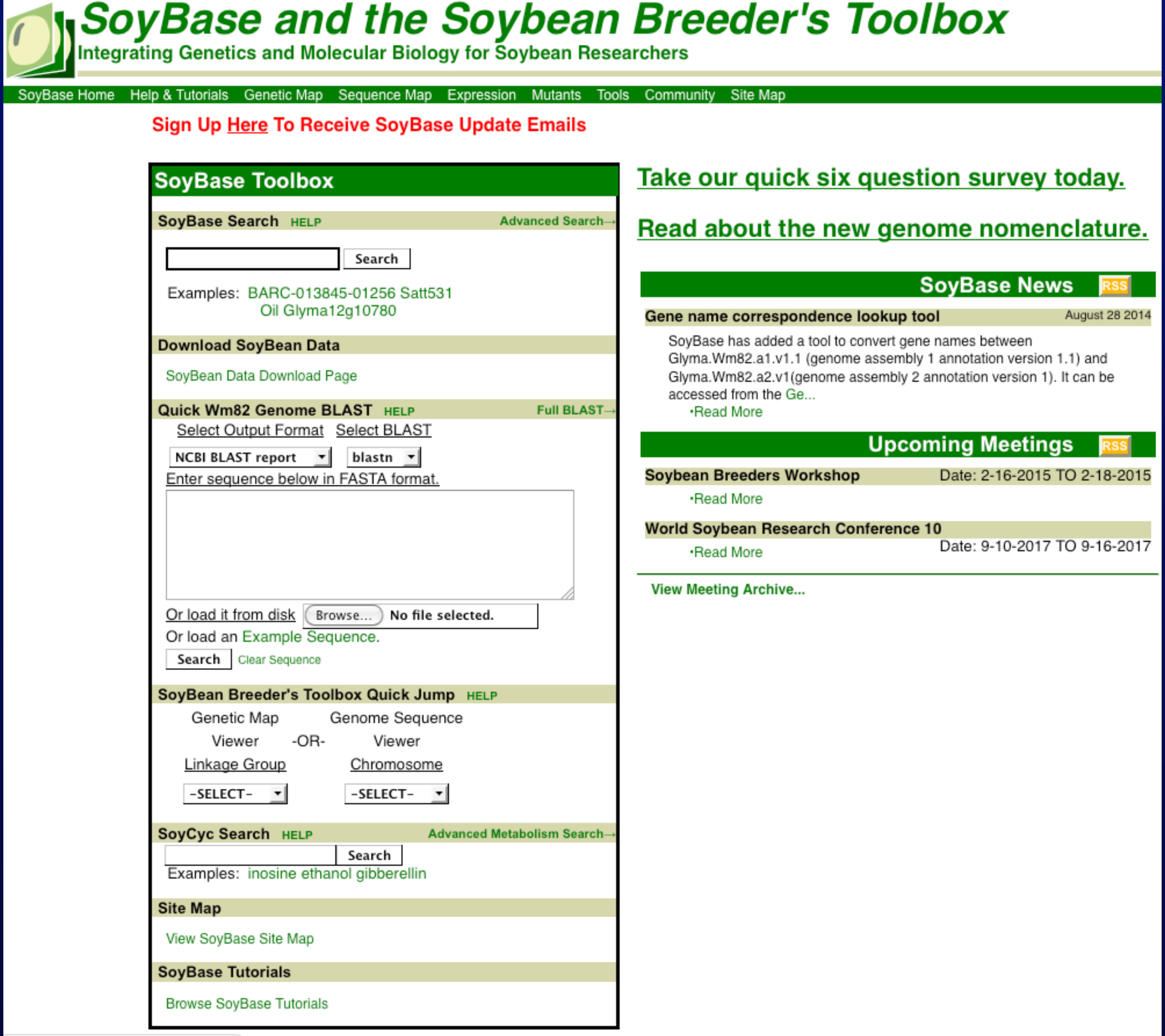
SoyBase, the USDA-ARS Soybean Genetics and Genomics Database

David Grant

Victoria Carollo Blake
Steven B. Cannon
Kevin Feeley
Rex T. Nelson
Nathan Weeks



SoyBase and the Soybean Breeder's Toolbox

Integrating Genetics and Molecular Biology for Soybean Researchers

Sign Up Here To Receive SoyBase Update Emails

SoyBase Toolbox

SoyBase Search HELP Advanced Search→

Examples: BARC-013845-01256 Satt531
Oil Glyma12g10780

Download SoyBean Data

SoyBean Data Download Page

Quick Wm82 Genome BLAST HELP Full BLAST→

Select Output Format: NCBI BLAST report
Select BLAST: blastn
Enter sequence below in FASTA format.

Or load it from disk Browse... No file selected.
Or load an Example Sequence.

SoyBean Breeder's Toolbox Quick Jump HELP

Genetic Map Viewer -OR- Genome Sequence Viewer
Linkage Group: -SELECT-
Chromosome: -SELECT-

SoyCyc Search HELP Advanced Metabolism Search→

Examples: inosine ethanol gibberellin

Site Map

View SoyBase Site Map

SoyBase Tutorials

Browse SoyBase Tutorials

Take our quick six question survey today.

Read about the new genome nomenclature.

SoyBase News

Gene name correspondence lookup tool August 28 2014

SoyBase has added a tool to convert gene names between Glyma.Wm82.a1.v1.1 (genome assembly 1 annotation version 1.1) and Glyma.Wm82.a2.v1(genome assembly 2 annotation version 1). It can be accessed from the Ge...

·Read More

Upcoming Meetings

Soybean Breeders Workshop Date: 2-16-2015 TO 2-18-2015

·Read More

World Soybean Research Conference 10 Date: 9-10-2017 TO 9-16-2017

·Read More

View Meeting Archive...

SoyBase Site Map and Navigation

SoyBase and the Soybean Breeder's Toolbox

Integrating Genetics and Molecular Biology for Soybean Researchers

SoyBase Home   Help & Tutorials   Genetic Map   Sequence Map   Expression   Mutants   Tools   Community   Site Map

SoyBase Site Map

Table of Contents

- SoyBase Home Page
- Genetic and Physical Map Resources
- Sequence Map Resources
- Anaylsis Tools
- Community Resources
- SoySeq Expression Atlas
- Download SoyBean Data

SoyBase Home Page

Genetic and Physical Map Resources

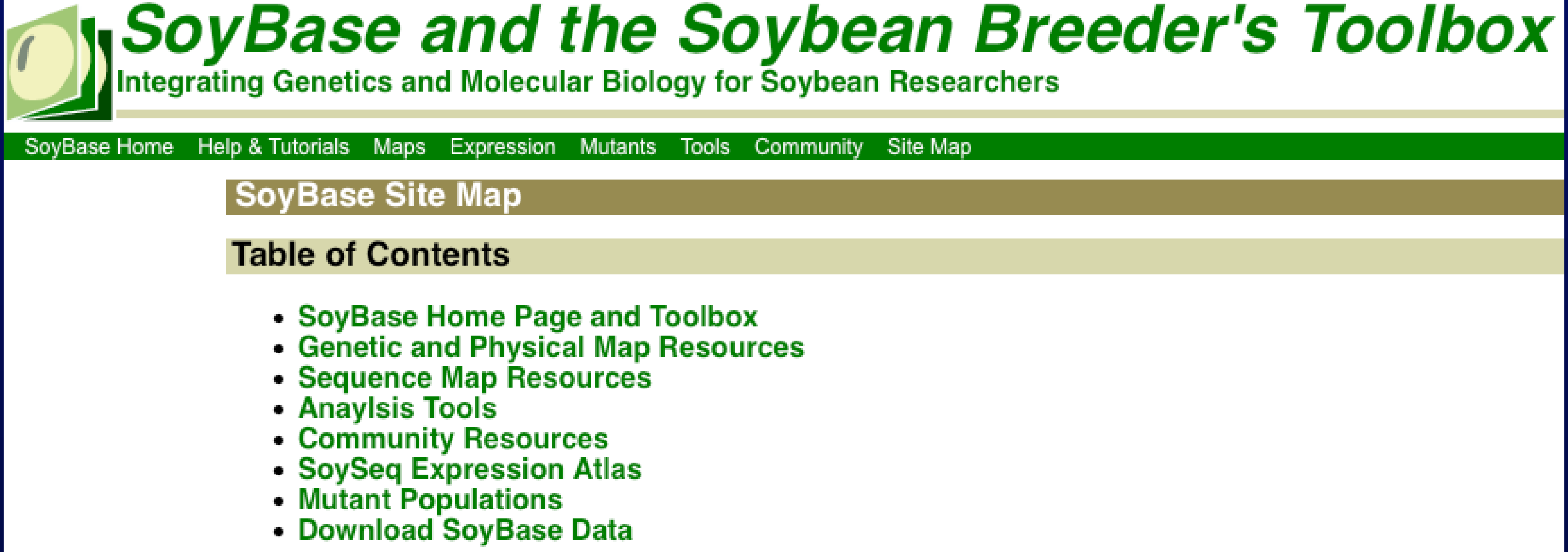
Sequence Map Resources

Anaylsis Tools

Community Resources

SoySeq Expression Atlas

Download SoyBean Data



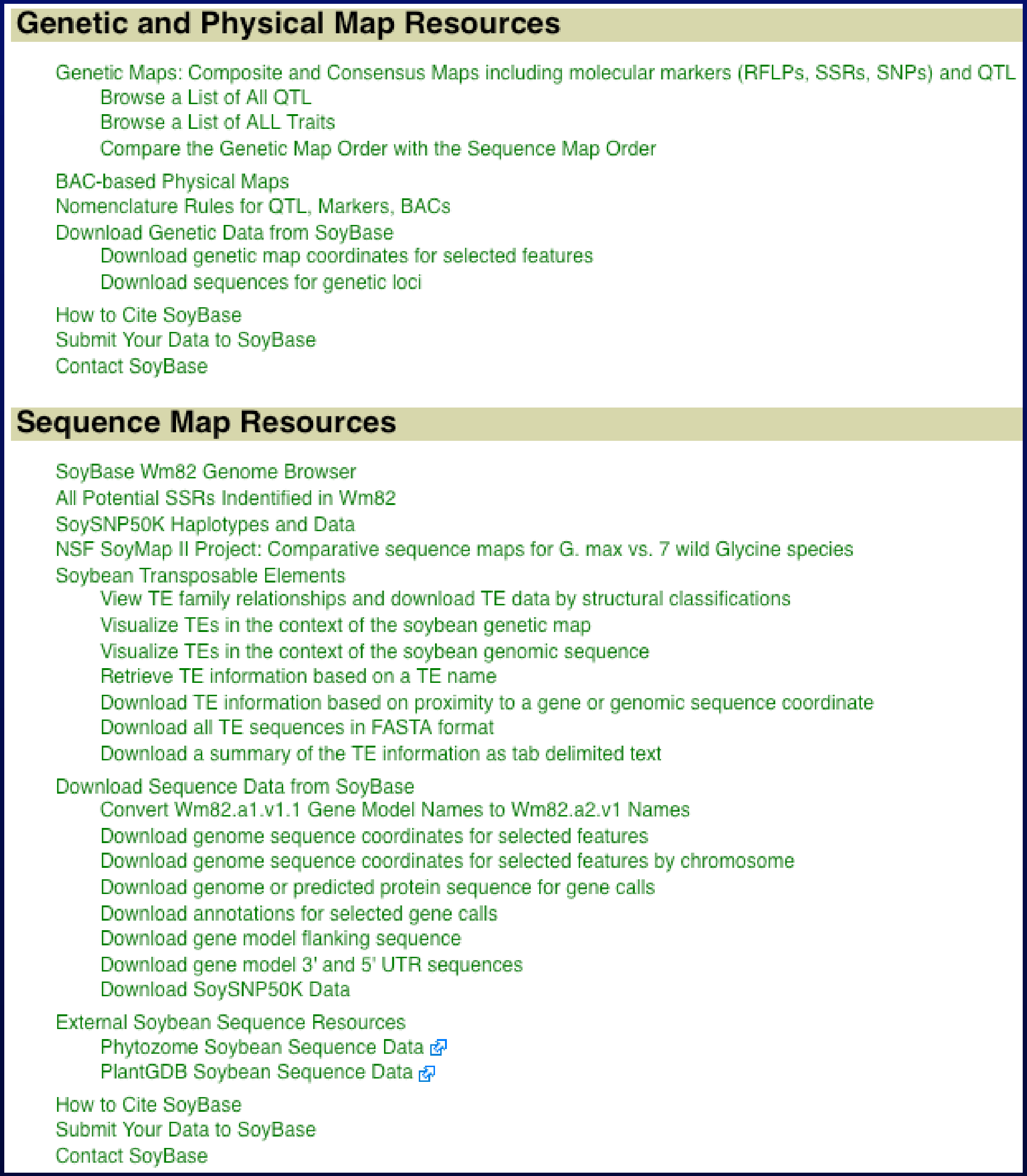
SoyBase and the Soybean Breeder's Toolbox

Integrating Genetics and Molecular Biology for Soybean Researchers

SoyBase Site Map

Table of Contents

- SoyBase Home Page and Toolbox
- Genetic and Physical Map Resources
- Sequence Map Resources
- Anaylsis Tools
- Community Resources
- SoySeq Expression Atlas
- Mutant Populations
- Download SoyBase Data



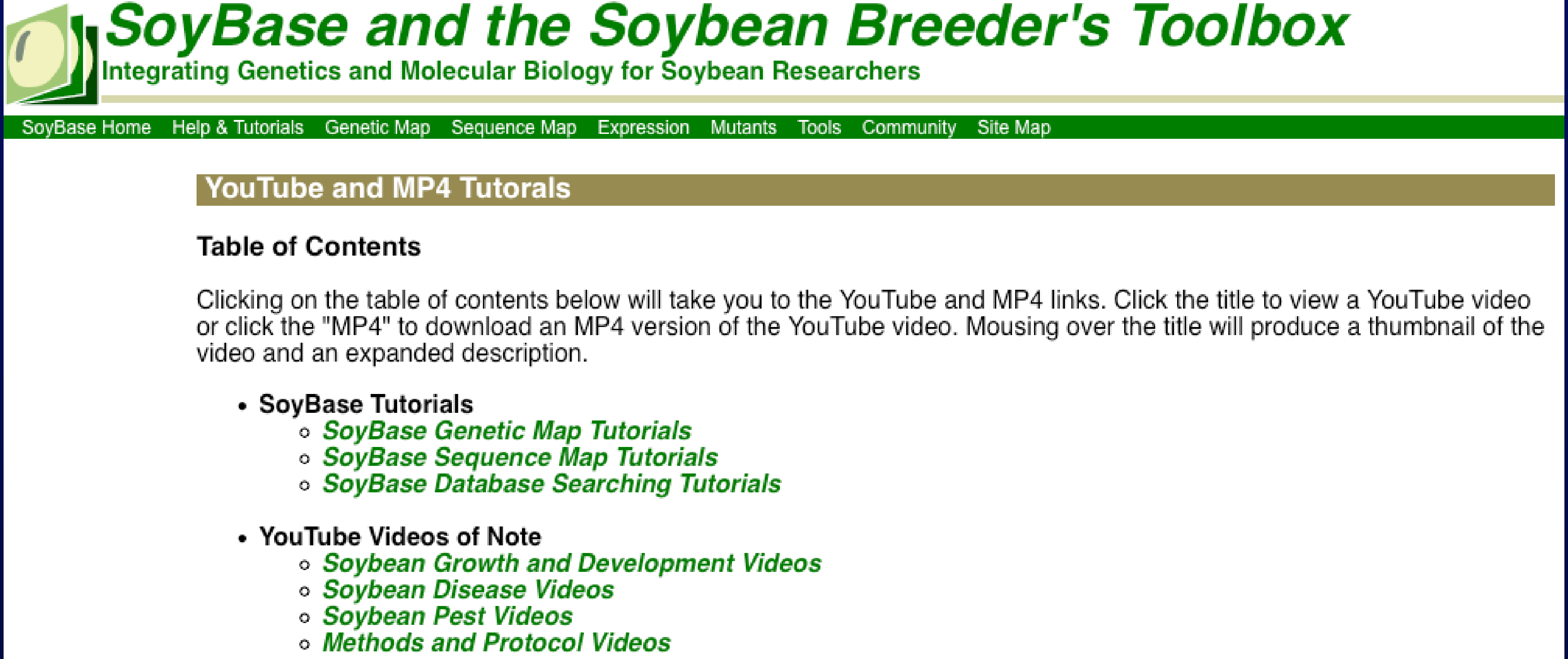
Genetic and Physical Map Resources

- Genetic Maps: Composite and Consensus Maps including molecular markers (RFLPs, SSRs, SNPs) and QTL
    - Browse a List of All QTL
    - Browse a List of ALL Traits
    - Compare the Genetic Map Order with the Sequence Map Order
- BAC-based Physical Maps
- Nomenclature Rules for QTL, Markers, BACs
- Download Genetic Data from SoyBase
    - Download genetic map coordinates for selected features
    - Download sequences for genetic loci
- How to Cite SoyBase
- Submit Your Data to SoyBase
- Contact SoyBase

Sequence Map Resources

- SoyBase Wm82 Genome Browser
- All Potential SSRs Indentified in Wm82
- SoySNP50K Haplotypes and Data
- NSF SoyMap II Project: Comparative sequence maps for G. max vs. 7 wild Glycine species
- Soybean Transposable Elements
    - View TE family relationships and download TE data by structural classifications
    - Visualize TEs in the context of the soybean genetic map
    - Visualize TEs in the context of the soybean genomic sequence
    - Retrieve TE information based on a TE name
    - Download TE information based on proximity to a gene or genomic sequence coordinate
    - Download all TE sequences in FASTA format
    - Download a summary of the TE information as tab delimited text
- Download Sequence Data from SoyBase
    - Convert Wm82.a1.v1.1 Gene Model Names to Wm82.a2.v1 Names
    - Download genome sequence coordinates for selected features
    - Download genome sequence coordinates for selected features by chromosome
    - Download genome or predicted protein sequence for gene calls
    - Download annotations for selected gene calls
    - Download gene model flanking sequence
    - Download gene model 3' and 5' UTR sequences
    - Download SoySNP50K Data
- External Soybean Sequence Resources
    - Phytozome Soybean Sequence Data
    - PlantGDB Soybean Sequence Data
- How to Cite SoyBase
- Submit Your Data to SoyBase
- Contact SoyBase

Video Tutorials: Using SoyBase and Community How-Tos



YouTube and MP4 Tutorals

Table of Contents

Clicking on the table of contents below will take you to the YouTube and MP4 links. Click the title to view a YouTube video or click the "MP4" to download an MP4 version of the YouTube video. Mousing over the title will produce a thumbnail of the video and an expanded description.

- SoyBase Tutorials
  - *SoyBase Genetic Map Tutorials*
  - *SoyBase Sequence Map Tutorials*
  - *SoyBase Database Searching Tutorials*

- YouTube Videos of Note
  - *Soybean Growth and Development Videos*
  - *Soybean Disease Videos*
  - *Soybean Pest Videos*
  - *Methods and Protocol Videos*

SoyBase Genetic Map Tutorials

- How to zoom into a region on the genetic map
  (3:10, 118 Mb MP4)
- How to flip genetic maps to resolve corresponding marker positions
  (1:44, 57 Mb MP4)
- Turning QTL classes off/on
  (3:08, 78 Mb MP4)
- Removing and adding genetic maps
  (4:27, 115 Mb MP4)
- How to quickly bring up a genetic or sequence map at SoyBase
  (1:18, 28 Mb MP4)
- How do I find markers on a genetic map that are not visible
  (2:00, 57 Mb MP4)
- How do I compare the genetic marker order to the sequence marker order
  (3:45, 113 Mb MP4)

SoyBase Sequence Map Tutorials

- How to move tracks in the SoyBase sequence viewer
  (0:56, 31 Mb MP4)
- How to find information on tracks in the SoyBase sequence viewer
  (2:01, 35 Mb MP4)
- Modifying tracks on the SoyBase sequence viewer
  (1:12, 41 Mb MP4)
- Zooming into a region on the SoyBase sequence map
  (2:04, 44 Mb MP4)
- How to visualize the approximate genetic position of a gene call on the genetic map
  (2:19, 62 Mb MP4)

SoyBase Searching Tutorials

- How to do a quick BLAST against the soybean genome sequence
  (1:45, 46 Mb MP4)
- Advanced BLAST analysis at SoyBase
  (4:07, 95 Mb MP4)
- How to perform a quick search of SoyBase
  (3:17, 79 Mb MP4)
- How to perform a SoyCyc quick search
  (2:01, 44 Mb MP4)
- How to quickly get a list of all loci at SoyBase
  (2:10, 80 Mb MP4)
- How do I get a list of all QTL in SoyBase
  (3:08, 79 Mb MP4)
- How to search the Fast Neutron Mutant collection using a trait value
  (1:51, 44 Mb MP4)
- Explanation of a Fast Neutron search report
  (4:07, 120 Mb MP4)

YouTube Videos of Note

Soybean Growth and Development Videos

- Soybean Full Seed (R6) Growth Stage
- Soybean Stages, Development, and Management
- Understanding Soybean Emergence
- Soybean Emergence and Germination Common Issues
- Soybean Yellowing at V2 growth
- R5 Soybean Hail Damage
- R4 Soybean Hail Damage
- Later Season Soybean Growth Stages
- Potassium Deficiency in Soybean
- Potassium Deficient Corn and Soybean During Drought
- Determining Soybean Growth Stages - Tennessee Soybeans
- Early Drought Stress on Timely Planted Soybean
- Row spacing effect on yeild
- Planting date and water stress effect on soybean

Soybean Pest Videos

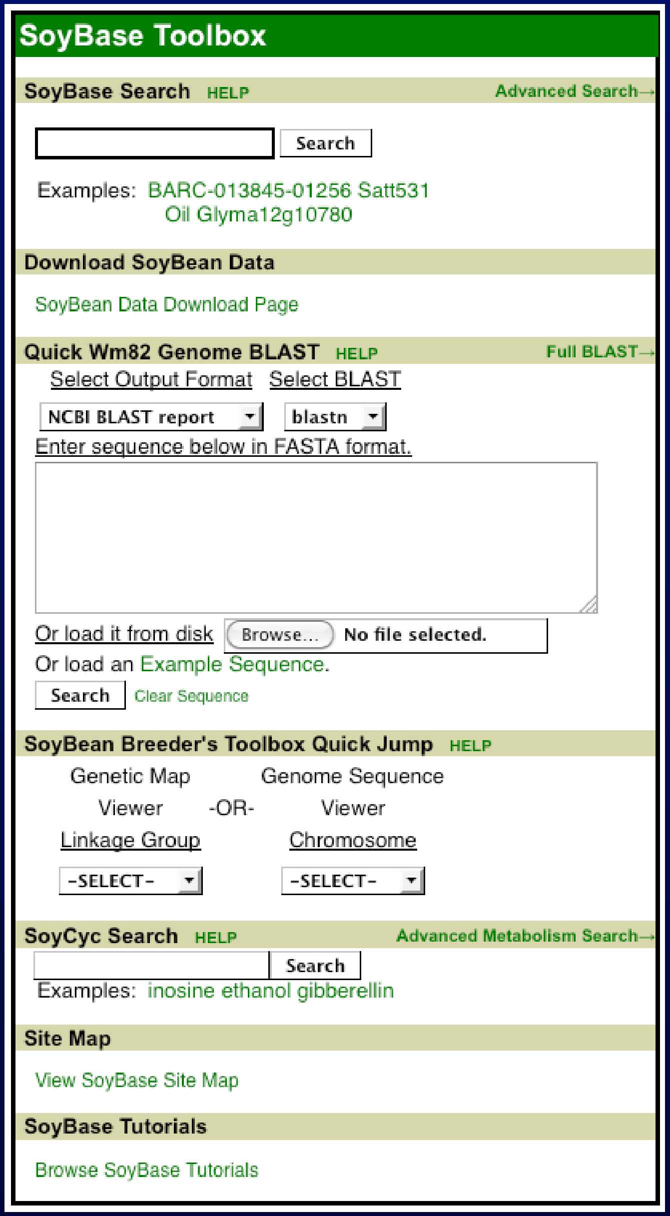
- Scouting Soybean Aphids in Wisconsin
- Soybean Aphid Natural Enemies
- Soybean Aphid Speed Scouting -- How To
- When to Speed Scout for soybean aphids
- Spider Mites in Soybean
- Soybean Insect Pest Management - Japanese Beetle and Soybean Aphid
- Soybean Cyst Nematode Identification and Management
- Soybean Cyst Nematode Fertilization
- Japanese beetle management in corn and soybean
- Soybean Cyst Nematode infestation management
- Scouting for Soybean Cyst Nematode Infestation
- Sampling soil for Soybean Cyst Nematode Infestation

Soybean Disease Videos

- Identifying Common Soybean Diseases
- Sudden death syndrome of soybean
- White Mold in Soybean
- Soybean Leaf Diseases
- Uncovering the root of soybean cyst nematode resistance
- Soybean Vein Necrosis Virus Research at IPFW
- Soybean Vein Necrosis Virus
- Soybean Vein Necrosis Disease (SVND)
- Soybean rust monitoring cuts SC farm losses
- Soybean Diseases Management and Fungicide Application Recommendations

Methods and Protocol Videos

- Estimating Soybean Yields - Simplified
- How to Take a Soybean Leaf Sample
- How to pollenate soybean
- How to use a two row harvester



SoyBase Toolbox

SoyBase Search HELP Advanced Search→

Search

Examples: BARC-013845-01256 Satt531
Oil Glyma12g10780

Download SoyBean Data

SoyBean Data Download Page

Quick Wm82 Genome BLAST HELP Full BLAST→

Select Output Format Select BLAST

NCBI BLAST report blastn

Enter sequence below in FASTA format.

Or load it from disk Browse... No file selected.

Or load an Example Sequence.

Search Clear Sequence

SoyBean Breeder's Toolbox Quick Jump HELP

Genetic Map Viewer -OR- Genome Sequence Viewer

Linkage Group Chromosome

-SELECT- -SELECT-

SoyCyc Search HELP Advanced Metabolism Search→

Search

Examples: inosine ethanol gibberellin

Site Map

View SoyBase Site Map

SoyBase Tutorials

Browse SoyBase Tutorials

Search all of SoyBase for a specific item or a general term. All items related to search term will be returned.

Tools for extracting and downloading all or subsets of the data in SoyBase.

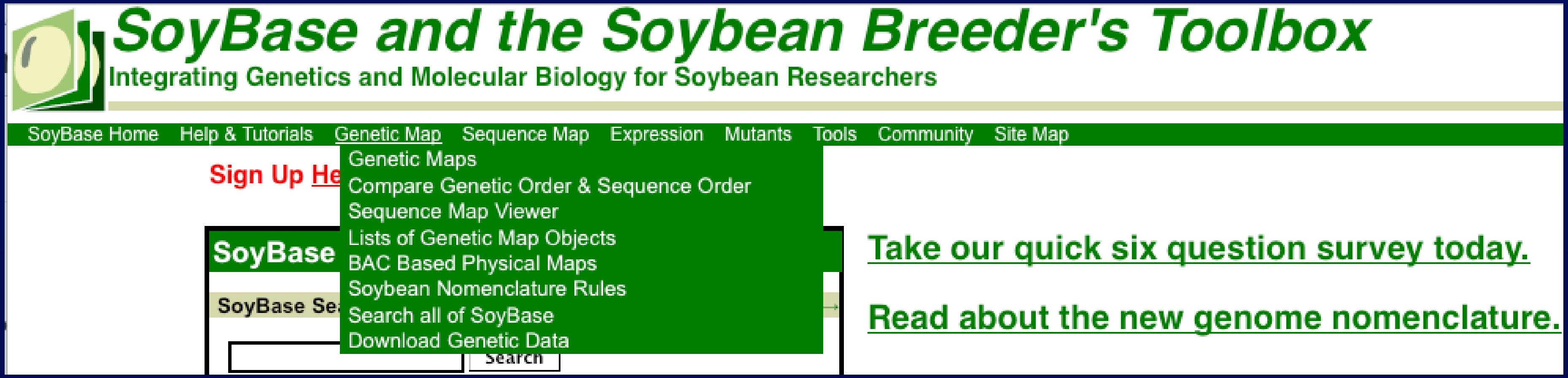
BLAST sequence similarity search against the Wm82.a2 genome sequence. The Full BLAST link allows sequence similarity searches against more than 30 selected subsets of soybean sequences.

Jump directly to viewer for a specific Linkage Group (genetic map) or Chromosome (genome sequence).

Search the SoyBase SoyCyc metabolism database.

Comprehensive overview of the data, map and sequence browsers, and analysis and reporting tools at SoyBase.

Short video tutorials on using SoyBase.

