

**ARTIFICIAL SELECTION AND THE GENOME: A DEEP
PEDIGREE ANALYSIS OF AN ELITE SOYBEAN CULTIVAR**

Chris M. GRAINGER C M AND RAJCAN I

Department of Plant Agriculture, University of Guelph, Guelph ON, N1G 2W1

E-mail: cgrainge@uoguelph.ca

Modern soybean [*Glycine max* (L.) Merr.] varieties are superior to their ancestors due to the effects of long-term artificial selection applied by breeders. The consequences of long-term breeder selection on the soybean genome are largely unknown in public soybean breeding programs. This knowledge is important for understanding various aspects of applied breeding such as genetic diversity, linkage and recombination and allele frequency changes due to selection. As pedigrees are a record of breeder manipulations, they represent selections that have become commercially successful. For Ontario (Canada) soybeans, one such elite variety was OAC Bayfield, which had exceptional commercial success. Furthermore, it is a parent of a number of successful varieties developed by multiple independent breeding programs. A total of 42 lines from six breeding programs, which comprise the multi-generational pedigree of OAC Bayfield, were genotyped with molecular markers to track chromosomal transmission throughout the pedigree. The genotype profile of the 20 chromosomes revealed substantial allelic structure that had been built up in certain chromosomes. This structure was in the form of specific linkage blocks, which were primarily a result of recombination involving ancestral linkage blocks or linkage blocks introduced from the cultivar Fiskeby-V. There were a number of instances where the allele composition remained conserved for entire chromosomes when tracking transmission from OAC Bayfield to various cultivars derived from it. A selective sweep analysis was performed using the members of OAC Bayfield's pedigree (both ancestors and progeny) to identify genomic regions that have retained a selective signature through OAC Bayfield in the varieties derived from it. Using a reverse-genetics strategy, these regions were compared to the genetic maps on SoyBase to determine what genes/quantitative trait loci (QTL) were present. Out of the 19 identified loci, 18 mapped to previously discovered QTL with yield being the trait with the highest number of QTL identified. Results from this study can aid soybean breeding efforts in various ways such as ensuring favorable linkages are maintained or identifying regions of fixation that may benefit from introgressive breeding strategies.

Artificial Selection and the Genome: A Deep Pedigree Analysis of an Elite Soybean Cultivar



Chris Grainger M.Sc.

Dr. Istvan Rajcan

**Soybean Breeding and Genetics
Department of Plant Agriculture
University of Guelph, Guelph ON Canada**



Presentation Outline

- Ontario Soybean Production
- Research Objectives
- Pedigree Genotyping
- Identifying Selection Signatures
- Current Research Activity

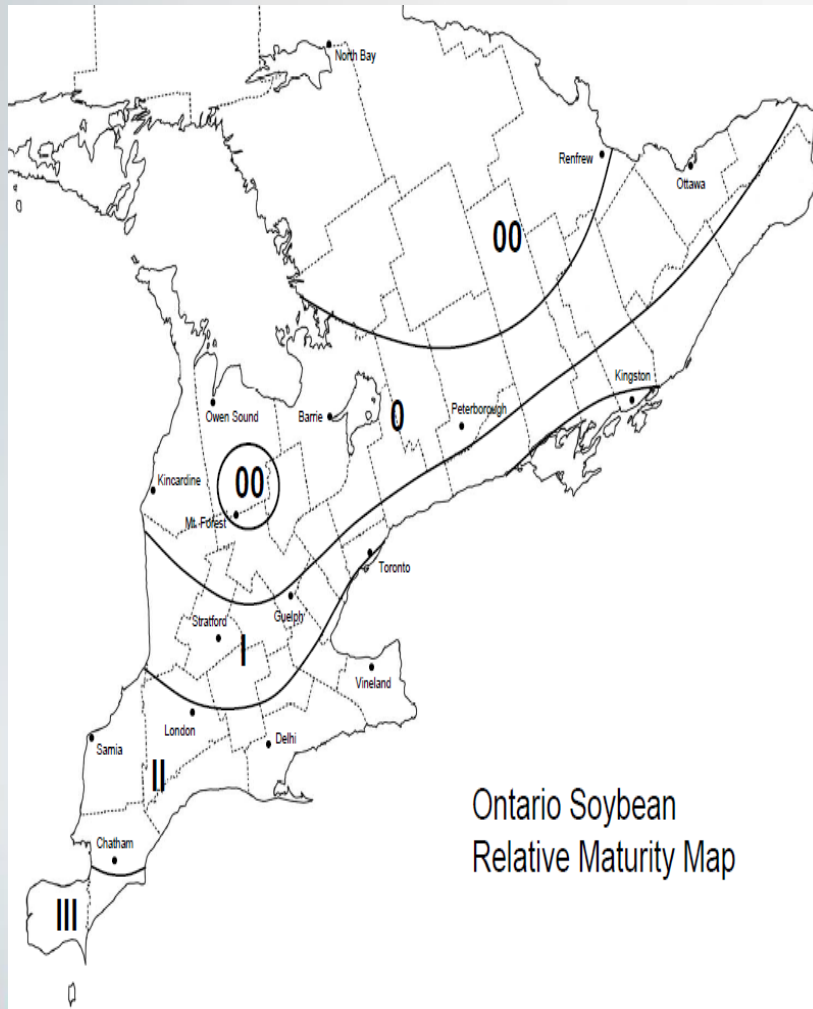


History of N.A. Soybean Breeding



- Introduced in North America in late 1700's.
- Charles Zavitz developed first variety in Canada in the 1920's (OAC 211).
- 35 ancestors contribute 95% of alleles in North American cultivars (Gizlice et al., 1994).
- 5 lines account for 65% of the genetic base for northern cultivars (Gizlice et al., 1994).

Soybean Breeding Progress

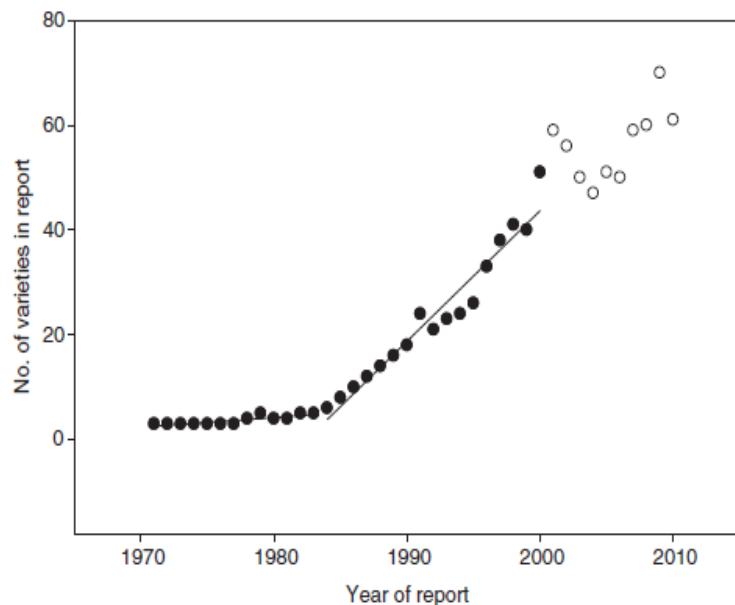


- Ontario ~65% of soybean production in Canada (CSC. 2011)
- From 1934 to 1992 yields have increased 0.5-0.7% per year (Voldeng et al. 1997).
- Concerns over genetic base and plateaus in genetic gain.
- Effects of long-term breeder selection within a breeding program?

OOPSCC, 2012

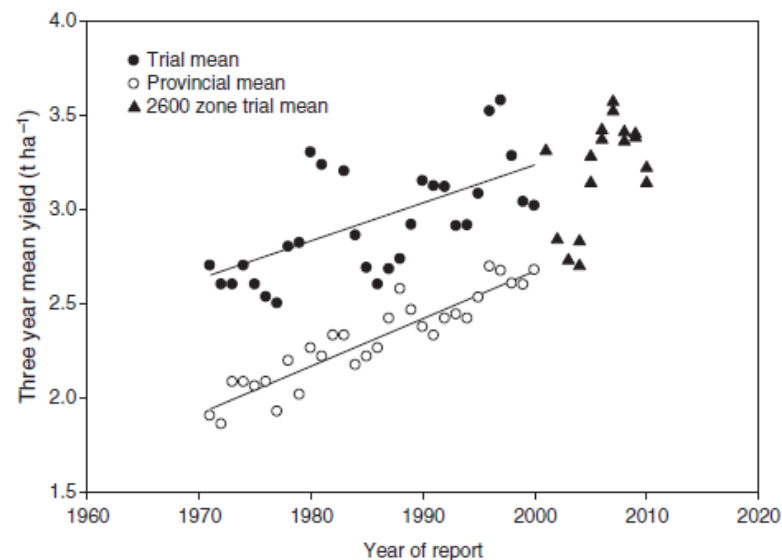
Retrospective Soybean Development

Breeding activity



36-fold increase

Yield improvement



1.4-fold increase

Cober and Voldeng. 2012. *Can. J. Plant Sci.*

The Success of OAC Bayfield

- OAC Bayfield released in 1994.
- In registration trials, out yielded check cultivars by 15%.
- 400 000 acres were grown in 1997 (20% of total crop).
- Contributed estimated \$750 million to Ontario's economy.
- Parent of OAC Kent/OAC Champion/OAC Wallace.

1998 ARIO Field Crops Review

Research Objectives

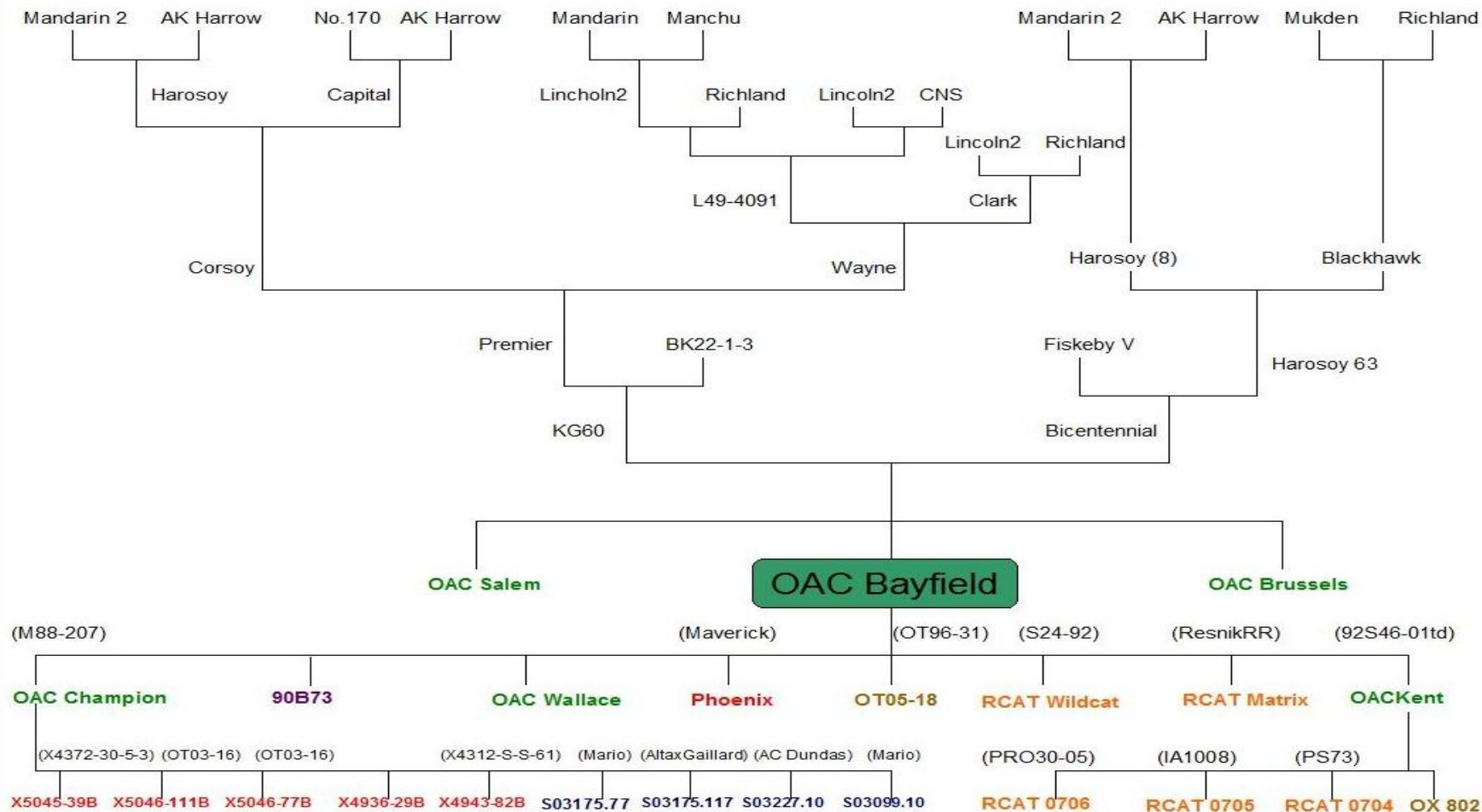


- Use molecular markers to track chromosome transmission throughout the pedigree of OAC Bayfield.
- Characterize the allelic structure that has been established over long-term artificial selection.
- Perform a selective sweep analysis to identify loci which may have experienced selection.
- Compare results to QTL studies reported in SoyBase.

