

GENOME WIDE ASSOCIATION ANALYSIS OF AGRONOMIC TRAITS

Samira Mafi Moghaddam

**Common Bean Coordinate Agricultural Project
2013 Advisory Board Meeting**



Introduction

- ❖ Why use Genome Wide Association Study(GWAS)?
 - ✓ Samples multiple alleles in a natural germplasm collection
 - ✓ No requirement of crossing or pedigree information
 - ✓ Higher resolution in mapping the quantitative trait loci (QTL) due to the recombination that has occurred over many generations
 - ✓ More precise positioning of the QTL
- ✓ **Goal:**
 - ✓ At this point GWAS is used to identify candidate regions of the genome that control the traits of interest

GWAS Procedure

❖ Genotypes

- ❑ Genotypes are from the Middle American gene pool (total of 300)

❖ Marker set

- ❑ 10,282 SNP markers
- ❑ Revised set 10,287 SNPs (exclude repeated markers and those with missing data =>20%) ★

❖ Imputation

- ❑ Guess the missing genotypes (fastPHASE, Scheet P and Stephens, M ,2006)


❖ GWAS

- ❑ Linear mixed model in GAPIT which runs in R (Lipka et. al 2012)

GWAS Procedure

❖ Population Structure

- ❑ Principal component analysis
- ❑ (Proc princomp in SAS)

- 
- 1) Number of components that control 25% and 50% of variation
 - 2) Velicer's minimum average partial test (1976) as the stopping rule (SAS) / BIC from GAPIT to determine the number of PC to be incorporated in the GWAS★

❖ Relatedness

- 1) Identity by state kinship matrix (EMMA in GAPIT, Kang et al, 2008)
 - 2) Identity by descent kinship matrix (Loiselle in GAPIT, Loiselle et al. 1995)
- Markers that are not in LD ($LD < 0.5$) are used for the two above steps (PLINK, Purcell et al, 2007)

GWAS Procedure

□ Models

- Naïve $y = X\alpha + \varepsilon$
- PC $y = X\alpha + P\beta + \varepsilon$
- EMMA $y = X\alpha + K1 + \varepsilon$
- Loiselle $y = X\alpha + K2 + \varepsilon$
- EMMA + PC $y = X\alpha + P\beta + K1 + \varepsilon$
- Loiselle + PC $y = X\alpha + P\beta + K2 + \varepsilon$

□ Significant P-value cutoff :

- Bonferroni correction (0.05/total number of markers)
- Modified Bonferroni (0.05/number of markers that are not in LD)
- Top 5-10% P-values

Summary of Analyzed Traits

❖ Agronomics traits in the national trial

- ❑ Seven phenotypes (DF, DM, PH, GH, LG, SW,SY)
- ❑ Three hundred genotypes (271 used in GWAS)

❖ Agronomic traits under drought stress and non drought

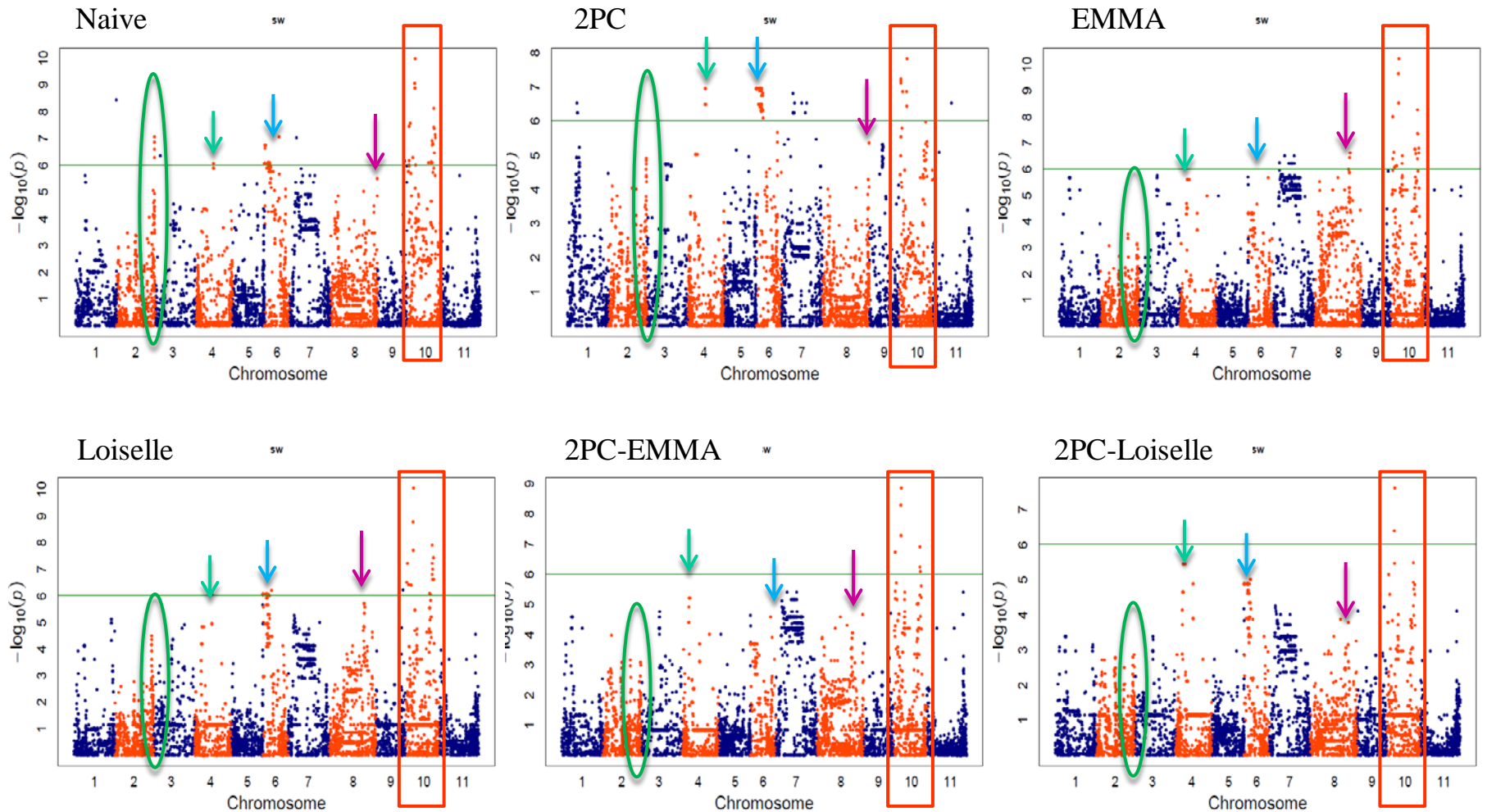
- ❑ Seven phenotypes (DF, DM, PH, GH, LG, SW,SY) + DSI, PR, GM
- ❑ Ninety six genotypes (87 used in GWAS)

