

Allele Mining with the Soybean Allele Catalog Tool

Kristin Bilyeu

Plant Genetics Research Unit

USDA/ARS

University of Missouri

Columbia, MO 65211 USA

kristin.bilyeu@usda.gov

February 14, 2024



Soybean Allele Catalog Tool

Based on whole genome sequence data for large accession panels

soykb.org

Or the Soyhub:

<https://soykb.org/soyhub.php/>

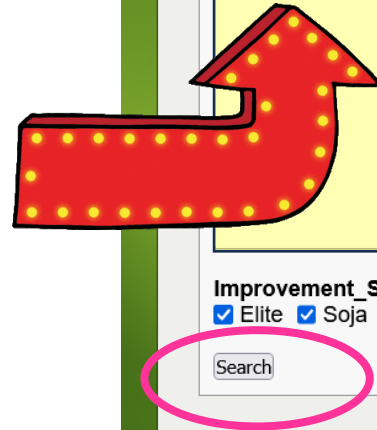
The screenshot shows the Soybean Knowledge Base (SoyKB) website. At the top left is the soykb logo. The top right has a login section with fields for Username and Password, and a Go button. Below the login is a navigation bar with links: Home, Search, Browse, Tools, Soy Hub, Data Files, Analytics, Information, and About. The main header features a green background with a soybean leaf and the text "SOYBEAN KNOWLEDGE BASE (SoyKB) A web resource for Soybean Translational Genomics". The main content area is titled "Soybean Allele Catalog Tool" and includes two search panels. The left panel, "Search by Gene IDs", has a Dataset dropdown set to "Soy1066 Allele Catalog" and a text input field containing "Glyma.09G235100". Below this are checkboxes for Improvement Status: Elite, Soja, Landrace, and Cultivar, all of which are checked. A Search button is at the bottom. The right panel, "Search by Accessions and Gene ID", has a Dataset dropdown set to "Soy2939 Allele Catalog" and a text input field for Accessions. Below this is a Gene ID input field. A Search button is at the bottom. At the bottom of the page, there are two buttons: "Download Accession Information" and "View Demo", with the latter circled in pink. A footer section contains a citation: "If you use the Soybean Allele Catalog Tool in your work, please cite: Chan YO, Dietz N, Zeng S, Wang J, Flint-Garcia S, Salazar-Vidal MN, Škrabišová M, Bilyeu K, Joshi T: The Allele Catalog Tool: a web-based interactive tool for allele discovery and analysis. BMC Genomics 2023, 24(1):107."

Soybean Allele Catalog Tool

Based on whole genome sequence data for large accession panels

soykb.org

Or the Soyhub:
<https://soykb.org/soyhub.php/>



The screenshot shows the Soybean Knowledge Base (SoyKB) website. At the top, there is a navigation bar with links for Home, Search, Browse, Tools, Soy Hub, Data Files, Analytics, Information, and About. Below the navigation bar is a header section with the SoyKB logo and the text "SOYBEAN KNOWLEDGE BASE (SoyKB) A web resource for Soybean Translational Genomics".

The main content area features the "Soybean Allele Catalog Tool" interface. It includes a search bar with a "Quick Search" button and a "Gene Card" dropdown menu. The search results are displayed in two columns:

- Search by Gene IDs:** The "Dataset" is set to "Soy1066 Allele Catalog". The "Gene IDs" field contains "Glyma.09G235100". Below this field is a "Search" button, which is circled in pink.
- Search by Accessions and Gene ID:** The "Dataset" is set to "Soy2939 Allele Catalog". The "Accessions" field contains "HN005_PI404166 HN006_PI407788A". Below this field is a "Search" button.

At the bottom of the page, there is a "Download Accession Information" button and a "View Demo" button. A footer section contains a citation: "If you use the Soybean Allele Catalog Tool in your work, please cite: Chan YO, Dietz N, Zeng S, Wang J, Flint-Garcia S, Salazar-Vidal MN, Škrabišová M, Bilyeu K, Joshi T: The Allele Catalog Tool: a web-based interactive tool for allele discovery and analysis. BMC Genomics 2023, 24(1):107."

Results for *R* gene controlling black or brown seed coat/hilum pigments

	Soja	Landrace	Elite	Total	Cultivar	Gene	Chromosome	45758816	45758833	45758856	45759100	45759137	45759165	45760553	45760555	
<input type="checkbox"/>	107	258	199	626	93	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>
<input type="checkbox"/>	0	99	94	239	31	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	T frameshift_variant R75fs	G Ref	A Ref	G Ref	<input type="checkbox"/>
<input type="checkbox"/>	0	44	46	90	5	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	T intron_variant&splice_donor_variant	A Ref	G Ref	<input type="checkbox"/>
<input type="checkbox"/>	0	37	18	55	9	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	C W32S	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>
<input type="checkbox"/>	0	26	24	50	2	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	C frameshift_variant G63fs	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>
<input type="checkbox"/>	2	1	0	3	0	Glyma.09G235100	Chr09	A L19I	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>
<input type="checkbox"/>	1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	C E166Q	<input type="checkbox"/>
<input type="checkbox"/>	1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	C E165A	G Ref	<input type="checkbox"/>
<input type="checkbox"/>	0	1	0	1	0	Glyma.09G235100	Chr09	C Ref	G frameshift_variant Q25fs	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>

[Download \(Accession Counts\)](#)

[Download \(All Accessions\)](#)

[Download Accession Information](#)

[Download All \(Accession Counts\)](#)

[Download All \(All Accessions\)](#)

Reference

missense

frameshift

splice

Each row of the results is an allele with gene and variant position details

Gene and variant position information

	Soja	Landrace	Elite	Total	Cultivar	Gene	Chromosome	45758816	45758833	45758856	45759100	45759137	45759165	45760553	45760555
<input type="checkbox"/>	107	258	199	626	93	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref
<input type="checkbox"/>	0	99	94	239	31	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	T frameshift_variant R75fs	G Ref	A Ref	G Ref
<input type="checkbox"/>	0	44	46	90	5	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	T intron_variant&splice_donor_variant	A Ref	G Ref
<input type="checkbox"/>	0	37	18	55	9	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	C W32S	CG Ref	TC Ref	G Ref	A Ref	G Ref
<input type="checkbox"/>	0	26	24	50	2	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	C frameshift_variant G63fs	TC Ref	G Ref	A Ref	G Ref
<input type="checkbox"/>	2	1	0	3	0	Glyma.09G235100	Chr09	A L19I	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref
<input type="checkbox"/>	1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	C E166Q
<input type="checkbox"/>	1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	C E165A	G Ref
<input type="checkbox"/>	0	1	0	1	0	Glyma.09G235100	Chr09	C Ref	G frameshift_variant Q25fs	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref

[Download \(Accession Counts\)](#) [Download \(All Accessions\)](#)

[Download Accession Information](#)

[Download All \(Accession Counts\)](#)

[Download All \(All Accessions\)](#)

Reference

missense

frameshift

splice

A frequency table has improvement status with most frequent (Total) on top

Frequency table

Soja	Landrace	Elite	Total	Cultivar	Gene	Chromosome	45758816	45758833	45758856	45759100	45759137	45759165	45760553	45760555	
107	258	199	626	93	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>
0	99	94	239	31	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	T frameshift_variant R75fs	G Ref	A Ref	G Ref	<input type="checkbox"/>
0	44	46	90	5	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	T intron_variant&splice_donor_variant	A Ref	G Ref	<input type="checkbox"/>
0	37	18	55	9	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	C W32S	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>
0	26	24	50	2	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	C frameshift_variant G63fs	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>
2	1	0	3	0	Glyma.09G235100	Chr09	A L19I	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>
1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	C E166Q	<input type="checkbox"/>
1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	C E165A	G Ref	<input type="checkbox"/>
0	1	0	1	0	Glyma.09G235100	Chr09	C Ref	G frameshift_variant Q25fs	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>

[Download \(Accession Counts\)](#) [Download \(All Accessions\)](#)

[Download Accession Information](#) [Download All \(Accession Counts\)](#) [Download All \(All Accessions\)](#)

Reference

missense

frameshift

splice

Mutations in the *R* gene cause normally black pigments in the seed coat/hilum to be brown

	Soja	Landrace	Elite	Total	Cultivar	Gene	Chromosome	45758816	45758833	45758856	45759100	45759137	45759165	45760553	45760555	
<input type="checkbox"/>	107	258	199	626	93	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	R
<input type="checkbox"/>	0	99	94	239	31	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	T frameshift_variant R75fs	G Ref	A Ref	G Ref	r
<input type="checkbox"/>	0	44	46	90	5	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	T intron_variant&splice_donor_variant	A Ref	G Ref	r
<input type="checkbox"/>	0	37	18	55	9	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	C W32S	CG Ref	TC Ref	G Ref	A Ref	G Ref	r
<input type="checkbox"/>	0	26	24	50	2	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	C frameshift_variant G63fs	TC Ref	G Ref	A Ref	G Ref	r
<input type="checkbox"/>	2	1	0	3	0	Glyma.09G235100	Chr09	A L19I	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>
<input type="checkbox"/>	1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	C E166Q	<input type="checkbox"/>
<input type="checkbox"/>	1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTTTA Ref	G Ref	CG Ref	TC Ref	G Ref	C E165A	G Ref	<input type="checkbox"/>
<input type="checkbox"/>	0	1	0	1	0	Glyma.09G235100	Chr09	C Ref	G frameshift_variant Q25fs	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>

Download (Accession Counts) Download (All Accessions)

Download Accession Information Download All (Accession Counts) Download All (All Accessions)

Reference

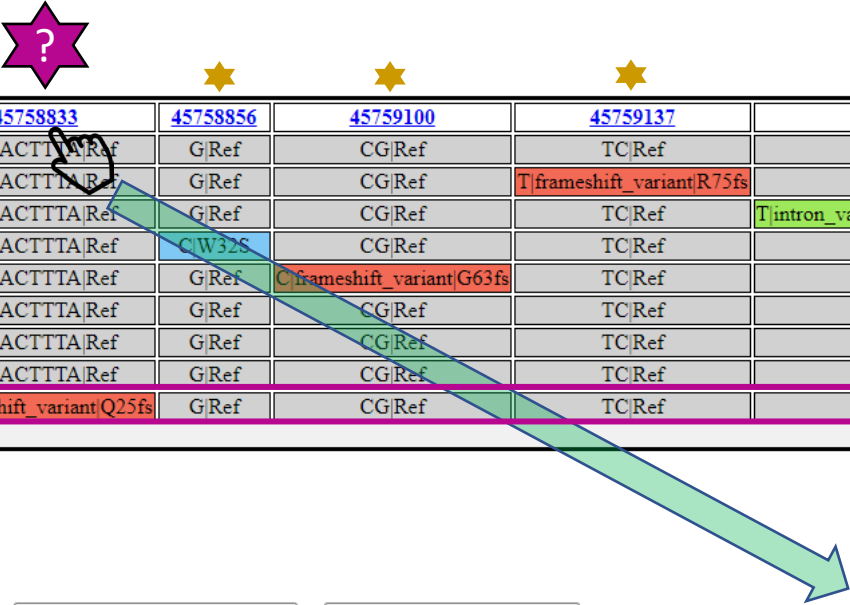
missense

frameshift

splice

Gillman, J., Tetlow, A., Lee, J.-D., Shannon, J.G., and Bilyeu, K., *Loss-of-function mutations affecting a specific Glycine max R2R3 MYB transcription factor result in brown hilum and brown seed coats. BMC Plant Biology, 2011. 11(1): p. 155.*

Allele mining-is 09:4578833 variant a new *r* allele?



Soja	Landrace	Elite	Total	Cultivar	Gene	Chromosome	45758816	45758833	45758856	45759100	45759137	45759165	45760553	45760555	
107	258	199	626	93	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	R
0	99	94	239	31	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	CG Ref	T frameshift_variant R75fs	G Ref	A Ref	G Ref	r
0	44	46	90	5	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	CG Ref	TC Ref	T intron_variant&splice_donor_variant	A Ref	G Ref	r
0	37	18	55	9	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	C W32S	CG Ref	TC Ref	G Ref	A Ref	G Ref	r
0	26	24	50	2	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	C frameshift_variant G63fs	TC Ref	G Ref	A Ref	G Ref	r
2	1	0	3	0	Glyma.09G235100	Chr09	A L19I	GCAACTT A Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>
1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	C E166Q	<input type="checkbox"/>
1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	CG Ref	TC Ref	G Ref	C E165A	G Ref	<input type="checkbox"/>
0	1	0	1	0	Glyma.09G235100	Chr09	C Ref	G frameshift_variant Q25fs	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>

[Download \(Accession Counts\)](#) [Download \(All Accessions\)](#)

[Download Accession Information](#) [Download All \(Accession Counts\)](#) [Download All \(All Accessions\)](#)

Reference

missense

frameshift

splice

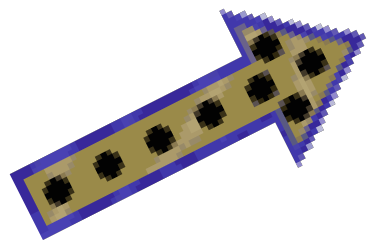
A new *r* frameshift allele for brown color



Soja	Landrace	Elite	Total	Cultivar	Gene	Chromosome	45758816	45758833	45758856	45759100	45759137	45759165	45760553	45760555	
107	258	199	626	93	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	R
0	99	94	239	31	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	CG Ref	T frameshift_variant R75fs	G Ref	A Ref	G Ref	r
0	44	46	90	5	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	CG Ref	TC Ref	T intron_variant&splice_donor_variant	A Ref	G Ref	r
0	37	18	55	9	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	C W32S	CG Ref	TC Ref	G Ref	A Ref	G Ref	r
0	26	24	50	2	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	C frameshift_variant G63fs	TC Ref	G Ref	A Ref	G Ref	r
2	1	0	3	0	Glyma.09G235100	Chr09	A L19I	GCAACTT A Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	<input type="checkbox"/>
1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	CG Ref	TC Ref	G Ref	A Ref	C E166Q	<input type="checkbox"/>
1	0	0	1	0	Glyma.09G235100	Chr09	C Ref	GCAACTT A Ref	G Ref	CG Ref	TC Ref	G Ref	C E165A	G Ref	<input type="checkbox"/>
0	1	0	1	0	Glyma.09G235100	Chr09	C Ref	G frameshift_variant Q25fs	G Ref	CG Ref	TC Ref	G Ref	A Ref	G Ref	r

[Download \(Accession Counts\)](#) [Download \(All Accessions\)](#)

Chromosome	Position	Accession	SoyKB_Accession	GRIN_Accession	Improvement_Status	Classification	Genotype	Functional_Effect	Imputation	HILUMCOLOR	SCOATCOLOR
Chr09	45758833	SRR1533319	PI_567258	PI567258	Landrace	Other	G	frameshift_variant		Br	Br



Annual updates to allele catalog data panel

kristin.bilyeu@usda.gov

soykb.org

- Contact me about adding your data to the Soybean Allele Catalog Tool (fall deadline?)
- Private data? Pipelines available on Github (<https://github.com/yenon18>)
- Send feedback about tool or let us know if you want to collaborate!

Welcome to Soy Hub
A hub for soybean-applied genomics predictions based on a curated panel of diverse soybean resequenced accessions ([Soy1066](#)).

Explore variation:

- Allele Catalog**
 - Find accessions with certain allele
 - Find new alleles in known genes
- GenVarX**
 - Explore variation in promoters
 - Search TFs
 - Explore CNV

Predict new causal mutations:

- AccuTool**
 - Use GWAS results for prediction
 - Calculate Accuracy for your markers or candidate causative mutations (CM) based on Soy775 35.7M variant positions
- SNPViz**
 - Check genomic context of your variant positions in empowered haplotype viewer on various resequenced data sets
- Reference Interassembly Gene Browser**
 - Search between reference genotypes, genome assemblies, or annotation versions

Meet the team and funders



Dr. Mária Škrabišová
[Palacký University Olomouc](#)



Dr. Nick Dietz (2021)



Yen On Chan*



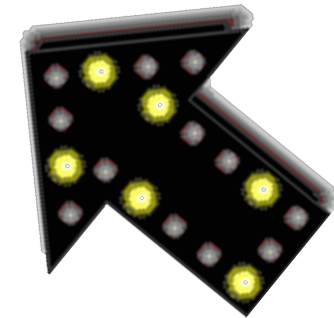
Dr. Trupti Joshi



Anser Mahmood*

Dr. Nathan Grant

Jana Biová*



Allele Mining with the Soybean Allele Catalog Tool

Thanks for your attention!

kristin.bilyeu@usda.gov

February 14, 2024