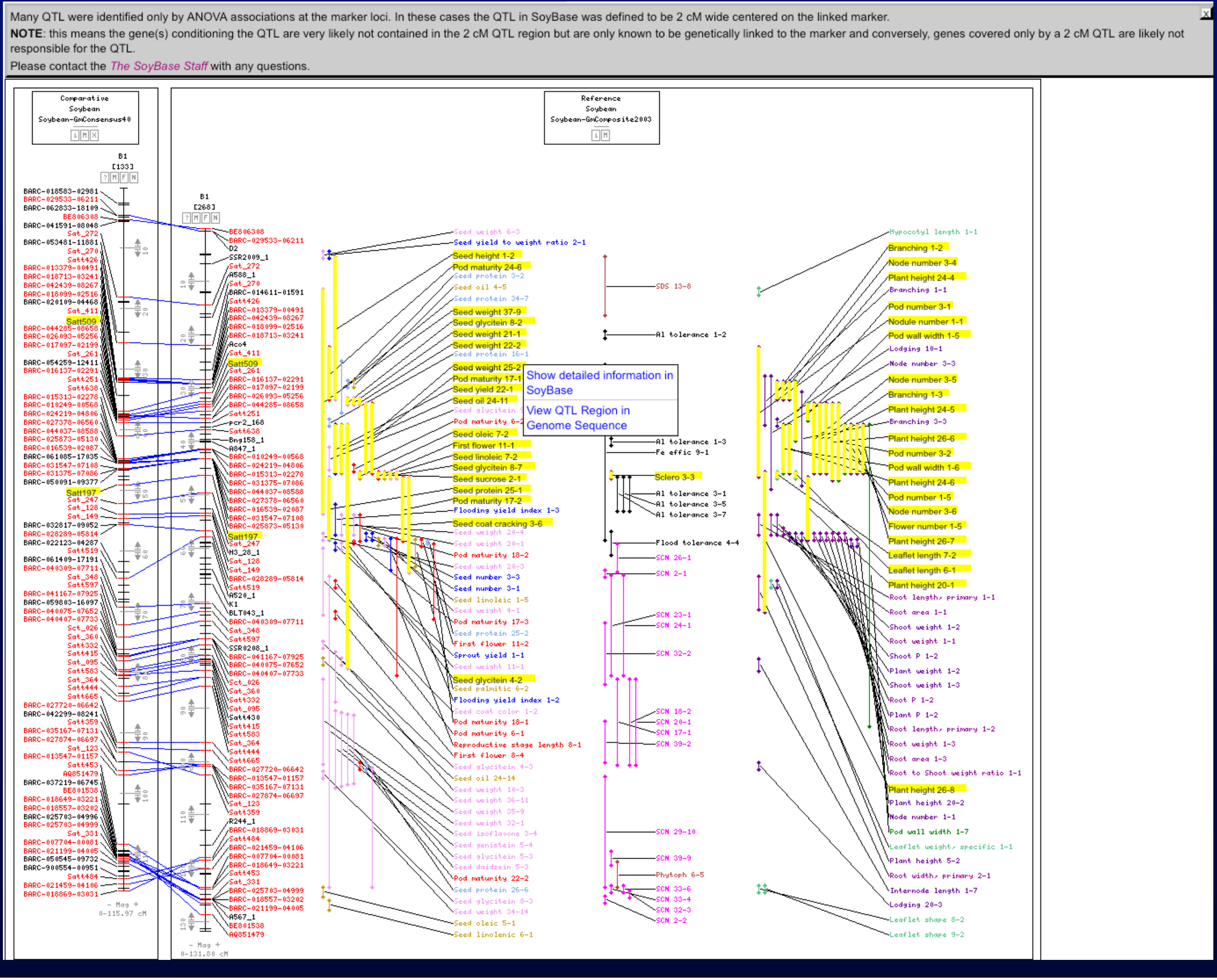


SoyBase and the Soybean Breeder's Toolbox

Integrating Genetics and Molecular Biology for Soybean Researchers

Take our quick six question survey today.

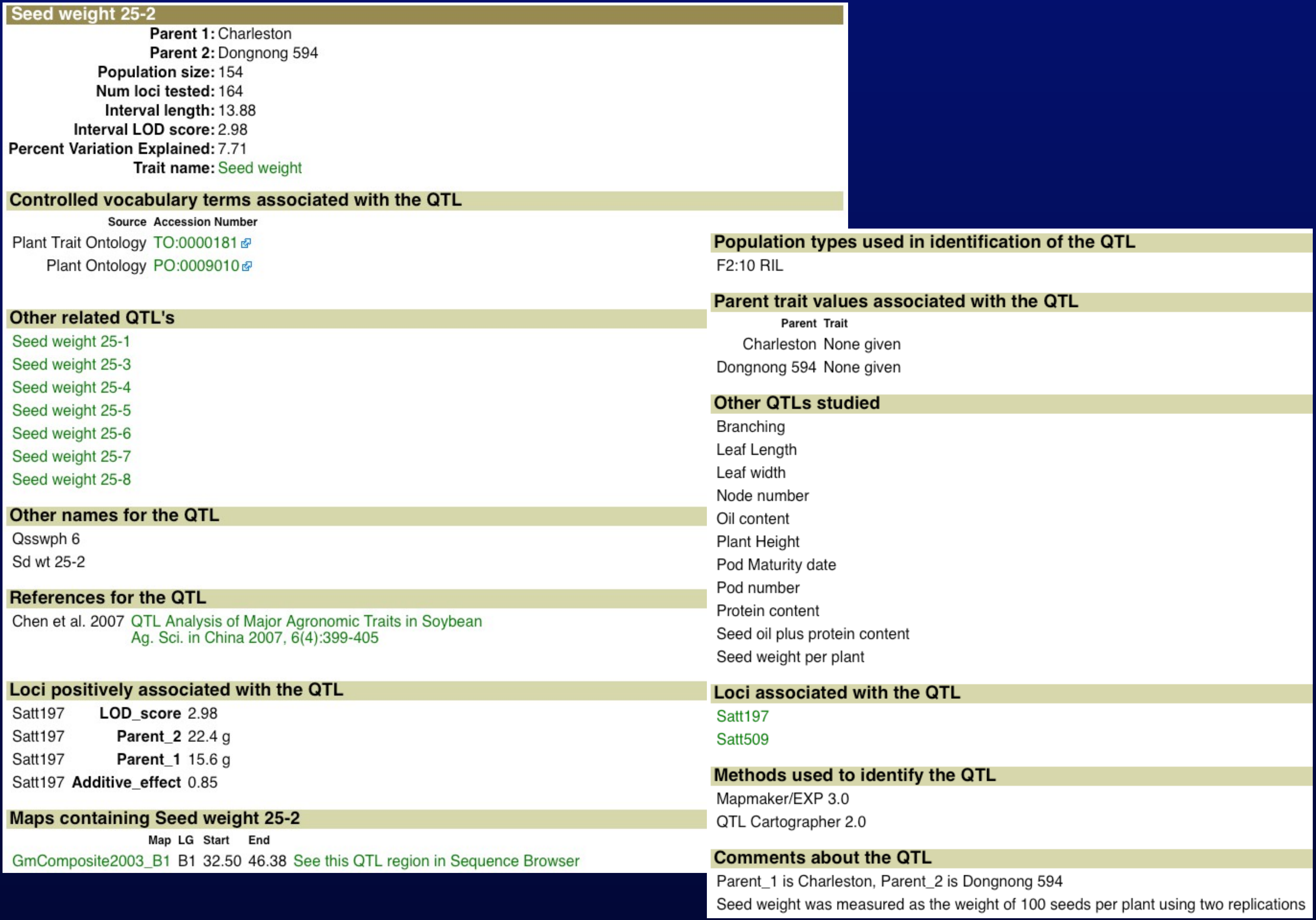
Read about the new genome nomenclature.



Many QTL were identified only by ANOVA associations at the marker loci. In these cases the QTL in SoyBase was defined to be 2 cM wide centered on the linked marker.

NOTE: this means the gene(s) conditioning the QTL are very likely not contained in the 2 cM QTL region but are only known to be genetically linked to the marker and conversely, genes covered only by a 2 cM QTL are likely not responsible for the QTL.

Please contact the *The SoyBase Staff* with any questions.



Seed weight 25-2

Parent 1: Charleston
Parent 2: Dongnong 594
Population size: 154
Num loci tested: 164
Interval length: 13.88
Interval LOD score: 2.98
Percent Variation Explained: 7.71
Trait name: Seed weight

Controlled vocabulary terms associated with the QTL

Source	Accession Number
Plant Trait Ontology	TO:0000181
Plant Ontology	PO:0009010

Other related QTL's

Seed weight 25-1
Seed weight 25-3
Seed weight 25-4
Seed weight 25-5
Seed weight 25-6
Seed weight 25-7
Seed weight 25-8

Other names for the QTL

Qsswph 6
Sd wt 25-2

References for the QTL

Chen et al. 2007 QTL Analysis of Major Agronomic Traits in Soybean
Ag. Sci. in China 2007, 6(4):399-405

Loci positively associated with the QTL

Satt197	**LOD_score**	2.98
Satt197	**Parent_2**	22.4 g
Satt197	**Parent_1**	15.6 g
Satt197	**Additive_effect**	0.85

Maps containing Seed weight 25-2

Map	LG	Start	End	
GmComposite2003_B1	B1	32.50	46.38	See this QTL region in Sequence Browser

Population types used in identification of the QTL

F2:10 RIL

Parent trait values associated with the QTL

Parent	Trait
Charleston	None given
Dongnong 594	None given

Other QTLs studied

Branching
Leaf Length
Leaf width
Node number
Oil content
Plant Height
Pod Maturity date
Pod number
Protein content
Seed oil plus protein content
Seed weight per plant

Loci associated with the QTL

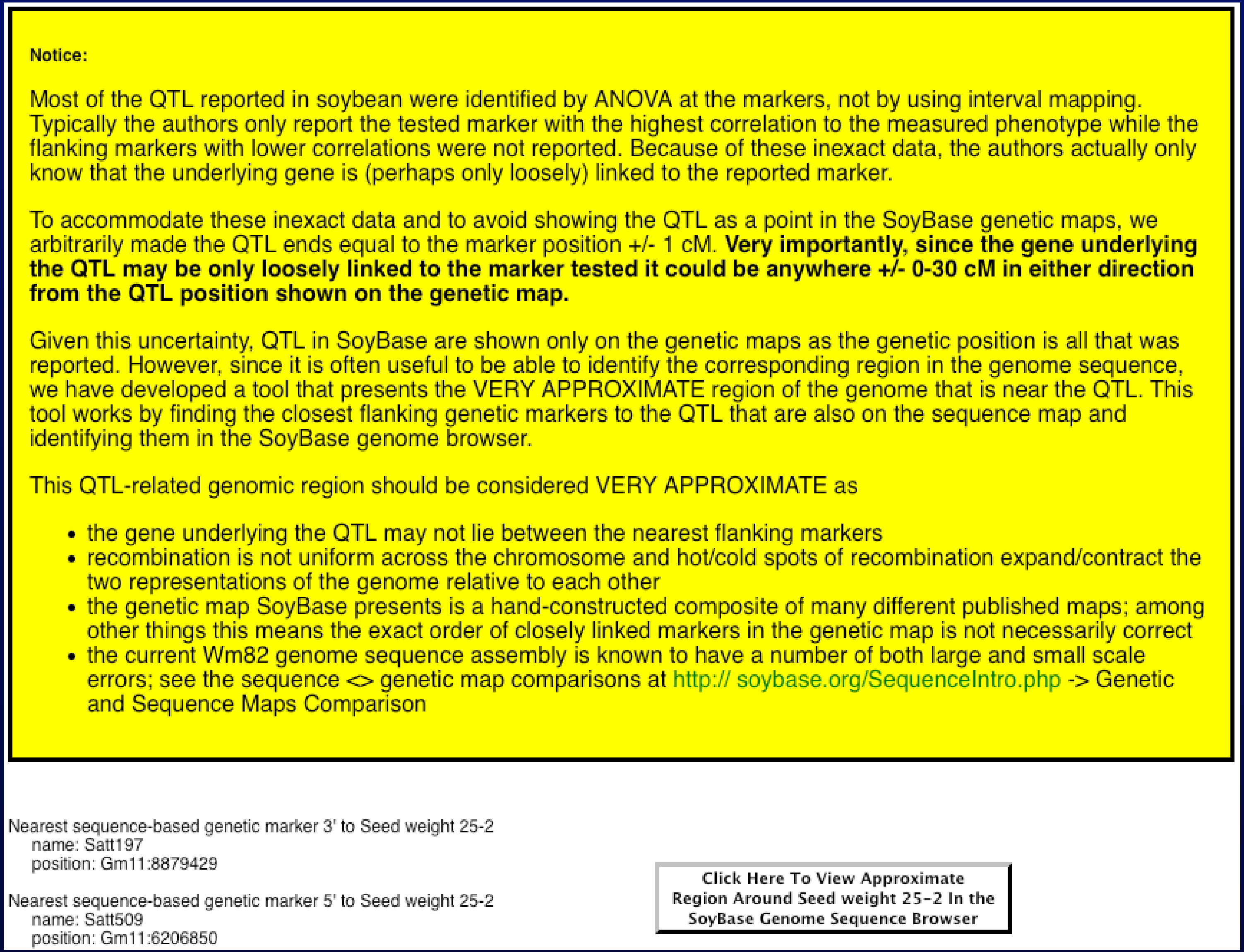
Satt197
Satt509

Methods used to identify the QTL

Mapmaker/EXP 3.0
QTL Cartographer 2.0

Comments about the QTL

Parent_1 is Charleston, Parent_2 is Dongnong 594
Seed weight was measured as the weight of 100 seeds per plant using two replications



Notice:

Most of the QTL reported in soybean were identified by ANOVA at the markers, not by using interval mapping. Typically the authors only report the tested marker with the highest correlation to the measured phenotype while the flanking markers with lower correlations were not reported. Because of these inexact data, the authors actually only know that the underlying gene is (perhaps only loosely) linked to the reported marker.

To accommodate these inexact data and to avoid showing the QTL as a point in the SoyBase genetic maps, we arbitrarily made the QTL ends equal to the marker position +/- 1 cM. **Very importantly, since the gene underlying the QTL may be only loosely linked to the marker tested it could be anywhere +/- 0-30 cM in either direction from the QTL position shown on the genetic map.**

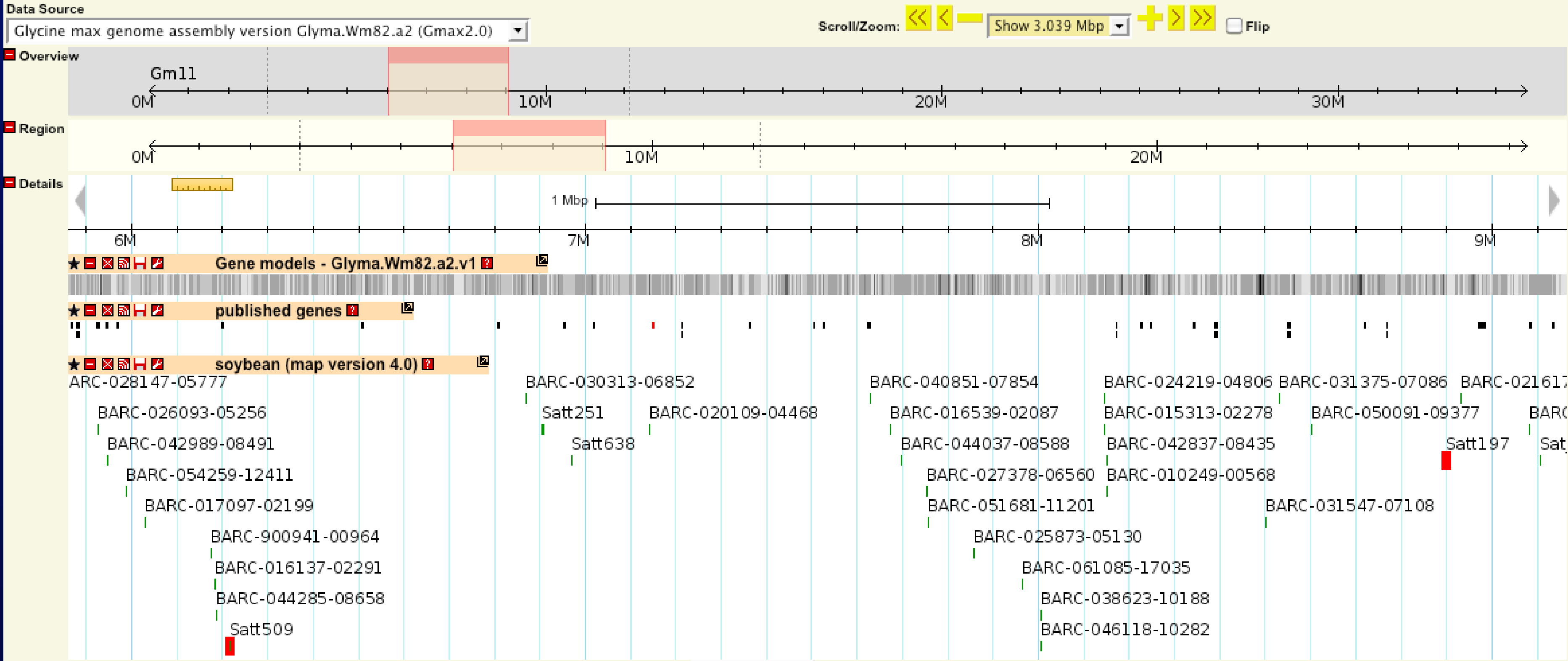
Given this uncertainty, QTL in SoyBase are shown only on the genetic maps as the genetic position is all that was reported. However, since it is often useful to be able to identify the corresponding region in the genome sequence, we have developed a tool that presents the VERY APPROXIMATE region of the genome that is near the QTL. This tool works by finding the closest flanking genetic markers to the QTL that are also on the sequence map and identifying them in the SoyBase genome browser.

This QTL-related genomic region should be considered VERY APPROXIMATE as

- the gene underlying the QTL may not lie between the nearest flanking markers
- recombination is not uniform across the chromosome and hot/cold spots of recombination expand/contract the two representations of the genome relative to each other
- the genetic map SoyBase presents is a hand-constructed composite of many different published maps; among other things this means the exact order of closely linked markers in the genetic map is not necessarily correct
- the current Wm82 genome sequence assembly is known to have a number of both large and small scale errors; see the sequence <> genetic map comparisons at http:// soybase.org/SequenceIntro.php -> Genetic and Sequence Maps Comparison

Nearest sequence-based genetic marker 3' to Seed weight 25-2
name: Satt197
position: Gm11:8879429

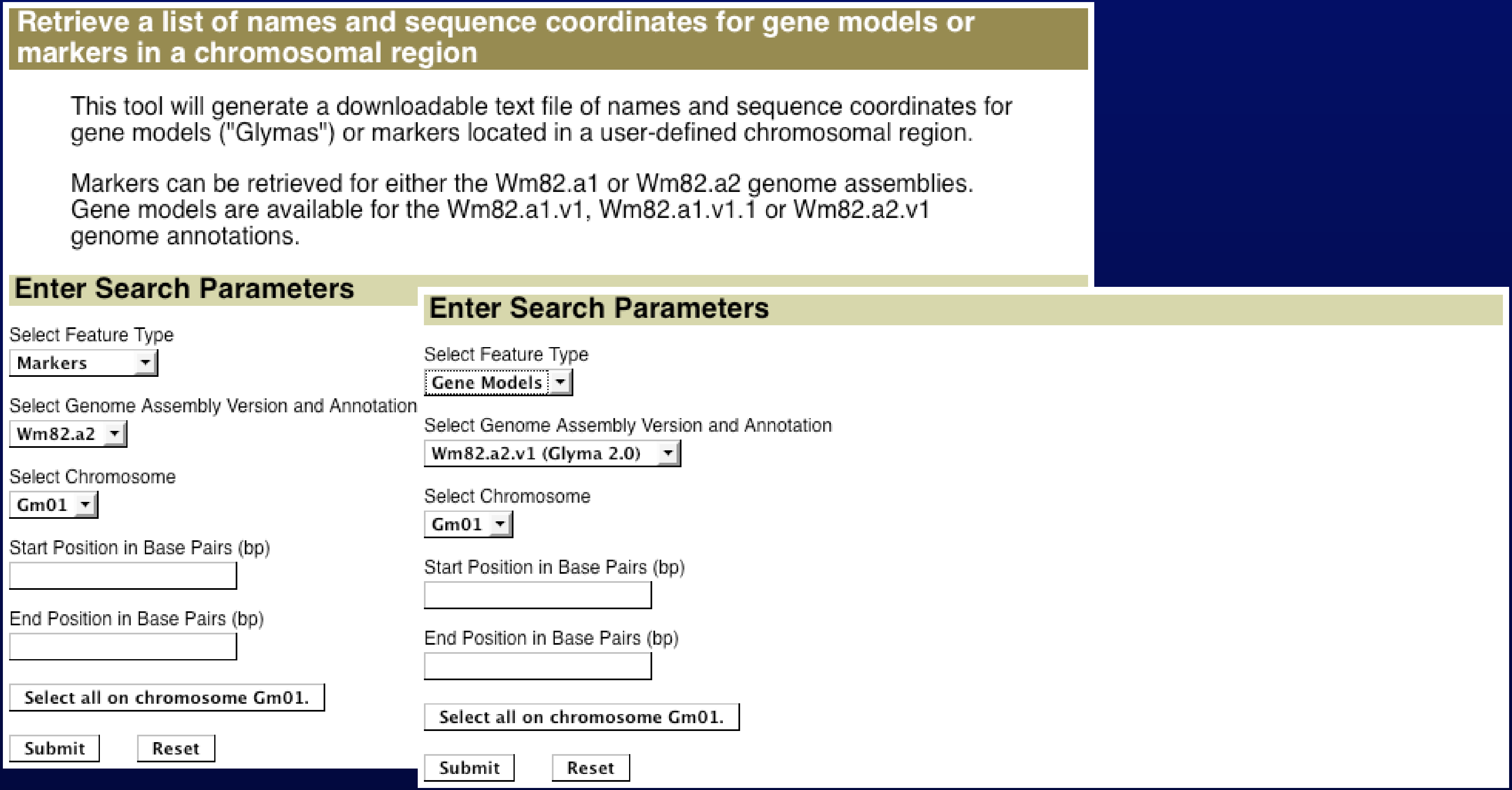
Nearest sequence-based genetic marker 5' to Seed weight 25-2
name: Satt509
position: Gm11:6206850



Data Source: Glycine max genome assembly version Glyma.Wm82.a2 (Gmax2.0)

Scroll/Zoom: Show 3.039 Mbp

☐ Flip



Retrieve a list of names and sequence coordinates for gene models or markers in a chromosomal region

This tool will generate a downloadable text file of names and sequence coordinates for gene models ("Glymas") or markers located in a user-defined chromosomal region.

Markers can be retrieved for either the Wm82.a1 or Wm82.a2 genome assemblies. Gene models are available for the Wm82.a1.v1, Wm82.a1.v1.1 or Wm82.a2.v1 genome annotations.

Enter Search Parameters

Select Feature Type
Markers

Select Genome Assembly Version and Annotation
Wm82.a2

Select Chromosome
Gm01

Start Position in Base Pairs (bp)

End Position in Base Pairs (bp)

Enter Search Parameters

Select Feature Type
Gene Models

Select Genome Assembly Version and Annotation
Wm82.a2.v1 (Glyma 2.0)

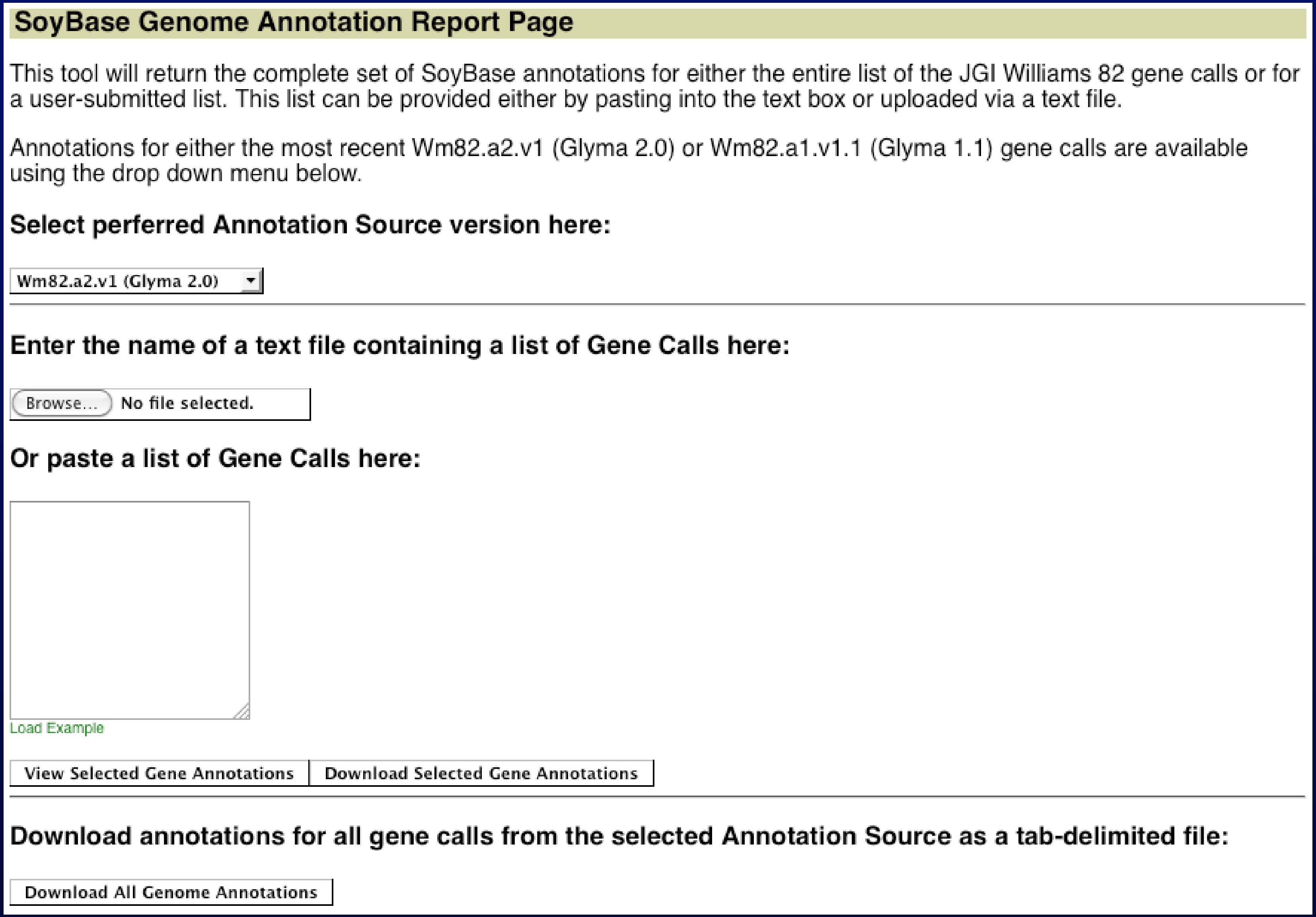
Select Chromosome
Gm01

Start Position in Base Pairs (bp)

End Position in Base Pairs (bp)

Gene models on Genome build Wm82.a2.v1
Chromosome Gm11 from 6216988 to 8899050 bp

Gene Model	Start	Stop
Glyma.11g082800	6221273	6229354
Glyma.11g082900	6232129	6233067
Glyma.11g083000	6237429	6238491
Glyma.11g083100	6239967	6242268
Glyma.11g083200	6245591	6246466
Glyma.11g083300	6254033	6254848
Glyma.11g083400	6256998	6257711
Glyma.11g083500	6260432	6261499
Glyma.11g083600	6265374	6277820
Glyma.11g083700	6279503	6280405
Glyma.11g083800	6285020	6295194
Glyma.11g083900	6304911	6305972
Glyma.11g084000	6306788	6310937
Glyma.11g084100	6320131	6325155
Glyma.11g084200	6327560	6337749
Glyma.11g084300	6340163	6340882
Glyma.11g084400	6342649	6344632
Glyma.11g084500	6349265	6352603
Glyma.11g084600	6358863	6365321
Glyma.11g084700	6367696	6374677
Glyma.11g084800	6379315	6381377
Glyma.11g084900	6393085	6394280
Glyma.11g085000	6397113	6399993
…		
…		
…		



SoyBase Genome Annotation Report Page

This tool will return the complete set of SoyBase annotations for either the entire list of the JGI Williams 82 gene calls or for a user-submitted list. This list can be provided either by pasting into the text box or uploaded via a text file.

Annotations for either the most recent Wm82.a2.v1 (Glyma 2.0) or Wm82.a1.v1.1 (Glyma 1.1) gene calls are available using the drop down menu below.

Select perferred Annotation Source version here:

Wm82.a2.v1 (Glyma 2.0)

Enter the name of a text file containing a list of Gene Calls here:

No file selected.

Or paste a list of Gene Calls here:

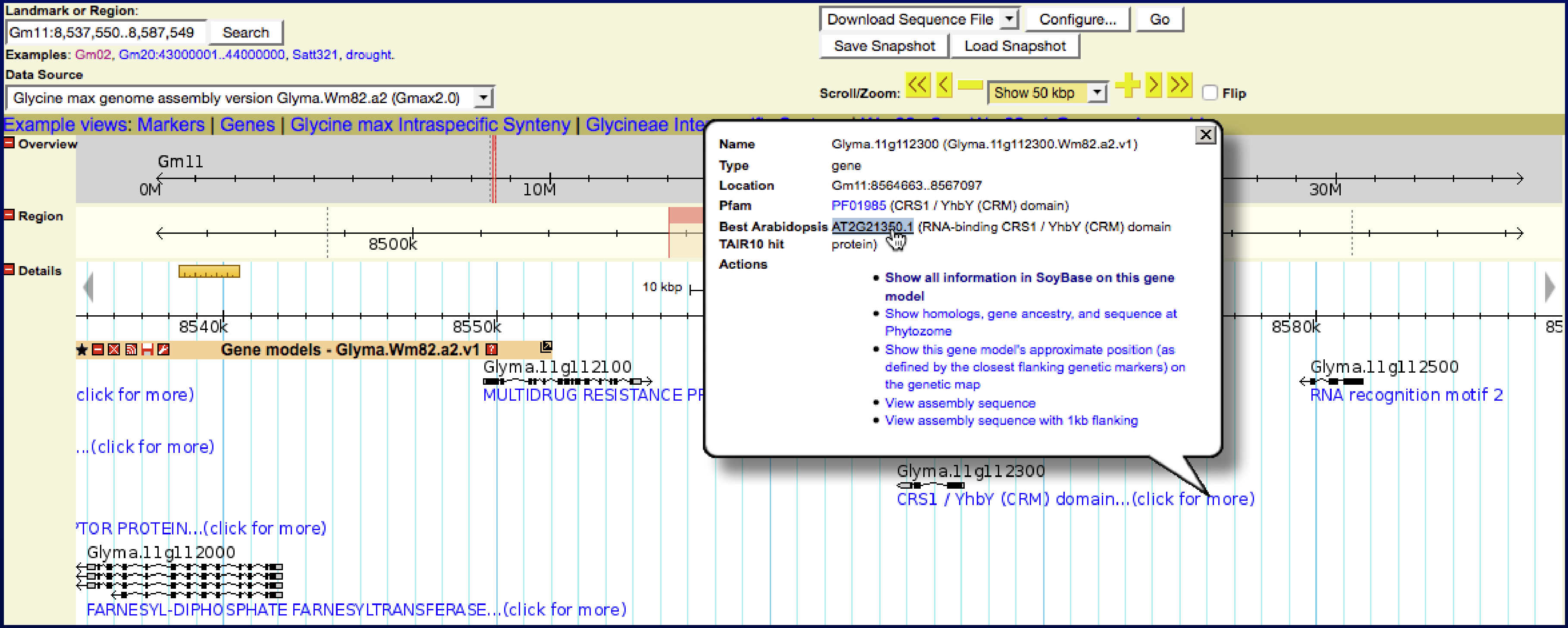
Download annotations for all gene calls from the selected Annotation Source as a tab-delimited file:

Feature	Source	Annotation ID	Annotation Text
Glyma.11g082800	KOG	KOG2473	RNA polymerase III transcription factor (TF)IIIC subunit
Glyma.11g082800	PFAM	PF09734	RNA polymerase III transcription factor (TF)IIIC subunit
Glyma.11g082800	Panther	PTHR13230	GENERAL TRANSCRIPTION FACTOR IIIC, POLYPEPTIDE 5
Glyma.11g082800	Panther	PTHR13230:SF5	GENERAL TRANSCRIPTION FACTOR 3C POLYPEPTIDE 5
Glyma.11g082800	AT	AT3G49410.1	Transcription factor IIIC, subunit 5
Glyma.11g082900	PFAM	PF03087	Arabidopsis protein of unknown function
Glyma.11g082900	AT	AT4G35690.1	Arabidopsis protein of unknown function (DUF241)
Glyma.11g083000	PFAM	PF03087	Arabidopsis protein of unknown function
Glyma.11g083000	AT	AT4G35690.1	Arabidopsis protein of unknown function (DUF241)
Glyma.11g083100	PFAM	PF03087	Arabidopsis protein of unknown function
Glyma.11g083100	AT	AT4G35690.1	Arabidopsis protein of unknown function (DUF241)
Glyma.11g083200	PFAM	PF03087	Arabidopsis protein of unknown function
Glyma.11g083200	AT	AT4G35690.1	Arabidopsis protein of unknown function (DUF241)
Glyma.11g083300	KOG	KOG4473	Uncharacterized membrane protein
Glyma.11g083300	PFAM	PF01988	VIT family
Glyma.11g083300	AT	AT3G43660.1	Vacuolar iron transporter (VIT) family protein
Glyma.11g083400	PFAM	PF03087	Arabidopsis protein of unknown function
Glyma.11g083400	AT	AT4G35690.1	Arabidopsis protein of unknown function (DUF241)
Glyma.11g083500	PFAM	PF03087	Arabidopsis protein of unknown function
Glyma.11g083500	AT	AT4G35680.1	Arabidopsis protein of unknown function (DUF241)
Glyma.11g083600	GO	GO:0003677	DNA binding
Glyma.11g083600	GO	GO:0006306	DNA methylation
Glyma.11g083600	GO	GO:0008168	methyltransferase activity
Glyma.11g083600	PFAM	PF00145	C-5 cytosine-specific DNA methylase
Glyma.11g083600	PFAM	PF00385	Chromo (CHRromatin Organisation MOdifier) domain
Glyma.11g083600	PFAM	PF01426	BAH domain
Glyma.11g083600	Panther	PTHR10629	CYTOSINE-SPECIFIC METHYLTRANSFERASE

…

…

…



Landmark or Region:

Gm11:8,537,550..8,587,549

Examples: Gm02, Gm20:43000001..44000000, Satt321, drought.

