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?

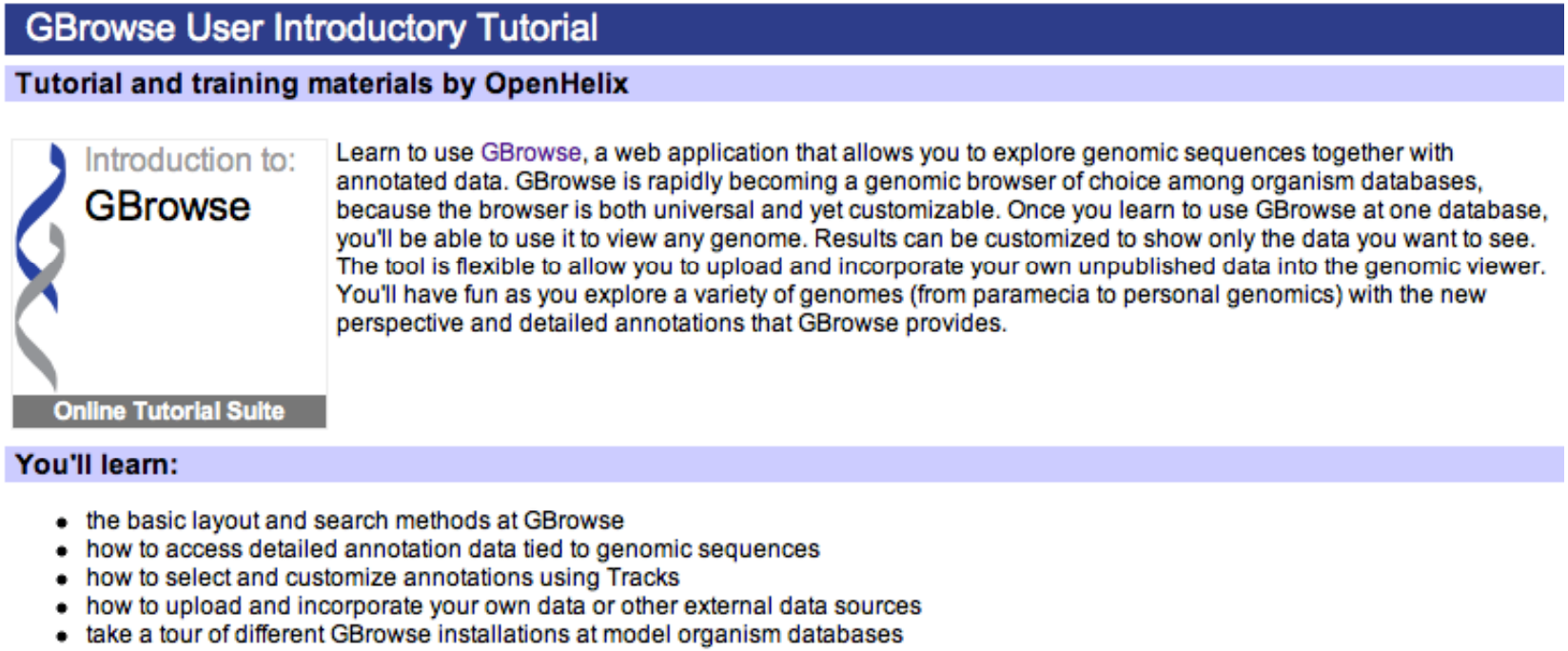
- Wikipedia Definition: “A graphical interface for display of information from a biological database for genomic data. Genome browsers enable researchers to visualize and browse entire genomes with annotated data including gene prediction and structure, proteins, expression, regulation, variation, comparative analysis, etc.” – en.wikipedia.org/wiki/Genome_browser (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

Introduction to:
GBrowse
Online Tutorial Suite

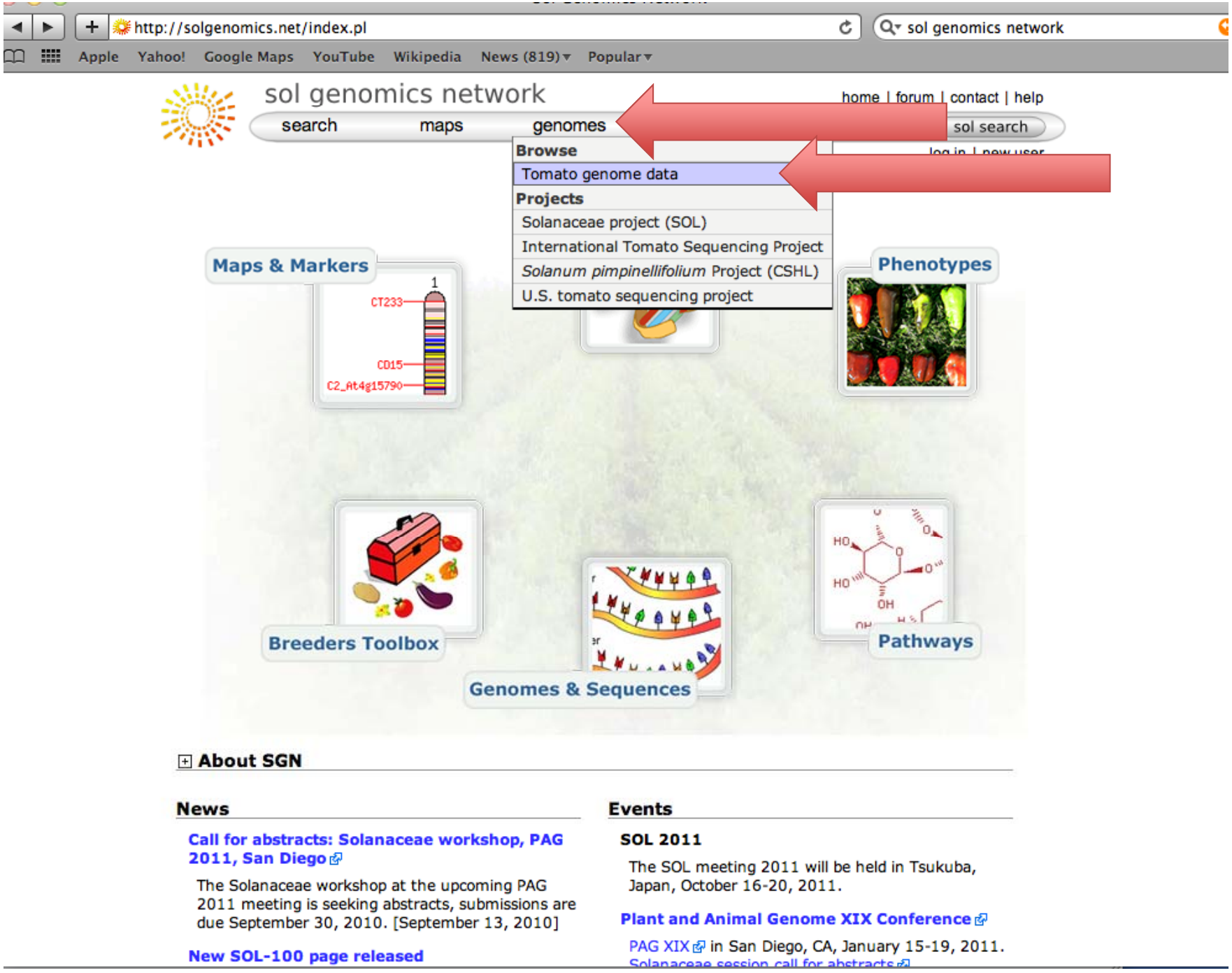
Learn to use GBrowse, a web application that allows you to explore genomic sequences together with annotated data. GBrowse is rapidly becoming a genomic browser of choice among organism databases, because the browser is both universal and yet customizable. Once you learn to use GBrowse at one database, 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.

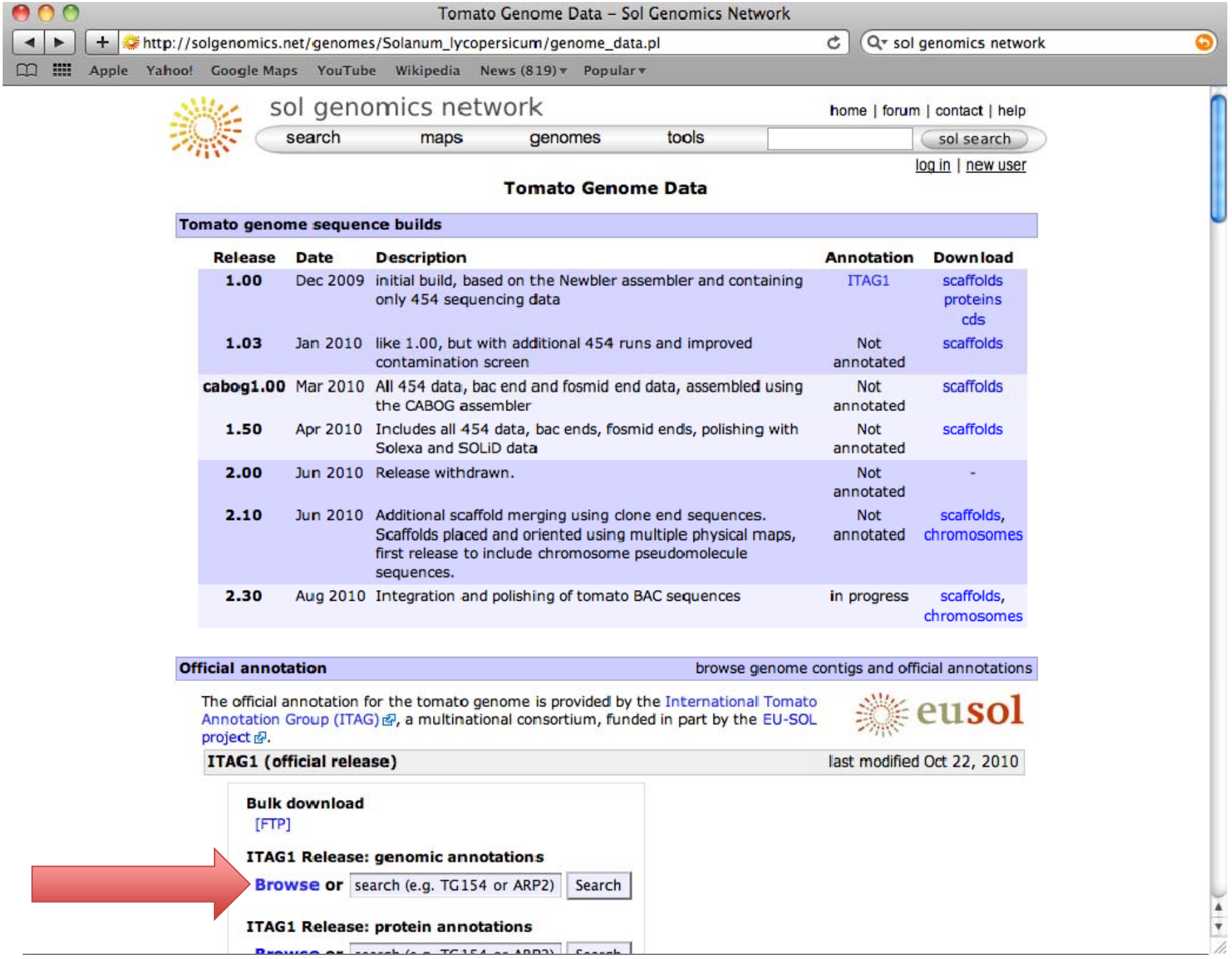
You'll learn:

- 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

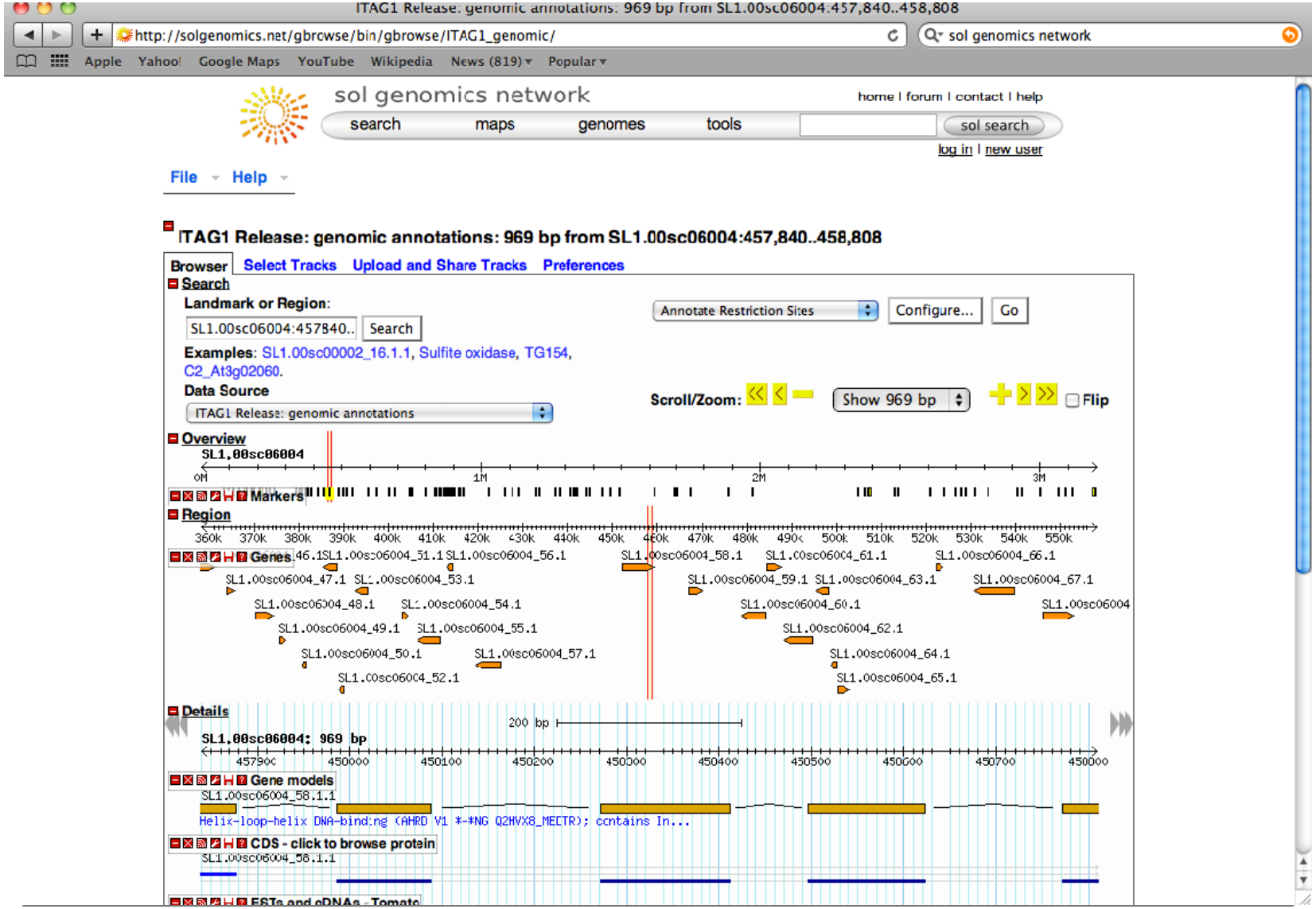
Where can I find the tomato genome browser?

http://solgenomics.net

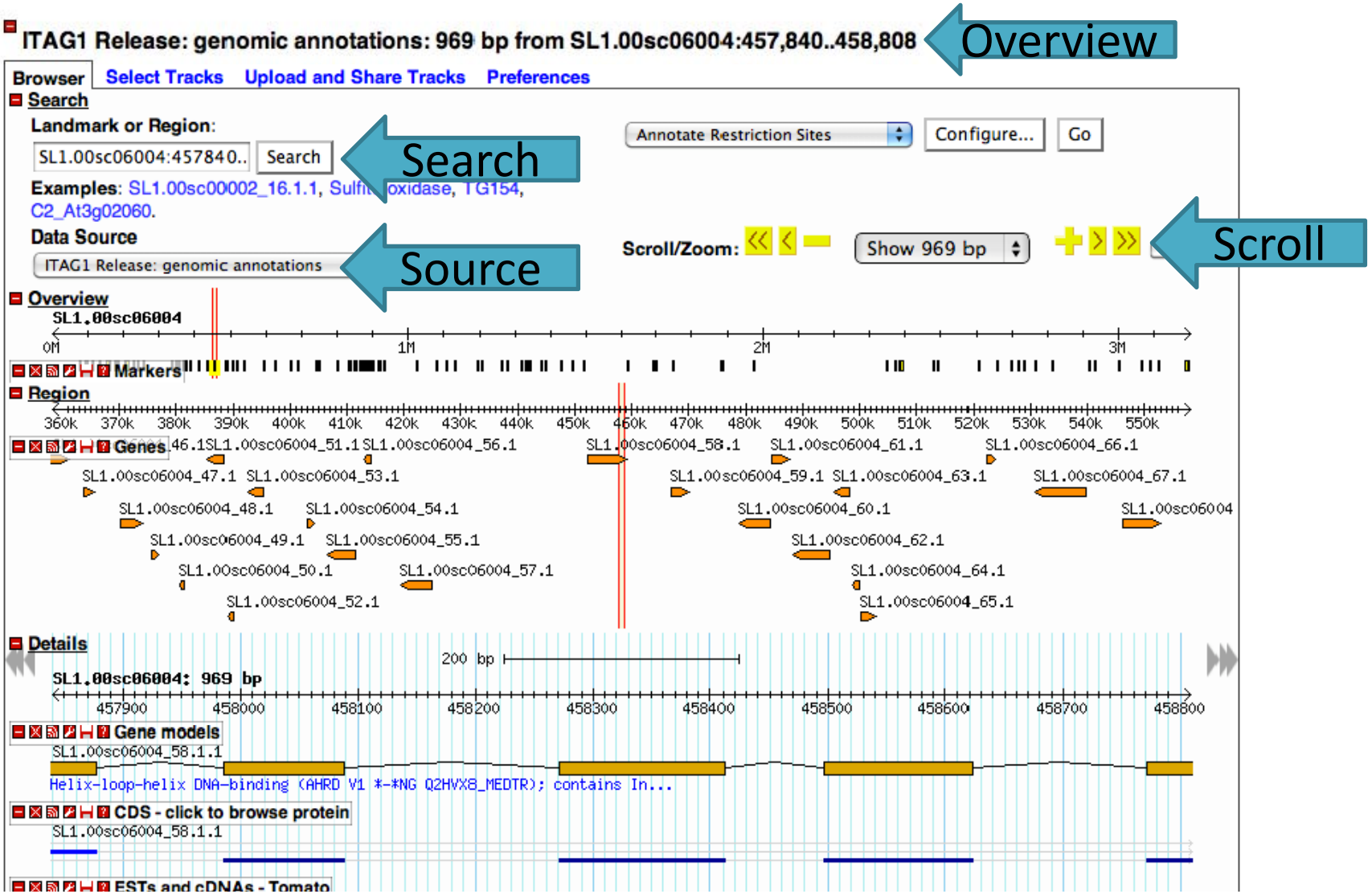

http://solgenomics.net/index.pl
sol genomics network
Apple
Yahoo!
Google Maps
YouTube
Wikipedia
News (819)
Popular
sol genomics network
home | forum | contact | help
search
maps
genomes
sol search
log in | new user
Browse
Tomato genome data
Projects
Solanaceae project (SOL)
International Tomato Sequencing Project
Solanum pimpinellifolium Project (CSHL)
U.S. tomato sequencing project
Maps & Markers
1
CT233
CD15
C2_At4g15790
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 – Sol Genomics Network
http://solgenomics.net/genomes/Solanum_lycopersicum/genome_data.pl
sol genomics network
Apple Yahoo! Google Maps YouTube Wikipedia News (819) Popular
sol genomics network
home | forum | contact | help
search maps genomes tools
sol search
log in | new user
Tomato Genome Data
Tomato genome sequence builds
Release Date Description Annotation Download
1.00 Dec 2009 initial build, based on the Newbler assembler and containing only 454 sequencing data ITAG1 scaffolds proteins cds
1.03 Jan 2010 like 1.00, but with additional 454 runs and improved contamination screen Not annotated scaffolds
cabog1.00 Mar 2010 All 454 data, bac end and fosmid end data, assembled using the CABOG assembler Not annotated scaffolds
1.50 Apr 2010 Includes all 454 data, bac ends, fosmid ends, polishing with Solexa and SOLiD data Not annotated scaffolds
2.00 Jun 2010 Release withdrawn. Not annotated -
2.10 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
2.30 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.
eusol
ITAG1 (official release)
last modified Oct 22, 2010
Bulk download
[FTP]
ITAG1 Release: genomic annotations
Browse or search (e.g. TG154 or ARP2) Search
ITAG1 Release: protein annotations

At Last… The Tomato Genome Browser

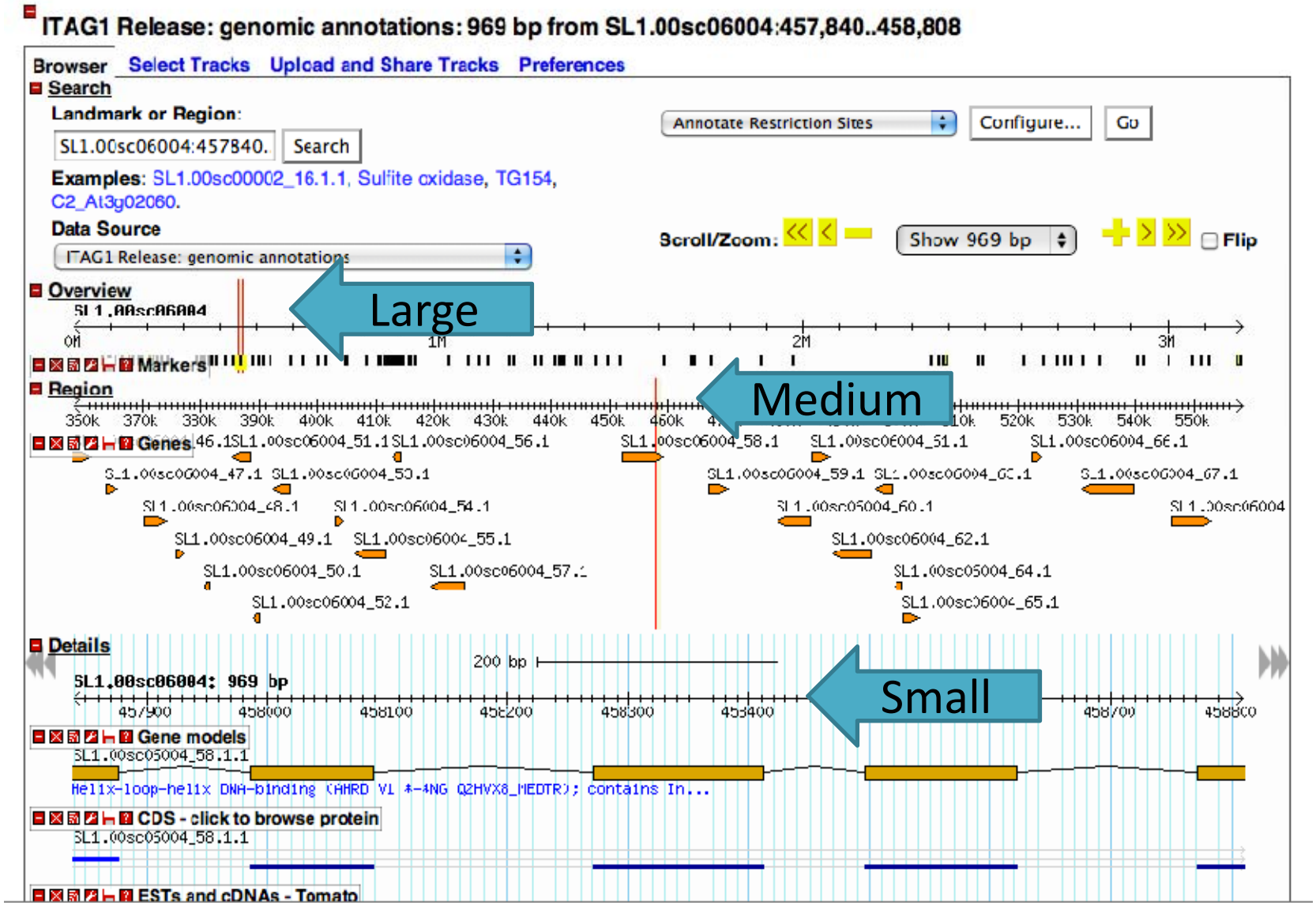


There's a lot of information! What am I looking at?

