From Genomes to Phenomes to Breeding

Lukas Mueller
Boyce Thompson Institute

Boyce Thompson Institute
for Plant Research

Sol Genomics Network

http://solgenomics.net/

sol genomics network

home | forum | contact | help

search maps genomes tools sol search

log in | new user

Genes

Search loci
Search unigenes
Expression
Search/browse tomato genome
Become an SGN locus editor

Phenotypes

Maps & Markers

Breeders Toolbox

Genomes & Sequences

Pathways

About SGN

News

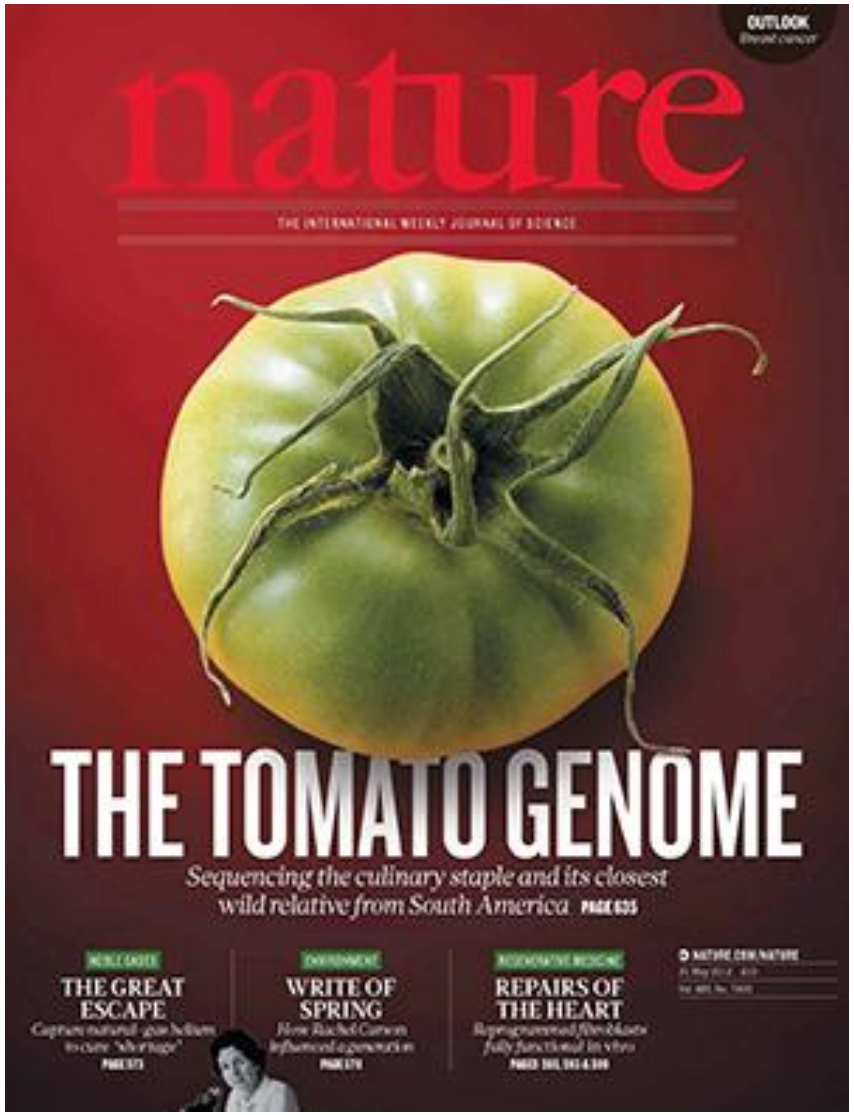
New *Solanum lycopersicum* assembly 2.10 released

Events

5th EPSO Conference

http://solgenomics.net/

Go to "http://solgenomics.net/search


OUTLOOK
nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
THE TOMATO GENOME
Sequencing the culinary staple and its closest wild relative from South America PAGE 635
THE GREAT ESCAPE
WRITE OF SPRING
REPAIRS OF THE HEART

Tomato Genome

- BAC by BAC approach 2004-2009 (~1500 BACs)
- Whole genome shotgun of entire tomato genome, started in 2009
- Technologies used: 454, Solexa (Syngenta), SOLiD
- 454 data assembled using Newbler
- Homopolymer correction using Solexa data
- Integration of BAC sequences
- Ordering and orientation of scaffolds based on

sol genomics network

Why Gh13, a begomivirus-resistant inbred?

- Begomoviruses are a major threat in subtropical and tropical countries
- Gh13 highly resistant to monopartite and bipartite begomoviruses
- Gh13 used in association studies of molecular marker with resistance
  - F3 family experiments
  - RIL population available
- SolCAP SNP analysis available
- NSF requires that seed be available for distribution

Boyce Thompson Institute for Plant Research

Origin OF Gh13

TYLCV virus resistance: HUJI

(Vidavski and Czosnek, 1994)

L. hirsutum

S. Habrochaites

LA1777 & LA0386

6 years

Ih902 x S line, FAVI 9

Hybrids sent to Guatemala (1998)

Boyce Thompson Institute for Plant Research

sol genomics network

sol genomics network

Gh13 inbred: known introgressions

Ty3 ï chromosome 6, introgressed from wild species (*S. chilense?*)

I2 ï chromosome 11 (*S. pimpinellifolium)*

Other introgressions from *S. habrochaites.*
+ Other species?

Boyce Thompson Institute for Plant Research

sol genomics network

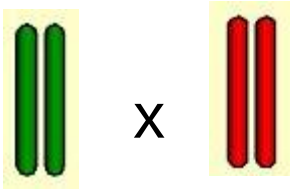
Gh13 inbred: known introgressions

Ty3 ï chromosome 6, introgressed from wild species (*S. chilense?*)

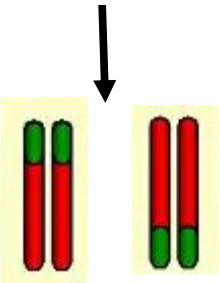
I2 ï chromosome 11 (*S. pimpinellifolium)*

Other introgressions from *S. habrochaites.*
+ Other species?

Disease resistance alleles can be found in wild species
Need to find introgression **regions**
And define introgression **contents**



F1, F2, BC . . .

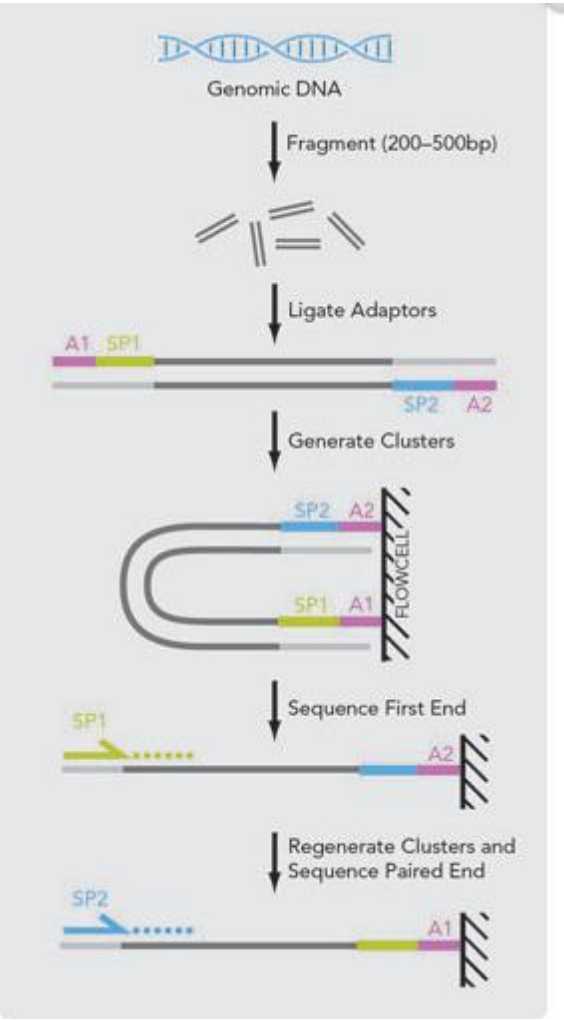


Boyce Thompson Institute for Plant Research

sol genomics network

Gh13 inbred: Whole genome sequencing

- Illumina HiSeq 2000
- One lane paired-ends = 20X tomato genome coverage
- Cost in 2012 : 2,400$

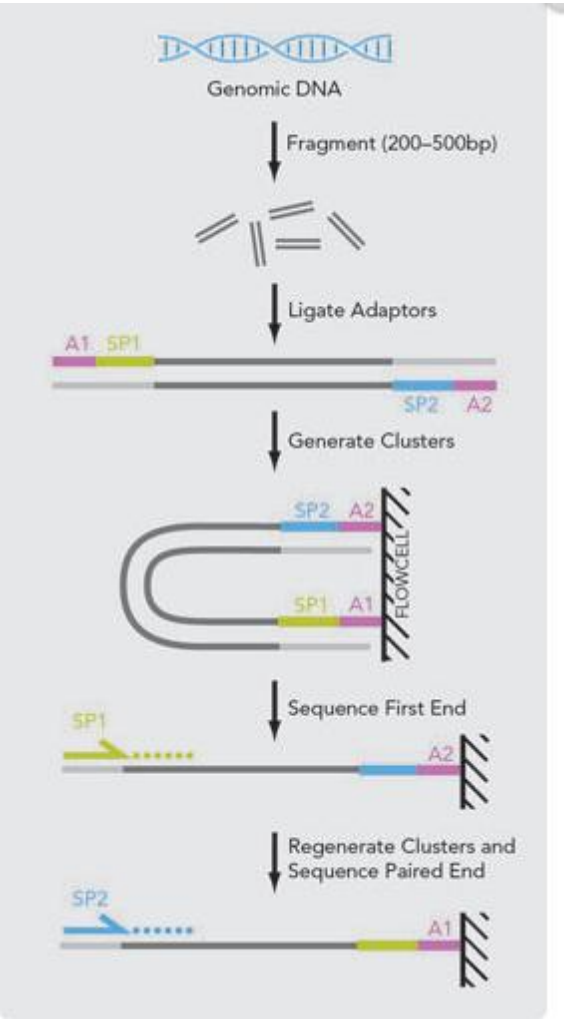


Boyce Thompson Institute for Plant Research

sol genomics network

Gh13 inbred: Whole genome sequencing

- Illumina HiSeq 2000
- One lane paired-ends = 20X tomato genome coverage
- Cost in 2012 : 2,400$



Output: high number of reads
Relatively simple to align to a reference genome.

Challenges:

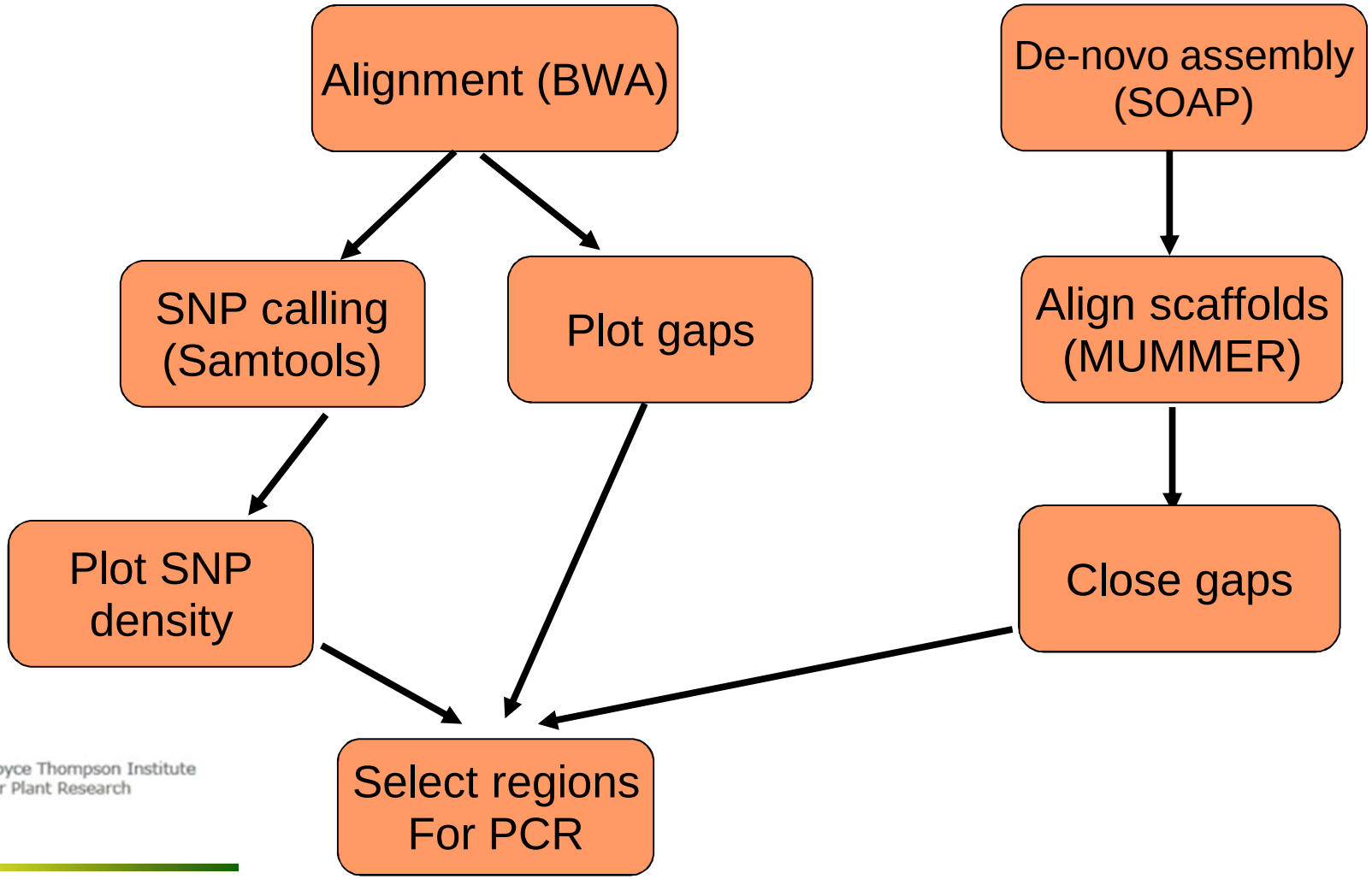
- Low coverage regions
- Regions different from Heinz1706