Data Source

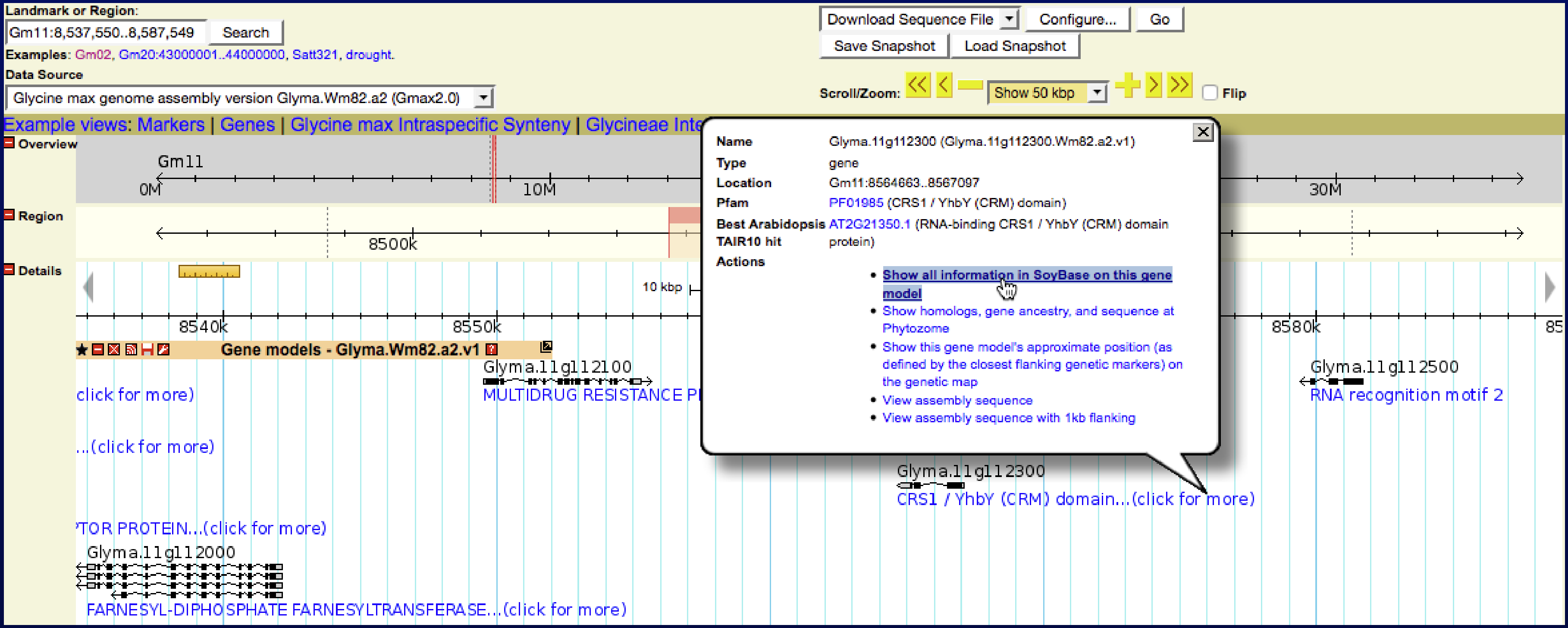
Glycine max genome assembly version Glyma.Wm82.a2 (Gmax2.0)

Scroll/Zoom: Show 50 kbp

Example views: Markers | Genes | Glycine max Intraspecific Synteny | Glycineae Inte

Name	Glyma.11g112300 (Glyma.11g112300.Wm82.a2.v1)
Type	gene
Location	Gm11:8564663..8567097
Pfam	PF01985 (CRS1 / YhbY (CRM) domain)
Best Arabidopsis TAIR10 hit	AT2G21350.1 (RNA-binding CRS1 / YhbY (CRM) domain protein)
Actions	

- **Show all information in SoyBase on this gene model**
- Show homologs, gene ancestry, and sequence at Phytozome
- Show this gene model's approximate position (as defined by the closest flanking genetic markers) on the genetic map
- View assembly sequence
- View assembly sequence with 1kb flanking



Landmark or Region: Gm11:8,537,550..8,587,549

Examples: Gm02, Gm20:43000001..44000000, Satt321, drought.

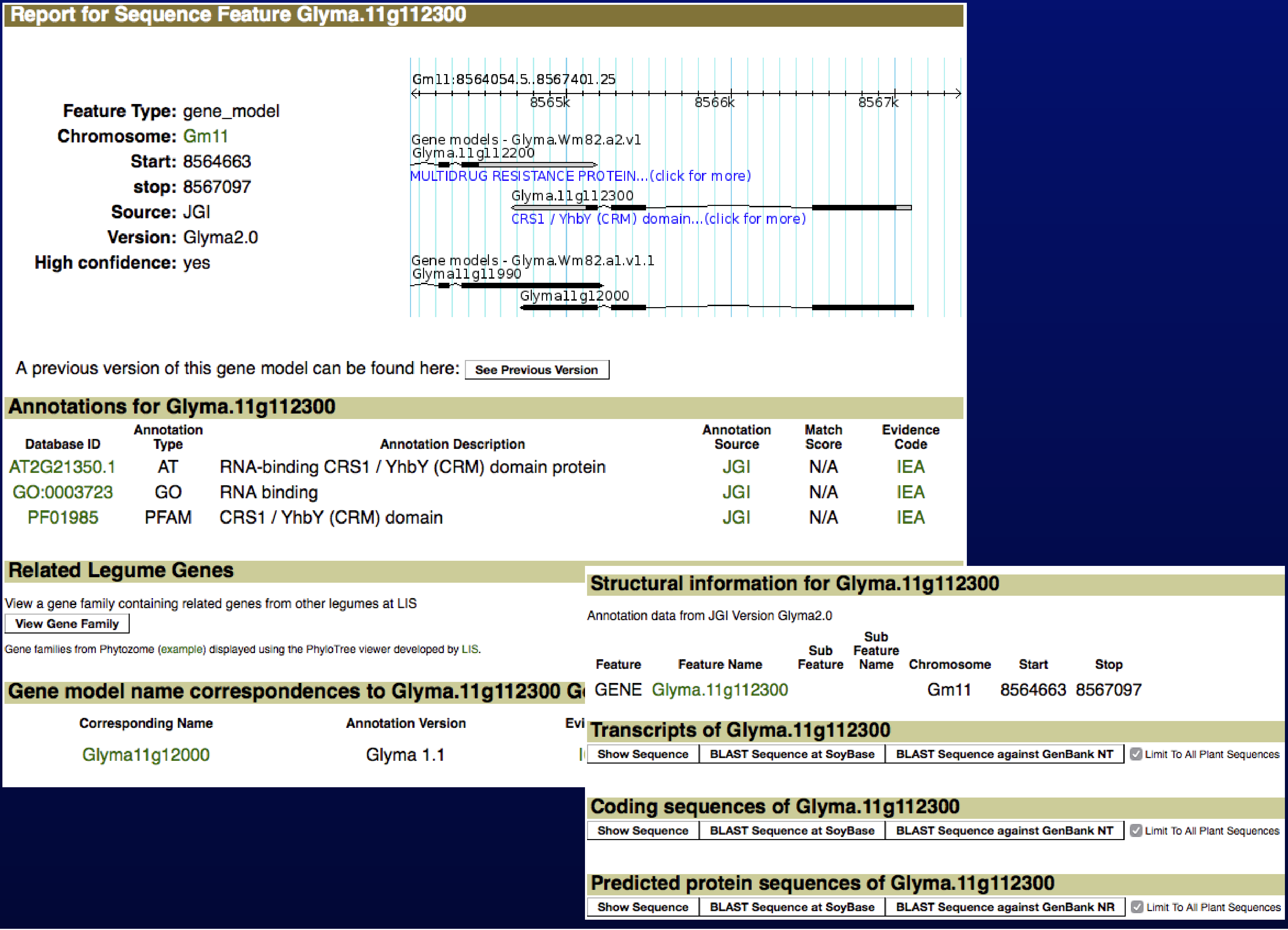
Data Source: Glycine max genome assembly version Glyma.Wm82.a2 (Gmax2.0)

Example views: Markers | Genes | Glycine max Intraspecific Synteny | Glycineae Inte

Name	Glyma.11g112300 (Glyma.11g112300.Wm82.a2.v1)
Type	gene
Location	Gm11:8564663..8567097
Pfam	PF01985 (CRS1 / YhbY (CRM) domain)
Best Arabidopsis TAIR10 hit	AT2G21350.1 (RNA-binding CRS1 / YhbY (CRM) domain protein)

Actions

- Show all information in SoyBase on this gene model
- Show homologs, gene ancestry, and sequence at Phytozome
- Show this gene model's approximate position (as defined by the closest flanking genetic markers) on the genetic map
- View assembly sequence
- View assembly sequence with 1kb flanking



Report for Sequence Feature Glyma.11g112300

Feature Type: gene_model
Chromosome: Gm11
Start: 8564663
stop: 8567097
Source: JGI
Version: Glyma2.0
High confidence: yes

A previous version of this gene model can be found here: See Previous Version

Annotations for Glyma.11g112300

Database ID	Annotation Type	Annotation Description	Annotation Source	Match Score	Evidence Code
AT2G21350.1	AT	RNA-binding CRS1 / YhbY (CRM) domain protein	JGI	N/A	IEA
GO:0003723	GO	RNA binding	JGI	N/A	IEA
PF01985	PFAM	CRS1 / YhbY (CRM) domain	JGI	N/A	IEA

Related Legume Genes

View a gene family containing related genes from other legumes at LIS

View Gene Family

Gene families from Phytozome (example) displayed using the PhyloTree viewer developed by LIS.

Gene model name correspondences to Glyma.11g112300 G

Corresponding Name	Annotation Version	Evi
Glyma11g12000	Glyma 1.1	I

Structural information for Glyma.11g112300

Annotation data from JGI Version Glyma2.0

Feature	Feature Name	Sub Feature	Sub Feature Name	Chromosome	Start	Stop
GENE	Glyma.11g112300			Gm11	8564663	8567097

Transcripts of Glyma.11g112300

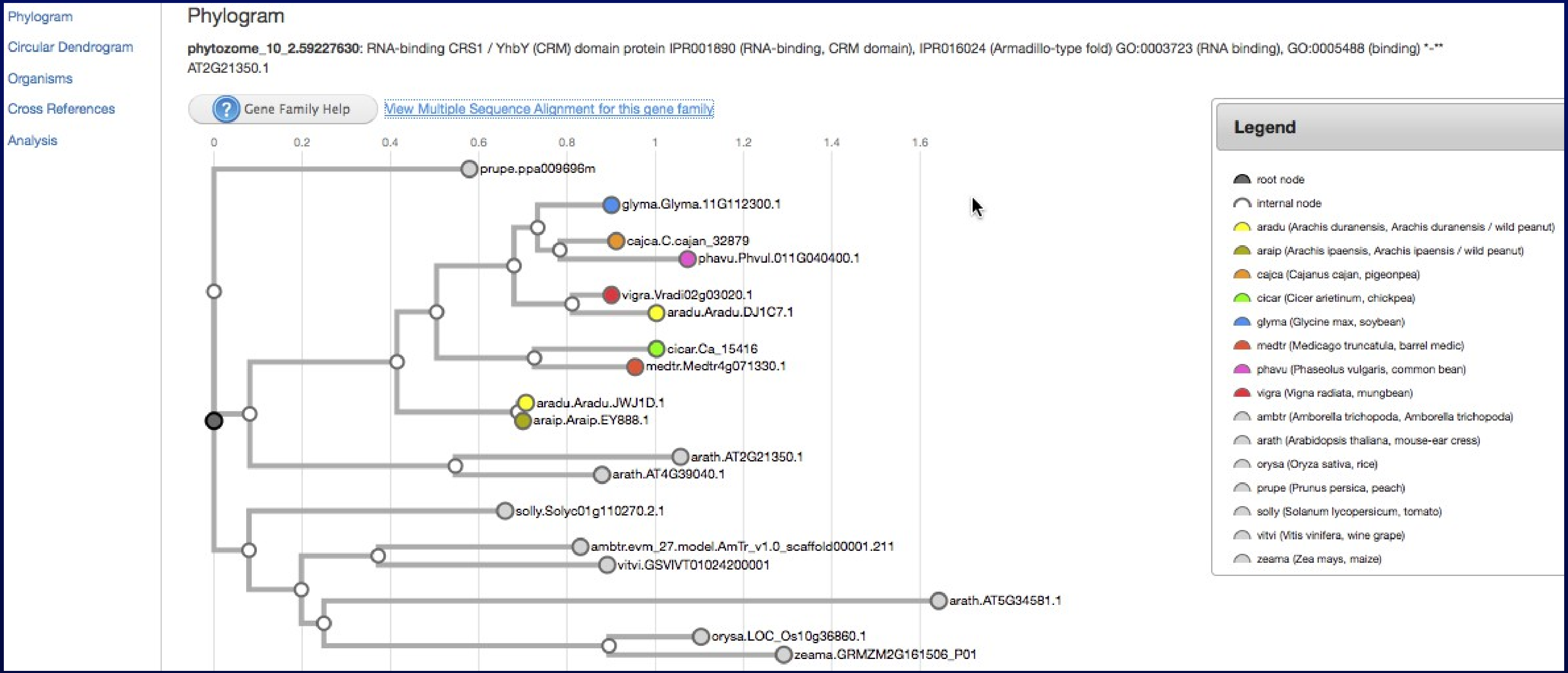
Show Sequence | BLAST Sequence at SoyBase | BLAST Sequence against GenBank NT | ☑ Limit To All Plant Sequences

Coding sequences of Glyma.11g112300

Show Sequence | BLAST Sequence at SoyBase | BLAST Sequence against GenBank NT | ☑ Limit To All Plant Sequences

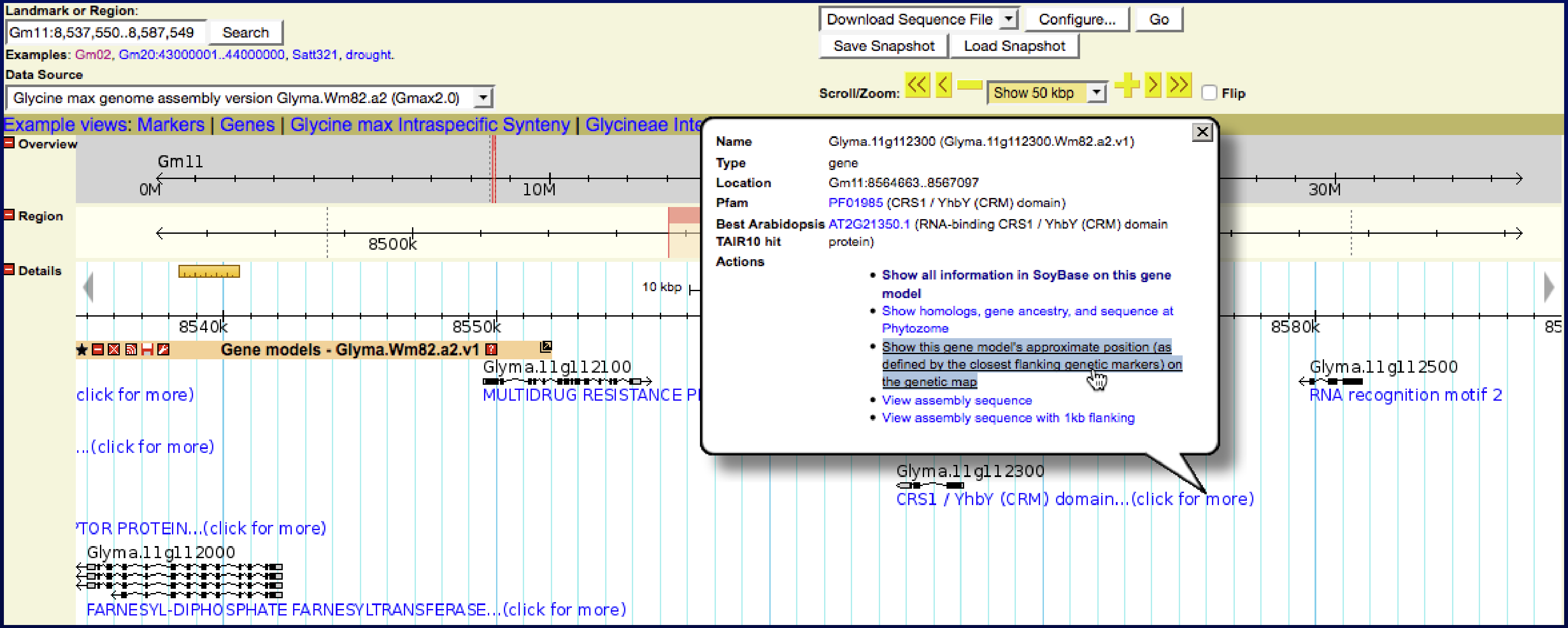
Predicted protein sequences of Glyma.11g112300

Show Sequence | BLAST Sequence at SoyBase | BLAST Sequence against GenBank NR | ☑ Limit To All Plant Sequences



Phylogram

phytozome_10_2.59227630: RNA-binding CRS1 / YhbY (CRM) domain protein IPR001890 (RNA-binding, CRM domain), IPR016024 (Armadillo-type fold) GO:0003723 (RNA binding), GO:0005488 (binding) *-** AT2G21350.1



Landmark or Region: Gm11:8,537,550..8,587,549

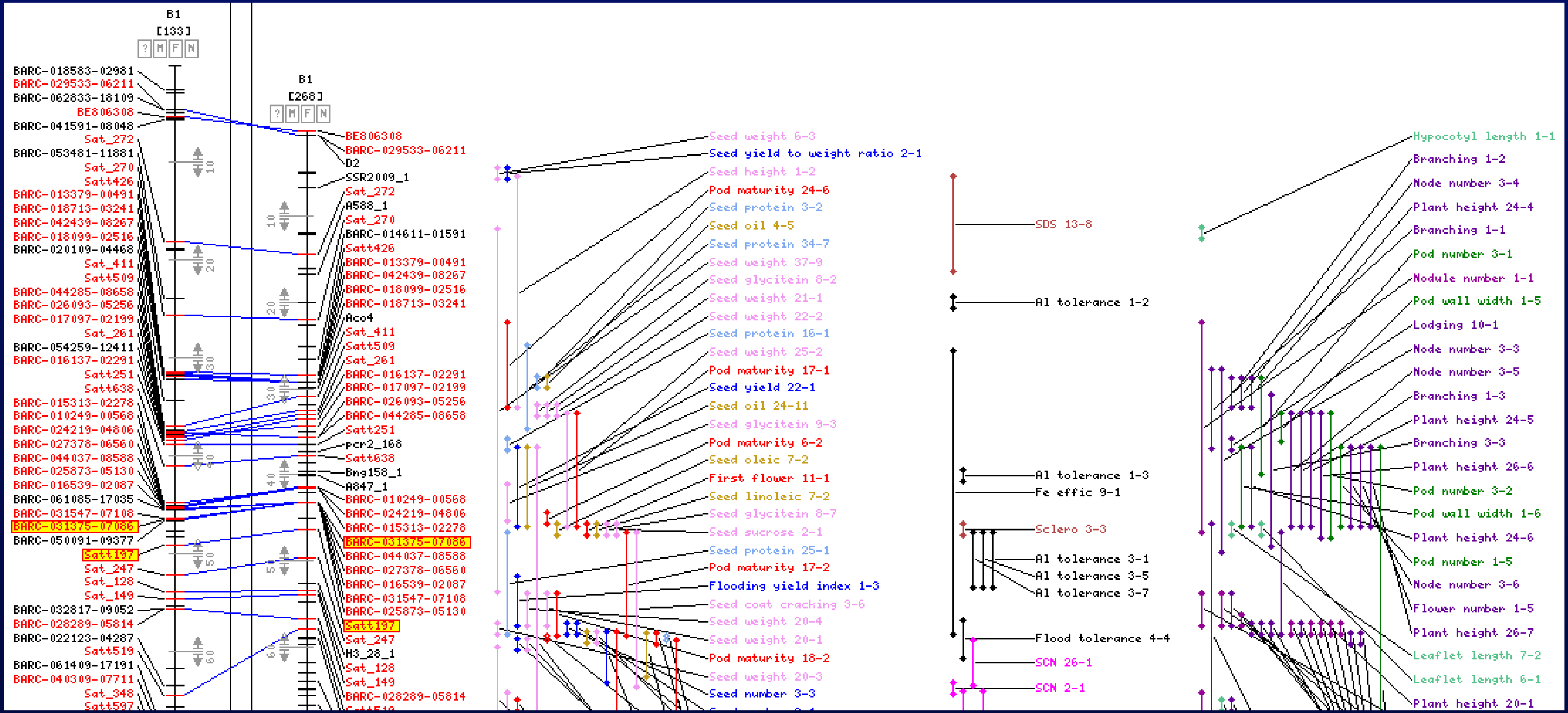
Examples: Gm02, Gm20:43000001..44000000, Satt321, drought.

Data Source: Glycine max genome assembly version Glyma.Wm82.a2 (Gmax2.0)

Scroll/Zoom: Show 50 kbp

Name	Glyma.11g112300 (Glyma.11g112300.Wm82.a2.v1)
Type	gene
Location	Gm11:8564663..8567097
Pfam	PF01985 (CRS1 / YhbY (CRM) domain)
Best Arabidopsis TAIR10 hit	AT2G21350.1 (RNA-binding CRS1 / YhbY (CRM) domain protein)
Actions	

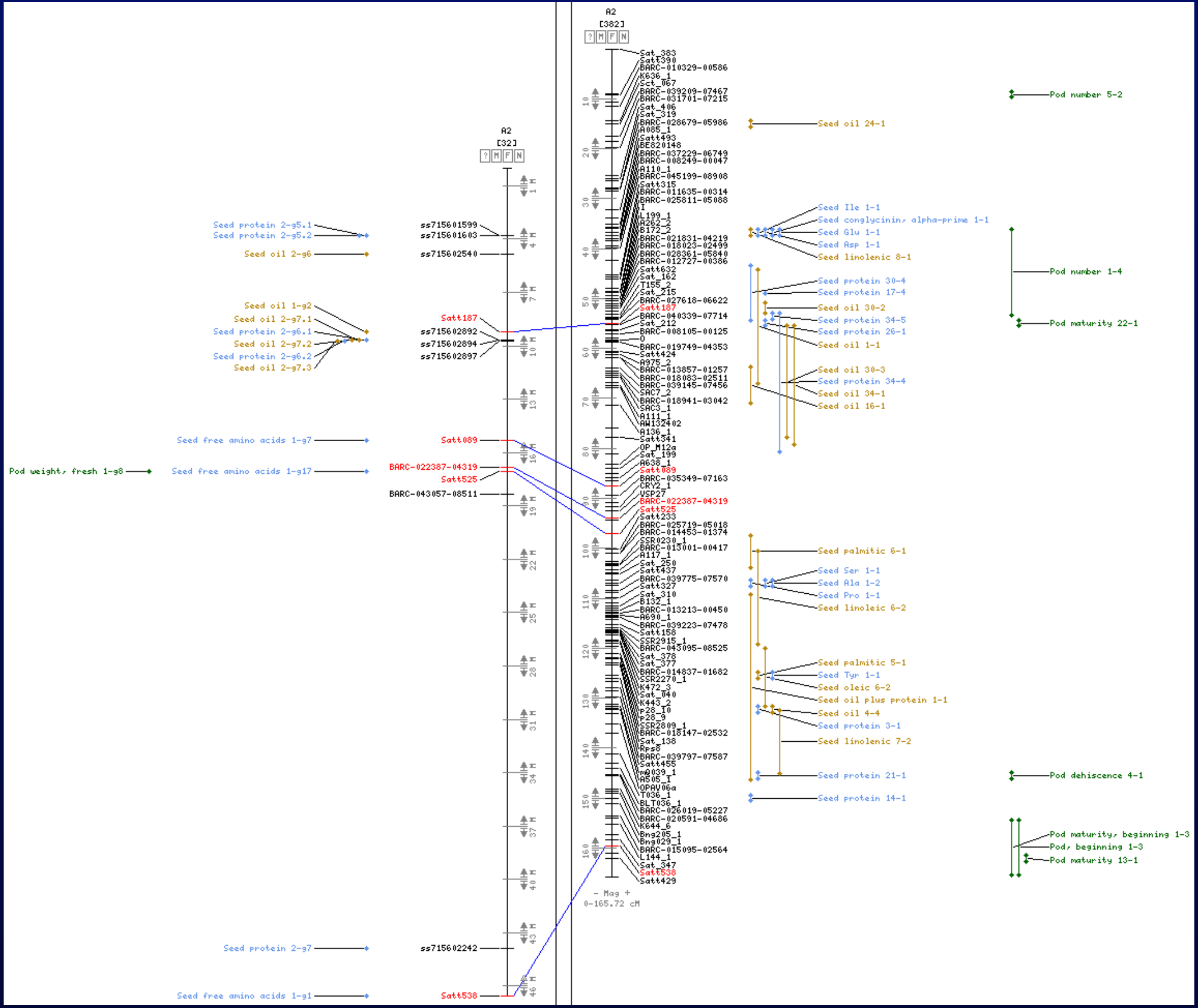
- **Show all information in SoyBase on this gene model**
- Show homologs, gene ancestry, and sequence at Phytozome
- Show this gene model's approximate position (as defined by the closest flanking genetic markers) on the genetic map
- View assembly sequence
- View assembly sequence with 1kb flanking


B1
[133]
? M F N
BARC-018583-02981
BARC-029533-06211
BARC-062833-18109
BE806308
BARC-041591-08048
Sat_272
BARC-053481-11881
Sat_270
Satt426
BARC-013379-00491
BARC-018713-03241
BARC-042439-08267
BARC-018099-02516
BARC-020109-04468
Sat_411
Satt509
BARC-044285-08658
BARC-026093-05256
BARC-017097-02199
Sat_261
BARC-054259-12411
BARC-016137-02291
Satt251
Satt638
BARC-015313-02278
BARC-010249-00568
BARC-024219-04806
BARC-027378-06560
BARC-044037-08588
BARC-025873-05130
BARC-016539-02087
BARC-061085-17035
BARC-031547-07108
BARC-031375-07086
BARC-050091-09377
Satt197
Sat_247
Sat_128
Sat_149
BARC-032817-09052
BARC-028289-05814
BARC-022123-04287
Satt519
BARC-061409-17191
BARC-040309-07711
Sat_348
Satt597
10
20
30
40
50
60
B1
[268]
? M F N
BE806308
BARC-029533-06211
D2
SSR2009_1
Sat_272
A588_1
Sat_270
BARC-014611-01591
Satt426
BARC-013379-00491
BARC-042439-08267
BARC-018099-02516
BARC-018713-03241
Aco4
Sat_411
Satt509
Sat_261
BARC-016137-02291
BARC-017097-02199
BARC-044285-08658
Satt251
pcr2_168
Satt638
Bng158_1
A847_1
BARC-010249-00568
BARC-024219-04806
BARC-015313-02278
BARC-031375-07086
BARC-044037-08588
BARC-027378-06560
BARC-016539-02087
BARC-031547-07108
BARC-025873-05130
Satt197
Sat_247
H3_28_1
Sat_128
Sat_149
BARC-028289-05814
10
20
30
40
50
60
Seed weight 6-3
Seed yield to weight ratio 2-1
Seed height 1-2
Pod maturity 24-6
Seed protein 3-2
Seed oil 4-5
Seed protein 34-7
Seed weight 37-9
Seed glycitein 8-2
Seed weight 21-1
Seed weight 22-2
Seed protein 16-1
Seed weight 25-2
Pod maturity 17-1
Seed yield 22-1
Seed oil 24-11
Seed glycitein 9-3
Pod maturity 6-2
Seed oleic 7-2
First flower 11-1
Seed linoleic 7-2
Seed glycitein 8-7
Seed sucrose 2-1
Seed protein 25-1
Pod maturity 17-2
Flooding yield index 1-3
Seed coat cracking 3-6
Seed weight 20-4
Seed weight 20-1
Pod maturity 18-2
Seed weight 20-3
Seed number 3-3
SDS 13-8
Al tolerance 1-2
Al tolerance 1-3
Fe effic 9-1
Sclero 3-3
Al tolerance 3-1
Al tolerance 3-5
Al tolerance 3-7
Flood tolerance 4-4
SCN 26-1
SCN 2-1
Hypocotyl length 1-1
Branching 1-2
Node number 3-4
Plant height 24-4
Branching 1-1
Pod number 3-1
Nodule number 1-1
Pod wall width 1-5
Lodging 10-1
Node number 3-3
Node number 3-5
Branching 1-3
Plant height 24-5
Branching 3-3
Plant height 26-6
Pod number 3-2
Pod wall width 1-6
Plant height 24-6
Pod number 1-5
Node number 3-6
Flower number 1-5
Plant height 26-7
Leaflet length 7-2
Leaflet length 6-1
Plant height 20-1