Characterization of Genetic Diversity and Chromosome Composition and Transmission Through the Pedigree of OAC Bayfield



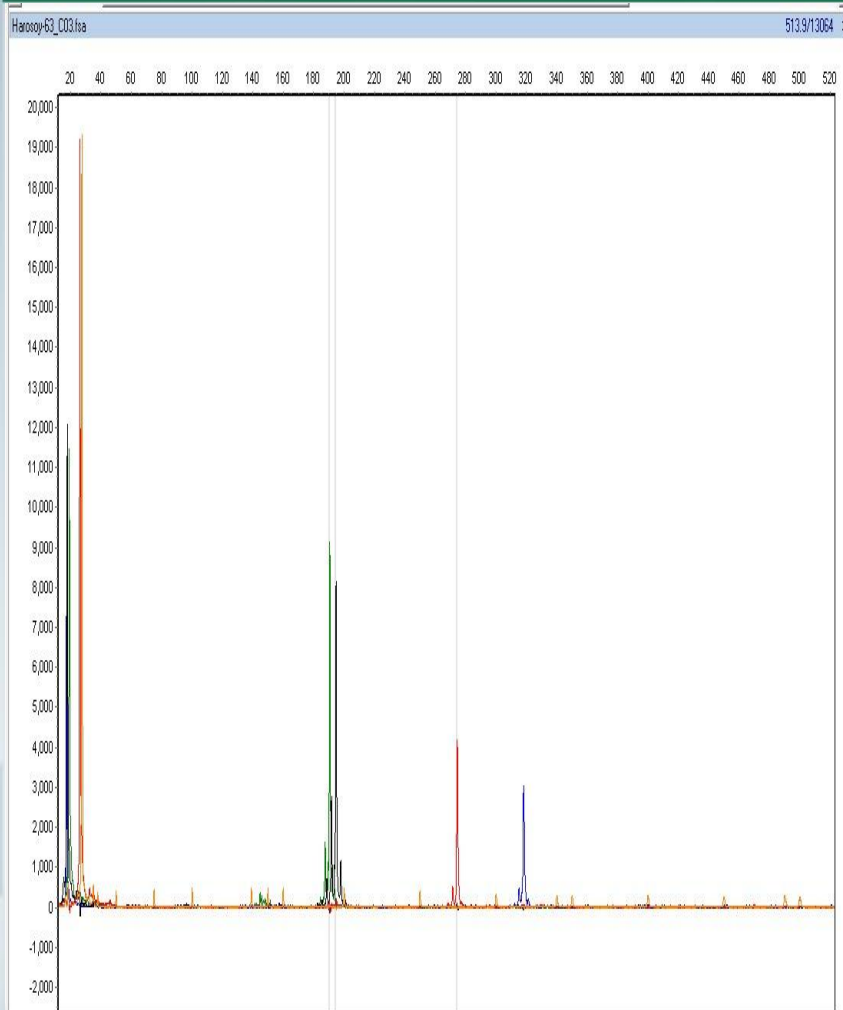
Pedigree of OAC Bayfield



University of Guelph
Pioneer Hi-Bred
Agriculture Canada

Ridgetown Campus
La coop Fédérée
Semences Prograin

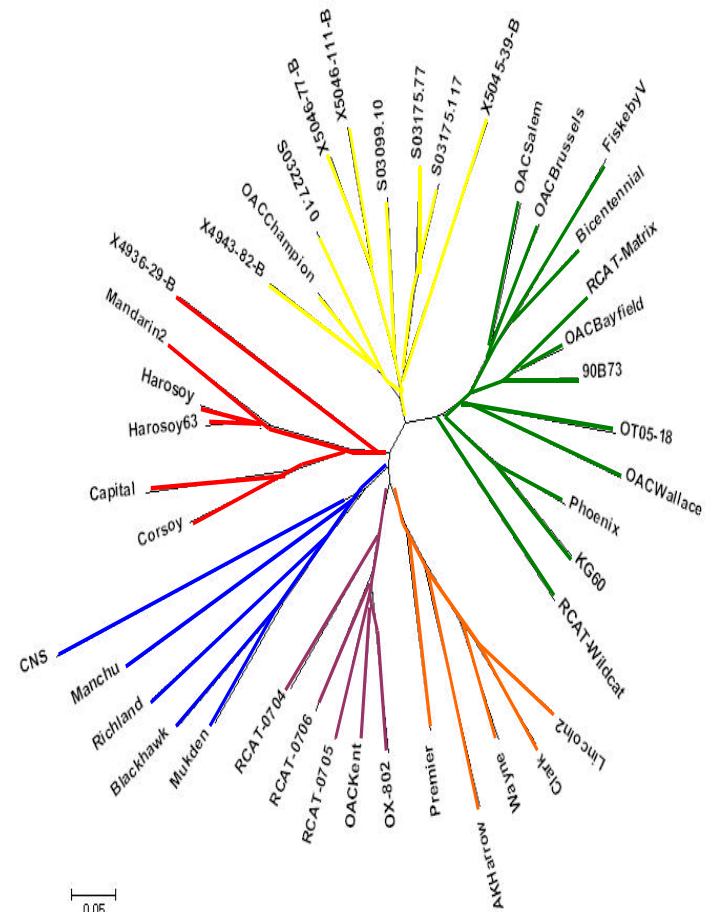
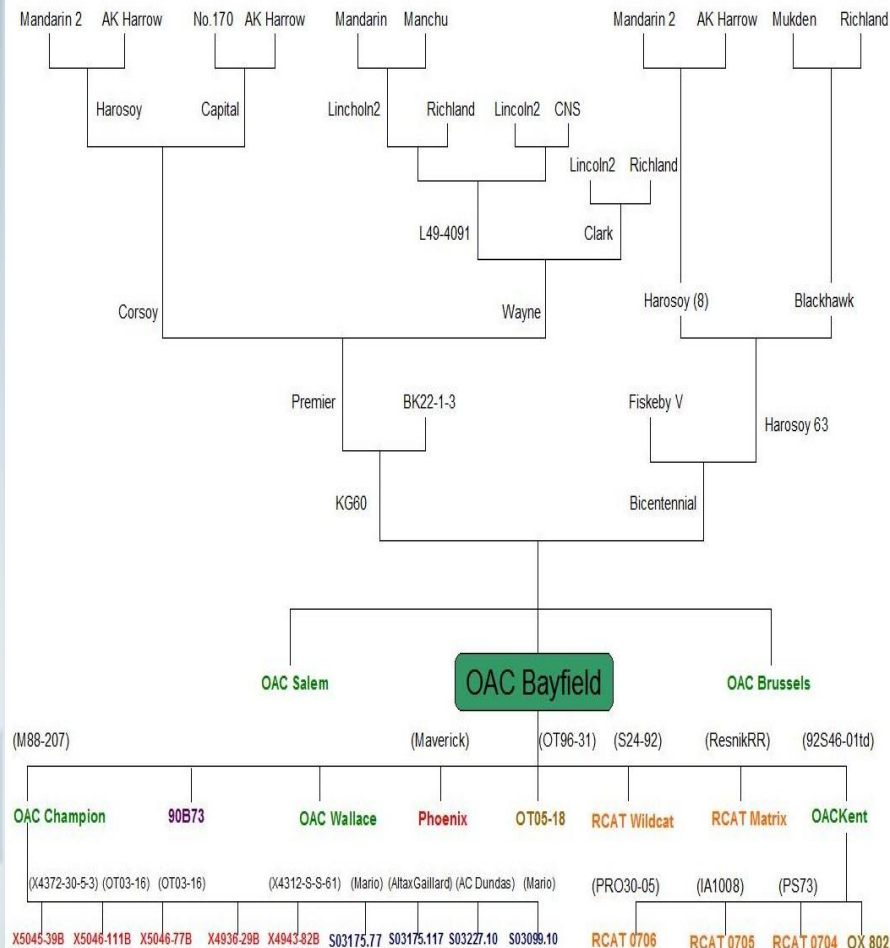
SSR Genome Scan



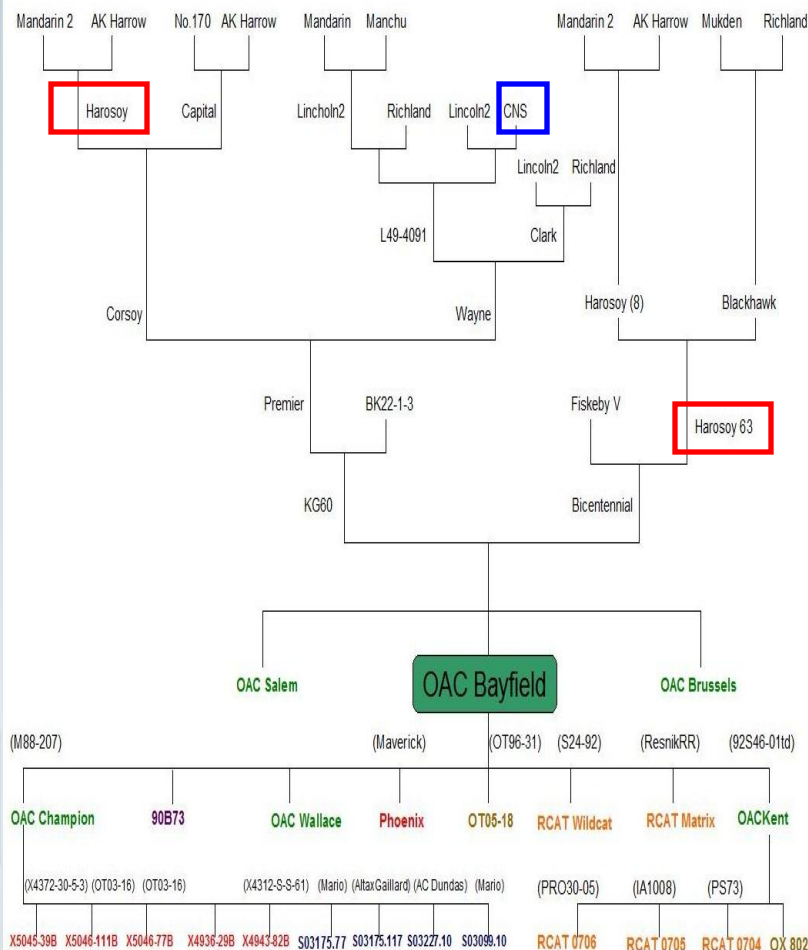
- SSR Genotyping done using ABI 3730.
- Allele calls performed using GeneMarker software.
- Genotype profiles were visualized using GGT software.
- Various statistics calculated using PowerMarker v3.25.



