

# Protein and Oil QTL and their Associated Impact on Agronomic Traits

Brian Diers, Lillian Brzostowski, Russell  
Ward, Tim Pruski, Myungsik Kim,  
Christina Fliege, and Matt Hudson

# Outline

- History and background
- cqSeed protein-003 on chr 20 (LG I)
  - Fine mapping
  - Impact of protein and oil and associated effects on agronomic traits
- cqSeed protein-001 on chr 15 (LG E)
  - Fine mapping
  - Impact of protein and oil and associated effects on agronomic traits
- Implications on breeding
- Conclusions

## RFLP analysis of soybean seed protein and oil content

B. W. Diers<sup>1</sup>, P. Keim<sup>2</sup>, W. R. Fehr<sup>1</sup>, and R. C. Shoemaker<sup>3,4,\*</sup>

<sup>1</sup> Department of Agronomy, Iowa State University, Ames, IA 50011, USA

<sup>2</sup> Department of Biology, Northern Arizona University, Flagstaff, AZ 86011, USA

<sup>3</sup> Departments of Agronomy and Genetics, Iowa State University, Ames, IA 50011, USA

<sup>4</sup> USDA-Agricultural Research Service, Field Crops Research Unit, Iowa State University, Ames, IA 50011, USA

Received May 25, 1990; Accepted July 9, 1991

Communicated by A. L. Kahler

**Summary.** The objectives of this study were to present an expanded soybean RFLP map and to identify quantitative trait loci (QTL) in soybean [*Glycine max* (L.) Merr.] for seed protein and oil content. The study population was formed from a cross between a *G. max* experimental line (A81-356022) and a *G. soja* Sieb. and Zucc. plant introduction (PI 468916). A total of 252 markers was mapped in the population, forming 31 linkage groups. Protein and oil content were measured on seed harvested from a replicated trial of 60 F<sub>2</sub>-derived lines in the F<sub>3</sub> generation (F<sub>2:3</sub> lines). Each F<sub>2:3</sub> line was genotyped with 243 RFLP, five isozyme, one storage protein, and three morphological markers. Significant ( $P < 0.01$ ) associations were found between the segregation of markers and seed protein and oil content. Segregation of individual markers explained up to 43% of the total variation for specific traits. All *G. max* alleles at significant loci for oil content were associated with greater oil content than *G. soja* alleles. All *G. soja* alleles at significant loci for protein content were associated with greater protein content than *G. max* alleles.

crops will continue to make increased seed protein and oil major breeding objectives in the future. Both protein and oil content are quantitatively inherited in soybean (Burton 1985; Wilcox 1985). Breeders have been successful in manipulating these traits, but their underlying genetic controls have not been elucidated.

Genetic markers have allowed researchers to systematically map and characterize genes that are important in conferring quantitative traits. These genes have been mapped to what has become known as quantitative trait loci (QTL). The use of restriction fragment length polymorphism (RFLP) markers has increased the efficiency of mapping QTLs, because of the greater number of markers that can be scored in a single population relative to other markers used such as isozyme or morphological markers. Genetic mapping of QTL has been documented in maize and tomato (Edwards et al. 1987; Osborn et al. 1987; Paterson et al. 1988). QTL mapping has led to an increased understanding of genes involved in the inheritance of quantitative traits, and may improve genetic gains in breeding programs through marker-assisted se-

**Table 1.** Markers significantly ( $P < 0.01$ ) associated with variation for protein and oil content

Marker	$R^2$	$P > F$	Means of genotypic classes <sup>a</sup>			Linkage group
			MM	SM	SS	
g (kg seed) <sup>-1</sup>						
<i>Protein</i>						
pK-11a	0.42	0.0001	450	465	474	K
pA-407a	0.39	0.0001	451	464	473	K
pA-144	0.24	0.0007	454	464	472	K
pA-688	0.25	0.001	454	465	472	K
pSAC-7a <sup>b</sup>	0.24	0.003	455	464	472	A
pA-242b	0.19	0.004	456	465	468	A
pA-23	0.16	0.01	457	466	468	A
pA-245a <sup>c</sup>	0.12	0.01	455		465	C
<i>Total oil</i>						
pSAC-7a	0.43	0.0001	153	145	136	A
pA-242b	0.39	0.0001	154	143	140	A
pA-23	0.32	0.0001	153	144	139	A
pb <sup>d</sup>	0.27	0.0001	152	144	139	A
pK-11A	0.27	0.0002	155	144	140	K
pA-407a	0.28	0.0005	154	144	140	K
pA-454	0.23	0.0008	152	143	142	A
pK-229	0.22	0.001	151	144	140	A
pA-203	0.18	0.006	152	145	142	A

