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?
Retrospective Soybean Development

Breeding activity

- 36-fold increase

Yield improvement

- 1.4-fold increase

The Success of OAC Bayfield

- 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

(M88-207)
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
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.
Impact of Fiskeby-V

- Swedish variety developed in 1950’s by Sven A. Holmberg.
- MG 000 (very early maturity).
- Novel alleles/linkage blocks found in many of the MG 00-1 lines, absent in MG II lines.
- Pod maturity and flowering QTL found in linkage blocks in SoyBase.
## Allele Inheritance

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

NS: Non significant  
* Significant at $\alpha = 0.01$  
** One backcross (expected ratio is 75:25)
Conserved Allelic Structure
Conserved Allelic Structure
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

Experimental Design

Bi-Parental Populations

Association Panels

F2/RIL

Either method requires segregating genotypic and phenotypic variation
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.
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
Pedigree-Based Groups

Ancestral

Current

OAC Bayfield

OAC Salem

OAC Wallace

Phoenix

OT05-18

RCAT Wildcat

RCAT Matrix

OACKent

Ancient

Bayfield

Current
Selection Statistics: LnRH

\[ \ln[(RH)] = \ln \left[ \left( \frac{1}{1 - He_{\text{Top1}}} - 1 \right) \right] \left( \frac{1}{1 - He_{\text{Top2}}} - 1 \right) \]

Current Group

Ancestral Group

\[ He = \text{Expected Heterozygosity/Gene Diversity} \]

Schlötterer, 2002
Selection Statistics: Fst

A

\[ F(A1) = 0.5 \]
\[ F(A2) = 0.5 \]

A1/A2

\[ F(A1) = 0.5 \]
\[ F(A2) = 0.5 \]

Greater the F_{st}, Greater the allelic differentiation

B

\[ F(A2) = 0 \]

\[ F(A2) = 1.0 \]
Genome Scan for Selection
## What’s in the Genomic Region?

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

* 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.

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

| QTL Total | 7 | 8 | 4 | 10 | 5 | 7 | 4 |

*Considered as component traits for yield
92 traits which have had QTL studies conducted
QTL numbers from SoyBase

**Found in SoyBase:**

<table>
<thead>
<tr>
<th>Trait</th>
<th>Plant Architecture</th>
<th>Protein</th>
<th>Oil</th>
<th>Yield</th>
<th>Maturity</th>
<th>Disease</th>
<th>Lodging</th>
</tr>
</thead>
<tbody>
<tr>
<td>SoyBase QTL Total</td>
<td></td>
<td>328*</td>
<td>109</td>
<td>128</td>
<td>117</td>
<td>108</td>
<td>200</td>
</tr>
</tbody>
</table>

*Plant hieght* 140  
*Leaf width* 58  
*Leaflet shape* 63  
*Leaf length* 67

**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.

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

QTL Total 7 8 4 10 5 7 4
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.