Tomato Breeder's Roundtable 2013, Chiangmai, Thailand

EAST-WEST SEED

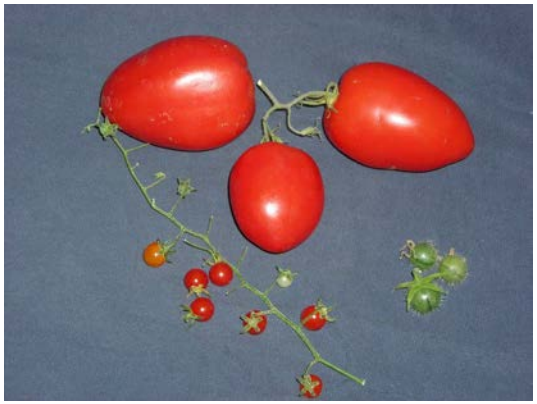
Population Level Analysis to Move from Massive Sequence Data Sets to Application

David Francis, Sung-Chur Sim , Heather Merk ,
Horticulture and Crop Science, The Ohio State University, Wooster, OH
Allen Van Deynze & Kevin Stoffel ,
Seed Biotechnology Center, University of California, Davis, CA
C. Robin Buell and John Hamilton ,
Biology, Michigan State University, East Lansing, MI
David Douches & Dan Zarka ,
Crop and Soil Sciences, Michigan State University, East Lansing, MI

Sol CAP

The Scientific Questions/issues

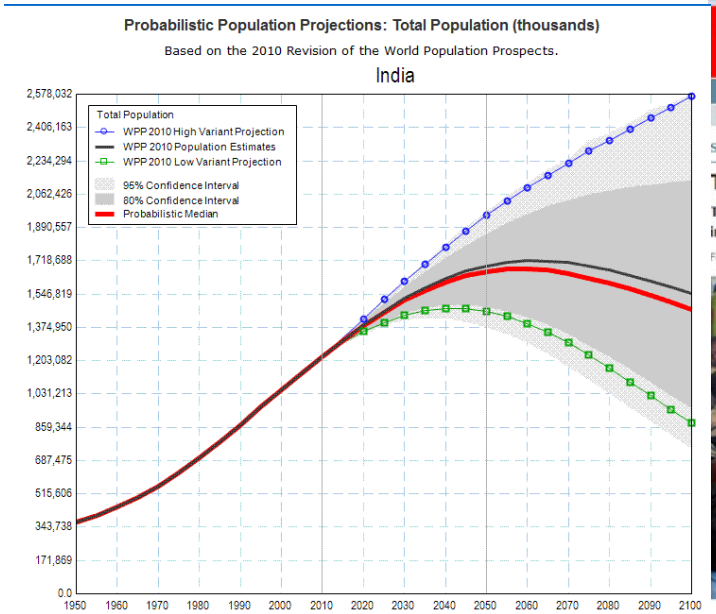
Domestication = Genetic Change

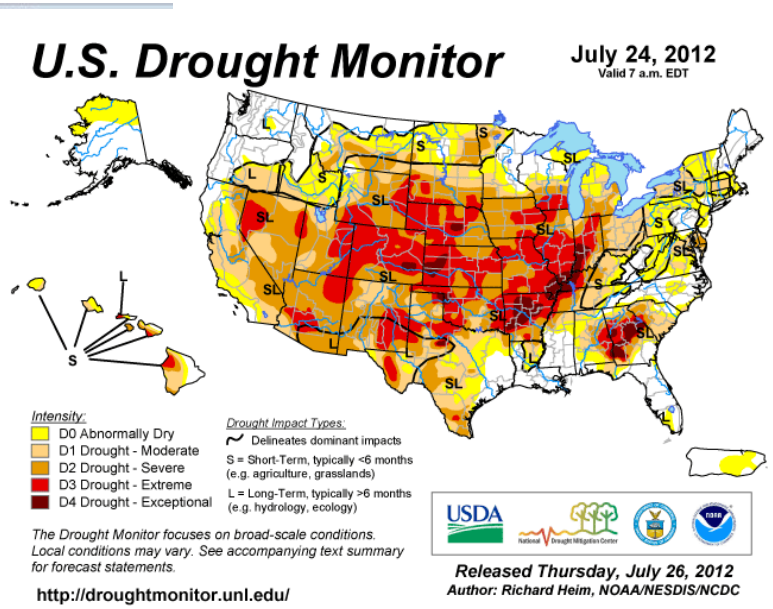


Environmental adaptation = Genetic Change

Market differentiation = Genetic Change

The “real” issue: How do we harness the power of genomics to breed high quality food crops that contribute to sustainable agricultural production and human health?





http://esa.un.org/unpd/ppp/index.htm

Bayesian Probabilistic Population Projections for 2045 = median: 9.0 billion; 95% interval: 7.8-10.3 billion

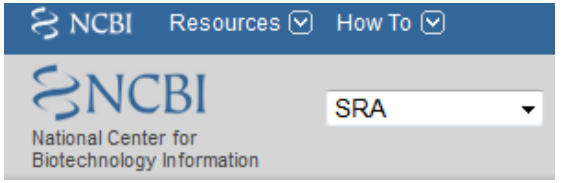
Sequence Resources for Tomato

Community resources

- H1706 genome (Processing)
- LA1589 genome (wild current tomato)
- TA496 EST (Processing)
- Micro-Tom EST (Novelty)

SolCAP resources

- OH068-6405 (fresh-market)
- FL7600 (fresh-market)
- NC84173 (fresh-market)
- OH9242 (processing)
- PI114490 (cherry tomato)
- PI128216 (wild current tomato)



Focus on cultivated types with direct relevance to breeding programs

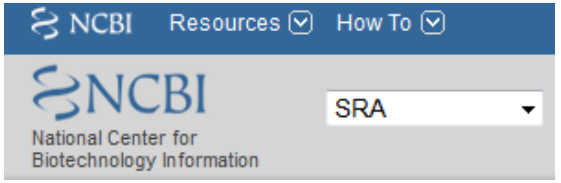
Sol CAP

Sequence Resources for Tomato

Community resources

- H1706 genome (Processing)
- LA1589 genome (wild current tomato)
- TA496 EST (Processing)
- Mi
- FL
- NC84173 (fresh-market)
- OH9242 (processing)
- PI114490 (cherry tomato)
- PI128216 (wild current tomato)

+150 Genomes



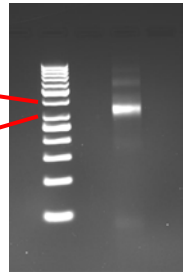
Focus on cultivated types with direct relevance to breeding programs

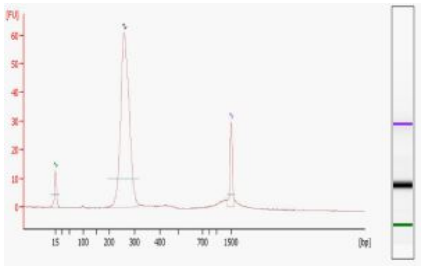
Sol CAP

mRNA Library creation/QC

400

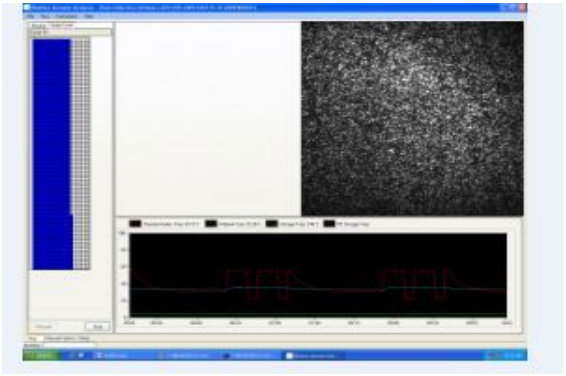
300



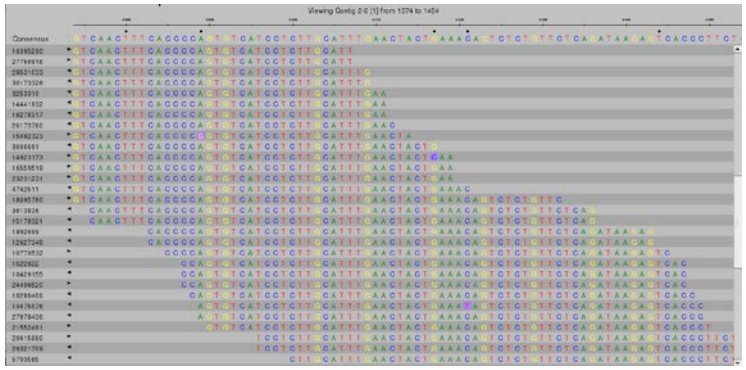


GAII sequencing by synthesis (single and paired end)

Data Collection



Assembly



~2-4 Gb of transcriptome sequence for 6 tomato accessions (normalized libraries)

Detail on sequencing, selection and validation of SNPs

17 Gb of sequence, assembly into 30.6 -34.9 Mb of sequence for six accessions, representation of 55.3 -59.6% of the predicted tomato gene set

Hamilton et al. 2012. The Plant Genome 2012 5:17-29
https://www.crops.org/publications/tpg/articles/5/1/17

Sequence reads deposited into NCBI Short Read Archive (SRA)

Accession: SRX111862
Accession: SRX111861
Accession: SRX111859
Accession: SRX111858
Accession: SRX111857
Accession: SRX111853
Accession: SRX111850
Accession: SRX111849
Accession: SRX111848
Accession: SRX111845
Accession: SRX111558
Accession: SRX111557
Accession: SRX111556

Total: 1 run, 18.5M spots, 1.6G bases, 1.1Gb

#	Run	# of Spots	# of Bases	Size
1.	SRR390340	18,500,235	1.6G	1.1Gb

#	Run	# of Spots	# of Bases	Size
1.	SRR390335	8,211,421	895M	526.9Mb

Sol CAP

Single Nucleotide (SNP) Discovery "Pipeline"

Align RNA-Seq reads to tomato scaffolds using Bowtie and call SNPs using SAMTools: **3,730,151 SNPs**

↓

Filter SNPs for read depth and density using the SAMTools utility samtools.pl varFilter: **113,493 SNPs**

↓

Filter SNPs for min SNP quality, consensus quality, mapping quality, and max percentage 3' bases at SNP positions: **86,556 SNPs**

↓

Create list of unique quality filtered SNP positions on tomato scaffolds
57,996 SNPs
(High Confidence SNPs)

Remove multi-allelic SNPs and SNPs within 50bp of an intron:
33,394 SNPs

↓

Filter out SNPs with a BeadXpress score < 0.75 and an Infinium score < 0.9 : **27,238 SNPs**
(Illumina compatible SNPs)

↓

Filter based on frequency among 8 accessions; add spatial criteria; cross validated w/ INRA and Trait Genetics

↓

8,784 SNPs (**7,720 SNPs** passed manufacture/ production QC and are "scoreable")

illumina

Hamilton et al., The Plant Genome 2012 5:17-29

Tomato Infinium SNP Array

- 522 community candidate genes
- Highly filtered
- 3,503 PM (Sl x LA716)
- 4,491 SNPs (Sl x Spimp)
- 7,720 total & 6,167 SNPs in cultivated (~758 Mb)

SNP design/annotation file: **http://solcap.msu.edu/**
Cluster File: **http://www.extension.org/pages/61007**

extension
Putting research into practice

Sol CAP

TraitGenetics

INRA

illumina

Genotyping: distribution of variation in cultivated tomato

Accessions

- 144 Fresh Market
- 144 Processing
- 2 Greenhouse
- 40 land races
- 3 unimproved
- 48 vintage (heirloom)
- 48 *cerasiformae*
- 64 wild species
- 11 hybrids

Parents of key resource Populations

- M82 x LA716
- Hunt 100 x LA407
- NC 23E X LA1269
- Sun1642 x LA1589
- OH 88119 x PI128216
- OH 8245 x PI365914
- (Fla7600 xPI114490) xOH9242
- OH 88119 x Ha 7998
- OH 9242 x OH 8245
- OH 9241 x OH 8245

Genotyping: distribution of variation in cultivated tomato

Accessions

- 144 Fresh Market
- 144 Processing
- 2 Greenhouse
- 40 land races
- 3 unimproved
- 48 vintage (heirloom)
- 48 *cerasiformae*
- 64 wild species
- 11 hybrids

