

# What's New at SoyBase

Feb. 2014

David Grant  
Rex Nelson  
Kevin Feeley  
Robert Baker  
Nathan Weeks  
Steven Cannon

# Improved QTL Names

Asian Soybean Rust	Asian Soybean Rust
Branching	Branching
BSR	BSR
Cnpy ht	Canopy height
Cnpy wdth	Canopy width
CID	Carbon isotope discrimination
CWP	Cellwall polysacch composition
CES	Chlorimuron sensitivity
CCW	Common cutworm
CEW	Corn earworm
Lflt chlorophyll	Leaflet chlorophyll
Lf lgth	Leaflet length
Specific lflt N	Leaflet N, specific
Lf Phos	Leaflet P
Lflt shape	Leaflet shape
TRG(DW)	Leaflet trigonelline, dry
TRG(FW)	Leaflet trigonelline, fresh
Lf wdth	Leaflet width
NitR7	Seed N at R7
Sd num	Seed number
Oil	Seed oil
Sd wt/plant	Seed weight per plant
Sd yld	Seed yield
Yld/Ht	Seed yield/Plant height
Yld/SW	Seed yield/Seed weight
R5	Seed, beginning

# SoyBase and the Soybean Breeder's Toolbox

Integrating Genetics and Molecular Biology for Soybean Researchers

SoyBase

Maps

Genome

Analysis Tools

Resources

SoySeq

Search Data Tools SoyCyc Legume Research Programs Soybean Breeders Workshop Download Data Submit Your Data Contact Us

## Community Resources and Communications

### Tutorials

#### SoyBase Video Tutorials

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

#### YouTube Videos

- Soybean Growth and Development
- Soybean Diseases
- Soybean Pests
- Methods and Protocols

Making Crosses with Soybean  
Harvesting Research Plots

## YouTube and MP4 Tutorials

### 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

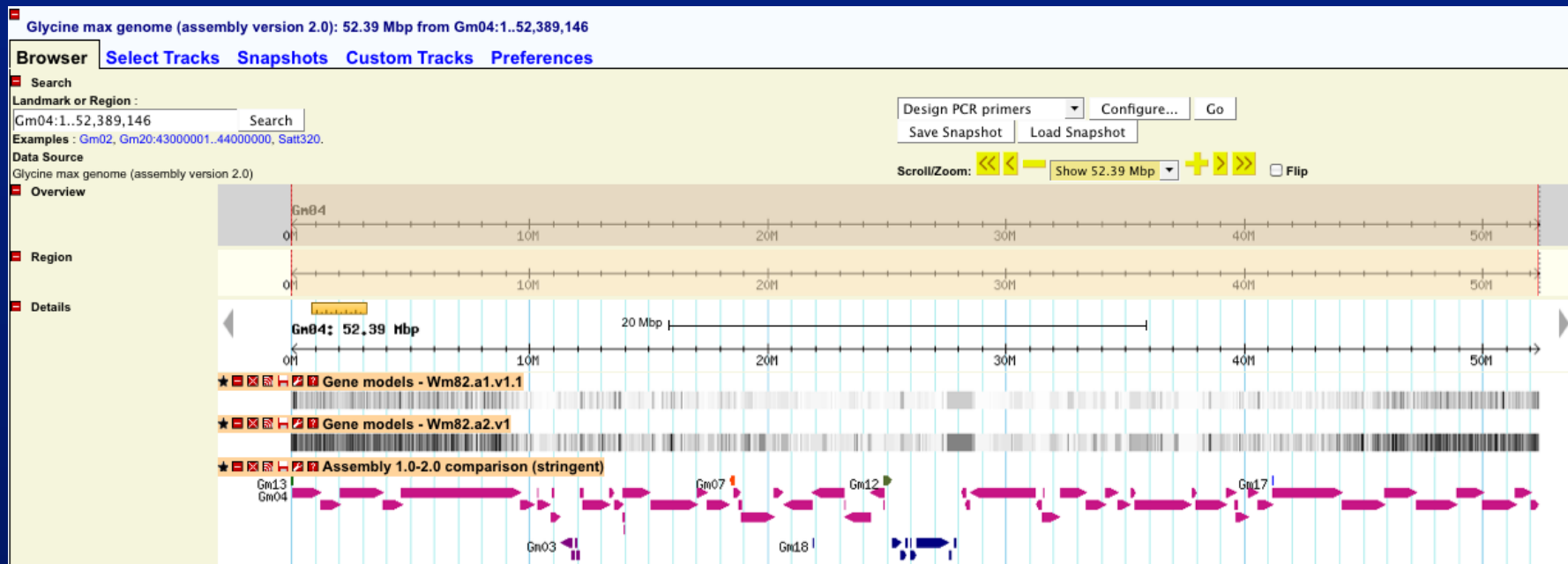
- How to zoom into a region on the genetic map (3:10, 118 Mb MP4)
- How to flip genetic maps to resolve corresponding marker positions (1:44, 57 Mb MP4)
- Turning QTL classes off/on (3:08, 78 Mb MP4)
- Removing and adding genetic maps (4:27, 115 Mb MP4)
- How to quickly bring up a genetic or sequence 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)

### 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)



## New Wm82 genome assembly from JGI

- corrects known problems with the v1 assembly
- incorporates additional transcript data resulting in improved gene predictions
- revised gene model and annotation nomenclature

