

RESEARCH ARTICLE

Open Access

# Association mapping in sunflower for sclerotinia head rot resistance

Corina M Fusari<sup>1</sup>, Julio A Di Rienzo<sup>2</sup>, Carolina Troglia<sup>3</sup>, Verónica Nishinakamasu<sup>1</sup>, María Valeria Moreno<sup>4</sup>, Carla Maringolo<sup>3</sup>, Facundo Quiroz<sup>3</sup>, Daniel Álvarez<sup>4</sup>, Alberto Escande<sup>3</sup>, Esteban Hopp<sup>1,5</sup>, Ruth Heinz<sup>1,5</sup>, Verónica V Lia<sup>1,5</sup> and Norma B Paniego<sup>1\*</sup>

## Abstract

**Background:** Sclerotinia Head Rot (SHR) is one of the most damaging diseases of sunflower in Europe, Argentina, and USA, causing average yield reductions of 10 to 20 %, but leading to total production loss under favorable environmental conditions for the pathogen. Association Mapping (AM) is a promising choice for Quantitative Trait Locus (QTL) mapping, as it detects relationships between phenotypic variation and gene polymorphisms in existing germplasm without development of mapping populations. This article reports the identification of QTL for resistance to SHR based on candidate gene AM.

**Results:** A collection of 94 sunflower inbred lines were tested for SHR under field conditions using assisted inoculation with the fungal pathogen *Sclerotinia sclerotiorum*. Given that no biological mechanisms or biochemical pathways have been clearly identified for SHR, 43 candidate genes were selected based on previous transcript profiling studies in sunflower and *Brassica napus* infected with *S. sclerotiorum*. Associations among SHR incidence and haplotype polymorphisms in 16 candidate genes were tested using Mixed Linear Models (MLM) that account for population structure and kinship relationships. This approach allowed detection of a significant association between the candidate gene *HaRIC\_B* and SHR incidence ( $P < 0.01$ ), accounting for a SHR incidence reduction of about 20 %.

**Conclusions:** These results suggest that AM will be useful in dissecting other complex traits in sunflower, thus providing a valuable tool to assist in crop breeding.

## Background

Sunflower (*Helianthus annuus* L.) is one of the most important sources of vegetable oil grown worldwide [1]. Fungal infections represent one of the main constraints for crop yield and productivity, having a detrimental impact on quality components. *Sclerotinia sclerotiorum* (Lib) de Bary is a worldwide distributed necrotrophic pathogen, attacking more than 400 plant species including sunflower, soybean and rapeseed [2,3]. The fungus can attack several plant organs causing diverse symptoms in leaves, stalks and flowers, with Sclerotinia Head Rot (SHR) being the most damaging for sunflower crop production. SHR is considered a major disease in Europe,

Argentina, and USA, causing average yield reductions of 10 to 20 %, and can even result in loss of the entire harvest [4,5]. Disease control on *S. sclerotiorum* is difficult, since the fungus persists in soils for long periods and at high inoculum levels [6-8].

Resistance to *S. sclerotiorum* has been described as quantitatively inherited with predominantly additive gene action, and medium heritability [3]. Classical linkage mapping based on biparental populations was used to dissect Quantitative Trait Loci (QTL) for SHR resistance. These analyses have rendered QTL with small effects, explaining only a minor proportion of the phenotypic variance [9-15]. In addition, a number of studies have been done in other *S. sclerotiorum* host species to understand the defense mechanisms triggered in resistant genotypes [16-25]. One of the most comprehensive studies has been conducted in *Brassica napus* by performing microarray analysis in resistant and susceptible genotypes infected with *S. sclerotiorum*

\* Correspondence: npaniego@cni.inta.gov.ar

<sup>1</sup>Instituto de Biotecnología, Centro Investigación en Ciencias Veterinarias y Agronómicas (CICVyA), Instituto Nacional de Tecnología Agropecuaria (INTA), 1686, Hurlingham, Buenos Aires, Argentina

Full list of author information is available at the end of the article

[25]. A total of 686 and 1,586 genes were found to be differentially expressed after infection in the resistant and susceptible genotypes, respectively. The number of differentially expressed genes increased over infection time, most of them being up-regulated. The putative functions of these genes included pathogenesis related proteins, proteins involved in oxidative burst, protein kinase and molecule transporters, among others [25]. However, extrapolating this information to sunflower, and using it to evaluate new sources of resistance requires the identification of orthologous genes between *B. napus* and *H. annuus*.

Association Mapping (AM) was suggested as a promising alternative to classical linkage mapping to elucidate the genetic basis of complex traits [26]. The AM approach is based on the extent of Linkage Disequilibrium (LD) observed in a set of accessions that are not closely related. In contrast to classical biparental population mapping, AM is a method that detects relationships between phenotypic variation and gene polymorphisms in existing germplasm, without development of mapping populations. This method incorporates the effects of recombination occurring in many past generations into a single analysis and thus, it is complementary to the classical biparental approach [27]. The main drawback of AM is the possibility of false-positive results due to unrecognized population structure. In order to avoid population stratification effects, information on the relatedness among genotypes is commonly included through the recognition of the population structure and/or as a kinship matrix between genotypes [26].

AM has been successfully applied in mapping genes involved in several traits in different plant species (e.g. flowering time and aluminum tolerance in maize, resistance to late blight in potato, kernel size and milling quality in wheat, resistance to dieback in lettuce) [28-32]. Even though two possible strategies have been proposed, Genome Wide Association (GWA) and candidate gene approaches, the latter has been the most widely used in plants, mostly due to the lack of complete genome data for many plant species.

Despite the fact that both nucleotide diversity and decay of LD have been assessed in sunflower germplasm from different origins [33-35], no AM studies have been published to date either for SHR or any other trait.

This paper reports the identification of resistance QTL for SHR based on candidate gene AM. Given that no biological mechanisms or biochemical pathways have been positively identified for SHR, selection of candidate genes was based on previous transcript profiling studies in sunflower [36-38] and *B. napus* [25]. This approach resulted successful in detecting a significant association between one of the candidate genes evaluated and SHR incidence. These results suggest that AM is a useful

strategy for dissecting complex traits in sunflower, thus providing a valuable tool to assist in crop breeding.

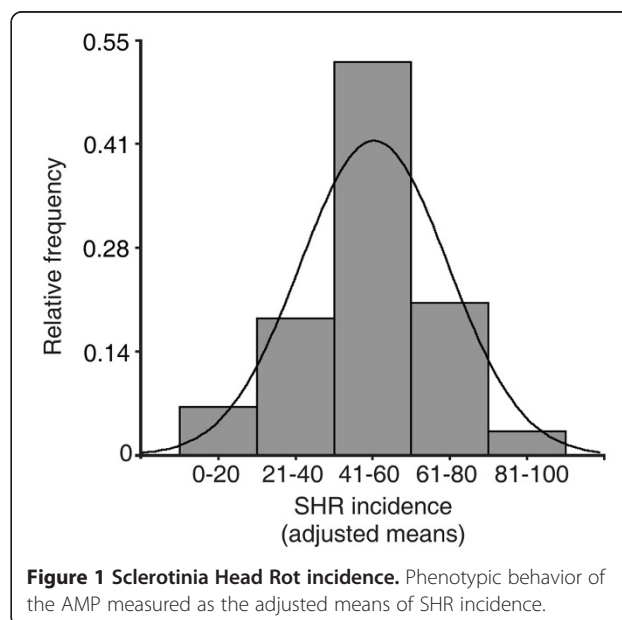
## Results

### Phenotypic data

Ninety-four inbred lines belonging to the "Sunflower Breeding Program" of INTA were evaluated for SHR incidence during 2008/2009 and 2009/2010 in replicated trials. Because the experiments were conducted as randomized complete blocks, and repeated in two consecutive trials, the model used to obtain the adjusted line means, included the trial, and the blocks within trials, as random effects. All inoculation days were suitable to produce disease. The line effect was significant ( $P < 0.0001$ ). The adjusted SHR incidence means varied from 0 % to 100 %, with an average of 50.4 %. In fact, 52 % of the Association Mapping Population (AMP) showed an intermediate behavior against the disease, i.e. between 40 % and 60 % (Figure 1 and Additional file 1).