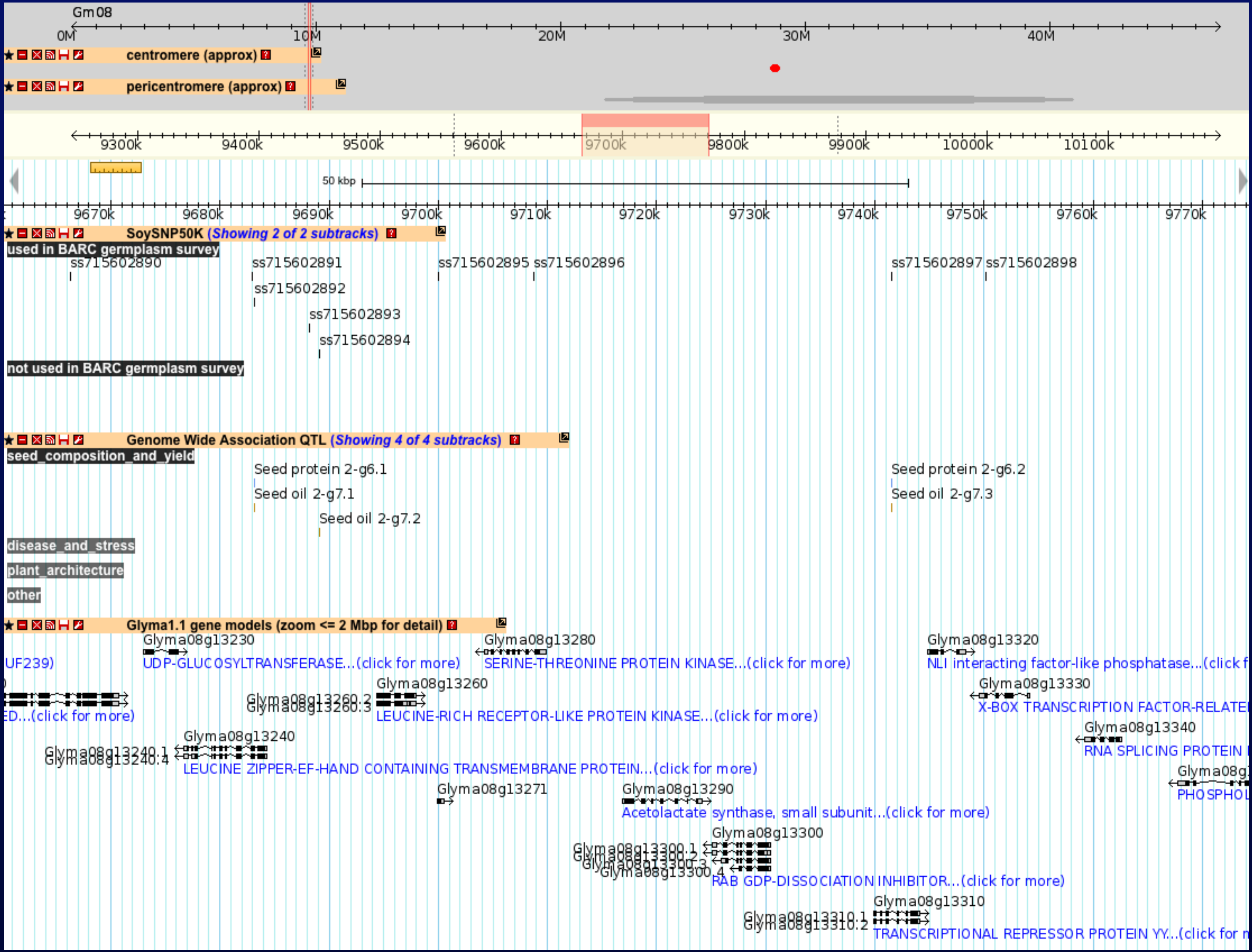
Some Traits Analyzed by GWAS

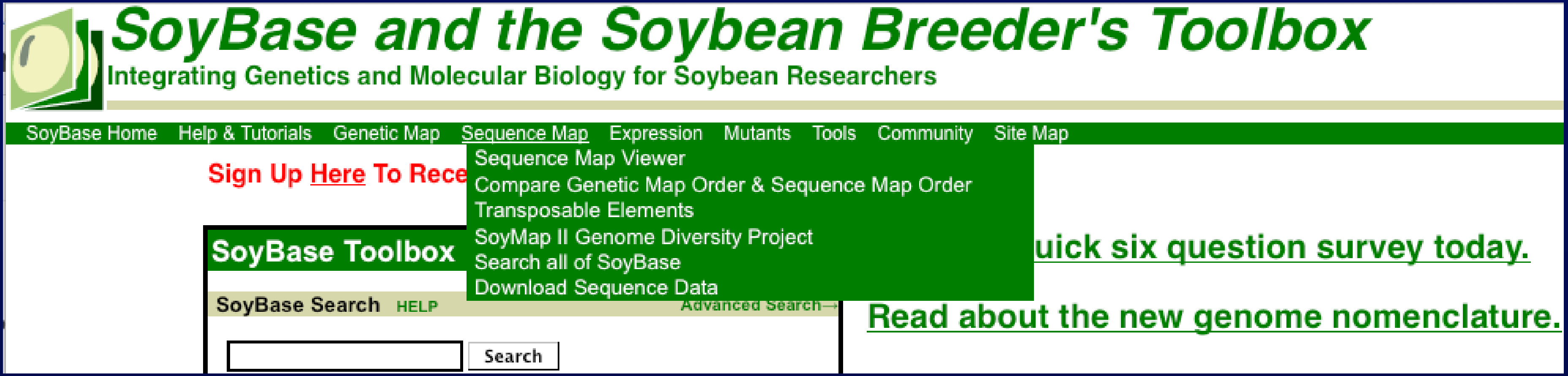
Al tolerance
Chlorophyll content
Chlorophyll fluorescence ABS/RC
Chlorophyll fluorescence ETo/ABS
Chlorophyll fluorescence ETo/TR
Chlorophyll fluorescence Fv/Fm
Chlorophyll fluorescence Piabs
Days from flowering to maturity
Days to flowering
Days to maturity
Drought tolerance
Iron deficiency chlorosis
Pod weight, fresh
Pods per plant
SCN race 1
SCN race 3
Seed free amino acids
Seed length
Seed length to thickness ratio
Seed length to width ratio
Seed oil
Seed protein
Seed sucrose
Seed thickness
Seed weight
Seed weight, fresh
Seed width
Seed width to thickness ratio
Seed yield
Seeds per plant
Soybean mosaic virus

View GWAS QTL in Context of Genetic Map



View GWAS QTL in Context of Genome Sequence





SoyBase and the Soybean Breeder's Toolbox

Integrating Genetics and Molecular Biology for Soybean Researchers

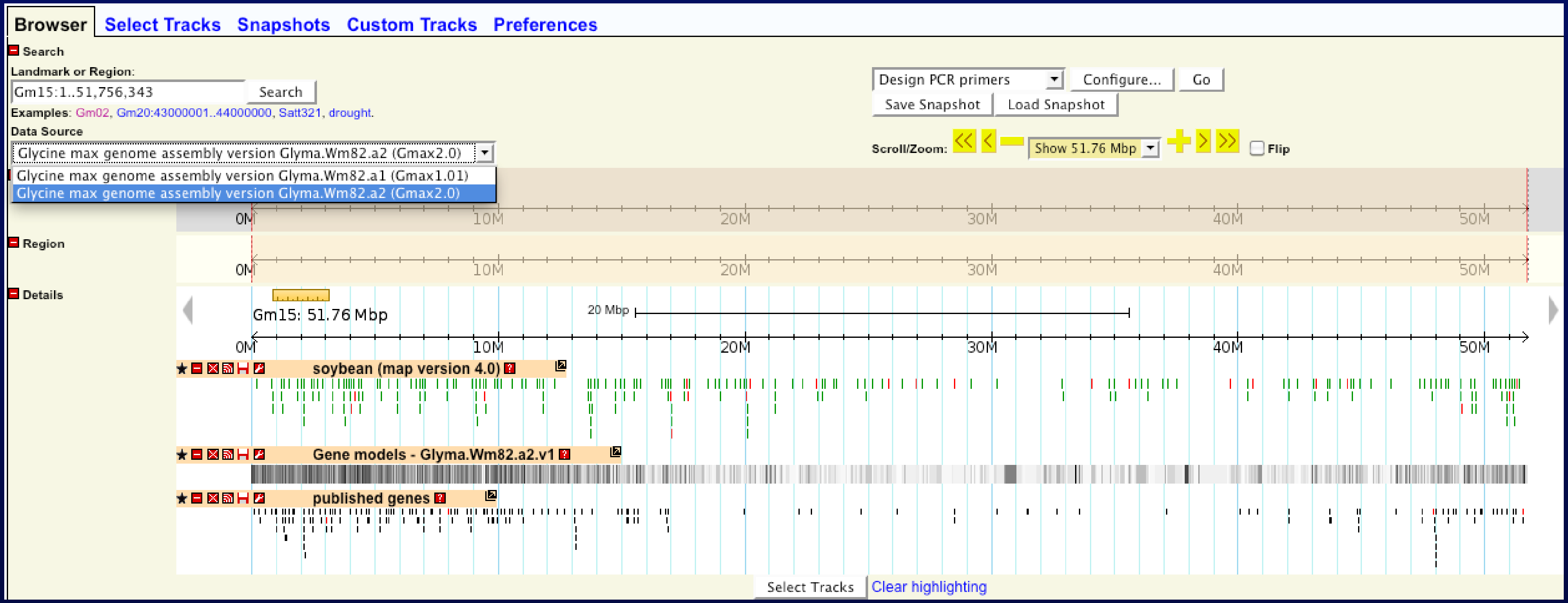
Sign Up Here To Rece

uick six question survey today.

Read about the new genome nomenclature.

SoyBase Toolbox

Genome Browser Available for Both Wm82.a1 and Wm82.a2



Landmark or Region:
Gm15:1..51,756,343

Examples: Gm02, Gm20:43000001..44000000, Satt321, drought.

Data Source
Glycine max genome assembly version Glyma.Wm82.a2 (Gmax2.0)

- Glycine max genome assembly version Glyma.Wm82.a1 (Gmax1.01)
- Glycine max genome assembly version Glyma.Wm82.a2 (Gmax2.0)

Gm15: 51.76 Mbp

Browser | **Select Tracks** | Snapshots | Custom Tracks | Preferences

<< Back to Browser   Show Active Tracks Only   Show Favorites Only ☆   Clear All Favorites ☆

Tracks

General soybean features ☐ All on ☐ All off (assembly supercontigs, centromere (approx) [?], pericentromere (approx) [?])

Genes ☐ All on ☑ All off
- ☆ ☐ Glyma1.0 gene models [?]
- ☆ ☐ Glyma1.1 exon density [?]
- ☆ ☐ NCBI Glycine max annotation [?]
- ☆ ☐ Glyma1.0 gene models (transposon-like) [?]
- ☆ ☐ Glyma1.1 gene models [?]
- ☆ ☐ published genes [?]

Genome structure ☐ All on ☐ All off
- ☆ ☐ Assembly 1.0-2.0 comparison (relaxed) [?]
- ☆ ☐ old duplication blocks (soy-soy 58 Mya) [?]
- ☆ ☐ synteny blocks (with Cajanus) [?]
- ☆ ☐ Assembly 1.0-2.0 comparison (stringent) [?]
- ☆ ☐ recent duplication blocks (soy-soy 13 Mya) [?]
- ☆ ☐ synteny blocks (with Medicago) [?]

Mutants ☐ All on ☐ All off
- ☆ ☐ fast neutron [?] [*showing 2/2 subtracks*]

Naturally occurring sequence variants ☐ All on ☐ All off
- ☆ ☐ SoySNP50K (zoom <= 2 Mbp) [?] [*showing 2/2 subtracks*]

BAC clones ☐ All on ☐ All off
- ☆ ☐ Wm82 FPC contigs [?]
- ☆ ☐ *Phaseolus vulgaris* (common bean) [?]
- ☆ ☐ Wm82 BACs (minimum tiling path) [?]
- ☆ ☐ Wm82 sequenced BACs [?]
- ☆ ☐ Wm82 BACs [?]

Expression - microarray ☐ All on ☐ All off
- ☆ ☐ Affy SoyChip1 [?]

Markers ☐ All on ☑ All off
- ☆ ☐ *Phaseolus vulgaris* (common bean) [?]
- ☆ ☐ soybean (map version 3.0) [?]
- ☆ ☐ Wm82_potential_SSR [?]
- ☆ ☐ Conserved Orthologous Sequence [Doug Cook] [?]
- ☆ ☐ soybean (map version 4.0) [?]
- ☆ ☐ Pigeonpea SSR markers [?]
- ☆ ☐ Williams 82 x PI479752 [Hyten 2010] [?]

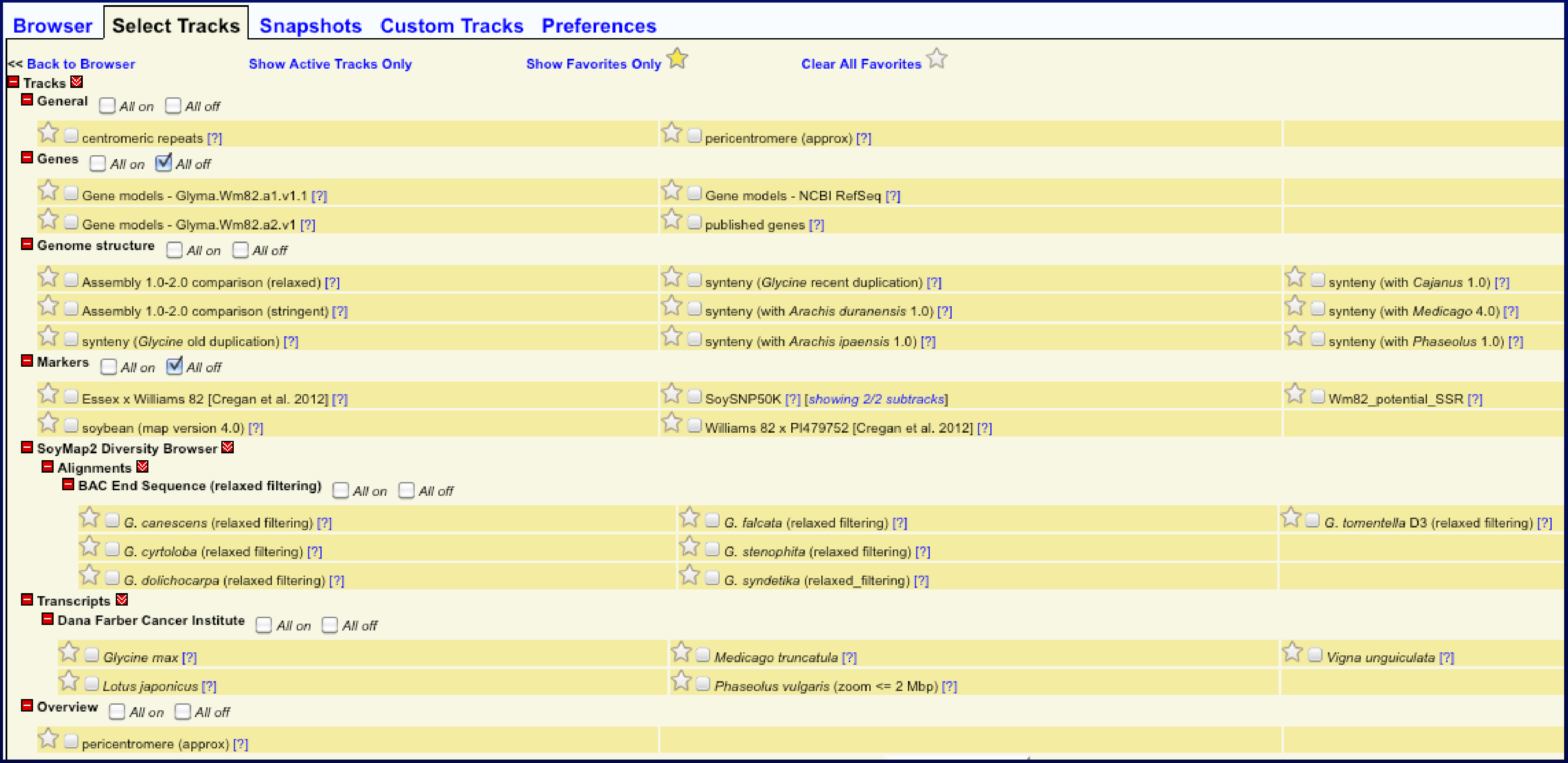
RNA-Seq
- **RNA-Seq** ☐ All on ☐ All off
  - ☆ ☐ Prima 2000 [?] [*showing 3/3 subtracks*]
- **454**
  - **Hypocotyl early response to Phytophthora sojae [Cannon, Bhattacharyya, Sandhu et al. 2010]** ☐ All on ☐ All off
    - ☆ ☐ alignments [Cannon, Bhattacharyya, Sandhu et al. 2010] [?] [*showing 5/5 subtracks*]
    - ☆ ☐ transcript count by tissue [?]
- **Illumina**
  - **Seed development [Bolon, Joseph et al. 2010]** ☐ All on ☐ All off
    - ☆ ☐ transcript count by tissue [?]
    - ☆ ☐ transcript density [?] [*showing 8/8 subtracks*]
  - **Soy atlas [Severin et al. 2010]** ☐ All on ☐ All off
    - ☆ ☐ transcript density [?] [*showing 14/14 subtracks*]

Repetitive sequence ☐ All on ☐ All off
- ☆ ☐ TE1 LTR V2 (all) [?]
- ☆ ☐ TE1 LTR V2 (LTR) [?]
- ☆ ☐ telomeric repeats [?]
- ☆ ☐ TE1 LTR V2 (Helitron_ukn) [?]
- ☆ ☐ TE1 LTR V2 (TIR) [?]
- ☆ ☐ TE1 LTR V2 (LINE_ukn) [?]
- ☆ ☐ TE1 LTR V2 density (all) [?]

SoyMap2 Diversity Browser
- **Alignments**
  - **BAC End Sequence** ☐ All on ☐ All off (*G. canescens* [?], *G. falcata* [?], *G. syndetika* [?], *G. cyrtoloba* [?], *G. soja* [?], *G. tomentella D3* [?], *G. dolichocarpa* [?], *G. stenophita* [?])
  - **BAC End Sequence (abnormally long insert size)** ☐ All on ☐ All off (*G. canescens* [?], *G. falcata* [?], *G. syndetika* [?], *G. cyrtoloba* [?], *G. soja* [?], *G. tomentella D3* [?], *G. dolichocarpa* [?], *G. stenophita* [?])
  - **BAC End Sequence (relaxed filtering)** ☐ All on ☐ All off (*G. canescens* [?], *G. falcata* [?], *G. tomentella D3* [?], *G. cyrtoloba* [?], *G. stenophita* [?], *G. dolichocarpa* [?], *G. syndetika* [?])

Transcripts
- **Dana Farber Cancer Institute** ☐ All on ☐ All off
  - ☆ ☐ *Glycine max* [?]
  - ☆ ☐ *Lotus japonicus* [?]
  - ☆ ☐ *Medicago truncatula* [?]
- **JCVI** ☐ All on ☐ All off
  - ☆ ☐ *Arachis hypogaea* [?]
  - ☆ ☐ *Lotus japonicus* [?]
  - ☆ ☐ *Vigna unguiculata* [?]
  - ☆ ☐ *Arachis stenosperma* [?]
  - ☆ ☐ *Phaseolus coccineus* [?]
  - ☆ ☐ Lupinus albus transcripts [JCVI] [?]
  - ☆ ☐ *Glycine max* [?]
  - ☆ ☐ *Phaseolus vulgaris* [?]
  - ☆ ☐ *Glycine soja* [?]
  - ☆ ☐ *Pisum sativum* [?]
- **Various sources** ☐ All on ☐ All off
  - ☆ ☐ *Cajanus cajan* v.1 [Varshney et al.] [?]
  - ☆ ☐ *Cajanus cajan* v.2 [Varshney et al.] [?]
  - ☆ ☐ *Chamaecrista fasciculata* - clean [Singer et al. 2010] [?]
  - ☆ ☐ *Cajanus cajan* v.1 primers [Varshney et al.] [?]
  - ☆ ☐ *Cajanus cajan* v.2 primers [Varshney et al.] [?]
  - ☆ ☐ *Phaseolus vulgaris* [Phil McClean] [?]

Overview ☐ All on ☑ All off
- ☆ ☐ centromere (approx) [?]
- ☆ ☐ pericentromere (approx) [?]



Tracks

General ☐ All on ☐ All off

☆ ☐ centromeric repeats [?]	☆ ☐ pericentromere (approx) [?]	

Genes ☐ All on ☑ All off

☆ ☐ Gene models - Glyma.Wm82.a1.v1.1 [?]	☆ ☐ Gene models - NCBI RefSeq [?]	
☆ ☐ Gene models - Glyma.Wm82.a2.v1 [?]	☆ ☐ published genes [?]	

Genome structure ☐ All on ☐ All off

☆ ☐ Assembly 1.0-2.0 comparison (relaxed) [?]	☆ ☐ synteny (*Glycine* recent duplication) [?]	☆ ☐ synteny (with *Cajanus* 1.0) [?]
☆ ☐ Assembly 1.0-2.0 comparison (stringent) [?]	☆ ☐ synteny (with *Arachis duranensis* 1.0) [?]	☆ ☐ synteny (with *Medicago* 4.0) [?]
☆ ☐ synteny (*Glycine* old duplication) [?]	☆ ☐ synteny (with *Arachis ipaensis* 1.0) [?]	☆ ☐ synteny (with *Phaseolus* 1.0) [?]

Markers ☐ All on ☑ All off

☆ ☐ Essex x Williams 82 [Cregan et al. 2012] [?]	☆ ☐ SoySNP50K [?] [*showing 2/2 subtracks*]	☆ ☐ Wm82_potential_SSR [?]
☆ ☐ soybean (map version 4.0) [?]	☆ ☐ Williams 82 x PI479752 [Cregan et al. 2012] [?]	

SoyMap2 Diversity Browser

Alignments

BAC End Sequence (relaxed filtering) ☐ All on ☐ All off

☆ ☐ *G. canescens* (relaxed filtering) [?]	☆ ☐ *G. falcata* (relaxed filtering) [?]	☆ ☐ *G. tomentella* D3 (relaxed filtering) [?]
☆ ☐ *G. cyrtoloba* (relaxed filtering) [?]	☆ ☐ *G. stenophita* (relaxed filtering) [?]	
☆ ☐ *G. dolichocarpa* (relaxed filtering) [?]	☆ ☐ *G. syndetika* (relaxed_filtering) [?]	

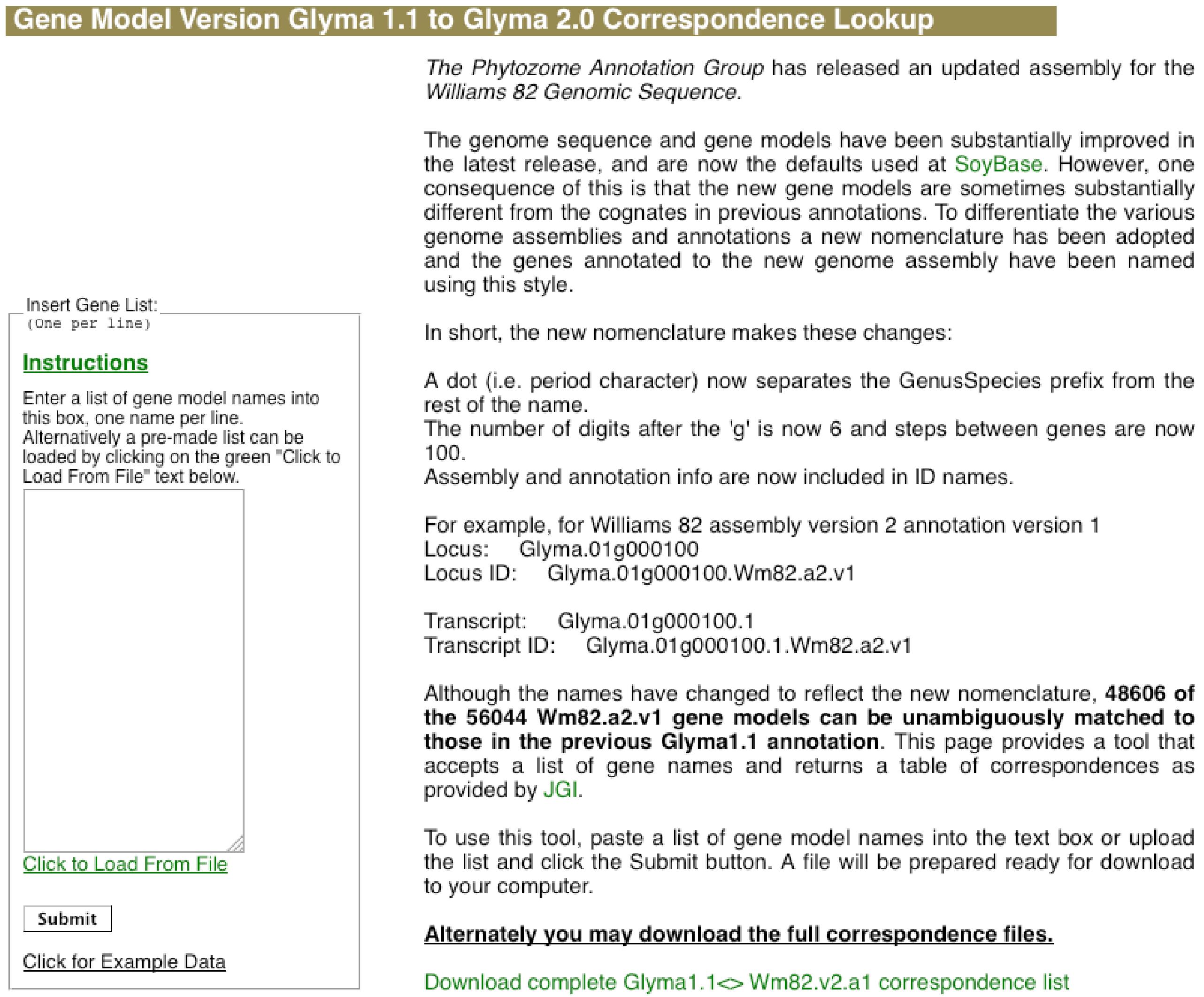
Transcripts

Dana Farber Cancer Institute ☐ All on ☐ All off

☆ ☐ *Glycine max* [?]	☆ ☐ *Medicago truncatula* [?]	☆ ☐ *Vigna unguiculata* [?]
☆ ☐ *Lotus japonicus* [?]	☆ ☐ *Phaseolus vulgaris* (zoom <= 2 Mbp) [?]	

Overview ☐ All on ☐ All off

☆ ☐ pericentromere (approx) [?]		



Gene Model Version Glyma 1.1 to Glyma 2.0 Correspondence Lookup

The Phytozome Annotation Group has released an updated assembly for the *Williams 82 Genomic Sequence.*

The genome sequence and gene models have been substantially improved in the latest release, and are now the defaults used at SoyBase. However, one consequence of this is that the new gene models are sometimes substantially different from the cognates in previous annotations. To differentiate the various genome assemblies and annotations a new nomenclature has been adopted and the genes annotated to the new genome assembly have been named using this style.

In short, the new nomenclature makes these changes:

A dot (i.e. period character) now separates the GenusSpecies prefix from the rest of the name.
The number of digits after the 'g' is now 6 and steps between genes are now 100.
Assembly and annotation info are now included in ID names.

For example, for Williams 82 assembly version 2 annotation version 1
Locus: Glyma.01g000100
Locus ID: Glyma.01g000100.Wm82.a2.v1

Transcript: Glyma.01g000100.1
Transcript ID: Glyma.01g000100.1.Wm82.a2.v1

Although the names have changed to reflect the new nomenclature, **48606 of the 56044 Wm82.a2.v1 gene models can be unambiguously matched to those in the previous Glyma1.1 annotation**. This page provides a tool that accepts a list of gene names and returns a table of correspondences as provided by JGI.

To use this tool, paste a list of gene model names into the text box or upload the list and click the Submit button. A file will be prepared ready for download to your computer.

Alternately you may download the full correspondence files.

