Genomic tools enhance power and precision of hazelnut breeding

Vidyasagar R Sathuvalli, Shawn A Mehlenbacaher, Brooke C Peterschmidt

Department of Horticulture, Oregon State University, Corvallis OR 97331

Overview

- Hazelnut breeding
- Eastern filbert blight (EFB) disease
  - Disease resistance breeding – Marker assisted selection
  - Sources of resistance to EFB
  - Map-based cloning of EFB resistance ‘Gasaway’ gene
- Sporophytic incompatibility
- Development of SSR markers from hazelnut transcriptome

Hazelnut Breeding

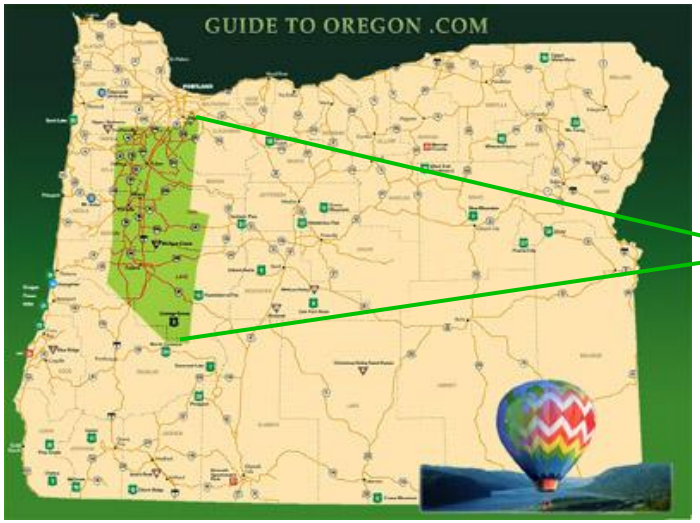
European Hazelnut
(*Corylus avellana* L.)

- Origin - Europe and Asia minor
- Family: Betulaceae
- National Clonal Germplasm Repository
  - 825 accessions of *Corylus*
  - 429 of *C. avellana*
- ~ 100 more accessions at OSU

http://caliban.mpiz-koeln.mpg.de/thome/band2/tafel_005.jpg

Hazelnut Production (MT)

1.	Turkey	500,000	65.2%
2.	Italy	104,900	13.7%
3.	United States	42,640	5.6%
4.	Azerbaijan	30,430	4.0%
5.	Georgia	21,800	2.8%



Oregon's Willamette Valley – home to 99 % of the U.S. hazelnut industry

faostat.fao.org

Hazelnut Breeding Objectives

A. Blanched kernel market (for chocolate, baked goods)

1. Bud mite resistance
2. Round nut shape
3. High percent kernel
4. Precocity
5. High yield
6. Easy pellicle removal
7. Few defects
8. Early maturity
9. Free-falling nuts

B. Resistance to eastern filbert blight (EFB)

1. Simply inherited resistance (‘Gasaway’ & others)
2. Quantitative resistance (e.g. ‘Tonda di Giffoni’)

Mehlenbacher, 2011

Hazelnut Breeding Flow Chart

1. Choose parents, make crosses
2. Grow in greenhouse
3. Seedlings in field
4. 〃
5. 〃
6. Evaluate a few nuts
7. Evaluate nuts
8. Layer, Evaluate nuts
9. Nursery, Evaluate nuts
10. Plant replicated trials
11. Trees in field
12. Evaluate nuts
13. 〃
14. 〃
15. 〃
16. Evaluate nuts, Summarize data
17. Release new cultivar

Breeding cycle: 8 years from seed to seed

Mehlenbacher, 2011

Obtaining Hybrid Seed

Mehlenbacher, 2011

Growing Hybrid Seedlings (4000 planted per year)

Mehlenbacher, 2011

Evaluation of Original Seedlings

Stage-I with removal of discards

Mehlenbacher, 2011

Propagation of Selections

Tie-Off Layerage (suckers form every year)

Harvested layers are weak. They are held in the nursery for one year, and then planted in the orchard.

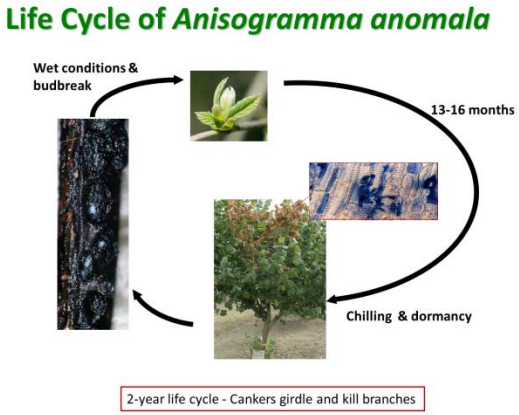
Mehlenbacher, 2011

Second Stage of Evaluation

in Replicated Trials

Eastern Filbert Blight in Hazelnut

- Caused by ascomycete *Anisograma anomala*
- Native to eastern North America
- First discovered in Willamette Valley – 1986
- 16 – 24 months between inoculation and canker development
- More than 60 % of Oregon's hazelnut orchards are affected or in close proximity to diseased orchards



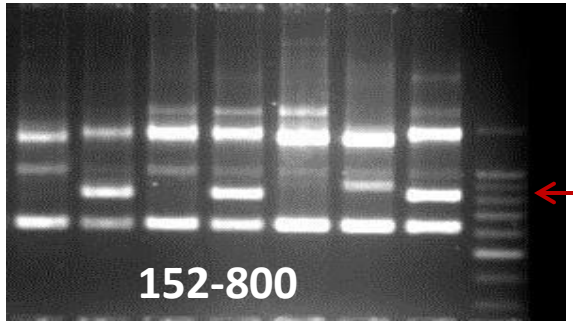
Disease Resistance Breeding

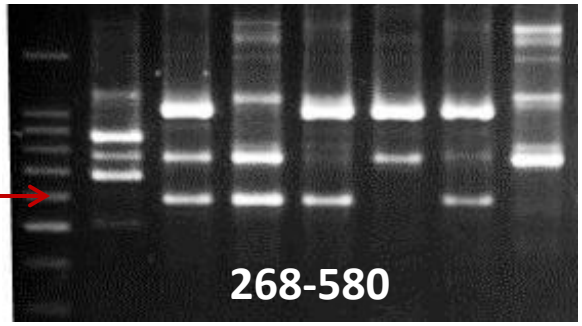
- Host genetic resistance was first identified in ‘Gasaway’
- An obsolete pollinizer
- Resistance controlled by dominant allele at single locus
- RAPD markers closely linked to resistance were identified
- Markers assisted selection (MAS) is being carried out for ‘Gasaway’ resistance

Barcelona **Jefferson**
Susceptible **Resistant**

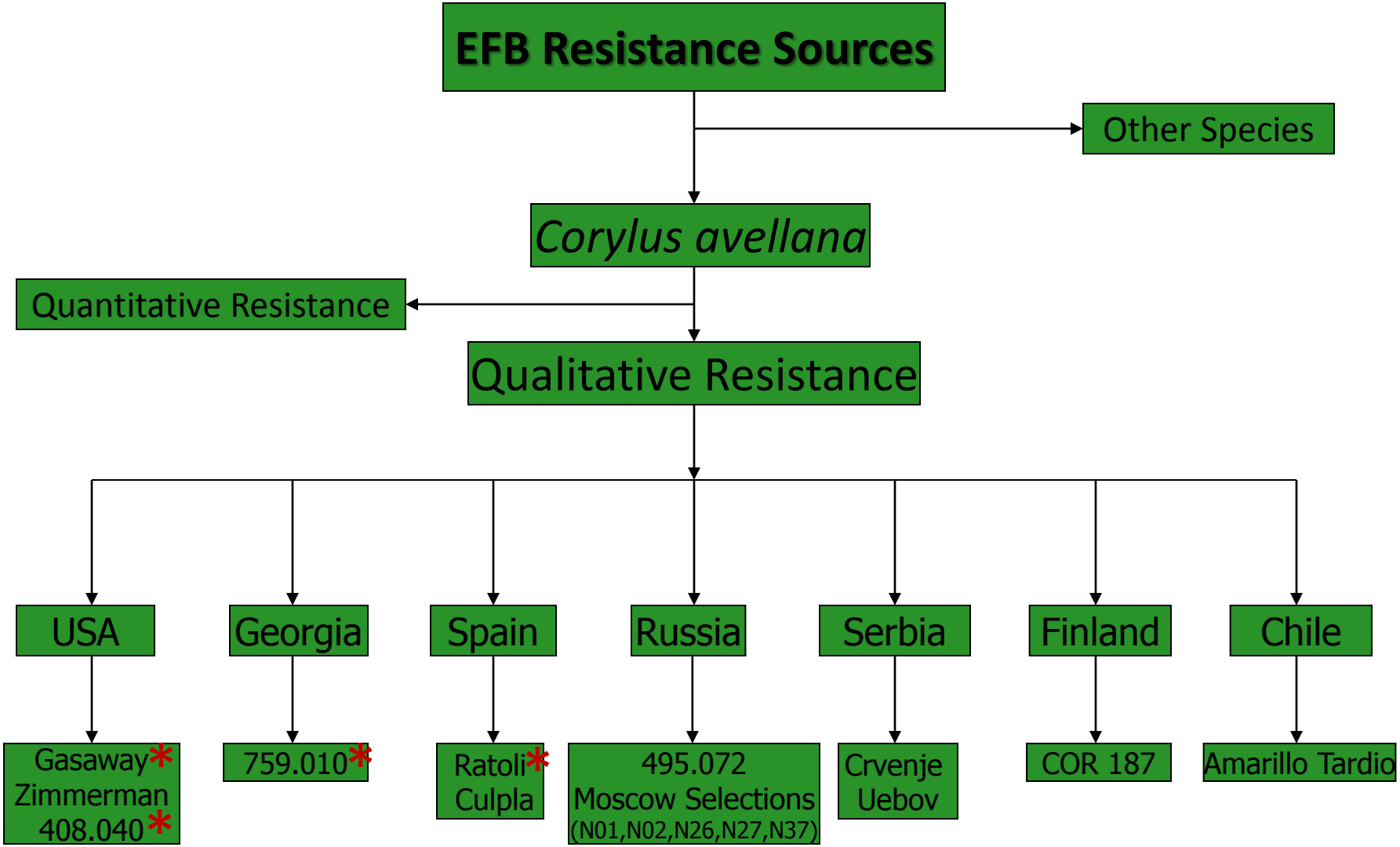
Marker Assisted Selection (MAS)

- Extract DNA from ~ 3000 seedlings per year, 192 per day
- Amplify using flanking RAPD markers 152-800 and 268-580
- Seedlings that lack markers are discarded





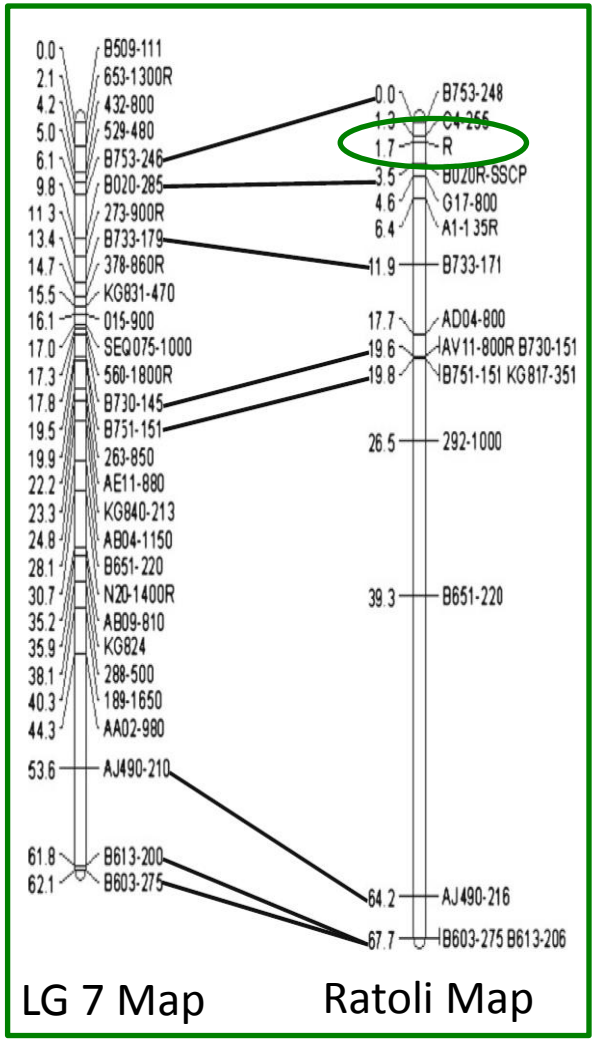
New sources of resistance to EFB

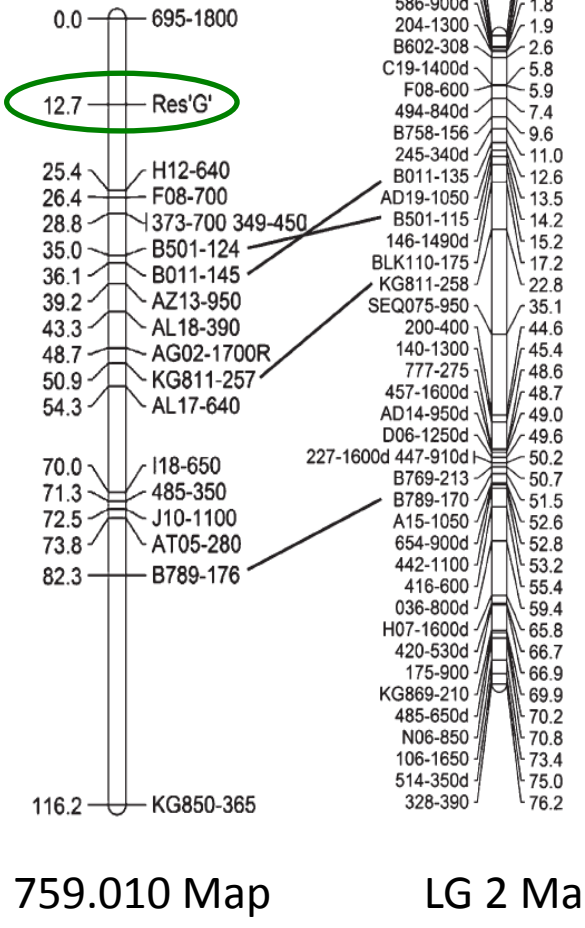


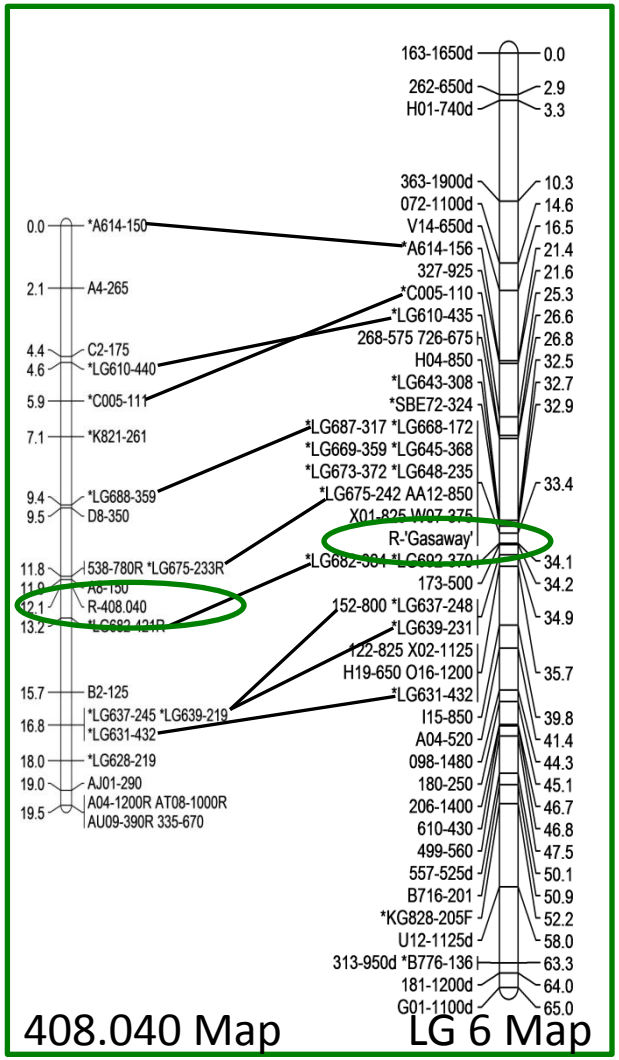
*Markers linked to resistance were identified

Assignment to Linkage Groups

(based on co-segregation with SSR markers)







Assignment to Linkage Groups

Gasaway	**LG6**
Ratoli	**LG7**
Georgian 759.010	**LG2**
408.040	**LG6**
Culpla	**LG6?**
Russian 495.072	**LG6?**

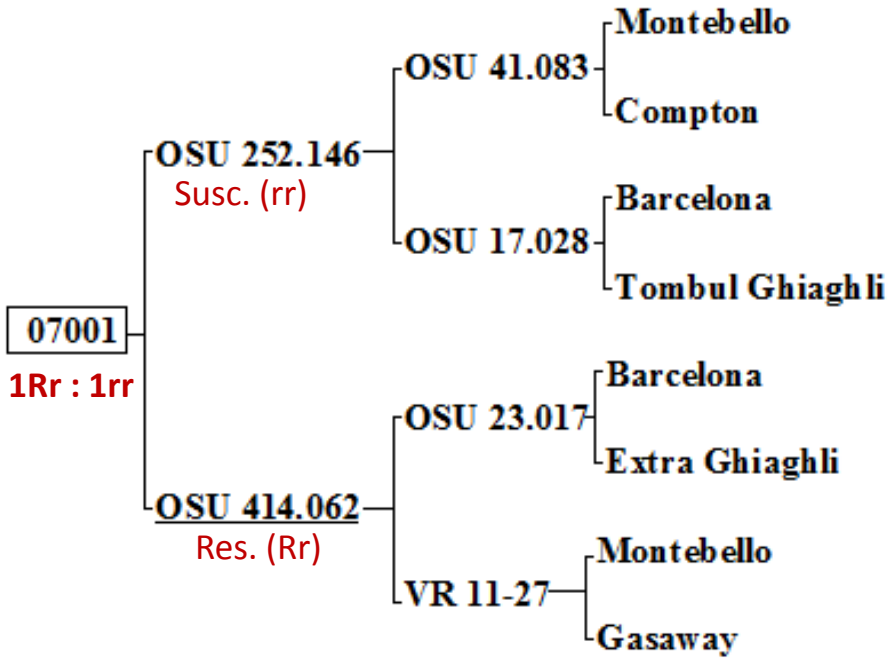
Are the *R* genes from 'Gasaway' 408.040, 'Culpla' and 495.072 the same or part of a cluster?

Map-based Cloning of the EFB Resistance Gene from 'Gasaway'

- Requirements of map based cloning
  - A large mapping population
  - Genetic map
  - Recombinants between flanking markers
  - A Bacterial Artificial Chromosome (BAC) library
  - Probes to screen BAC library for chromosome walking

Mapping Population

- In 2007, controlled crosses between OSU 252.146 and OSU 414.062 generated 1488 seedlings
  - 07001 – 1080 seedlings
  - 07002 – 408 seedlings

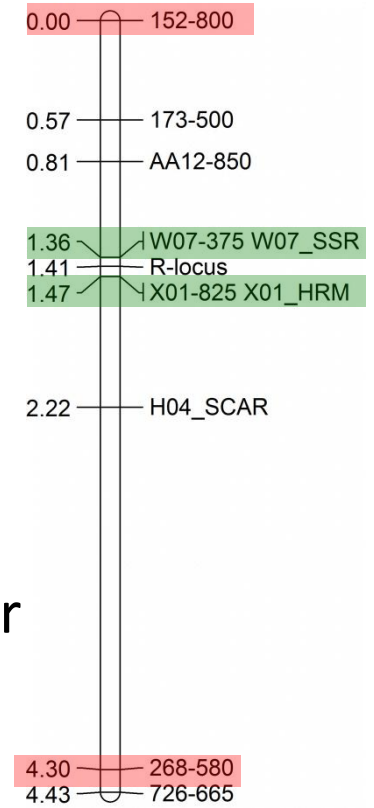


Pedigree of progeny 07001

- 07002 is from reciprocal cross, OSU 414.062 x OSU 252.146

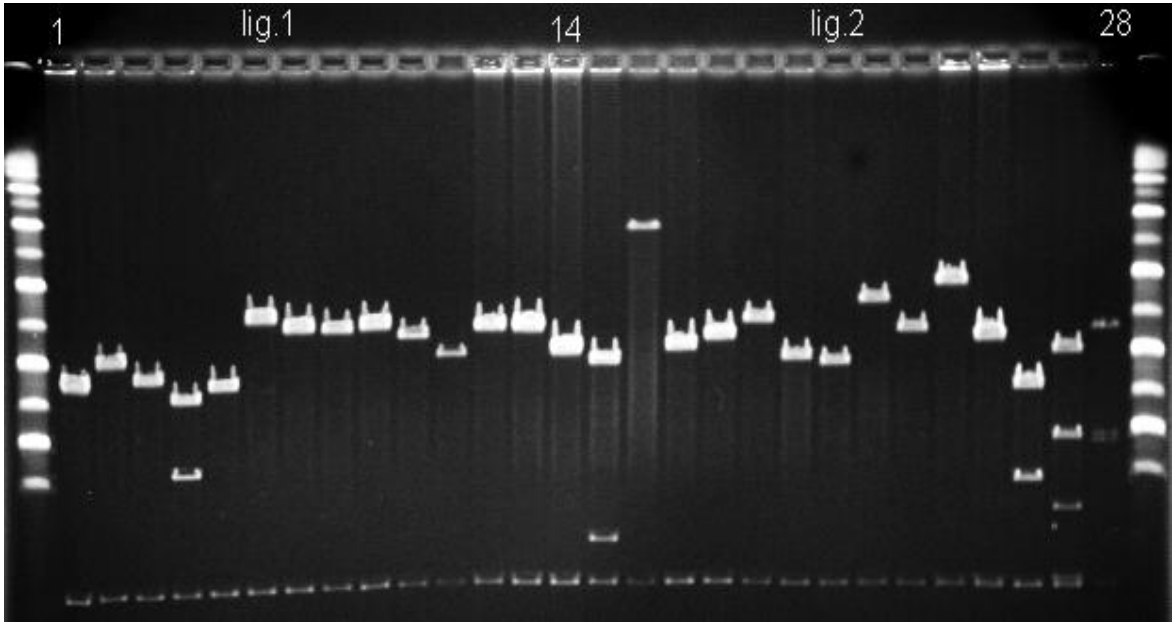
Fine Scale Genetic Mapping

- DNA was extracted from 1488 seedlings in 2008
- Screened for the presence of robust RAPD markers 152-800 and 268-580
- Both markers present – Assumed to be resistant
- Both markers absent – Assumed to be susceptible
- 87 recombinants (one marker present and the other absent) were identified – inoculated in the greenhouse
- Resistance is flanked by two RAPD markers W07-375 and X01-825



BAC Library of Hazelnut

- Constructed for ‘Jefferson’
- Average insert size – 117kb
- Genome coverage – 12x
- 39,936 clones arrayed in 104 384-well plates



Sathuvalli and Mehlenbacher (2011) Genome 54:862-867

Screening of BAC Library

Pooling and Screening by PCR

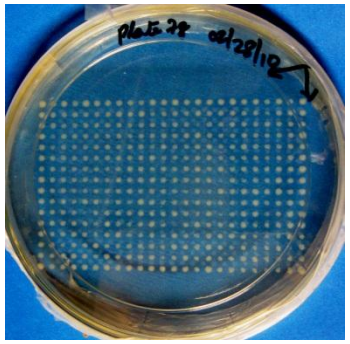


Plate pools - 104
Row pools - 16 per plate
Column pools - 24 per plate

Extract DNA

18 SCARs from 9 RAPD markers

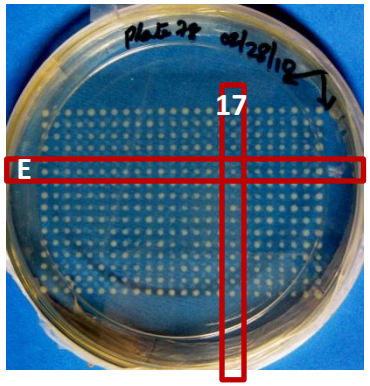
Screen Plate pools

Screen Row and Column pools

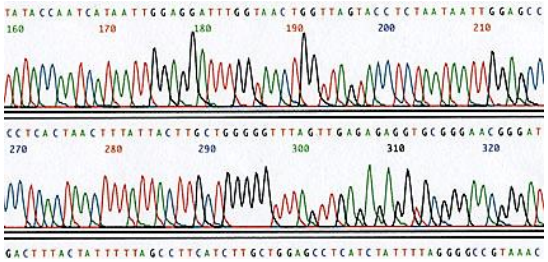
Identify the positive clones

Sequence the BAC ends

Screening twice helps avoid false positives and other PCR artifacts



Hit:78-Ex17



Chromosome Walking

Sequence the BAC ends

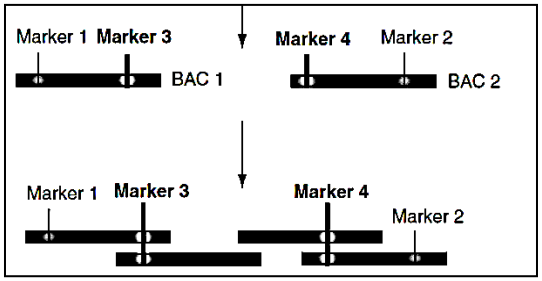
Design new primers from BAC ends

Map the new markers

Screen the BAC library with new BAC end probes

Identify new BACs

Identify the BACs carrying resistance



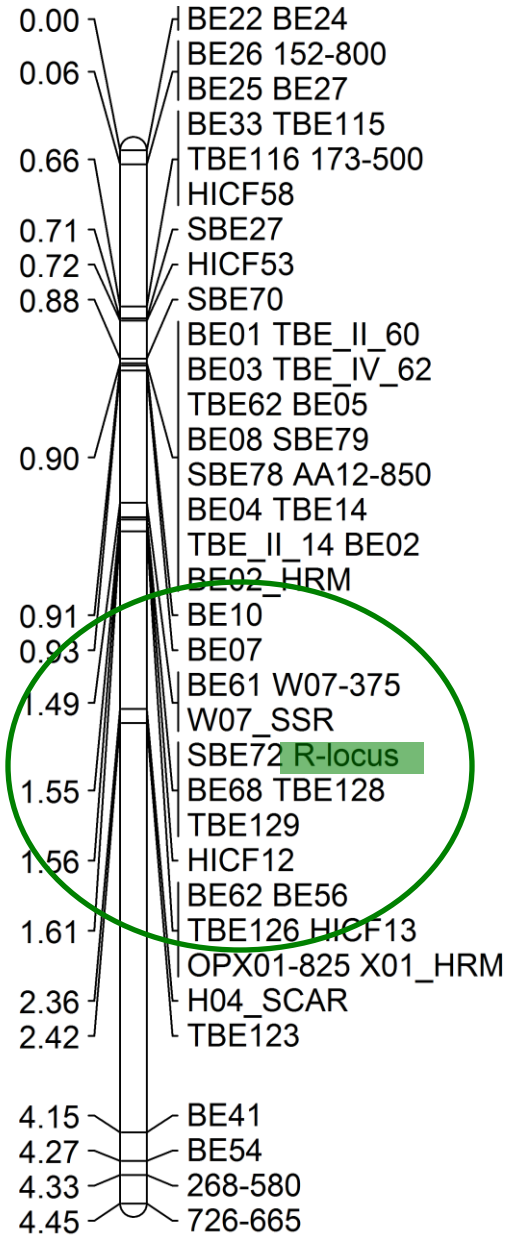
3 rounds of screening identified 93 BACs

Markers from BAC end Sequences

- Few polymorphic BAC end markers – a major constraint
- Markers were developed from BAC end sequences
- 41 markers were mapped to the resistance region
  - 23 - Sequence Characterized Amplified Region (SCAR)
  - 7 - Single Stranded Conformational Polymorphism (SSCP)
  - 4 - Simple Sequence Repeat (SSR)
  - 7 - High Resolution Melting (HRM)

Fine Mapping

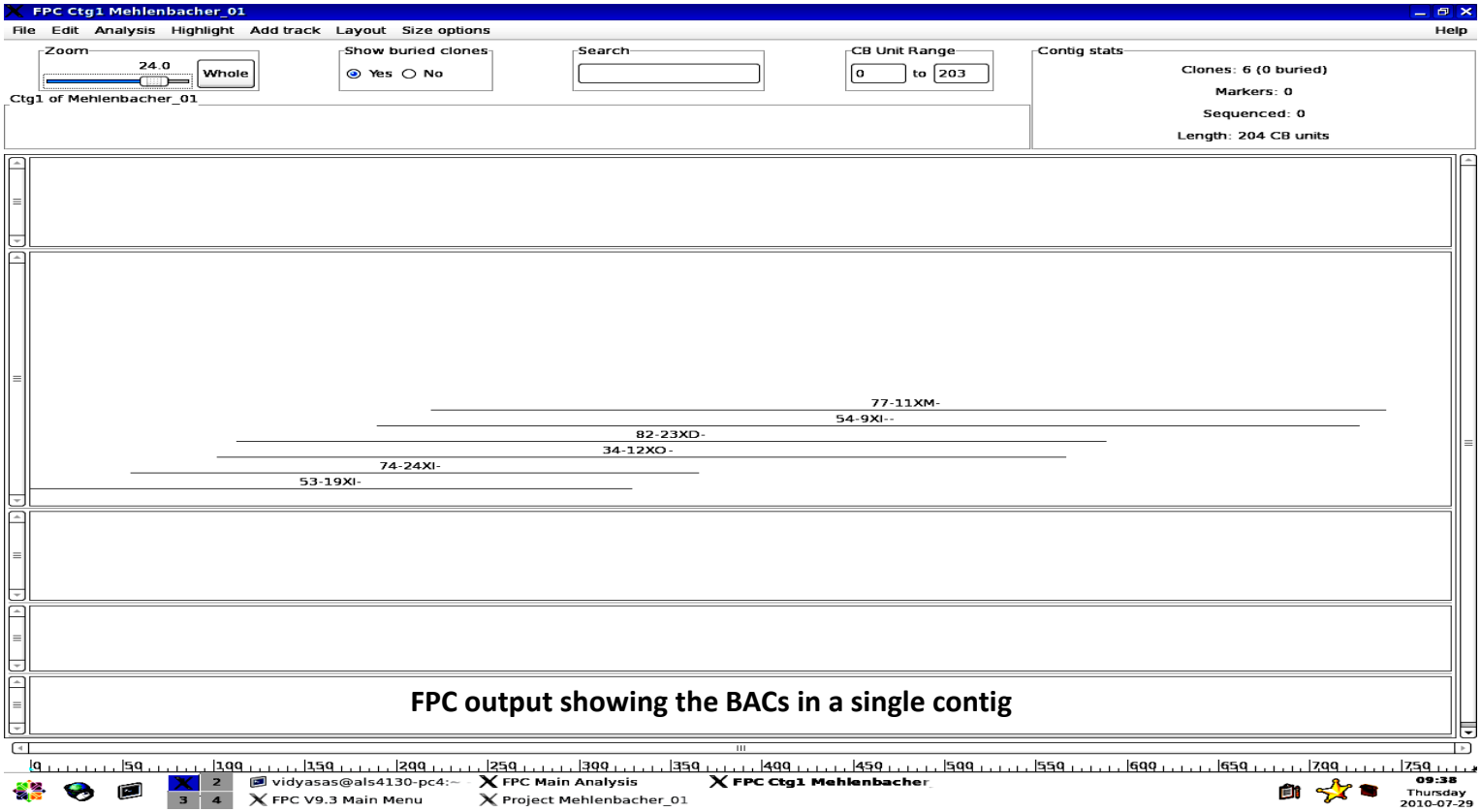
- High density map
  - 41 BAC end markers
  - 8 RAPDs
  - 2 HRMs from RAPDs
  - Resistance phenotype
- Map spans 4.45 cM
- Averages 0.03cM between markers
- 34 markers placed < 1cM from resistance



Sathuvalli and Mehlenbacher, in preparation

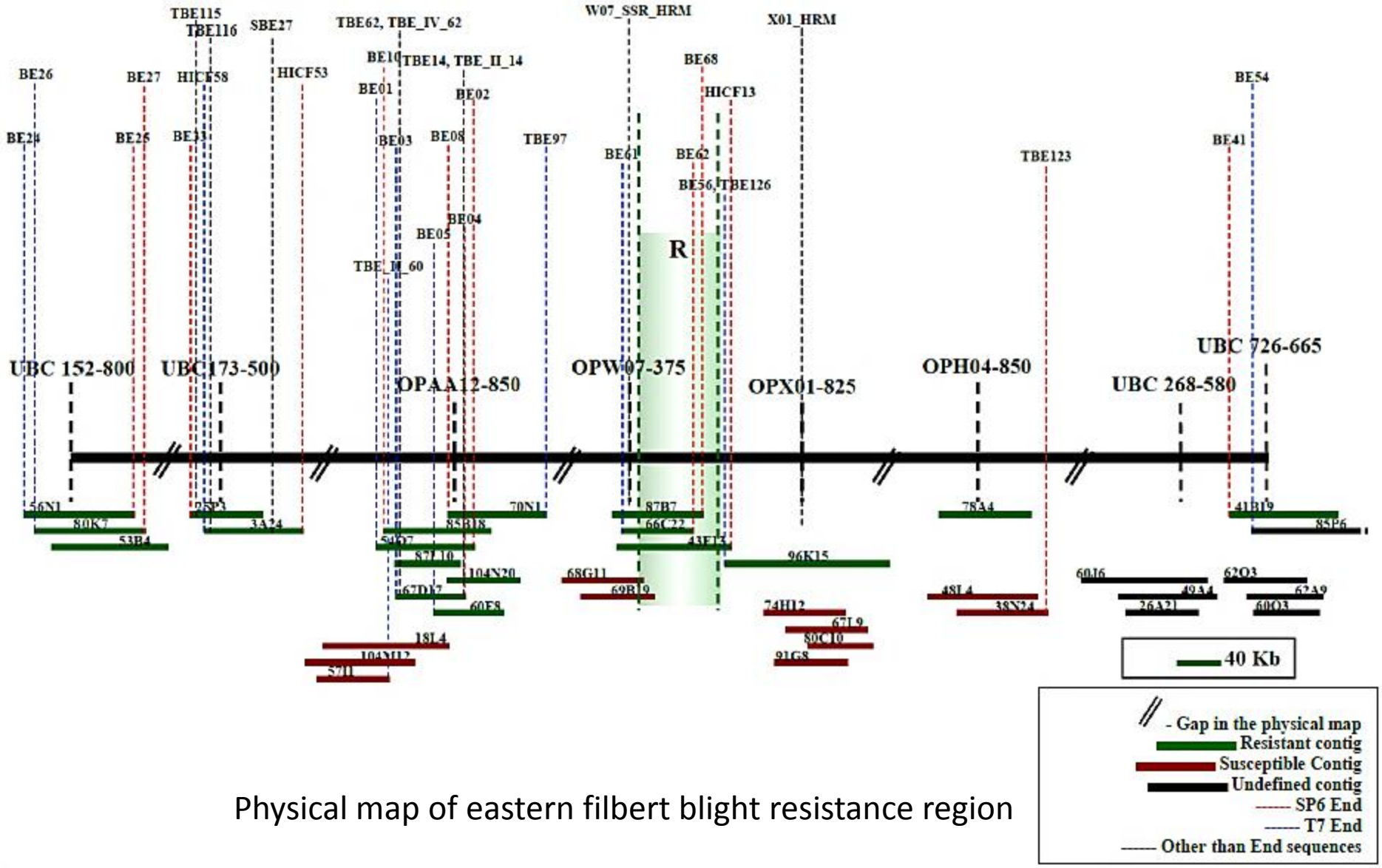
BAC Fingerprinting

- Each BAC screening provided more than one BAC hit
- HICF to merge BACs into contigs



FPC output showing the BACs in a single contig

Physical Map

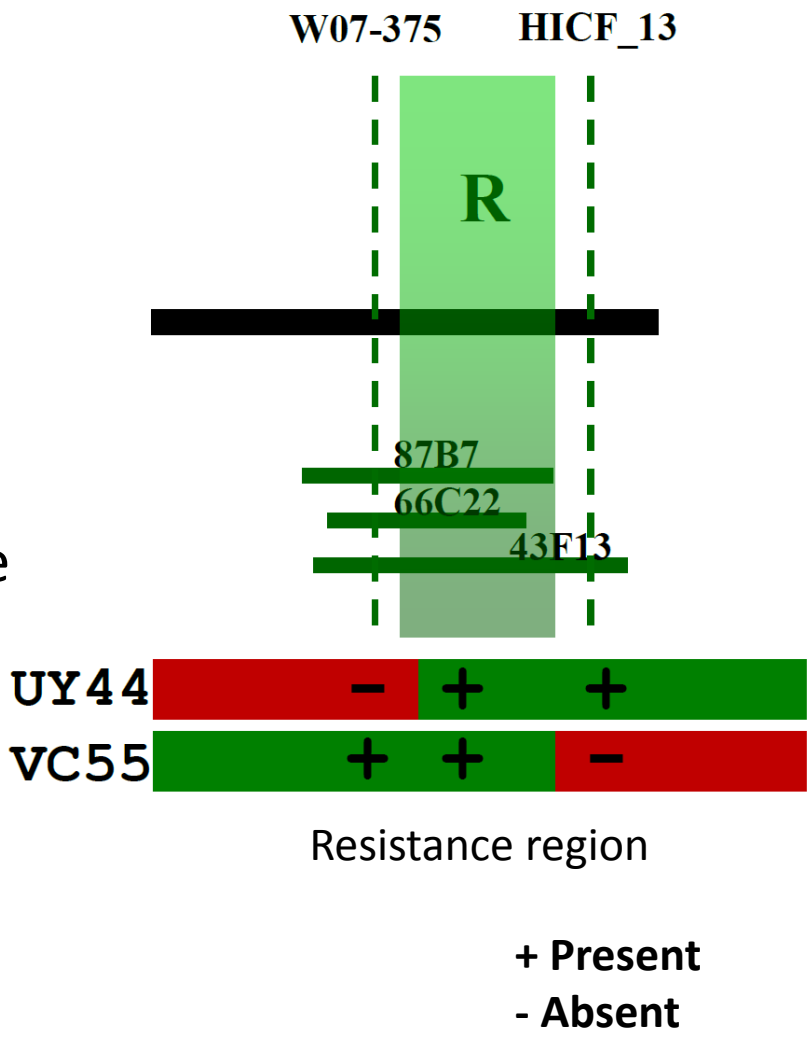


Physical map of eastern filbert blight resistance region

Sathuvalli and Mehlenbacher, in preparation

Identification of Resistant Contig

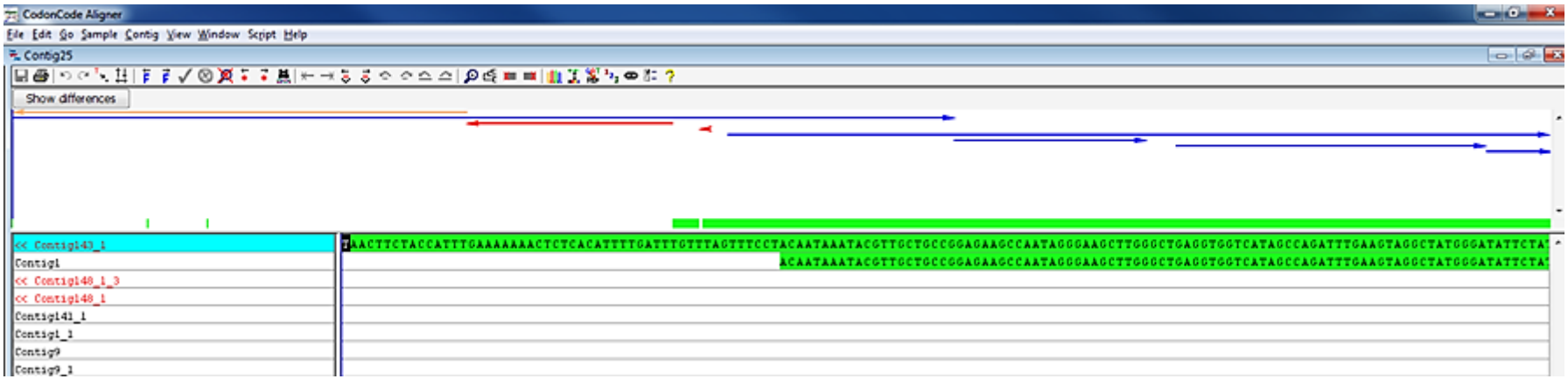
- A single recombination event was observed between W07-375 and resistance
- A single recombination event was observed between resistance and HICF13
- W07-375 and HICF13 are from the same contig
- Contig size is ~ 150kb
- Resistant contig consists of 3 overlapping clones



Sathuvalli and Mehlenbacher, in preparation

Sequencing of Whole BACs

- BACs in the resistance region were sequenced using Illumina technology
- Velvet and SOPRA programs were used to assemble Illumina reads into contigs
- CodonCode Aligner was used to align, trim and correct the contigs from Velvet and SOPRA



Sathuvalli and Mehlenbacher (2011), in dept. review

BAC Sequencing

BAC clone	RAPD marker	Illumina system	Barcode	No. of contigs	Longest contig (bp)	Shortest contig (bp)	Total length (bp)	Estimated size (bp)	Coverage (%)
49A4	H19-650	HiSeq	GAC	12	34,129	1,833	151,300	125,000	100*
56N1	152-800	HiSeq	CGC	6	53,963	3,150	120,427	125,000	96.34
53B4	152-800	HiSeq	CTC	15	23,972	1,735	123,510	130,000	95.01
25P3	173-500	GAIIx	ACG	3	45,211	7,373	92,393	90,000	100*
3A24	173-500	HiSeq	GGG	7	42,945	2,203	115,648	125,000	92.52
18L4	AA12-850	GAIIx	GGG	8	35,993	1,897	120,953	150,000	80.64
54O7	AA12-850	GAIIx	CCC	5	54,064	5,244	110,897	120,000	92.41
85B18	AA12-850	GAIIx	CCC	4	65,897	5,948	124,801	125,000	99.84
67D17	AA12-850	GAIIx	TGC	2	51,448	41,556	93,004	90,000	100*
87L10	AA12-850	GAIIx	CTG	5	45,632	393	90,035	85,000	100*
60F8	AA12-850	GAIIx	GCT	5	65,438	2,807	94,236	85,000	100*
70N1	AA12-850	GAIIx	CAC	5	77,983	4,307	117,921	120,000	98.27
72F19	AA12-850	GAIIx	GGG	1	108,194	0	108,194	100,000	100*
43F13	W07-375	GAIIx	CGT	8	31,223	2,815	101,433	135,000	75.14
65G23	W07-375	GAIIx	TGC	4	51,029	708	100,335	110,000	90.80
66C22	W07-375	GAIIx	CTG	5	40,853	1,013	68,934	90,000	76.59
68G11	W07-375	GAIIx	ACG	2	59,391	35,367	94,758	100,000	94.76
69B19	W07-375	GAIIx	CAC	4	45,817	2,609	82,878	90,000	92.09
85B7	W07-375	GAIIx	CGT	8	44,096	1,536	74,289	115,000	64.60
67L9	X01-825	GAIIx	GCT	4	69,962	1,743	94,012	100,000	94.01
96K15	X01-825	GAIIx	AGC	19	70,280	2,898	270,199	225,000	100*
48L4	H04-850	HiSeq	GGC	11	16,929	1,083	73,471	115,000	63.89
784A	H04-850	HiSeq	TCG	8	27,094	1,285	95,533	120,000	79.61
38N24	H04-850	HiSeq	CAC	11	31,170	1,307	98,093	130,000	75.46
60J6	268-580	HiSeq	CGG	15	30,411	1,672	121,356	150,000	80.90
54O21	726-665	GAIIx	AGC	7	52,147	1,951	125,069	125,000	100*
41B19	726-665	HiSeq	CCC	15	20,958	1,280	106,603	100,000	100*
62A9	726-665	HiSeq	CCA	13	22,638	651	95,618	95,000	100*

Sequence coverage:
60-100%

Sathuvalli and Mehlenbacher (2012) Tree Genetics & Genomes (Submitted)

Ab initio Gene Annotation

- Gene prediction carried out using AUGUSTUS
  - *Arabidopsis* as the gene prediction model
  - RNA-Seq data from 'Jefferson' (Data from Mockler lab)
  - Amino acid sequences of predicted genes were BLAST(P) searched for protein homology

Contig	RAPD	Contig Source	No. of BACs sequenced	No. of BACs in the contig	No. of genes at four percentages of RNA-Seq transcript support			
					0%	60%	80%	100%
01	173-500	Resistant	2	2	34	13	8	7
02	AA12-850	Resistant	7	7	30	11	6	5
03	AA12-850	Susceptible	3	3	62	16	10	6
04	W07-375	Resistant	3	3	37	12	6	5
05	W07-375	Susceptible	2	2	15	4	1	0
06	X01-825	Resistant	1	1	37	9	6	6
07	X01-825	Susceptible	1	4	18	8	6	3
				Total	233	73	43	32

Potential Candidate Genes (2 of 5)

Predicted gene Contig4_g19

- Belongs to p-loop NTPase superfamily
- Includes NBS-LRR type disease resistance proteins
- NBS-LRR is the major class of *R* genes so far identified

Predicted gene Contig4_g25

- Belongs to F-box superfamily
- Plant F-box genes – one of the largest multigene superfamilies
- Controls many important biological functions including disease resistance

Sathuvalli et al., in preparation

Incompatibility in Hazelnut

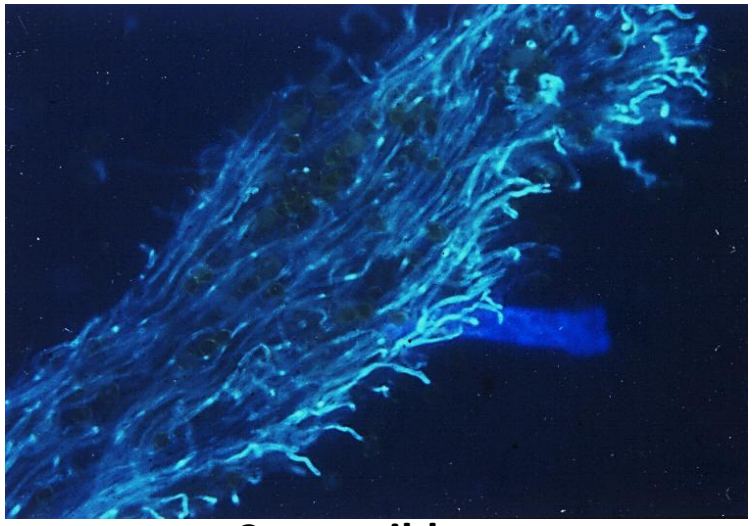
- Sporophytic, one *S*-locus, 33 alleles
- Dominance or co-dominance in pollen
- Co-dominance in stigmas
- Determined by fluorescence microscopy

If the same allele is expressed by the stigma and the pollen, the cross is incompatible.

Incompatibility Testing

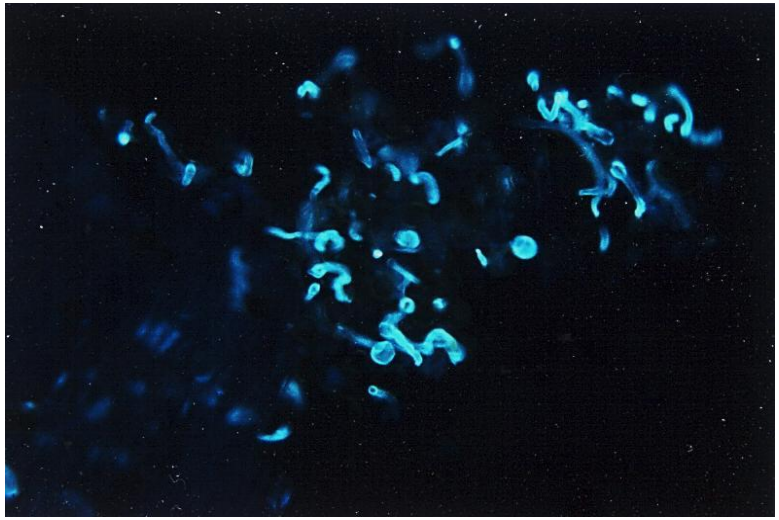
using Fluorescence Microscopy



Compatible

- Excellent germination
- Long parallel tubes

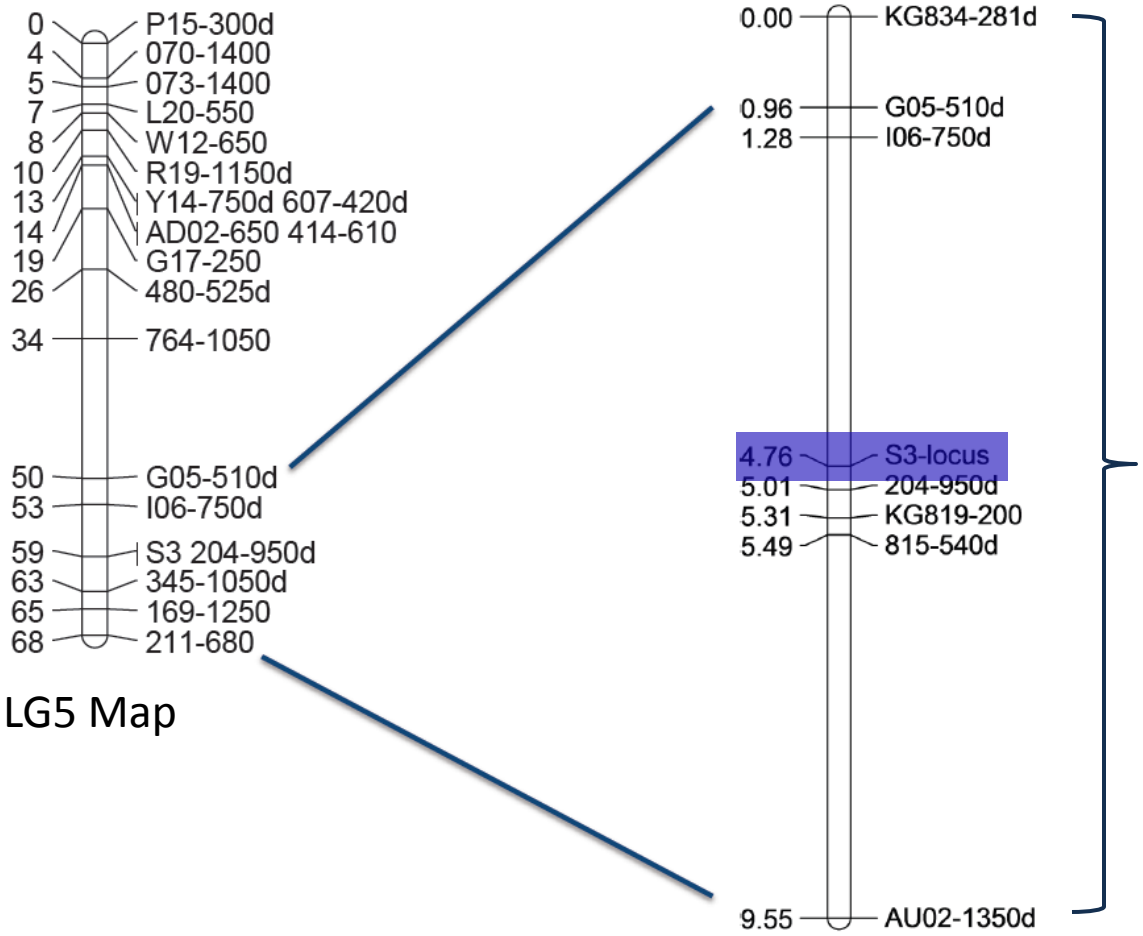


Incompatible

- Poor germination
- Short tubes, bulbs

Mehlenbacher, 2011

S-locus Mapping



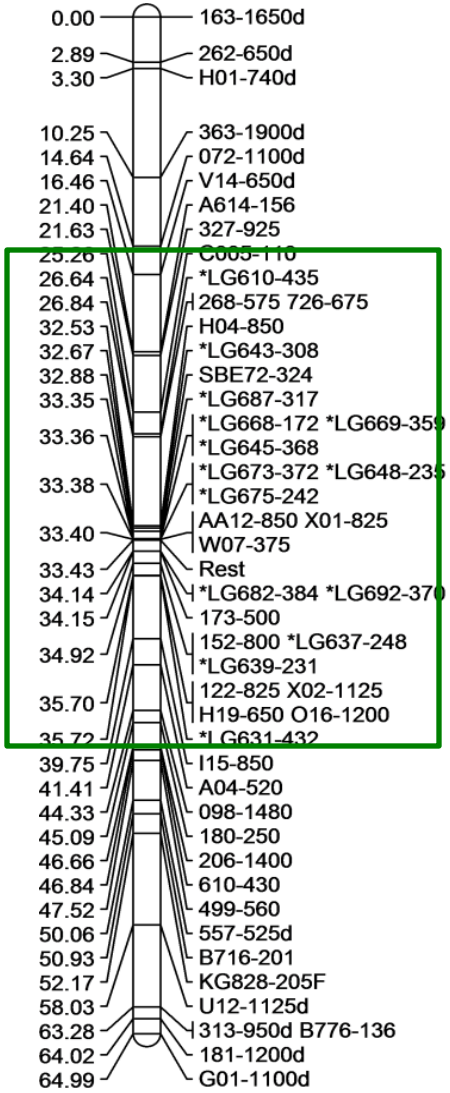
LG5 Map

Fine mapping of *S*-locus region

Sathuvalli and Mehlenbacher (2011) Genome 54:862-867

Development of SSRs

- 24 SSRs developed from BACs sequenced in the EFB resistance region
- SSRs (3-5 bp repeats) are being developed from the hazelnut transcriptome (Data from Mockler lab)
  - Leaf : 158 SSRs
  - Bark : 129 SSRs
  - Catkins : 81 SSRs
- Screening 368 SSRs with 24 diverse genotypes identified 158 (43%) as polymorphic
- Mapping and characterization studies are underway for these SSRs



Sathuvalli and Mehlenbacher (2012) Tree Genetics & Genomes (Submitted)

Summary

- Genomic tools are being effectively used in hazelnut breeding
- Map-based cloning identified BACs in the EFB resistance region
    - 5 candidate genes identified
- Map-based cloning of *S*-locus initiated
- SSR markers were developed in the EFB resistance region
- SSR markers are being developed from transcriptome data

Acknowledgements

David C Smith
Todd C Mockler (Genome Sequencing)
Nahla V Bassil
Cristino Montes
Becky McCluskey
Kahraman Gurcan
Erik Rowley
Brian Knaus (Bioinformatics)
Rich Cronn
Chris Sullivan (CGRB)
Wayne Wood

Funding Sources

Oregon Hazelnut Commission

Specialty Crops Research Initiative Grant
2009-51181-06028