<sup>a</sup> MM designates homozygous *G. max* class, SM heterozygous class, and SS homozygous *G. soja*

<sup>b</sup> Actin gene probe provided by Dr. R. Meagher (University of Georgia)

<sup>c</sup> *G. soja* alleles dominant to *G. max* alleles

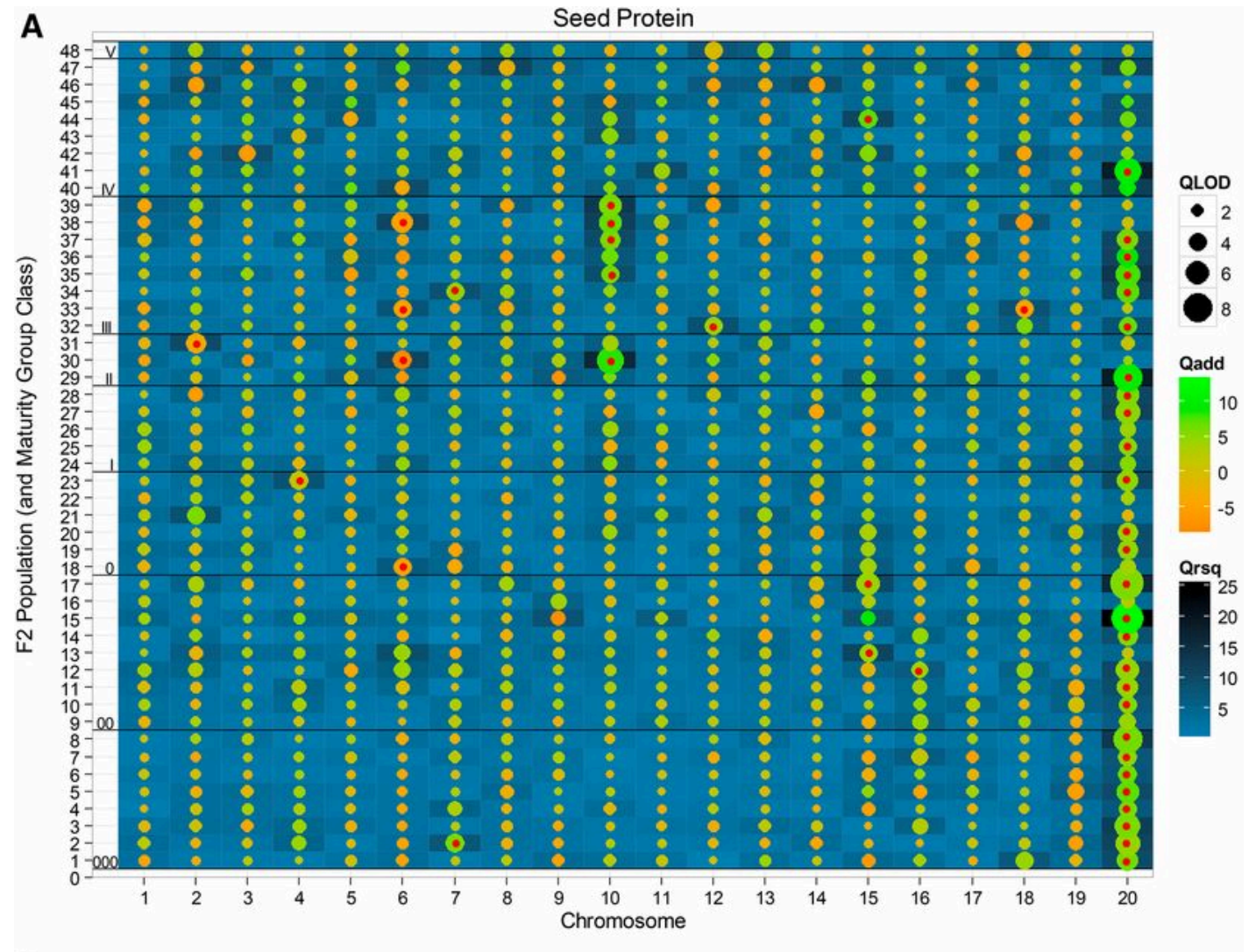
<sup>d</sup> Morphological marker blunt-sharp pubescence tip (Palmer and Kilen 1987)

LG K = chr  
20 (LG I)