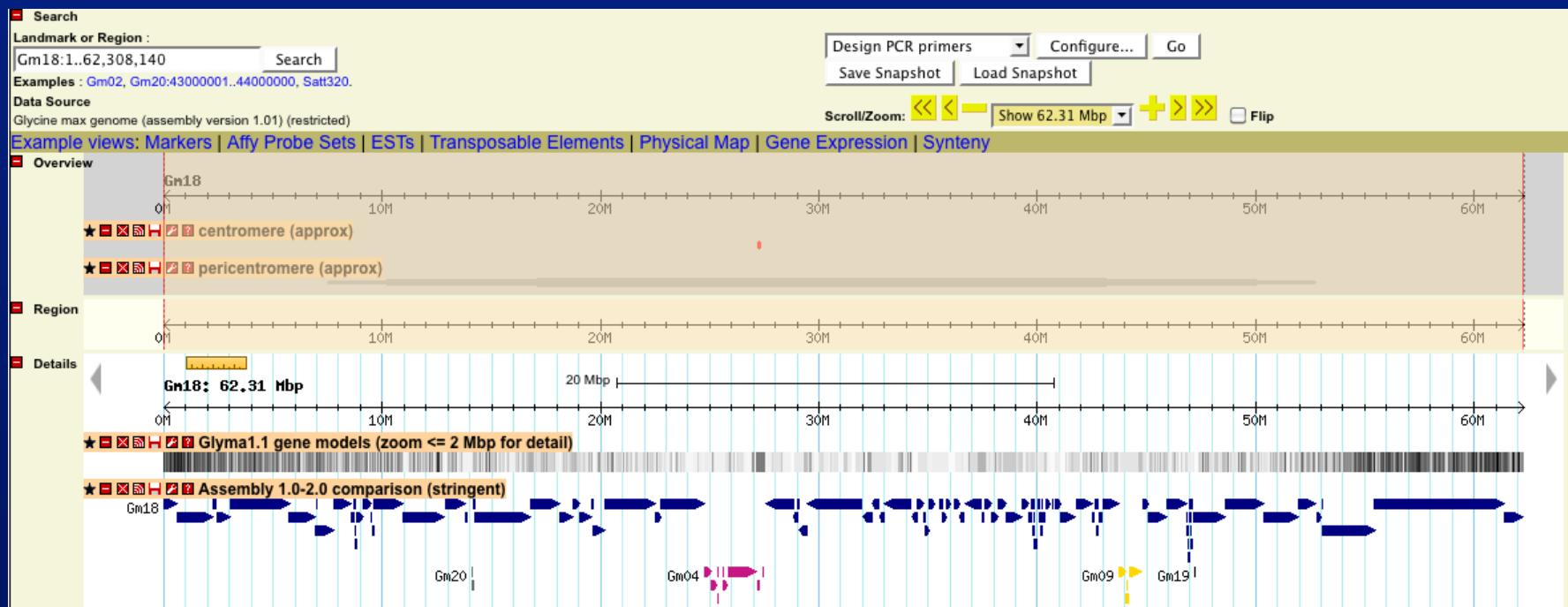
locus: **Glyma.01g000100**

transcript: **Glyma.01g000100.1**

for annotation version Wm82.a2.v1

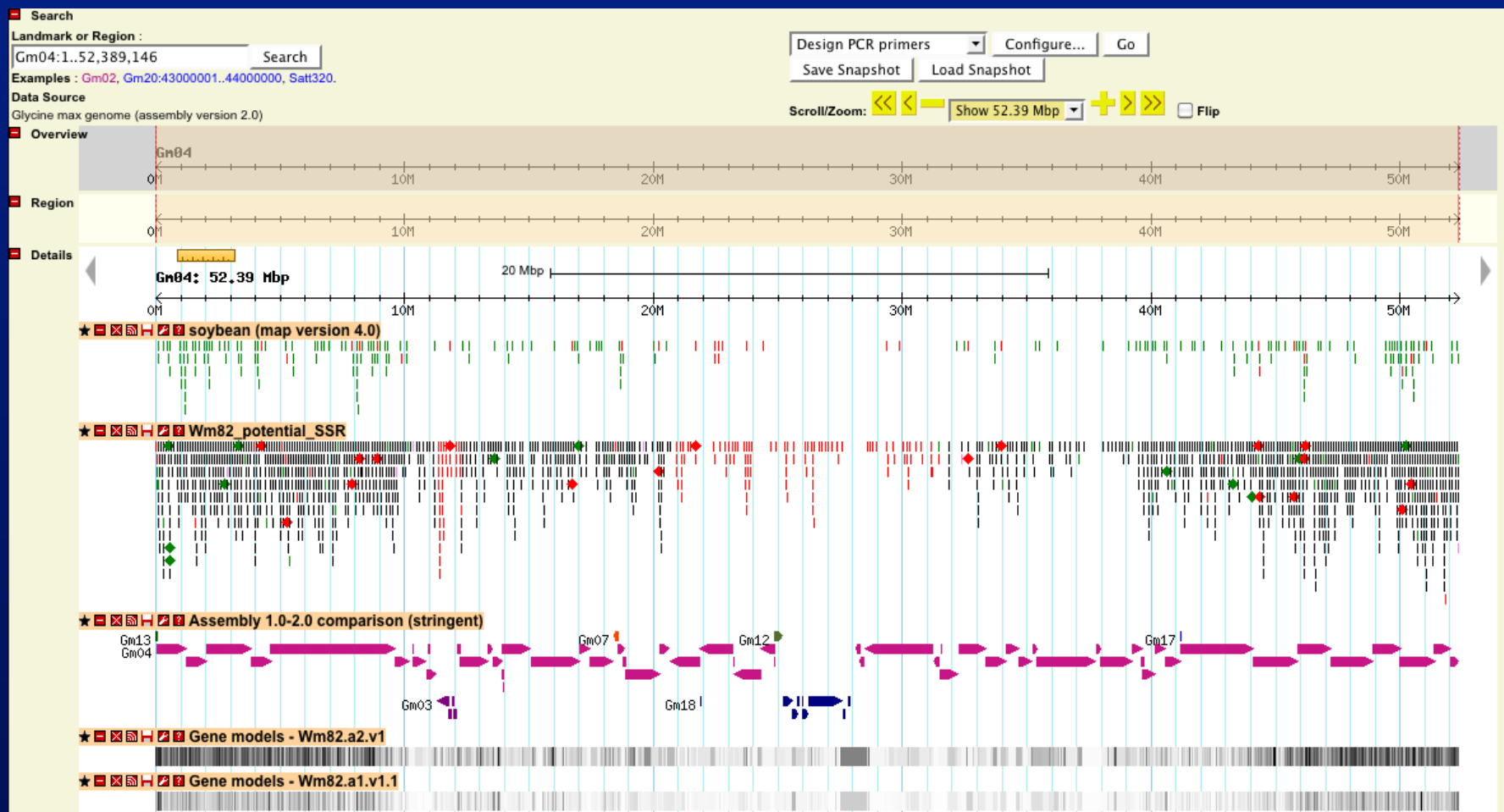
locus ID: **Glyma.01g000100.Wm82.a2.v1**

transcript ID: **Glyma.01g000100.1.Wm82.a2.v1**



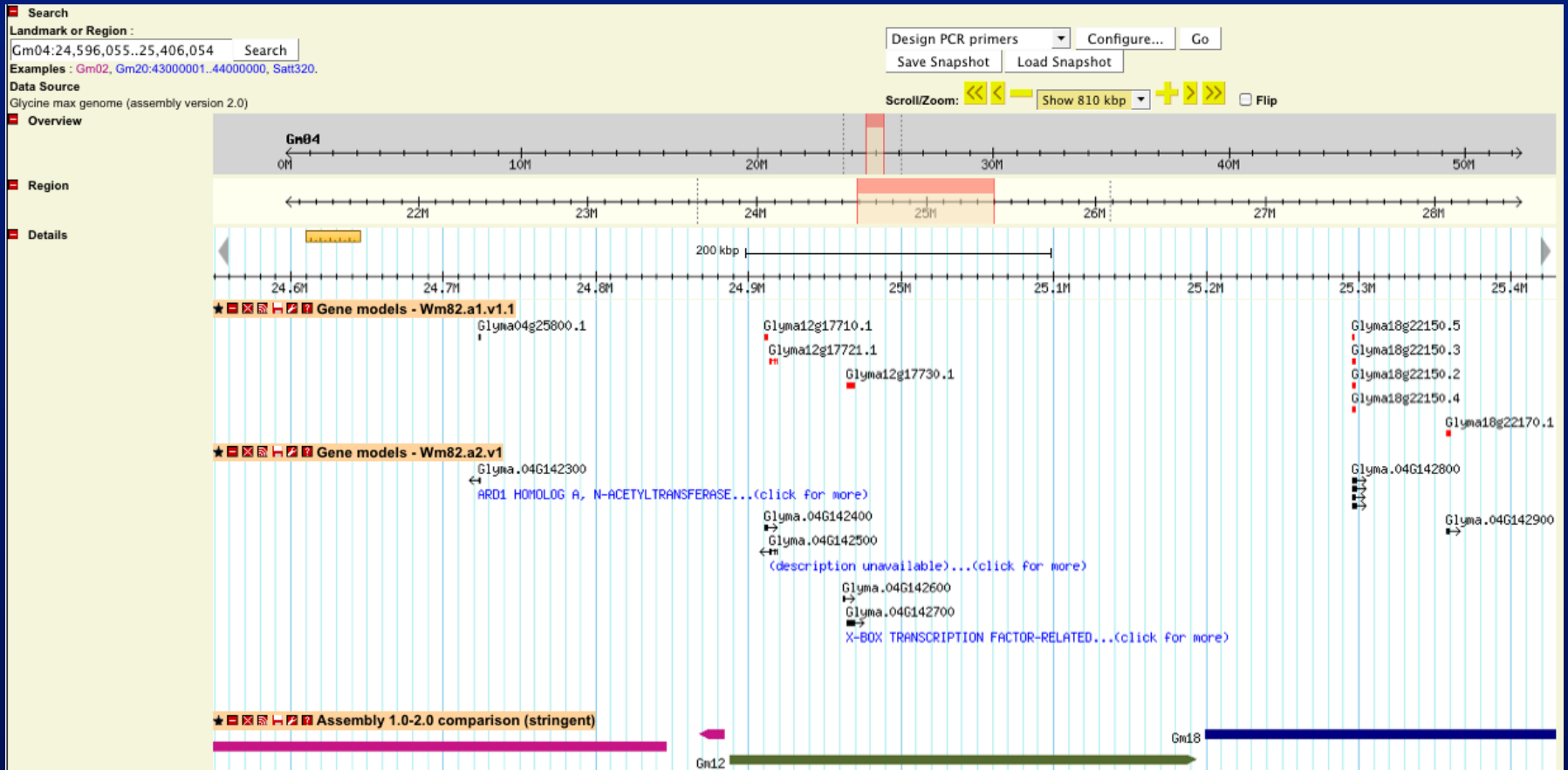
The current sequence browser now has a track showing synteny between the v1 and v2 assemblies.

That the v1 and v2 synteny tracks are reciprocal for shared regions.



Wm82.a1.v1.1 gene models and markers mapped onto new assembly when possible.