❖ Nutritional traits

- ❖ Phytate, Fiber, protein, Caco2, Zinc and Iron
- ❖ Hundred and fifty six genotypes (only Middle American gene pool)/ 194 genotypes (Middle American and Andean gene pool)
- ❖ Zn and Fe : 154 genotypes (only Middle American gene pool)

GWAS Results- Agronomic Traits

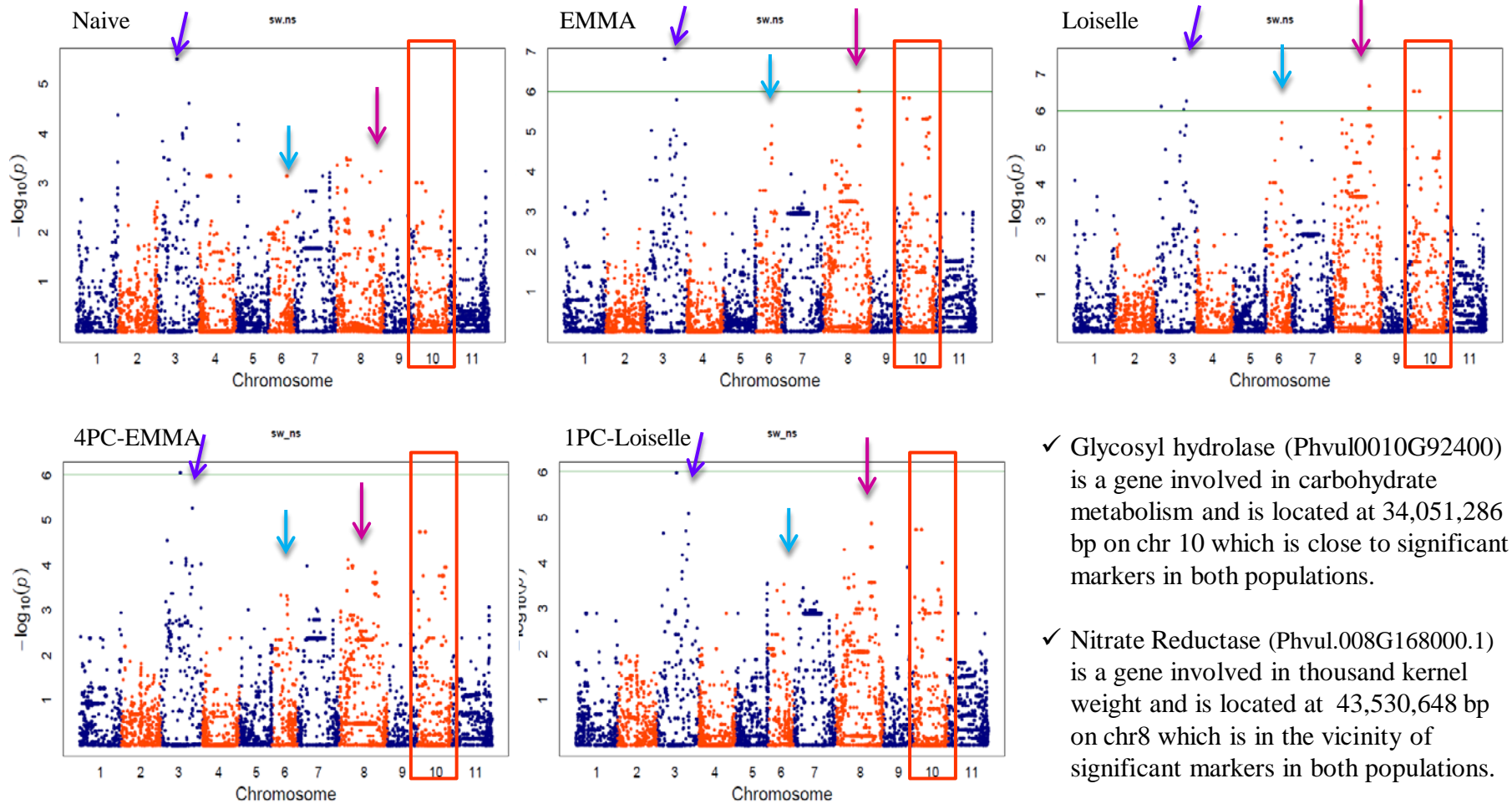
Manhattan plot – Seed weight (Large national trial)