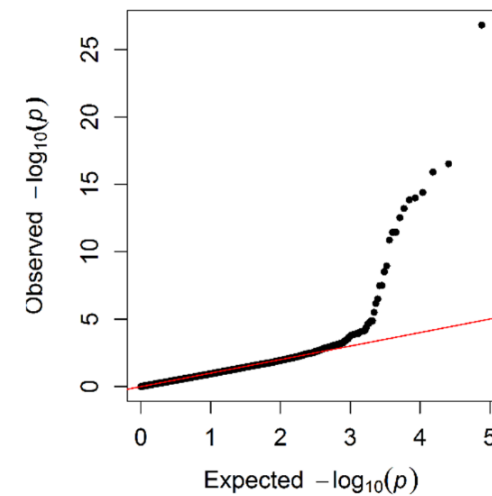
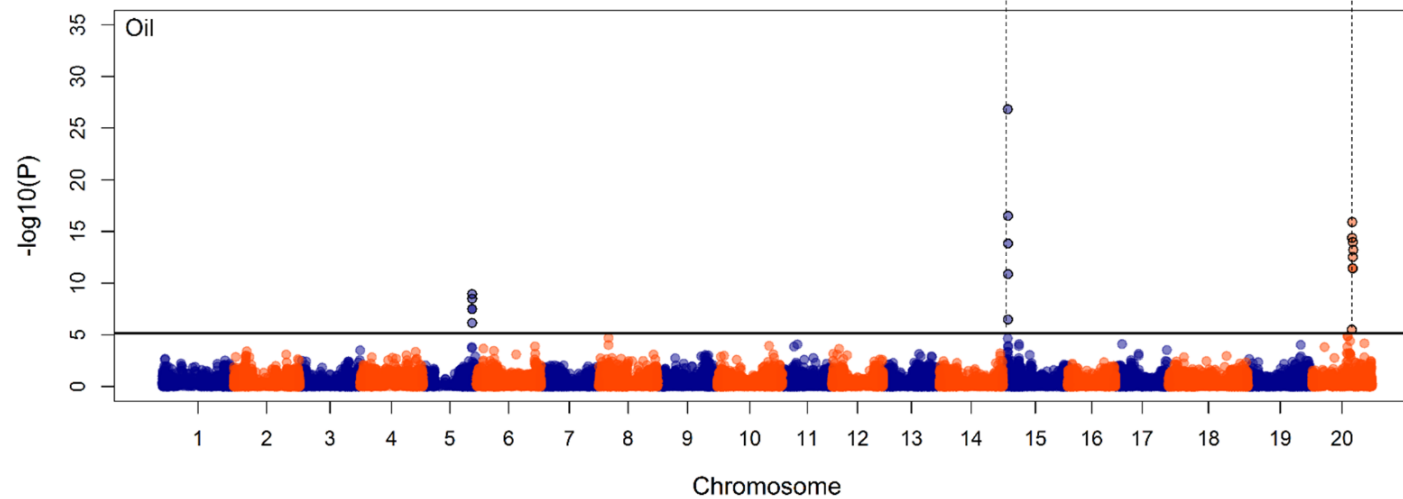
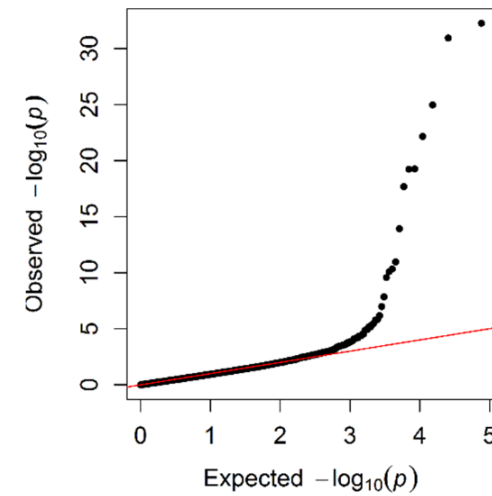
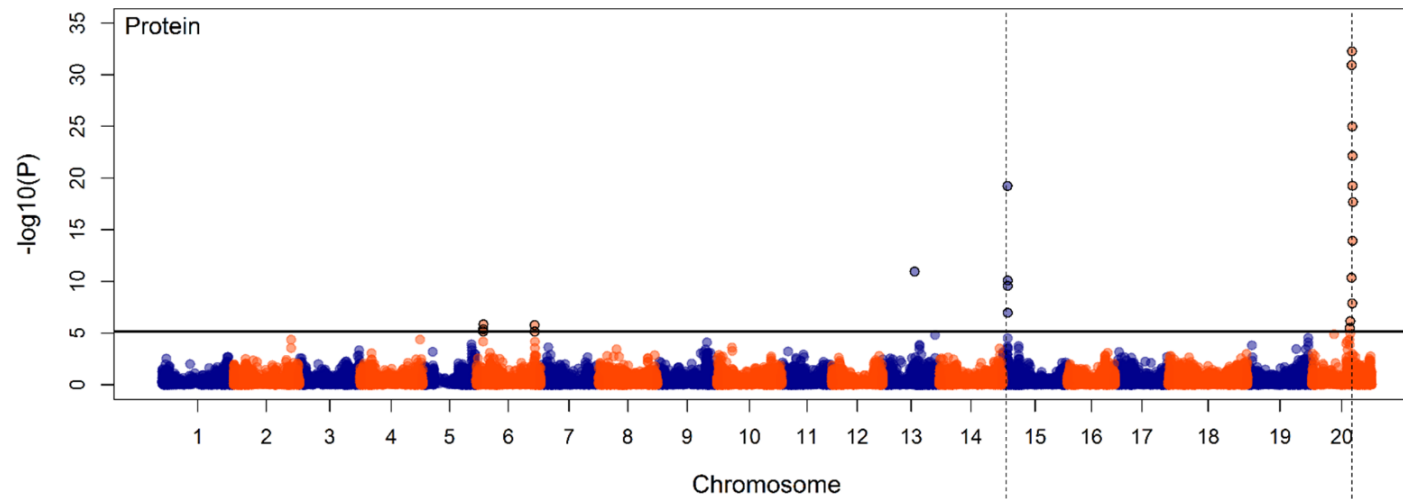
LG A = chr 15  
(LG E)

