



United States  
Department of  
Agriculture

National Institute  
of Food and  
Agriculture



Solanaceae Coordinated  
Agricultural Project



# Introduction to the Tomato Genome Browser Part I

Heather L. Merk

The Ohio State University, OARDC

SolCAP workshop

# Overview

## Part 1

- 🍅 What is a genome browser?
- 🍅 Where can I find the tomato genome browser?
- 🍅 What kinds of information can I see and where does it come from?

## Part 2

- 🍅 Demonstrate use of browser to identify markers linked to bacterial spot resistance

# What is a Genome Browser?

- 🍅 We have a LOT of public tomato sequence data
  - > 2500 molecular markers on high density tomato map (*S. lycopersicum* LA925 x *S. pennellii* LA716)
  - > 300 000 ESTs (GenBank – an NIH public database)
  - 2 draft genome sequences (*S. lycopersicum*, *S. pimpinellifolium*)
- 🍅 We also have a lot of data from other species e. g. Arabidopsis, potato
- 🍅 Data rapidly generated
- 🍅 How can we integrate this data?

# What is a Genome Browser?

- 🍅 Wikipedia Definition: “A graphical interface for display of information from a biological database for genomic data. Genome browsers enable researchers to visualize and browse entire genomes with annotated data including gene prediction and structure, proteins, expression, regulation, variation, comparative analysis, etc.” – [en.wikipedia.org/wiki/Genome\\_browser](http://en.wikipedia.org/wiki/Genome_browser) (verified 26 Oct 2010)
- 🍅 There is a LOT of information available!
- 🍅 Easy to get overwhelmed!

# The Tomato Genome Browser

- 🍅 Gbrowse – generic genome browser
- 🍅 Can easily be customized
- 🍅 Helpful introductory tutorial:  
<http://www.openhelix.com/gbrowse>

## GBrowse User Introductory Tutorial

Tutorial and training materials by OpenHelix



Learn to use [GBrowse](#), a web application that allows you to explore genomic sequences together with annotated data. GBrowse is rapidly becoming a genomic browser of choice among organism databases, because the browser is both universal and yet customizable. Once you learn to use GBrowse at one database, you'll be able to use it to view any genome. Results can be customized to show only the data you want to see. The tool is flexible to allow you to upload and incorporate your own unpublished data into the genomic viewer. You'll have fun as you explore a variety of genomes (from paramecia to personal genomics) with the new perspective and detailed annotations that GBrowse provides.

### You'll learn:

- the basic layout and search methods at GBrowse
- how to access detailed annotation data tied to genomic sequences
- how to select and customize annotations using Tracks
- how to upload and incorporate your own data or other external data sources
- take a tour of different GBrowse installations at model organism databases

# Where can I find the tomato genome browser?

<http://solgenomics.net>

The screenshot shows the Sol Genomics Network (SGN) website. At the top, there is a navigation bar with the text "sol genomics network" and a search bar. The search bar contains the text "sol search" and a "search" button. To the right of the search bar are links for "home", "forum", "contact", and "help". Below the search bar are links for "log in" and "new user".

The main content area features several interactive icons and panels:

- Maps & Markers:** A panel showing a chromosome map with markers labeled "CT233", "CO15", and "C2\_ft4g15790".
- Genes:** A panel showing a 3D DNA double helix structure.
- Phenotypes:** A panel showing images of various tomato fruits.
- Breeders Toolbox:** A panel showing a toolbox icon and various tomato fruits.
- Genomes & Sequences:** A panel showing a chromosome map with various colored segments.
- Pathways:** A panel showing a chemical structure diagram of a sugar molecule.

At the bottom of the page, there is a section titled "About SGN" and two columns of news and events:

**News**

**Call for abstracts: Solanaceae workshop, PAG 2011, San Diego**

The Solanaceae workshop at the upcoming PAG 2011 meeting is seeking abstracts, submissions are due September 30, 2010. [September 13, 2010]

**Events**

**SOL 2011**

The SOL meeting 2011 will be held in Tsukuba, Japan, October 16-20, 2011.

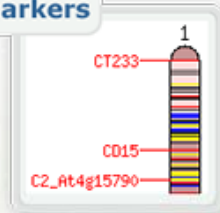
**Plant and Animal Genome XIX Conference**

PAG XIX in San Diego, CA, January 15-19, 2011.

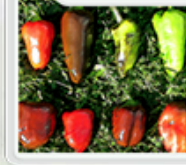


- Browse**
- Tomato genome data
- Projects**
- Solanaceae project (SOL)
- International Tomato Sequencing Project
- Solanum pimpinellifolium* Project (CSHL)
- U.S. tomato sequencing project

Maps & Markers



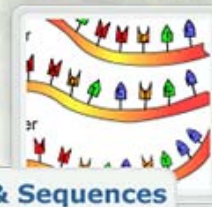
Phenotypes



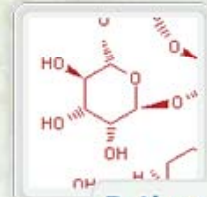
Breeders Toolbox



Genomes & Sequences



Pathways



About SGN

News

Call for abstracts: Solanaceae workshop, PAG 2011, San Diego

The Solanaceae workshop at the upcoming PAG 2011 meeting is seeking abstracts, submissions are due September 30, 2010. [September 13, 2010]

New SOL-100 page released

Events

SOL 2011

The SOL meeting 2011 will be held in Tsukuba, Japan, October 16-20, 2011.

Plant and Animal Genome XIX Conference

PAG XIX in San Diego, CA, January 15-19, 2011. Solanaceae session call for abstracts



### Tomato Genome Data

#### Tomato genome sequence builds

Release	Date	Description	Annotation	Download
<b>1.00</b>	Dec 2009	initial build, based on the Newbler assembler and containing only 454 sequencing data	ITAG1	scaffolds proteins cds
<b>1.03</b>	Jan 2010	like 1.00, but with additional 454 runs and improved contamination screen	Not annotated	scaffolds
<b>cabog1.00</b>	Mar 2010	All 454 data, bac end and fosmid end data, assembled using the CABOG assembler	Not annotated	scaffolds
<b>1.50</b>	Apr 2010	Includes all 454 data, bac ends, fosmid ends, polishing with Solexa and SOLID data	Not annotated	scaffolds
<b>2.00</b>	Jun 2010	Release withdrawn.	Not annotated	-
<b>2.10</b>	Jun 2010	Additional scaffold merging using clone end sequences. Scaffolds placed and oriented using multiple physical maps, first release to include chromosome pseudomolecule sequences.	Not annotated	scaffolds, chromosomes
<b>2.30</b>	Aug 2010	Integration and polishing of tomato BAC sequences	in progress	scaffolds, chromosomes

#### Official annotation

[browse genome contigs and official annotations](#)

The official annotation for the tomato genome is provided by the [International Tomato Annotation Group \(ITAG\)](#), a multinational consortium, funded in part by the [EU-SOL project](#).