GWAS Results- Agronomic Traits

Manhattan plot – Seed weight (Small trial)

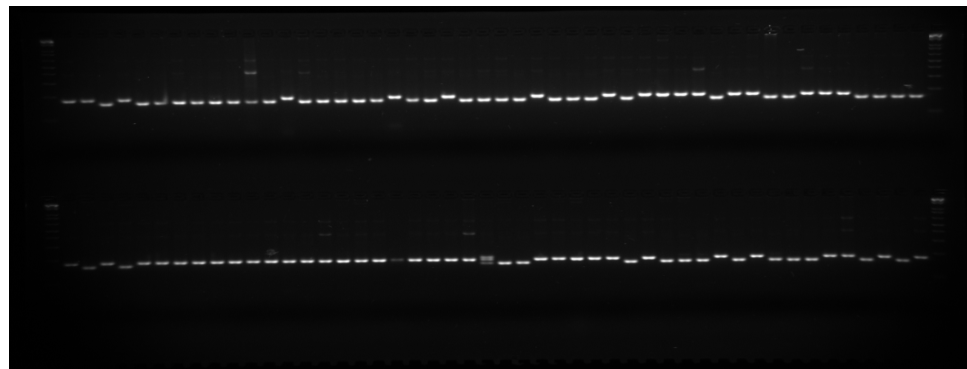
- ✓ The peak on chromosome 3 is weak in the large trial
- ✓ The peak on chromosome 2 is weak in the small trial
- ✓ The peaks on chromosomes 6, 8 and 10 are strong in both populations
- ✓ There is a weak signal on chromosome 11 in both populations



- ✓ Glycosyl hydrolase (Phvul0010G92400) is a gene involved in carbohydrate metabolism and is located at 34,051,286 bp on chr 10 which is close to significant markers in both populations.
- ✓ Nitrate Reductase (Phvul.008G168000.1) is a gene involved in thousand kernel weight and is located at 43,530,648 bp on chr8 which is in the vicinity of significant markers in both populations.

Incorporation of Indel Markers

- Goal: test the performance of Indels in association mapping
- Region of interest: around the nitrate reductase gene (43,530,648 bp)
- Twenty two Indel markers (from 2,687 Indel panel) between 42 Mb to 46 Mb region were screened on available genotypes from the large National trial (252genotypes)
- Twelve Markers were polymorphic and scorable.
- GWAS results using Indels and SNPs match in both populations.

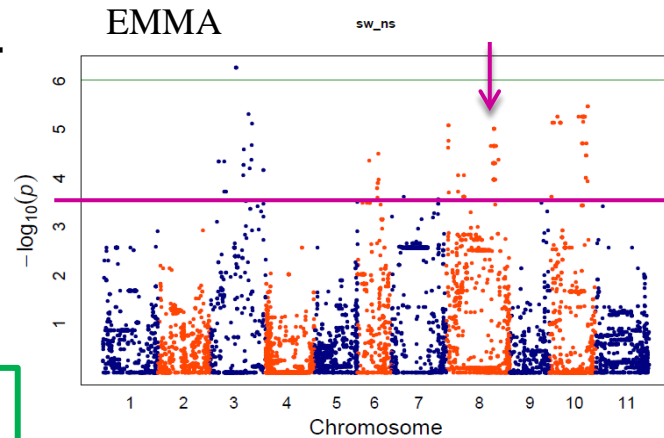


NDSU-Ind-8-42.4672

Seed Weight –Small Population

81genotypes-12 SNPs

SNP	Chr	Position	P.value	maf	R-square (Naïve)
ndsu_8_42.0202	8	42,020,183	8.64E-01	0.42	0.47
ndsu_8_42.3026	8	42,302,570	3.98E-01	0.3	0.02
ndsu_8_42.4672	8	42,467,187	8.68E-01	0.47	0.08
ndsu_8_42.4676	8	42,467,576	7.92E-01	0.46	0.06
ndsu_8_43.5710	8	43,570,952	6.19E-02	0.37	0.67
ndsu_8_44.7084	8	44,708,437	6.19E-02	0.37	0.67
ndsu_8_44.8451	8	44,845,148	6.19E-02	0.37	0.67
ndsu_8_45.2120	8	45,211,996	3.28E-01	0.38	0.61



