Theor Appl Genet (2014) 127:193–209
DOI 10.1007/s00122-013-2210-x

ORIGINAL PAPER

Candidate gene association mapping of Sclerotinia stalk rot resistance in sunflower (*Helianthus annuus* L.) uncovers the importance of *COI1* homologs

Zahirul I. Talukder · Brent S. Hulke · Lili Qi · Brian E. Scheffler · Venkatramana Pegadaraju · Kevin McPhee · Thomas J. Gulya

Received: 2 May 2013 / Accepted: 3 October 2013 / Published online: 6 November 2013

Abstract

Key message **Functional markers for Sclerotinia basal stalk rot resistance in sunflower were obtained using gene-level information from the model species *Arabidopsis thaliana*.**

Abstract Sclerotinia stalk rot, caused by *Sclerotinia sclerotiorum*, is one of the most destructive diseases of sunflower (*Helianthus annuus* L.) worldwide. Markers for genes controlling resistance to *S. sclerotiorum* will enable efficient marker-assisted selection (MAS). We sequenced eight candidate genes homologous to *Arabidopsis thaliana* defense genes known to be associated with Sclerotinia disease resistance in a sunflower association mapping population evaluated for Sclerotinia stalk rot resistance. The total candidate gene sequence regions covered a concatenated length of 3,791 bp per individual. A total of 187 polymorphic sites were detected for all candidate gene sequences, 149 of which were single nucleotide polymorphisms (SNPs) and 38 were insertions/deletions. Eight SNPs in the coding regions led to changes in amino acid codons. Linkage disequilibrium decay throughout the candidate gene regions declined on average to an $r^2 = 0.2$ for genetic intervals of 120 bp, but extended up to 350 bp with $r^2 = 0.1$. A general linear model with modification to account for population structure was found the best fitting model for this population and was used for association mapping. Both *HaCOI1-1* and *HaCOI1-2* were found to be strongly associated with Sclerotinia stalk rot resistance and explained 7.4 % of phenotypic variation in this population. These SNP markers associated with Sclerotinia stalk rot resistance can potentially be applied to the selection of favorable genotypes, which will significantly improve the efficiency of MAS during the development of stalk rot resistant cultivars.

Communicated by M. J. Sillanpaa.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2210-x) contains supplementary material, which is available to authorized users.

Z. I. Talukder · K. McPhee
Department of Plant Sciences, North Dakota State University, 166 Loftsgard Hall, Fargo, ND 58108-6050, USA

B. S. Hulke (✉) · L. Qi · T. J. Gulya
USDA-ARS, Northern Crop Science Laboratory, 1307 18th St. N., Fargo, ND 58102, USA
e-mail: brent.hulke@ars.usda.gov

B. E. Scheffler
USDA-ARS, Genomics and Bioinformatics Research Unit, 141 Experiment Station Rd., Stoneville, MS 38776, USA

V. Pegadaraju
BioDiagnostics Inc., 507 Highland Drive, River Falls, WI 54022, USA

Introduction

Sclerotinia sclerotiorum (Lib.) de Bary is an Ascomycete, necrotrophic pathogen with a host range of more than 400 plant species, including sunflower (Boland and Hall 1994). This fungus has been a significant agricultural problem, and causes crop losses of up to $200 million annually in the United States alone (Bolton et al. 2006). *S. sclerotiorum* causes the most damage to sunflower in cool and humid production regions, and it has been reported in all important sunflower producing regions of the world (Masirevic and Gulya 1992). Unlike other hosts, sunflower is vulnerable to infection by *S. sclerotiorum* both via floral and root infection. The fungus overwinters as sclerotia in the soil

or in plant debris. If the sclerotia germinate myceliogenically, it can infect nearby sunflower roots leading to a basal stalk rot and wilt, a disease unique to sunflower. Head rot occurs when sclerotia germinate carpogenically to produce apothecia, and the subsequent airborne ascospores infect senescing florets, similar to other susceptible crops.

The genetics of resistance to head and stalk rot are reasoned to be different based on lack of correlation between stalk rot and head rot resistance in field experiments surveying germplasm collections (Gulya et al. 1989), and each disease is under polygenic control (Castaño et al. 1993; Gentzbittel et al. 1998; Bert et al. 2002, 2004; Fusari et al. 2012). No vertical resistance against *S. sclerotiorum* is known in cultivated sunflower, and all heritable resistance is horizontal or quantitatively inherited (Hahn 2002; Rönicke et al. 2005). A number of Sclerotinia resistant sunflower inbred lines were released in the past decade from the US Department of Agriculture-Agricultural Research Service (USDA-ARS), exploiting this quantitative variation (Miller and Gulya 1999, 2006; Miller et al. 2006). In France, substantial increases in Sclerotinia head rot resistance were reported from a long-term recurrent selection program (Vear et al. 2007).

The advent of high-throughput DNA marker technology greatly facilitates the mapping of quantitative trait loci (QTLs) associated with quantitative disease resistance. Linkage analysis-based QTL mapping has been demonstrated as a successful tool to dissect the genetic bases of complex traits in several important crops since the 1990s (Bernardo 2008; Kearsey and Farquhar 1998). QTL mapping of Sclerotinia head and stalk rot resistance has been reported in several bi-parental sunflower mapping populations in the past decade (Bert et al. 2002, 2004; Gentzbittel et al. 1998; Mestries et al. 1998; Davar et al. 2010; Micic et al. 2004, 2005a, b; Yue et al. 2008). In these studies, the co-segregation of genetic markers and phenotype is identified within families generated by controlled crosses with known ancestry. There are few opportunities for recombination to occur within families and pedigrees, resulting in low mapping resolution (Hall et al. 2010; Zhu et al. 2008). Only QTLs that differ between two parental lines could be detected in each mapping population. In contrast, linkage disequilibrium (LD)-based association mapping (AM) overcomes these limitations, and has received increased attention by plant geneticists as an alternative tool to resolve the genetics of complex traits. Examples include mapping resistance genes for soybean cyst nematode in soybean (Li et al. 2009), late blight in potato (Gebhardt et al. 2004), powdery mildew in barley (Ivandic et al. 2003), Verticillium wilt in potato (Simko et al. 2004a, b), and dieback in lettuce (Simko et al. 2009); as well as flowering time in maize (Thornsberry et al. 2001), in barley (Stracke et al. 2009), and in pearl millet (Saïdou et al. 2009); starch-related traits in maize (Wilson et al. 2004); and anthocyanin biosynthesis in maize (Szalma et al. 2005). AM, initially developed for human linkage studies, exploits historical and evolutionary recombination events at the population level (Mackay and Powell 2007; Nordborg and Tavare 2002), therefore, effectively mining QTL from a greater portion of the variation that exists within a species. Mapping resolution of AM populations composed of a diverse set of germplasm is expected to be higher than bi-parental mapping populations of the same size, because the many recombination events have an effect of shrinking LD (Hall et al. 2010). Mutation, mating system, genetic drift, population admixture, and selection also affect resolution as well (Flint-Garcia et al. 2003).

In contrast to linkage mapping, AM has an advantage that existing germplasm resources, such as germplasm collections and breeding programs, can be used directly. However, linkage mapping has a significant advantage over AM in that population structure and kinship may exist in an AM population, generating spurious marker-trait associations, while this is not a factor in linkage analysis (Speed et al. 2012; Myles et al. 2009; Wang et al. 2005). The importance of population structure or kinship to an AM model is a function of the germplasm included in the study population, and can differ by population within species. Population model generating software, such as STRUCTURE, and genetic distance measures have been applied as covariates to reduce the number of false positive associations due to population structure (Pritchard et al. 2000b). The STRUCTURE method estimates the population structure using a Bayesian clustering method and members of the population are assigned to subpopulations on the basis of multilocus genotype data (Falush et al. 2003; Pritchard et al. 2000a). Accounting for shared common ancestry or kinship among the members of the association mapping population may also be required, particularly for populations derived from breeding programs. Yu et al. (2006) developed a unified mixed AM model and accommodated both the population structure and kinship to control the sources of Type I error in a marker-trait association study in maize.

In AM, a whole genome may be scanned with thousands of markers to identify QTL that are associated with a particular phenotype, or alleles at a few selected functional candidate genes already suspected of being involved in the trait of interest may be tested for association (Rafalski 2002a, b; Zhu et al. 2008). One of the simplest methods of identifying candidate gene markers relies on the resequencing of amplicons from several genetically distinct individuals of an association population (Rafalski 2002b; Zhu et al. 2008). Functional analysis of defense signal pathways in the model species *Arabidopsis thaliana* provides a potential source of candidate genes for studying disease resistance in sunflower and other crops (Guo and Stotz 2007;

Hernandez-Blanco et al. 2007; Tang et al. 2007; Zhou et al. 1999). Recently, 21 candidate genes homologous with *Arabidopsis* and *Brassica napus* genes identified in a plant-Sclerotinia transcript profiling study were evaluated for association with Sclerotinia head rot resistance in sunflower, with some success (Fusari et al. 2012). In their study, a collection of 94 sunflower inbred lines was evaluated and one highly significant association was found between the *HaRIC_B* gene and head rot incidence. While the work of Fusari et al. (2012) deals with Sclerotinia head rot, it does provide us with some certainty that similar methods could be successful for the Sclerotinia stalk rot disease in sunflower, our disease of interest.

In the present study, we have resequenced functional candidate genes in cultivated sunflower that may be associated with resistance to stalk rot caused by *S. sclerotiorum.* Our goal is to identify DNA sequence variation within the candidate genes in the form of SNPs and InDels and find association of these markers with the disease phenotype in sunflower to develop marker-based selection tools for breeding.