#### ITAG1 (official release)

last modified Oct 22, 2010

#### Bulk download

[\[FTP\]](#)

#### ITAG1 Release: genomic annotations

**Browse or**

#### ITAG1 Release: protein annotations

**Browse or**



# At Last... The Tomato Genome Browser

The screenshot displays the Sol Genomics Network browser interface. The browser window title is "ITAG1 Release: genomic annotations: 969 bp from SL1.00sc06004:457,840..458,808". The address bar shows the URL "http://solgenomics.net/gbrowse/bin/gbrowse/ITAG1\_genomic/" and the search bar contains "sol genomics network".

The page header includes the Sol Genomics Network logo and navigation links: "home | forum | contact | help". Below the logo are tabs for "search", "maps", "genomes", and "tools", along with a "sol search" button and links for "log in" and "new user".

The main content area is titled "ITAG1 Release: genomic annotations: 969 bp from SL1.00sc06004:457,840..458,808". It features a "Browser" menu with options: "Select Tracks", "Upload and Share Tracks", and "Preferences".

The "Search" section includes a "Landmark or Region:" input field with "SL1.00sc06004:457840.." and a "Search" button. Below it are "Examples: SL1.00sc00002\_16.1.1, Sulfite oxidase, TG154, C2\_At3g02060." and a "Data Source" dropdown menu set to "ITAG1 Release: genomic annotations".

The "Overview" section shows a genomic map with a scale from 0M to 3M. The "Region" section provides a detailed view of the 969 bp region, with a scale from 360k to 550k. The "Genes" section lists several genes, including SL1.00sc06004\_46.1, SL1.00sc06004\_47.1, SL1.00sc06004\_48.1, SL1.00sc06004\_49.1, SL1.00sc06004\_50.1, SL1.00sc06004\_51.1, SL1.00sc06004\_52.1, SL1.00sc06004\_53.1, SL1.00sc06004\_54.1, SL1.00sc06004\_55.1, SL1.00sc06004\_56.1, SL1.00sc06004\_57.1, SL1.00sc06004\_58.1, SL1.00sc06004\_59.1, SL1.00sc06004\_60.1, SL1.00sc06004\_61.1, SL1.00sc06004\_62.1, SL1.00sc06004\_63.1, SL1.00sc06004\_64.1, SL1.00sc06004\_65.1, SL1.00sc06004\_66.1, and SL1.00sc06004\_67.1.

The "Details" section shows a zoomed-in view of the 969 bp region, with a scale from 457900 to 450000. It includes "Gene models" for SL1.00sc06004\_58.1.1, "CDS - click to browse protein" for SL1.00sc06004\_58.1.1, and "ESTs and cDNAs - Tomato".

# There's a lot of information!

## What am I looking at?

ITAG1 Release: genomic annotations: 969 bp from SL1.00sc06004:457,840..458,808 ← Overview

Browser [Select Tracks](#) [Upload and Share Tracks](#) [Preferences](#)

**Search**

Landmark or Region:   Annotate Restriction Sites

Examples: [SL1.00sc00002\\_16.1.1](#), [Sulfite oxidase](#), [TG154](#), [C2\\_At3g02060](#).

Data Source:  Scroll/Zoom:

**Overview**

SL1.00sc06004

0M 1M 2M 3M

**Region**

360k 370k 380k 390k 400k 410k 420k 430k 440k 450k 460k 470k 480k 490k 500k 510k 520k 530k 540k 550k

**Genes**

46.1SL1.00sc06004\_51.1 SL1.00sc06004\_56.1 SL1.00sc06004\_58.1 SL1.00sc06004\_61.1 SL1.00sc06004\_66.1

SL1.00sc06004\_47.1 SL1.00sc06004\_53.1 SL1.00sc06004\_59.1 SL1.00sc06004\_63.1 SL1.00sc06004\_67.1

SL1.00sc06004\_48.1 SL1.00sc06004\_54.1 SL1.00sc06004\_60.1 SL1.00sc06004\_62.1

SL1.00sc06004\_49.1 SL1.00sc06004\_55.1 SL1.00sc06004\_64.1

SL1.00sc06004\_50.1 SL1.00sc06004\_57.1 SL1.00sc06004\_65.1

SL1.00sc06004\_52.1

**Details**

SL1.00sc06004: 969 bp

200 bp

457900 458000 458100 458200 458300 458400 458500 458600 458700 458800

**Gene models**

SL1.00sc06004\_58.1.1

Helix-loop-helix DNA-binding (AHRD V1 \*-\*NG Q2HVX8\_MEDTR); contains In...

**CDS - click to browse protein**

SL1.00sc06004\_58.1.1

**ESTs and cDNAs - Tomato**

← Search

← Source

← Scroll

**I'm scared! Get me outta here!**



# Scales

ITAG1 Release: genomic annotations: 969 bp from SL1.00sc06004:457,840..458,808

Browser [Select Tracks](#) [Upload and Share Tracks](#) [Preferences](#)

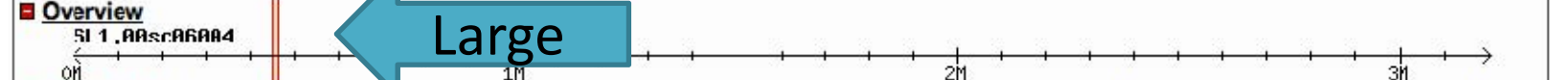
Search

Landmark or Region:

Examples: [SL1.00sc00002\\_16.1.1](#), [Sulfite oxidase](#), [TG154](#), [C2\\_A13g02060](#).

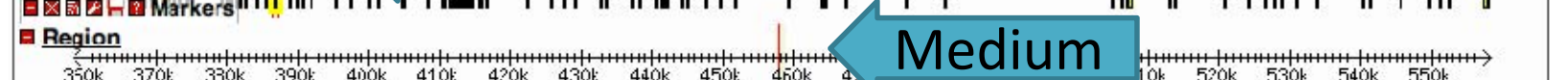
Data Source:     Flip

Overview



Markers


Region



Genes

SL1.00sc06004\_47.1 SL1.00sc06004\_50.1 SL1.00sc06004\_58.1 SL1.00sc06004\_51.1 SL1.00sc06004\_66.1  
SL1.00sc06004\_48.1 SL1.00sc06004\_54.1 SL1.00sc06004\_59.1 SL1.00sc06004\_60.1 SL1.00sc06004\_67.1  
SL1.00sc06004\_49.1 SL1.00sc06004\_55.1 SL1.00sc06004\_62.1 SL1.00sc06004\_64.1  
SL1.00sc06004\_50.1 SL1.00sc06004\_57.1 SL1.00sc06004\_65.1  
SL1.00sc06004\_52.1