Download complete Glyma1.1<> Wm82.v2.a1 correspondence list

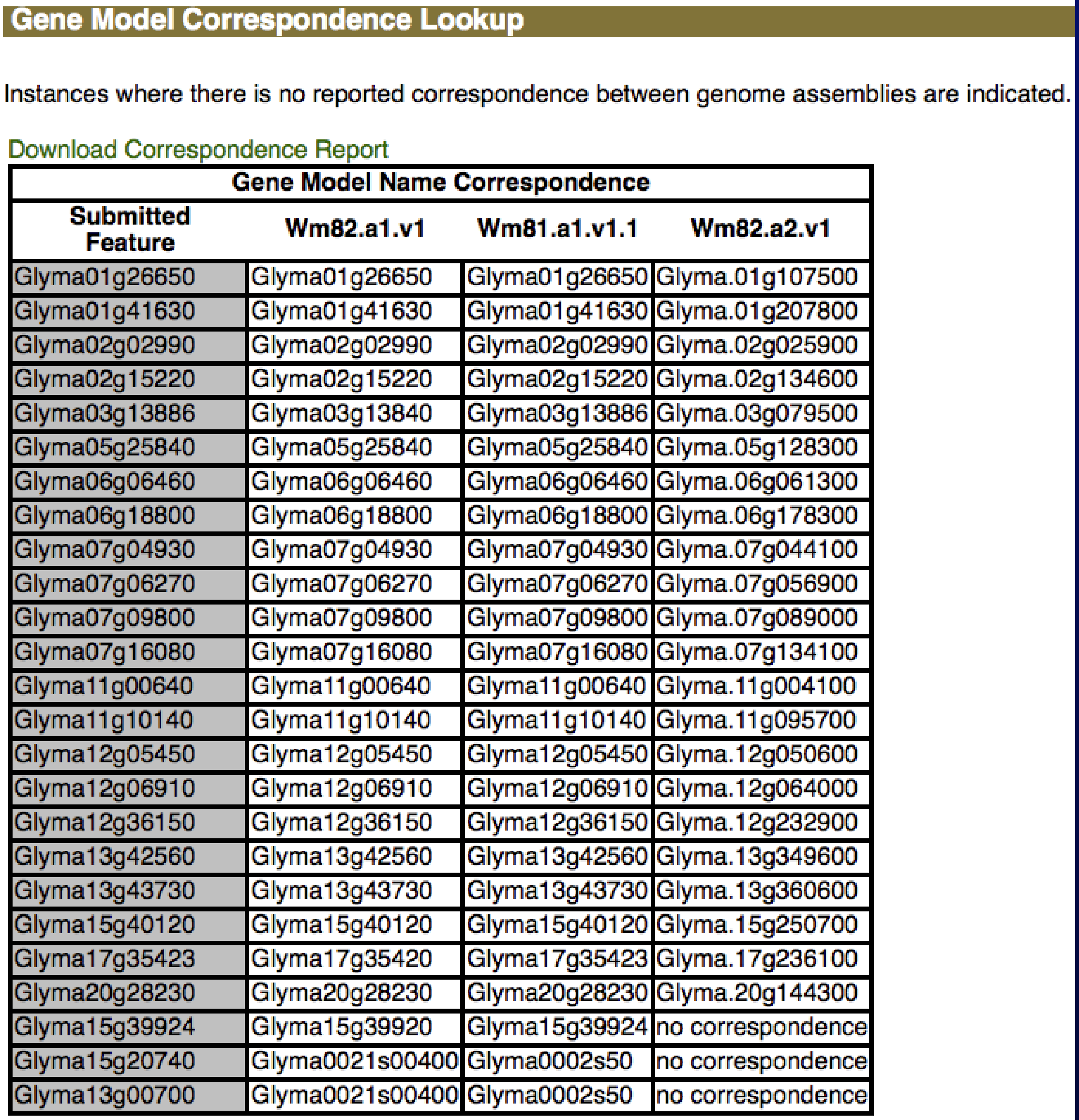
Insert Gene List:
`(One per line)`

Instructions

Enter a list of gene model names into this box, one name per line. Alternatively a pre-made list can be loaded by clicking on the green "Click to Load From File" text below.

Click to Load From File

Click for Example Data



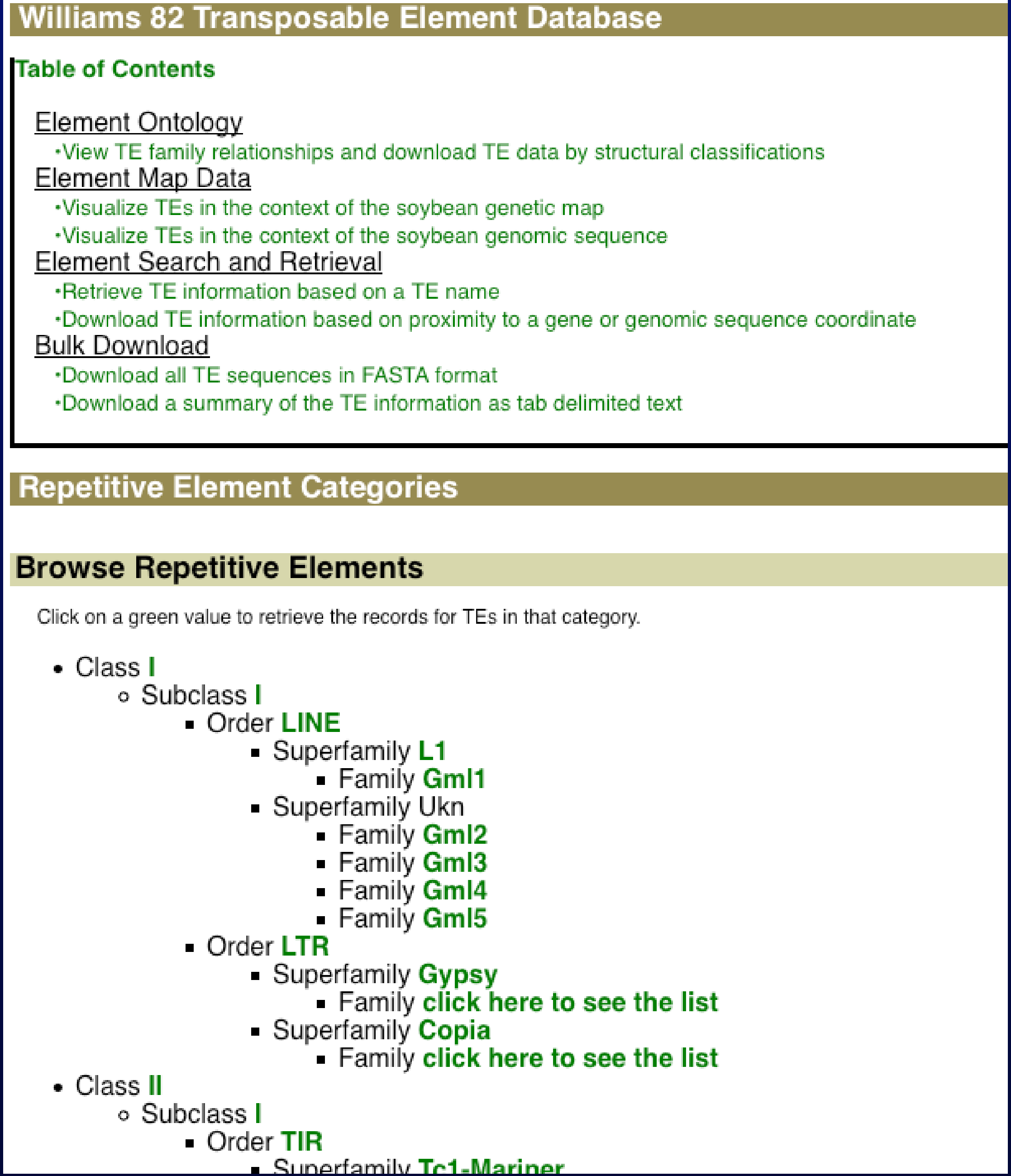
Gene Model Correspondence Lookup

Instances where there is no reported correspondence between genome assemblies are indicated.

Download Correspondence Report

Gene Model Name Correspondence

Submitted Feature	Wm82.a1.v1	Wm81.a1.v1.1	Wm82.a2.v1
Glyma01g26650	Glyma01g26650	Glyma01g26650	Glyma.01g107500
Glyma01g41630	Glyma01g41630	Glyma01g41630	Glyma.01g207800
Glyma02g02990	Glyma02g02990	Glyma02g02990	Glyma.02g025900
Glyma02g15220	Glyma02g15220	Glyma02g15220	Glyma.02g134600
Glyma03g13886	Glyma03g13840	Glyma03g13886	Glyma.03g079500
Glyma05g25840	Glyma05g25840	Glyma05g25840	Glyma.05g128300
Glyma06g06460	Glyma06g06460	Glyma06g06460	Glyma.06g061300
Glyma06g18800	Glyma06g18800	Glyma06g18800	Glyma.06g178300
Glyma07g04930	Glyma07g04930	Glyma07g04930	Glyma.07g044100
Glyma07g06270	Glyma07g06270	Glyma07g06270	Glyma.07g056900
Glyma07g09800	Glyma07g09800	Glyma07g09800	Glyma.07g089000
Glyma07g16080	Glyma07g16080	Glyma07g16080	Glyma.07g134100
Glyma11g00640	Glyma11g00640	Glyma11g00640	Glyma.11g004100
Glyma11g10140	Glyma11g10140	Glyma11g10140	Glyma.11g095700
Glyma12g05450	Glyma12g05450	Glyma12g05450	Glyma.12g050600
Glyma12g06910	Glyma12g06910	Glyma12g06910	Glyma.12g064000
Glyma12g36150	Glyma12g36150	Glyma12g36150	Glyma.12g232900
Glyma13g42560	Glyma13g42560	Glyma13g42560	Glyma.13g349600
Glyma13g43730	Glyma13g43730	Glyma13g43730	Glyma.13g360600
Glyma15g40120	Glyma15g40120	Glyma15g40120	Glyma.15g250700
Glyma17g35423	Glyma17g35420	Glyma17g35423	Glyma.17g236100
Glyma20g28230	Glyma20g28230	Glyma20g28230	Glyma.20g144300
Glyma15g39924	Glyma15g39920	Glyma15g39924	no correspondence
Glyma15g20740	Glyma0021s00400	Glyma0002s50	no correspondence
Glyma13g00700	Glyma0021s00400	Glyma0002s50	no correspondence



Williams 82 Transposable Element Database

Table of Contents

Element Ontology
- View TE family relationships and download TE data by structural classifications

Element Map Data
- Visualize TEs in the context of the soybean genetic map
- Visualize TEs in the context of the soybean genomic sequence

Element Search and Retrieval
- Retrieve TE information based on a TE name
- Download TE information based on proximity to a gene or genomic sequence coordinate

Bulk Download
- Download all TE sequences in FASTA format
- Download a summary of the TE information as tab delimited text

Repetitive Element Categories

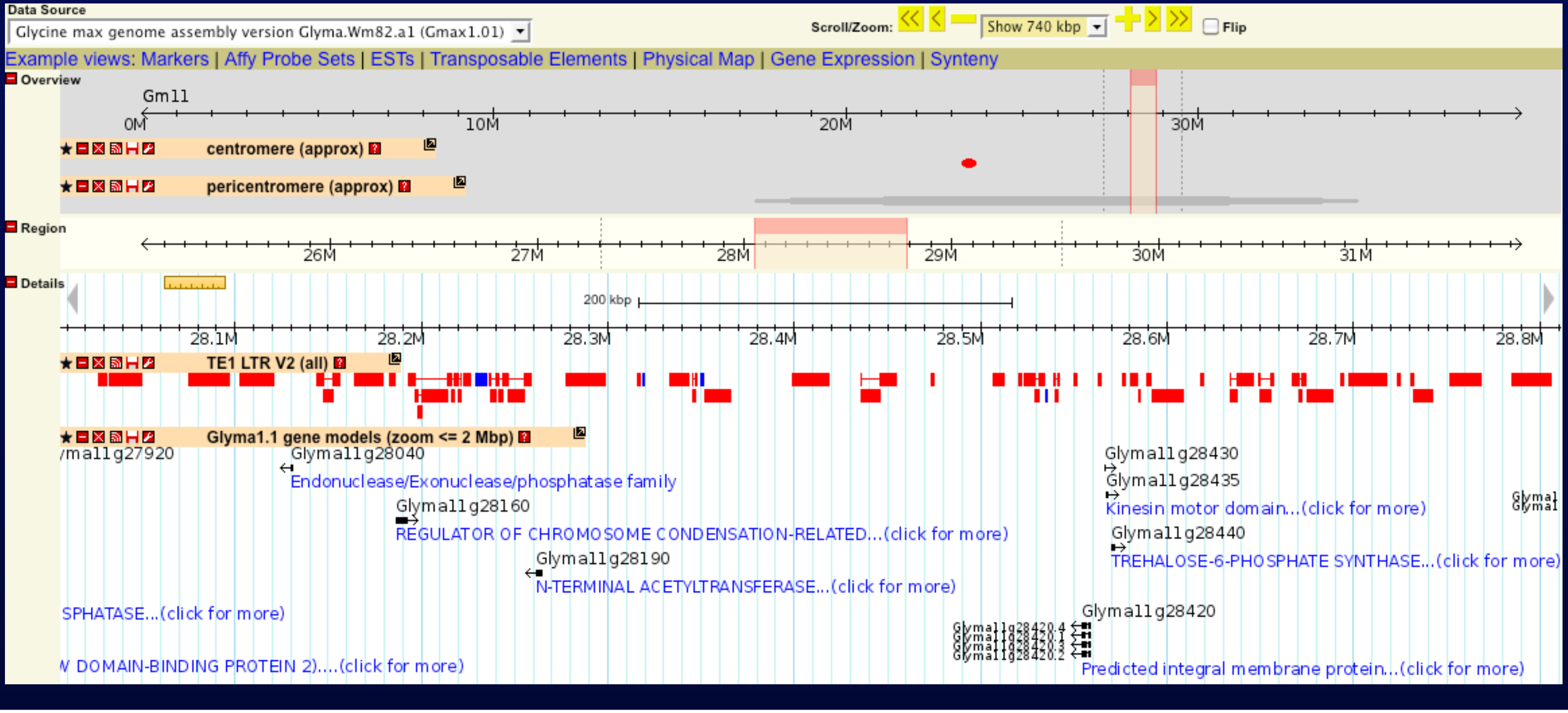
Browse Repetitive Elements

Click on a green value to retrieve the records for TEs in that category.

- Class **I**
  - Subclass **I**
    - Order **LINE**
      - Superfamily **L1**
        - Family **Gml1**
      - Superfamily Ukn
        - Family **Gml2**
        - Family **Gml3**
        - Family **Gml4**
        - Family **Gml5**
    - Order **LTR**
      - Superfamily **Gypsy**
        - Family **click here to see the list**
      - Superfamily **Copia**
        - Family **click here to see the list**
- Class **II**
  - Subclass **I**
    - Order **TIR**
      - Superfamily **Tc1-Mariner**

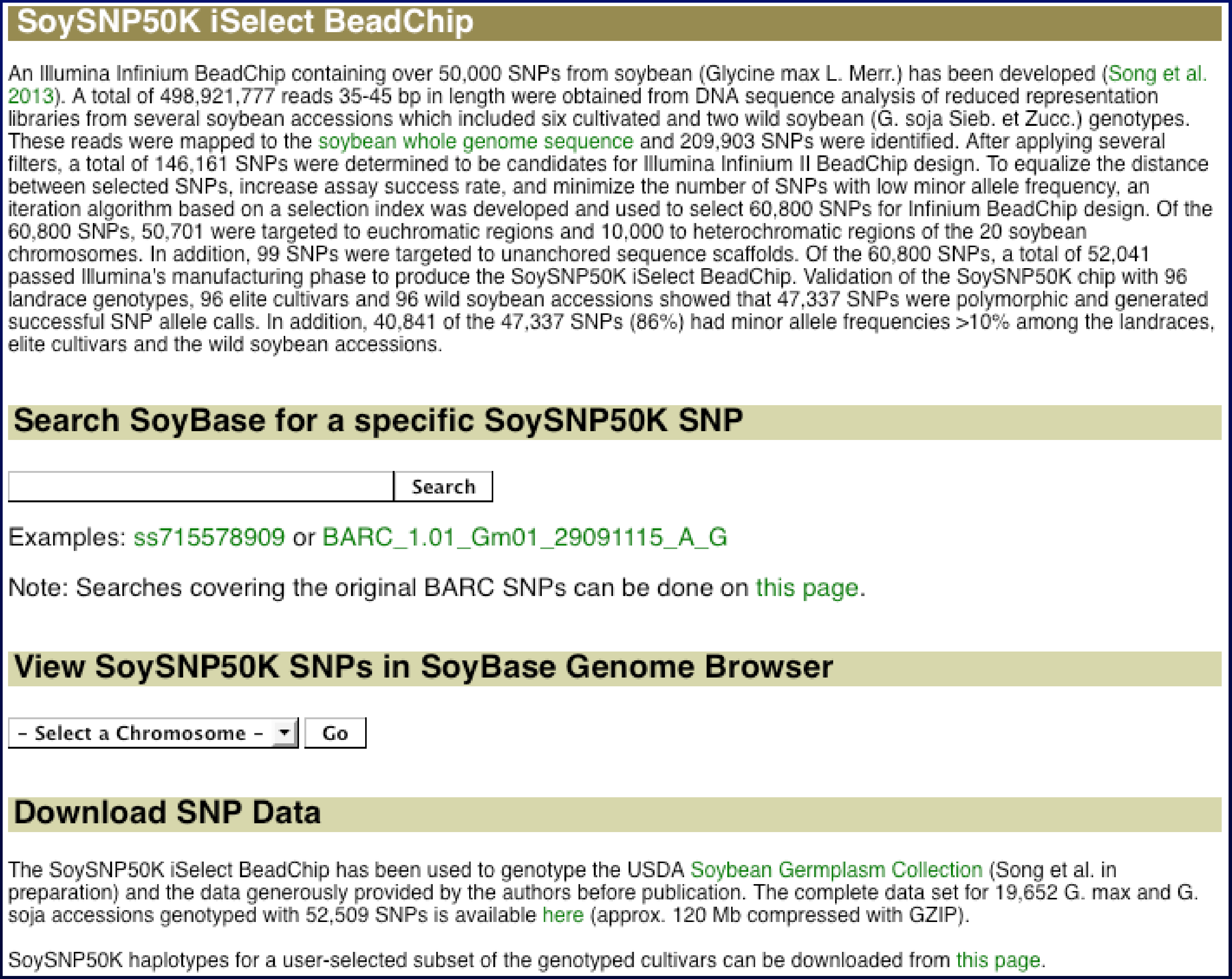
Data Source: Glycine max genome assembly version Glyma.Wm82.a1 (Gmax1.01)

Scroll/Zoom: Show 39.17 Mbp ☐ Flip



Data Source: Glycine max genome assembly version Glyma.Wm82.a1 (Gmax1.01)

Scroll/Zoom: Show 740 kbp ☐ Flip



SoySNP50K iSelect BeadChip

An Illumina Infinium BeadChip containing over 50,000 SNPs from soybean (Glycine max L. Merr.) has been developed (Song et al. 2013). A total of 498,921,777 reads 35-45 bp in length were obtained from DNA sequence analysis of reduced representation libraries from several soybean accessions which included six cultivated and two wild soybean (G. soja Sieb. et Zucc.) genotypes. These reads were mapped to the soybean whole genome sequence and 209,903 SNPs were identified. After applying several filters, a total of 146,161 SNPs were determined to be candidates for Illumina Infinium II BeadChip design. To equalize the distance between selected SNPs, increase assay success rate, and minimize the number of SNPs with low minor allele frequency, an iteration algorithm based on a selection index was developed and used to select 60,800 SNPs for Infinium BeadChip design. Of the 60,800 SNPs, 50,701 were targeted to euchromatic regions and 10,000 to heterochromatic regions of the 20 soybean chromosomes. In addition, 99 SNPs were targeted to unanchored sequence scaffolds. Of the 60,800 SNPs, a total of 52,041 passed Illumina's manufacturing phase to produce the SoySNP50K iSelect BeadChip. Validation of the SoySNP50K chip with 96 landrace genotypes, 96 elite cultivars and 96 wild soybean accessions showed that 47,337 SNPs were polymorphic and generated successful SNP allele calls. In addition, 40,841 of the 47,337 SNPs (86%) had minor allele frequencies >10% among the landraces, elite cultivars and the wild soybean accessions.

Search SoyBase for a specific SoySNP50K SNP

Examples: ss715578909 or BARC_1.01_Gm01_29091115_A_G

Note: Searches covering the original BARC SNPs can be done on this page.

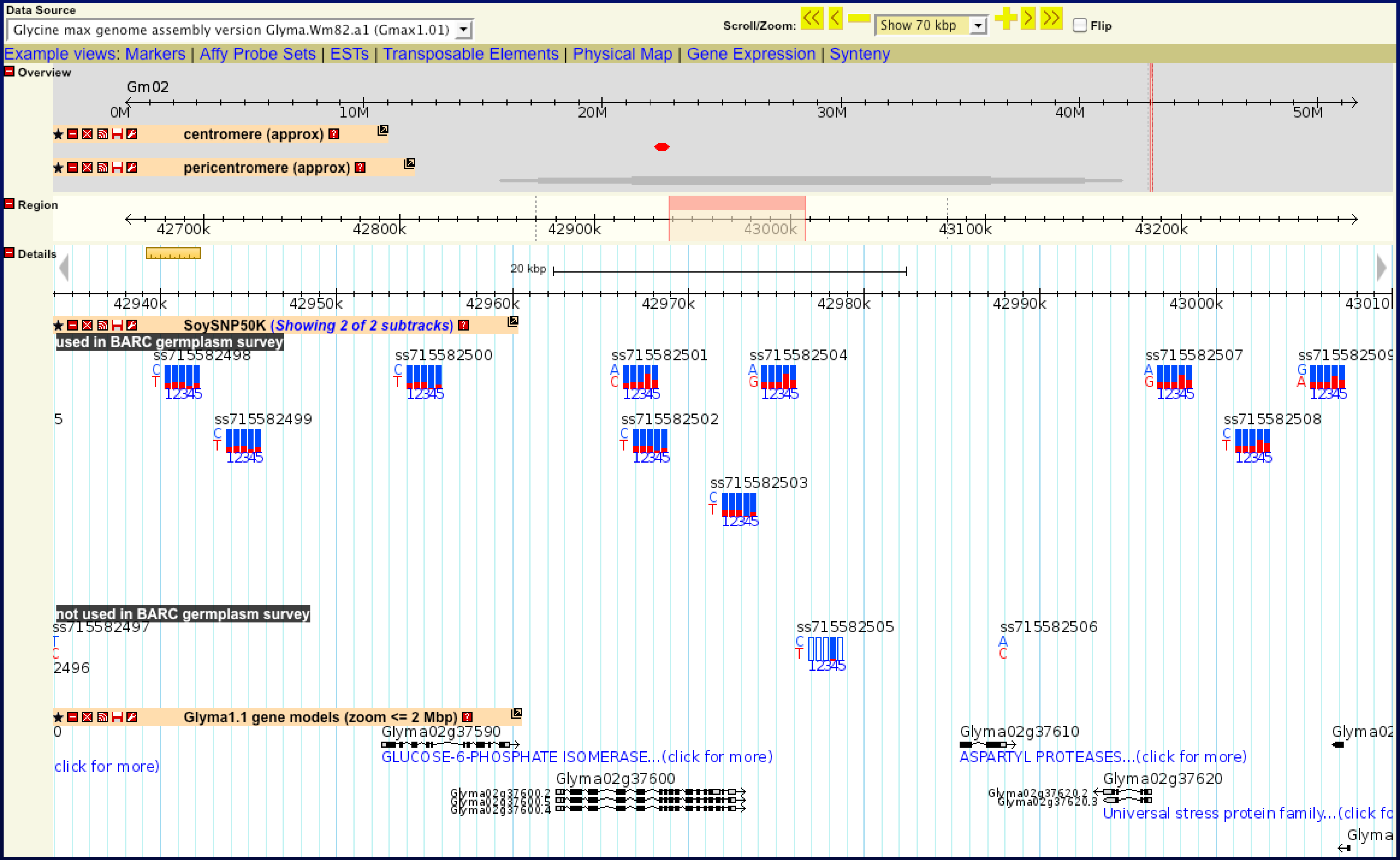
View SoySNP50K SNPs in SoyBase Genome Browser

Download SNP Data

The SoySNP50K iSelect BeadChip has been used to genotype the USDA Soybean Germplasm Collection (Song et al. in preparation) and the data generously provided by the authors before publication. The complete data set for 19,652 G. max and G. soja accessions genotyped with 52,509 SNPs is available here (approx. 120 Mb compressed with GZIP).

SoySNP50K haplotypes for a user-selected subset of the genotyped cultivars can be downloaded from this page.

SoySNP50K SNPs in Genome Browser

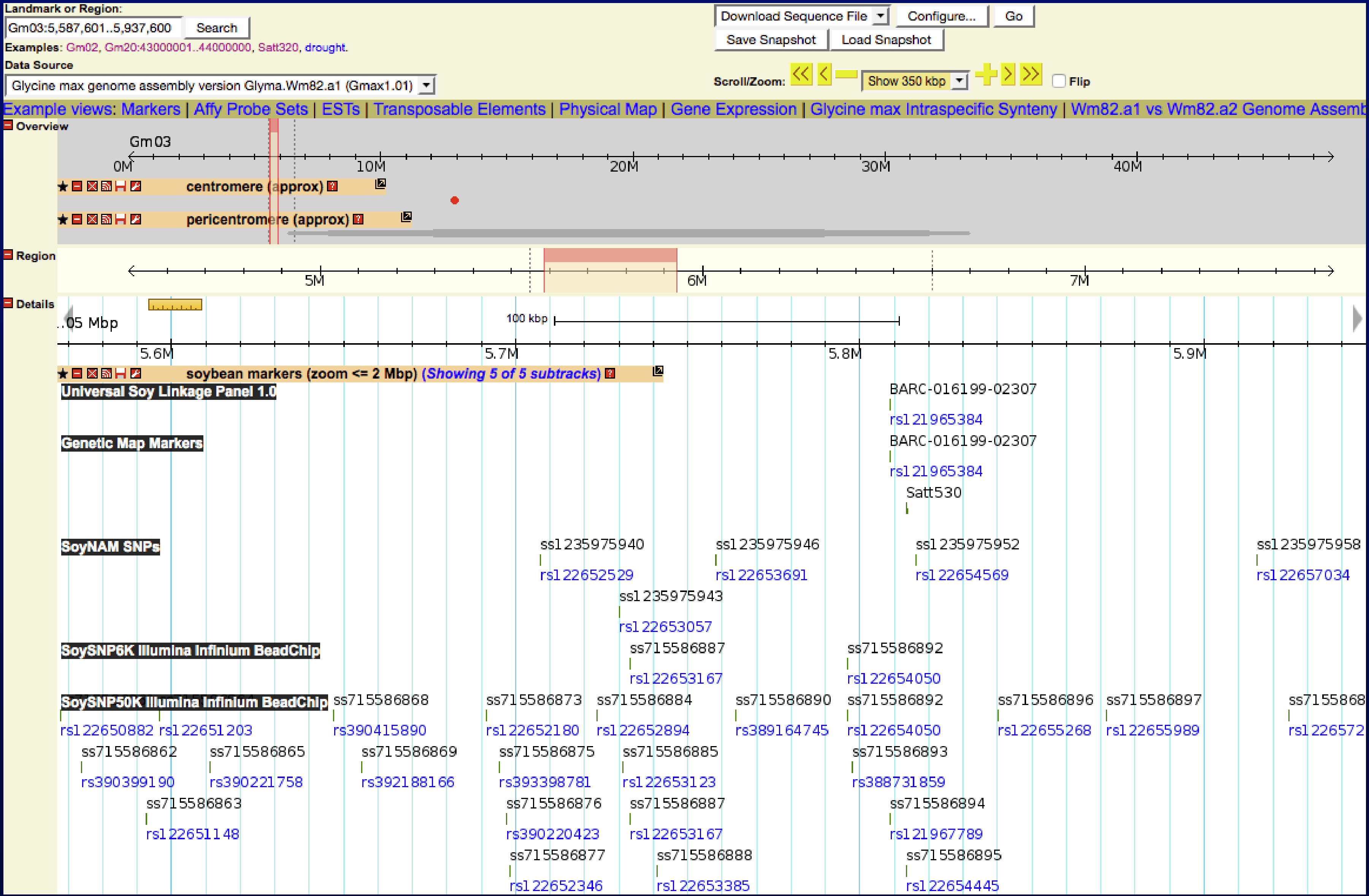


Data Source: Glycine max genome assembly version Glyma.Wm82.a1 (Gmax1.01)

Scroll/Zoom: Show 70 kbp

☐ Flip

Markers in Genome Browser



Landmark or Region: Gm03:5,587,601..5,937,600

Examples: Gm02, Gm20:43000001..44000000, Satt320, drought.

Data Source: Glycine max genome assembly version Glyma.Wm82.a1 (Gmax1.01)

Scroll/Zoom: Show 350 kbp ☐ Flip

Example views: Markers | Affy Probe Sets | ESTs | Transposable Elements | Physical Map | Gene Expression | Glycine max Intraspecific Synteny | Wm82.a1 vs Wm82.a2 Genome Assemb

soybean markers (zoom <= 2 Mbp) *(Showing 5 of 5 subtracks)*

Universal Soy Linkage Panel 1.0

BARC-016199-02307 rs121965384

Genetic Map Markers

BARC-016199-02307 rs121965384

Satt530

SoyNAM SNPs

ss1235975940 rs122652529

ss1235975946 rs122653691

ss1235975952 rs122654569

ss1235975958 rs122657034

ss1235975943 rs122653057

SoySNP6K Illumina Infinium BeadChip

ss715586887 rs122653167

ss715586892 rs122654050

SoySNP50K Illumina Infinium BeadChip

rs122650882

rs122651203

ss715586868 rs390415890

ss715586873 rs122652180

ss715586884 rs122652894

ss715586890 rs389164745

ss715586892 rs122654050

ss715586896 rs122655268

ss715586897 rs122655989

ss7155868 rs1226572

ss715586862 rs390399190

ss715586865 rs390221758

ss715586869 rs392188166

ss715586875 rs393398781

ss715586885 rs122653123

ss715586893 rs388731859

ss715586863 rs122651148

ss715586876 rs390220423

ss715586887 rs122653167

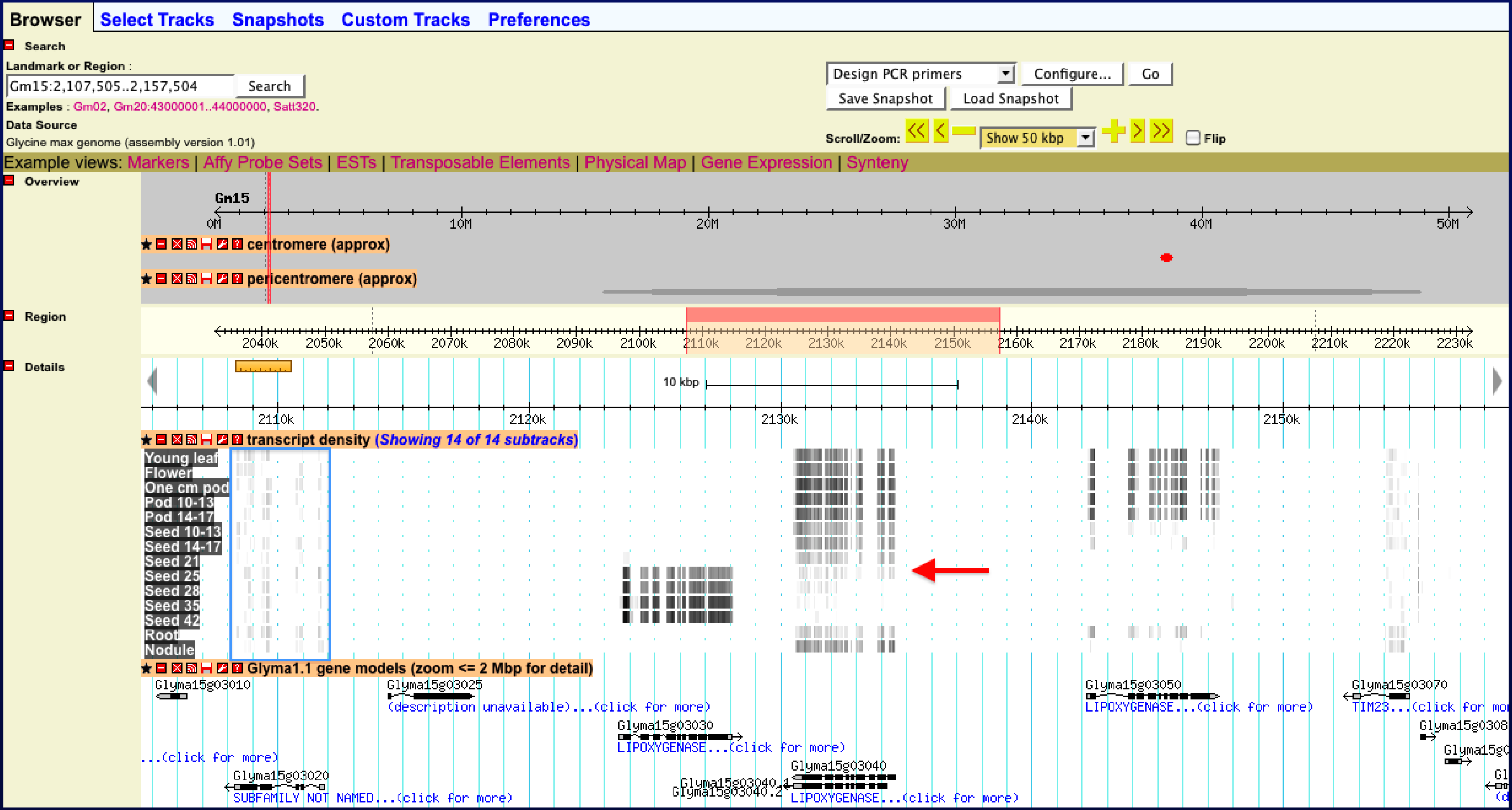
ss715586894 rs121967789

ss715586877 rs122652346

ss715586888 rs122653385

ss715586895 rs122654445

RNA-seq Expression Data on the Genome Sequence



Landmark or Region : Gm15:2,107,505..2,157,504

Examples : Gm02, Gm20:43000001..44000000, Satt320.