Parents of key resource Populations

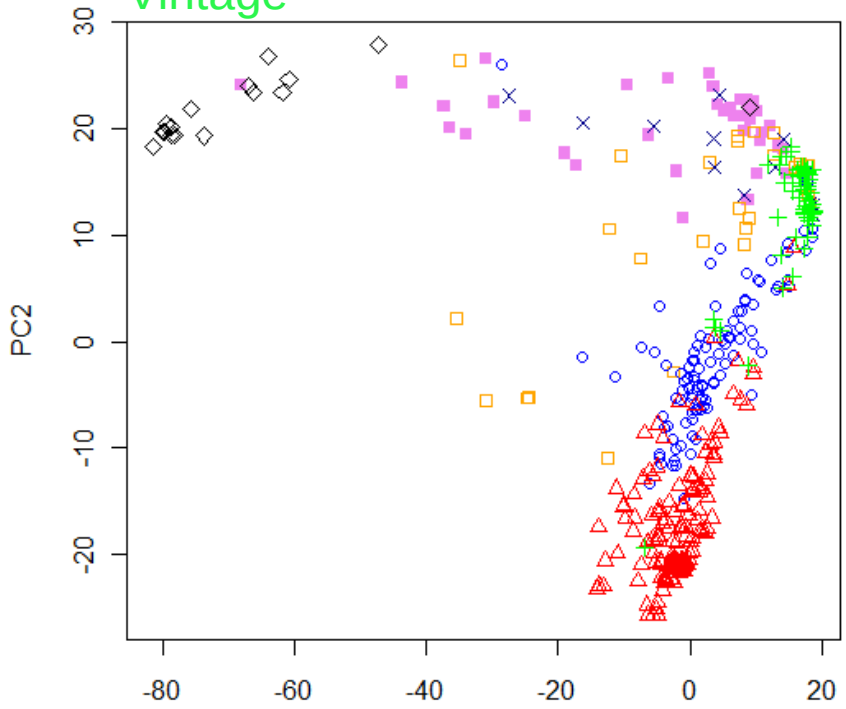
- M82 x LA716
- Hunt
-
-
-
- OH 9242 x OH 8245
- OH 9241 x OH 8245

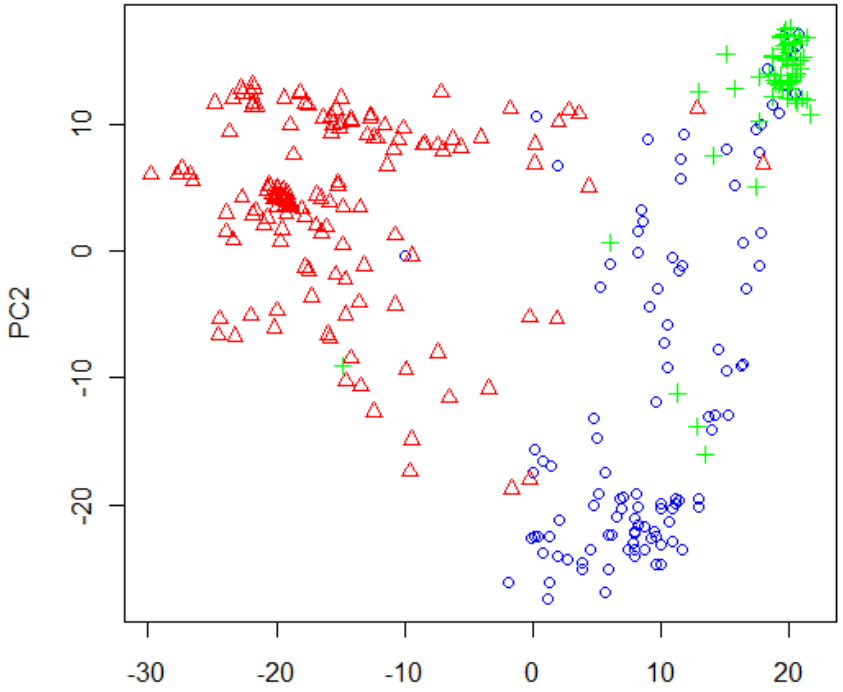
Selection of germplasm to sequence and genotype is where hypotheses are stated

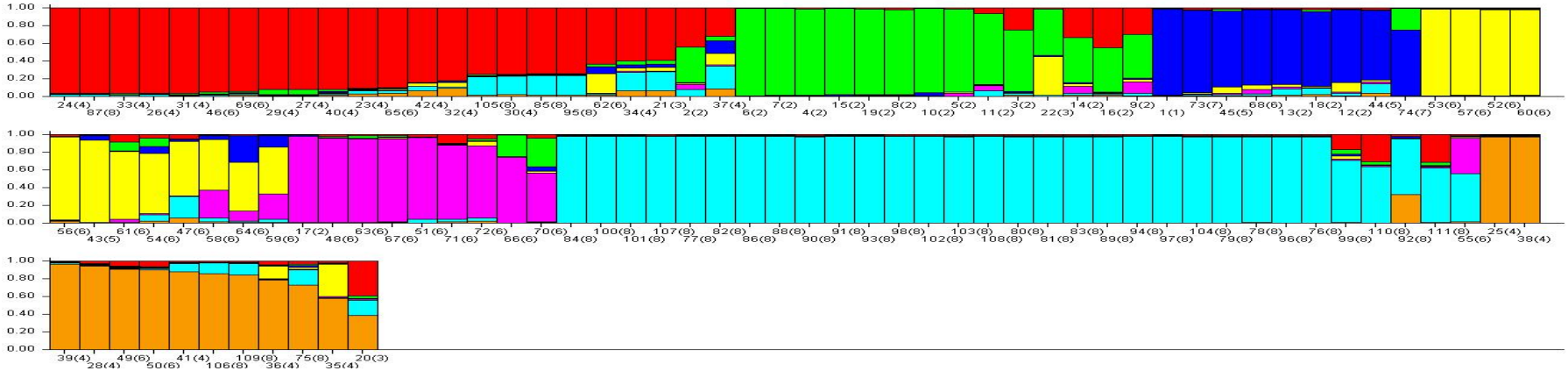
S. pimpinellifolium, Cherry,

(A) cerasiformae, LR, FM, Processing, Vintage

(B) Vintage, FM, Processing







Population analysis

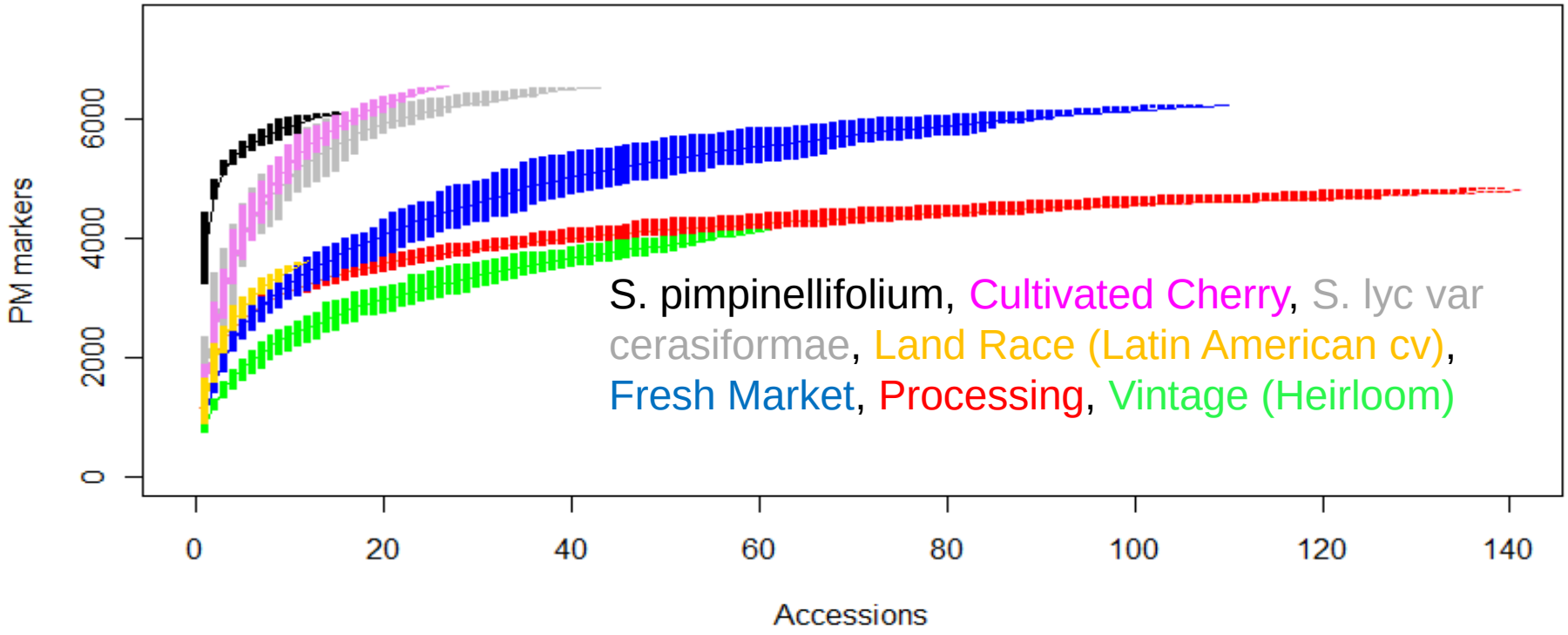
386 cultivated varieties

Original classification	Genetically defined Sub-populations
Processing	Arid Processing
	Humid Processing
Fresh-Market	Arid Fresh-Market
	Humid Fresh-Market
Vintage (Heirloom)	Large-fruited vintage
	Cultivated Cherry
	Landrace

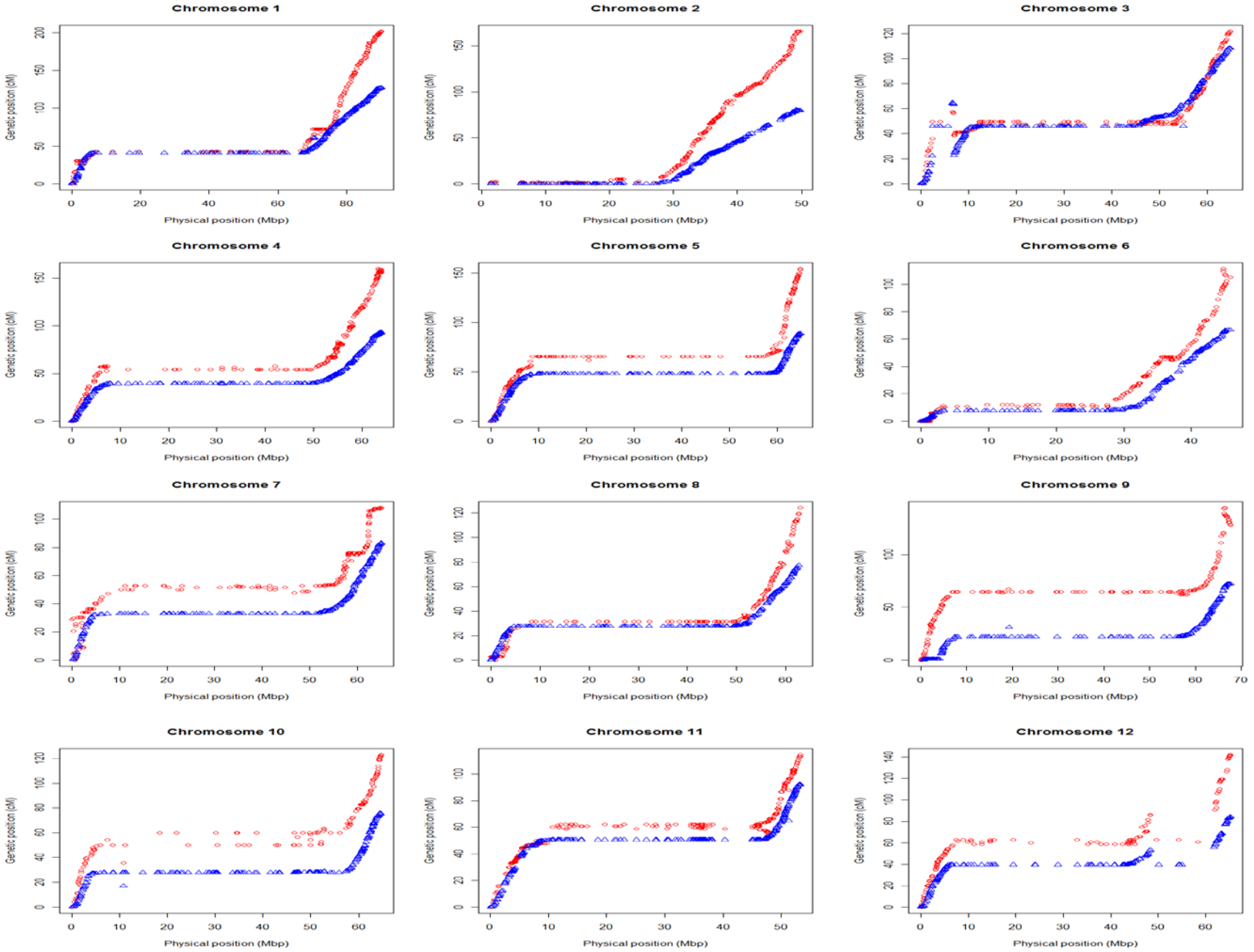
Supported by PCA, STRUCTURE, F_{ST}, Patterns of LD decay

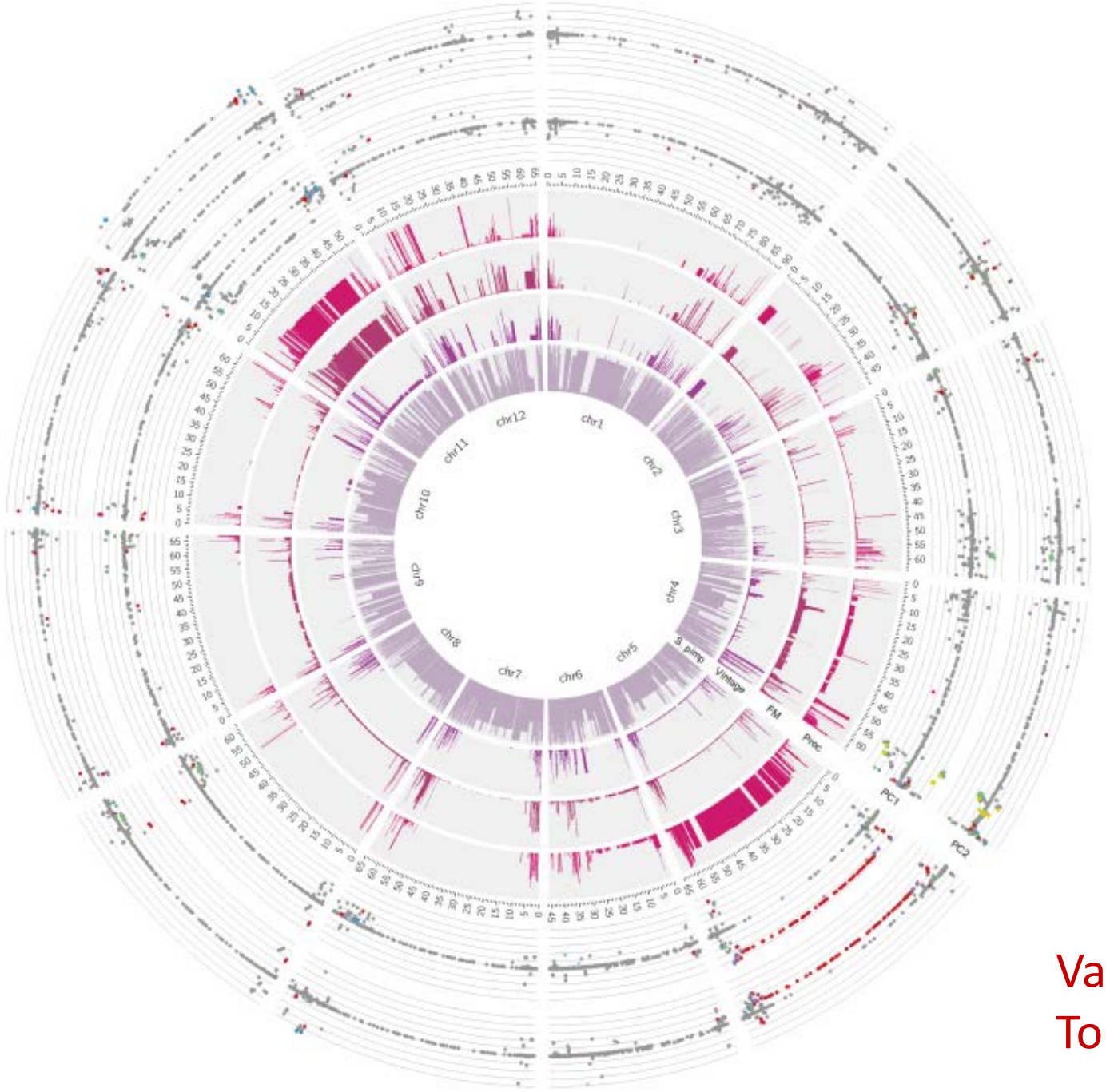
Rarefaction analysis - used in this context to identify the number of variable SNP sites based on the number of accessions sampled

Conclusion: contemporary tomato varieties have more genetic diversity than vintage varieties (F_{ST}, LD decay curves, STRUCTURE support sub-population classifications; PM rate, allelic diversity supports statements about “genetic diversity”).



Comparison of genetic maps with sequence assembly





MAF

Processing

Fresh Market

Vintage

S. pimpinellifolium

Loci under selection

PC1 & PC2 Lositan

No Positive

Proc vs Vin

FM vs Vin

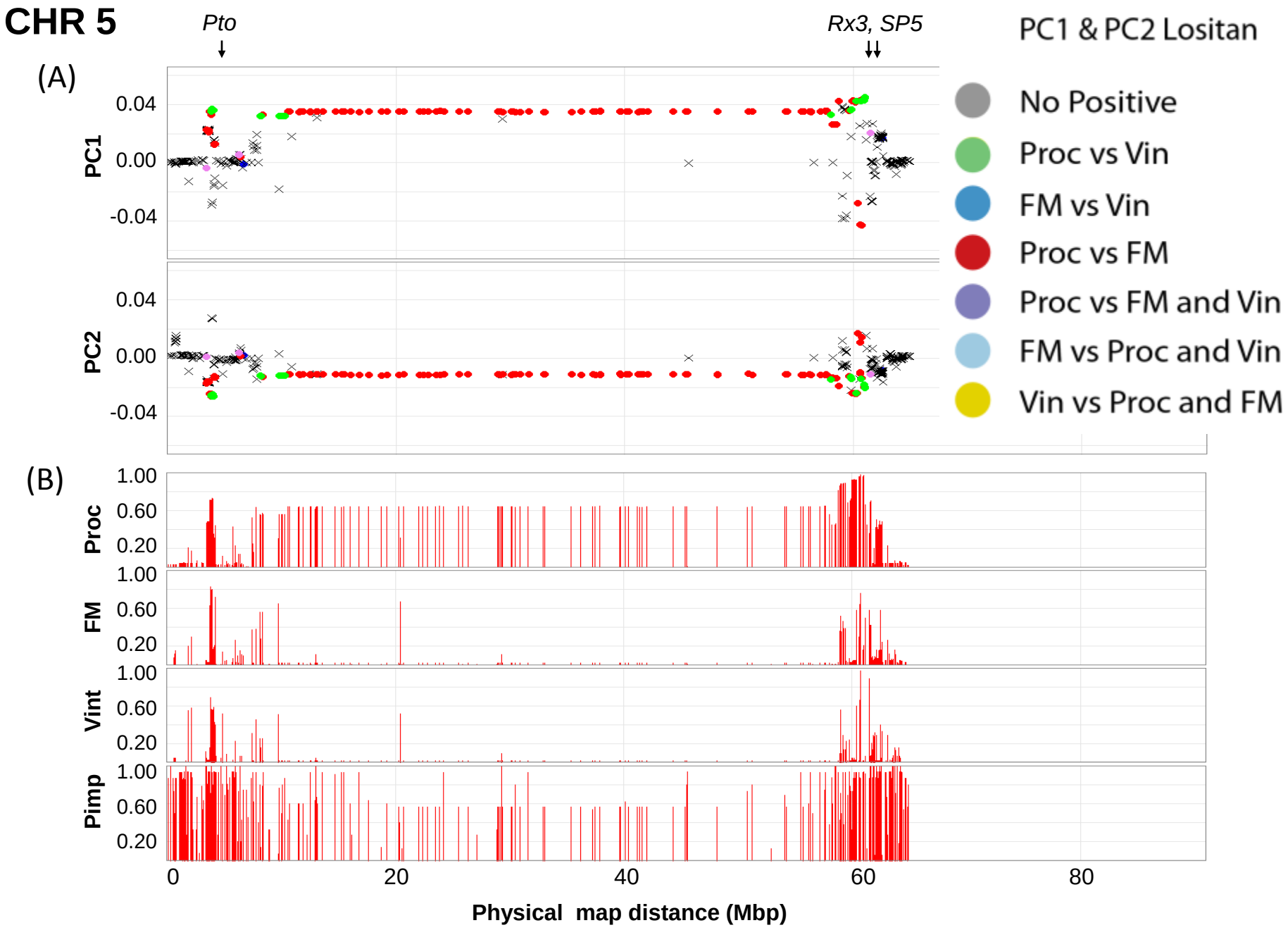
Proc vs FM

Proc vs FM and Vin

FM vs Proc and Vin

Vin vs Proc and FM

Variation in the Tomato Genome

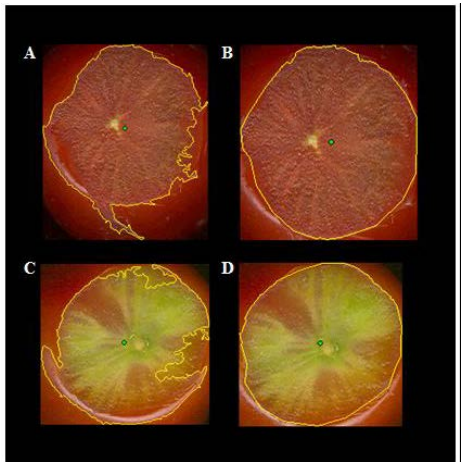

CHR 5
Pto
Rx3, SP5
(A)
PC1
PC2
0.04
0.00
-0.04
PC1 & PC2 Lositan
No Positive
Proc vs Vin
FM vs Vin
Proc vs FM
Proc vs FM and Vin
FM vs Proc and Vin
Vin vs Proc and FM
(B)
Proc
FM
Vint
Pimp
1.00
0.60
0.20
0
20
40
60
80
Physical map distance (Mbp)

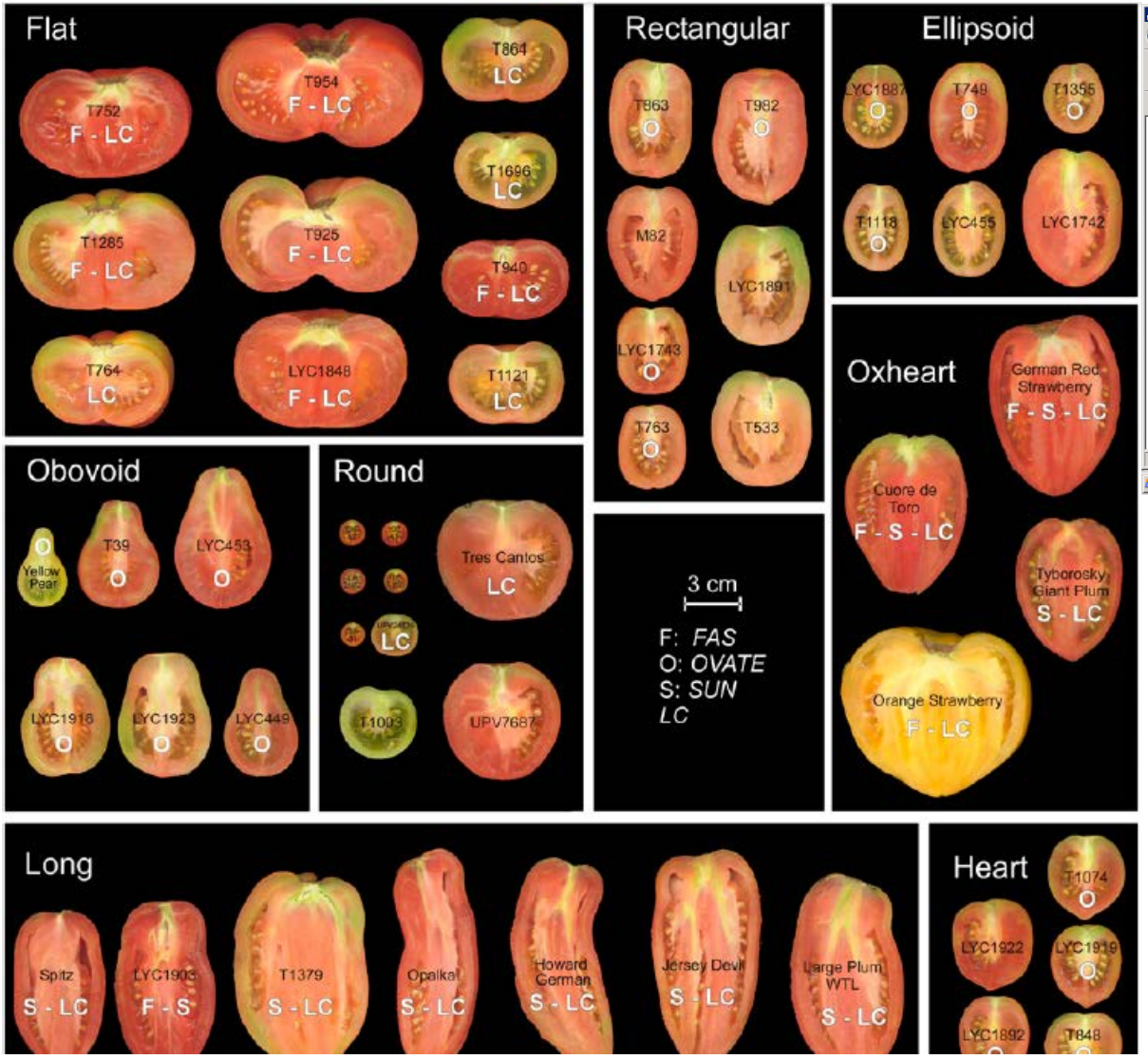
Traits:

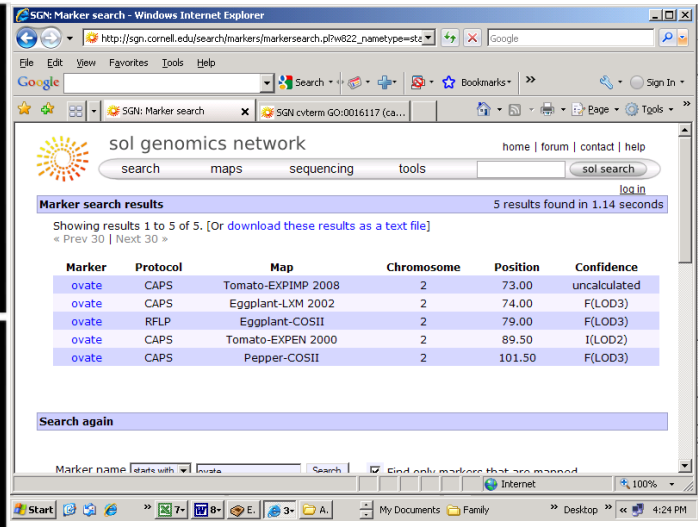
Total traits measured: 52

Reduced to 22 most informative
(h^2, PCA and other methods)

Yield (total and marketable)
Color and Color uniformity
BRIX
pH
Vitamin C
Fruit Size and Shape



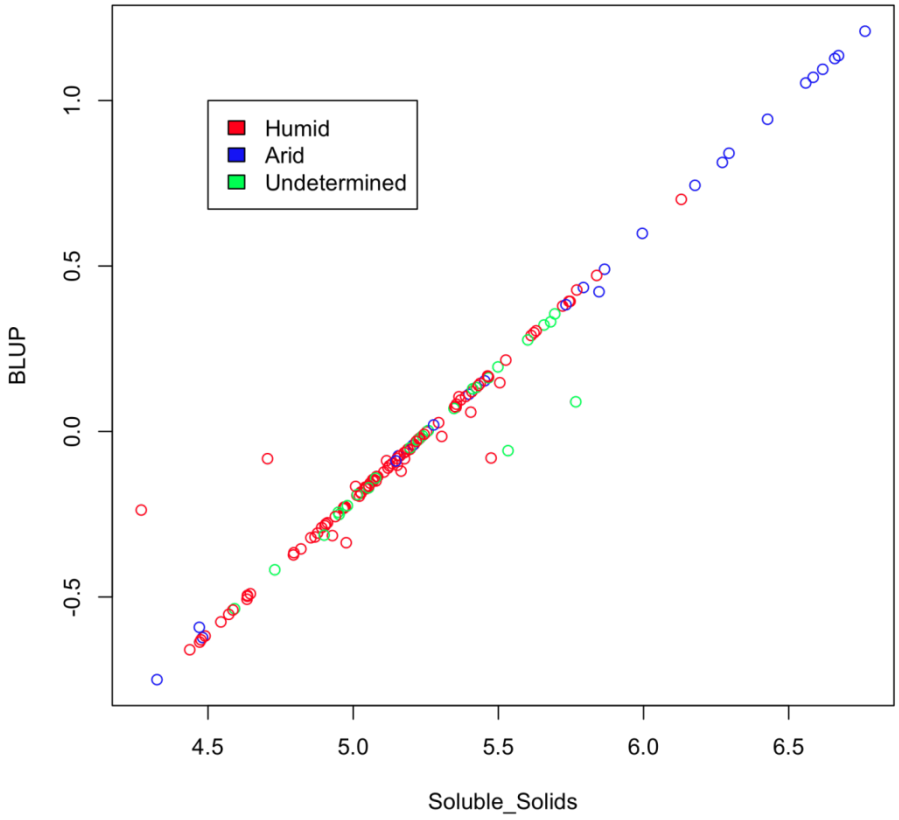


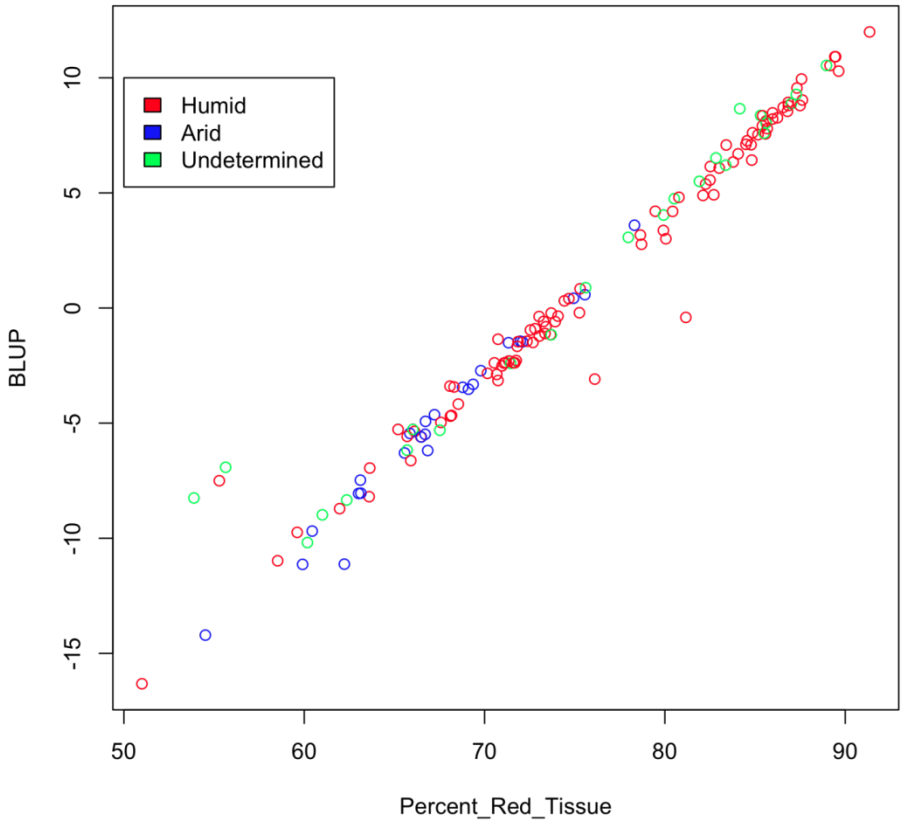


phenotype ontologies are being linked to gene ontologies

Rodríguez et al., 2011
http://www.plantphysiol.org/content/156/1/275

Phenotypic analysis





Evaluation of SolCAP processing tomato collection.

- Comparison of average value with Best Linear Unbiased Predictor (BLUP); color coded by origin of accessions.
- Germplasm from different regional breeding programs contains significant variation for traits of value.

Community projects

Projects

- Cornell, NC State, IU, UFL, PENN State, OH State

SNP optimization

- For most tomato breeding programs, there is insufficient recombination for 7,720 SNPs to be uniquely informative (LD decays over cM).
- Optimized sets of 384 SNPs for Processing & Fresh Market
- PIC in target population, genome coverage (cM and physical)

http://www.extension.org/pages/61007/

Genotyping by sequencing

- reduced representation

illumina

LGC Genomics

incorporating KBioscience

These resources are brought to you by the Cooperative Extension System and your Local Institution

OHIO STATE UNIVERSITY EXTENSION

Local Extension offices near you
Learn more about this institution

Select a different institution

extension
more mind reach

Home | About | Resource Areas | News | Articles | Answers | Learning Lessons

Search

Plant Breeding and Genomics Home

Downstream Analysis of SNPs from the SolCAP Tomato Infinium Array (II)

Last Updated: December 30, 2011

Have a question?
Try asking one of our Experts

Print

Share / Save

This resource area was created by the:
Plant Breeding and Genomics community

Resource Area Feeds

- Track all new content

Authors: Sung-Chur Sim, The Ohio State University
Heather L. Merk, The Ohio State University

SolCAP Cluster File

Please read before accessing the pre-publication tomato cluster file.

The Solanaceae Coordinated Agricultural Project (SolCAP) is pleased to share a pre-publication tomato cluster file for use by research teams representing public and private interests as a resource to facilitate high-throughput genotyping.

The cluster file is based on genotyping the SolCAP tomato collection (solcap.msu.edu/tomato_germplasm_data.shtml). We caution that the cluster file is limited by the germplasm analyzed to date. The SolCAP collection is heavily biased toward inbred lines. Thus heterozygous calls were underrepresented in the data used to create this version. We encourage you to independently verify the genotype calls based on this file.

This cluster file is made available pre-publication with the understanding that users will respect the rights of those who developed the array and cluster file for publication. Consistent with current practices for release of sequence data (www.genome.gov/page.cfm?pageID=10506537), we are making this cluster file available to the community to use for research purposes. By downloading the file, you agree to not publish a global analysis or description of the SolCAP Infinium array prior to its publication by SoLCAP in a peer reviewed journal. Any use of the cluster file should acknowledge and credit the "Solanaceae Coordinated Agricultural Project". SolCAP is in the process of writing a manuscript on the array.
–The SolCAP team

I AGREE (click to download file)

Attachments:

PMsearch_SolCAP.xls (992.5 KB)

FASTAmultiquery_SolCAP.txt (12.71 KB)

ProcessingData_SolCAP.xlsx (10.97 MB)

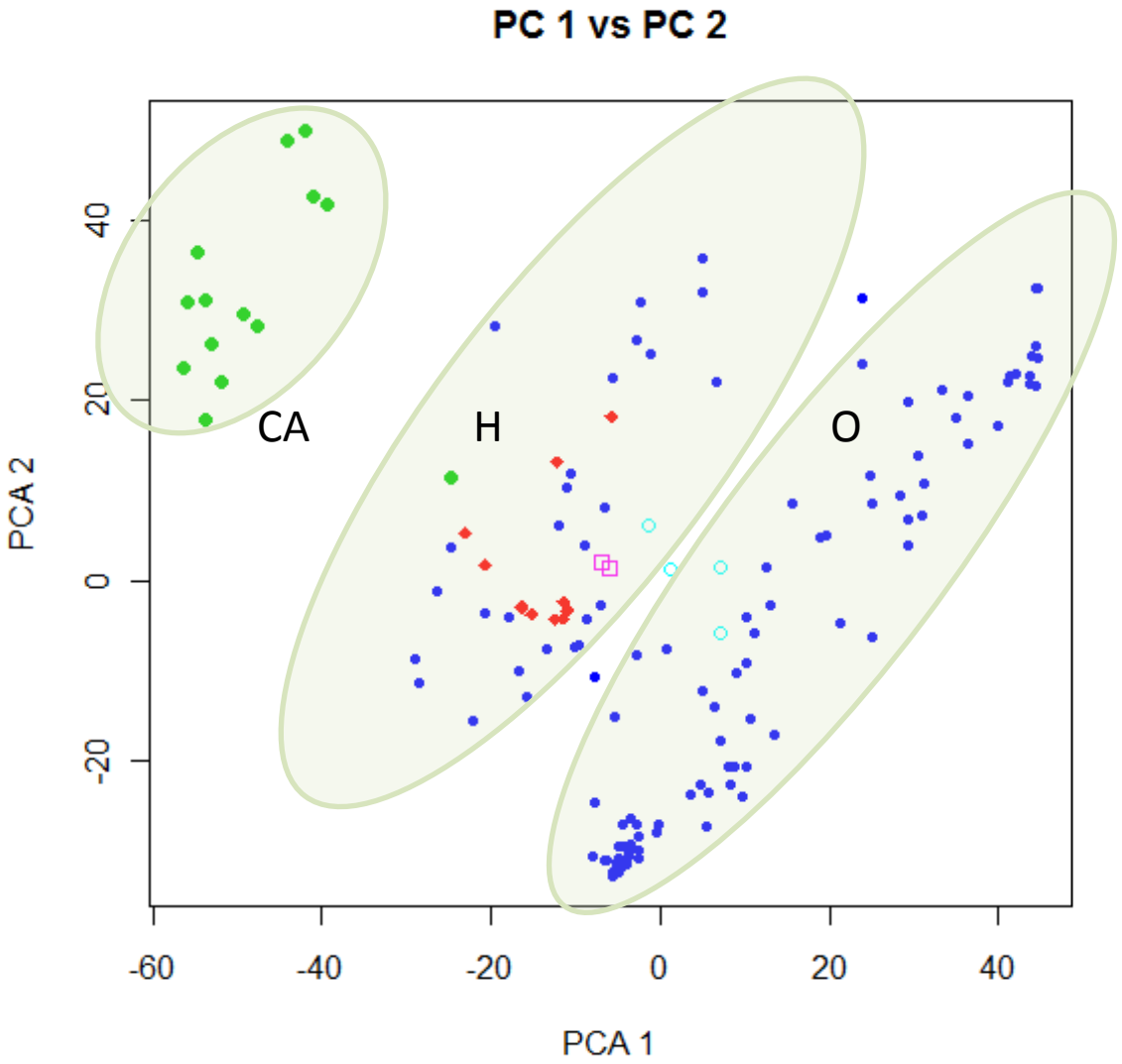
FM_ 384SNPs_SolCAP_10032011.csv (100.21 KB)

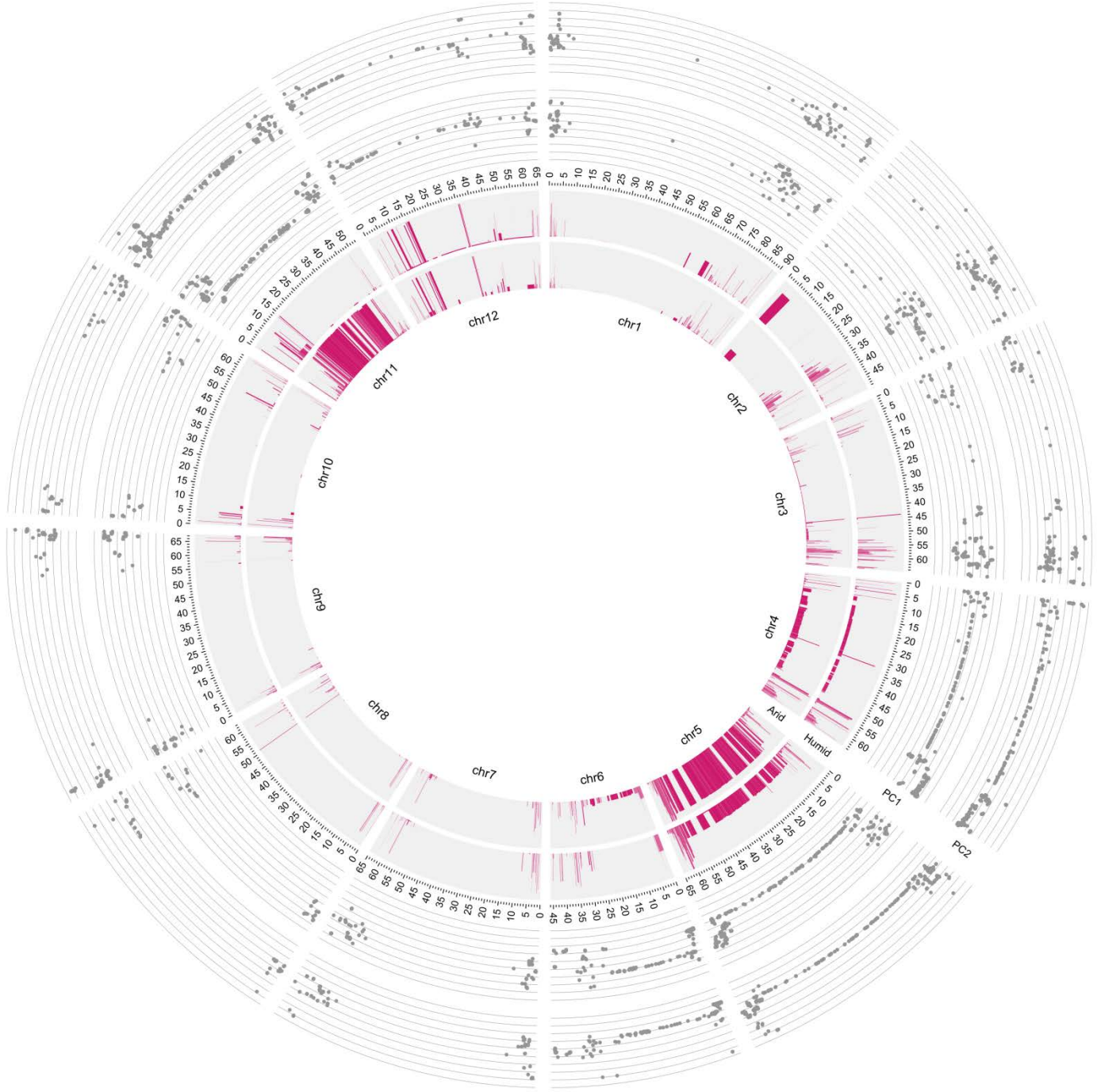
Proc_384SNPs_SolCAP_10122011.csv (100.42 KB)

2011SolCAP Workshop_Sung-Chur_F v2.pdf (2.16 MB)

OH State Community Project

California
Ohio
Undeterimined
Ridgetown College, Ontario
Oregon




chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
Arid
Humid
PC1
PC2

Linking phenotype and SNP in a nested RIL

AxB; CXB; AXD

(OH2641 x OH987034; OH7814 x OH987034; OH2641 x OH981136)

(O x H; O x H, O x CA)

288 progeny

Augmented Experimental design (2 year, 2 locations)

http://www.extension.org/pages/60430/

SNP optimization for “community projects”

For most tomato breeding programs, there is insufficient recombination for 7,720 SNPs to be highly informative.

Optimized sets of 384 SNPs for Processing & Fresh Market

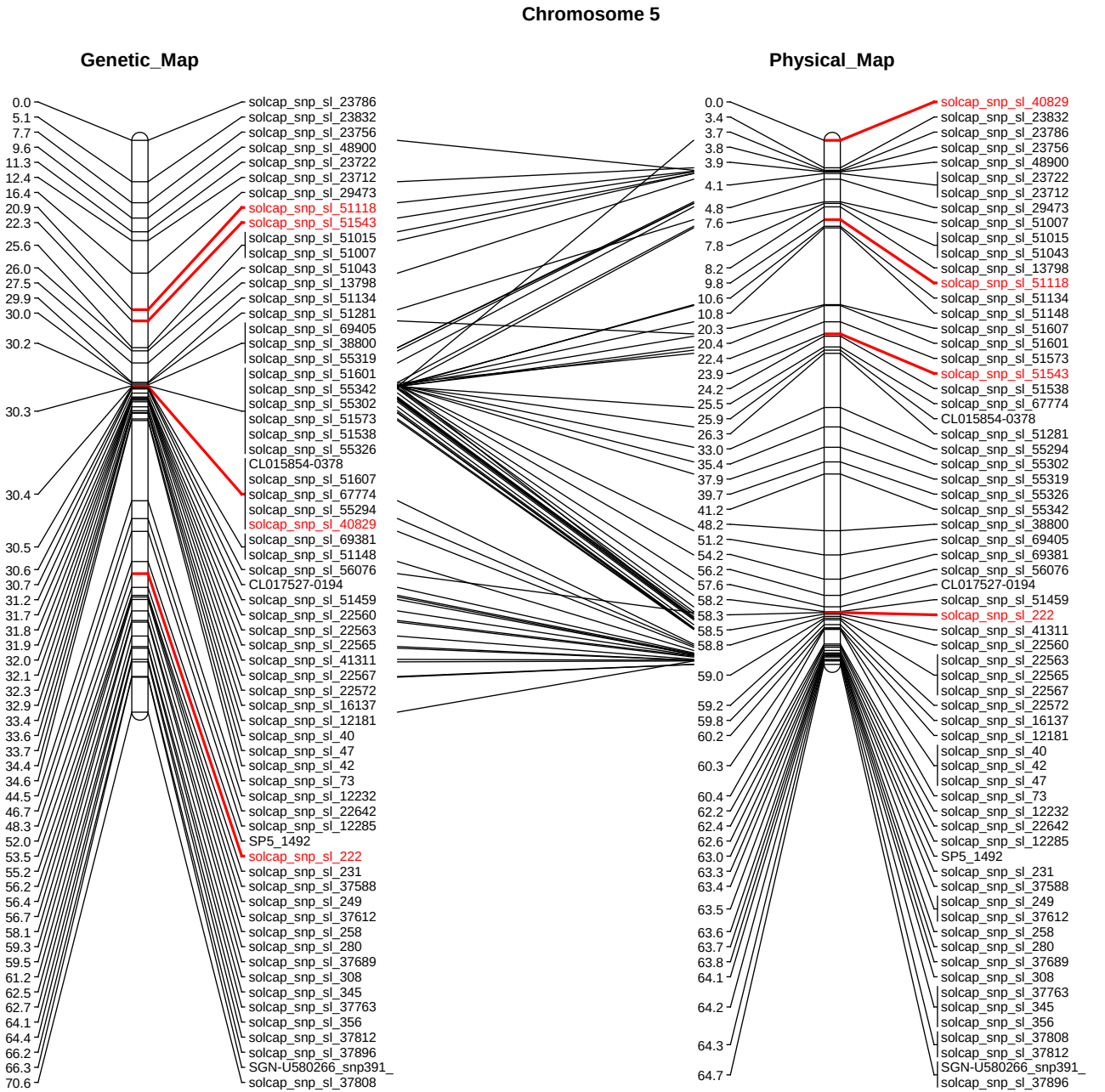
PIC in target population, genome coverage (cM and physical)

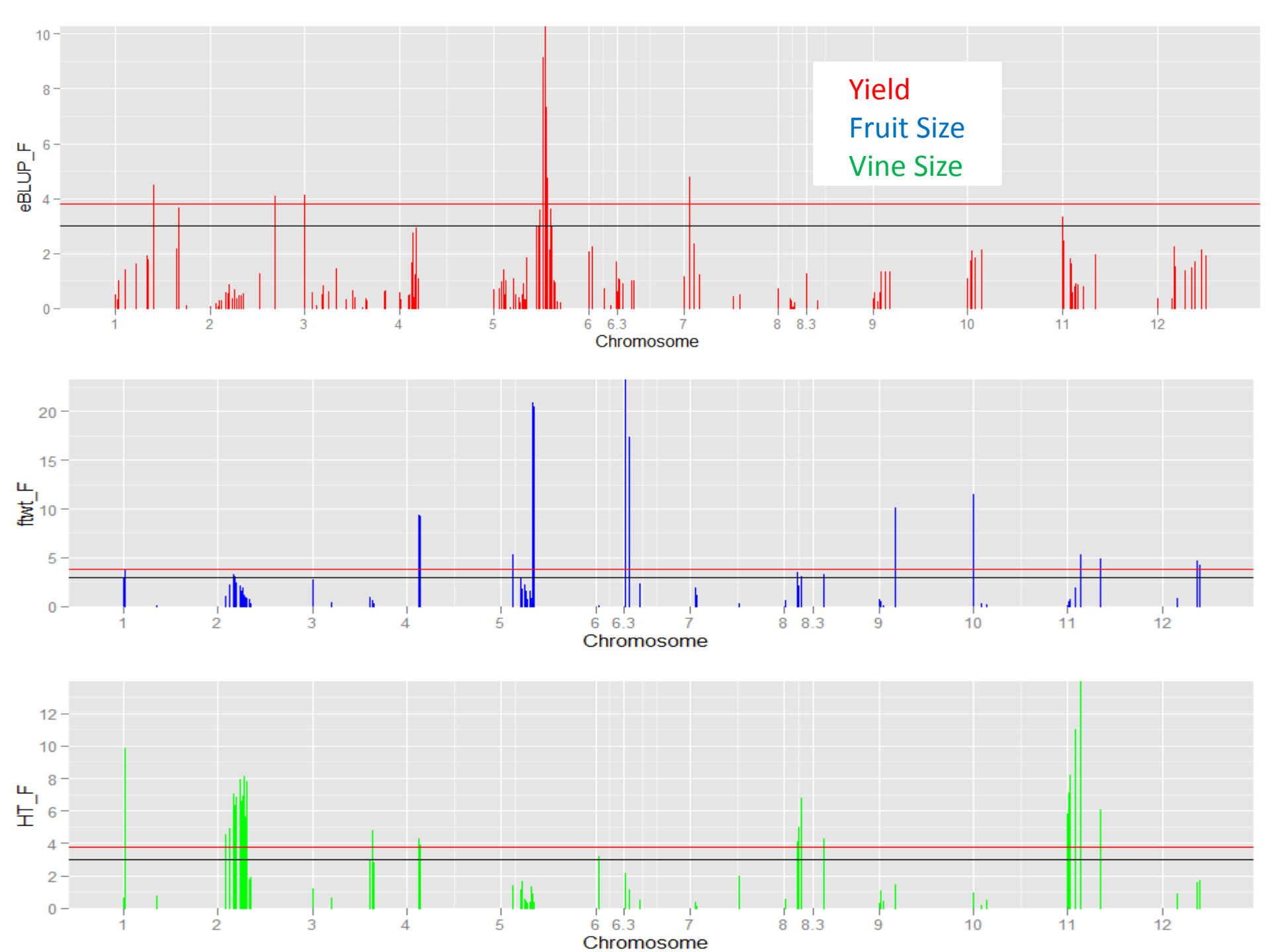
http://www.extension.org/pages/61007/

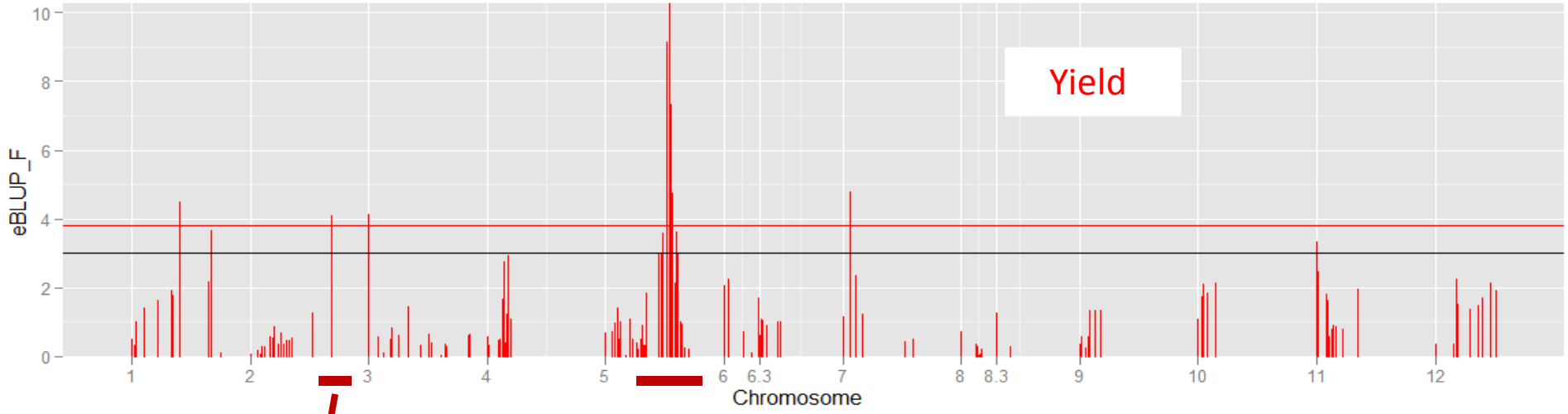
- Cluster file
- Design files for optimized SNP sets (fresh market and processing)

SNPs work well as markers in elite x elite populations (158-289 PM)

>95% co-linearity between physical map and recombination map




Yield
Fruit Size
Vine Size
eBLUP_F
ftwt_F
HT_F
Chromosome
1
2
3
4
5
6
6.3
7
8
8.3
9
10
11
12



```
      430        440        450        460        470        480        490
CTTCAAGAATATGGTCAAGTAGGTGATGCTAATGTGTTTTACTGGTTTCAAAACAGGAAATCAAGAAGTA
CTTCAAGAATATGGTCAAGTAGGTGATGCTAATGTGTTTTACTGGTTTCAAAACAGGAAATCAAGAAGTA
CTTCAAGAATATGGTCAAGTAGGTGATGCTAATGTGTTTTACTGGTTTCAAAACAGGAAATCAAGAAGTA
CTTCAAGAATATGGACAAGTAGGTGATGCTAATGTGTTTTACTGGTTTCAAAACAGGAAATCAAGAAGTA
CTTCAAGAATATGGACAAGTAGGTGATGCTAATGTGTTTTACTGGTTTCAAAACAGGAAATCAAGAAGTA
CTTCAAGAATATGGACAAGTAGGTGATGCTAATGTGTTTTACTGGTTTCAAAACAGGAAATCAAGAAGTA
CTTCAAGAATATGGTCAAGTAGGTGATGCTAATGTGTTTTACTGGTTTCAAAACAGGAAATCAAGAAGTA
CTTCAAGAATATGGTCAAGTAGGTGATGCTAATGTGTTTTACTGGTTTCAAAACAGGAAATCAAGAAGTA
CTTCAAGAATATGGTCAAGTAGGTGATGCTAATGTGTTTTACTGGTTTCAAAACAGGAAATCAAGAAGTA
```

"novel" alleles at *Simple inflorescence* (S) may increase the number of fruit

```
     1480       1490       1500       1510       1520       1530       1540
GCATTTCATGGAAGTGATTGATGCTATGCTAGGCCTGTTCACTTTTTACACAGTGTTCTTTTTAACAAAG
GCATTTCATGGAAGTGATTGATGCTATGCTAGGCCTGTTCACTTTTTACACAGTGTTCTTTTTAACAAAG
GCATTTCATGGAAGTGATTGATGCTAAGCTAGGCCTGTTCACTTTTTACACAGTGTTCTTTTTAACAAAG
GCATTTCATGGAAGTGATTGATGCTAAGCTAGGCCTGTTCACTTTTTACACAGTGTTCTTTTTAACAAAG
GCATTTCATGGAAGTGATTGATGCTAAGCTAGGCCTGTTCACTTTTTACACAGTGTTCTTTTTAACAAAG
GCATTTCATGGAAGTGATTGATGCTAAGCTAGGCCTGTTCACTTTTTACACAGTGTTCTTTTTAACAAAG
GCATTTCATGGAAGTGATTGATGCTATGCTAGGCCTGTTCACTTTTTACACAGTGTTCTTTTTAACAAAG
GCATTTCATGGAAGTGATTGATGCTATGCTAGGCCTGTTCACTTTTTACACAGTGTTCTTTTTAACAAAG
GCATTTCATGGAAGTGATTGATGCTAAGCTAGGCCTGTTCACTTTTTACACAGTGTTCTTTTTAACAAAG
```

Xanthomonas gardneri

Xanthomonas euvesicatoria

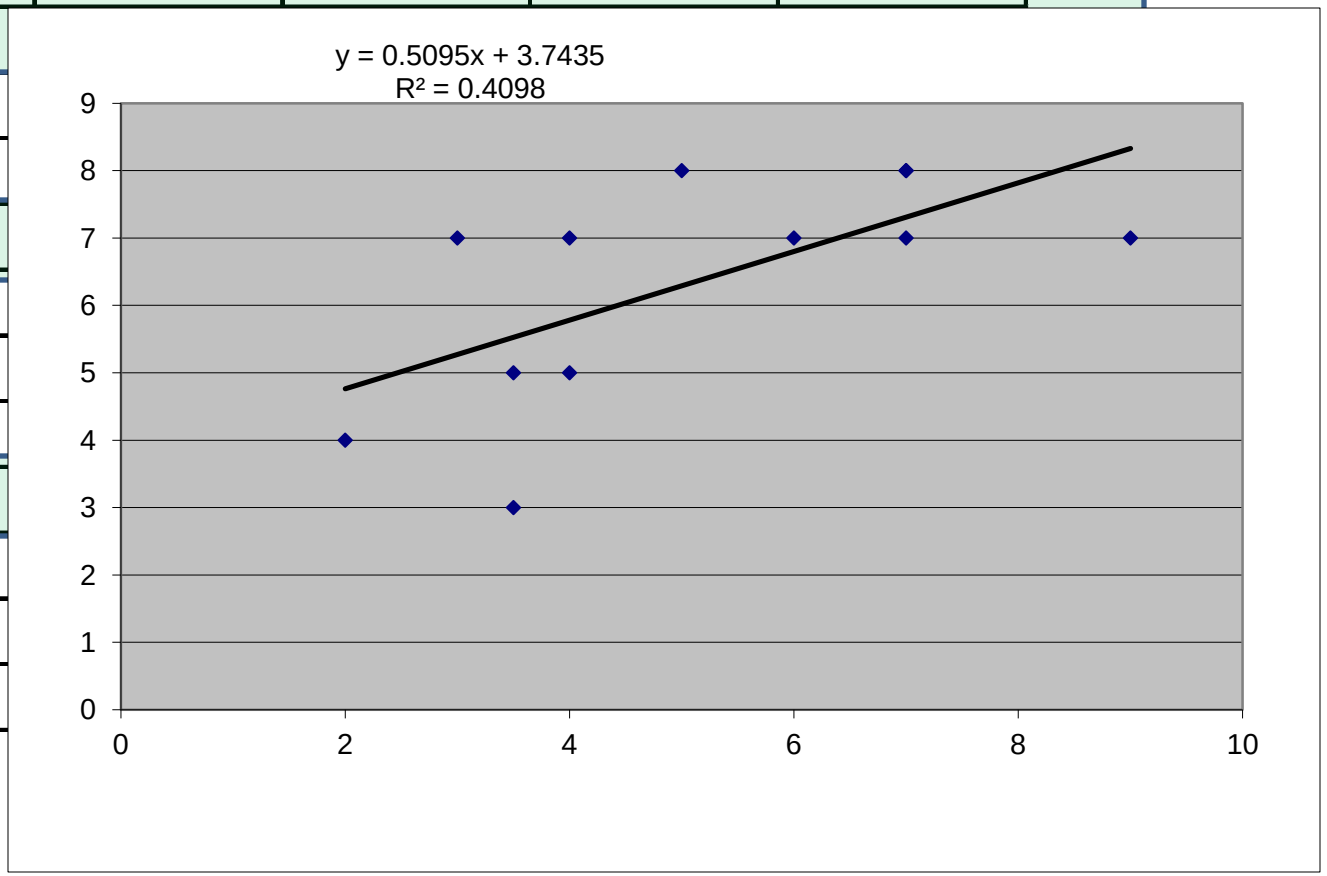
Comparison

	Disease rating			
Variety	**STEM**	**FOLIAGE**	**FRUIT**	**OVERALL**
SG07-015	7	4	5	5
SG07-002	2	3.5	3	3
SG07-003	2	3.5	5	4
SG07-011	4	6	7	5
SG07-004	2	2	4	3
SG07-030	7	7	8	7
SG07-017	4	4	7	5
SG07-018	5	9	7	8
SG07-028	4	3	7	4
SG07-027	6	7	7	7
SG07-019	6	7	8	7
SG07-020	4	5	8	6

	Disease rating			
Variety	STEM	FOLIAGE	FRUIT	OVERALL
SG07-015	7	4	5	5
SG07-002	2	3.5	3	3
SG07-003	2	3.5	5	4
SG07-011	4	6	7	5
SG07-004	2	2	4	3
SG07-030	7	7	8	7
SG07-017	4	4	7	5
SG07-018	5	9	7	8
SG07-028	4	3	7	4
SG07-027	6	7	7	7
SG07-019	6	7	8	7
SG07-020	4	5	8	6

These have PI128216 in their pedigree

Variety	Disease rating			
	STEM	FOLIAGE	FRUIT	OVERALL
SG07-015	7	4	5	5
SG07-002				
SG07-003				
SG07-011				
SG07-004				
SG07-030				
SG07-017				
SG07-018				
SG07-028				
SG07-027				
SG07-019				
SG07-020				



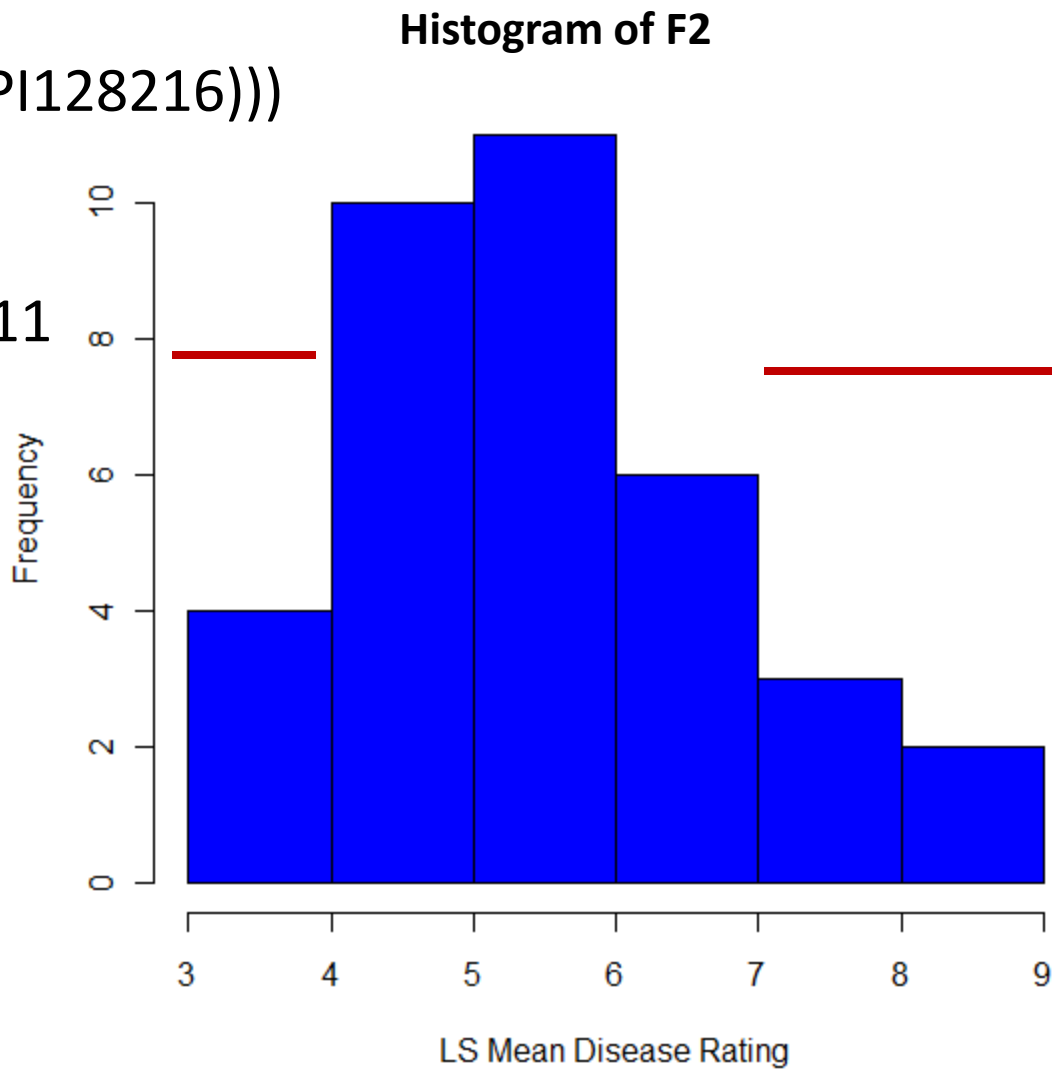
Populations:

OH8245 x (OH88119 x (OH88119 x (OH88119 x PI128216)))

- OH06-8021
- OH06-8040

Select high and low in 2011

Evaluate plots in 2012

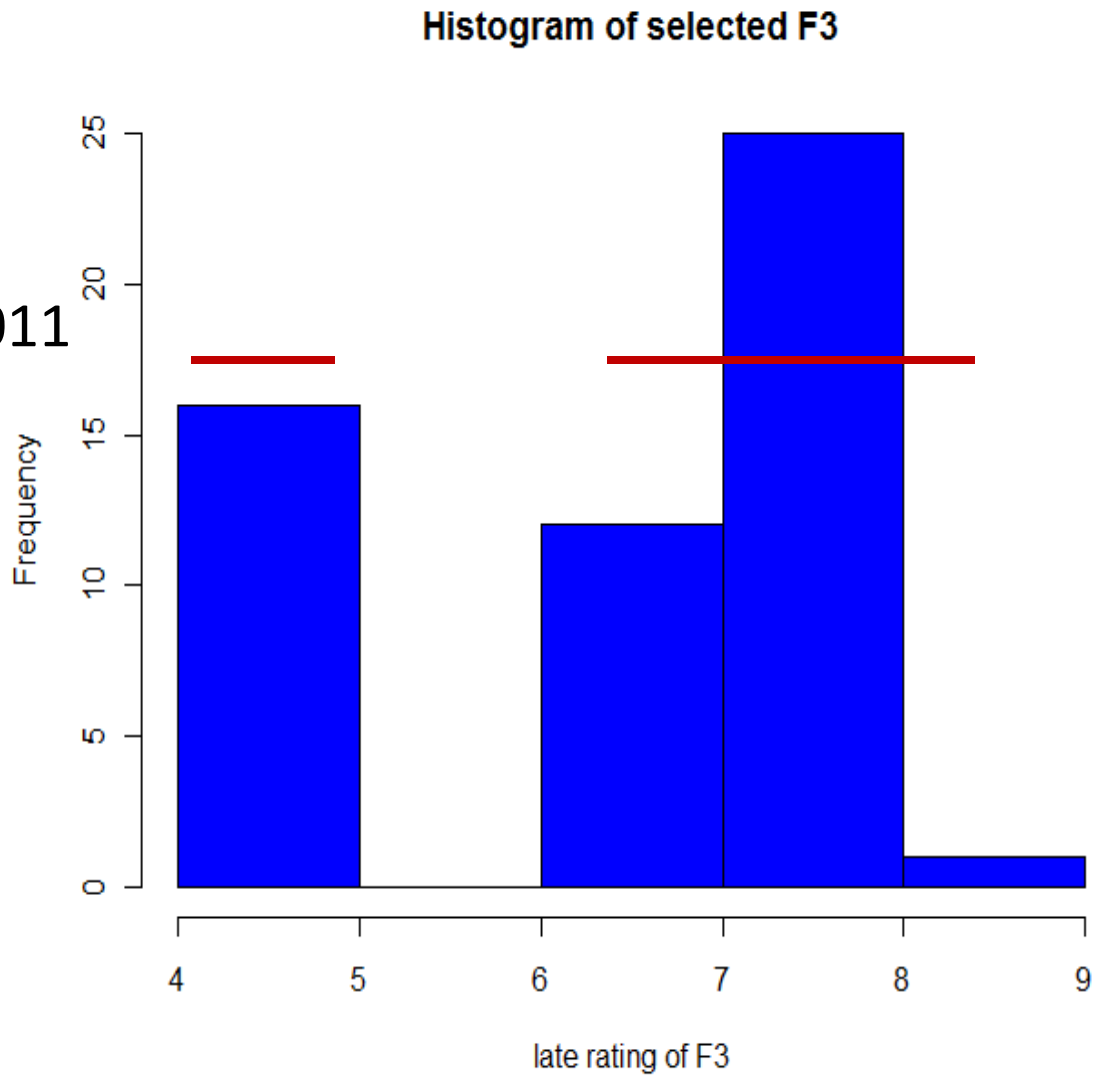


Populations:

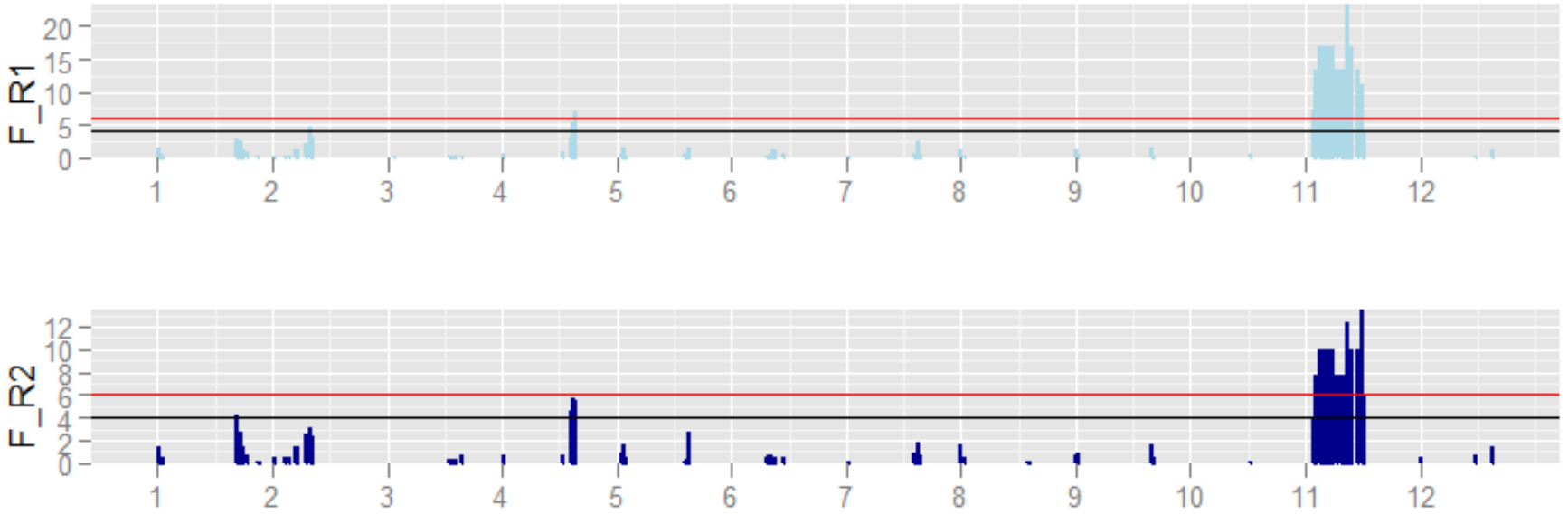
OH8245 x OH06-8021

OH8245 x OH06-8040

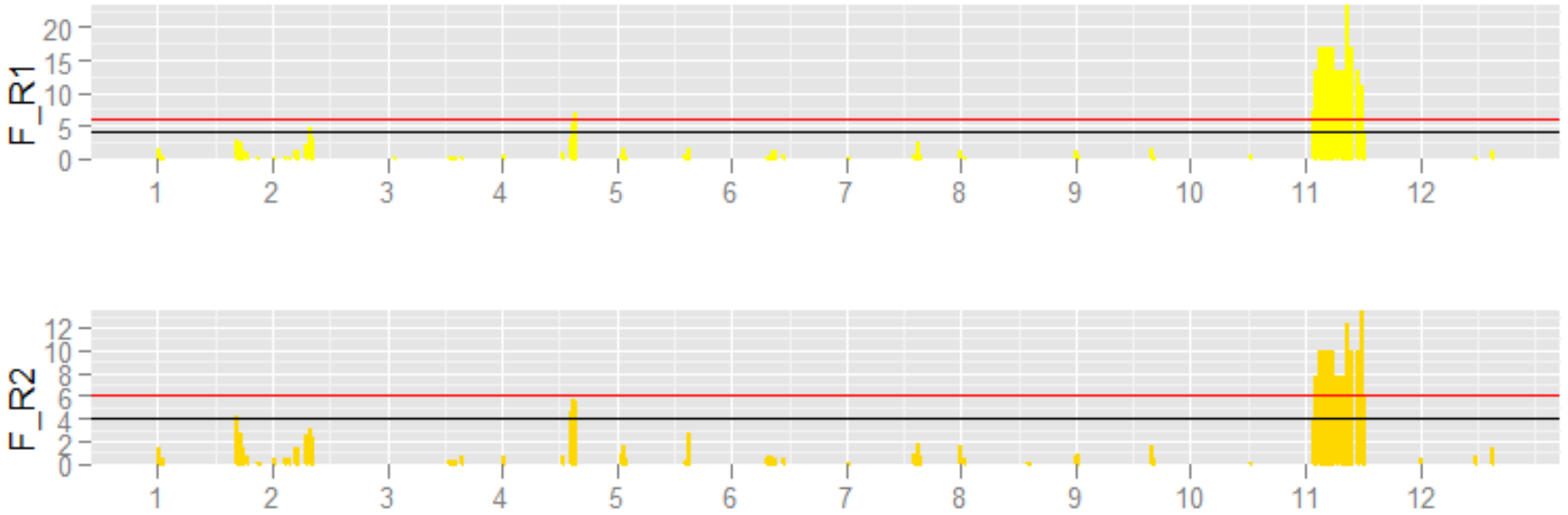
Select high and low in 2011
Evaluate plots in 2012

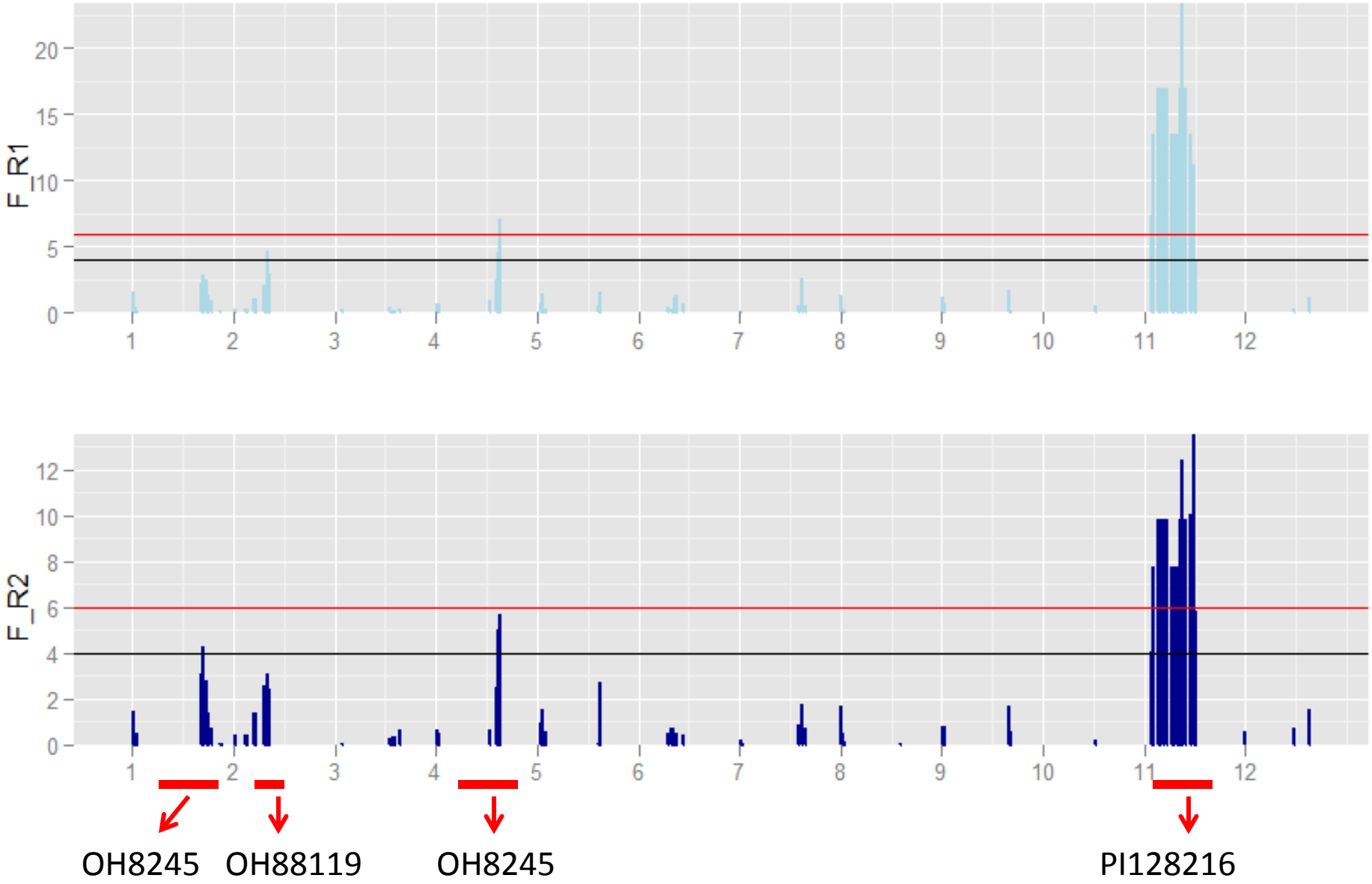


Combined populations



OH8245 x OH8021





Origin of alleles associated with resistance

“non-blighting resistance”
C28 → OH7870 → OH8245

Complex Breeding Population

Six advanced lines

Parent	Resistance
FL8233 (P1)	QTL-11and *Rx-4/Xv3*
OH7536 (P2)	*Rx-3*
OH8614 (P3)	QTL-11 ?
OH03-6439 (P4)	QTL-11 ?
MR13 (P5)	?
OH03-7463 (P6)	*Rx-3*

The first 9 crosses among six parents

	P2	P3	P4	P5	P6
P1	X	X	X	X	
P2			X		X
P3			X	X	
P4				X	

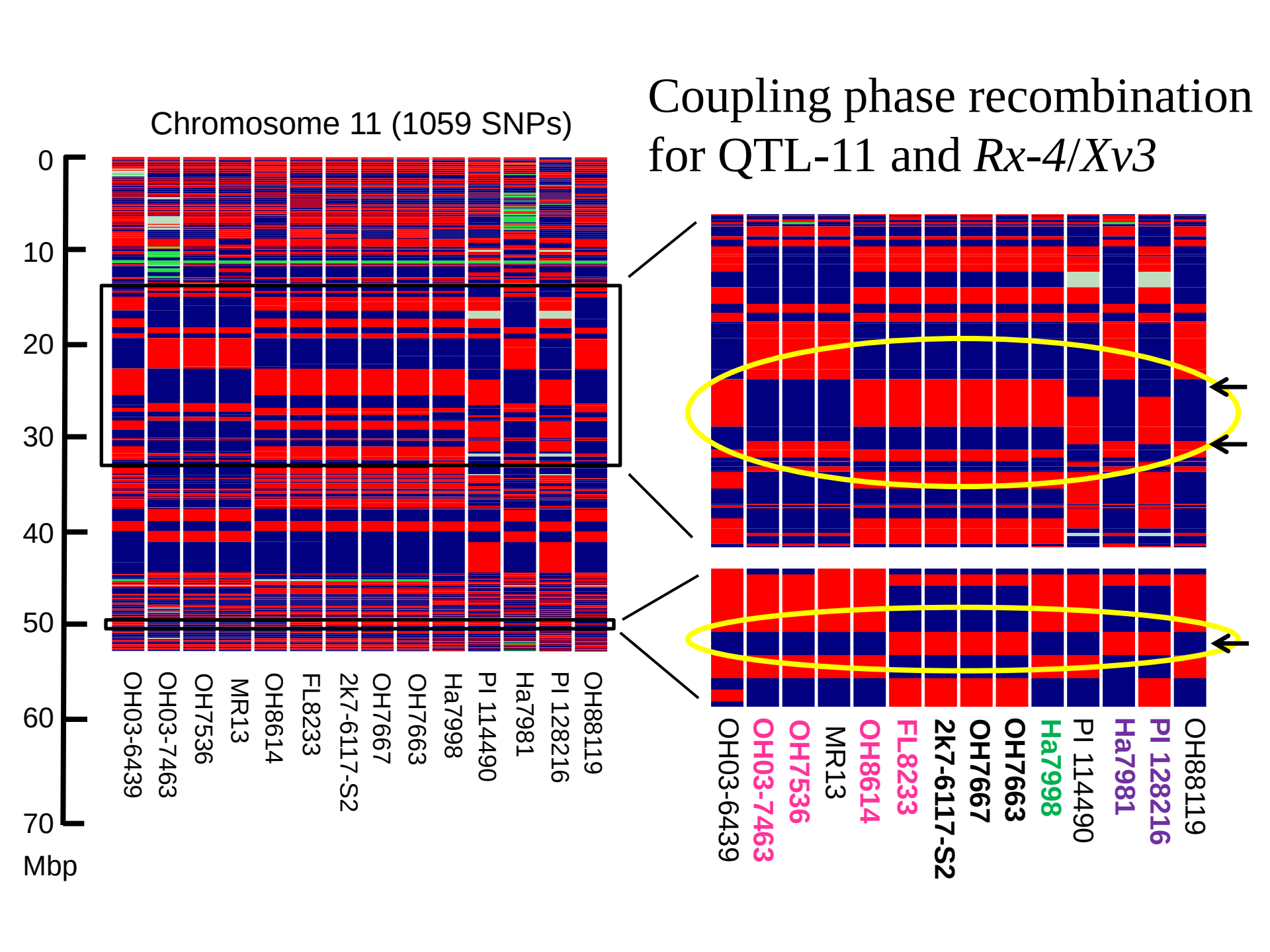
11 sequential crosses among the F_1 individuals from the first crosses

A complex breeding population consisting of over 1,100 progeny

Field trails with multiple species of bacterial spot in 2012

Resistance against Multiple Species

Variety	T1		T2		T3		Xg	
	Early	Late	Early	Late	Early	Late	Early	Late
PI 114490	2.0	2.0	2.5	3.0	1.5	3.0	3.0	4.0
OH7663	**2.5**	**3.0**	**4.0**	**4.0**	**2.0**	**4.0**	**3.0**	**4.5**
2K7-6117-S2	**3.0**	**3.5**	**4.0**	**4.0**	**2.0**	**4.5**	**3.0**	**5.0**
FL8233	3.0	3.5	4.0	4.0	2.0	4.5	2.0	4.0
H7981	3.0	3.0	4.0	4.5	1.0	4.0	4.0	4.5
H7998	3.0	3.0	4.0	6.0	5.0	5.0	5.0	6.0
OH7667	**4.0**	**4.0**	**4.0**	**4.0**	**2.5**	**4.0**	**3.0**	**4.0**
PI 128216	4.0	4.0	5.0	5.5	3.0	4.5	3.5	5.5
OH8245	5.0	5.0	5.0	6.0	4.0	5.5	4.0	7.0
OH88119	**7.5**	**9.0**	**6.5**	**9.0**	**7.0**	**9.5**	**7.0**	**9.5**
LSD (α=0.05)	1.3	1.6	1.4	1.2	0.9	1.2	1.4	1.4


Coupling phase recombination for QTL-11 and *Rx-4*/*Xv3*
Chromosome 11 (1059 SNPs)
0
10
20
30
40
50
60
70
Mbp
OH03-6439
OH03-7463
OH7536
MR13
OH8614
FL8233
2k7-6117-S2
OH7667
OH7663
Ha7998
PI 114490
Ha7981
PI 128216
OH8119
OH03-6439
OH03-7463
OH7536
MR13
OH8614
FL8233
2k7-6117-S2
OH7667
OH7663
Ha7998
PI 114490
Ha7981
PI 128216
OH8119

Resistant Selection

OH7663

OH7663 Fruit

OH88119

Conclusions

Sequencing “relevant” germplasm rather than “model” accessions was essential to develop a tool relevant to breeding programs.

Human Selection has led to differentiation

- Geographic/environmental adaptation & market use
- Supported by PCA, STRUCTURE, LD, F_{ST}
- Recombination is limited in cultivated populations

Contemporary Breeding has increased genetic variation relative to “vintage” or “Heirloom” varieties

- Allelic variation/No. of polymorphic markers
- Variation reflects both introgression and breeding history

Science:

- Insights into the effect of human selection on the tomato genome during domestication, adaptation to new environments, and breeding for market specialization.

Application:

- Optimized SNPs for community projects has been cost effective/productive.
- Establishing linkage to traits of fundamental economic importance (Yield QTL in elite x elite populations; bacterial spot resistance)
- Sufficient variation exists for MAS and GWS strategies

Acknowledgments

Collaborators, OSU

Heather Merk
Shawn Yarnes
Sung-Chur Sim
Matt Robbins
Troy Aldrich

Collaborators, MSU

David Douches
C Robin Buell
John Hamilton
Dan Zarka
Kelly Zarka

Collaborators, Cornell

Walter de Jong
Lucas Mueller
Joyce van Eck
Naama Menda

Collaborators, UCD

Allen Van Deynze
Kevin Stoffel

Collaborators, INRA

Mathilde Causse

Industry Collaborators

Cindy Lawley, Illumina
Martin Ganal, Trait Genetics

Collaborators, CAU

Wencai Yang
Hui Wang

Collaborators, UIB

Hipolito Medrano
Pep Cifre
Josefina Bota
Miquel Angel Conesa

Funding

USDA/AFRI

This project is supported by the Agriculture and Food Research Initiative of USDA's National Institute of Food and Agriculture.

Caleb Orchard

Debora Liabeuf

A few of the other people that deserve credit….

MICHIGAN STATE UNIVERSITY

UCDAVIS AGGIES

OHIO STATE