Details



SL1.00sc06004: 969 bp

Gene models

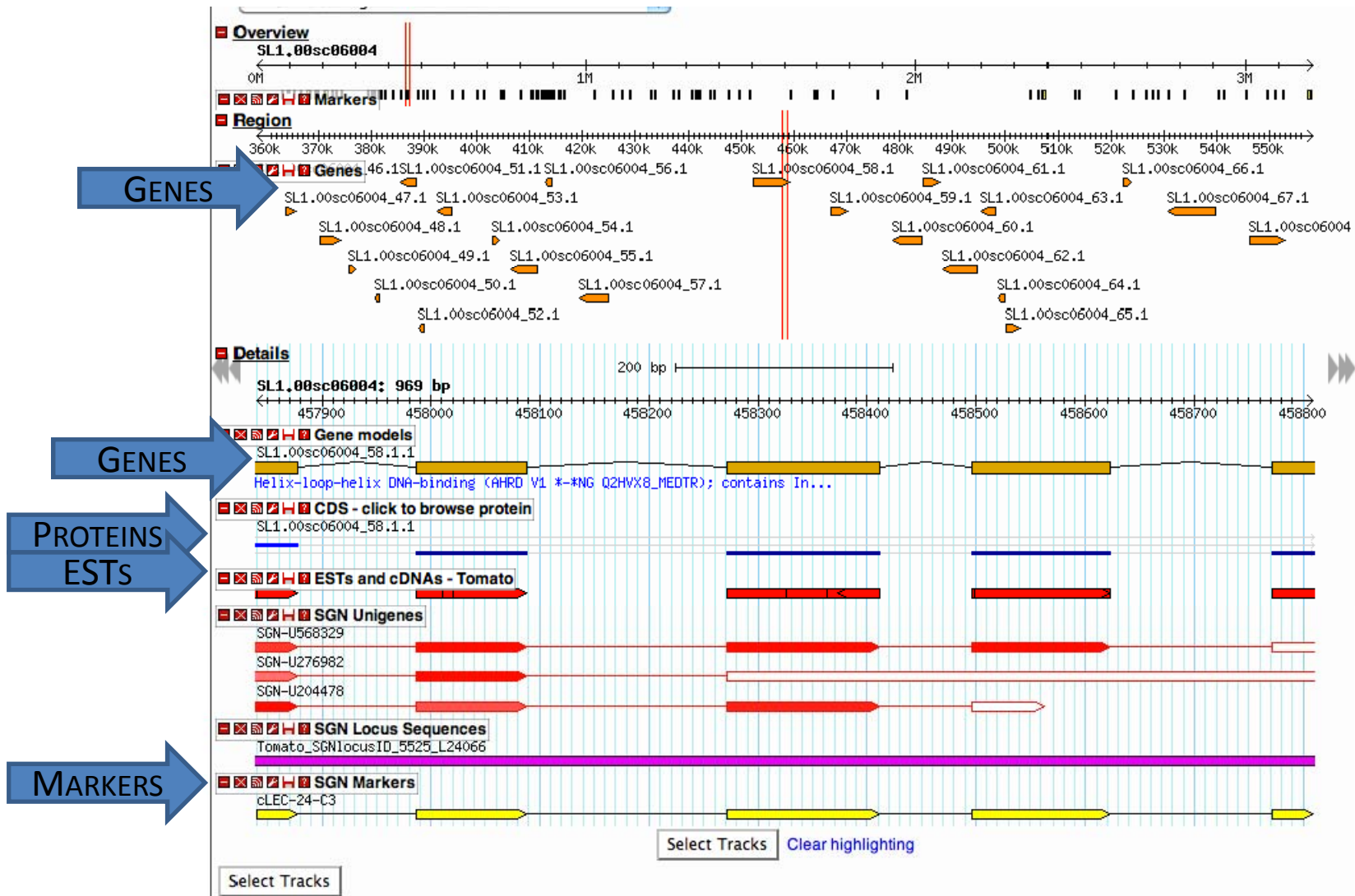
SL1.00sc05004\_58.1.1  
Helix-loop-helix DNA-binding (HRD VL \*-4NG Q2HVX8\_MEDTR); contains In...

CDS - click to browse protein

SL1.00sc05004\_58.1.1

ESTs and cDNAs - Tomato

# Tracks – Types of Data to View



# Wait! – Where Does the Data Come From?

## ITAG – International Tomato Annotation Group

- Predicted genes and proteins – using prediction software
- Based on tomato, other Solanaceae, Arabidopsis