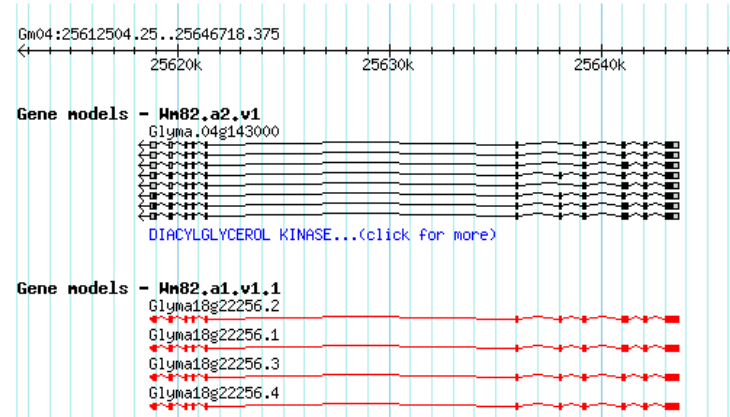
Markers and gene models with different chromosomal locations or DNA strand between the two assemblies are indicated in red.





## Report for Sequence Feature Glyma.04g143000

**Feature Type:** gene\_model  
**Chromosome:** Gm04  
**Start:** 25618725  
**stop:** 25643608  
**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.04g143000

Database ID	Annotation Type	Annotation Description	Annotation Source	Match Score	Evidence Code
AT2G18730.1	AT	diacylglycerol kinase 3	JGI	N/A	IEA
AT4G30340.1	AT	diacylglycerol kinase 7	JGI	N/A	IEA
AT5G57690.1	AT	diacylglycerol kinase 4	JGI	N/A	IEA
GO:0004143	GO	diacylglycerol kinase activity	JGI	N/A	IEA
GO:0007205	GO	protein kinase C-activating G-protein coupled receptor signaling pathway	JGI	N/A	IEA
KOG1169	KOG	Diacylglycerol kinase	JGI	N/A	IEA
PTHR11255	Panther	DIACYLGLYCEROL KINASE	JGI	N/A	IEA
PTHR11255:SF32	Panther	DIACYLGLYCEROL KINASE BETA	JGI	N/A	IEA
PF00609	PFAM	Diacylglycerol kinase accessory domain	JGI	N/A	IEA
PF00781	PFAM	Diacylglycerol kinase catalytic domain	JGI	N/A	IEA

### Gene model name correspondences to Glyma.04g143000 Gene Call Version Glyma2.0

Corresponding Name	Annotation Version	Evidence	Comments
Glyma18g22256	Glyma 1.1	IGC	As supplied by JGI

Glycine max genome (assembly version 2.0): 100 kbp from Gm04:48,560,000..48,659,999

Browser [Select Tracks](#) [Snapshots](#) [Custom Tracks](#) [Preferences](#)

Search

Landmark or Region:

Gm04:48,560,000..48,659,999

Examples: [Gm02](#), [Gm20:43000001..44000000](#), [Sat320](#).

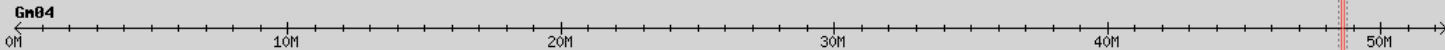
Data Source

Glycine max genome (assembly version 2.0)

