

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

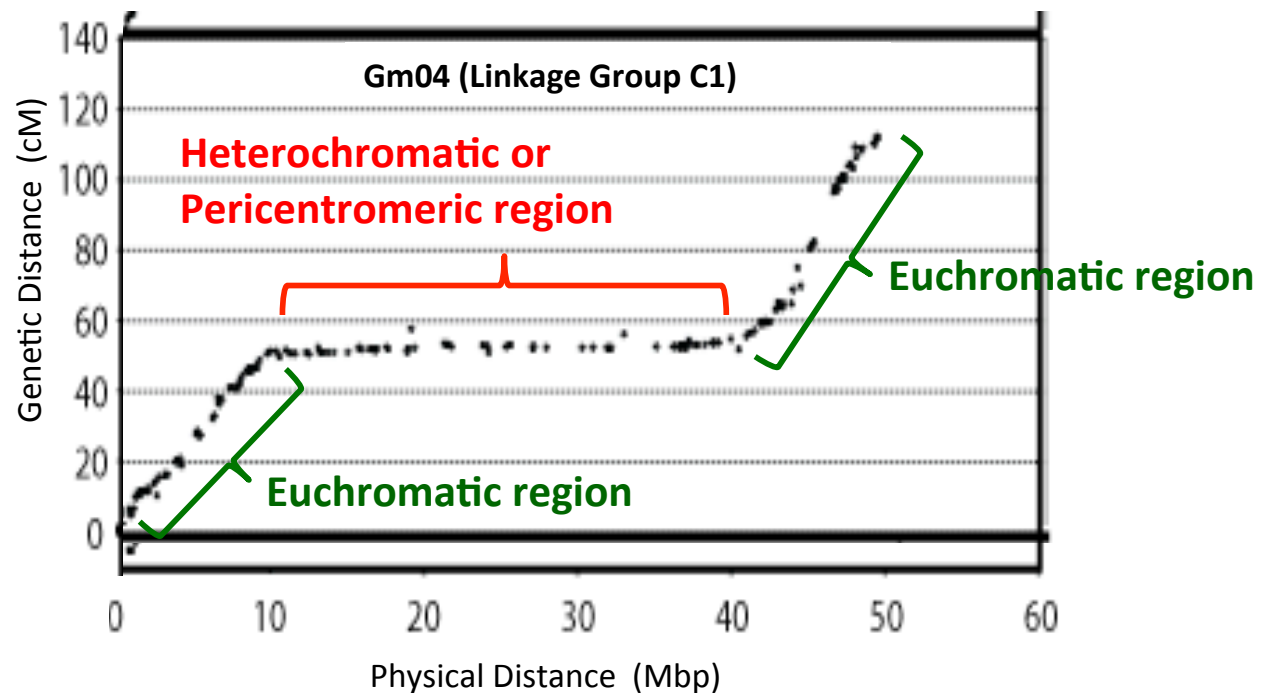
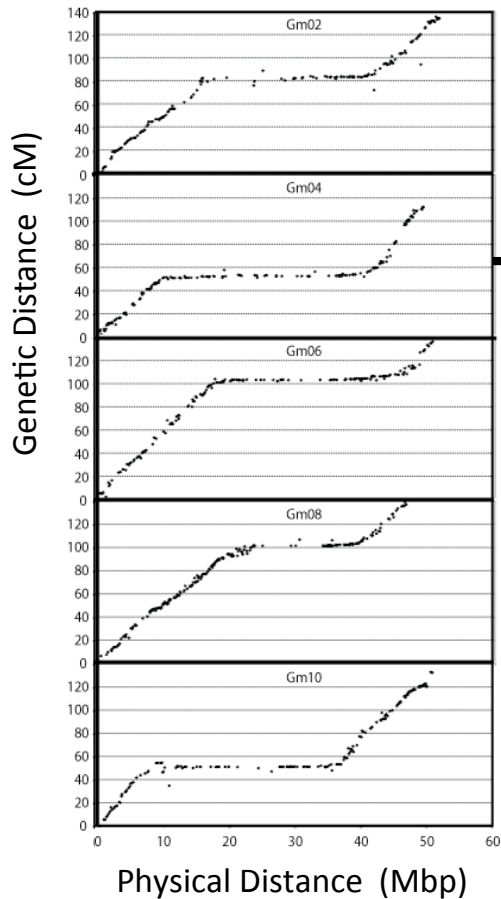
Soybean Genomics and Improvement Laboratory, USDA, ARS, Beltsville. MD