Materials and methods

Plant materials and DNA extraction

The candidate gene association mapping (CG-AM) population consists of 260 cultivated sunflower lines. Eleven genotypes of the CG-AM population are USDA-ARS inbred lines developed specifically for Sclerotinia resistance, while the rest of the genotypes are USDA Plant Introductions (PIs) previously untested for Sclerotinia reaction and obtained from the USDA-ARS North Central Regional Plant Introduction Station, Ames, Iowa, USA (USDA 2013). The PI accessions originated from Eastern Europe ($n = 27$), Western Europe ($n = 25$), North America ($n = 19$), South America ($n = 10$), Africa ($n = 10$), Asia ($n = 4$), and one with uncertain origin. Young sunflower leaves from 3-week-old seedlings were collected from replicated field trials, lyophilized, and stored in a −80 °C freezer. Total genomic DNA was extracted from frozen leaf tissue in duplicate with DNeasy 96 Plant Extraction kits (Qiagen, Valencia, USA). A few sunflower PIs did not yield good quality DNA even with repeated extraction, so a modified Qiagen protocol (Horne et al. 2004) was used to obtain good quality DNA.

Assessment of stalk rot resistance

The CG-AM population was evaluated for Sclerotinia stalk rot resistance for two years at three locations. Field trials were conducted at Davenport, North Dakota, and Crookston, Minnesota, USA, in 2008, and in 2009 at Grandin, North Dakota, and Crookston, Minnesota, USA. The CG-AM population was planted in a sets-within-replications experimental design (Hallauer and Miranda 1988) with two replications as complete blocks in each location. Commercial hybrids 'Croplan 305' and 'Cargill 270' were used as resistant and susceptible controls, respectively, in all sets in all stalk rot phenotyping experiments. The seeds were planted in 6 m long, single-row plots with 75 cm row spacing, yielding approximately 25 plants per row. Each trial was artificially inoculated 5–6 weeks after planting when the plants were approximately at the V-6 growth stage, or no more than 45 cm tall. Inoculum consisting of mycelium on millet was produced and applied by the method of Gulya (2004). Plots were evaluated for disease incidence at least twice, with the first evaluation 7–9 weeks after inoculation, and the second evaluation 2 weeks later. Disease incidence was expressed as the percent of plants showing wilting and/or basal stem rot lesions. All plants (maximum 25) were scored to get the disease incidence data for each accession in each replication.

All statistical analyses of the phenotypic data were carried out using SAS 9.2 (SAS Institute 2008). The analysis of variance (ANOVA) was conducted using PROC MIXED. Environments, genotypes, sets, replications, and their interactions were considered random effects. The relative efficiency of the sets-in-replications design over the randomized complete block design (RCBD) was estimated for each of the environments individually. The raw phenotypic trait values of individual genotypes were adjusted for variation among sets within replications following Schutz and Cockerham (1966) and adjusted means were used in the AM analysis.

Candidate gene selection and primer design

Eight *Arabidopsis* genes, *COI1* (Coronatine Insensitive 1), *NPR1* (Nonexpresser of PR genes 1), *EIN2* (Ethylene Insensitive 2), *ABI1* (ABA Insensitive 1), *ABI2* (ABA Insensitive 2), *DET3* (De-Etiolated 3), *PAD3* (Phytoalexin Deficient 3), and *LACS2* (Long-Chain Acyl-CoA Synthetase 2), were selected based on a defense response study in *Arabidopsis* against *Sclerotinia sclerotiorum* (Guimarães and Stotz 2004; Guo and Stotz 2007; Stotz, personal communication, 2008).

Nucleotide sequences of the selected *Arabidopsis thaliana* defense genes were retrieved from the TAIR (*Arabidopsis* Information Resource) database. Each gene sequence was searched against the NCBI sunflower EST database using the nucleotide–nucleotide BLAST (blastn) algorithm (Altschul et al. 1997). Sunflower ESTs with a *e*-value no greater than 10^{-06} were then selected for each gene. No significant hit was found for *Arabidopsis* genes

NPR1 and *PAD3* in the sunflower EST database. The ABI2 protein shows 80 % amino acid sequence identity to ABI1 in *Arabidopsis* and, as expected, common sunflower EST sequences represented both *ABI* genes. The GenBank accession numbers were then cross-referenced with contig assembly sequences in the Compositae Genome Project database (http://cgpdb.ucdavis.edu/). The selected contig sequences for each gene were compared against Genbank *Arabidopsis* sequence using pairwise sequence alignment to visualize gene identity and remove contigs with poor alignments. Phylogenic relationships (orthology, paralogy) of the *Arabidopsis* candidate genes to the sunflower cDNAs were determined using the procedure described by Fusari et al. (2012). The sunflower and Arabidopsis genes, together with GenBank EST data of 50 additional plant species, were translated into logical open reading frames and aligned using the MAFFT routine L-INS-I (Katoh et al. 2002). Phylogenetic relationships were obtained by fitting a Maximum Parsimony model to the protein alignments using the software TNT (Goloboff et al. 2008). Heuristic searches started from 20 Wagner trees and were subjected to tree bisection reconnection (TBR) branch swapping. Node support values were calculated using 100 jackknife replicates. After construction, the gene tree was compared to the known species tree generated through the NCBI Common Taxonomy Tree tool (http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi). Orthologous relationships were inferred for those genes which followed known speciation events, and genes were considered paralogous if they did not follow speciation.

Amplification of candidate genes, DNA sequencing, and SNP discovery

Overlapping primer pairs were designed from contig sequences using the Primer3 software (http://frodo.wi.mit.edu/primer3/) with an average length of 20 nucleotides, and a melting temperature around 60 °C. A list of primers used for amplification of each candidate gene along with corresponding sunflower contig sequence names are presented in Table 1. Primers were used to amplify a portion of the candidate gene genomic regions within a subset of 104 sunflower genotypes from our CG-AM population. Half of these 104 genotypes were selected from genotypes that exhibited the best resistance response in our two year multi-location trials, while the remaining half of the genotypes consisted of the most susceptible lines. The polymerase chain reaction (PCR) amplification was performed using Applied Biosystems® GeneAmp PCR System 9700 (Perkin-Elmer, Norwalk, CT, USA) and a touchdown protocol under the following conditions: initial denaturation of the template DNA at 95 °C for 5 min, 12 cycles of 95 °C for 45 s, annealing at 64–59 °C for 45 s (0.5 °C decrease in each cycle) and 72 °C for 1 min, followed by another 30 cycles of 95 °C for 45 s, 59 °C for 45 s and 72 °C for 1 min, with a final extension of 10 min. PCR was performed in 15 µl volumes containing 10 ng genomic DNA, 1 × PCR buffer, 0.2 mM each dNTP, 0.5 µM of both forward and reverse primers, 2.0 mM $MgCl_2$, 0.5 U JumpStart *Taq* DNA polymerase (Sigma), and 0.6× PVP 40 (Sigma). PCR products were analyzed in 1.5 % agarose gels buffered in 1 × TBE (90 mM Trisborate, 2 mM EDTA) and visualized under UV-light after staining with ethidium bromide (1 µg/ml). PCR products were purified with ExoSAP-IT (USB Corporation, Cleveland, USA) according to the manufacturer's instructions. All DNA fragments were sequenced in both directions using the Applied Biosystems ABI Prism 3730XL DNA sequencer with Big Dye Terminator v3.1 chemistry (Rosenblum et al. 1997). Primers used to amplify PCR fragments were also employed for the sequencing reactions. We aligned bidirectional sequences for each individual separately and visually inspected the chromatograms using BioEdit version 7.0.9 (Hall 1999). Heterozygotes were evident by the presence of a double peak at the SNP site in the sequence chromatogram. Consensus sequences for all lines were aligned using the ClustalW (Thompson et al. 1994) alignment option in the BioEdit program. Polymorphisms appearing in only one or two lines were rechecked on chromatograms to distinguish true polymorphisms from probable polymerase or scoring errors. Direct comparison of assembled cDNA sequence and its corresponding genomic DNA sequences was used to predict exon and intron regions. Sequence polymorphisms were deduced from comparisons of the alignments with each other and the assembled cDNA from the Compositae Genome Project using BioEdit (Hall 1999) and SNiPlay (Dereeper et al. 2011).

Population structure and kinship

Analyses of the population structure and kinship were performed with all 260 genotypes of the CG-AM mapping panel. For structure analysis, we used 136 SNP markers that originate from all 17 linkage groups of the sunflower genome with a minimum of 5 cM separation, which were selected from a panel of more than 8700 SNP markers developed by National Sunflower Association (NSA) SNP Consortium (Mandan, North Dakota, USA). The software STRUCTURE version 2.3.3 (Pritchard et al. 2000a) was used to determine the presence of population structure and assign sunflower lines to subpopulations. STRUCTURE was run for $K = 1$–10 clusters using the ancestry model that allows population admixture with no information a priori. Five replications were performed for each K value. Each run consisted of a burn-in period of 100,000 iterations followed by 200,000 Markov Chain Monte Carlo iterations,

Table 1 *Arabidopsis thaliana* defense-related genes, gene functions, corresponding sunflower contigs retrieved from Compositae Genome Project database (CGPDB), sunflower candidate genes, and PCR primers for amplification and sequencing