### Candidate gene selection

A total of 43 genes were used as starting point for candidate gene selection. Putative orthologous sequences, either from sunflower or from other Asteraceae species were identified for 18 out of 19 *A. thaliana* loci selected from the work of Zhao et al. [25] using the phylogenetic approach detailed in Methods. After PCR amplification using sunflower genomic DNA as template, products with the expected sequence were obtained for 10/19 loci. For six of them, paralogous sequences were also identified in *Helianthus*. Thus, the initial 10 loci of *A. thaliana* correlated with 17 genes in sunflower (*HaMT*, *HaPRP1*, *HaPRP2*, *HaCYP450A*, *HaCYP450B*, *HaCYP450C*,



**Figure 1 Sclerotinia Head Rot incidence.** Phenotypic behavior of the AMP measured as the adjusted means of SHR incidence.

*HaROPGEF1*, *HaROPGEF2*, *HaROPGEF3*, *HaPP2C*, *HaRIC\_A*, *HaRIC\_B*, *HaHP1*, *HaHP2*, *HaFbox1*, *HaFbox2* and *HaZFHD*) (Table 1).

Specific PCR products were obtained for 11/17 genes selected from the RHA801 EST library (*HaDRP*, *HaCP*, *HaGDPDI*, *HaCWP*, *HaPI*, *HaPAL*, *HaRNAHe*, *HaHP3*, *HaKIV*, *HaPEP*, *HaTF1*). Conversely, PCR amplification of *Ha26SPS*, *HaTF2* and *HaLTP1* yielded unspecific products, whereas little or no amplification was observed for *HaTHI*, *HaTE*, *HaPK*. Introns of variable length (500–1200 pb) were found for *HaRNAHe*, *HaHP3*, *HaKIV*, *HaPEP* and *HaTF1*.

Five germin-like proteins (*HaGLP1* to *HaGLP5*) and two transcription factors from the WRKY family (*HaWRKY5* and *HaWRKY7*) were amplified successfully. Summarizing, a total of 30 candidate genes were suitable for amplification in the 10 sunflower inbred lines selected as Core Set (CS) for initial polymorphism development. The candidates encompassed 17 genes derived from the expression analysis of Zhao et al. [25], six genes from the RHA801 EST library [38] and seven genes chosen based on their putative role in defense mechanisms in sunflower [36,37] (Table 1).

#### SNP identification and genotyping of candidate genes

Polymorphisms were found in 28/30 candidate genes, and 21 were further selected to be genotyped in the AMP. The seven genes that were not included in subsequent analyses were discarded due to (1) minor allele frequencies lower than 0.2 in the CS (*HaCYP450A*, *HaCYP450B* and *HaHP1*), (2) failure of DNA sequencing in some members of the CS (*HaROPGEF1*, *HaFbox1* and *HaFbox2*) and (3) the instability of polymorphic sites (e.g. indels made of poly-T tracts in *HaGLP1*, Table 1). Genes *HaDRP*, *HaCP*, *HaGLP2*, *HaGLP3* and *HaGLP5*, selected to be genotyped by denaturing High Performance Liquid Chromatography (dHPLC) in the AMP, did not achieve the expected resolution and were also excluded from further analyses. Hence, 16/28 polymorphic candidate genes were genotyped in the AMP. Three candidate genes were genotyped by dHPLC (*HaPRP2*, *HaCYP450C*, *HaPAL*), nine by Fluorescent Capillary Electrophoresis (FCE) (*HaMT*, *HaROPGEF2*, *HaRIC\_A*, *HaRIC\_B*, *HaCWP*, *HaPI*, *HaWRKY5*, *HaWRKY*, *HaGLP4*) and four were typified by direct sequencing (*HaPP2C*, *HaHP2*, *HaZFHD*, *HaGDPDI*) (Table 1). Haplotypes not present in the CS were found in the AMP for *HaCYP450C*, *HaPP2C*, *HaRIC\_B*, *HaHP2*, *HaZFHD*, *HaGDPDI* and *HaWRKY5*. Regarding the 16 candidate genes genotyped in the AMP, the number of haplotypes ranged from two to nine, with an average of 3.6 haplotypes per gene. The average haplotype frequency was 27.6 %. The lowest haplotype frequency (1.1 %) was found for *HaCYP450C* and *HaWRKY5*, whereas the highest haplotype frequency (93.4 %) was found for *HaPI* gene.

#### Population structure

The AMP was characterized using eight SSRs, resulting in the detection of 47 alleles, ranging from four alleles at locus HA4103 to nine at locus HA991 (average of 6.6 alleles per locus).

Population structure was evaluated through Principal Coordinate Analysis (PCO) and the Bayesian method implemented in STRUCTURE software [39]. The first two principal coordinates of the PCO explained 12.6 % and 9.8 % of the molecular variance, with no clear clustering of individuals being detected among the 94 inbred lines analyzed (Additional file 2). According to the STRUCTURE analysis, the log-likelihood values of the data conditional on  $k$  reached a plateau at  $k=9$  (Additional file 2: Figure S1B). Likewise, the *ad hoc* procedure proposed by Evanno et al. [40] showed peaks of  $\Delta k$  at  $k=9$ ,  $k=5$  and  $k=2$ . Thus,  $k=9$  was selected as the most likely number of ideal populations, since its signal was detected by both the log-likelihood and the Evanno et al. criteria [40]. Although most of the inbred lines (62 %) showed an inferred ancestry higher than 60 % to one of the  $k=9$  ideal populations, the other 38 % of individuals showed substantial admixture (Additional file 2).

#### Association analysis

The MLM accounts for multiple levels of relatedness, including population structure (**Q** or **P** matrixes) and kinship relationships (**K** matrix). The **K** matrix was calculated following Bernardo [41], as  $K_{ij}^T = \frac{S_{ij}-1}{1-T} + 1$ . The  $T$  parameter represents the probability that two alleles are alike in state, given that they are not identical by descent. In practice,  $T$  is unknown, and should be assessed, usually, by a maximum likelihood estimation procedure. However, despite several  $T$  values ranging from 0.2 to 0.8 were empirically tested to evaluate its effect on results, the latter were unaffected by changes of  $T$ . The  $P$ -values of the association analysis using either the **P** matrix (based on PCO) or the **Q** matrix (based on the Bayesian approach of STRUCTURE) and two  $T$  values for the kinship matrix were found to be highly stable (Table 2). A significant association was found between *HaRIC\_B* and SHR incidence ( $P < 0.01$ , Table 2), with the maximum family wise error rate being below 14 % for the number of genes analyzed and a significance level of 0.01.

In order to identify which haplotype was involved in the association detected, the averages of the SHR incidence adjusted means were plotted against the three different haplotypes found in *HaRIC\_B* (Figure 2A). The haplotype 3 is associated with the lowest levels of SHR incidence and is present in four inbred lines (GP365, C192-1, 5381 and 5393-E), which showed SHR incidence of 0.0, 22.5, 25.5 and 46.3 %, respectively. The effect of haplotype 3 was estimated as an incidence reduction of 19 % relative