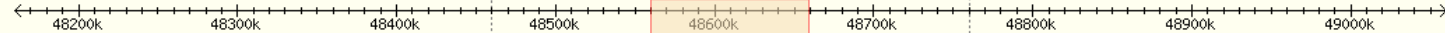
Design PCR primers

Scroll/Zoom:         Flip

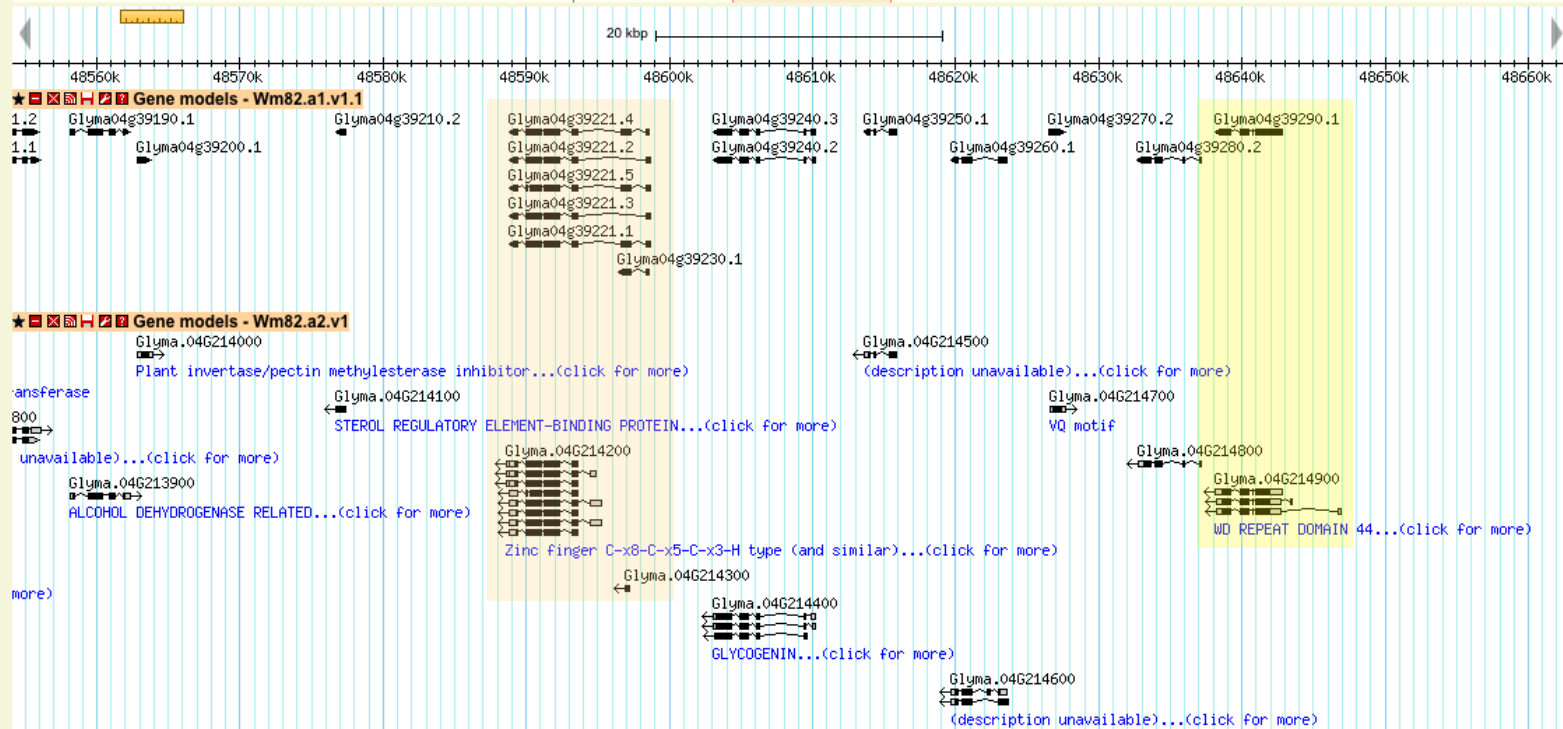
Overview



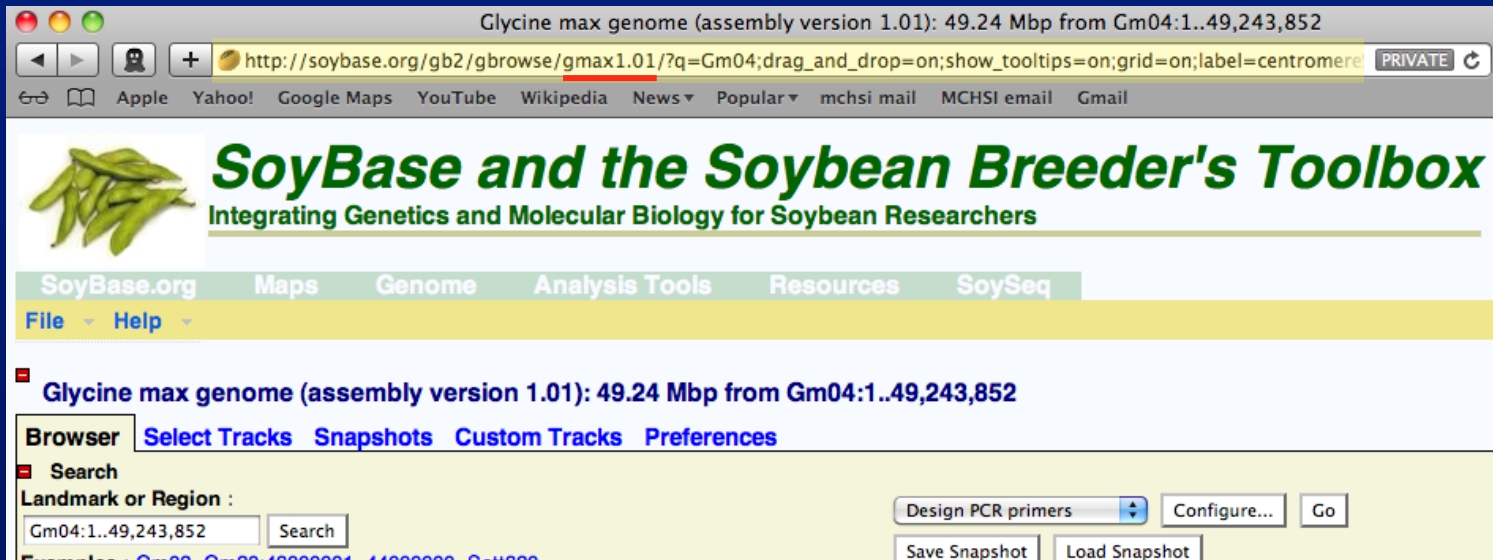
Region



Details



# Switching Between v1 and v2 Genome Browsers



Glycine max genome (assembly version 1.01): 49.24 Mbp from Gm04:1..49,243,852

http://soybase.org/gb2/gbrowse/gmax1.01/?q=Gm04;drag\_and\_drop=on;show\_tooltips=on;grid=on;label=centromere PRIVATE

**SoyBase and the Soybean Breeder's Toolbox**  
Integrating Genetics and Molecular Biology for Soybean Researchers

SoyBase.org Maps Genome Analysis Tools Resources SoySeq

File Help

**Glycine max genome (assembly version 1.01): 49.24 Mbp from Gm04:1..49,243,852**

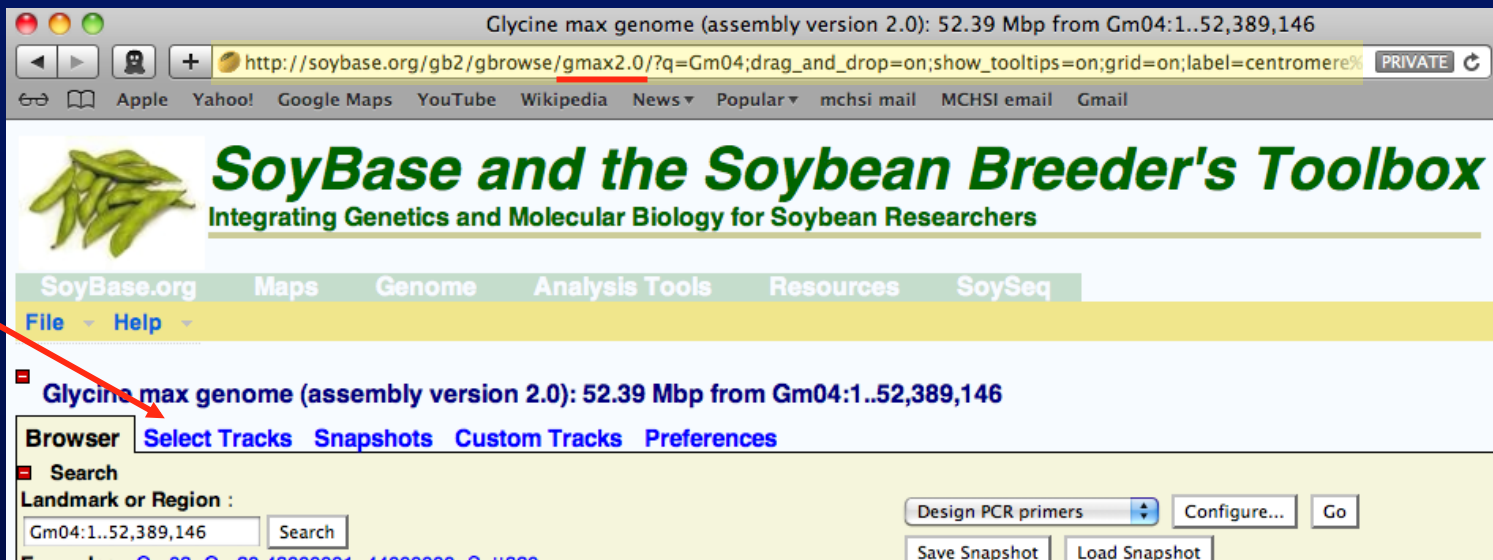
Browser Select Tracks Snapshots Custom Tracks Preferences

Search

Landmark or Region : Gm04:1..49,243,852 Search

Design PCR primers Configure... Go

Save Snapshot Load Snapshot



Glycine max genome (assembly version 2.0): 52.39 Mbp from Gm04:1..52,389,146

http://soybase.org/gb2/gbrowse/gmax2.0/?q=Gm04;drag\_and\_drop=on;show\_tooltips=on;grid=on;label=centromere% PRIVATE

**SoyBase and the Soybean Breeder's Toolbox**  
Integrating Genetics and Molecular Biology for Soybean Researchers

SoyBase.org Maps Genome Analysis Tools Resources SoySeq

File Help

**Glycine max genome (assembly version 2.0): 52.39 Mbp from Gm04:1..52,389,146**

Browser Select Tracks Snapshots Custom Tracks Preferences

Search

Landmark or Region : Gm04:1..52,389,146 Search

Design PCR primers Configure... Go

Save Snapshot Load Snapshot

# SoyBase and the Soybean Breeder's Toolbox

Integrating Genetics and Molecular Biology for Soybean Researchers

SoyBase

Maps

Genome

Analysis Tools

Resources

SoySeq

[About the Sequence Map](#)

[SoyBase Browser](#)

[Transposable Elements](#)

[Soybean at Phytozome \(JGI\)](#)

[PlantGDB Browser](#)

[Download Data](#)

## SoyBase Soybean Sequence Map Resources

### SoyBase browser

#### Choose a Chromosome

Chromosome

Number 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Linkage Group D1a D1b N C1 A1 C2 M A2 K O B1 H F B2 E J D2 G L I

View the JGI Gma1.01 WGS soybean genomic sequence, as pseudomolecules, integrating features from the soybean composite genetic and physical maps with the genomic sequence. Links from GBrowse features (BACs, markers) to locations in the SoyBase genetic and physical maps are available from a contextual menu. The preliminary 7x assembly is still available [here](#).

[View or download](#) a list of markers or other sequence map features by chromosomes here.

[Download gene call sequences](#) from the Williams 82 genomic sequence.

### Search for Sequence Features and Annotations

#### SoyBase Search

Perform searches of our entire data set.

[See haplotype data for 19,652 cultivars analyzed with the SoySNP50K chip](#)

[View or download a list](#) of genetic markers or other features by linkage group.

### 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 52,041 SNPs is available [here](#) (approx. 170 Mb compressed with GZIP) .

SoySNP50K haplotypes for a user-selected subset of the genotyped cultivars can be downloaded from [this page](#).



### Download SoySNP50K haplotypes for a subset of the cultivars

Paste a list of cultivar names, one per line, into the box below or use the Browse button to navigate to a plain text file on your computer that contains a list of cultivar names.

The submit button will initiate a download to your computer of the SoySNP50K haplotypes for the cultivars you provide.

```
PI086046
PI090208
PI219698
PI253651A
PI347550A
PI398807
PI408055A
PI408069
```

Browse...

Submit Load Examples

Or use our [shopping cart](#) tool to select cultivars of interest from the complete list of cultivars haplotyped by Song et al.

#### Readme Text:

(Included in download)

Including this READ ME there are four files returned for a cultivar-haplotype search. All data are in plain text files with tab delimiters between columns.

#### NOTE:

We have observed that on some computers opening these files directly in Excel results in each line being placed in a single cell rather than 1 cell per tab-separated column. This can be solved by opening the file in Word or another text editor and using copy/paste to put the data into Excel.

#### Downloaded Data Files:

[cultivar\\_snps](#)

This is a 2 column tab-delimited file, 1 line per cultivar.

Cultivar Name	Cultivar Haplotype
PI442007B	TAATTAATAAGTAGTAATAACCCAGGA.....