Arabidopsis genes	TAIR accession number[a]	Gene functions	Contig in CGPDB	Sunflower candidate genes[b]	Primers (5′–3′)
ABI1 and ABI2	AT4G26080 AT5G57050	Involved in abscisic acid (ABA) signal transduction. Negative regulator of ABA promotion of stomatal closure	HELI_ANNU.CSA1.3103	*HaABI1-1*	F-CCACGGCTGTGGTGGCGTTA R-GTATAACCTTGCCCCCTGCT
			HELI_ANNU.CSA1.831	*HaABI1-2*	F-CCTGATCCGGAAGTCACATT R-CGCTAGATCCGGCCCACCGA
COI1	AT2G39940	Required for wound and jasmonate -induced transcriptional regulation. Encodes a protein containing Leu-rich repeats and a degenerate F-box motif	HELI_ANNU.CSA1.2994	*HaCOI1-1*	F-CCGATTTGCCACTGGATAAC R-ACACGCTGGATAGTCGTTCC
			HELI_ANNU.CSA1.2945	*HaCOI1-2*	F-ATTTCTTTAGCGCGAGGTTG R-ATTCGCAGGTGGGTGTAGAG
DET3	AT1G12840	Encodes subunit C of the vacuolar H^+–ATPase (V-ATPase). V-ATPases constitute a family of highly conserved ATP-dependent proton pumps responsible for acidification of endomembrane compartments in eukaryotic cells	HELI_ANNU.CSA1.503	*HaDET3-1*	F-GATACCAGTTCCGGTGATCG R-GGTTACACCAAGGGTGATGG
EIN2	AT5G03280	Encodes a protein with similarities to Nramp metal-ion transporters that is a central regulator of ET signaling	HELI_ANNU.CSA1.189	*HaEIN2-1*	F-GCCCGAATGCGGGGAAGGTT R- CGGGAACGCCACGTCACCAG
			HELI_ANNU.CSA1.4831	*HaEIN2-2*	F-GCCGAAAAAGCGAATAACAG R-CCGTTCATCTCCTTGCTCTC
LACS2	AT1G49430	Encodes a long-chain acyl-CoA synthetase that catalyzes the synthesis of omega-hydroxy fatty acyl-CoA intermediates in the pathway to cutin biosynthesis	HELI_ANNU.CSA1.6756	*HaLACS2-1*	F-GGAATGGGTTATGAGCATGG R- CCAGCTGCGGGAAATGTT

[a] TAIR: The Arabidopsis Information Resource, available at: http://arabidopsis.org/

[b] Candidate genes names were given with an *Ha* prefix and a numeric suffix indicating different *Helianthus* homologs of the original *Arabidopsis* gene

assuming that allele frequencies are correlated across clusters (Falush et al. 2003). To determine the most probable K, three factors were considered. The ΔK statistic (Evanno et al. 2005) a genetic distance dendrogram, and germplasm passport data, and breeding records (USDA 2013) were compared to determine a sensible number of K. The resulting Q matrix was visualized with the software DISTRUCT version 1.1 (Rosenberg 2002). The relative kinship matrix (K) (Loiselle et al. 1995) was estimated in the SPAGeDi version 1.3a software (Hardy and Vekemans 2002) using 5,244 highly polymorphic SNP markers with <5 % missing data from NSA-SNP Consortium panel. Negative kinship values between individuals were set to 0, as this indicates there was less relationship than that expected between two random individuals (Hardy and Vekemans 2002). Both the Q matrix from structure analysis and K matrix from kinship analysis were considered for inclusion as covariates in the AM analyses.

Linkage disequilibrium and SNP marker complexity reduction

Linkage disequilibrium (LD) between all SNPs and InDels within each candidate gene was measured in terms of r^2 (Weir 1996) in HaploView version 4.2 (Barrett et al. 2005). The decay of LD with physical distance was estimated using a nonlinear regression analysis of LD between polymorphic sites (r^2) versus the genetic distance between sites in base pairs as suggested by Kolkman et al. (2007) for sunflower (Remington et al. 2001). The expected value of r^2 under drift–recombination equilibrium is $E\left(r^2\right) = 1/(1+C)$, where $C = 4Nc$, N is the effective population size and c is the recombination fraction between sites (Hill and Weir 1988). Under the assumption of a low level of mutation and an adjustment of sample size n, the expectation (Hill and Weir 1988) becomes:

$$E\left(r^2\right) = \left[\frac{10+C}{(2+C)(11+C)}\right]\left[1+\frac{(3+C)(12+12C+C^2}{n(2+C)(11+C)}\right].$$

We pooled our data across candidate genes and fit this equation into a nonlinear regression model using the statistical package R version 2.13.1 (R Development Core Team 2011). We also measured the rate of LD decay individually for six candidate genes.

Before AM analysis, the number of independent markers was determined based on the output of the Tagger function within HaploView, with a threshold of $r^2 < 0.8$ for independence. Complexity of correlated markers was reduced by selecting one marker from each group that directly caused a change to the putative protein sequence, when it was possible to do so. AM analyses were then carried out with TASSEL software version 3.0 (Bradbury et al. 2007).

We employed four different models for AM analysis. First, the AM analysis was performed using the general linear model (GLM) function in TASSEL with and without the Q matrix as a covariate. Next, two mixed linear model (MLM) analyses were performed (Yu et al. 2006). One method included the K matrix as a random effect, and a second included both the K matrix as a random effect and the Q matrix as a fixed effect. Quantile–quantile plots of the raw P values were constructed using the R statistical package to determine the best fitting model (R Development Core Team 2011). The P values were corrected for multiple tests using the permutation algorithm within TASSEL, which was performed for 10,000 iterations. Based on these results, we selected ten SNP markers that showed association in at least one test at a relaxed $P \leq 0.40$ threshold in our 104 accession subset.

SNP genotyping assay and association mapping analysis

Genotyping of the remaining 156 sunflower lines with the ten selected markers from the preliminary analysis was performed using the KASP (KBiosciences Competitive Allele-Specific PCR, KBiosciences, UK) genotyping system. Allele specific primers were designed from flanking SNP sequences for each selected SNP using proprietary protocols. The full dataset of ten selected markers in all 260 lines were combined and tested using the best fitting AM model, GLM + Q. The P values were corrected for multiple tests using the permutation algorithm within TASSEL, which was performed for 10,000 iterations.

Results

Sclerotinia stalk rot resistance of 260 Plant Introductions and inbred lines

Stalk rot in sunflower is a complex trait controlled by both genetic and environmental factors requiring extensive phenotypic scoring. Therefore, each entry was scored for stalk rot incidence in four environments in replicated field trials. As expected with entries of diverse origin, the analysis of variance (ANOVA) revealed significant variation among genotypes for stalk rot resistance in all environments (data not shown). The error variance of the Davenport location data was high and not comparable to the other three locations. We kept Davenport data separate, while the data for the other three environments were combined in a multi-environment analysis. This combined analysis showed that the effect of individual environment was non-significant. However, there was a significant genotype by environment interaction effect ($P \leq 0.05$). Much of this may be explained by differences in the broadness of the phenotypic distribution

between locations and its effect on relative performance among the genotypes, but changes in rank are also present, indicating the importance of multienvironment data.

In general, the stalk rot incidence for the AM population was higher in 2008 than in 2009 (USDA 2013). In 2008, the stalk rot incidence ranged from 3 to 100 % in the Crookston location with an average of 57 %, while for the Davenport location, the mean was 50 %, and the range ran from 0 to 100 %. In contrast, the stalk rot incidence in Crookston during 2009 season ranged from 0 to 84 % with an average of 23 %, while the average disease incidence was 32 % in Grandin in 2009 and ranged from 1 to 88 %.

Sequence polymorphism, SNPs and InDels for candidate genes

A total of seven sunflower candidate genes homologous to one of six *A. thaliana* defense response genes and one 5` UTR region were sequenced in a set of 104 cultivated PIs and inbred lines (Table 2). Half the genotypes of this set (52) were sampled from the best resistant lines (average disease incidence ranges from 4.7 to 14.5 %) based on the phenotypic evaluation in our field study, while the other half of this set were sampled from the most susceptible lines (average disease incidence ranges from 45 to 83 %) (USDA 2013). The candidate gene sequence regions analyzed covered a concatenated length of 3,791 bp per individual. The fragment size of each gene ranged from 189 to 808 bp, having 3–85 polymorphic sites. Three of the candidate gene sequences possessed only exon sequence, while the remainder included both exon and intron sequence. All of the candidate genes were given unique names with an *Ha* prefix and a numeric suffix indicating different *Helianthus* homologs of the original *Arabidopsis* gene.

A total of 187 polymorphisms were detected for eight candidate gene sequences, an average of one polymorphism in every 20 bp sequence. Among the polymorphic sites, 149 were SNPs and 38 were InDels. The InDels varied in length from 1 to 34 bp, and were located in non-coding regions. Coding regions of the candidate genes included 54 SNPs with a frequency of 1 SNP per 43 bp, while the non-coding regions included 95 SNPs with a frequency of 1 SNP per 15 bp. Eight of the 54 SNPs in the coding regions led to changes in the putative amino acid sequence. We filtered out highly correlated SNPs from each candidate gene using the Tagger function implemented within HaploView. A total of 61 tag SNPs explained most of the variation in polymorphic sites identified for all candidate genes.

Linkage disequilibrium

We investigated LD between pairs of segregating sites to better understand LD in sunflower genes and the probability of correlation of CG-SNPs with other adjacent genes. Three strong LD blocks were observed in the eight candidate gene loci sequenced, one in *HaABI1-1* and two in *HaDET3-1* (Supplemental Figure 1). The strong LD block in *HaABI1-1* was in the interval from SNP 'HaABI1-1_141' to SNP 'HaABI1-1_159' spanning 19 bp (mean r^2 within LD block = 0.85). In *HaDET3-1*, two strong LD blocks were in the interval from SNP 'HaDET3-1_227' to SNP 'HaDET3-1_244' spanning 18 bp (mean r^2 within LD block = 1.0), and one from SNP 'HaDET3-1_292' to SNP 'HaDET3-1_317i', spanning 30 bp (mean r^2 within LD block = 1.0).