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

From Schmutz et al. 2010.
Nature, 463:178-183

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

SNP Minor Allele Frequencies (MAF) in 96 Diverse Elite Cultivars, 96 Diverse Asian Landraces and 101 Wild Soybean Accessions				
	MAF>0.05		MAF>0.10	
Population	No. of SNPs	Percent of Total	No. of SNPs	Percent of Total
Elite	33,878	72.1%	30,520	65.0%
Landrace	38,932	82.8%	34,500	73.4%
Wild	36,642	80.0%	31,463	68.7%
All genotypes	44,029	92.8%	40,747	85.9%

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

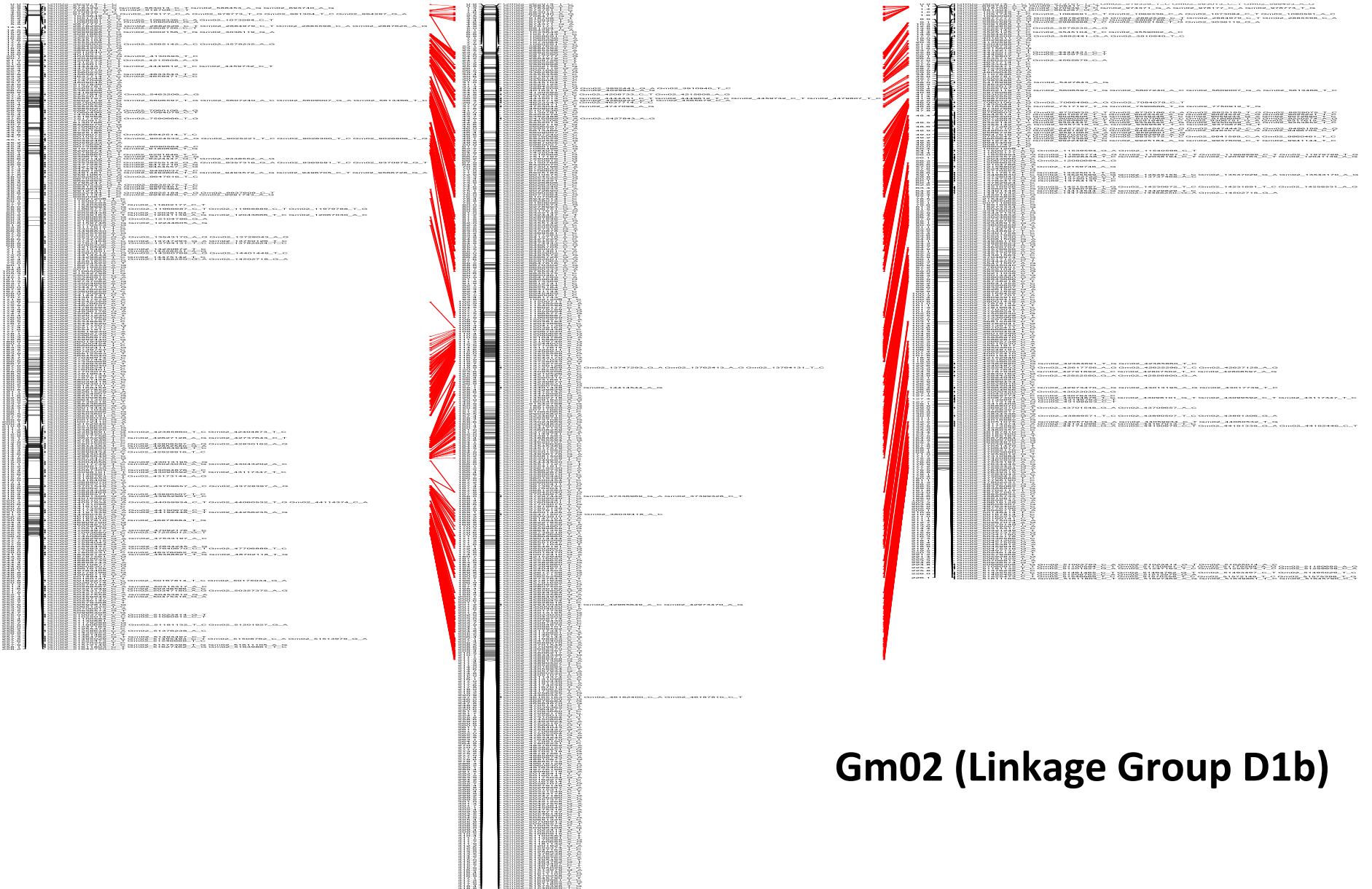


**Provide the basis for refinements of the
Soybean Whole Genome Sequence Glyma1.01**

Physical Position (Glyma1.01)

ExW82 Linkage map

W82xP Linkage map



Gm02 (Linkage Group D1b)

