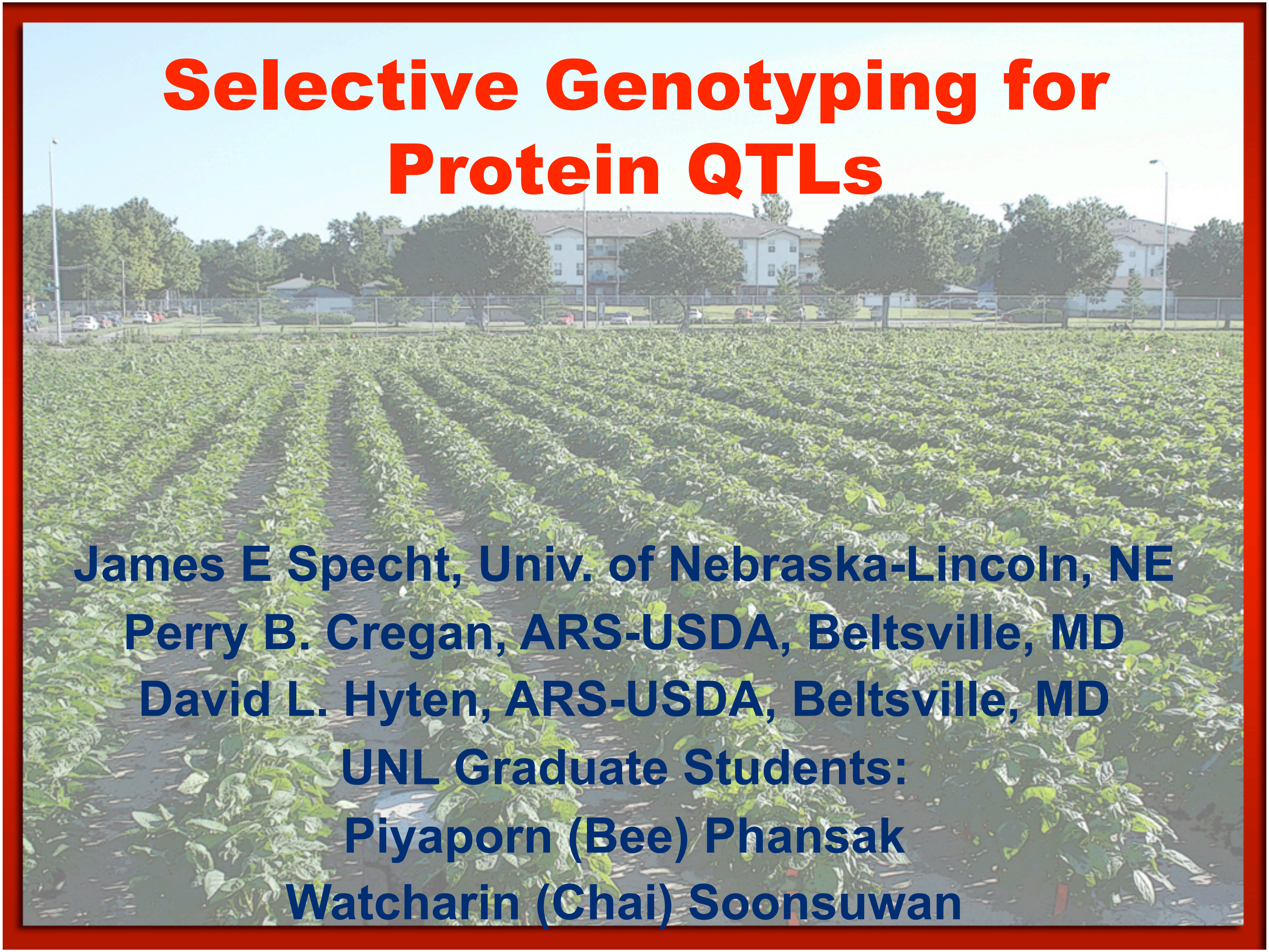
Selective Genotyping for Protein QTLs



James E Specht, Univ. of Nebraska-Lincoln, NE

Perry B. Cregan, ARS-USDA, Beltsville, MD

David L. Hyten, ARS-USDA, Beltsville, MD

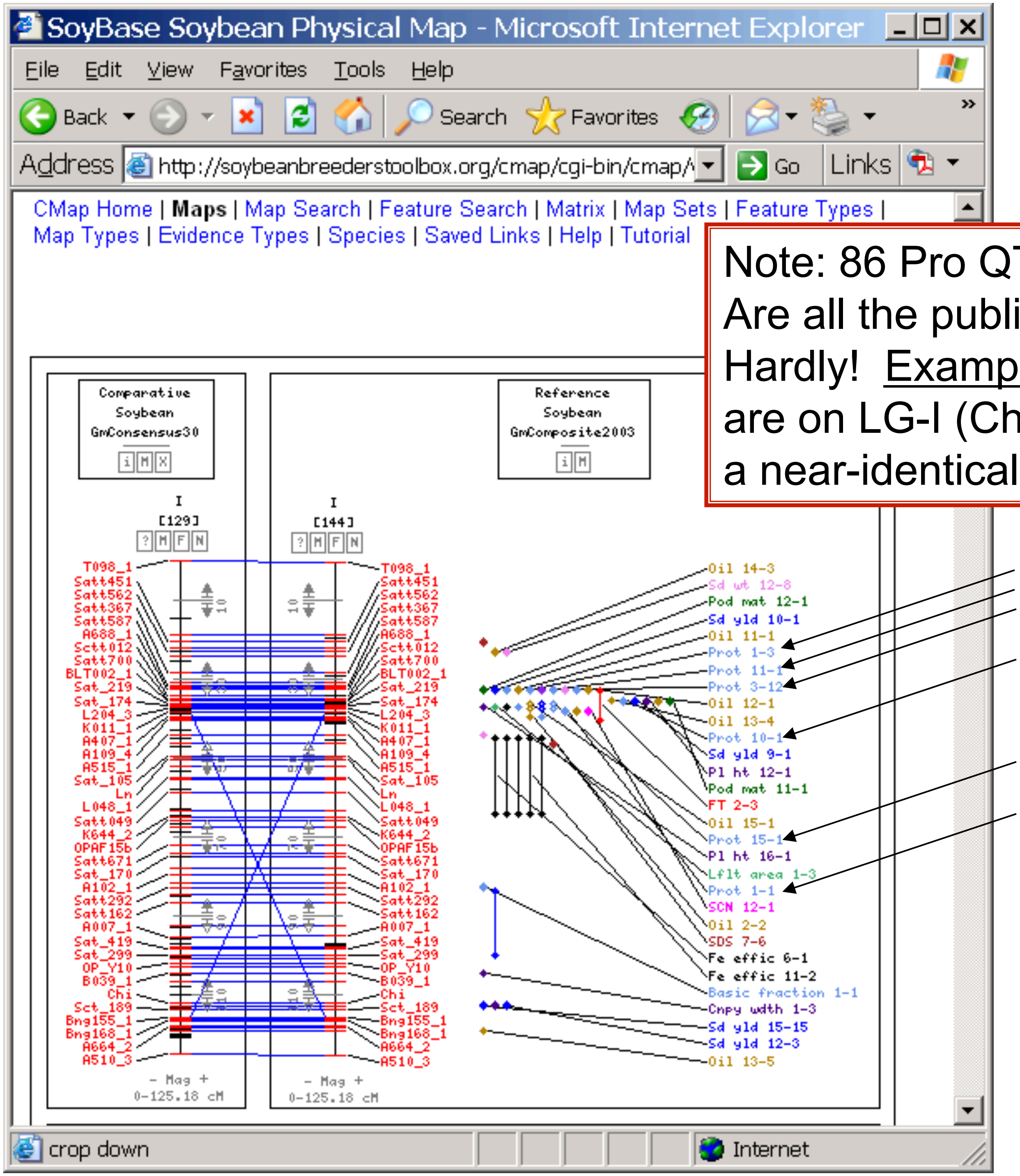
UNL Graduate Students:

Piyaporn (Bee) Phansak

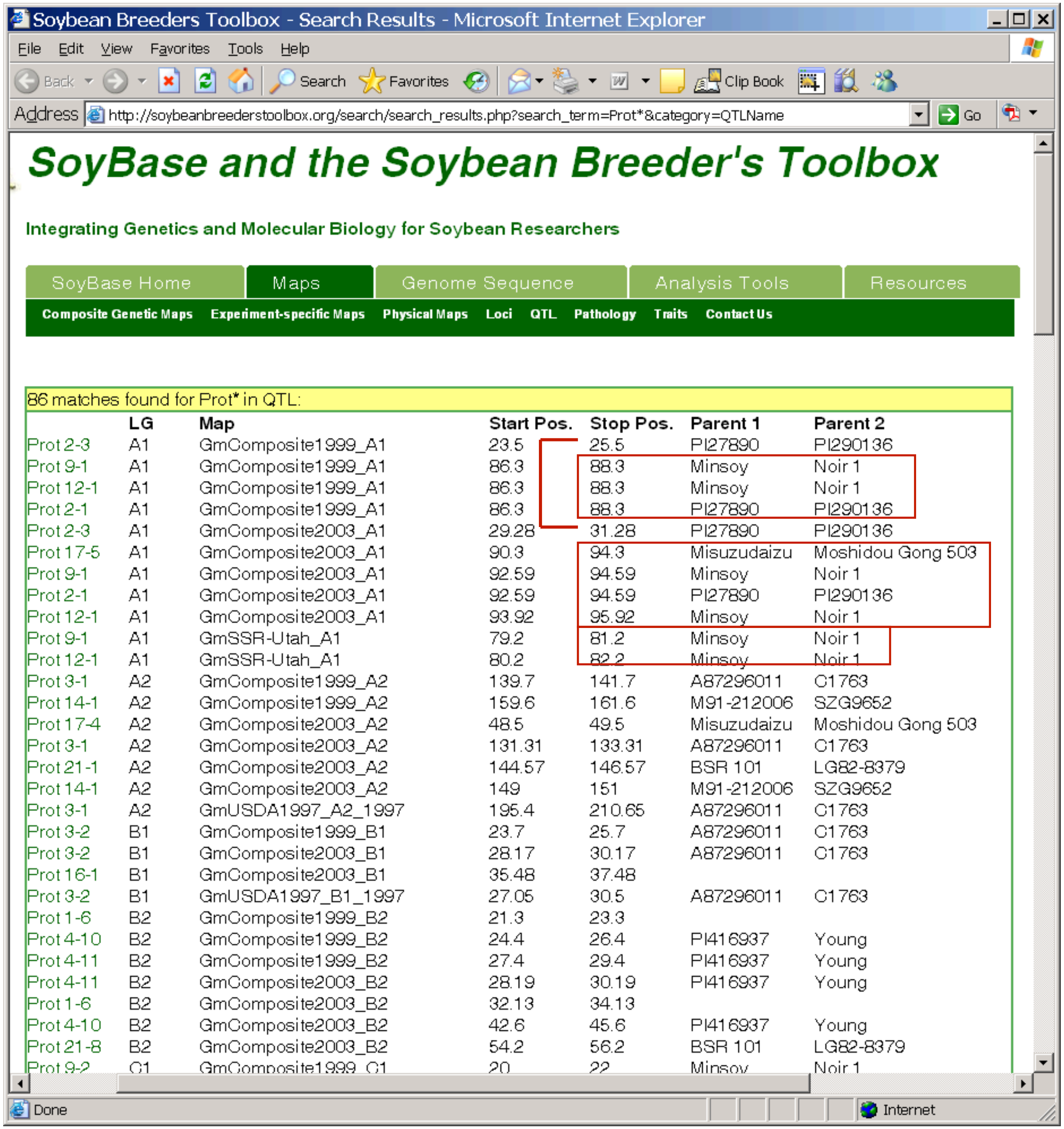
Watcharin (Chai) Soonsuwan

Introduction

- High seed protein germplasm accessions can be successfully used as parents for enhancing seed protein in a soybean breeding program.
- There are hundreds of such accessions in the USDA germplasm collection.
- How many different QTLs contribute to the high protein accessions? Conventional biparental F2 (or RIL) population analysis or use association analysis of high vs. low protein populations?



Note: 86 Pro QTLs in SOYBASE! Are all the published 86 unique? Hardly! <u>Example</u>: <u>Six</u> of the 86 are on LG-I (Chr 20) and map to a near-identical map position.


Soybean Breeders Toolbox - Search Results - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Clip Book
Address http://soybeanbreederstoolbox.org/search/search_results.php?search_term=Prot*&category=QTLName
Go
SoyBase Home
Maps
Genome Sequence
Analysis Tools
Resources
Composite Genetic Maps
Experiment-specific Maps
Physical Maps
Loci
QTL
Pathology
Traits
Contact Us
Done
Internet

SoyBase and the Soybean Breeder's Toolbox

Integrating Genetics and Molecular Biology for Soybean Researchers

86 matches found for Prot* in QTL:

	LG	Map	Start Pos.	Stop Pos.	Parent 1	Parent 2
Prot 2-3	A1	GmComposite1999_A1	23.5	25.5	PI27890	PI290136
Prot 9-1	A1	GmComposite1999_A1	86.3	88.3	Minsoy	Noir 1
Prot 12-1	A1	GmComposite1999_A1	86.3	88.3	Minsoy	Noir 1
Prot 2-1	A1	GmComposite1999_A1	86.3	88.3	PI27890	PI290136
Prot 2-3	A1	GmComposite2003_A1	29.28	31.28	PI27890	PI290136
Prot 17-5	A1	GmComposite2003_A1	90.3	94.3	Misuzudaizu	Moshidou Gong 503
Prot 9-1	A1	GmComposite2003_A1	92.59	94.59	Minsoy	Noir 1
Prot 2-1	A1	GmComposite2003_A1	92.59	94.59	PI27890	PI290136
Prot 12-1	A1	GmComposite2003_A1	93.92	95.92	Minsoy	Noir 1
Prot 9-1	A1	GmSSR-Utah_A1	79.2	81.2	Minsoy	Noir 1
Prot 12-1	A1	GmSSR-Utah_A1	80.2	82.2	Minsoy	Noir 1
Prot 3-1	A2	GmComposite1999_A2	139.7	141.7	A87296011	C1763
Prot 14-1	A2	GmComposite1999_A2	159.6	161.6	M91-212006	SZG9652
Prot 17-4	A2	GmComposite2003_A2	48.5	49.5	Misuzudaizu	Moshidou Gong 503
Prot 3-1	A2	GmComposite2003_A2	131.31	133.31	A87296011	C1763
Prot 21-1	A2	GmComposite2003_A2	144.57	146.57	BSR 101	LG82-8379
Prot 14-1	A2	GmComposite2003_A2	149	151	M91-212006	SZG9652
Prot 3-1	A2	GmUSDA1997_A2_1997	195.4	210.65	A87296011	C1763
Prot 3-2	B1	GmComposite1999_B1	23.7	25.7	A87296011	C1763
Prot 3-2	B1	GmComposite2003_B1	28.17	30.17	A87296011	C1763
Prot 16-1	B1	GmComposite2003_B1	35.48	37.48		
Prot 3-2	B1	GmUSDA1997_B1_1997	27.05	30.5	A87296011	C1763
Prot 1-6	B2	GmComposite1999_B2	21.3	23.3		
Prot 4-10	B2	GmComposite1999_B2	24.4	26.4	PI416937	Young
Prot 4-11	B2	GmComposite1999_B2	27.4	29.4	PI416937	Young
Prot 4-11	B2	GmComposite2003_B2	28.19	30.19	PI416937	Young
Prot 1-6	B2	GmComposite2003_B2	32.13	34.13		
Prot 4-10	B2	GmComposite2003_B2	42.6	45.6	PI416937	Young
Prot 21-8	B2	GmComposite2003_B2	54.2	56.2	BSR 101	LG82-8379
Prot 9-2	C1	GmComposite1999_C1	20	22	Minsoy	Noir 1

Another Example: Note that 11 of the 86 QTLs on LG-A1 (Chr 5) mapped to only 4? or 3? or 2? different positions. So, far fewer than 86 QTLs available to breeders.

Another Issue: In most cases, the lines used in the bi-parental matings listed here did not have much of a contrast in seed protein content.

With 86 Seed Protein QTLs why search for more?

The mean map position for a QTL typically has a plus or minus confidence interval of ~10 cM, so many of the 86 QTLs are likely repeat discoveries of fewer existent QTLs!

Of the mapped QTLs that truly have different map positions, none have an additive effect greater that the respective 1.2 g/kg or 0.85 g/kg additive effects of the protein QTLs on LG-I and LG-E discovered long ago by Diers et al. (1992). All other protein QTLs documented in SoyBase have much smaller additive effects.

The high protein germplasm accessions in the USDA have not been systematically evaluated for protein QTLs. Do protein QTLs beyond those already reported in SoyBase exist in this germplasm?