Boyce Thompson Institute for Plant Research

sol genomics network

Gh13 inbred: Whole genome sequencing

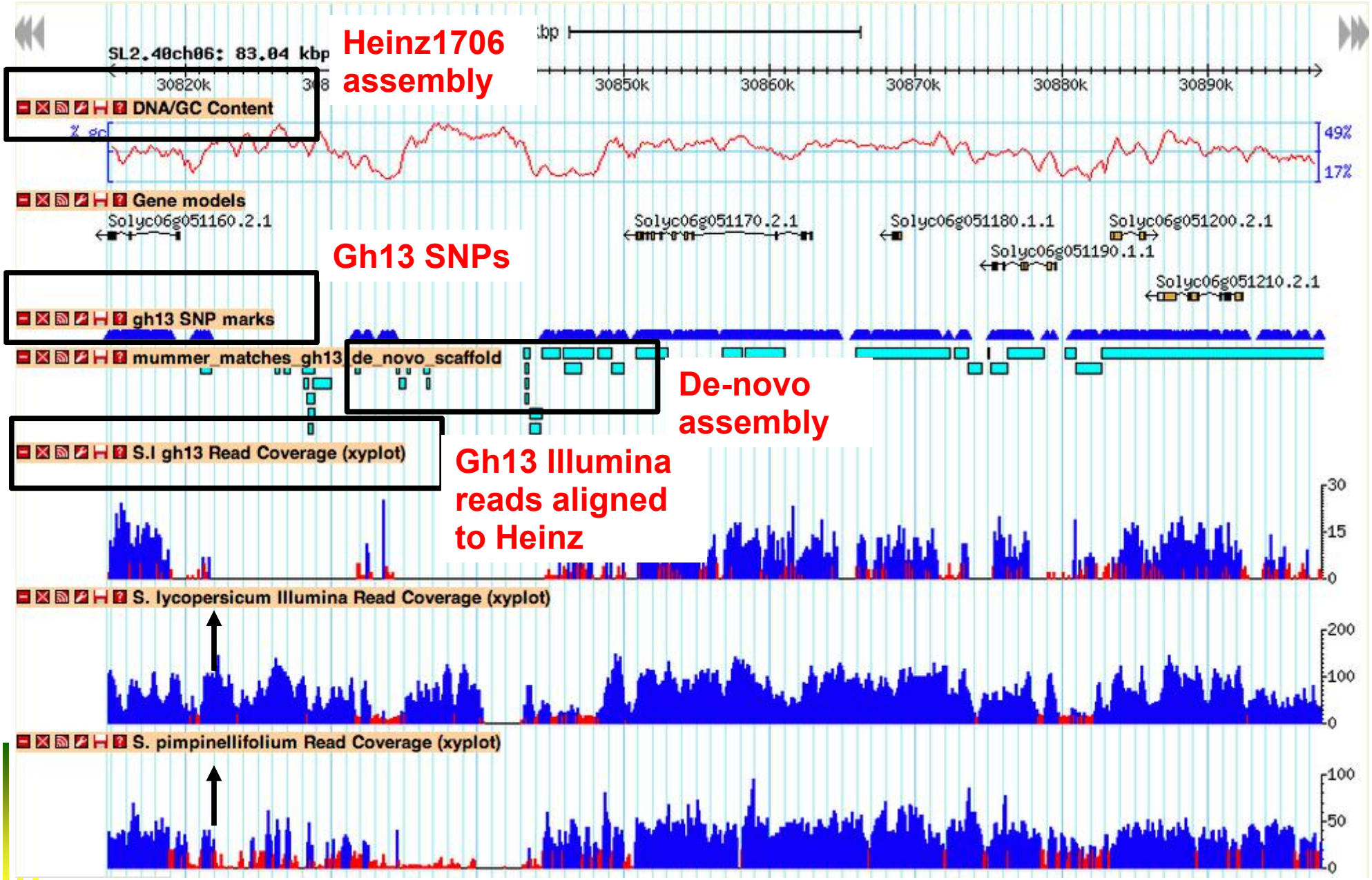
Assembly: Heinz1706 is the reference genome



Boyce Thompson Institute for Plant Research

Assembly: Heinz1706 is the reference genome

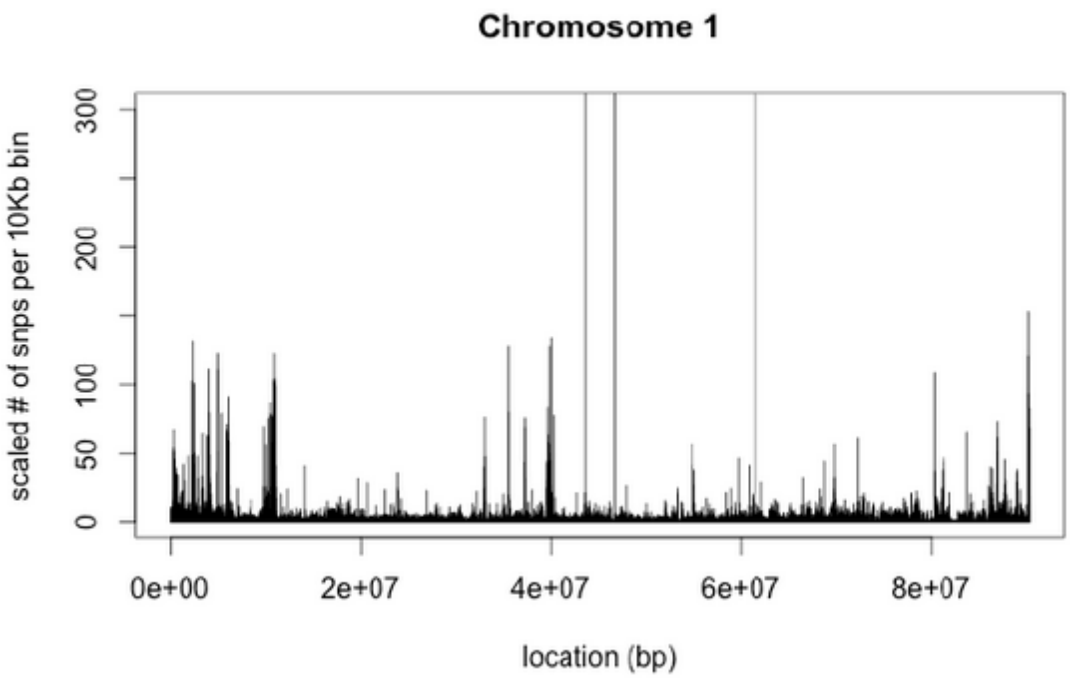
sol genomics network

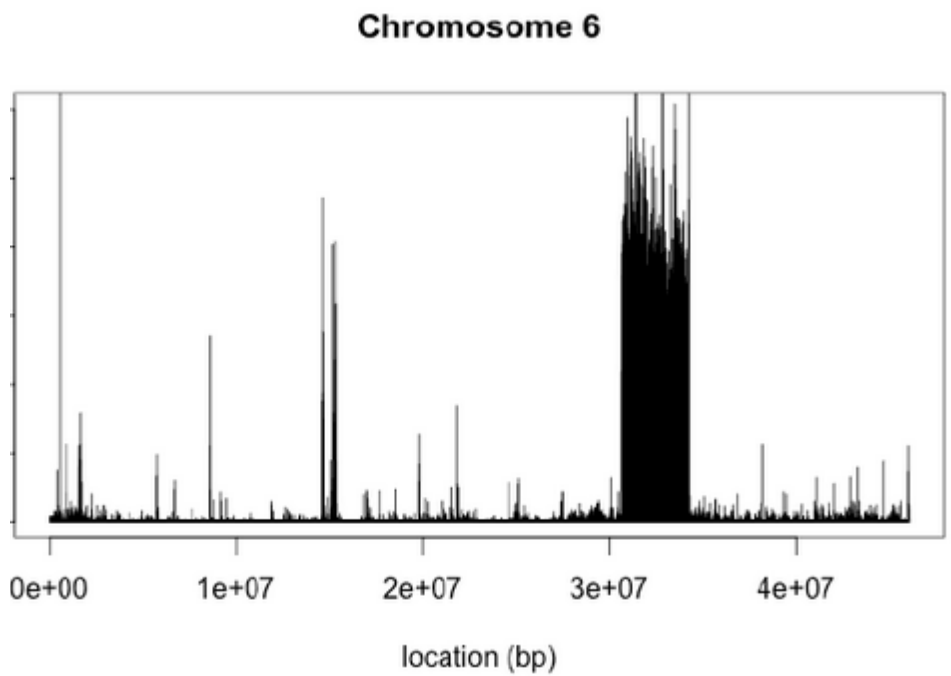


sol genomics network

Gh13 inbred: SNP distribution

Hypothesis: SNPs are denser in introgression regions.



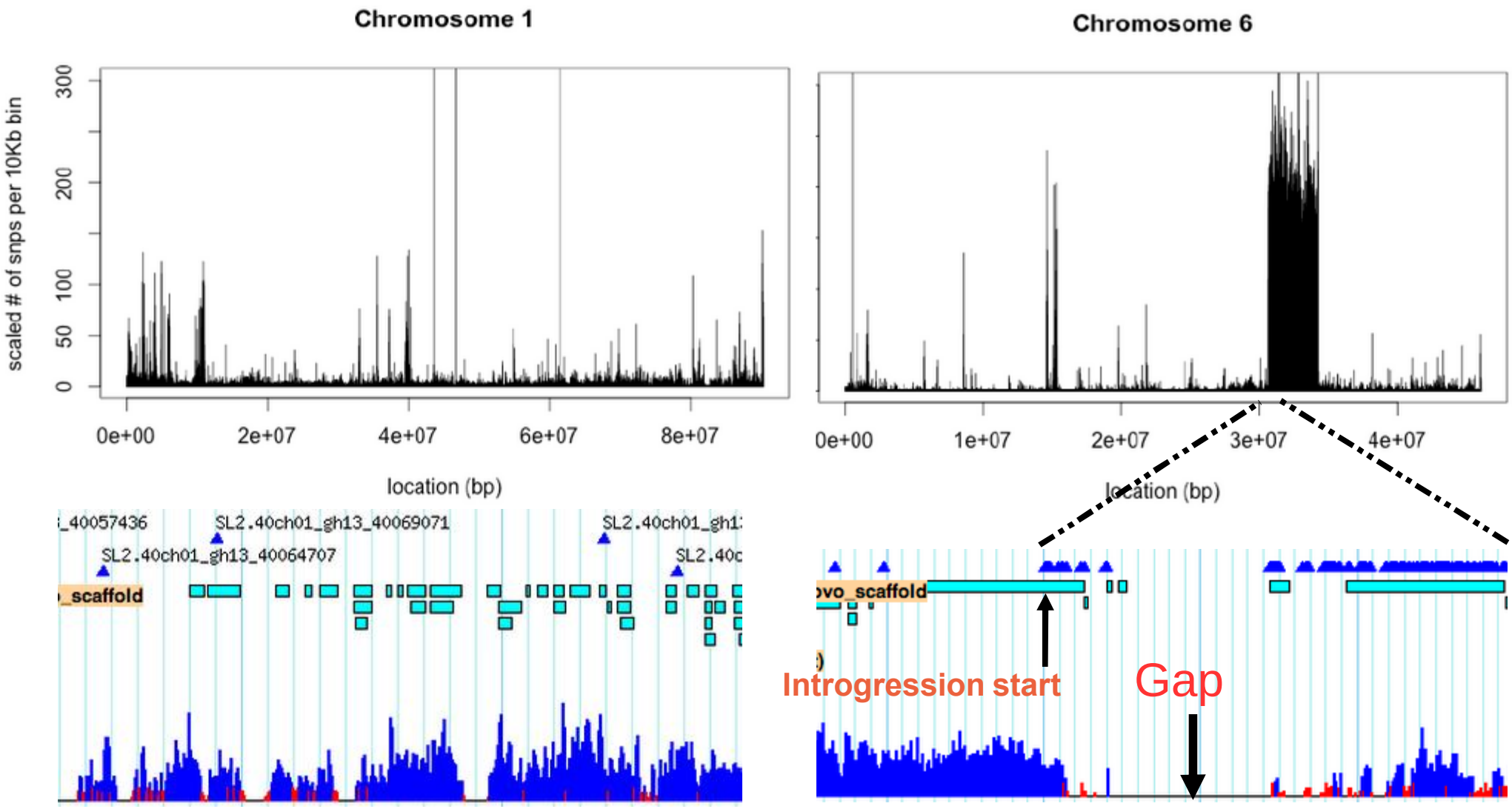


Boyce Thompson Institute for Plant Research

sol genomics network

Gh13 inbred: SNP distribution

Hypothesis: SNPs are denser in introgression regions.

