SGN, a community-based, clade oriented database for the Solanaceae

Lukas Mueller

Naama Menda

Isaak Tecle

Robert Buels

Adri Mills

Dean Eckstrom

Anuradha Pujar

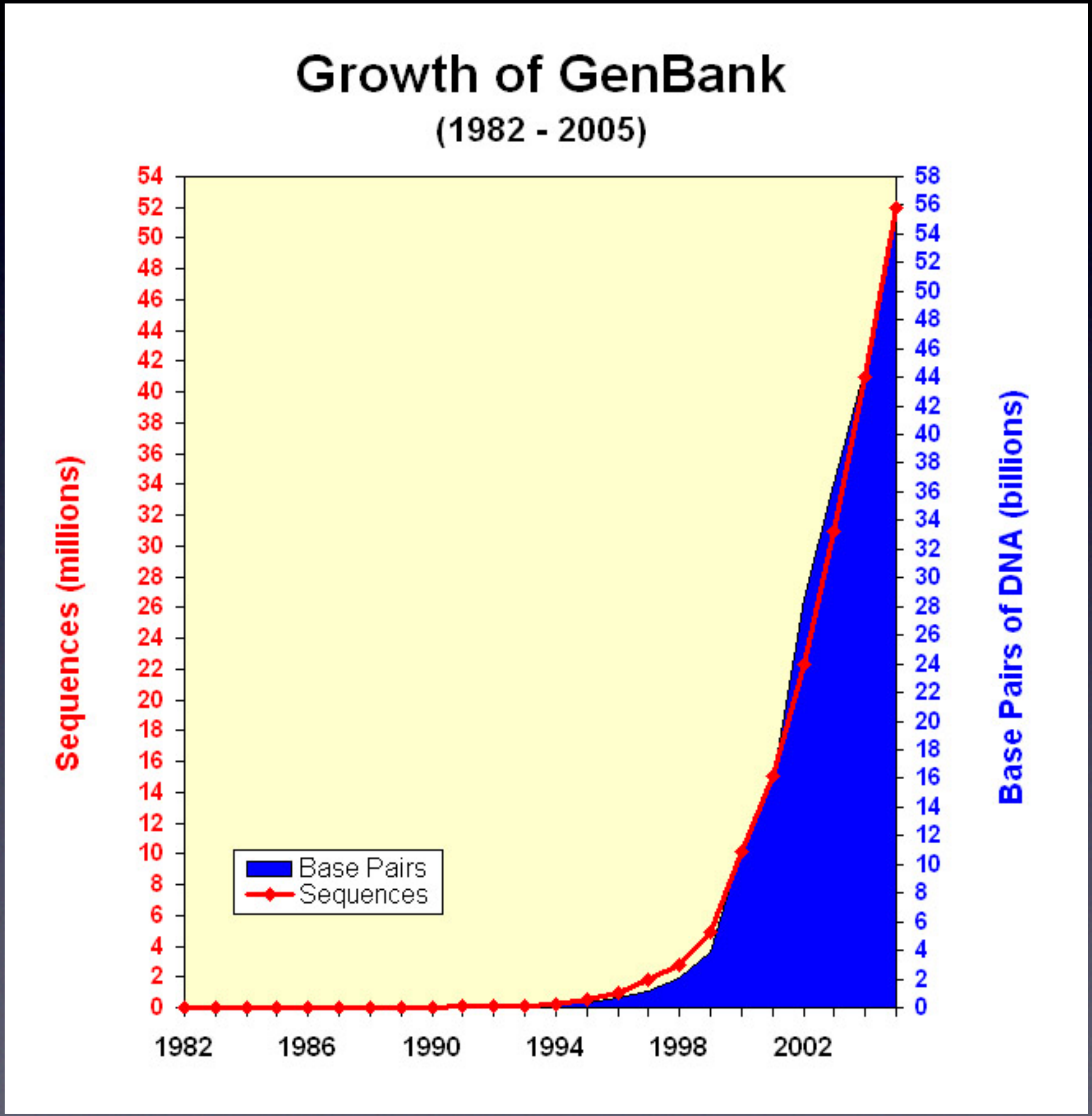
Tom York

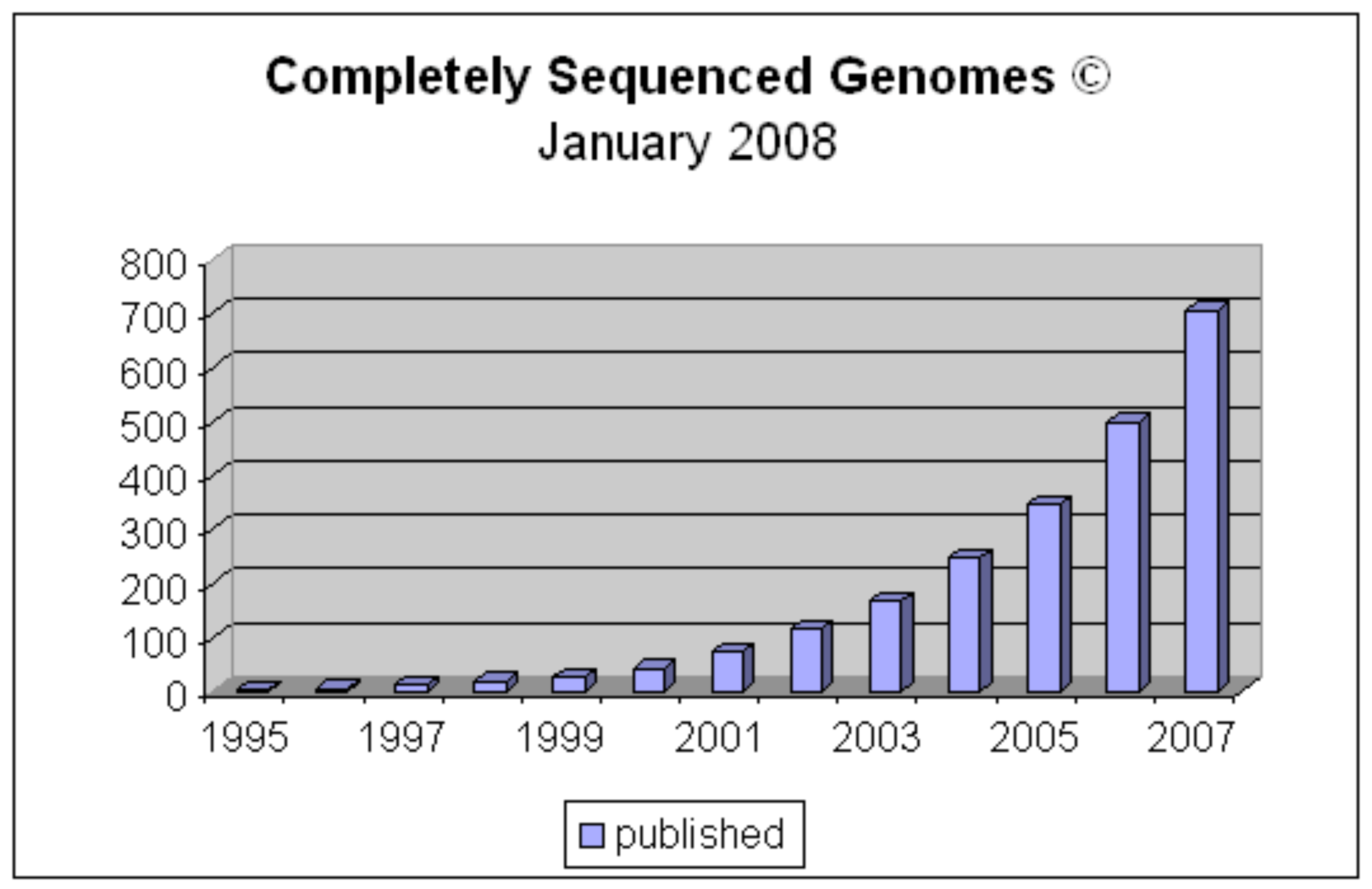
Chris Carpita

Aureliano Bombarley

The Genomic Era

- Rapid growth of sequence, SNP, expression data





Completely Sequenced Genomes ©
January 2008

The (post-) genomic era

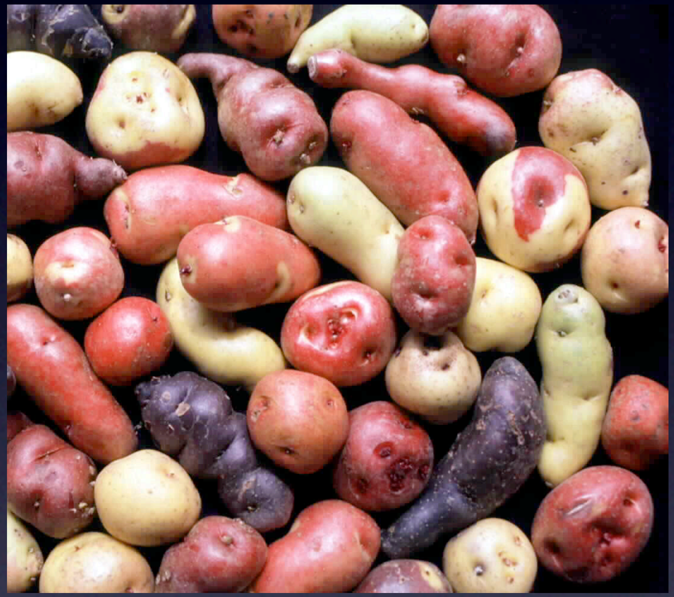
- Massive amounts of sequence data
- Biology has become an information-based science
- Data infrastructure, access, querying and visualization
- Curation & Annotation

Solanaceae as a model for phenotypic diversity

Similar genomes can lead to many profoundly outcomes

Tomato

Eggplant



Potato

Pachytene chromosomes

The International Solanaceae Genome Project (SOL):
Systems Approach to Diversity and Adaptation

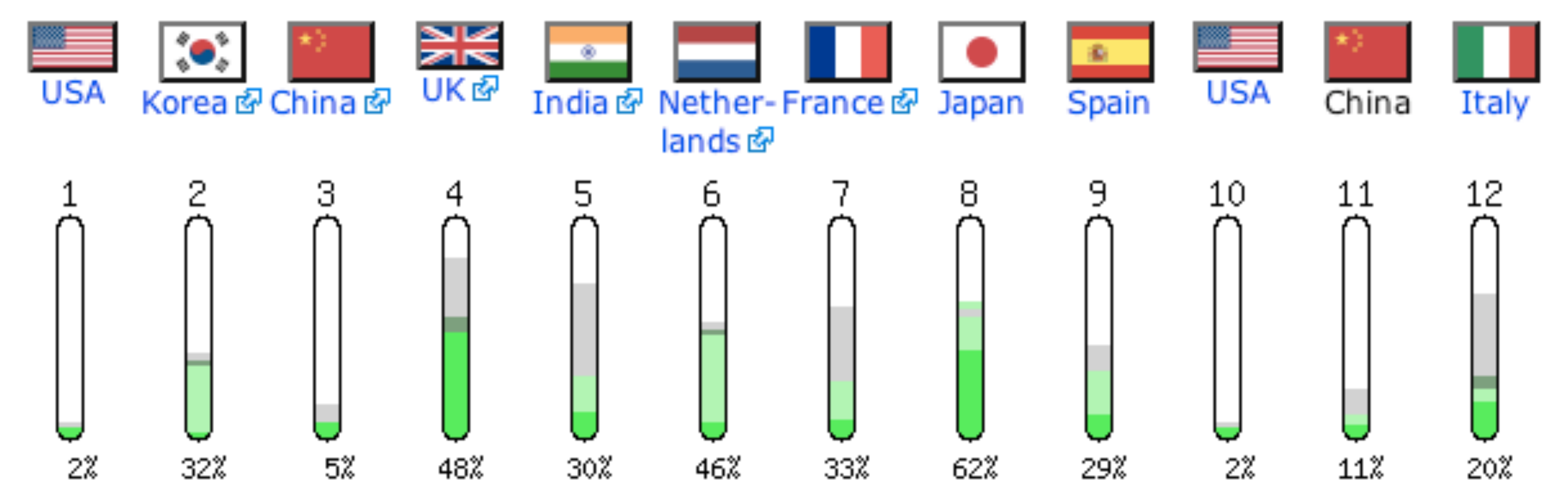
SOL whitepaper, 2004

How can a common set of genes give rise a wide range of morphologically and ecologically distinct organisms that occupy our planet?

How a deeper understanding of genetic basis of plant diversity can be harnessed to better meet the needs of society in an environmentally-friendly and sustainable manner?

SOL Project

- Cornerstone: Sequencing of tomato genome as a reference genome for the clade



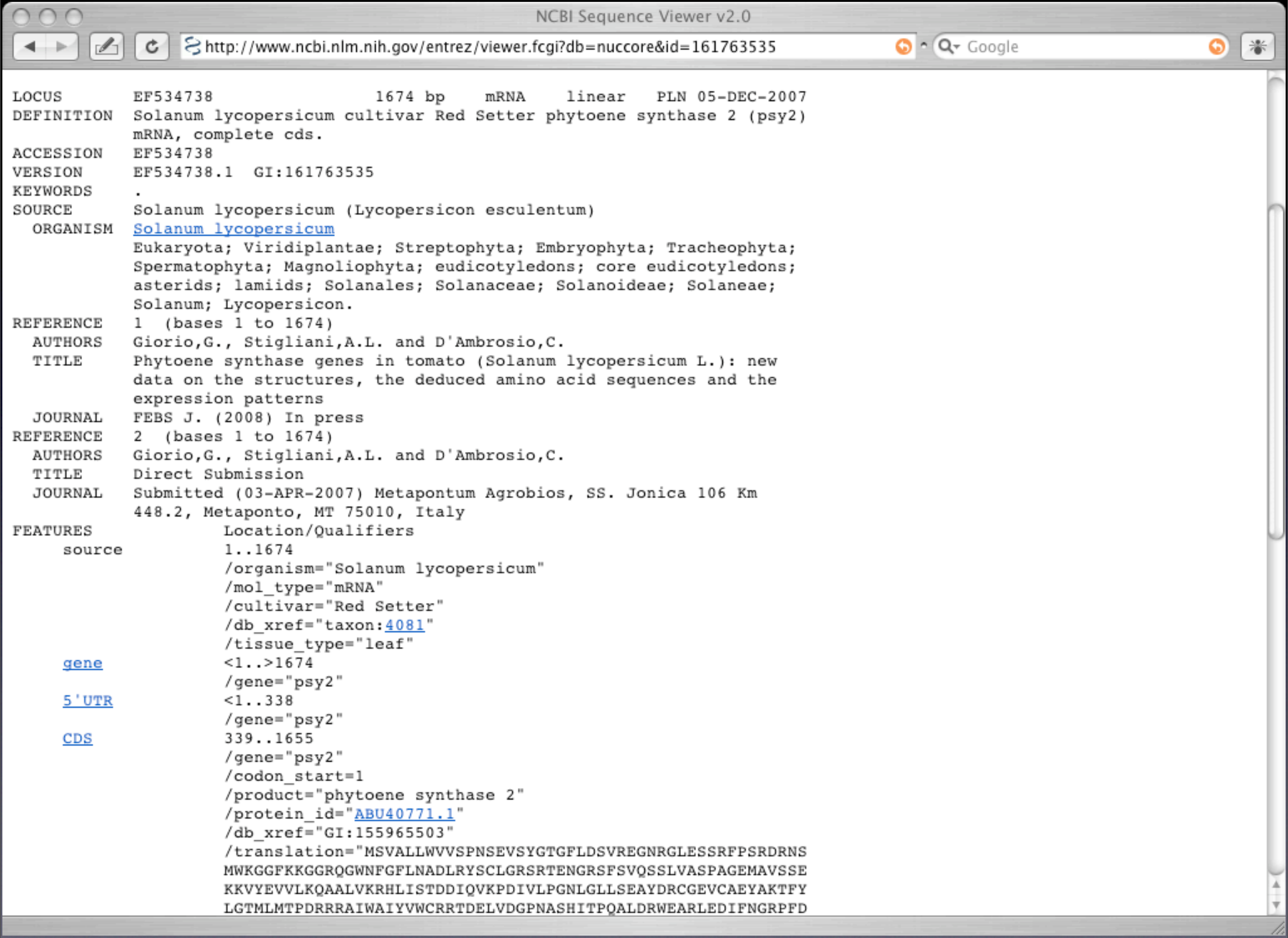
Plant Databases

tair

MaizeGDB
Maize Genetics and Genomics Database

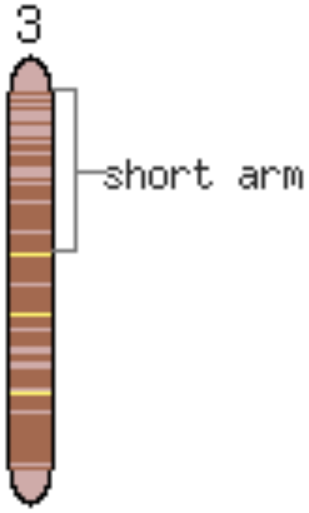
GRAMENE

SGN

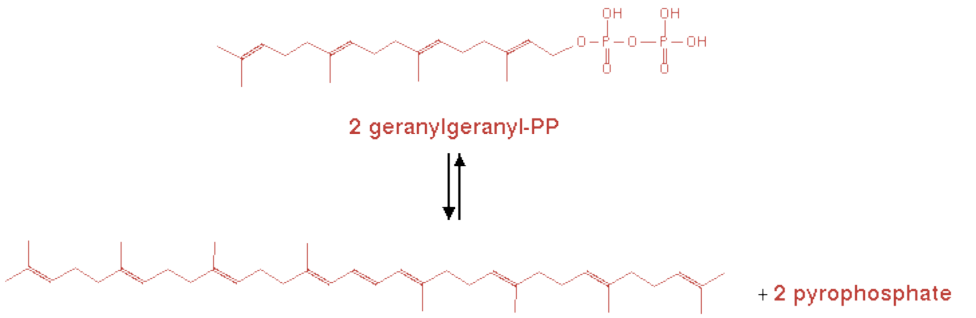


```
LOCUS       EF534738                1674 bp    mRNA    linear   PLN 05-DEC-2007
DEFINITION  Solanum lycopersicum cultivar Red Setter phytoene synthase 2 (psy2)
            mRNA, complete cds.
ACCESSION   EF534738
VERSION     EF534738.1  GI:161763535
KEYWORDS    .
SOURCE      Solanum lycopersicum (Lycopersicon esculentum)
  ORGANISM  Solanum lycopersicum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanoideae; Solaneae;
            Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 1674)
  AUTHORS   Giorio,G., Stigliani,A.L. and D'Ambrosio,C.
  TITLE     Phytoene synthase genes in tomato (Solanum lycopersicum L.): new
            data on the structures, the deduced amino acid sequences and the
            expression patterns
  JOURNAL   FEBS J. (2008) In press
REFERENCE   2  (bases 1 to 1674)
  AUTHORS   Giorio,G., Stigliani,A.L. and D'Ambrosio,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2007) Metapontum Agrobios, SS. Jonica 106 Km
            448.2, Metaponto, MT 75010, Italy
FEATURES             Location/Qualifiers
     source          1..1674
                     /organism="Solanum lycopersicum"
                     /mol_type="mRNA"
                     /cultivar="Red Setter"
                     /db_xref="taxon:4081"
                     /tissue_type="leaf"
     gene            <1..>1674
                     /gene="psy2"
     5'UTR           <1..338
                     /gene="psy2"
     CDS             339..1655
                     /gene="psy2"
                     /codon_start=1
                     /product="phytoene synthase 2"
                     /protein_id="ABU40771.1"
                     /db_xref="GI:155965503"
                     /translation="MSVALLWVVSPNSEVSYGTGFLDSVREGNRGLESSRFPSRDRNS
                     MWKGGFKKGGRQGWNFGFLNADLRYSCLGRSRTENGRSFSVQSSLVASPAGEMAVSSE
                     KKVYEVVLKQAALVKRHLISTDDIQVKPDIVLPGNLGLLSEAYDRCGEVCAEYAKTFY
                     LGTMLMTPDRRRAIWAIYVWCRRTDELVDGPNASHITPQALDRWEARLEDIFNGRPFD
```

Locus name	**yellow flesh / phytoene synthase**
Locus symbol	**r**
Gene activity	**phytoene synthase**
Description	**Psy1 encodes phytoene synthase. It is predominantly expressed in early seedling and late fruit ripening stages.**
Chromosome	**3**
Arm	**short**



Tomato-EXPEN 2000



Biological process

involved in carotenoid biosynthesis (GO:0016117)

Cellular component

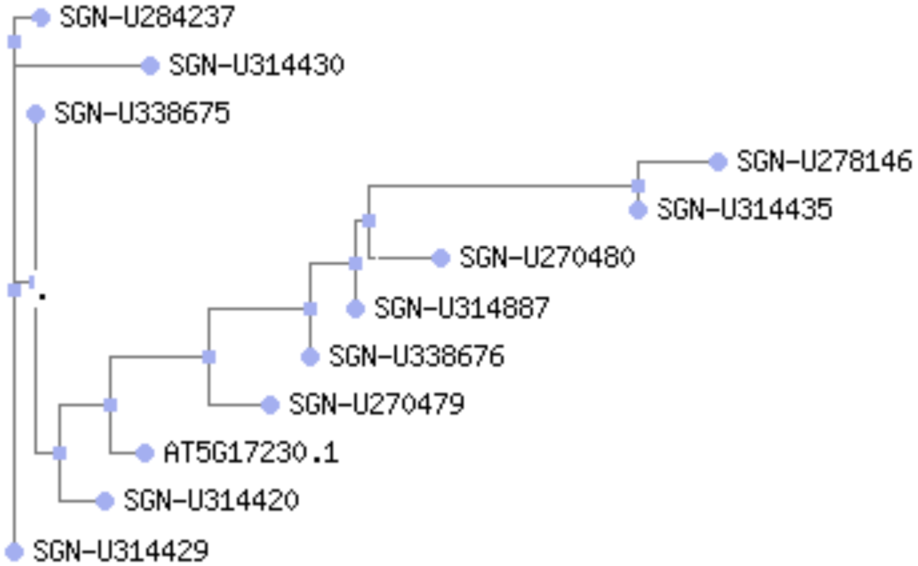
located in chloroplast (GO:0009507)

Plant structure

expressed in fruit (PO:0009001)