Marker-Pedigree Congruence



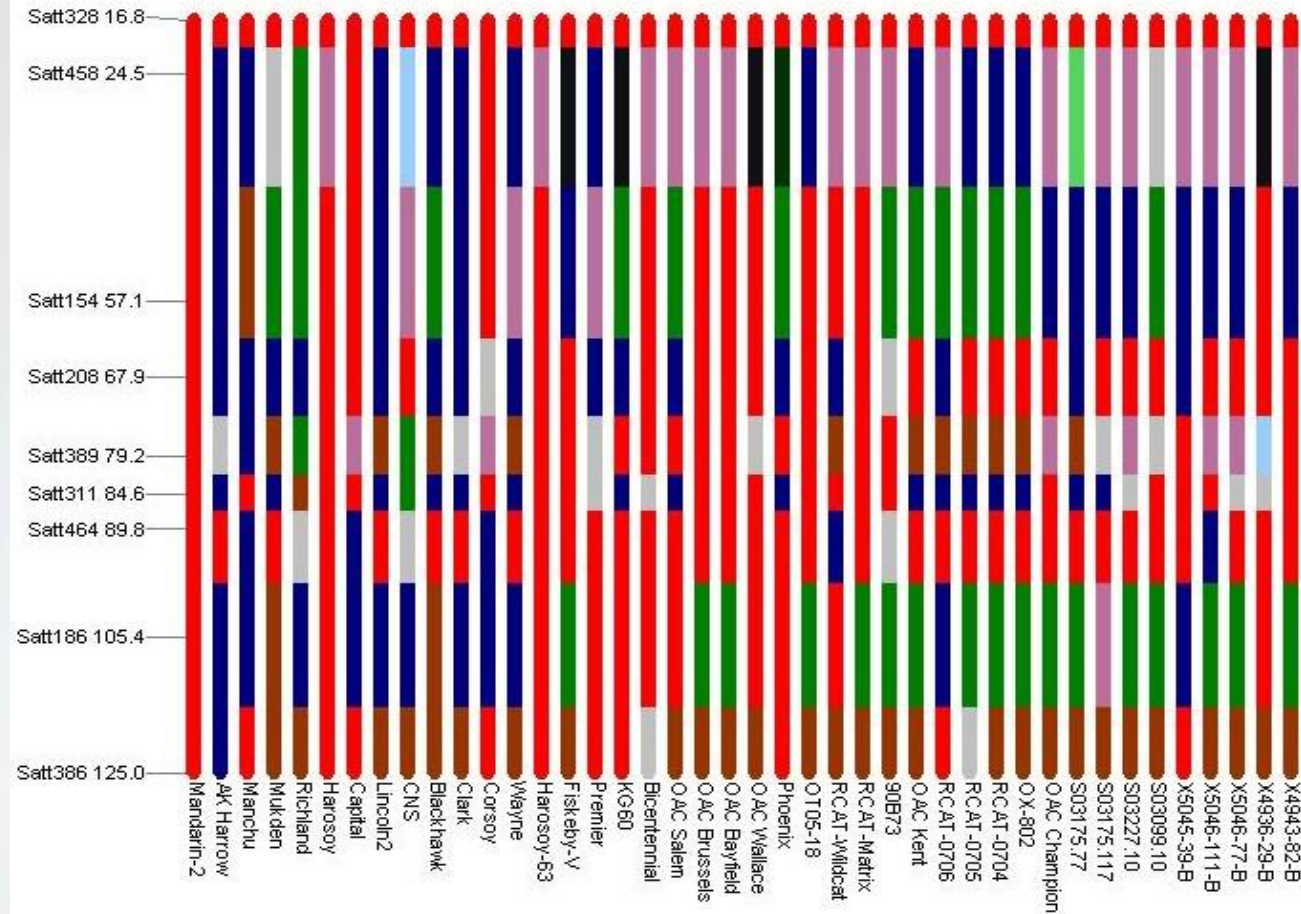
Genetic Diversity Summary



- Average of 4 alleles/marker.
- 14 loci with frequency > 0.95 .
- Most diverse line was CNS.
- Most genetically similar lines were Harosoy and Harosoy 63 (91% shared alleles).

Graphical Genotyping

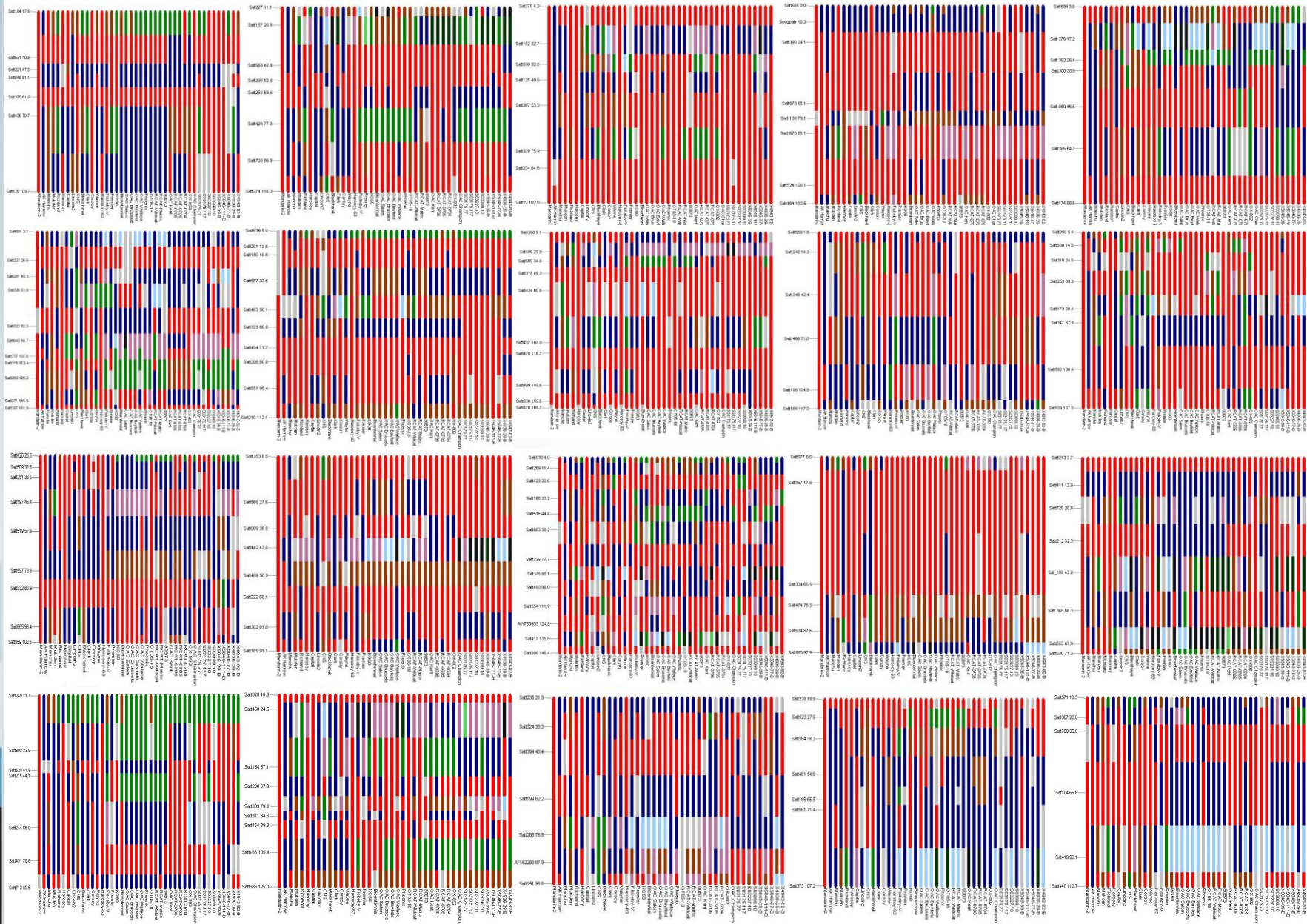
SSR Markers



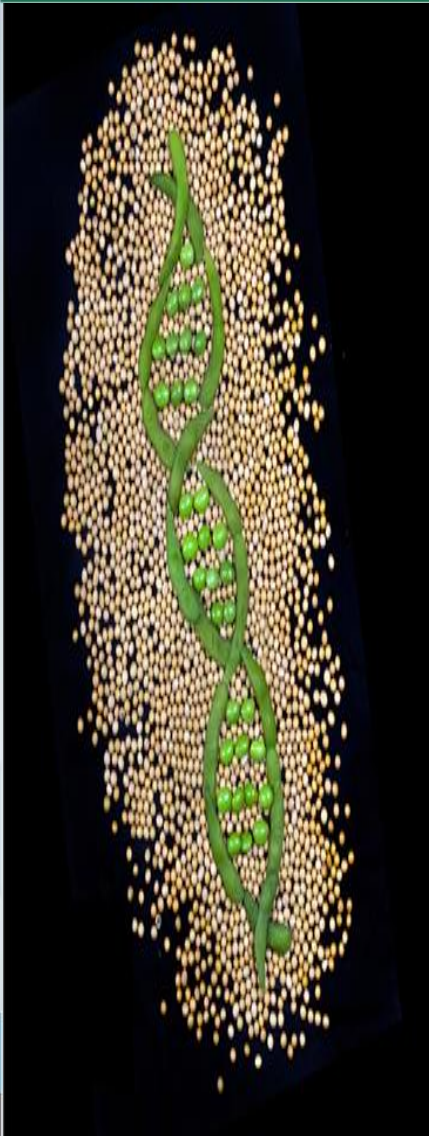
Genotypes



Graphical Genotyping



Linkage and Recombination: Soybean



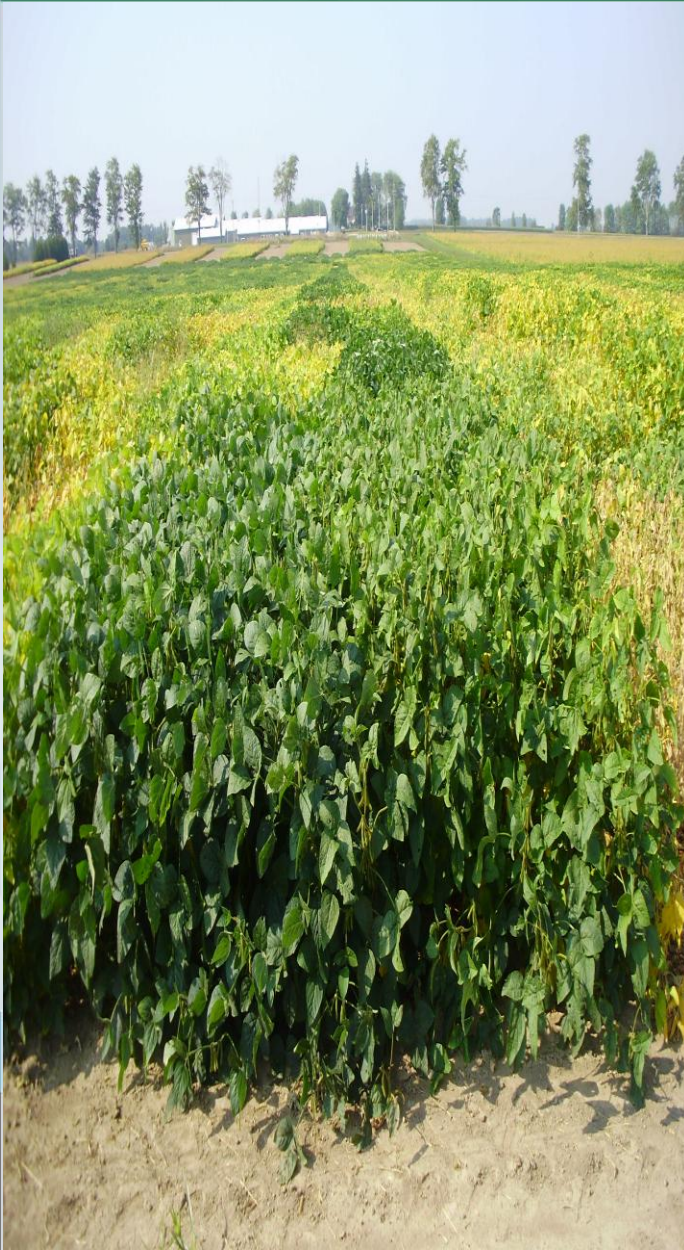
Ott et al. (2011) Physical distribution of recombination in soybean

- Telomeric 25% of chromosome contained average of 50.2% of the genes.
- Centromeric 25% contained average of 6.7% of the genes.
- Majority of recombination occurs in telomeric regions.
- SSR marker location and gene location highly correlated ($R^2=0.83$).