Frequency of Seed Protein / Oil Values in the *Glycine max* Collection

Descriptor: PROTEIN   obtype: NUMERIC   CGC: YES
Protein percent of dry weight of seed. This descriptor is a numeric field. Blank value means no data. Ex: 26.8, 49.9

Value	Freq.	Value	Freq.	Value	Freq.
31.70 - 34.72	2	40.76 - 43.78	5601	49.82 - 52.84	383
34.72 - 37.74	216	43.78 - 46.80	6484	52.84 - 55.86	56
37.74 - 40.76	1537	46.80 - 49.82	2027	55.86 - 58.88	6

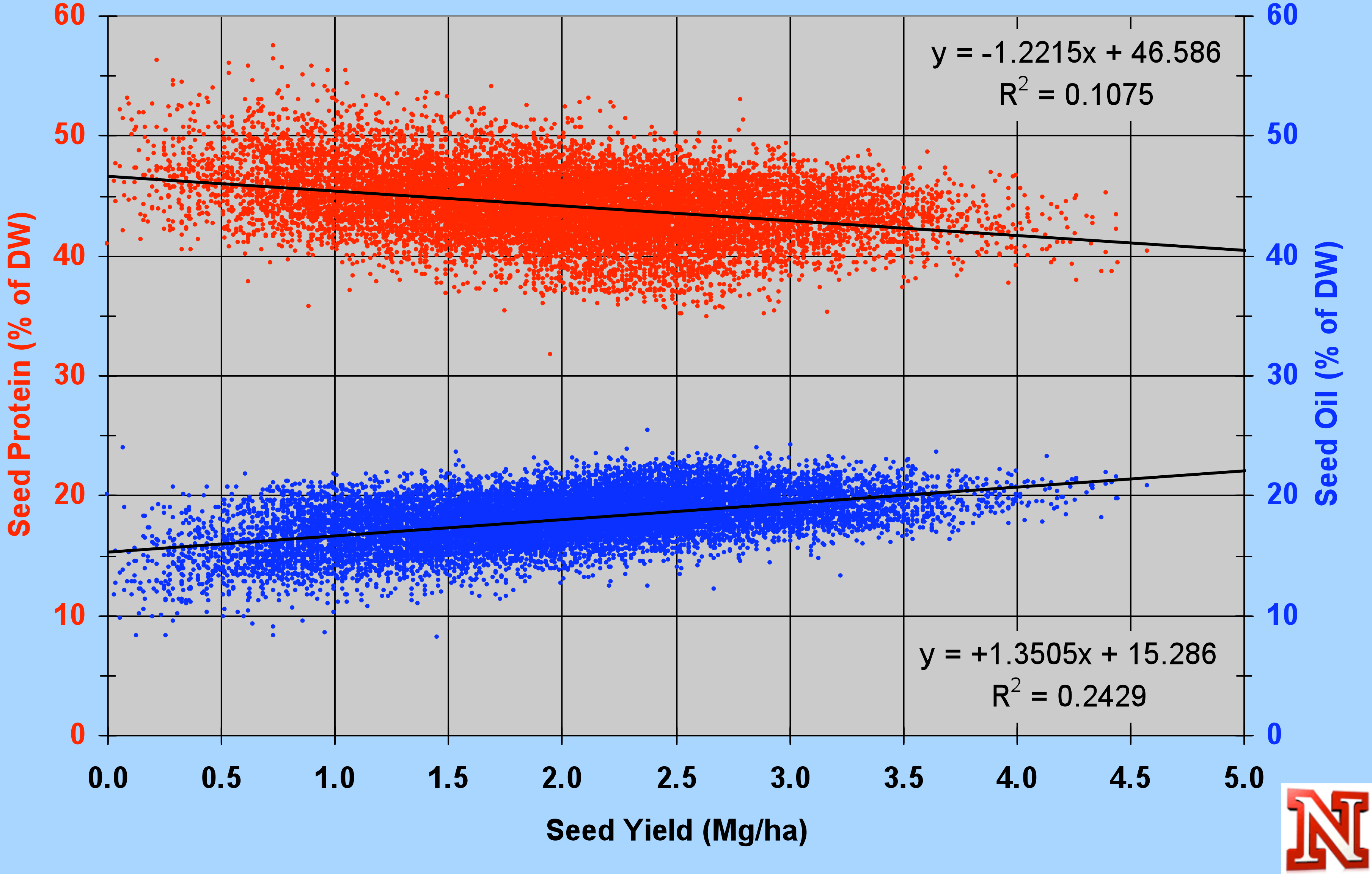
Descriptor: OIL   obtype: NUMERIC   CGC: YES
Oil percent of dry weight of seed. Descriptor is a numeric field. Blank value means no data. Ex: 16.4, 19.2

Value	Freq.	Value	Freq.	Value	Freq.
6.50 - 8.71	42	13.13 - 15.34	1348	19.76 - 21.97	2858
8.71 - 10.92	147	15.34 - 17.55	4828	21.97 - 24.18	274
10.92 - 13.13	275	17.55 - 19.76	6538	24.18 - 26.39	2

Source: GRIN, March 2009

UNIVERSITY OF Nebraska Lincoln

Seed Protein/Yield/Oil - Germplasm Resources Information Network (GRIN) Values for ca. 15,000 of the 18,000 *Glycine max* Accessions



Objectives

- Identify the primary protein QTL(s) governing the high seed protein content in each of ~50 high protein germplasm accessions of MG 000, 00, 0, I, II, III, & IV.
- Use selective genotyping (i.e., phenotypic tail analysis) with the new 1536-SNP USLP 1.0 linkage panel (Hyten et al., 2010) to map the seed protein QTLs in 240-plant F2 populations derived from the mating of each of the 50 high protein PI with an elite high-yielding cultivar of equivalent MG.

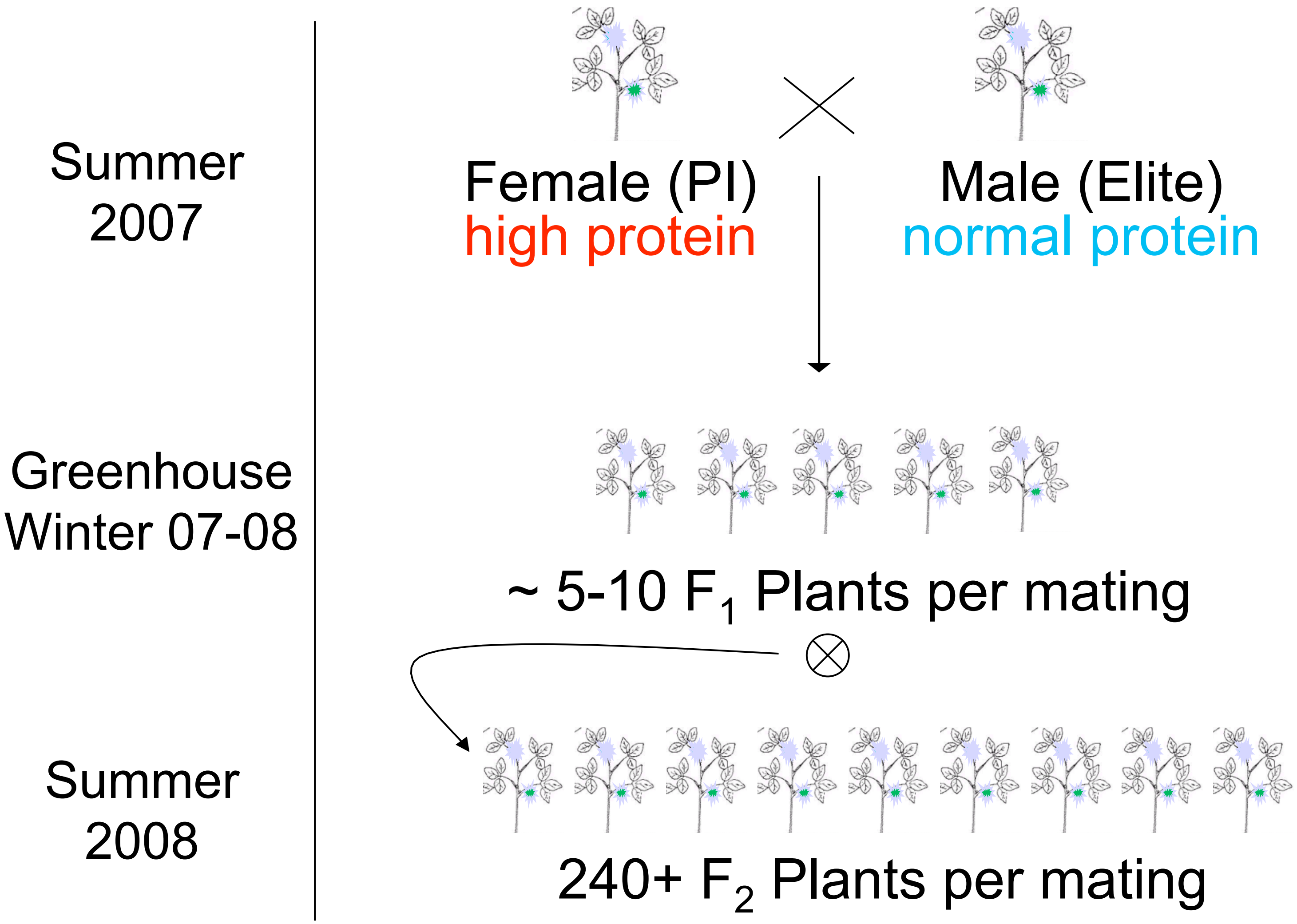
Parental germplasm description. Each accession was mated to an agronomic cultivar of similar maturity (blue highlighted).

There were 52 matings, but four did not produce at least 240 F2 plants.

Maturity group	Parent code	Plant Introduction	Name (if any)	Protein	Oil	Flower color	Hilum color	Pod color	Pub color	Pub form	Scoat color	Scoat lust	Shatearly
000	1001	PI153296	V-4	529	151	P	Bl	Br	T	E	Gn	S	2.0
000	1002	PI 189963	Geant Vert	504	158	P	Bl	Br	T	E	Gn	D	2.0
000	1003	PI 548399	Pando	522	155	P	Bl	Br	T	E	Gn	S	1.0
000	1004	PI 372423	Ronset 4	477	156	P	Bl	Br	T	E	Lgn	I	
000	1005	FC 30687	Kosodiguri Extra Early	512	157	P	Bl	Br	T	E	Gn	I	3.0
000	1006	PI 153293	N-34	511	158	P	Bl	Br	T	E	Gn	S	2.0
000	1007	PI 372412	Hercumft	478	161	P	Bl	Tn	T	E	Lgn	S	
000	1008	PI 548341	Hidatsa	509	163	P	Bl	Br	T	E	Gn	S	2.0
000	1009	PI 548414	Siox	522	159	P	Bl	Br	T	E	Gn	S	2.0
000	1021M	PI 567787	OAC Vision	430/427	199/186	P	Tn	Br	T	E	Y	D	3.0
00	1022	PI 153302	V-16	507	158	P	Bl	Br	T	E	Gn	S	1.5
00	1023	PI 159764		526	157	P	Bl	Br	T	E	Gn	S	1.5
00	1024	PI 438415	Ronest 4	485	164	P	Bl	Br	T	E	Gn	I	
00	1025	PI 153301	V-14	508	147	P	Bl	Br	T	E	Gn	S	1.5
00	1026	PI 189880	Bitterhof	489	173	P	Y	Br	G	E	Y	S	1.0
00	1027	PI 153297	V-6	510	148	P	Bl	Br	T	E	Gn	S	1.5
00	1038M	PI 602897	Jim	416	185	P	Y	Br	G	E	Y	I	2.0
	2211		HHP				Bl				Bl		
	2212		AC Proteus			P	Br	Br	T		Y	D	
	2213		AC Proteina			P	Br	Br	T		Y		
0	1039	PI 427138	Choseng No. 1	480	144	W	Bf	Br	G	A	Y	D	
0	1040	PI 261469	Wasedaizue No. 1	488	195	W	Bf	Br	G	A	Y	D	1.0
0	1041	PI 181571	No. 58	485	177	W	Bf	Br	G	A	Y	D	2.0
0	1042	PI 424148	Shirome (Korea)	483	150	W	Bf	Br	G	A	Y	I	2.0
0	1043	PI 423954	Shirome (Japan)	473	156	W	Bf	Br	G	Sa	Y	D	
0	1044	PI 154196	No. 51	494	160	P	Bl	Br	T	E	Gn	D	1.0
0	1053M	PI 602594	MN 301	403	196	P	Y	Br	G	E	Y	I	2.0
i	1054	PI 437088A	DV-147	484	155	P	Br	Br	T	E	Y	D	1.0
i	1055	PI 423949	Saikai 20	514	144	Lp	Bf	Br	G	A	Y	I	1.0
i	1056	PI 427141	Seuhae No. 20	495	141	P	Br	Br	T	E	Y	D	1.5
i	1057	PI 437716A	Sjuj-dja-pyn-da-do	482	138	P	Bf	Br	G	Sa	Y	I	3.0
i	1058	PI 423942	Saikai 1	489	149	P	Bf	Tn	G	A	Y	I	2.0
i	1074M	PI 602593	MN1301	378/407	207/195	W	Y	Br	G	E	Y	D	1.0

ii	1075	PI 423948A	Saikai 18	499	157	Lp	Bf	Br	G	E	Y	S	1.0
ii	1076	PI437112A		482	154	W	Y	Tn	G	E	Y	S	1.0
ii	1098	PI 548608	Provar	484	191	P	Br	Br	T	E	Y	D	1.0
ii	1106M	PI 597386	Dwight	382	195	P	Bl	Tn	T	E	Y	D	1.0
iii	1107	PI 445845	Szu yueh pa	504	132	W	Bf	Tn	G	A	Y	D	3.0
iii	1108	PI 398516		494	167	P	Y	Br	Lt	E	Y	D	1.0
iii	1109	PI 91725-4	Akazu	477	170	W	Bf	Br	G	Sa	Y	D	1.0
iii	1110	PI 340011		493	165	P	Y	Br	G	E	Y	D	2.5
iii	1111	PI 243532	Kariho-takiya	478	162	W	Br	Dbr	T	E	Y	S	5.0
iii	1112	PI 438427		493	139	P	Y	Br	G	E	Y	D	1.0
iii	1113	PI 408138C		497	168	P	Y	Br	G	E	Y	D	1.5
iii	1121	PI 398672		494	177	Dp	Rbr	Br	T	E	Rbr	S	1.0
iii	1122	PI 360843	Oshimashirome	484	184	W	Y	Br	G	E	Y	I	1.0
iii	1137M	PI 597387	Pana	411	194	P	Bf	Br	G	E	Y	D	1.0
iv	1138	PI 253666		479	157	W	Bf	Br	G	Sa	Y	I	1.0
iv	1139	PI 407788A	ORD 8113	507	151	P	Bf	Tn	G	E	Y	S	1.5
iv	1140	PI 424286		493	155	P	Bf	Tn	G	E	Y	D	1.5
iv	1141	PI 404177	Tiu sen jan lj gu	514	155	P	Y	Lbr	G	E	Y	D	1.0
iv	1142	PI 407877B	KAREI 511-11	488	166	P	Bf	Br	G	E	Y	D	1.5
iv	1143	PI 398704		488	158	P	Bf	Br	G	E	Y	I	1.0
iv	1145	PI 398970		491	160	P	Lbf	Tn	G	E	Y	D	1.0
iv	1146	PI 407823		493	159	P	Bf	Tn	G	E	Y	I	1.5
iv	1148	PI 407845A		491	166	P	Y	Tn	G	E	Y	S	1.5
iv	1152	PI 407773B		492	161	W	Bl	Tn	T	E	Y	I	1.0
iv	1181M	PI 606748	Rend	424	180	W	Bf	Br	G	E	Y	D	1.0
v	1183	PI 458.256		476	195	P	Y	Br	G	Sa	Y	I	2.5
v	1183		Essex	Qui et al. reported Essex had an allele that gave both high protein AND high oil									

Population Development



Summer
2007

Female (PI)
high protein

×

Male (Elite)
normal protein

Greenhouse
Winter 07-08

~ 5-10 F_1 Plants per mating

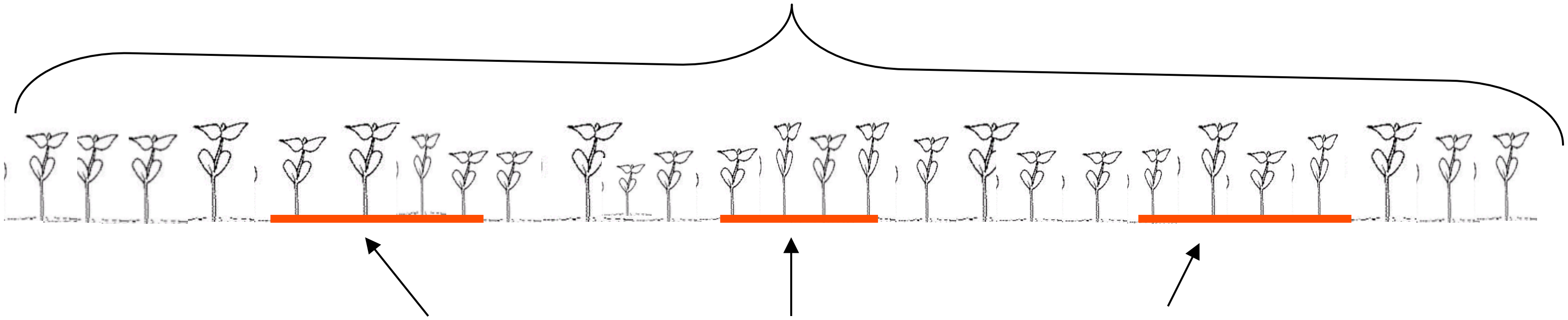
⊗

Summer
2008

240+ F_2 Plants per mating

Field Layout (one F_2 pop)

Population of 240 F_2 plants
spaced 3”apart in the row

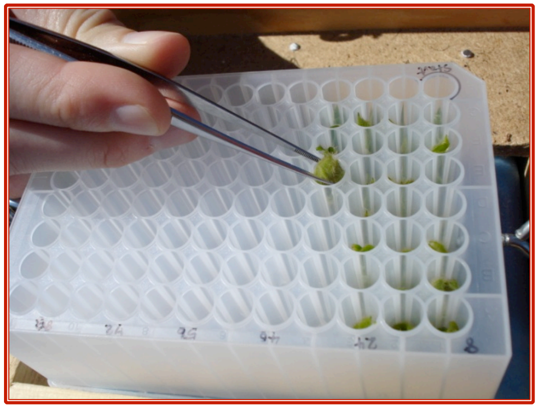


Four female and Four male parents (also 3” apart) were repeated at five positions within each F_2 plant row to obtain a measure of field variation on the protein content of a homozygous genotype

Collecting leaf samples

- Leaf samples were collected 3-4 weeks after planting date (starting with earliest MG matings first).
- One trifoliolate leaf was collected from each F_2 plant.
- The leaf collection plates (3, for up to 288 F2 plants per population) kept on dry ice in the field.