81genotypes-10287 SNPs

SNP	Chr	Position	P_value	maf
m9223	8	41,742,018	2.17E-05	0.345679
m4922	8	43,827,895	4.97E-05	0.37037
m6145	8	44,237,412	4.97E-05	0.37037
m6144	8	44,246,478	4.97E-05	0.37037
m6143	8	44,270,482	4.97E-05	0.37037
m6142	8	44,272,503	4.97E-05	0.37037
m6141	8	44,273,042	4.97E-05	0.37037
m6139	8	44,287,718	4.97E-05	0.37037
m6174	8	44,338,667	1.06E-04	0.358025
m6173	8	44,399,862	1.06E-04	0.358025
m4856	8	44,425,316	1.06E-04	0.358025
m4864	8	44,470,617	2.17E-05	0.345679
m4863	8	44,479,849	1.06E-04	0.358025
m4862	8	44,480,227	2.17E-05	0.345679
m4861	8	44,494,182	2.17E-05	0.345679
m4860	8	44,526,077	2.17E-05	0.345679
m4859	8	44,532,123	2.17E-05	0.345679
m4858	8	44,535,382	1.06E-04	0.358025
m4029	8	44,605,263	4.97E-05	0.37037
m4051	8	44,712,020	9.70E-06	0.358025
m4050	8	44,724,115	9.70E-06	0.358025
m4048	8	44,754,521	9.70E-06	0.358025
m4044	8	44,792,096	9.70E-06	0.358025
m4046	8	44,804,652	9.70E-06	0.358025
m7104	8	44,833,382	2.17E-05	0.345679
m7103	8	44,839,102	1.06E-04	0.358025
m7102	8	44,842,481	1.06E-04	0.358025
m7101	8	44,853,371	1.06E-04	0.358025
m7100	8	44,858,735	2.17E-05	0.345679
m7099	8	44,862,000	1.06E-04	0.358025
m9586	8	44,891,515	1.06E-04	0.358025
m4752	8	44,921,533	2.17E-05	0.345679
m7248	8	45,096,248	1.06E-04	0.358025
m8871	8	45,109,668	1.06E-04	0.358025
m8872	8	45,117,316	1.06E-04	0.358025
m8870	8	45,129,565	1.06E-04	0.358025
m2503	8	45,152,224	4.97E-05	0.37037
m2502	8	45,232,332	4.97E-05	0.37037
m2504	8	45,319,371	4.97E-05	0.37037
m6457	8	45,870,532	4.97E-05	0.37037
m9472	8	46,032,428	1.06E-04	0.358025
m3465	8	46,377,730	2.17E-05	0.345679
m3464	8	46,433,050	2.17E-05	0.345679

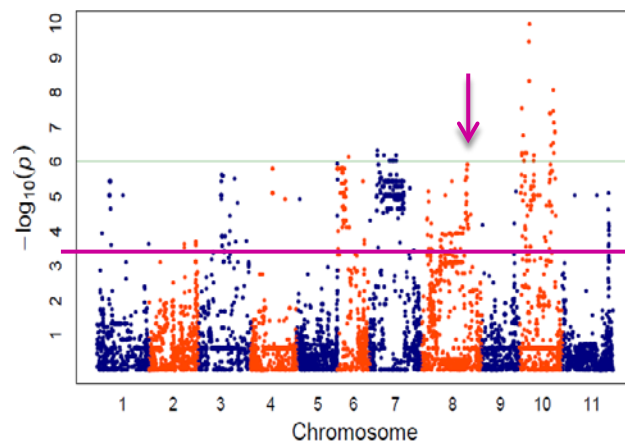
- Most of the significant markers in this region are located on position 44MB
- The smallest P-value belongs to highlighted markers which are in LD and are spread from 44.71 to 44.80Mb
- The same pattern and locations are observed when Indel markers are used
- The R-square of Highlighted SNP= 0.69 in Naïve and 0.14 in PC model

Seed Weight –National Trial

252genotypes-12 Indels

SNP	Chr	Position	P.value	maf	R-square (Naive)
ndsu_8_42.0202	8	42,020,183	6.96E-03	0.42	0.55
ndsu_8_42.3026	8	42,302,570	2.95E-01	0.32	0.08
ndsu_8_42.4672	8	42,467,187	2.03E-01	0.42	0.14
ndsu_8_42.4676	8	42,467,576	1.97E-01	0.42	0.14
ndsu_8_43.0997	8	43,099,724	4.32E-02	0.00	0.00
ndsu_8_43.5710	8	43,570,952	3.11E-03	0.40	0.59
ndsu_8_44.5633	8	44,563,293	4.50E-01	0.01	0.00
ndsu_8_44.7084	8	44,708,437	9.01E-04	0.40	0.64
ndsu_8_44.8451	8	44,845,148	9.01E-04	0.40	0.64
ndsu_8_45.2120	8	45,211,996	4.61E-04	0.42	0.61

EMMA



252genotypes-10287 SNPs

SNP	chr	Position	P_value	maf
m4922	8	43,827,895	3.51E-06	0.42
m6145	8	44,237,412	2.97E-05	0.42
m6144	8	44,246,478	2.97E-05	0.42
m6143	8	44,270,482	2.97E-05	0.42
m6142	8	44,272,503	2.97E-05	0.42
m6141	8	44,273,042	2.97E-05	0.42
m6139	8	44,287,718	2.97E-05	0.42
m6174	8	44,338,667	4.64E-05	0.41
m6173	8	44,399,862	4.64E-05	0.41
m4856	8	44,425,316	4.64E-05	0.41
m4864	8	44,470,617	8.35E-06	0.40
m4863	8	44,479,849	4.64E-05	0.41
m4862	8	44,480,227	8.35E-06	0.40
m4861	8	44,494,182	8.35E-06	0.40
m4860	8	44,526,077	8.35E-06	0.40
m4859	8	44,532,123	8.35E-06	0.40
m4858	8	44,535,382	4.64E-05	0.41
m4029	8	44,605,263	1.73E-05	0.41
m4051	8	44,712,020	2.81E-06	0.40
m4050	8	44,724,115	2.81E-06	0.40
m4048	8	44,754,521	2.81E-06	0.40
m4044	8	44,792,096	2.81E-06	0.40
m4046	8	44,804,652	2.81E-06	0.40
m7104	8	44,833,382	8.35E-06	0.40
m7103	8	44,839,102	4.64E-05	0.41
m7102	8	44,842,481	4.64E-05	0.41
m7101	8	44,853,371	4.64E-05	0.41
m7100	8	44,858,735	8.35E-06	0.40
m7099	8	44,862,000	4.64E-05	0.41
m9586	8	44,891,515	1.45E-05	0.40
m4752	8	44,921,533	8.35E-06	0.40
m7248	8	45,096,248	4.64E-05	0.41
m8871	8	45,109,668	4.64E-05	0.41
m8872	8	45,117,316	4.64E-05	0.41
m8870	8	45,129,565	4.64E-05	0.41
m2503	8	45,152,224	1.68E-06	0.42
m2502	8	45,232,332	1.68E-06	0.42
m2504	8	45,319,371	1.22E-06	0.41
m6457	8	45,870,532	1.22E-06	0.41
m9472	8	46,032,428	3.78E-05	0.40