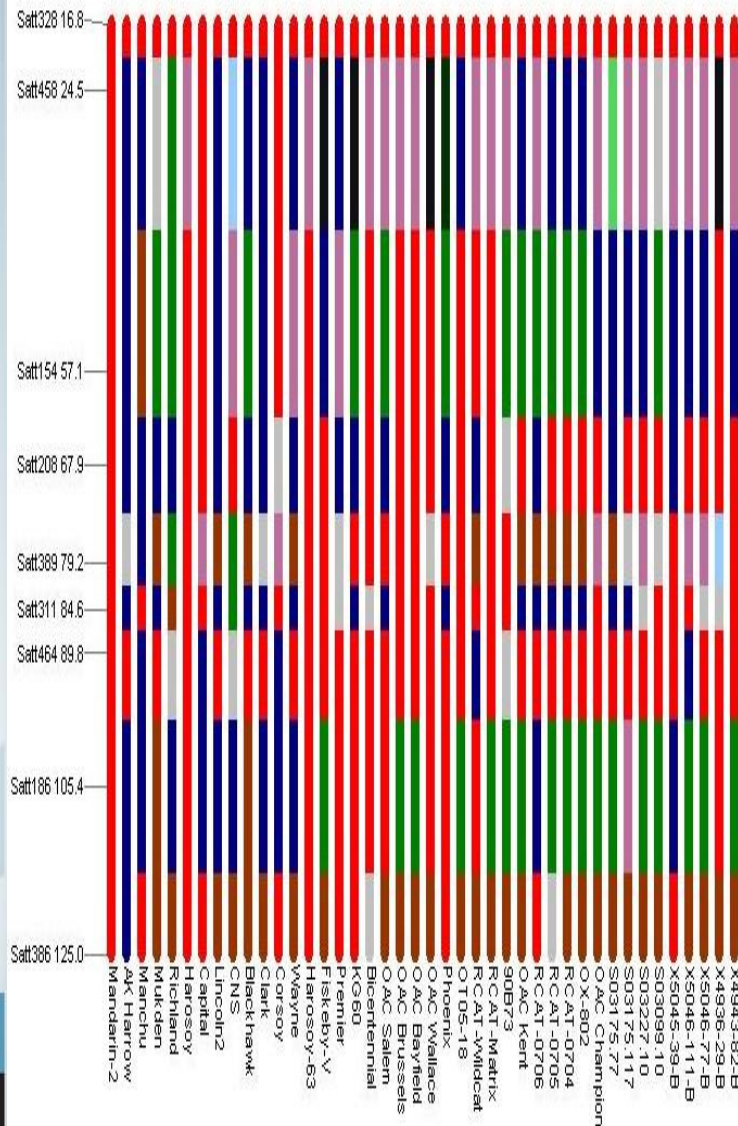
Linkage and Recombination: Soybean

Lorenzen et al. (1996)

- Study on the role of recombination during cultivar development.
- Linkage groups inherited without change between cultivars independently developed by different breeding programs.
- Suggest breeder selection is maintaining highly conserved linkage block transmission.



Chromosome Transmission



- Characterize allelic changes that have occurred through long-term phenotypic selection.
- 161 markers at density of 1 marker/10cM.
- Define and track specific linkage block transmission.
- Distinct maturity classes represented in lines.



- UNIVERSITY
-
- of GUELPH

Allele Inheritance

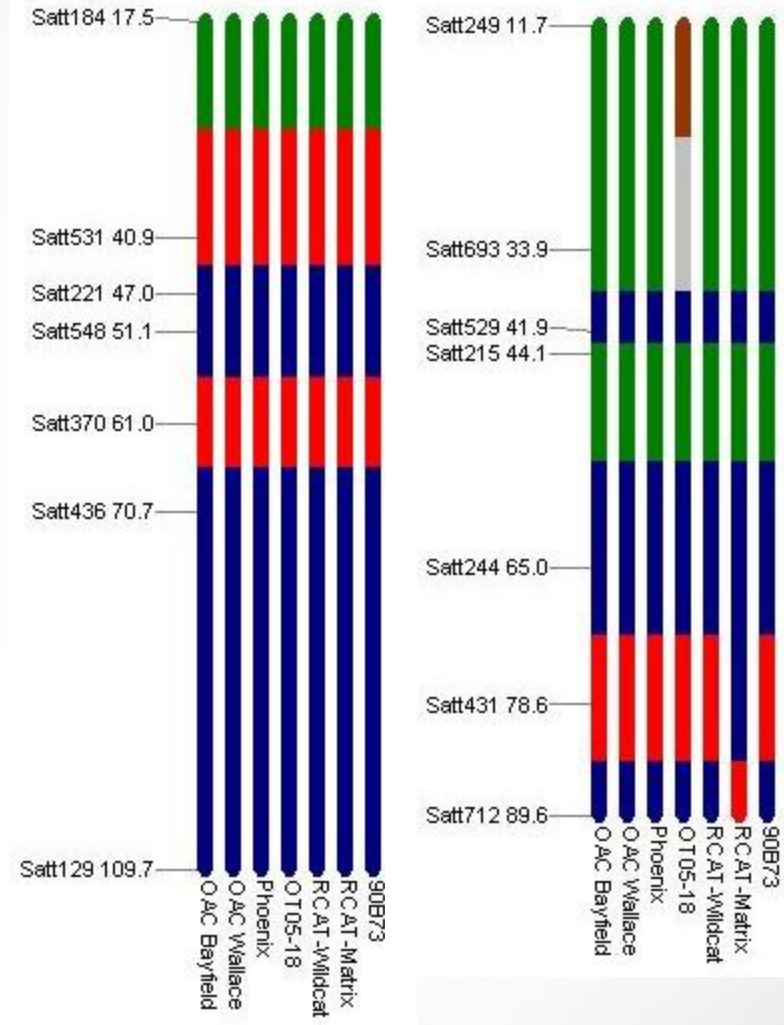
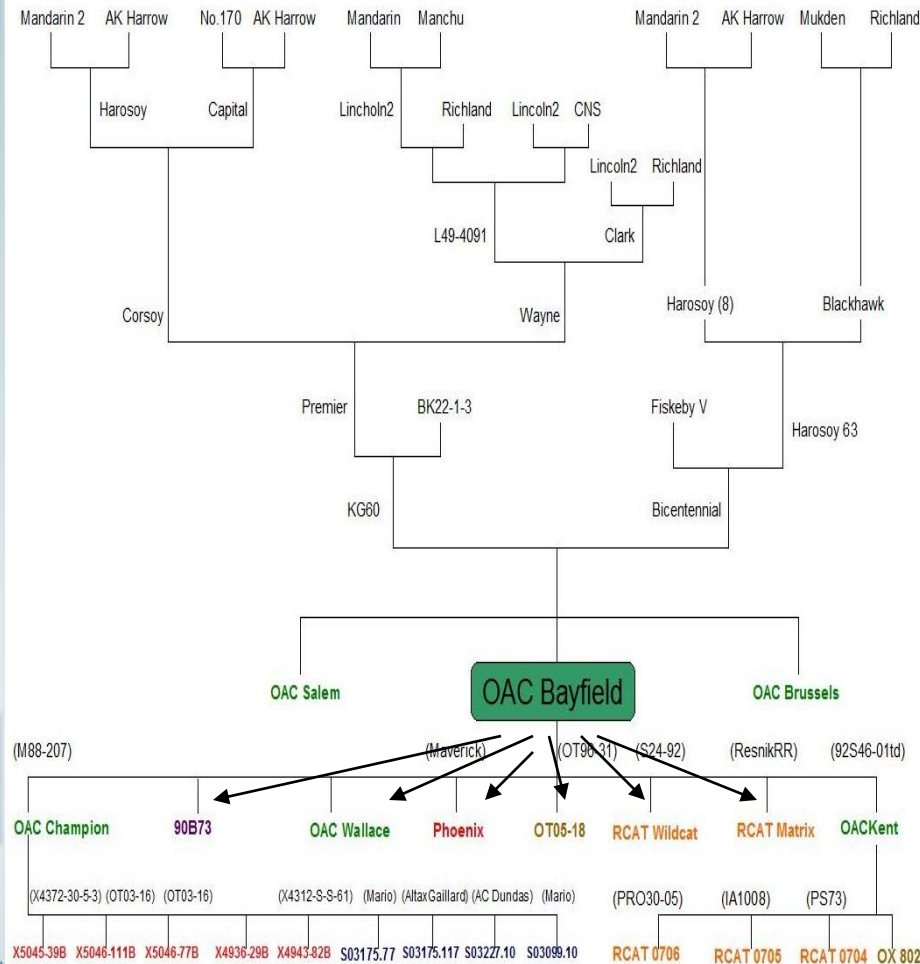
Cultivar	Number of Loci	Number loci Parent 1	Number loci Parent 2	χ^2	P Value
Harosoy**	87	Mandarin-2 (67)	AK Harrow (20)	0.19	NS
Blackhawk	78	Mukden (45)	Richland (33)	1.85	NS
Clark**	80	Lincoln 2 (64)	Richland (16)	1.07	NS
Premier	57	Corsoy (14)	Wayne (43)	14.75	<0.01*
Bicentennial	86	Harosoy 63 (29)	Fiskeby-V (57)	9.12	<0.01*
OAC Salem	63	KG 60 (28)	Bicentennial (35)	0.78	NS
OAC Brussels	63	KG 60 (15)	Bicentennial (48)	17.29	<0.01*
OAC Bayfield	66	KG 60 (30)	Bicentennial (36)	0.71	NS

NS: Non significant

* Significant at $\alpha = 0.01$

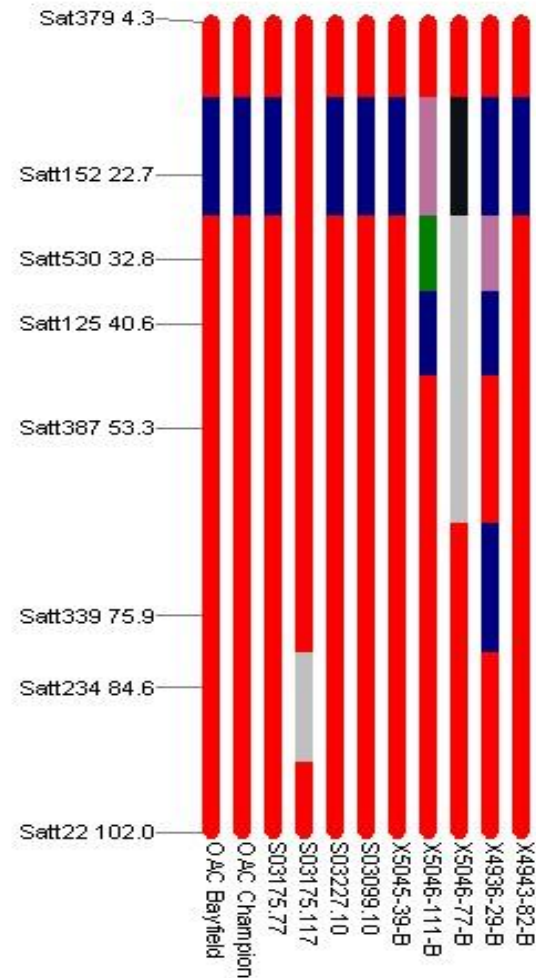
** One backcross (expected ratio is 75:25)