The leaf collection plates (3 per pop), when full, were put immediately into a -80C freezer.



Harvesting & Threshing

- Mature F_2 plants were individually harvested into bags
- F_2 plants were individually threshed to create packets of F_2-derived F_3 seed progenies
- Parental and F_1 plants planted among F_2 plants were also individually threshed

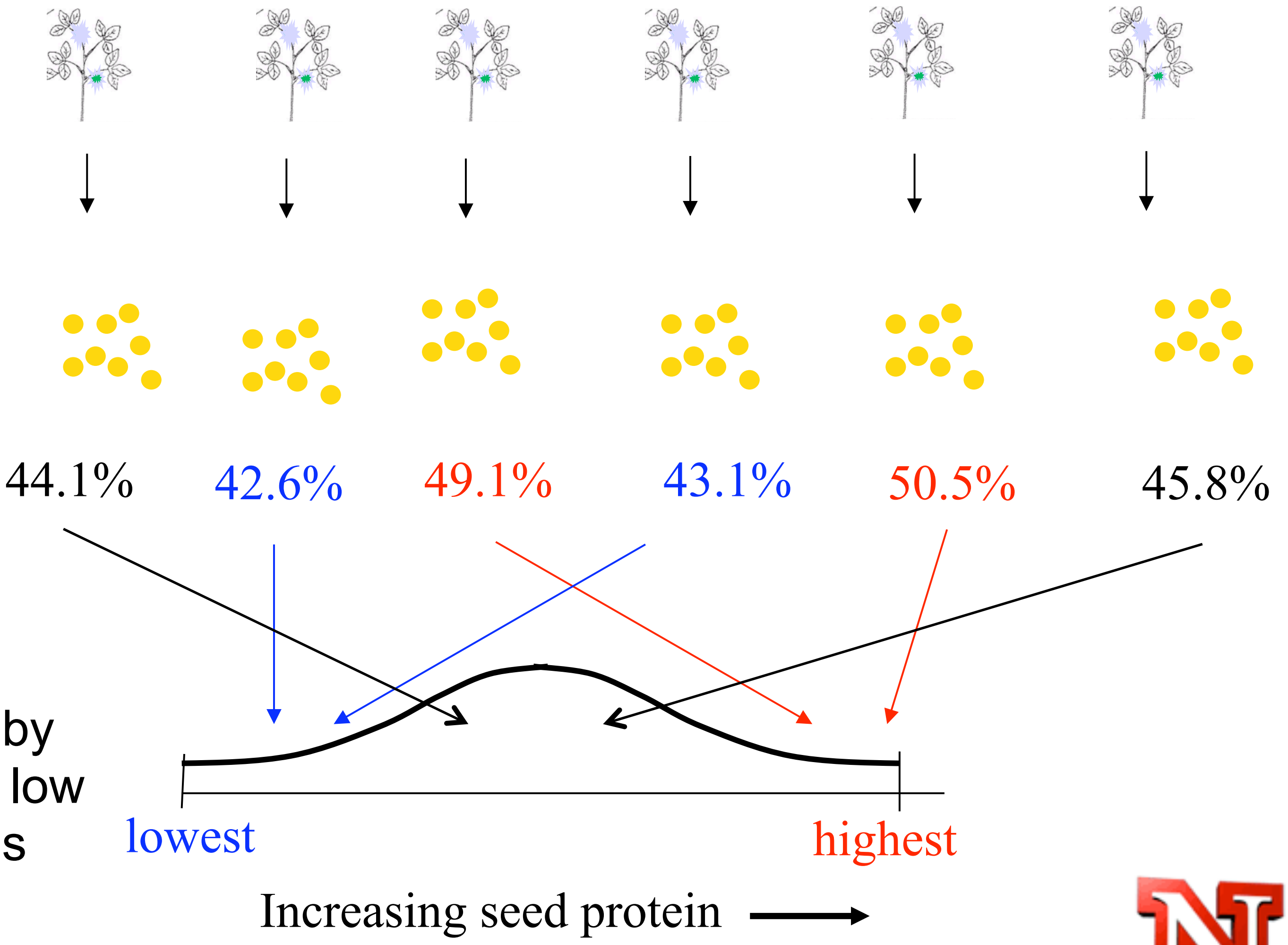
Phenotyping

F$_{2.3}$ seed NIR Analysis

~240 F$_2$ plants

NIR-based Protein of F$_{2.3}$ seed

Rank progenies by protein - Identify low and high quintiles



44.1% 42.6% 49.1% 43.1% 50.5% 45.8%

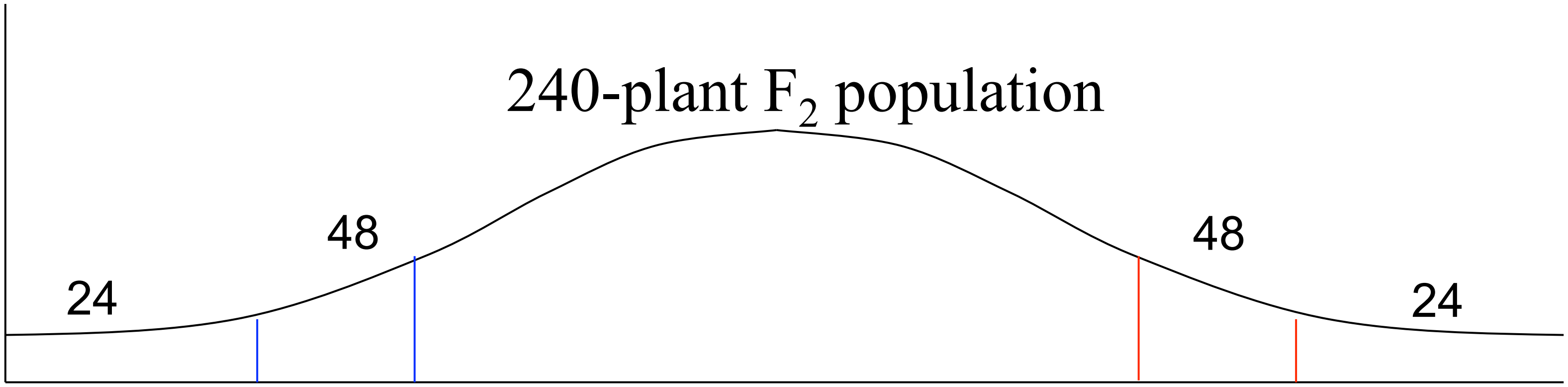
lowest

highest

Increasing seed protein

$F_{2.3}$ seed NIR Analysis

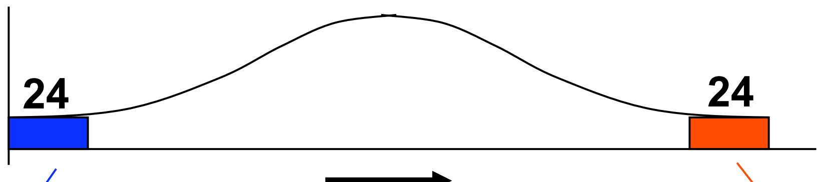
Redo (2^{nd} rep) NIR on low and high quintiles (i.e., validate the phenotypic values in the tails) to increase the precision of decile selection!



$F_{2.3}$ - lowest decile (24) of total protein range

$F_{2.3}$ - highest decile (24) of total protein range

240-plant F_2 population

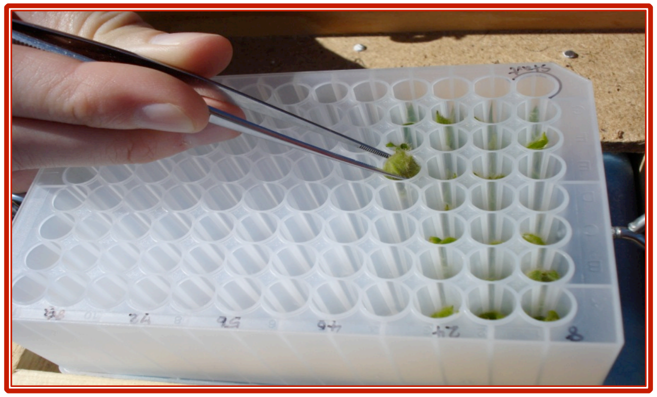


Increasing seed protein

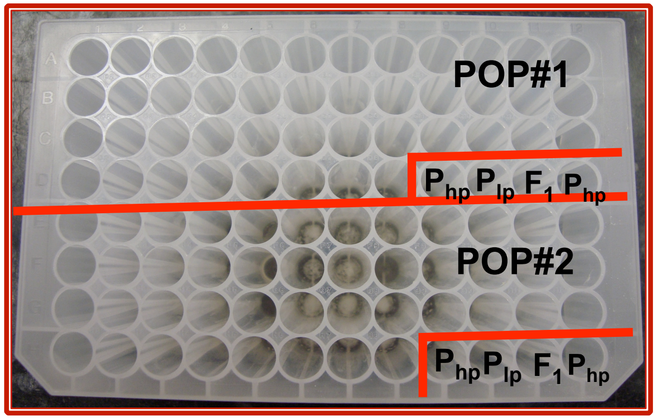
F_2 plants producing lowest protein F_3 seed (Lowest 10%)

F_2 plants producing highest protein F_3 seed (Highest 10%)

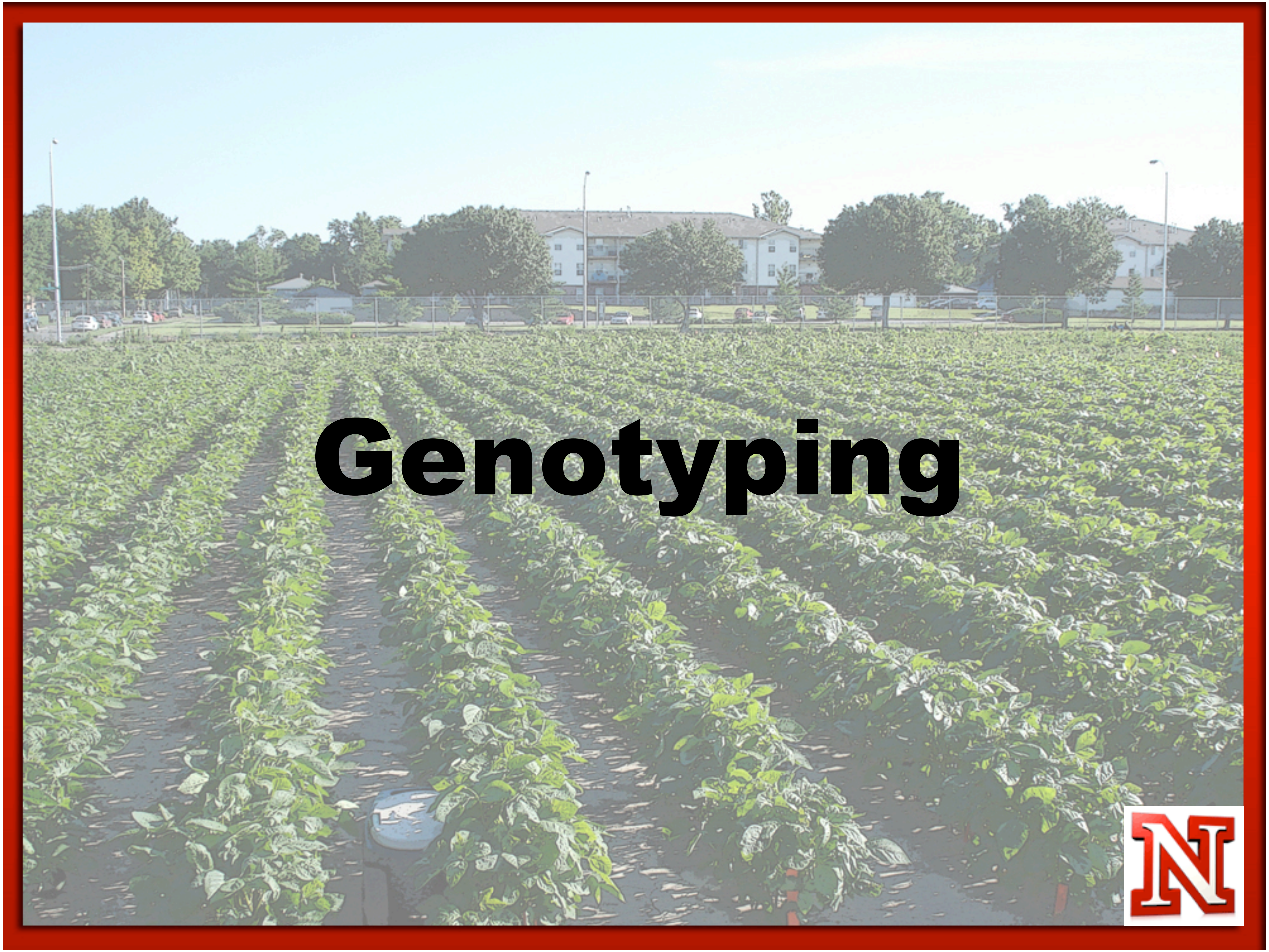
F_2 leaflet retrieved from collection plates for each respective $F_{2.3}$ seed progeny in each decile fraction (22L:22H = 44 wells).