# QTL for Protein (Phansak et al. 2016)

- Selective genotyping was done in 48 populations
- Significant protein or oil QTL mapped in 35 matings
- Chromosome 20 significant for 27 matings and chr 15 for 3 matings.



# QTL for Protein and Oil Identified through Association Mapping (Bandillo et al., 2015)



# cqSeed Protein-003 (chr 20)

- This is the largest effect protein QTL.
- Confirmed and the high protein allele show to be associated with reduced seed oil concentration, less yield, smaller seed, earlier maturity, and earlier plants (Sebolt et al., 2000; Nichols et al., 2006).
- Many sources of the high protein QTL allele.
  - Our source has been *G. soja* PI 468916.
  - Danbaekkong has a protein QTL allele in the same region that was found to have less impact on agronomic traits in the southern USA.

# cqSeed Protein-003 (chr 20) map position

Nichols et al., 2006

Bolon et al., 2010

838

CROP SCIENCE, VOL. 46, MARCH-APRIL 2006

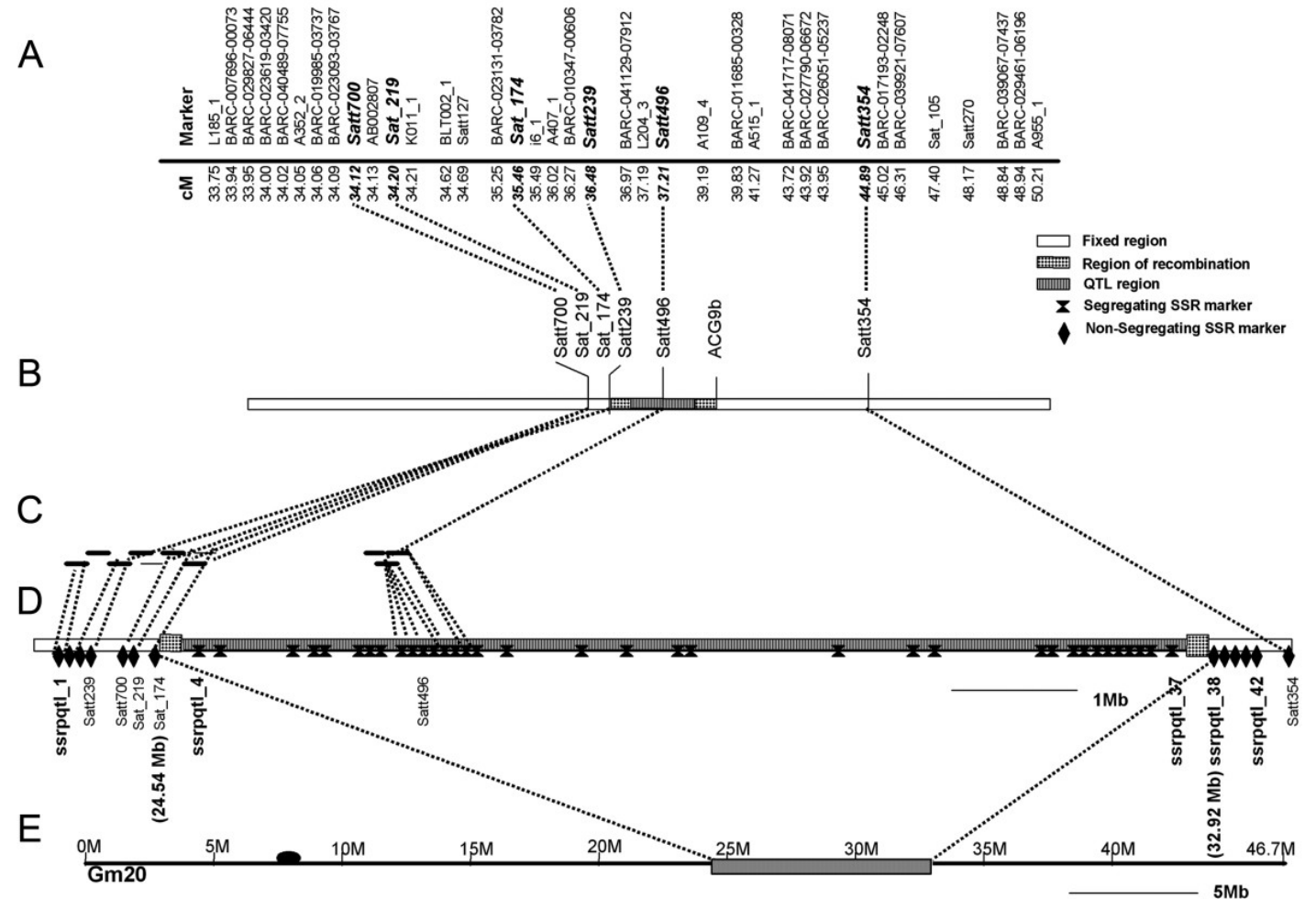
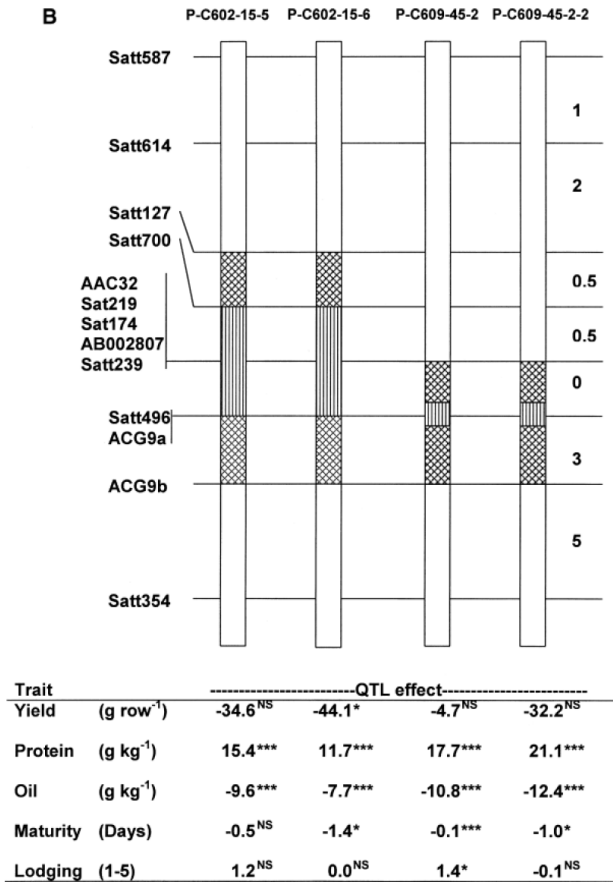
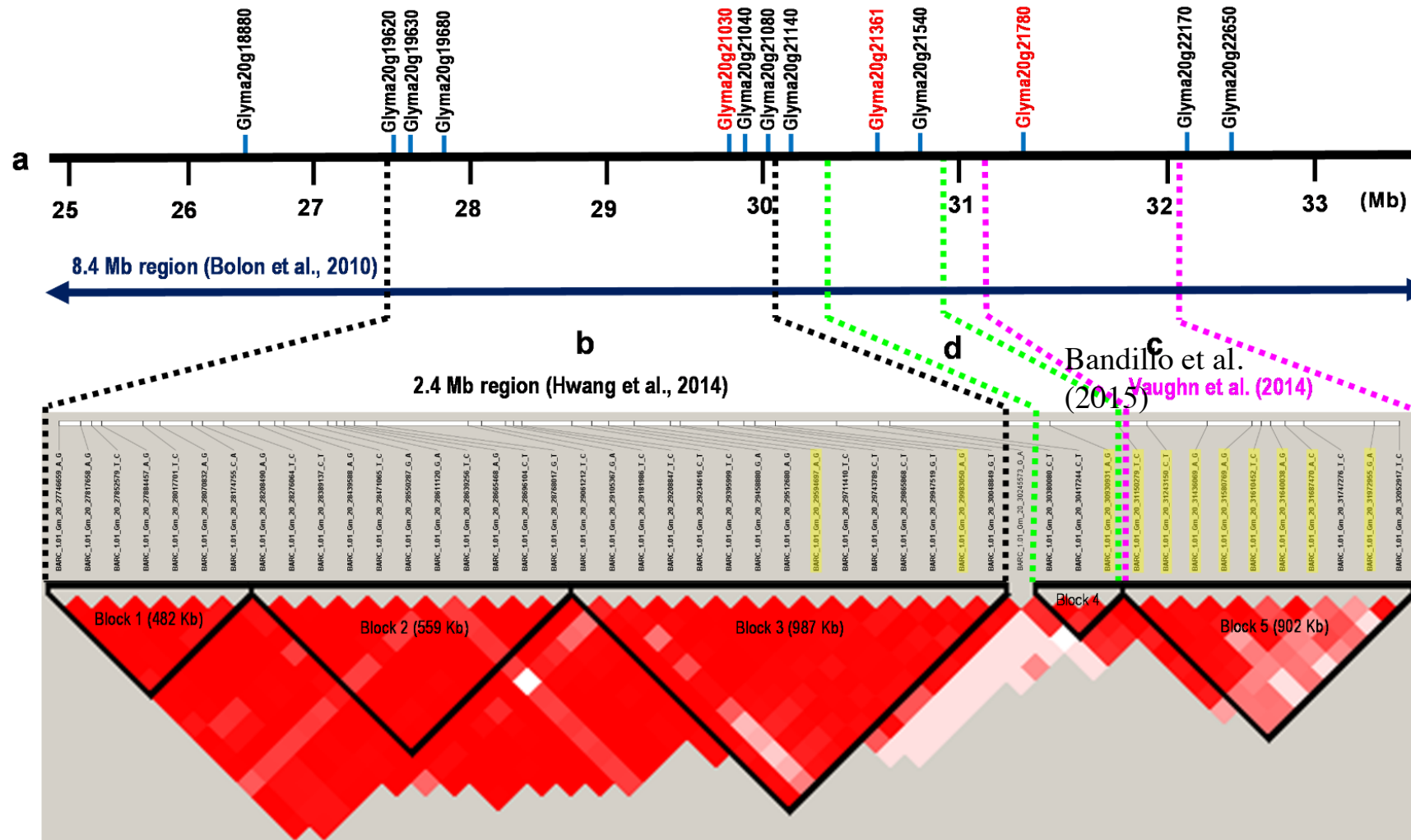