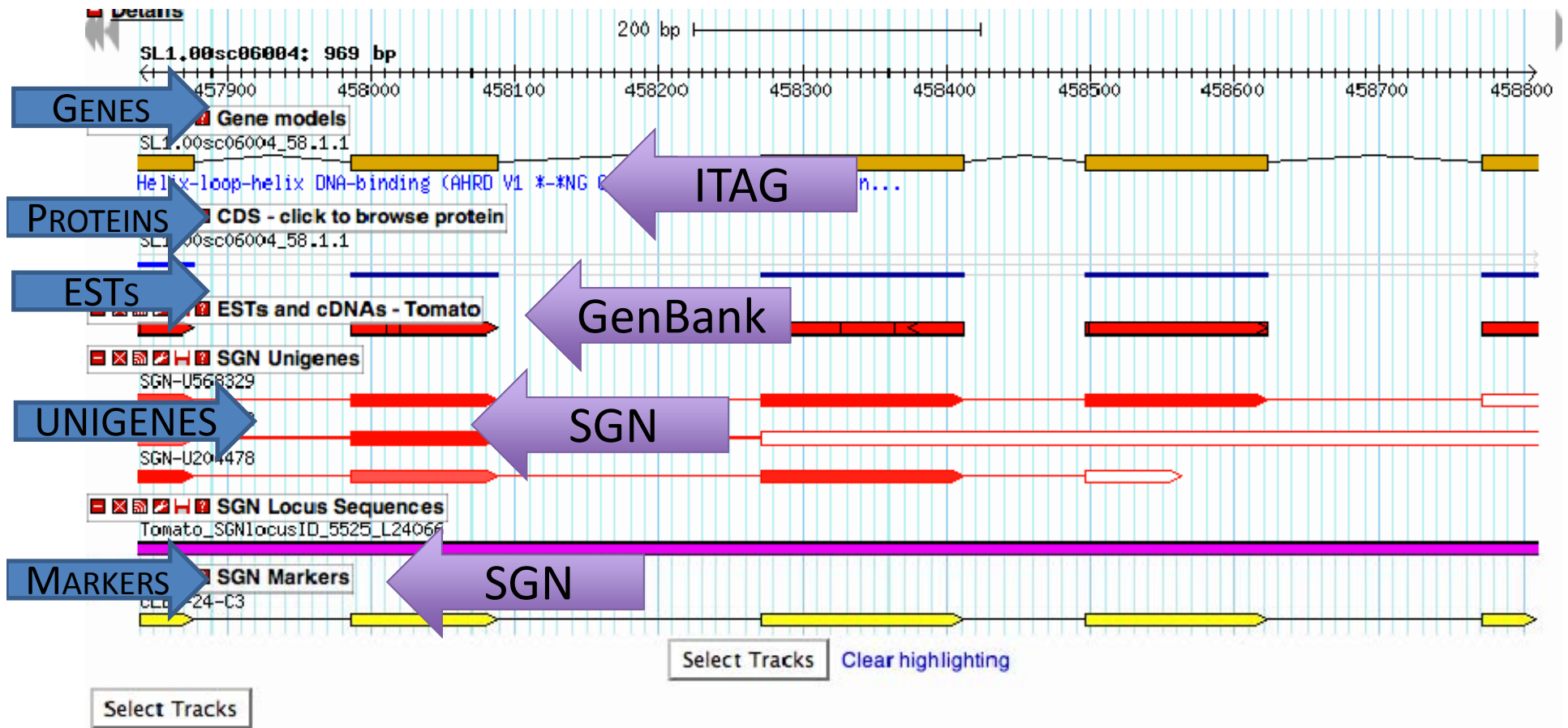
## GenBank

- ESTs
- May or may not support predicted genes/proteins

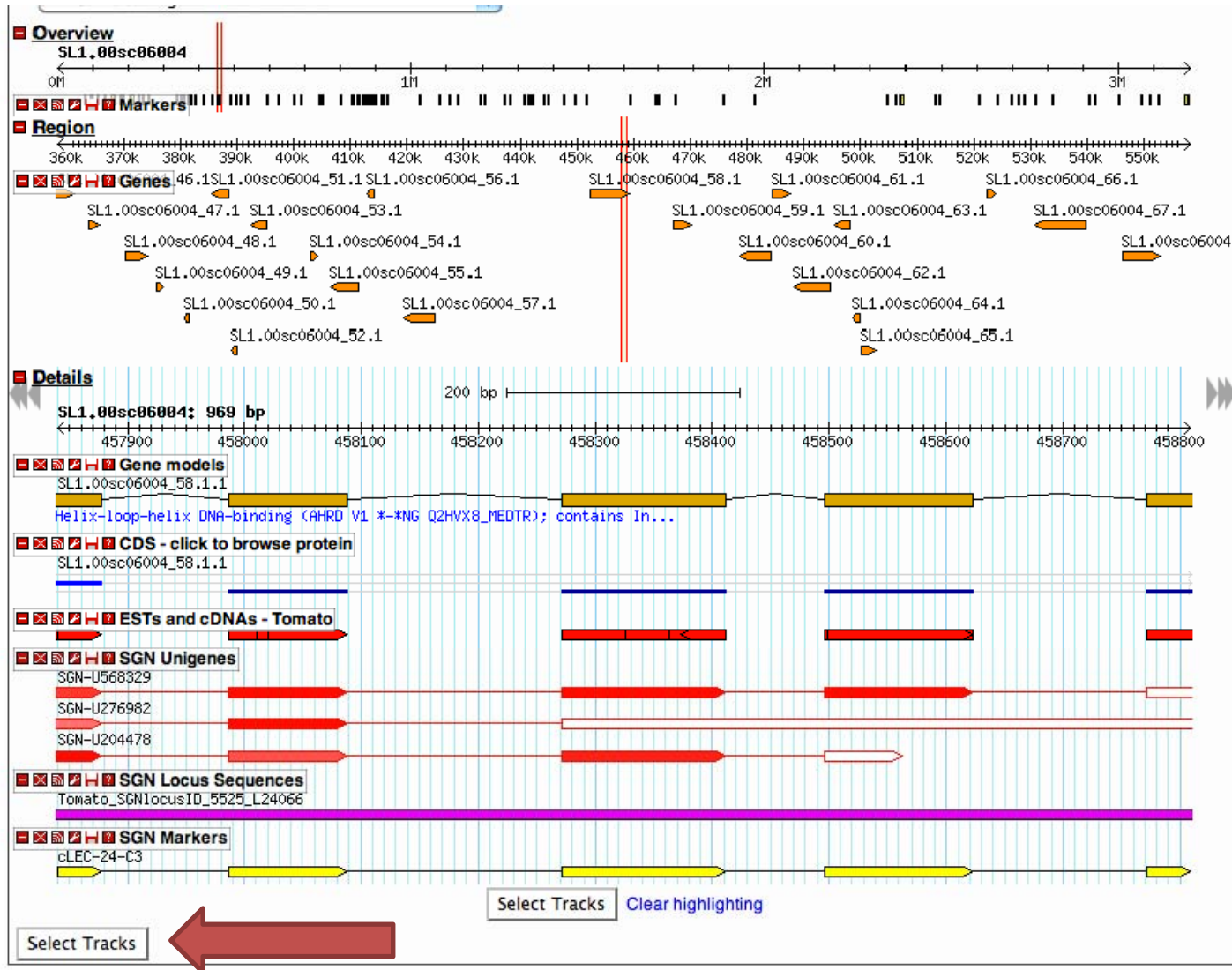
## SGN – Sol Genomics Network

- Markers

# Tracks - Data Sources



# Select Tracks



# Select Tracks

ITAG1 Release: genomic annotations: 20 kbp from SL1.00sc06004:490,001..510,000

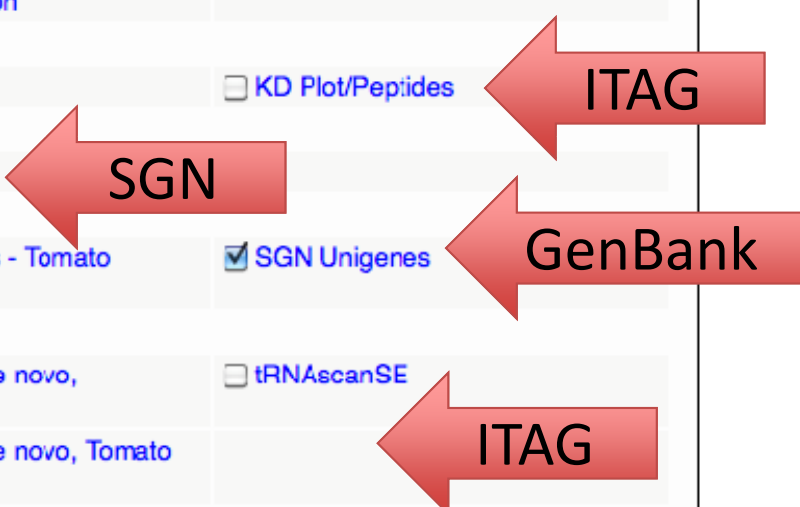
Browser **Select Tracks** Upload and Share Tracks Preferences

<< Back to Browser

**Tracks**

- Overview**  All on  All off
  - Markers  tRNAscanSE
- Region**  All on  All off
  - GC Content  Genes
- DNA**  All on  All off
  - DNA/GC Content  6-frame translation
- Gene models**  All on  All off
  - CDS - click to browse protein  Gene models  KD Plot/Peptides
- Genetic Loci**  All on  All off
  - SGN Locus Sequences  SGN Markers
- Genome data and reagents**  All on  All off
  - ESTs and cDNAs - Other Solanaceae  ESTs and cDNAs - Tomato  SGN Unigenes
- Prediction Features**  All on  All off
  - AUGUSTUS (de novo, Tomato trained)  GlimmerHMM (de novo, Arabidopsis trained)  tRNAscanSE
  - GeneID (de novo, Tomato trained)  GlimmerHMM (de novo, Tomato trained)
  - GeneMark (de novo, Arabidopsis trained)  Infernal
- Analysis**  All on  All off
  - Restriction Sites