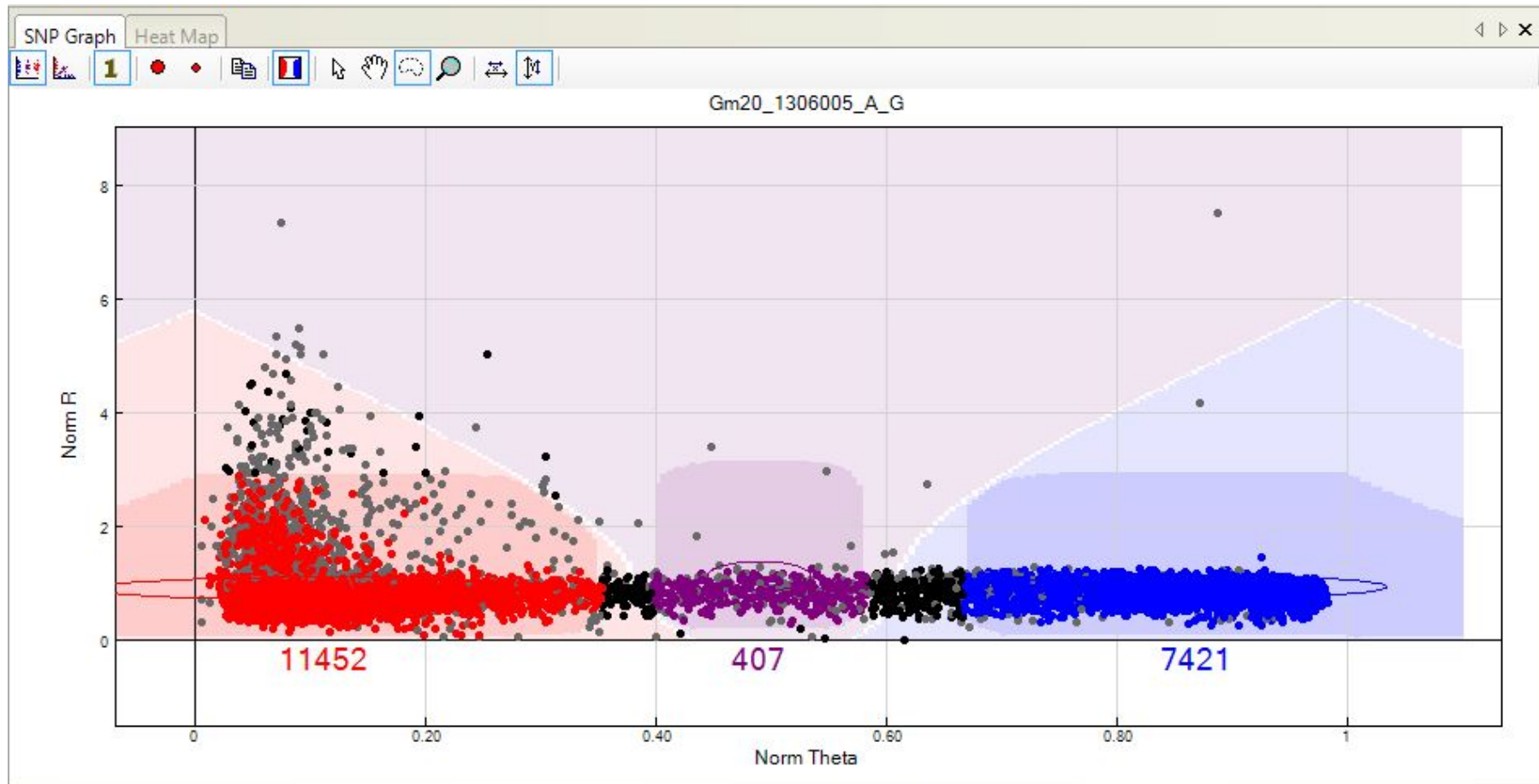
Summary of Refinements of the Soybean Whole Genome Sequence Glyma1.01 based upon the W82xPI and ExW82 Mapping Populations

- ◆ 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)

