Update on SNP Genotyping of the USDA Soybean Germplasm Collection

United Soybean Board Project #8265:

Whole Genome Analysis of the USDA Soybean Germplasm Collection and Applications for New Gene Discovery (50,000 SNPs)

David Hyten, Qijian Song, Gaofeng Jia, Charles Quigley, Edward Fickus, Randall Nelson, Vincent Pantalone, James Specht, and Perry Cregan

Soybean Genomics and Improvement Laboratory, USDA, ARS, Beltsville. MD
Project Objectives:

- Development of a 50,000 SNP Illumina iSelect Genechip (SoySNP50K)
- Genotyping of 1000+ RILs of Williams 82 x G. soja PI479752 and 900+ RILs of Essex x Williams 82 with the SoySNP50K Genechip
- Development of a 50,000 SNP consensus genetic map
- Analysis of the entire USDA Soybean Germplasm Collection with the SoySNP50K Genechip
- Haplotype Map development and application based upon the SoySNP50K analysis of 96 Elite cultivars, 96 Asian Landraces and 101 G. soja genotypes
- Haplotype Map development and application based upon the SNP analysis of the entire USDA Soybean Germplasm Collection
- Make all data available to the user community via SoyBase
Euchromatic and Heterochromatic Regions of the Soybean Genome and Development of the Illumina iSelect SoySNP50K Genechip


- Euchromatic regions makeup 459 Mbp of the genome sequence
- Heterochromatic regions makeup 491 Mbp with very low recombination, thus much larger haplotype block structure
SNP Selection for the SoySNP50K Genechip

60,800 Selected SNPs

Euchromatic regions:
50,701 SNPs, 110.5 per Mbp
Avg. distance between SNPs: 9.05 Kbp

Heterochromatic regions:
10,000 SNPs, 20.5 Mbp
Avg. distance between SNPs: 49 Kbp

99 SNPs selected in unanchored scaffolds in the Glyma1.01 genome build
The Illumina iSelect SoySNP50K Genechip

Of 51,041 SNPs ultimately included in the SoySNP50K beadpool, 47,446 produced successful allele calls among the 96 Elite cultivars, 96 Asian Landraces and 101 *G. soja* genotypes.

<table>
<thead>
<tr>
<th>Population</th>
<th>MAF&gt;0.05</th>
<th>MAF&gt;0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of SNPs</td>
<td>Percent of Total</td>
</tr>
<tr>
<td>Elite</td>
<td>33,878</td>
<td>72.1%</td>
</tr>
<tr>
<td>Landrace</td>
<td>38,932</td>
<td>82.8%</td>
</tr>
<tr>
<td>Wild</td>
<td>36,642</td>
<td>80.0%</td>
</tr>
<tr>
<td>All genotypes</td>
<td>44,029</td>
<td>92.8%</td>
</tr>
</tbody>
</table>
Two Ultra-high Resolution Soybean Genetic Maps
Created with the SoySNP50K Genechip

Williams 82 x PI479752: 21,481 loci x 1083 RILs
Essex x Williams 82: 13,157 loci x 922 RILs

5,963 SNP loci are common to the two mapping populations

Provide the basis for refinements of the Soybean Whole Genome Sequence Glyma1.01
Summary of Refinements of the Soybean Whole Genome Sequence Glyma1.01 based upon the W82xPI and ExW82 Mapping Populations

- 14 regions were identified where the current Soybean Whole Genome Sequence disagrees with both the W82xPI and ExW82 maps
- 7 unanchored sequence scaffolds were positioned on the Soybean Whole Genome Sequence
- HOWEVER, we still have some disagreements in the marker orders based upon the genetic mapping software that is used (MSTMap vs. JoinMap vs. R/qtl)
A total of 19,798 accessions including 1116 *G. soja* accessions

DNA was extracted from all accessions and analyzed with the SoySNP50K Genechip

Allele calling was completed for all genotypes using the Illumina GenomeStudio software

- Allele calling could not be done on all 19,798 genotypes at one time

- Initially sets of 1824 genotypes were analyzed one at a time
An example of a SNP with 19,768 accessions
A Set of 1824 genotypes
A Second Set of 1824 genotypes
A Set of 60 plates (5760 genotypes)
Results from the Initial Analysis of the USDA Soybean Germplasm Collection

- 2035 genotypes need to be re-analyzed because of chip failures or DNA quality issues
- 641 - Genotypes with greater than 10% heterozygous allele calls that will be re-analyzed
- 194 - New accessions that have been received from the Soybean Germplasm Collection that need to be analyzed
- 2870 genotypes to be re-analyzed with the SoySNP50K Genechip
Analyses of the 96 Elite cultivars, 96 Asian Landraces and 101 G. soja genotypes

- **$F_{st}$ analysis of a few interesting regions** – Asian Landraces vs. Elite cultivars and G. max vs. G. soja

  **$F_{st}$** (fixation index) is a measure of population differentiation. **$F_{st}$** values range from 0 to 1 with zero representing no allele differentiation and 1 representing complete allele differentiation.

- **Haplotype analysis of the Elite cultivars, Asian landraces and G. soja lines**
$F_{st}$ $G. \text{max} \ (\text{Landraces} + \text{Elite lines}) \ versus \ G. \ soja$

QTL:
- Seed oil and protein concentration
- Seed size
Many QTL have been reported in these two regions but all have been based upon the analysis of *G. max* x *G. max* populations.
Haplotype Map Development and Application Based upon the SoySNP50K Analysis of 96 Elite cultivars, 96 Asian Landraces and 101 *G. soja* genotypes

<table>
<thead>
<tr>
<th>Population</th>
<th>Number of blocks</th>
<th>Average block size (kb)</th>
<th>Mbp in LD Blocks and % of region in LD blocks</th>
<th>Number of blocks</th>
<th>Average block size (kb)</th>
<th>Mbp in LD Blocks and % of region in LD blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild</td>
<td>3,257</td>
<td>13.1</td>
<td>42.7 (9.3%)</td>
<td>303</td>
<td>340.5</td>
<td>103.2 (21.0%)</td>
</tr>
<tr>
<td>Landrace</td>
<td>4,218</td>
<td>55</td>
<td>232 (50.5%)</td>
<td>320</td>
<td>1200</td>
<td>384 (78.2%)</td>
</tr>
<tr>
<td>Elite</td>
<td>2,013</td>
<td>149</td>
<td>300 (65.4%)</td>
<td>180</td>
<td>2300</td>
<td>211.3 (84.3%)</td>
</tr>
</tbody>
</table>

**Proportion of blocks of various sizes in euchromatic regions of Wild, Landrace and Elite genotypes**

**Proportion of blocks of various sizes in heterochromatic regions Wild, Landrace and Elite genotypes**
Haplotype Analysis of 805 lines with Soybean Rust Resistance* for 5 Known \textit{Rpp} Genes, \textit{Rpp1} – \textit{Rpp5}

<table>
<thead>
<tr>
<th>\textit{Rpp} locus</th>
<th>Number of lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Rpp1}</td>
<td>120</td>
</tr>
<tr>
<td>\textit{Rpp2}</td>
<td>315</td>
</tr>
<tr>
<td>\textit{Rpp3}</td>
<td>59</td>
</tr>
<tr>
<td>\textit{Rpp4}</td>
<td>99</td>
</tr>
<tr>
<td>\textit{Rpp5}</td>
<td>548</td>
</tr>
<tr>
<td>None</td>
<td>151</td>
</tr>
</tbody>
</table>

Lines with various numbers of \textit{Rpp} resistance genes

<table>
<thead>
<tr>
<th>No. of \textit{Rpp} genes per line</th>
<th>Number of lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>304</td>
</tr>
<tr>
<td>2</td>
<td>249</td>
</tr>
<tr>
<td>3</td>
<td>98</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

* From: Miles, Frederick and Hartman. 2006. Plant Health Progress.
Update on SNP Genotyping of the USDA Soybean Germplasm Collection (USB Project #8265)

Current Status: “Allele calling” for all of the 19,000+ soybean accessions analyzed with the 50,000 SNP SoySNP50K Genechip has been completed.

DNA Isolation from approximately 2,900 accessions that need to be rerun is completed. Most genotypes have been re-analyzed with the SoySNP50K.

Results from the analysis of Wild soybean, Soybean Landraces and Elite Cultivars have been obtained.

Two ultra-high resolution soybean genetic maps have been created.

Haplotype Map development and application based upon the SNP analysis of the entire USDA Soybean Germplasm Collection is beginning.

All data will be available to the user community via SoyBase.
Many Thanks!!!!

- Dave Hyten, Qijian Song, Gaofeng Jia, Chuck Quigley, Ed Fickus, Karen Williams, and Sophie Zebell, USDA-ARS, Beltsville, MD
- Randall Nelson, USDA-ARS, Urbana, IL
- Vince Pantalone, Univ. of Tennessee
- James Specht, Univ. of Nebraska

Funding Support

United Soybean Board, USDA-ARS