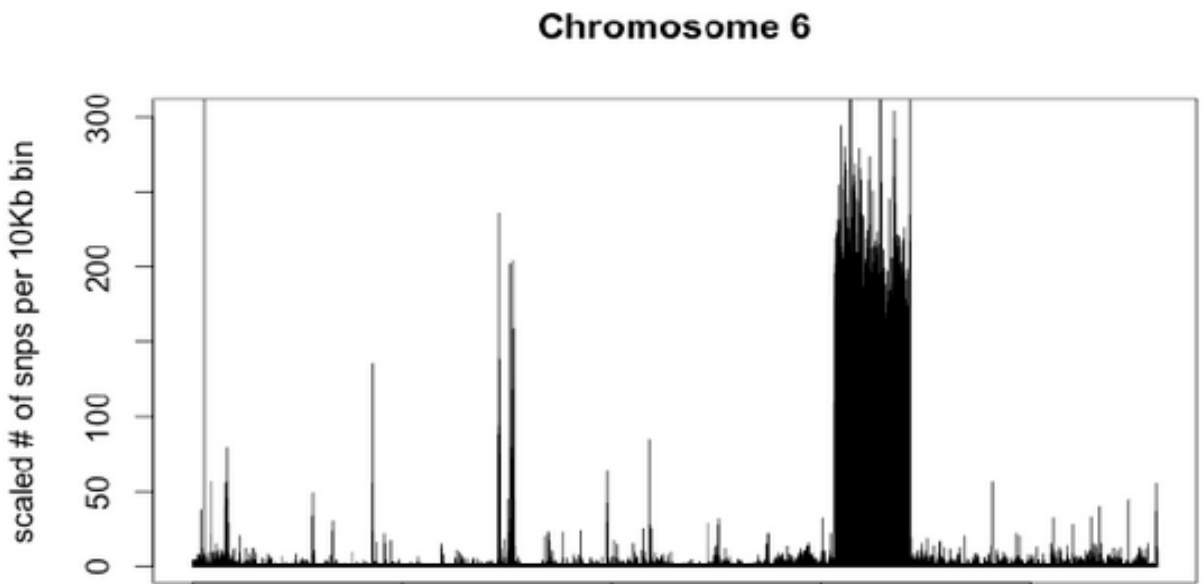


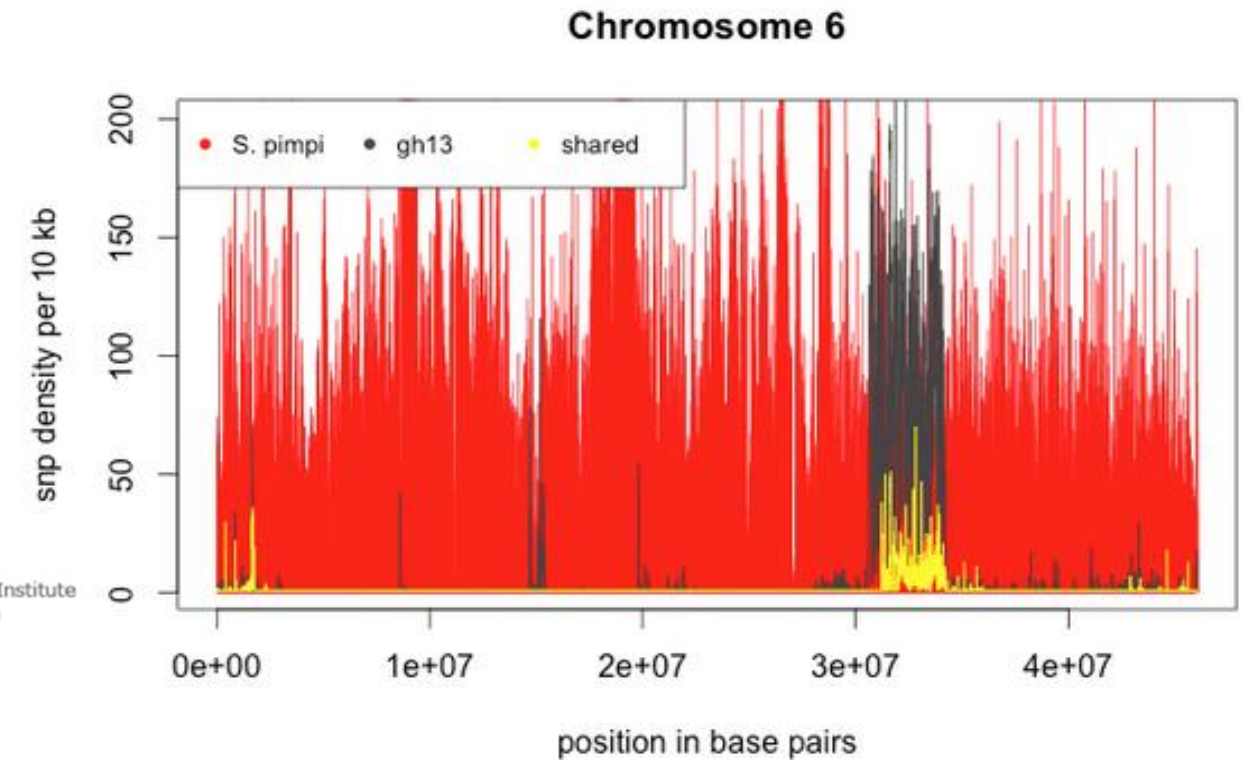
~50kb , Chr 1

~50kb , Chr 6

SNP distribution: Gh13, *S.pimpinellifolium*

sol genomics network

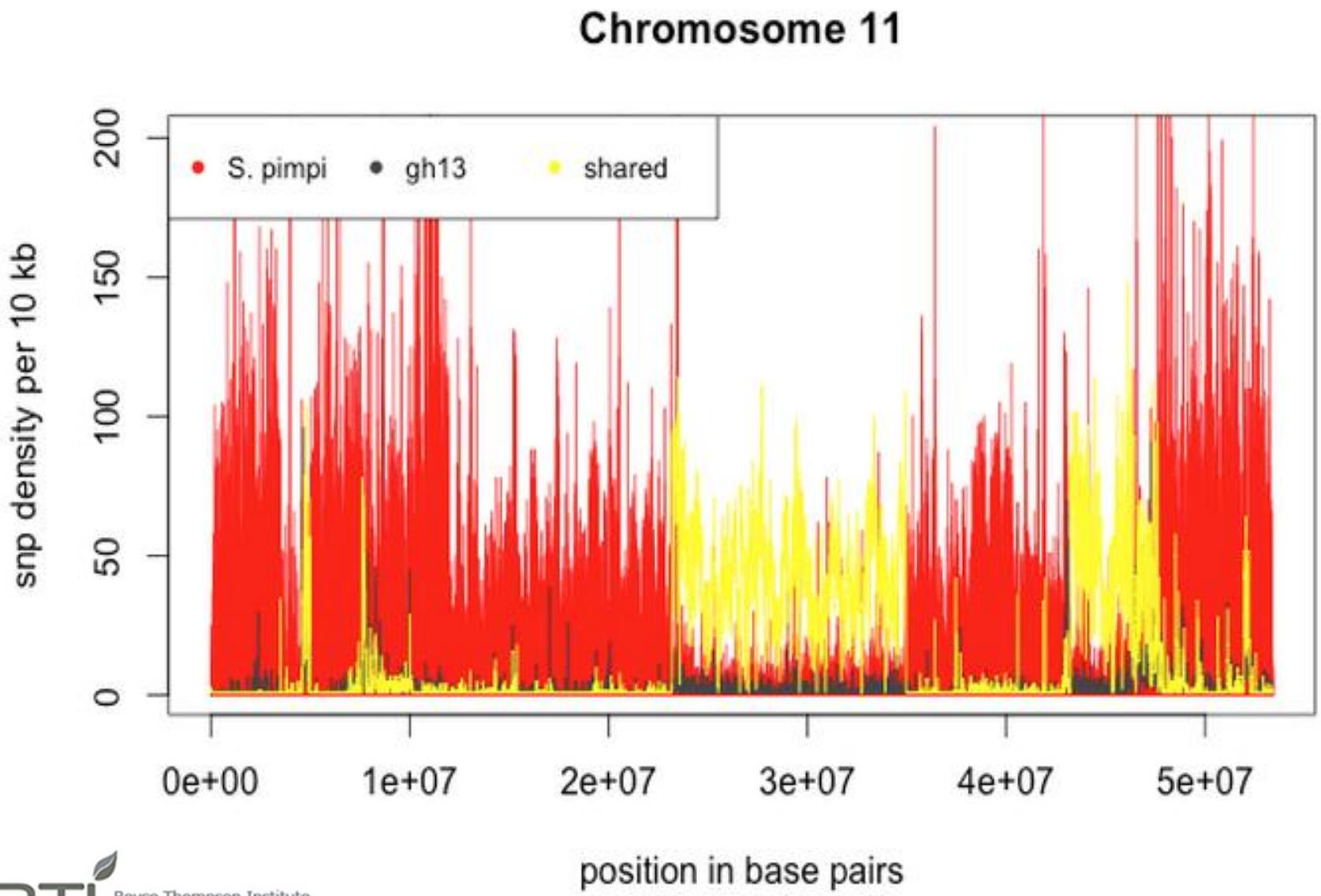




Boyce Thompson Institute for Plant Research

SNP distribution: Gh13, *S.pimpinellifolium*

sol genomics network



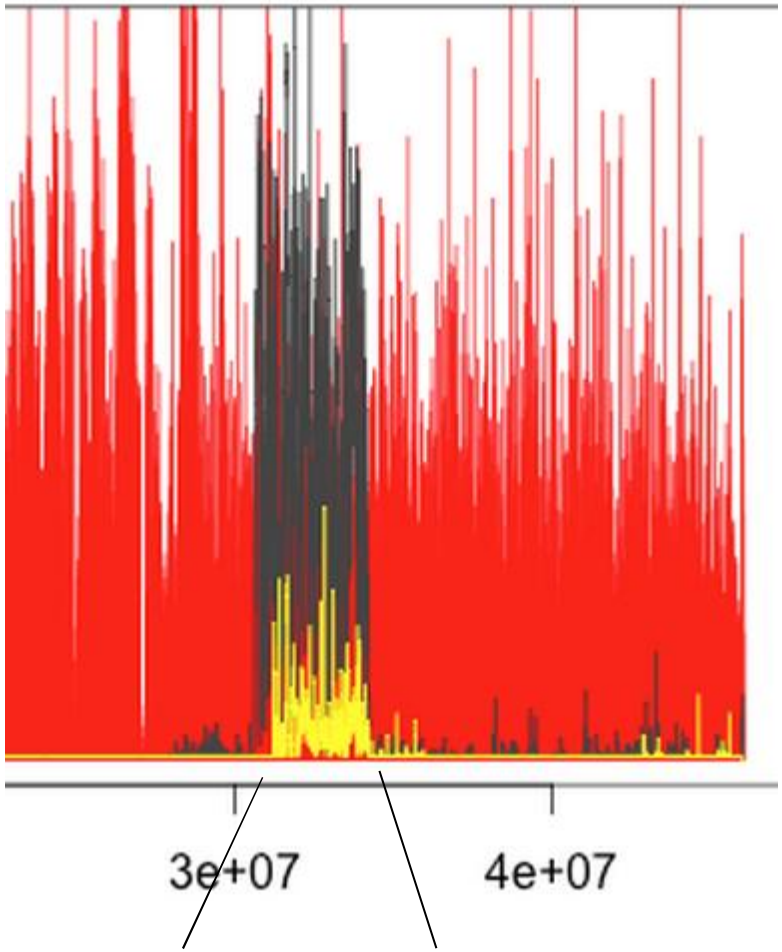
Boyce Thompson Institute for Plant Research

S.pimpinellifolium introgressions?

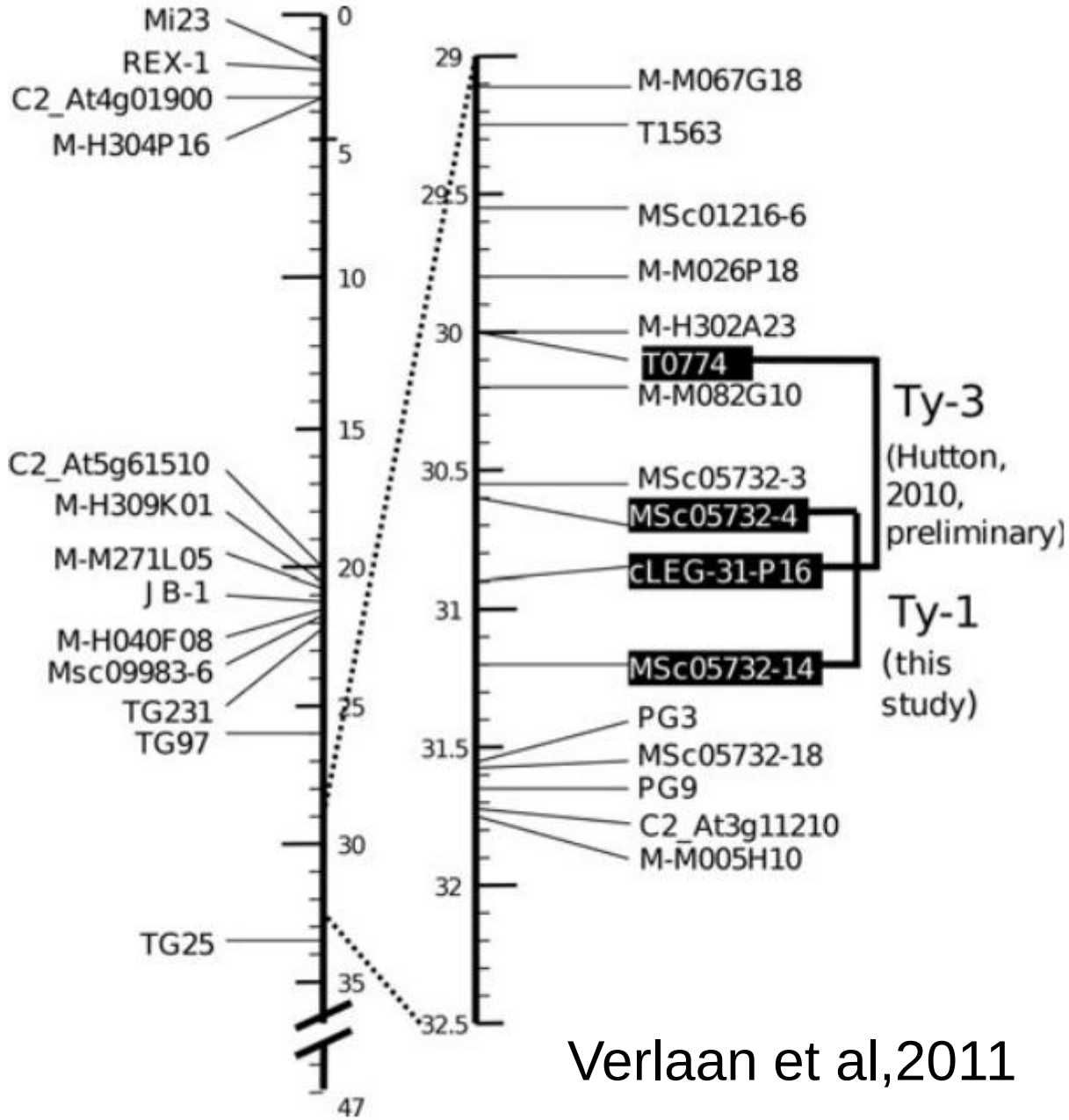
SNP distribution: Gh13, *S.pimpinellifolium*