**Table 1 Candidate genes for SHR: sources of selection and features for genotyping the association mapping population**

Source of selection	Target locus	Sunflower gene ID <sup>a</sup>	Sequence length <sup>b</sup>	SNPs in CS <sup>c</sup>	Indels in CS <sup>d</sup>	Haplotypes in CS <sup>e</sup>	Genotyping method <sup>f</sup>	Primers for AMP genotyping
<i>B. napus</i> expression profile	At1g34580	<u>HaMT</u>	580 (191)	8 (8)	3 (11)	3 (2)	FCE	F:5'GAAAGGACATGCTACTTTATGG3' R:5'CTTTACTTGAATTAAGTTACT3'
	At5g51550	<u>HaPRP1</u> <u>HaPRP2</u>	431 437 (166)	0 11 (11)	- -	1 3 (2)	dHPLC	F: 5'GGACGGGAACGTAAAATAATG3' R: 5'CCGTCTGTCCGTACAATCG3'
	At3g48310	<u>HaCYP450A</u> <u>HaCYP450B</u> <u>HaCYP450C</u>	786 310 400 (260)	1 (0) 2 (0) 2 (2)	- - 1 (1)	2 2 2 (6)	dHPLC	F: 5'AAGTGACTTTAGCAACGTCC3' R: 5'GAGTTGGTATGGGTGGATGAA3'
	At5g05940	<u>HaROPGEF1</u> <u>HaROPGEF2</u>	447 403 (197)	12 (11) 2 (2)	3 (50) 2 (11)	4 2 (2)	FCE	F: 5'TGCGTAGTGGTTCTAAAATTGG3' R: 5'CGTCAATCATTACCCCAACC3'
	At1g48040	<u>HaROPGEF3</u> <u>HaPP2C</u>	801 514 (398)	0 4 (4)	- -	1 2 (4)	Direct sequencing	F: 5'ACTGGGACTACGGCATTGAC3' R: 5'TGCTGAATTTCTGGCTCTGA3'
	At1g04450	<u>HaRIC_A</u> <u>HaRIC_B</u>	940 (261) 1424 (537)	8 (8) 12 (0)	4 (6) 2 (15)	3 (2) 3 (3)	FCE FCE	F: 5'GCACGAATAGTGACATTGAAAC3' R: 5'ACATAAAACAGTTTTCCGGTCC3' F: 5'GGCTTGCGTTACATCTCTGA3' R: 5'CCCACTAGGAGCATTGGAA 3'
	At1g03687	<u>HaHP1</u>	198	0	1 (1)	1		
	At4g09180	<u>HaHP2</u>	309 (256)	1 (1)	-	2 (5)	Direct sequencing	F: 5'CTGCTATCCAGGCTCATTCA3' R: 5'AGAATGGCAGGGCGACCAAG3'
	At1g13200	<u>HaFbox1</u> <u>HaFbox2</u>	265 262	11 (11) 8 (5)	- 1 (1)	3 5		
	At1g14687	<u>HaZFHD</u>	166 (166)	9 (7)	-	4 (9)	Direct sequencing	F: 5'TCATGCCCTCACTAACATGC 3' R: 5'TTTGTCCGGAATCTTTTTTCG 3'
Sunflower EST library	TC49193	<u>HaDRP</u>	587	14 (12)	1 (1)	2		
	BQ973243	<u>HaCP</u>	542	14 (10)	-	3		
	-	<u>HaGDPDI</u>	620 (707)	18 (14)	2 (2)	3 (10)	Direct sequencing	F: 5'CAGAACTGATCAACCCGAAA 3' R: 5'TGCATGCATCTTGAAAAATAG 3'
	-	<u>HaCWP</u>	351 (177)	4 (4)	3 (10)	2 (2)	FCE	F: 5'CAGGAATCACGGTCCCTAGT3' R: 5'TGAAACATGAGGGATGAGCA3'
	TC42391	<u>HaPI</u>	658 (235)	14 (9)	3 (4)	5 (2)	FCE	F: 5'TCCAACAGTGTGTGACCTTTG3' R: 5'CATTAGTTACGTTACAAGCTAT3'
	TC57179	<u>HaPAL</u>	343 (343)	2 (2)	-	2 (2)	dHPLC	F: 5'TGTGGTCTCAAATTCATTAATAAC3' R: 5'GGCCATTCCTAACAGGATCA3'
Defense responses	CF088675	<u>HaWRKY5</u>	1395 (227)	39 (33)	14 (56)	8 (6)	FCE	F: 5'CCGATCAAAGGCTCAATCTA3' R: 5'CACATCCGCTAGTTCACACC3'

**Table 1 Candidate genes for SHR: sources of selection and features for genotyping the association mapping population (Continued)**

BU016906	<u>HaWRKY7</u>	1163 (174)	1 (1)	2 (22)	3 (2)	FCE	F: 5'CATTGTTGGTCAACCCTGTG3' R: 5'AGGGAAGCATAACCATGACG3'
EU112647	<i>HaGLP1</i>	876	0	3 (3)	3		
AJ540203	<i>HaGLP2</i>	794	6 (6)	-	2		
TC17527	<i>HaGLP3</i>	621	13 (9)	-	6		
TC18217	<u>HaGLP4</u>	761 (209)	9 (2)	3 (11)	4 (4)	FCE	F: 5'TGGCTGCAACAACCTTCCTT3' R: 5'TTCAATCCAGAAACAACTTCTAA3'
TC17648	<i>HaGLP5</i>	1710	2 (1)	2 (3)	3		

<sup>a</sup> Acronyms used throughout this article, candidate genes genotyped in the AMP are underlined.

<sup>b</sup> Length of PCR products sequenced in the CS. Length of PCR products genotyped in the AMP are given in parentheses.

<sup>c</sup> Parsimony informative sites are given in parentheses.

<sup>d</sup> Base pairs of indels are given in parentheses.

<sup>e</sup> Number of haplotypes detected in the AMP are given in parentheses.

<sup>f</sup> FCE: fluorescent capillary electrophoresis, dHPLC: denaturing high liquid performance chromatography.

<sup>g</sup> Primers 5'-end 6-FAM labeled.

to the average incidence estimated for the whole set of inbred lines involved in the experiment.

*HaRIC\_B* PCR amplification using genomic DNA from inbred lines carrying haplotypes 1, 2 and 3 yielded products of different lengths. Haplotype 1 generated the shortest PCR product (1565 bp), haplotype 2 produced an intermediate length fragment correlating with a 10 bp insertion present in the first intron (1575 bp), and haplotype 3

**Table 2 P-values of associations between candidate genes and SHR incidence in sunflower**

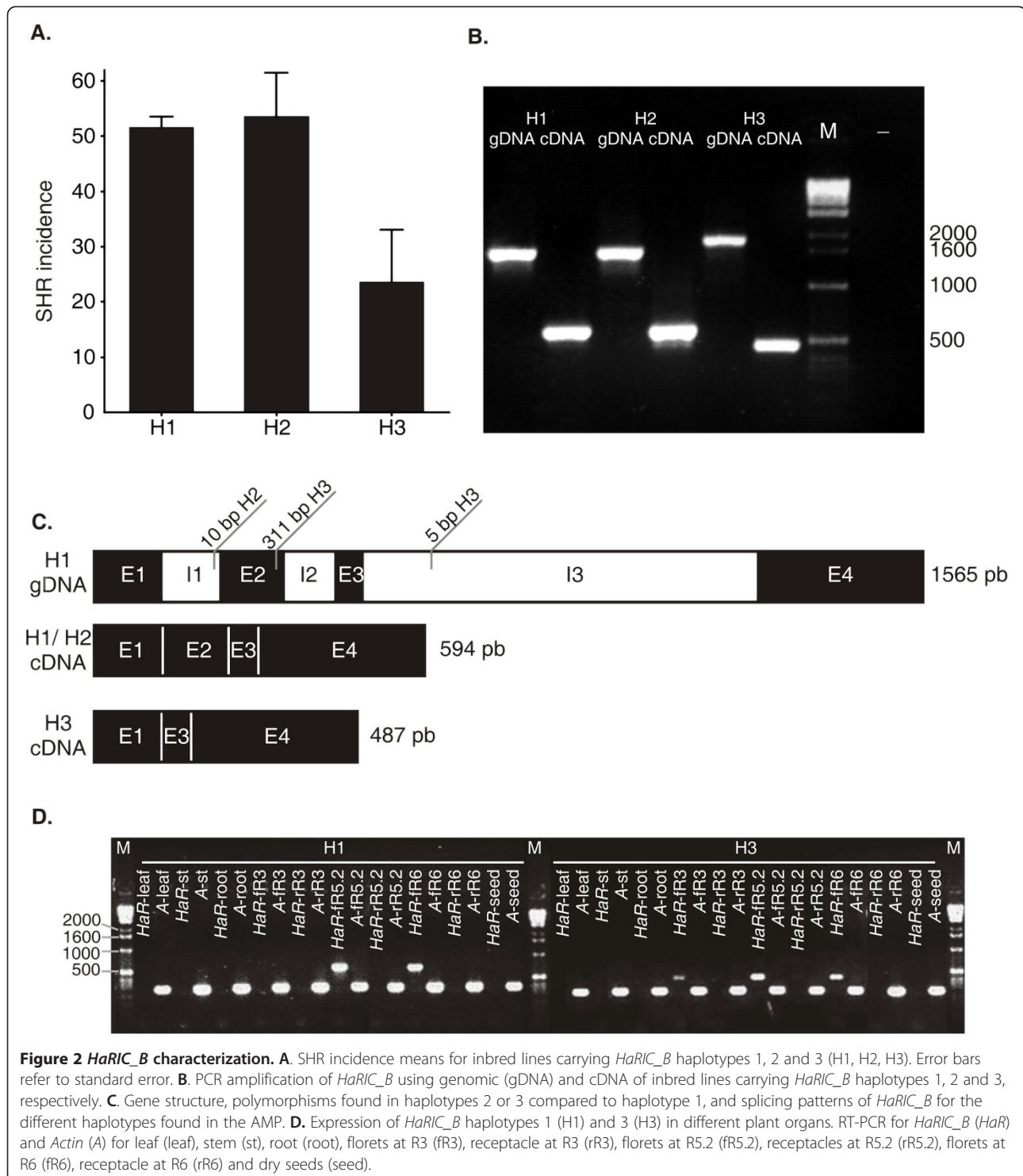
SUNFLOWER GENE ID	P MATRIX		Q MATRIX	
	(T = 0.2)	(T = 0.8)	(T = 0.2)	(T = 0.8)
<i>HaMT</i>	0.8400	0.8356	0.8237	0.8192
<i>HaPRP2</i>	0.9493	0.9225	0.4897	0.9136
<i>HaCytP450C</i>	0.7856	0.7820	0.4897	0.7488
<i>HaROPGEF2</i>	0.9142	0.9486	0.9087	0.9454
<i>HaPP2C</i>	0.7153	0.7256	0.6797	0.6918
<i>HaRIC_A</i>	0.2094	0.2112	0.2015	0.2036
<i>HaRIC_B</i>	0.0098**	0.0096**	0.0066**	0.0066**
<i>HaHP2</i>	0.5818	0.5993	0.7514	0.5113
<i>HaZFHD</i>	0.4526	0.4539	0.3443	0.3472
<i>HaGDPDI</i>	0.9807	0.9811	0.9770	0.9775
<i>HaWP</i>	0.4699	0.4645	0.4501	0.4449
<i>HaTRP</i>	0.3764	0.3707	0.3501	0.9142
<i>HaPAL</i>	0.1635	0.1646	0.1342	0.1359
<i>HaWRKY5</i>	0.2255	0.2249	0.1398	0.1408
<i>HaWRKY7</i>	0.1069	0.1019	0.0933	0.0892
<i>HaGLP4</i>	0.3487	0.3567	0.3128	0.3219