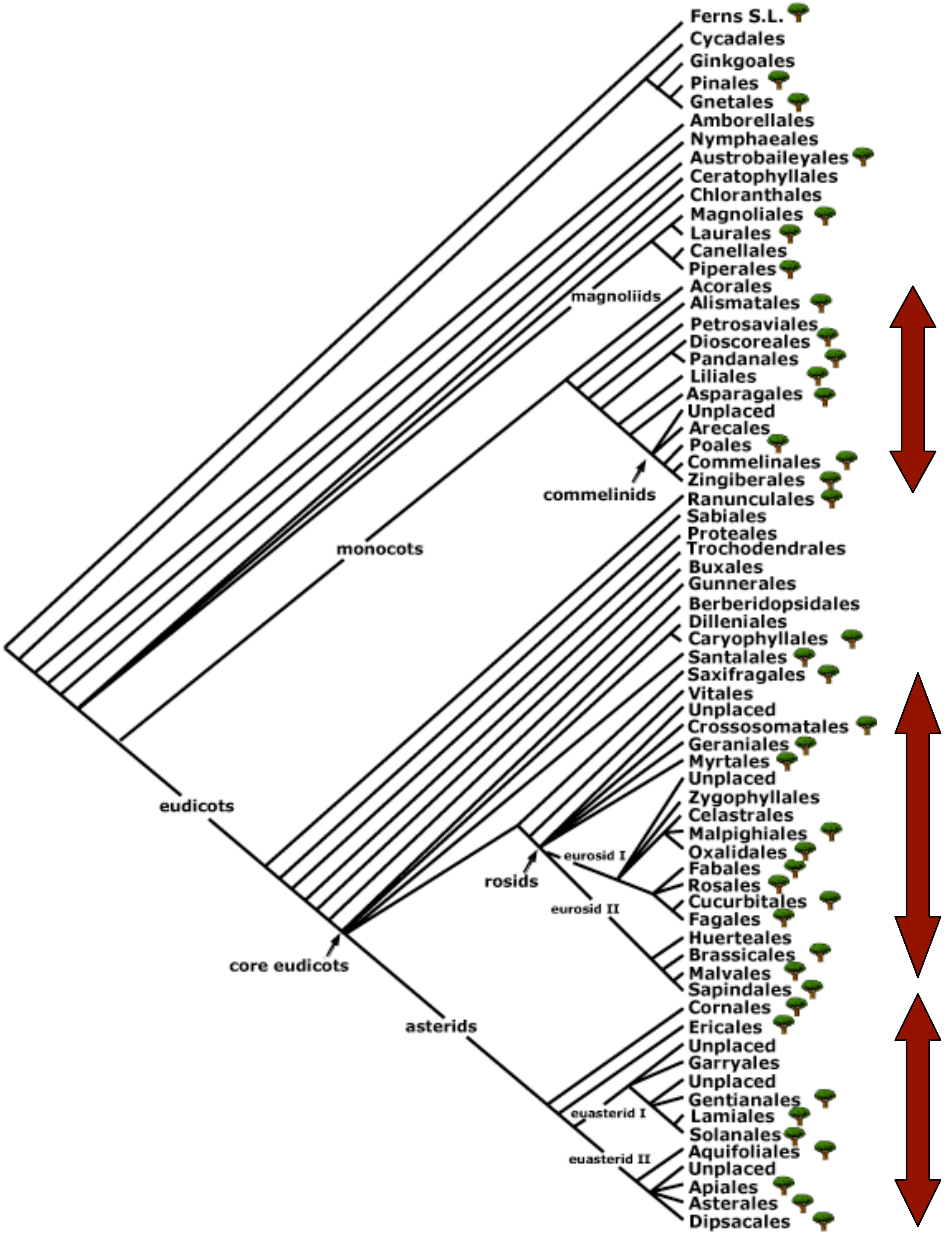
Solanaceae phenotypes

has phenotype yellow fruit (SP:0000121)



Model Organism Databases (MOD)

- “vertically” integrate all the information for that species
- Create a new database for every sequenced species
- Limitations of the MODs: Comparative queries difficult, expensive to scale to many organisms



Monocots
rice, maize,
wheat

Rosids
Arabidopsis,
poplar, *Medicago*

Asterids
tomato, potato,
pepper, eggplant
coffee, tea,
snapdragon,
morning glory,
olives, lettuce,
carrots Gerbera

Clade Oriented Database (COD)

- The model developed at SGN
- Relate each plant of a clade to as many other plants in the clade as possible
- Relate to important model systems outside the clade (Arabidopsis, rice)
- “horizontal” integration of data from species in the clade in the same database
- Advantages: facilitates comparative queries, more scalable and affordable

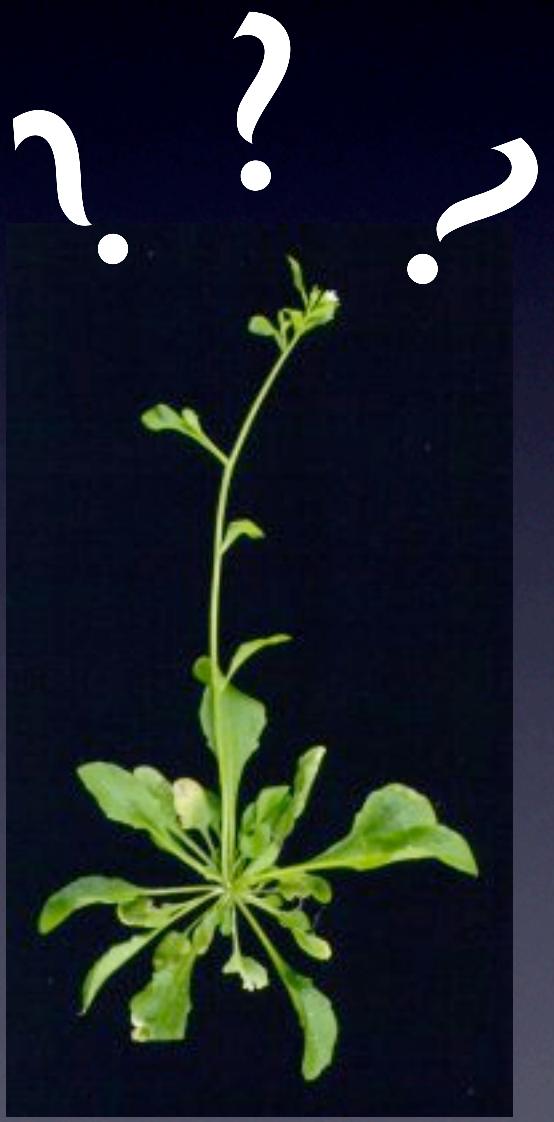
Outline

- Why Solanaceae? - Unique Solanaceae Tools
- From phenotypes to genotypes
- Biochemical Pathways
- Tomato Genome Sequencing Project
- Solanaceae Gene Families
- Locus community annotation
- Future directions

Why Solanaceae?

- What makes Solanaceae so interesting?
  - Unique biology
  - Unique tools
- How can we accommodate the Solanaceae specific tools in a database?

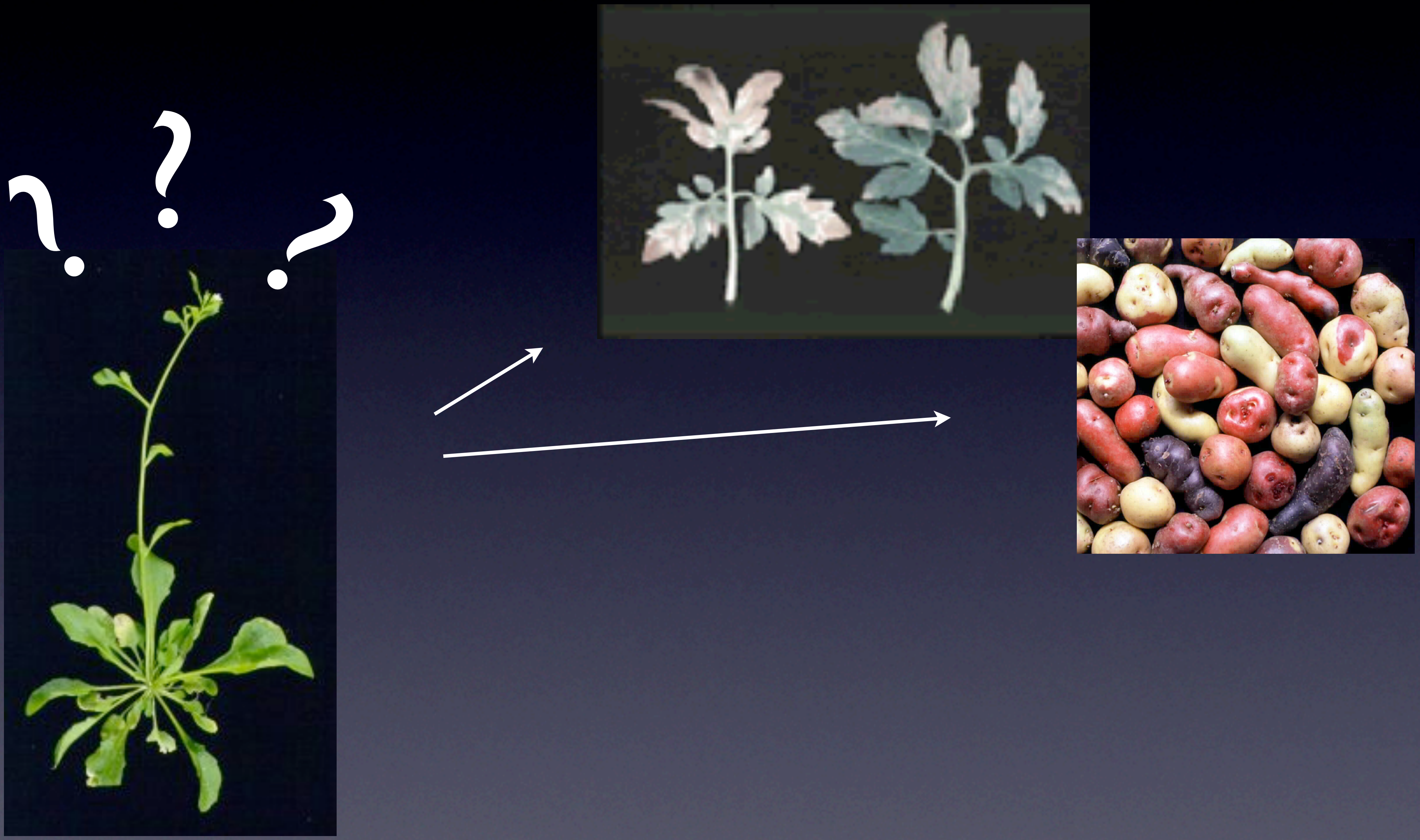
Model systems are limited to their biology



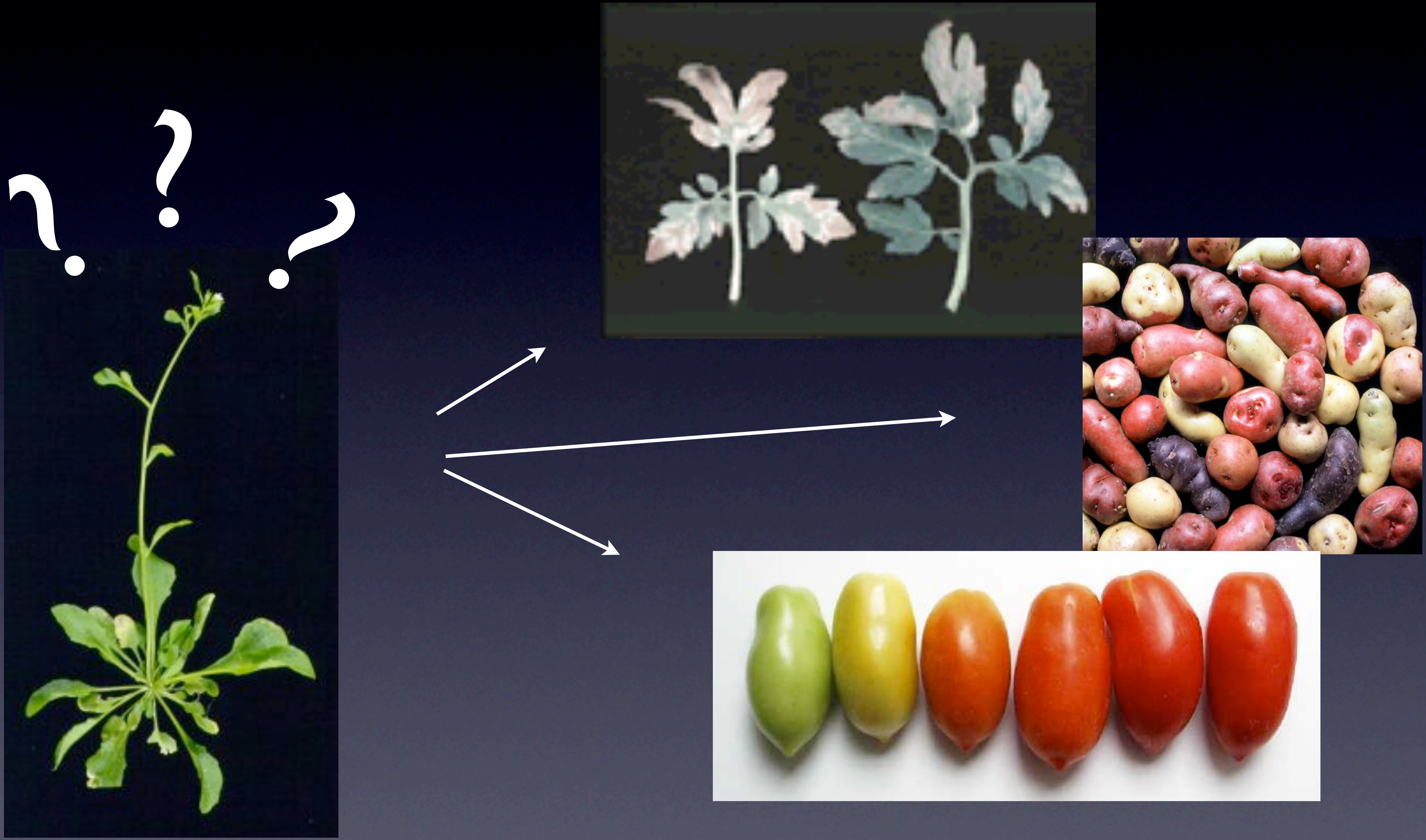
Model systems are limited to their biology

? ? ?

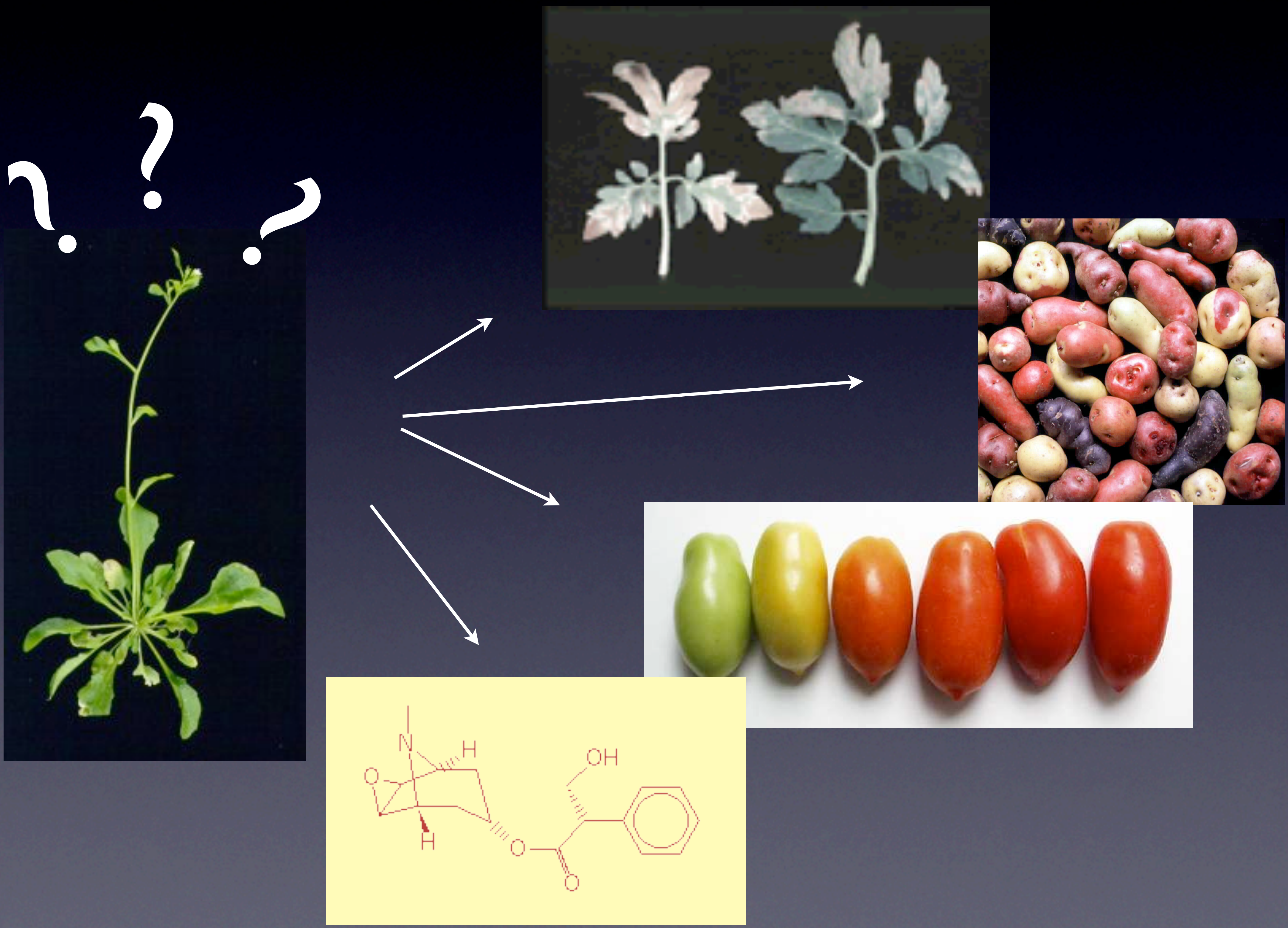
Model systems are limited to their biology



Model systems are limited to their biology



Model systems are limited to their biology



Solanaceae toolbox

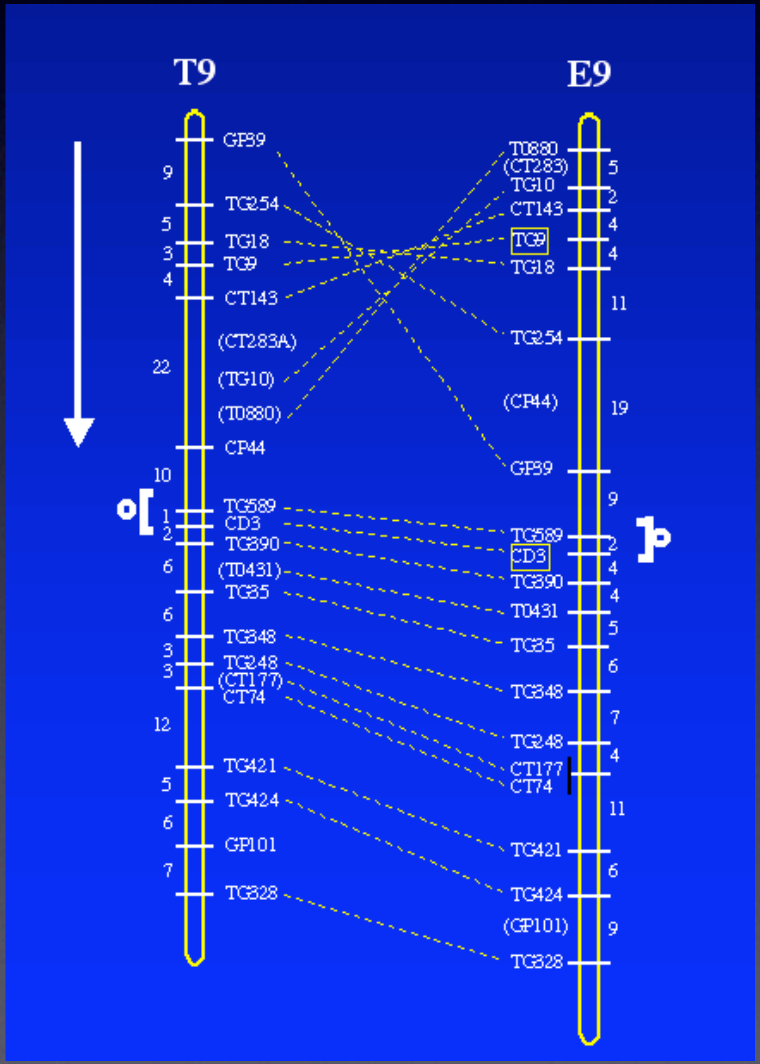
- Solanaceae community
- Make unique Solanaceae tools and infrastructure available to researchers
- Allow researchers expand the content of the toolbox

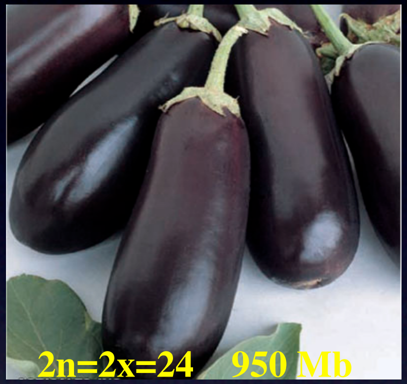
Solanaceae genomes are tied together by comparative genetic maps

2n=2x=24 950 Mb

2n=4x=24 1800 Mb

2n=2x=24 1300 Mb





2n=2x=24 950 Mb

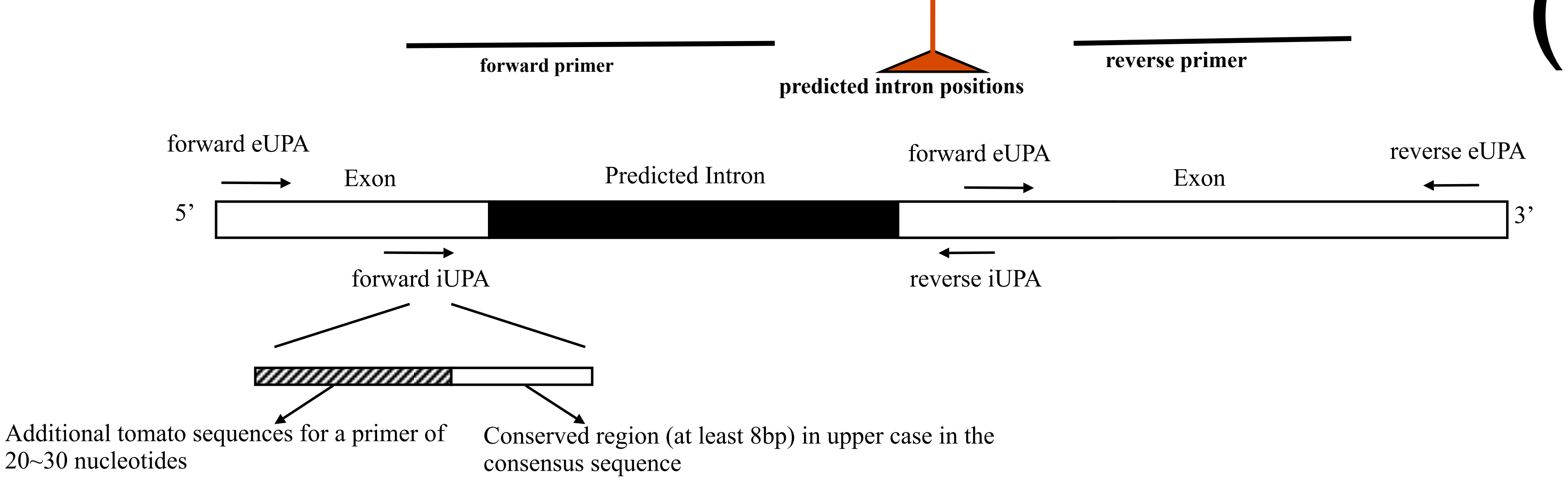
2n=2x=24 3000 Mb

2n=2x= 22 ca. 1000 Mb

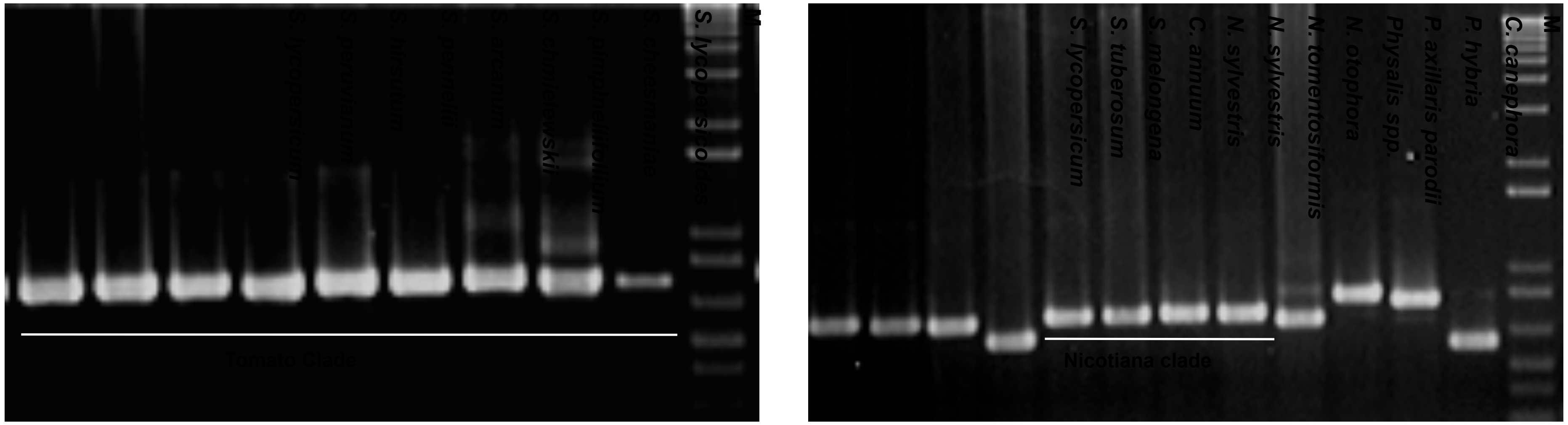
COSII Markers for Comparative Mapping in Euasterid Species

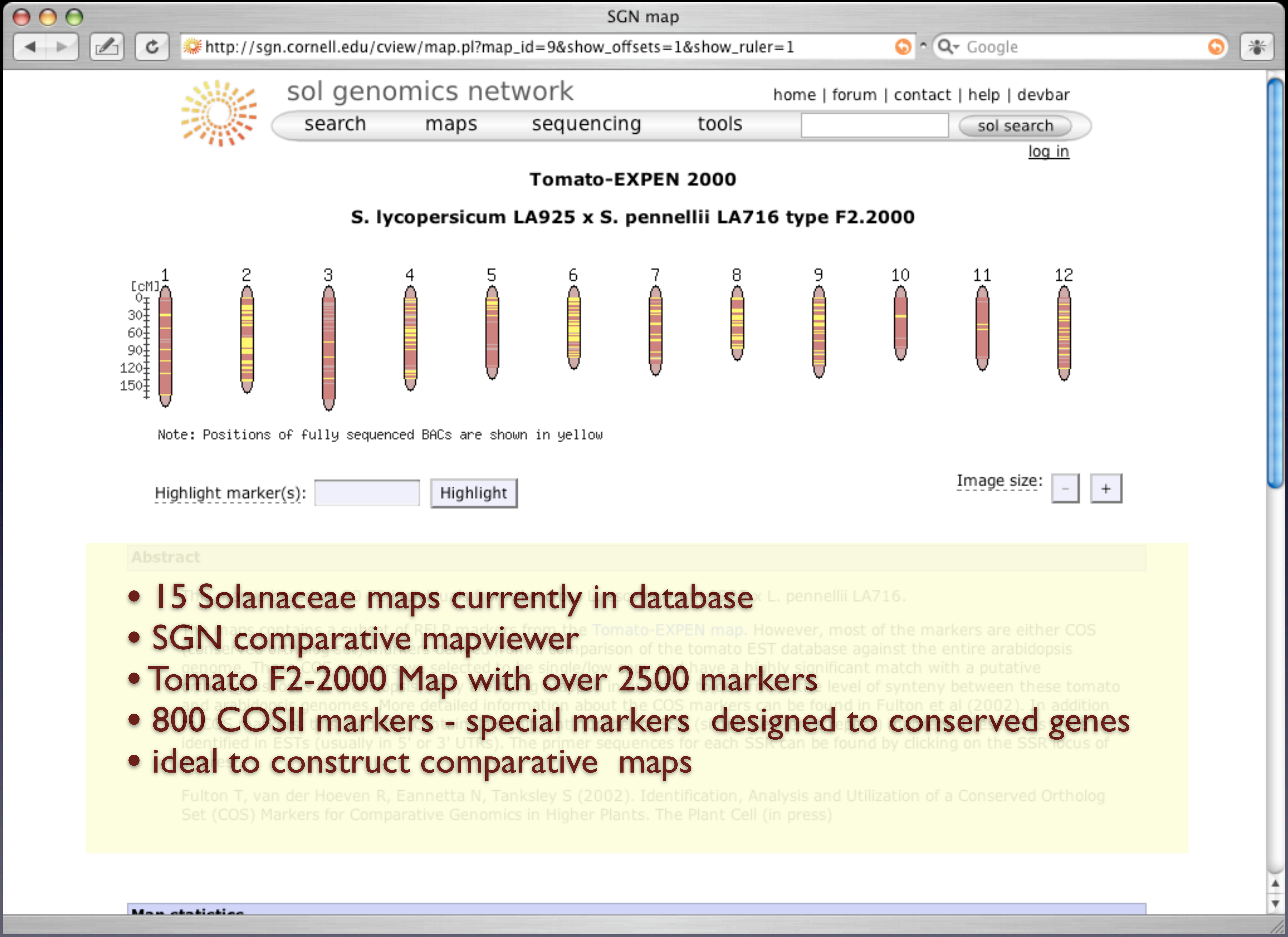
over 2500
CAPS-based
COS markers
(Feinan Wu)

```
At5g06360.1        AAAACATTGAAAATGCATGAAGAGTCATCATCAGCCAAGGTTCTTAAAGCTGGCAAATGGGAGGTCCCTCTTCCA
At5g06360.1        K--T--L--K--M--H--E--E--S--S--S--A--K--V--L--K--A--G--K--W--E--V--P--L--P--
tomato214584       AAAACATTGGCTATGCATGAAGAGTCATCATCCGCTAAGATTCTTAAAGCTGGAAAATGGGAGGTGCCTCTGCCC
tomato214584       K--T--L--A--M--H--E--E--S--S--S--A--K--I--L--K--A--G--K--W--E--V--P--L--P--
potato175718       AAAACATTGGCTATGCATGAAGAGTCATCATCCGCTAAGATTCTTAAAGCTGGAAAATGGGAGGTGCCTCTACCT
potato175718       K--T--L--A--M--H--E--E--S--S--S--A--K--I--L--K--A--G--K--W--E--V--P--L--P--
pepper197506       AAAACATTGGCTATGCATGAAGAGTCATCAGCCGCCAAGATTCTCAAAGCTGGAAAATGGGAGGTGCCTCTGCCT
pepper197506       K--T--L--A--M--H--E--E--S--S--A--A--K--I--L--K--A--G--K--W--E--V--P--L--P--
coffee120708       AAAACTTTGGCTATGCATGAGGAGTCATCAGCCGCCAAGGTTCTCAAAGCTGGAAAATGGGATGTACCTCTTCCC
coffee120708       K--T--L--A--M--H--E--E--S--S--A--A--K--V--L--K--A--G--K--W--E--V--P--L--P--
consensus          AAAACWTTGGCTATGCATGARGAGTCATCAKCCGCYAAGRTTCTYAAAGCTGGAAAATGGGAKGTRCCTCTDCCY
Modified consensus AAAACaTTGGCTATGCATGAaGAGTCATCAtCCGCtAAGaTTCTtAAAGCTGGAAAATGGGAgGTgCCTCTgCCc
```



iUPAs for C2_At1g13380





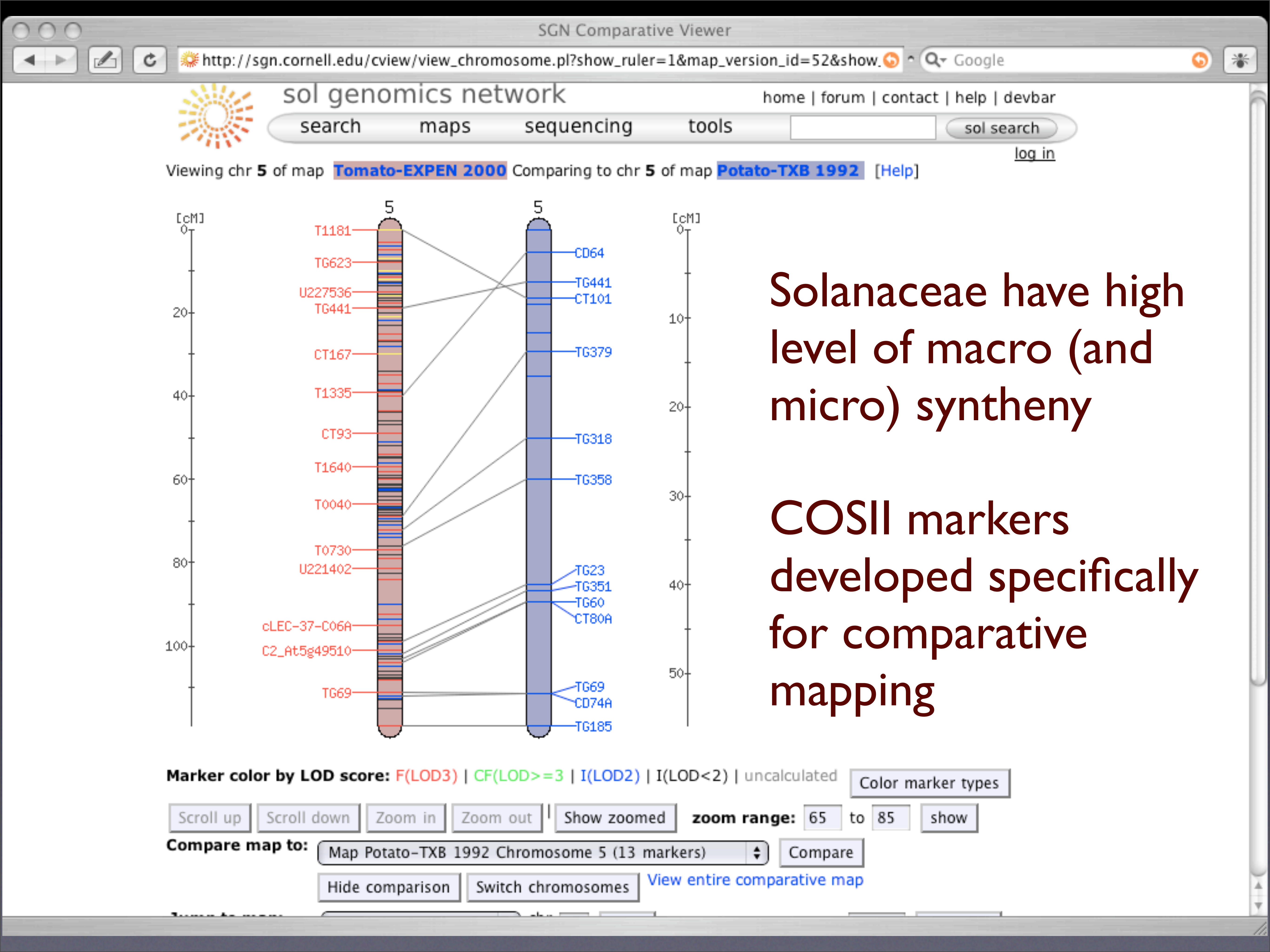
Tomato-EXPEN 2000

S. lycopersicum LA925 x S. pennellii LA716 type F2.2000

Note: Positions of fully sequenced BACs are shown in yellow

- 15 Solanaceae maps currently in database
- SGN comparative mapviewer
- Tomato F2-2000 Map with over 2500 markers
- 800 COSII markers - special markers designed to conserved genes
- ideal to construct comparative maps

Fulton T, van der Hoeven R, Eannetta N, Tanksley S (2002). Identification, Analysis and Utilization of a Conserved Ortholog Set (COS) Markers for Comparative Genomics in Higher Plants. The Plant Cell (in press)



SGN Comparative Viewer

http://sgn.cornell.edu/cview/view_chromosome.pl?show_ruler=1&map_version_id=52&show.

sol genomics network

home | forum | contact | help | devbar

search maps sequencing tools sol search

log in

Viewing chr **5** of map **Tomato-EXPEN 2000** Comparing to chr **5** of map **Potato-TXB 1992** [Help]

Marker color by LOD score: F(LOD3) | CF(LOD>=3 | I(LOD2) | I(LOD<2) | uncalculated

Color marker types

Scroll up Scroll down Zoom in Zoom out Show zoomed **zoom range:** 65 to 85 show

Compare map to: Map Potato-TXB 1992 Chromosome 5 (13 markers) Compare

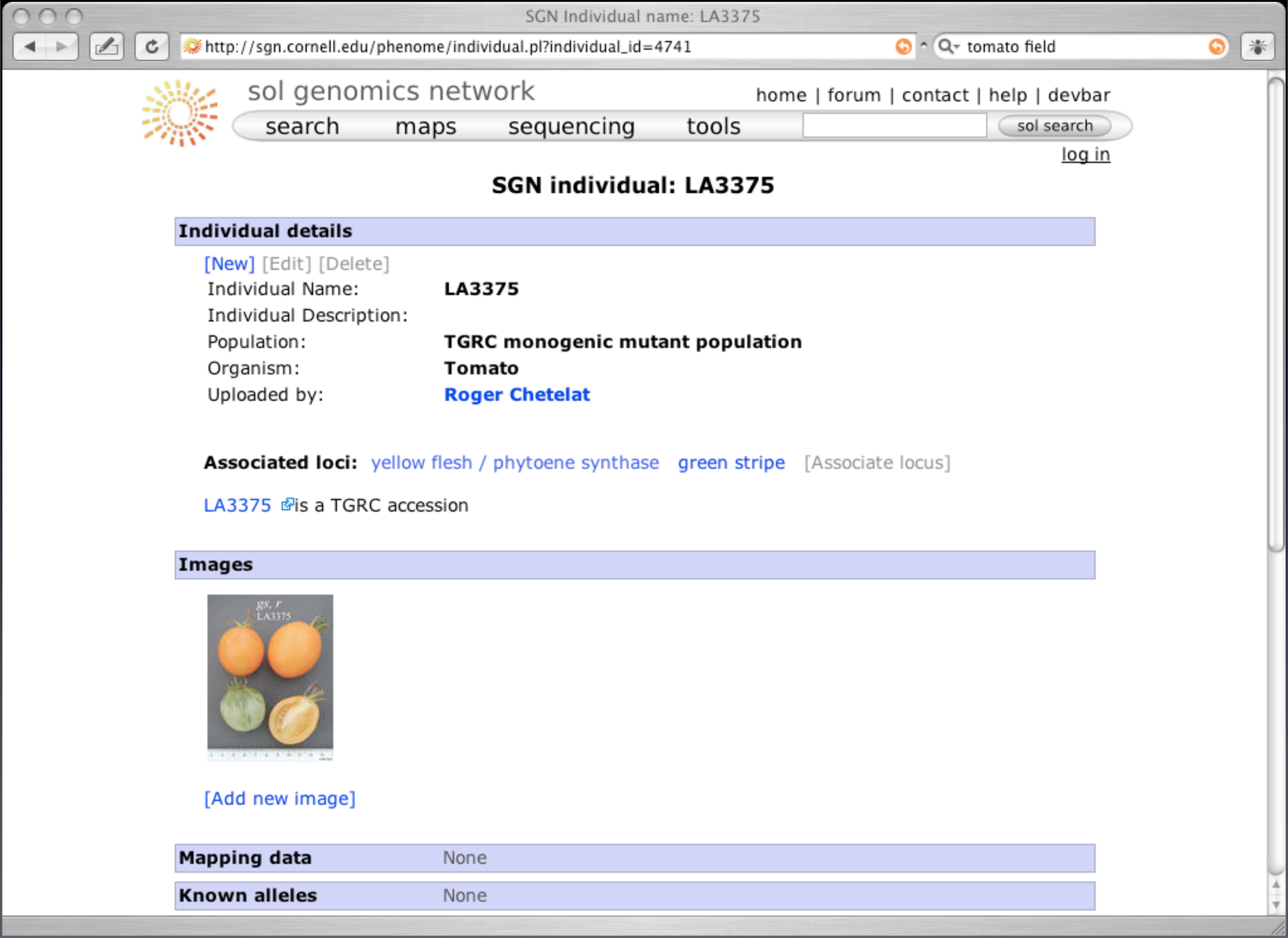
Hide comparison Switch chromosomes View entire comparative map

Solanaceae have high level of macro (and micro) syntheny

COSII markers developed specifically for comparative mapping

Mutant Populations

- 2500 tomato EMS-mutant lines
- 800 tomato fast neutron mutant lines
- 250 eggplant EMS-mutant lines
- 3800 TGRC (stock center) single mutant lines



SGN individual: LA3375

Individual details

Individual Name:	**LA3375**
Individual Description:	
Population:	**TGRC monogenic mutant population**
Organism:	**Tomato**
Uploaded by:	**Roger Chetelat**

Associated loci: yellow flesh / phytoene synthase green stripe

LA3375 is a TGRC accession

Images

Mapping data None

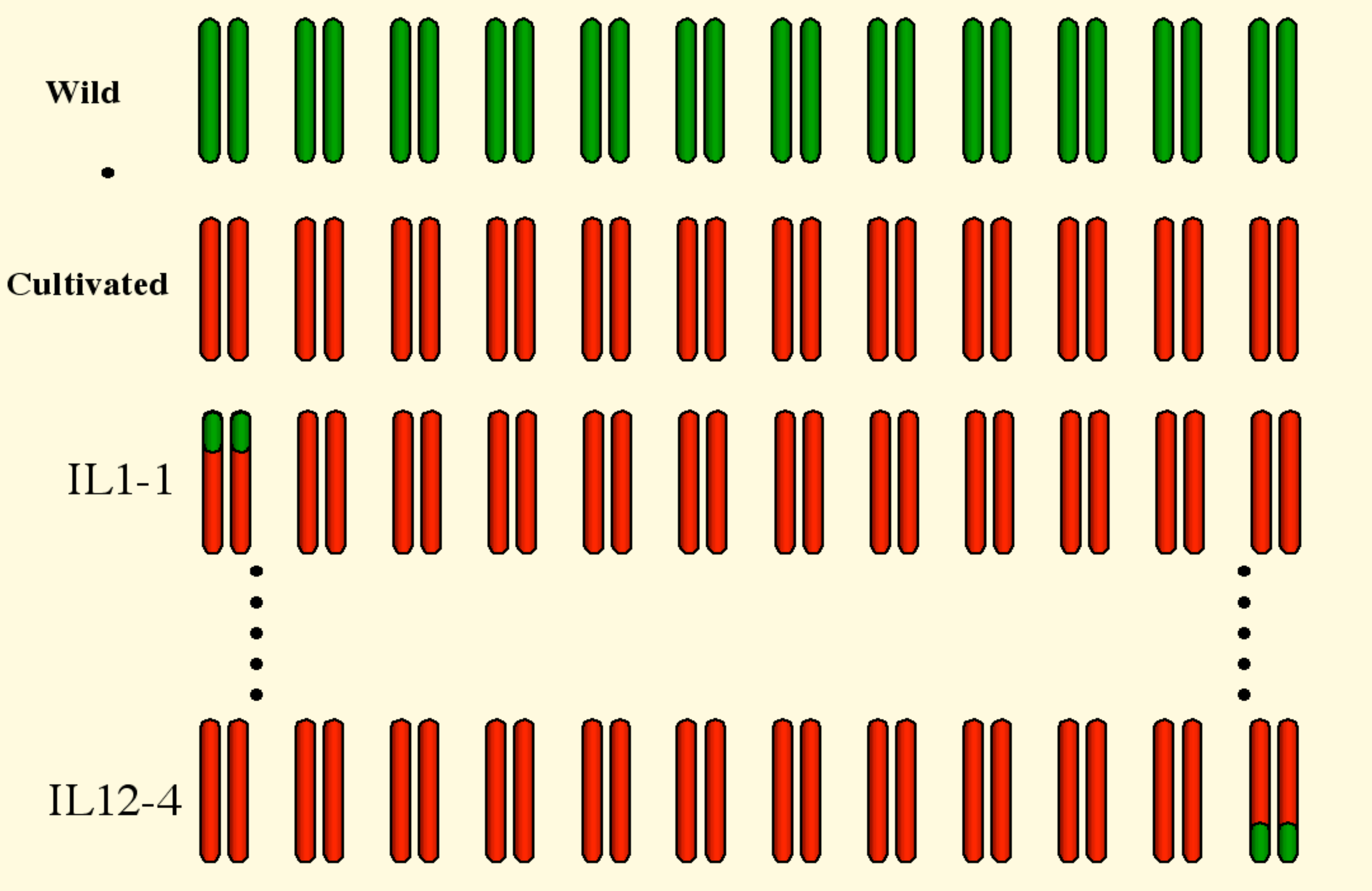
Known alleles None

Phenotypic descriptors

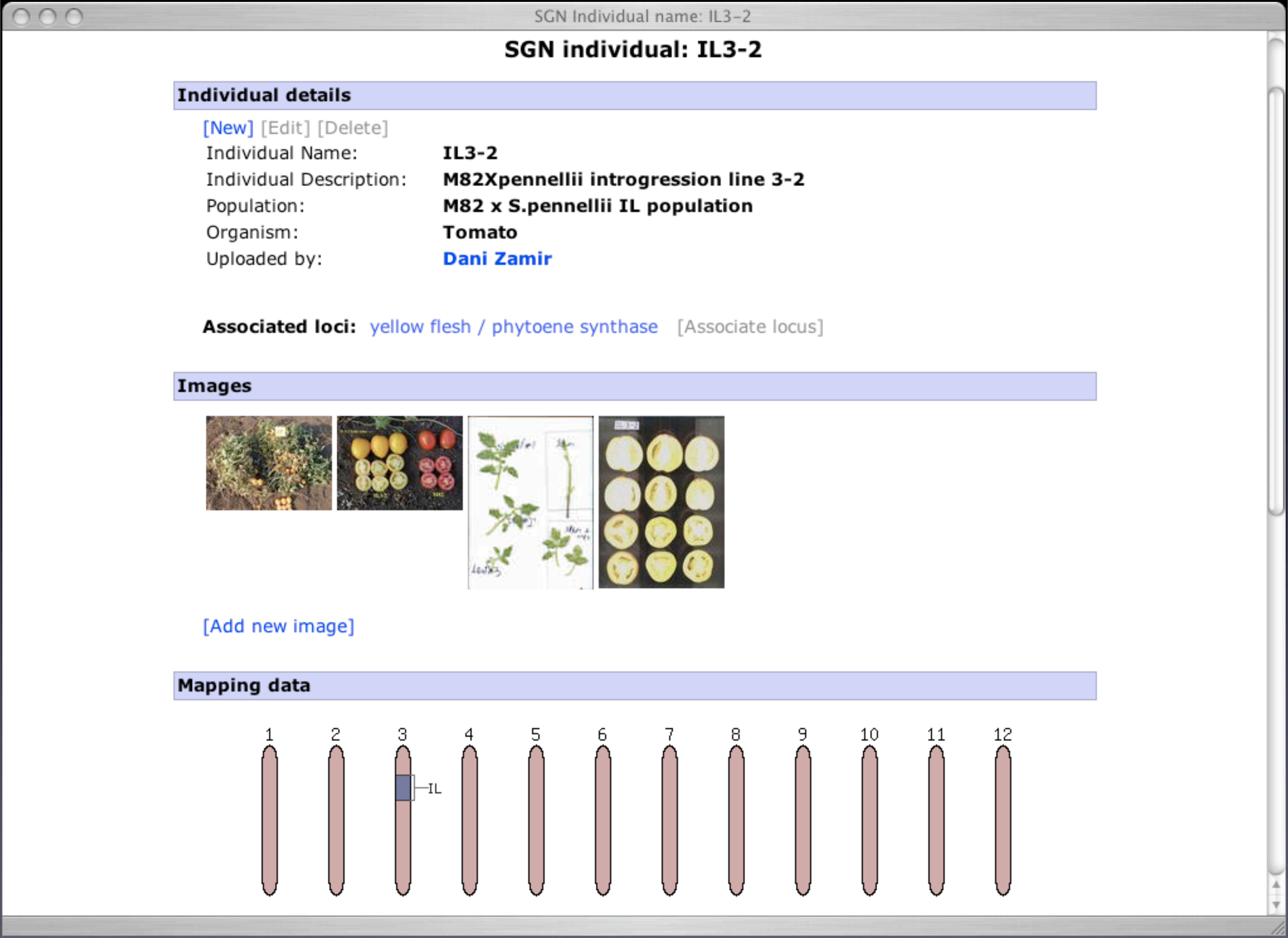
Standard Tools for biological databases

- **Plant Ontology** (PO)
  - ✓ describes plant structure
- **Solanaceae Phenotype** Ontology (SP)
  - ✓ describes plant phenotypes
- **Gene Ontology** (GO)
  - ✓ describes biological processes, cellular location, and molecular function

Zamir Introgression Lines



- used in many phenotypic projects
- provide direct link to a section of the genome



SGN Individual name: IL3-2

SGN individual: IL3-2

Individual details

[New] [Edit] [Delete]

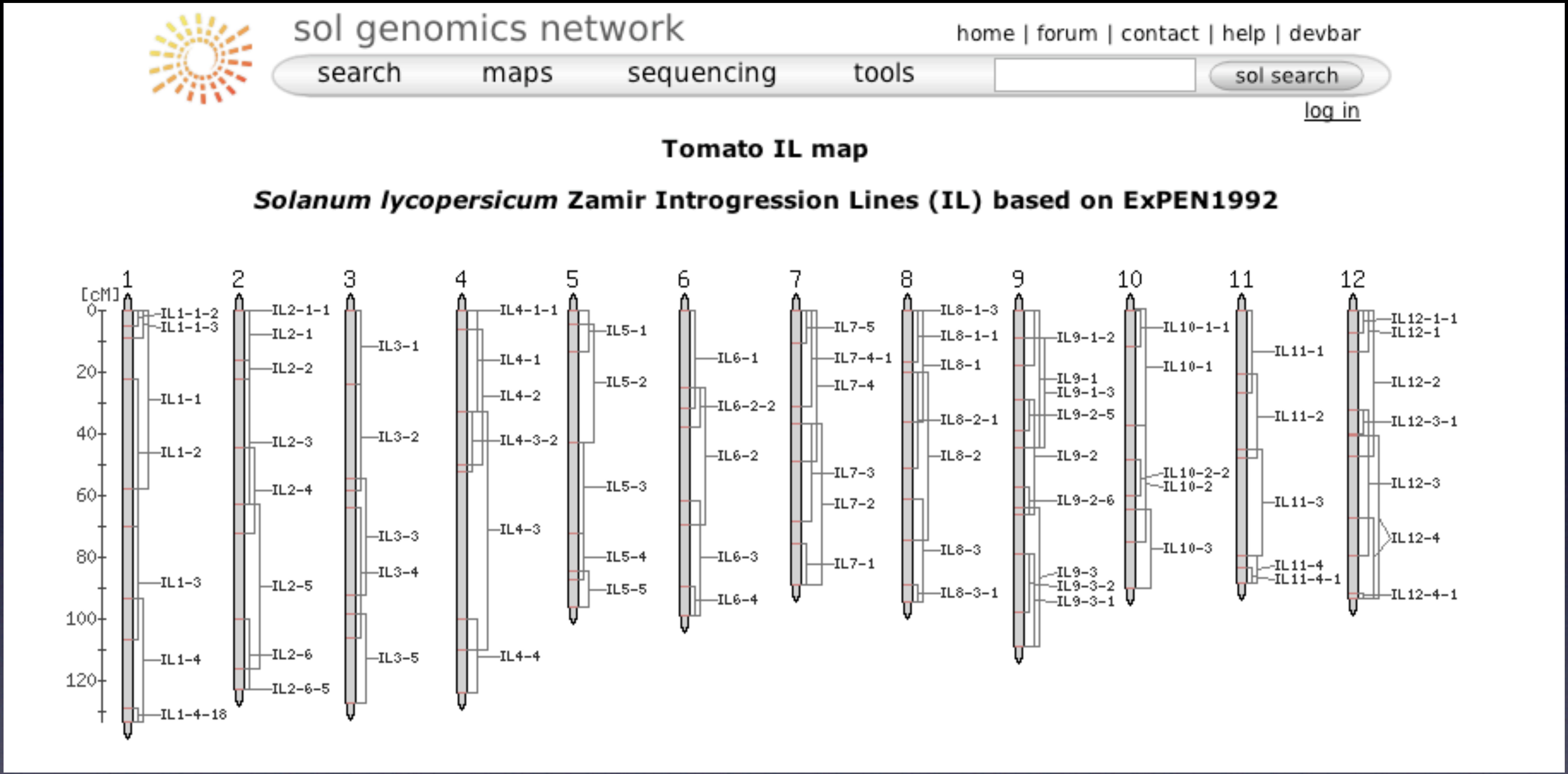
Individual Name:	**IL3-2**
Individual Description:	**M82Xpennellii introgression line 3-2**
Population:	**M82 x S.pennellii IL population**
Organism:	**Tomato**
Uploaded by:	**Dani Zamir**

Associated loci: yellow flesh / phytoene synthase [Associate locus]

Images

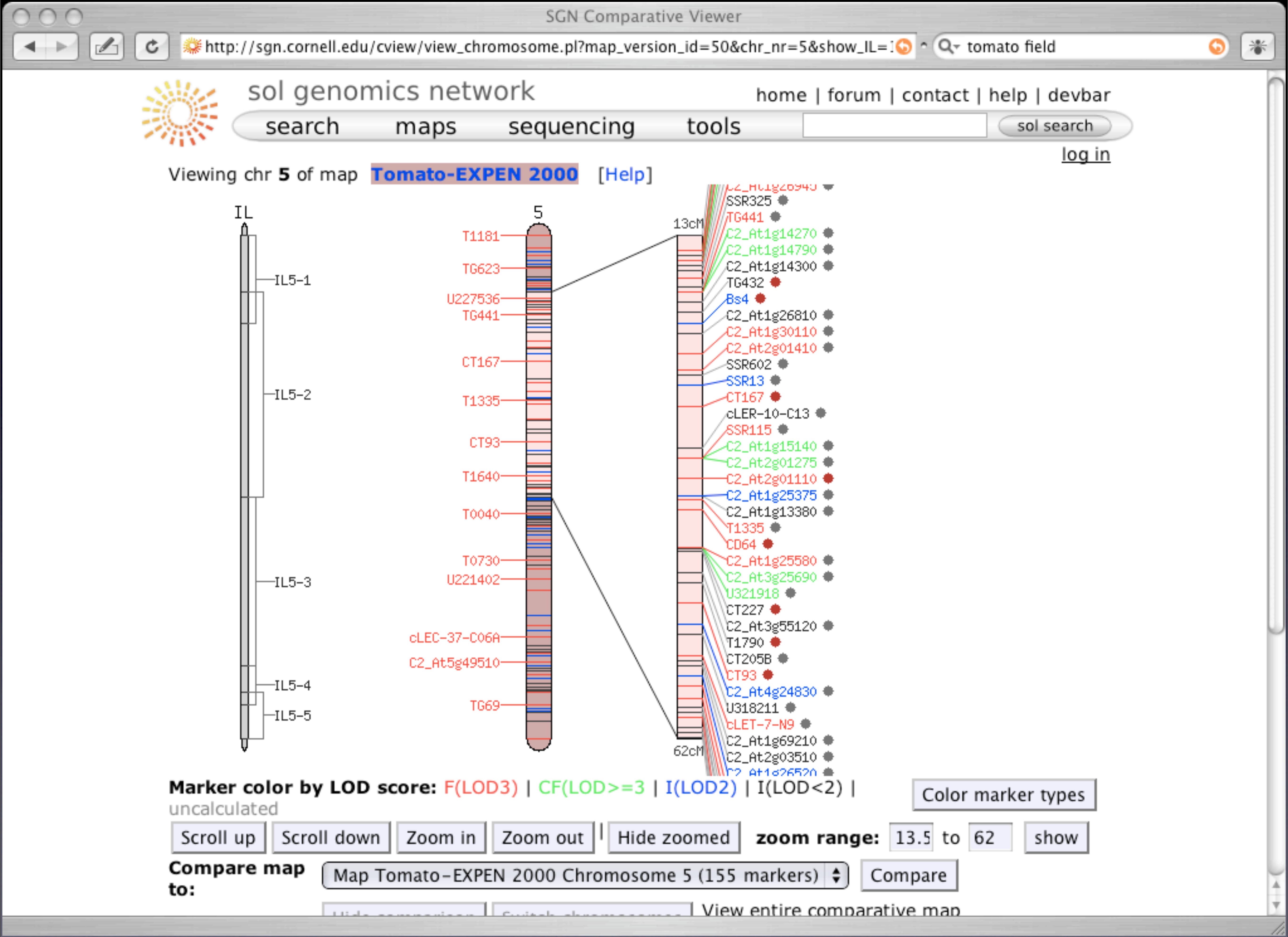
[Add new image]

Mapping data



Tomato IL map

Solanum lycopersicum Zamir Introgression Lines (IL) based on ExPEN1992



SGN Comparative Viewer

http://sgn.cornell.edu/cview/view_chromosome.pl?map_version_id=50&chr_nr=5&show_IL=

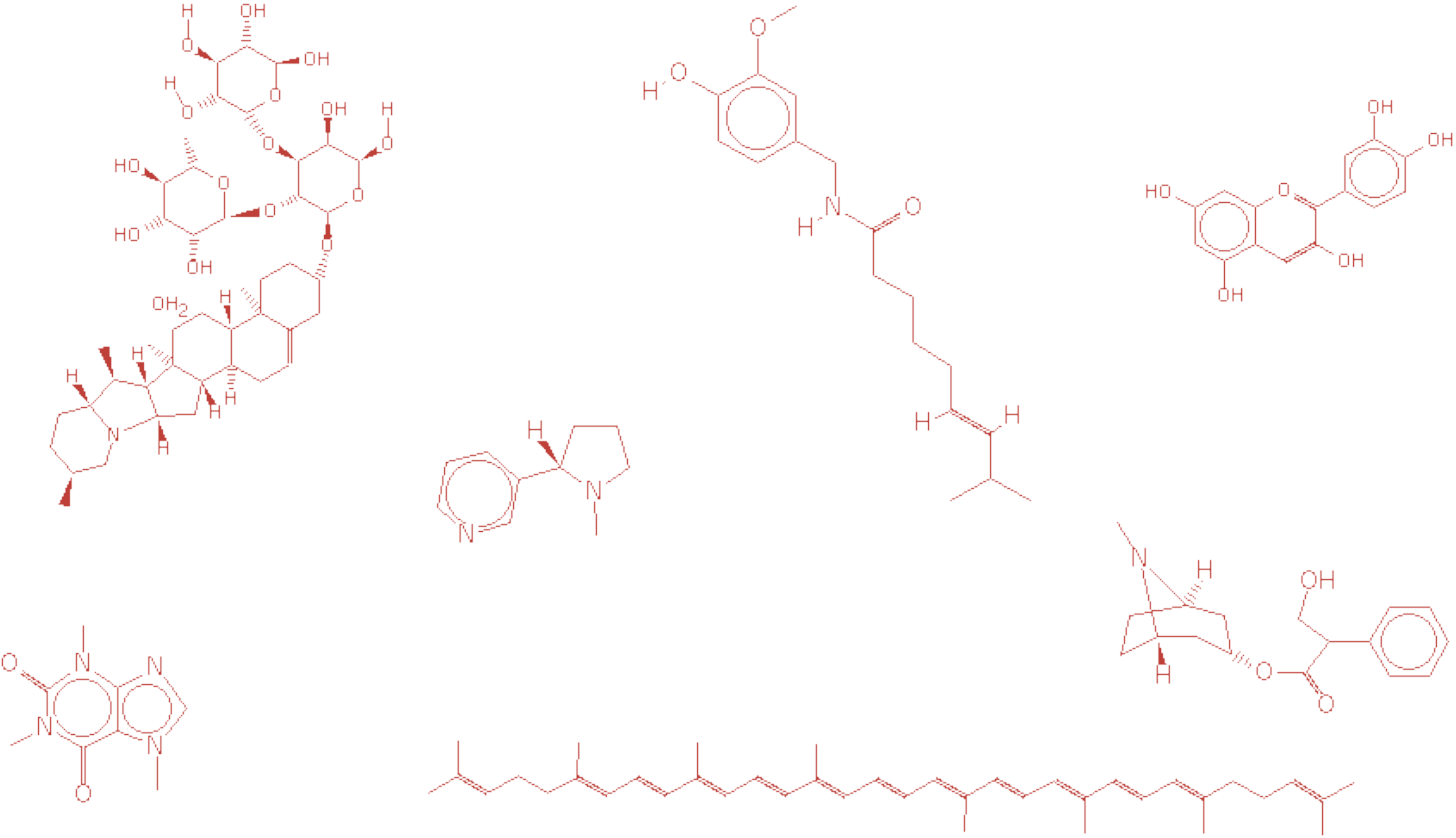
Viewing chr **5** of map **Tomato-EXPEN 2000** [Help]

Marker color by LOD score: F(LOD3) | CF(LOD>=3 | I(LOD2) | I(LOD<2) | uncalculated

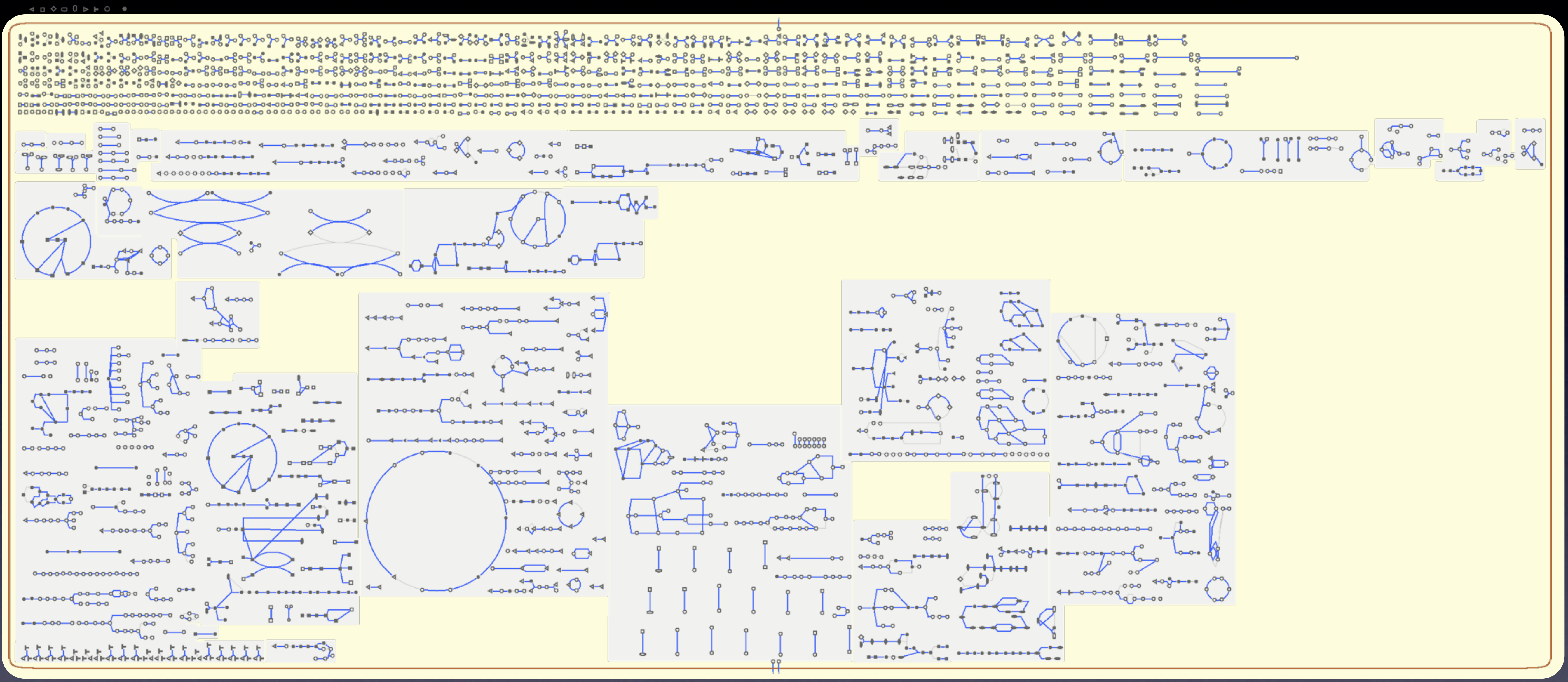
zoom range: 13.5 to 62

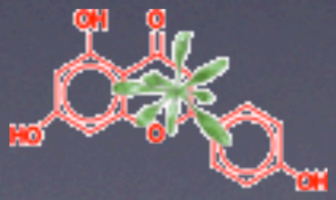
Compare map to: Map Tomato-EXPEN 2000 Chromosome 5 (155 markers)

Solanaceae produce diverse metabolites



Pathway Genome DBs



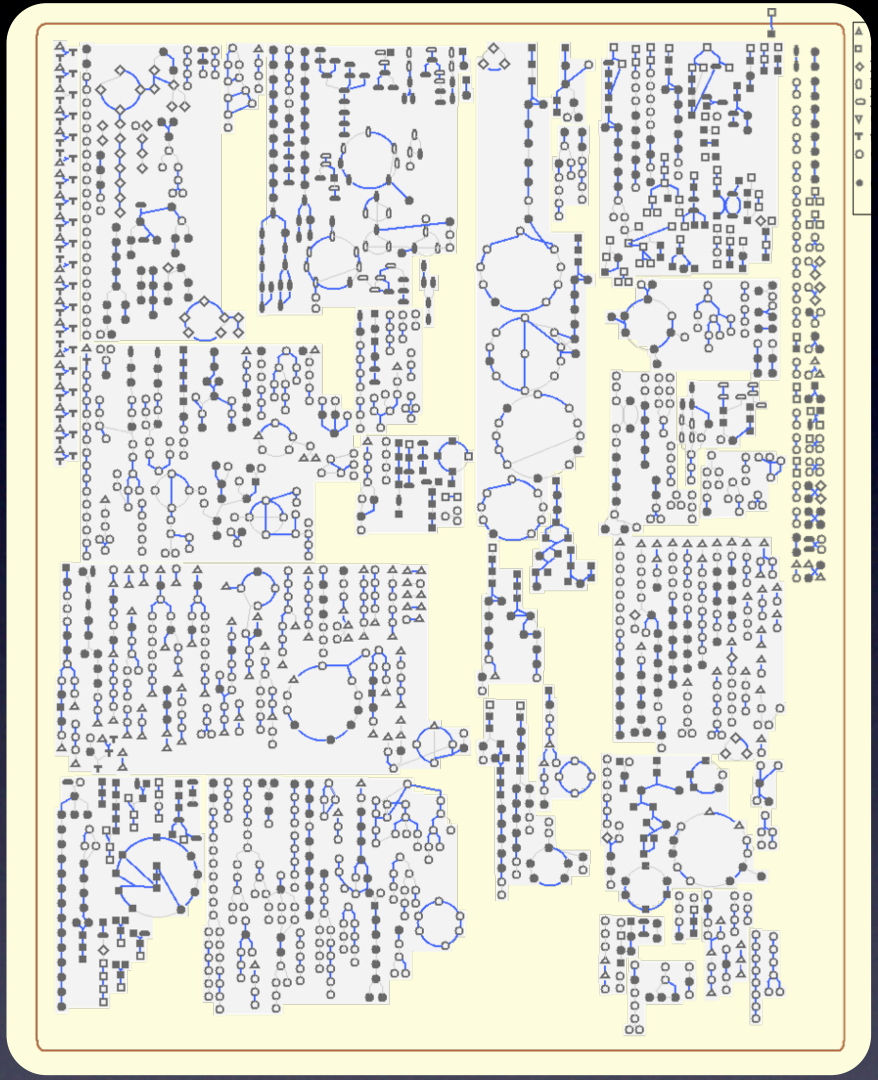


AraCyc

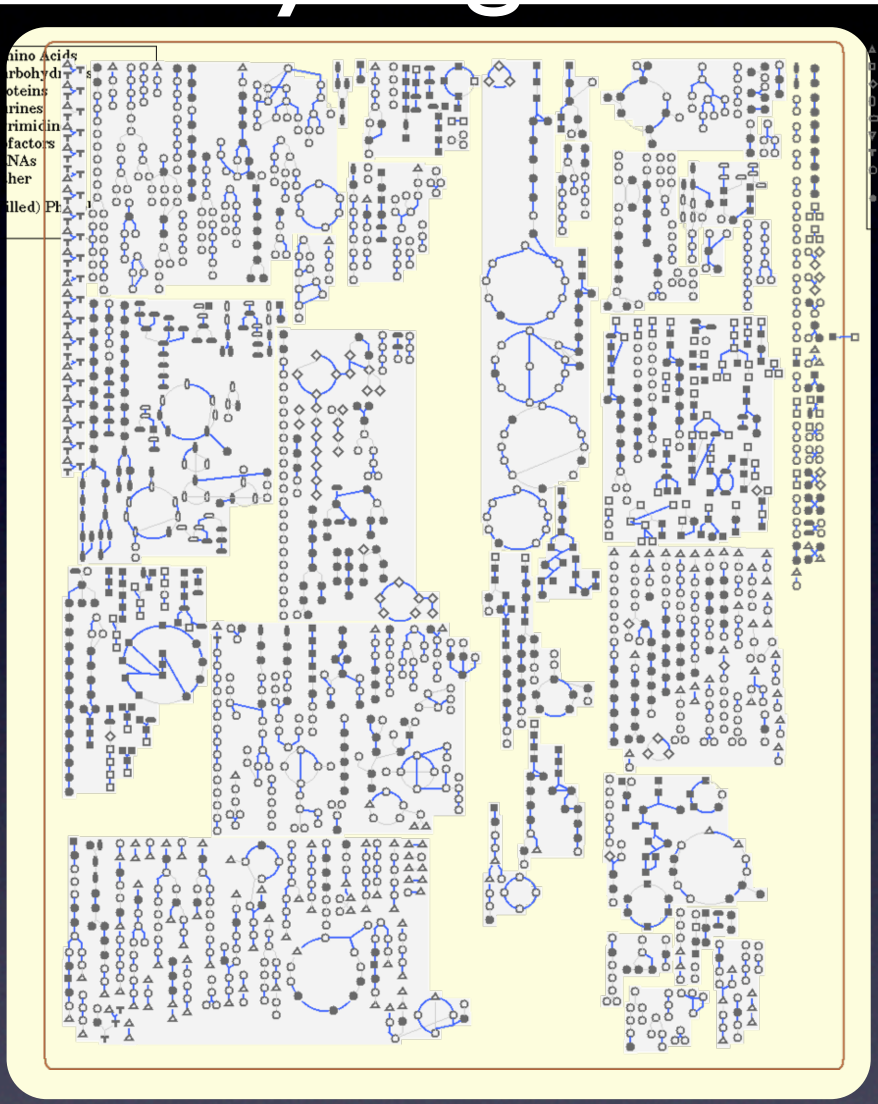
283 pathways
1901 loci
1908 compounds

SolCyc

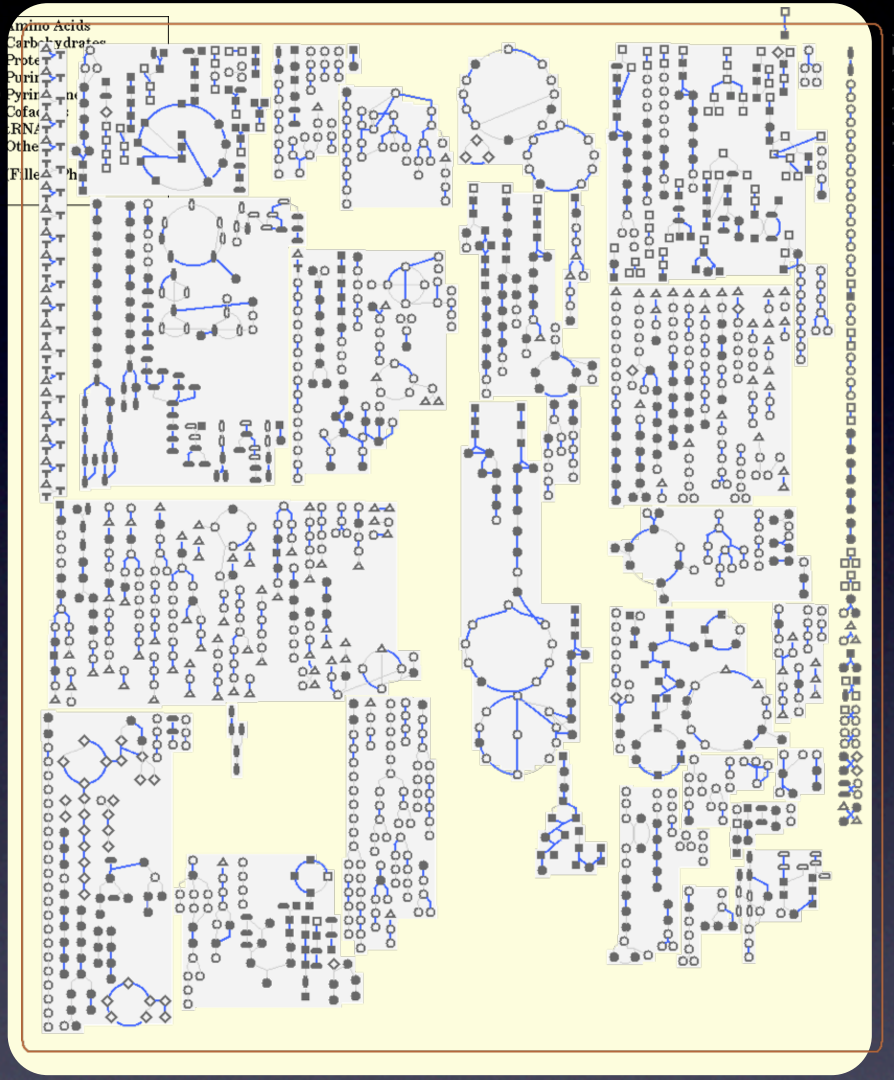
http://solcyc.sgn.cornel.edu



Tomato

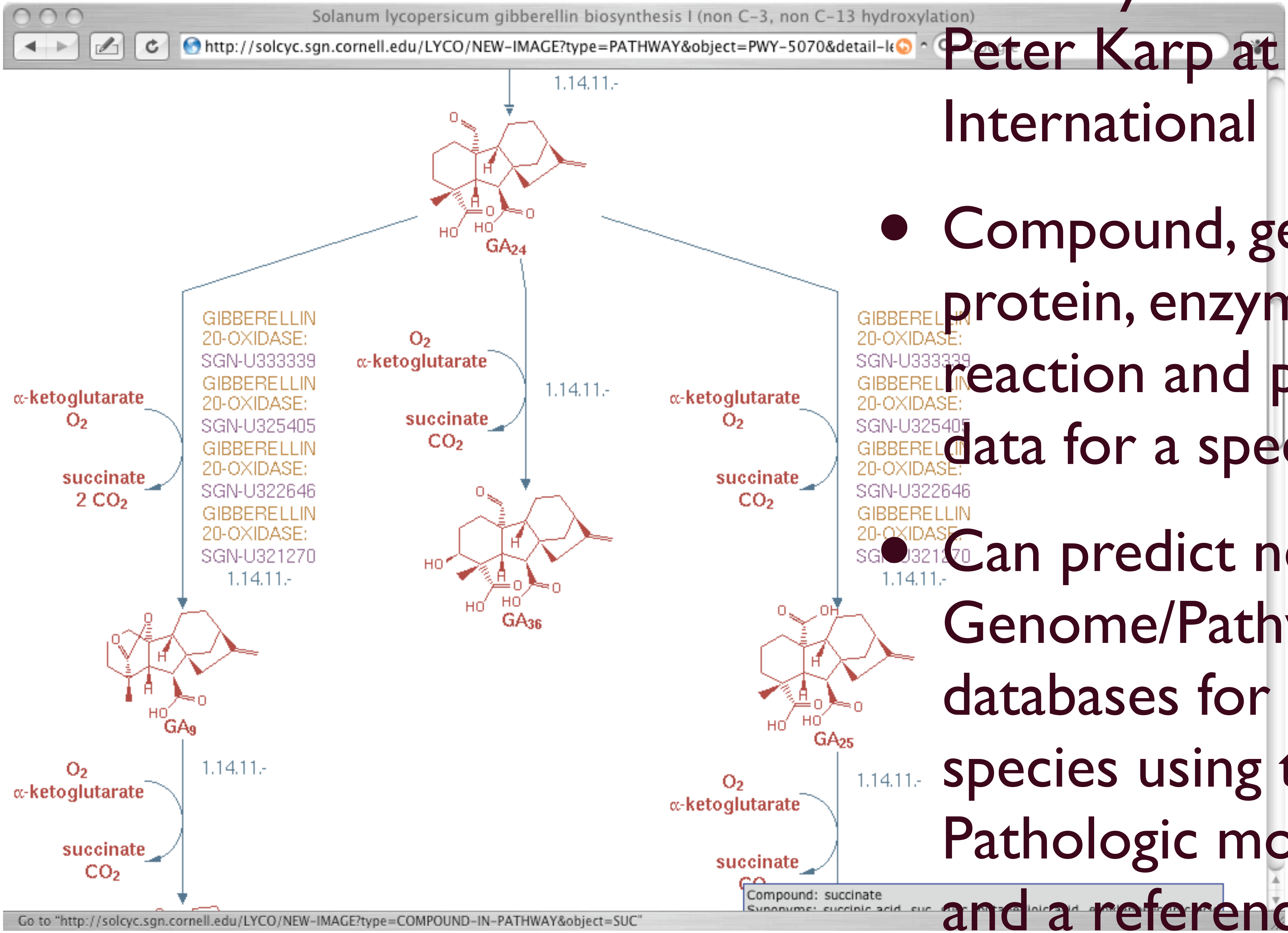


Potato



Pepper

+ Eggplant, Petunia, Coffee



Solanum lycopersicum gibberellin biosynthesis I (non C-3, non C-13 hydroxylation)

http://solcyc.sgn.cornell.edu/LYCO/NEW-IMAGE?type=PATHWAY&object=PWY-5070&detail-le

Go to "http://solcyc.sgn.cornell.edu/LYCO/NEW-IMAGE?type=COMPOUND-IN-PATHWAY&object=SUC"

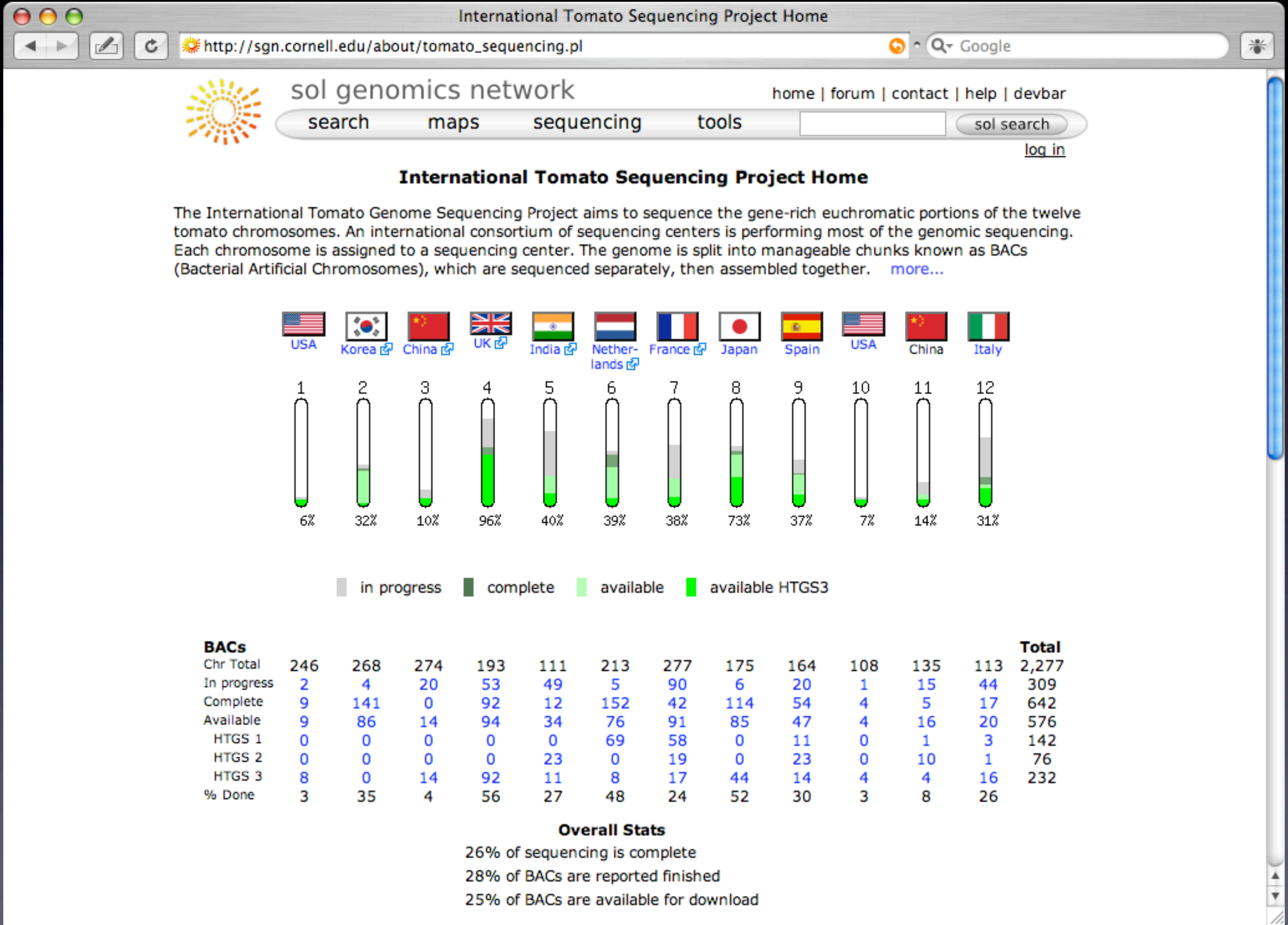
Compound: succinate

- Pathway Tools by Peter Karp at SRI International
- Compound, gene, protein, enzyme, reaction and pathway data for a species
- Can predict new Genome/Pathway databases for a species using the Pathologic module and a reference database (MetaCyc)

MetaCyc

- Genome Pathway Database of all known pathways
- Focus on microorganisms and plants
- SRI & SGN grant to annotate pathways
- SGN Focus on plants (mostly Asterid) pathways

Tomato Genome Sequencing



International Tomato Sequencing Project Home

The International Tomato Genome Sequencing Project aims to sequence the gene-rich euchromatic portions of the twelve tomato chromosomes. An international consortium of sequencing centers is performing most of the genomic sequencing. Each chromosome is assigned to a sequencing center. The genome is split into manageable chunks known as BACs (Bacterial Artificial Chromosomes), which are sequenced separately, then assembled together. more...

▨ in progress ■ complete ▨ available ■ available HTGS3

BACs													**Total**
Chr Total	246	268	274	193	111	213	277	175	164	108	135	113	2,277
In progress	2	4	20	53	49	5	90	6	20	1	15	44	309
Complete	9	141	0	92	12	152	42	114	54	4	5	17	642
Available	9	86	14	94	34	76	91	85	47	4	16	20	576
HTGS 1	0	0	0	0	0	69	58	0	11	0	1	3	142
HTGS 2	0	0	0	0	23	0	19	0	23	0	10	1	76
HTGS 3	8	0	14	92	11	8	17	44	14	4	4	16	232
% Done	3	35	4	56	27	48	24	52	30	3	8	26	

Overall Stats

26% of sequencing is complete
28% of BACs are reported finished
25% of BACs are available for download

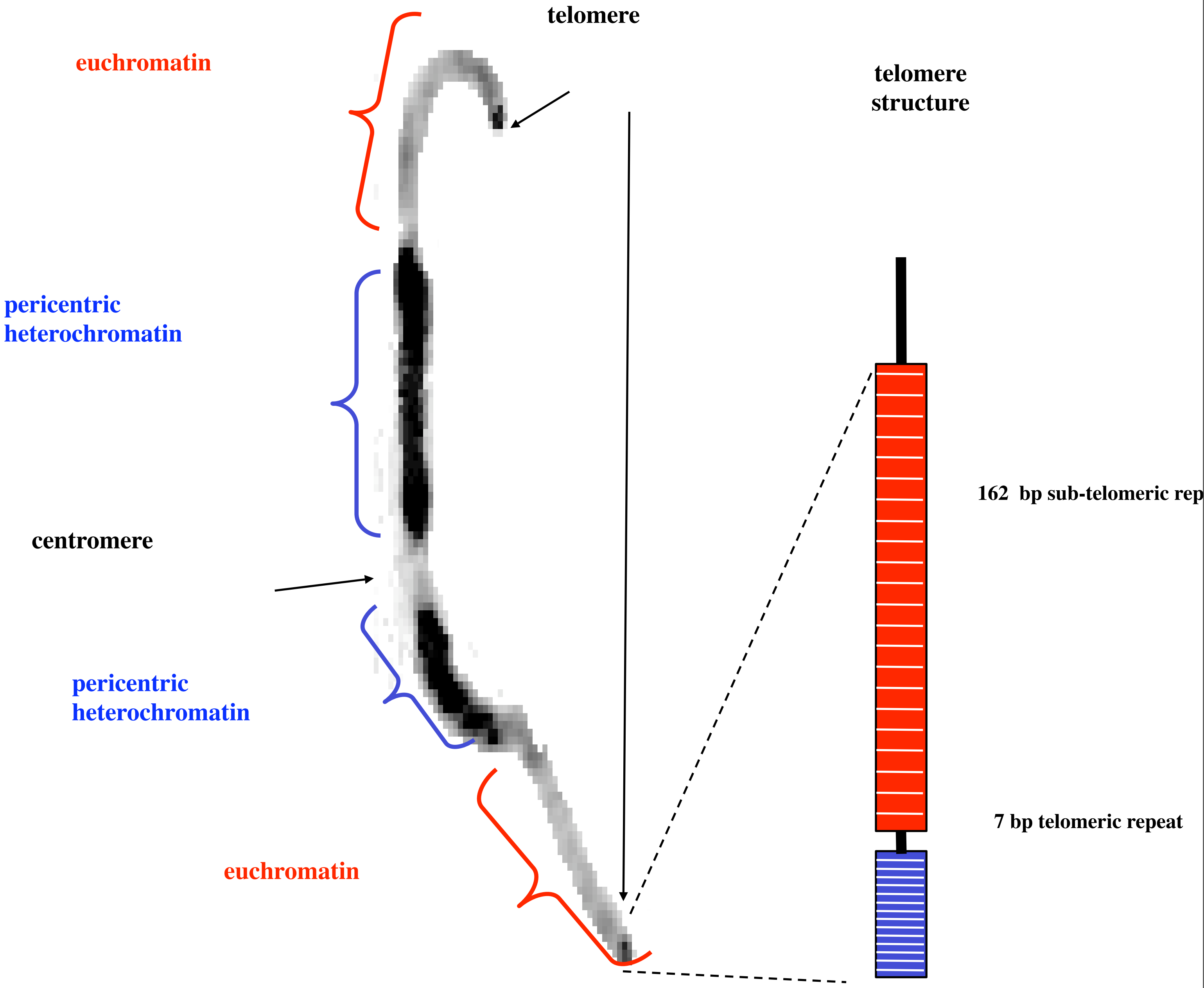
SGN is a central informatics hub for tomato sequencing

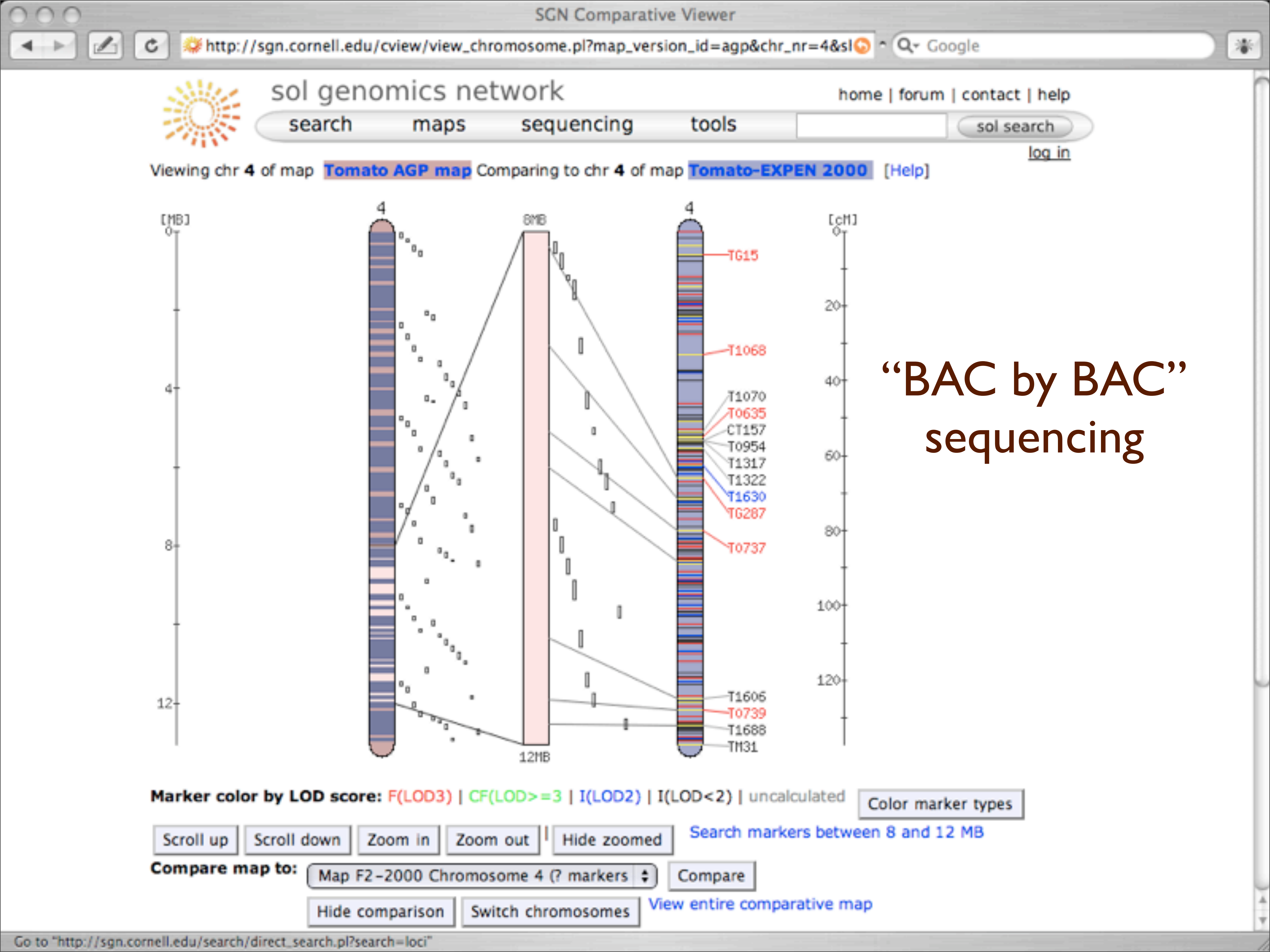
- BAC registry: Status of every BAC in the sequence pipeline
- International partners upload BAC sequences, chromatograms and assembly information
- SGN runs quality checks, annotation pipelines, and publishes data for download and browsing

850MB Genome

240MB euchromatin

estimated >85% of genes





http://sgn.cornell.edu/cview/view_chromosome.pl?map_version_id=agp&chr_nr=4&sl

Viewing chr **4** of map **Tomato AGP map** Comparing to chr **4** of map **Tomato-EXPEN 2000** [Help]

Marker color by LOD score: F(LOD3) | CF(LOD>=3) | I(LOD2) | I(LOD<2) | uncalculated

Compare map to: Map F2-2000 Chromosome 4 (? markers

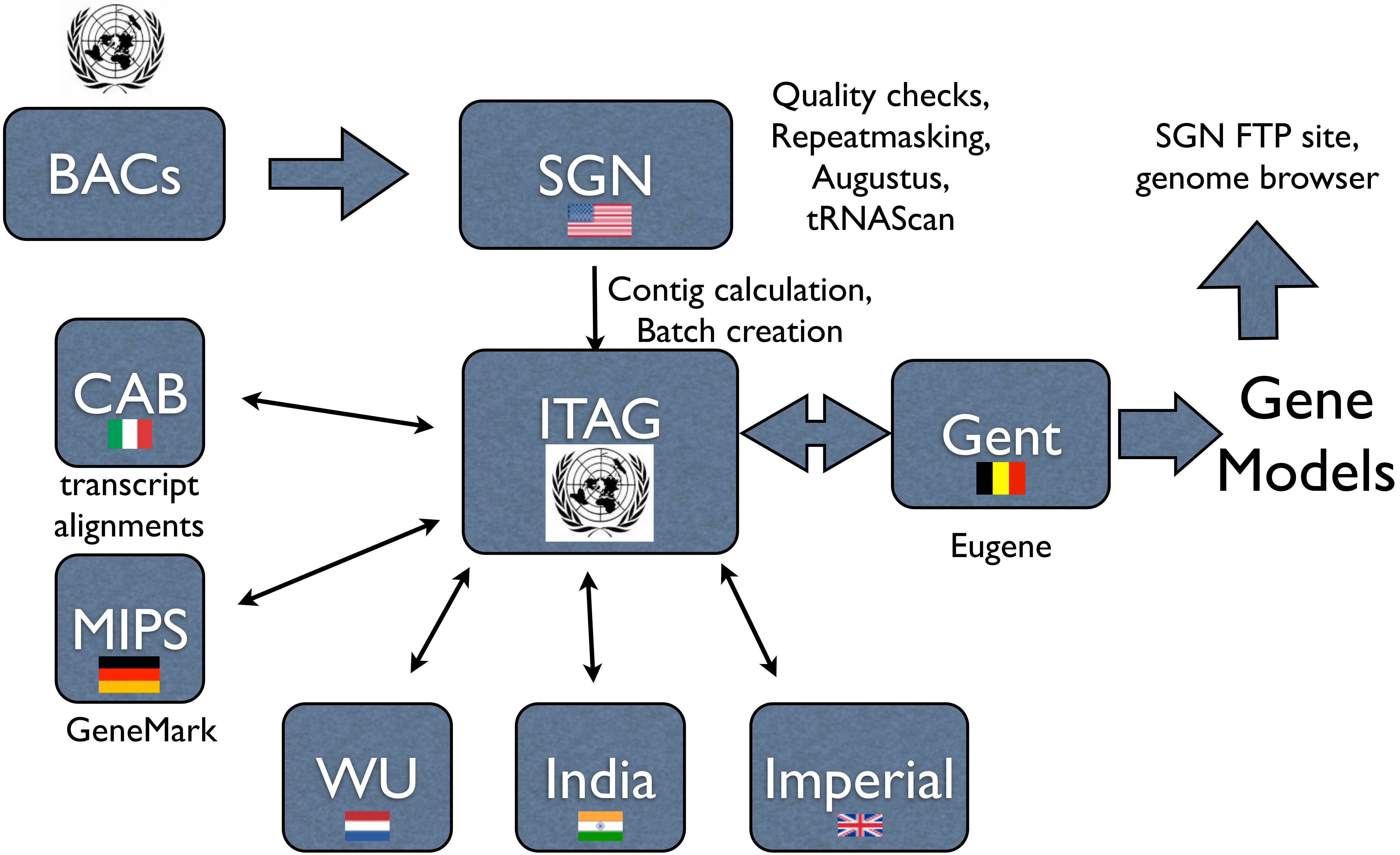
Go to "http://sgn.cornell.edu/search/direct_search.pl?search=loci"

“BAC by BAC” sequencing

ITAG

- International Tomato Annotation Group
- High quality automated annotation of the tomato genome
- distributed annotation pipeline among members (USA, GB, Italy, Germany, India, Belgium, etc.)

Structural Annotation Pipeline



BACs

SGN

Quality checks,
Repeatmasking,
Augustus,
tRNAScan

SGN FTP site,
genome browser

Contig calculation,
Batch creation

CAB

transcript
alignments

ITAG

Gent

Gene
Models

Eugene

MIPS

GeneMark

WU

GlimmerHMM

India

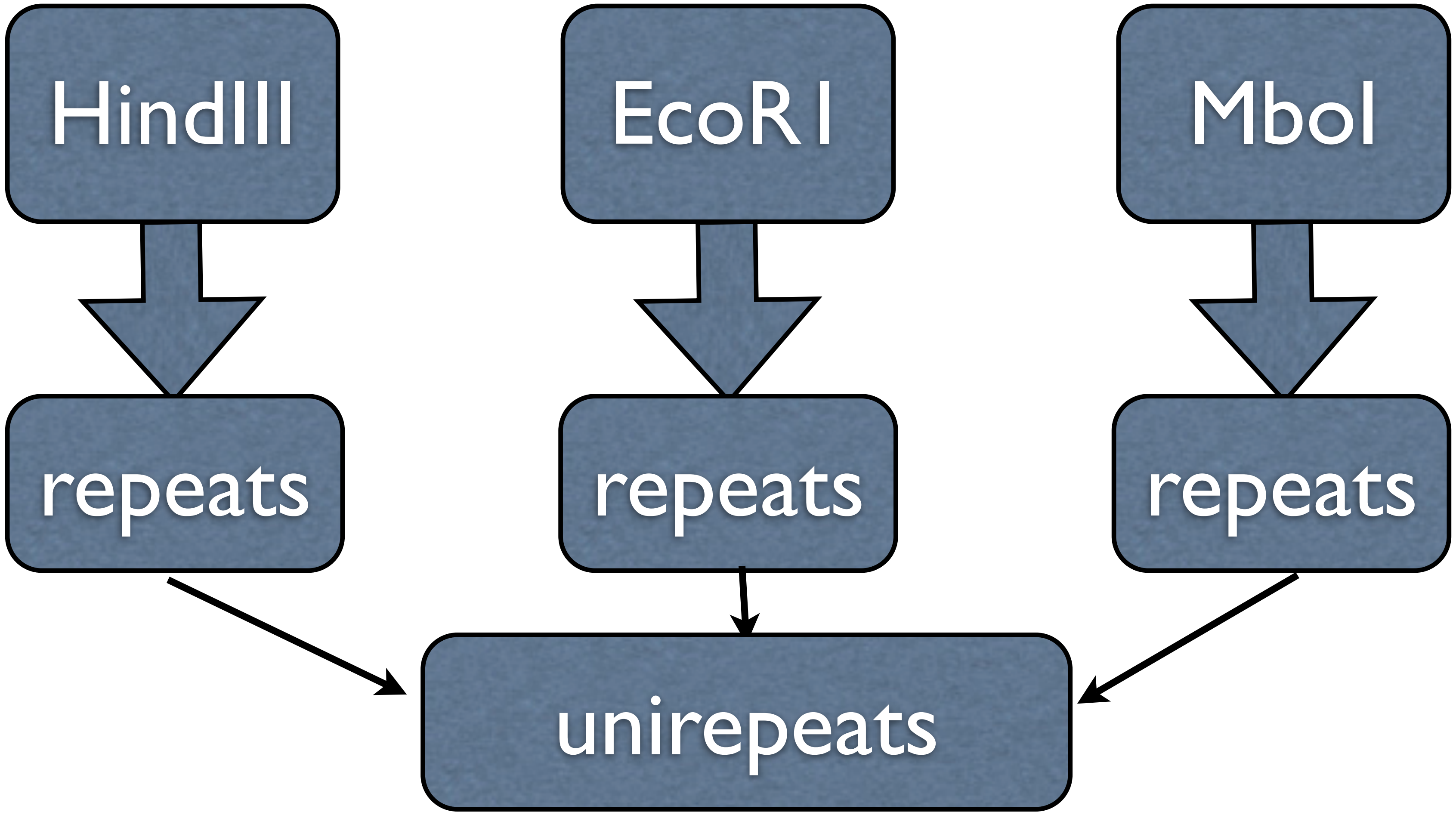
BLAST

Imperial

infernal

Repeat Annotation

- Generated a de-novo repeat dataset using RepeatScout
- Using 340,000 BAC end sequences from 3 BAC libraries


HindIII

EcoR I

MboI

repeats

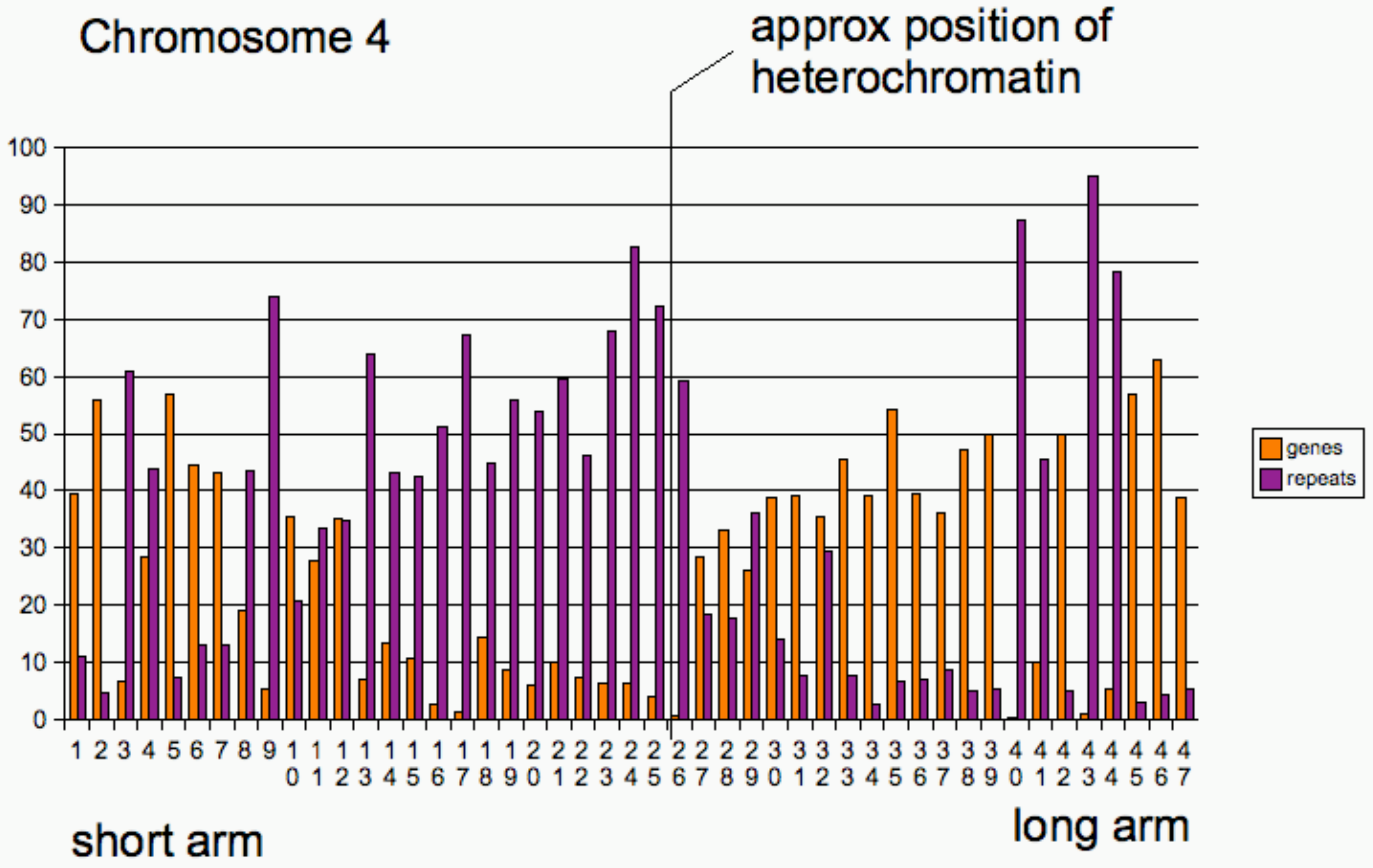
repeats

repeats

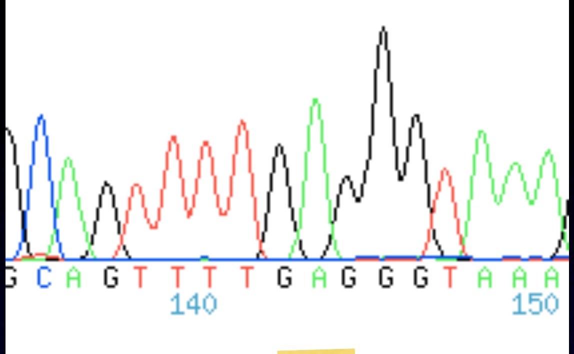
unirepeats

GBrowse on SGN
http://sgn.cornell.edu/gbrowse/gbrowse/tomato_bacs/?ref=C01HBa0088L02.2&start=1&end
Google
Overview
Overview of C01HBa0088L02.2
0k
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
110k
Details
0k
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
110k
GenomeThreader vs SGN tomato and potato ESTs predicted gene locations
GenomeThreader vs SGN tomato and potato ESTs supporting alignments
SGN markers
T1893
T1306
TG159-F
FGENESH (Tomato) predictions
RepeatMasker repeat alignments
tRNAscan-SE predicted tRNAs
Tomato BAC Ends
LE_HBa0088L02_T7_56211
SL_EcoRI0040H03_SP6_242511
LE_HBa0090A07_SP6_25740
LE_HBa0009A08_SP6_3854
SL_MboI0144M05_SP6_365372
LE_HBa0219
LE_HBa0038H14_T7_92907
SL_EcoRI0115E02_SP6_323126
LE_HB
LE_HBa0036I22_T7_142112
LE_HBa0015G05_SP6_129693
LE_HBa0219E11_T7_395854

Chromosome 4

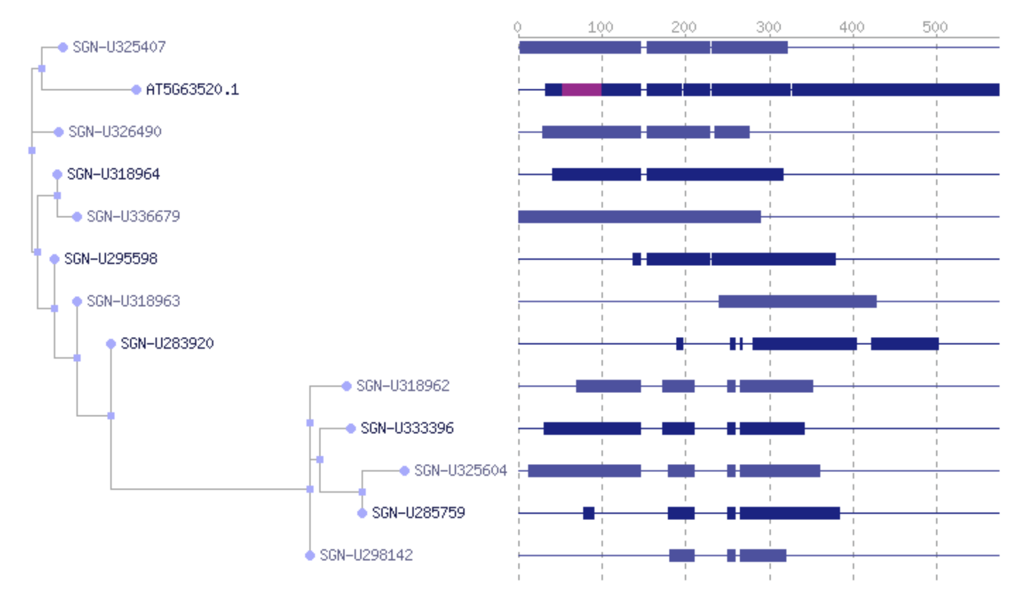


COD approach to Sequence



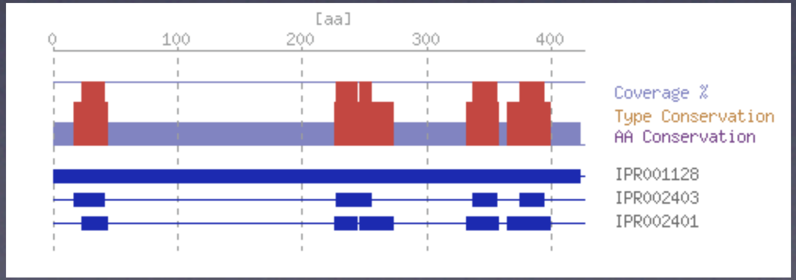
Tomato, Potato, Pepper, Eggplant, Tobacco, Coffee, Snapdragon ESTs (900,000 sequences)

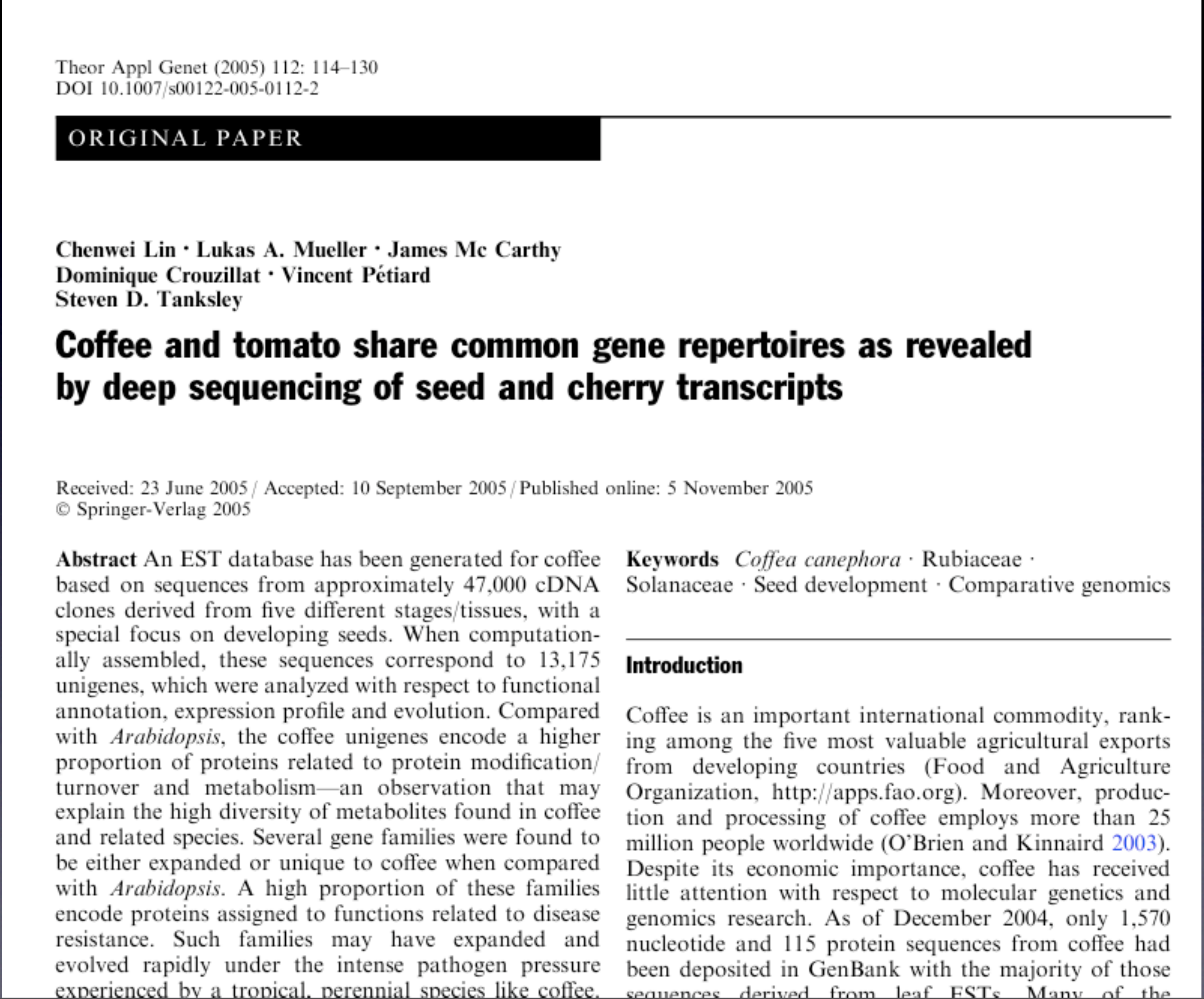
Unigenes



Gene families

Annotation, Domains, Protein Prediction, SNPs





Theor Appl Genet (2005) 112: 114–130
DOI 10.1007/s00122-005-0112-2

ORIGINAL PAPER

Chenwei Lin · Lukas A. Mueller · James Mc Carthy
Dominique Crouzillat · Vincent Pétiard
Steven D. Tanksley

Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts

Received: 23 June 2005 / Accepted: 10 September 2005 / Published online: 5 November 2005

Abstract An EST database has been generated for coffee based on sequences from approximately 47,000 cDNA clones derived from five different stages/tissues, with a special focus on developing seeds. When computationally assembled, these sequences correspond to 13,175 unigenes, which were analyzed with respect to functional annotation, expression profile and evolution. Compared with *Arabidopsis*, the coffee unigenes encode a higher proportion of proteins related to protein modification/turnover and metabolism—an observation that may explain the high diversity of metabolites found in coffee and related species. Several gene families were found to be either expanded or unique to coffee when compared with *Arabidopsis*. A high proportion of these families encode proteins assigned to functions related to disease resistance. Such families may have expanded and evolved rapidly under the intense pathogen pressure experienced by a tropical, perennial species like coffee.

Keywords *Coffea canephora* · Rubiaceae · Solanaceae · Seed development · Comparative genomics

Introduction

Coffee is an important international commodity, ranking among the five most valuable agricultural exports from developing countries (Food and Agriculture Organization, http://apps.fao.org). Moreover, production and processing of coffee employs more than 25 million people worldwide (O'Brien and Kinnaird 2003). Despite its economic importance, coffee has received little attention with respect to molecular genetics and genomics research. As of December 2004, only 1,570 nucleotide and 115 protein sequences from coffee had been deposited in GenBank with the majority of those sequences derived from leaf ESTs. Many of the

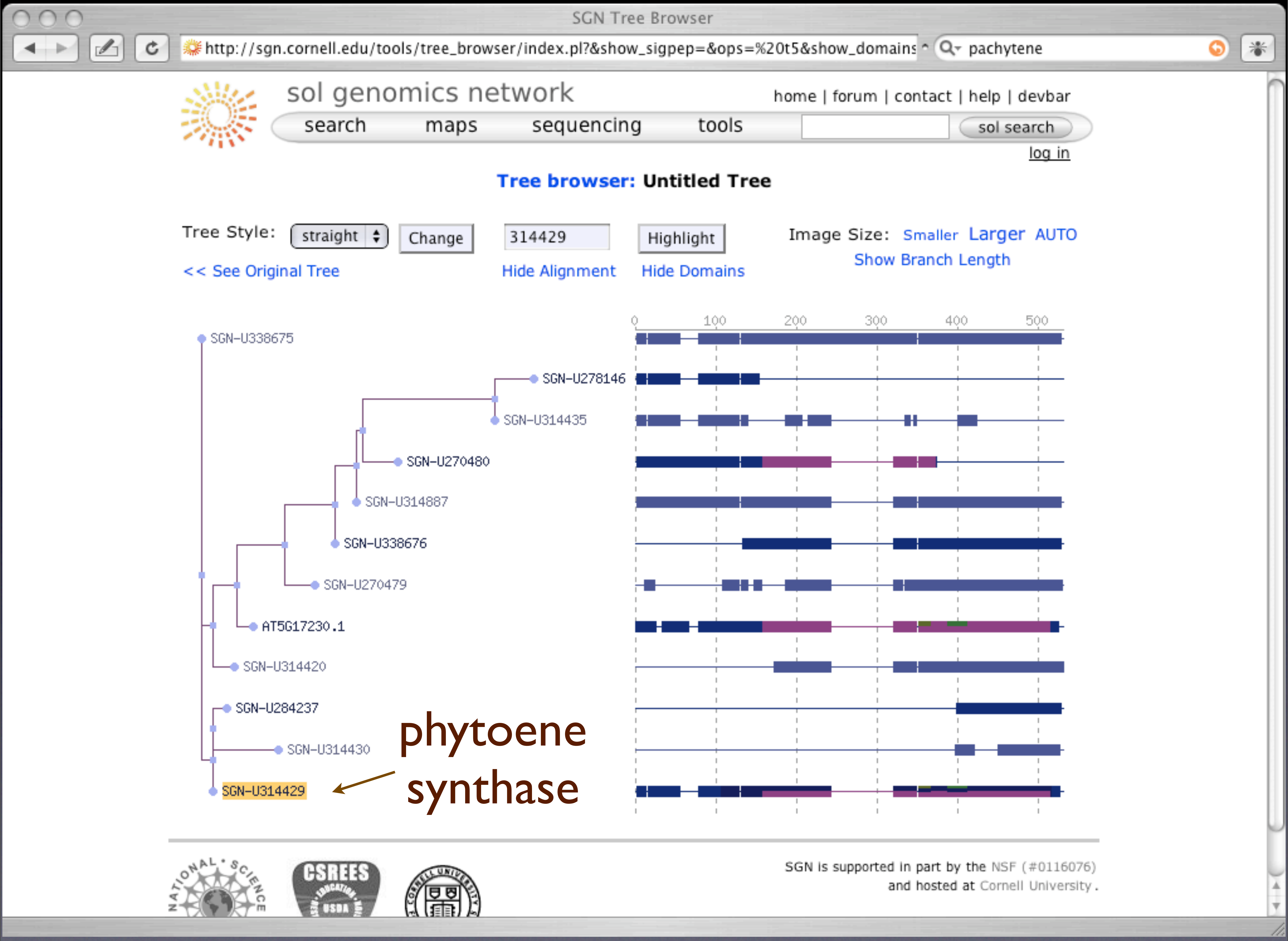
Gene Families

- Calculated using MCL with I-value of 1.1, 2 and 5
- Alignments (MUSCLE)
- Trees generated (modeltest)

- Determination of orthologs and paralogs
- Gene family expansion and contraction in evolution
- Determination of sites under selection

Gene Family Stats

i-value	families	total members
1.1	5169	66814
2	10,519	61751
5	11,851	55995



SGN Tree Browser

http://sgn.cornell.edu/tools/tree_browser/index.pl?&show_sigpep=&ops=%20t5&show_domains

pachytene

sol genomics network

Tree browser: **Untitled Tree**

Tree Style: straight

314429

Image Size: Smaller Larger AUTO

SGN is supported in part by the NSF (#0116076) and hosted at Cornell University.

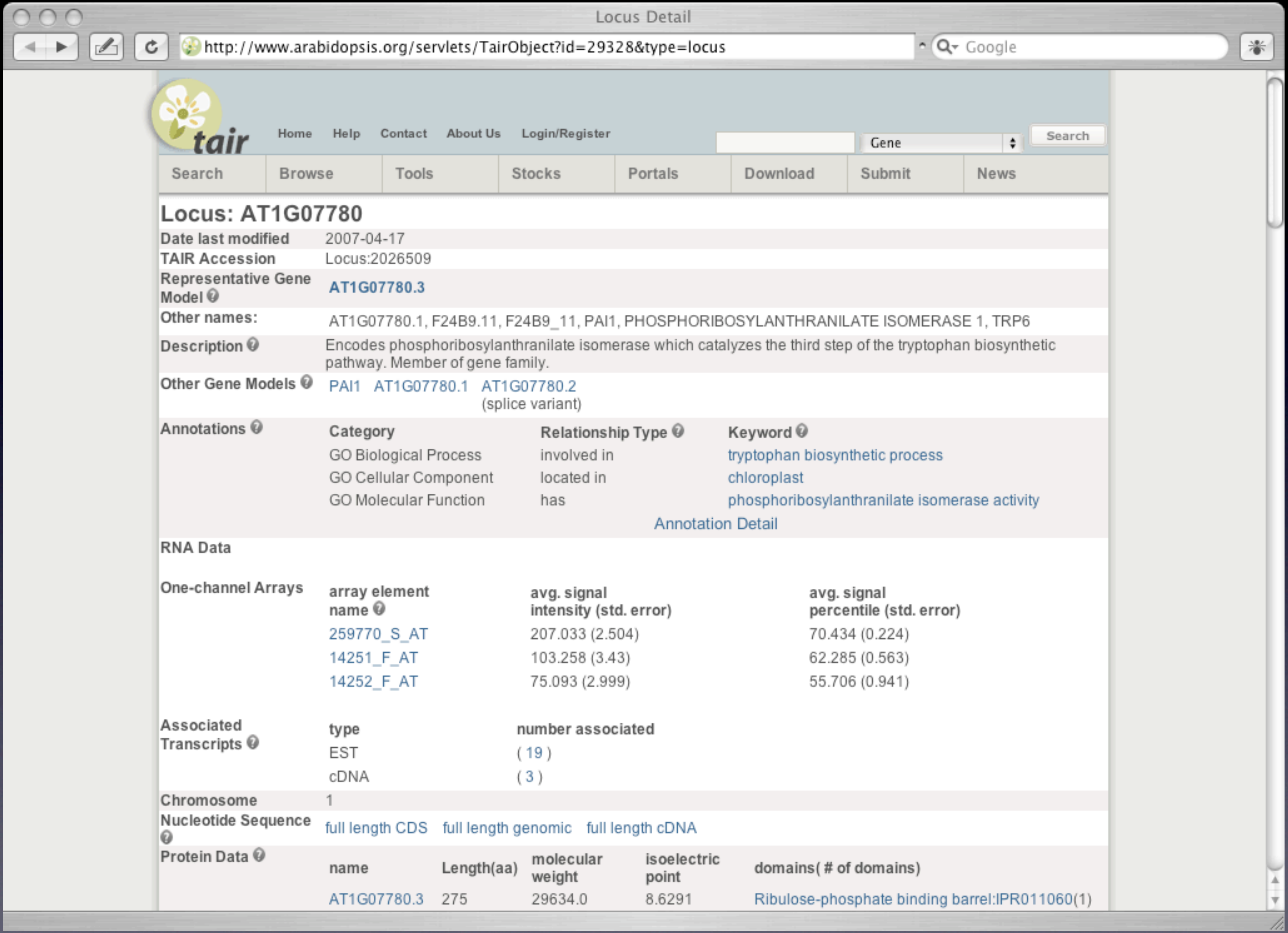
phytoene synthase

Analyses

- Identified unique and expanded gene families in *Coffea robusta*, incl. putative disease resistance genes (Lin et al).
- Secretom Project: Mapped signal peptides to Arabidopsis families, deduced events in signal peptide evolution (in prep).
- Calculated orthologs from all trees: 3375 orthologous relationships
- Identified 114 Solanaceae families not found in Arabidopsis

Outline

- Why Solanaceae? - Unique Solanaceae Tools
- From phenotypes to genotypes
- Biochemical Pathways
- Tomato Genome Sequencing Project
- Solanaceae Gene Families
- Locus community annotation
- Future directions



Locus: AT1G07780

Date last modified	2007-04-17
TAIR Accession	Locus:2026509
Representative Gene Model	**AT1G07780.3**
Other names:	AT1G07780.1, F24B9.11, F24B9_11, PAI1, PHOSPHORIBOSYLANTHRANILATE ISOMERASE 1, TRP6
Description	Encodes phosphoribosylanthranilate isomerase which catalyzes the third step of the tryptophan biosynthetic pathway. Member of gene family.
Other Gene Models	PAI1 AT1G07780.1 AT1G07780.2 (splice variant)

Annotations

Category	Relationship Type	Keyword
GO Biological Process	involved in	tryptophan biosynthetic process
GO Cellular Component	located in	chloroplast
GO Molecular Function	has	phosphoribosylanthranilate isomerase activity

Annotation Detail

RNA Data

One-channel Arrays

array element name	avg. signal intensity (std. error)	avg. signal percentile (std. error)
259770_S_AT	207.033 (2.504)	70.434 (0.224)
14251_F_AT	103.258 (3.43)	62.285 (0.563)
14252_F_AT	75.093 (2.999)	55.706 (0.941)

Associated Transcripts

type	number associated
EST	(19)
cDNA	(3)

Chromosome	1
Nucleotide Sequence	full length CDS full length genomic full length cDNA

Protein Data

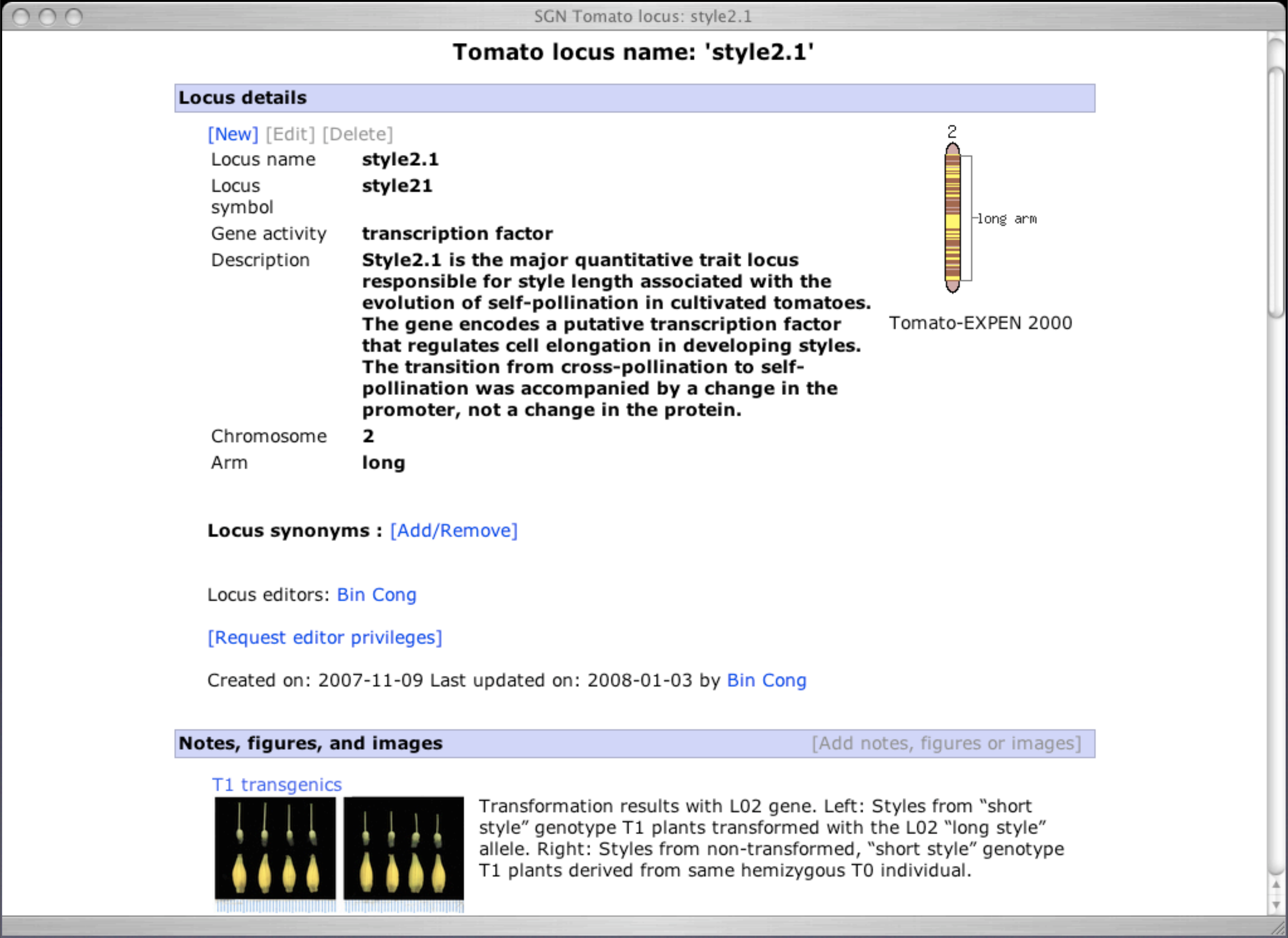
name	Length(aa)	molecular weight	isoelectric point	domains(# of domains)
AT1G07780.3	275	29634.0	8.6291	Ribulose-phosphate binding barrel:IPR011060(1)

SGN Locus database

- Pulls phenotype and sequence information together
- “facebook” for loci
- Updated through simple web interfaces by Locus Editors from the research community
- Prevents propagation of false annotations
- Rich datatypes: Mapping, sequence, ontology, literature, mutants, alleles, etc.

Loci by species

species	*# of loci*
Tomato	1628
Potato	1028
Pepper	373
Coffee	112
Eggplant	48



Tomato locus name: 'style2.1'

Locus details

[New] [Edit] [Delete]

Locus name	**style2.1**
Locus symbol	**style21**
Gene activity	**transcription factor**
Description	**Style2.1 is the major quantitative trait locus responsible for style length associated with the evolution of self-pollination in cultivated tomatoes. The gene encodes a putative transcription factor that regulates cell elongation in developing styles. The transition from cross-pollination to self-pollination was accompanied by a change in the promoter, not a change in the protein.**
Chromosome	**2**
Arm	**long**

Tomato-EXPEN 2000

Locus synonyms : [Add/Remove]

Locus editors: Bin Cong

[Request editor privileges]

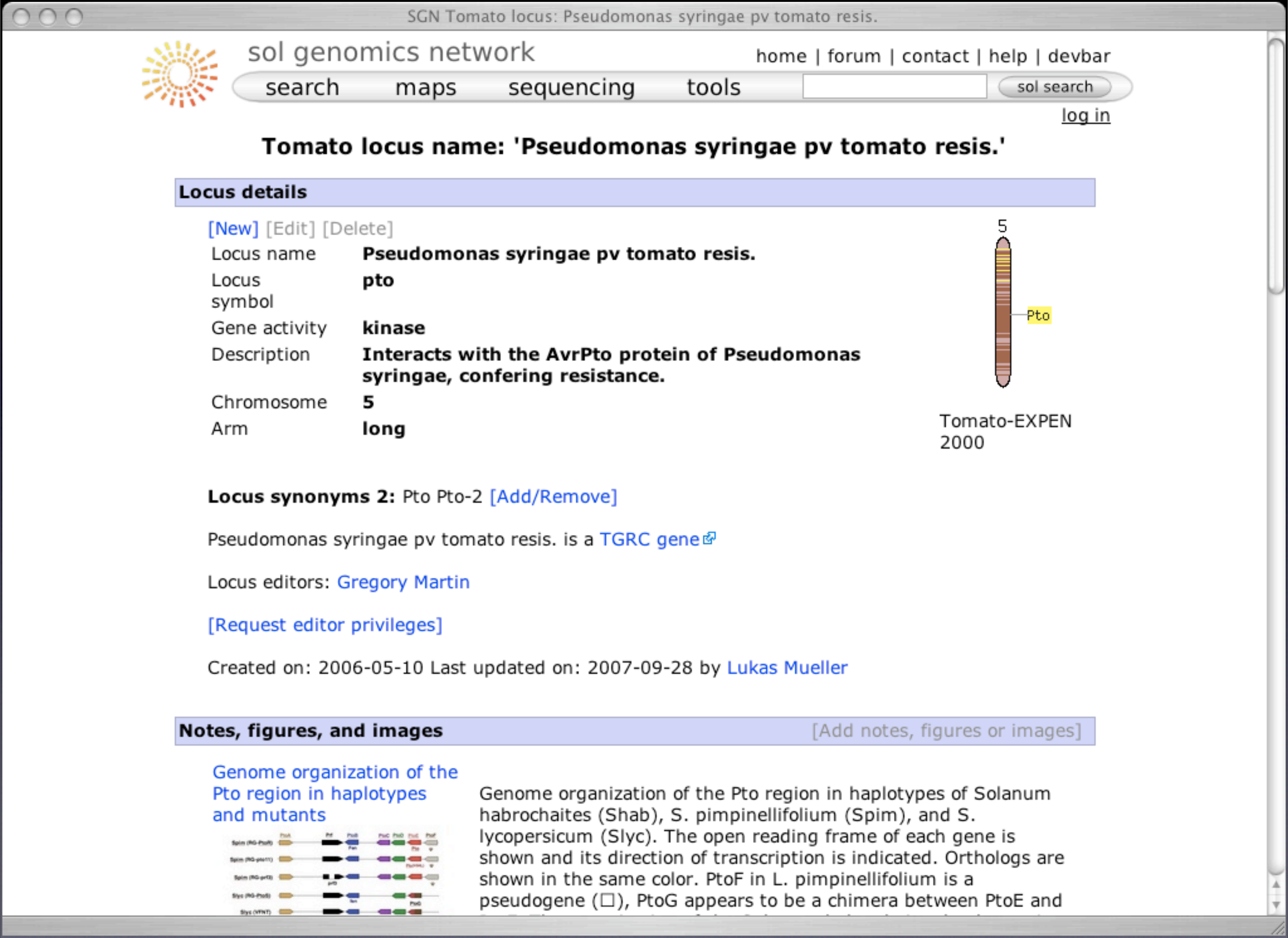
Created on: 2007-11-09 Last updated on: 2008-01-03 by Bin Cong

Notes, figures, and images

[Add notes, figures or images]

T1 transgenics

Transformation results with L02 gene. Left: Styles from “short style” genotype T1 plants transformed with the L02 “long style” allele. Right: Styles from non-transformed, “short style” genotype T1 plants derived from same hemizygous T0 individual.



Tomato locus name: 'Pseudomonas syringae pv tomato resis.'

Locus details

[New] [Edit] [Delete]

Locus name	**Pseudomonas syringae pv tomato resis.**
Locus symbol	**pto**
Gene activity	**kinase**
Description	**Interacts with the AvrPto protein of Pseudomonas syringae, confering resistance.**
Chromosome	**5**
Arm	**long**

Locus synonyms 2: Pto Pto-2 [Add/Remove]

Pseudomonas syringae pv tomato resis. is a TGRC gene

Locus editors: Gregory Martin

[Request editor privileges]

Created on: 2006-05-10 Last updated on: 2007-09-28 by Lukas Mueller

Notes, figures, and images

[Add notes, figures or images]

Genome organization of the Pto region in haplotypes and mutants

Genome organization of the Pto region in haplotypes of Solanum habrochaites (Shab), S. pimpinellifolium (Spim), and S. lycopersicum (Slyc). The open reading frame of each gene is shown and its direction of transcription is indicated. Orthologs are shown in the same color. PtoF in L. pimpinellifolium is a pseudogene (☐), PtoG appears to be a chimera between PtoE and

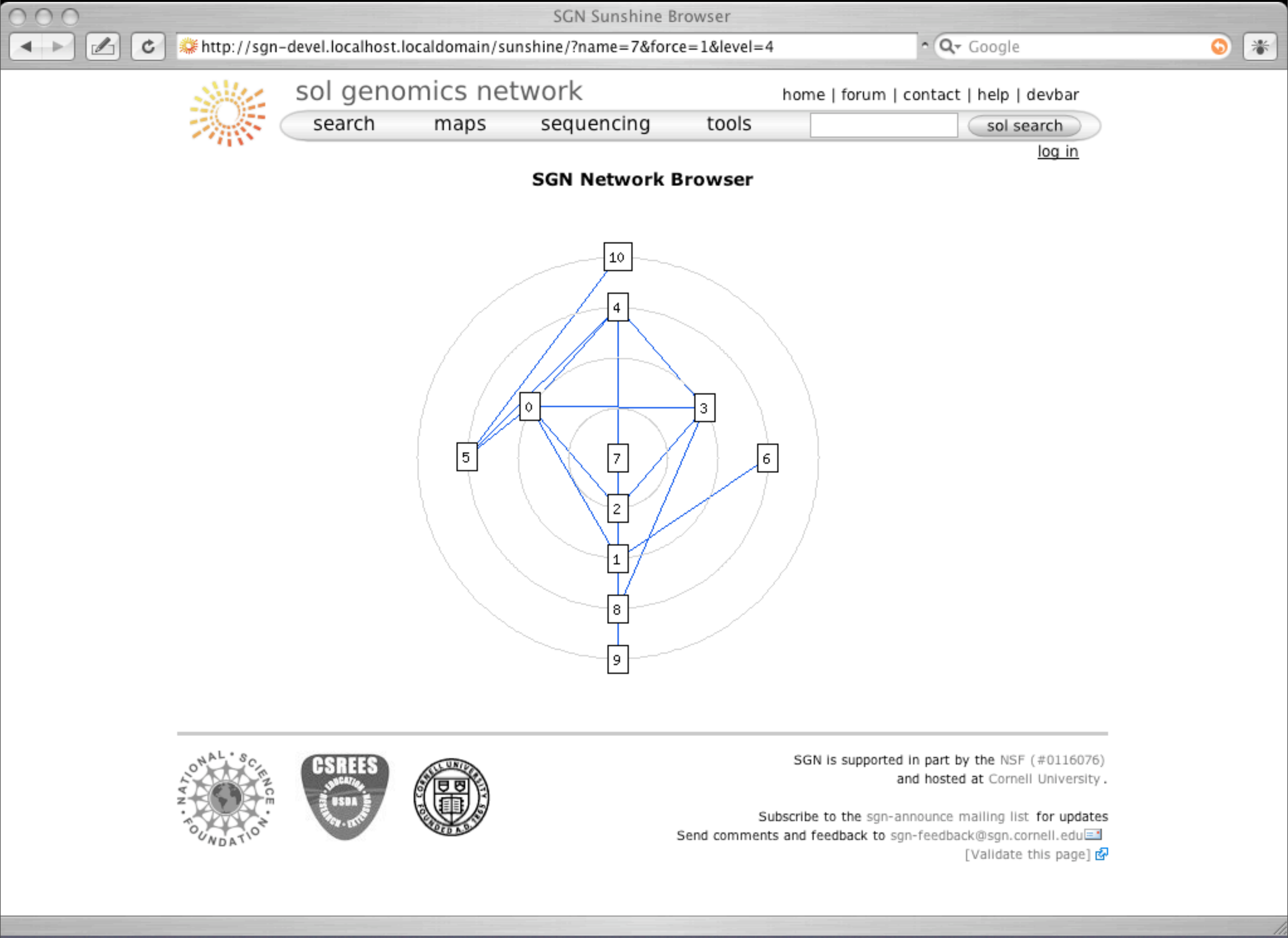
Locus Stats

- >3000 loci
- 32 locus editors
- 82 loci annotated
- 226 manual annotations (10294 automatic)
- Users can request editor privileges using a link on the desired locus page
- Log in to add and edit locus information
- Goal: 100 locus editors in 1 year

To annotate your Solanaceae genes!

Annotation of Networks

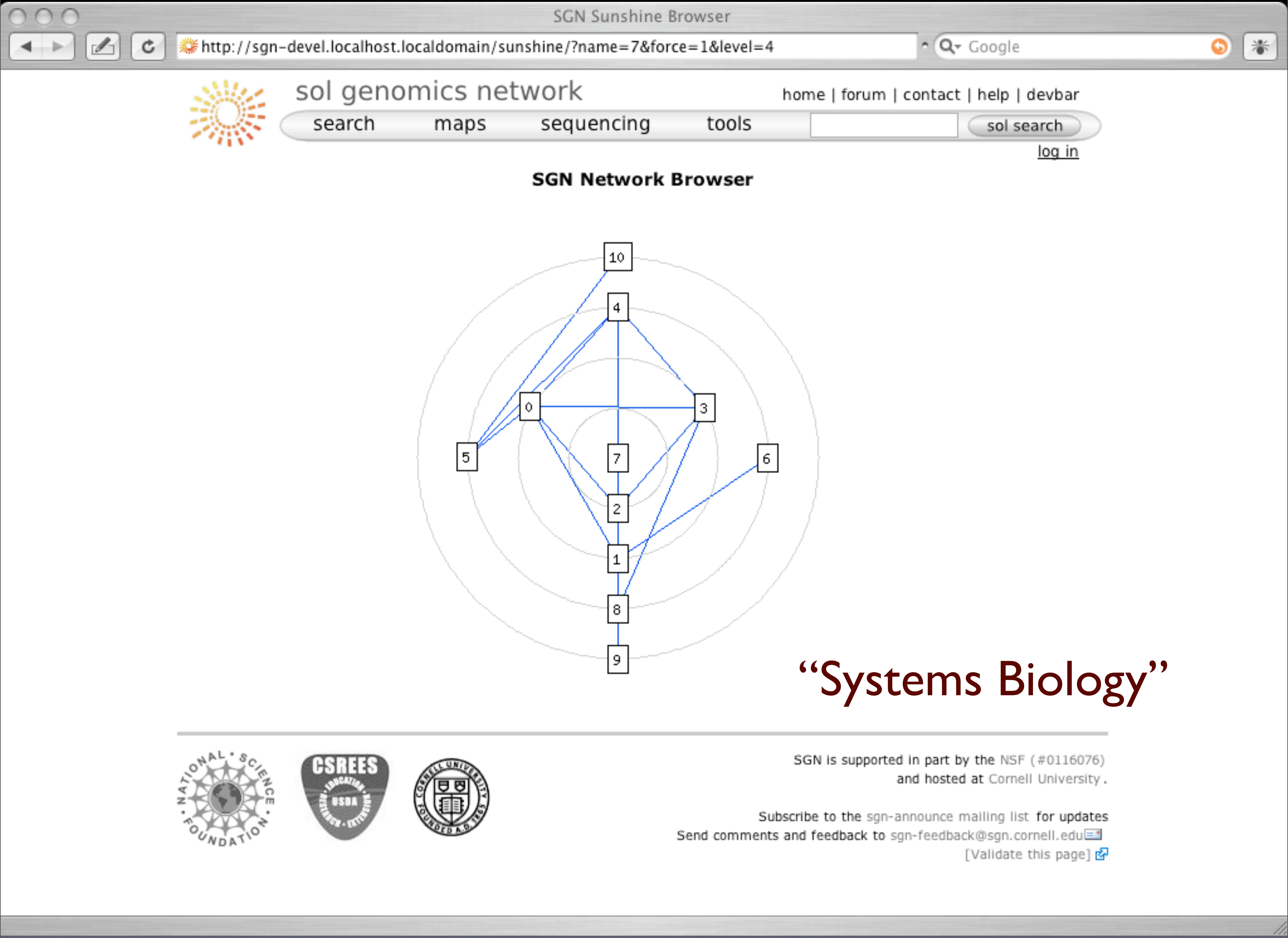
- Loci act as networks
- Locus-locus associations in the database
- Developed an ontology to describe locus-locus relationships
- Query tools and viewers



SGN Network Browser

SGN is supported in part by the NSF (#0116076)
and hosted at Cornell University.

Subscribe to the sgn-announce mailing list for updates
Send comments and feedback to sgn-feedback@sgn.cornell.edu
[Validate this page]



SGN Network Browser

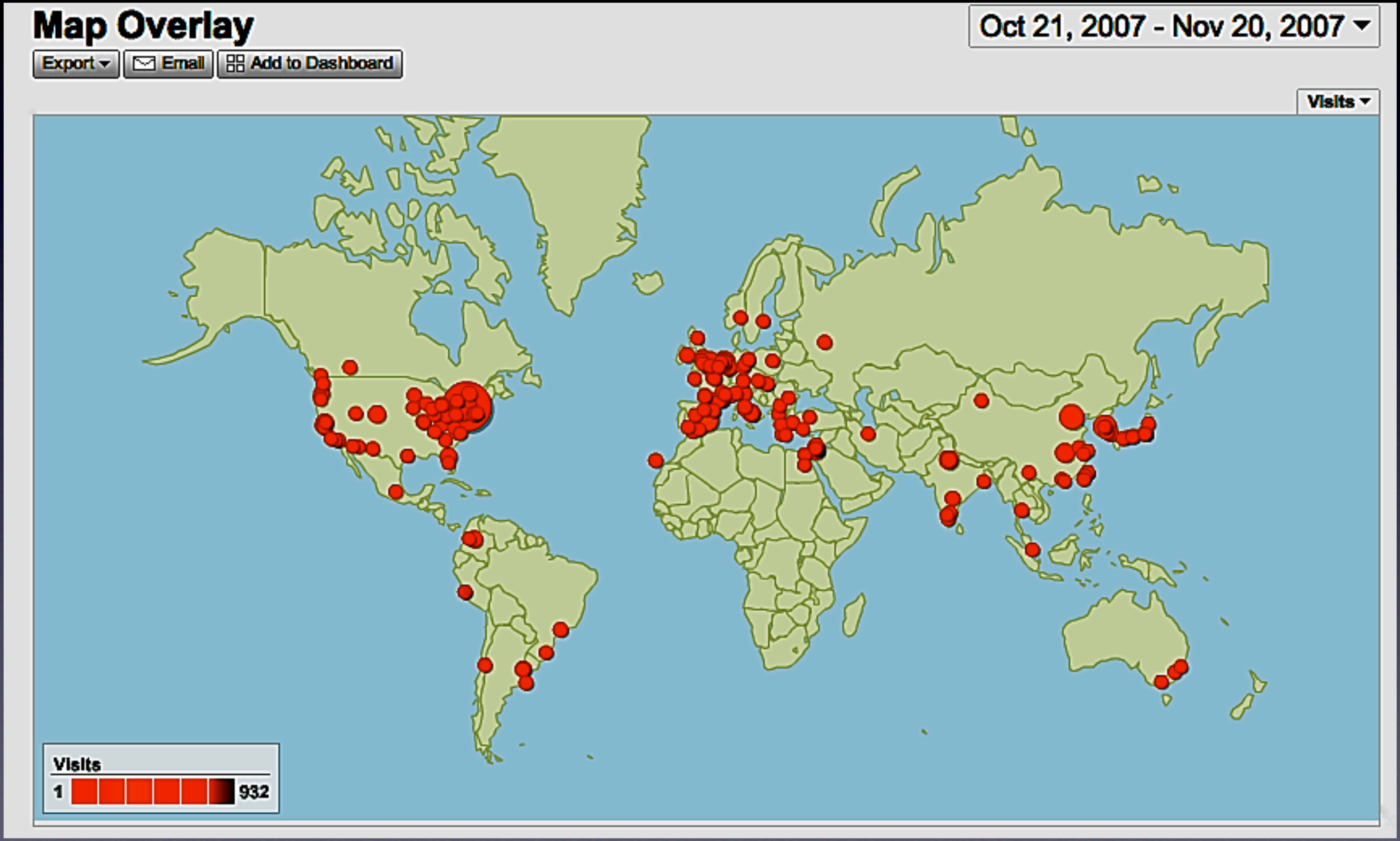
SGN is supported in part by the NSF (#0116076)
and hosted at Cornell University.

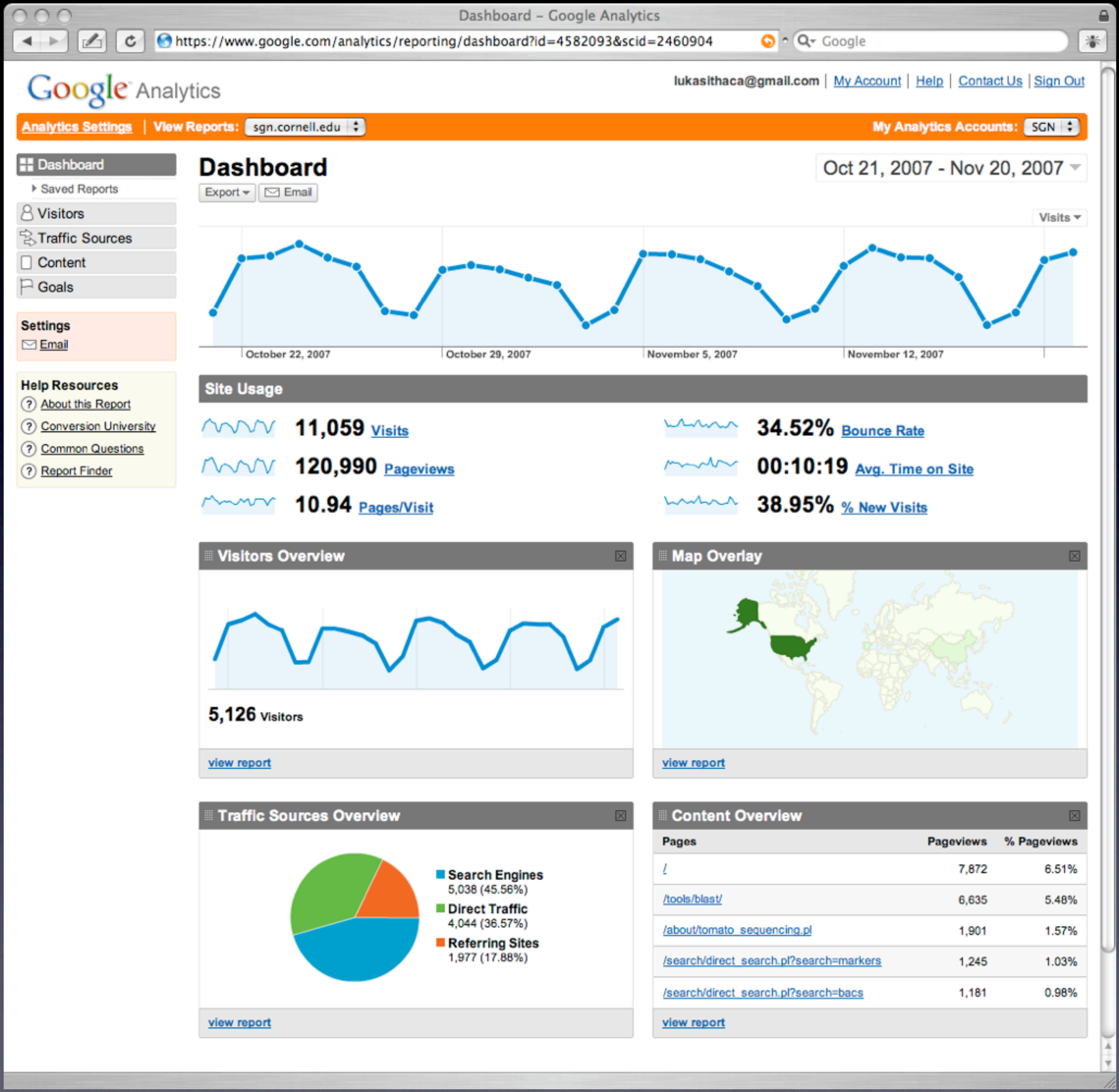
Subscribe to the sgn-announce mailing list for updates
Send comments and feedback to sgn-feedback@sgn.cornell.edu

[Validate this page]

"Systems Biology"

Site usage


Map Overlay
Export
Email
Add to Dashboard
Oct 21, 2007 - Nov 20, 2007
Visits
Visits
1
932



Dashboard

Oct 21, 2007 - Nov 20, 2007

Site Usage

11,059 Visits

120,990 Pageviews

10.94 Pages/Visit

34.52% Bounce Rate

00:10:19 Avg. Time on Site

38.95% % New Visits

Visitors Overview

5,126 Visitors

Traffic Sources Overview

- 🟦 Search Engines 5,038 (45.56%)
- 🟩 Direct Traffic 4,044 (36.57%)
- 🟧 Referring Sites 1,977 (17.88%)

Content Overview

Pages	Pageviews	% Pageviews
/	7,872	6.51%
/tools/blast/	6,635	5.48%
/about/tomato_sequencing.pl	1,901	1.57%
/search/direct_search.pl?search=markers	1,245	1.03%
/search/direct_search.pl?search=bacs	1,181	0.98%

Future directions

(short-mid term)

- Tomato genome chromosome 1 & 10
- MetaCyc pathway annotation
- Secretom Project
- Expand and improve SGN site, open community annotation for all interested communities

Future directions (Related Genomes)

- Potato (Intl. consortium),
- tobacco (PMI)
- *Mimulus guttatus* (Phrymaceae)

Future directions
SOL-100

(Sequence 100 Solanaceae genomes!)

454 sequencer

- New sequencing technologies will increase throughput to enable sequencing of 100s of genomes at low cost
- Need for bioinformatics will rise tremendously

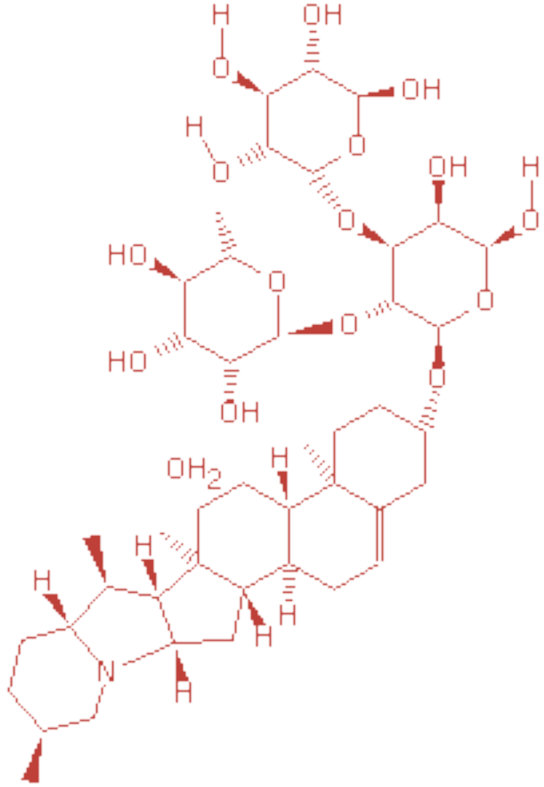
Future directions

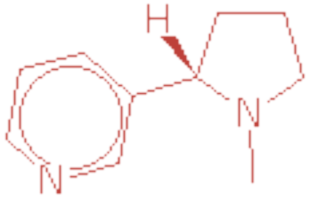
- Plants don’t exist in isolation
- Integrate Solanaceae pathogens or link to resources such as Pseudomonas database

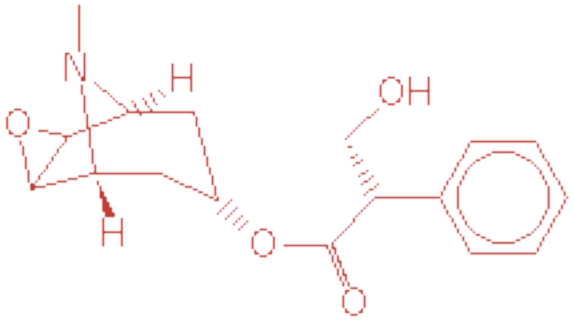
Future directions

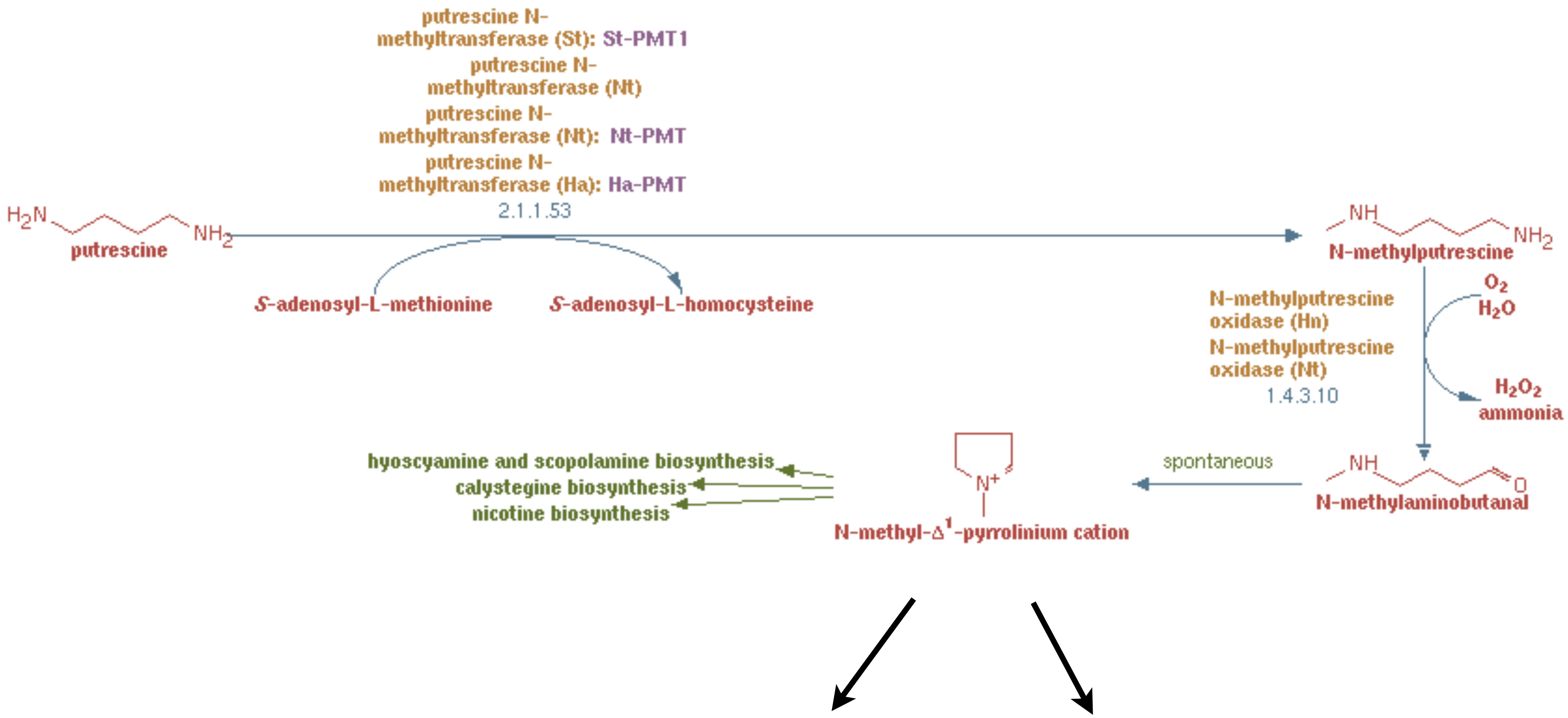
- Finished tomato genome sequence **&** SGN annotation data **&** SolCyc data
- Evolution of secondary metabolites in the Solanaceae
- Opportunities for collaboration

Future directions

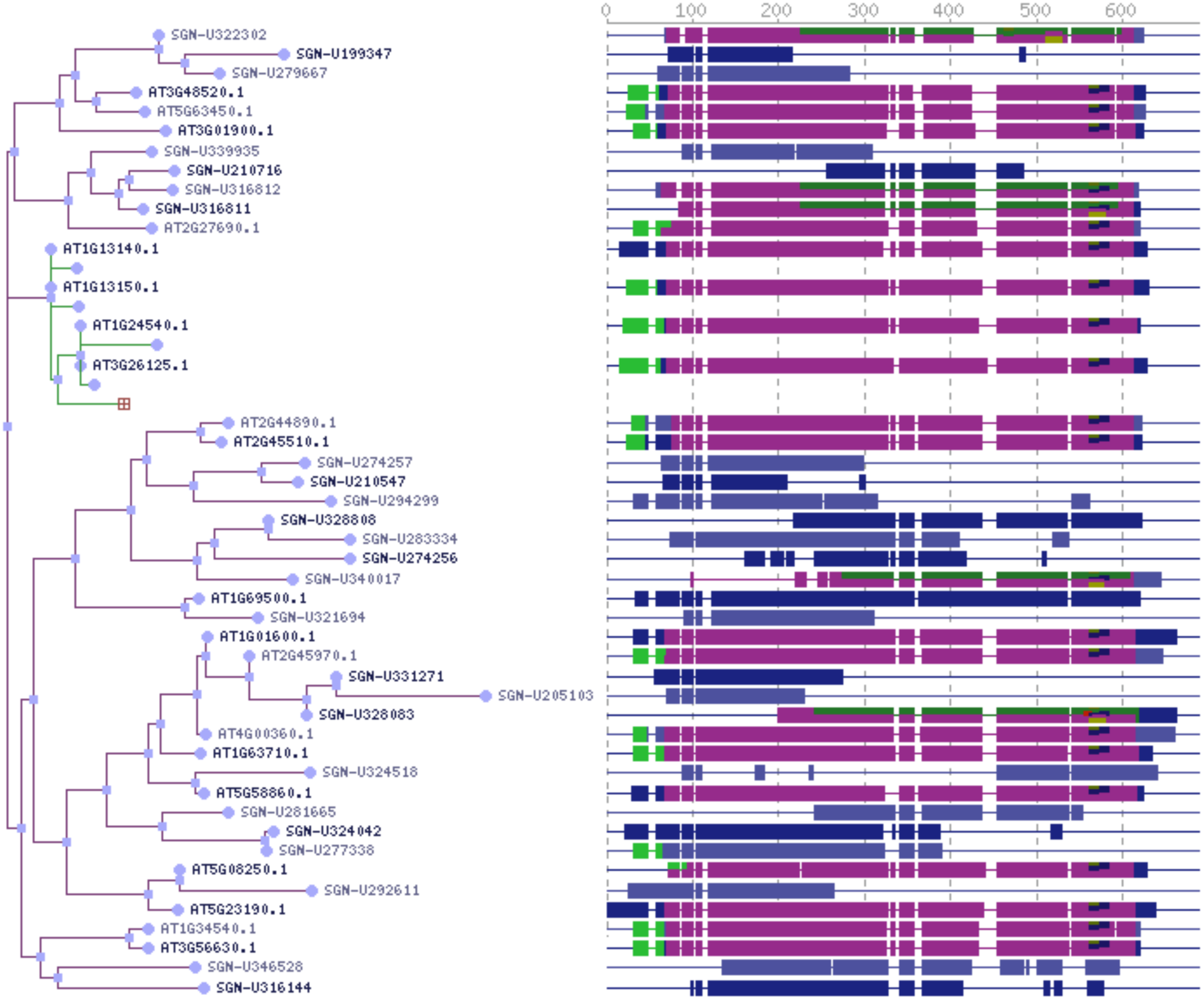








Nicotine scopolamine


0
100
200
300
400
500
600
SGN-U322302
SGN-U199347
SGN-U279667
AT3G48520.1
AT5G63450.1
AT3G01900.1
SGN-U339935
SGN-U210716
SGN-U316812
SGN-U316811
AT2G27690.1
AT1G13140.1
AT1G13150.1
AT1G24540.1
AT3G26125.1
AT2G44890.1
AT2G45510.1
SGN-U274257
SGN-U210547
SGN-U294299
SGN-U328808
SGN-U283334
SGN-U274256
SGN-U340017
AT1G69500.1
SGN-U321694
AT1G01600.1
AT2G45970.1
SGN-U331271
SGN-U205103
SGN-U328083
AT4G00360.1
AT1G63710.1
SGN-U324518
AT5G58860.1
SGN-U281665
SGN-U324042
SGN-U277338
AT5G08250.1
SGN-U292611
AT5G23190.1
AT1G34540.1
AT3G56630.1
SGN-U346528
SGN-U316144

Summary

- SGN has innovated in several areas
- Clade-oriented approach
- Community annotation

Acknowledgments

Naama Menda

Isaak Tecle

Robert Buels

Adri Mills

Dean Eckstrom

Anuradha Pujar

Tom York

Chris Carpita

Aureliano Bombarley

Previous SGN staff

Beth Skwarecki
Marty Kreuter
Chenwei Lin
John Binns
Teri Solow
Nicholas Taylor
Dan Ilut
Robert Ahrens
Koni Wright

Interns

(internships organized with the help of Joyce van Eck)
Outreach - undergrads & high school students

Tim Jacobs
Sasha Naydich
Jessica Reuter
Matthew Crumb
Bob Albright
Emily Hart
Scott Bland
Amarachukwu Enemuo
Benjamin Cole

Caroline Nyenke
Tyler Simmons
Evan Herbst
Emil Keyder
Aseem Kohli
Igor Dolgalev
Miriam Wallace
Jay Gadgil
Jennifer Lee

BTI

- Jim Giovannoni
- Joyce van Eck

Cornell

- Steven Tanksley
- Jocelyn Rose

USDA

- Ted Tannhauser

Danforth Center

- Brad Barbazuk

U Colorado

- Steven Stack

U Wisconsin

- David Spooner

CIP

- Meredith Bonierbale

Nestle

- James McCarthy
- Dominique Crouzillat

ATC

- Steve Coates

Users & Submitters
Dani Zamir

C.M. Rick
TGRC
Tomato Genetics Resource Center

Funding

National Science Foundation
WHERE DISCOVERIES BEGIN

USDA CSREES

BARD

National Institutes of Health
The Nation's Medical Research Agency

Nestlé Good Food, Good Life

Thanks!

(c) the Internet