Fig. 1. Depiction of intervals on linkage group (LG) I that were (1) segregating (shaded with vertical markings), (2) contained a crossover (shaded with cross markings), or (3) were fixed for A81-356022 (no shading), the recurrent parent, for each of the Set 1 (A) and Set 2 (B) populations. Marker order (left) and map distances in centimorgans (right) are based on a linkage analysis across the three Set 1 populations. AAC32, ACG9a, and ACG9b are amplified fragment length polymorphism markers, and all others are simple sequence repeat markers. Shown below the linkage maps are the mean differences between lines in each population that were homozygous for *G. soja* or *G. max* alleles at Satt239 or Satt496, which were the genetic markers with the largest or nearly the largest  $R^2$  value in each population. Positive allele values are used when the *G. soja* alleles had a positive numerical effect on the trait. Not significant and significant at 0.05, 0.01, and 0.001 probability levels are shown by NS, \*, \*\*, and \*\*\*, respectively.

# Haplotype blocks in the cqSeed Protein-003 (Chr 20) protein QTL interval (Bandillo et al., 2015)





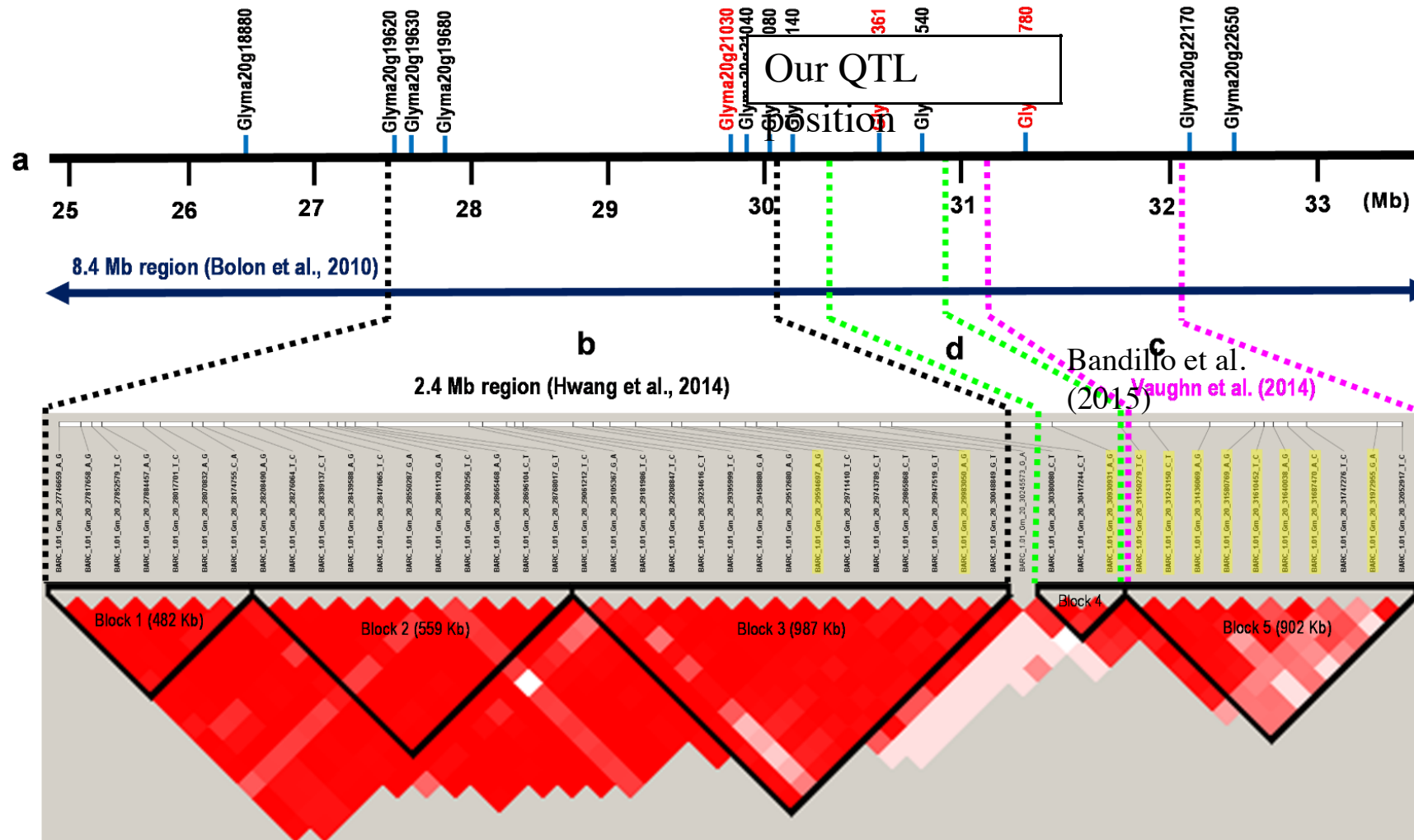