Back to Browser



**Much Better! I can do this!**



# Conclusion – What We Learned

- 🍅 What a genome browser is
- 🍅 Where to find the tomato genome browser
- 🍅 The kinds of information we can look at using the tomato genome browser
- 🍅 Where information in the browser comes from



United States  
Department of  
Agriculture

National Institute  
of Food and  
Agriculture



Solanaceae Coordinated  
Agricultural Project



# Using the Tomato Genome Browser: An Example Part 2

Heather L. Merk

The Ohio State University, OARDC

SolCAP workshop



# Overview

- 🍅 How can the genome browser help ME?
- 🍅 Example – finding markers associated with bacterial spot resistance

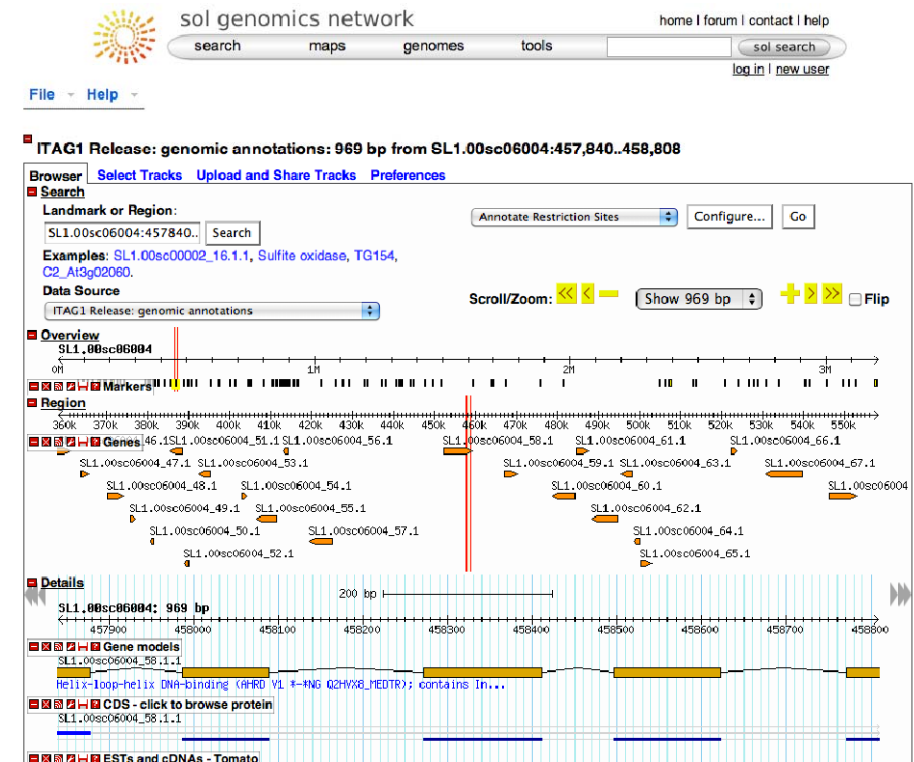


Photo courtesy of D. Francis, Ohio State University

# How can the genome browser help ME?

- 🍅 Information from multiple databases/data sources is combined in browser
- 🍅 What does the genome look like around a marker, EST, BAC, other sequence, etc?
- 🍅 Tool to help guide further research – e.g. identify candidate genes, identify potential molecular markers

# Tomato Bacterial Spot Resistance

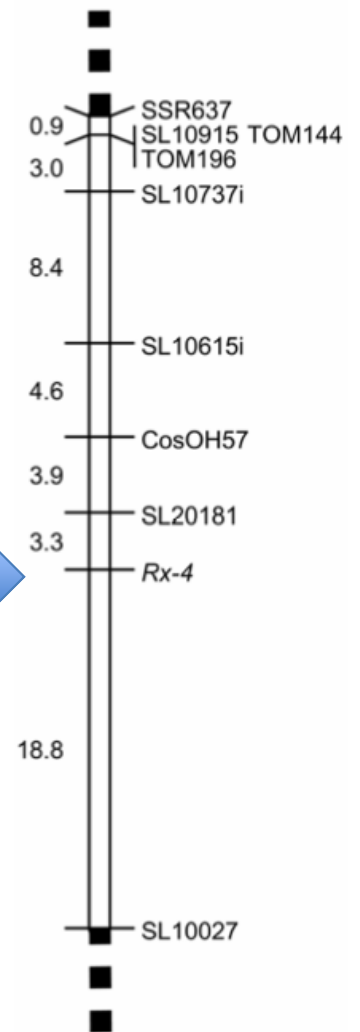
🍅 Robbins et al. (2009)

🍅 PI 128216 confers resistance to race T3

🍅 Rx-4 mapped to chr. 11

🍅 SL20181 – SNP in tomato unigene SGN-U567870

🍅 We can use genome browser to find more markers – for fine mapping, marker-assisted selection



# Step 1: Search

- 🍅 A. Type SGN-U567870 in search box
- 🍅 B. Data source - ITAG1 Release: Genomic Annotation
- 🍅 C. Select tracks to view
- 🍅 D. Search

[File](#) ▾ [Help](#) ▾

## ITAG1 Release: genomic annotations

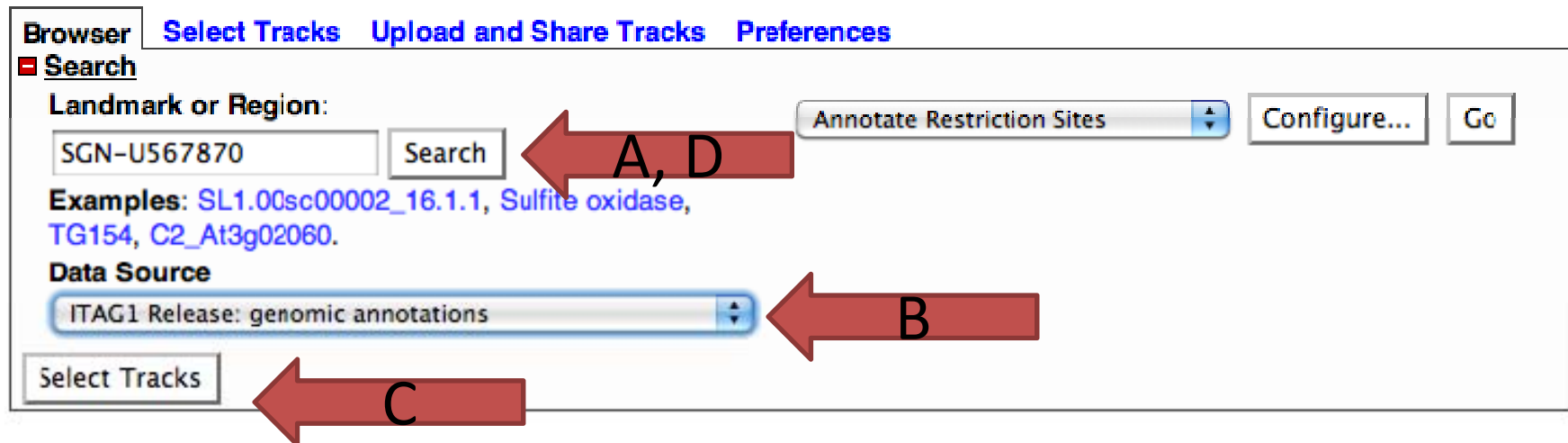
**Browser** [Select Tracks](#) [Upload and Share Tracks](#) [Preferences](#)

**Search**

Landmark or Region:

Examples: [SL1.00sc00002\\_16.1.1](#), [Sulfite oxidase](#), [TG154](#), [C2\\_At3g02060](#).