Kinship relationships were obtained as suggested by Bernardo [41]. The K matrix was calculated as  $K_{ij}^T = \frac{S_{ij} - 1}{T - 1} + 1$ , where  $T$  is the probability that two alleles are alike in state, given that they are not identical by descent. In practice,  $T$  is unknown, but different  $T$  values (0.2, 0.3, 0.7, 0.8, only showing data for 0.2 and 0.8) were evaluated to reach the maximum likelihood in model (1) explained in Methods Section. \*\* $P < 0.01$ .

amplified the longest PCR product (1871 bp, Figure 2B and 2 C). *HaRIC\_B* was also analyzed at the cDNA level to investigate its gene structure in sunflower. According to this analysis, *HaRIC\_B* genomic structure consists of four exons and three introns (Figure 2C). Haplotype 1 and 2 did not show length variation at the cDNA level. The length difference between haplotype 1/2 and 3 corresponds to a 311-bp insertion located 31 bp before the end of exon two and to a 5-bp indel detected in the third intron (Figure 2C). The 311-bp insertion leads to the splicing of exon two and to a shorter RT-PCR product (487 bp vs. 596 bp, Figure 2C). At the protein level, this splicing pattern causes a frame-shift which results in a premature stop codon (data not shown). Accumulation of *HaRIC\_B* transcripts was examined by RT-PCR in different organs and capitula developmental stages (R3, R5.2, R6) of inbred lines carrying haplotypes 1 and 3, respectively. *HaRIC\_B* is accumulated specifically in florets, in contrast to the housekeeping gene, which was found in all the tested tissues (Figure 2D).

## Discussion

The genetic determinants of SHR resistance were investigated following a candidate gene approach intended to provide fine-scale resolution to QTL mapping efforts for this trait. Selection of candidate genes is one of the fundamental challenges of candidate gene AM, particularly when there is limited knowledge about the molecular mechanisms underlying the trait under study. For SHR, the most comprehensive experimental evidence comes from the study of gene expression changes in two genotypes of the oilseed *B. napus* infected with *S. sclerotiorum* using a whole genome microarray from *A. thaliana* [25]. However, transferring this information to sunflower was hindered by ortholog identification between highly divergent species. Commonly, most researchers use pairwise distance comparisons algorithms, such as BLAST, COG



(Clusters of Orthologous Groups), RBH (Reciprocal BLAST Hits), RSD (Reciprocal Smallest Distance Algorithm) and INPARANOID, to determine gene orthology [42]. In contrast, although they have been shown to exhibit greater accuracy and lower error rates than pairwise comparison methods, phylogeny based approaches have been only

partially exploited due to the complexity of the automation of sequence alignment and the choice of appropriate genes and species to be included in the analysis [42,43]. Notwithstanding, it is only by phylogenetic reconstruction that ortholog and paralog relationships can be distinguished, especially for species with incomplete genome sequence data

where the best hit of pairwise comparison methods is often not the nearest neighbor [44].

The phylogenetic approach devised here allowed identification of 17 genes from Asteraceae that are orthologs or paralogs to the *A. thaliana* loci detected as over-expressed at 24 hpi in the *B. napus* resistant genotype. Particularly, for At5g51550, At3g48310, At5g05940, At1g04450 and At1g13200, two to three loci were identified as paralogs to the *A. thaliana* sequence used as query, providing wider coverage of their putative functional spectrum (Table 1).

Two additional sources served to increase to 30 the number of candidate genes, an EST library obtained through suppressive subtractive hybridization from sunflower capitula of the genotype RHA801 infected with *S. sclerotiorum* at 48 hpi, and previous literature reports [36-38]. After identification of polymorphisms (SNPs and indels) within the core set of 10 inbred lines, a total of 21 candidate genes were selected to be further genotyped in the AMP. This number of candidate genes has proved adequate to find significant genotype-phenotype associations for traits with different degrees of complexity. Examples include the studies carried out in *A. thaliana* for flowering time [45], in potato for late blight [30] and in maize for aluminum tolerance [31]. Moreover, in this study the authors used information sources similar to those described here to select the candidate genes, finding significant associations for six of them.

The AMP analyzed here is representative of the elite breeding pool used by INTA in the "Sunflower Breeding Program". It encompasses germplasm from different geographical origins, with some of the lines being derived from introgressions with wild *Helianthus* species (Additional file 1). In agreement with the morphological diversity and intricate ancestry of the AMP, the SSR markers revealed high levels of genetic variability. Indeed, the mean number of alleles per locus was even higher than the estimate obtained for a set of compounds, populations and lines conserved at the INTA Germplasm Bank (6.6 vs. 5.71) [46].

As suggested by Stich et al. [47], marker-trait associations were assessed using MLM, which took into account trials, blocks, population structure and kinship relationships to control type I error rates. Both PCO and the Bayesian method (STRUCTURE software) were used to infer population structure for the 94 elite inbred lines. While PCO analysis showed no clear grouping pattern by drawing the two first Principal Coordinates, STRUCTURE suggested the presence of nine different gene pools (Additional file 2). The lack of defined groups in PCO is not unexpected given the restricted genetic of cultivated sunflower [1]. However, either using **P** or **Q** matrixes in the association analysis did not have an impact on the *P*-values (Table 2). Moreover, varying the probability *T* for the calculation of the **K** matrix did not modify the association results.

A significant association was detected between a lower SRH incidence and the haplotype 3 of *HaRIC\_B* ( $P < 0.01$ ). The nature of the mutation found in the haplotype 3 of *HaRIC\_B*, the highly specific expression pattern for this gene in sunflower and the experimental evidence on the role of this protein family in *A. thaliana* [48,49] lend support to the biological significance of the detected association. *HaRIC\_B* was selected from the study of Zhao et al. [25] following ortholog identification via phylogenetic analysis. Haplotype 3 showed an insertion of 311-bp at the 3' end of exon two, which alters the RNA splicing and, consequently, leads to both the generation of a frame-shift and a premature stop codon in the mRNA. Only ca. 15 % of mutations that have been identified as being associated with genetic variation in plant quantitative traits involve changes in the amino acid composition of proteins. However, it has been shown that many of the associations that correspond to noncoding mutations located in introns, untranslated regions or intergenic regions show up as significant because they are in LD with untyped causal mutations that in turn are nonsynonymous substitutions [50].