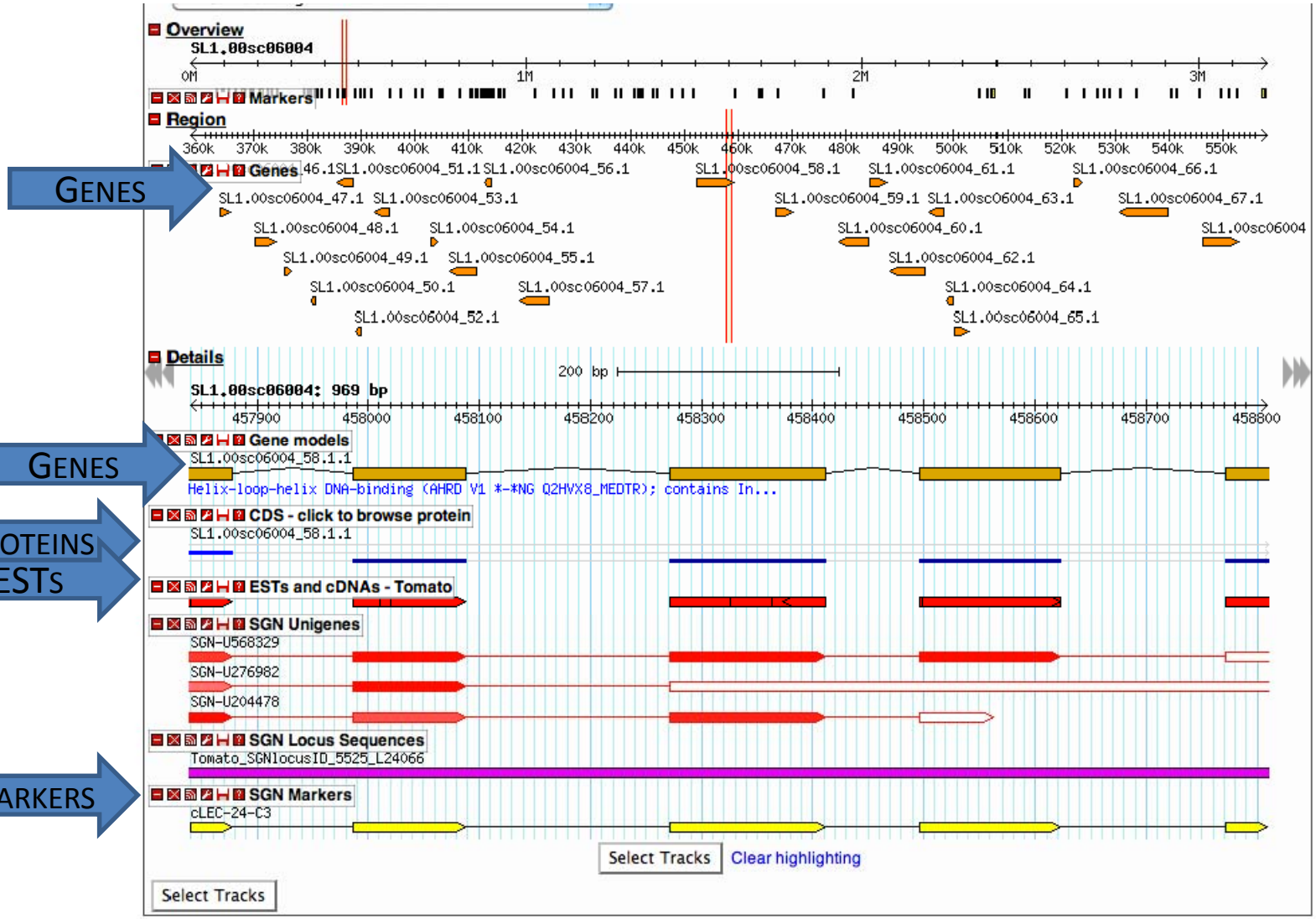


I’m scared! Get me outta here!

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:
SL1.00sc06004:457840.
Search
Examples: SL1.00sc00002_16.1.1, Sulfite oxidase, TG154, C2_At3g02060.
Data Source
ITAG1 Release: genomic annotations
Annotate Restriction Sites
Configure...
Go
Scroll/Zoom:
Show 969 bp
Flip
Overview
Markers
Region
Genes
Details
200 bp
SL1.00sc06004: 969 bp
Gene models
SL1.00sc05004_58.1.1
Helix-loop-helix DNA-binding (AHRD V1 *-4NG Q2HVX8_MEDTR); contains In...
CDS - click to browse protein
SL1.00sc05004_58.1.1
ESTs and cDNAs - Tomato
Large
Medium
Small

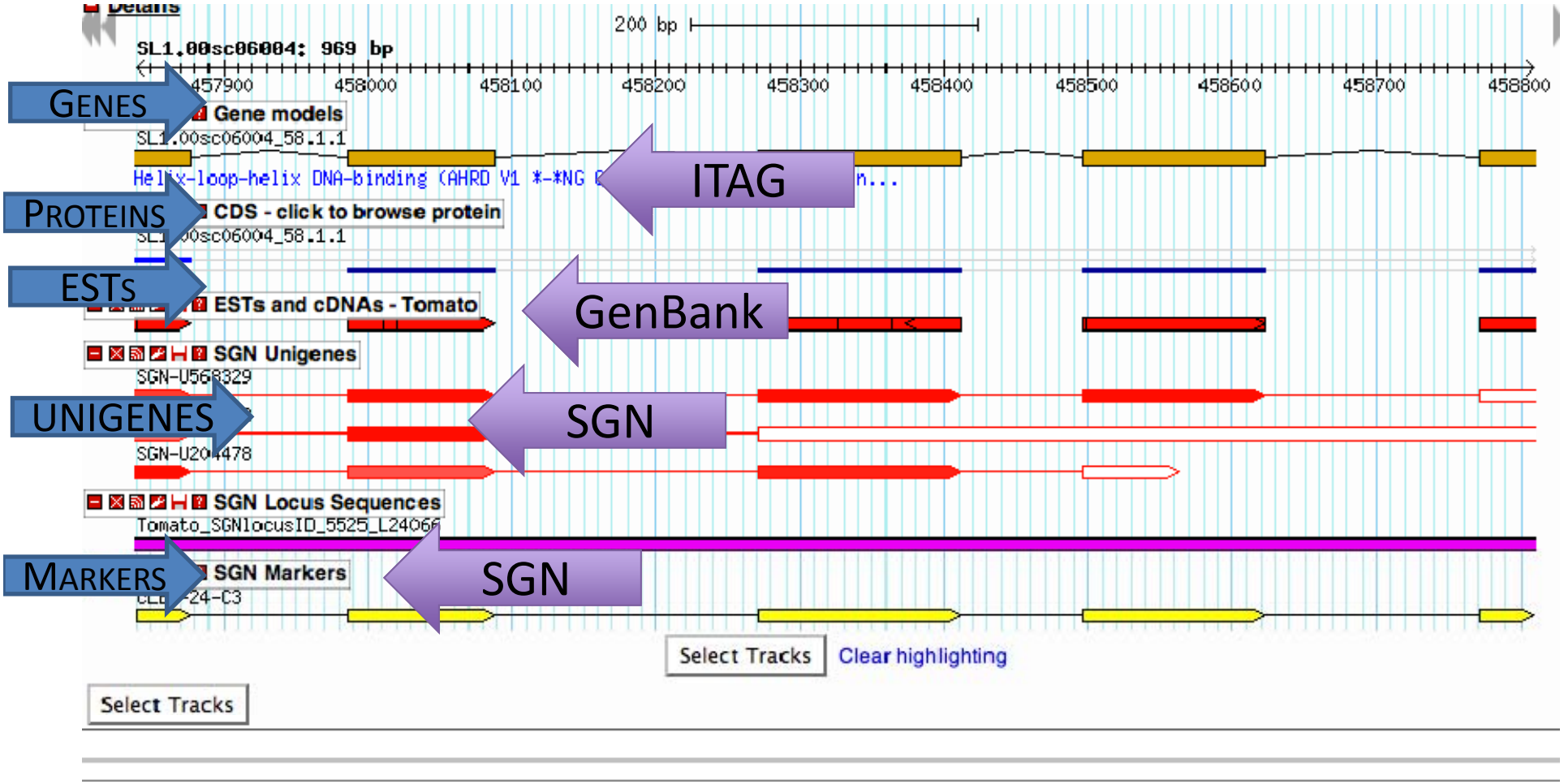
Tracks – Types of Data to View



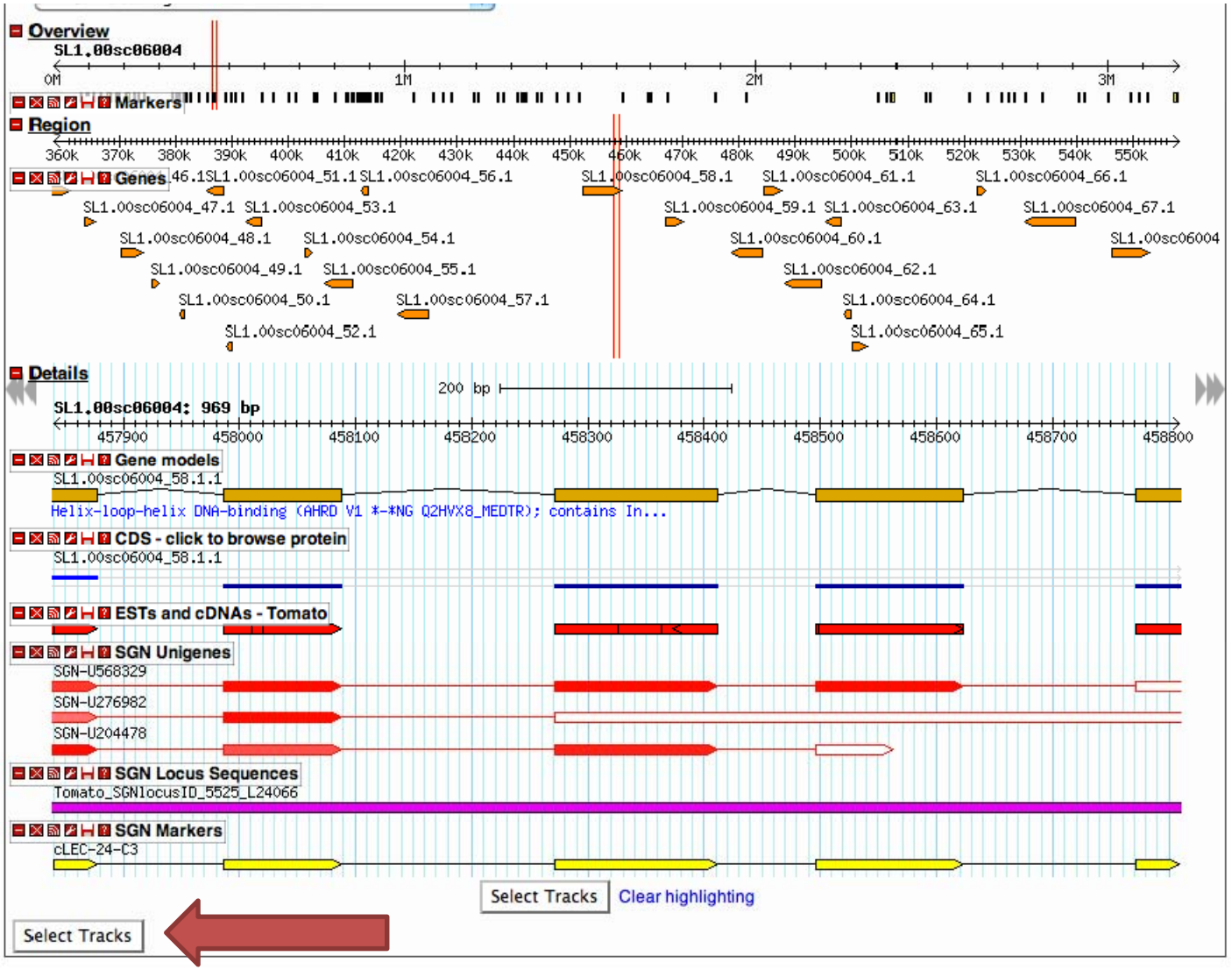
Wait! – Where Does the Data Come From?