Conserved Allelic Structure



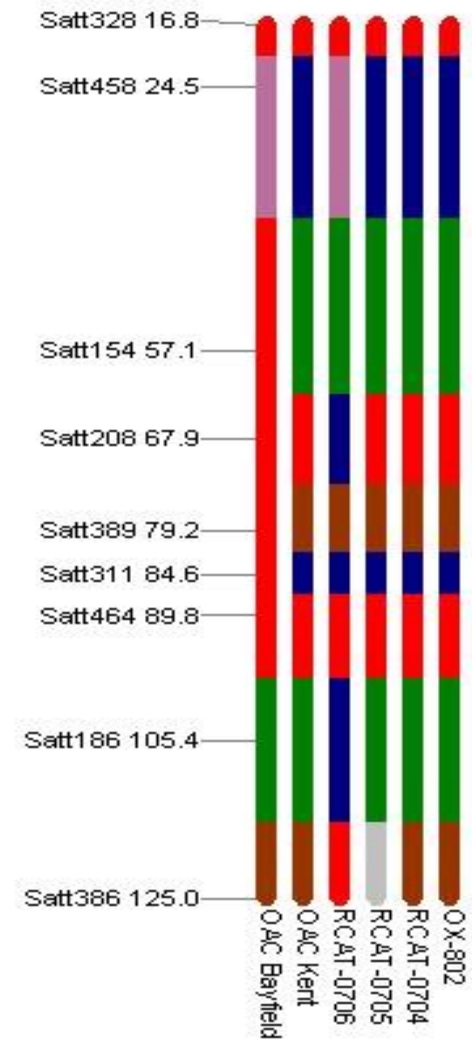
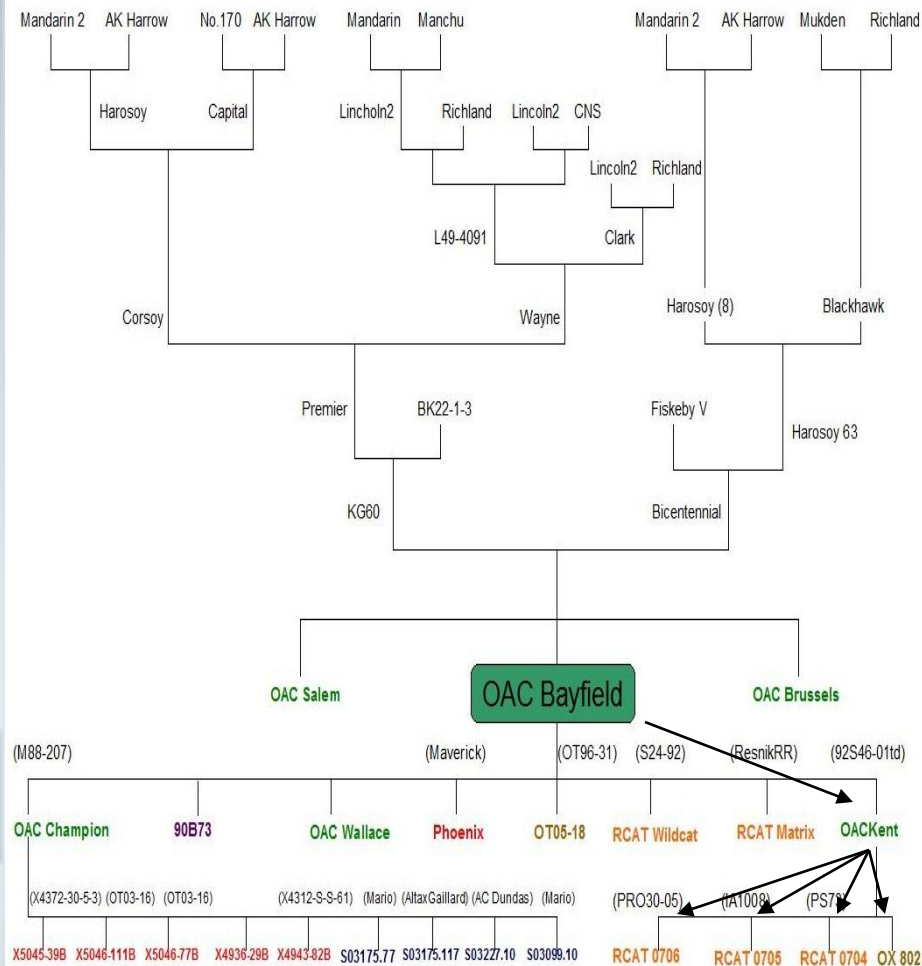
Chr 1

Chr 16



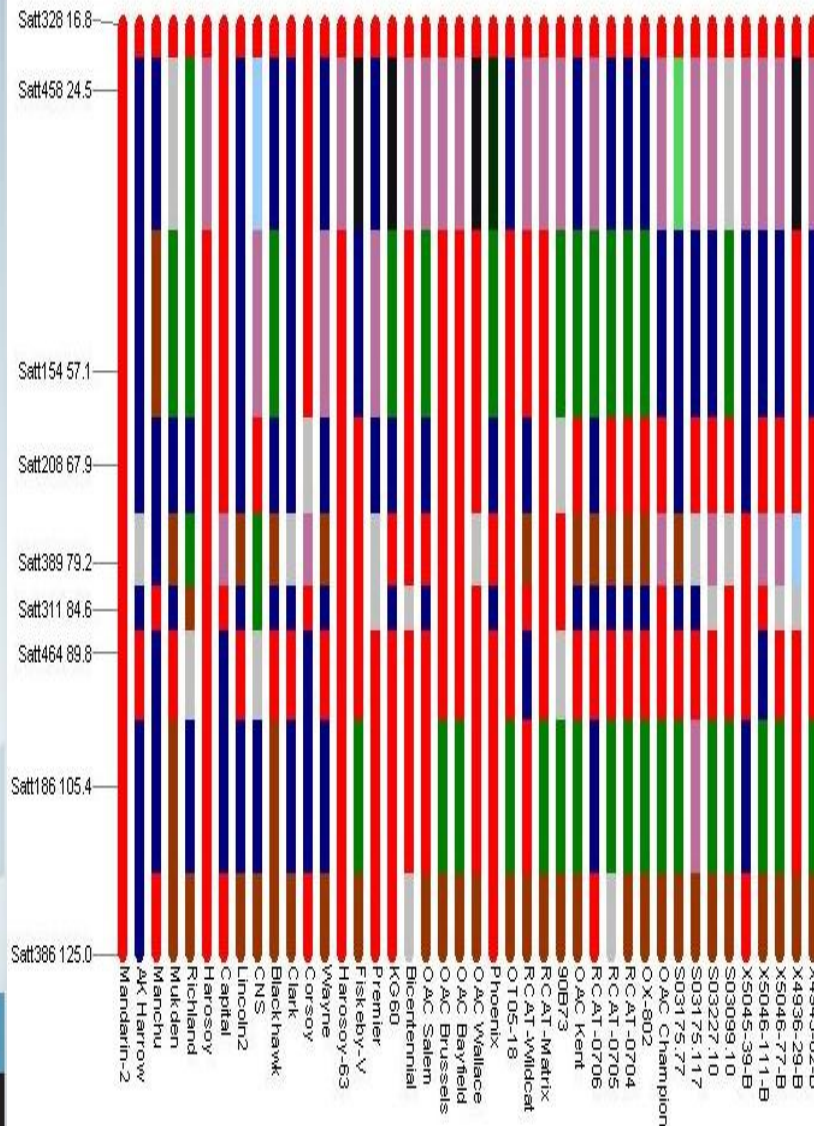
Chr 3

Lack of Allelic Structure



Chr 17

Breeding Implications



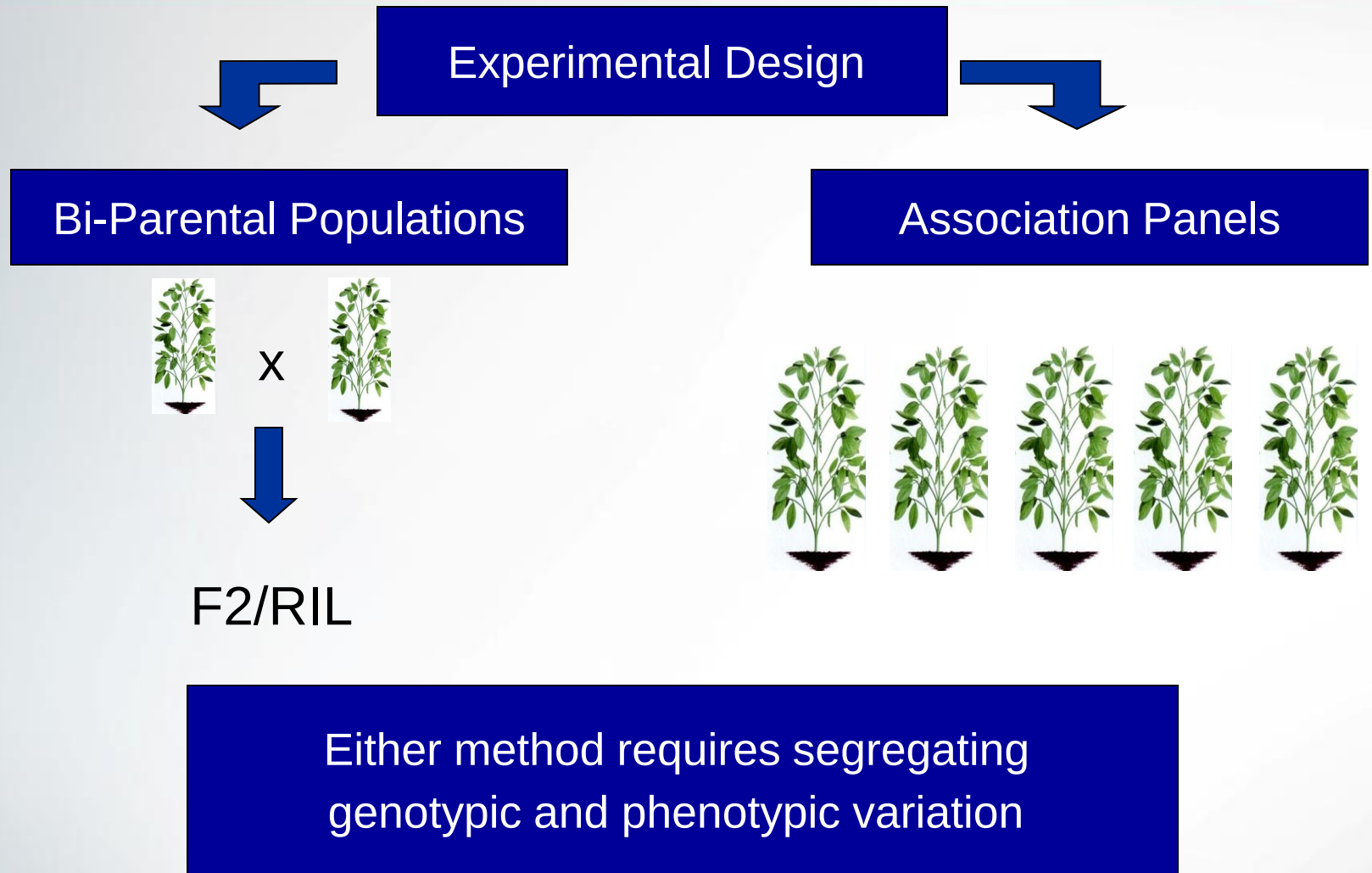
- Allelic structure observed across multiple breeding programs.
- Important where fixation is taking place.
- Breeding theory: Major QTL most likely fixed first.
- Breeding “paradox” created in terms of commercial cultivar development.