Although *HaRIC\_B* molecular function has been inferred by homology, the orthologous *A. thaliana* gene (AT1G04450, RIC3) has been experimentally described as a small binding-protein that interacts with ROP1 (Rho family GTPase). There are 11 Arabidopsis RIC proteins (for Rop-interactive CRIB motif-containing proteins) involved in pollen tube growth and other functions [49]. Wu et al. [49] have shown that RIC3 transcripts are found only in the flowers and inflorescences of *A. thaliana*. In agreement with its proposed ortholog relationship, the expression pattern of *HaRIC\_B* matches that found for RIC3, with transcripts being present in florets from R3 to R6 developmental stage (Figure 2D). It is noteworthy that R5.2 is the period of maximum susceptibility to *S. sclerotiorum* infection in the sunflower cultivation areas of South America, and florets are the main entry point for the pathogen [4].

Different studies suggest that the role of RIC3 involves elevation of cytosolic  $Ca^{2+}$  and regulation of (1) depolymerization of actin filaments, (2) exo-cytosis, and (3)  $Ca^{2+}$  mediated signals [48,51]. Interestingly, the Rop-interactive domain of *HaRIC\_B* is located in exon two, which is missing in plants carrying haplotype 3. Thus, if a functional protein can be synthesized from it, its regulation by Rop GTPases would seemingly be not possible.

Recently, two necrosis and ethylene-induced peptides (NEPs) have been described in *S. sclerotiorum* (*SsNep1* and *SsNep2*). *SsNep2* expression is highly dependent on  $Ca^{2+}$  concentration, and compounds increasing calcium levels (i.e. caffeine and lanthanum chloride) greatly reduced *S. sclerotiorum* virulence and expression of *SsNep2* [52]. Thus, consistent with its putative role of intracellular

calcium elevation, the haplotype 3 of *HaRIC\_B* might be participating in the defense against SHR by repressing expression of the necrosis factor *SsNep2*, through the de-regulation and elevation of cytosolic  $Ca^{2+}$  concentrations in the target organ for pathogen attack.

The power of association mapping greatly depends not only on the allele frequency distribution but also on the magnitude of the effect that can be ascribed to a locus, relative to other loci present in the population [53]. Thus, the detection of association between *HaRIC\_B* (haplotype 3) and a lower SHR incidence, despite the fact that haplotype 3 was in low frequency, suggests that it has a strong effect on the phenotype. Indeed, considering that resistance is the result of the interaction of multiple factors, having found a single haplotype that accounts for a SHR incidence reduction of about 20 % in average, emphasizes the importance of the finding. However, beyond the experimental evidence presented here, and the biological considerations that support the role of haplotype 3 on the resistance to SHR, validation of these results will require the re-evaluation of this candidate polymorphism in a larger AMP and different field testing environments. A wider association study is currently underway using large-scale gene sampling and high-throughput genotyping methods.

It has been shown that the causal polymorphism for a QTL can be distant from the functional gene under analysis, particularly in species with high levels of LD, such as sunflower [33-35,54]. While it cannot be ruled out that the polymorphism responsible for a lower SHR incidence resides in a linked ungenotyped region, the evidences discussed here suggest that *HaRIC\_B* can be considered a strong candidate to be directly involved in SHR resistance. To assess its relationships with QTL previously identified in biparental populations, *HaRIC\_B* was genotyped in the sunflower RIL mapping population PAC2 X RHA266. It was mapped to LG11 between the markers E36M59\_9 and E38M50\_17 (data not shown), a region for which no QTL have been reported to date [3,7,9-14,55]. This is not unexpected, as the parental lines used to investigate SHR resistance may not carry the haplotype 3 of *HaRIC\_B*, especially considering its low frequency in the AMP. In this context, *HaRIC\_B* may be thought of as a new and highly delimited QTL for SHR resistance.

## Conclusions

Identification of genes and alleles underlying resistance to SHR has been a major concern for sunflower research groups. The present work contributes to previous QTL studies by identifying the most highly delimited QTL for SHR reported to date. Validation experiments are currently ongoing to determine the specific role of *HaRIC\_B*, however, the association found is highly supported by (1) its specific expression pattern in the target organs of

*S. sclerotiorum* infection, (2) the role of its orthologous gene in Arabidopsis in altering cytosolic  $Ca^{2+}$  concentration, (3) the fact that  $Ca^{2+}$  plays an important part in reducing *S. sclerotiorum* virulence. Finally, our results demonstrate that association genetics via candidate genes is a valuable approach for elucidating the molecular basis of complex agronomic traits in sunflower, and for developing DNA-based markers for “precision breeding” of improved varieties.

## Methods

### Plant material

The sunflower Association Mapping Population (AMP) was composed of 94 inbred lines. Ten of them are public lines and have been previously used in the development of biparental populations, while the remaining 84 belong to the “Sunflower Breeding Program” of INTA. The inbred lines are described in Additional file 1 and additional information is available from the Sunflower Germplasm Bank.

### Genomic DNA extraction

The DNA was extracted with NucleoSpin® Plant II or NucleoSpin® 96 Plant II (Macherey-Nagel, Argentina) from lyophilized leaves from 3-week old plants grown in the experimental field.

### Field trials

Field experiments were conducted at Balcarce Experimental Station-INTA (37° 50' 0" S, 58° 15' 33" W, Province of Buenos Aires, Argentina) during growing seasons 2008/2009 (sowing date November 21<sup>st</sup>, 2008) and 2009/2010 (sowing date December 11<sup>th</sup>, 2009). Seeds were sown in typical Argiudol soil containing 5 % of organic matter at pH 6.2.

A randomized complete block design with two replications was used. Each experimental unit had one row 6.0 m length by 0.7 m wide, containing 30 plants. Sunflower capitula were sprayed with inoculum (1 cm<sup>3</sup>) at the R5.2 stage from the Schreiner and Miller's scale [56]. Capitula were immediately covered with paper bags up to 15 days post inoculation. A control to check the efficacy of the inoculation procedure (one susceptible cultivar inoculated at once with the tested inbred lines) was included. SHR incidence, i.e. the percentage of infected capitula in each row was evaluated at R9 stage [56].

### Inoculum preparation

A population of *S. sclerotiorum* from Balcarce (Buenos Aires, Argentina) was used for inoculum preparation. The sclerotia were collected in the field and stored in paper bags at 13° ± 5° for three months. For ascospore production, Escande et al. [57] procedure was followed. Shortly,



sclerotia were exposed at  $-18^{\circ} \pm 2^{\circ}$  for seven days and buried 1 cm deep in humid pasteurized soil until stipe emergence. Cultures were incubated at  $16^{\circ}$  and approximately 2500 lux of continuous daylight. Mature apothecia were harvested and positioned upside down in glass Petri dishes for 4 h to favor ascospore releasing. Ascospores were stored in Petri dishes at  $-18^{\circ}$  until use. For capitula inoculation, ascospores were suspended in water to a concentration of 2500 spores/cm<sup>3</sup>.

### Candidate gene selection

Candidate genes for SHR resistance were selected from three different data sources: (1) 19 loci, at least twofold over-expressed in a *B. napus* resistant genotype after 24 hours post-infection (hpi), were selected from the study of Zhao et al. [25]; (2) 17 candidate genes were selected according to their annotated function (e.g. biotic and abiotic stress related proteins, cell wall and cell membrane proteins) from an EST library obtained through suppressive subtractive hybridization from capitula of the genotype RHA801 infected with *S. sclerotiorum* at 48 hpi [38]; (3) seven candidate genes, i.e. two WRKY transcription factors and five germin-like proteins, previously identified in sunflower and reported to participate in the defense response against *S. sclerotiorum* infections in sunflower and other species were also selected [36,37]. The general features of the candidate genes are described in Additional file 3.