[snp\\_id\\_and\\_position](#)

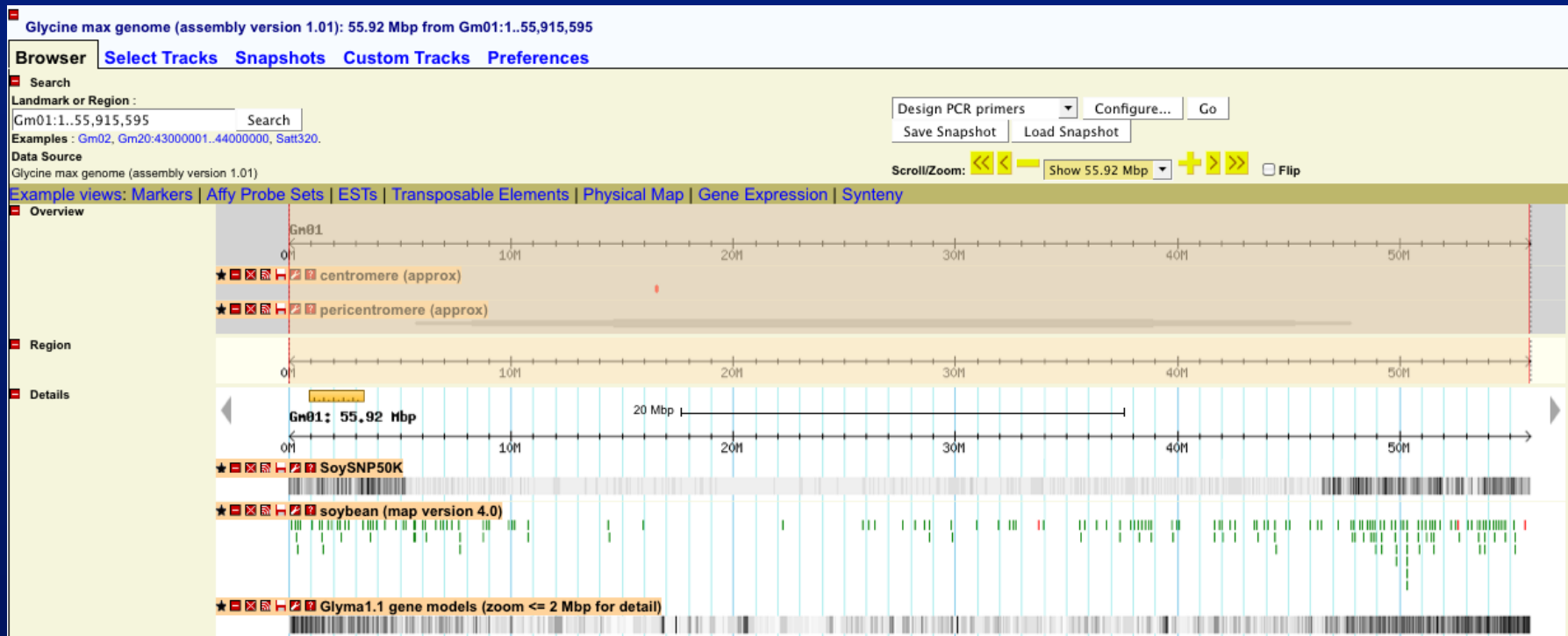
This file contains what amounts to the column header information for 'cultivar\_snps'. It is provided as a separate file so that the haplotype data in 'cultivar\_snps' doesn't need to have tabs between each SNP allele value.

SNP Name	ss715578401	ss715578402	ss715578403	ss715578404	ss715578405
Chromosome	Gm01	Gm01	Gm01	Gm01	Gm01
Position	10015701	10052330	1006696	10074809	1009248

[snp\\_id\\_position\\_and\\_value](#)

This is a more human readable version that includes both the haplotype data and the SNP descriptions. The order of SNP lines is the same as the order of SNP allele values in 'cultivar\_snps'.

SNP ID	Chromosome	Position	PI442007B	PI365426	PI587703A	PI228064
ss715578401	Gm01	10015701	A	A	A	A
ss715578402	Gm01	10052330	C	C	C	C



SoySNP50K data currently only on genome assembly Wm82.a1.v1.1

Glycine max genome (assembly version 1.01): 220 kbp from Gm01:7,239,844..7,459,843

**Browser** [Select Tracks](#) [Snapshots](#) [Custom Tracks](#) [Preferences](#)

**Search**

Landmark or Region :

Gm01:7,239,844..7,459,843

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

**Data Source**

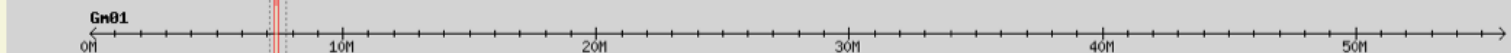
Glycine max genome (assembly version 1.01)

Design PCR primers

ScrollZoom:         Flip

Example views: [Markers](#) | [Affy Probe Sets](#) | [ESTs](#) | [Transposable Elements](#) | [Physical Map](#) | [Gene Expression](#) | [Synteny](#)