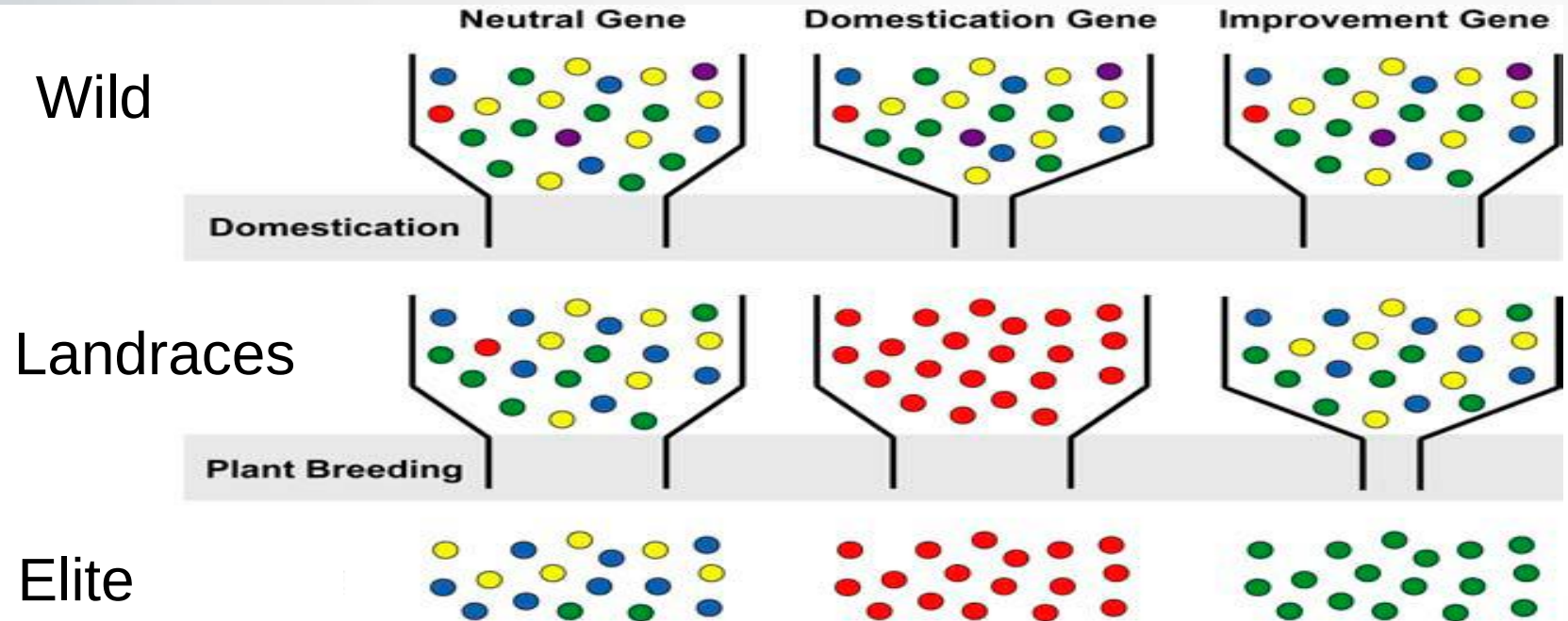
***In silico* QTL Mapping Using a Microsatellite Genome-Wide Scan for Genomic Selection Signatures Retained Through the Pedigree of OAC Bayfield**



Mapping Studies in Plant Breeding



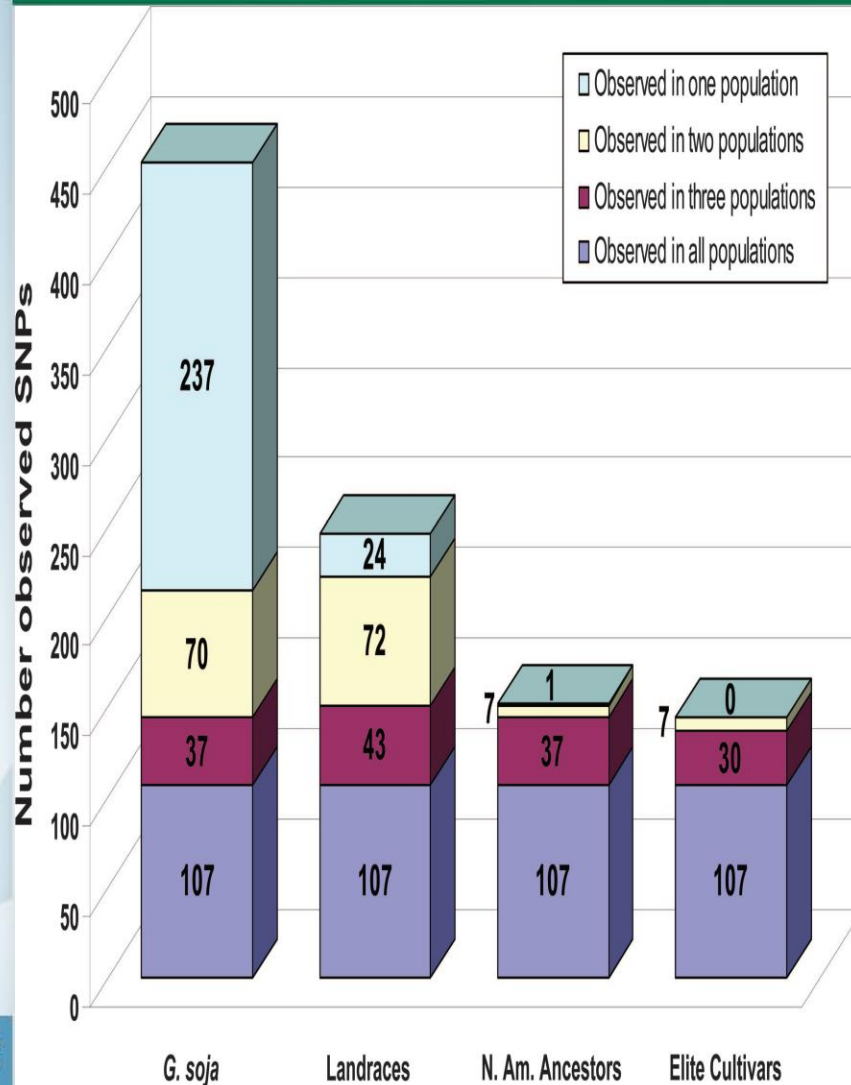
Effects of Selection



Could be missing an interesting class of genes due to fixation in elite germplasm.

Modified from Yamasaki et al. 2005

Genetic Bottlenecks

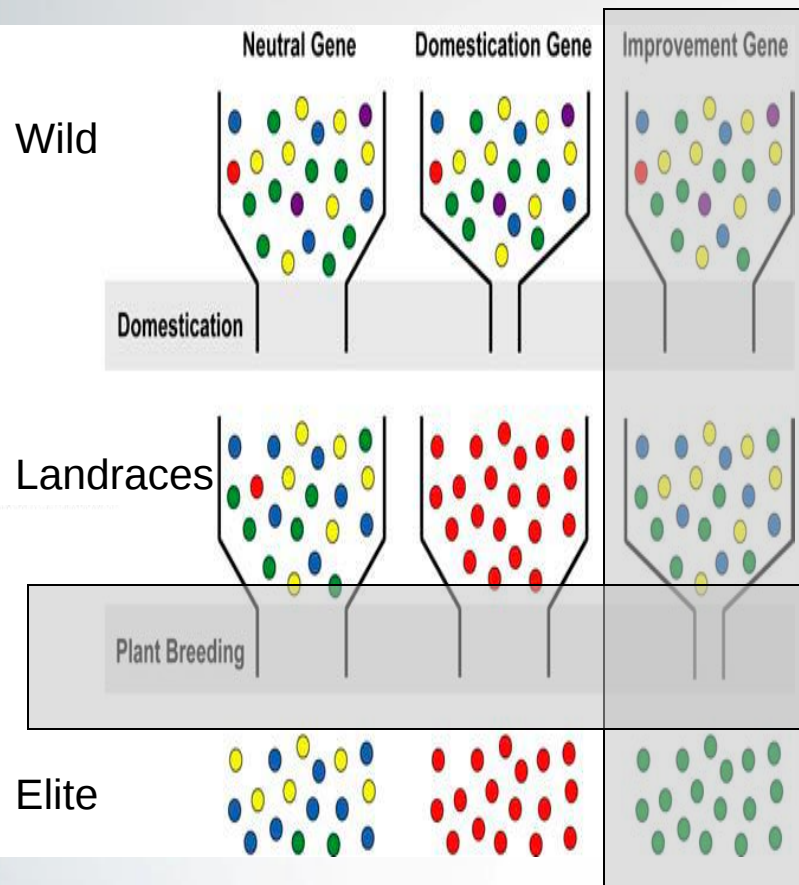


Hyten et al. (2006)

- Impacts of genetic bottlenecks on soybean genome diversity.
- Domestication from *G. Soja* to *G. Max*, 81% of rare alleles lost.
- 40 of 102 genes fixed in elite germplasm. Only 7 in *G. soja* population.

Hyten et al. 2006

Selection Signatures

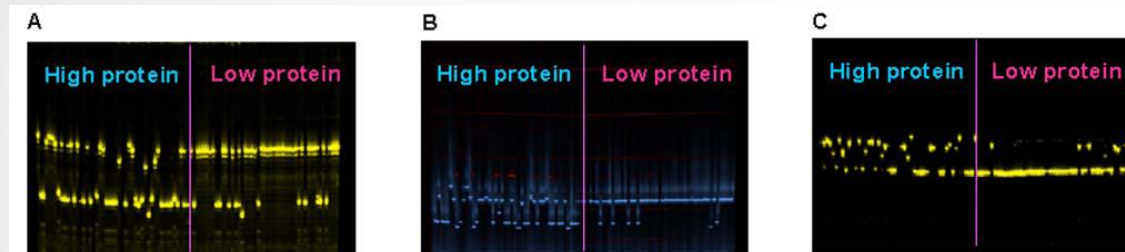


- Reduction in diversity beyond general effect of drift/bottleneck.
- Have been shown to be found in regions of agronomic importance.
- What genes/QTL are breeders selecting for during cultivar improvement.
- Using pedigree to identify these type of selective signatures.



Selection Signatures: Soybean

Uncovering signatures of selection in the soybean genome using SSR diversity near QTLs of agronomic importance
Jun et al. 2011. Genes & Genomics



Artificial selection for determinate growth habit in soybean
Tian et al. 2010. PNAS

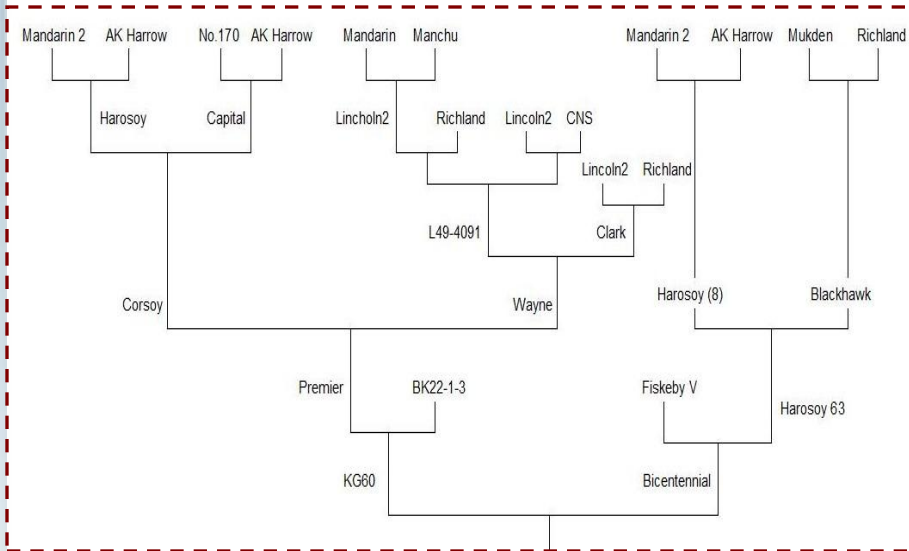
	10	20	30	40	50	60	70	80	90
Dt1	MAKMPLEPLIVGRVIGEVLDSTTTSTKMTVSYNKKQVYNGHELFPSTVNTKPKVEIEGGDMRSFFTLIMTDPDVPGPSDPYLRHLHWIV								
dt1-1	MAKMPLEPLIVGRVIGEVLDSTTTSTKMTVSYNKKQVYNGHELFPSTVNTKPKVEIEGGDMRSFFTLIMTDPDVPGPSDPYLRHLHWIV								
dt1-2a	MAKMPLEPLIVGRVIGEVLDSTTTSTKMTVSYNKKQVYNGHELFPSTVNTKPKVEIEGGDMRSFFTLIMTDPDVPGPSDPYLRHLHWIV								
dt1-2b	MAKMPLEPLIVGRVIGEVLDSTTTSTKMTVSYNKKQVYNGHELFPSTVNTKPKVEIEGGDMRSFFTLIMTDPDVPGPSDPYLRHLHWIV								
dt1-2c	MAKMPLEPLIVGRVIGEVLDSTTTSTKMTVSYNKKQVYNGHELFPSTVNTKPKVEIEGGDMRSFFTLIMTDPDVPGPSDPYLRHLHWIV								