2-pop 96-well plates with frozen leaflets shipped to Beltsville for DNA extraction and SNP genotyping


POP#1
P_{hp} P_{lp} F_1 P_{hp}
POP#2
P_{hp} P_{lp} F_1 P_{hp}

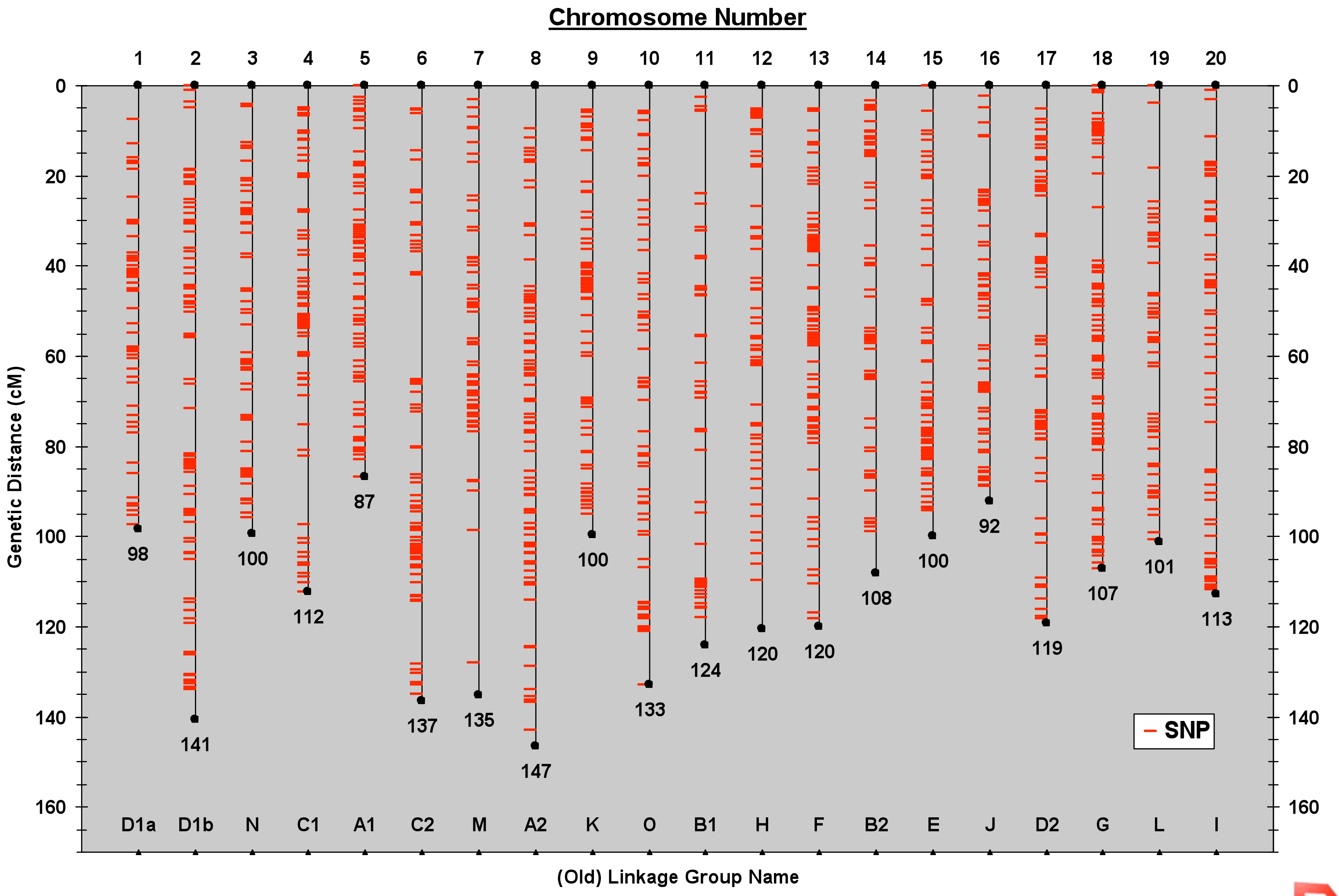
Genotyping



Single Nucleotide Polymorphism (SNP) Analysis

- Screen the 24 high protein and 24 low protein $F_{2.3}$ progenies with 1536 SNP loci in USLP 1.0 (Hyten et al., 2010, forthcoming in Crop Science).
- SNP allele analyses were conducted using GoldenGate assay on the Illumina® BeadStation 500 Genotyping Platform
- The 1536 SNP analysis on the first 20 populations were conducted by personnel at the Genomics and Improvement Laboratory, USDA, Beltsville, MD.
- The other 28 populations are not yet analyzed.

Soybean Genetic Map - USLP 1.0 - 1536 SNP Markers

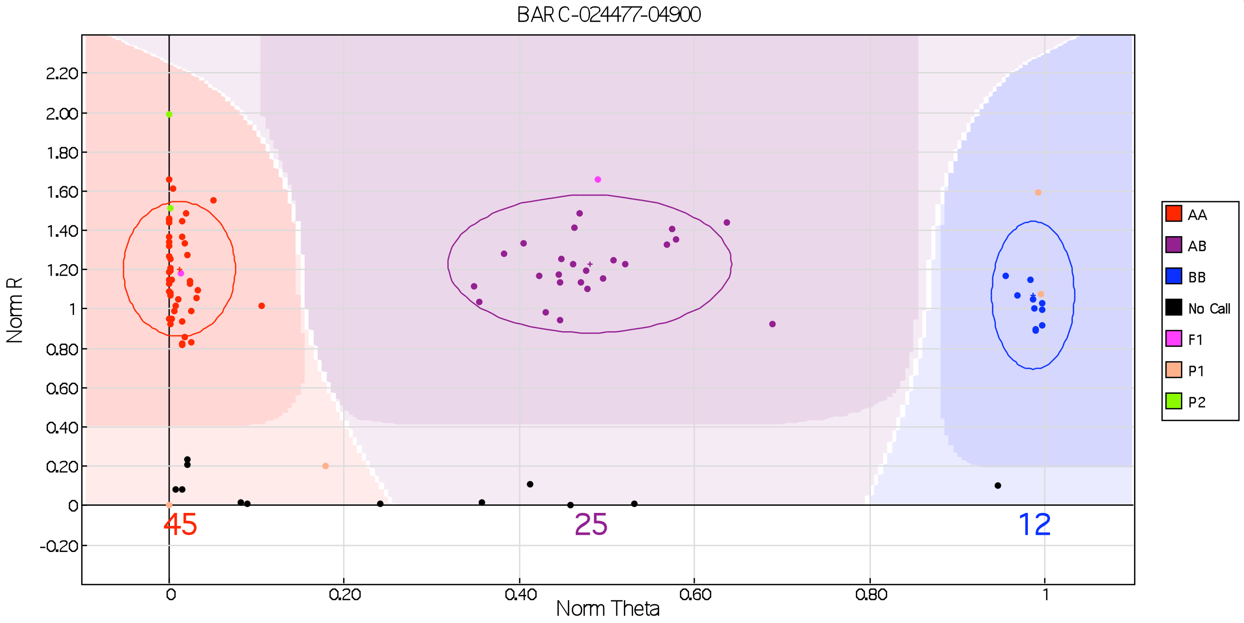


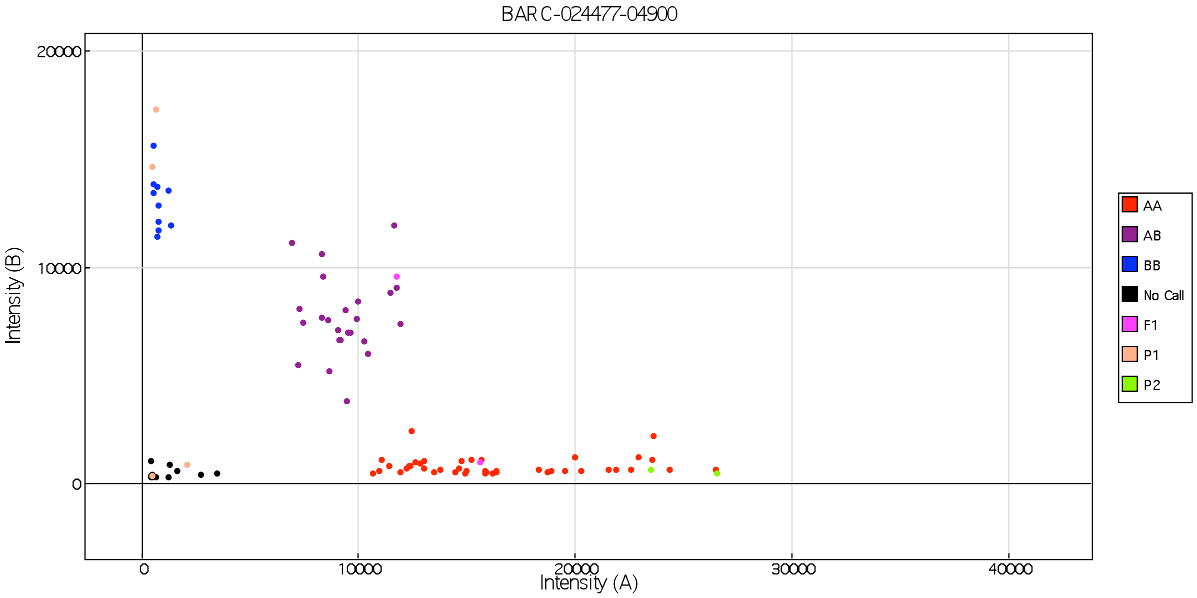
(Hyten et al. (2010); Crop Sci (in press)

SNP Analysis

- BeadStudio software was the program used for SNP allele scoring (next slide).
- Not all 1536 SNPs can be parentally bimorphic in any given population, but in elite x PI matings, one can expect bimorphism for ~500 SNPs.
- Goal: Genotyping the lowest and highest deciles of each F_2 protein distribution, SNP loci with a skewed allele frequency between the two decile groups is suggestive of a SNP locus linkage to a protein QTL.

SNP Analysis – Bead Studio

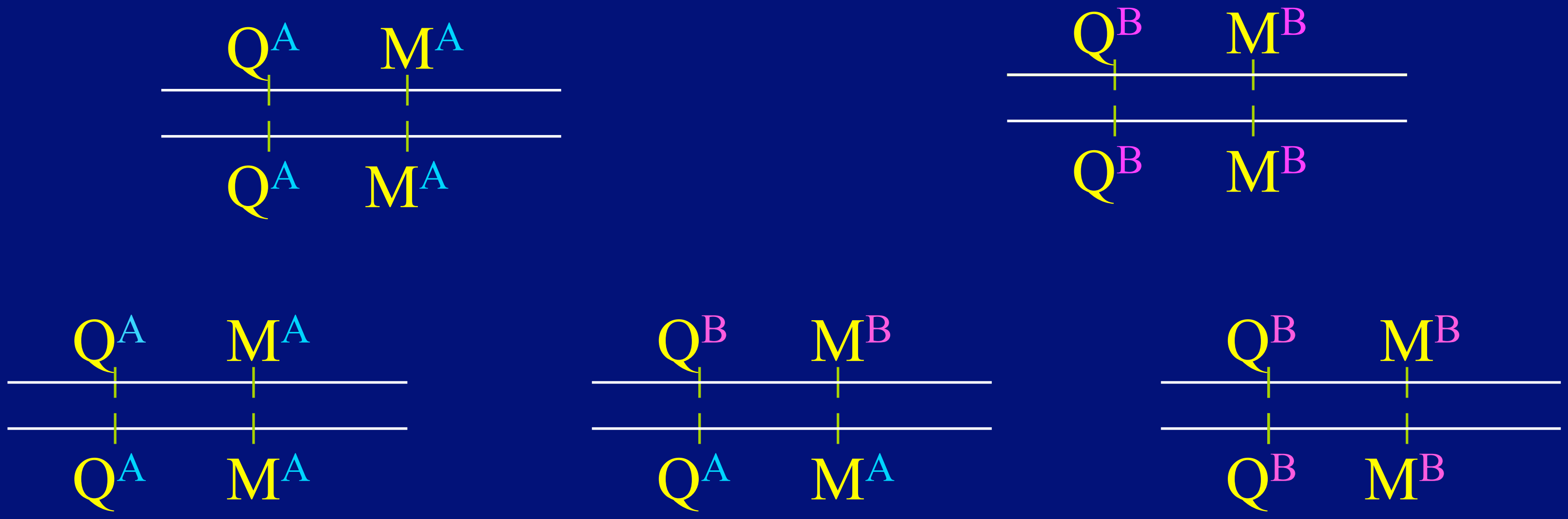




Detecting QTL in Selective Genotyping

Theoretical Concept - 1 Marker

Cultivar (normal pro AA) x **PI** (high pro BB)



Q^A M^A

Q^A M^A

Q^B M^B

Q^B M^B

Q^A M^A

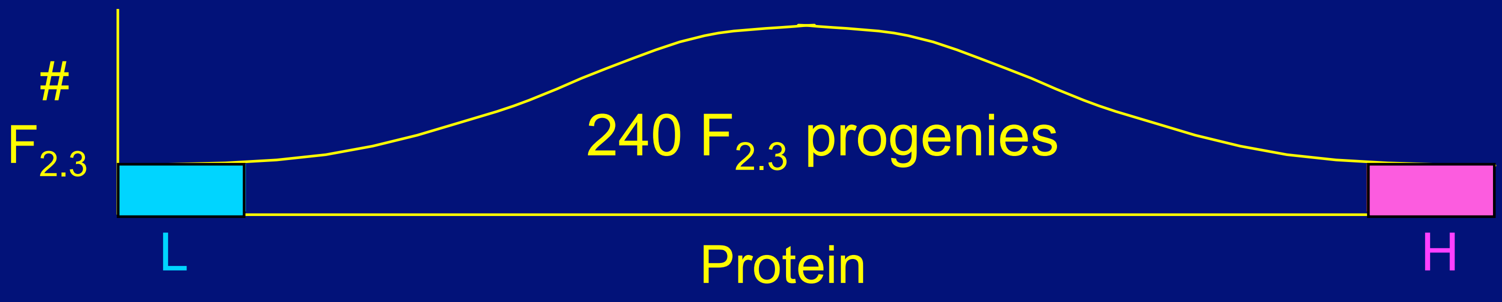
Q^A M^A

Q^B M^B

Q^A M^A

Q^B M^B

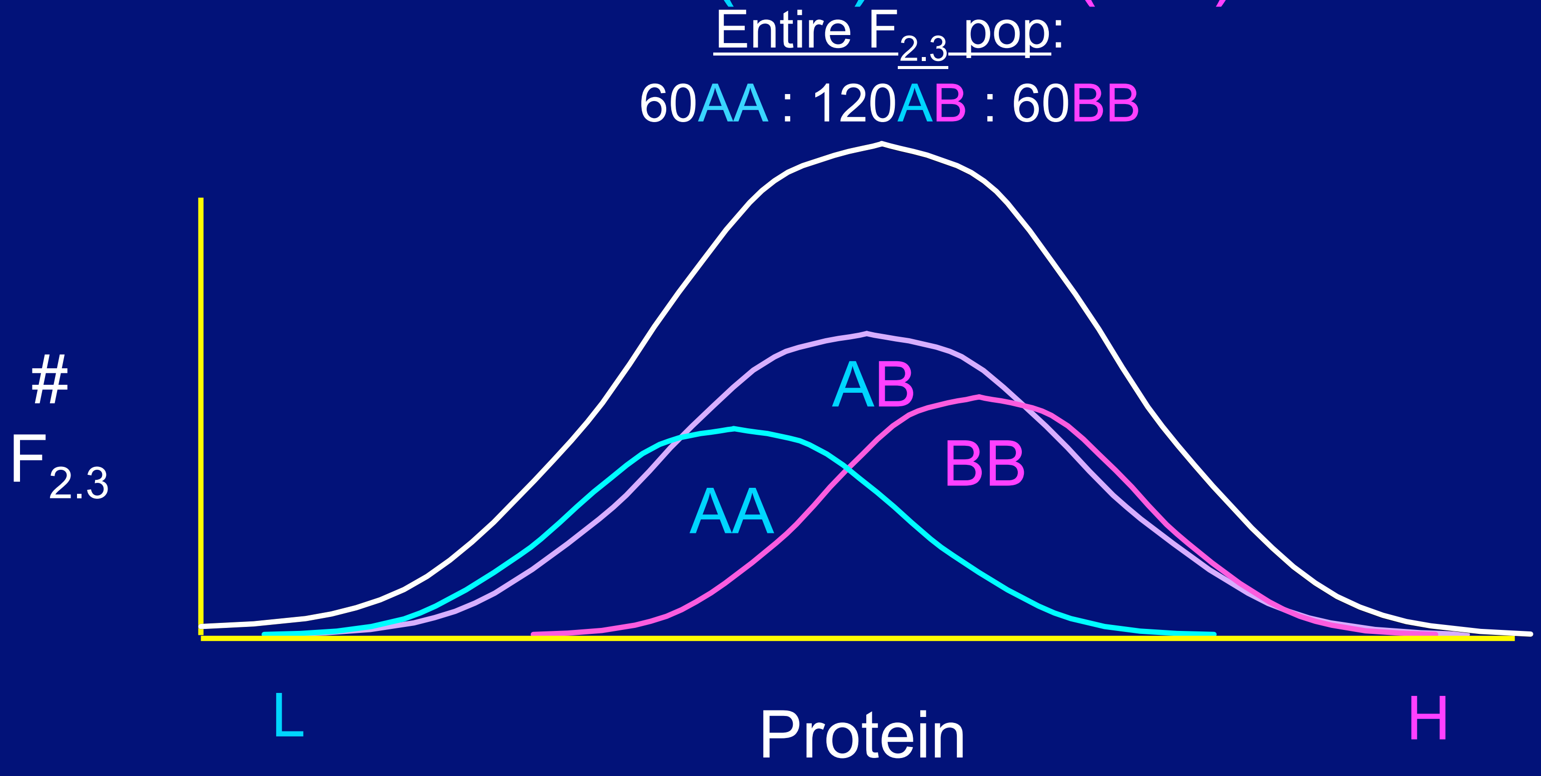
Q^B M^B



Assuming complete linkage of marker allele with protein QTL allele

$F_{2.3}$ Protein distribution by genotype of a linked Marker

Cultivar (AA) x PI (BB)