sol genomics network

Chromosome 6



30.6Mb - 34Mb



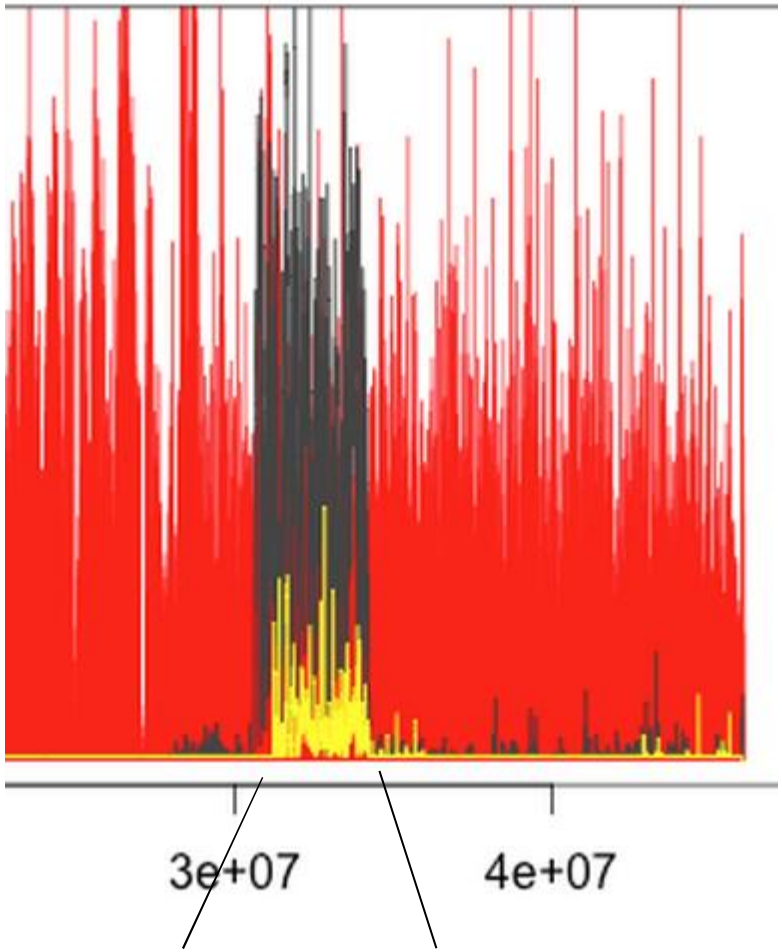
Verlaan et al,2011

Boyce Thompson Institute for Plant Research

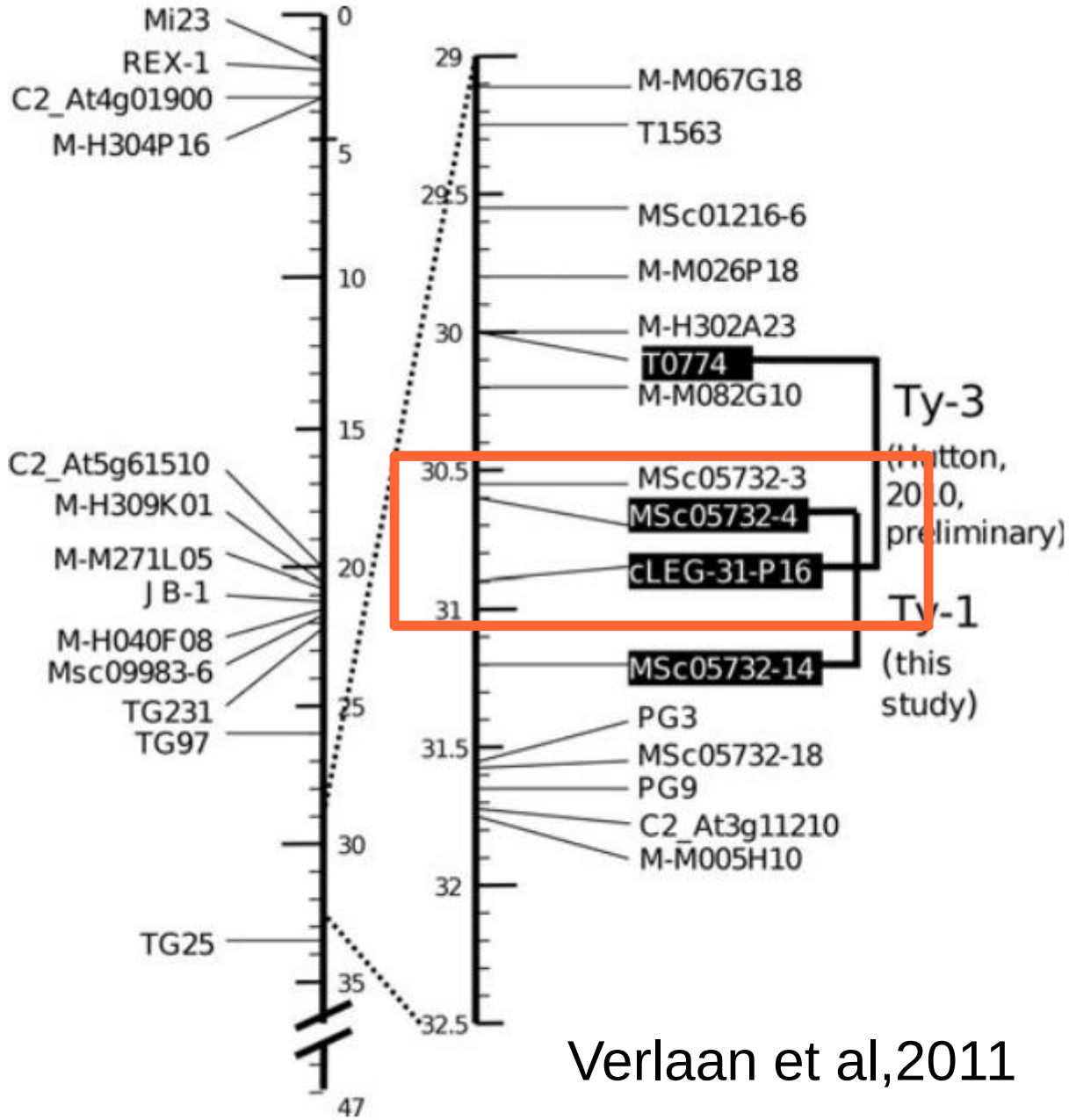
SNP distribution: Gh13, *S.pimpinellifolium*

sol genomics network

Chromosome 6



30.6Mb - 34Mb



Verlaan et al,2011

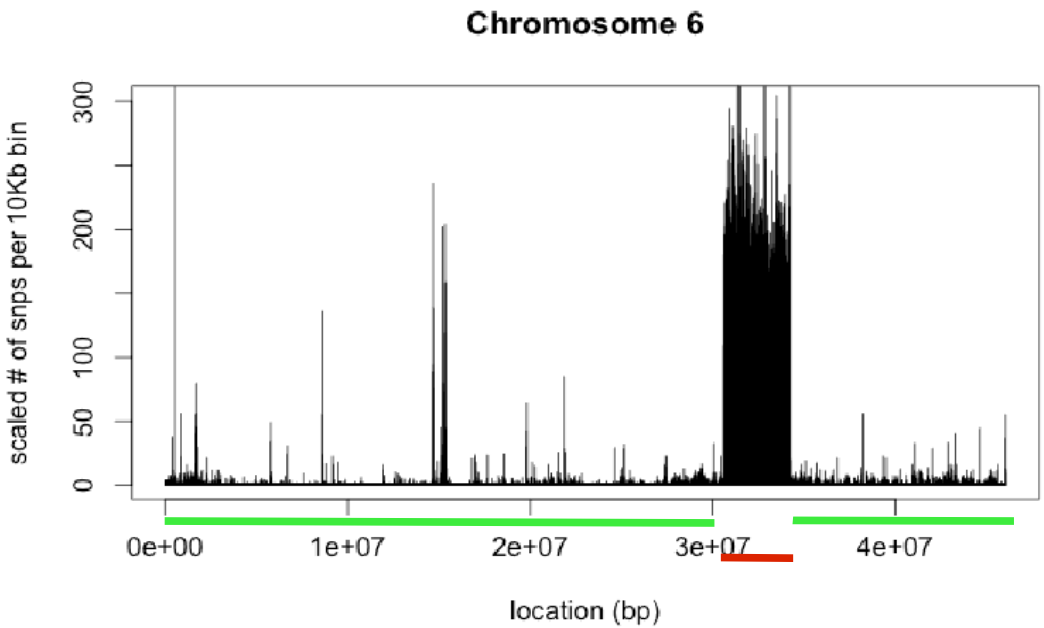
Boyce Thompson Institute for Plant Research

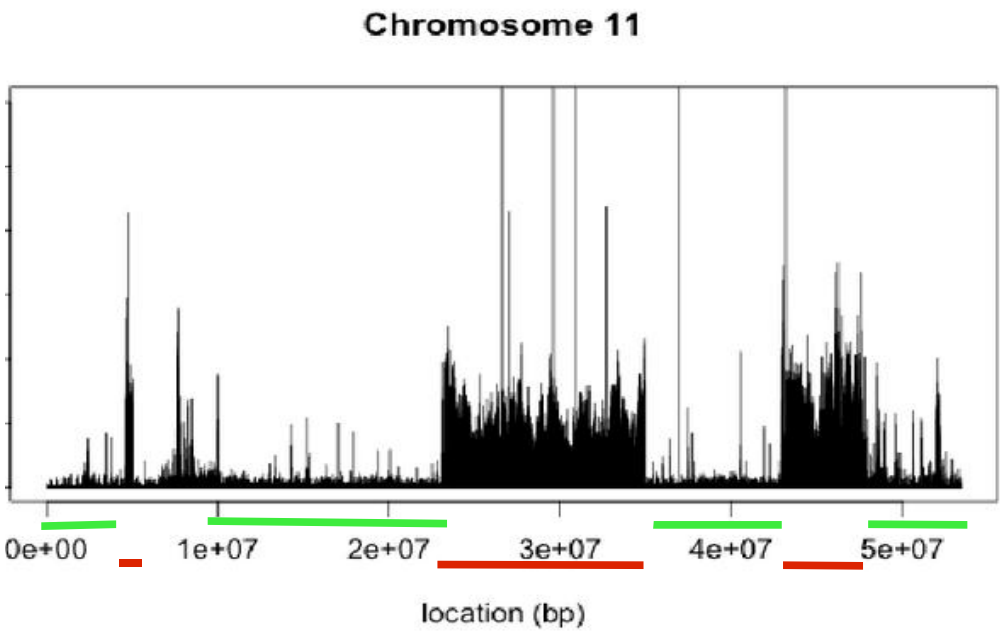
PCR design: Gh13 chr. 6 and 11

sol genomics network

Hypothesis:

1. SNP non-peak regions are closest to Heinz1706
2. SNP Peak regions come from wild introgressions

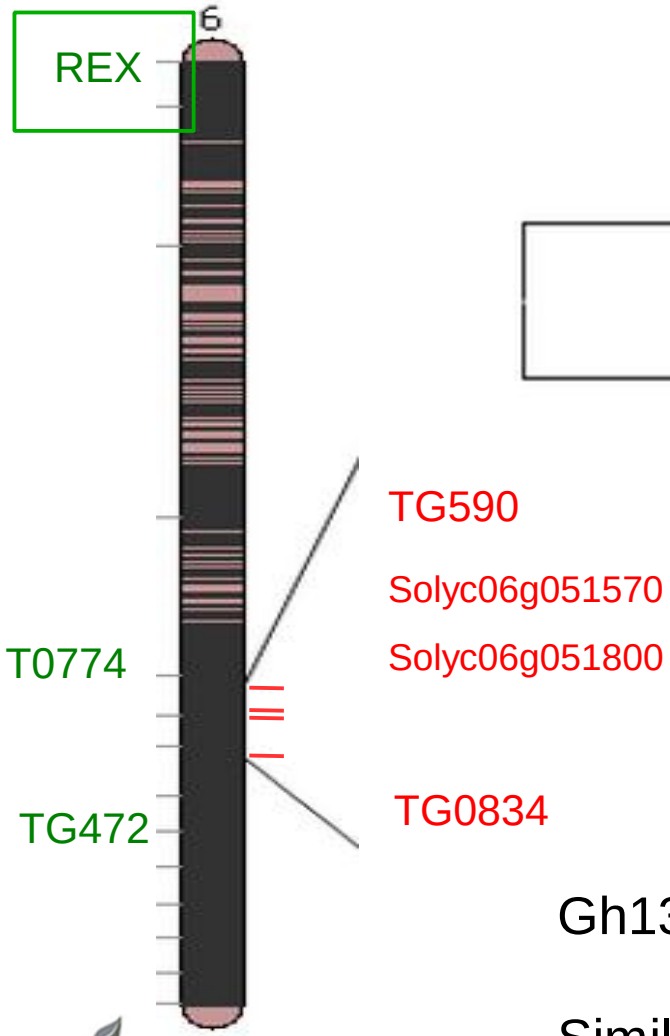


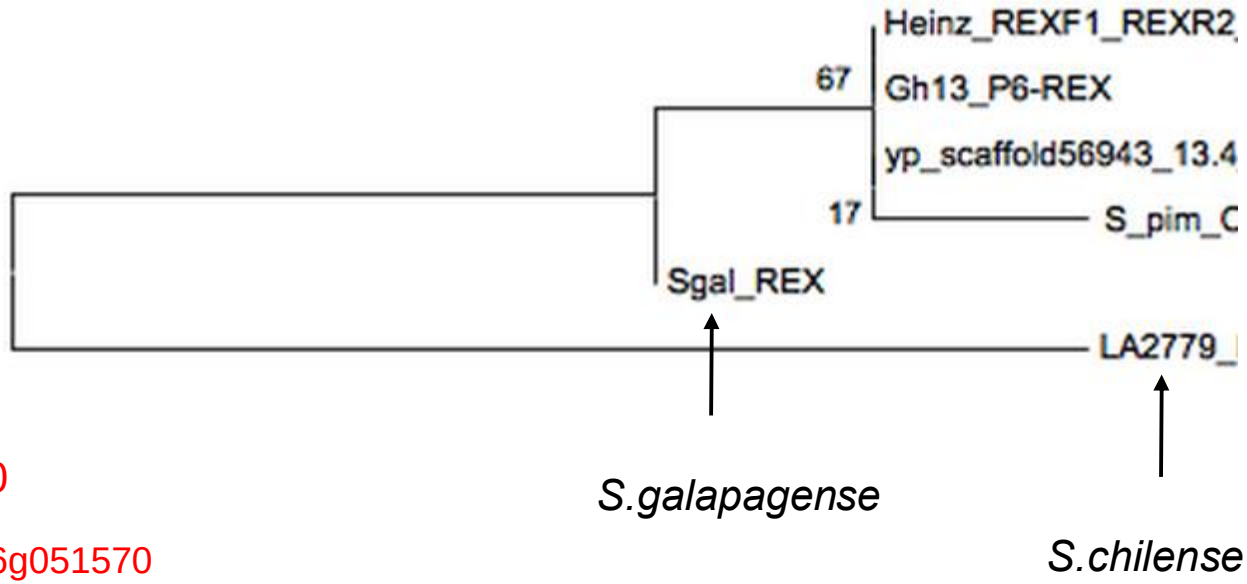


Boyce Thompson Institute for Plant Research

PCR design: Gh13 chr. 6

1. SNP non-peak regions are closest to Heinz1706



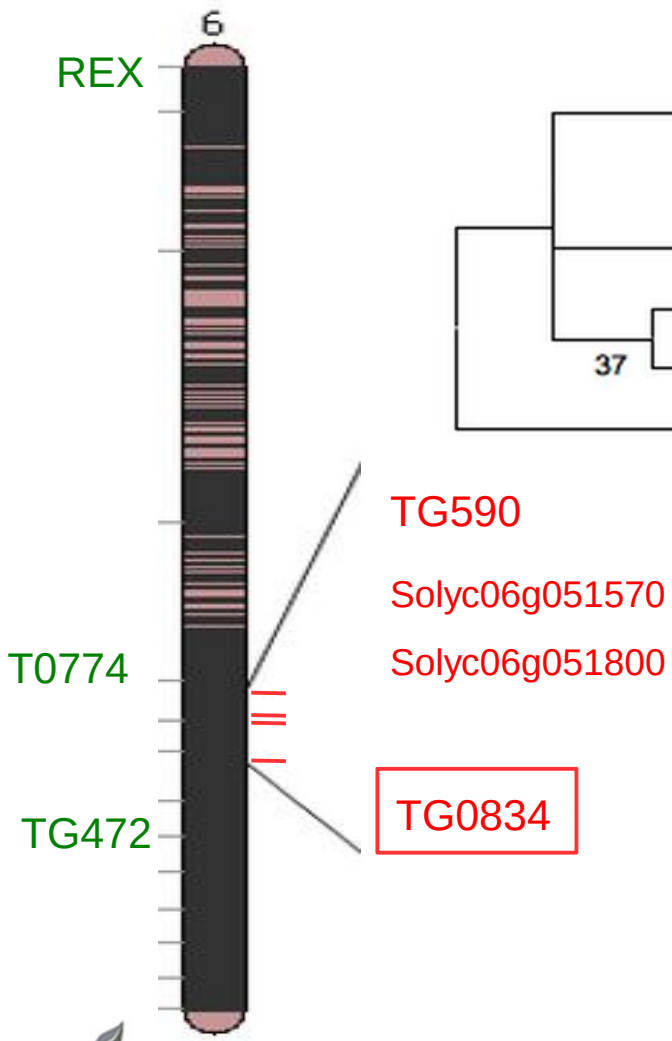


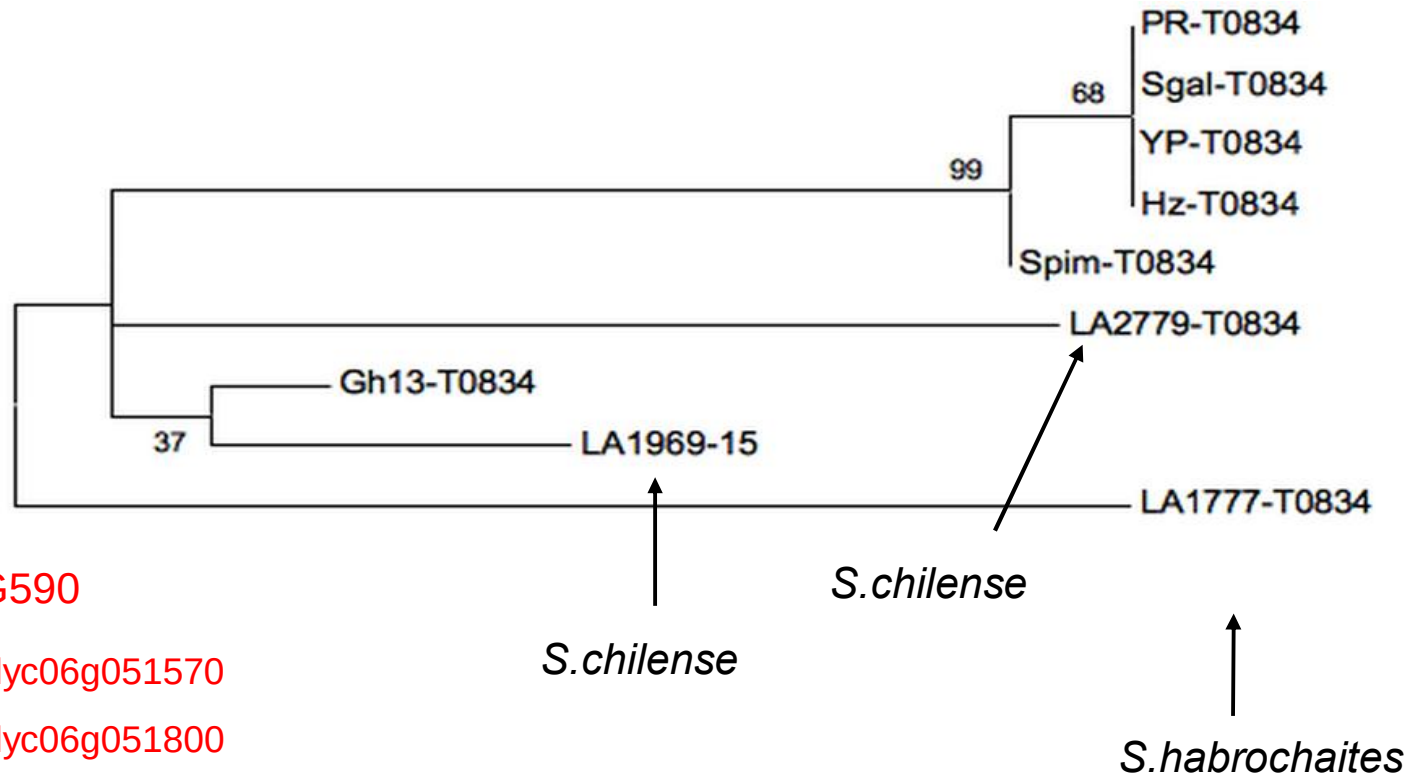
Gh13 closest to Heinz, Yellow Pear, *S.pimpinellifolium*

Similar results for all other non-peak markers!
(trees built with MEGA, maximum likelihood 500 bootstrap replicates)

PCR design: Gh13 chr. 6

2. SNP Peak regions come from wild introgressions



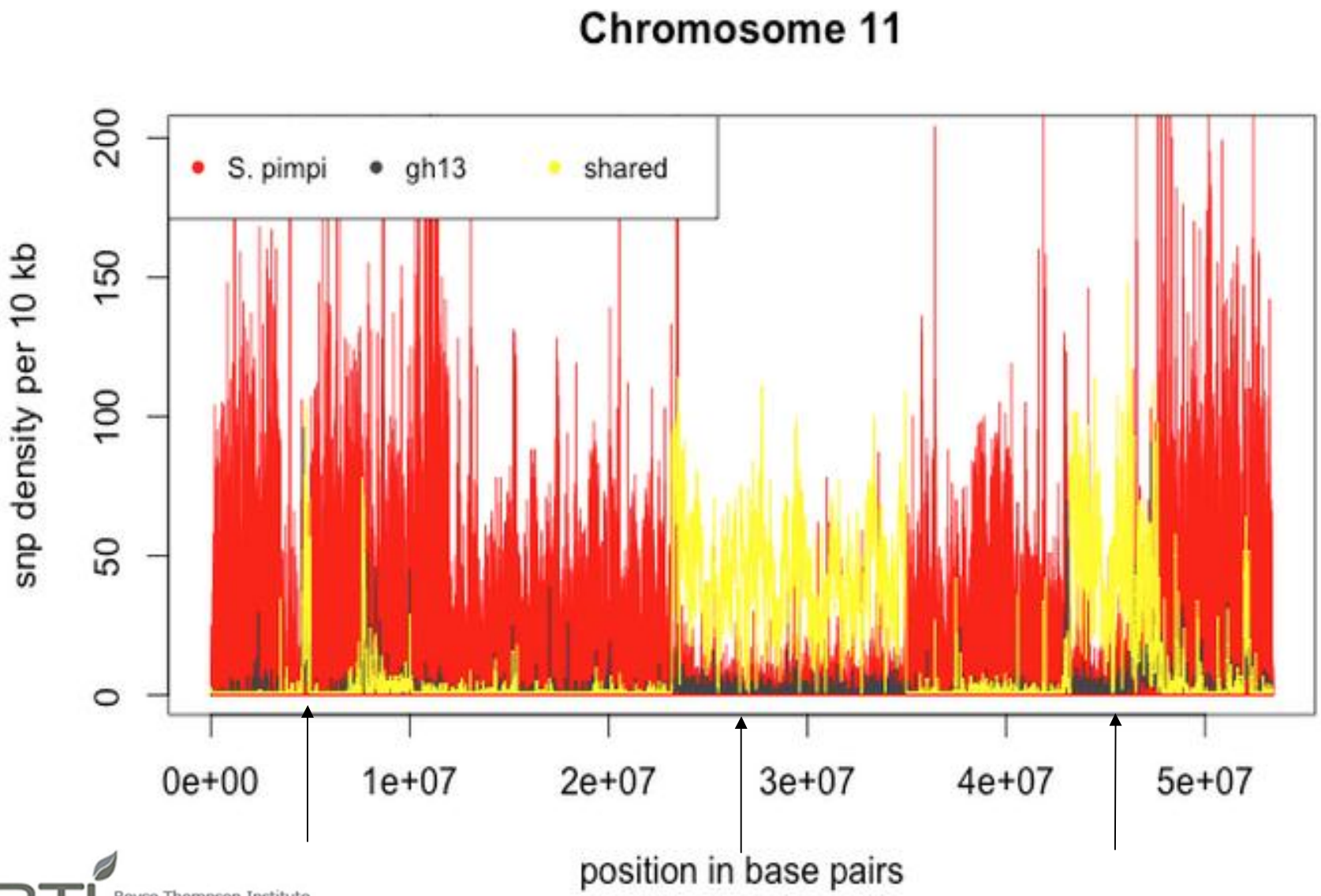


Gh13 closest to *S.chilense*.
Purple Russian, Yellow Pear, Heinz, *S.galapagense* cluster together.

Similar results for all other SNP-peak markers!