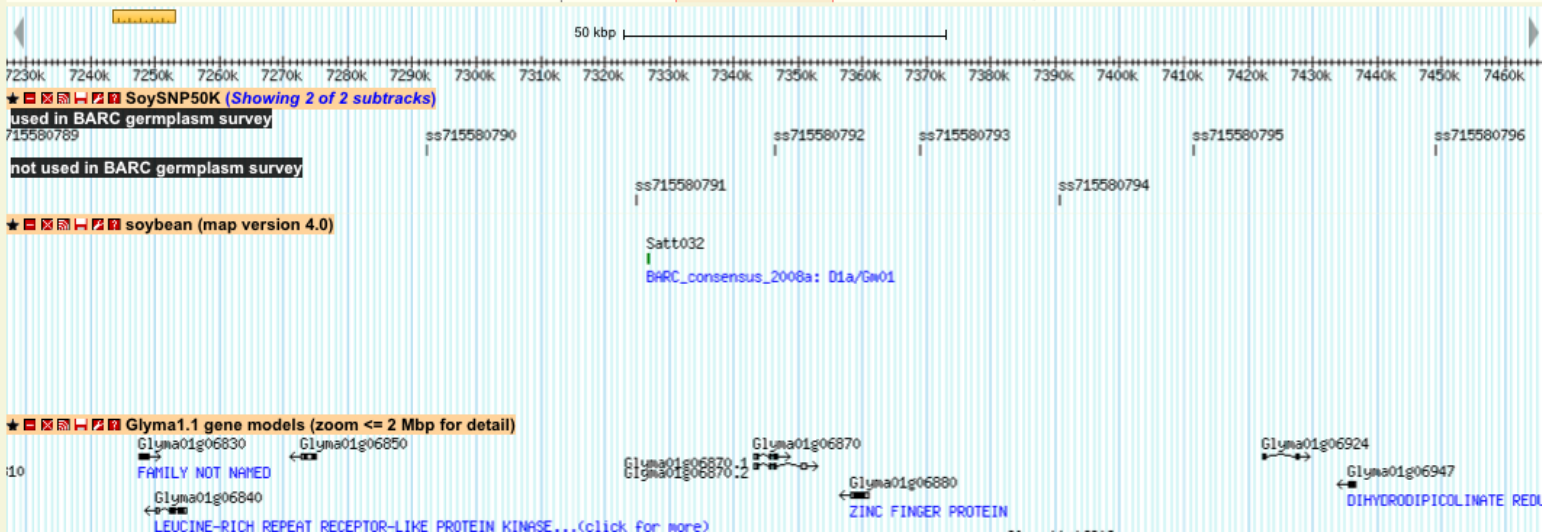
**Overview**



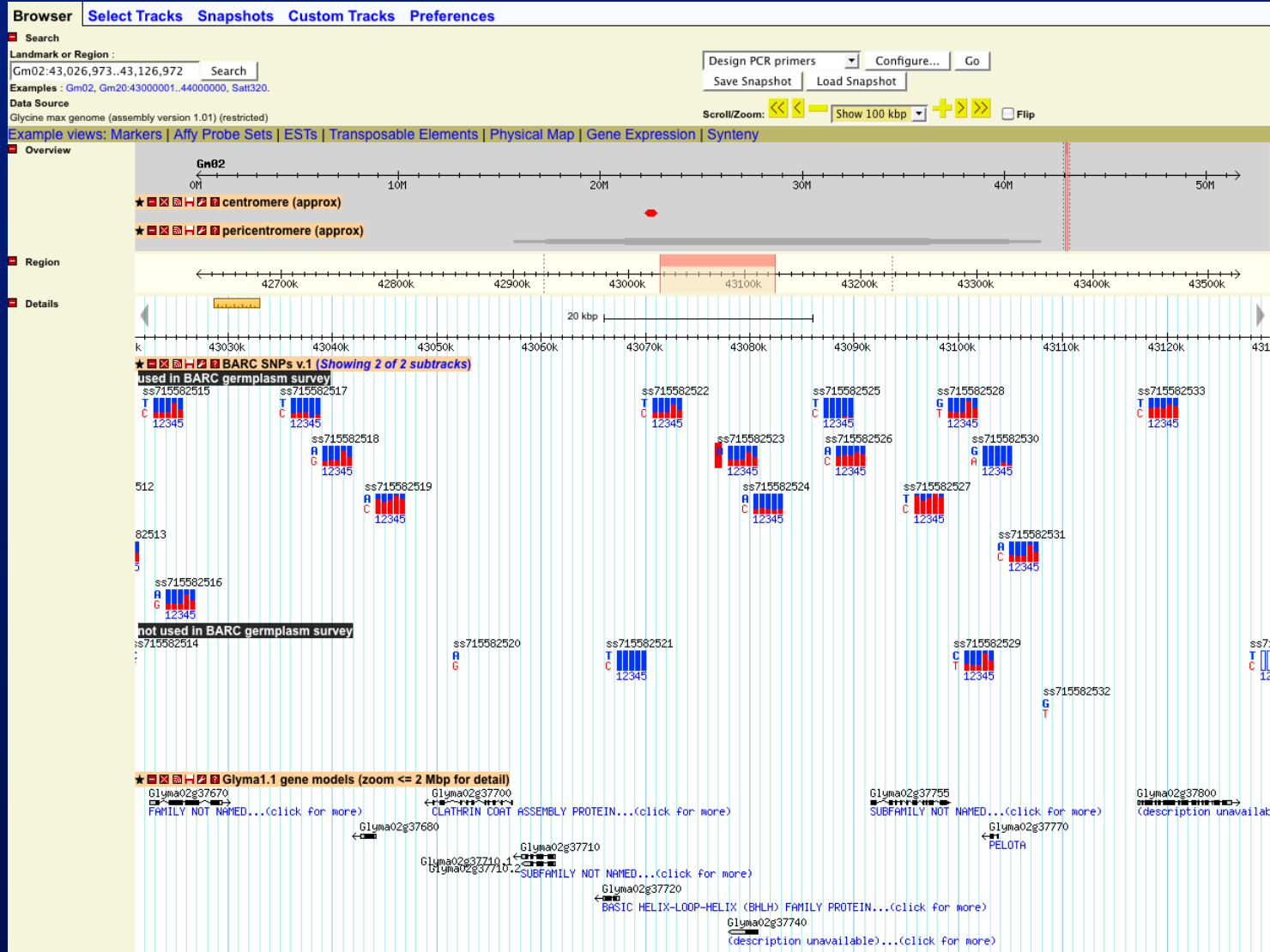
**Region**



**Details**







Glycine max genome (assembly version 1.01): 33 kbp from Gm01:7,341,000..7,373,999

**Browser** [Select Tracks](#) [Snapshots](#) [Custom Tracks](#) [Preferences](#)

**Search**

Landmark or Region:

Gm01:7,341,000..7,373,999

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

**Data Source**

Glycine max genome (assembly version 1.01)

Design PCR primers

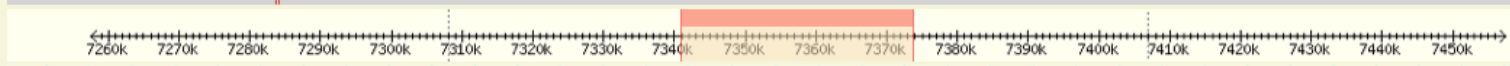
Scroll/Zoom:     Show 33 kbp  Flip

[Example views: Markers](#) | [Affy Probe Sets](#) | [ESTs](#) | [Transposable Elements](#) | [Physical Map](#) | [Gene Expression](#) | [Synteny](#)

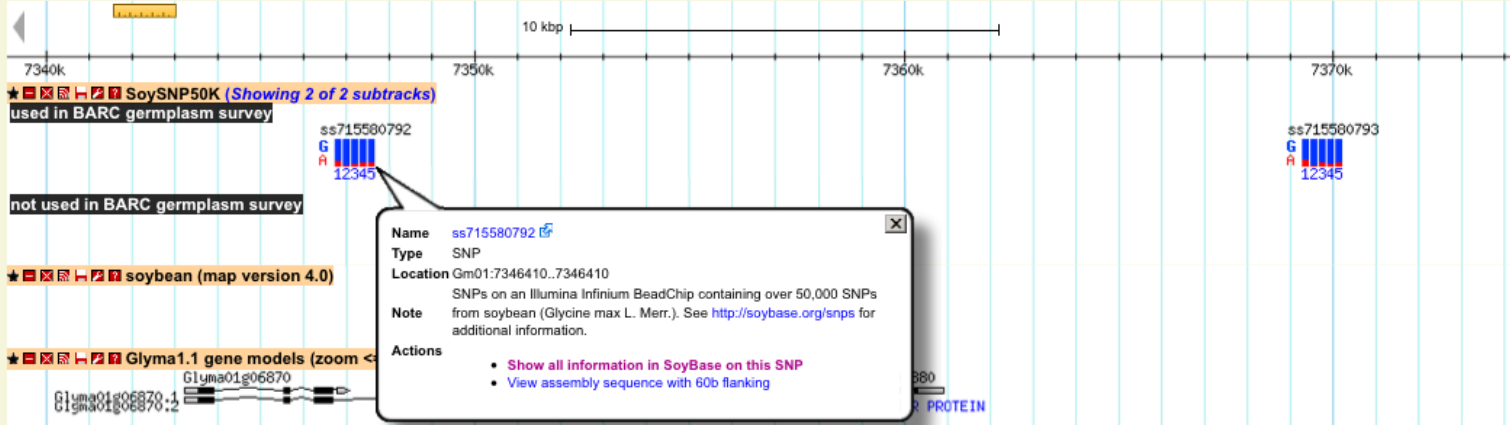
**Overview**



**Region**



**Details**



**Name** [ss715580792](#)

**Type** SNP

**Location** Gm01:7346410..7346410

**Note** SNPs on an Illumina Infinium BeadChip containing over 50,000 SNPs from soybean (Glycine max L. Merr.). See <http://soybase.org/snps> for additional information.

**Actions**

- Show all information in SoyBase on this SNP
- View assembly sequence with 60b flanking

### Report for ss715578909

**Locus type:** SNP  
**Organism:** Glycine max  
**dbSNP name:** ss715578909  
**Alternate name:** BARC\_1.01\_Gm01\_29091115\_A\_G

### Genomic sequence information

**Chromosome:** Gm01  
**SNP position:** 29091115  
**SNP alleles:** A/G  
**Location in gene:** intergenic  
**Genomic context:**     **Start:** 29091055  
                              **End:** 29091175

**Sequence:**

```
GTCCGAGATCGTAGTTTTACTGGAACAAGTTTTGGAGAGGAAGAAGGGAGAATGAATTG
[A/G]GAGGAAGAGAGGGGTGACAAACATAACATATCTTAAATGTAAAACATGGACCTAA
TATGT
```

[Select All](#)

**Genome Assembly Version:** Glyma1.0

### Allele frequency

Population	Minor Allele	Minor Allele Frequency
96 G. max landraces	G	0.427
96 G. max elite cultivars	G	0.431
<b>Gmax192 G. max landrace + elite</b>	G	0.429
96 G. soja accessions	G	0.219
<b>ALL288 G. max landrace + elite cultivars + G. soja</b>	G	0.359

### Allele distribution

Accession	Allele
Wm82	A

The allelic composition for the 19,652 cultivars and accessions used in the BARC Germplasm Survey can be downloaded [here](#).