- The R-square of SNPs in Blue box= 0.65 in Naïve and 0.15 in PC model
- The same regions are detected in the larger population using both SNPs and Indels which indicates that the population size in this study did not affect the power to detect a strong signal

Chr	Df			Dm			Lg			Ph			SW			Sy			Identified QTL and Nearest Marker Position	The Nearest Gene		
	Nat	Ns	Ds	Nat	Ns	Ds	Nat	Ns	Ds	Nat	Ns	Ds	Nat	Ns	Ds	Nat	Ns	Ds				
1																						
2																					dm2.1, dm2.2, dm2.3 (unknown position)	
3			24 Mb	46 Mb									22-25Mb	24Mb	24Mb	22-23 Mb					PhvuI.003G104600 (FPF1) df-gene 25,967,515bp	
4		1 Mb & 44Mb	1 Mb	1 Mb									8-11Mb			2Mb & 34Mb					dm 4.1(unknown position)	
5							5 Mb	5 Mb	5 Mb													
6			10 Mb										0-8Mb	19Mb	19Mb	25Mb	25 Mb				1)sw6.2(20,254,184 bp) 2)sw6.1(28,199,940 bp)	phvuI.006g171100(tryptophan biosynthesis protein) sw-gene 28,198,235bp
7							43-48 Mb	45-48 Mb	45-48 Mb						24 & 11 Mb		2 Mb				1)sw7.1(31,123,850 bp) 2)sw7.1(4,959,794 bp)	sw-gene 1)PhvuI.007G130500(oligosaccharide biosynthesis) 31,123,624bp 2)PhvuI.007G059700 (nutrient reservoir activity) 4,959,769bp
8													44-45 Mb	44-45 Mb	44-45 Mb						sw8.1(7,858,729 bp)	PhvuI.008G168000.1 Nitrate reductase gene 43,530,648bp
9		21-22 Mb	21-22 Mb	16-20 Mb	21&25&29 Mb	21&25&29 Mb							16-17Mb only in 2PC model		29-35Mb/16.5Mb						1)sw9.1(17,485,625 bp) 2) dm9.1 and 9.2 (unknown position) 3)df9.1(22,665,991bp)	1) PhvuI.009G130100 (EFS df-gene) 19,158,508bp 2) PhvuI.009G159400 (FRI df-gene) 23,114,173bp 3)PhvuI.009G116700 sw-gene (chitin catabolism) 17,485,614bp
10													9-10 Mb & 32-36 Mb	2-10 Mb & 31-35 Mb	9-10 Mb & 32-36 Mb	2-10 Mb & 31-35 Mb						PhvuI0010G92400 (sw-gene in Carbohydrate metabolism) 34,055,353bp
11		7 Mb	7 Mb	7 Mb	8 Mb		9-10 Mb	9-10 Mb	9-10 Mb	8-10 Mb	8-10 Mb	8-10 Mb	43-44 Mb	43-44 Mb	43Mb						F17R8.420 ph-QTL (unknown position)	PhvuI.011G070700 (APRR3 df-gene) 6,218,052bp

Conclusion



- Indel markers can be used in GWAS at medium throughput levels.
- The GWAS results can be different in each population due to changes in allele frequencies. However, some QTL were found in both population sizes in our study such as those related to seed weight and days to flower traits.