Data from all eight candidate gene loci were pooled to estimate the overall decay of LD which is shown by plots of r^2 as a function of physical distance in bp between the

Table 2 Sequence length, number and distribution of polymorphic sites, SNPs, InDels and SNP frequency throughout coding, non-coding and overall entire sequenced regions of eight sunflower candidate genes

Candidate genes	Total sequence				Coding region					Non-coding region		
	Length (bp)	Polymorphic site	SNPs	InDels	Length/total predicted exon length (bp)	SNPs	Synonymous substitution	Non-synonymous substitution	InDels	Length (bp)	SNPs	InDels
HaABI1-1	279	16	12	4	171/1,065	1	1	0	0	108	11	4
HaABI1-2	189	5	5	0	189/639	5	1	4	0	0	–	–
HaCOI1-1	668	24	23	1	512/678	18	17	1	0	156	5	1
HaCOI1-2	660	12	12	0	660/660	12	11	1	0	0	–	–
HaDET3-1	808	85	61	24	0/771	–	–	–	–	808	61	24
HaEIN2-1	483	18	15	3	403/513	13	11	2	0	80	2	3
HaEIN2-2	239	3	3	0	239/1,158	3	3	0	0	0	–	–
HaLACS2-1	465	24	18	6	174/1,044	2	2	0	0	291	16	6
Total	3,791	187	149	38	2,348/6,528	54	46	8	0	1,443	95	38

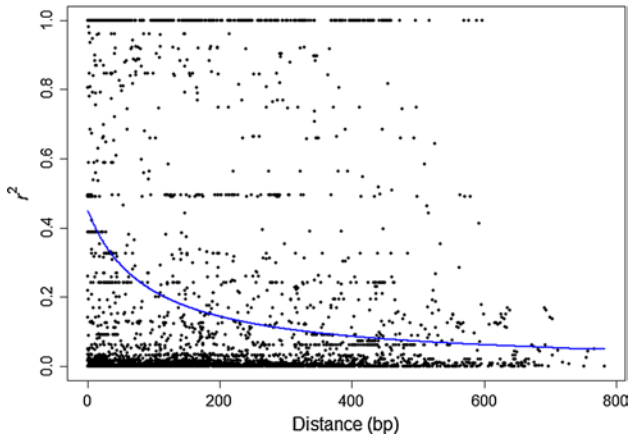

Fig. 1 Linkage disequilibrium (LD) decay of sunflower stalk rot resistance candidate genes. LD was measured by squared correlations of allele frequency (r^2) values against genetic distance (bp) between pairs of polymorphic sites within candidate gene sequences. The trendline is the expected decline in linkage disequilibrium based on a nonlinear regression of r^2 against distance, using a mutation–recombination–drift model (Hill and Weir 1988), which explained $R^2 = 0.058$ or 5.8 % of the total variation

SNPs (Fig. 1). The average LD throughout the candidate gene loci decayed to $r^2 = 0.2$ at 120 bp and $r^2 = 0.1$ at 350 bp.

Population structure and kinship

In our association mapping panel, population structure was assessed using a model-based Bayesian clustering method for determining the number of clusters, *K*. Analysis revealed a steadily increasing curve for Ln P(D) without showing any clear plateau, although a slight tendency of plateauing was observed between runs $K = 4$ and 5 (Fig. 2a). Thus, it was difficult to deduce the optimal number of clusters based on the values of Ln P(D) probability. Pritchard et al. (2000a) state "the problem of inferring the number of clusters, *K*, present in a data set is notoriously difficult," and further, "clusters may not necessarily correspond to 'real' populations", so "there are biological reasons to be careful interpreting *K*". Evanno et al. (2005) suggested the summary statistic ΔK based on the rate of change of the estimated likelihood between successive *K*

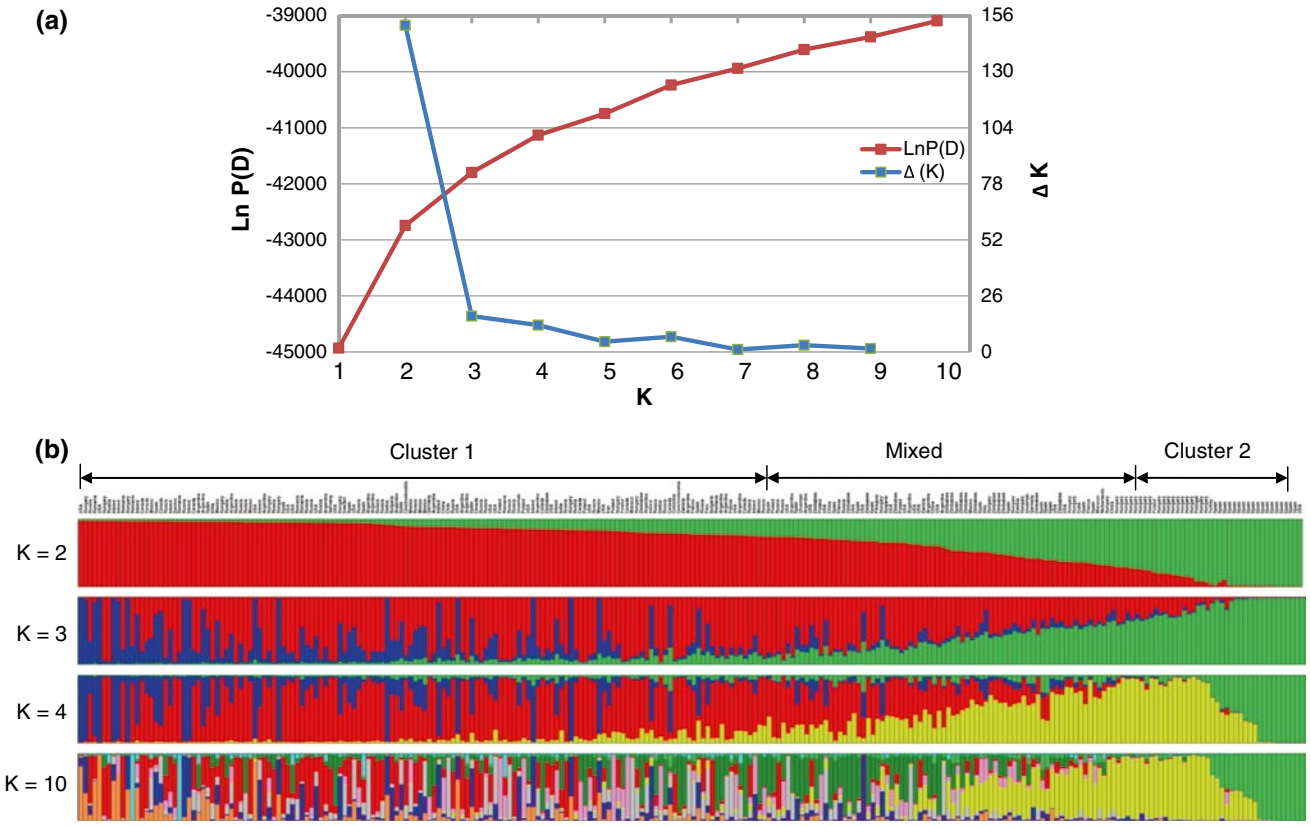

Fig. 2 Estimation of number of subpopulations in the candidate gene association mapping population. **a** The log probability, Ln P(D) for each value of *K* ($K = 1$–10) averaged over 5 runs of STRUCTURE analysis with 100,000 burn-in steps and 200,000 simulation steps is plotted with the ΔK values for each of the successive *K* runs. **b** *Bar plot* of the STRUCTURE analysis. Each of the 260 genotypes is represented by a *vertical bar*, which is partitioned into *K* colored segments that represent the individual's estimated membership to the *K* clusters. The separation of cluster 1 and cluster 2 was done by the threshold for membership coefficient, $Q \geq 0.75$

values provided by STRUCTURE to find a salient break in the slope of the distribution of the Ln P(D) probability function at the true value K. When plotted against respective K values, the ΔK attained a clear maximum at $K = 2$ (Fig. 2a), suggesting that a true subdivision was found at $K = 2$ (Fig. 2b). Since there was some indication that $K = 3$ and $K = 4$ might be valid as well, we compared a genetic distance dendrogram of the association population to their subdivisions based on $K = 2$, 3, and 4 clusters. Clusters were clearly separated by genetic distance on the dendrogram for $K = 2$, but pure individuals of the $K = 3$ and 4 clusters were scattered across the dendrogram and did not follow a genetic distance trend (data not shown). Some of the USDA released inbred lines were shown to be pure individuals in the $K = 3$ and 4 models, and the others admixed, but these designations did not make sense in terms of the pedigrees of these lines. The non-sensical splitting of clusters is a symptom of model overfitting, as described by Pritchard et al. (2000a) in their examples. Therefore, we determined that $K = 2$ is the correct number K for our Q matrix.

Pairwise kinship coefficients were calculated following the procedures of Loiselle et al. (1995). The kinship analysis indicates complex ancestral relationships among the 260 sunflower lines. The distribution of the relative kinship estimates showed that 58.4 % of the pairwise kinship estimates were equal to or less than 0, indicating that the lines were unrelated. The remaining kinship estimates ranged from 0.05 to 0.5, with a continuously decreasing number of pairs falling in higher estimate categories (data not shown). Sunflower genotypes in cluster 2 of the STRUCTURE results, which is comprises 19 Spanish lines, have a high mean relative kinship value of 0.467, while the 11 USDA-ARS released inbred lines have only a mean relative kinship value of 0.175.

Association mapping

AM analysis was performed in two stages. First, we analyzed the marker-trait association with SNP/InDels identified directly from resequenced genes in the subset of 104 lines. Liberal thresholds ($P \leq 0.40$ after multiple test corrections) and use of the GLM model without covariates were chosen to decrease the likelihood of Type II error during marker number reduction (Bernardo 2010). This preliminary analysis revealed a total of 27 SNP/InDel markers of the six candidate genes *HaCOI1-1*, *HaCOI1-2*, *HaABI1-2*, *HaEIN2-1*, *HaEIN2-2*, and *HaDET3-1* significantly associated with stalk rot resistance, which ultimately reduced down to 10 independent or tag SNPs. Among them, three were from *HaABI1-2*, two each were from *HaCOI1-1* and *HaCOI1-2*, and one each were from *HaEIN2-1, HaEIN2-2* and *HaDET3-1* candidate genes. To identify the best fitting AM analysis model for this population, quantile–quantile plots of estimated $-\log10\ (P)$ were constructed from the four AM models (Fig. 3). The observed P values from the GLM and MLM + K model greatly deviated from the expected P value, assuming a normal distribution for P. The observed P values from the GLM + Q model and the MLM + Q + K model were close to the expected P values indicating a clear influence of population structure on the association analysis. To further analyze model fit of the GLM + Q and the MLM + Q + K models, we examined two competing model comparison statistics: the Bayesian Information Criterion (BIC) in Proc Mixed (SAS Institute 2008), and posterior predictive loss using the method of Waldmann et al. (2008) and Gelfand and Ghosh (1998) in Proc MCMC (SAS Institute 2008). Both of these statistics suggested that the MLM + Q + K model overfit the data and that GLM + Q is the preferred, parsimonious model for our final analysis.