### Data source

**Data provided by:** Cregan, Perry and Song, Qijian  
**Reference:** [Song et al. 2013a](#)  
**Assay Information:** Illumina Infinium BeadChip

# Conceptual Haplotype Viewer Using A Condensed Graphical Form

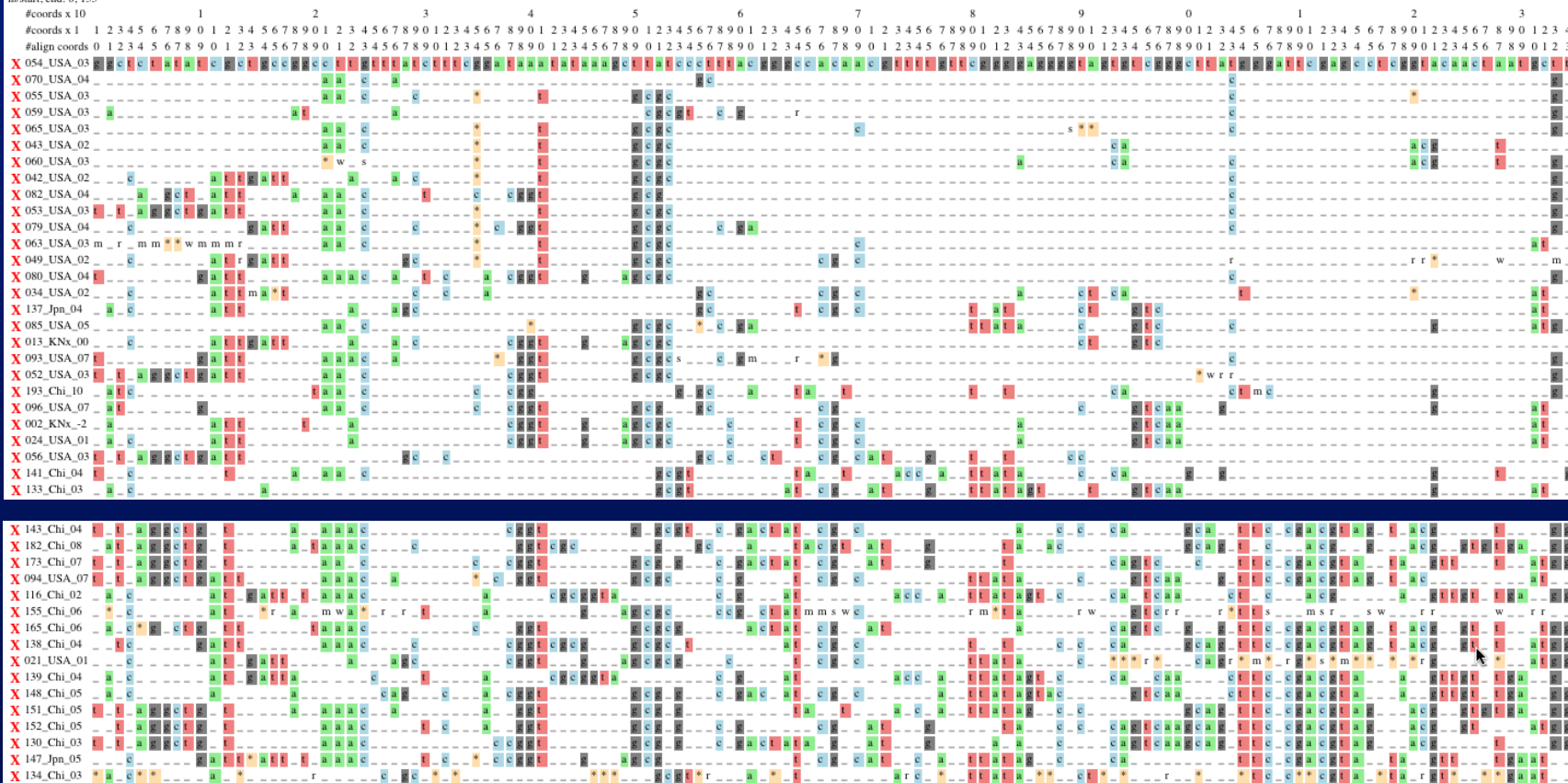
## Hap viewer

Download files: [output file](#) [input file](#) Short URL: <http://bit.ly/16tEu>

Views: [show fasta](#) [show input](#) [Show output](#) [show tree view](#)

[\[Display Settings\]](#)

nlstart\_end: 0..133



# Conceptual Report for a GWAS Experiment

## GWAS-1\_Proten 1

**Experiment Name:** [GWAS Experiment 1](#)  
**Manhattan plot view:** [View Manhattan Plot](#)  
**Trait name:** [Protein content](#)

## SNPs involved in protein content

GWAS ID	SNP ID	Chromosome	Position	Linkage Grp.	Position	Score
<a href="#">GWAS-1_protein 1</a>	<a href="#">BARC-013379-00491</a>	Bn03	31593562	A1	28.5	6.5
<a href="#">GWAS-1_protein 2</a>	<a href="#">BARC-028289-05814</a>	Bn03	45160253	A1	58.9	6.8
<a href="#">GWAS-1_protein 3</a>	<a href="#">BARC-040309-07711</a>	Bn03	48593762	A1	70.2	6.1
<a href="#">GWAS-1_protein 4</a>	<a href="#">BARC-027874-06697</a>	Bn03	54628741	A1	100.9	6.9
<a href="#">GWAS-1_protein 5</a>	<a href="#">BARC-013547-01157</a>	Bn03	56593562	A1	101.1	6.9

## References for the Analysis

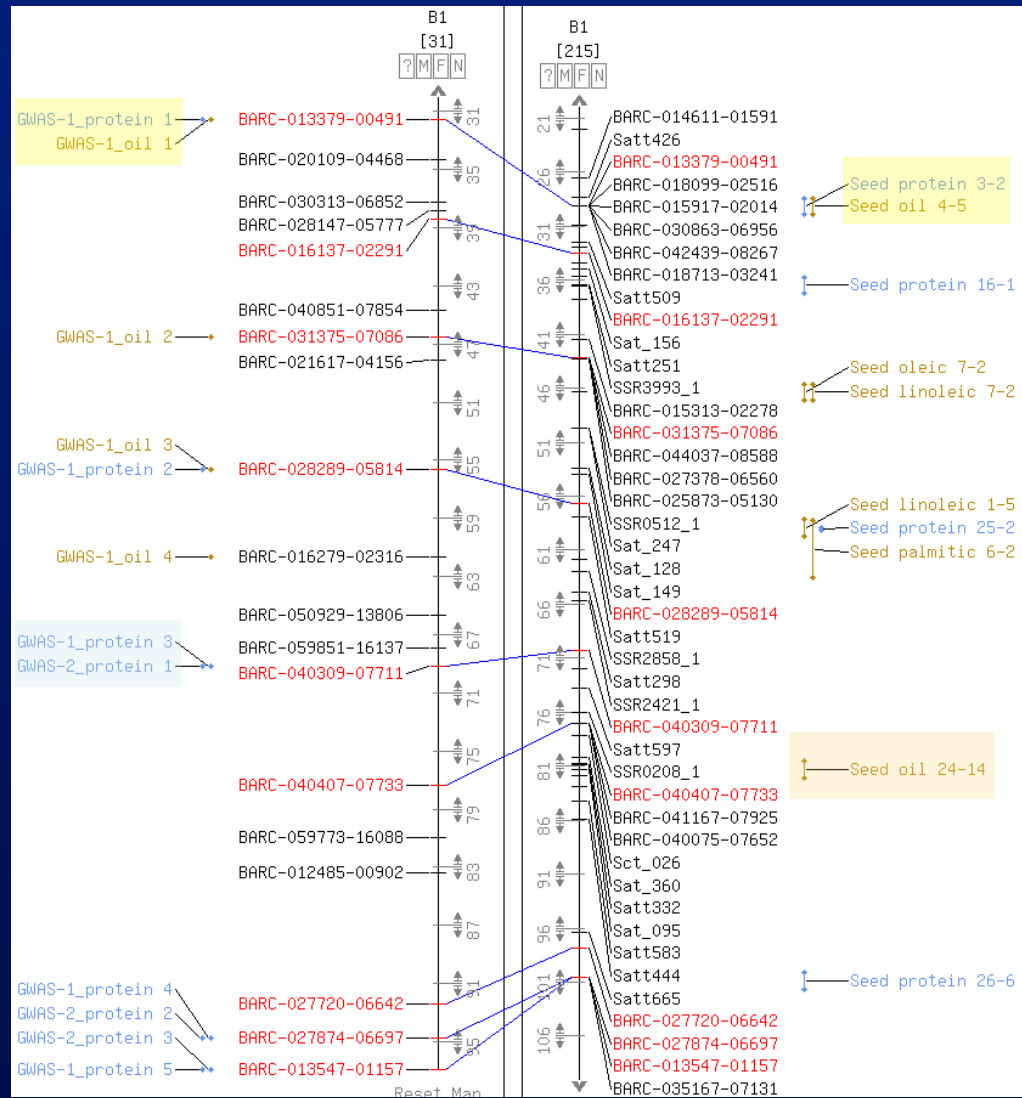
Gore et al. 2014 [Genome wide association study of seed oil in Brassica napus](#)  
[Euphyt. 2014, 152\(1\):67-73](#)

