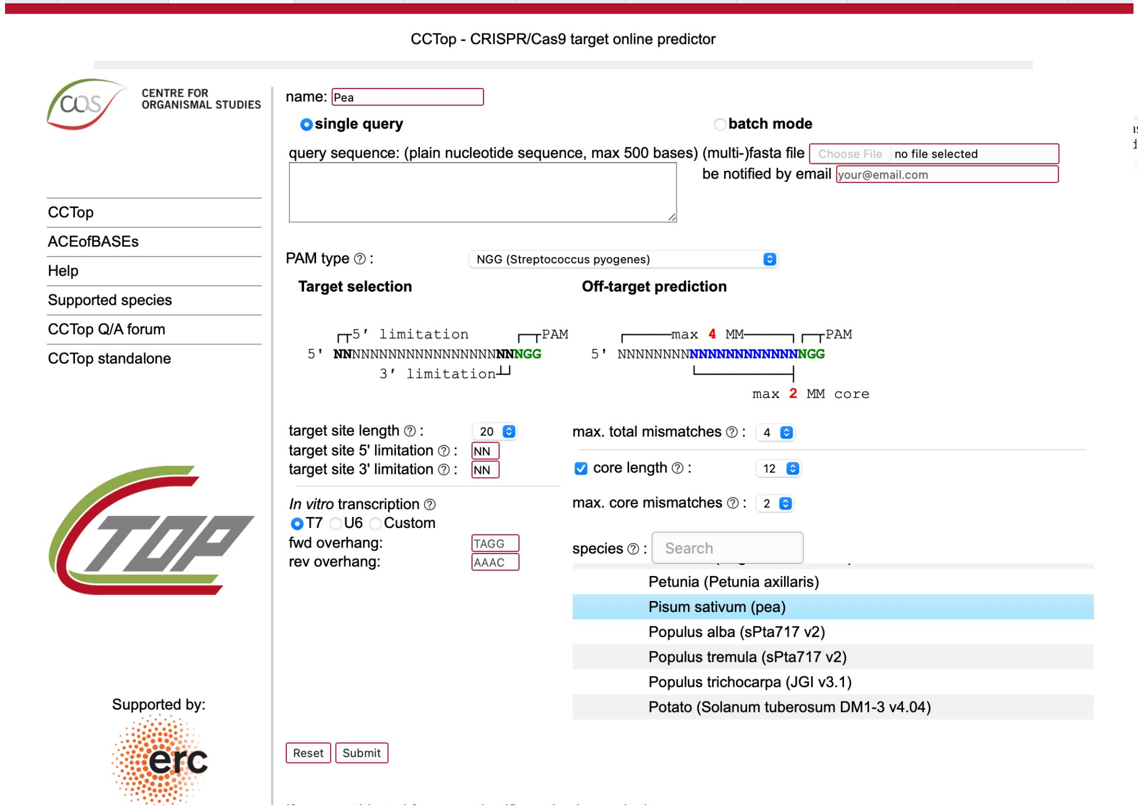
Select a Protospacer Adjacent Motif (PAM)

20bp-NGG - Sp Cas9, SpCas9-HF1, eSpCas9 1.1

See notes on enzymes in the manual.

SUBMIT

Version 5.01 - Documentation - Contact us - Downloads/local installation - Citation - License


CCTop - CRISPR/Cas9 target online predictor

CENTRE FOR ORGANISMAL STUDIES

- CCTop
- ACEofBASEs
- Help
- Supported species
- CCTop Q/A forum
- CCTop standalone

name: Pea

● single query ○ batch mode

query sequence: (plain nucleotide sequence, max 500 bases) (multi-)fasta file Choose File no file selected

be notified by email your@email.com

PAM type ⓘ : NGG (Streptococcus pyogenes)

Target selection

```
   ┌┐5' limitation       ┌┐PAM
5' NNNNNNNNNNNNNNNNNNNNNNNGG
       3' limitation┘
```

Off-target prediction

```
     ┌──max 4 MM──┐ ┌┐PAM
5' NNNNNNNNNNNNNNNNNNNNNGG
           └──────┘
           max 2 MM core
```

target site length ⓘ : 20

target site 5' limitation ⓘ : NN

target site 3' limitation ⓘ : NN

In vitro transcription ⓘ

● T7 ○ U6 ○ Custom

fwd overhang: TAGG

rev overhang: AAAC

max. total mismatches ⓘ : 4

☑ core length ⓘ : 12

max. core mismatches ⓘ : 2

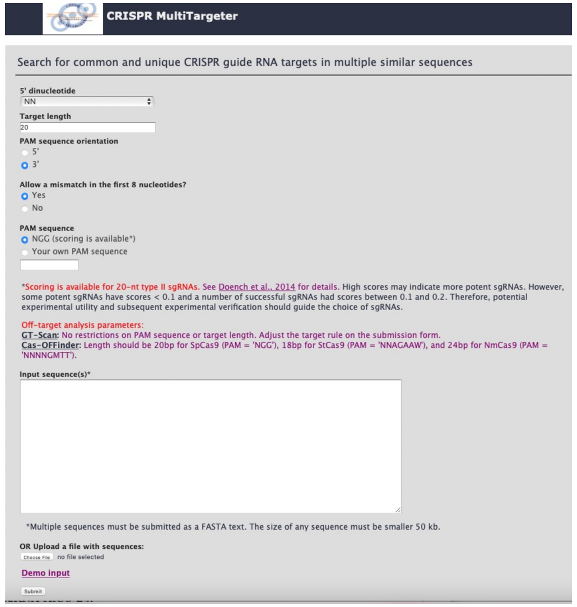
species ⓘ : Search

- Petunia (Petunia axillaris)
- Pisum sativum (pea)
- Populus alba (sPta717 v2)
- Populus tremula (sPta717 v2)
- Populus trichocarpa (JGI v3.1)
- Potato (Solanum tuberosum DM1-3 v4.04)

Reset Submit

Supported by:

erc


CRISPR MultiTargeter

Search for common and unique CRISPR guide RNA targets in multiple similar sequences

5' dinucleotide: NN

Target length: 20

PAM sequence orientation

○ 5'
● 3'

Allow a mismatch in the first 8 nucleotides?

● Yes
○ No

PAM sequence

● NGG (scoring is available*)
○ Your own PAM sequence

*Scoring is available for 20-nt type II sgRNAs. See Doench et al., 2014 for details. High scores may indicate more potent sgRNAs. However, some potent sgRNAs have scores < 0.1 and a number of successful sgRNAs had scores between 0.1 and 0.2. Therefore, potential experimental utility and subsequent experimental verification should guide the choice of sgRNAs.

Off-target analysis parameters:

GT-Scan: No restrictions on PAM sequence or target length. Adjust the target rule on the submission form.

Cas-OFFinder: Length should be 20bp for SpCas9 (PAM = 'NGG'), 18bp for StCas9 (PAM = 'NNAGAAW'), and 24bp for NmCas9 (PAM = 'NNNNGMTT').

Input sequence(s)*

*Multiple sequences must be submitted as a FASTA text. The size of any sequence must be smaller 50 kb.