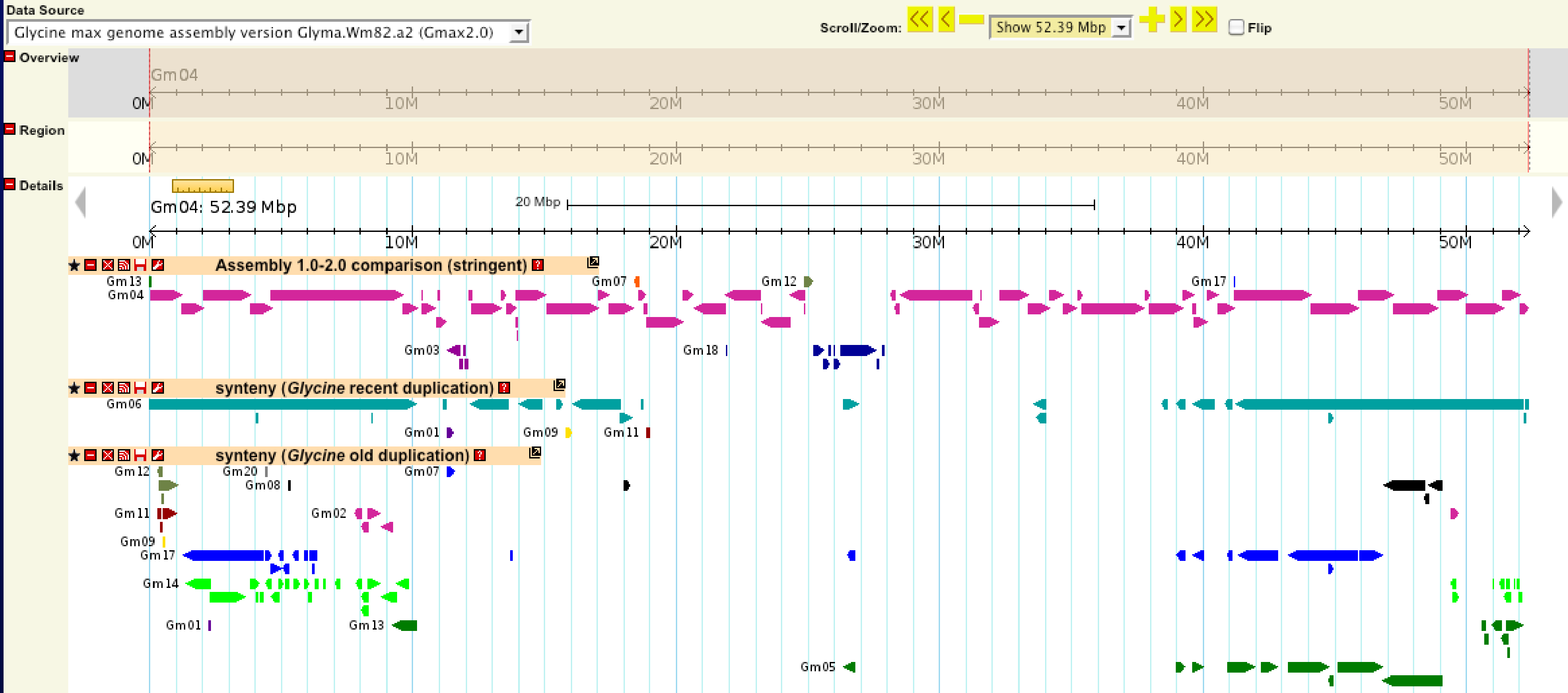
Data Source

Glycine max genome (assembly version 1.01)

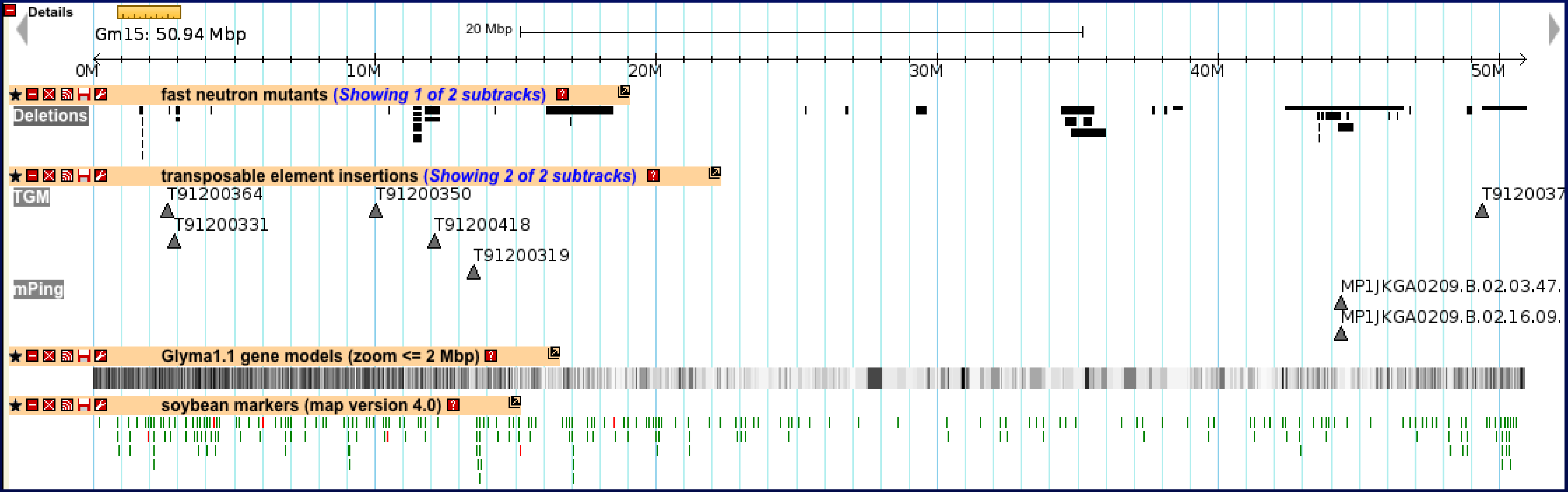
Example views: Markers | Affy Probe Sets | ESTs | Transposable Elements | Physical Map | Gene Expression | Synteny

Scroll/Zoom: Show 50 kbp

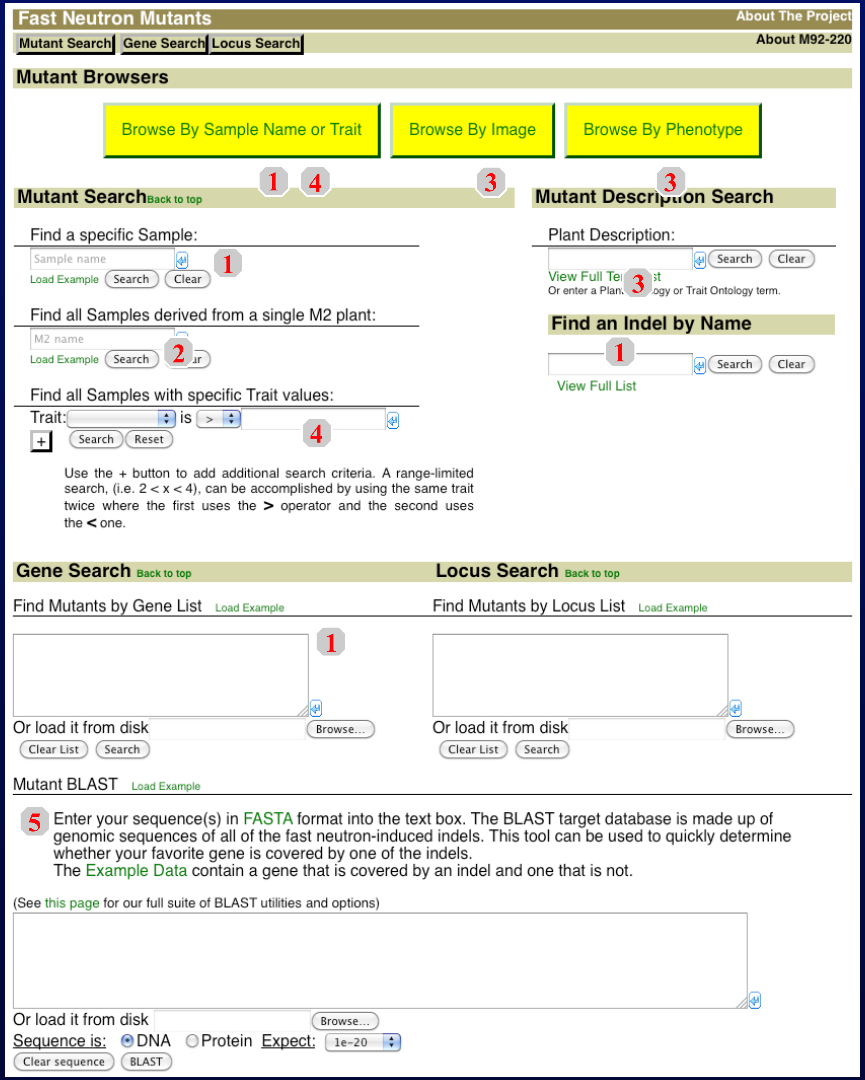
Show Large-scale Genome Structure Relationships



Transposable Element Insertions and Fast Neutron-induced Indels Displayed in Genome Browser



This shows the primary entry to the Fast Neutron Mutant Population page. From here mutants can be selected by sample name, gene model, locus name or indel name[1], pedigree[2], visual[3] or biochemical[4] phenotype, or sequence similarity to a query sequence[5].

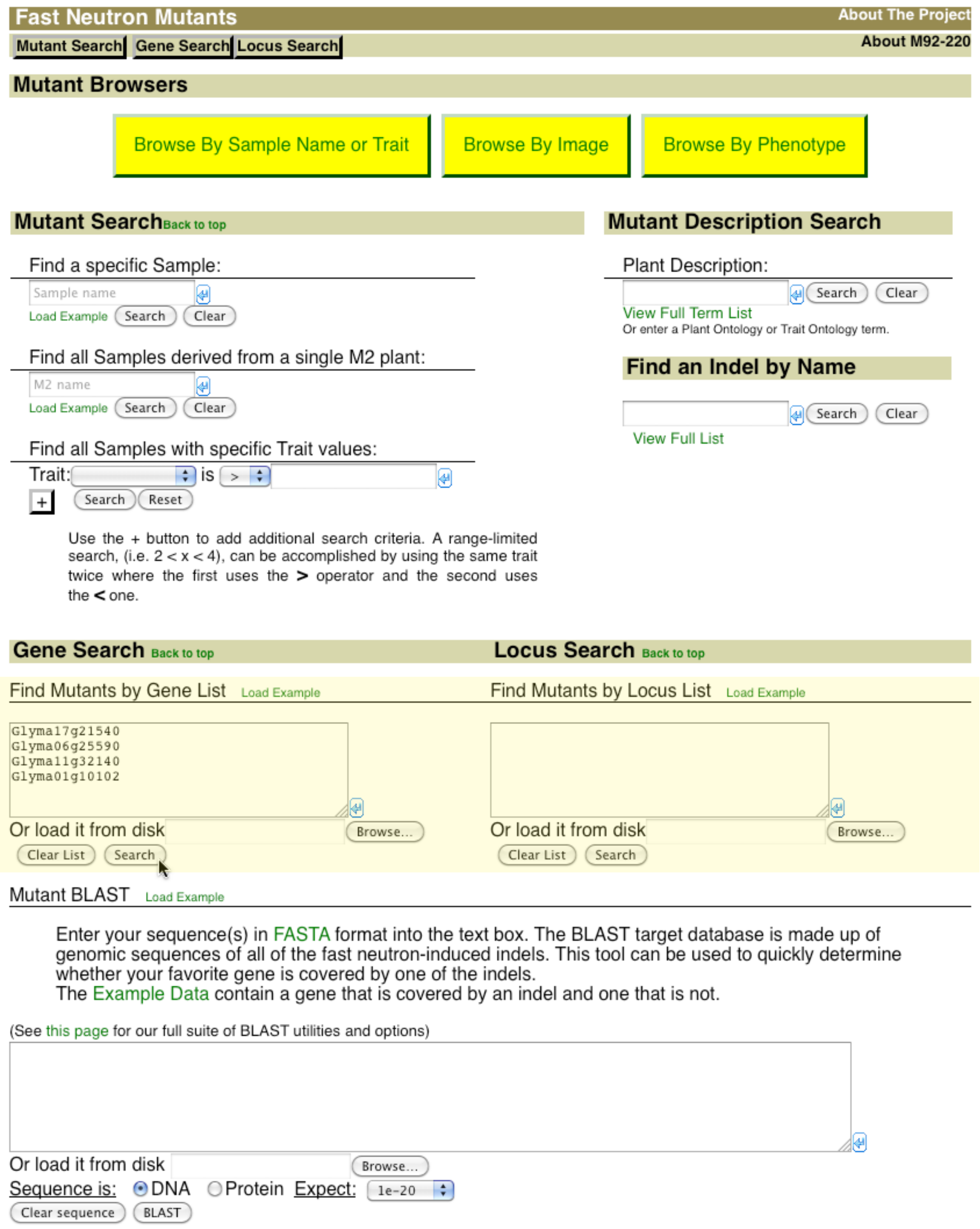


Use the + button to add additional search criteria. A range-limited search, (i.e. 2 < x < 4), can be accomplished by using the same trait twice where the first uses the **>** operator and the second uses the **<** one.

Enter your sequence(s) in FASTA format into the text box. The BLAST target database is made up of genomic sequences of all of the fast neutron-induced indels. This tool can be used to quickly determine whether your favorite gene is covered by one of the indels.
The Example Data contain a gene that is covered by an indel and one that is not.

(See this page for our full suite of BLAST utilities and options)

A list of gene or locus names can be submitted to determine if any are covered by an indel.



Or enter a Plant Ontology or Trait Ontology term.

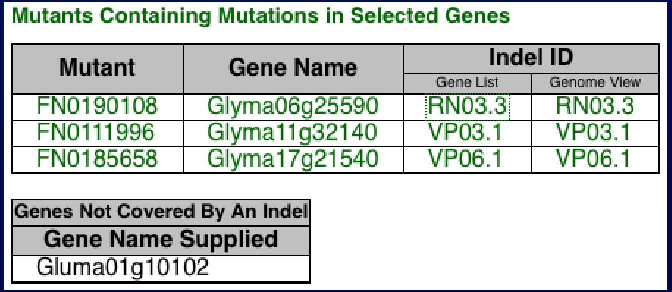
Use the + button to add additional search criteria. A range-limited search, (i.e. 2 < x < 4), can be accomplished by using the same trait twice where the first uses the **>** operator and the second uses the **<** one.

```
Glyma17g21540
Glyma06g25590
Glyma11g32140
Glyma01g10102
```

Enter your sequence(s) in FASTA format into the text box. The BLAST target database is made up of genomic sequences of all of the fast neutron-induced indels. This tool can be used to quickly determine whether your favorite gene is covered by one of the indels.
The Example Data contain a gene that is covered by an indel and one that is not.

(See this page for our full suite of BLAST utilities and options)

Sequence is: ● DNA ○ Protein Expect: 1e-20

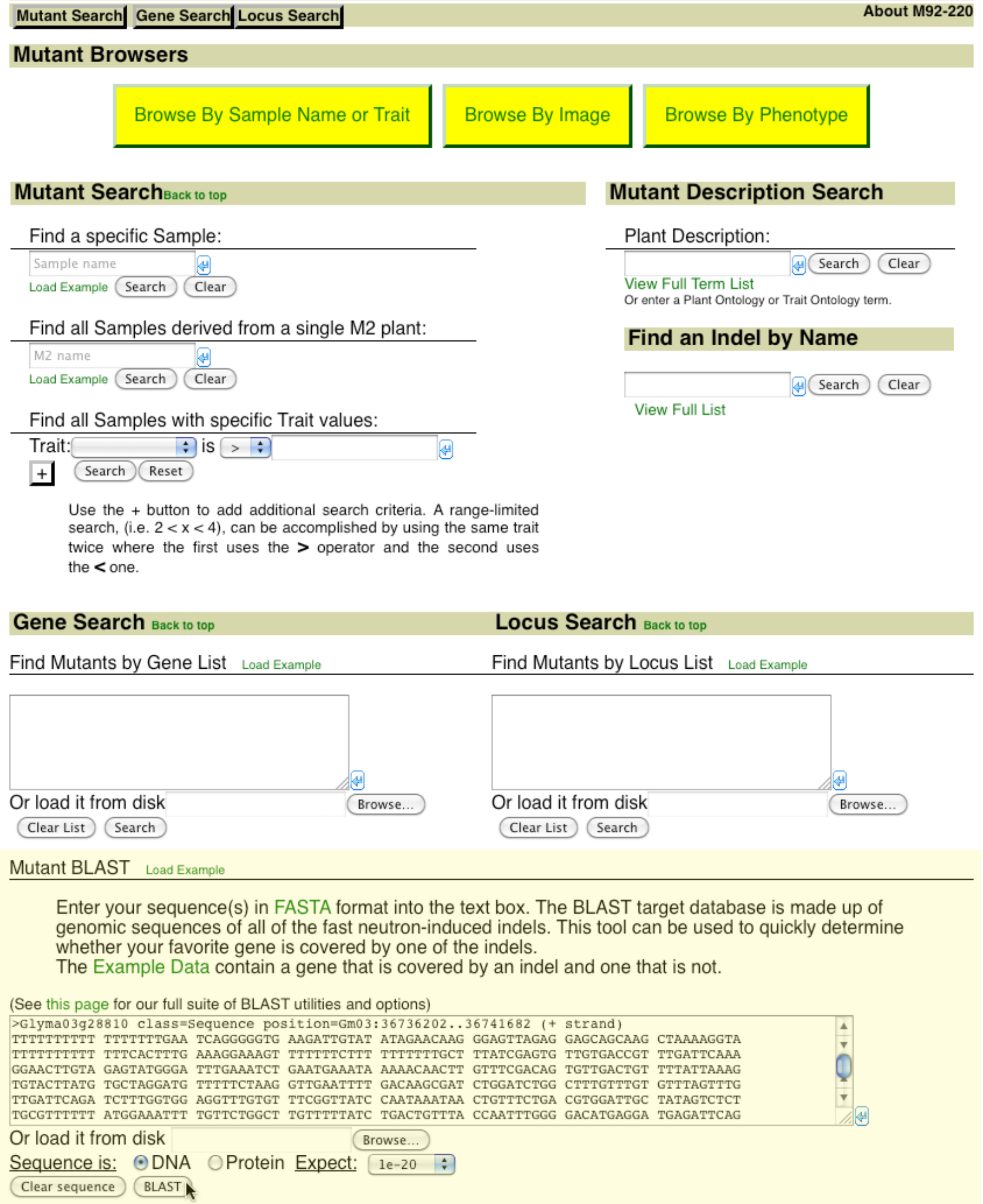


Mutants Containing Mutations in Selected Genes

Mutant	Gene Name	Indel ID	
		Gene List	Genome View
FN0190108	Glyma06g25590	RN03.3	RN03.3
FN0111996	Glyma11g32140	VP03.1	VP03.1
FN0185658	Glyma17g21540	VP06.1	VP06.1

Genes Not Covered By An Indel
Gene Name Supplied
Gluma01g10102

One or more FASTA sequences can be BLASTed against the indel sequences to determine if any are covered by an indel.



Fast Neutron Mutants

Mutant Browsers

Mutant Search

Find a specific Sample:

Find all Samples derived from a single M2 plant:

Find all Samples with specific Trait values:

Trait: is >

Use the + button to add additional search criteria. A range-limited search, (i.e. 2 < x < 4), can be accomplished by using the same trait twice where the first uses the **>** operator and the second uses the **<** one.

Mutant Description Search

Plant Description:

Or enter a Plant Ontology or Trait Ontology term.

Find an Indel by Name

Gene Search

Find Mutants by Gene List

Or load it from disk

Locus Search

Find Mutants by Locus List

Or load it from disk

Mutant BLAST

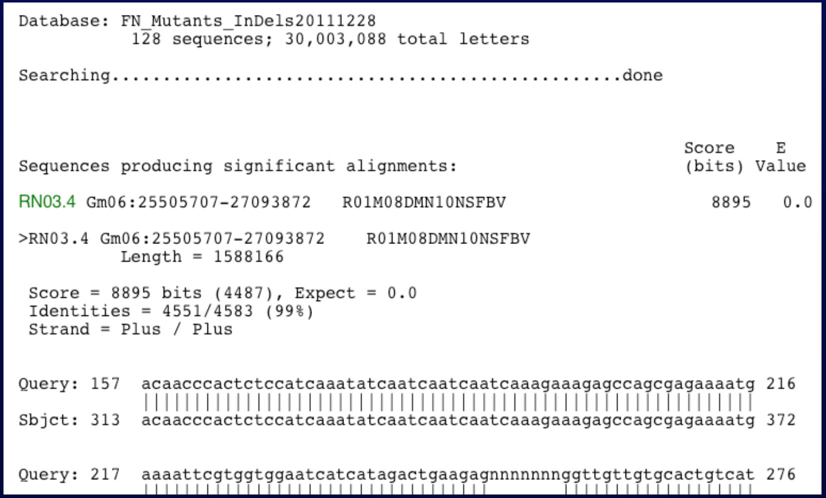
Enter your sequence(s) in FASTA format into the text box. The BLAST target database is made up of genomic sequences of all of the fast neutron-induced indels. This tool can be used to quickly determine whether your favorite gene is covered by one of the indels.
The Example Data contain a gene that is covered by an indel and one that is not.

(See this page for our full suite of BLAST utilities and options)

```
>Glyma03g28810 class=Sequence position=Gm03:36736202..36741682 (+ strand)
TTTTTTTTTT TTTTTTTGAA TCAGGGGGTG AAGATTGTAT ATAGAACAAG GGAGTTAGAG GAGCAGCAAG CTAAAAGGTA
TTTTTTTTTT TTTCACTTTG AAAGGAAAGT TTTTTTCTTT TTTTTTTGCT TTATCGAGTG TTGTGACCGT TTGATTCAAA
GGAACTTGTA GAGTATGGGA TTTGAAATCT GAATGAAATA AAAACAACTT GTTTCGACAG TGTTGACTGT TTTATTAAAG
TGTACTTATG TGCTAGGATG TTTTTCTAAG GTTGAATTTT GACAAGCGAT CTGGATCTGG CTTTGTTTGT GTTTAGTTTG
TTGATTCAGA TCTTTGGTGG AGGTTTGTGT TTCGGTTATC CAATAAATAA CTGTTTCTGA CGTGGATTGC TATAGTCTCT
TGCGTTTTTT ATGGAAATTT TGTTCTGGCT TGTTTTTATC TGACTGTTTA CCAATTTGGG GACATGAGGA TGAGATTCAG
```

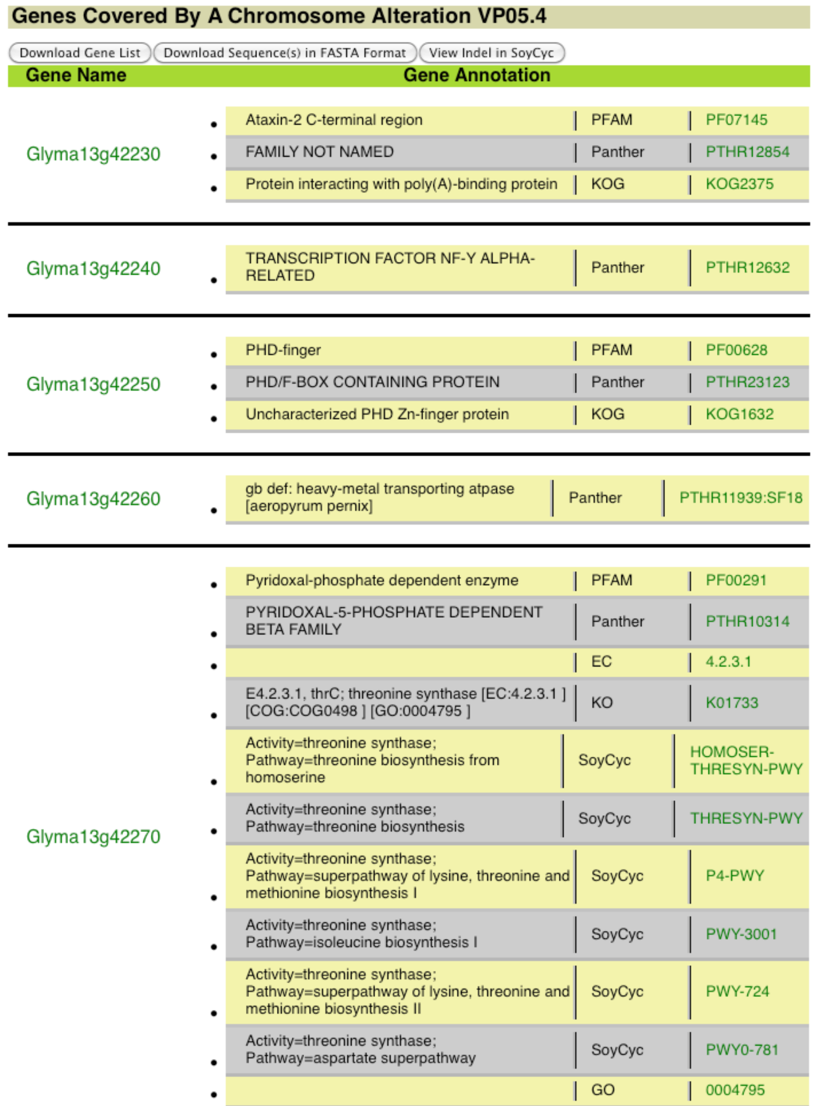
Or load it from disk

Sequence is: ● DNA ○ Protein Expect: 1e-20



```
Database: FN_Mutants_InDels20111228
           128 sequences; 30,003,088 total letters

Searching..................................................done



                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

RN03.4 Gm06:25505707-27093872    R01M08DMN10NSFBV                   8895    0.0

>RN03.4 Gm06:25505707-27093872    R01M08DMN10NSFBV
          Length = 1588166

 Score = 8895 bits (4487), Expect = 0.0
 Identities = 4551/4583 (99%)
 Strand = Plus / Plus


Query: 157   acaacccactctccatcaaatatcaatcaatcaatcaaagaaagagccagcgagaaaatg 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 313   acaacccactctccatcaaatatcaatcaatcaatcaaagaaagagccagcgagaaaatg 372


Query: 217   aaaattcgtggtggaatcatcatagactgaagagnnnnnnnggttgttgtgcactgtcat 276
              ||||||||||||||||||||||||||||||||||       |||||||||||||||||||
```