Second, primers were designed for the 10 tag SNPs to genotype the remaining 156 germplasm lines of the AM population and to analyze the complete set of 260 lines with the GLM + Q AM model. The most significant polymorphism was found at SNP position HaCOI1-1_251 (Table 3). This SNP caused an amino acid substitution of Asparagine to Lysine, which could be a functional polymorphism associated with the resistance mechanism, because it is predicted to result in a change of the charge, pH, size, and hydrophobicity at that amino acid site. All of the other SNPs in high LD with HaCOI1-1_251 are synonymous or beyond the translated region. The other significant marker, HaCOI1-2_408 of the *HaCOI1-2* candidate gene, did not result in an amino acid change, but may be in LD with a nearby mutation that does result in an important amino acid change or may have a regulatory effect on the gene. Phylogenetic analysis indicated that the *HaCOI1-1* gene is orthologous to *COI1* in *Arabidopsis thaliana*, while *HaCOI1-2* is paralogous to the *Arabidopsis* gene (Supplemental Figure 3). The consensus DNA sequence around both SNPs is available in Supplemental Table 1.

Discussion

Candidate gene selection

Candidate gene AM is a hypothesis-driven approach for dissecting complex traits that directly tests individual genes for association with a phenotype (Neale and Savolainen 2004). The gene can be either a positional candidate that results from a prior linkage study, or a functional candidate that is based on homology with a gene of known function in a model species (Balding 2006; Zhu et al. 2008). In the present study, we selected *S. sclerotiorum* disease

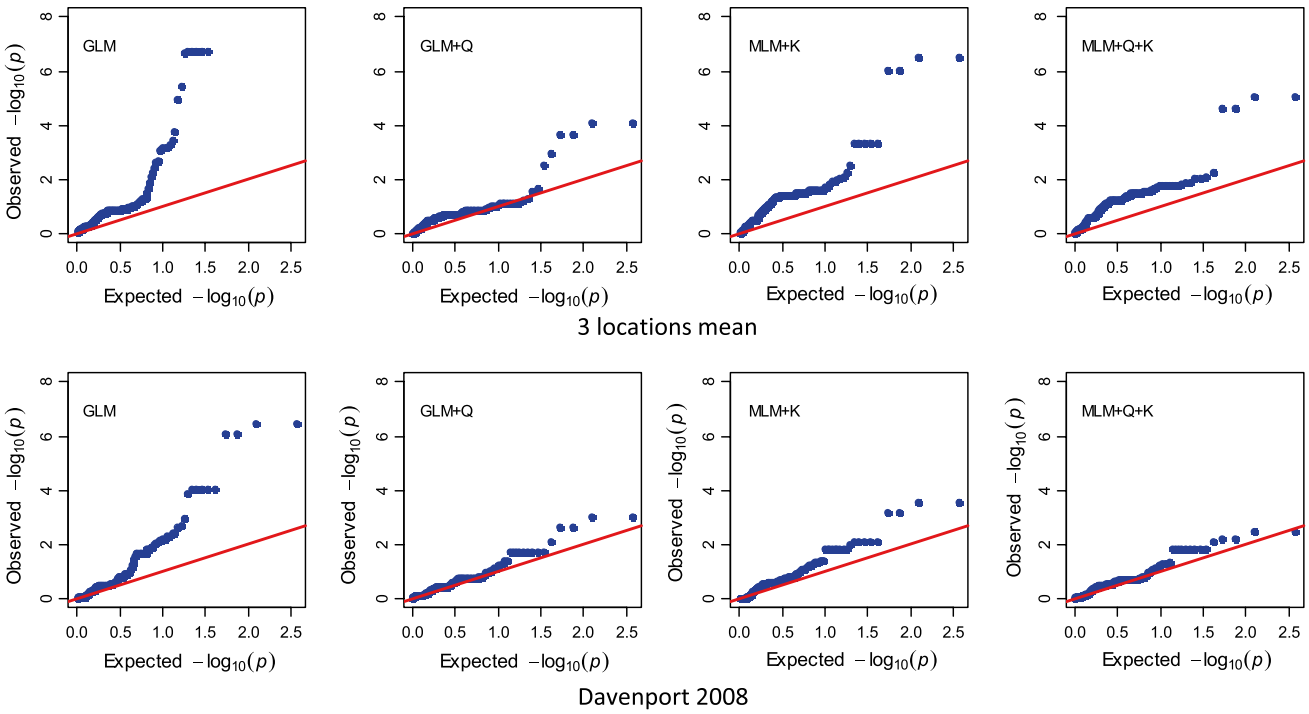

Fig. 3 Quantile–quantile plots of estimated $-\log_{10}(p)$ from four different association analysis models, GLM, GLM + *Q*, MLM + *K* and MLM + *Q* + *K* involving 104 sunflower genotypes. The upper panel is for 3 locations combined data and the *lower panel* is for the Davenport 2008 data. The *red lines* represent the expected values under the null distribution and the *blue dots* represent the observed *P* values (color figure online)

Table 3 Association of candidate gene SNP markers and stalk rot resistance in a sunflower AM population (260 genotypes) using a GLM + *Q* statistical model

Candidate gene	Tag SNP	Alleles		Amino acid		3 Environment mean		Effect[c] estimate	Davenport 2008		Effect[c] estimate
		Major	Minor	Major	Minor	*P*-value[a]	R^{2b}		*P*-value[a]	R^{2b}	
HaABI1-2	HaABI1-2_32	G	C	Glycine	Alanine	0.999	0.36	−1.26	1.000	0.38	0.42
	HaABI1-2_155	C	G	Alanine	Glycine	0.746	0.70	−3.65	1.000	0.07	1.44
	HaABI1-2_ 163	G	A	Glycine	Arginine	0.658	1.40	−1.27	0.900	1.06	−1.75
HaCOI1-1	HaCOI1-1_251	C	A	Asparagine	Lysine	0.009[d]	4.52	−7.20	0.495	2.00	−6.37
	HaCOI1-1_312	A	C	Synonymous (Arginine)		0.998	0.43	−2.23	0.379	2.30	−4.12
HaCOI1-2	HaCOI1-2_72	G	A	Synonymous (Alanine)		0.995	0.57	−1.81	0.999	0.49	−2.03
	HaCOI1-2_408	T	C	Synonymous (Serine)		0.102[d]	2.87	−3.20	0.774	1.39	−1.76
HaDET3-1	HaDET3-1_25	G	A	5′ UTR		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	0.876	1.33	−3.36
HaEIN2-2	HaEIN2-2_208	C	T	Synonymous (Leucine)		0.999	0.34	0.99	0.979	0.78	−1.37

[a] *P* value adjusted for multiple comparisons using the permutation method in TASSEL with 10,000 replications

[b] Marker R^2, percent phenotypic variation explained by the marker in the population

[c] Effect of major allele assuming minor allele is assigned a value of zero

[d] The raw *P* values for HACOI1-1_251 and HACOI1-2_408 were 0.00091 and 0.0105, respectively

resistance candidate genes from well-characterized metabolic pathways and prior evidence in the control of the phenotype. A comprehensive study involving *Arabidopsis thaliana* mutants showed that jasmonic acid (JA), salicylic acid (SA), and ethylene (ET) hormone signaling pathways are involved in defense against *S. sclerotiorum* (Guo and

Stotz 2007). Knockout mutants of the jasmonate receptor *coi1*, and central regulators of SA and ET signaling, *npr1* and *ein2*, respectively, were all hypersusceptible to *S. sclerotiorum*. Thus, defense against *S. sclerotiorum* in *Arabidopsis* is implicated by the activation of these three defense pathways (Guo and Stotz 2007). *ABI* genes are involved in abscisic acid (ABA) biosynthesis, a regulator of stomatal opening apparently required for optimal fungal colonization during infection. The ABA-insensitive *abi1* and *abi2* *Arabidopsis* mutants consistently showed more susceptibility to oxalate deficient *S. sclerotiorum* than wild-type plants (Guimarães and Stotz 2004). *Arabidopsis LACS2* gene encodes a long-chain acyl-CoA synthetase that is involved in the pathway to cutin biosynthesis (Schnurr et al. 2004). Tang et al. (2007) showed that mutations of *LACS2* make *Arabidopsis* more resistant to a virulent strain of the necrotrophic fungus *Botrytis cinerea*.

Lignin biosynthesis is not only important for plant support and water transport, but also for plant defense (Hammond-Kosack and Jones 1996).The *Arabidopsis DET3* gene encodes the C-subunit of the vacuolar H^+-ATPase (V-ATPase) that controls cell elongation and regulation of meristem activity (Schumacher et al. 1999). The *det3* mutation resulted in dark grown seedlings with morphological characteristics of light-grown seedlings (Cabrera y Poch et al. 1993). Later it was found that the *DET3* knockout leads to ectopical lignin synthesis (lignin deposited in cells other than xylem vessels and sclerified parenchyma) and defense-related gene expression (Caño-Delgado et al. 2003; Newman et al. 2004). The *Arabidopsis* gene *PAD3 e*ncodes an enzyme required for biosynthesis of camalexin, a low molecular weight phytoalexin involved in resistance to fungal pathogens (Schuhegger et al. 2006; Zhou et al. 1999). *Arabidopsis* mutant lines deficient in camalexin, indole, or aliphatic glucosinolate biosynthesis showed hypersusceptibility to *S. sclerotiorum* (Stotz et al. 2011a).

