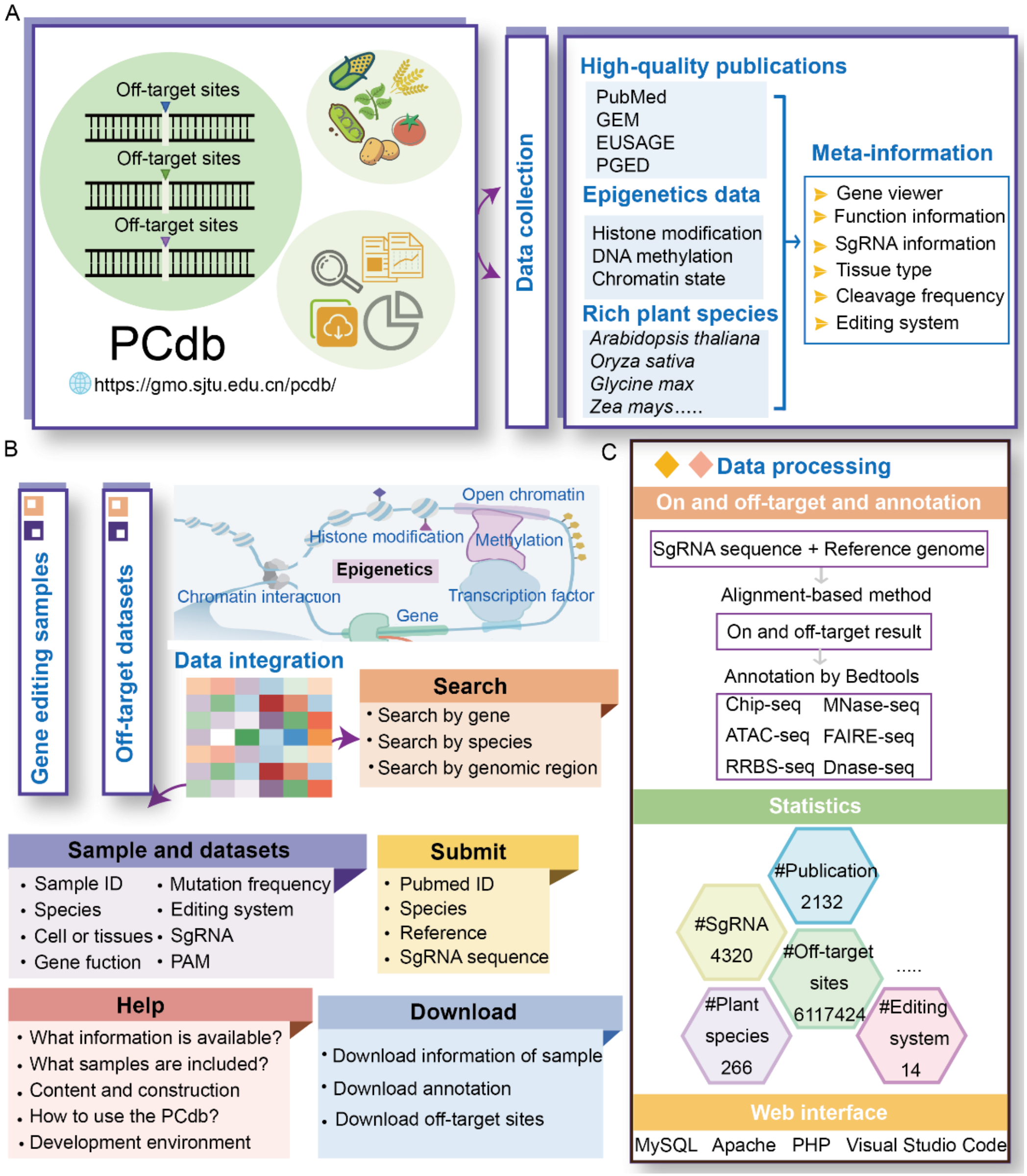
PCdb tutorial

1. Description of PCdb

We present PCdb (Plant CRISPR Database; https://gmo.sjtu.edu.cn/pcdb), a comprehensive plant-focused database by integrating experimentally validated sgRNAs, their annotated genomic contexts, genome-wide off-target predictions, and multi-layered epigenomic annotations. PCdb encompasses 6,172 manually curated editing records from 2,132 publications, covering 4,320 unique sgRNAs and 6,117,424 predicted off-target sites across nine major plant species. Uniquely, PCdb contextualizes potential editing outcomes-both on- target and off-target-within the chromatin landscape by incorporating DNA methylation profiles, chromatin accessibility data, and histone modification patterns. The database features an intuitive web interface supporting flexible queries, interactive visualization tools, and comprehensive analytical modules for both sgRNA efficiency assessment and off-target analysis. A case study reanalysis of rice yield-related genes demonstrates PCdb's capability to provide a comprehensive performance profile, evaluating both on-target characteristics and off-target risks within their native epigenomic context. Through systematic analysis of the database, we reveal critical sequence and chromatin features influencing editing outcomes, providing novel insights for improved gene editing efficacy and specificity.



1.1 Data acquisition

Plant genome editing data were collected through systematic PubMed literature using keywords related to CRISPR/Cas and plants. After manual curation, 2,132 original English research articles with complete sgRNA sequences, clearly defined editing systems, target genes, species, and phenotypes were included. To expand coverage, data from 609 additional publications were integrated from curated databases such as GEM and EUSAGE, focusing on major plant species. For each record, essential metadata-including species, sgRNA and PAM sequences, Cas variants, target genes, and tissue or cell types were extracted and manually validated. Records with incomplete or ambiguous information were excluded to ensure data quality and consistency.

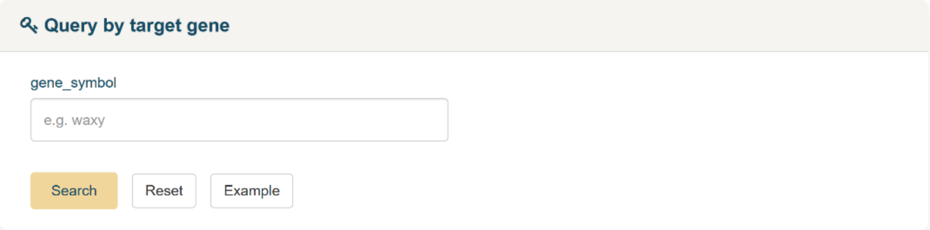
1.2 Data processing

PCdb employs a unified and standardized data processing pipeline to ensure high accuracy, consistency, and cross-study comparability of plant genome editing data. All sgRNA sequences are uniformly mapped to RefSeq reference genomes to precisely define on-target sites and to computationally predict potential off-target regions using Cas-OFFinder, which should be considered candidates for experimental validation. Each on-target and off-target site is annotated with genomic coordinates, functional elements (e.g., coding regions, introns, and UTRs), epigenomic context including histone modifications, chromatin accessibility, and DNA methylation, as well as sgRNA efficiency and specificity scores derived from established algorithms. In addition, PCdb integrates curated functional annotations of target genes from the literature and links them to external resources, providing users with comprehensive biological and regulatory insights to support sgRNA evaluation and experimental design.

2. Search by target

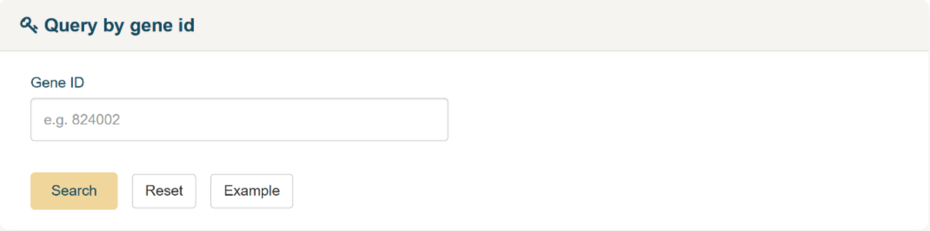
2.1 Target information

Provides gene symbol or ID, function, and a pie chart that visually represents the species distribution involved in gene editing studies related to this gene.



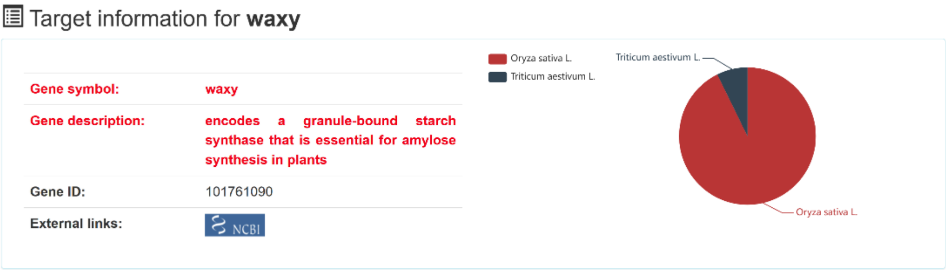
Query by target gene

gene_symbol: e.g. waxy



Query by gene id

Gene ID: e.g. 824002

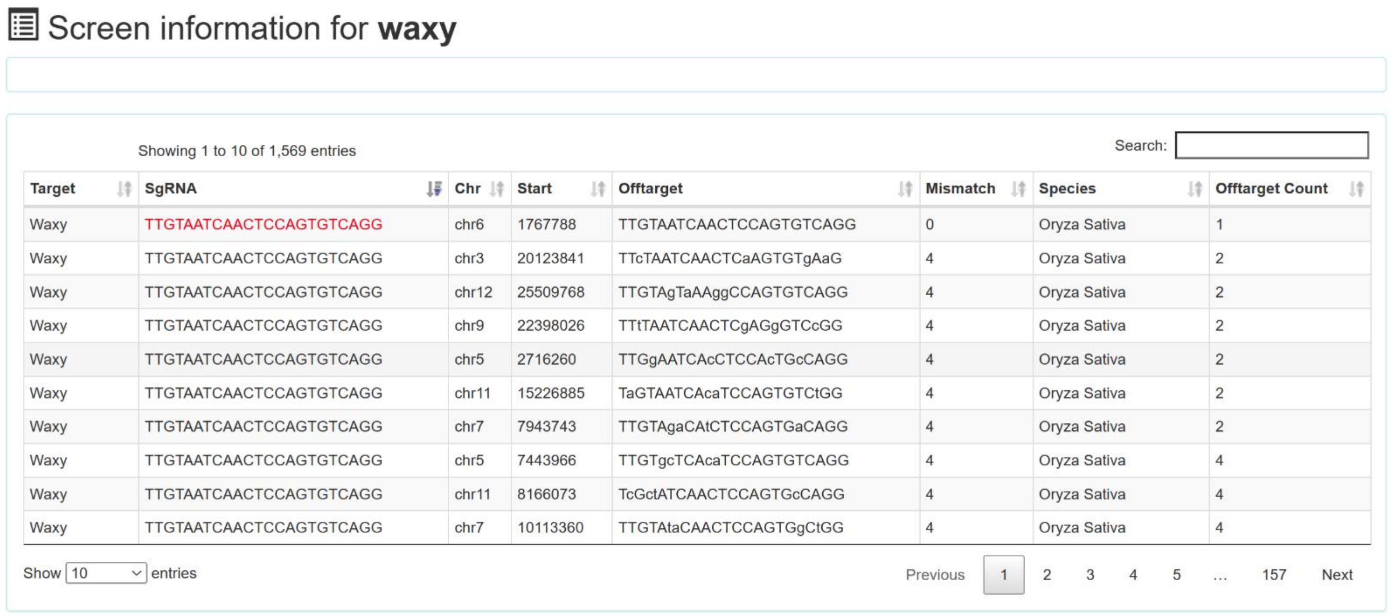


Target information for **waxy**

Gene symbol:	**waxy**
Gene description:	**encodes a granule-bound starch synthase that is essential for amylose synthesis in plants**
Gene ID:	101761090
External links:	NCBI

2.2 Summary table of this gene-related sgRNAs

Lists all sgRNAs associated with the gene in a tabular format, including sgRNA sequences, PAM sequences, the editing systems used, and references to related studies. And shows detailed off-target analysis for the sgRNAs, including off-target sequences, number of mismatches, species, and the total number of predicted off-target sites for each sgRNA.



Screen information for **waxy**

Showing 1 to 10 of 1,569 entries

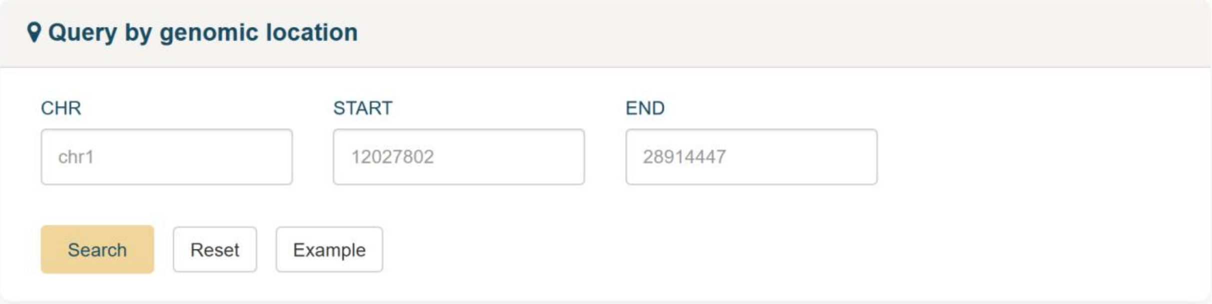
Search:

Target	SgRNA	Chr	Start	Offtarget	Mismatch	Species	Offtarget Count
Waxy	TTGTAATCAACTCCAGTGTCAGG	chr6	1767788	TTGTAATCAACTCCAGTGTCAGG	0	Oryza Sativa	1
Waxy	TTGTAATCAACTCCAGTGTCAGG	chr3	20123841	TTcTAATCAACTCaAGTGTgAaG	4	Oryza Sativa	2
Waxy	TTGTAATCAACTCCAGTGTCAGG	chr12	25509768	TTGTAgTaAAggCCAGTGTCAGG	4	Oryza Sativa	2
Waxy	TTGTAATCAACTCCAGTGTCAGG	chr9	22398026	TTtTAATCAACTCgAGgGTCcGG	4	Oryza Sativa	2
Waxy	TTGTAATCAACTCCAGTGTCAGG	chr5	2716260	TTGgAATCAcCTCCAcTGcCAGG	4	Oryza Sativa	2
Waxy	TTGTAATCAACTCCAGTGTCAGG	chr11	15226885	TaGTAATCAcaTCCAGTGTCtGG	4	Oryza Sativa	2
Waxy	TTGTAATCAACTCCAGTGTCAGG	chr7	7943743	TTGTAgaCAtCTCCAGTGaCAGG	4	Oryza Sativa	2
Waxy	TTGTAATCAACTCCAGTGTCAGG	chr5	7443966	TTGTgcTCAcaTCCAGTGTCAGG	4	Oryza Sativa	4
Waxy	TTGTAATCAACTCCAGTGTCAGG	chr11	8166073	TcGctATCAACTCCAGTGcCAGG	4	Oryza Sativa	4
Waxy	TTGTAATCAACTCCAGTGTCAGG	chr7	10113360	TTGTAtaCAACTCCAGTGgCtGG	4	Oryza Sativa	4

Show 10 entries

3. Search by genomic region

This search approach enables a focused exploration of a particular genomic locus, helping users to evaluate sgRNA sequences within a defined region and assess the associated off-target risks. A table listing all sgRNAs located within the specified genomic region, along with their corresponding off-target sequences. Each sgRNA is annotated with multiple epigenetic features across different species, including ATAC-seq signal S, DNase sensitivity, histone modifications (H3K4me1, H3K4me3, H3K27ac), and DNA methylation levels.

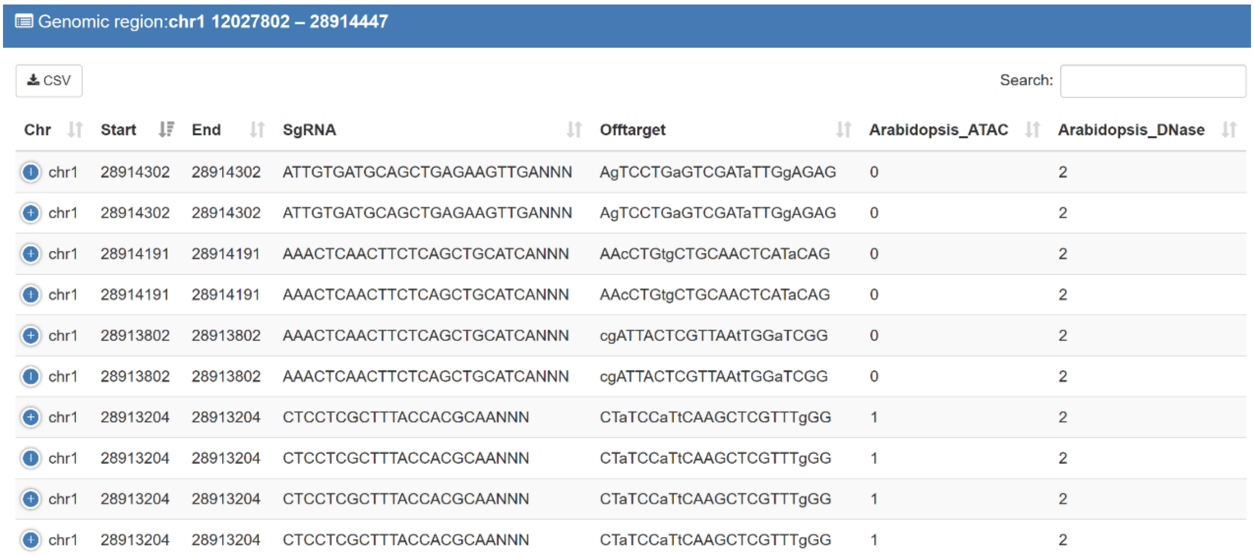


Query by genomic location

CHR: chr1

START: 12027802

END: 28914447



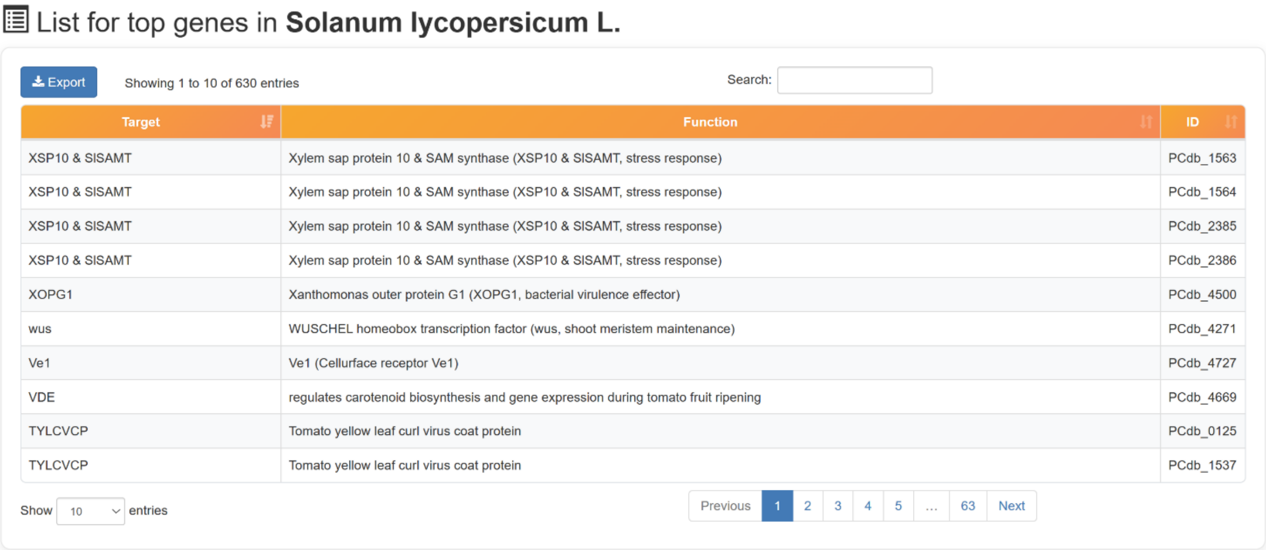
Genomic region:**chr1 12027802 – 28914447**

Chr	Start	End	SgRNA	Offtarget	Arabidopsis_ATAC	Arabidopsis_DNase
chr1	28914302	28914302	ATTGTGATGCAGCTGAGAAGTTGANNN	AgTCCTGaGTCGATaTTGgAGAG	0	2
chr1	28914302	28914302	ATTGTGATGCAGCTGAGAAGTTGANNN	AgTCCTGaGTCGATaTTGgAGAG	0	2
chr1	28914191	28914191	AAACTCAACTTCTCAGCTGCATCANNN	AAcCTGtgCTGCAACTCATaCAG	0	2
chr1	28914191	28914191	AAACTCAACTTCTCAGCTGCATCANNN	AAcCTGtgCTGCAACTCATaCAG	0	2
chr1	28913802	28913802	AAACTCAACTTCTCAGCTGCATCANNN	cgATTACTCGTTAAtTGGaTCGG	0	2
chr1	28913802	28913802	AAACTCAACTTCTCAGCTGCATCANNN	cgATTACTCGTTAAtTGGaTCGG	0	2
chr1	28913204	28913204	CTCCTCGCTTTACCACGCAANNN	CTaTCCaTtCAAGCTCGTTTgGG	1	2
chr1	28913204	28913204	CTCCTCGCTTTACCACGCAANNN	CTaTCCaTtCAAGCTCGTTTgGG	1	2
chr1	28913204	28913204	CTCCTCGCTTTACCACGCAANNN	CTaTCCaTtCAAGCTCGTTTgGG	1	2
chr1	28913204	28913204	CTCCTCGCTTTACCACGCAANNN	CTaTCCaTtCAAGCTCGTTTgGG	1	2

4 Search by species

4.1 Target gene list for the species

A table listing key target genes studied in the selected species, including gene names, functional annotations, and database entry IDs.



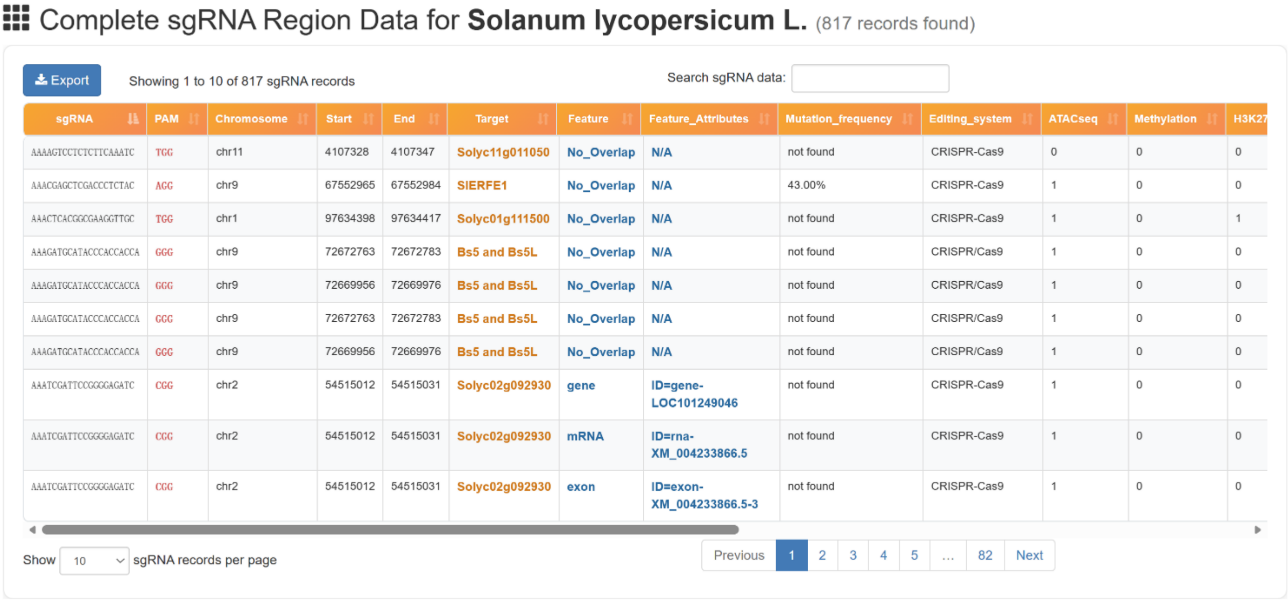
List for top genes in **Solanum lycopersicum L.**

Showing 1 to 10 of 630 entries

Target	Function	ID
XSP10 & SlSAMT	Xylem sap protein 10 & SAM synthase (XSP10 & SlSAMT, stress response)	PCdb_1563
XSP10 & SlSAMT	Xylem sap protein 10 & SAM synthase (XSP10 & SlSAMT, stress response)	PCdb_1564
XSP10 & SlSAMT	Xylem sap protein 10 & SAM synthase (XSP10 & SlSAMT, stress response)	PCdb_2385
XSP10 & SlSAMT	Xylem sap protein 10 & SAM synthase (XSP10 & SlSAMT, stress response)	PCdb_2386
XOPG1	Xanthomonas outer protein G1 (XOPG1, bacterial virulence effector)	PCdb_4500
wus	WUSCHEL homeobox transcription factor (wus, shoot meristem maintenance)	PCdb_4271
Ve1	Ve1 (Cellurface receptor Ve1)	PCdb_4727
VDE	regulates carotenoid biosynthesis and gene expression during tomato fruit ripening	PCdb_4669
TYLCVCP	Tomato yellow leaf curl virus coat protein	PCdb_0125
TYLCVCP	Tomato yellow leaf curl virus coat protein	PCdb_1537

4.2 Complete sgRNA information for the species

For each sgRNA, the table provides its sequence, PAM type, genomic location, target gene information, and functional annotation (e.g., gene, exon, mRNA). It also reports the associated genome editing system, observed mutation frequency (if available), and integrated epigenomic features such as chromatin accessibility, DNA methylation, and histone modification signals. Together, these fields enable users to evaluate sgRNA genomic context, regulatory environment, and potential editing performance in a unified view.



Complete sgRNA Region Data for **Solanum lycopersicum L.** (817 records found)

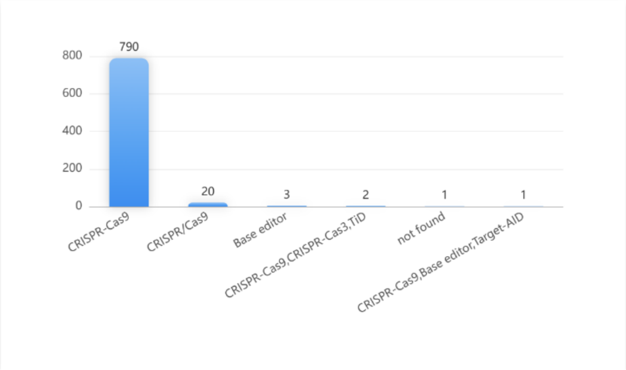
Showing 1 to 10 of 817 sgRNA records

sgRNA	PAM	Chromosome	Start	End	Target	Feature	Feature_Attributes	Mutation_frequency	Editing_system	ATACseq	Methylation	H3K27
AAAAGTCCTCTCTTCAAATC	TGG	chr11	4107328	4107347	Solyc11g011050	No_Overlap	N/A	not found	CRISPR-Cas9	0	0	0
AAACGAGCTCGACCCTCTAC	AGG	chr9	67552965	67552984	SlERFE1	No_Overlap	N/A	43.00%	CRISPR-Cas9	1	0	0
AAACTCACGGCGAAGGTTGC	TGG	chr1	97634398	97634417	Solyc01g111500	No_Overlap	N/A	not found	CRISPR-Cas9	1	0	1
AAAGATGCATACCCACCACCA	GGG	chr9	72672763	72672783	Bs5 and Bs5L	No_Overlap	N/A	not found	CRISPR/Cas9	1	0	0
AAAGATGCATACCCACCACCA	GGG	chr9	72669956	72669976	Bs5 and Bs5L	No_Overlap	N/A	not found	CRISPR/Cas9	1	0	0
AAAGATGCATACCCACCACCA	GGG	chr9	72672763	72672783	Bs5 and Bs5L	No_Overlap	N/A	not found	CRISPR/Cas9	1	0	0
AAAGATGCATACCCACCACCA	GGG	chr9	72669956	72669976	Bs5 and Bs5L	No_Overlap	N/A	not found	CRISPR/Cas9	1	0	0
AAATCGATTCCGGGGAGATC	CGG	chr2	54515012	54515031	Solyc02g092930	gene	ID=gene-LOC101249046	not found	CRISPR-Cas9	1	0	0
AAATCGATTCCGGGGAGATC	CGG	chr2	54515012	54515031	Solyc02g092930	mRNA	ID=rna-XM_004233866.5	not found	CRISPR-Cas9	1	0	0
AAATCGATTCCGGGGAGATC	CGG	chr2	54515012	54515031	Solyc02g092930	exon	ID=exon-XM_004233866.5-3	not found	CRISPR-Cas9	1	0	0

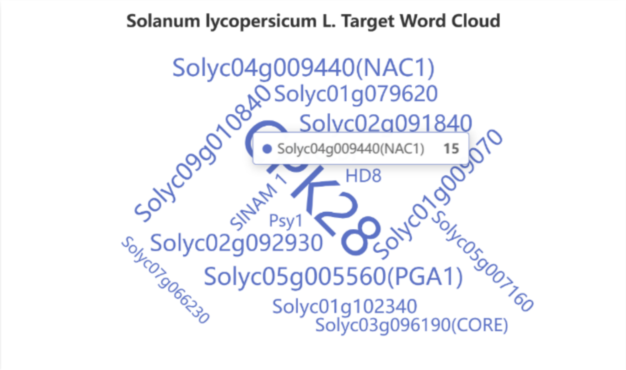
4.3 Statistics of gene editing systems and targets related to the species

A bar chart showing the distribution of different genome editing systems used in research on this species. Word cloud chart of target genes in this species. The visualization helps users understand the current trends and technology preferences.

Editing System Distribution

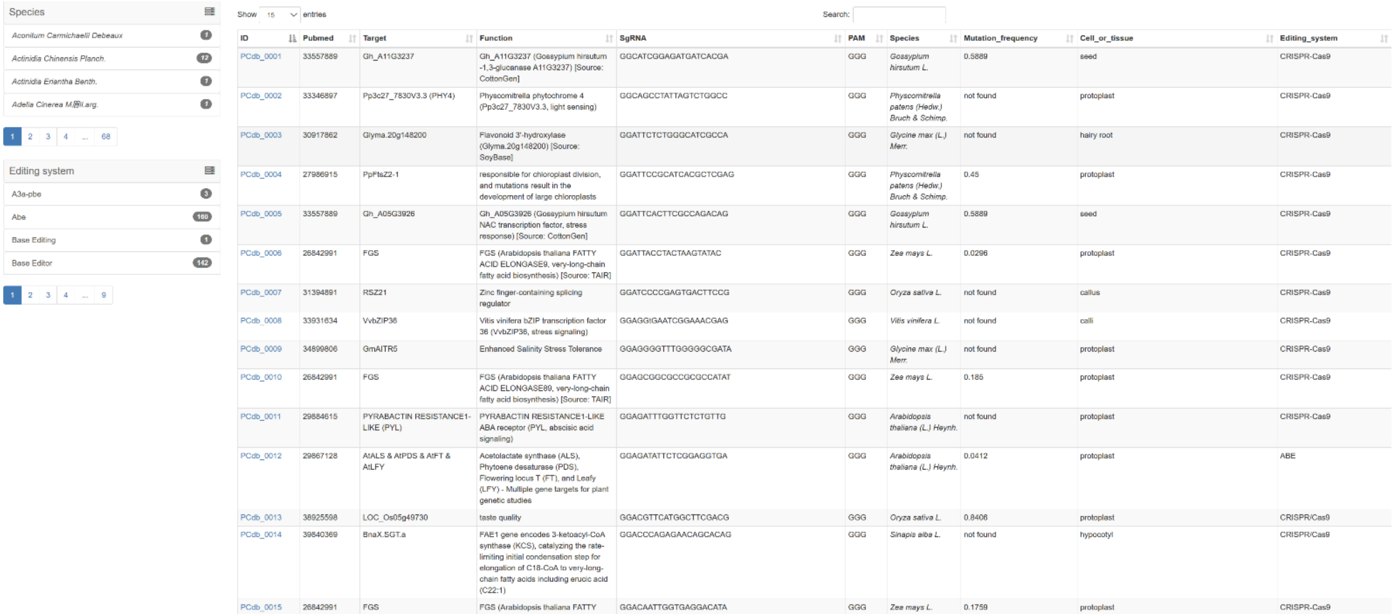


Word Cloud Chart of Target Genes



5. Browse by source

The ‘Browse’ page is organized as an interactive and alphanumerically sortable table that allows users to quickly browse samples and customize filters through ‘Species’, ‘Editing system’. Users can use the ‘Show entries’ drop-down menu to get different numbers of records per page.

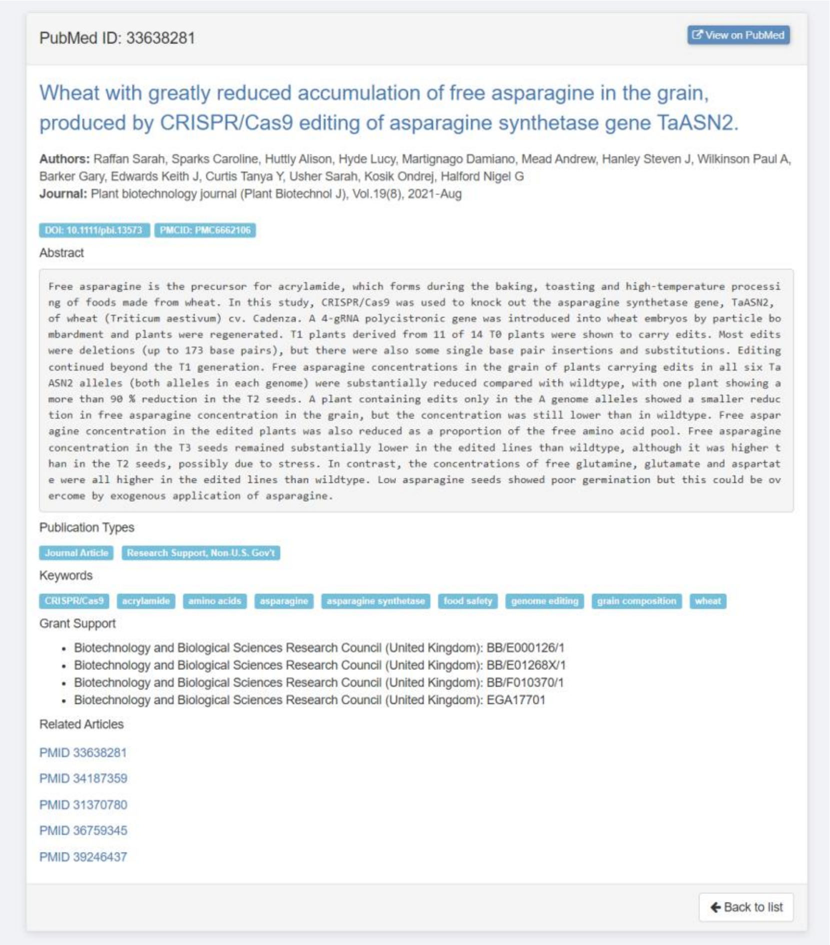


Species	
Aconitum Carmichaelii Debeaux	1
Actinidia Chinensis Planch.	12
Actinidia Eriantha Benth.	1
Adelia Cinerea M.圖ll.arg.	1

Editing system	
A3a-pbe	3
Abe	160
Base Editing	1
Base Editor	142

ID	Pubmed	Target	Function	SgRNA	PAM	Species	Mutation_frequency	Cell_or_tissue	Editing_system
PCdb_0001	33557889	Gh_A11G3237	Gh_A11G3237 (Gossypium hirsutum -1,3-glucanase A11G3237) [Source: CottonGen]	GGCATCGGAGATGATCACGA	GGG	*Gossypium hirsutum L.*	0.5889	seed	CRISPR-Cas9
PCdb_0002	33346897	Pp3c27_7830V3.3 (PHY4)	Physcomitrella phytochrome 4 (Pp3c27_7830V3.3, light sensing)	GGCAGCCTATTAGTCTGGCC	GGG	*Physcomitrella patens (Hedw.) Bruch & Schimp.*	not found	protoplast	CRISPR-Cas9
PCdb_0003	30917862	Glyma.20g148200	Flavonoid 3'-hydroxylase (Glyma.20g148200) [Source: SoyBase]	GGATTCTCTGGGCATCGCCA	GGG	*Glycine max (L.) Merr.*	not found	hairy root	CRISPR-Cas9
PCdb_0004	27986915	PpFtsZ2-1	responsible for chloroplast division, and mutations result in the development of large chloroplasts	GGATTCCGCATCACGCTCGAG	GGG	*Physcomitrella patens (Hedw.) Bruch & Schimp.*	0.45	protoplast	CRISPR-Cas9
PCdb_0005	33557889	Gh_A05G3926	Gh_A05G3926 (Gossypium hirsutum NAC transcription factor, stress response) [Source: CottonGen]	GGATTCACTTCGCCAGACAG	GGG	*Gossypium hirsutum L.*	0.5889	seed	CRISPR-Cas9
PCdb_0006	26842991	FGS	FGS (Arabidopsis thaliana FATTY ACID ELONGASE9, very-long-chain fatty acid biosynthesis) [Source: TAIR]	GGATTACCTACTAAGTATAC	GGG	*Zea mays L.*	0.0296	protoplast	CRISPR-Cas9
PCdb_0007	31394891	RSZ21	Zinc finger-containing splicing regulator	GGATCCCCGAGTGACTTCCG	GGG	*Oryza sativa L.*	not found	callus	CRISPR-Cas9
PCdb_0008	33931634	VvbZIP36	Vitis vinifera bZIP transcription factor 36 (VvbZIP36, stress signaling)	GGAGGtGAATCGGAAACGAG	GGG	*Vitis vinifera L.*	not found	calli	CRISPR-Cas9
PCdb_0009	34899806	GmAITR5	Enhanced Salinity Stress Tolerance	GGAGGGGTTTGGGGGCGATA	GGG	*Glycine max (L.) Merr.*	not found	protoplast	CRISPR-Cas9
PCdb_0010	26842991	FGS	FGS (Arabidopsis thaliana FATTY ACID ELONGASE89, very-long-chain fatty acid biosynthesis) [Source: TAIR]	GGAGCGGCGCCGCGCCATAT	GGG	*Zea mays L.*	0.185	protoplast	CRISPR-Cas9
PCdb_0011	29884615	PYRABACTIN RESISTANCE1-LIKE (PYL)	PYRABACTIN RESISTANCE1-LIKE ABA receptor (PYL, abscisic acid signaling)	GGAGATTTGGTTCTCTGTTG	GGG	*Arabidopsis thaliana (L.) Heynh.*	not found	protoplast	CRISPR-Cas9
PCdb_0012	29867128	AtALS & AtPDS & AtFT & AtLFY	Acetolactate synthase (ALS), Phytoene desaturase (PDS), Flowering locus T (FT), and Leafy (LFY) - Multiple gene targets for plant genetic studies	GGAGATATTCTCGGAGGTGA	GGG	*Arabidopsis thaliana (L.) Heynh.*	0.0412	protoplast	ABE
PCdb_0013	38925598	LOC_Os05g49730	taste quality	GGACGTTCATGGCTTCGACG	GGG	*Oryza sativa L.*	0.8406	protoplast	CRISPR/Cas9
PCdb_0014	39840369	BnaX.SGT.a	FAE1 gene encodes 3-ketoacyl-CoA synthase (KCS), catalyzing the rate-limiting initial condensation step for elongation of C18-CoA to very-long-chain fatty acids including erucic acid (C22:1)	GGACCCAGAGAACAGCACAG	GGG	*Sinapis alba L.*	not found	hypocotyl	CRISPR/Cas9
PCdb_0015	26842991	FGS	FGS (Arabidopsis thaliana FATTY	GGACAATTGGTGAGGACATA	GGG	*Zea mays L.*	0.1759	protoplast	CRISPR-Cas9

To view a given sample, users only need to click on the ‘Sample ID’. The information on the sample details page includes the literature source and a summary of the literature.


PubMed ID: 33638281

Wheat with greatly reduced accumulation of free asparagine in the grain, produced by CRISPR/Cas9 editing of asparagine synthetase gene TaASN2.

Authors: Raffan Sarah, Sparks Caroline, Huttly Alison, Hyde Lucy, Martignago Damiano, Mead Andrew, Hanley Steven J, Wilkinson Paul A, Barker Gary, Edwards Keith J, Curtis Tanya Y, Usher Sarah, Kosik Ondrej, Halford Nigel G
Journal: Plant biotechnology journal (Plant Biotechnol J), Vol.19(8), 2021-Aug

DOI: 10.1111/pbi.13573 PMCID: PMC6662106

Abstract

Free asparagine is the precursor for acrylamide, which forms during the baking, toasting and high-temperature processing of foods made from wheat. In this study, CRISPR/Cas9 was used to knock out the asparagine synthetase gene, TaASN2, of wheat (Triticum aestivum) cv. Cadenza. A 4-gRNA polycistronic gene was introduced into wheat embryos by particle bombardment and plants were regenerated. T1 plants derived from 11 of 14 T0 plants were shown to carry edits. Most edits were deletions (up to 173 base pairs), but there were also some single base pair insertions and substitutions. Editing continued beyond the T1 generation. Free asparagine concentrations in the grain of plants carrying edits in all six TaASN2 alleles (both alleles in each genome) were substantially reduced compared with wildtype, with one plant showing a more than 90 % reduction in the T2 seeds. A plant containing edits only in the A genome alleles showed a smaller reduction in free asparagine concentration in the grain, but the concentration was still lower than in wildtype. Free asparagine concentration in the edited plants was also reduced as a proportion of the free amino acid pool. Free asparagine concentration in the T3 seeds remained substantially lower in the edited lines than wildtype, although it was higher than in the T2 seeds, possibly due to stress. In contrast, the concentrations of free glutamine, glutamate and aspartate were all higher in the edited lines than wildtype. Low asparagine seeds showed poor germination but this could be overcome by exogenous application of asparagine.

Publication Types

Journal Article Research Support, Non-U.S. Gov't

Keywords

CRISPR/Cas9 acrylamide amino acids asparagine asparagine synthetase food safety genome editing grain composition wheat

Grant Support

- Biotechnology and Biological Sciences Research Council (United Kingdom): BB/E000126/1
- Biotechnology and Biological Sciences Research Council (United Kingdom): BB/E01268X/1
- Biotechnology and Biological Sciences Research Council (United Kingdom): BB/F010370/1
- Biotechnology and Biological Sciences Research Council (United Kingdom): EGA17701

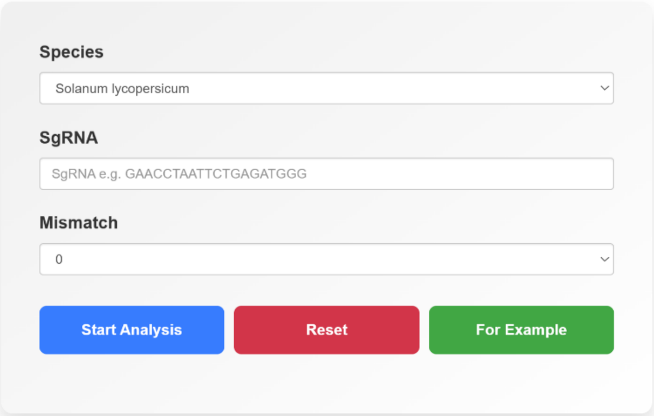
Related Articles

PMID 33638281
PMID 34187359
PMID 31370780
PMID 36759345
PMID 39246437

6. Analysis function

Users can specify the number of mismatches to filter results and instantly retrieve detailed information about potential off-target sites, including chromosome location, mismatch count, and genomic start and end positions. The results are presented in a searchable, sortable table for easy interpretation.

Analyze the SgRNA and Off-Target Efficiency


Species: Solanum lycopersicum

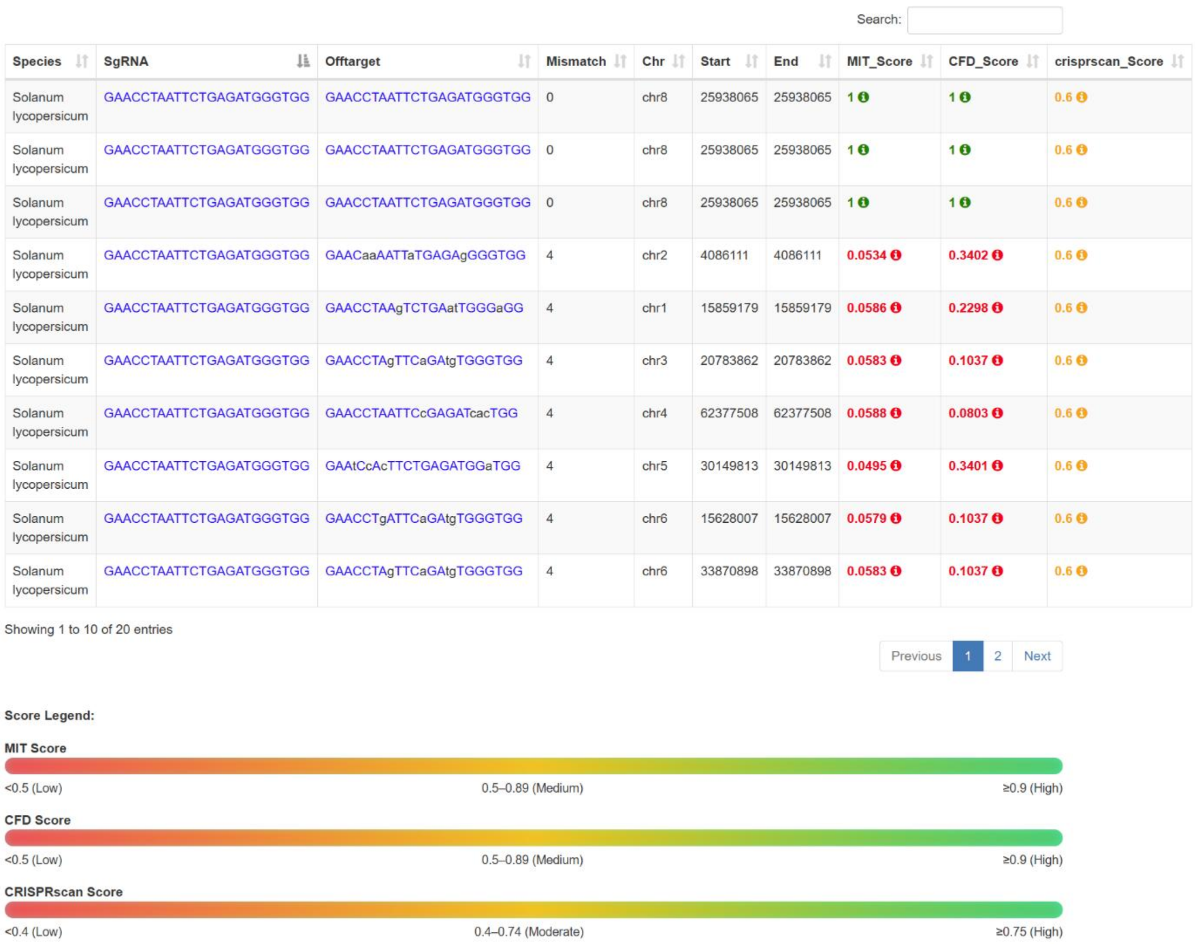
SgRNA: SgRNA e.g. GAACCTAATTCTGAGATGGG

Mismatch: 0


Function Introduction

This module allows users to analyze **sgRNA efficiency** and **off-target risks** across various plant genomes. By selecting a species, entering the sgRNA sequence, and setting a mismatch threshold (0–4), users can predict **off-target sites** and assess editing precision. Results include **off-target alignments**, MIT, CFD, and CRISPRscan scores, plus **mutation heatmaps** and **nucleotide distributions**, supporting optimized **sgRNA design** and risk evaluation for plant genome editing.

Each sgRNA using three distinct scoring systems: MIT Score, CFD Score, and CRISPRscan Score. This multi-metric strategy provides a more robust and nuanced assessment of potential off-target effects. Unlike approaches that rely on a single metric, our integrated framework ensures a balanced consideration of both safety and efficacy, enhancing the reliability of sgRNA selection for genome editing.



Search:

Species	SgRNA	Offtarget	Mismatch	Chr	Start	End	MIT_Score	CFD_Score	crisprscan_Score
Solanum lycopersicum	GAACCTAATTCTGAGATGGGTGG	GAACCTAATTCTGAGATGGGTGG	0	chr8	25938065	25938065	1	1	0.6
Solanum lycopersicum	GAACCTAATTCTGAGATGGGTGG	GAACCTAATTCTGAGATGGGTGG	0	chr8	25938065	25938065	1	1	0.6
Solanum lycopersicum	GAACCTAATTCTGAGATGGGTGG	GAACCTAATTCTGAGATGGGTGG	0	chr8	25938065	25938065	1	1	0.6
Solanum lycopersicum	GAACCTAATTCTGAGATGGGTGG	GAACaaAATTaTGAGAgGGGTGG	4	chr2	4086111	4086111	0.0534	0.3402	0.6
Solanum lycopersicum	GAACCTAATTCTGAGATGGGTGG	GAACCTAAgTCTGAatTGGGaGG	4	chr1	15859179	15859179	0.0586	0.2298	0.6
Solanum lycopersicum	GAACCTAATTCTGAGATGGGTGG	GAACCTAgTTCaGAtgTGGGTGG	4	chr3	20783862	20783862	0.0583	0.1037	0.6
Solanum lycopersicum	GAACCTAATTCTGAGATGGGTGG	GAACCTAATTCcGAGATcacTGG	4	chr4	62377508	62377508	0.0588	0.0803	0.6
Solanum lycopersicum	GAACCTAATTCTGAGATGGGTGG	GAAtCcAcTTCTGAGATGGaTGG	4	chr5	30149813	30149813	0.0495	0.3401	0.6
Solanum lycopersicum	GAACCTAATTCTGAGATGGGTGG	GAACCTgATTCaGAtgTGGGTGG	4	chr6	15628007	15628007	0.0579	0.1037	0.6
Solanum lycopersicum	GAACCTAATTCTGAGATGGGTGG	GAACCTAgTTCaGAtgTGGGTGG	4	chr6	33870898	33870898	0.0583	0.1037	0.6

Showing 1 to 10 of 20 entries

Previous 1 2 Next

Score Legend:

MIT Score

<0.5 (Low) 0.5–0.89 (Medium) ≥0.9 (High)

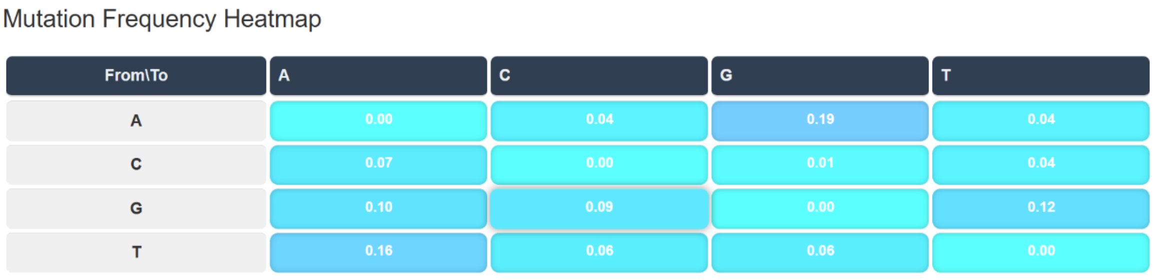
CFD Score

<0.5 (Low) 0.5–0.89 (Medium) ≥0.9 (High)

CRISPRscan Score

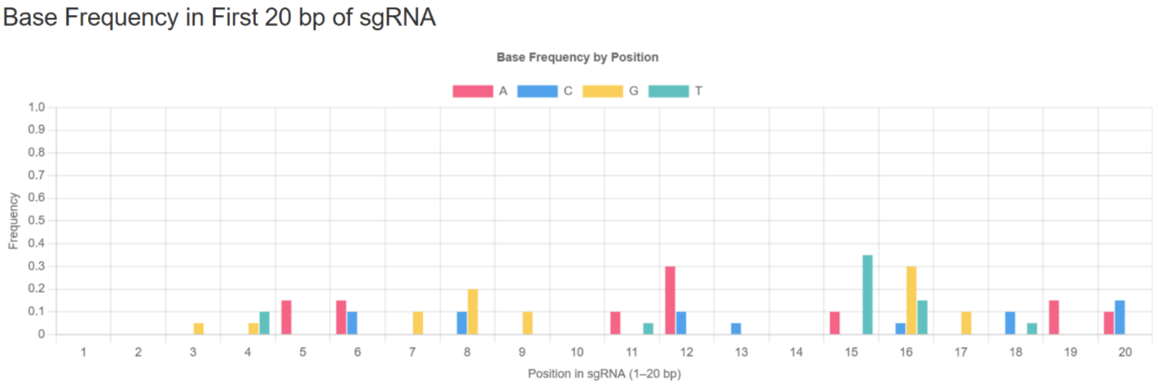
<0.4 (Low) 0.4–0.74 (Moderate) ≥0.75 (High)

To further enhance the accuracy and rational design of sgRNAs, we performed a detailed nucleotide-level analysis, including a mutation frequency heatmap and a base frequency distribution across the first 20 bp of sgRNAs. This level of granularity in sequence analysis represents an innovative step beyond conventional sgRNA design pipelines, enabling more informed refinement of guide sequences based on both mutational trends and positional base preferences.



Mutation Frequency Heatmap

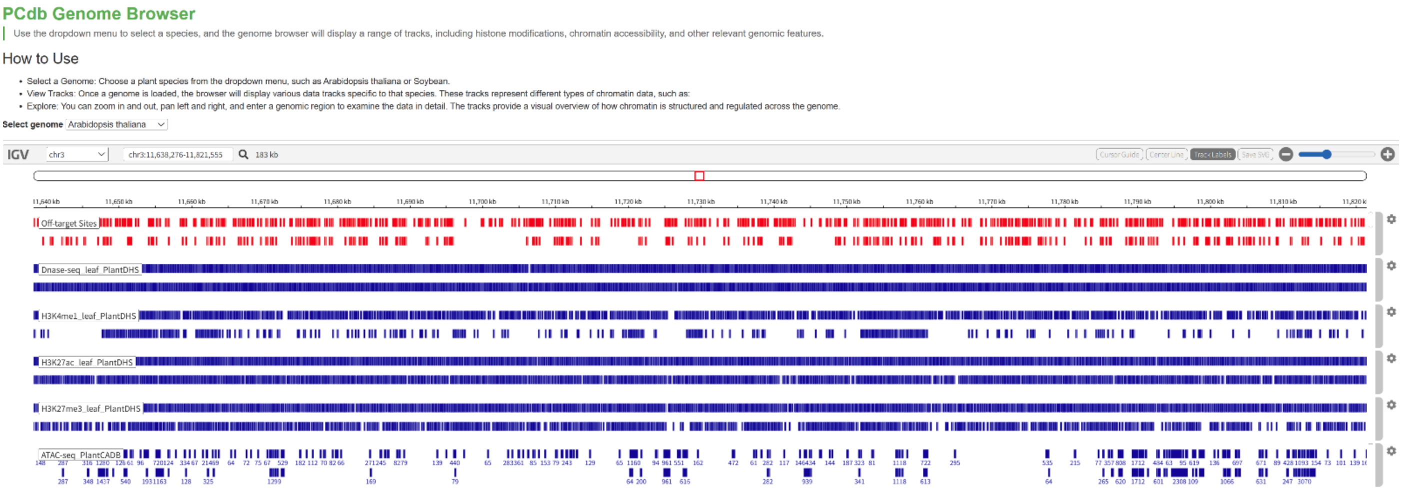
From\To	A	C	G	T
A	0.00	0.04	0.19	0.04
C	0.07	0.00	0.01	0.04
G	0.10	0.09	0.00	0.12
T	0.16	0.06	0.06	0.00



Base Frequency in First 20 bp of sgRNA

7. Genome browser

Users can view the regulatory information of off-target by using the 'Genome-Browser' page. Through useful tracks, users can get information including ChIP-seq, accessible chromatin regions and so on.



PCdb Genome Browser

Use the dropdown menu to select a species, and the genome browser will display a range of tracks, including histone modifications, chromatin accessibility, and other relevant genomic features.

How to Use

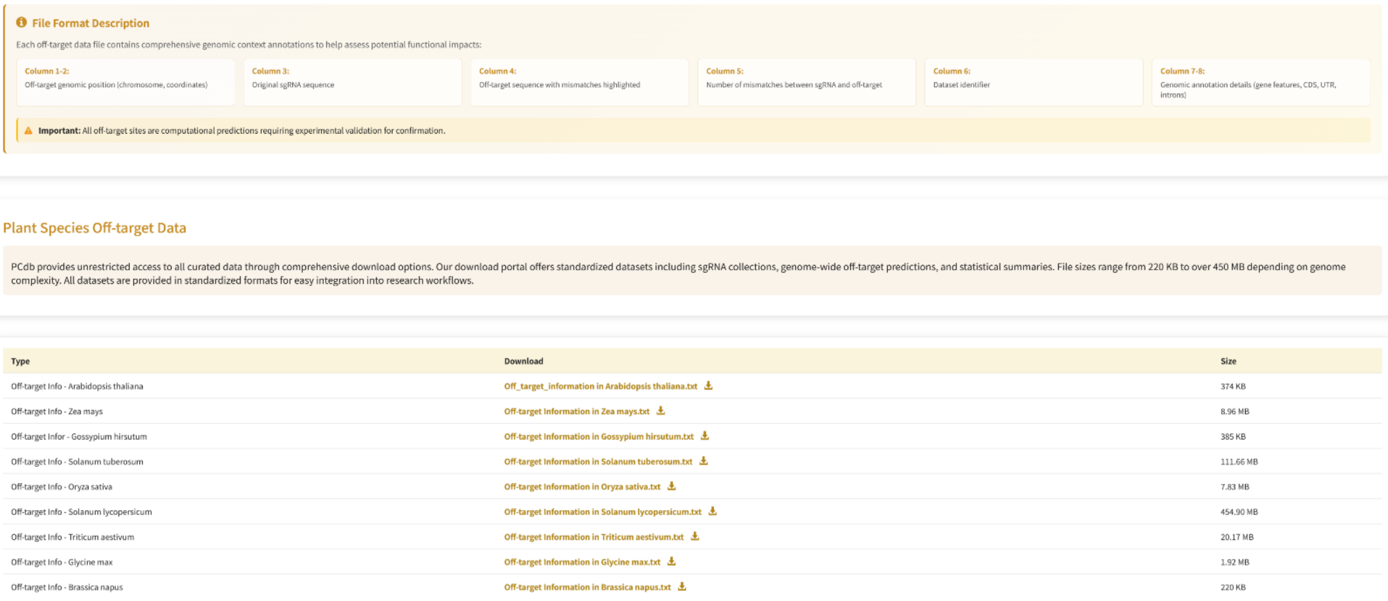
- Select a Genome: Choose a plant species from the dropdown menu, such as Arabidopsis thaliana or Soybean.
- View Tracks: Once a genome is loaded, the browser will display various data tracks specific to that species. These tracks represent different types of chromatin data, such as:
- Explore: You can zoom in and out, pan left and right, and enter a genomic region to examine the data in detail. The tracks provide a visual overview of how chromatin is structured and regulated across the genome.

Select genome: Arabidopsis thaliana

chr3:11,638,276-11,821,555

8. Download data

The Download section provides standardized PCdb datasets, including sgRNA collections, genome-wide off-target predictions, annotated genomic contexts, and summary reports. File sizes range from small single-species datasets to large genome-wide prediction files. All files include detailed metadata for reproducibility and proper attribution. Users can also export search results in multiple formats for further analysis.



File Format Description

Each off-target data file contains comprehensive genomic context annotations to help assess potential functional impacts:

Column 1-2:	Column 3:	Column 4:	Column 5:	Column 6:	Column 7-8:
Off-target genomic position (chromosome, coordinates)	Original sgRNA sequence	Off-target sequence with mismatches highlighted	Number of mismatches between sgRNA and off-target	Dataset identifier	Genomic annotation details (gene features, CDS, UTR, introns)

Important: All off-target sites are computational predictions requiring experimental validation for confirmation.

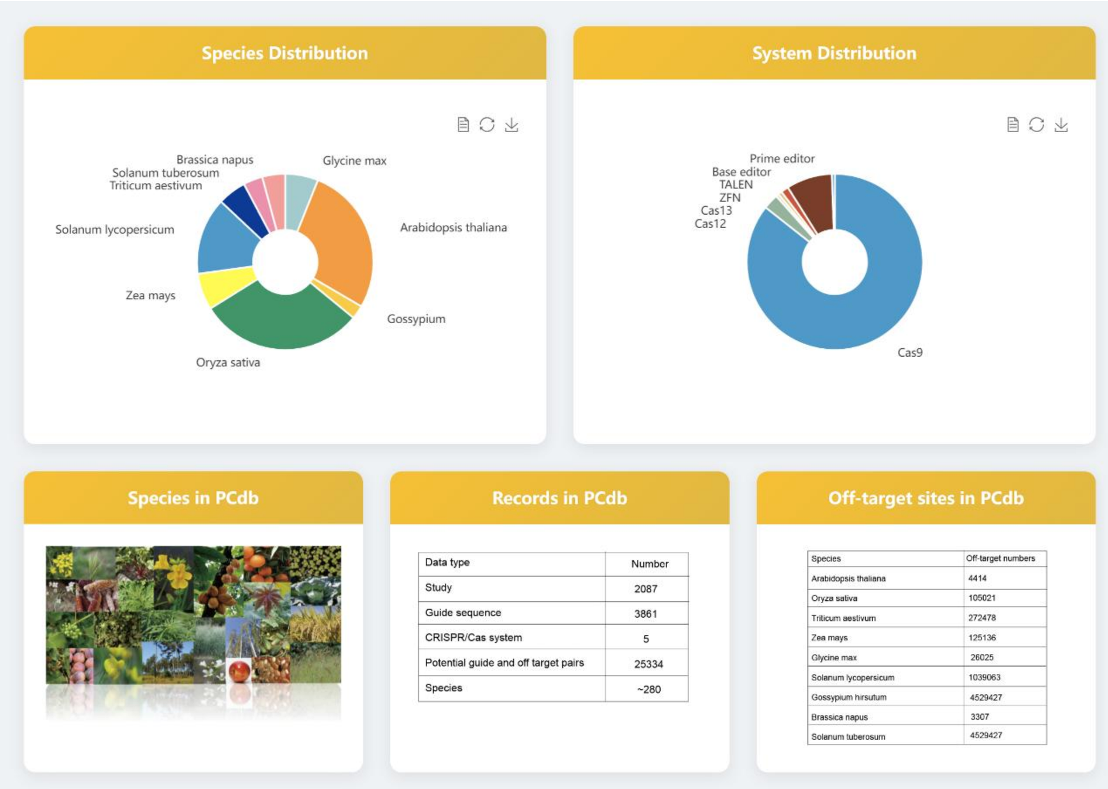
Plant Species Off-target Data

PCdb provides unrestricted access to all curated data through comprehensive download options. Our download portal offers standardized datasets including sgRNA collections, genome-wide off-target predictions, and statistical summaries. File sizes range from 220 KB to over 450 MB depending on genome complexity. All datasets are provided in standardized formats for easy integration into research workflows.

Type	Download	Size
Off-target Info - Arabidopsis thaliana	Off_target_information in Arabidopsis thaliana.txt	374 KB
Off-target Info - Zea mays	Off-target Information in Zea mays.txt	8.96 MB
Off-target Infor - Gossypium hirsutum	Off-target Information in Gossypium hirsutum.txt	385 KB
Off-target Info - Solanum tuberosum	Off-target Information in Solanum tuberosum.txt	111.66 MB
Off-target Info - Oryza sativa	Off-target Information in Oryza sativa.txt	7.83 MB
Off-target Info - Solanum lycopersicum	Off-target Information in Solanum lycopersicum.txt	454.90 MB
Off-target Info - Triticum aestivum	Off-target Information in Triticum aestivum.txt	20.17 MB
Off-target Info - Glycine max	Off-target Information in Glycine max.txt	1.92 MB
Off-target Info - Brassica napus	Off-target Information in Brassica napus.txt	220 KB

9. Statistics overview

The statistical page provides a comprehensive overview of its breadth and diversity. The Species Distribution chart illustrates broad taxonomic coverage, with data spanning key model and crop species such as *Arabidopsis thaliana, Oryza sativa, Zea mays, and Solanum lycopersicum*. The System Distribution highlights the dominant use of Cas9, while also including emerging genome editing platforms such as CRISPR-Cas12, Base Editor, and Prime Editor.

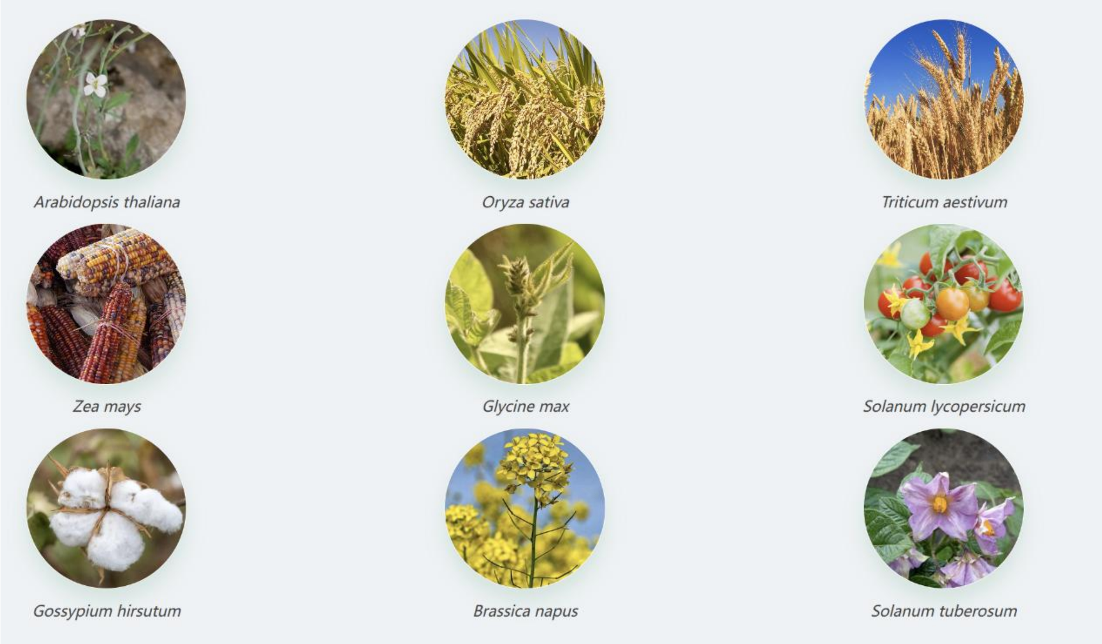


Records in PCdb

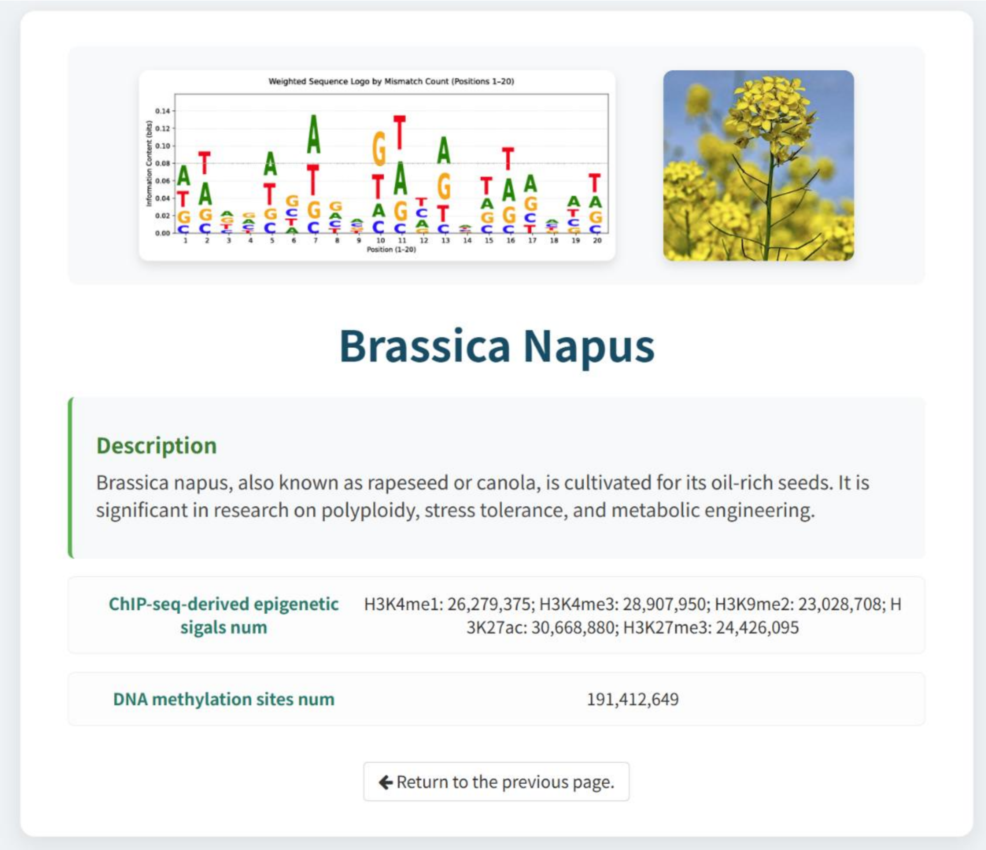
Data type	Number
Study	2087
Guide sequence	3861
CRISPR/Cas system	5
Potential guide and off target pairs	25334
Species	~280

Off-target sites in PCdb

Species	Off-target numbers
Arabidopsis thaliana	4414
Oryza sativa	105021
Triticum aestivum	272478
Zea mays	125136
Glycine max	26025
Solanum lycopersicum	1039063
Gossypium hirsutum	4529427
Brassica napus	3307
Solanum tuberosum	4529427



Clicking on a species image will display a summary of the data available for that species.



Brassica Napus

Description

Brassica napus, also known as rapeseed or canola, is cultivated for its oil-rich seeds. It is significant in research on polyploidy, stress tolerance, and metabolic engineering.

ChIP-seq-derived epigenetic sigals num	H3K4me1: 26,279,375; H3K4me3: 28,907,950; H3K9me2: 23,028,708; H3K27ac: 30,668,880; H3K27me3: 24,426,095
DNA methylation sites num	191,412,649

10. Platform implementation

The PCdb web application is deployed on a Linux-based Apache Web Server 2.4.62. The backend database utilizes MySQL 5.7.43 for efficient data storage and retrieval. Server-side scripting is implemented in PHP 8.3.12, while the frontend interface employs Bootstrap v3.3.7 and jQuery v2.1.1 for responsive design. Interactive visualizations are powered by ECharts framework. The database is fully compatible with major web browsers including Mozilla Firefox, Google Chrome, and Internet Explorer, and is freely accessible at https://gmo.sjtu.edu.cn/pcdb without registration requirements.