# Diers Lab Fine Mapping cqSeed Protein-003

- Tested over 7,000 plants for recombination in the chr 20 interval and tested populations from recombinant plants (over 14 years of work).

	Position†		Recombinant Line						
b10_3	26,649,365		↓ H§	H	H	A	A	B	
QTL mapped to a <80 kb interval								↓	
			SSR3		A	H	H	A	A
SSR4			A	H	H	A	A	H	
SSR5			A	H	↑ H	↑ A	A	H	
SSR6			A	B	H	A	H	H	

# Haplotype Blocks in the cqSeed Protein-003 Protein OTL Interval (Bandillo et al., 2015)



# Impact on composition of cqSeed Protein-003

- Tested backcross lines with the G. soja allele or Danbaekkong allele segregating.

	Protein g kg <sup>-1</sup>	Oil g kg <sup>-1</sup>	Yield kg ha <sup>-1</sup>
<u>G. soja PI 468916 allele (4-5 environments)</u>			
Loda	18***	-9***	-134 ns
Dwight	19***	-10***	-319 **
LS93-0375	22***	-12***	-147 ns
C1981	23***	-13***	-270***
<u>Danbaekkong allele (4 environments)</u>			
Dwight	19***	-7***	-363***
LS93-5005	20***	0***	-155***

\*, \*\*, \*\*\* Significant at P<0.05, 0.01 and 0.001, respectively

# cqSeed Protein-001 (chr 15)

- Effect not as large as cqSeed Protein-003.
- Confirmed for composition, but little work done to test its impact on agronomic traits.
- Identified from multiple germplasm sources including *G. max* and *G. soja*.
- Fine mapped and evaluated the impact of the QTL allele from PI 407788A (MG IV line from Korea).