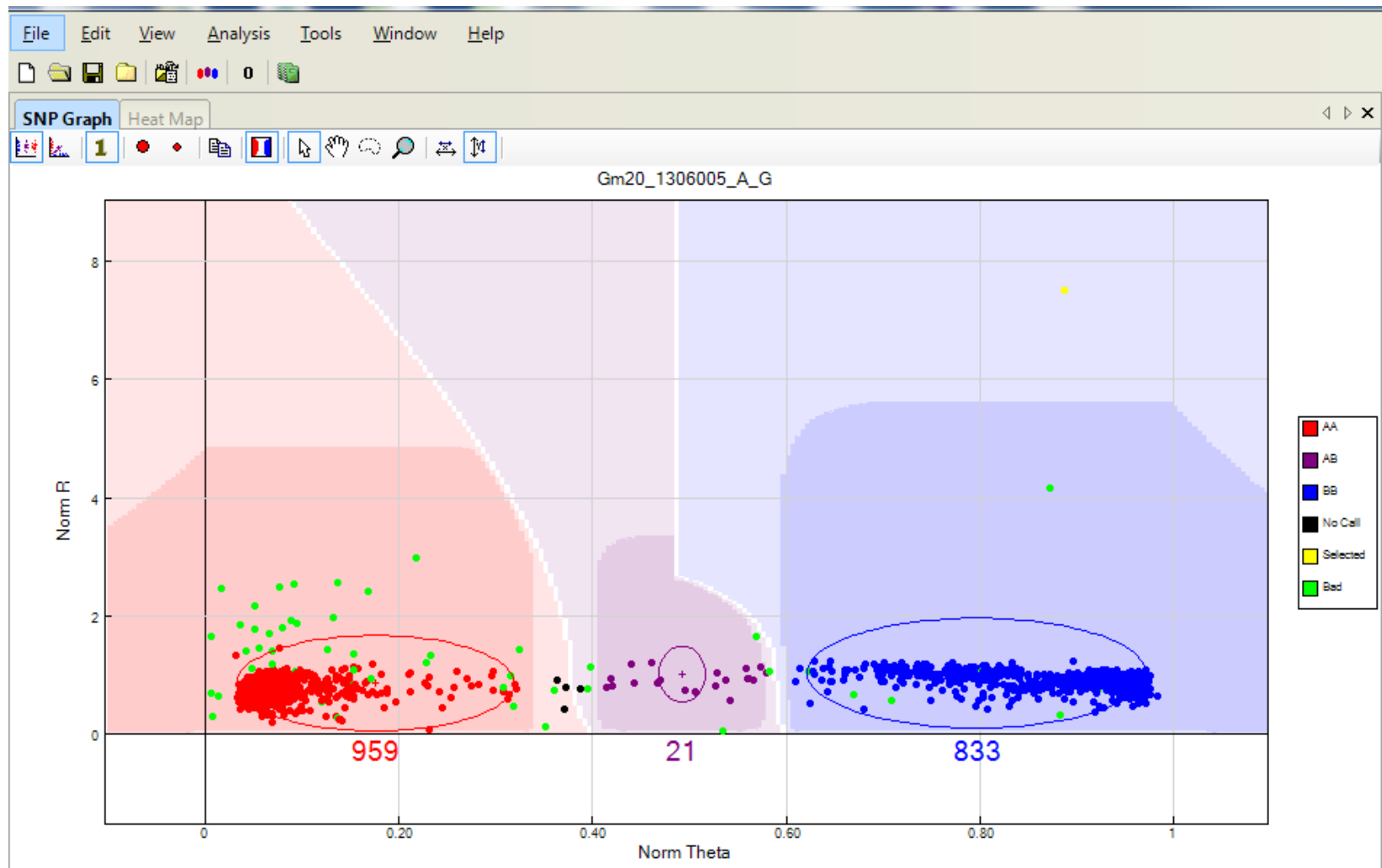
Analysis of the Entire USDA Soybean Germplasm Collection with the SoySNP50K Genechip