While several candidate genes were addressed in this paper, our work does not take into consideration all of the possible candidate genes that have been proposed for resistance to necrotrophic pathogens, any of which may be appropriate candidate genes for Sclerotinia stalk or head rot resistance (reviewed in Laluk and Mengiste 2010). Not surprisingly, a recent paper by Fusari et al. (2012) on Sclerotinia head rot resistance mapping studied a completely different list of candidate genes than our study, in a completely different population of sunflower genotypes derived from their breeding program, at the same time our work was being conducted on the US sunflower germplasm collection. Although sunflower line reactions to head rot and stalk rot have been shown to be different, suggesting different underlying genetics (Gulya et al. 1989), it is impossible to determine whether or not resistance genes are mutually exclusive for the two diseases until additional candidate gene work is conducted by the sunflower community, with the goal of testing most or all logical candidate genes against both stalk rot and head rot data sets. Complete elucidation of all resistance genes involved in one or both of these diseases is a long-term goal, and will require the attention of multiple investigators.

Linkage disequilibrium

The pattern of LD within a species is critically important to the design and success of association mapping studies. The number of marker-trait association tests may be reduced considerably using a minimal set of SNPs (tag-SNPs) that are not in high LD to each other. Levels of LD decay varied between loci, persisting further along the resequenced portions of *HaABI1-1*, *HaCOI1-1*, *HaCOI1-2,* and *HaLACS2-1* than was observed in the *HaDET3-1* and *HaEIN2-1* loci (Supplemental Figure 2). Varying levels of nucleotide diversity and LD decay have previously been observed between maize genes (Andersen et al. 2008) which could reflect different levels of selection pressure put on individual loci during domestication. The trendline of LD across these eight candidate loci and the overall population decayed to $r^2 = 0.1$ at about 350 bp (Fig. 1). Previous studies in sunflower observed the LD decayed to $r^2 = 0.1$ within 200 bp in wild populations and about 1,100 bp in exotic OP cultivars (Liu and Burke 2006). However, a much slower LD decay was reported for elite inbred sunflower lines, to $r^2 = 0.32$ within 550 kbp (Kolkman et al. 2007) and $r^2 = 0.1$ within 100 kbp (Fusari et al. 2008). The discrepancy of LD decay reports in earlier studies in sunflower with our study could be explained by the size and population history of the study materials used, as well as the genes selected for study. LD decay is only valid relative to the germplasm in which it is investigated, and rates of LD decay decrease as the germplasm diversity broadens (Caldwell et al. 2006; Morrell et al. 2005). All the previous studies involved only a few sunflower genotypes. Smaller sample sizes tend to show increased LD distance and, therefore, require caution while declaring LD decay (Yan et al. 2009).

Association mapping

In our CG-AM population, the stalk rot disease scores showed a distribution that shows some correlation to the underlying population structure. A model consisting of population structure alone explains about 22 % of the phenotypic variation. The genotypes in cluster 1 include many resistant lines (mean disease score is 33, $n = 148$ lines), while cluster 2 contains more susceptible lines (mean disease score is 53, $n = 34$ lines). Again, the kinship estimate of the nineteen Spanish lines in cluster 2 was found to be

relatively high, but this was not generally so in cluster 1, indicating that there was more relatedness among lines in the small second cluster than the general population. Use of a mixed model strategy (Kang et al. 2008; Yu et al. 2006; Zhao et al. 2007) allowed us to consider different levels of relatedness in the AM analysis. Although this approach effectively decreases Type I error, it can mask true associations that are strongly correlated with population structure and kinship, increasing Type II error, so model fit must be carefully considered (Zhao et al. 2007, 2011). In the present study, we explored the impact of population structure (GLM + Q model), kinship coefficients (MLM + K model), or both (MLM + Q + K model) in the association models. Addition of the Q matrix resulted in a reduction of the number of marker-trait associations in both GLM + Q and MLM + Q + K models, indicating that covariates were necessary to reduce false positive associations. This was confirmed by quantile–quantile plots depicted for the four AM analysis models (Fig. 3). Further analysis of the GLM + Q and MLM + Q + K models suggested that addition of K actually overfits the model, resulting in greater deviation of the posterior predictive loss and BIC statistics from zero. These results indicate a low level of kinship effect on association analysis in this population, which is likely due to the broad range of origin of the materials. The GLM + Q model appears to be the best fit for this population.

Jasmonic acid (JA) and its derivates have long been known to have a crucial role in plant defense against biotic and abiotic challenges (Guo and Stotz 2007; Ralhan et al. 2012; Stotz et al. 2011b; Xu et al. 2002). The *Arabidopsis thaliana COI1* gene encodes a protein containing leucine-rich repeats and a degenerate F-box motif which physically associates with AtCUL1, AtRbx1, and either of the *Arabidopsis* Skp1-like proteins ASK1 or ASK2 to assemble ubiquitin-ligase complexes, SCF(COI1) (Xu et al. 2002). The SCF(COI1) complex is essential for the JA-signaling pathway. Functional mutants of *COI1* are insensitive or have reduced sensitivity to JA (Xie et al. 1998). A single amino acid substitution in the F-box motif of COI1 abolishes the formation of the SCF(COI1) complexes and results in loss of the JA response (Xu et al. 2002). One specific mutant, *coi1-1* in *Arabidopsis*, results in inviable pollen, as well as reduced defense against insect and pathogens (Xie et al. 1998). This result was repeated in the work of Guo and Stotz (2007) with the *Sclerotinia-Arabidopsis* pathosystem. The *HaCOI1-1* and *HaCOI1-2* candidate genes in sunflower are orthologous and paralogous, respectively, to the *Arabidopsis thaliana COI1* gene and have been genetically mapped at 50.9 and 7.0 cM distance, respectively, in linkage group 14 of the Tang et al. (2002) reference map (Bowers et al. 2012). The significant SNP at HaCOI1-1_251 is interesting because it results in a major amino acid change just beyond the final conserved leucine-rich repeat, as defined by Xie et al. (1998). This leucine-rich repeat in *Arabidopsis* is the only one abolished by the *coi1-1* mutant for *Sclerotinia* susceptibility analyzed by Guo and Stotz (2007), underlining the importance of this region to resistance (Fig. 4). The substitution to asparagine due to the HaCOI1-1_251 mutation results in greater Sclerotinia stalk rot resistance (Table 3) and is a functionally different residue in size, pH, charge, and hydrophobicity than either the lysine residue in the minor sunflower allele or the arginine residue of the *COI1* predicted protein in Arabidopsis from cDNA AF036340 (Fig. 4; Xie et al. 1998). This, combined with the observation that asparagine is present at that position only in *Helianthus annuus* (based on our phylogeny data of 52 plant species) suggests that this mutation may be an enhancement of function from the ancestral gene.

Our study reveals that the two significant candidate genes explain a total of 7.4 % of the phenotypic variation in this population, which is only a part of the total quantitative variation for basal stalk rot resistance. The strength and magnitude of associations with these genes robustly support their role in Sclerotinia stalk rot resistance in sunflower. We expect that stalk rot resistance in sunflower is influenced by many additional, yet still unknown genes each contributing a small part to the expression of the phenotype, as observed in the case of many other quantitative traits (Haseneyer et al. 2010; Ghavami et al. 2011). Very few comparisons of our results can be made to previous mapping studies because of the lack of published information for Sclerotinia stalk rot resistance QTL in sunflower. Micic et al. (2004) performed QTL mapping of Sclerotinia midstalk rot resistance in F_2/F_3 families developed from a cross between sunflower lines, $NDBLOS_{sel}$ × CM625 and identified stem lesion and speed of fungal growth QTLs in linkage group 16. Using selective genotyping of F_2/F_3 families developed from a cross between CM625 × TUB-5-3234, Micic et al. (2005b) then identified QTLs for stem lesion, leaf lesion, and speed of fungal growth in linkage group 10. Davar et al. (2010) identified basal stem rot resistance QTLs in linkage groups 1, 2, 4, 6, 8, 14, and 17 using a recombinant inbred population developed from the cross between the sunflower parental lines PAC2 and RHA266. Although interesting correlations were detected between the position of the *HaCOI1* candidate genes and the Davar et al. (2010) QTLs, it is difficult to be certain that our candidate genes are the same as the QTLs found in that study because of the lack of complete precision in QTL maps.

The candidate gene associations reported here represent an important first step toward identifying the genes underlying quantitative resistance to Sclerotinia stalk rot in sunflower, and more importantly, to a more streamlined and cost-effective approach to breeding for yield and