Data Source:



# Step 1c: Select Tracks



sol genomics network

[home](#) | [forum](#) | [contact](#) | [help](#)

[search](#)

[maps](#)

[genomes](#)

[tools](#)

[log in](#) | [new user](#)

[File](#) ▾ [Help](#) ▾

ITAG1 Release: genomic annotations: 6.94 kbp from SL1.00sc06004:224,489..231,428

[Browser](#) [Select Tracks](#) [Upload and Share Tracks](#) [Preferences](#)

<< [Back to Browser](#)

Tracks

**Overview**  All on  All off

Markers

tRNAscanSE

**Region**  All on  All off

GC Content

Genes

**DNA**  All on  All off

DNA/GC Content

6-frame translation

**Gene models**  All on  All off

CDS - click to browse protein

Gene models

KD Plot/Peptides

**Genetic Loci**  All on  All off

SGN Locus Sequences

SGN Markers

MARKERS - SGN

**Genome data and reagents**  All on  All off

ESTs and cDNAs - Other  
Solanaceae

ESTs and cDNAs - Tomato

SGN Unigenes

ESTs - GENBANK

**Prediction Features**  All on  All off

AUGUSTUS (de novo, Tomato trained)

GlimmerHMM (de novo, Arabidopsis trained)

tRNAscanSE

GeneID (de novo, Tomato trained)

GlimmerHMM (de novo, Tomato trained)

GeneMark (de novo, Arabidopsis trained)

Infernal

**Analysis**  All on  All off

Restriction Sites

# Step 2: Orientation

🍅 Look at scales for orientation

ITAG1 Release: genomic annotations: 6.94 kbp from SL1.00sc06004:224,489..231,428

Browser [Select Tracks](#) [Upload and Share Tracks](#) [Preferences](#)

Search

Landmark or Region:

Examples: [SL1.00sc00002\\_16.1.1](#), [Sulfite oxidase](#), [TG154](#), [C2\\_At3g02060](#).

Data Source:     Flip

Overview

SL1.00sc06004 **LARGE**

Region

130k 140k 150k 160k 170k 180k 190k 200k 210k 220k 230k **MEDIUM** 30k 290k 300k 310k 320k

Details

SL1.00sc06004: 6.94 kbp

2 kbp **SMALL**

SGN Unigenes

SGN-U567870 **UNIGENE** SGN-U282751 SGN-U2

SGN Locus Sequences

SGN Markers

TG286 TG286

ESTs and cDNAs - Tomato

# Step 3: Zoom Out


ITAG1 Release: genomic annotations: 6.94 kbp from SL1.00sc06004:224,489..231,428

Browser [Select Tracks](#) [Upload and Share Tracks](#) [Preferences](#)

Search

Landmark or Region:

Examples: [SL1.00sc00002\\_16.1.1](#), [Sulfite oxidase](#), [TG154](#), [C2\\_At3g02060](#).

Data Source   Scroll/Zoom:       Flip

Overview

SL1.00sc06004

Region

130k 140k 150k 160k 170k 180k 190k 200k 210k 220k 230k 240k 250k 260k 270k 280k 290k 300k 310k 320k

Details

SL1.00sc06004: 6.94 kbp

2 kbp

225k 226k 227k 228k 229k 230k 231k

SGN Unigenes

SGN-U567870

SGN-U282751

SGN-U2

SGN Locus Sequences

SGN Markers

TG286

Select Tracks

Select Tracks

# Step 4: Locate Markers

ITAG1 Release: genomic annotations: 289.9 kbp from SL1.00sc06004:105,056..394,945

Browser [Select Tracks](#) [Upload and Share Tracks](#) [Preferences](#)

Search

Landmark or Region:

Examples: [SL1.00sc000C2\\_16.1.1](#), [Sulfite oxidase](#), [TG154](#), [C2\\_At3g0206C](#).

Data Source:

Scroll/Zoom:

Overview

Region

Details

SL1.00sc06004: 289.9 kbp

SGN Unigenes

SGN Locus Sequences

SGN Markers

← UNIGENE →

22 MARKERS →

Select Tracks Clear highlighting

Select Tracks

# Step 5: Click on Marker

ITAG1 Release: genomic annotations: 289.9 kbp from SL1.00sc06004:105,056..394,945

Browser [Select Tracks](#) [Upload and Share Tracks](#) [Preferences](#)

Search

Landmark or Region:

Examples: [SL1.00sc00002\\_16.1.1](#), [Sulfite oxidase](#), [TG154](#), [C2\\_At3g02060](#).

Data Source:     Flip

Overview

SL1.00sc06004

Region

Details

SL1.00sc06004: 289.9 kbp

SGN Unigenes

SGN Locus Sequences

SGN Markers

SSR406

Potato\_SGNlocusID\_2191\_SGN-U273550

Tobacco\_SGNlocusID\_8183

C2\_At2g27470

C2\_At2g27450

TG121

TG46

TG286

T1297

C2\_At2g27310

C2\_At2g27290

cLEF-51-69

C2\_At2g27730

C2\_At3g44380

TG46

TG286

C2\_At5g22340

cLET-10-011

T1094

cLEX-14-

SSR406

T0979

C2\_At1g30825

C2\_At5g22350

C2\_At5g22210

C2\_At5g40950

C2\_At3g54470

Select Tracks

Clear highlighting



# Step 7: Locate primers



sol genomics network

[home](#) | [forum](#) | [contact](#) | [help](#)

[search](#)

[maps](#)

[genomes](#)

[tools](#)

[sol search](#)

[log in](#) | [new user](#)

## SGN Marker SSR406

SGN-M1079

### SSR info

**Annealing temperatures: Low: 50 High: 55**  
**Repeat motif: TC Repeat number: 15**  
**Forward primer: ACCTGTGGGATCGACCTAGT**  
**Reverse primer: GCTTGTGGGTGCATAACCTT**  
**Predicted size: 198**



### Derivations

This marker was derived from [EST read SGN-E8107](#)

### Mapped locations

None

### Other PCR data

PCR data	Exp. ID 32092	
<b>Forward primer (5'-3')</b> ACCTGTGGGATCGACCTAGT	<b>Accessions and product sizes</b>	<b>Approximate temperature</b> 55°C
<b>Reverse primer (5'-3')</b> GCTTGTGGGTGCATAACCTT		<b>Mg<sup>2+</sup> concentration</b> Unknown

### Genomic locations

#### Annotation Dataset

ITAG1 Release: genomic annotations

#### Location(s)


SL1.00sc06741:390261..390359

# BUT...

- 🍅 We still need to determine whether these markers are polymorphic in our population
- 🍅 We can use BLAST (Basic Local Alignment Search Tool) to look for sequence differences between our sequence and others
  - May give us an indication of likelihood of polymorphism in our population

# Locating sequence

Marker was derived from an EST  
Click on EST read to obtain sequence

 sol genomics network [home](#) | [forum](#) | [contact](#) | [help](#)

[search](#) [maps](#) [genomes](#) [tools](#)  [sol search](#)

[log in](#) | [new user](#)


### SGN Marker SSR406

SGN-M1079

#### SSR info

**Annealing temperatures: Low: 50 High: 55**  
**Repeat motif: TC Repeat number: 15**  
**Forward primer: ACCTGTGGGATCGACCTAGT**  
**Reverse primer: GCTTGTGGGTGCATAACCTT**  
**Predicted size: 198**

#### Derivations

This marker was derived from [EST read SGN-E8107](#) 

#### Mapped locations

None

#### Other PCR data

PCR data	Exp. ID 32092	
<b>Forward primer (5'-3')</b> ACCTGTGGGATCGACCTAGT	<b>Accessions and product sizes</b>	<b>Approximate temperature</b> 55°C
<b>Reverse primer (5'-3')</b> GCTTGTGGGTGCATAACCTT		<b>Mg<sup>2+</sup> concentration</b> Unknown

#### Genomic locations

Annotation Dataset	Location(s)
ITAG1 Release: genomic annotations	<a href="#">SL1.00sc06741:390261..390359</a>



### EST details - SGN-E8107

#### Search information

**Request:** SGN-E8107  
**Request From:** web user

**Match:** SGN-E8107  
**Match Type:** EST sequence internal identifier

#### Clone information

**SGN ID:** SGN-C9975  
**Library Name:** cLEC

**Clone name:** cLEC-67-I15  
**Organism:** *Solanum lycopersicum* (formerly *Lycopersicon esculentum*)

[Order Clone](#)

**Tissue:** combined undifferentiated and shooting callus  
**Development Stage:** 7-10 days post germination

**Microarray:** This clone is not found on any microarray  
There is no map position defined on SGN for this EST or others in the same unigene.

#### Additional sequencing

**Clone:** SGN-C9975 [cLEC-67-I15] **Trace:** SGN-T33613 **EST:** SGN-E211387 **Direction:** 5' **Facility:** TIGR

[\[Show information hierarchy\]](#)

#### Sequence

## SEQUENCE

**Sequence ID:** SGN-E8107  
**Status:** Legacy Chimera not assessed

**Length:** 703 bp (called/trimmed by facility)  
**Direction:** 5'

```
>SGN-E8107 [] (called/trimmed by facility)
CATCACCACCACCGCGCCGCGCTGCGCCACCACGAGTTCAGAGAAGCGCCTGAATTCTACAACCTCAGCTGAATGTCCTTCAATTGAAAATA
CCCATGAATTAATTTGCTCCGATGAAGCTGTTTCATGCGAGCAATGACACTCGATTCGCGTTATATCCGTTGATCAATGGCGGCTATTCTCTCAATT
CTTCAACATTTCTTGGCCACAAAACATAATTTTCACTTCGTTGCTTCCGATCCGCGAGACGGTCACATCTACGCGCCACCATTTTAGCTTC
ATTCCTTATTTGAAATTCGAAGTTTATCGATTCGACGATTCCTCTGTTTCCGGTTTAAATTTCAACTTCTATTCGTTTCAGCTCTCGATTGCCCTT
TGAACATGCGGAGAAGTTATTTAGCTAATATTTTACCTTTATGTTGTTCCGAAAGTTGTTGTTATTTGGATTCCGATTTGGTTTATTTAGTTGACGATATT
CGGAAGCTATCAGAACTCCACTCGGGGAAGATCATATACTTGCTGCTGTAATCTGTAATGCGAATTTCACTTCGTATTTTACTCCGACTTT
CTGGTCCAACCCATCTCTATCTTACATTTGCGGATCGGAAAGCTTGTATTTTAAATACAGGGGTAATGGTAATCGATCTCGATCGATGGGAGA
ATGGTGATTATACCAGGAAGATTGAAGAATGGATGGAA
```

[\[BLAST\]](#)

## BLAST

# Consider what to BLAST against...



sol genomics network

[home](#) | [forum](#) | [contact](#) | [help](#)

[search](#)

[maps](#)

[genomes](#)

[tools](#)

[sol search](#)

[log in](#) | [new user](#)

## NCBI BLAST

Simple

Advanced

Sequence Set

SGN Tomato Combined - WGS, BAC, and unigene sequences

[db details](#)

Program

BLASTN (nucleotide to nucleotide)

### Query sequence

single sequence only, use Advanced for multiple

```
>SGN-E8107
CATCACCACCACCGCGGCCGCCCTGGCCACCAGTTCAGAGAAGCGCCTGAATTCTACAAC
TCAGCTGAATG
TCCTTCAATTGAAAATACCCATGAATTAATTTGCTCCGATGAACCTGTTTCATGCAGCAATGACTCTCG
ATTCCGCTTA
TATCCGTGGATCAATGGCGGCTATTCTCTCAATCTTCAACATTCTTCTTGCCACAAAACATAATTTT
TCACTTCGT
TGCTCCGCATCCGCAGACCGGCTCACATCTACGGCCACCATTTAGCTTCATTTCTTATTGAAATT
```

Expect (e-value) Threshold

1e-10

Clear

Search

Substitution Matrix

BLOSUM62 (default)

Show Graphics

all

Max. hits to show

100

# There are a LOT of options I chose Lycopersicon mRNA

The screenshot shows the 'sol genomic' search interface. On the left, there are several configuration options: 'Sequence Set' is set to 'SGN Tom', 'Program' is 'BLASTN', 'Expect (e-value) Threshold' is '1e-10', 'Substitution Matrix' is 'BLOSUM62 (default)', and 'Max. hits to show' is '100'. The main search area contains a list of database categories and their sub-entries. The 'Subgenus: Lycopersicon (all tomato species) SGN mRNA sequences' option is highlighted in blue. On the right, there are navigation links for 'home', 'forum', 'contact', and 'help', along with a 'sol search' button and 'log in'/'new user' links. Below the search options, there are two text boxes: the left one contains a sequence fragment starting with '>SGN-E8107' and the right one contains a sequence fragment starting with 'TTCTACAAC'. At the bottom right, there are 'Clear' and 'Search' buttons.

**sol genomic** search

home | forum | contact | help

sol search

log in | new user

**Sequence Set** SGN Tom

**Program** BLASTN

**Expect (e-value) Threshold** 1e-10

**Substitution Matrix** BLOSUM62 (default)

**Max. hits to show** 100

Potato BAC Sequences  
 S. penellii Genome  
 S. penellii BAC Ends - LpenBAC  
 S. penellii Cosmid Ends - LpenCOS

Tobacco Genome  
 Assembly of tobacco methylation filtered genome sequences (from TGI)  
 Tobacco genome sequences (methylation filtered from TGI)

Organelle Genomes  
 Arabidopsis thaliana Mitochondria  
 Nicotiana tabacum Chloroplast  
 Tomato chloroplast genome, cultivar IPA-6

SGN ESTs  
 All SGN mRNA sequences  
 Genus: Nicotiana SGN mRNA sequences  
 Species: Antirrhinum majus SGN mRNA sequences  
 Species: Capsicum annuum (pepper) SGN mRNA sequences  
 Species: Coffea arabica SGN mRNA sequences  
 Species: Coffea canephora SGN mRNA sequences  
 Species: Hedyotis terminalis SGN mRNA sequences  
 Species: Ipomoea batatas (sweet potato) SGN mRNA sequences  
 Species: Kadua centranthoides SGN mRNA sequences  
 Species: Nicotiana benthamiana SGN mRNA sequences  
 Species: Nicotiana sylvestris (wood tobacco) SGN mRNA sequences  
 Species: Nicotiana tabacum (tobacco) SGN mRNA sequences  
 Species: Petunia hybrida SGN mRNA sequences  
 Species: Solanum cheesmaniae SGN mRNA sequences  
 Species: Solanum habrochaites SGN mRNA sequences  
 Species: Solanum lycopersicoides SGN mRNA sequences  
 Species: Solanum lycopersicum SGN mRNA sequences  
 Species: Solanum melongena (eggplant) SGN mRNA sequences  
 Species: Solanum pennellii SGN mRNA sequences  
 Species: Solanum peruvianum SGN mRNA sequences  
 Species: Solanum pimpinellifolium SGN mRNA sequences  
 Species: Solanum tuberosum (potato) SGN mRNA sequences  
**Subgenus: Lycopersicon (all tomato species) SGN mRNA sequences**

SGN Unigenes  
 All SGN Unigene sequences  
 Antirrhinum majus Unigenes  
 Capsicum annuum (pepper) Unigenes  
 Coffea arabica (arabica coffee) Unigenes  
 Coffea canephora (robusta coffee) Unigenes  
 Coffee species Unigenes  
 Hedyotis species Unigenes  
 Ipomoea batatas (sweet potato) Unigenes  
 Lycopersicon Combined (Tomato) Unigenes  
 Nicotiana benthamiana Unigenes

>SGN-E8107  
 CATCACCACCACCGCG  
 TCAGCTGAATG  
 TCCTTCAATTGAAAATA  
 ATTCCGCTTA  
 TATCCGTGGATCAATGC  
 TCACTTCGT  
 TGCTTCCGCATCCGCAC

TTCTACAAC  
 TGACACTCG  
 GCATAATTTT  
 ATTTGAAATT

db details

Clear Search

# BLAST Results

BLAST Search Report - Sol Genomics Network

http://solgenomics.net/tools/blast/view\_result.pl?output\_graphs=bioperl\_histogram&filterq: Google

log in | new user

### BLAST Results

Note: Please **do not bookmark** this page. BLAST results are automatically deleted after 7 days.

#### Graphics

##### Alignment Summary

All hits shown.

Mouse-overs require JavaScript

Fwd: >> Rev: << Neg P Exponent: -> 0-10 10-50 50-200 200-> 4/Nov/2010

#### Conservedness Histogram

##### Conservedness Histogram

The histogram shows a count of hits for each base in the query sequence, but counts only the domains BLAST finds, meaning this is really more a function of region than of individual base. Within the graph, green shows exact base matches within conserved regions; blue shows non-matching bases within conserved regions. Gaps introduced into the query by BLAST are ignored; gaps introduced into target sequences are not.

Query region [388 -> 676]: conservedness in (3, 9)

Query 3 351 703 Conservedness

#### BLAST Report

[View / download raw report] (9.6K)

CONSERVEDNESS HISTOGRAM

Our EST is highly conserved

# BLAST Results

```
http://solgenomics.net/static/documents/tempfiles/blast/Xejwrt
http://solgenomics.net/static/documents/tempfiles/blast/Xejwrt&flv
Apple Yahoo! Google Maps YouTube Wikipedia News (376) Popular v

Sequences producing significant alignments:
Score      E
(bits)     Value

SGN-E211387 #Solanum lycopersicum (formerly Lycopersicon esculen... 1318 0.0
SGN-E1255387 #Solanum lycopersicum (formerly Lycopersicon esculen... 821 0.0
SGN-E205127 #Solanum lycopersicum (formerly Lycopersicon esculen... 593 e-168

>SGN-E211387 #Solanum lycopersicum (formerly Lycopersicon esculentum)
   [cLEC-67-I15](5')
   Length = 730

Score = 1318 bits (665), Expect = 0.0
Identities = 674/677 (99%)
Strand = Plus / Plus

Query: 1   catcaccaccaccgcccggccgcccgcctgcccaccgcaacagttcagagaagcgccgaatt 60
          |||
Sbjct: 54  catcaccaccaccgcccggccgcccgcctgcccaccgcaacagttcagagaagcgccgaatt 113


Query: 61  ctacaactcagctgaatgtccttcaattgaaaatacccatgaattaatttgcctccatga 120
          |||
Sbjct: 114 ctacaactcagctgaatgtccttcaattgaaaatacccatgaattaatttgcctccatga 173

Query: 121  agctgttcacgcagccatgacactcagattccgcttatatccgtggatcaatggcgctat 180
          |||
Sbjct: 174  agctgttcacgcagccatgacactcagattccgcttatatccgtggatcaatggcgctat 233


Query: 181  tctctcaattcttcaacattcttcttgcaccacaaaacataaattttcacttctgttctt 240
          |||
Sbjct: 234  tctctcaattcttcaacattcttcttgcaccacaaaacataaattttcacttctgttctt 293

Query: 241  cgcacccgcagacgcctcaccatctacggccaccatcttagcttcatttccttattttaa 300
          |||
Sbjct: 294  cgcacccgcagacgcctcaccatctacggccaccatcttagcttcatttccttattttaa 353

Query: 301  attcgaagtttatcgatttcgacgattcctctgtttccggtttaatttcaacttctattcg 360
          |||
Sbjct: 354  attcgaagtttatcgatttcgacgattcctctgtttccggtttaatttcaacttctattcg 413
```

 Look for mismatches between query and subject sequences

 No mismatches here

 Doesn't provide evidence for marker polymorphism

# Conclusion – What We Learned

- 🍅 How to use the tomato genome browser to identify molecular markers that may be useful in fine-mapping and marker-assisted selection
- 🍅 Important to remember that the browser is a work in progress – e.g. don't have all marker data currently

sol genomics network home | forum | contact | help

search maps genomes tools sol search

File Help

ITAG1 Release: genomic annotations: 969 bp from SL1.00sc06004:457,840..458,808

Browser Select Tracks Upload and Share Tracks Preferences

Search Landmark or Region: SL1.00sc06004:457840.. Search Annotate Restriction Sites Configure... Go

Examples: SL1.00sc00002\_16.1.1, Sulfite oxidase, TG154, C2\_A13g02060.

Data Source ITAG1 Release: genomic annotations Scroll/Zoom: Show 969 bp Flip

Overview SL1.00sc06004

Region

Genes

Details

Gene models

CDS - click to browse protein

ESTs and cDNAs - Tomato



Photo courtesy of D. Francis, Ohio State University