- ◆ 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

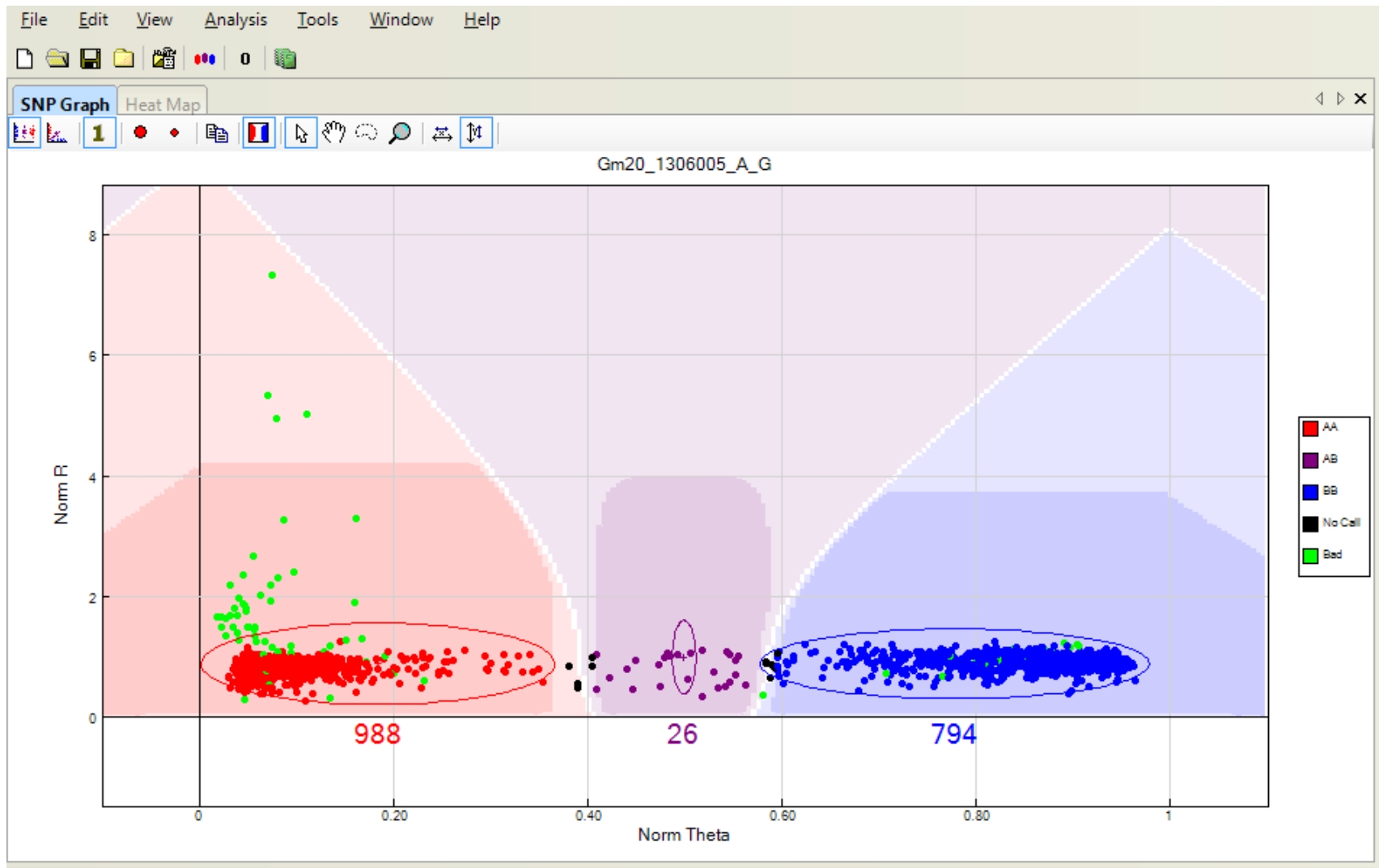
An example of a SNP with 19,768 accessions