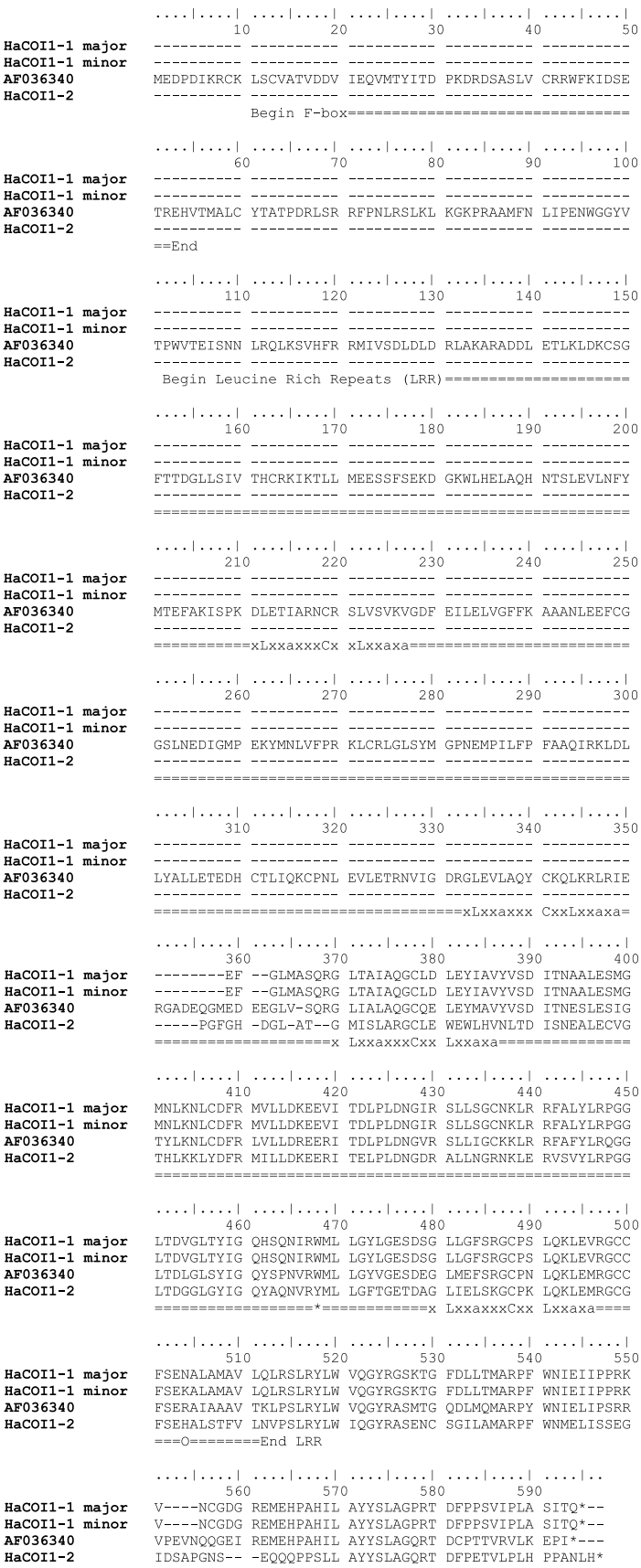

Fig. 4 Predicted protein sequence alignment for *COI1*, *HaCOI1-1*, and *HaCOI1-2*. The two alleles of COI1-1_251 are included as separate lines. The final line includes annotation of the functional portions of the protein, the F-box and Leucine-Rich Regions, as well as the stop codon that is the result of the *coi1-1* mutant in *Arabidopsis* (indicated by *asterisk*) and site of the amino acid residue change due to HaCOI1-1_251 (indicated by '*O*'). Conserved Leucine-Rich Repeat sequences as defined by Xie et al. (1998) for *COI1* are shown as 'xLxx-axxxCxxLxxaxa' where 'L' and 'C' are leucine and cysteine, respectively, 'x' is any amino acid, and 'a' is any hydrophobic amino acid

Sclerotinia resistance simultaneously. Owing to the quantitative nature of the target trait, genome-wide association mapping with sufficiently dense marker coverage could provide a more comprehensive picture of loci involved in the regulation of stalk rot resistance. Studies designed to test this hypothesis are currently underway. Nevertheless, the significant markers in this study may be immediately applied in marker-assisted selection (MAS) breeding for rapid pre-screening of sunflower lines for the potential to express stalk rot resistance before further advancement or field evaluation.

Acknowledgments We thank the staff of the USDA-ARS North Central Regional Plant Introduction Station, Ames, Iowa, USA for providing us with germplasm and National Sunflower Association for providing SNP markers. We are very grateful to Drs. John Burke and John Bowers for providing map positions of the candidate genes and critically reviewing the manuscript. We also thank Dr. Z. Liu, M. Ramsett, N. Balbyshev, A. Hogness, A. Jani, and D. Weiskopf for technical assistance. This research was supported by the USDA-ARS National Sclerotinia Initiative, grant number 58-5442-8-230.

Conflict of interest All experiments detailed herein were conducted in compliance with the applicable laws of the United States and the state of North Dakota. The authors declare that they have no conflict of interest.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Andersen JR, Zein I, Wenzel G, Darnhofer B, Eder J, Ouzunova M, Lübberstedt T (2008) Characterization of phenylpropanoid pathway genes within European maize (*Zea mays* L.) inbreds. BMC Plant Biol 8:2. doi:10.1186/1471-2229-8-2

Balding DJ (2006) A tutorial on statistical methods for population association studies. Nat Rev Genet 7:781–791

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

Bernardo R (2008) Molecular markers and selection for complex Traits in plants: learning from the last 20 years. Crop Sci 48:1649–1664

Bernardo R (2010) Breeding for quantitative traits in plants, 2nd edn. Stemma Press, Minnesota

Bert P-F, Jouan I, Tourvieille de Labrouhe D, Serre F, Nicolas P, Vear F (2002) Comparative genetic analysis of quantitative traits in sunflower (*Helianthus annuus* L.): 1. QTL involved in resistance to *Sclerotinia sclerotiorum* and *Diaporthe helianthi*. Theor Appl Genet 105:985–993

Bert PF, Dechamp-Guillaume G, Serre F, Jouan I, Tourvieille de Labrouhe D, Nicolas P, Vear F (2004) Comparative genetic analysis of quantitative traits in sunflower (*Helianthus annuus* L.) 3. Characterisation of QTL involved in resistance to *Sclerotinia sclerotiorum* and *Phoma macdonaldi*. Theor Appl Genet 109:865–874

Boland GJ, Hall R (1994) Index of plant hosts of *Sclerotinia sclerotiorum*. Can J Plant Pathol 16:94–108

Bolton MD, Thomma BPHJ, Nelson BD (2006) *Sclerotinia sclerotiorum* (Lib.) de Bary: biology and molecular traits of a cosmopolitan pathogen. Mol Plant Pathol 7:1–16

Bowers JE, Nambeesan S, Corbi J, Barker MS, Rieseberg LH, Knapp SJ, Burke JM (2012) Development of an ultra-dense genetic map of the sunflower genome based on single-feature polymorphisms. PLoS ONE 7:e51360

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Cabrera y Poch HL, Peto CA, Chory J (1993) A mutation in the *Arabidopsis* DET3 gene uncouples photoregulated leaf development from gene expression and chloroplast biogenesis. Plant J 4:671–682

Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics 172:557–567

Caño-Delgado A, Penfield S, Smith C, Catley M, Bevan M (2003) Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana*. Plant J 34:351–362

Castaño F, Vear F, Tourvieille de Labrouhe D (1993) Resistance of sunflower inbred lines to various forms of attack by *Sclerotinia sclerotiorum* and relations with some morphological characters. Euphytica 68:85–98

Davar R, Darvishzadeh R, Majd A, Ghosta Y, Sarrafi A (2010) QTL mapping of partial resistance to basal stem rot in sunflower using recombinant inbred lines. Phytopathol Mediterr 49:330–341

Dereeper A, Nicolas S, Lecunff L, Bacilieri R, Doligez A, Peros JP, Ruiz M, This P (2011) SNiPlay: a web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects. BMC Bioinformatics 12:134

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Fusari CM, Lia VV, Hopp HE, Heinz RA, Paniego NB (2008) Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. BMC Plant Biol 8:7. doi:10.1186/1471-2229-8-7

Fusari CM, Di Rienzo JA, Troglia C, Nishinakamasu V, Valeria Moreno M, Maringolo C, Quiroz F, Álvarez D, Escande A, Hopp E, Heinz R, Lia VV, Paniego NB (2012) Association mapping in sunflower for Sclerotinia head rot resistance. BMC Plant Biol 12:93

Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Schüler K (2004) Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol Breeding 13:93–102

Gelfand AE, Ghosh SK (1998) Model choice: a minimum posterior predictive loss approach. Biometrika 85:1–11

Gentzbittel L, Mouzeyar S, Badaoui S, Mestries E, Vear F, Tourvieille De Labrouhe D, Nicolas P (1998) Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus* L. Theor Appl Genet 96:519–525

Ghavami F, Elias ME, Mamidi S, Ansari O, Sargolzaei M, Adhikari T, Mergoum M, Kianian SF (2011) Mixed model association mapping for Fusarium head blight resistance in Tunisian-derived durum wheat populations. G3-Genes Genomes Genet 1:209–218

Goloboff PA, Farris JS, Nixon KC (2008) TNT, a free program for phylogenetic analysis. Cladistics 24:774–786

Guimarães RL, Stotz HU (2004) Oxalate production by *Sclerotinia sclerotiorum* deregulates guard cells during infection. Plant Physiol 136:3703–3711

Gulya TJ (2004) An inoculation method for Sclerotinia stalk rot. In: Proceedings of the 26th Sunflower Research Workshop, Fargo, ND, USA, January 14-15, 2004. Available: http://www.sunflowernsa.com/research/research-workshop/documents/Gulya_StalkRot_04.PDF. Accessed 15 August 2012

Gulya TJ, Vick BA, Nelson BD (1989) Sclerotinia head rot of sunflower in North Dakota: 1986 incidence, effect on yield and oil components, and sources of resistance. Plant Dis 73:504–507

Guo X, Stotz HU (2007) Defense against *Sclerotinia sclerotiorum* in *Arabidopsis* is dependent on Jasmonic Acid (JA), Salicylic Acid (SA), and Ethylene (ET) signaling. Mol Plant Microbe In 20:1384–1395

Hahn V (2002) Genetic variation for resistance to Sclerotinia head rot in sunflower inbred lines. Field Crops Res 77:153–159

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hall D, Tegström C, Ingvarsson PK (2010) Using association mapping to dissect the genetic basis of complex traits in plants. Brief Funct Genomics 9:157–165

Hallauer AR, Miranda JB (1988) Quantitative genetics in maize breeding, 2nd edn. Iowa State Univ. Press, Ames

Hammond-Kosack KE, Jones JD (1996) Resistance gene-dependent plant defense responses. Plant Cell 8:1773–1791

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620

Haseneyer G, Stracke S, Piepho H-P, Sauer S, Geiger HH, Graner A (2010) DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. BMC Plant Biol 10:5. doi:10.1186/1471-2229-10-5

Hernandez-Blanco C, Dong Hu XFJ, Sanchez-Vallet A, Deslandes L, Llorente F, Berrocal-Lobo M, Keller H, Barlet X, Sanchez-Rodrıguez C, Anderson LK, Somerville S, Marco Y, Molina A (2007) Impairment of cellulose synthases required for Arabidopsis secondary cell wall formation enhances disease resistance. Plant Cell 19:890–903

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54–78

Horne EC, Kumpatla SP, Patterson KA, Gupta M, Thompson SA (2004) Improved high-throughput sunflower and cotton genomic DNA extraction and PCR fidelity. Plant Mol Biol Rep 22: 83a–83i

Ivandic V, Thomas WTB, Nevo E, Zhang Z, Forster BP (2003) Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in *Hordeum spontaneum*. Plant Breed 122:300–304