In the analysis of Zhao et al. [25], the transcriptional differences were detected using a 70-mer oligo-microarray from *Arabidopsis thaliana*, therefore, a phylogenetic approach was devised to identify the corresponding orthologous genes in sunflower. Briefly, a tBLASTn search [58] was performed against several data bases (GenBank Viridiplantae EST data base, KEGG *Helianthus annuus* data base, DFCI Sunflower Gene Index) using as query the *A. thaliana* protein sequences. A group of EST sequences with E-values  $< 10^{-20}$  and high similarity to each query ( $\geq 50\%$  of identity in amino acid sequence) was selected to represent different plant families, and was used to determine the position and minimum length of the regions to be included in the phylogenetic reconstruction. ESTs with hits over two regions with different Open Reading Frames (ORF) were excluded from the analysis. All alignments were required to include at least 150 amino acids. To avoid over representation of particular groups, a maximum of four ESTs was allowed per species, except for species of the genera *Helianthus* and *Lactuca*, for which all detected hits were considered for analysis. Once the sequences were selected, they were translated into the appropriate ORF and aligned using the MAFFT routine L-INS-i [59]. The protein alignments were used to obtain the phylogenetic relationships through Maximum Parsimony (MP) using the software TNT [60]. Heuristic searches were performed starting from 20

Wagner trees with tree bisection-reconnection branch-swapping (TBR). All characters were treated as equally weighted. Only when a low number of replicates found the shortest trees, an additional TBR branch-swapping search over the trees found previously was performed. Node support values in MP analyses were assessed using 100 jackknife replicates. Finally, sunflower orthologous and paralogous sequences were determined based on the phylogenetic relationships found for each locus. Orthology was assigned comparing gene trees with species trees to infer speciation/duplication events. In the absence of sunflower sequences orthologous to the *A. thaliana* query, sunflower paralogous and/or orthologous sequences from either *Lactuca* or other Asteraceae species were selected for subsequent analysis. The sequences selected for each locus were used as template for primer design.

### Primer design and PCR amplification

For those candidate genes for which *H. annuus* sequences were available, primer design was conducted using software Primer3 [61]. In those instances in which only *Lactuca* or other Asteraceae sequences were available, degenerated primers were designed either manually or using the software iCO-DEHOP [62].

Specific amplification products obtained with degenerated primers were purified with QIAquick PCR Purification Kit (QIAGEN, Germany). Afterwards, they were cloned on pGEMT-easy (Promega, USA) and at least ten colonies were sequenced. Sequences that yielded high similarity by BLAST searches [58] to the target sequence were used to design new non-degenerated primers with software Primer3 [61].

Primer functionality was assayed using RHA801 and HA89 genomic DNA as templates in the PCR amplification. PCR reactions were performed as described previously in Fusari et al. [35] using touchdown programs ranging from  $68^{\circ}$  to  $55^{\circ}$ .

Primer pairs yielding the PCR products of expected sizes were then used to amplify a total of ten inbred lines, the CS, to detect polymorphisms (SNPs and indels). PCR products were purified either as described previously or using ExoSAP (Exonuclease I & Shrimp Alkaline Phosphatase, USB, USA). Sequencing reactions, sequence analysis and polymorphism identification were done as described in Fusari et al. [35].

### Genotyping of candidate genes in the AMP

DNA polymorphisms found in the CS were used to define haplotypes for each candidate gene (Table 1). According to the complexity and number of haplotypes in each case, different genotyping methods were selected. When needed, new primers were designed with

Primer3 software [61] for genotyping purposes. Those genes which showed only two haplotypes in the CS were genotyped with denaturing dHPLC, those which had indels were genotyped by FCE and those with more than two haplotypes determined by both SNPs and indels were genotyped by direct sequencing (Table 1). The dHPLC and FCE genotyping methods were carried out as described in Fusari et al. [63]. For dHPLC, PCR products that did not match any of the haplotypes tested or displayed a heteroduplex-like profile in the homoduplex controls were subjected to direct sequencing. For FCE, PCR products of different length than those found in the CS were amplified with unlabeled primers and sequenced. All sequencing reactions and SNP analyses were performed as described previously [35].

In all cases, three individuals per inbred line were genotyped separately, making a total of 282 genotyped individuals.

#### Genotyping of SSR in the AMP

To study the genetic structure of the AMP, eight SSR loci located in different linkage groups of the sunflower genetic map were chosen from a preliminary survey of 35, based on their power to distinguish among individuals [46]. The selected loci (HA77, HA293, HA928, HA991, HA2063, HA2920, HA3239, HA4103) were genotyped in the 94 lines of the AMP (3 individuals per inbred line, 282 individuals in total) using FCE as previously described [63].

LG position and size data of SSR loci are available in Poormohammad Kiani et al. [64]. Although most of the lines were highly inbred, less than 10 % of the accessions exhibited a small proportion of heterozygosity. In such cases, a consensus genotype was established for the corresponding loci-line combination based on the genotypes of the three individuals scored. The criteria applied to designate the consensus genotype was as follows: (a) if two individuals carried the same homozygous genotype and the third individual was heterozygous carrying the allele present in the other two individuals, the inbred line was considered homozygous for such allele; (b) if two individuals were heterozygous and the third one was homozygous for any of the alleles detected in the other individuals, the inbred line was scored as heterozygous for those alleles; (c) if two out of three individuals were homozygous for different alleles, the inbred lines was considered heterozygous for those alleles. In those loci for which more than three alleles were present among the individuals, no consensus genotype was determined, and it was considered as missing data. The probability of identity (PI) and the probability of identity considering the genetic similarity among siblings (PI sibs) were computed from the AMP data as described by Waits et al. [65] using GenAlEx6 [66]. These statistics are widely used as an indication of the minimum number of

loci required for reliable genetic tagging [66]. The estimates obtained for the present AMP (PI = 0.000006; PI sibs = 0.004) are well within the ranged accepted in population genetics studies [65].

#### Statistical analysis

The statistical analysis was conducted according to the two-step method described in Stich et al. [47]. Haplotype-based tests were preferred to single-marker-based tests to increase power to detect associations.

The phenotypic data were analyzed on the basis of a Mixed Linear Model (MLM) encompassing inbred lines as fixed effect and the “trial” and “blocks within trials” as random effects:

$$Y_{ijk} = \mu + Line_i + Trial_j + Blocks(Trial)_{jk} + e_{ijk}$$

where  $Y_{ijk}$  represents Sclerotinia Head Rot incidence at the R9 stage,  $\mu$  a common mean value for every observation,  $Line_i$  the effect of the  $i$ th inbred line,  $Trial_j$  the random effect of the  $j$ th trial (2008/2009 and 2009/2010),  $Block(Trial)_{jk}$  the random effect of the  $k$ th block in the  $j$ th trial and  $e_{ijk}$  the random error of each observation.

Over the two trials, the adjusted inbred line mean was calculated for the 94 inbred lines:

$$\hat{M}_i = \hat{\mu} + \hat{Line}_i$$

The association analysis was performed with a MLM, where the adjusted inbred line mean was modeled for the  $j$ th candidate gene by:

$$\mathbf{M} = \mu\mathbf{1} + \mathbf{Xm}^{(j)} + (\mathbf{P} \text{ or } \mathbf{Q})\boldsymbol{\beta} + \mathbf{U}\mathbf{b} + \mathbf{e} \quad (1)$$

Where  $\mathbf{M}$  is the vector of length  $\ell$  (# inbred lines) of estimated SHR incidence (94 adjusted means),  $\mu$  is the common mean,  $\mathbf{1}$  is a  $\ell$ -length vector of ones,  $\mathbf{X}$  is an  $\ell \times h_j$  incidence matrix,  $\mathbf{m}^{(j)}$  is the haplotype effect vector of length  $h_j$ , of the  $j$ th candidate gene,  $\mathbf{P}$  (or  $\mathbf{Q}$ ) are  $\ell \times p$  matrices which account for the population structure,  $\boldsymbol{\beta}$  is a  $p$ -length vector of regression coefficients associated to  $\mathbf{P}$  (or  $\mathbf{Q}$ ) and  $\mathbf{U}$  is a  $\ell \times \ell$  matrix of weights for the  $\ell$ -length vector random effects  $\mathbf{b}$  of inbred lines effects such as  $\mathbf{K} = \mathbf{U}\mathbf{U}'$ . The assumptions for the random components of the model are:  $\mathbf{b} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{K})$  being  $\mathbf{K}$  the kinship matrix between inbred lines,  $\mathbf{e} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I})$ , and  $\text{cov}(\mathbf{b}, \mathbf{e}) = \mathbf{0}$ .

Population structure was taken into account by two different approaches: Principal Coordinates Analysis (PCO), which generates the  $\mathbf{P}$  matrix (94 inbred lines x 5 Principal Coordinates), and the Bayesian method implemented in STRUCTURE software, which generates the  $\mathbf{Q}$  matrix (94 inbred lines x membership coefficients to each of the 9 ideal populations detected) [39].

PCO was performed based on Roger's distance [67] obtained from the SSR allele frequency matrix (94 inbred

lines x 47 alleles) using InfoStat software [68]. The first five principal coordinates, which explain altogether 43 % of the variance, were used in the **P** matrix.