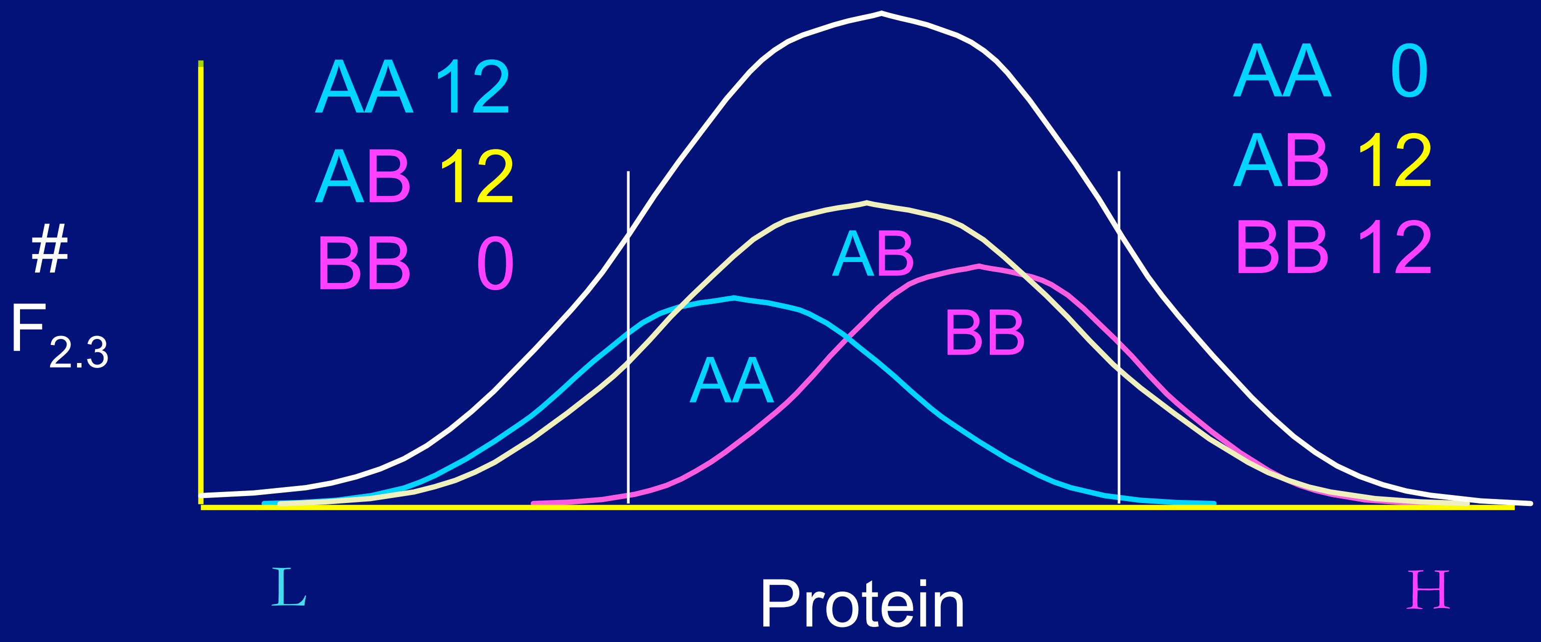


Marker genotypes in protein deciles

Cultivar (AA) x PI (BB)

Entire $F_{2.3}$ pop:

60AA : 120AB : 60BB



Chi-square test of observed vs. expected

$\alpha = 0.05$

Allele genotypes in protein deciles

Cultivar (AA) x PI (BB)

P_{Al} = 36/48

= 0.75

A 36

B 12

Entire $F_{2.3}$ pop:

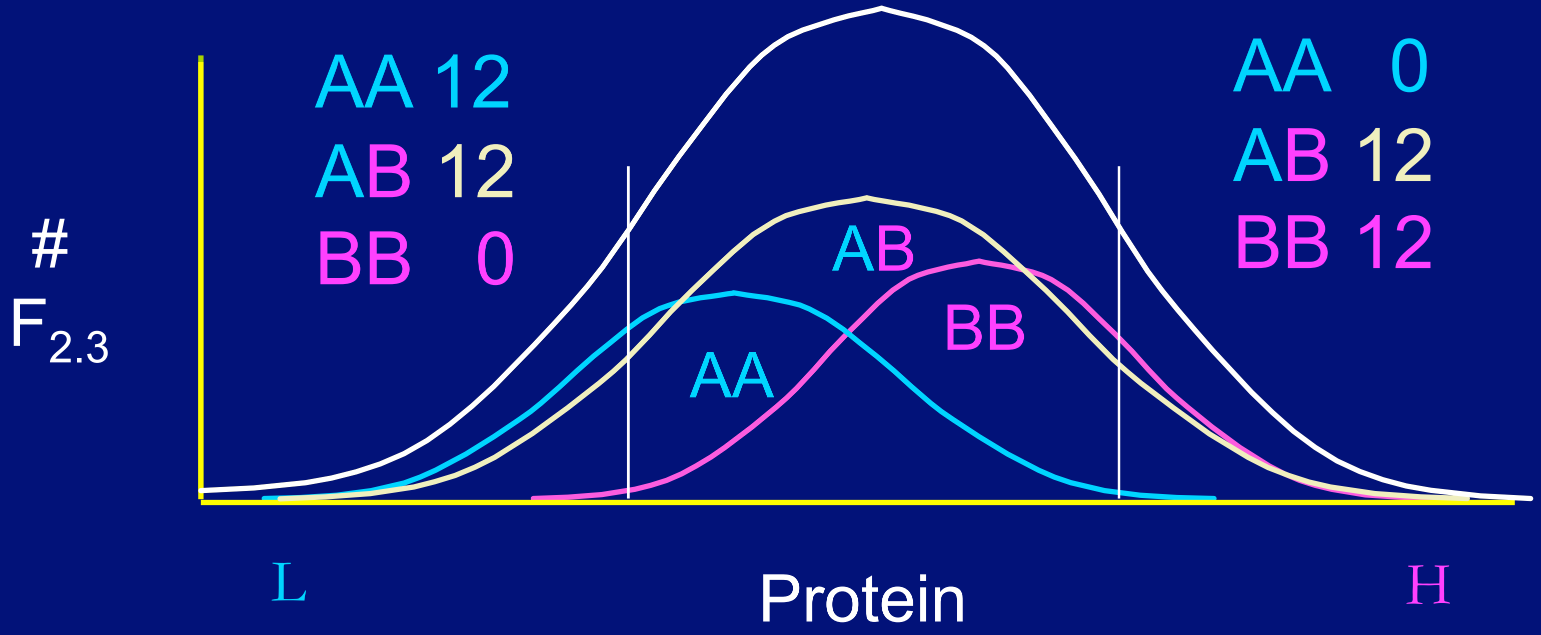
240A : 240B

A 12

B 36

P_{Ah} = 12/48

= 0.25



AA 12

AB 12

BB 0

AA 0

AB 12

BB 12

A two-sample t-test of SNP allele frequency

Detecting QTL in Selective Genotyping

If a marker is unlinked to a QTL for seed protein content, then the expected marker allele A frequencies the low quintile group (N_{low}) and the high quintile group (N_{high}) should have null hypothesis values of 0.5 and 0.5, respectively.

$$t = \frac{P_{Alow} - P_{Ahigh}}{\sqrt{\frac{p_{A0}(1-p_{A0})}{2N_{low}} + \frac{p_{A0}(1-p_{A0})}{2N_{high}}}}$$

P_{Alow} = an A allele frequency of low protein decile

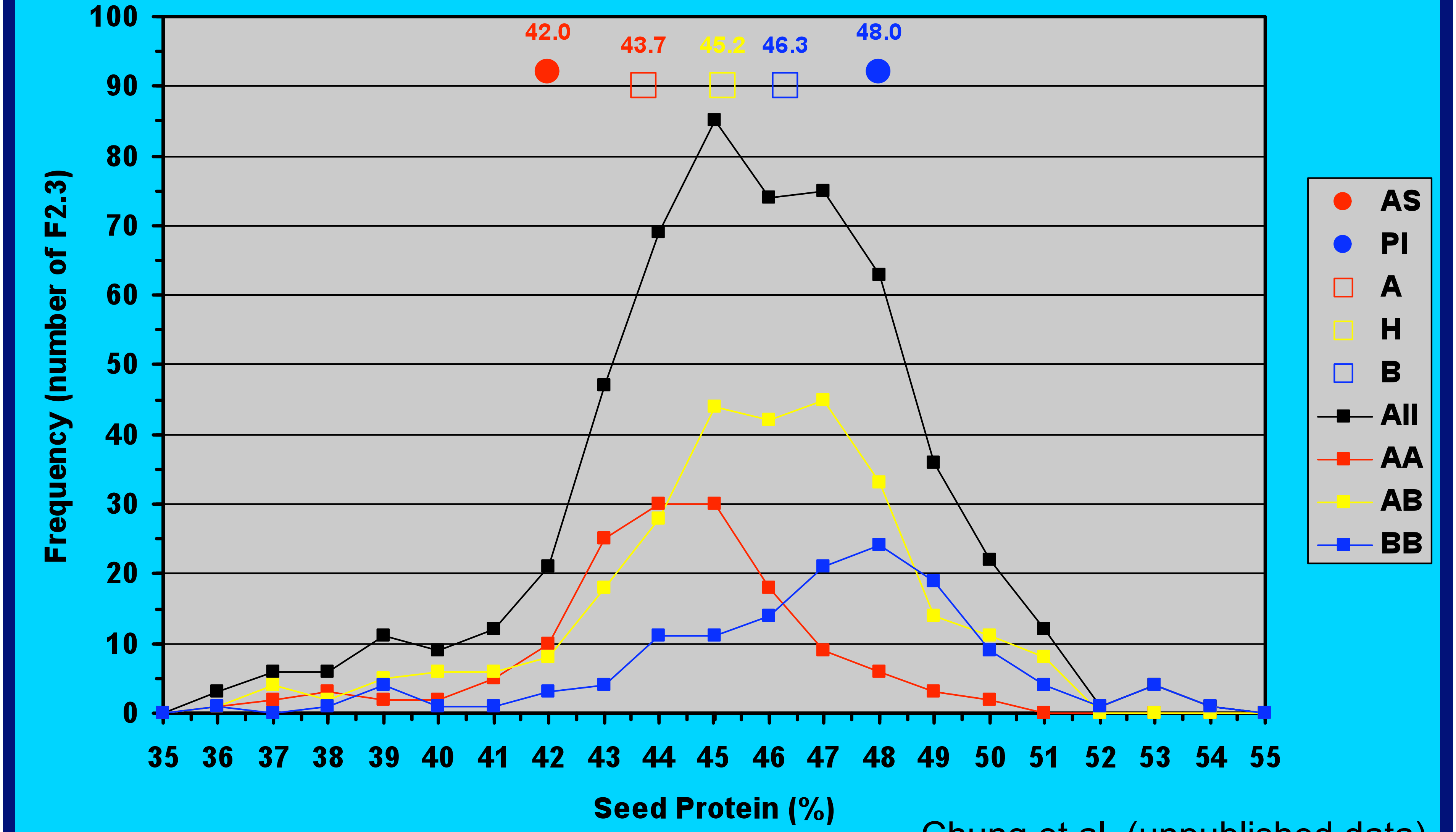
P_{Ahigh} = an A allele frequency of high protein decile

N_{low} = # of low protein $F_{2.3}$ progenies = 24

N_{high} = # of high protein $F_{2.3}$ progenies = 24

Bernardo (2002)

Seed Protein Distribution for 557 F2.3 Progenies of an F2 Plant Population derived
from a Low x High Protein Mating of Asgrow A3733 (42.0%) x PI 437.088A (48.0%)
(Note: The 1AA:2AB:1BB F2 Genotypes are those for the LG-I SSR Marker Satt496)



Chung et al. (unpublished data)

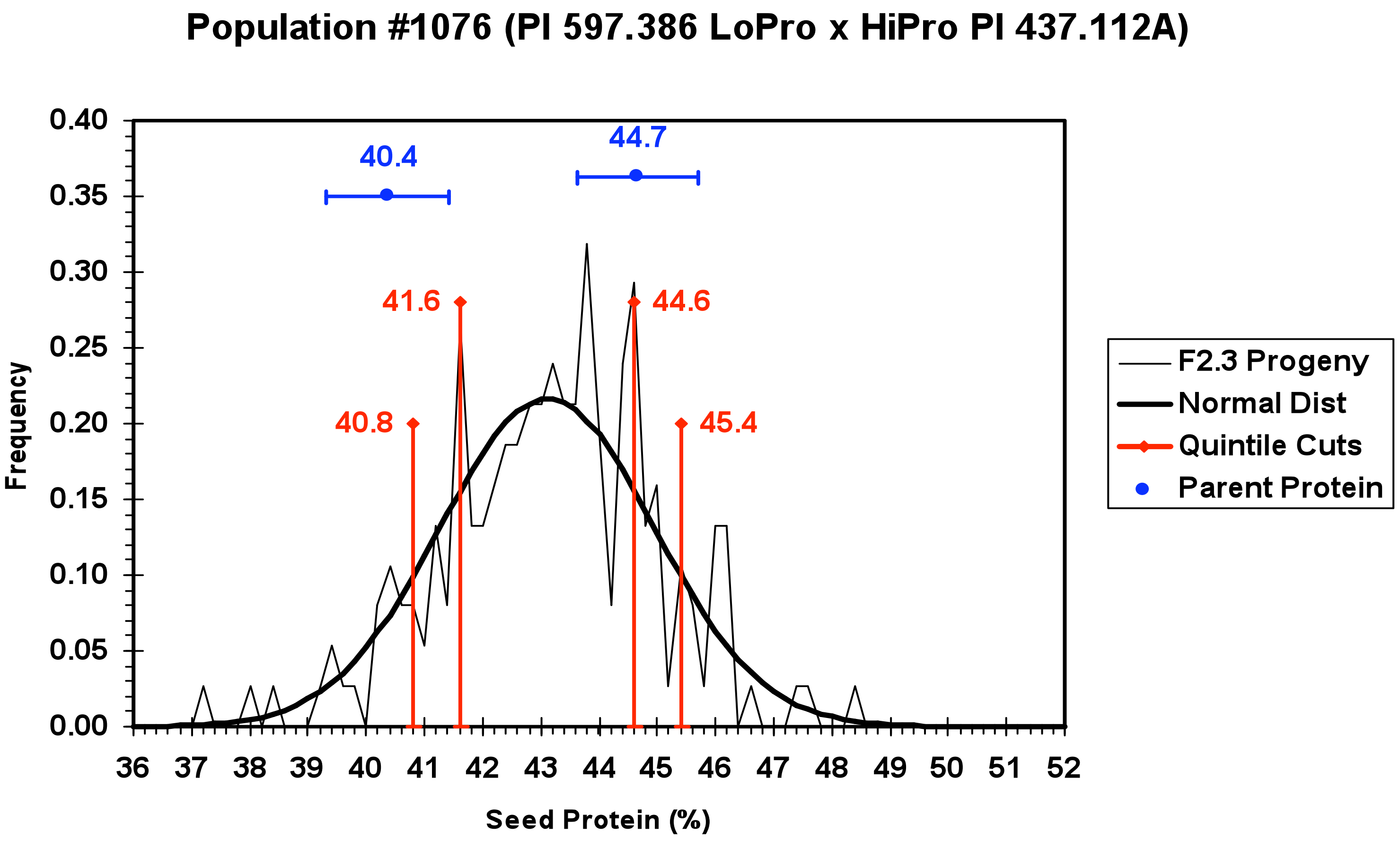
QTL Analysis

- QTL analysis with R/QTL software
  - Marker regression (MR)
  - Interval Mapping Analysis
    - Maximum Likelihood method using Expectation-Maximization (EM) algorithm
    - Multiple Imputation Method (IMP)
  - Stratified permutation used to derive a genome-wide significance criterion (alpha = 0.05)