SNP distribution: Gh13, *S.pimpinellifolium*

sol genomics network

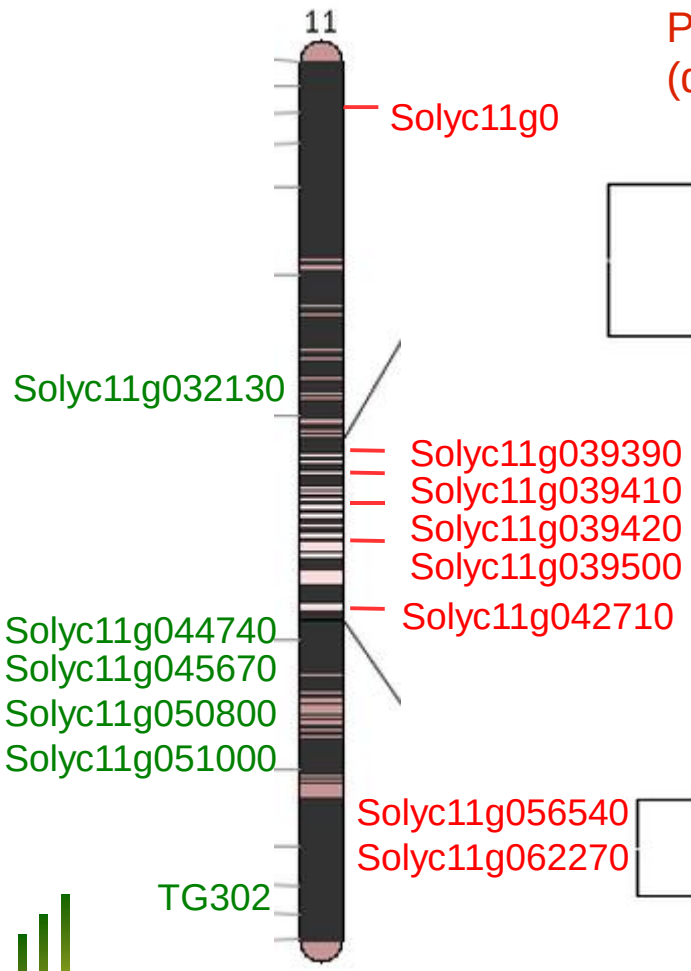


Boyce Thompson Institute for Plant Research

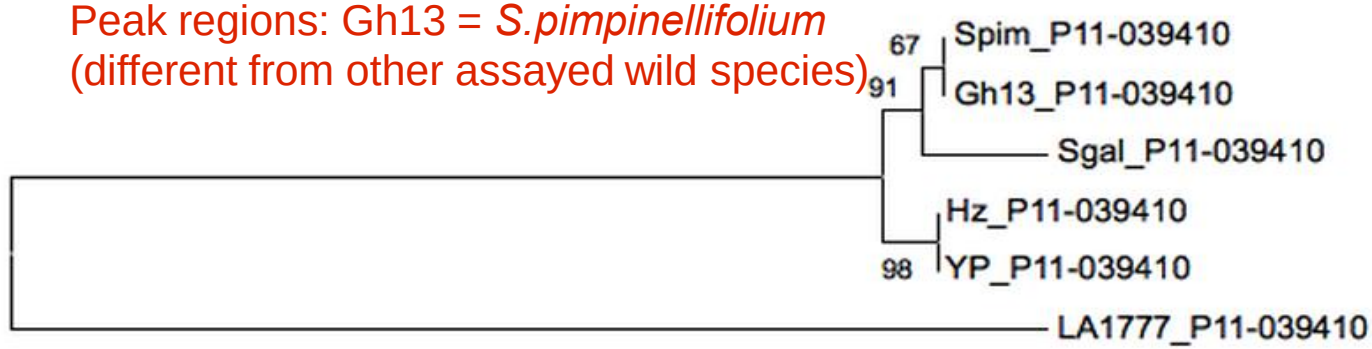
S.pimpinellifolium introgressions?

PCR design: Gh13 chr. 11-
SNPs shared with *S.pimpinellifolium*

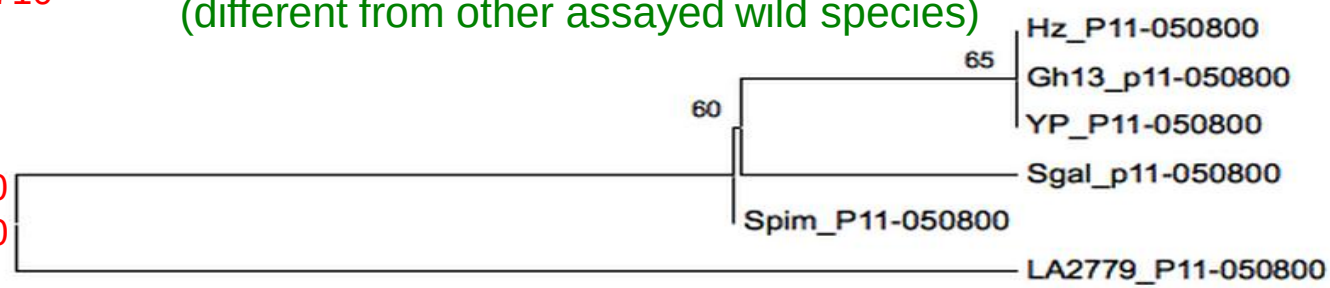
sol genomics network



Peak regions: Gh13 = *S.pimpinellifolium*
(different from other assayed wild species)



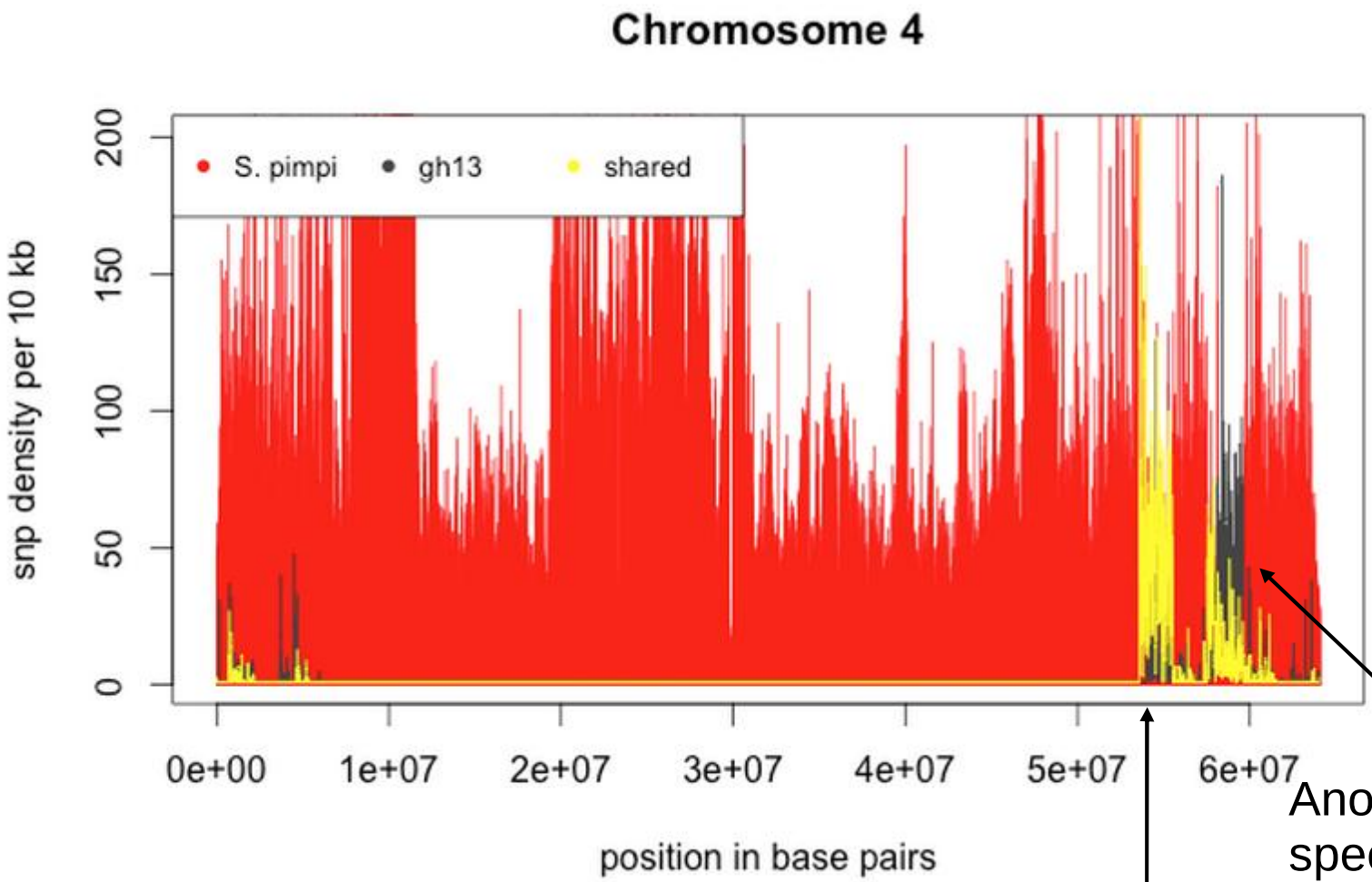
Non-peak regions: Gh13 = *S.lycopersicum*
(different from other assayed wild species)



Boyce Thompson Institute for Plant Research

Gh13 wild introgresions ï more to explore

sol genomics network



Another wild species?

S.pimpinellifolium?

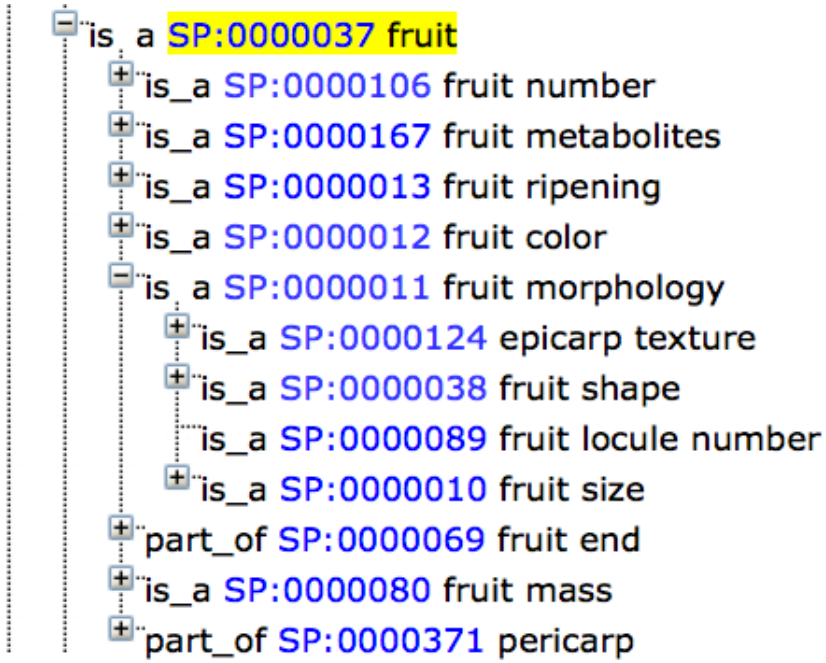
Boyce Thompson Institute for Plant Research

ñWe eat phenotypesò

Phenotypes

Phenotyping is hard

- Labor intensive, expensive
- Standardization of phenotypic measurements
- Ontology-based systems for databases

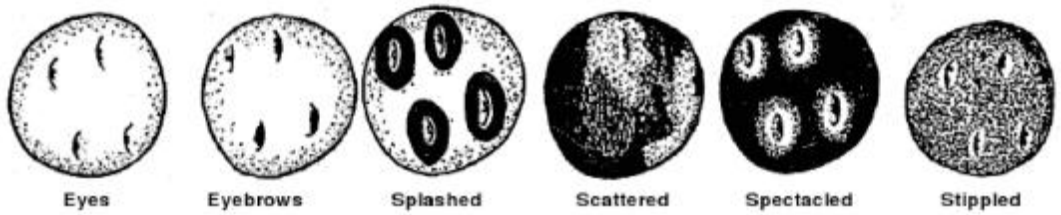


Ontologies

Many ontologies are currently developed:

- http://www.cropontology.org/
  - http://www.bioversityinternational.org/
- http://plantontology.org & PATO

- Difficult to apply one ontology to all plants!



Crop Ontology Curation Tool

http://www.cropontology.org/

Crop Ontology Curation Tool

Generation Challenge Programme
CULTIVATING PLANT DIVERSITY FOR THE RESOURCE POOR

Home | About | Users | Feedback

Plant Trait Ontology and Crop Annotation Workshop, at Oregon State University, US, 12th-15th September 2012 - consult the wiki!

Search

Add New Terms | API | Help | Agtrials | Annotation Tool | Register | Login

Latest

OBO Ontology | Trait Dictionary

General Germplasm Ontology

FAO/IPGRI Multi-Crop Passport Descriptor 88 terms BIOVERSITY
FAO/IPGRI Multi-Crop Passport Descriptor

Germplasm 386 terms SHRESTHA
germplasm

ICIS germplasm method 166 terms SHRESTHA
ICIS germplasm methods

Identifier Test Ontology 58 terms SHRESTHA
Test ontology for identifier functionality.

Phenotype and Trait Ontology

Cassava 126 terms BAKARE
Cassava Trait Ontology

Chickpea 253 terms PRASAD
Chickpea Traits

Common Bean 437 terms AGUERRERO

Location and Environmental Ontology

Country and Location 1118 terms SHRESTHA
Describes official ISO 3166-1 alpha-2, alpha-3 and numeric country codes along with location names.

Crop Research 256 terms SHRESTHA
Describes experimental design, environmental conditions and methods associated with the crop study/experiment/trail and their evaluation.

Plant Anatomy & Development Ontology

Musa Anatomy 149 terms CHANNELIERE
Musa Anatomy

Sol Genomics Network

http://solgenomics.net/stock/6799/view

sol genomics network

home | forum | contact | help | faq

search maps genomes tools sol search

log in | new user

Accession: 313-100

Stock details New QTL population | Back to stock search

[New] [Edit] [Delete]

Organism	**Solanum lycopersicum**
Stock type	**accession**
Stock name	**313-100**
Uniquename	**313-100**
Description	

Stock editors: Esther van der Knaap

Synonyms	None
Pedigree data	None
Additional information	None

Associated loci (0) [log-in to associate new locus]

Experimental data None

Related stocks

Accessions this accession is a member of

Type	Name
f2 population	QTL Tomato Sausage x LA1589 F2

Images (1) [Add new image]

Literature annotation (0) None [Associate publication]

Ontology annotation () [Add ontology annotations]

Phenotype data [Download phenotypes]

Experiment: phenotypes recorded for population QTL Tomato Sausage x LA1589 F2 by Esther van der Knaap

Phenotype data

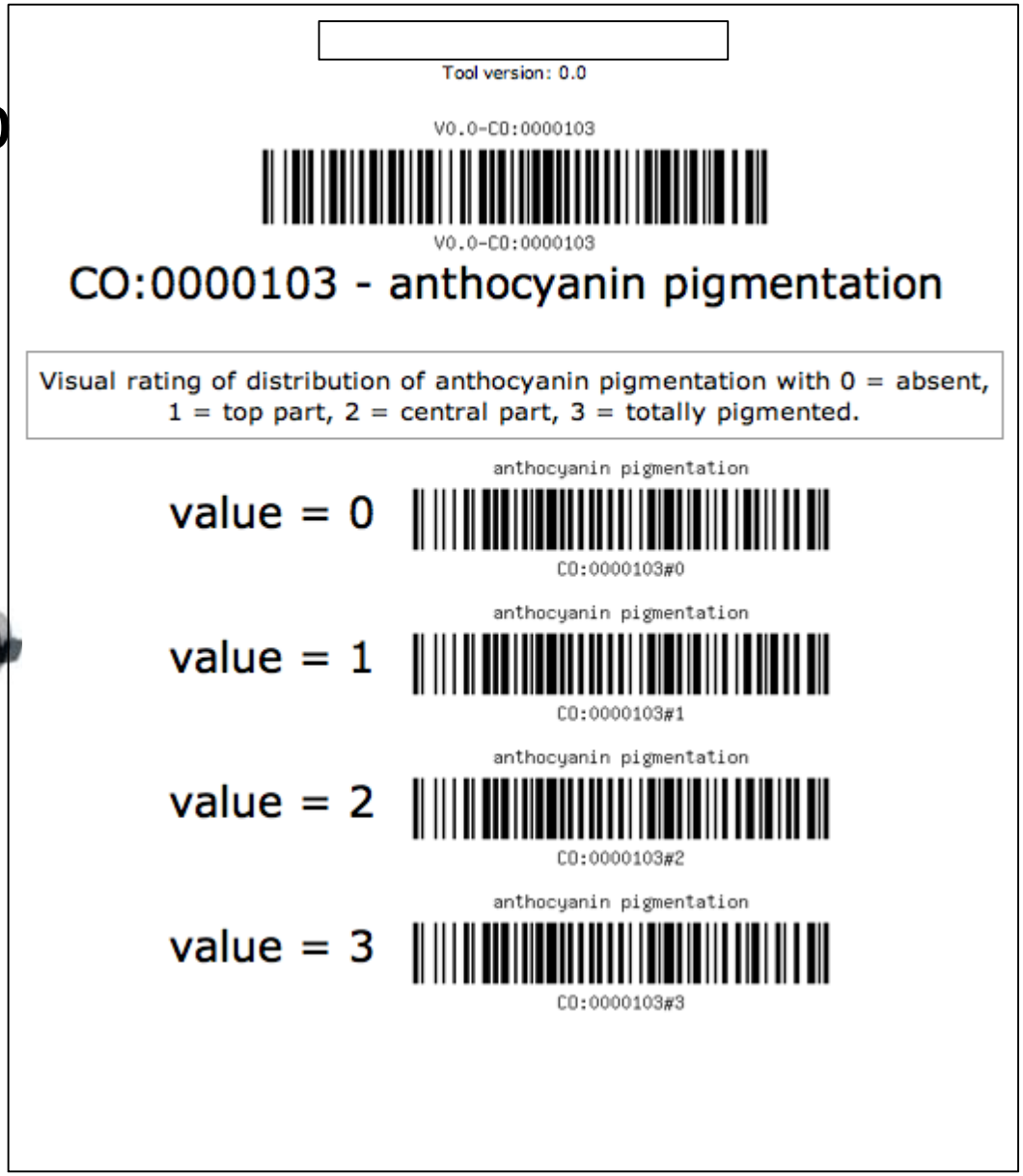
[Download phenotypes]

Experiment: phenotypes recorded for population QTL Tomato Sausage x LA1589 F2 by Esther van der Knaap

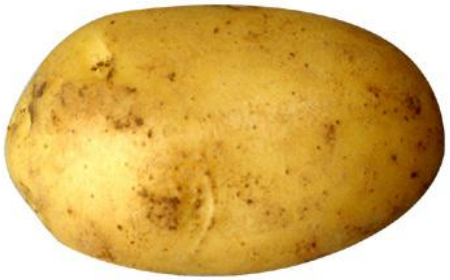
Trait	Average	Min	Max	Lines/repeats
distal angle macro 10% (distal angle macro 10%)	151.30	151.30	151.30	1
distal angle macro 15% (distal angle macro 15%)	131.60	131.60	131.60	1
distal angle macro 20% (distal angle macro 20%)	108.55	108.55	108.55	1
distal angle micro 2% (distal angle micro 2%)	173.07	173.07	173.07	1
distal angle micro 3% (distal angle micro 3%)	168.79	168.79	168.79	1
distal angle micro 5% (distal angle micro 5%)	167.66	167.66	167.66	1
distal eccentricity index (distal eccentricity index)	0.98	0.98	0.98	1
distal fruit end blockiness 10% (distal fruit end blockiness 10%)	0.60	0.60	0.60	1
distal fruit end blockiness 20% (distal fruit end blockiness 20%)	0.79	0.79	0.79	1
distal fruit end blockiness 30% (distal fruit end blockiness 30%)	0.91	0.91	0.91	1
distal fruit end blockiness 5% (distal fruit end blockiness 5%)	0.44	0.44	0.44	1
distal fruit end indentation (distal fruit end indentation)	0.00	0.00	0.00	1
distal fruit end protrusion (distal fruit end protrusion)	0.00	0.00	0.00	1
eccentricity area index (eccentricity area index)	0.09	0.09	0.09	1
fruit area (fruit area)	35566.63	35566.63	35566.63	1
fruit length mid-width (fruit length mid-width)	202.00	202.00	202.00	1
fruit longest length (fruit longest length)	204.25	204.25	204.25	1
fruit mid-height width (fruit mid-height width)	215.50	215.50	215.50	1
fruit perimeter (fruit perimeter)	708.60	708.60	708.60	1
fruit shape circular (fruit shape circular)	0.99	0.99	0.99	1
fruit shape eccentric (fruit shape eccentric)	0.97	0.97	0.97	1
fruit shape ellipsoid (fruit shape ellipsoid)	0.99	0.99	0.99	1
fruit shape index external (fruit shape index external)	0.94	0.94	0.94	1
fruit shape index external 2 (fruit shape index external 2)	0.94	0.94	0.94	1

Trait scoring

- Use ñbarcode toolsò



Sol CAP



Tomato panel (~400 accessions)
Incl. Processing, fresh market, heirloom, wild relatives

Potato panel (~400 accessions)

Phenotyping for breeder traits:
Tomato Analyzer
Fruit shape
Color
PH
Brix
Vitamin C
Lycopene
sugars

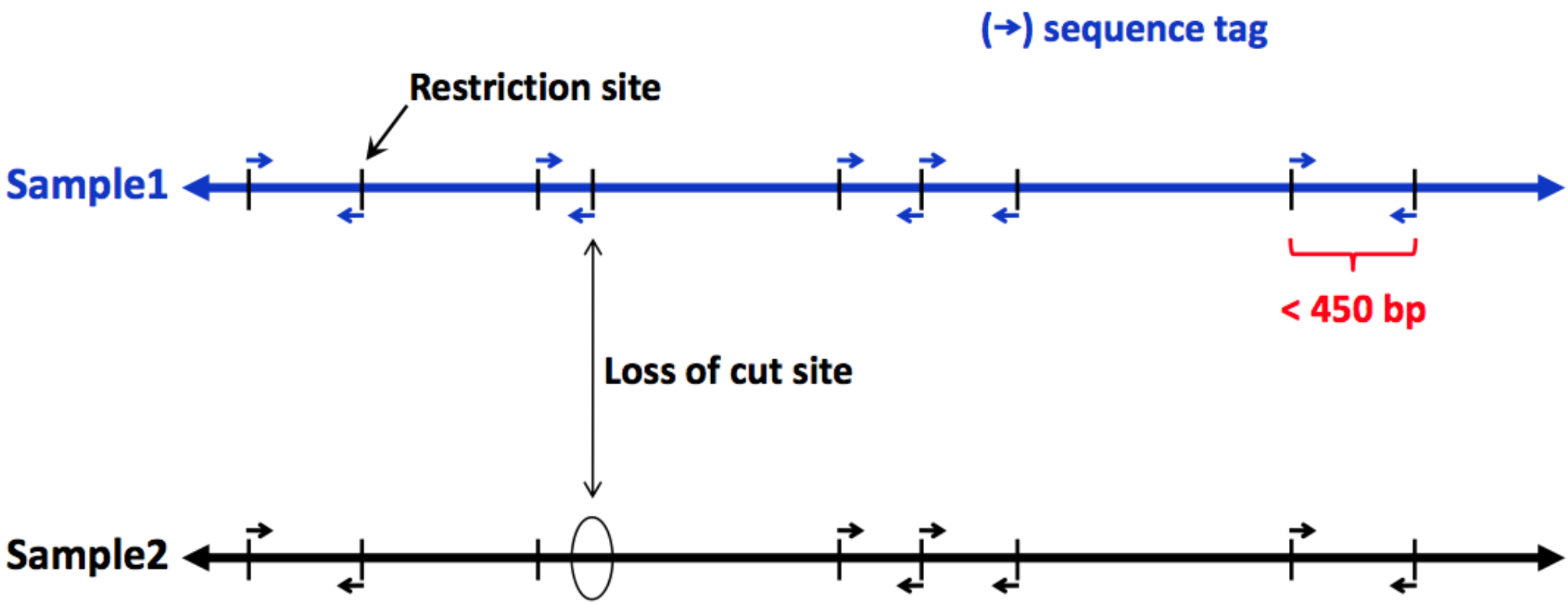
Genotyping
(Illumina Infinium chip)

Phenotyping:
Specific gravity
chip color after cold storage
sucrose/glucose
Skin texture
tuber shape(l/w/h)
eyedepth
skincolor
Flower color
Flesh color
growth habit
total yield etc.

Genotyping
(Illumina Infinium chip)

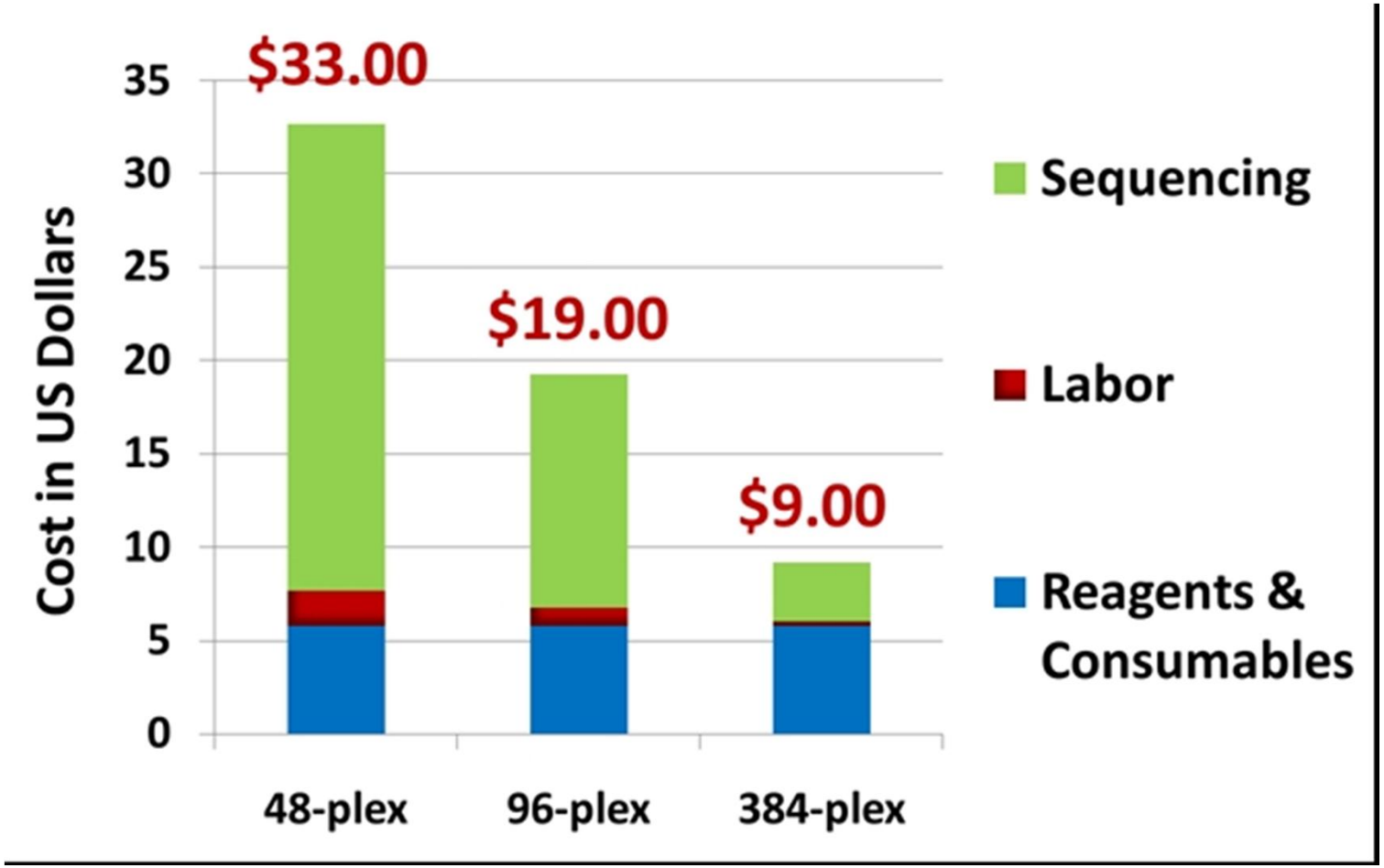
Genotyping by Sequencing (GBS)