OR Upload a file with sequences: Choose File no file selected

Demo input

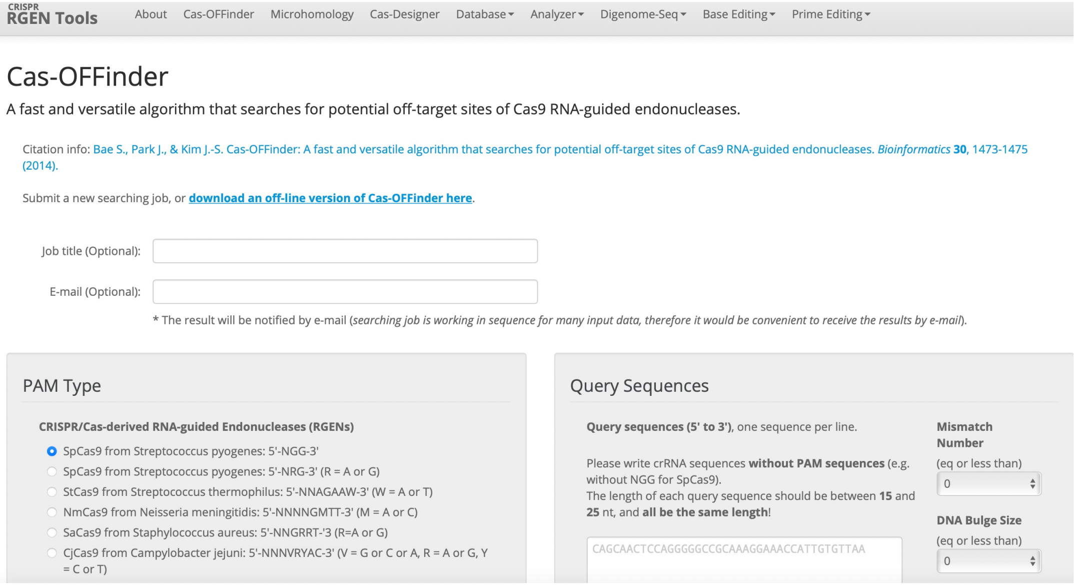
Submit

gRNAs targeting Multisequence

https://multicrispr.net/multalign_input.html

Off Target Prediction

http://www.rgenome.net/cas-offinder/



Cas-OFFinder

A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases.

Citation info: Bae S., Park J., & Kim J.-S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473-1475 (2014).

Submit a new searching job, or **download an off-line version of Cas-OFFinder here**.

Job title (Optional):

E-mail (Optional):

* The result will be notified by e-mail (*searching job is working in sequence for many input data, therefore it would be convenient to receive the results by e-mail*).

PAM Type

CRISPR/Cas-derived RNA-guided Endonucleases (RGENs)

- ● SpCas9 from Streptococcus pyogenes: 5'-NGG-3'
- ○ SpCas9 from Streptococcus pyogenes: 5'-NRG-3' (R = A or G)
- ○ StCas9 from Streptococcus thermophilus: 5'-NNAGAAW-3' (W = A or T)
- ○ NmCas9 from Neisseria meningitidis: 5'-NNNNGMTT-3' (M = A or C)
- ○ SaCas9 from Staphylococcus aureus: 5'-NNGRRT-'3 (R=A or G)
- ○ CjCas9 from Campylobacter jejuni: 5'-NNNVRYAC-3' (V = G or C or A, R = A or G, Y = C or T)