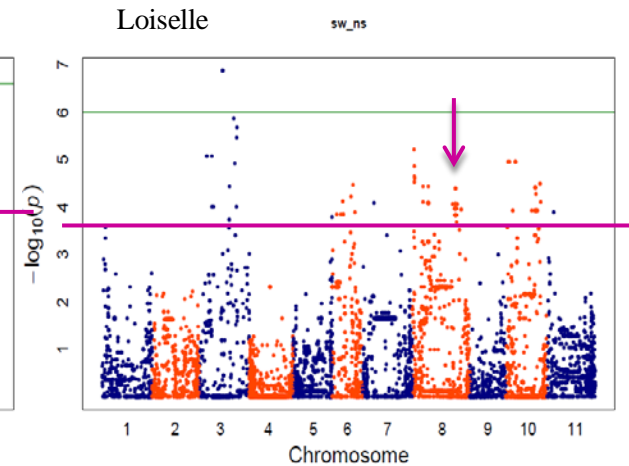
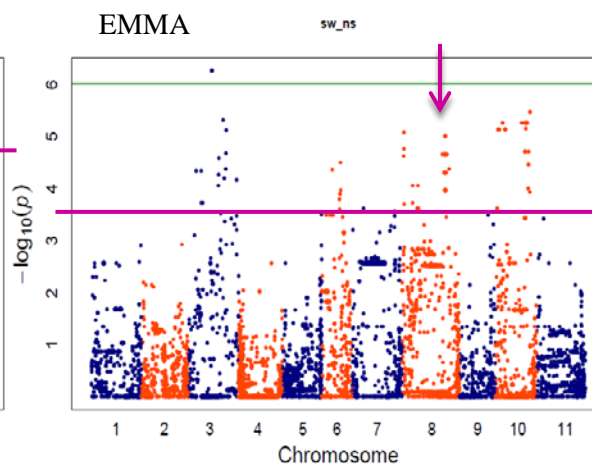
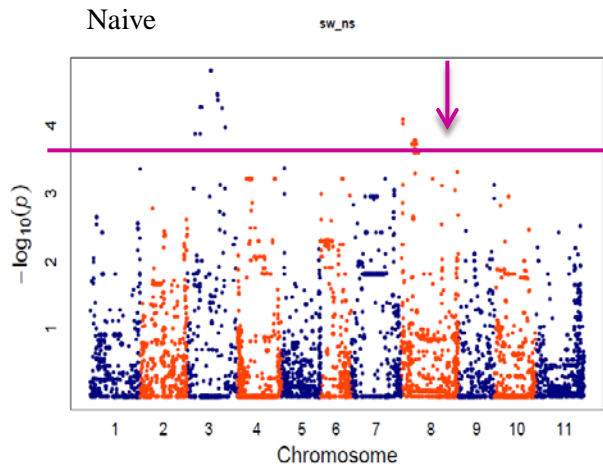
THANK YOU



Reference of QTL

- Checa, O. E., & Blair, M. W. (2012). Inheritance of Yield-Related Traits in Climbing Beans (L.). *Crop Science*, 52(5), 1998-2013.
- Blair, M. W., Galeano, C. H., Tovar, E., Muñoz Torres, M. C., Castrillón, A. V., Beebe, S. E., & Rao, I. M. (2012). Development of a Mesoamerican intra-genepool genetic map for quantitative trait loci detection in a drought tolerant × susceptible common bean (*Phaseolus vulgaris* L.) cross. *Molecular Breeding*, 1-18.
- (Blair et al. 2003,2006)
- (Gaitan-Solis et al. 2002)
- Pérez-Vega, E., Pañeda, A., Rodríguez-Suárez, C., Campa, A., Giraldez, R., & Ferreira, J. J. (2010). Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). *TAG Theoretical and Applied Genetics*, 120(7), 1367-1380.
- Beattie, A. D., Larsen, J., Michaels, T. E., & Pauls, K. P. (2003). Mapping quantitative trait loci for a common bean (*Phaseolus vulgaris* L.) ideotype. *Genome*, 46(3), 411-422.
- Wright, E. M., & Kelly, J. D. (2011). Mapping QTL for seed yield and canning quality following processing of black bean (*Phaseolus vulgaris* L.). *Euphytica*, 179(3), 471-484.

Seed weight -Drought-ns-SNP



Sig	SNP	Chr	Position (bp)	P-value	MAF
m4051		8	44,712,020	1.28E-01	0.35
m4050		8	44,724,115	1.28E-01	0.35
m4048		8	44,754,521	1.28E-01	0.35
m4044		8	44,792,096	1.28E-01	0.35
m4046		8	44,804,652	1.28E-01	0.35

Sig	SNP	Chr	Position (bp)	P-value	MAF
m4051		8	44,712,020	9.70E-06	0.35
m4050		8	44,724,115	9.70E-06	0.35
m4048		8	44,754,521	9.70E-06	0.35
m4044		8	44,792,096	9.70E-06	0.35
m4046		8	44,804,652	9.70E-06	0.35

Sig	SNP	Chr	Position (bp)	P-value	MAF
m4050		8	44,724,115	3.95E-05	0.35
m4048		8	44,754,521	3.95E-05	0.35
m4051		8	44,712,020	3.95E-05	0.35
m4044		8	44,792,096	3.95E-05	0.35
m4046		8	44,804,652	3.95E-05	0.35