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl Acids Res 30:3059–3066

Kearsey MJ, Farquhar AG (1998) QTL analysis in plants; where are we now? Heredity 80:137–142

Kolkman JM, Berry ST, Leon AJ, Slabaugh MB, Tang S, Gao W, Shintani DK, Burke JM, Knapp SJ (2007) Single nucleotide polymorphisms and linkage disequilibrium in sunflower. Genetics 177:457–468

Laluk K, Mengiste T (2010) Necrotroph attacks on plants: Wanton destruction or covert extortion? The Arabidopsis Book e0136

Li Y-H, Zhang C, Gao Z-S, Smulders MJM, Ma Z, Liu Z-X, Nan H-Y, Chang R-Z, Qiu L-J (2009) Development of SNP markers and haplotype analysis of the candidate gene for *rhg1*, which confers resistance to soybean cyst nematode in soybean. Mol Breed 24:63–76

Liu A, Burke JM (2006) Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics 173:321–330

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). Am J Bot 82:1420–1425

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 12:57–63

Masirevic S, Gulya TJ (1992) *Sclerotinia* and phomopsis—two devastating sunflower pathogens. Field Crops Res 30:271–300

Mestries E, Gentzbittel L, Tourvieille de Labrouhe D, Nicolas P, Vear F (1998) Analyses of quantitative trait loci associated with resistance to *Sclerotinia sclerotiorum* in sunflowers (*Helianthus annuus* L.) using molecular markers. Mol Breeding 4:215–226

Micic Z, Hahn V, Bauer E, Schön CC, Knapp SJ, Tang S, Melchinger AE (2004) QTL mapping of Sclerotinia midstalk-rot resistance in sunflower. Theor Appl Genet 109:1474–1484

Micic Z, Hahn V, Bauer E, Melchinger AE, Knapp SJ, Tang S, Schön CC (2005a) Identification and validation of QTL for Sclerotinia midstalk-rot resistance in sunflower by selective genotyping. Theor Appl Genet 111:233–242

Micic Z, Hahn V, Bauer E, Schön CC, Melchinger AE (2005b) QTL mapping of resistance to Sclerotinia midstalk-rot in RIL of sunflower population NDBLOSsel × CM625. Theor Appl Genet 110:1490–1498

Miller JF, Gulya TJ (1999) Registration of eight Sclerotinia-tolerant sunflower germplasm lines. Crop Sci 39:301–302

Miller JF, Gulya TJ (2006) Registration of two restorer (RHA 439 and RHA 440) and one maintainer (HA 441) Sclerotinia tolerant oilseed sunflower germplasms. Crop Sci 46:482–483

Miller JF, Vick BA, Gulya TJ (2006) Registration of two maintainer (HA 451 and HA 452) and three restorer (RHA 453-RHA 455) Sclerotinia-tolerant oilseed sunflower germplasms. Crop Sci 46:2727–2728

Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. Proc Natl Acad Sci 102:2442–2447

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. Trends Plant Sci 9:325–330

Newman LJ, Perazza DE, Juda L, Campbell MM (2004) Involvement of the R2R3-MYB, *At*MYB61, in the ectopic lignification and dark-photomorphogenic components of the *det3* mutant phenotype. Plant J 37:239–250

Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18:83–90

Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Pritchard JK, Stephens MN, Rosenberg N, Donnelly P (2000b) Association mapping in structured populations. Am J Hum Genet 67:170–181

Rafalski A (2002a) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94

Rafalski JA (2002b) Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Sci 162:329–333

Ralhan A, Schoettle S, Thurow C, Iven T, Feussner I, Polle A, Gatz C (2012) The vascular pathogen *Verticillium longisporum*

requires a jasmonic acid-independent COI1 function in roots to elicit disease symptoms in Arabidopsis shoots. Plant Physiol 159:1192–1203

R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, available at http://www.R-project.org/

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Rönicke S, Hahn V, Friedt W (2005) Resistance to *Sclerotinia sclerotiorum* of 'high oleic' sunflower inbred lines. Plant Breed 124:376–381

Rosenberg NA (2002) Distruct: a program for the graphical display of structure results. http://www.cmb.usc.edu/»noahr/distruct. Accessed 10 May 2012

Rosenblum BB, Lee LG, Spurgeon SL, Khan SH, Menchen SM, Heiner CR, Chen S-M (1997) New dye-labeled terminators for improved DNA sequencing patterns. Nucleic Acids Res 25:4500–4504

Saïdou A–A, Mariac C, Luong V, Pham J-L, Bezanc¸on G, Vigouroux Y (2009) Association studies identify natural variation at PHYC linked to flowering time and morphological variation in pearl millet. Genetics 182:899–910

SAS Institute (2008) The SAS System for Windows 7, release 9.2. SAS Institute, Cary, NC, USA

Schnurr J, Shockey J, Browse J (2004) The acyl-CoA synthetase encoded by *LACS2* is essential for normal cuticle development in Arabidopsis. Plant Cell 16:629–642

Schuhegger R, Nafisi M, Mansourova M, Petersen BL, Olsen CE, Svatoš A, Halkier BA, Glawischnig E (2006) CYP71B15 (*PAD3*) catalyzes the final step in camalexin biosynthesis. Plant Physiol 141:1248–1254

Schumacher K, Vafeados D, McCarthy M, Sze H, Wilkins T, Chory J (1999) The *Arabidopsis* det3 mutant reveals a central role for the vacuolar H(+)-ATPase in plant growth and development. Genes Dev 13:3259–3270

Schutz WM, Cockerham CC (1966) The effect of field blocking on gains from selection. Biometrics 22:843–863

Simko I, Costanzo S, Haynes KG, Christ BJ, Jones RW (2004a) Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. Theor Appl Genet 108:217–224

Simko I, Haynes KG, Ewing EE, Costanzo S, Christ BJ, Jones RW (2004b) Mapping genes for resistance to *Verticillium albo-atrum* in tetraploid and diploid potato populations using haplotype association tests and genetic linkage analysis. Mol Genet Genomics 271:522–531

Simko I, Pechenick DA, McHale LK, Truco MJ, Ochoa OE, Michelmore RW, Scheffler BE (2009) Association mapping and marker-assisted selection of the lettuce dieback resistance gene *Tvr1*. BMC Plant Biol 9:135. doi:10.1186/1471-2229-9-135

Speed D, Hemani G, Johnson MR, Balding DJ (2012) Improved heritability estimation from genome-wide SNPs. Am J Hum Genet 91:1011–1021

Stotz HU, Jikumaru Y, Shimada Y, Sasaki E, Stingl N, Mueller MJ, Kamiya Y (2011a) Jasmonate-dependent and COI1-independent defense responses against *Sclerotinia sclerotiorum* in *Arabidopsis thaliana*: Auxin is part of *COI1*-independent defense signaling. Plant Cell Physiol 52:1941–1956

Stotz HU, Sawada Y, Shimada Y, Hirai MY, Sasaki E, Krischke M, Brown PD, Saito K, Kamiya Y (2011b) Role of camalexin, indole glucosinolates, and side chain modification of glucosinolate-derived isothiocyanates in defense of Arabidopsis against *Sclerotinia sclerotiorum*. Plant J 67:81–93

Stracke S, Haseneyer G, Veyrieras JB, Geiger HH, Sauer S, Graner A, Piepho HP (2009) Association mapping reveals gene action and interactions in the determination of flowering time in barley. Theor Appl Genet 118:259–273

Szalma SJ, Buckler ES, Snook ME, McMullen MD (2005) Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theor Appl Genet 110:1324–1333

Tang S, Yu J-K, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105:1124–1136

Tang D, Simonich MT, Innes RW (2007) Mutations in *LACS2*, a long-chain Acyl-coenzyme A synthetase, enhance susceptibility to avirulent *Pseudomonas syringae* but confer resistance to *Botrytis cinerea* in Arabidopsis. Plant Physiol 144:1093–1103

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

USDA (2013) ARS, National Genetic Resources Program. Germplasm Resources Information Network (GRIN). [Online Database] National Germplasm Resources Laboratory, Beltsville, Maryland. Available: http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?495348. Accessed 11 September 2013

Vear F, Serre F, Roche S, Walser P, Tourvieille de Labrouhe D (2007) Improvement of *Sclerotinia sclerotiorum* head rot resistance in sunflower by recurrent selection of a restorer population. Helia 30:1–12

Waldmann P, Hallander J, Hoti F, Sillanpää MJ (2008) Efficient Markov Chain Monte Carlo implementation of Bayesian analysis of additive and dominance genetic variances in noninbred pedigrees. Genetics 179:1101–1112

Wang WYS, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6:109–118

Weir BS (1996) Genetic data analysis II: Methods for discrete population genetic data. Sinauer Associates Inc, Sunderland

Wilson LM, Whitt SR, Ibáñez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733

Xie DX, Feys BF, James S, Nieto-Rostro M, Turner JG (1998) *COI1*: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. Science 280:1091–1094

Xu L, Liu F, Lechner E, Genschik P, Crosby WL, Ma H, Peng W, Huang D, Xie D (2002) The SCF(COI1) ubiquitin-ligase complexes are required for jasmonate response in Arabidopsis. Plant Cell 14:1919–1935

Yan JB, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS ONE 4:e8451. doi:10.1371/journal.pone.0008451

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Yue B, Radi SA, Vick BA, Cai X, Tang S, Knapp SJ, Gulya TJ, Miller JF, Hu J (2008) Identifying quantitative trait loci for resistance to Sclerotinia head rot in two USDA sunflower germplasms. Phytopathology 98:926–931

Zhao K, Aranzana MJ, Kim S, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An Arabidopsis example of

association mapping in structured samples. PLoS Genet 3:e4. doi:10.1371/journal.pgen.0030004

Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2:467. doi:10.1038/ncomms1467

Zhou N, Tootle TL, Glazebrook J (1999) Arabidopsis *PAD3*, a gene required for camalexin biosynthesis, encodes a putative cytochrome P450 monooxygenase. Plant Cell 11:2419–2428

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20