	100	110	120	130	140	150	160	170
Dt1	TDIPGTTDATFGKELVSYEVKPNIGIHRFVFLFKQKRRCQVTPPTS RDHFNTRKFAAENDLGLPVAAVYFNAQRETAARRR							
dt1-1	TDIPGTTDATFGKELVSYEVKPNIGIHRFVFLFKQKRRCQVTPPTS RDHFNTRKFAAENDLGLPVAAVYFNAQRETAARRR							
dt1-2a	TDIPGTTDATFGKELVSYEVKPNIGIHRFVFLFKQKRRCQVTPPTS RDHFNTRKFAAENDLGLPVAAVYFNAQRETAARRR							
dt1-2b	TDIPGTTDATFGKELVSYEVKPNIGIHRFVFLFKQKRRCQVTPPTS RDHFNTRKFAAENDLGLPVAAVYFNAQRETAARRR							
dt1-2c	TDIPGTTDATFGKELVSYEVKPNIGIHRFVFLFKQKRRCQVTPPTS RDHFNTRKFAAENDLGLPVAAVYFNAQRETAARRR							

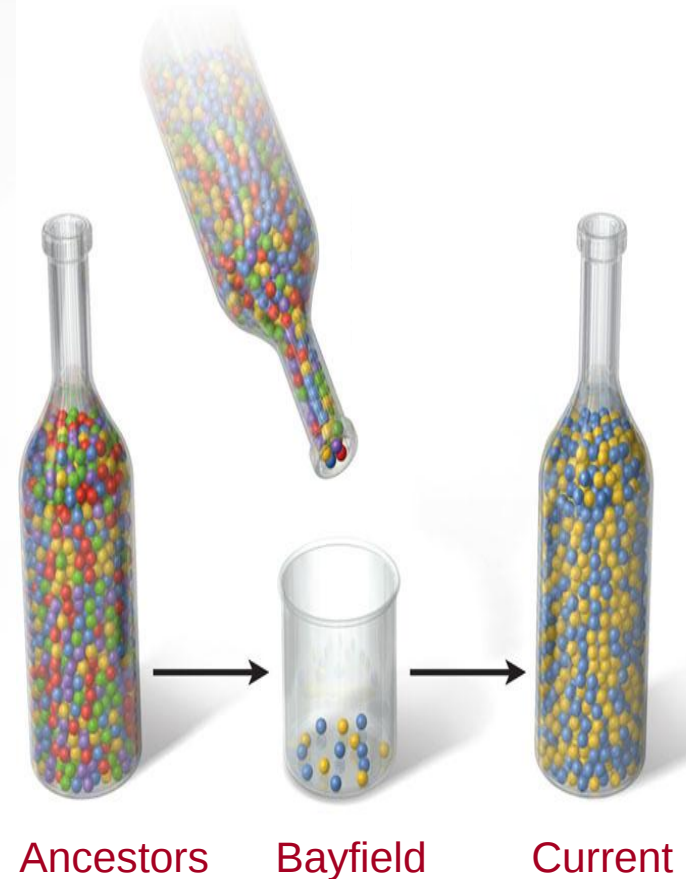
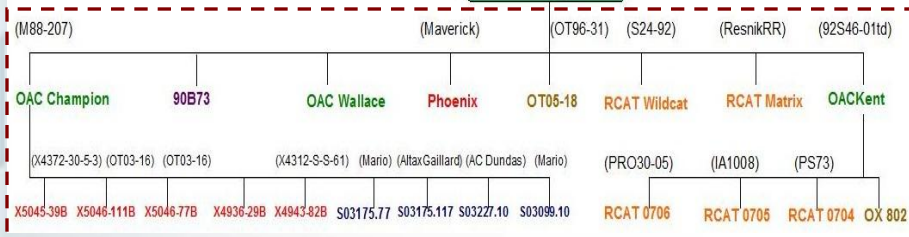


Pedigree-Based Groups

Ancestral



Current



Selection Statistics: LnRH

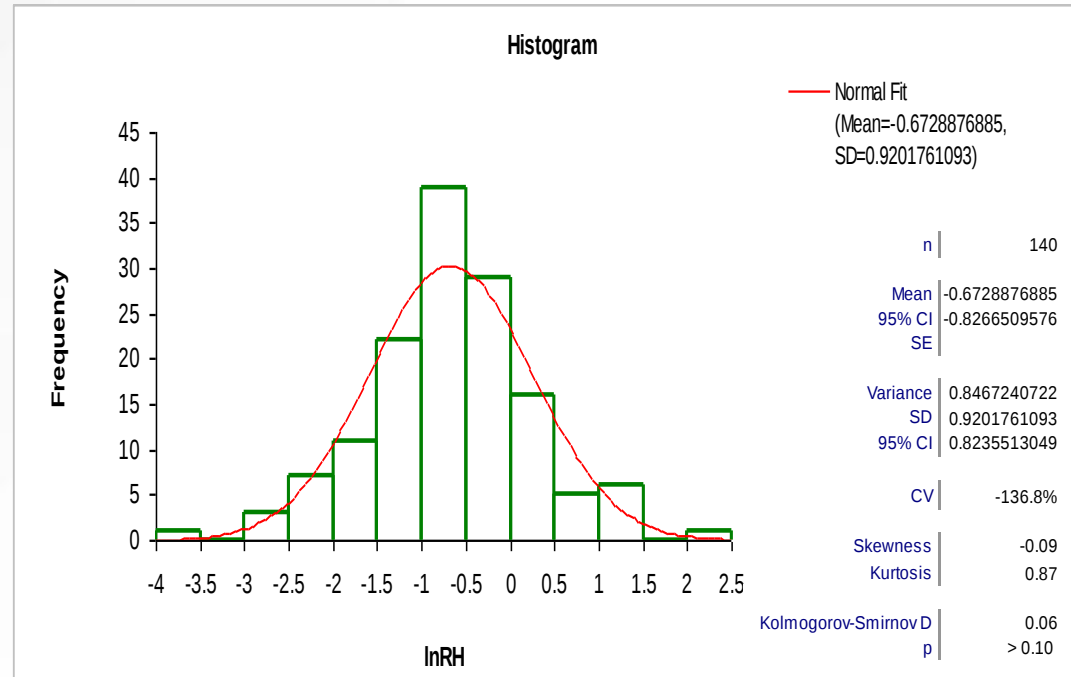
Current Group



$$\ln[RH] = \ln \left[\frac{\left(\left(\frac{1}{1-He_{Pop1}} \right)^2 - 1 \right)}{\left(\left(\frac{1}{1-He_{Pop2}} \right)^2 - 1 \right)} \right]$$



Ancestral Group

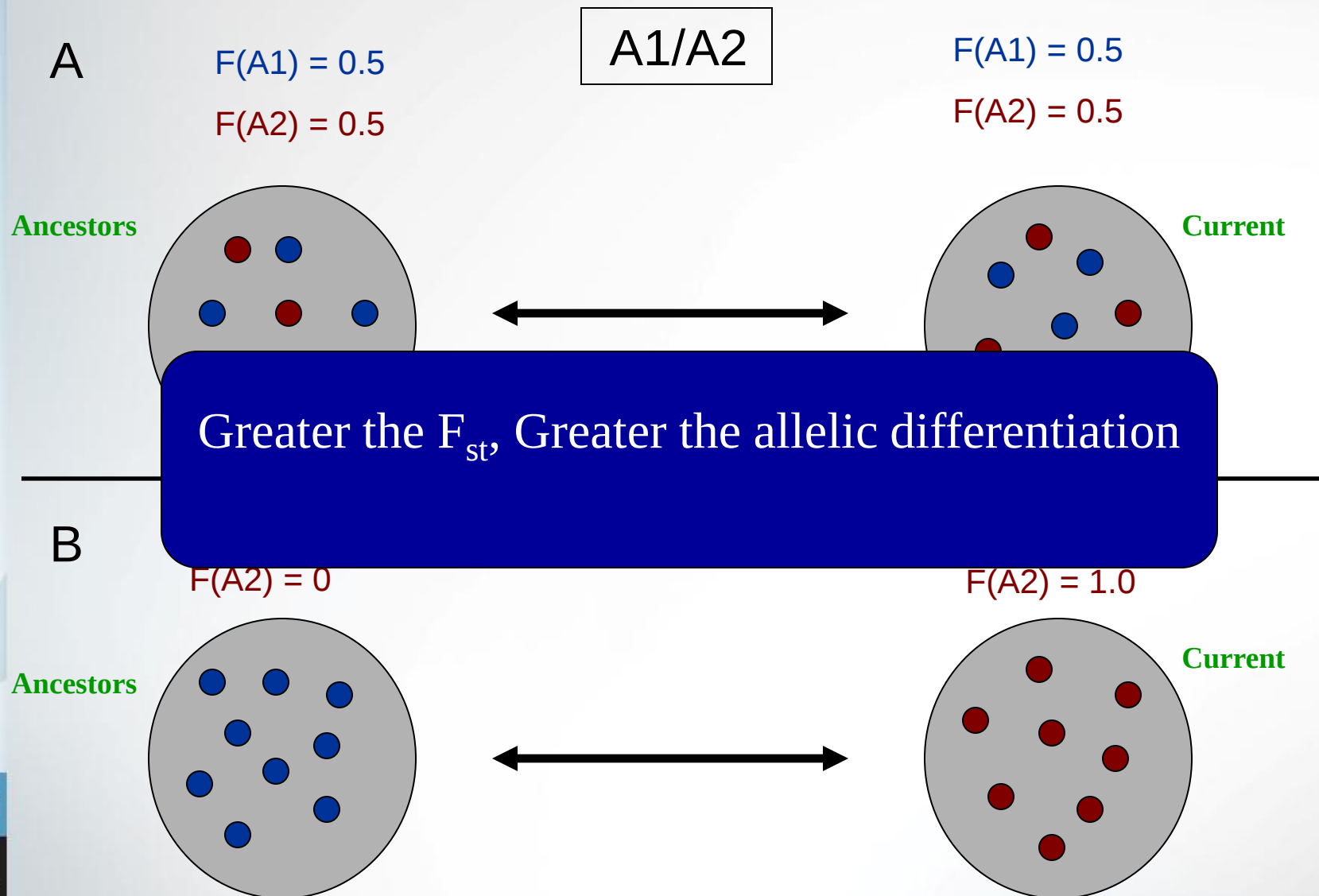


He = Expected Heterozygosity/Gene Diversity

Schlötterer, 2002

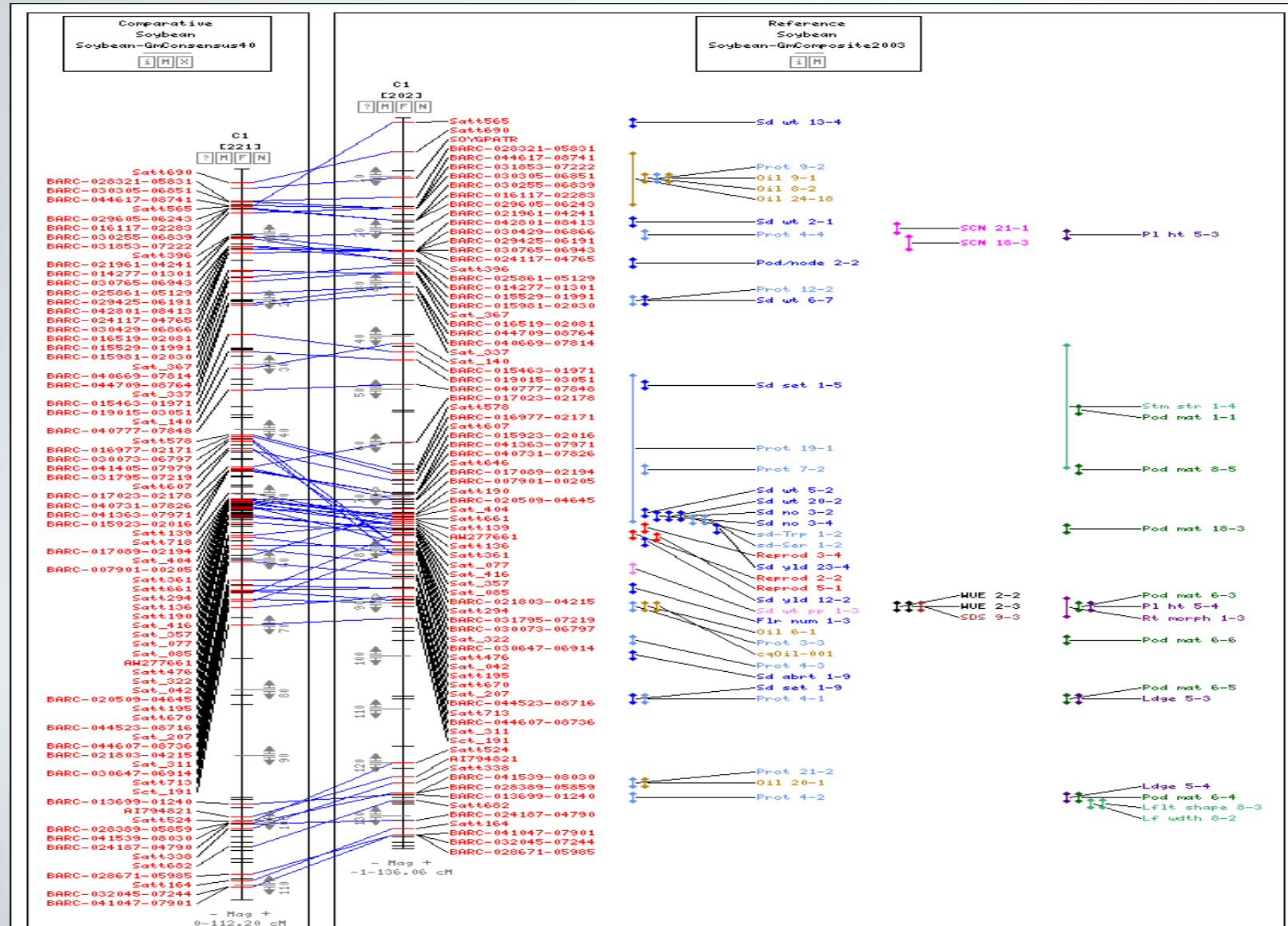


Selection Statistics: F_{st}





SoyBase: Using Genetic Resources



What's in the Genomic Region?

Chromosome	SSR Marker	Detected with LnRH/Fst/Both	Marker Map Position (cM)	Trait QTL
1	Satt 531	lnRH	40.86	Protein
2	Satt 274**	lnRH	116.34	Multi QTL*
3	Satt 387**	lnRH	53.25	Multi QTL*, <i>Rpg 4</i> gene
4	Satt 578**	lnRH	65.08	Multi QTL*
5	Satt 050**	lnRH	46.45	Leaf width
6	Satt 319	Fst	113.4	Multi QTL, <i>E1</i> gene, <i>T</i> gene
6	Satt 357	Both	151.91	Sudden Death Syndrome
9	Satt 242	Both	14.35	Protein/Oil
11	Satt 426	Fst	28.3	Protein/Oil
11	Satt 332	Fst	80.9	Seed weight/Oil
12	Satt 469**	Fst	58.9	Corn ear worm
15	Satt 369**	Fst	56.3	Leaf shape
15	Satt 230**	Fst	71.3	Plant height
16	Satt 249	Both	11.74	Multi QTL*
17	Satt 186**	Fst	105.4	Multi QTL*
17	Satt 386	Fst	125	Sclerotina
19	Satt 561**	Fst	71.4	Yeild
20	Sat 104	Both	65.6	N/A
20	Sat 419**	Both	98.1	Yeild

* Multi QTL refers to regions where three or more QTL map to the same position.

**Marker associated with Trait QTL according to SoyBase (www.soybase.org)

Trait QTL Breakdown

Trait classification for QTL identified via selective sweep analysis in the pedigree of OAC Bayfield.

Specific QTL designations are according to SoyBase (www.soybase.org) composite genetic maps.

Trait	Plant Architecture	Protein	Oil	Yield	Maturity	Disease	Other
SoyBase QTL Reference	PI ht 17-5	Prot 3-5	Oil 19-1	Sd yld 15-12	Pod mat 16-4	Sclero 3-16	NitR5 1-4
	Lf width 9-1	sd-Leu 1-2	Oil 1-2	Sd wt 10-3	Pod mat 8-5	SDS 2-5	Daidzein 2-7
	Lf width 8-1	sd-Tyr 1-2	Oil 4-5	Yld/Ht 4-1	Pod mat 19-6	SDS 1-1	Isoflv 1-4
	Lflr shape 9-6	Prot 7-2	Oil 24-14	Sd yld 23-6	Pod mat 19-2	CEW 1-2	Drought index 1-5
	Lflr shape 8-10	Prot 24-3		Sd yld 15-7	Pod mat 8-3	Sclero 6-4	
	PI ht 13-5	Prot 3-2		Yld/Ht 2-5		Sclero 3-7	
	PI ht 8-3	sd-Glu 1-4		Sd yld 8-1		Sclero 2-10	
		Prot 17-2		Sd yld 22-5			
				Flr num 1-11*			
				Pod num 1-10*			
QTL Total	7	8	4	10	5	7	4
*Considered as component traits for yield							

Trait QTL Breakdown

 **SoyBase and the Soybean Breeder's Toolbox**
Integrating Genetics and Molecular Biology for Soybean Researchers

[SoyBase](#) [Maps](#) [Genome](#) [Analysis Tools](#) [Resources](#) [SoySeq](#)

[Composite Genetic Maps](#) [Experiment-specific Maps](#) [Physical Maps](#) [QTL](#) [Download Data](#) [Contact Us](#)

Matching Traits	
Abnormal seedling	Peanut root-knot nematode resistance
Aluminum Tolerance	Phomopsis seed decay
Arabinose	Photoperiod insensitivity
Beginning pod	Phytophthora
Beginning	
P	
Linoleic	Stem terpen
Linolenic	Sucrose
Linolenic Acid	Sudden death syndrome resistance
Lodging	Tobacco budworm resistance
Maturity Date	Tobacco ringspot virus
Oil	Trigonelline concentration
Oil/Protein	Water use efficiency
Oleate	Yield
Oleic Acid	Yield/Height
Palmitate	Yield/Seed weight
Palmitic	

92 traits which have had
QTL studies conducted

QTL numbers from SoyBase

Found in SoyBase:

Trait	Plant Architecture	Protein	Oil	Yield	Maturity	Disease	Lodging
SoyBase QTL Total	328*	109	128	117	108	200	63**
*Plant height	140						
*Leaf width	58						
*Leaflet shape	63						
*Leaf length	67						
**None detected							

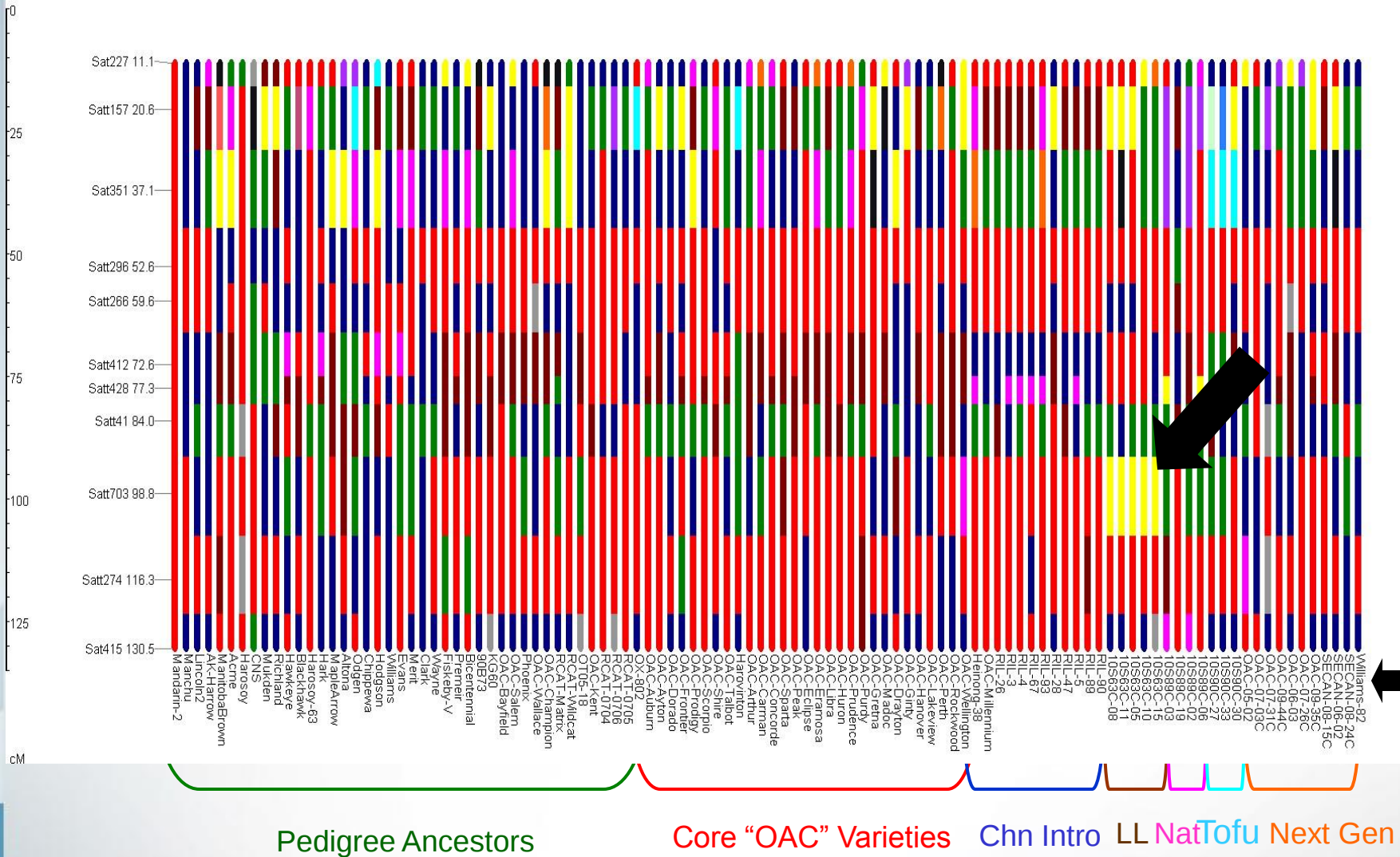
Found in Study:

Trait classification for QTL identified via selective sweep analysis in the pedigree of OAC Bayfield.

Specific QTL designations are according to SoyBase (www.soybase.org) composite genetic maps.

Trait	Plant Architecture	Protein	Oil	Yield	Maturity	Disease	Other
SoyBase QTL Reference	PI ht 17-5	Prot 3-5	Oil 19-1	Sd yld 15-12	Pod mat 16-4	Sclero 3-16	NitR5 1-4
	Lf width 9-1	sd-Leu 1-2	Oil 1-2	Sd wt 10-3	Pod mat 8-5	SDS 2-5	Daidzein 2-7
	Lf width 8-1	sd-Tyr 1-2	Oil 4-5	Yld/Ht 4-1	Pod mat 19-6	SDS 1-1	Isoflv 1-4
	Lft shape 9-6	Prot 7-2	Oil 24-14	Sd yld 23-6	Pod mat 19-2	CEW 1-2	Drought index 1-5
	Lft shape 8-10	Prot 24-3		Sd yld 15-7	Pod mat 8-3	Sclero 6-4	
	PI ht 13-5	Prot 3-2		Yld/Ht 2-5		Sclero 3-7	
	PI ht 8-3	sd-Glu 1-4		Sd yld 8-1		Sclero 2-10	
		Prot 17-2		Sd yld 22-5			
				Flr num 1-11*			
				Pod num 1-10*			
QTL Total	7	8	4	10	5	7	4

Next Phase Pedigree Genotyping



Chr 2

High Density Genome Scan: GBS



An Improved Genotype by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping

Sonah et al. 2013. PLOS ONE



Haplotype Structure Comparison

U of G Germplasm



USDA Genebank



Conclusions

- DNA fingerprinting pedigrees to understand effects on long-term breeder selection.
- Allele composition of specific chromosomes conserved *across* breeding programs depending on target environment.
- Where allele fixation occurs is of importance for continued genetic gain. Candidate regions for germplasm introgression
- Identification of unique genetic contribution from key ancestors.
- Markers identified by selective sweep analysis map to known QTL and in many cases are associated with QTL.



Acknowledgements

Project Funding



Breeding Programs



UNIVERSITY
of GUELPH
RIDGETOWN CAMPUS



Agriculture et
Agroalimentaire Canada

Agriculture and
Agri-Food Canada

Canada