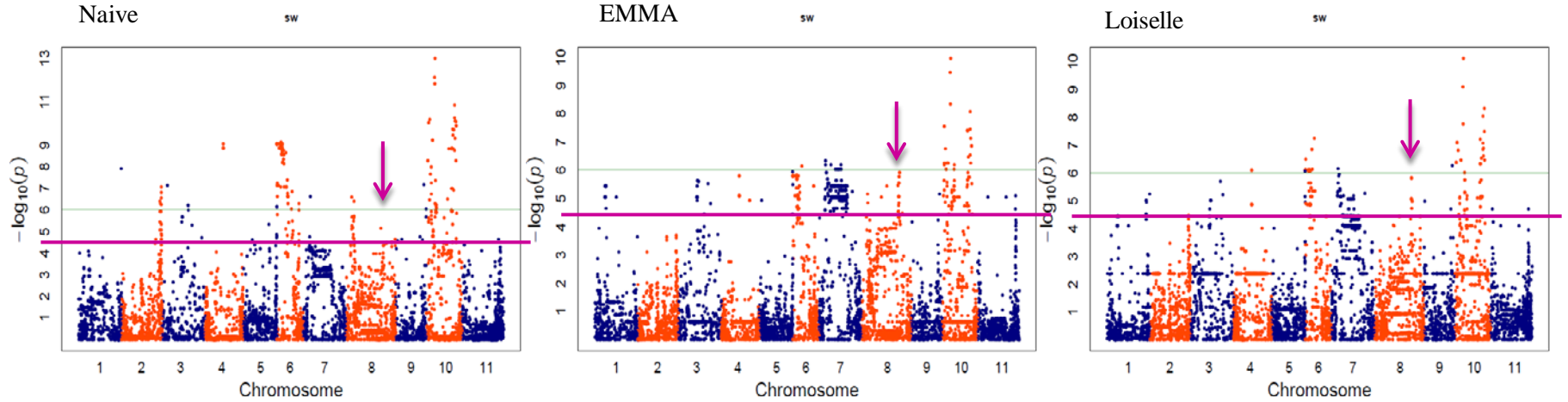
Seed weight -Drought trial-Indels

Indel	chr	Position	P_value	maf
ndsu_8_43.5710	8	43,570,952	3.47E-04	0.37
ndsu_8_44.7084	8	44,708,437	3.47E-04	0.37
ndsu_8_44.8451	8	44,845,148	3.47E-04	0.37
ndsu_8_42.3026	8	42,302,570	1.78E-03	0.30
ndsu_8_42.4672	8	42,467,187	1.25E-02	0.47
ndsu_8_45.2120	8	45,211,996	1.18E-01	0.38
ndsu_8_42.4676	8	42,467,576	1.68E-01	0.46
ndsu_8_42.0202	8	42,020,183	4.49E-01	0.42

Indel	chr	Position	P_value	maf
ndsu_8_43.5710	8	43,570,952	6.19E-02	0.37
ndsu_8_44.7084	8	44,708,437	6.19E-02	0.37
ndsu_8_44.8451	8	44,845,148	6.19E-02	0.37
ndsu_8_45.2120	8	45,211,996	3.28E-01	0.38
ndsu_8_42.3026	8	42,302,570	3.98E-01	0.30
ndsu_8_42.4676	8	42,467,576	7.92E-01	0.46
ndsu_8_42.0202	8	42,020,183	8.64E-01	0.42
ndsu_8_42.4672	8	42,467,187	8.68E-01	0.47

Indel	chr	Position	P_value	maf
ndsu_8_43.5710	8	43,570,952	2.89E-03	0.37
ndsu_8_44.7084	8	44,708,437	2.89E-03	0.37
ndsu_8_44.8451	8	44,845,148	2.89E-03	0.37
ndsu_8_45.2120	8	45,211,996	2.66E-02	0.38
ndsu_8_42.0202	8	42,020,183	3.24E-01	0.42
ndsu_8_42.3026	8	42,302,570	3.45E-01	0.30
ndsu_8_42.4672	8	42,467,187	5.24E-01	0.47
ndsu_8_42.4676	8	42,467,576	7.77E-01	0.46

Seed weight -National trial-252 Genotypes-SNP



Sig	SNP	Chr	Position (bp)	P-value	MAF
m2504		8	45,319,371	4.24E-05	0.41
m6457		8	45,870,532	4.24E-05	0.41
.					
.					
m4051		8	44,712,020	1.65E-04	0.40
m4050		8	44,724,115	1.65E-04	0.40
m4048		8	44,754,521	1.65E-04	0.40
m4044		8	44,792,096	1.65E-04	0.40
m4046		8	44,804,652	1.65E-04	0.40

Sig	SNP	Chr	Position (bp)	P-value	MAF
m2504		8	45,319,371	1.22E-06	0.41
m6457		8	45,870,532	1.22E-06	0.41
m2503		8	45,152,224	1.68E-06	0.42
m2502		8	45,232,332	1.68E-06	0.42
m4051		8	44,712,020	2.81E-06	0.40
m4050		8	44,724,115	2.81E-06	0.40
m4048		8	44,754,521	2.81E-06	0.40
m4044		8	44,792,096	2.81E-06	0.40
m4046		8	44,804,652	2.81E-06	0.40
m4922		8	43,827,895	3.51E-06	0.42

Sig	SNP	Chr	Position (bp)	P-value	MAF
m4051		8	44,712,020	1.50E-06	0.40
m4050		8	44,724,115	1.50E-06	0.40
m4048		8	44,754,521	1.50E-06	0.40
m4044		8	44,792,096	1.50E-06	0.40
m4046		8	44,804,652	1.50E-06	0.40
m4864		8	44,470,617	8.35E-06	0.40
m4862		8	44,480,227	8.35E-06	0.40
m4861		8	44,494,182	8.35E-06	0.40
m4860		8	44,526,077	8.35E-06	0.40
m4859		8	44,532,123	8.35E-06	0.40

Seed weight –National trial-Indels

Naive

Indel	Chr	Position (bp)	P-value	MAF
ndsu_8_44.7084	8	44,708,437	1.38E-07	0.40
ndsu_8_44.8451	8	44,845,148	1.38E-07	0.40
ndsu_8_42.3026	8	42,302,570	3.75E-07	0.32
ndsu_8_43.5710	8	43,570,952	7.49E-07	0.40
ndsu_8_45.2120	8	45,211,996	8.28E-05	0.41
ndsu_8_42.4672	8	42,467,187	2.00E-04	0.42
ndsu_8_42.4676	8	42,467,576	7.95E-04	0.42
ndsu_8_42.0202	8	42,020,183	1.91E-02	0.42
ndsu_8_44.5633	8	44,563,293	2.81E-02	0.01