Population structure matrix **Q** was calculated with STRUCTURE software [39]. The number of *k* populations evaluated ranged from 0 to 12. The analysis was performed using five replicates per *k*, a burn-in period of 10<sup>5</sup> and a run length of 5\*10<sup>5</sup>. Allelic frequencies were kept correlated. No prior information on the origin of individuals was used to define groups. The run showing the highest posterior probability was considered for each *k* value. Estimation of the number of populations (*k*) was conducted following the software documentation and the *ad hoc* criterion proposed by Evanno et al. [40].

Kinship relationships among inbred lines were accounted for by the **K** matrix (94 x 94) proposed by Bernardo [41]. The matrix **K** is defined as  $\mathbf{K} = \left\{ K_{ij}^T \right\}$ , where  $K_{ij}^T = \frac{S_{ij}-1}{1-T} + 1$ , *S<sub>ij</sub>* is the proportion of shared alleles (SSR) between inbred lines *i* and *j*, and *T* the probability that two alleles are alike in state, given that they are not identical by descent. In practice, *T* is unknown, but different *T* values (0.2, 0.3, 0.7, 0.8) were evaluated to assess their effect on the likelihood of model (1). All MLM estimations were performed with the *lme* routine of R software [69].

FDR procedures to adjust *P*-values were not applied for three reasons: (a) we were working with a small number of genes which were selected *a priori* as candidates to be associated with the disease, (b) the inclusion of **P** (or **Q**) and **K** matrixes in the model has the purpose to control for the occurrence of false positives due to population structure and kinship relationships, and (c) FDR procedures tend to produce high false-negative error rates. Notwithstanding the maximum family wise error rate was estimated as a reference.

#### Characterization of genes associated to SHR resistance

Seeds from three inbred lines showing *HaRIC\_B* haplotypes 1, 2 and 3, respectively, were grown in the greenhouse. Leaves, stems and roots were harvested in liquid nitrogen at V8 developmental stage. Additionally, florets and receptacles of sunflowers at R3, R5.2 and R6 stages were harvested in liquid nitrogen separately. The RNA was extracted with RNAqueous<sup>®</sup> kit (Ambion<sup>®</sup>, USA), according to manufacturer's instructions. To remove polysaccharides and polyphenolics the tissue was homogenized in a mixture of RNA Isolation Aid (Ambion<sup>®</sup>, USA) and RNAqueous<sup>®</sup> Lysis/Binding solution. One μg of DNase-treated RNA was used to perform RT-PCR with the SuperScript III Reverse Transcriptase (Invitrogen, USA) using random primers according to the kit instructions.

To determine *HaRIC\_B* gene structure, genomic DNA and cDNA from R5.2 capitula of inbred lines carrying

haplotypes 1, 2 or 3 were used as template in PCR. To determine *HaRIC\_B* expression pattern cDNA from leaf, stem, root, dry seeds, florets and receptacles from capitula at the R3, R5.2 and R6 developmental stages were used for the RT-PCR. Amplifications were carried out as previously described using a touchdown program (60°–55°) with primers F: 5' TTGAGGGATTCTAATTG TTATAGTTGA 3' and R: 5' TTCGGGTGTTTCGTCCT TTT 3' for *HaRIC\_B* and F: 5' GGAGCAGAGAGATT CCGTTG 3' and R: 5' GAAGGTGCTGAGTGATGC AA 3' for *Actin*.

#### Additional files

**Additional file 1: Inbred Lines used for Association Mapping in Sunflower and Adjusted Means for Sclerotinia Head Rot Incidence.**

**Additional file 2: Analyses of population Structure in the AMP. A.**

Principal coordinate analysis of the AMP based on Rogers' distance estimates. Percentages in parentheses refer to the proportion of variance explained by the principal coordinate (PC). **B.** Data posterior probability (Ln *x/k*) and rate of change in the log probability of data between successive *k* values ( $\Delta k$ ). The values for *k*=9 are indicated by arrows. **C.** Estimated population structure at *k*=9. Each inbred line (bar) is partitioned into *k* colored segments that represent the individual's estimated membership fractions in *k* clusters.

**Additional file 3: Candidate Genes Selected for Sclerotinia Head Rot Resistance.**

#### Competing interests

The Authors declare no competing financial interests.

#### Authors' contributions

CMF selected the candidate genes along with VWL and NBP. CMF amplified the regions, carried out SNPs and indel development in the CS and performed candidate gene genotyping of the AMP. CT, CM and FQ planned and coordinate the phenotypic trials. CT collected the phenotypic data. DA selected the inbred lines to be included in the AMP. MVM and VN performed microsatellite genotyping. CMF and VL performed STRUCTURE and PCO analyses. JDR implemented the MLM and supervised the statistical analysis. RAH provided EST sequence information. CMF, VWL and NBP accomplished data and results interpretation. CMF, VWL and NBP wrote the manuscript. RAH and JDR helped to draft the manuscript. VWL and NBP coordinated the study. HEH and AE initiated the project and contributed to the work by the interpretation and discussion of the data. All authors read and approved the manuscript.

#### Authors' information

Dr. VWL; Dr. RAH and Dr. NBP are career members of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Dr. HEH is a career member of the Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC). CMF is current Alexander Von Humboldt Foundation's fellow at the Max Planck Institute of Molecular Plant Physiology (Germany). Verónica V. Lia and Norma B. Paniago co-directed the work and share last authorship.

Sequence Data is available at GenBank: JN231331 to JN231540.

#### Acknowledgments

This research was supported by ANPCyT/FONCYT, PID 2007 00073, INTA-PRR AEBIO 241001 and 245001. We thank Silvio Giuliano, Carlos Antonelli and Mauro Zabaleta for supporting field experiments at INTA Balcarce, and Martín Fernandez for support in greenhouses at INTA Castelar. We also want to thank Ing. Agr. Nora Mancuso, Ing. Agr. Pedro Ludueña and Ing. Agr. Julio Horacio González from Germplasm Bank-Estación Experimental Agropecuaria INTA Pergamino for providing breeding materials. We are also grateful to Dr. Andrea Puebla (Genomic Unit) for gene sequencing and Alberto Maligne for support in dHPLC runs.

#### Author details

<sup>1</sup>Instituto de Biotecnología, Centro Investigación en Ciencias Veterinarias y Agronómicas (CICVyA), Instituto Nacional de Tecnología Agropecuaria (INTA), 1686, Hurlingham, Buenos Aires, Argentina. <sup>2</sup>Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba, 5000, Córdoba, Argentina. <sup>3</sup>Estación Experimental Agropecuaria Balcarce, INTA, 7620, Balcarce, Buenos Aires, Argentina. <sup>4</sup>Estación Experimental Agropecuaria Manfredi, INTA, 5988, Manfredi, Córdoba, Argentina. <sup>5</sup>Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina.