Query Sequences

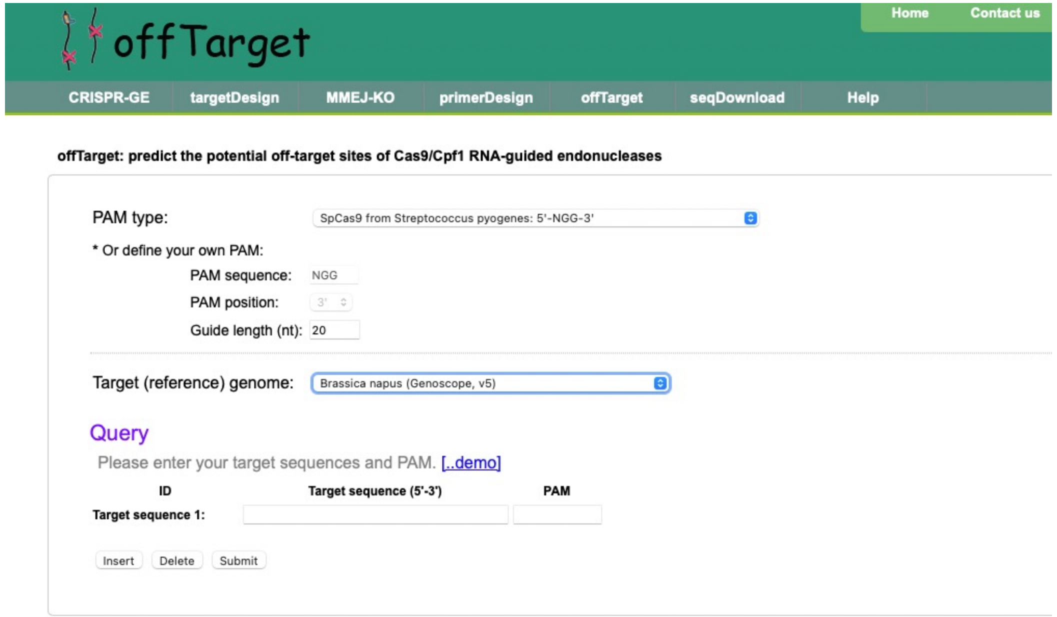
Query sequences (5' to 3'), one sequence per line.

Please write crRNA sequences **without PAM sequences** (e.g. without NGG for SpCas9).
The length of each query sequence should be between **15** and **25** nt, and **all be the same length**!

CAGCAACTCCAGGGGGCCGCAAAGGAAACCATTGTGTTAA

Mismatch Number
(eq or less than)
0

DNA Bulge Size
(eq or less than)
0



offTarget: predict the potential off-target sites of Cas9/Cpf1 RNA-guided endonucleases

PAM type: SpCas9 from Streptococcus pyogenes: 5'-NGG-3'

* Or define your own PAM:
PAM sequence: NGG
PAM position: 3'
Guide length (nt): 20

Target (reference) genome: Brassica napus (Genoscope, v5)

Query

Please enter your target sequences and PAM. [..demo]

ID	Target sequence (5'-3')	PAM
Target sequence 1:		

Target Genome

Organism Type
Plants

Genomes

- ○ Arabidopsis thaliana (TAIR10) - Thale cress
- ○ Oryza sativa (OSv4) - Rice
- ○ Solanum lycopersicum (SL2.4) - Tomato
- ○ Zea mays (AGPv3) - Corn
- ○ Chlamydomonas reinhardtii (Chlre4)
- ○ Solanum tuberosum (PGSC v4.03) - Potato
- ○ Glycine max (v1.0) - Soybean
- ○ Vitis vinifera (IGGP_12X/Ensplant26) - European grapevine
- ○ Manihot esculenta (JGI 4.1) - Cassava
- ○ Malus domestica (JGI 1.0) - Apple
- ○ Hordeum Vulgare (Ensembl Plants 28) - Barley
- ○ Nicotiana benthamiana (v1.0.1)
- ○ Fragaria vesca (1.0) - Wild strawberry
- ○ Citrus sinensis (1.0) - Sweet orange
- ○ Theobroma cacao (CIRAD 1.0) - Cacao
- ○ Theobroma cacao (CGD 1.1) - Cacao
- ○ Solanum lycopersicum (SL2.5) - Tomato
- ○ Musa acuminata (MA1) - Banana
- ○ Arachis ipaensis (PeanutBase v1.0)
- ○ Arachis duranensis (PeanutBase v1.0)
- ○ Actinidia chinensis (from IKGC) - Kiwifruit
- ● Brassica napus (v4.1) - Rapeseed
- ○ Glycine max (v2.0) - Soybean
- ○ Sorghum bicolor (v1.0) - Cereal grass

http://skl.scau.edu.cn/offtarget/

Questions?

selvaraju.kanagarajan@slu.se