Advances towards a Marker Assisted Selection (MAS) breeding program in sunflower for Sclerotinia disease resistance

Zahirul Talukder[1], Brent Hulke[2], Lili Qi[2], and Thomas Gulya[2]
[1]Department of Plant Sciences, NDSU
[2]USDA-ARS Northern Crop Science Laboratory
Fargo, North Dakota

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Introduction

- *Sclerotinia sclerotiorum* causes two serious diseases in sunflower
  - stalk rot, incited by root infection (unique to sunflower), and head rot, caused by airborne ascospores
  - genetics of resistance is different for the two diseases
  - Resistance is polygenic, no major resistant gene is known
  - **Resistance breeding relies on incorporating genetic factors from various partially-tolerant breeding lines**

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Goal of the project

- Identification of DNA markers associated with the Sclerotinia disease resistance
- Develop an integrated MAS breeding program for sunflower using high throughput SNP markers associated with resistance to Sclerotinia and other diseases, and also to important agronomic and oil traits

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Mapping approach

Linkage disequilibrium (LD) based Association Mapping

– Advantages

- no need to develop mapping population
- exploits historical and evolutionary recombination events at the population level
- mapping resolution is higher than bi-parental mapping population of the same size

– Disadvantages

- suffer from risk of incurring false positives due to population structure and kinship among individuals

Association Mapping

- Two approaches
  - Candidate-Gene Association (CGA) Mapping
    - alleles at a few selected functional candidate genes thought to be involved in controlling the trait of interest may be tested for association
  - Genome-Wide Association (GWA) Mapping
    - whole genome may be scanned to identify markers that are associated with a particular phenotype
- Candidate-Gene Association Mapping study is more hypothesis-driven than a Genome-Wide study

Materials and Methods

Stalk rot AM population	Head rot AM population
Total sunflower lines = 260	Total sunflower lines = 230
-- Plant introductions (PIs) = 249 -- Elite USDA inbred lines = 11	-- Plant introductions (PIs) = 196 -- Elite USDA inbred lines = 34
Inoculated field trials in 2008 & 2009	Inoculated field trials in 2011 & 2012
Total dataset = 4	Total dataset = 3
All field trials were conducted in a 'sets-in-reps' field design with 2 reps	
Car 270 (susceptible) & Croplan 305 (resistant) hybrids were used as checks	

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Candidate Gene Association Mapping

- Six *Arabidopsis thaliana* defense genes :
  - *ABI1* (ABA Insensitive 1), and
  - *ABI2* (ABA Insensitive 2) -involved in abscisic acid (ABA) signal transduction
  - *COI1* (Coronatine Insensitive 1) jasmonate receptor
  - *DET3* (De-Etiolated 3) - involved in oxalic acid signaling
  - *EIN2* (Ethylene Insensitive 2) - central regulator of ethylene signaling, and
  - *LACS2* (Long-chain Acyl-CoA Synthetase 2) - involved in cutin biosynthesis pathway
- Primer design:
  - Searched candidate gene sequences in the NCBI EST database for sunflower EST
  - Sunflower ESTs with high e-value were used to search for contig assembly sequences in the Compositae Genome Project database
  - *Primer3* software was used to design primer from the contig sequences

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Candidate Gene Association Mapping

- DNA extraction, PCR amplification & sequencing
- Sequence analysis and SNP survey:
  - *DnaSP* v.5.1 and *SNiPlay* softwares were used
- Population structure:
  - 135 SNP markers were used in *Structure* v.2.3.3 software
- Kinship analysis:
  - 5244 SNP markers were used in *SPAGeDi* v1.3a software
- Linkage disequilibrium and Association mapping:
  - *Haploview* v.4.2 and *TASSEL* v.3.0 software were used for LD and association mapping analysis, respectively

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Result: stalk rot CGAM

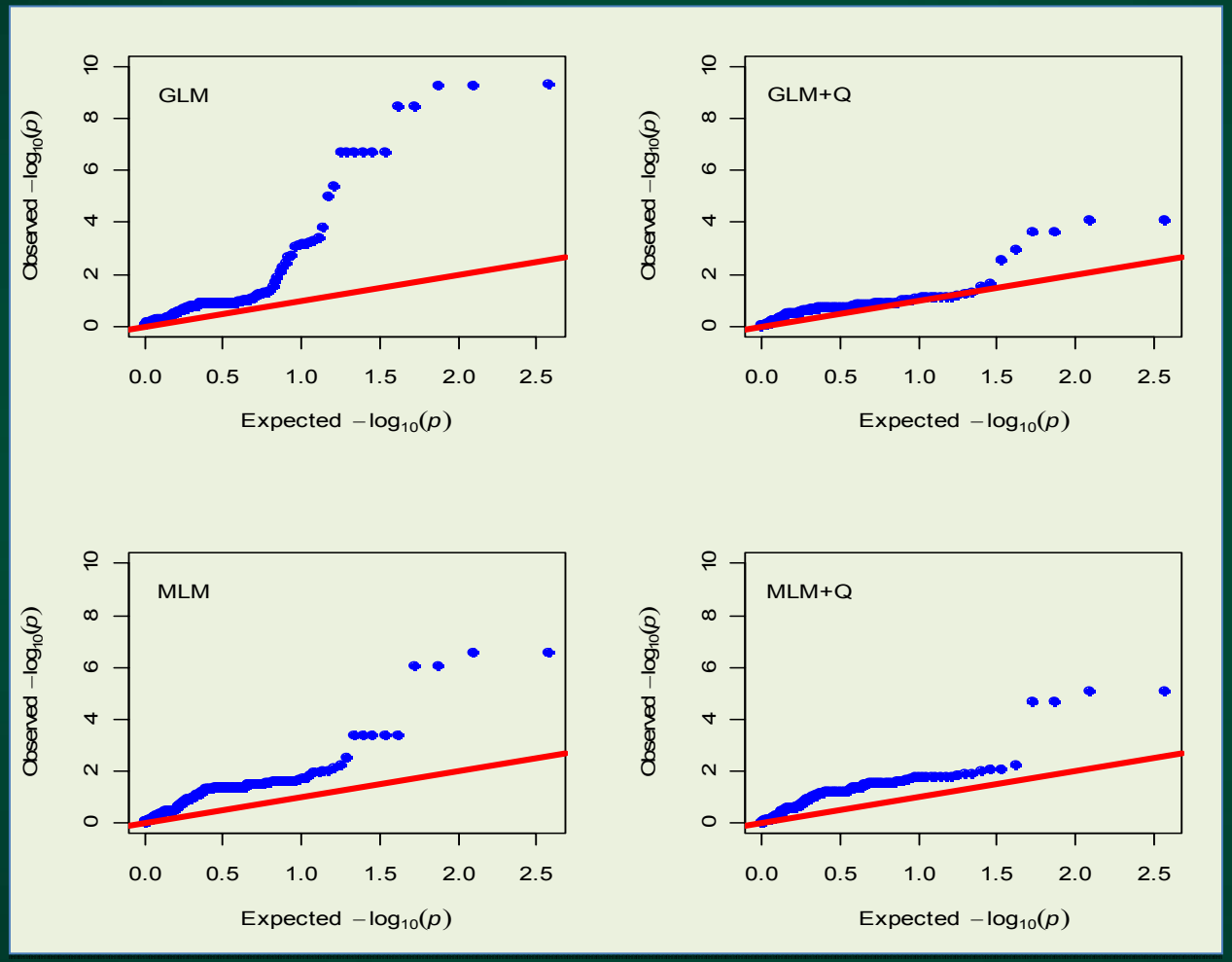

Figure 1. Quantile – quantile plots for both general linear model and mixed linear model for stalk rot association analysis

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Result: stalk rot CGAM

Association of candidate gene tagged SNP markers with stalk rot resistance using GLM+Q model

Candidate gene	Tag SNP	Alleles		Amino acid		3 Locations Mean			Davenport		
		Major	Minor	Major	Minor	*P*-value†	R^2 ‡	Effect §	*P*-value†	R^2 ‡	Effect §
HaABI1-2	HaABI1-2_32	G	C	Valine	Leucine	0.999	0.36	-1.26	1.000	0.38	0.42
	HaABI1-2_155	C	G	Proline	Alanine	0.746	0.70	-3.65	1.000	0.07	1.44
	HaABI1-2_163	G	A	Synonymous	--	0.658	1.40	-1.27	0.900	1.06	-1.75
HaCOI1-1	HaCOI1-1_251	C	A	Asparagine	Lysine	0.009	4.52	-7.20	0.495	2.00	-6.37
	HaCOI1-1_312	A	C	Synonymous	--	0.998	0.43	-2.23	0.379	2.30	4.12
HaCOI1-2	HaCOI1-2_72	G	A	Synonymous	--	0.995	0.57	-1.81	0.999	0.49	-2.03
	HaCOI1-2_408	T	C	Synonymous	--	0.102	2.87	-3.20	0.774	1.39	-1.76
HaDAT3-1	HaDAT3-1_25	G	A	Methionine	Isoleucine	1.000	0.02	-0.27	0.972	0.80	-0.84
HaEIN2-1	HaEIN2-1_250	T	G	Serine	Alanine	0.999	0.46	-1.67	0.867	1.33	-3.36
HaEIN2-2	HaEIN2-2_208	C	T	Synonymous	--	0.999	0.34	0.99	0.979	0.78	-1.37

† *P*-value adjusted for multiple comparisons.
‡ Marker R^2, percent phenotypic variation explained by the marker in the population.
§Effect of major allele assuming minor allele is assigned a value of zero.

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Preliminary Result: head rot CGAM

Association of candidate gene tagged SNP markers with head rot resistance using GLM+Q model

Candidate gene	Tag SNP	Sabin 2011		Staples 2011		Staples 2012		3 Locations mean	
		P-value†	R^2 ‡	*P-value*	R^2	*P-value*	R^2	*P-value*	R^2
HaABI1-2	HaABI1-2_32	0.315	2.96	0.027	5.54	0.672	2.16	0.068	4.59
	HaABI1-2_155	1.000	0.01	1.000	0.06	0.952	0.67	1.000	0.05
	HaABI1-2_163	0.000	7.74	0.176	3.68	0.096	4.31	0.008	6.52
HaCOI1-1	HaCOI1-1_251	0.000	9.33	0.087	4.57	0.247	3.54	0.004	7.28
	HaCOI1-1_312	0.252	3.41	0.756	2.04	0.797	1.97	0.455	2.82
HaCOI1-2	HaCOI1-2_72	1.000	0.40	0.999	0.52	1.000	0.52	1.000	0.17
	HaCOI1-2_408	0.653	2.11	0.999	0.69	0.668	2.23	0.657	2.20
HaDAT3-1	HaDAT3-1_25	0.000	11.50	0.006	7.15	0.035	5.40	0.000	10.16
HaEIN2-1	HaEIN2-1_250	0.997	0.88	0.540	2.71	0.808	2.03	0.808	1.96
HaEIN2-2	HaEIN2-2_208	0.000	10.73	0.037	5.25	0.001	9.07	0.000	11.24

† *P*-value adjusted for multiple comparisons.
‡ Marker R^2, percent phenotypic variation explained by the marker in the population.

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Genome-Wide Association mapping

- ~ 8700 SNP markers used for genotyping the AM population
  - so far we mapped 5017 SNP markers in 17 LG of sunflower
  - markers with more than 20% missing data were removed from analysis
  - non-polymorphic marker data were removed
  - both mapped and non-mapped markers were used in the analysis
- Both TASSEL and ProcGLMselect of SAS were used for analysis
  - *Q* matrix from Structure analysis was used as fixed effect in the model

Preliminary Result: stalk rot GWAM

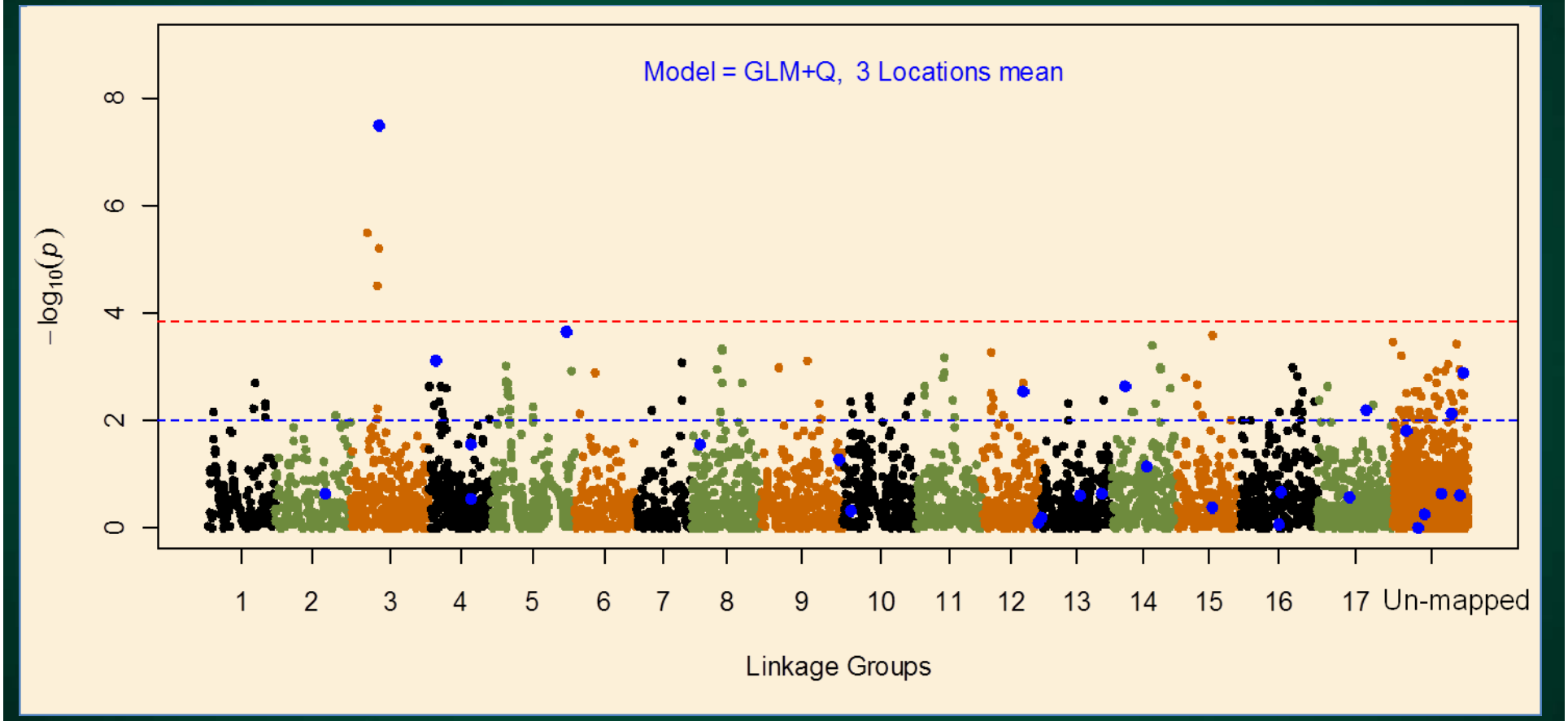

Figure 2. Manhattan plot showing genome-wide *P*-values of the GLM+*Q* model from TASSEL. x axis shows the SNPs along each chromosome; y axis is the −log10 (*P*-value) for the association. Red colored line is the threshold for *P*-value after Bonferroni multiple test correction and the blue line is the threshold for comparisonwise *P*–value. Blue dots indicate significant markers picked by Proc GLMselect analysis in SAS

Summary table of Genome-Wide AM of Sclerotinia stalk rot

Marker	LG	cM	ΔR^2
NSA003891	2	31.736	1.40
NSA005976	3	40.114	9.34
NSA007993	4	22.957	1.91
NSA005541		54.004	2.42
NSA005804		54.004	1.02
NSA006292	5	56.402	2.27
NSA004993	8	38.565	1.81
NSA008453	9	91.991	2.74
NSA004925	10	26.141	1.16
NSA005041	12	35.601	2.81
NSA004719		58.567	0.92
NSA006925		65.238	2.24
NSA007693	13	38.811	0.78
NSA008273		47.040	2.32
NSA005691	14	5.443	1.15
NSA003499		22.377	1.41
NSA002234	15	27.781	2.61
NSA004523		84.847	1.83
NSA006011	16	49.238	4.84
NSA003656		49.716	1.48
NSA006181	17	41.462	2.54
NSA005872		47.154	1.40
NSA001411	Un-mapped	-	3.63
NSA001682		-	0.29
NSA002808		-	0.69
NSA003872		-	1.20
NSA005952		-	0.57
NSA007386		-	0.94
NSA008550		-	0.73
NSA008970		-	0.92
NSA009400		-	0.27
Genetic variance explained by all 31 significant markers =			59.64

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Preliminary Result: Head rot GWAM

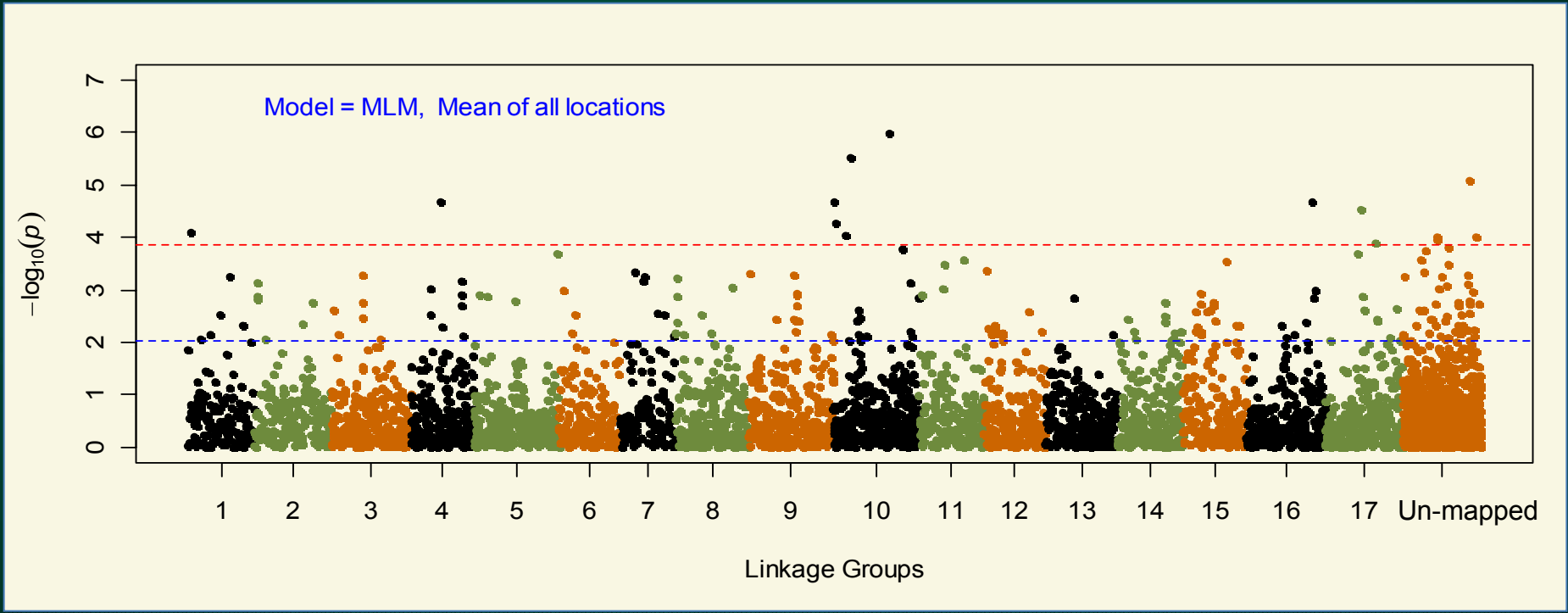

Figure 3. Manhattan plot showing genome-wide *P*-values from the MLM model. x axis shows the SNPs along each chromosome; y axis is the −log10 (*P*-value) for the association. Red colored line is the threshold for *P*-value after Bonferroni multiple test correction and the blue line is the threshold for comparisonwise *P*–value

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Preliminary Result: Head rot GWAM

Summary table of Genome-Wide AM of Sclerotinia head rot

Marker	LG	cM	*P*-value	R^2
NSA006224	1	9.372	8.26E-05	8.00
NSA001362	4	50.009	2.19E-05	9.07
NSA004172	10	1.530	2.07E-05	10.54
NSA007253		1.530	5.53E-05	9.88
NSA007975		39.058	9.28E-05	7.56
NSA004390		42.423	3.04E-06	12.30
NSA002571		45.505	1.08E-06	12.24
NSA004092	16	71.926	2.16E-05	9.52
NSA000436	17	42.329	2.95E-05	8.91
NSA006427		47.154	1.30E-04	7.63
NSA004113	Un-mapped	-	1.02E-04	7.29
NSA004248		-	1.11E-04	7.33
NSA008736		-	8.31E-06	10.58
NSA009919		-	9.58E-05	7.32

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Future Plans

- Complete GW-AM of Stalk rot resistance
- Continue working on association mapping of Head rot resistance and Phomopsis resistance
- Start working on GW-AM for important agronomic and oil traits

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS

Acknowledgement

- National Sclerotinia Initiative
- National Sunflower Association

•The authors thank following individuals for their technical assistance

Angelia Hogness
Leanne Matthiesen
Dana Weiskopf
Megan Ramsett
Nikolay Balbyshev
Arun Jani

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ARS