## Comments about the interval

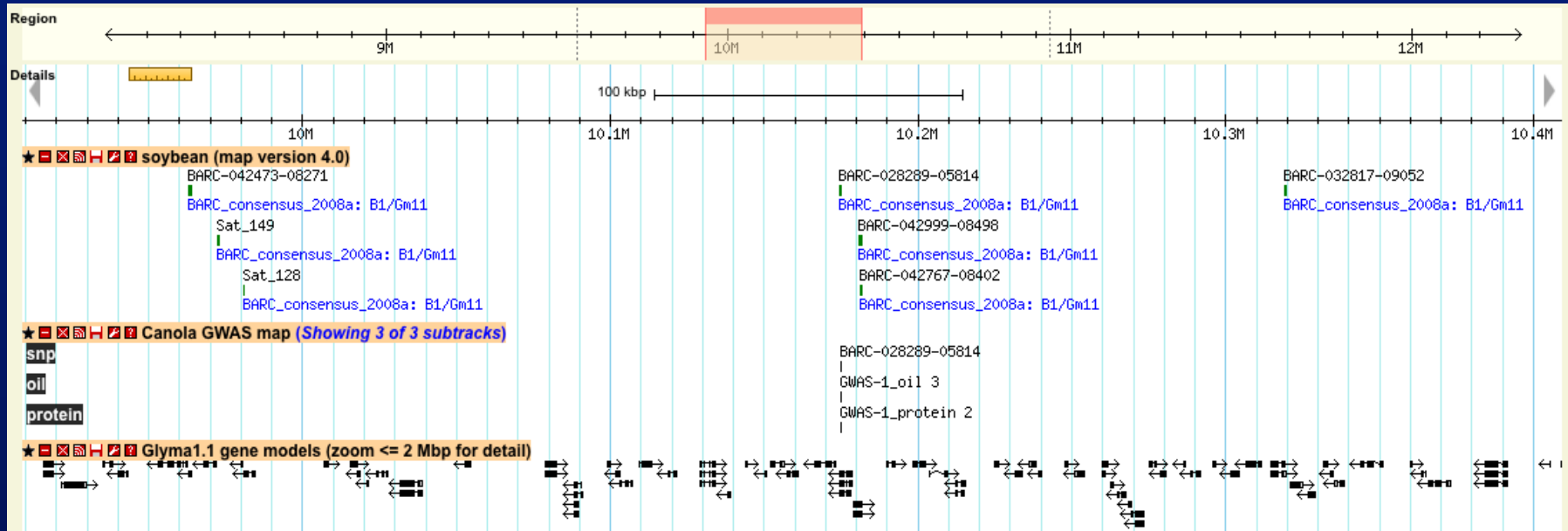
Authors identified Brana03g000200 and Brana03g000300 as candidate genes linked to [GWAS-1\\_protein 4](#) and [GWAS-1\\_protein 5](#)

Seed protein determined by NIR

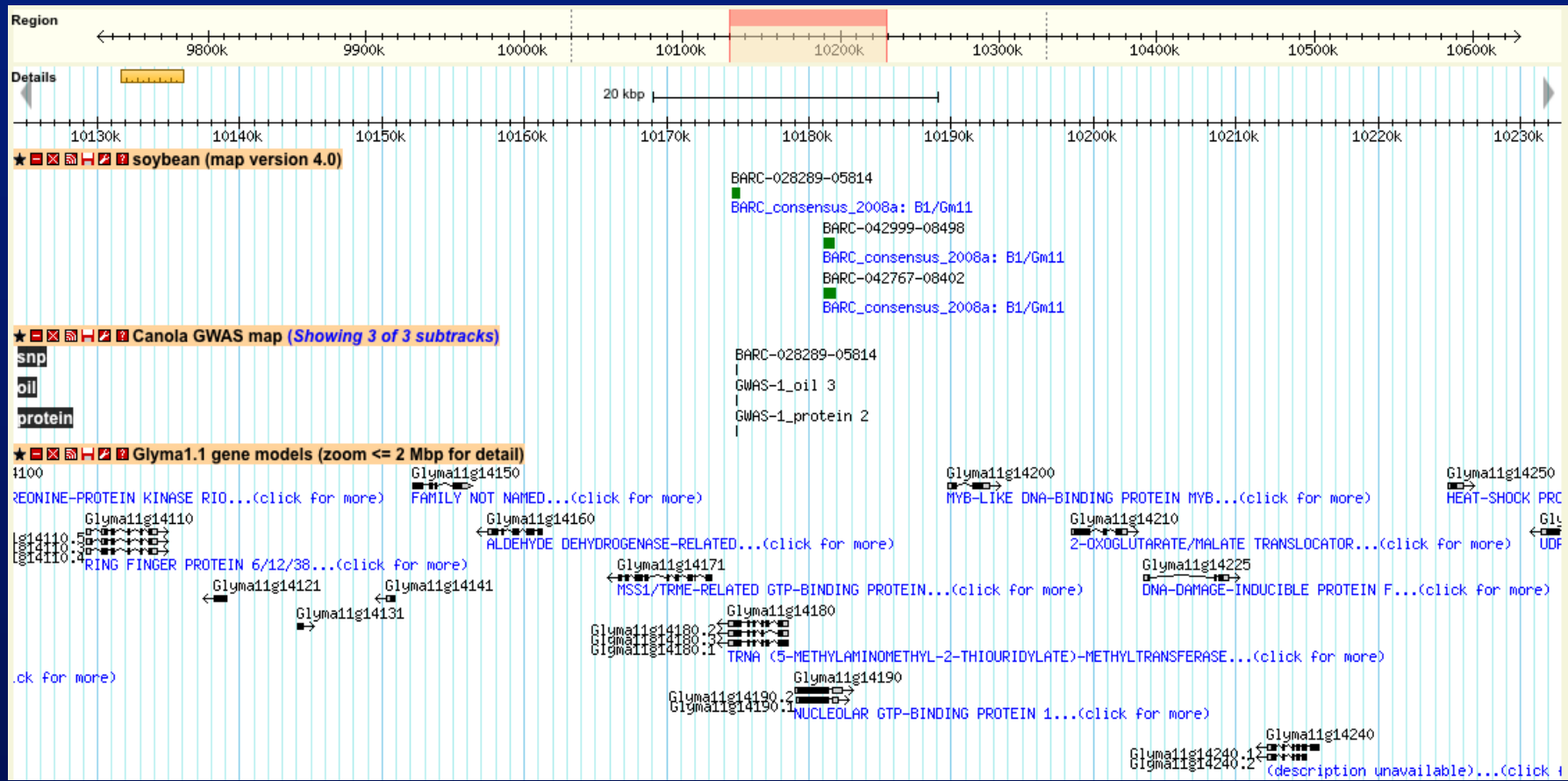
# Trait Associated Snps Shown Relative To The Genetic Map



# Trait Associated Snps Shown Relative To The Genome Sequence









# Questions?

<http://soybase.org>



Glycine max genome (assembly version 1.01): 33 kbp from Gm01:7,341,000..7,373,999

**Browser** [Select Tracks](#) [Snapshots](#) [Custom Tracks](#) [Preferences](#)

Search  
 Landmark or Region:  Search  
 Examples: Gm02, Gm20:43000001..44000000, Satt320.

Data Source  
 Glycine max genome (assembly version 1.01)

Design PCR primers    
 Save Snapshot

Scroll/Zoom:     Show 33 kbp    Flip

Example views: [Markers](#) | [Affy Probe Sets](#) | [ESTs](#) | [Transposable Elements](#) | [Physical Map](#) | [Gene Expression](#) | [Synteny](#)

**Overview**

**Region**

**Details**

Centromere (approx.)  
 Pericentromere (approx.)

7260k 7270k 7280k 7290k 7300k 7310k 7320k 7330k 7340k 7350k 7360k 7370k 7380k 7390k 7400k 7410k 7420k 7430k 7440k 7450k

10 kbp

7340k 7350k 7360k 7370k

SoySNP50K (Showing 2 of 2 subtracks)  
 used in BARC germplasm survey

ss715580792  
 G  
 A  
 12345

not used in BARC germplasm survey

ss715580793  
 G  
 A  
 12345

soybean (map version 4.0)

Glyma1.1 gene models (zoom <= 2 Mbp for detail)

Glyma01g06870  
 Glyma01g06870:1  
 Glyma01g06870:2

Glyma01g06880  
 ZINC FINGER PROTEIN