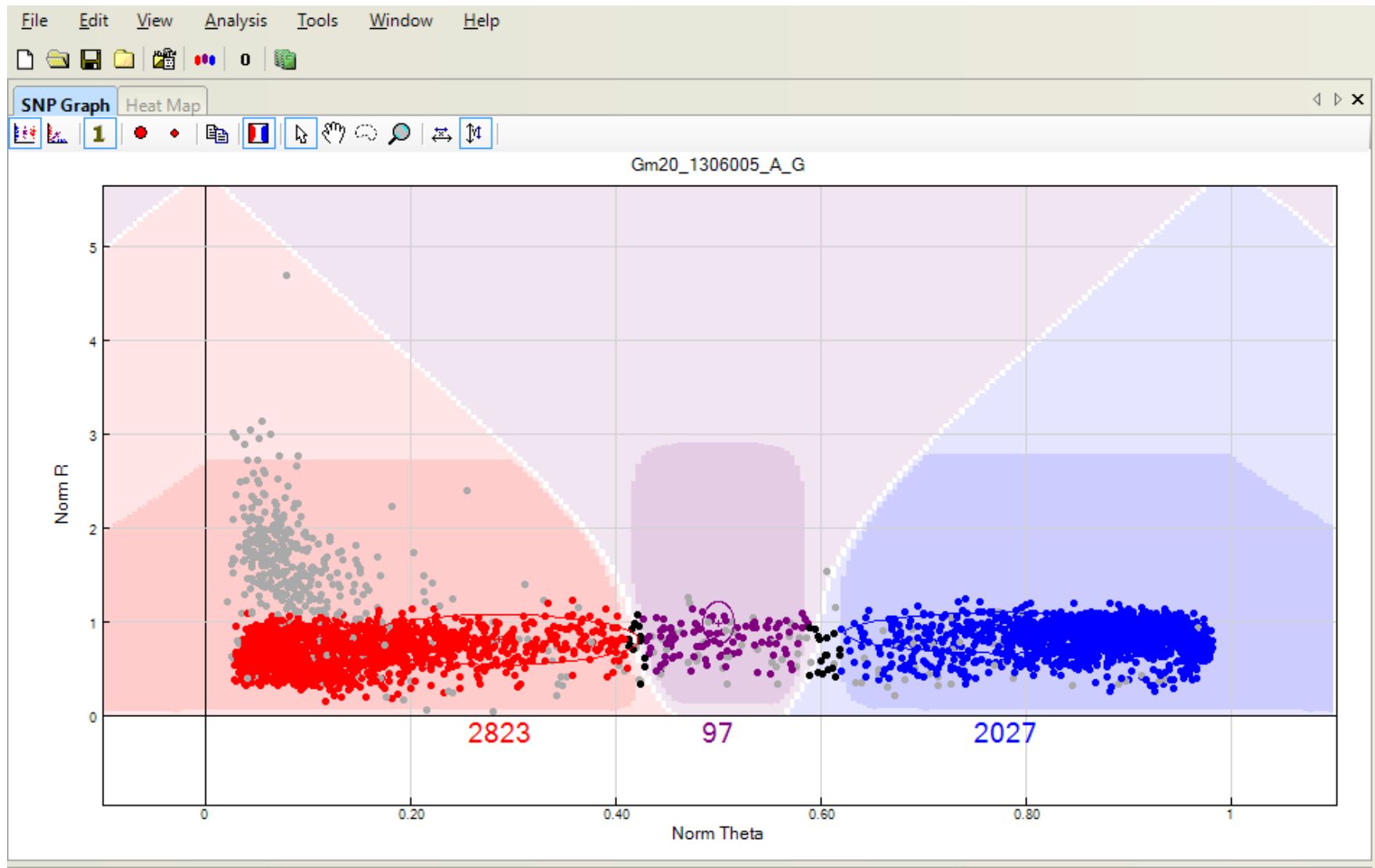
A Set of 1824 genotypes



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

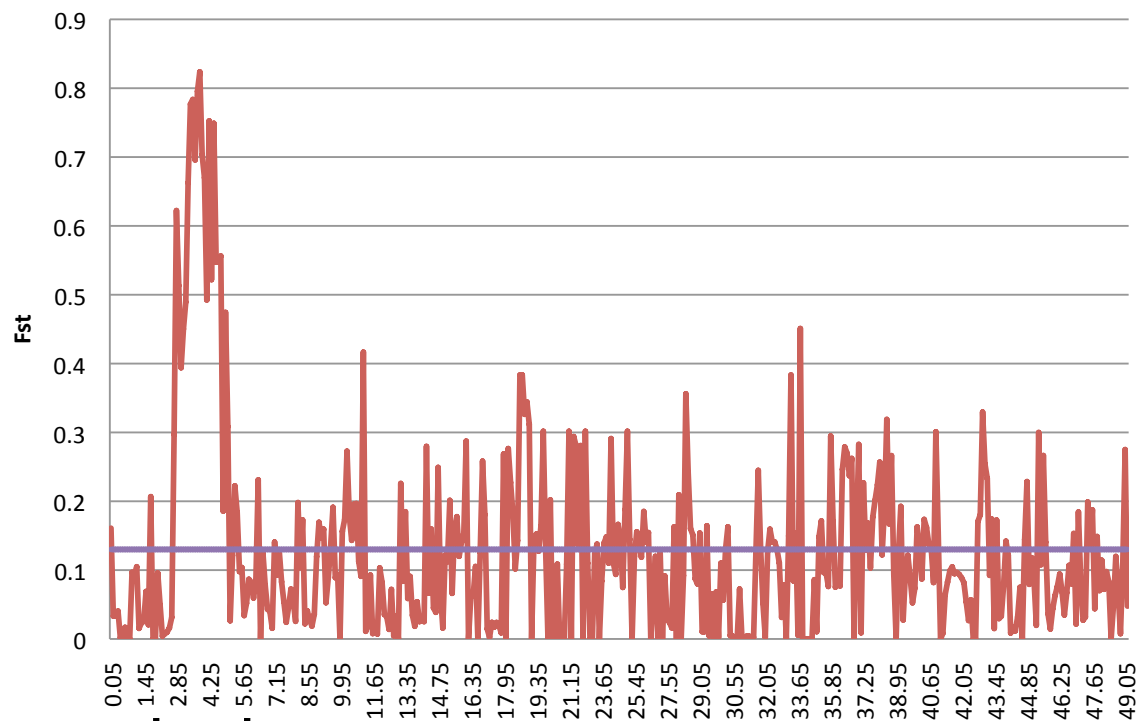
- ◆ F_{st} analysis of a few interesting regions – Asian Landraces vs. Elite cultivars and *G. max* vs. *G. soja*

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} Landraces versus Elite lines

Gm04 (Linkage Group C1)



QTL:

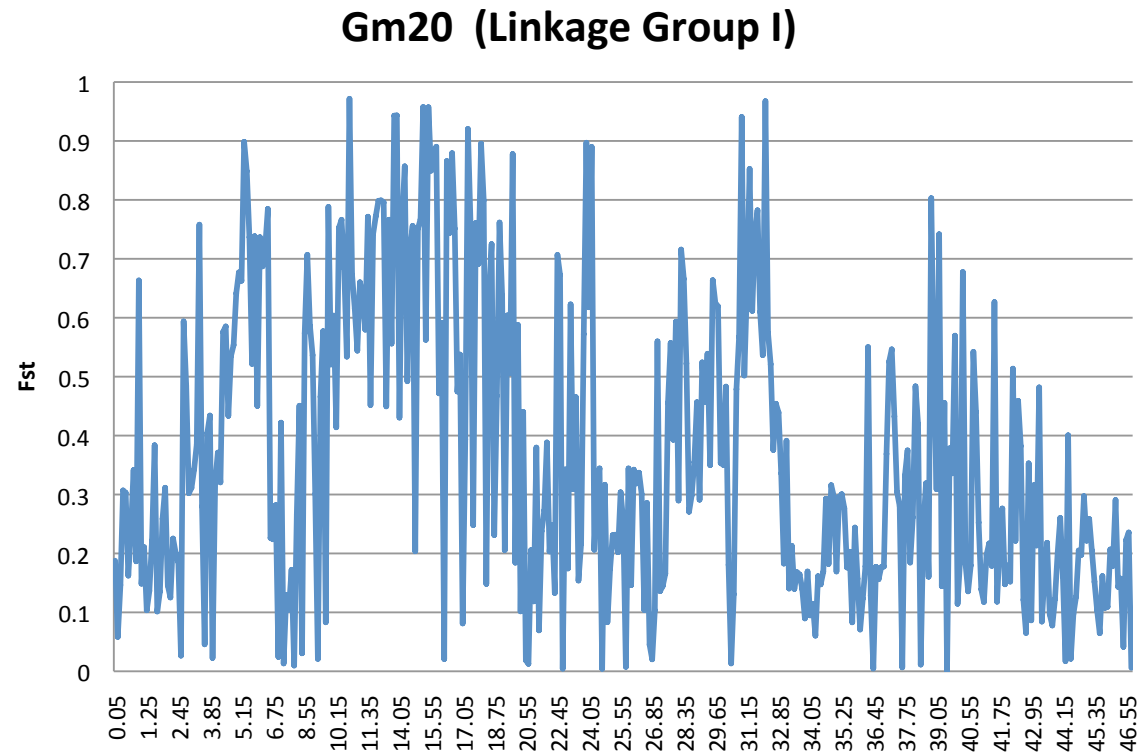
Plant height

Pods/node

Seed protein concentration

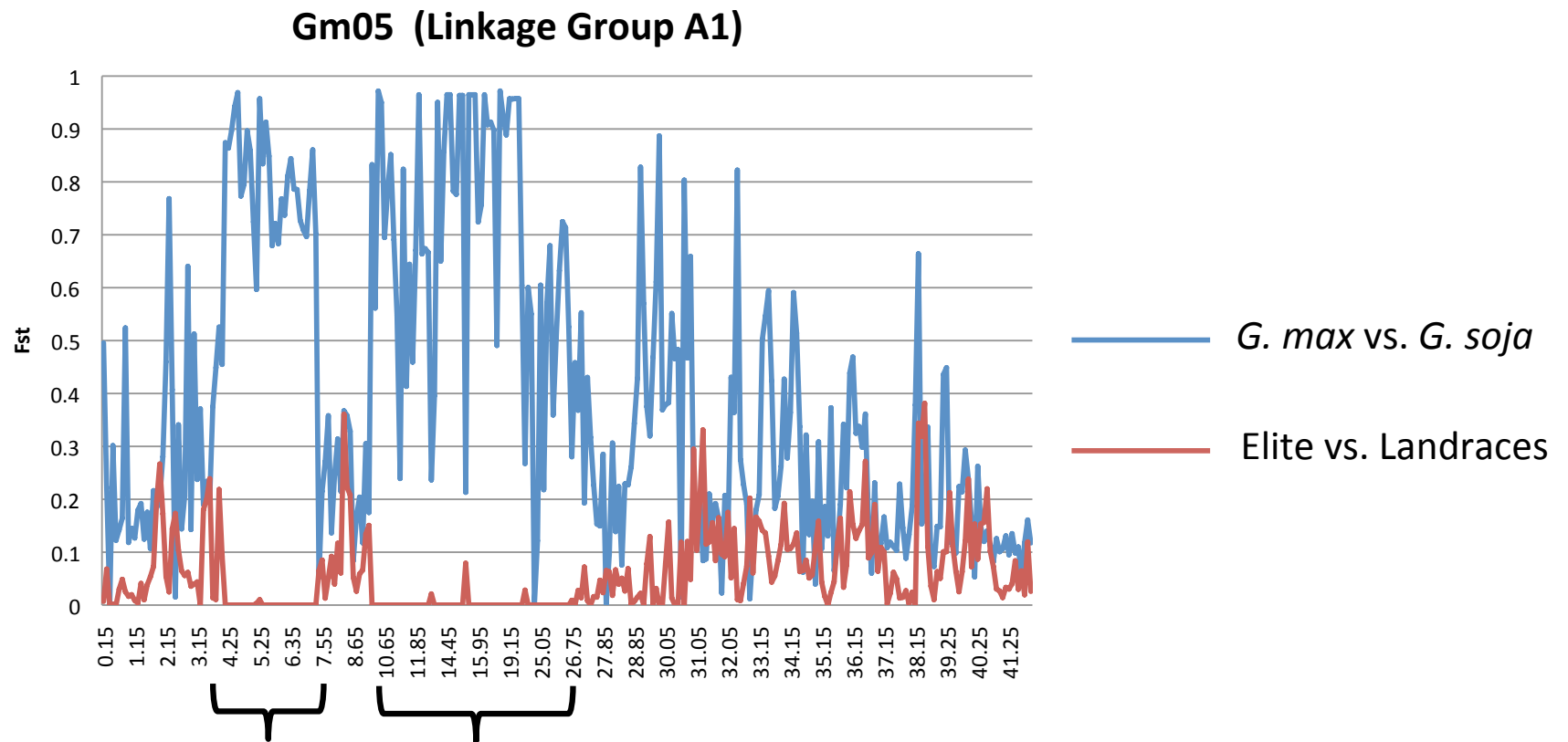
Seed size

F_{st} *G. max* (Landraces + Elite lines) versus *G. soja*



QTL:
Seed oil and protein concentration
Seed size

F_{st} of *G. max* Landraces vs. Elites lines and *G. max* versus *G. soja*

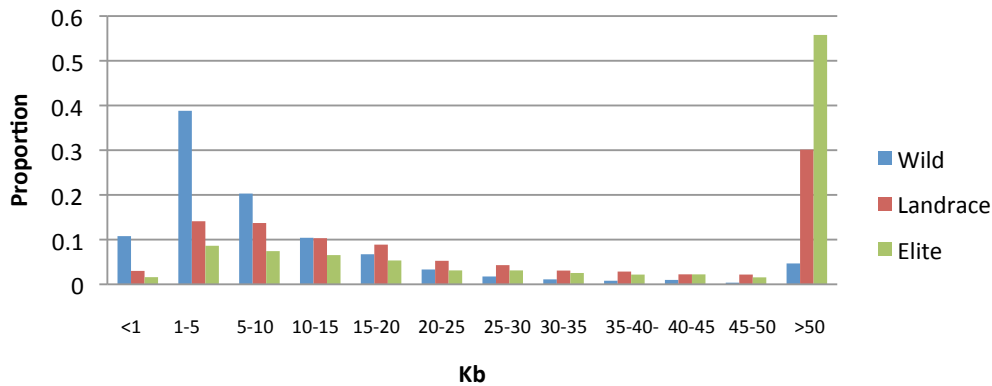


Many QTL have been reported in these two regions but all have been based upon the analysis of *G. max* x *G. max* populations.

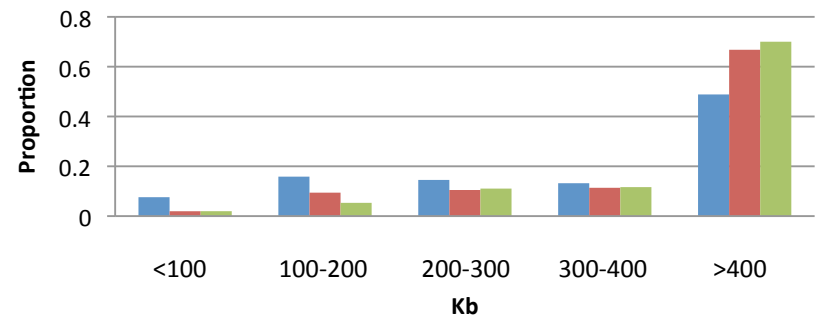
Haplotype Map Development and Application Based upon the SoySNP50K Analysis of 96 Elite cultivars, 96 Asian Landraces and 101 *G. soja* genotypes

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

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 *Rpp* Genes, *Rpp1* – *Rpp5*

Number of resistant lines with the indicated rust resistance allele(s)		Lines with various numbers of <i>Rpp</i> resistance genes	
<i>Rpp</i> locus	Number of lines	No. of <i>Rpp</i> genes per line	Number of lines
<i>Rpp1</i>	120	1	304
<i>Rpp2</i>	315	2	249
<i>Rpp3</i>	59	3	98
<i>Rpp4</i>	99	4	11
<i>Rpp5</i>	548	5	0
None	151		

* 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