Genes Covered by an Indel



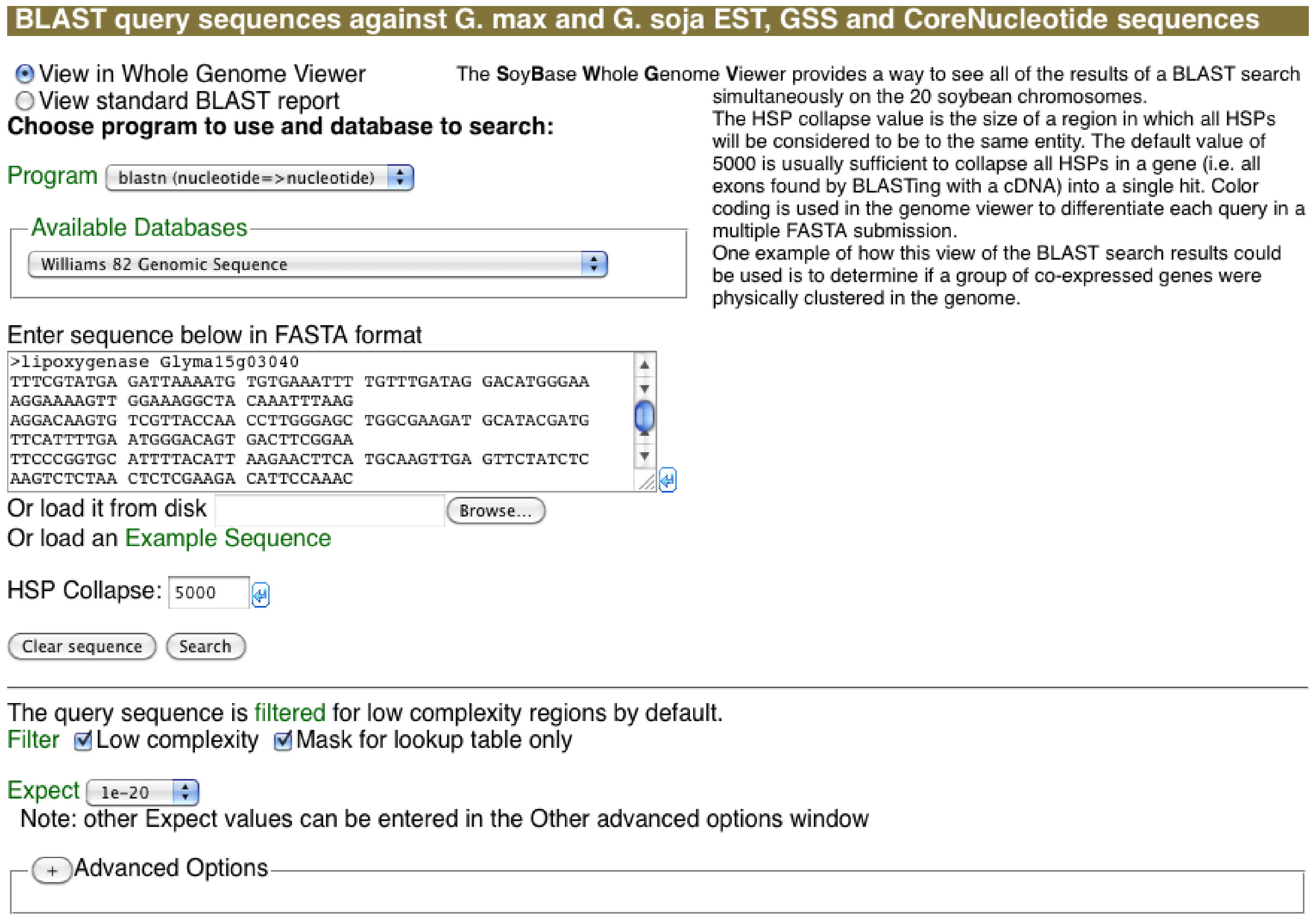
Genes Covered By A Chromosome Alteration VP05.4

Gene Name	Gene Annotation		
Glyma13g42230	Ataxin-2 C-terminal region	PFAM	PF07145
	FAMILY NOT NAMED	Panther	PTHR12854
	Protein interacting with poly(A)-binding protein	KOG	KOG2375
Glyma13g42240	TRANSCRIPTION FACTOR NF-Y ALPHA-RELATED	Panther	PTHR12632
Glyma13g42250	PHD-finger	PFAM	PF00628
	PHD/F-BOX CONTAINING PROTEIN	Panther	PTHR23123
	Uncharacterized PHD Zn-finger protein	KOG	KOG1632
Glyma13g42260	gb def: heavy-metal transporting atpase [aeropyrum pernix]	Panther	PTHR11939:SF18
Glyma13g42270	Pyridoxal-phosphate dependent enzyme	PFAM	PF00291
	PYRIDOXAL-5-PHOSPHATE DEPENDENT BETA FAMILY	Panther	PTHR10314
		EC	4.2.3.1
	E4.2.3.1, thrC; threonine synthase [EC:4.2.3.1] [COG:COG0498] [GO:0004795]	KO	K01733
	Activity=threonine synthase; Pathway=threonine biosynthesis from homoserine	SoyCyc	HOMOSER-THRESYN-PWY
	Activity=threonine synthase; Pathway=threonine biosynthesis	SoyCyc	THRESYN-PWY
	Activity=threonine synthase; Pathway=superpathway of lysine, threonine and methionine biosynthesis I	SoyCyc	P4-PWY
	Activity=threonine synthase; Pathway=isoleucine biosynthesis I	SoyCyc	PWY-3001
	Activity=threonine synthase; Pathway=superpathway of lysine, threonine and methionine biosynthesis II	SoyCyc	PWY-724
	Activity=threonine synthase; Pathway=aspartate superpathway	SoyCyc	PWY0-781
		GO	0004795
Glyma13g42280			

Selected Examples of Data Sets and Tools

- BLAST
- GO Enrichment Tool
- Data Downloads
- Searching SoyBase

Full BLAST Search



BLAST query sequences against G. max and G. soja EST, GSS and CoreNucleotide sequences

● View in Whole Genome Viewer
○ View standard BLAST report

Choose program to use and database to search:

The **S**oy**B**ase **W**hole **G**enome **V**iewer provides a way to see all of the results of a BLAST search simultaneously on the 20 soybean chromosomes.
The HSP collapse value is the size of a region in which all HSPs will be considered to be to the same entity. The default value of 5000 is usually sufficient to collapse all HSPs in a gene (i.e. all exons found by BLASTing with a cDNA) into a single hit. Color coding is used in the genome viewer to differentiate each query in a multiple FASTA submission.
One example of how this view of the BLAST search results could be used is to determine if a group of co-expressed genes were physically clustered in the genome.

Program: blastn (nucleotide=>nucleotide)

Available Databases: Williams 82 Genomic Sequence

Enter sequence below in FASTA format

```
>lipoxygenase Glyma15g03040
TTTCGTATGA GATTAAAATG TGTGAAATTT TGTTTGATAG GACATGGGAA
AGGAAAAGTT GGAAAGGCTA CAAATTTAAG
AGGACAAGTG TCGTTACCAA CCTTGGGAGC TGGCGAAGAT GCATACGATG
TTCATTTTGA ATGGGACAGT GACTTCGGAA
TTCCCGGTGC ATTTTACATT AAGAACTTCA TGCAAGTTGA GTTCTATCTC
AAGTCTCTAA CTCTCGAAGA CATTCCAAAC
```

Or load it from disk
Or load an Example Sequence

HSP Collapse: 5000

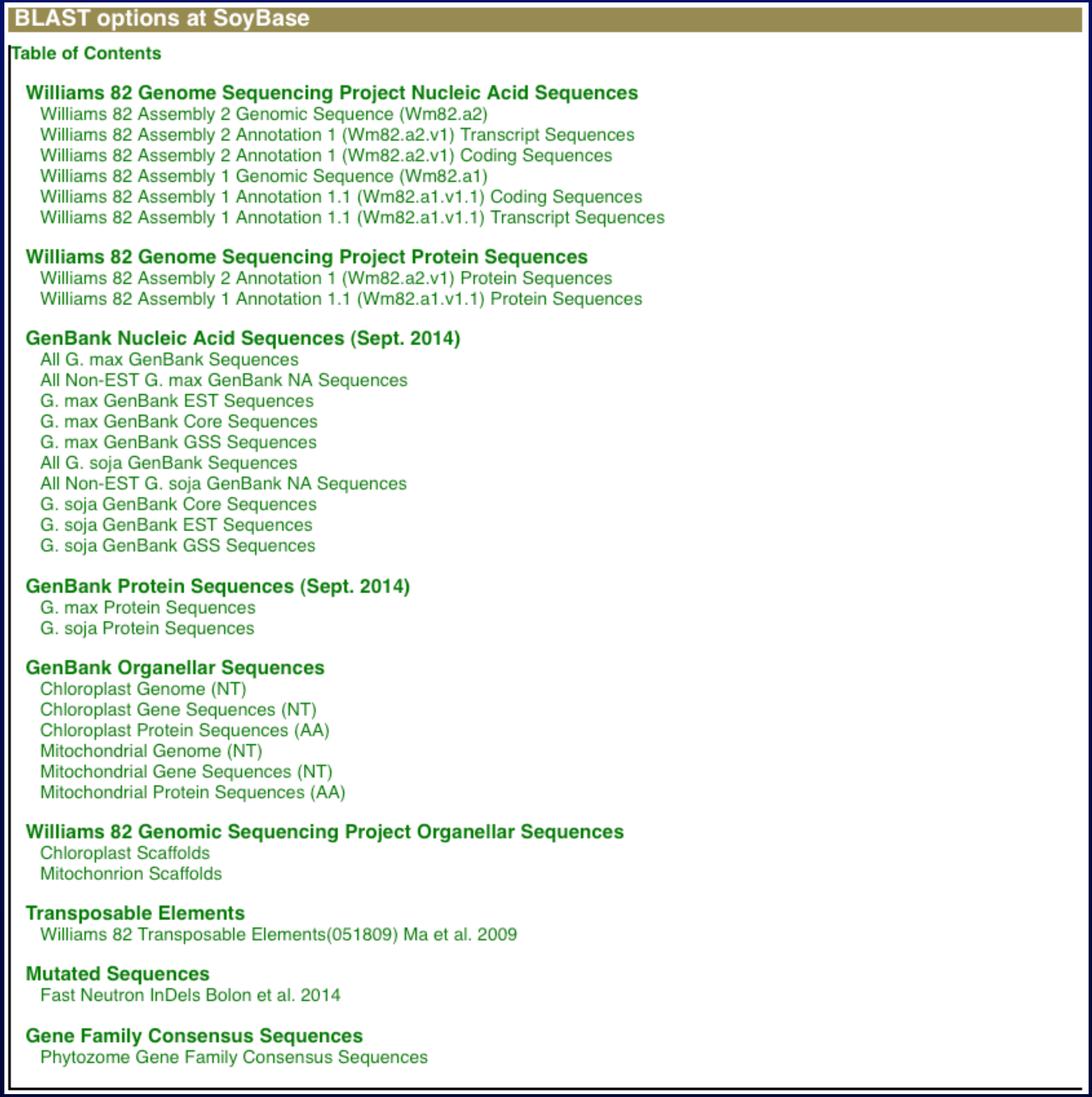
The query sequence is filtered for low complexity regions by default.

Filter ☑ Low complexity ☑ Mask for lookup table only

Expect: 1e-20

Note: other Expect values can be entered in the Other advanced options window

BLAST Targets at SoyBase



BLAST options at SoyBase

Table of Contents

Williams 82 Genome Sequencing Project Nucleic Acid Sequences
Williams 82 Assembly 2 Genomic Sequence (Wm82.a2)
*Williams 82 Assembly 2 Annotation 1 (Wm82.a2.v1) Transcript Sequences
Williams 82 Assembly 2 Annotation 1 (Wm82.a2.v1) Coding Sequences
Williams 82 Assembly 1 Genomic Sequence (Wm82.a1)
Williams 82 Assembly 1 Annotation 1.1 (Wm82.a1.v1.1) Coding Sequences
Williams 82 Assembly 1 Annotation 1.1 (Wm82.a1.v1.1) Transcript Sequences

Williams 82 Genome Sequencing Project Protein Sequences
*Williams 82 Assembly 2 Annotation 1 (Wm82.a2.v1) Protein Sequences
Williams 82 Assembly 1 Annotation 1.1 (Wm82.a1.v1.1) Protein Sequences

GenBank Nucleic Acid Sequences (Sept. 2014)
All G. max GenBank Sequences
All Non-EST G. max GenBank NA Sequences
G. max GenBank EST Sequences
G. max GenBank Core Sequences
G. max GenBank GSS Sequences
All G. soja GenBank Sequences
All Non-EST G. soja GenBank NA Sequences
G. soja GenBank Core Sequences
G. soja GenBank EST Sequences
G. soja GenBank GSS Sequences

GenBank Protein Sequences (Sept. 2014)
G. max Protein Sequences
G. soja Protein Sequences

GenBank Organellar Sequences
Chloroplast Genome (NT)
Chloroplast Gene Sequences (NT)
Chloroplast Protein Sequences (AA)
Mitochondrial Genome (NT)
Mitochondrial Gene Sequences (NT)
Mitochondrial Protein Sequences (AA)

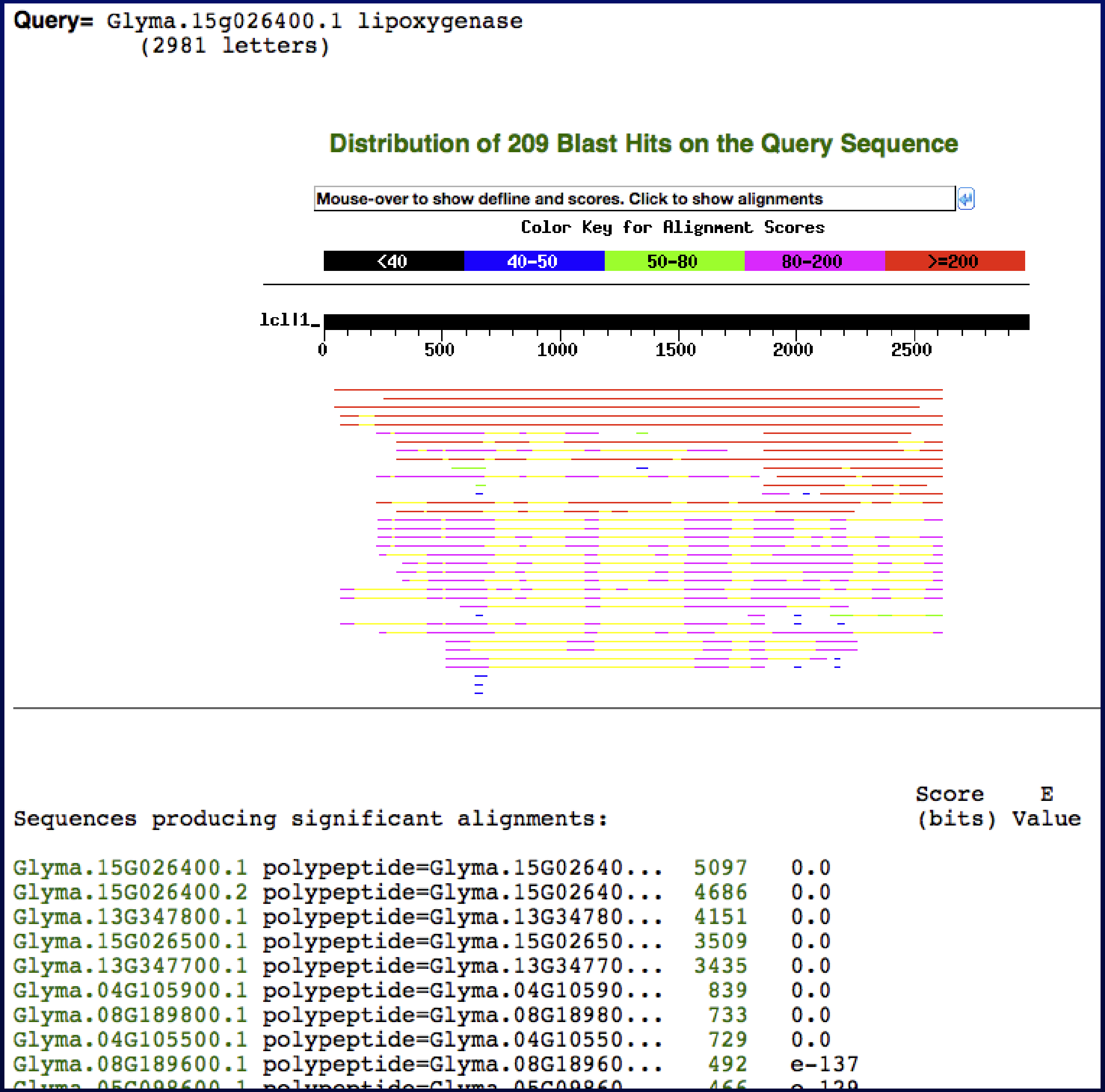
Williams 82 Genomic Sequencing Project Organellar Sequences
Chloroplast Scaffolds
Mitochonrion Scaffolds

Transposable Elements
Williams 82 Transposable Elements(051809) Ma et al. 2009

*Mutated Sequences**
Fast Neutron InDels Bolon et al. 2014

*Gene Family Consensus Sequences**
Phytozome Gene Family Consensus Sequences

BLAST Results

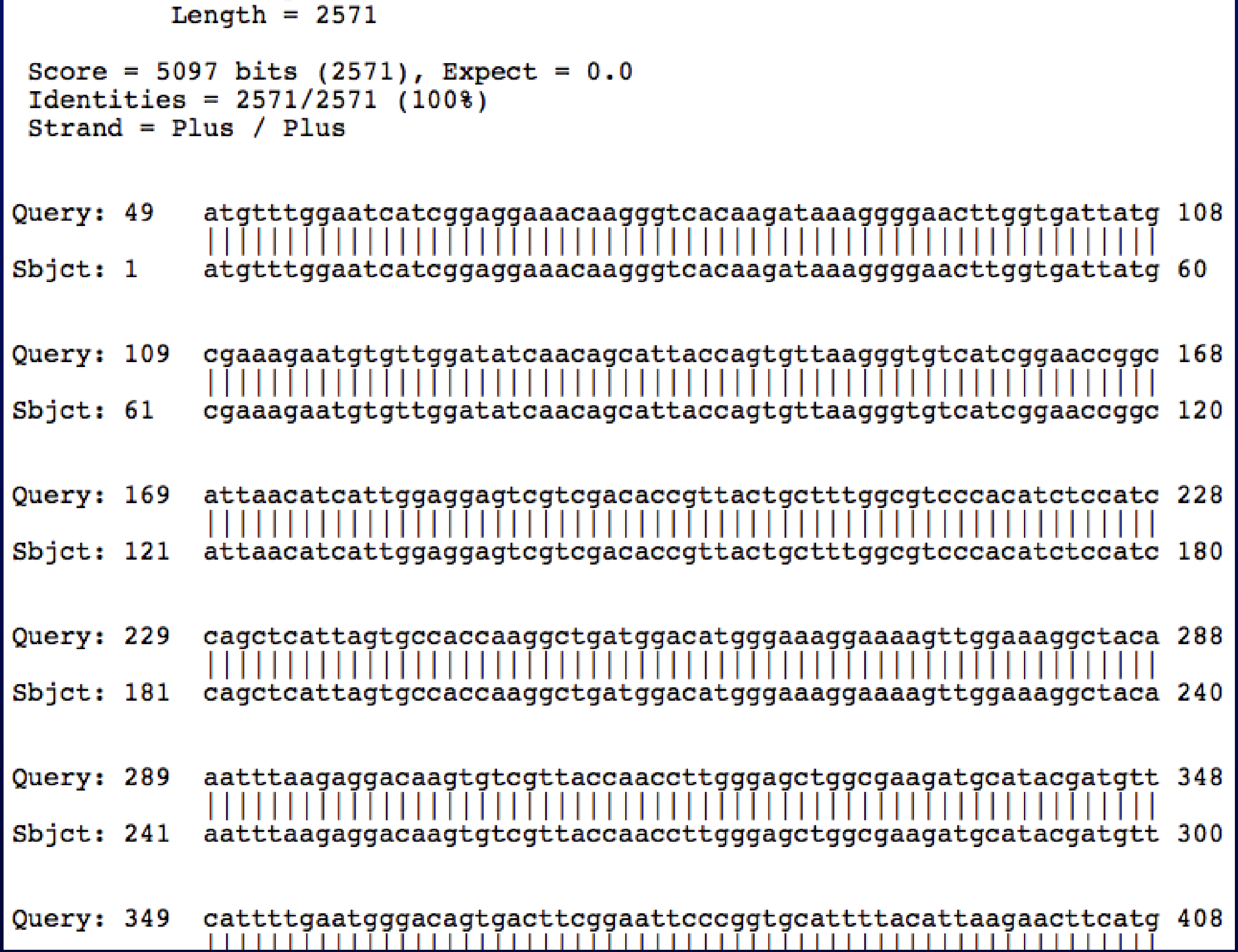


```
Query= Glyma.15g026400.1 lipoxygenase
         (2981 letters)
```

Distribution of 209 Blast Hits on the Query Sequence

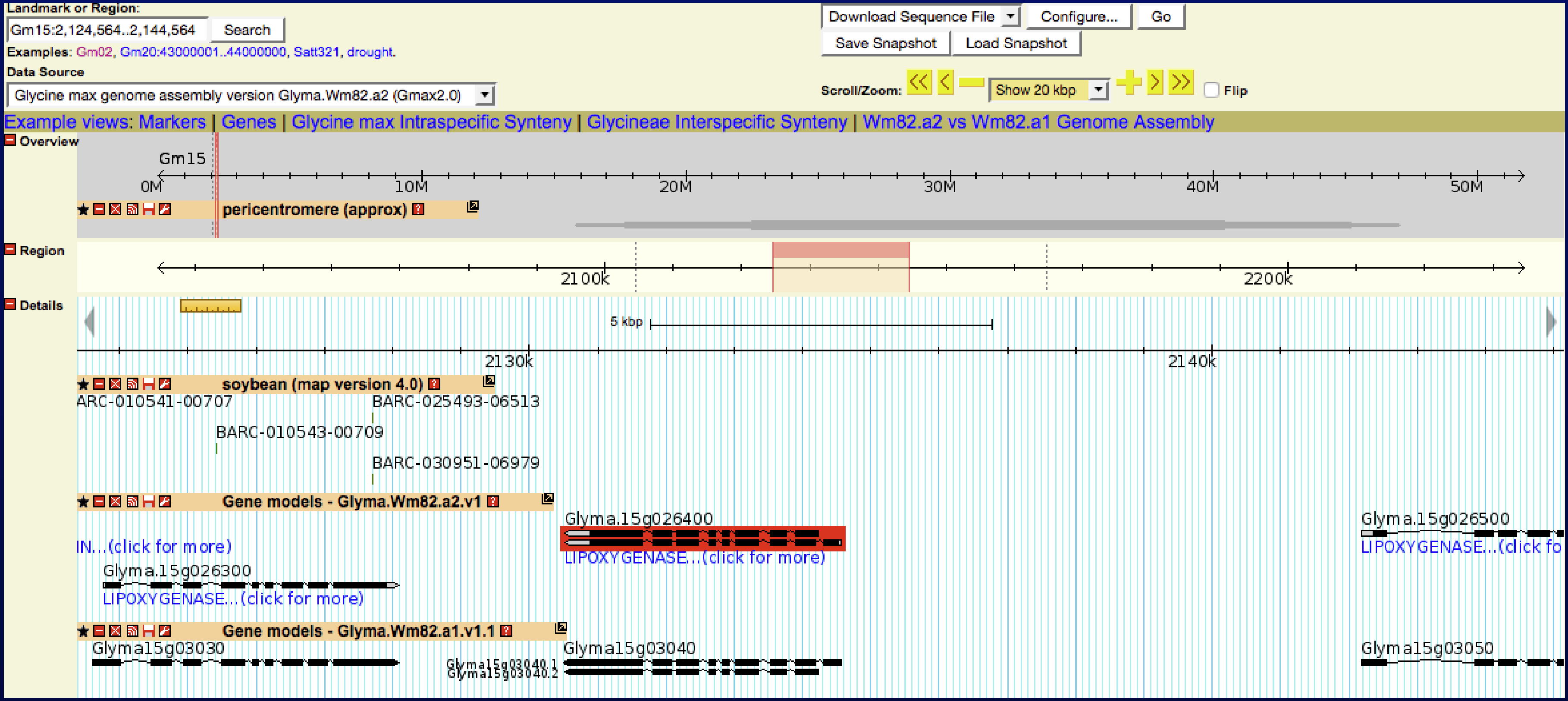
```
                                                               Score     E
Sequences producing significant alignments:                    (bits)  Value

Glyma.15G026400.1 polypeptide=Glyma.15G02640...  5097   0.0
Glyma.15G026400.2 polypeptide=Glyma.15G02640...  4686   0.0
Glyma.13G347800.1 polypeptide=Glyma.13G34780...  4151   0.0
Glyma.15G026500.1 polypeptide=Glyma.15G02650...  3509   0.0
Glyma.13G347700.1 polypeptide=Glyma.13G34770...  3435   0.0
Glyma.04G105900.1 polypeptide=Glyma.04G10590...   839   0.0
Glyma.08G189800.1 polypeptide=Glyma.08G18980...   733   0.0
Glyma.04G105500.1 polypeptide=Glyma.04G10550...   729   0.0
Glyma.08G189600.1 polypeptide=Glyma.08G18960...   492   e-137
Glyma.05G098600.1 polypeptide=Glyma.05G09860...   466   e-129
```