- Developed by Buckler lab (Elshire, 2011)
- Full genome sequencing too expensive
- Reduce sequence space using restriction
- Use highly multiplexed NGS approach



- Focuses NextGen sequencing power to ends of restriction fragments
- Scores both SNPs and presence/absence markers

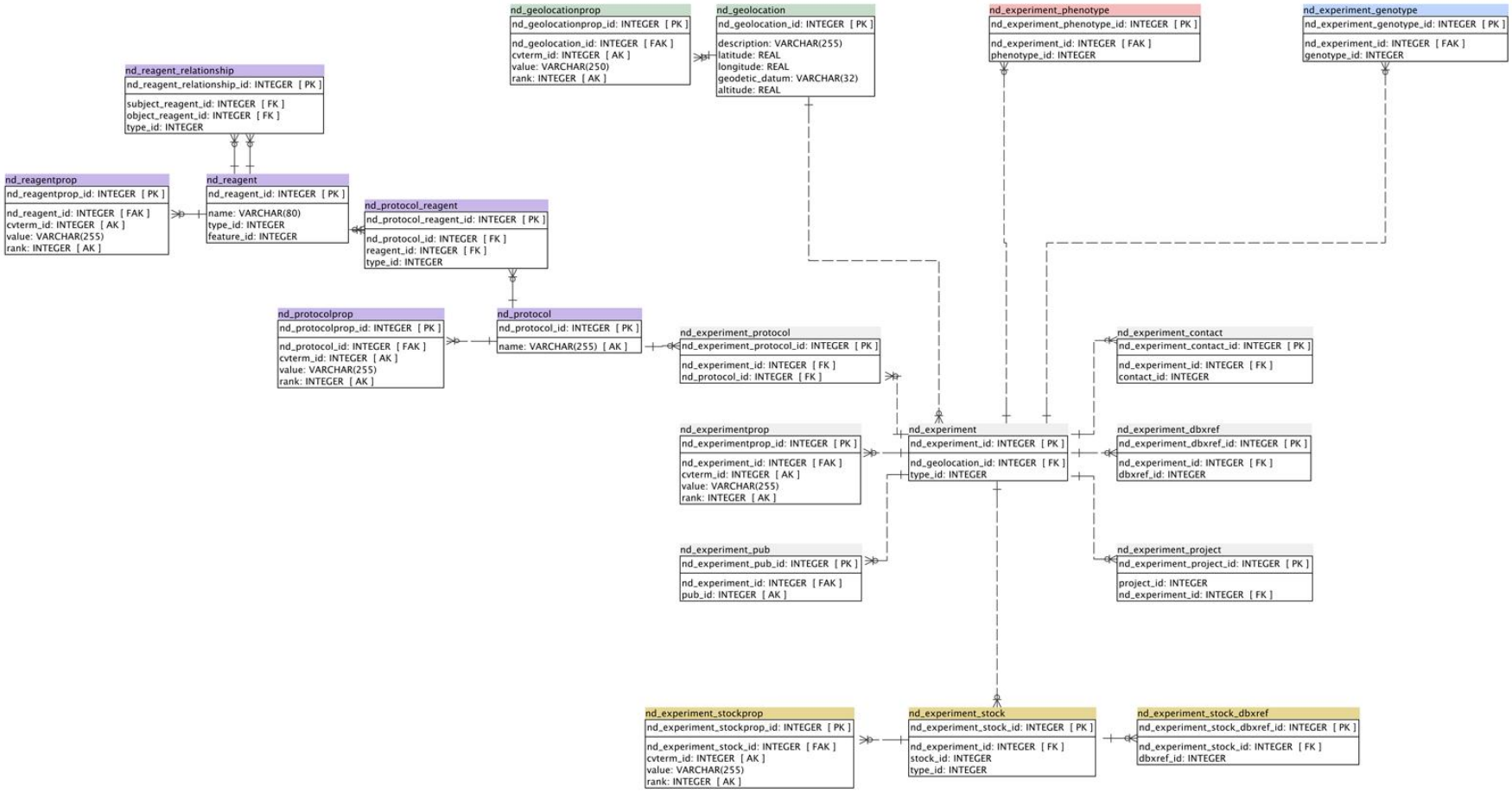
Genotyping by Sequencing (GBS)



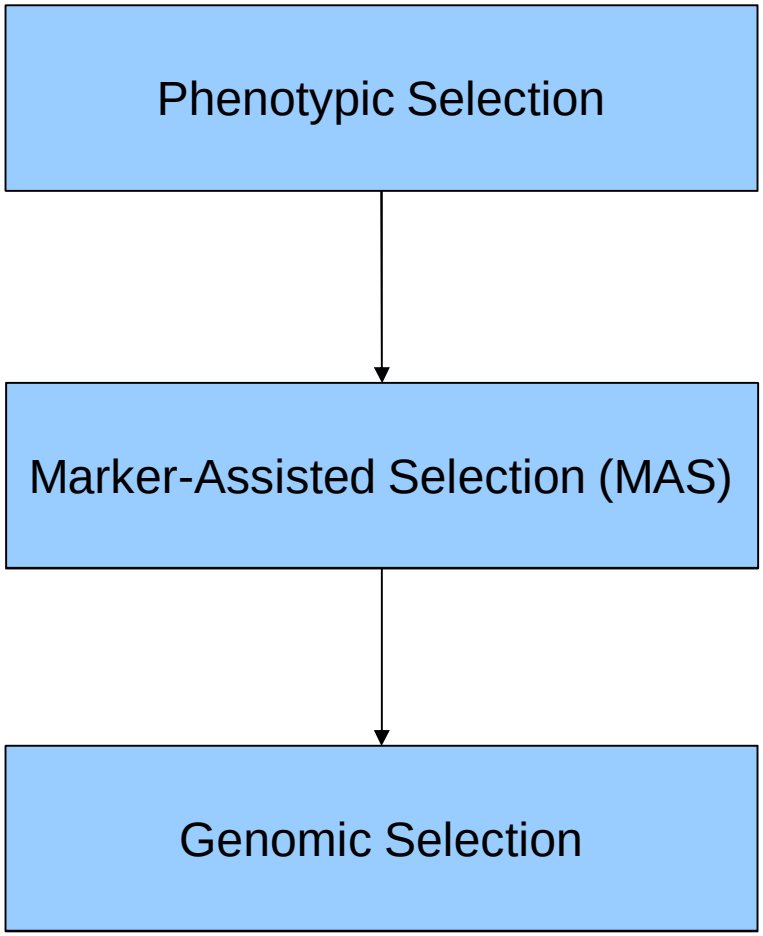
Storing genotypic data

- Challenge: Extremely voluminous
- 50,000 plants 20,000 markers = 1,000,000,000 datapoints
- Special techniques are needed to store data
  - Relational databases: Compress genotype data into strings
  - Non-relational databases: HDF5

Chado Natural Diversity Schema



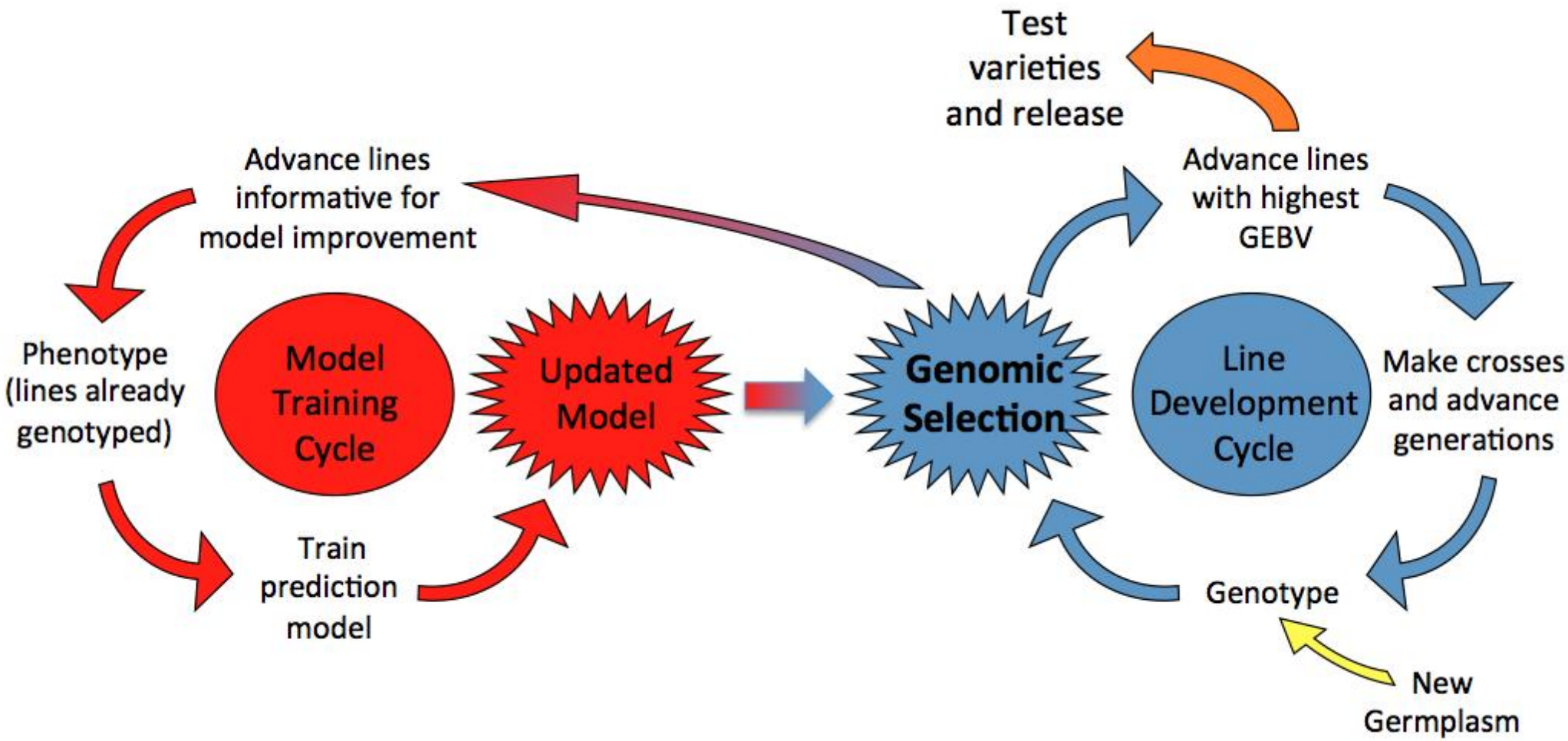
Breeding technologies



Genomic Selection

- Remove phenotyping from line development
- Use markers to model genetic relatedness between lines.
  - Use relatedness estimates to make predictions
- Use markers as predictors in regression-type models
  - Use estimated marker effects to make selections

Genomic Selection



(Jean-Luc Jannink)

Integrate Breeding functions

- Store genotypes and phenotypes in the database
  - Calculation of GS models
  - Prediction of phenotypes
- Manage breeding process:
  - Crosses
  - Pedigree tracking
  - Field planting
  - Sample collection
  - Data collection

Sol Genomics Network

http://solgenomics.net/breeders/home

sol genomics network

home | forum | contact | help | faq

search maps genomes tools sol search

Lukas Mueller (log out) **My SGN**

Breeder Tools

Trials

Locations

unknown	(0 plots)
OSU-OARDC Fremont, OH	(19739 plots)
Tidewater. Plymouth, NC	(0 plots)
UofI R&E Center, Aberdeen, Idaho	(404 plots)
Campbell's Soup Company	(10930 plots)
Mills River, North Carolina	(3041 plots)
Hutchinson Drive, Davis CA	(8921 plots)
University of Florida/ IFAS Gulf Coast Research & Education Center	(5777 plots)

Add new location

Crosses

Add new cross

Upload cross file

View all crosses

Phenotypes

Upload

Phenotype search

Accessions & plots

List of accessions:

Something wrong? Report a problem

SGN is supported by the NSF (#0116076), USDA CSREES and hosted at the Boyce Thompson Institute.

Conclusions

- Genome databases need to adapt to the needs of breeders
- Genomic technologies applicable to improvement of the breeding process
  - Genotyping by Sequencing
  - Genomic Selection
- Bioinformatics infrastructure required
  - Genome, phenome, & genotypic information, algorithms, breeder functions