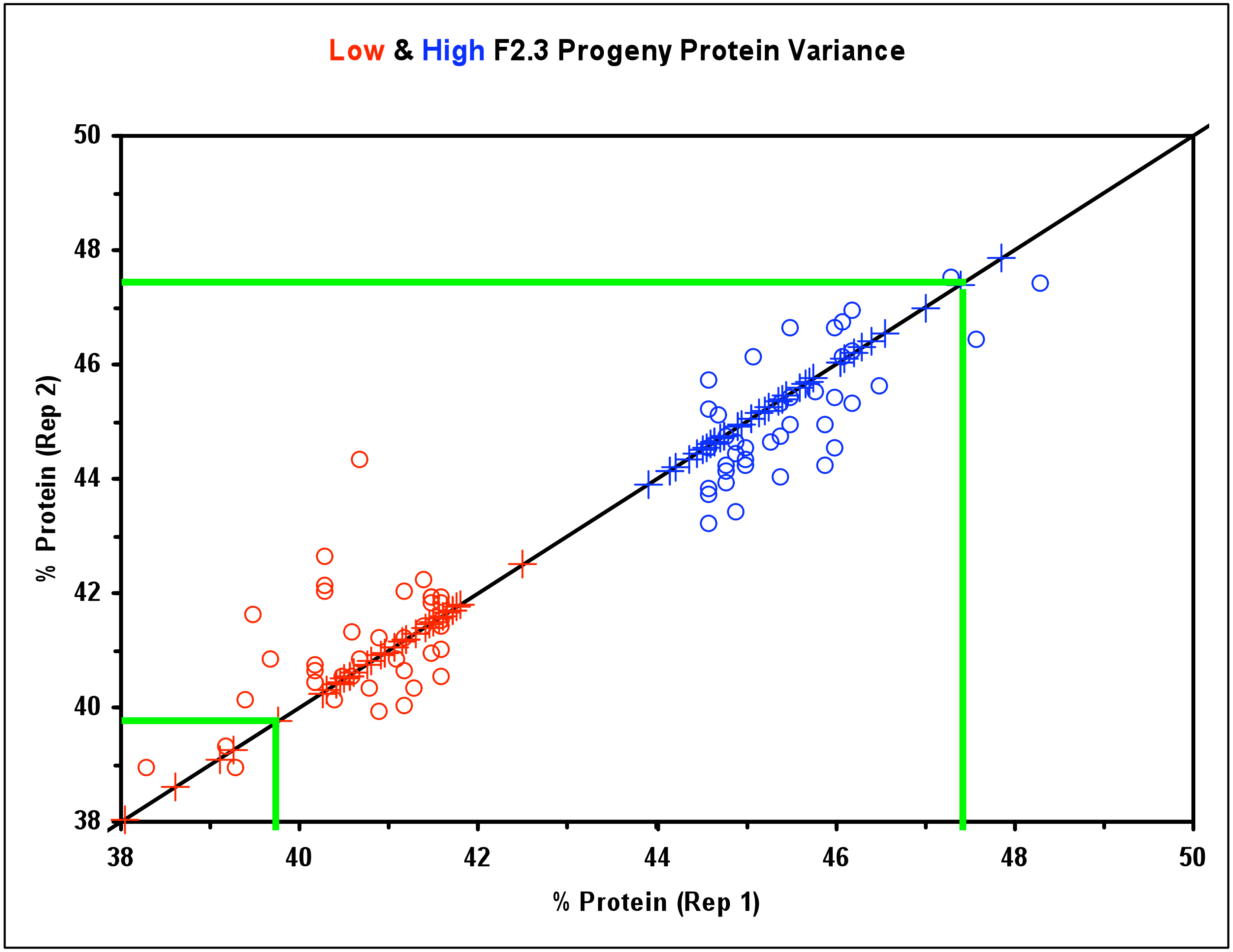
Results and Discussion

Phenotypic Distributions

- F_2 population from 1076 x Dwight mating
- Entire F_2 population NIR-phenotyped (rep 1) for F_3 seed protein content
- Lowest and highest quintiles re-NIR-phenotyped (rep 2)



Other Phenotypic Considerations



Comparison of replicate 1 & 2 NIR protein values in Population 1076

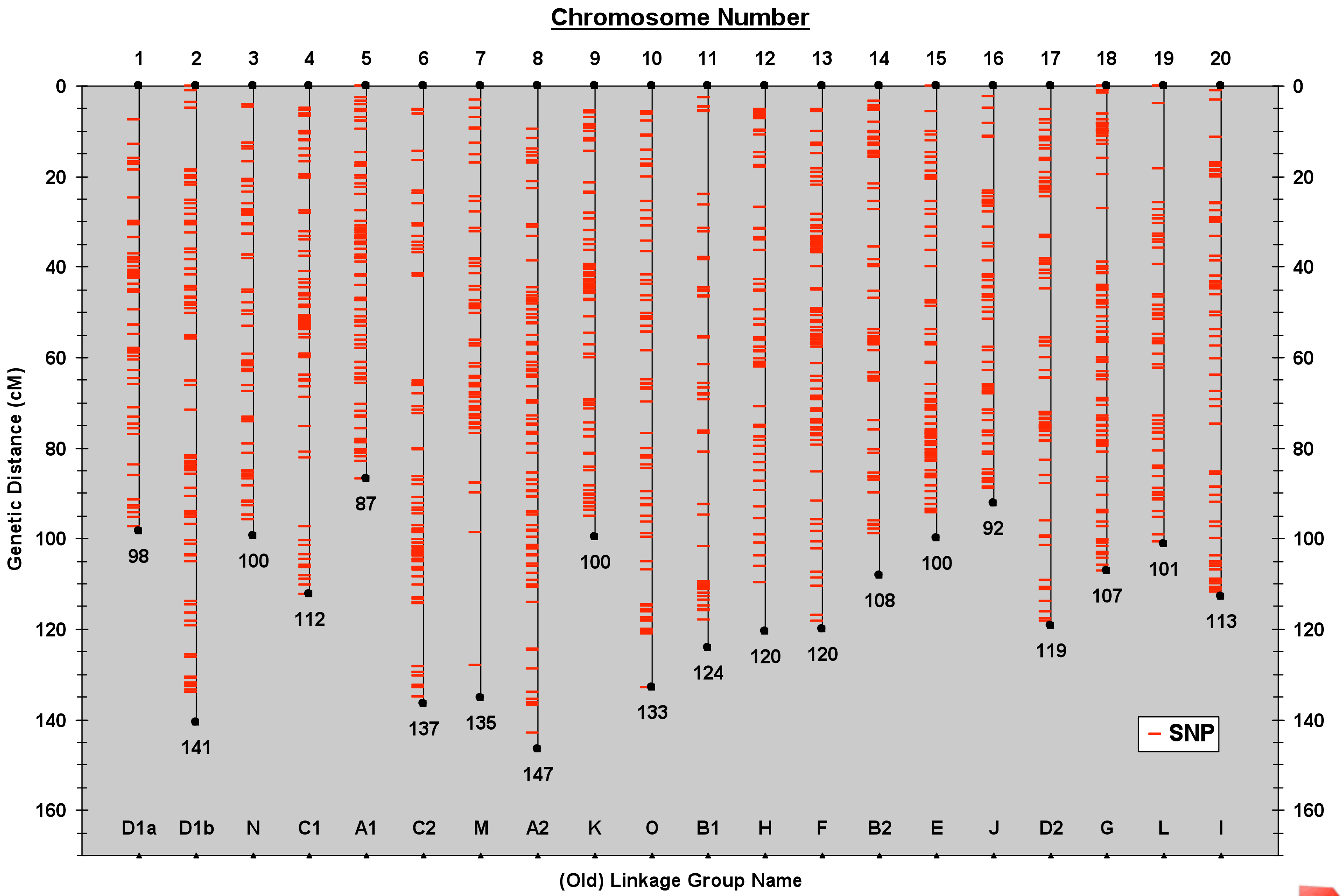
Genotypic Data

- Polymorphic markers in six populations

Population	Polymorphic marker	%
1076	497	39
1121	467	37
1122	425	33
1139	510	40
1143	472	37
1146	497	39

1272 SNP loci were bimorphic across all 20 populations.

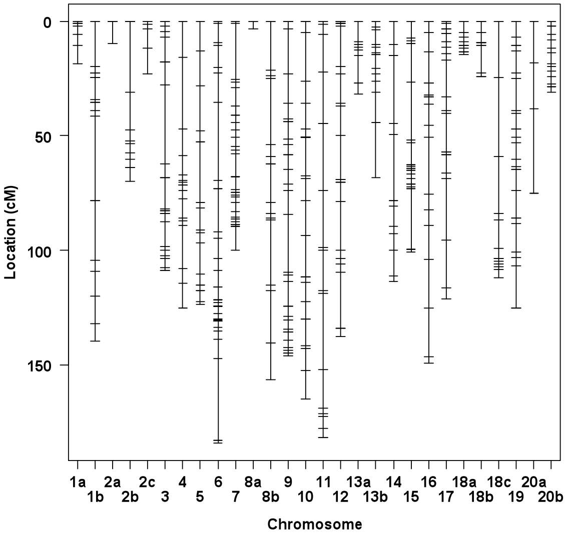
Soybean Genetic Map - USLP 1.0 - 1536 SNP Markers



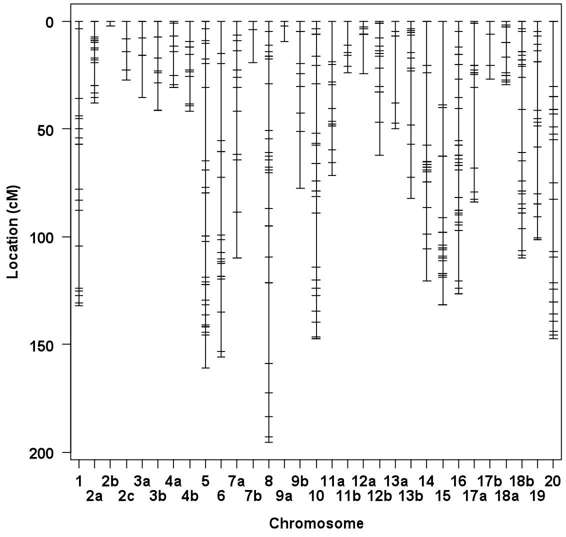
(Hyten et al. (2010); Crop Sci (in press)

Linkage Mapping

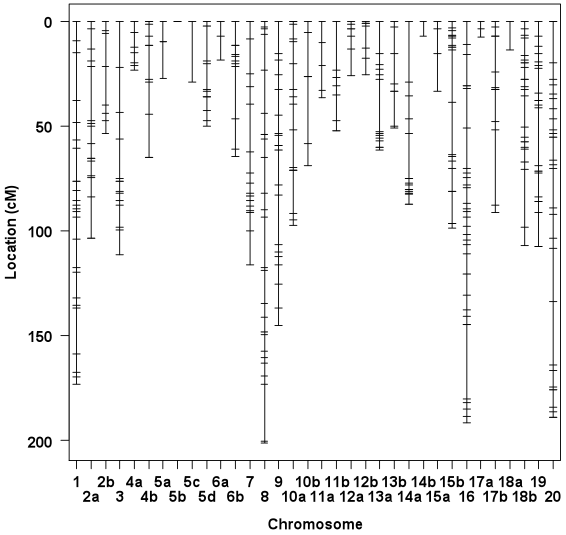
Genetic map
P1076



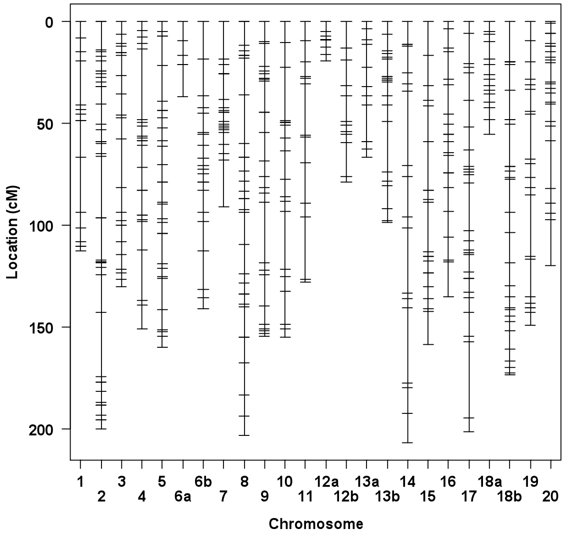
Genetic map
P1121



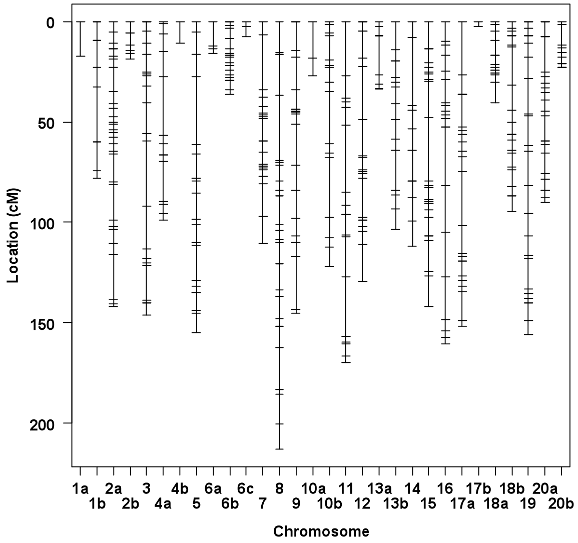
Genetic map
P1122



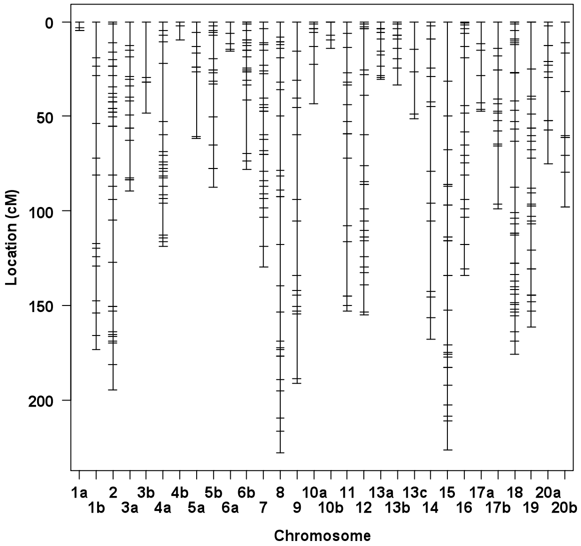
Genetic map
P1139



Genetic map
P1143

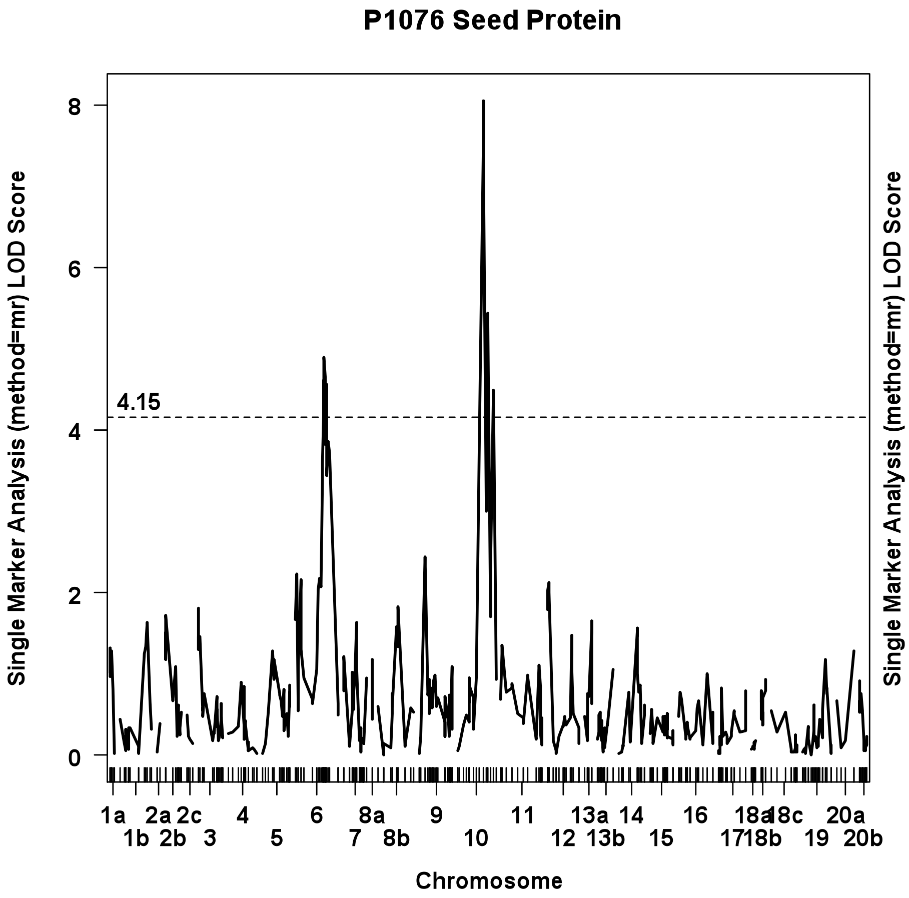


Genetic map
P1146

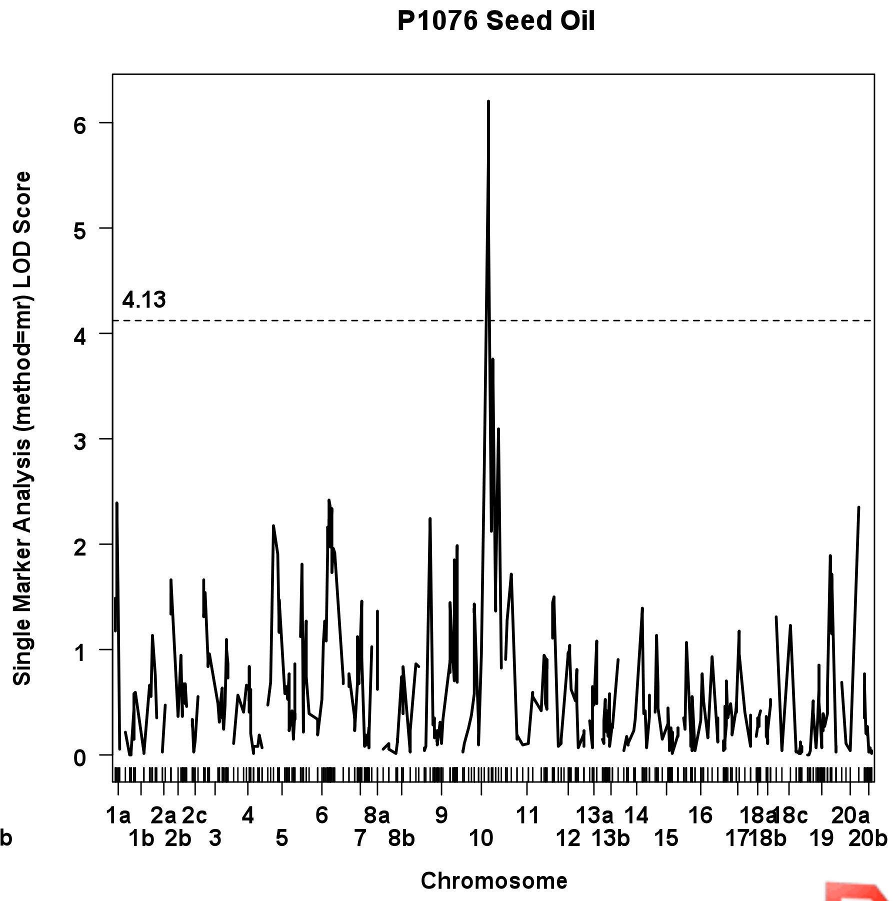


QTL Analysis

- Marker regression (MR)