View BLAST Alignments



```
>Glyma.15G026400.1 polypeptide=Glyma.15G026400.1.p locus=Glyma.15G026400
            ID=Glyma.15G026400.1.Wm82.a2.v1 annot-version=Wm82.a2.v1
          Length = 2571

 Score = 5097 bits (2571), Expect = 0.0
 Identities = 2571/2571 (100%)
 Strand = Plus / Plus


Query: 49   atgtttggaatcatcggaggaaacaagggtcacaagataaaggggaacttggtgattatg 108
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1    atgtttggaatcatcggaggaaacaagggtcacaagataaaggggaacttggtgattatg 60


Query: 109  cgaaagaatgtgttggatatcaacagcattaccagtgttaagggtgtcatcggaaccggc 168
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 61   cgaaagaatgtgttggatatcaacagcattaccagtgttaagggtgtcatcggaaccggc 120


Query: 169  attaacatcattggaggagtcgtcgacaccgttactgctttggcgtcccacatctccatc 228
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 121  attaacatcattggaggagtcgtcgacaccgttactgctttggcgtcccacatctccatc 180


Query: 229  cagctcattagtgccaccaaggctgatggacatgggaaaggaaaagttggaaaggctaca 288
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 181  cagctcattagtgccaccaaggctgatggacatgggaaaggaaaagttggaaaggctaca 240


Query: 289  aatttaagaggacaagtgtcgttaccaaccttgggagctggcgaagatgcatacgatgtt 348
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 241  aatttaagaggacaagtgtcgttaccaaccttgggagctggcgaagatgcatacgatgtt 300


Query: 349  cattttgaatgggacagtgacttcggaattcccggtgcattttacattaagaacttcatg 408
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

View BLAST Hit in Genome Browser



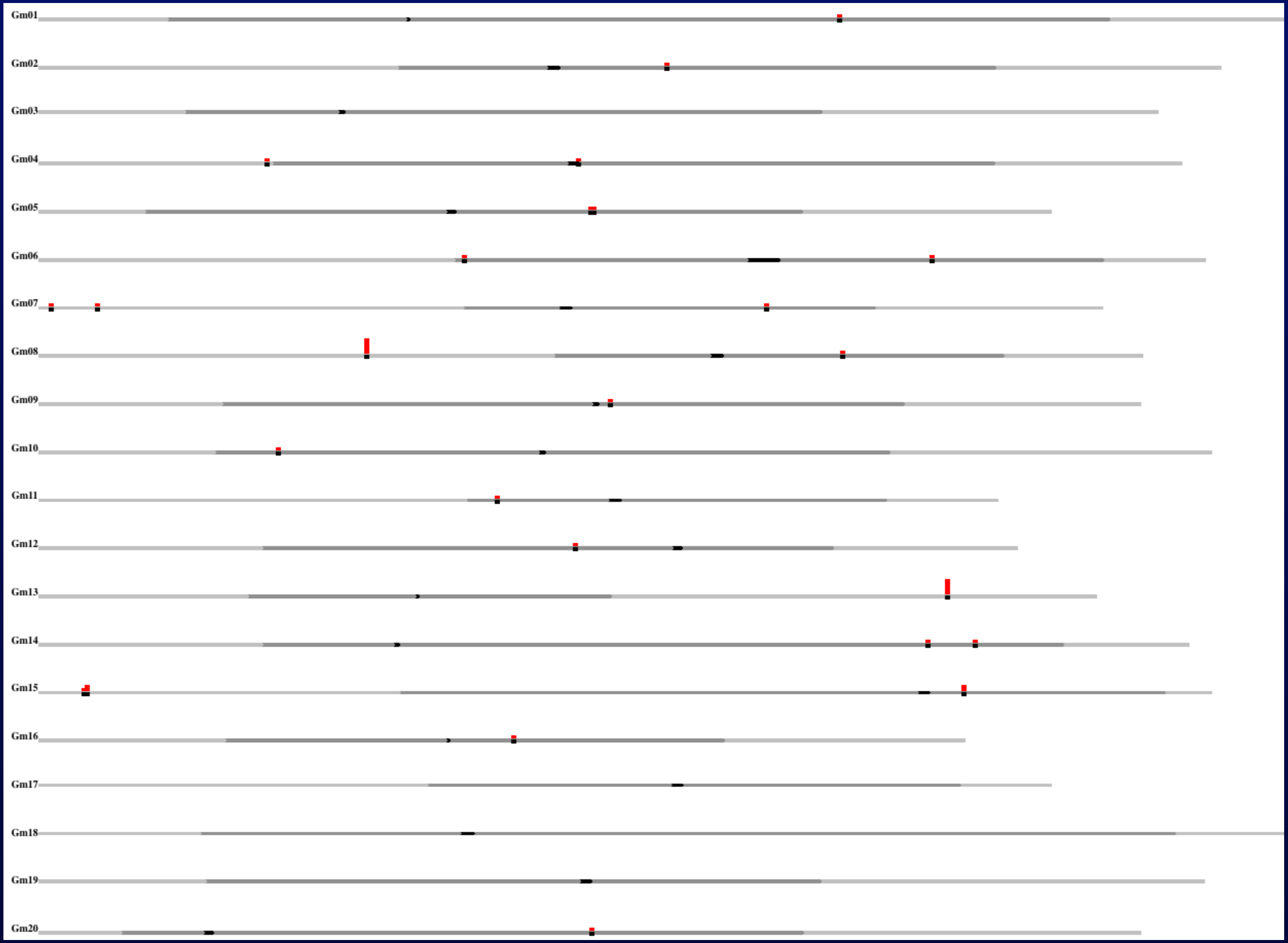
Landmark or Region: Gm15:2,124,564..2,144,564

Examples: Gm02, Gm20:43000001..44000000, Satt321, drought.

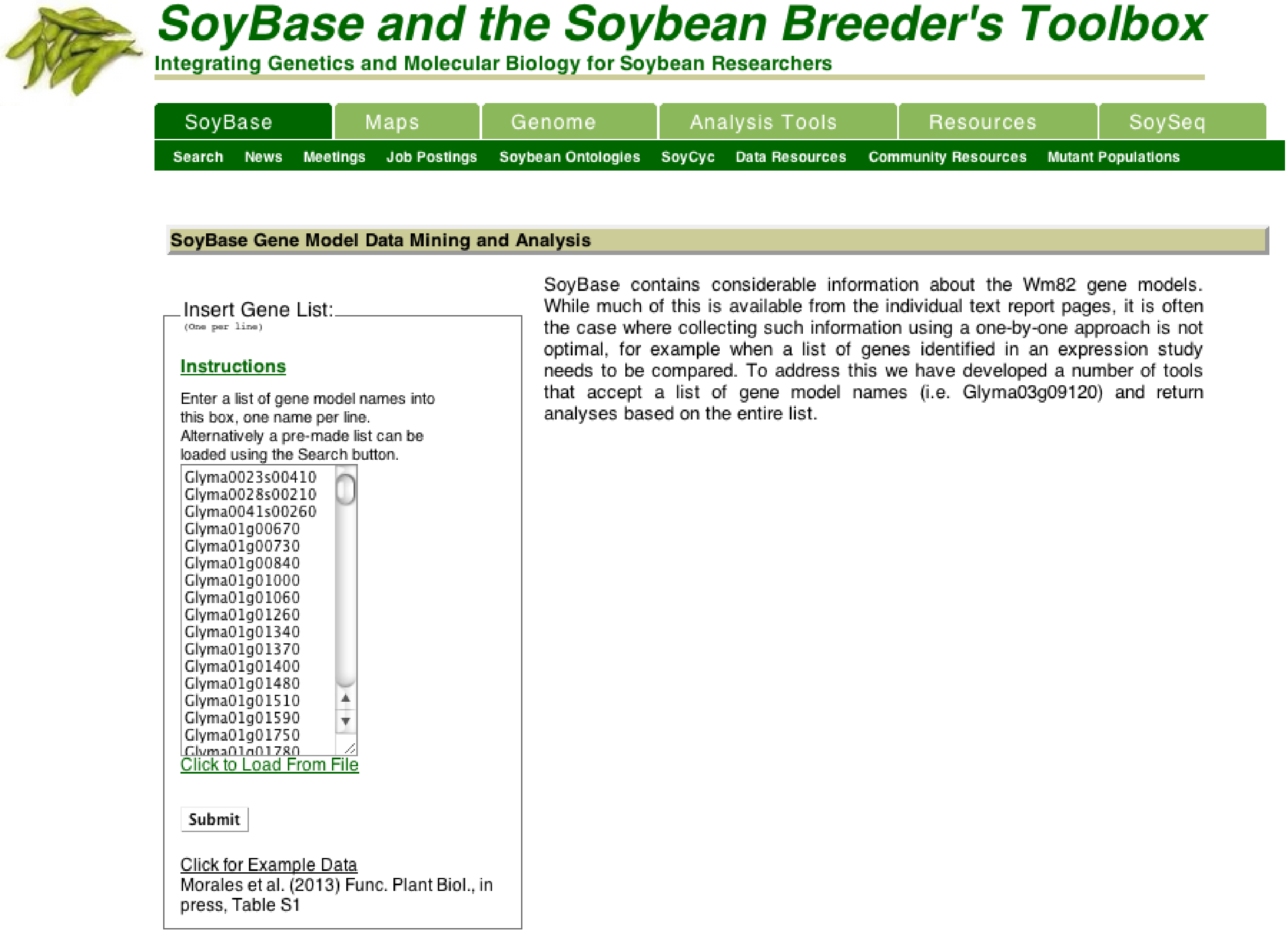
Data Source: Glycine max genome assembly version Glyma.Wm82.a2 (Gmax2.0)

Scroll/Zoom: Show 20 kbp

View All BLAST Hits for Whole Genome



GO Term Enrichment Tool



SoyBase Gene Model Data Mining and Analysis

Insert Gene List:
(One per line)

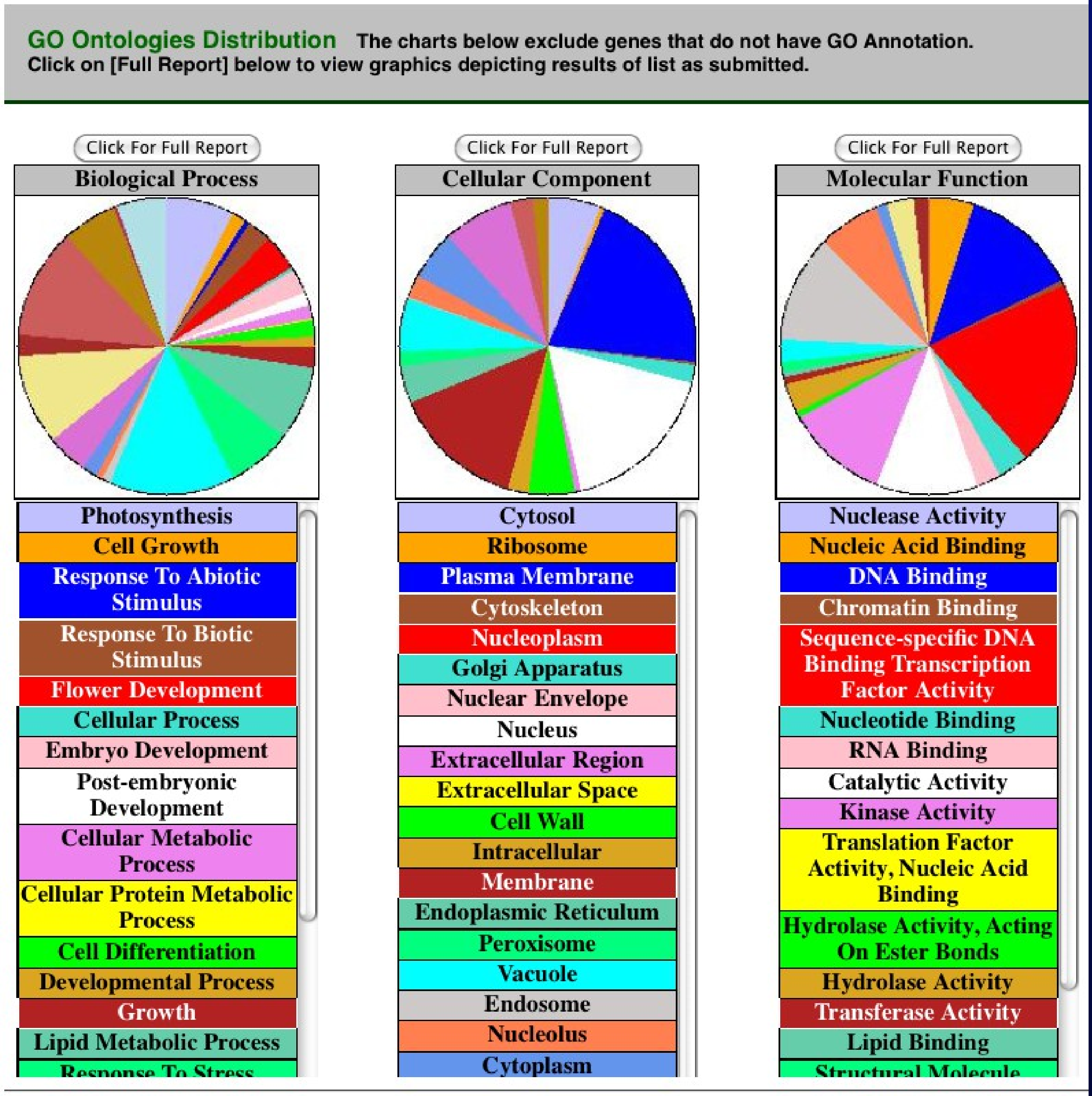
Instructions

Enter a list of gene model names into this box, one name per line. Alternatively a pre-made list can be loaded using the Search button.

Glyma0023s00410
Glyma0028s00210
Glyma0041s00260
Glyma01g00670
Glyma01g00730
Glyma01g00840
Glyma01g01000
Glyma01g01060
Glyma01g01260
Glyma01g01340
Glyma01g01370
Glyma01g01400
Glyma01g01480
Glyma01g01510
Glyma01g01590
Glyma01g01750
Glyma01g01780

Morales et al. (2013) Func. Plant Biol., in press, Table S1

SoyBase contains considerable information about the Wm82 gene models. While much of this is available from the individual text report pages, it is often the case where collecting such information using a one-by-one approach is not optimal, for example when a list of genes identified in an expression study needs to be compared. To address this we have developed a number of tools that accept a list of gene model names (i.e. Glyma03g09120) and return analyses based on the entire list.



GO Ontologies Distribution The charts below exclude genes that do not have GO Annotation.
Click on [Full Report] below to view graphics depicting results of list as submitted.

Biological Process

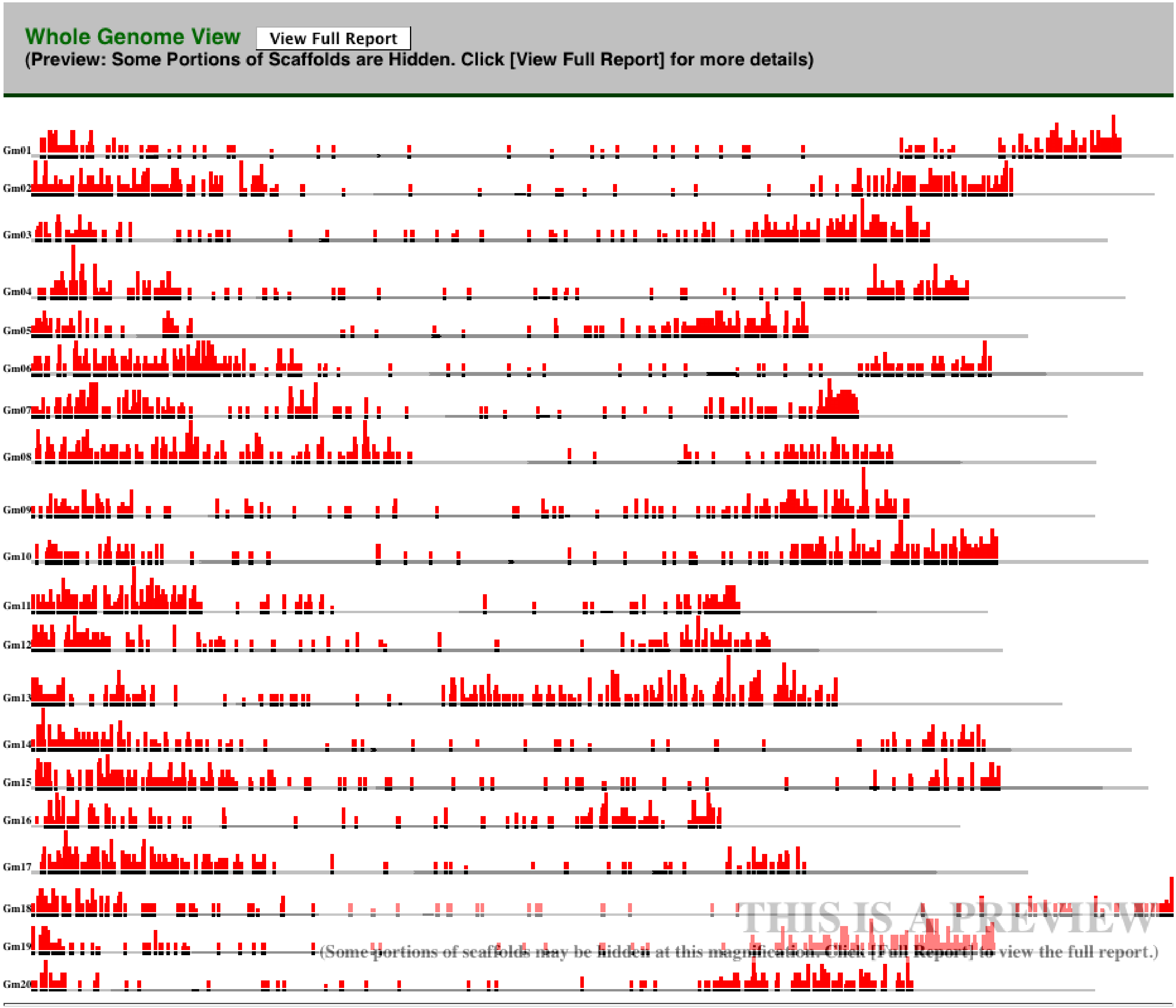
- Photosynthesis
- Cell Growth
- Response To Abiotic Stimulus
- Response To Biotic Stimulus
- Flower Development
- Cellular Process
- Embryo Development
- Post-embryonic Development
- Cellular Metabolic Process
- Cellular Protein Metabolic Process
- Cell Differentiation
- Developmental Process
- Growth
- Lipid Metabolic Process
- Response To Stress

Cellular Component

- Cytosol
- Ribosome
- Plasma Membrane
- Cytoskeleton
- Nucleoplasm
- Golgi Apparatus
- Nuclear Envelope
- Nucleus
- Extracellular Region
- Extracellular Space
- Cell Wall
- Intracellular
- Membrane
- Endoplasmic Reticulum
- Peroxisome
- Vacuole
- Endosome
- Nucleolus
- Cytoplasm

Molecular Function

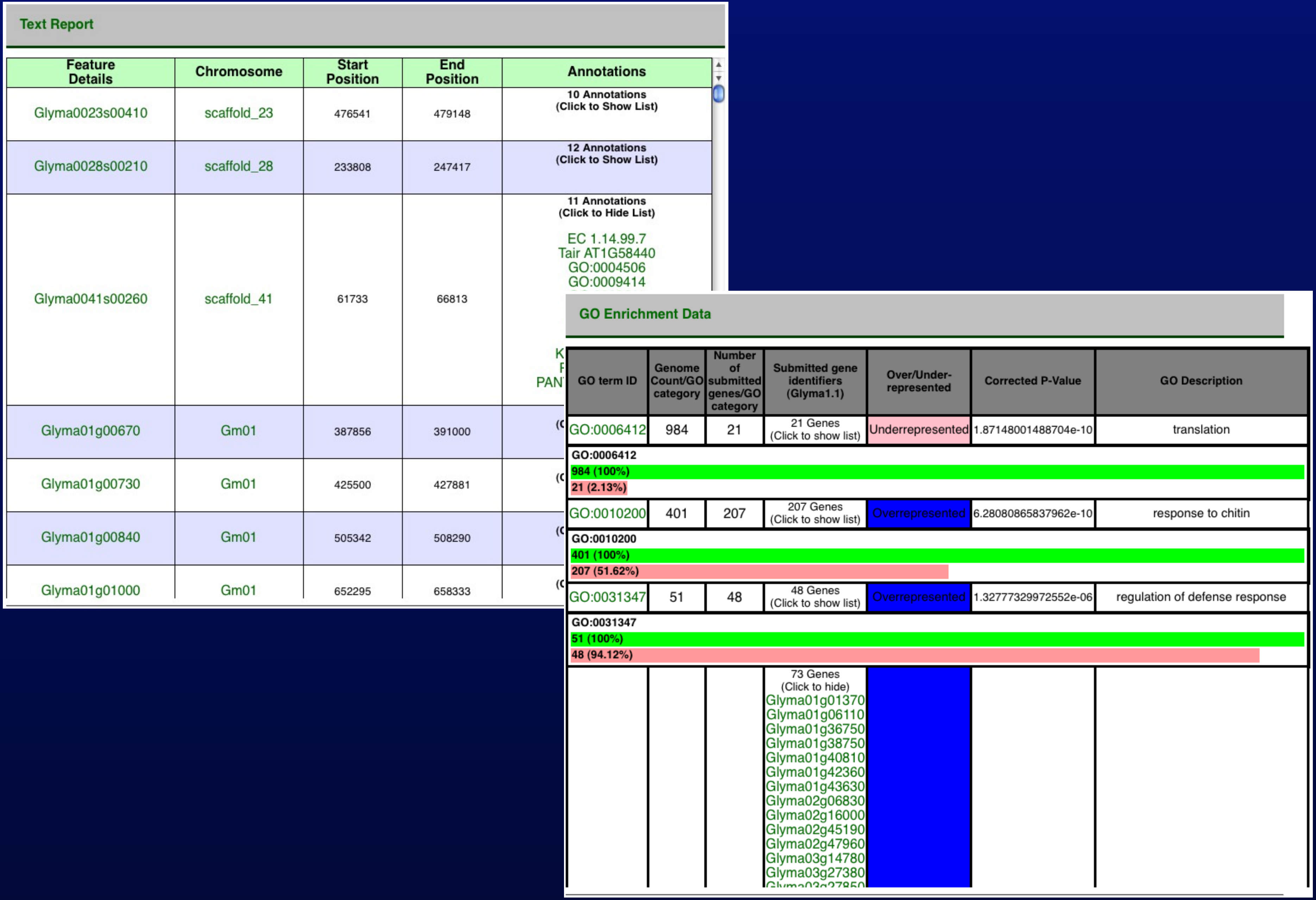
- Nuclease Activity
- Nucleic Acid Binding
- DNA Binding
- Chromatin Binding
- Sequence-specific DNA Binding Transcription Factor Activity
- Nucleotide Binding
- RNA Binding
- Catalytic Activity
- Kinase Activity
- Translation Factor Activity, Nucleic Acid Binding
- Hydrolase Activity, Acting On Ester Bonds
- Hydrolase Activity
- Transferase Activity
- Lipid Binding
- Structural Molecule


Whole Genome View

(Preview: Some Portions of Scaffolds are Hidden. Click [View Full Report] for more details)

THIS IS A PREVIEW

(Some portions of scaffolds may be hidden at this magnification. Click [Full Report] to view the full report.)



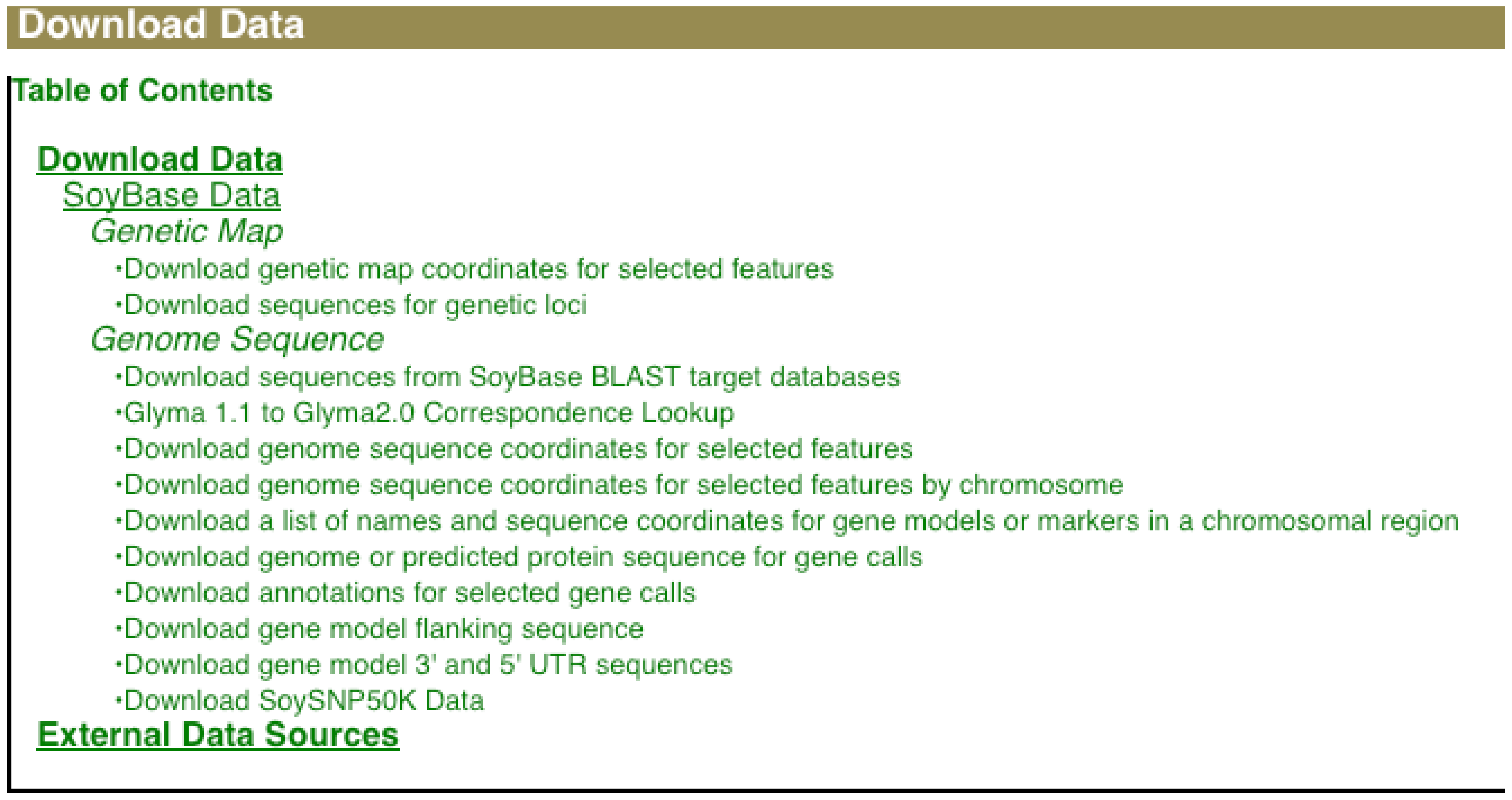
Text Report

Feature Details	Chromosome	Start Position	End Position	Annotations
Glyma0023s00410	scaffold_23	476541	479148	10 Annotations (Click to Show List)
Glyma0028s00210	scaffold_28	233808	247417	12 Annotations (Click to Show List)
Glyma0041s00260	scaffold_41	61733	66813	11 Annotations (Click to Hide List) EC 1.14.99.7 Tair AT1G58440 GO:0004506 GO:0009414 K F PAN
Glyma01g00670	Gm01	387856	391000	(C
Glyma01g00730	Gm01	425500	427881	(C
Glyma01g00840	Gm01	505342	508290	(C
Glyma01g01000	Gm01	652295	658333	(C

GO Enrichment Data

GO term ID	Genome Count/GO category	Number of submitted genes/GO category	Submitted gene identifiers (Glyma1.1)	Over/Under-represented	Corrected P-Value	GO Description
GO:0006412	984	21	21 Genes (Click to show list)	Underrepresented	1.87148001488704e-10	translation
GO:0006412 984 (100%) 21 (2.13%)						
GO:0010200	401	207	207 Genes (Click to show list)	Overrepresented	6.28080865837962e-10	response to chitin
GO:0010200 401 (100%) 207 (51.62%)						
GO:0031347	51	48	48 Genes (Click to show list)	Overrepresented	1.32777329972552e-06	regulation of defense response
GO:0031347 51 (100%) 48 (94.12%)						
			73 Genes (Click to hide) Glyma01g01370 Glyma01g06110 Glyma01g36750 Glyma01g38750 Glyma01g40810 Glyma01g42360 Glyma01g43630 Glyma02g06830 Glyma02g16000 Glyma02g45190 Glyma02g47960 Glyma03g14780 Glyma03g27380			

Data Downloads



Download Data

Table of Contents

Download Data

SoyBase Data

Genetic Map

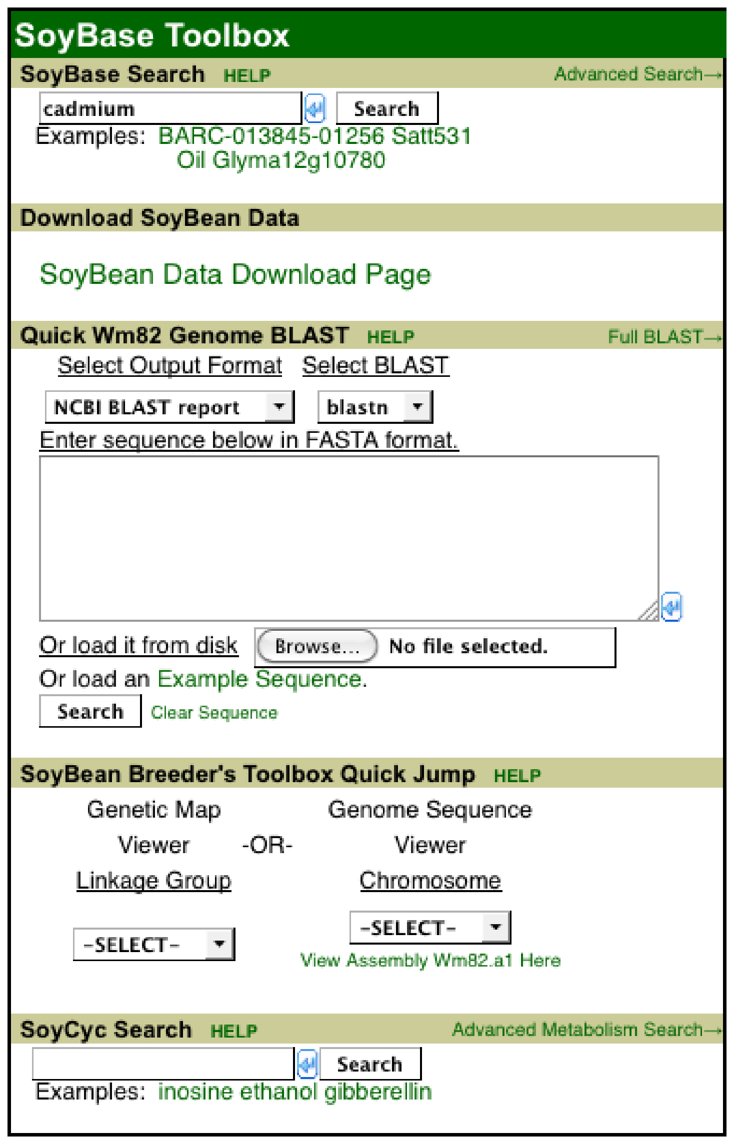
- Download genetic map coordinates for selected features
- Download sequences for genetic loci

Genome Sequence

- Download sequences from SoyBase BLAST target databases
- Glyma 1.1 to Glyma2.0 Correspondence Lookup
- Download genome sequence coordinates for selected features
- Download genome sequence coordinates for selected features by chromosome
- Download a list of names and sequence coordinates for gene models or markers in a chromosomal region
- Download genome or predicted protein sequence for gene calls
- Download annotations for selected gene calls
- Download gene model flanking sequence
- Download gene model 3' and 5' UTR sequences
- Download SoySNP50K Data

External Data Sources

Search All of SoyBase for Cadmium



SoyBase Toolbox

SoyBase Search HELP — Advanced Search→

cadmium [Search]

Examples: BARC-013845-01256 Satt531
Oil Glyma12g10780

Download SoyBean Data

SoyBean Data Download Page

Quick Wm82 Genome BLAST HELP — Full BLAST→

Select Output Format: NCBI BLAST report
Select BLAST: blastn

Enter sequence below in FASTA format.

Or load it from disk [Browse...] No file selected.

Or load an Example Sequence.

[Search] Clear Sequence

SoyBean Breeder's Toolbox Quick Jump HELP

Genetic Map Viewer -OR- Genome Sequence Viewer

Linkage Group: -SELECT-

Chromosome: -SELECT-

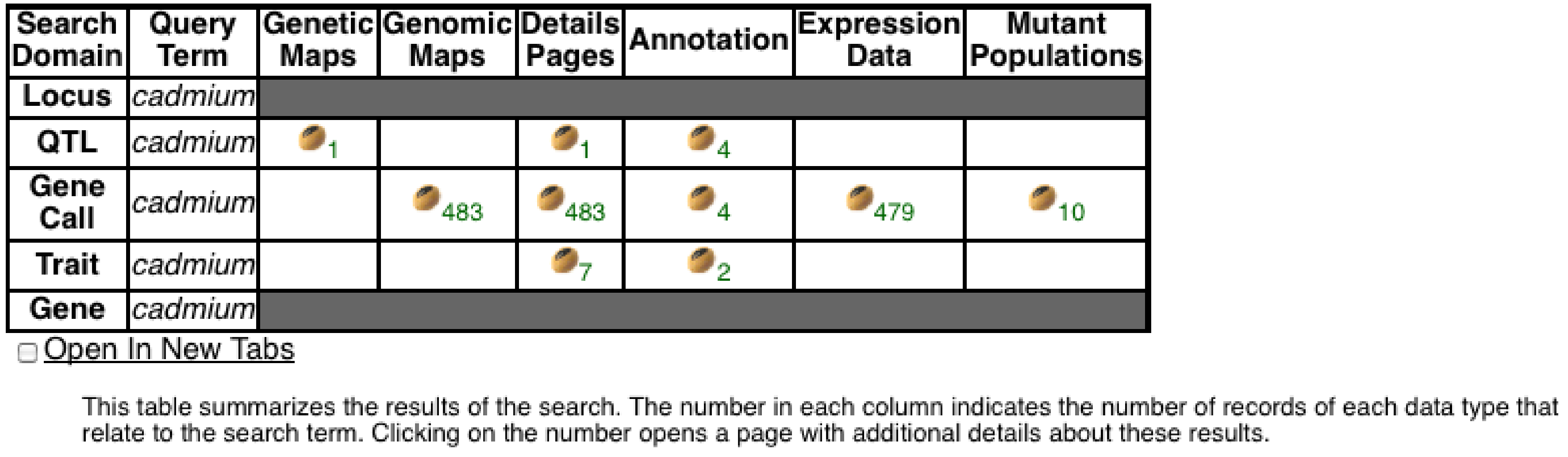
View Assembly Wm82.a1 Here

SoyCyc Search HELP — Advanced Metabolism Search→

[Search]

Examples: inosine ethanol gibberellin

Summary of Search Results



Search Domain	Query Term	Genetic Maps	Genomic Maps	Details Pages	Annotation	Expression Data	Mutant Populations
Locus	*cadmium*						
QTL	*cadmium*	1		1	4		
Gene Call	*cadmium*		483	483	4	479	10
Trait	*cadmium*			7	2		
Gene	*cadmium*						

☐ Open In New Tabs

This table summarizes the results of the search. The number in each column indicates the number of records of each data type that relate to the search term. Clicking on the number opens a page with additional details about these results.

SoyBase and the Soybean Breeder's Toolbox

Integrating Genetics and Molecular Biology for Soybean Researchers

Sign Up Here To Receive SoyBase Update Emails

Use this form to contact the SoyBase Curators when

- you have data you would like included in SoyBase
- have found an error on our site
- have a suggestion about improving SoyBase

We'll get back to you as soon as possible.

From: * (Type your name here)

Email: * (Type your email address here)

Subject: *

Comment: *

(Be sure to include the URL of the page if you are reporting an error)

Data Submission Templates and Instructions

Data Types

Bi-allelic QTL Data
  Excel spreadsheet for data entry

Genome Wide Association QTL Data
  Contact us for instructions

Re-sequencing Data (SNPs, CNV, etc.)
  Contact us for instructions

Expression or Transcriptomic Data (RNA-seq, GeneChip, custom chips, etc.)
  Contact us for instructions

Questions?

We value your opinion!!

Please take our quick six question survey using the link on the SoyBase home page or at http://www.soybase.org/survey/