- 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

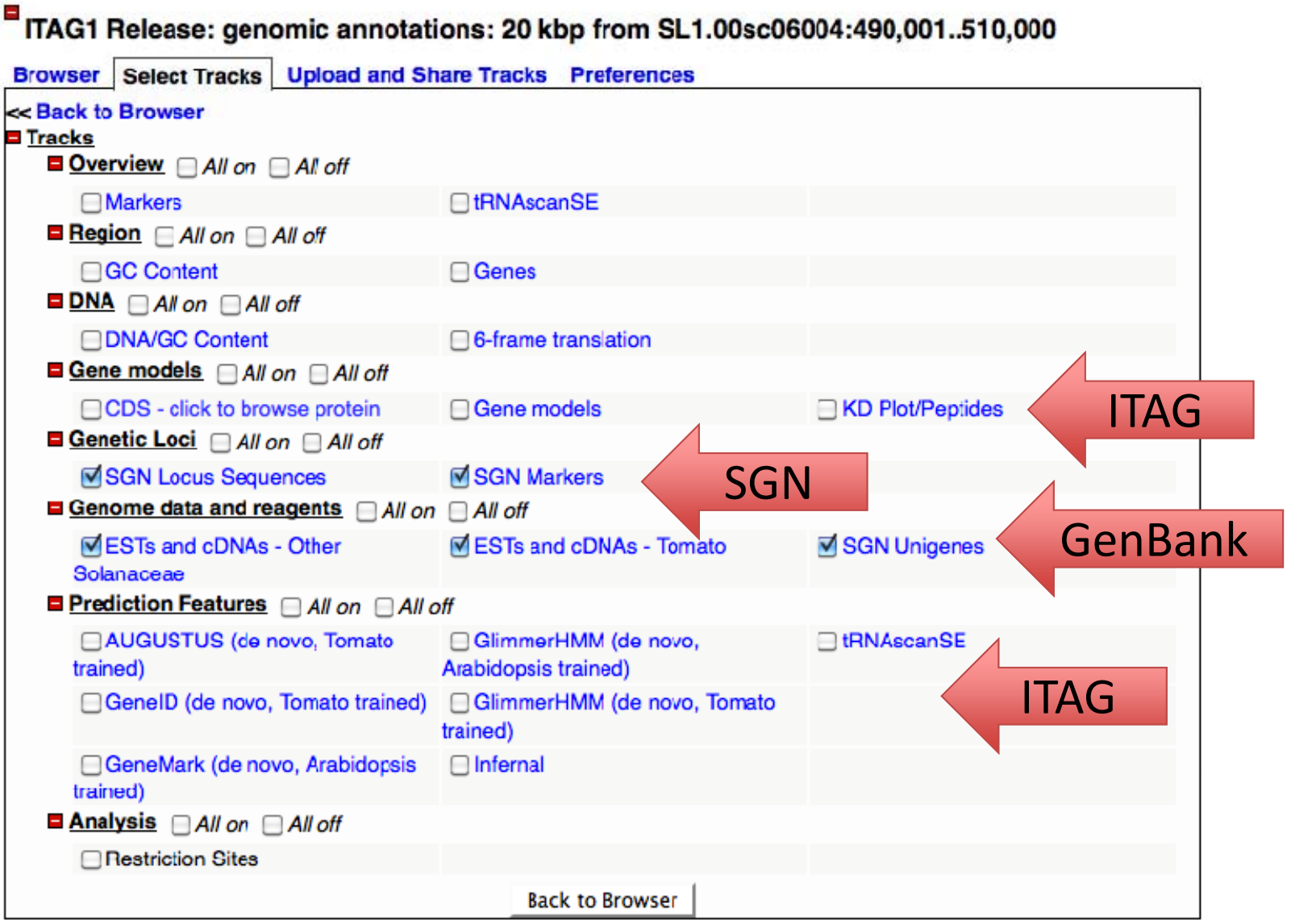
Tracks – Data Sources



Select Tracks



Select Tracks



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

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

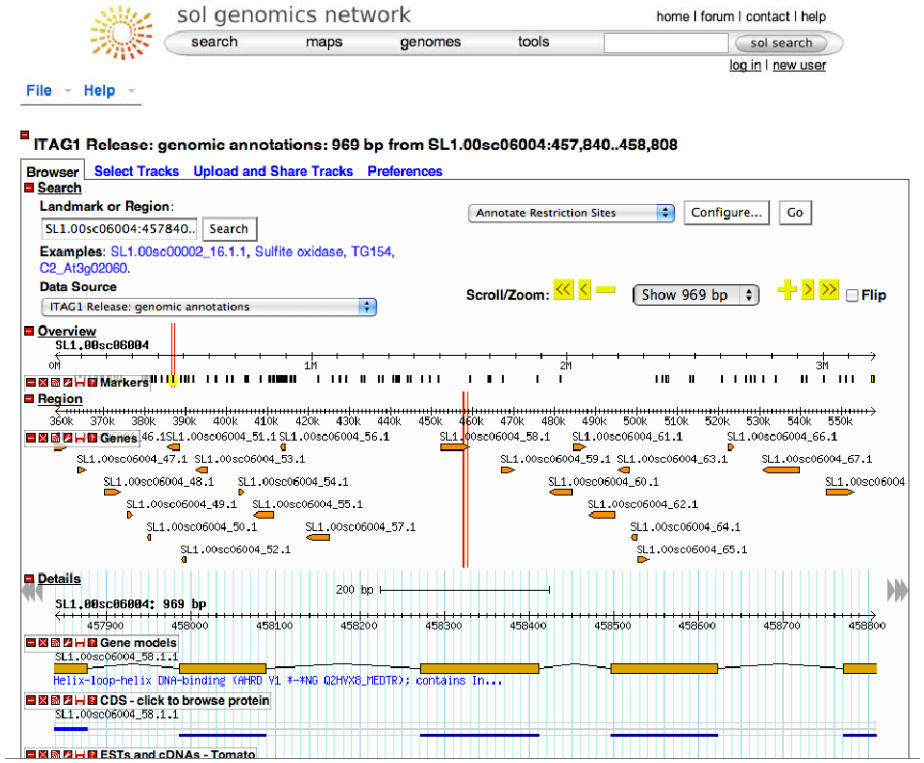


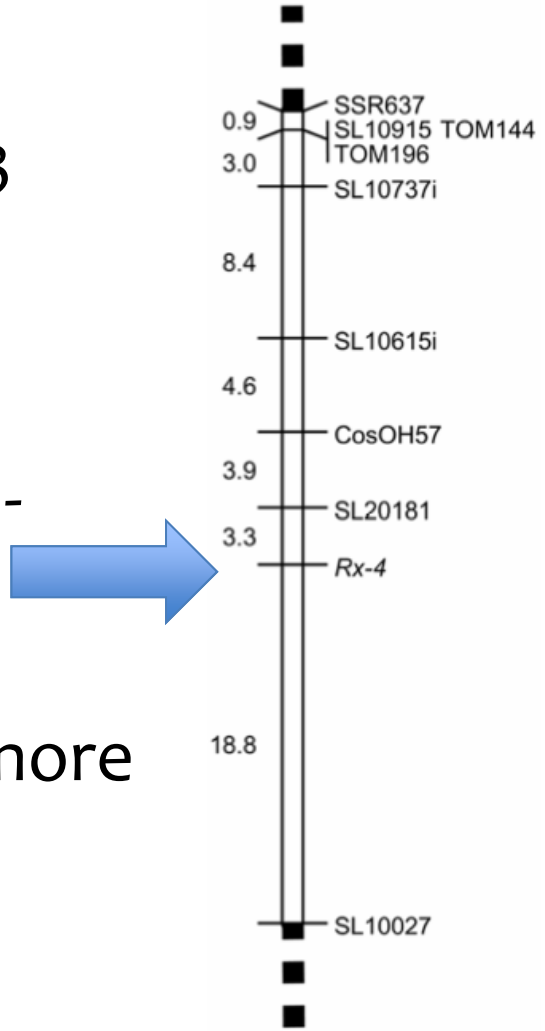
Photo courtesy of D. Francis, Ohio State University