Received: 30 November 2011 Accepted: 21 May 2012

Published: 18 June 2012

#### References

1. Jan C-C, Seiler GJ: **Sunflower**. In *Genetic resources, chromosome engineering, and crop improvement: oilseed crops*. Volume 4. Edited by Singh RJ. Boca Raton, FL: Taylor & Francis Group; 2007:166. 33487-2742.
2. Boland G, Hall R: **Index of plant hosts of Sclerotinia sclerotiorum**. *Can J Plant Pathol* 1994, **16**:93.
3. Mestries E, Gentzmittel L, de Labrouhe DT, Nicolas P, Vear F: **Analyses of quantitative trait loci associated with resistance to Sclerotinia sclerotiorum in sunflower (*Helianthus annuus* L.)**. *Mol Breed* 1998, **4**: 215-226.
4. Pereyra V, Escande A: *Enfermedades del girasol en la Argentina*. Manual de reconocimiento. Buenos Aires: INTA; 1994.
5. Van Becelaere G, Miller JF: **Combining ability for resistance to Sclerotinia head rot in sunflower**. *Crop Sci* 2004, **44**:1542-1545.
6. Rodriguez MA, Venedikian N, Bazzalo ME, Godeas A: **Histopathology of Sclerotinia sclerotiorum attack on flower parts of Helianthus annuus heads in tolerant and susceptible varieties**. *Mycopathologia* 2004, **157**:291-302.
7. Ronicke S, Hahn V, Vogler A, Friedt W: **Quantitative Trait Loci Analysis of Resistance to Sclerotinia sclerotiorum in Sunflower**. *Phytopathology* 2005, **95**:834-839.
8. Troglia C: *Factores ambientales que afectan la supervivencia y producción de inóculo de Sclerotinia sclerotiorum y umbral mínimo de inóculo para la podredumbre húmeda del capítulo de girasol*. In Master Thesis.: UNMDP, Facultad de Ciencias Agrarias, Unidad Integrada Balcarce; 2003.
9. Bert PF, Dechamp-Guillaume G, Serre F, Jouan I, de Labrouhe DT, Nicolas P, Vear F: **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* 2004, **109**:865-874.
10. Bert PF, Jouan I, De Labrouhe DT, Serre F, Nicolas P, Vear F: **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* 2002, **105**:985-993.
11. Micic Z, Hahn V, Bauer E, Melchinger AE, Knapp SJ, Tang S, Schon CC: **Identification and validation of QTL for Sclerotinia midstalk rot resistance in sunflower by selective genotyping**. *Theor Appl Genet* 2005, **111**:233-242.
12. Micic Z, Hahn V, Bauer E, Schon CC, Knapp SJ, Tang S, Melchinger AE: **QTL mapping of Sclerotinia midstalk-rot resistance in sunflower**. *Theor Appl Genet* 2004, **109**:1474-1484.
13. Micic Z, Hahn V, Bauer E, Schon CC, Melchinger AE: **QTL mapping of resistance to Sclerotinia midstalk rot in RIL of sunflower population NDBLOSSel x CM625**. *Theor Appl Genet* 2005, **110**:1490-1498.
14. Yue B, Radi SA, Vick BA, Cai X, Tang S, Knapp SJ, Gulya TJ, Miller JF, Hu J: **Identifying quantitative trait loci for resistance to Sclerotinia head rot in two USDA sunflower germplasms**. *Phytopathol* 2008, **98**:926-931.
15. Gentzmittel L, Mouzeyar S, Badaoui S, Mestries E, Vear F, Tourvielle De Labrouhe D, Nicolas P: **Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus* L.** *Theor Appl Genet* 1998, **96**:519-525.
16. Guo X, Stotz HU: **Defense against Sclerotinia sclerotiorum in Arabidopsis is dependent on jasmonic acid, salicylic acid, and ethylene signaling**. *Mol Plant Microbe Interact* 2007, **20**:1384-1395.
17. Hegedus DD, Li R, Buchwaldt L, Parkin I, Whitwill S, Coutu C, Bekkaoui D, Rimmer SR: **Brassica napus possesses an expanded set of polygalacturonase inhibitor protein genes that are differentially regulated in response to Sclerotinia sclerotiorum infection, wounding and defense hormone treatment**. *Planta* 2008, **228**:241-253.
18. Li R, Rimmer R, Buchwaldt L, Sharpe AG, Seguin-Swartz G, Coutu C, Hegedus DD: **Interaction of Sclerotinia sclerotiorum with a resistant Brassica napus cultivar: expressed sequence tag analysis identifies genes associated with fungal pathogenesis**. *Fungal Genet Biol* 2004, **41**:735-753.
19. Li R, Rimmer R, Buchwaldt L, Sharpe AG, Seguin-Swartz G, Hegedus DD: **Interaction of Sclerotinia sclerotiorum with Brassica napus: cloning and characterization of endo- and exo-polygalacturonases expressed during saprophytic and parasitic modes**. *Fungal Genet Biol* 2004, **41**:754-765.
20. Liu R, Zhao J, Xiao Y, Meng J: **Identification of prior candidate genes for Sclerotinia local resistance in Brassica napus using Arabidopsis cDNA microarray and Brassica-Arabidopsis comparative mapping**. *Sci China C Life Sci* 2005, **48**:460-470.
21. Perchepped L, Balague C, Riou C, Claudel-Renard C, Riviere N, Grezes-Besset B, Roby D: **Nitric oxide participates in the complex interplay of defense-related signaling pathways controlling disease resistance to Sclerotinia sclerotiorum in Arabidopsis thaliana**. *Mol Plant Microbe Interact* 2010, **23**:846-860.
22. Stotz HU, Sawada Y, Shimada Y, Hirai MY, Sasaki E, Krischke M, Brown PD, Saito K, Kamiya Y: **Role of Camalexin, Indole Glucosinolates, and Side Chain Modification of Glucosinolate-Derived Isothiocyanates in Defense of Arabidopsis against Sclerotinia sclerotiorum**. *Plant J* 2011, **67**(1):81-93.
23. Zhao J, Meng J: **Genetic analysis of loci associated with partial resistance to Sclerotinia sclerotiorum in rapeseed (*Brassica napus* L.)**. *Theor Appl Genet* 2003, **106**:759-764.
24. Zhao J, Udall JA, Quijada PA, Grau CR, Meng J, Osborn TC: **Quantitative trait loci for resistance to Sclerotinia sclerotiorum and its association with a homeologous non-reciprocal transposition in Brassica napus L.** *Theor Appl Genet* 2006, **112**:509-516.
25. Zhao J, Wang J, An L, Doerge RW, Chen ZJ, Grau CR, Meng J, Osborn TC: **Analysis of gene expression profiles in response to Sclerotinia sclerotiorum in Brassica napus**. *Planta* 2007, **227**:13-24.
26. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, et al: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness**. *Nat Genet* 2006, **38**:203-208.
27. Jorde L: **Linkage disequilibrium and the search for complex disease genes**. *Genome Res* 2000, **10**:1435.
28. Brescghello F, Sorrells ME: **Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars**. *Genetics* 2006, **172**: 1165-1177.
29. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES: **Dwarf8 polymorphisms associate with variation in flowering time**. *Nat Genet* 2001, **28**:286-289.
30. Pajerowska-Mukhtar K, Stich B, Achenbach U, Ballvora A, Lubeck J, Strahwald J, Tacke E, Hofferbert HR, Ilarionova E, Bellin D, et al: **Single nucleotide polymorphisms in the allene oxide synthase 2 gene are associated with field resistance to late blight in populations of tetraploid potato cultivars**. *Genetics* 2009, **181**:1115-1127.
31. Krill AM, Kirst M, Kochian LV, Buckler ES, Hoekenga OA: **Association and linkage analysis of aluminum tolerance genes in maize**. *PLoS One* 2010, **5**: e9958.
32. Simko I, Pechenick DA, McHale LK, Truco MJ, Ochoa OE, Michelmore RW, Scheffler BE: **Association mapping and marker-assisted selection of the lettuce dieback resistance gene Tvr 1**. *BMC Plant Biol* 2009, **9**:135.
33. Liu A, Burke JM: **Patterns of nucleotide diversity in wild and cultivated sunflower**. *Genetics* 2006, **173**:321-330.
34. Kolkman JM, Berry ST, Leon A, Slabaugh MB, Tang S, Gao W, Shintani DK, Burke JM, Knapp SJ: **Single Nucleotide Polymorphism and Linkage Disequilibrium in Sunflower**. *Genetics* 2007, **177**:457-468.
35. Fusari CM, Lia W, Hopp HE, Heinz RA, Paniago NB: **Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach**. *BMC Plant Biol* 2008, **8**:7.
36. Fernandez P, Paniago N, Lew S, Hopp HE, Heinz RA: **Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project**. *BMC Genomics* 2003, **4**:40.
37. Giacomelli JI, Ribichich KF, Dezar CA, Chan RL: **Expression analyses indicate the involvement of sunflower WRKY transcription factors in stress**

- responses, and phylogenetic reconstructions reveal the existence of a novel clade in the Asteraceae. *Plant Sci* 2010, **178**:398–410.
38. Peluffo L: Characterization of the defense mechanisms against *Sclerotinia sclerotiorum*, causal agent of Sunflower Head Rot, through metabolic and transcriptional profiles analyses. In *PhD thesis: University of Buenos Aires, Faculty of Exact and Natural Sciences*. 2010 [digital.bl.fcen.uba.ar/Download/Tesis/Tesis\_4765\_Peluffo.pdf]
39. Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000, **155**:945–959. <http://www.pritch.bsd.uchicago.edu>
40. Evanno G, Regnaut S, Goudet J: Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 2005, **14**:2611–2620.
41. Bernardo R: Estimation of coefficient of coancestry using molecular markers in maize. *Theor Appl Genet* 1993, **85**:1055–1062.
42. Chiu JC, Lee EK, Egan MG, Sarkar IN, Coruzzi GM, DeSalle R: OrthologID: automation of genome-scale ortholog identification within a parsimony framework. *Bioinformatics* 2006, **22**:699–707.
43. