Annotated EST-based SSR and SNP Genetic Maps for Chinese Chestnut

Genomics of Forest and Ecosystem Health in the Fagaceae (Beech Family)
November 10-13, 2009
North Carolina Biotechnology Center
Research Triangle Park, North Carolina, USA

Tom Kubisiak

Acknowledgments: much of the *Castanea* genetic marker work presented and/or discussed in this talk was possible due to the hard work of numerous people, in particular, the entire **Fagaceae Project Team:** Nick Wheeler, Dahlia Nielsen, Chris Smith, Jeff Tomkins, Bert Abbott, Michael Atkins, Barbara Blackmon, Meg Staton, Stephen Ficklin, Fred Hebard, Paul Sisco, John Carlson, Abdelali Barakat, William Powell, Kathleen Baier, Sandra Anagnostakis, Ronald Sederoff

Thank you to Charles Burdine for technical assistance and the USDA FS for continued support. Thank you to Jill Wegrzyn for SNP QA/QC assistance.

This project is supported by a grant from NSF (DBI-PGRP-TRPGR 0605135) awarded to Ronald Sederoff.

Please bookmark this website (www.fagaceae.org) to keep current on available data and progress on the Fagaceae Project

Chun-Huai Cheng

Please bookmark this website (http://genome.jgi-psf.org) to keep current on available data and progress on the *Cryphonectria* Genome Sequencing Project

JGI Tree of Life | Genome Projects | Login — **Cryphonectria parasitica,v1.0**

SEARCH | ADVANCED SEARCH | BLAST | BROWSE | GO | KEGG | KOG | DOWNLOAD | INFO | HOME | HELP!

photo: USDA Forest Service Archives

The ascomycete fungus *Cryphonectria parasitica* is the causal agent of chestnut blight. Introduced into North America from Asia before the turn of the 20th century, *C. parasitica* spread throughout the natural range of the American chestnut tree, destroying hundreds of millions of mature trees within a 50-year period. The discovery of viruses of *C. parasitica* that reduced the severity of the chestnut blight epidemic in Europe, a phenomenon termed hypovirulence, spurred research to develop hypovirulence-based biological control measures for North America. While the use of hypovirulent *C. parasitica* strains to treat individual disease cankers is highly effective, the spread of the virulence-attenuating viruses through North American *C. parasitica* populations has been hampered by a diverse self-nonself fungal recognition system termed vegetative incompatibility (*vic*). Current efforts to restore the American chestnut tree as a viable forest species emphasize classical breeding or genetic engineering of blight resistant trees, enhancing the efficacy of hypovirulence and integrating applications of both approaches.

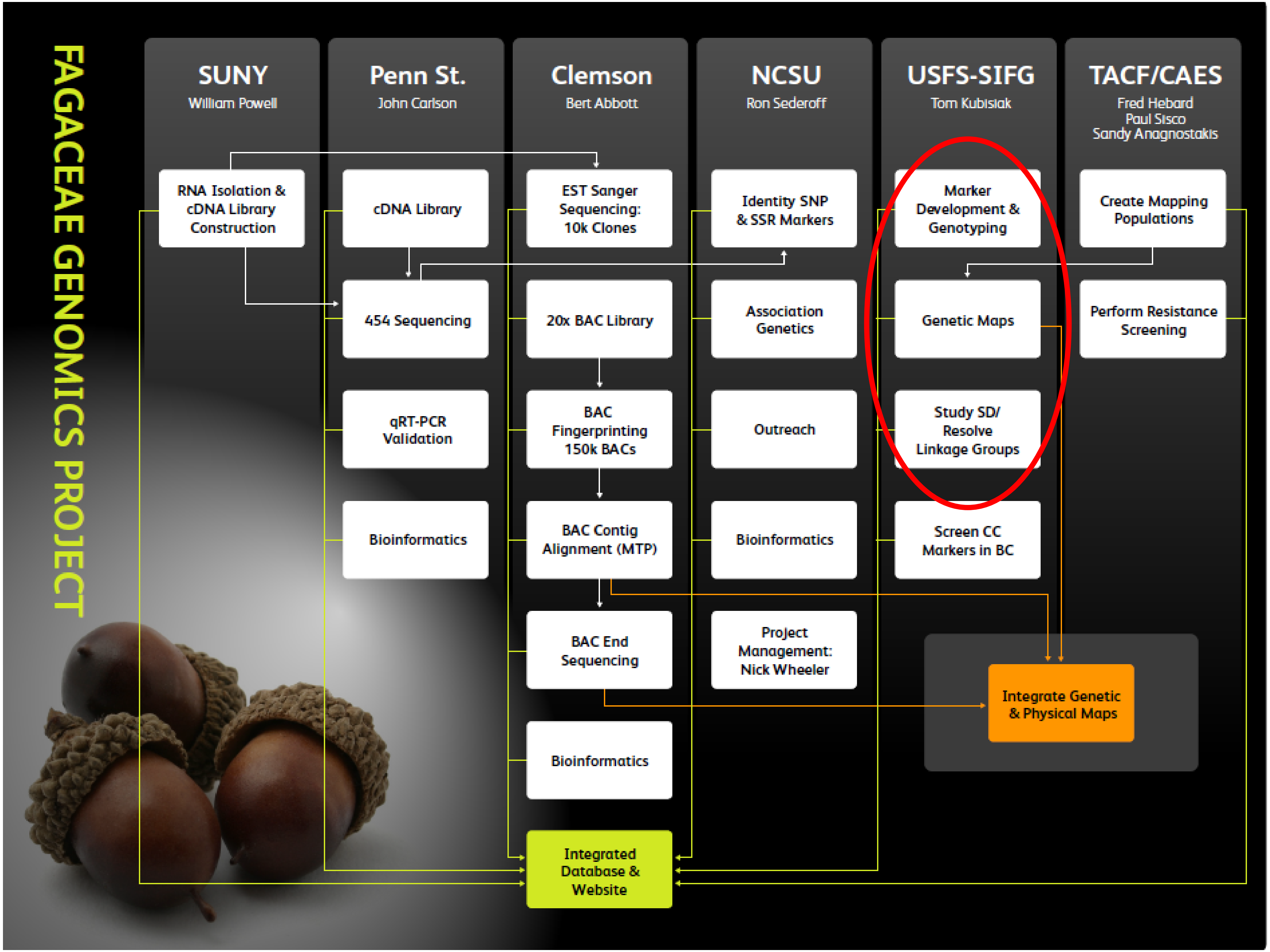
C. parasitica has a remarkable ability to colonize bark wounds on chestnut trees followed by the formation of a specialized structure called the hyphal fan, which efficiently penetrates through chemical and physical defense barriers and invades healthy cambial tissue under the bark. The fungus erupts from the bark after the formation of stromata, in which asexual and sexual fruiting bodies develop, releasing orange-pigmented spores to infect other trees. *C. parasitica* is also a tractable experimental organism for both classical and molecular genetics. It can mate in the laboratory, which facilitates genetic linkage analysis, and can be transformed with high efficiency. The level of homologous recombination during DNA-mediated transformation of spheroplasts has been increased to 85% with the development of a *C. parasitica* mutant strain defective in non-homologous end joining DNA repair. Because *C. parasitica* is haploid and asexual spores are uninucleate, gene disruption and selection of homokaryons after transformation is efficient. *C. parasitica* has been shown to support the replication of well-characterized viruses representing five different families: *Hypoviridae*, *Reoviridae*, *Narnaviridae*, *Partitiviridae* and *Chrysoviridae*. The development of infectious cDNA clones of several hypoviruses has provided a powerful system for the study of virus-host interactions and presented the means for engineering hypoviruses to enhance biological control potential.

Completion of the *C. parasitica* genome sequence will provide insights into the mechanisms underlying fungal colonization and penetration of tree host defense barriers. It will contribute to a molecular understanding of the vegetative incompatibility system, which will have significant implications for enhancing mycovirus-based biological control potential. It will also increase utility of a tractable experimental system for advancing studies on a wide variety of important biological processes. These range from signal transduction pathways underlying

Genomic Tool Development for the Fagaceae

Overall Goal: Develop genomic tools that will facilitate & accelerate the development of improved populations of Fagaceae species through tree breeding and genetic engineering

More specifically: perform large-scale gene and marker discovery, genetic and physical map construction & integration, and comparative genome analysis with a primary focus on *Castanea*


FAGACEAE GENOMICS PROJECT
SUNY
William Powell
Penn St.
John Carlson
Clemson
Bert Abbott
NCSU
Ron Sederoff
USFS-SIFG
Tom Kubisiak
TACF/CAES
Fred Hebard
Paul Sisco
Sandy Anagnostakis
RNA Isolation & cDNA Library Construction
cDNA Library
EST Sanger Sequencing: 10k Clones
Identity SNP & SSR Markers
Marker Development & Genotyping
Create Mapping Populations
454 Sequencing
20x BAC Library
Association Genetics
Genetic Maps
Perform Resistance Screening
qRT-PCR Validation
BAC Fingerprinting 150k BACs
Outreach
Study SD/ Resolve Linkage Groups
Bioinformatics
BAC Contig Alignment (MTP)
Bioinformatics
Screen CC Markers in BC
BAC End Sequencing
Project Management: Nick Wheeler
Integrate Genetic & Physical Maps
Bioinformatics
Integrated Database & Website

Genetic Mapping Pedigrees for *Castanea*

THE AMERICAN CHESTNUT FOUNDATION

Species Mapping populations

C. mollissima

	two crosses	**~800 progeny**

C. dentata

	two crosses	**~400 progeny**

QTL Mapping populations

Expanded F_2 population	**~135+ progeny**
BC_3 populations	**~1000+ progeny**

THE AMERICAN CHESTNUT FOUNDATION

Genetic Mapping Pedigrees for *C. mollissima*

Vanuxen ♀ x Nanking ♂ x Mahogany ♀

↓ ↓

Progeny Available	**VxN 440**	**MxN 367**
Progeny Mapped	**189**	**189**

Populations archived at CAES and TACF research farms

Summary of 454 transcriptome sequencing of Fagaceae species



Sample	Species	System	# Plates	# of Reads	# of bp	AL of Reads	# Contigs	AL of All Contigs	# Large Contigs
ACCanker	C dentata	GS20	1	129,508	13,080,308	101	7,171	168	247
ACHS1	C dentata	FLX	3/4	222,939	29,828,910	247	11,496	276	885
ACHS2	C dentata	FLX	3/4	254,810	38,165,054	246	9,431	271	691
ACWP1	C dentata	FLX	1/4	47653	11,380,607	238.8	4723		239
ACWP2	C dentata	FLX	1/4	33288	7,364,484	221	3793		239
CCCanker	C mollissima	GS20	1	235,635	23,799,135	101	14,308	168	436
CCMHS	C mollissima	FLX	3/4	228,594	56,051,191	246	21,828	344	3,074
CCNHS	C mollissima	FLX	3/4	259,859	64,271,926	247	28,784	339	4,451
CCWP1	C mollissima	FLX	1/4	60,445	14,643,040	242	4553		307
CCWP2	C mollissima	FLX	1/4	53,939	13,249,397	246	6348		485
Various	C mollissima	Sanger	NA	8101	4,348,356	526	NA		
ABWP1	F grandifolia	FLX	1/4	35,424	8,112,694	174	4163	229	174
ABWP2	F grandifolia	FLX	1/4	28,829	6,599,184	229	2467	229	83
WOA	Q alba	GS20+flx	1/4	119,385	28,669,562	241	14674		1058
WOB	Q alba	GS20+flx	1/2	83,821	19,297,054	231.5	8025		645
ROA	Q rubra	FLX	1/2	131,630	32,092,278	242	15494		1181
ROB	Q rubra	FLX	1/2	145,524	34,237,369	235	16138		1274
Total		both	5	2,184,941	428,799,020	198	93.018	261	9,784

88.4 Mb – 45,288 contigs

172 Mb – 48,335 contigs

14.6 Mb – 8,319

47.9 Mb- 22,102 contigs

56.3 Mb – 28,041 contigs

BMC Plant Biology

BioMed Central

Research article

Open Access

Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection

Abdelali Barakat*[1], Denis S DiLoreto[1], Yi Zhang[1], Chris Smith[2], Kathleen Baier[3], William A Powell[3], Nicholas Wheeler[2], Ron Sederoff[2] and John E Carlson*[1]

Address: [1]The School of Forest Resources, Department of Horticulture, The Huck Institutes of the Life Sciences, The Pennsylvania State University, 323 Forest Resources Building, University Park, PA 16802, USA, [2]Forest Biotechnology Group, North Carolina State University, Raleigh, North Carolina 27695, USA and [3]Department of Environmental Science and Forestry, State University of New York, Syracuse, NY, USA

Email: Abdelali Barakat* - aub14@psu.edu; Denis S DiLoreto - dsd134@psu.edu; Yi Zhang - yuz10@psu.edu; Chris Smith - chris@statgen.ncsu.edu; Kathleen Baier - kbaier@syr.edu; William A Powell - wapowell@esf.edu; Nicholas Wheeler - nickwheeler@scattercreek.com; Ron Sederoff - ron_sederoff@ncsu.edu; John E Carlson* - jec16@psu.edu

* Corresponding authors

Differential expression of the top 50 candidate genes is being validated with quantitative PCR (RT-PCR)

SSR Marker Development

in silico SSR polymorphisms were identified in stages using slightly differing Strategies and data:

1) combined partial 454 dataset from all Fagaceae species - potential for wide-use
486 primer pairs developed

2) Chinese chestnut 454 data + Sanger Data – specifically for identifying markers poly for mapping in *C. mollissima*
461 nonredundant primer pairs developed

SSRs for the Fagaceae

1) **combined (partial) 454 dataset**
   486 primer pairs

   Screened against Fagaceae species panel – 2 genotypes for each of six species

   by project completion all markers will be further characterized in five species using segregating populations and genotypes from natural populations

Number of candidate SSRs potentially useful within/across species

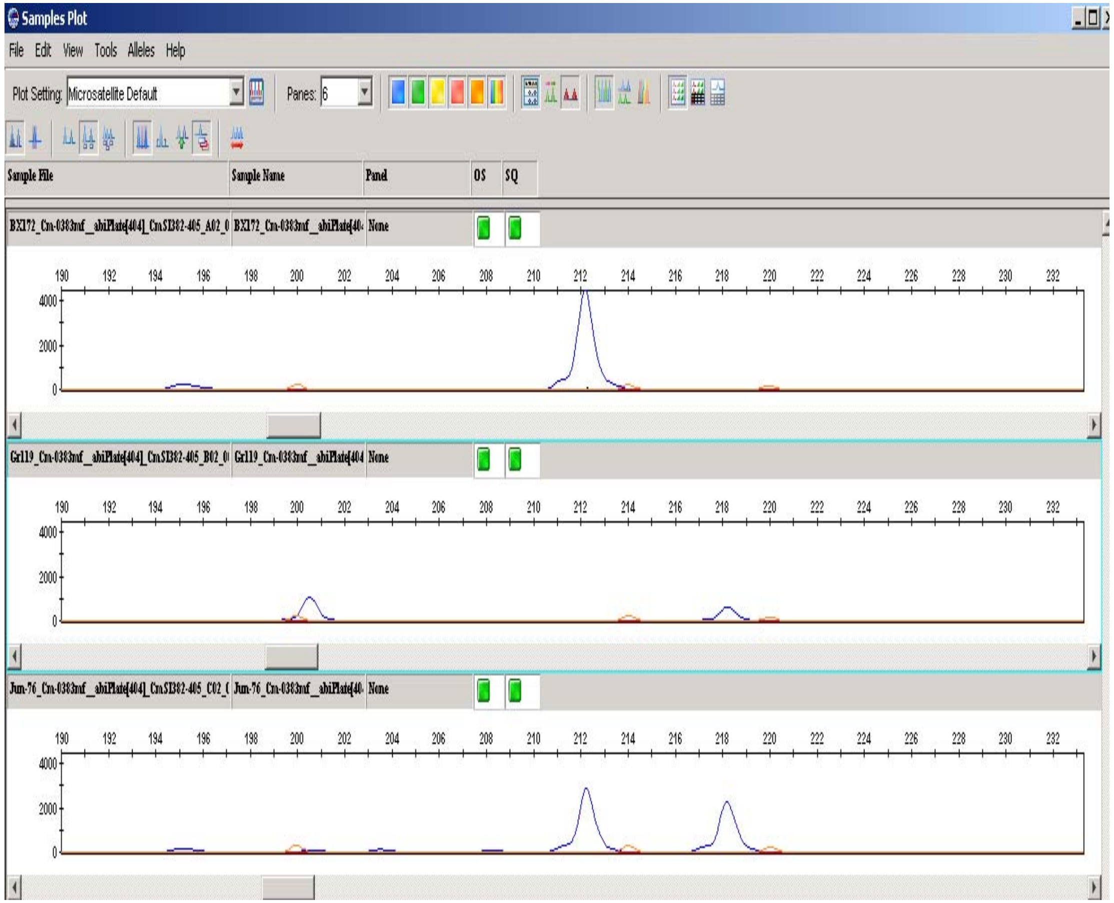
275 SSRs amplified	*Castanea mollissima*	*Castanea dentata*	*Castanea sativa*	*Quercus rubra*	*Quercus michauxii*	*Fagus grandifolia*
poly	114	109	128	57	77	38
mono	146	145	126	111	154	101

SSRs for Mapping in *C. mollissima*

2) Chinese chestnut 454 + Sanger data
461 nonredundant primer pairs

Characterized in Chinese chestnut mapping parents

381 Amplified - 295 heterozygous in at least one of our three Chinese parents



SSRs For Mapping in *C. mollissima*

1) Fagaceae combined partial 454 datasets
~115 markers heterozygous

2) Chinese chestnut 454 data plus Sanger data
~295 markers heterozygous

~400 Markers Useful for Mapping

Finishing Stages of Data Collection

SNPs for Mapping in *C. mollissima*

```
CCNH2_033757_2659_36 AACGCTGCCAcaa*cccgACCGCAAAGAGTTCACCAAGGTGGCTTTCCGTACGGCCA
CCNH2_089031_3765_03                        agAGTTCACCAAGGTGGCTTTCCGTACGGCCA
CCNH2_127719_3625_31 AACGCTGCCACAAACCCGACCGCAAAGAGTTCACCAAGGTGGCTTTCCGTACGGCCA
CCNH2_138069_3953_28 AGCGCTGCCACAAGCCCGATCGCAAAGAATTCACGAAAGTGGCGTTCCGTACAGCGA
CCNH2_175475_3494_32 AACGCTGCCACAAACCCGACCGCAAAGAGTTCACCAAGGTGGCTTTCCGTACGGCCA
CCNH2_199372_2304_05 AGCGCTGCCACAAGCCCGATCGCAAAGa*tTCACgaa*gTGGCGTTCCGTACAGCGA
CCNH2_208053_3428_33 AACGCTGCCACAAACCCGACCGCAAAGAGTTCACCAAGGTGGCTTTCCGTACGGCCA
CCNH2_285405_3694_37 AGCGCTGCCACAAGCCCGATCGcaa*ga**tcac**aagTGGCGTTCCGTACAGCGA
CCNH2_364692_2362_20 AACGCTGCCACAAACCCGACCGCAAAGAGTTCACCAAGGTGGCTTTCCGTACGGCCA
CCNH_017485_4011_272 AGCGCTGCCACAAGCCCGATCGCAAAGAATTCACGAAAGTGGCGTTCCGTACAGCGA
CCNH_028437_3647_262 AGCGCTGCCACAAGCCCGATCGCAAAGAATTCACGAAAGTGGCGTTCCGTACAGCGA
CCNH_159608_3820_129                                cgAAAGTGGCGTTCCGTACAGCGA
CCNH_183801_3502_318 AGCGCTGCCACAAGCCCGATCGCAAAGAATTCACGAAAGTGGCGTTCCGTACAGCGA
CCNH_200954_4011_245 AACGCTGCCACAAACCCGACCGCAAAGAGTTCACCAAGGTGGCTTTCCGTACGGCCA
```

POLYBAYES - PINESAP - ILLUMINA analysis

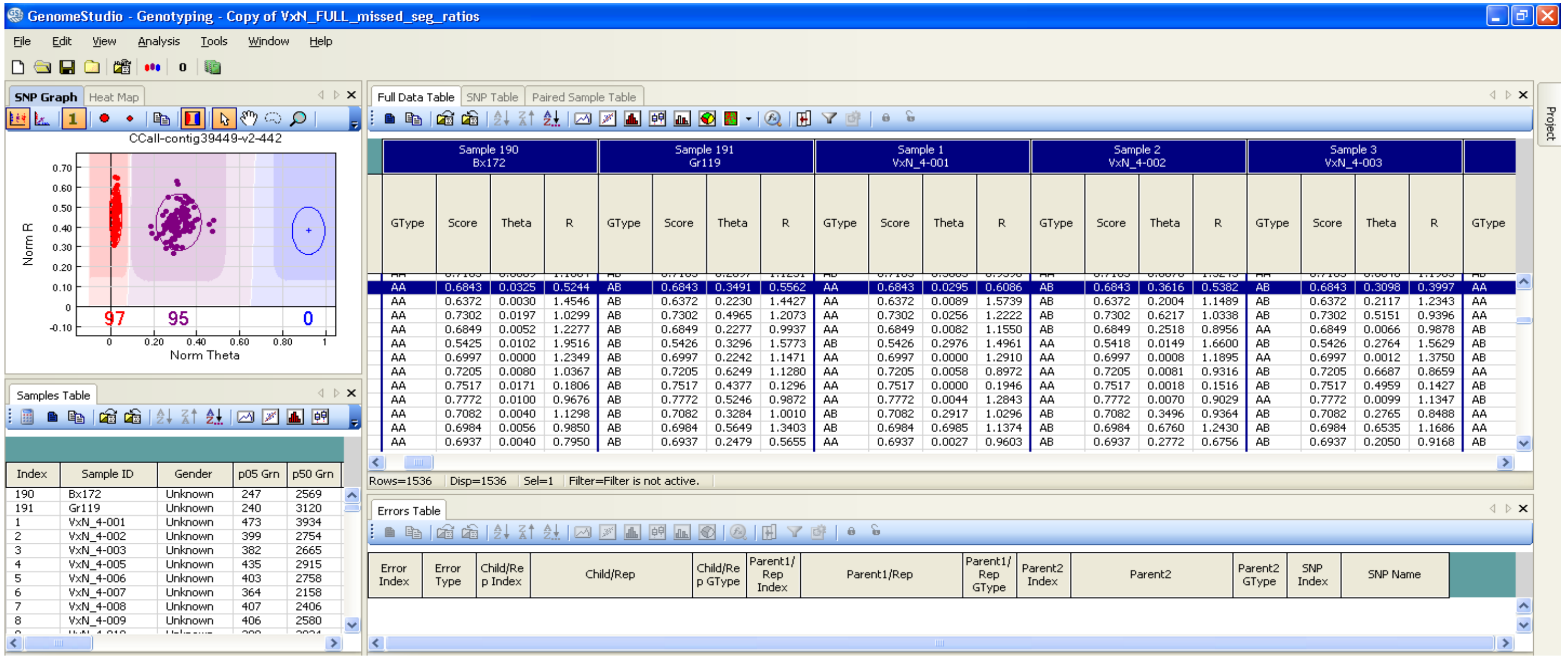
- **25,904 SNPs (prob{true SNP}>0.70)**
  - **Allele quality scores**
    - **Depth of reads**
    - **Base quality scores**
    - **Base composition in region of interest**

- **1731 chosen for SNP Array**
  - **205 in candidate defense response genes**
  - **1188 heterozygous in Nanking**
  - **338 other w/high allele quality scores**

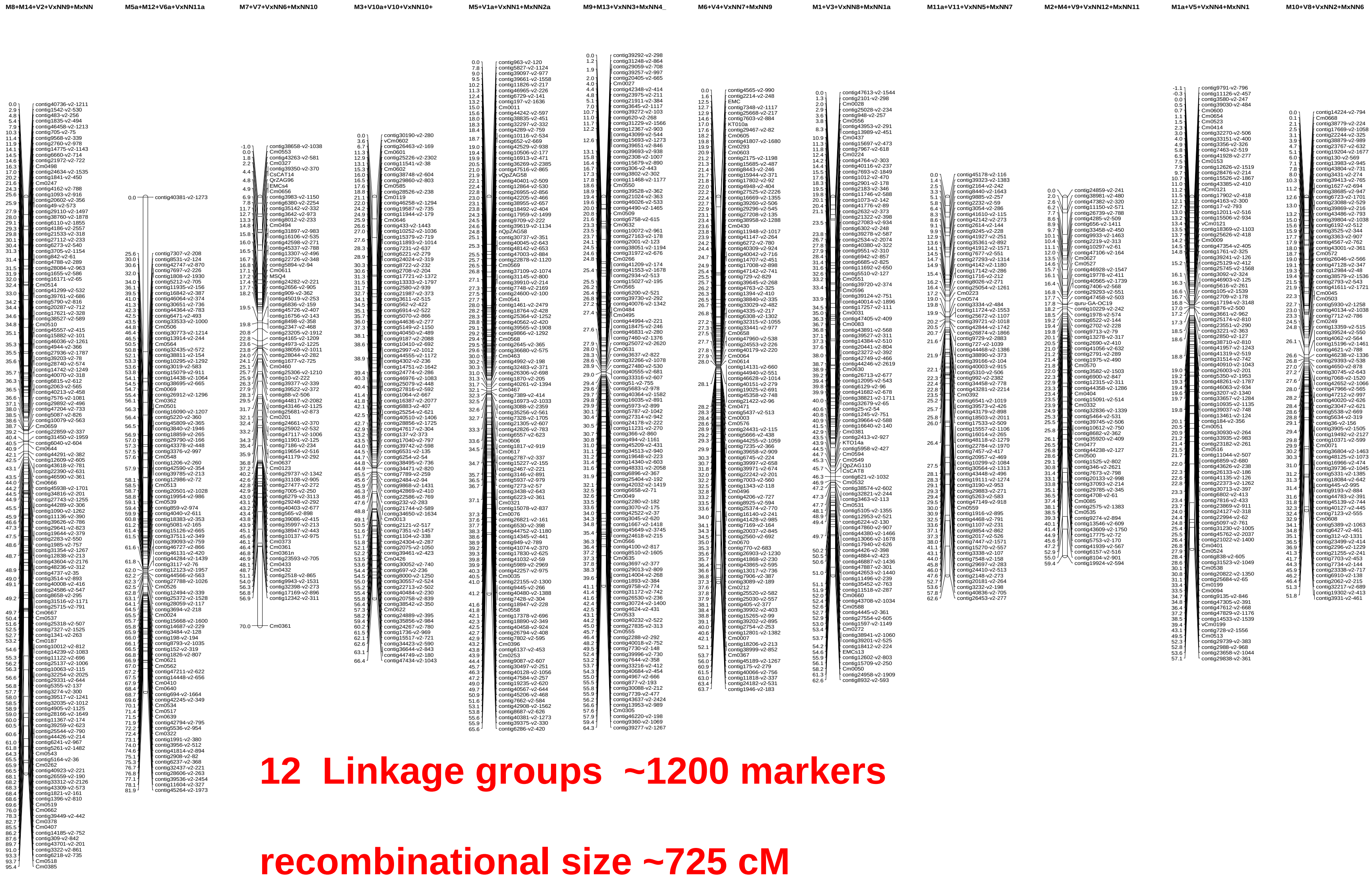
- **1536 chosen by Illumina for array design using the goldengate assay**

SNP markers Useful for Mapping in *C. mollissima*

VxN		MxN		Performance
1536		1536		
224	not clustered	224	not clustered	15% NC
275	Not Segregating	320	Not Segregating	19% NS
33	Fixed Alt Alleles	21	Fixed Alt Alleles	2% FA
141	Het in Vanuxem	99	Het in Mah	64% S
259	Het in Both	280	Het in Both	
604	Het in only Nank	592	Het in only Nank	



Current *C. mollissima* Map Stats



12 Linkage groups ~1200 markers

recombinational size ~725 cM

average marker spacing ~0.6 cM

Current Activities with Regards to Genetic Mapping in *Castanea*

Finishing SSR data scoring and performing post –analysis marker QA/QC

Annotate all markers placed on maps

Place previously identified rQTL within current marker framework

Combining genetic and physical maps (Eric Fang)

examining micro- and macro-scale synteny with other model plant species (Meg Staton)

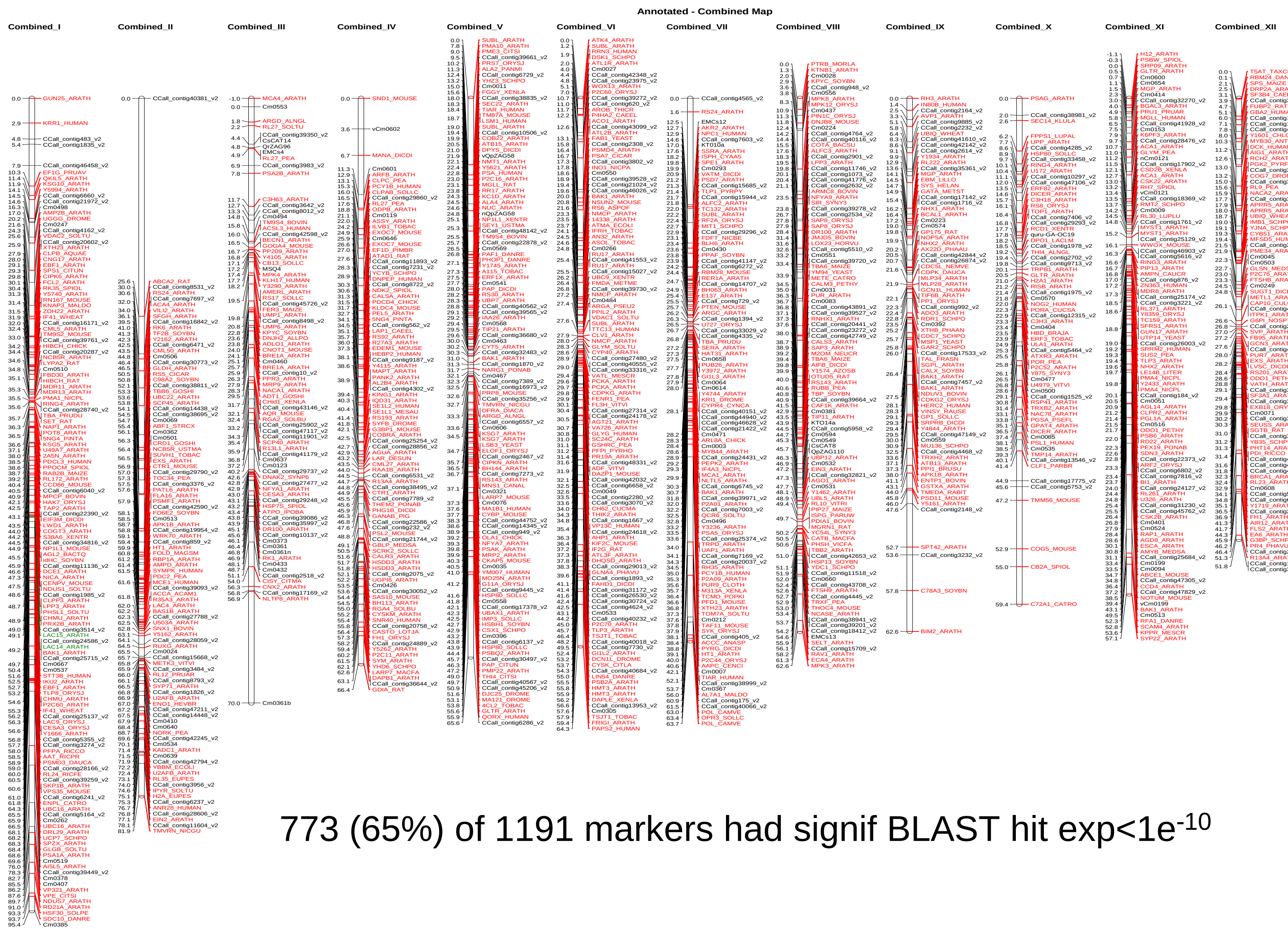
Testing Translocation Hypothesis in *Castanea*

Post –Analysis Marker QA/QC

-confirm scoring of markers with inconsistent marker order across parental maps

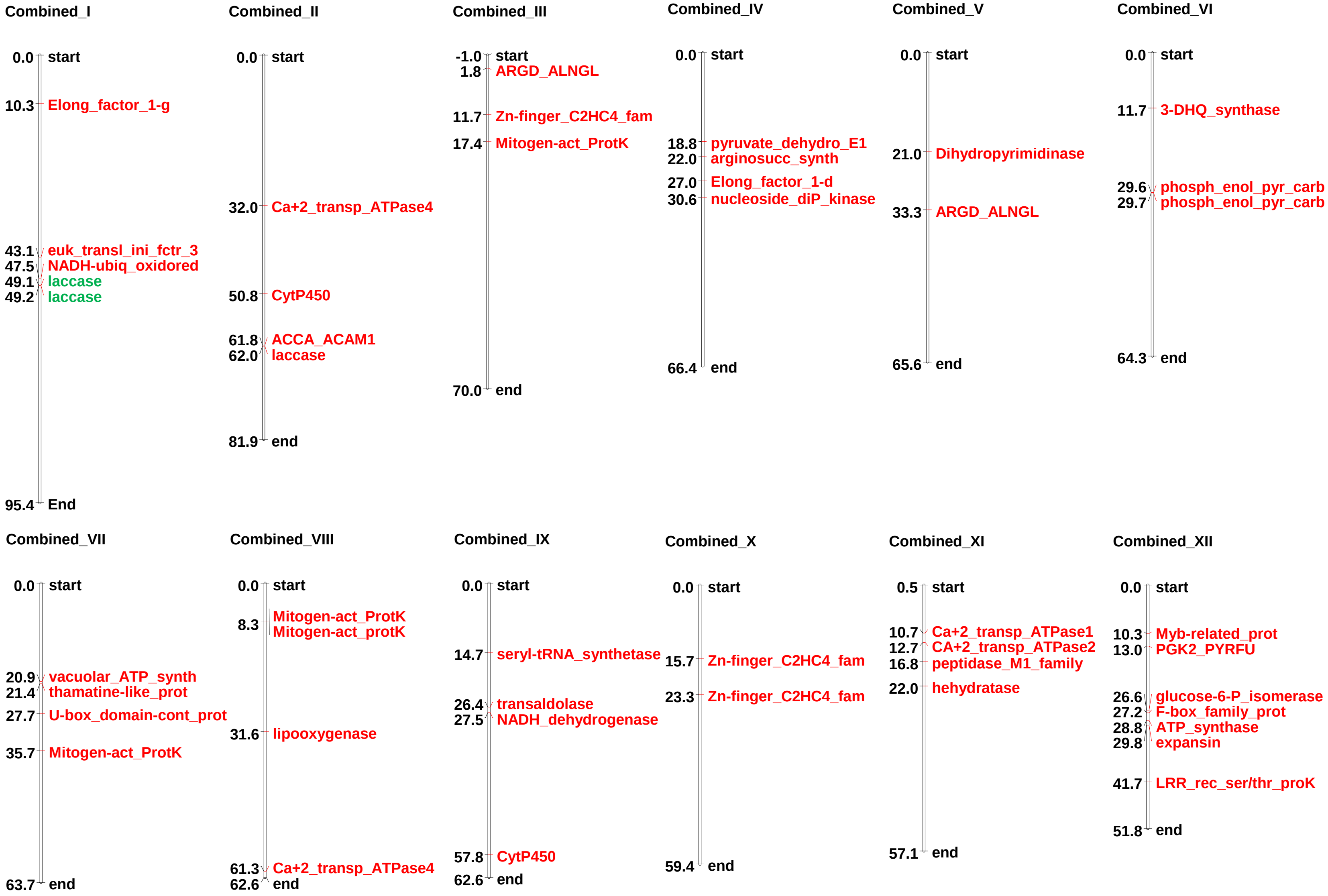
-confirm all observed double recombinants

Annotated Map for *C. mollissima*



773 (65%) of 1191 markers had signif BLAST hit exp<1e^{-10}

Candidate Defense-Response Genes (Barakat et al. 2009)



Place rQTL within existing marker framework

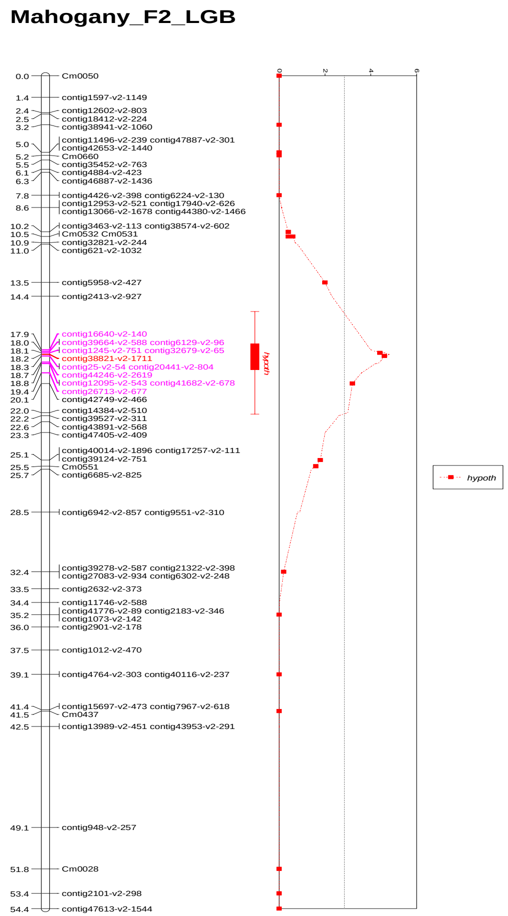
Place previously identified resistance QTL within current marker framework

F2 population (1997)

1536 SNP Array

	#Het Markers
4_31	52
4_52	99
Both	273

SSR ~400 Candidates



*hypothetical example

*****is there a correlation between rQTL and candidate disease response genes*****

Conserved Orthologous Sequences (cosII)

micro- and macro-scale synteny with other model plant species

COSII markers are PCR-based markers developed from a set of single-copy conserved orthologous genes (COSII genes) in Asterid species - Wu et al. Genetics 174(3):1407-20

2869 Arabidopsis COSII genes used to anchor COSII groups in Asterid species - we had significant BLAST hits for 1,555 contigs/genes in our *C. mollissima* data

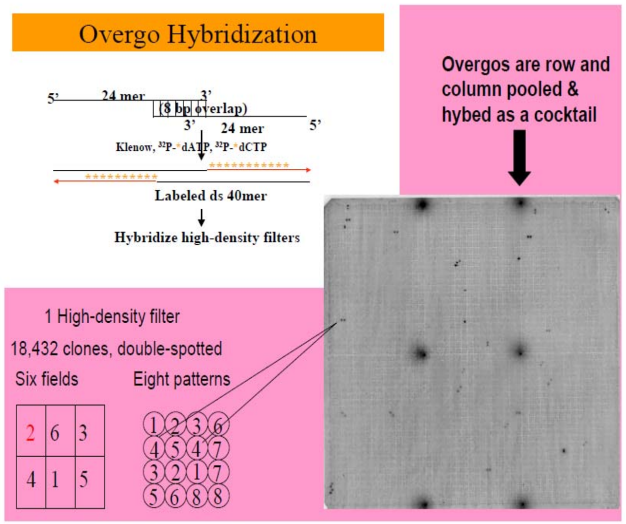
Of these 1,555 genes - 95 were placed on the *C. mollissima* genetic map