# Fine mapping cqSeed Protein- 003

Marker	Map position		Recombinant line†					
	Glyma 1.0‡	Glyma 2.0	LG05C-1863	LG05C-1782	LG05C-1824	LG05C-1801	LG05C-1870	LG05C-1804
BARCSOYSSR_15_0056	1,346,246	1,345,578	↓ A§	↓ B	↓ A	A	A	A
BARCSOYSSR_15_0134	2,938,929	¶	A	B	A	↓ A	A	A
			#					
BARCSOYSSR_15_0160	3,563,138	3,581,354	H	H	↓ H	A	A	A
BARCSOYSSR_15_0161	3,568,888	3,587,104	H	H	H	A	A	A
BARCSOYSSR_15_0164	3,650,821	3,669,036	H	↓ H	H	H	A	A
BARCSOYSSR_15_0186	3,910,706	3,928,658	H	H	H	H	A	A
Satt384	4,036,564	4,054,490	H	H	H	H	A	A
BARCSOYSSR_15_0190	4,062,653		H	H	H	H	A	A
			↓					
BARCSOYSSR_15_0194	4,103,668	4,122,592	H	H	A	H	↓ A	↓ A
BARCSOYSSR_15_0195	4,135,853		H	H	A	H	A	A
BARCSOYSSR_15_0196	4,136,520	4,155,446	H	H	A	H	A	A
Satt691	4,189,548	4,208,308	H	H	A	H	A	A

# Fine mapping cqSeed Protein- 003

535 kb interval

Kim et al. 2016

Marker	Map position		Recombinant line†					
	Glyma 1.0‡	Glyma 2.0	LG05C-1863	LG05C-1782	LG05C-1824	LG05C-1801	LG05C-1870	LG05C-1804
BARCSOYSSR_15_0056	1,346,246	1,345,578	↓ A§	↓ B	↓ A	A	A	A
BARCSOYSSR_15_0134	2,938,929	¶	A	B	A	↓ A	A	A
			#					
BARCSOYSSR_15_0160	3,563,138	3,581,354	H	H	↓ H	A	A	A
BARCSOYSSR_15_0161	3,568,888	3,587,104	H	H	H	A	A	A
BARCSOYSSR_15_0164	3,650,821	3,669,036	H	↓ H	H	H	A	A
BARCSOYSSR_15_0186	3,910,706	3,928,658	H	H	H	H	A	A
Satt384	4,036,564	4,054,490	H	H	H	H	A	A
BARCSOYSSR_15_0190	4,062,653		H	H	H	H	A	A
			↓					
BARCSOYSSR_15_0194	4,103,668	4,122,592	H	H	A	H	↓ A	↓ A
BARCSOYSSR_15_0195	4,135,853		H	H	A	H	A	A
BARCSOYSSR_15_0196	4,136,520	4,155,446	H	H	A	H	A	A
Satt601	4,180,548	4,208,308	H	H	A	H	A	A

# Impact on composition of cqSeed Protein-001

- Tested backcross lines with the high protein allele from PI 407788A.

	Protein	Oil	Yield
	g kg <sup>-1</sup>	g kg <sup>-1</sup>	kg ha <sup>-1</sup>
<u>G. soja PI 468916 allele (4 environments)</u>			
AR09- 192019	10***	-6***	-109 ns
*LD02-448 Significant at P < 0.05, 0.01, and 0.001, respectively.			



# Conclusion

- cqSeed Protein-003
  - Mapped to a <80 kb interval and a candidate gene identified.
  - Results in approximately a 20 g kg<sup>-1</sup> protein increase and a 10 g kg<sup>-1</sup> decrease in oil.
  - Negative impact on yield
  - The Danbaekkong allele found to also have a negative impact on yield.
- cqSeed Protein-001
  - Mapped to a 535 kb interval and a new round of mapping has been initiated.
  - Results in approximately a 10 g kg<sup>-1</sup> protein increase and a 6 g kg<sup>-1</sup> decrease in oil.
  - Negative trend for yield, but not significant.
- For each 10 g kg<sup>-1</sup> increase in protein
  - cqSeed Protein-003 = 67 – 228 kg ha<sup>-1</sup> yield reduction
  - cqSeed Protein-001 = 69 kg ha<sup>-1</sup> yield reduction