EMMA

Indel	Chr	Position (bp)	P-value	MAF
ndsu_8_45.2120	8	45,211,996	8.06E-04	0.41
ndsu_8_44.7084	8	44,708,437	1.51E-03	0.40
ndsu_8_44.8451	8	44,845,148	1.51E-03	0.40
ndsu_8_43.5710	8	43,570,952	3.11E-03	0.40
ndsu_8_42.0202	8	42,020,183	5.71E-03	0.42
ndsu_8_42.4676	8	42,467,576	1.97E-01	0.42
ndsu_8_42.4672	8	42,467,187	2.03E-01	0.42
ndsu_8_42.3026	8	42,302,570	2.95E-01	0.32

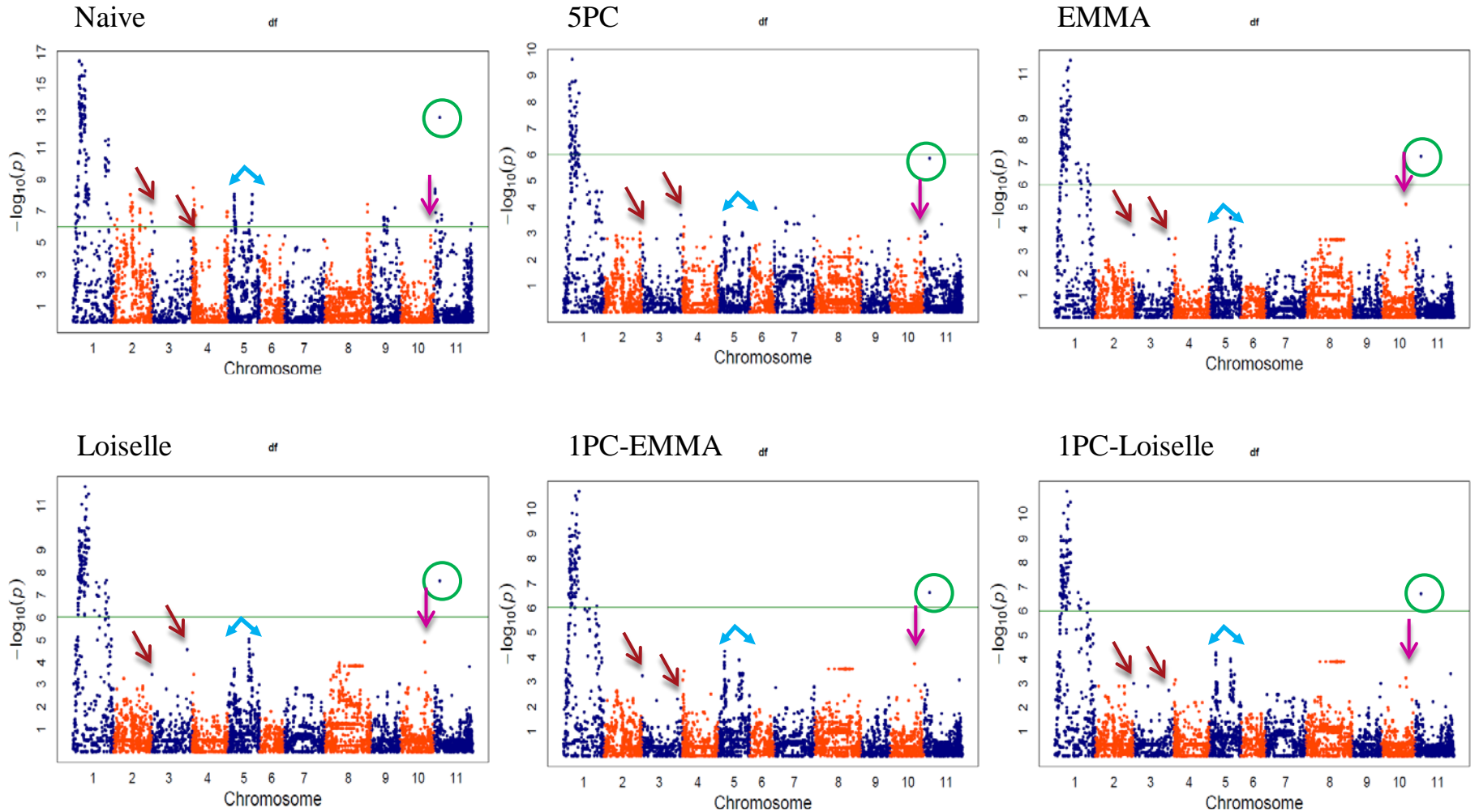
Loiselle

Indel	Chr	Position (bp)	P-value	MAF
ndsu_8_44.7084	8	44,708,437	8.91E-03	0.39
ndsu_8_44.8451	8	44,845,148	8.91E-03	0.39
ndsu_8_43.5710	8	43,570,952	2.06E-02	0.39
ndsu_8_45.2120	8	45,211,996	2.51E-02	0.43
ndsu_8_42.0202	8	42,020,183	9.97E-02	0.42
ndsu_8_42.3026	8	42,302,570	3.78E-01	0.32
ndsu_8_42.4672	8	42,467,187	6.54E-01	0.42
ndsu_8_42.4676	8	42,467,576	7.14E-01	0.41

- ✓ Position 44.7 -44.8 Mb is the significant region in proximity of Nitrate Reductase using both marker systems
- ✓ Indel markers can be used in GWAS at medium throughput level
- ✓ The population size did not affect the power to find three major signals on chr 6,8 and 10.
- ✓ The GWAS results can be different in each population due to changes in allele frequencies

GWAS Results- Agronomic traits

Manhattan plot – Days to flower (Large trial)



GWAS Results- Agronomic traits

Manhattan plot – Days to flower (Drought trial)