Of the 95 that mapped – only 36 appeared to be single copy in *C. mollissima*

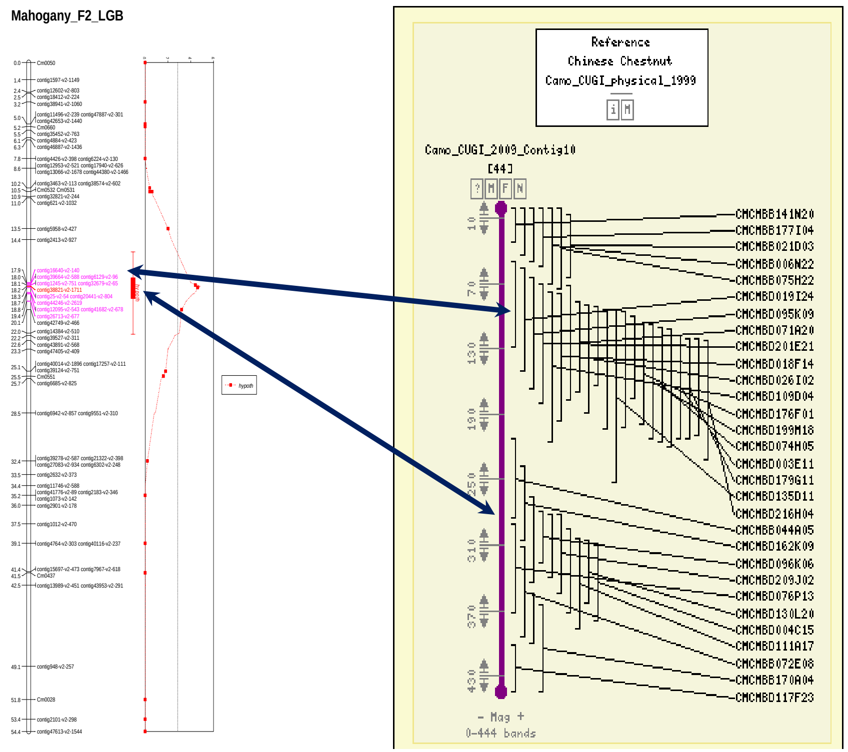
For these 36 contigs some level of microscale synteny was observed and needs to be further investigated

Combining Genetic and Physical Maps

-Overgos are being designed for as many mapped markers as possible and being used to probe BAC filters to combine the genetic and physical maps



Overgo hybridization scheme



rQTL

Physical BAC contigs

Testing the hypothesis for gross chromosomal rearrangements between American and Chinese chestnut

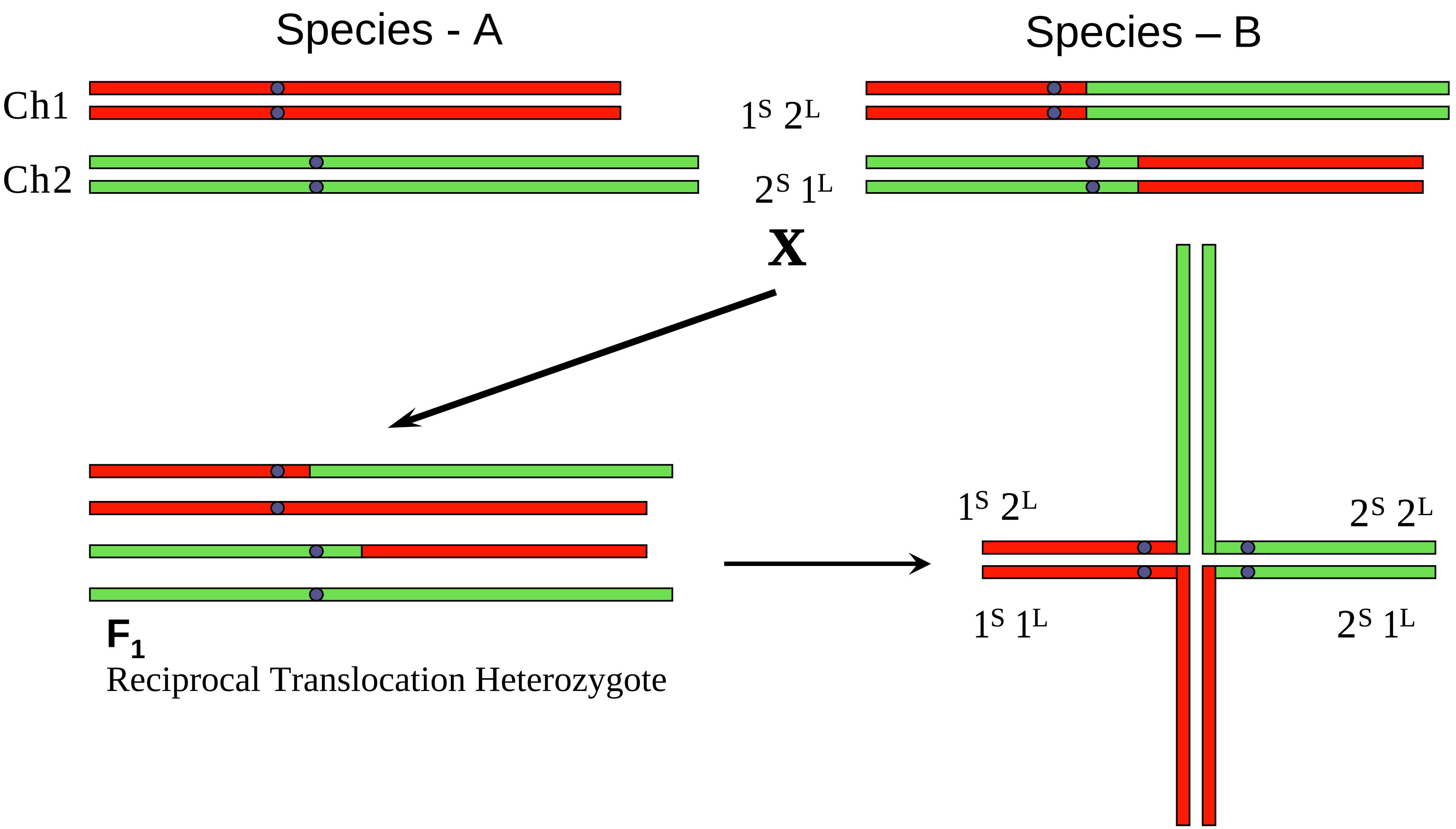
Previous genetic linkage data suggested that there may be gross chromosomal differences between American and Chinese chestnut genomes

1)Substantial segregation distortion on some LGs
2)High LODs required to obtain expected 12 LGs

Hypothesis proposed: chromosomal rearrangments

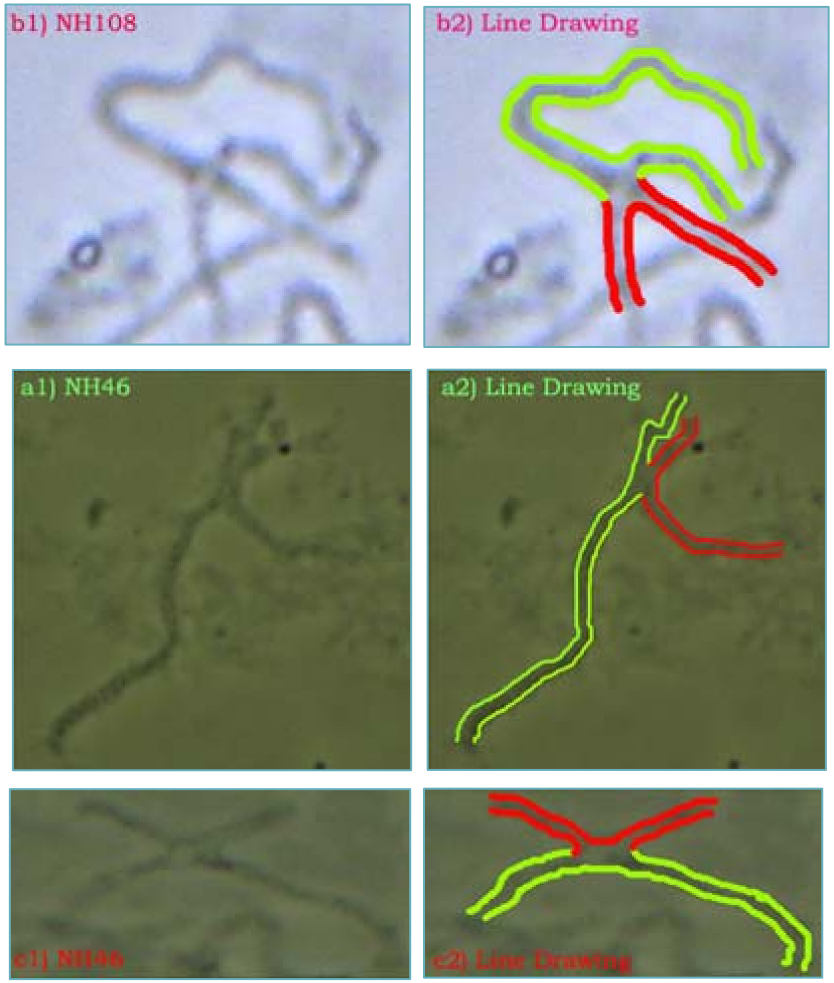
Importace: if translocations are occurring on chromosomes harboring rQTL will affect introgression of these regions into American chestnut

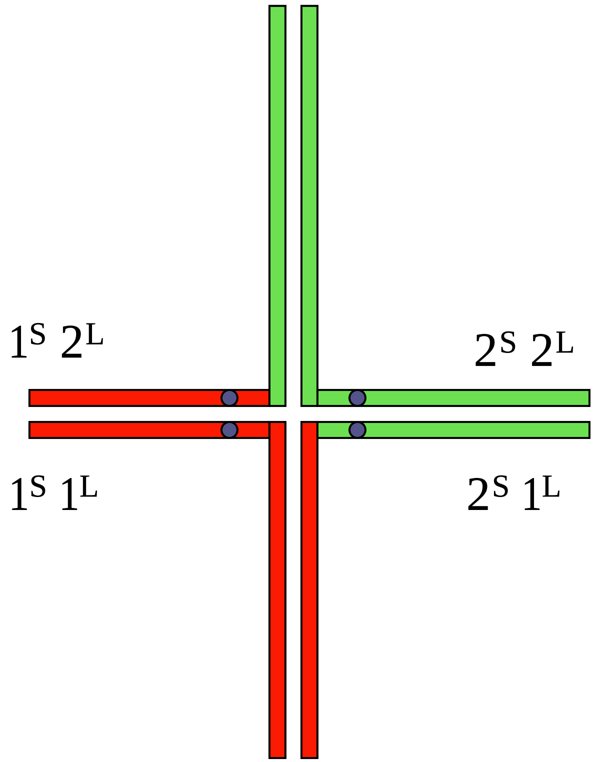
Reciprocal Translocation



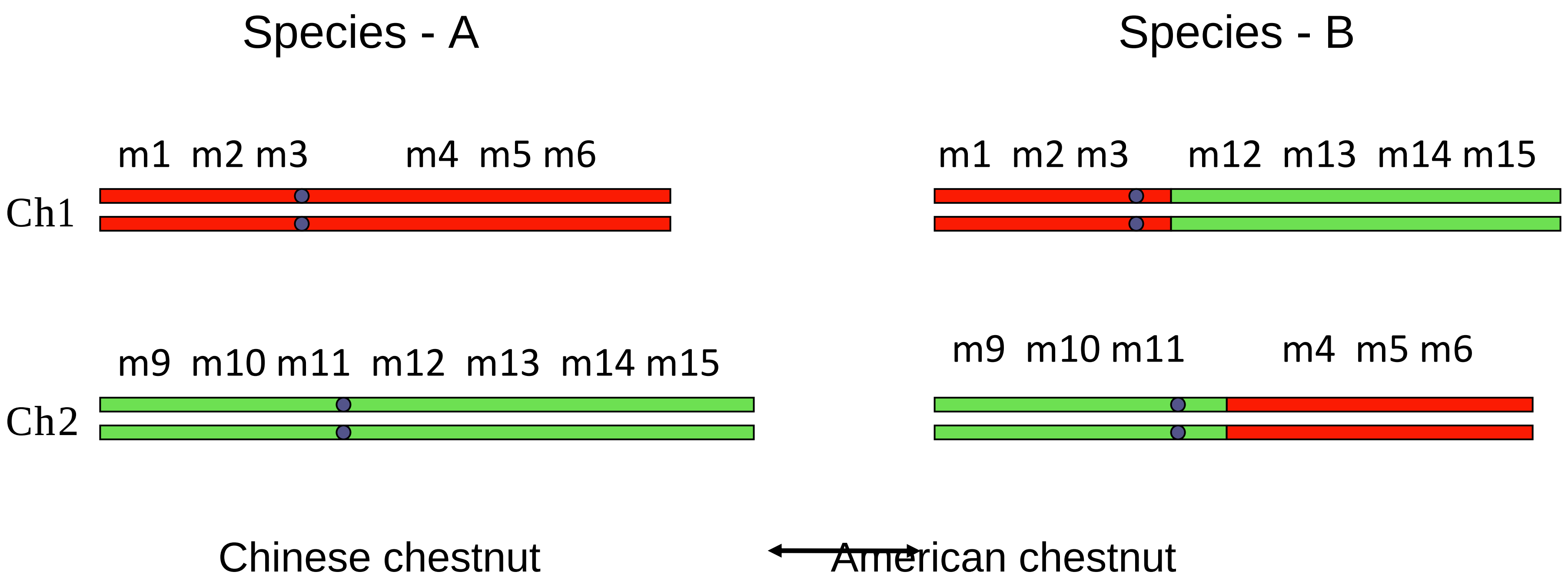
Diagrammatic representation of a reciprocal translocation heterozygote

Pachytene – Quadrivalent Association





Reciprocal Translocation
Diagrammatic Representation



Diagrammatic representation for evidence of a reciprocal translocation using genetic markers


THANK YOU FOR YOUR ATTENTION!