Chr. 6 (C2); marker S30557; LOD = 4.89
Chr. 10 (O); marker S19004; LOD = 8.05

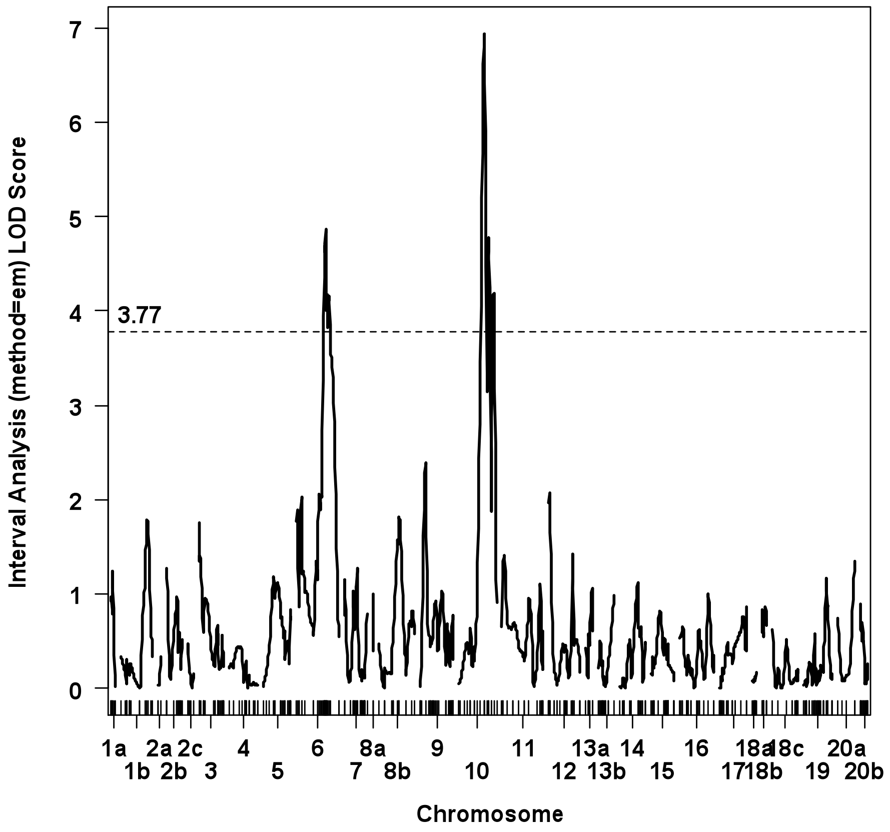


Chr. 10 (O); marker S19004; LOD = 6.21

QTL Analysis

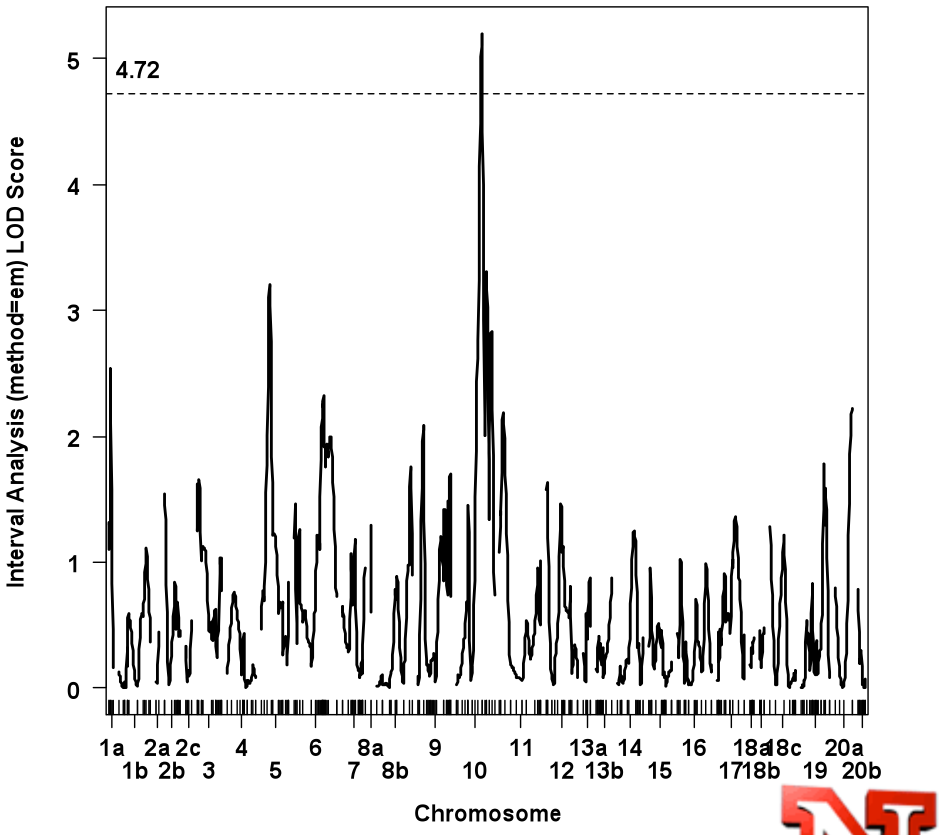
- Expectation-Maximization Algorithm (EM)

P1076 Seed Protein



Chr. 6 (C2); marker S12725; LOD = 4.86; a = -0.93; R^2 = 11
Chr. 10 (O); marker S19004; LOD = 6.94; a = 0.96; R^2 = 16

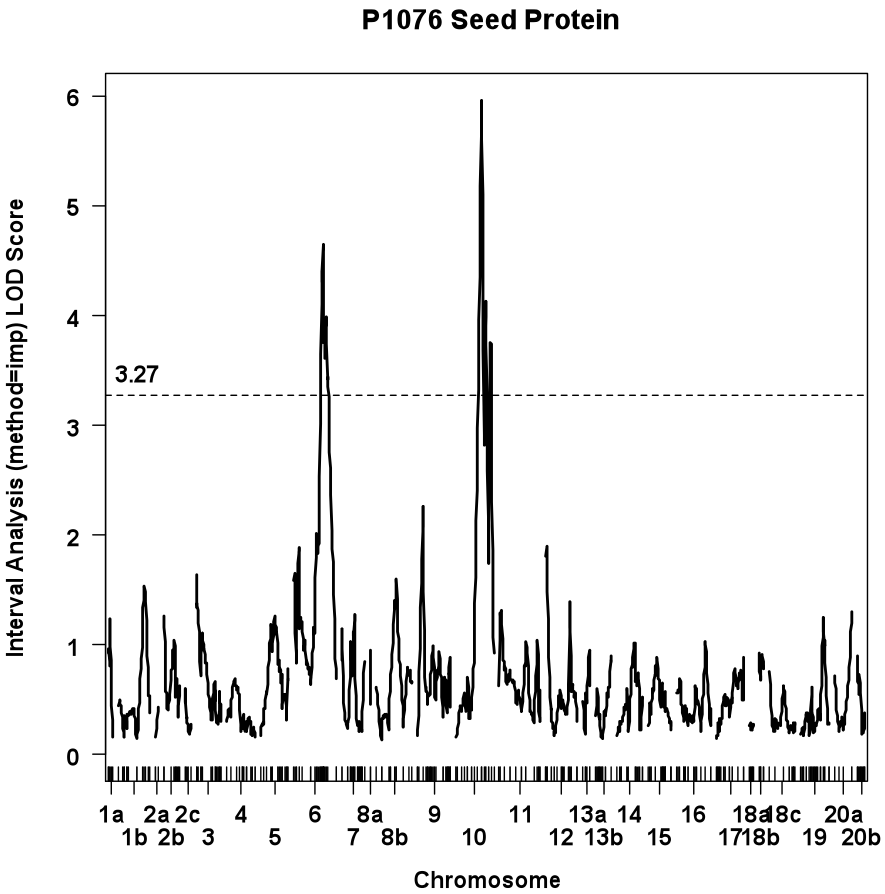
P1076 Seed Oil



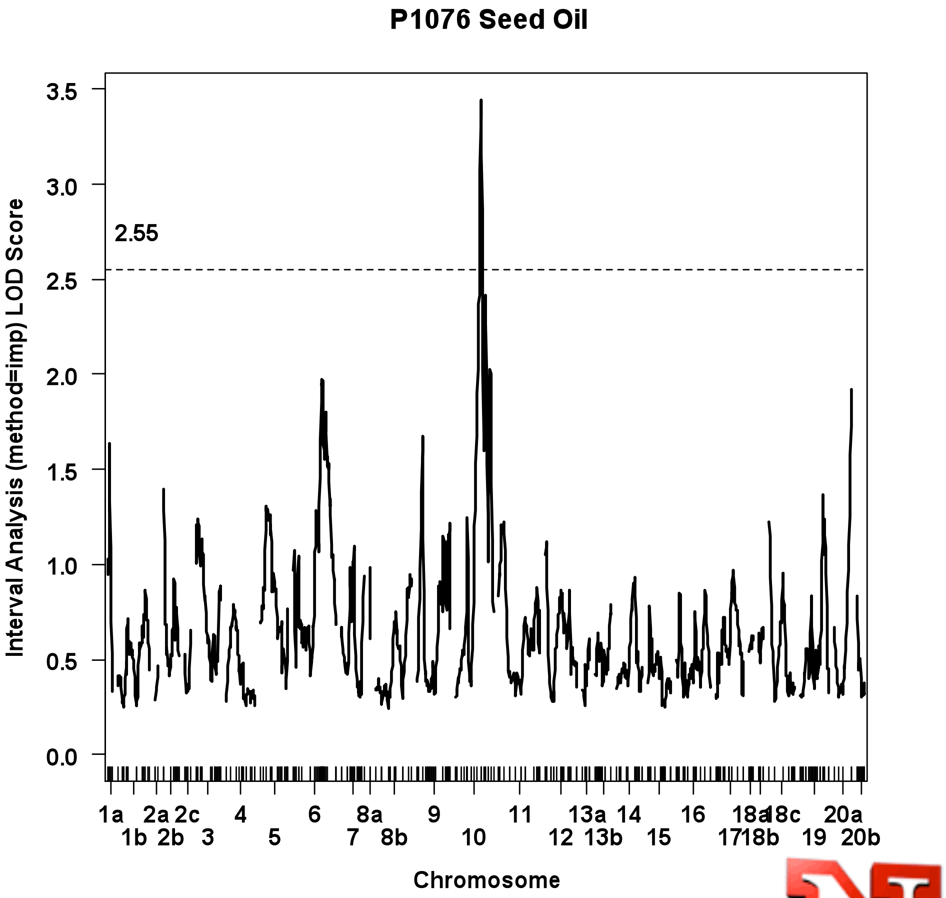
Chr. 10 (O); marker S19004; LOD = 6.94; a = 0.88; R^2 = 12

QTL Analysis

- Multiple Imputation Method (IMP)



P1076 Seed Protein



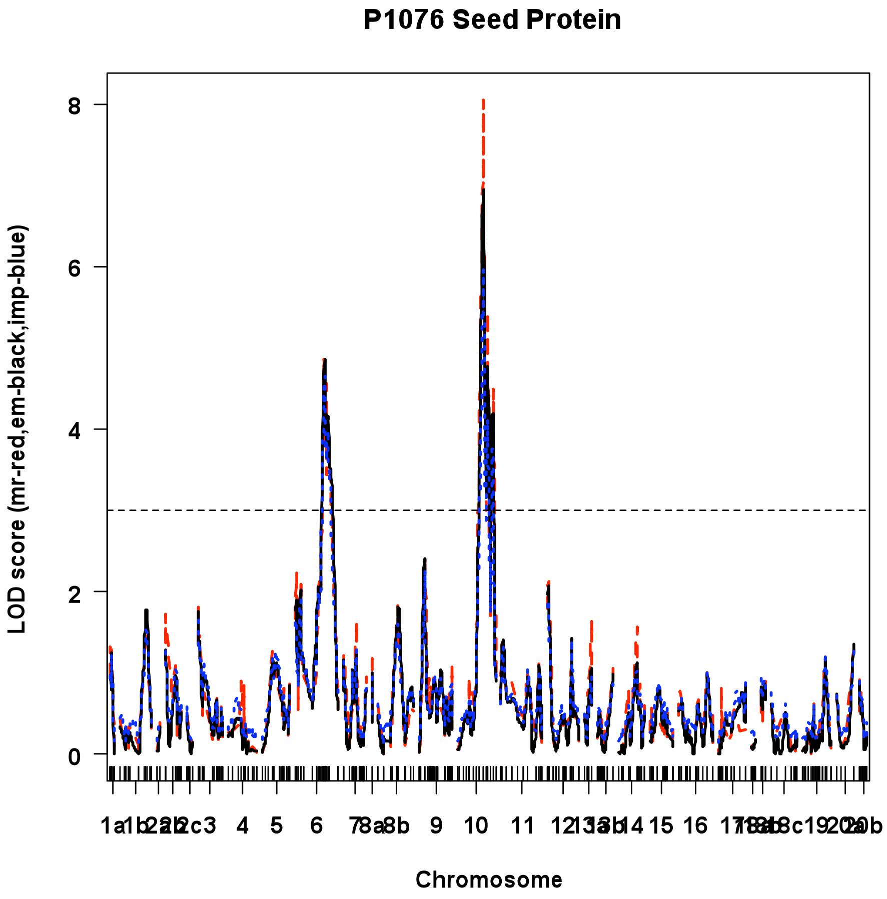
P1076 Seed Oil

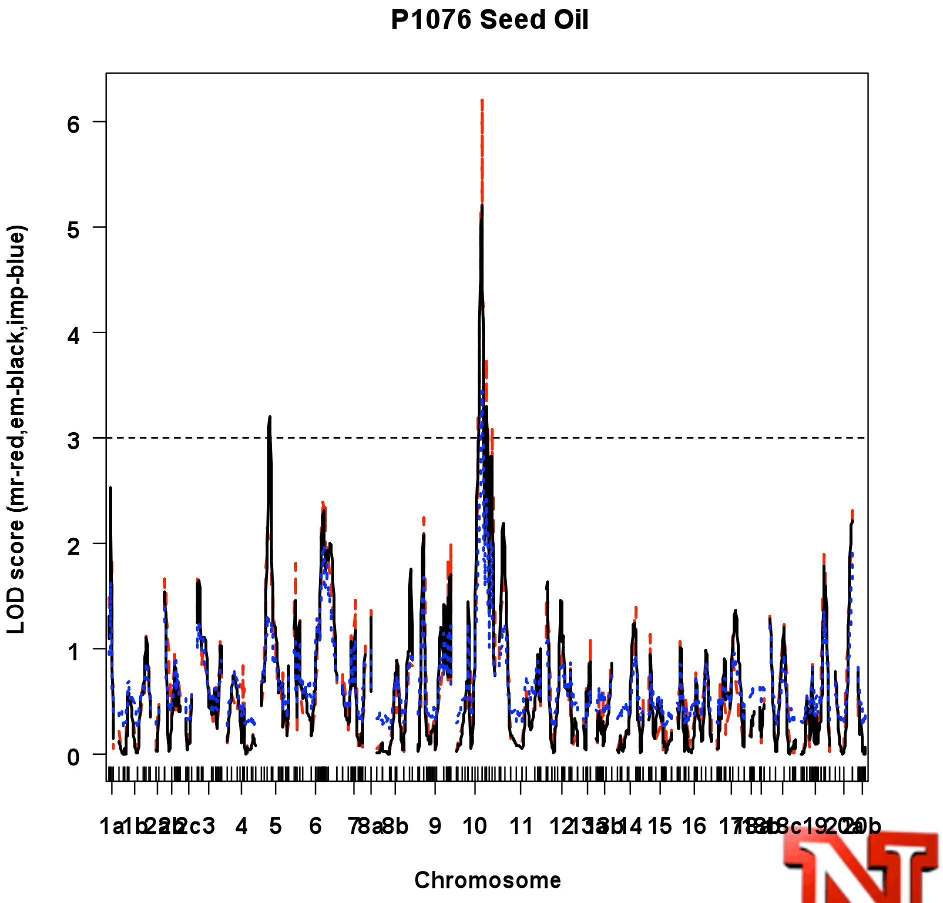
Chr. 6 (C2); marker S12725; LOD = 4.65; a = -0.93; R^2 = 11
Chr. 10 (O); marker S19004; LOD = 5.97; a = 0.96; R^2 = 14

Chr. 10 (O); marker S19004; LOD = 3.45; a = 0.88; R^2 = 8

QTL Analysis

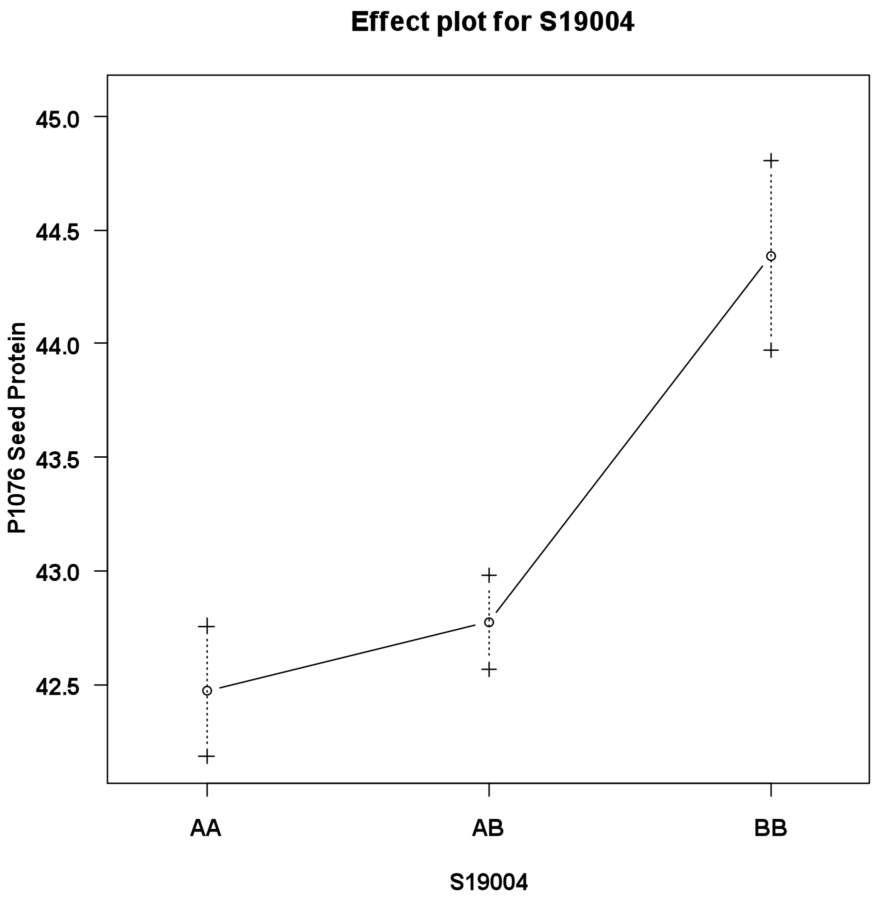
- Comparison of three methods

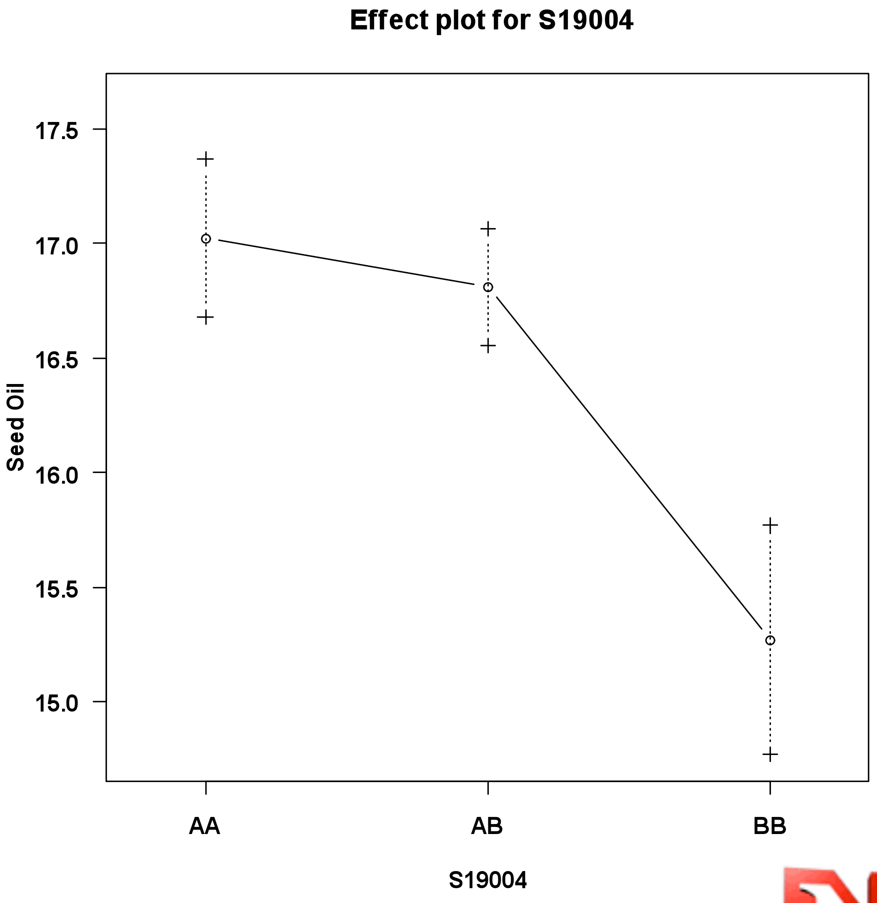




QTL Analysis

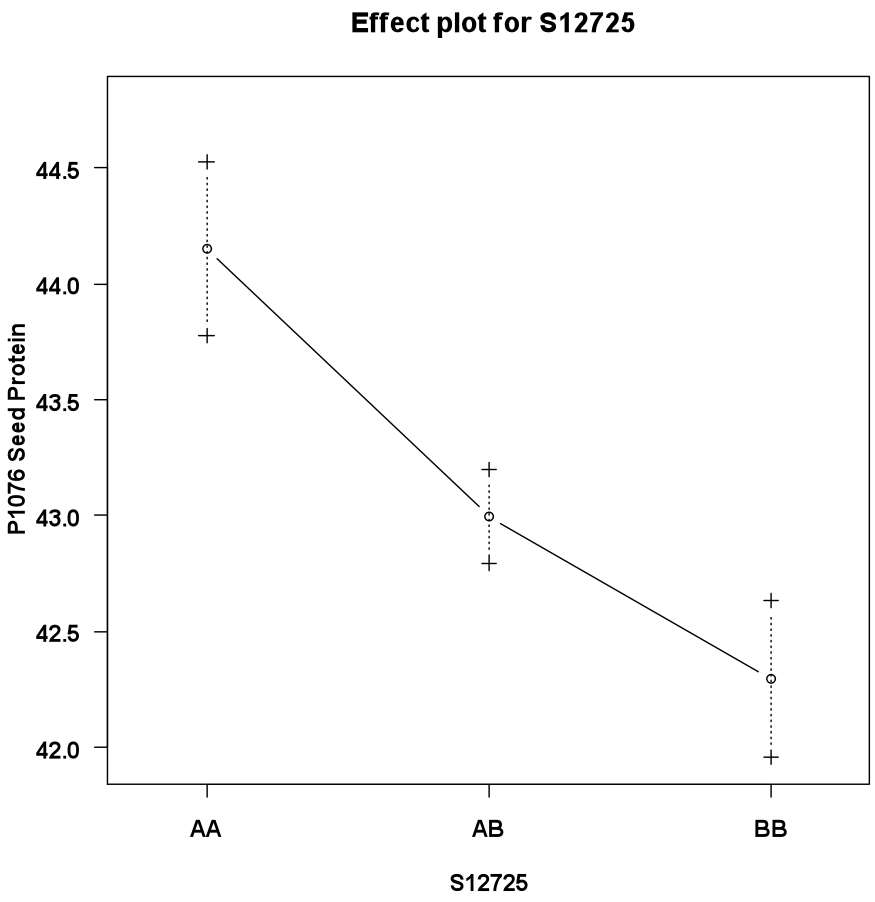
- Effect plot from EM method – Chr 10 (LG-O)

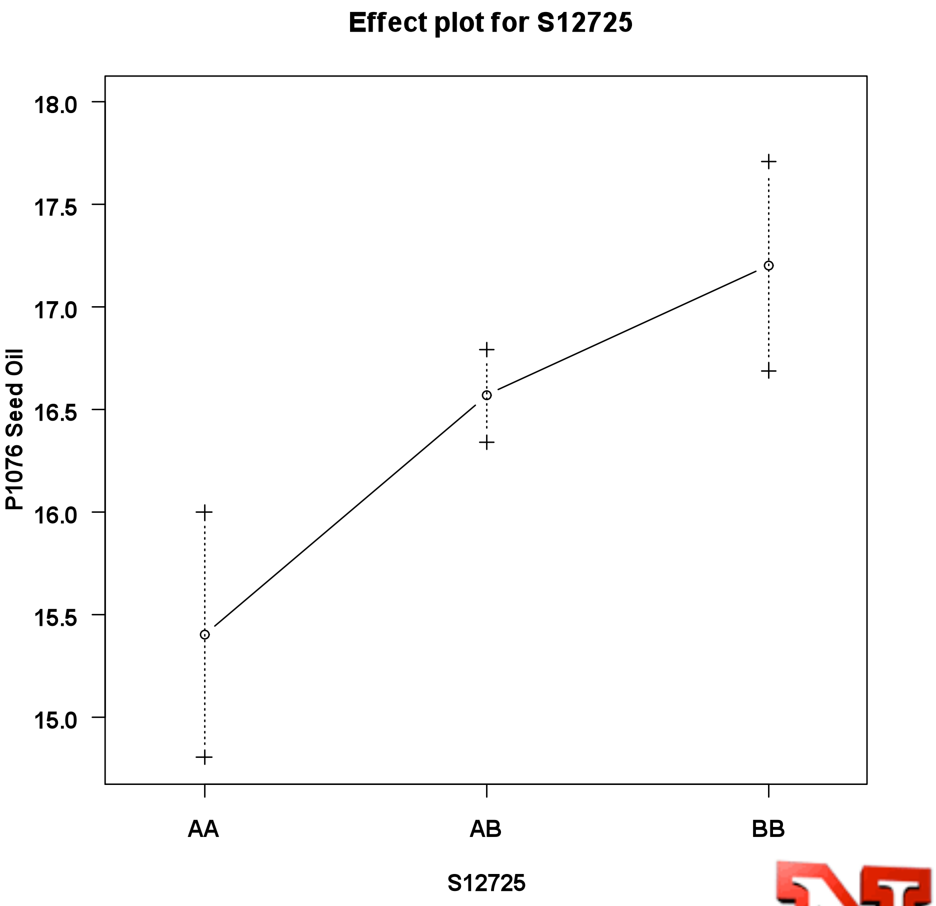

Effect plot for S19004


Effect plot for S19004

QTL Analysis

- Effect plot from EM method – Chr 6 (LG-C2)


Effect plot for S12725


Effect plot for S12725

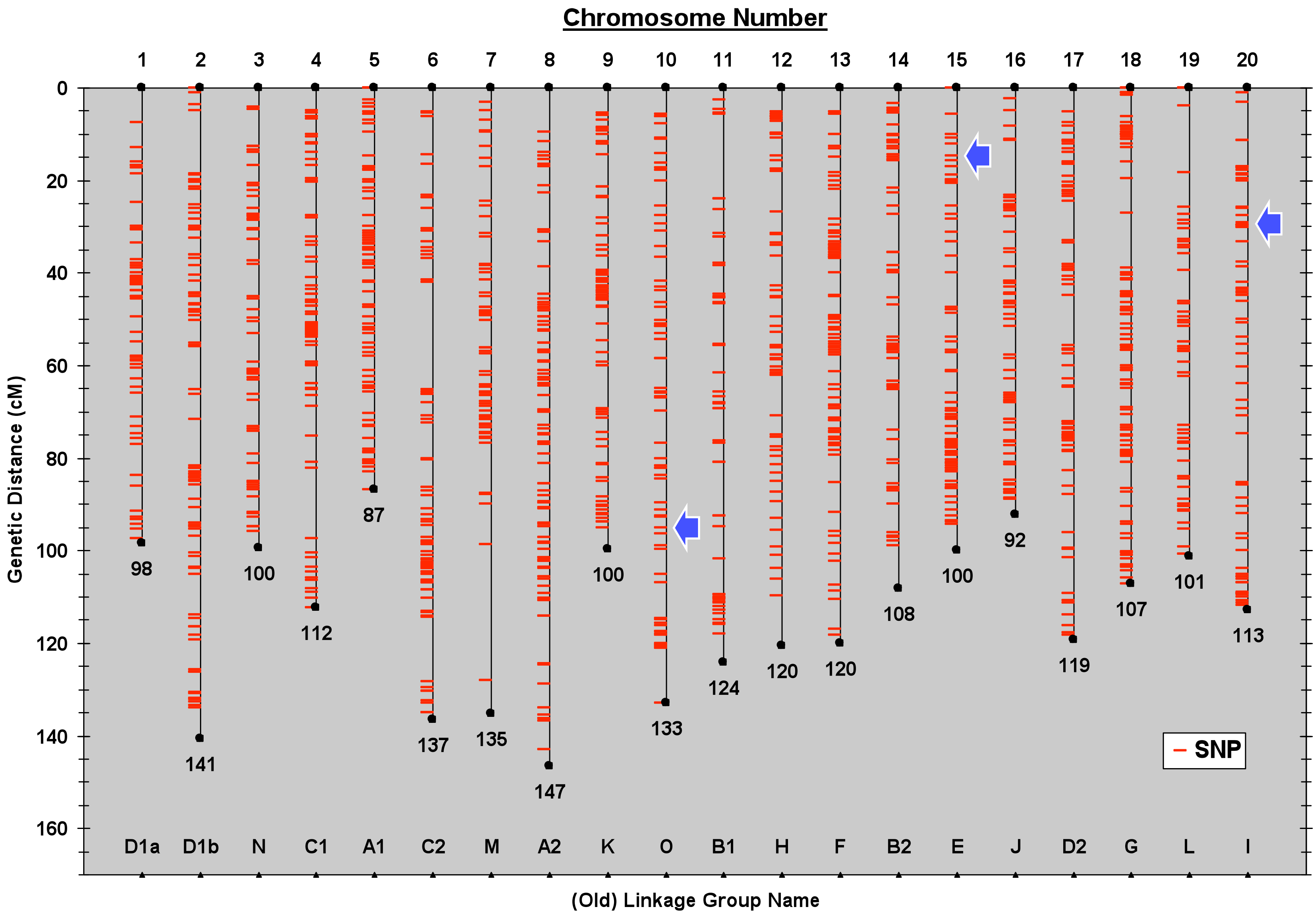
Conclusions

- The 22 lowest and 22 highest protein F2:3 progenies selected from ~240 total progeny in 48 of 52 populations (4 lost or discarded) were genotyped with 1536 SNPs distributed over the 20 chromosomes of the soybean genome.
- QTL analyses that have now been completed on 20 of the 48 populations. About 500 SNPs segregated in nearly every population.
- About 20 protein QTLs were detected in seven linkage groups in these 20 F2 populations.

Conclusions

- Over all 20 populations, statistically significant seed protein QTLs detected on Chr (LG-) in these pops:
  3 (LG-N) – 1140
  6 (LG-C2) – 1076, 1121, 1108
  10 (LG-O) – 1076, 1113, 1121, 1122, 1142
  14 (LG-B2) – 1146
  15 (LG-E) – 1140, 1143
  18 (LG-G) – 1108, 1121
  20 (LG-I) – 1024, 1025, 1110, 1113, 1138, 1139

Soybean Genetic Map - USLP 1.0 - 1536 SNP Markers



(Hyten et al. (2010); Crop Sci (in press)

Thanks for your attention!

Questions?