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

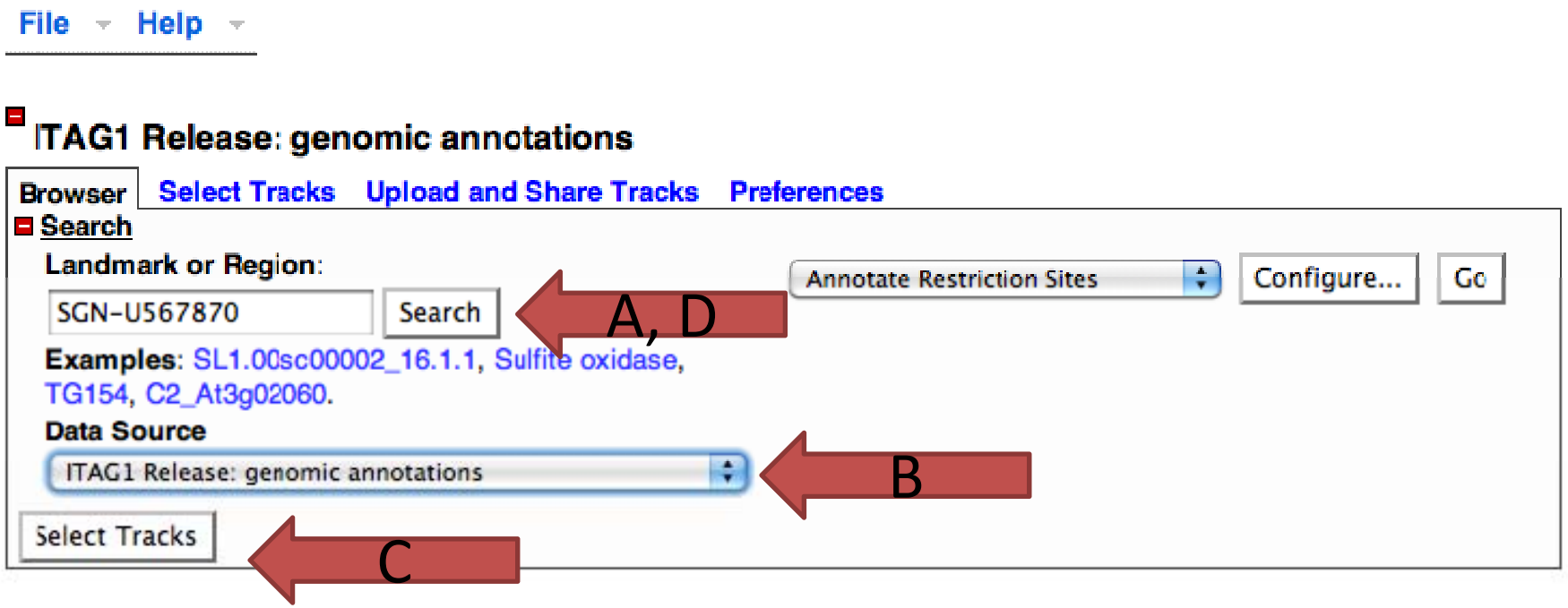
Tomato Bacterial Spot Resistance

- 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

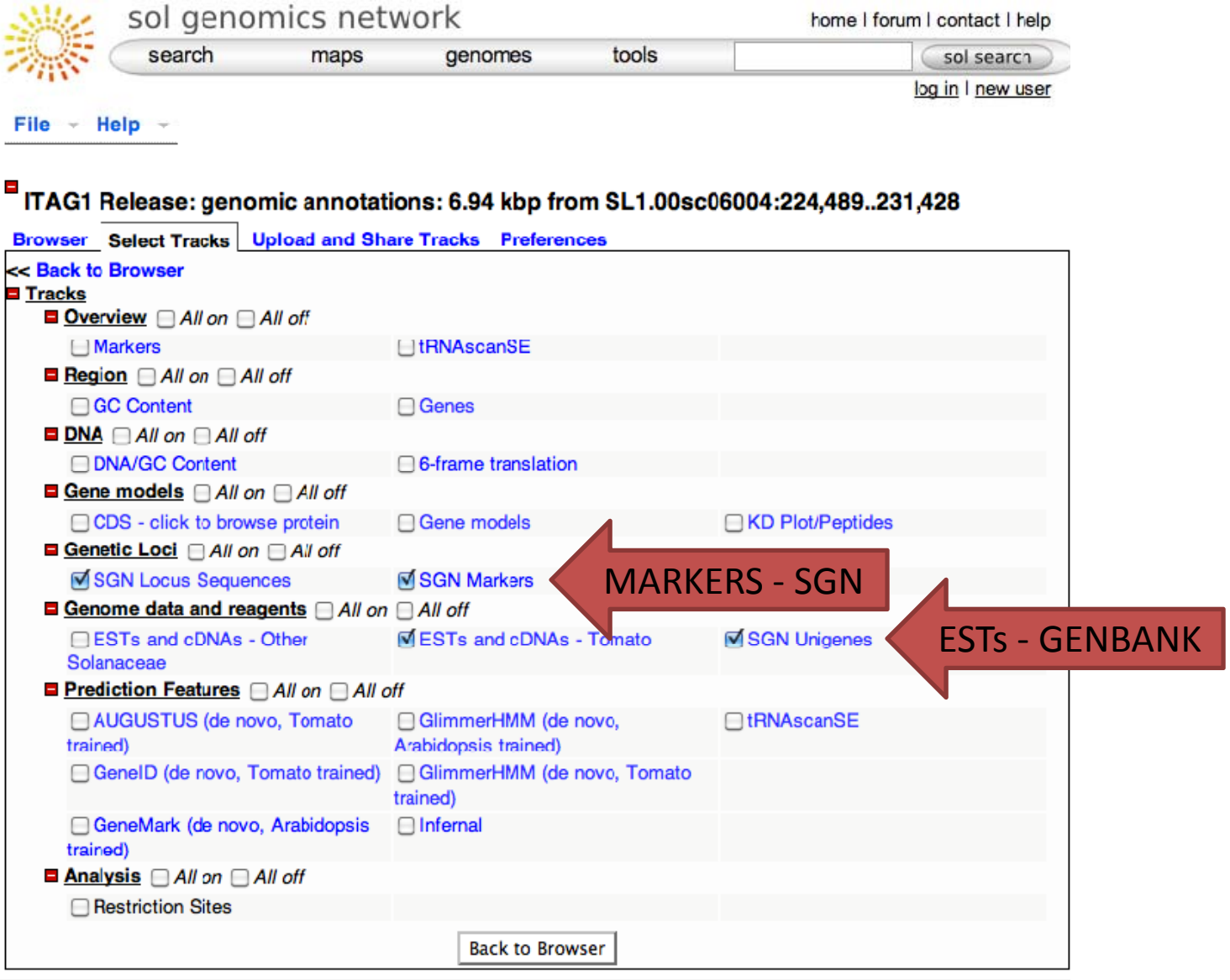


Step 1: Search

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

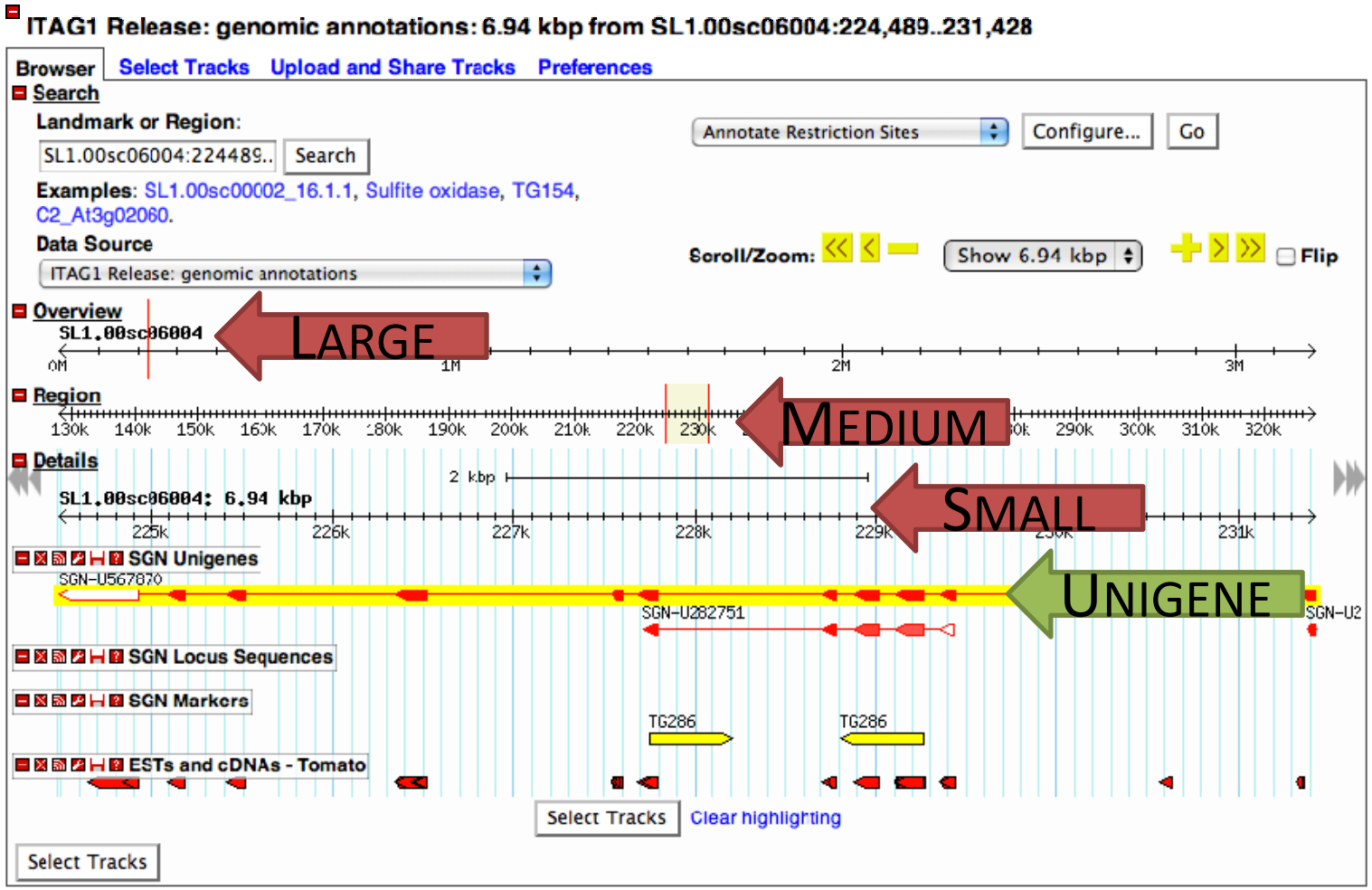


Step 1c: Select Tracks

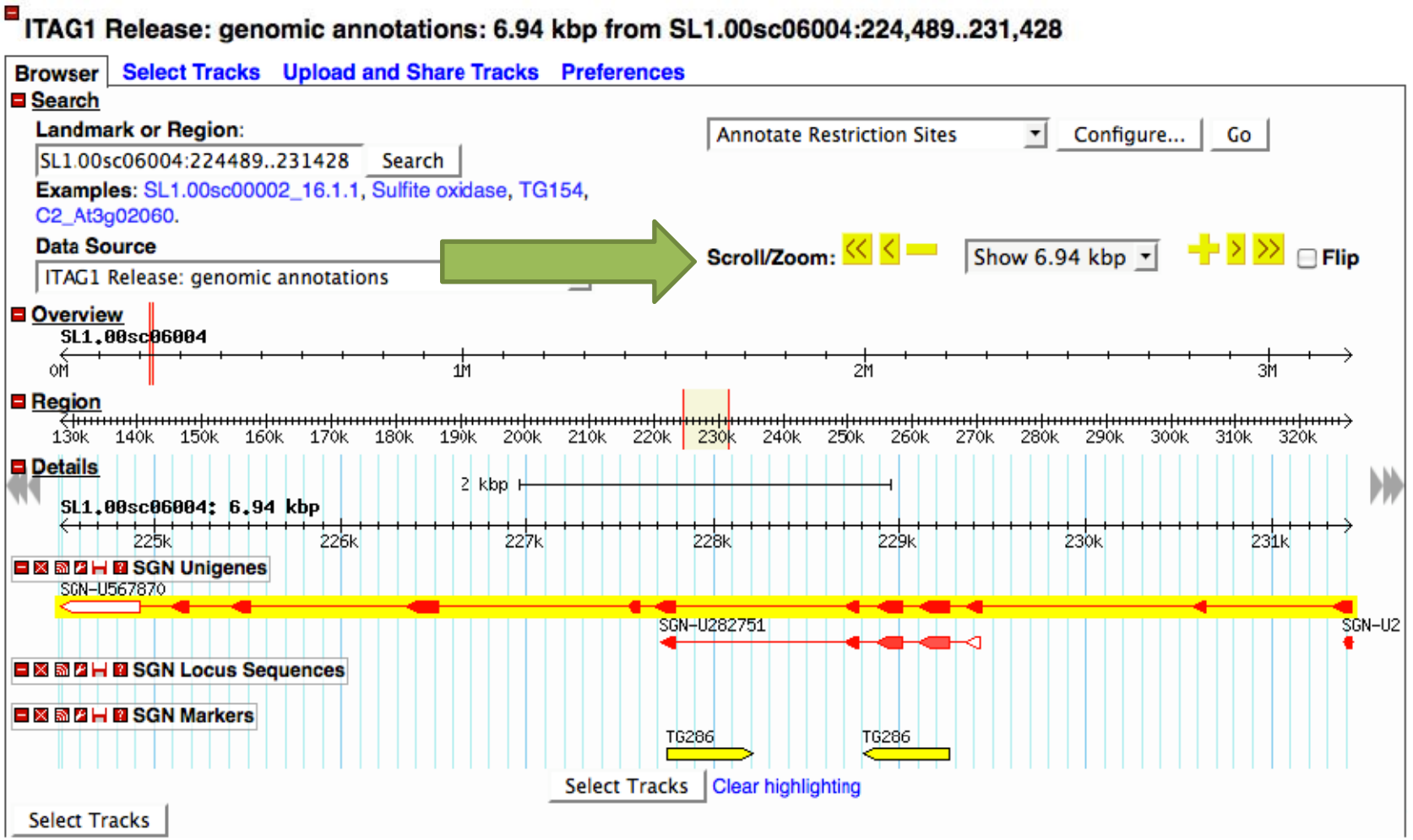


Step 2: Orientation

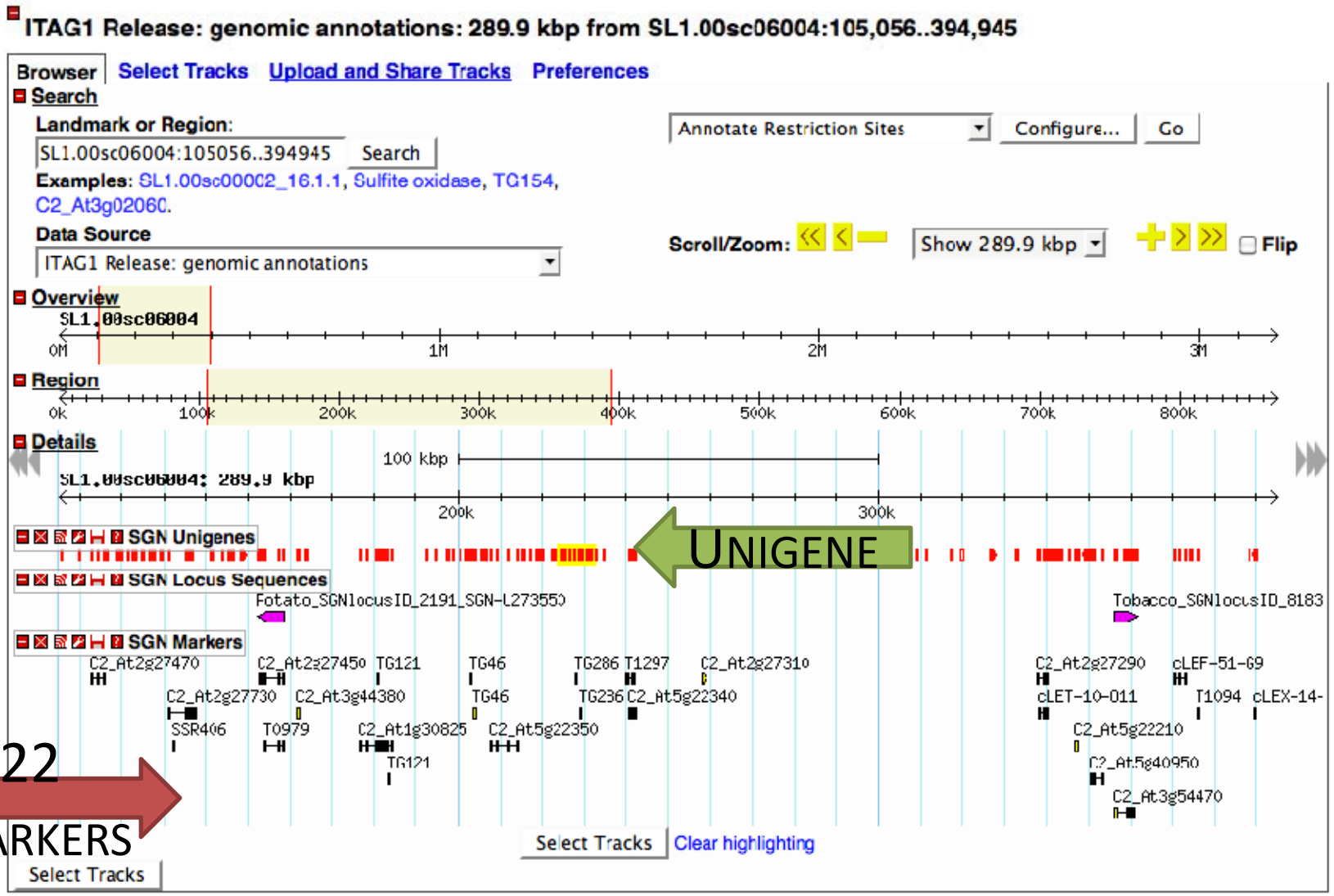
- Look at scales for orientation



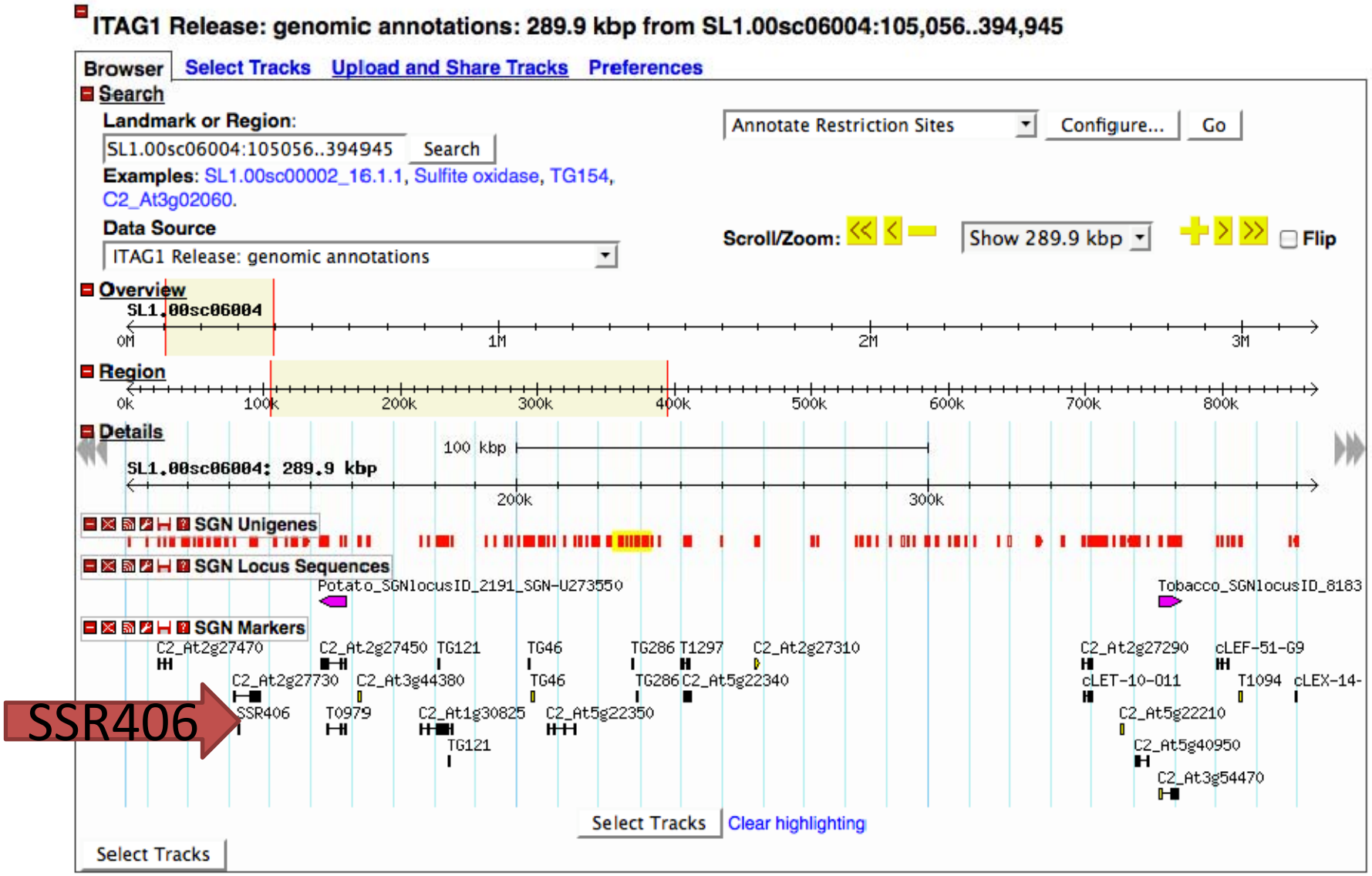
Step 3: Zoom Out



Step 4: Locate Markers

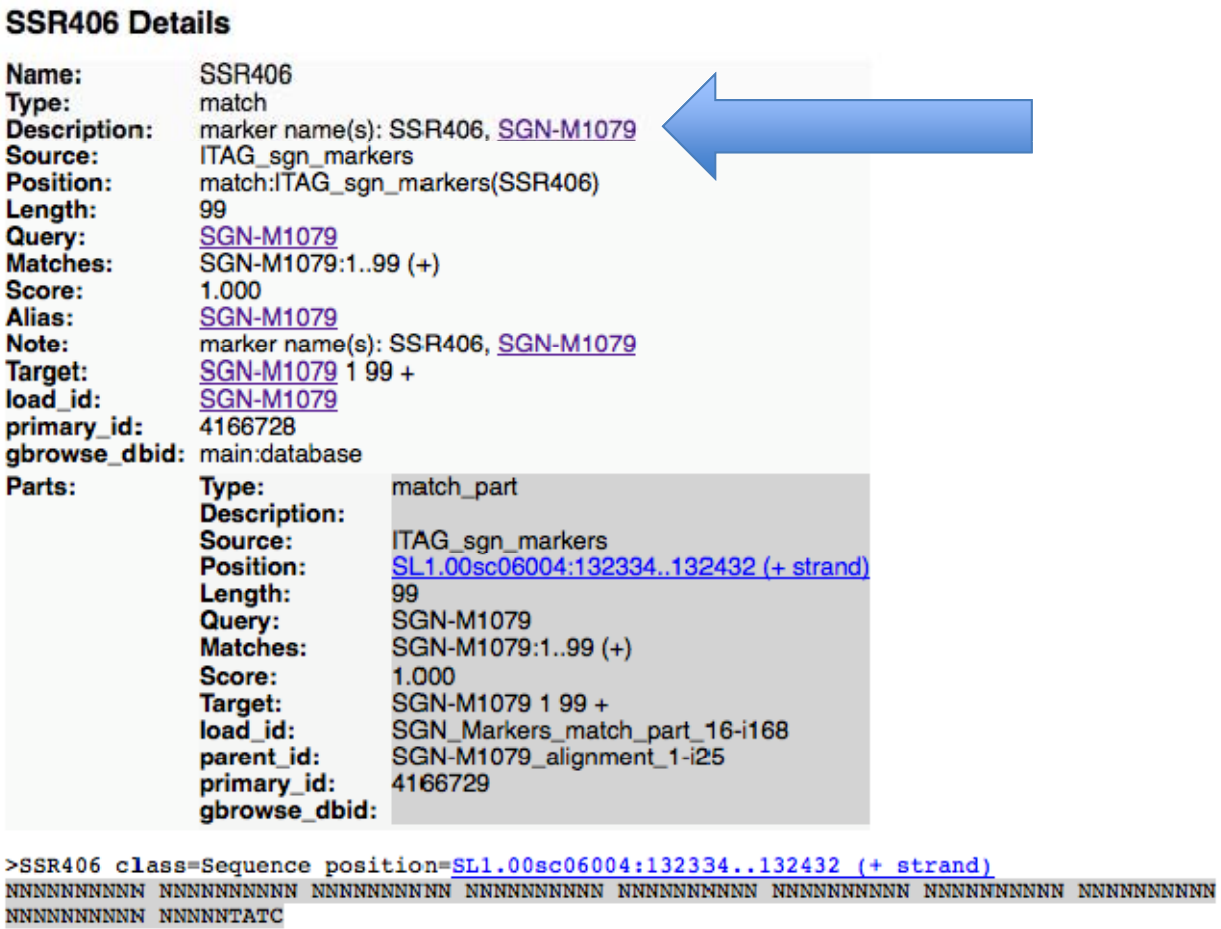


Step 5: Click on Marker



Step 6: Look at Marker Detail

- Click SGN-M1079 to view info on SGN



SSR406 Details

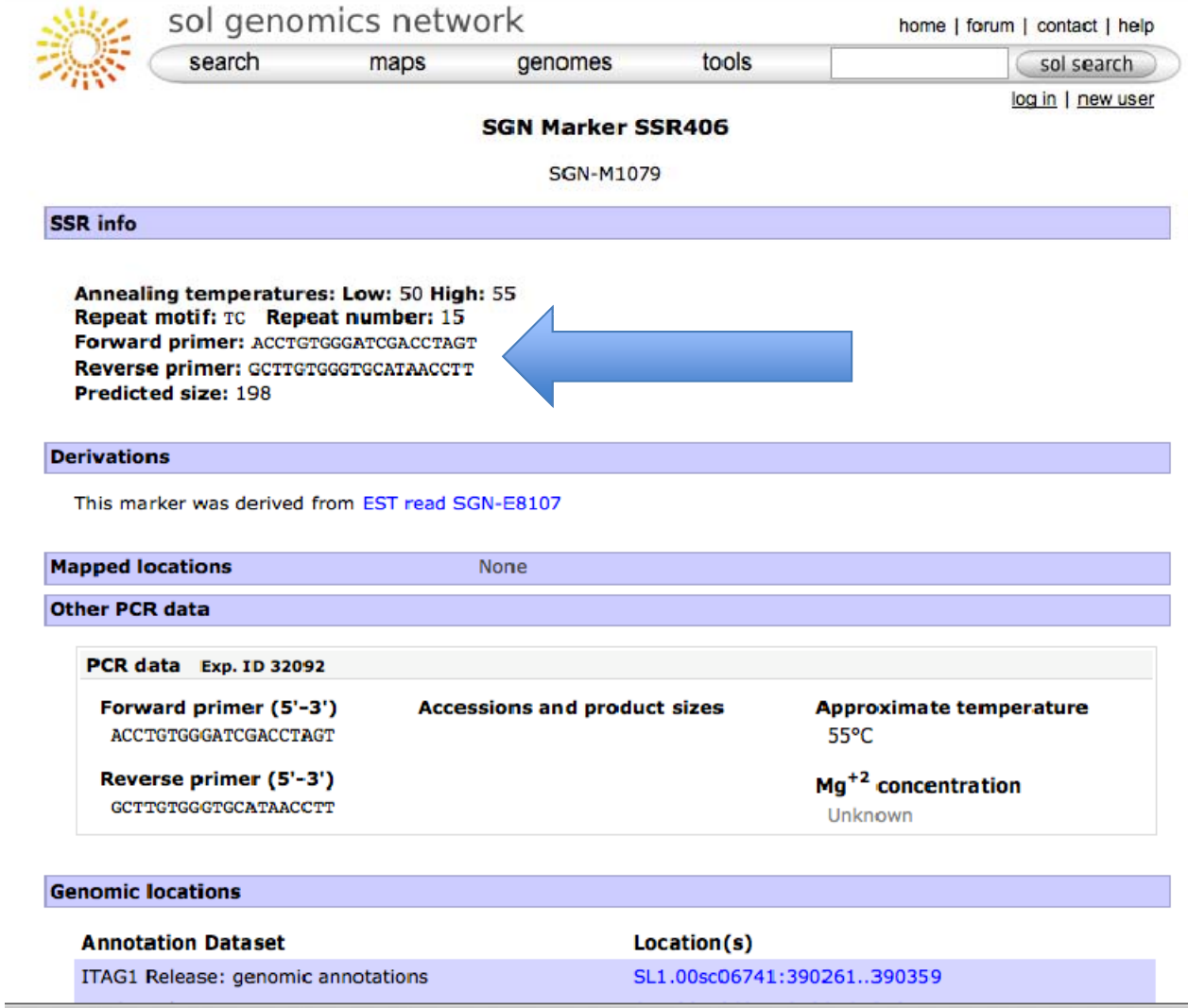
Name:	SSR406
Type:	match
Description:	marker name(s): SSR406, SGN-M1079
Source:	ITAG_sgn_markers
Position:	match:ITAG_sgn_markers(SSR406)
Length:	99
Query:	SGN-M1079
Matches:	SGN-M1079:1..99 (+)
Score:	1.000
Alias:	SGN-M1079
Note:	marker name(s): SSR406, SGN-M1079
Target:	SGN-M1079 1 99 +
load_id:	SGN-M1079
primary_id:	4166728
gbrowse_dbid:	main:database

Parts:

Type:	match_part
Description:	
Source:	ITAG_sgn_markers
Position:	SL1.00sc06004:132334..132432 (+ strand)
Length:	99
Query:	SGN-M1079
Matches:	SGN-M1079:1..99 (+)
Score:	1.000
Target:	SGN-M1079 1 99 +
load_id:	SGN_Markers_match_part_16-i168
parent_id:	SGN-M1079_alignment_1-i25
primary_id:	4166729
gbrowse_dbid:	

>SSR406 class=Sequence position=SL1.00sc06004:132334..132432 (+ strand)
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
NNNNNNNNNN NNNNNTATC

Step 7: Locate primers

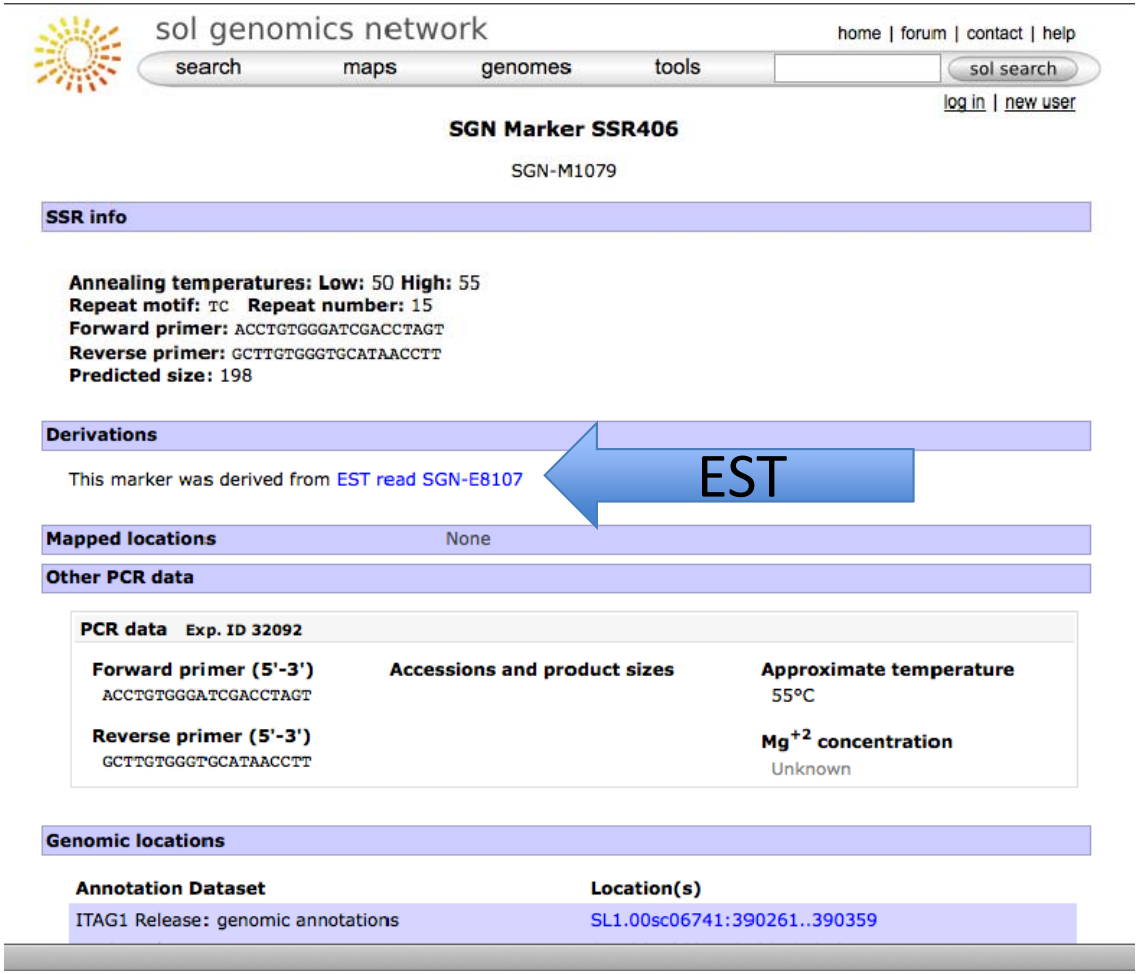


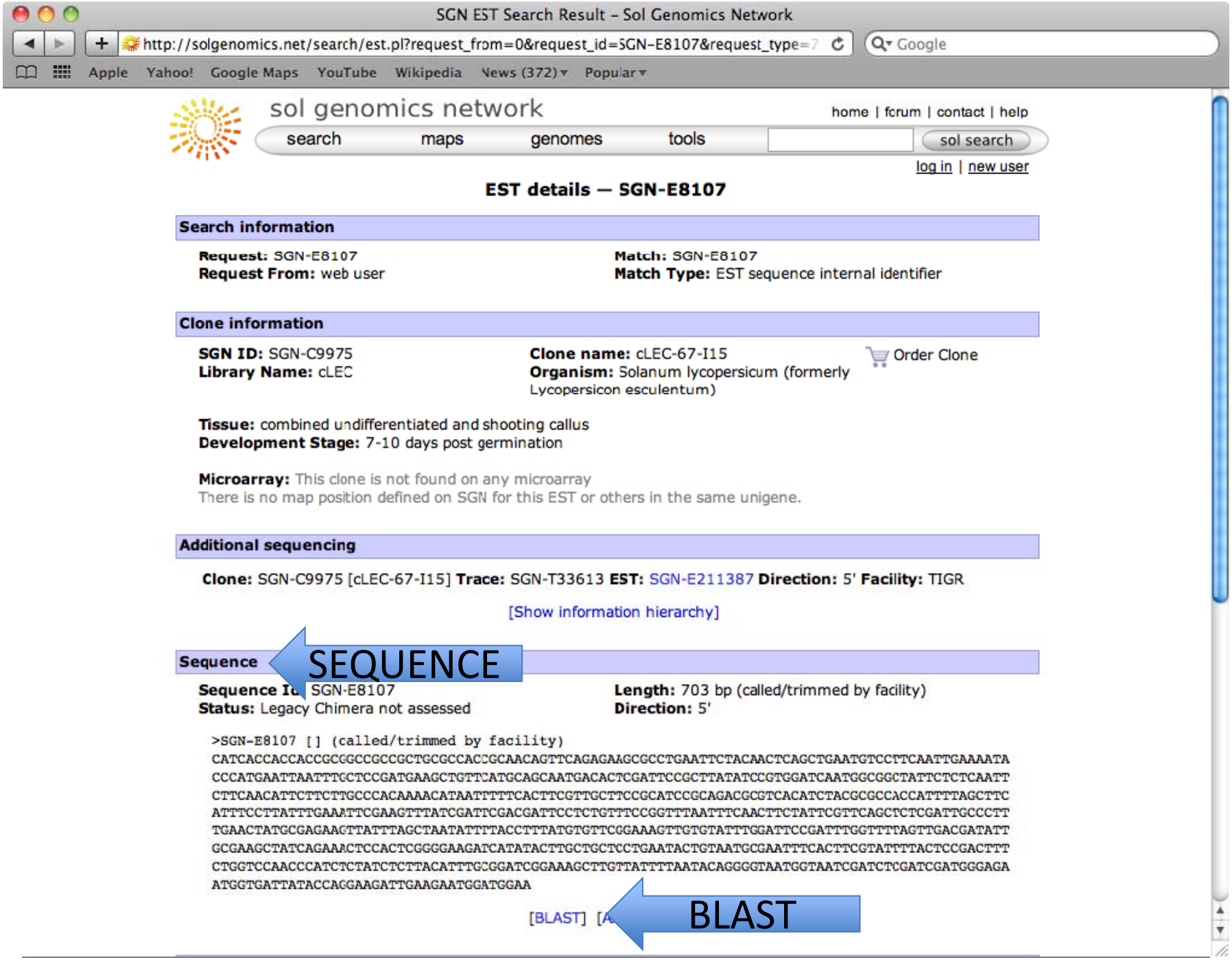
BUT….

- 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

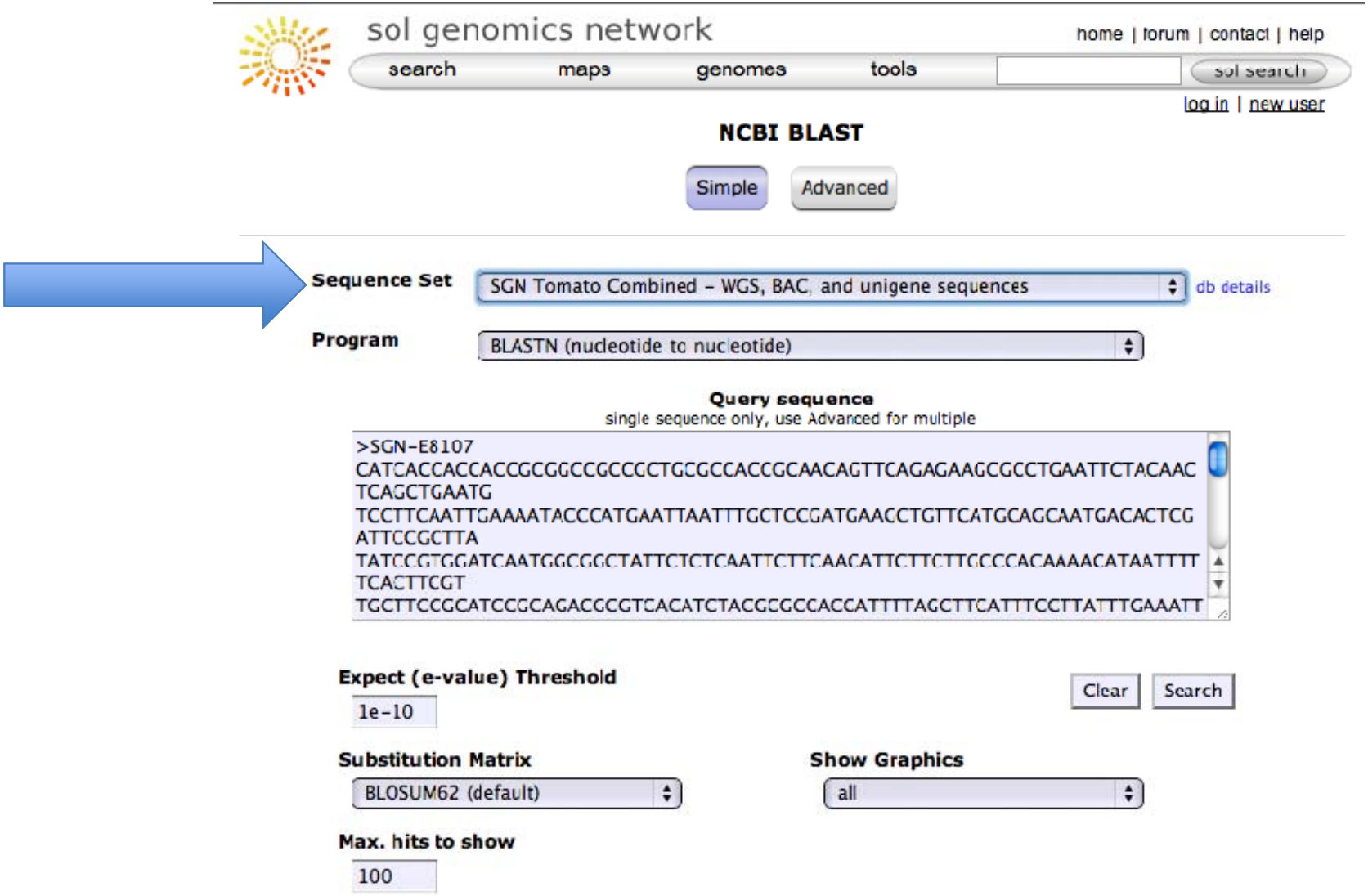
Locating sequence

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



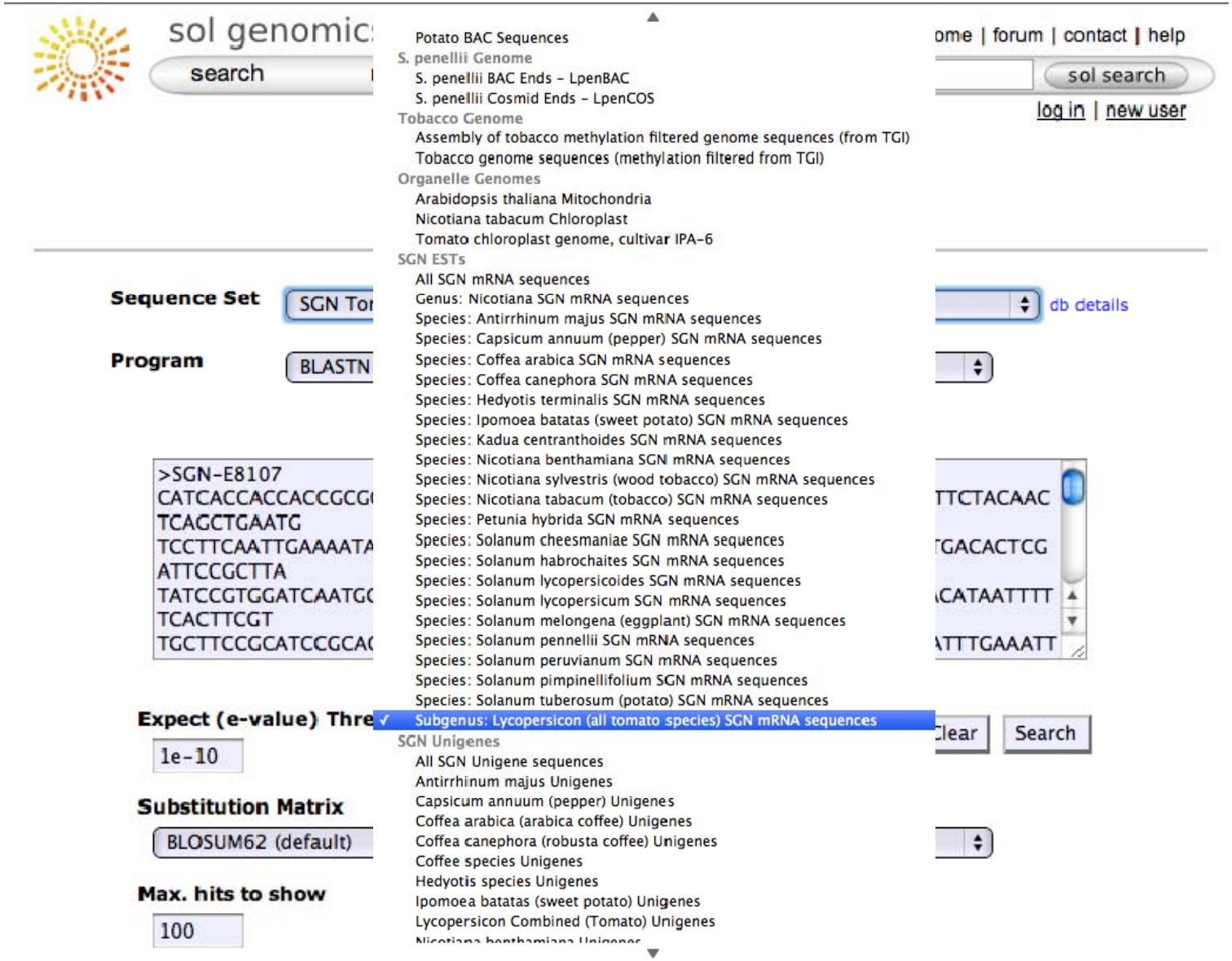

SGN EST Search Result – Sol Genomics Network
http://solgenomics.net/search/est.pl?request_from=0&request_id=SGN-E8107&request_type=7
Google
Apple
Yahoo!
Google Maps
YouTube
Wikipedia
News (372)
Popular
sol genomics network
home | forum | contact | help
search
maps
genomes
tools
sol search
log in | new user
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)
CATCACCACCACCGCGGCCGCCGCTGCGCCACCGCAACAGTTCAGAGAAGCGCCTGAATTCTACAACTCAGCTGAATGTCCTTCAATTGAAAATA
CCCATGAATTAATTTGCTCCGATGAAGCTGTTCATGCAGCAATGACACTCGATTCCGCTTATATCCGTGGATCAATGGCGGCTATTCTCTCAATT
CTTCAACATTCTTCTTGCCCACAAAACATAATTTTTCACTTCGTTGCTTCCGCATCCGCAGACGCGTCACATCTACGCGCCACCATTTTAGCTTC
ATTTCCTTATTTGAAATTCGAAGTTTATCGATTCGACGATTCCTCTGTTTCCGGTTTAATTTCAACTTCTATTCGTTCAGCTCTCGATTGCCCTT
TGAACTATGCGAGAAGTTATTTAGCTAATATTTTACCTTTATGTGTTCGGAAAGTTGTGTATTTGGATTCCGATTTGGTTTTAGTTGACGATATT
GCGAAGCTATCAGAAACTCCACTCGGGGAAGATCATATACTTGCTGCTCCTGAATACTGTAATGCGAATTTCACTTCGTATTTTACTCCGACTTT
CTGGTCCAACCCATCTCTATCTCTTACATTTGCGGATCGGAAAGCTTGTTATTTTAATACAGGGGTAATGGTAATCGATCTCGATCGATGGGAGA
ATGGTGATTATACCAGGAAGATTGAAGAATGGATGGAA
[BLAST] [A
BLAST

Consider what to BLAST against…

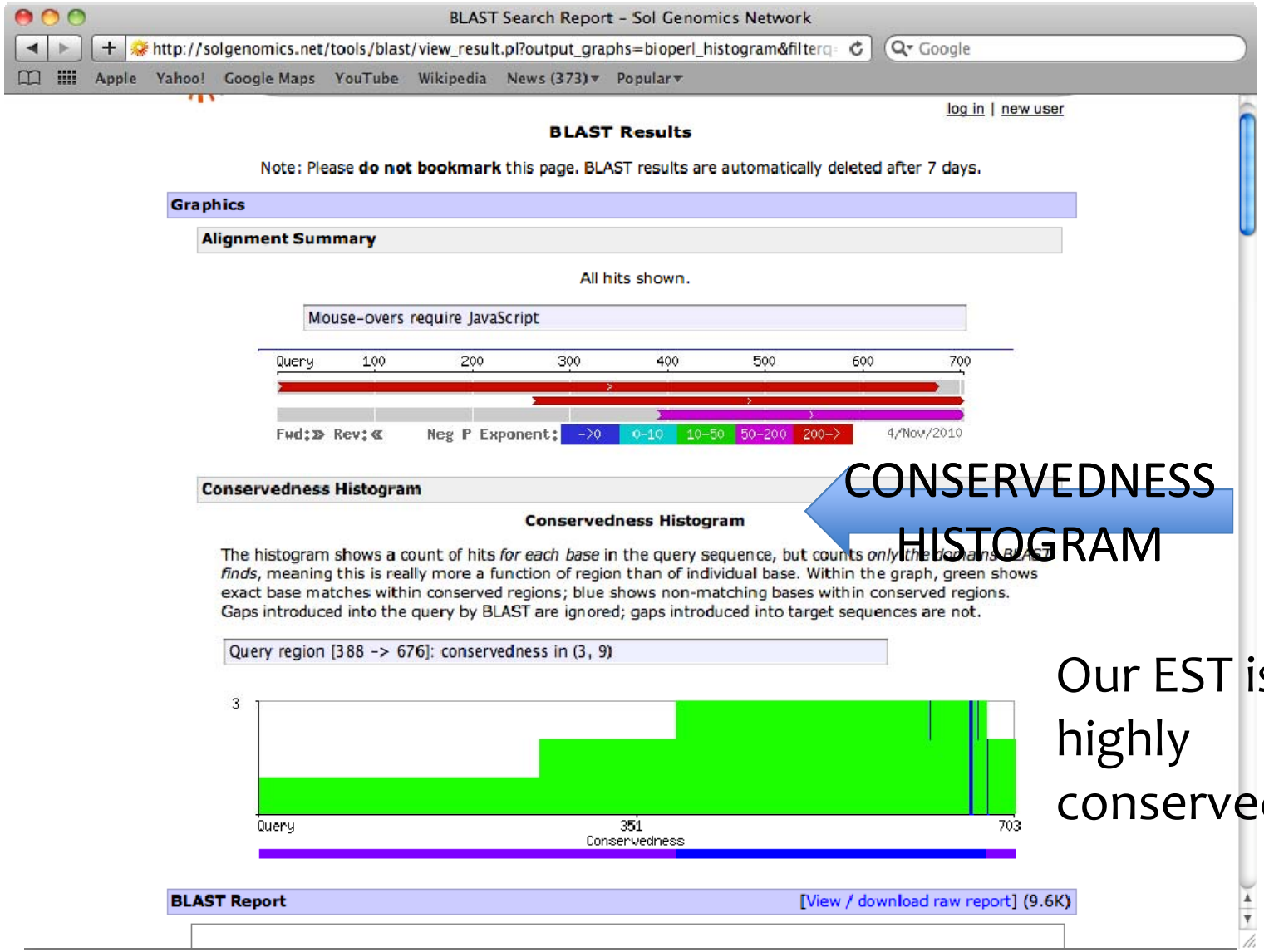


There are a LOT of options

I chose Lycopersicon mRNA



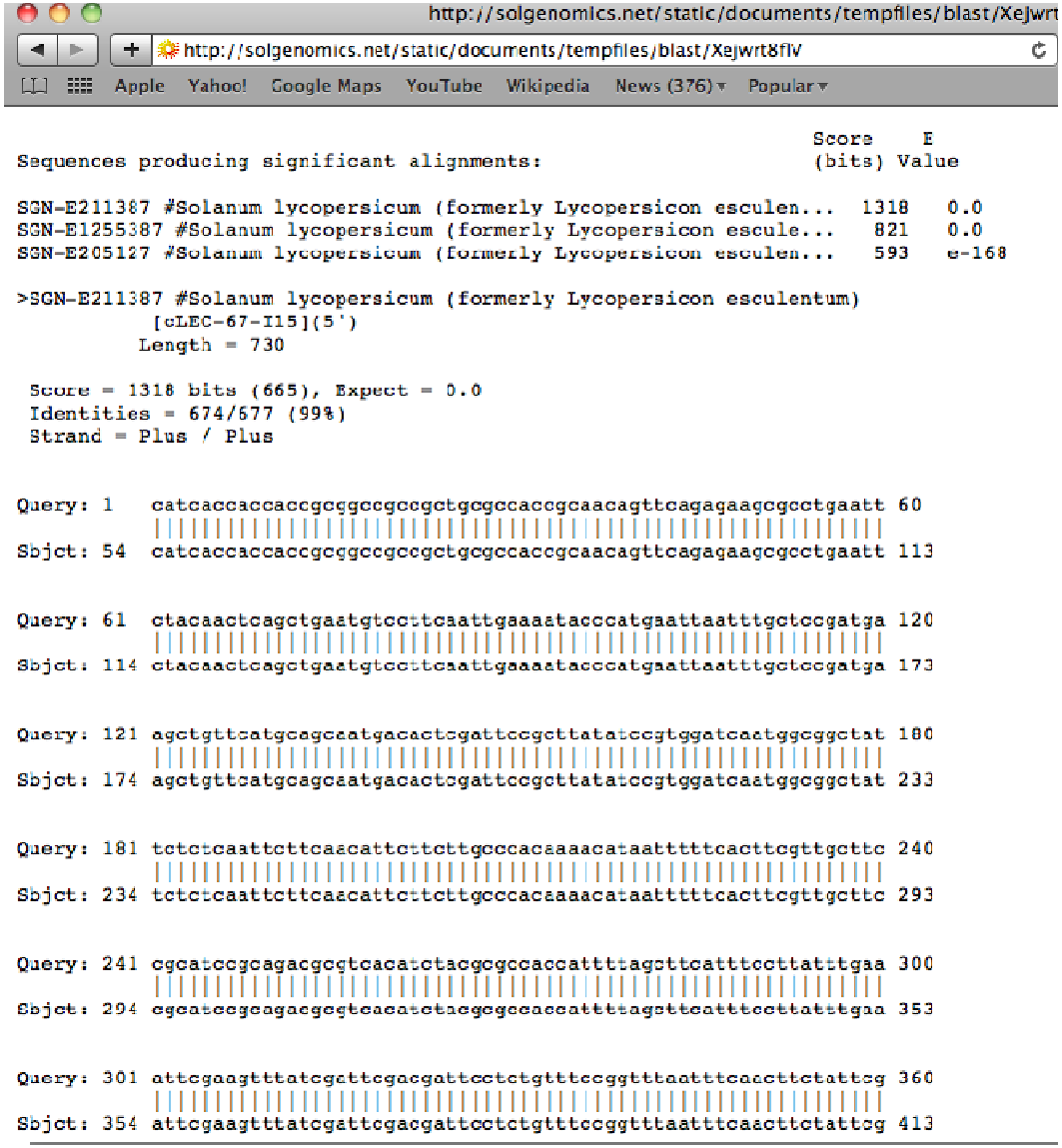
BLAST Results



CONSERVEDNESS HISTOGRAM

Our EST is highly conserved

BLAST Results



- 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

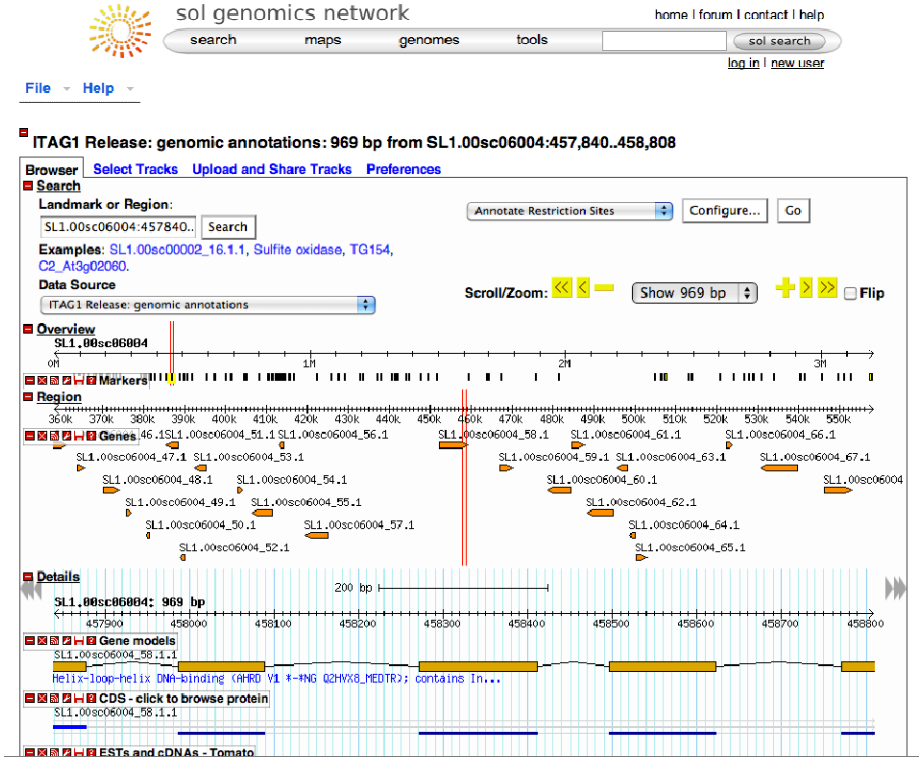


Photo courtesy of D. Francis, Ohio State University