



Jacqueline D. Campbell¹, Steven Cannon¹, Wei Huang¹, Nathan Weeks¹, Rex Nelson¹
¹ USDA-ARS-CICGRU, Ames, Iowa

SoyBase (<https://soybase.org>) is the USDA-ARS soybean resource hub for the soybean community, by providing a comprehensive collection of data, analysis tools, and links to external resources of interest to soybean researchers. A major goal of SoyBase is to assist researchers in discovering important trait, genomic, and genetic information within the vast amount of data available.

Choosing sequences from more than one genome assembly from the same species could lead to misleading sorting of hits in the display. Please check coverage and identity statistics carefully.

Nucleotide databases

- Cultivar_Fiskeby_III_(FiskebyIII)_Nucleotide_Sequence_Collection
 - Fiskeby III Assembly 1 Annotation 1 (FiskebyIII.a1.v1) Coding Sequence
 - Fiskeby III Assembly 1 Genomic Sequence (FiskebyIII.a1)
- Cultivar_Lee_Nucleotide_Sequence_Collection
 - Lee Assembly 1 Annotation 1 (Lee.a1.v1) Coding Sequences
 - Lee Assembly 1 Annotation 1 (Lee.a1.v1) Genomic Sequence
- Cultivar_Williams_82_(Wm82)_Nucleotide_Sequence_Collection
 - Williams 82 Assembly 1 Annotation 1.1 (Wm82.a1.v1.1) Coding Sequences
 - Williams 82 Assembly 1 Annotation 1.1 (Wm82.a1.v1.1) Genomic Sequence
 - Williams 82 Assembly 2 Annotation 1 (Wm82.a2.v1) Protein Sequences
 - Williams 82 Assembly 2 Annotation 1 (Wm82.a2.v1) Coding Sequences
 - Williams 82 Assembly 2 Genomic Sequence (Wm82.a2)
 - Williams 82 Assembly 4 Annotation 1 (Wm82.a4.v1) Coding Sequences
 - Williams 82 Assembly 4 Genomic Sequence (Wm82.a4)
- Cultivar_Williams_82_ISU01_(Wm82_ISU01)_Nucleotide_Sequence_Collection
 - Williams 82 ISU01 Assembly 2 Annotation 1 (Wm82_ISU.a2.v1) Coding Sequence
 - Williams 82 ISU01 Assembly 2 Genomic Sequence (Wm82_ISU01.a2)
- Cultivar_Zhonghuang_13_(ZH13)_Nucleotide_Sequence_Collections
 - ZH13 Assembly 1 Annotation 1 (ZH13.a1.v1) Genomic Sequence
- Glycine_max_Organelle_Nucleotide_Sequences
 - Chloroplast Genome
 - Mitochondrial genome
- Glycine_soja_Cultivar_W05_Nucleotide_Sequence_Collections
 - W05 Assembly 1 Annotation 1 (W05.a1.v1) Coding Sequences
 - W05 Assembly 1 Annotation 1 (W05.a1.v1) Genomic Sequence
- Glycine_soja_PI483463_Nucleotide_Sequence_Collection
 - PI483463 Assembly1 Annotation 1 (PI483463.a1.v1) Coding Sequences
 - PI483463 Assembly1 Annotation 1 (PI483463.a1.v1) Genomic Sequence
- Miscellaneous_Databases
 - PanSoy dataset: genomic sequence from 204 accessions but not in Wm82.a4

Protein databases

- Fiskeby III Assembly 1 Annotation 1 (FiskebyIII.a1.v1) Protein Sequence
- Glycine soja PI483463 Assembly 1 Annotation 1 (PI483463.a1.v1) Protein Sequence
- Glycine soja W05 Assembly 1 Annotation 1 (W05.a1.v1) Protein Sequence
- Lee Assembly 1 Annotation 1 (Lee.a1.v1) Protein Sequences
- Williams 82 Assembly 1 Annotation 1.1 (Wm82.a1.v1.1) Protein Sequences
- Williams 82 Assembly 2 Annotation 1 (Wm82.a2.v1) Protein Sequences
- Williams 82 Assembly 4 Annotation 1 (Wm82.a4.v1) Protein Sequences
- Williams 82 ISU01 Assembly 2 Annotation 1 (Wm82_ISU01.a2.v1) Protein Sequences
- ZH13 Assembly 1 Annotation 1 (ZH13.a1.v1) Protein Sequence

Download FASTA, XML, TSV

FASTA of all hits

FASTA of selected hit(s)

Alignment of all hits

Alignment of selected hit(s)

Standard tabular report

Full tabular report

Full XML report

Share results

Copy URL to clipboard

Send by email

Query= glyma.PI_548362.gnm1.ann1.SoyC10_01G000550.m1 query 2, length: 2,847

Graphical overview of hits

Length distribution of matching sequences

Sequences producing significant alignments

#	Similar sequences	Species	Query coverage (%)	Total score	E value	Identity (%)
1.	Glyma.08G360000.10 dsid=glyma.Wm82.gnm4.ann1.G...	Glycine max	99	5200	0	96.6%
2.	Glyma.08G360000.5 dsid=glyma.Wm82.gnm4.ann1.GL...	Glycine max	99	5200	0	96.6%
3.	Glyma.08G360000.3 dsid=glyma.Wm82.gnm4.ann1.GL...	Glycine max	99	5200	0	96.6%
4.	Glyma.08G360000.7 dsid=glyma.Wm82.gnm4.ann1.GL...	Glycine max	89	4615	0	96.6%
5.	Glyma.01G000900.2 dsid=glyma.Wm82.gnm4.ann1.GL...	Glycine max	100	5700	0	100%

hit 1 of query 2, length: 2,871

polyptide=Glyma.08G360000.10.p locus=Glyma.08G360000.ID=Glyma.08G360000.10.Wm82.a4.v1 annot-version=Wm82.a4.v1

Select | Sequence | FASTA | Alignment | LIS region linkouts

Graphical overview of aligning region(s)

SequenceServer BLAST searches

Wide variety of nucleotide and protein databases:

- Multiple soybean cultivars
- PanSoy dataset

Using SequenceServer for BLAST allows more output options:

- Downloading a full tabular or XML report
- Sending results to collaborators by email or sharing the URL
- Download alignment images in multiple formats

Pan Gene Search

Gene Name:

Example: Glyma.01G000100

Search

Assembly	Gene
Glyma.Lee.a1.v1	GlymaLee.02G000100
Glyma.Wm82.a1.v1	Glyma0002s50
Glyma.Wm82.a2.v1	Glyma.02G000100
Glyma.Wm82.a4.v1	Glyma.02G000100
Glyso.PI483463.a1.v1	GlysoPI483463.02G000100
Glyso.W05.a1.v1	Glysoja.02G002531
Glyma.Wm82.a2.v2	LOC100814813

Your gene Glyma.02G000100 belongs to panset_id: glysp.mixed.pan2.SoyPan036275

Pan-Gen Search Tool

- Returns the pan-gene set that query gene is from and other genomes
- Assembly and annotation version of the gene models will be returned
- If the query gene is unique to that genome; no pan-genes will be displayed

Data Source

Glycine max Lee genome assembly version 1 (glyma.Lee.gnm1)

Scroll/Zoom: Show 59 kbp Flip

Overview Region Details

20 kbp

28570k 28580k 28590k 28600k 28610k 28620k

GlymaLee.15G180800.1 GlymaLee.15G180900.1 GlymaLee.15G181000.1 GlymaLee.15G181100.1

PAV 0.7099 PAV 0.5270 PAV 0.3991 PAV 0.7369

Gene models - Glyma.Lee.gnm1

F-box/LRR-repeat protein At3g4... (click for more)

F-box/LRR-repeat protein At3g4... (click for more)

F-box/LRR-repeat protein At3g4... (click for more)

Pan-Genome Presence / Absence

Pan-genome representing over 1,000 soybean accessions from the USDA Soybean Germplasm Collection using Lee reference genome.

- Genome viewer representation of gene presence/absence
- Variant counts from soybean accessions in the study
- Download GFF3 and fasta file of genes not in Lee

SoybeanBase

- Partner of SoyBase and managed by Breeding Insights
- Integrate trait analysis, genotypic data and software tools to help researchers and breeders
- Assist in data management and data analysis
- Document crosses, field trails, genotyping protocols and more

SoybeanBase Search Manage Analyze About

Search

Wizard Accessions and plots Organisms Progenies, Parents, Crosses Field Trials Genotyping Plates Genotyping Data Projects Genotyping Protocols Traits Images People

Genotyping Data Project Search

Show 10 entries Search:

Genotyping Data Project Name	Description	Breeding program	Folder	Year	Location	Genotyping Facility
NAMParentsBARCSoySNP6K	NAM-NUST Common Accessions genotyped with BARCSoySNP6K BeadChip	NUST-Soybean (NUSSOY)		2022		None
NAMParentsBARCSoySNP6K001	NUST-NAM Common Accessions genotyped with BARCSoySNP6K BeadChip Illumina Infinium Assay	NUST-Soybean (NUSSOY)		2022	Lamberton-MN	None
NUST Soy	NUST Soy 2021	Breedbase		2022		None
NUSTBARCSoySNP6K001	NUST Population Subset 1 genotyped with BARCSoySNP6K BeadChip Illumina Infinium Assay (Song et al., 2020)	NUST-Soybean (NUSSOY)		2022	Lamberton-MN	None
OSU genotyping test	testing sequencing data upload with single sample	OSU-Soybean		2022		None
SoyNAM Genotyping	SoyNAM Genotyping Project using SoySNP6K	SoyNAM		2022		None

Download Data

- Genome, genes, protein sequences from multiple cultivated and wild species
- Pan-genome and Pan-genes sets
- From GWAS, expression and methylation studies

Glycine max Downloads				
Assembly	Cultivar	Genome	CDS	Proteins
Wm82.a1	Williams 82	Download Genome	Download CDS	Download Protein
		Download	Download	Download
Wm82.a2	Williams 82	Download Genome	Download CDS	Download Protein
		Download	Download	Download
Wm82.a4	Williams 82	Download Genome	Download CDS	Download Protein
		Download	Download	Download
Lee.a1	Lee	Download Genome	Download CDS	Download Protein
		Download	Download	Download
ZH13.a1	Zhonghuang 13	Download Genome	Download CDS	Download Protein
		Download	Download	Download

Glycine soja Downloads				
Assembly	Cultivar	Genome	CDS	Proteins
PI483463.a1	PI 483463	Download Genome	Download CDS	Download Protein
		Download	Download	Download
W05.a1	W05	Download Genome	Download CDS	Download Protein
		Download	Download	Download

How to submit your data to SoyBase

It is very helpful to consult with us early so that we can efficiently plan how SoyBase staff effort will be allocated. These early discussions will ensure that optimal decisions can be made about nomenclature, data file formats, etc. so there will be a minimal delay in making your data available to the community.



SoyBase.org is a member of the AgBioData Consortium.

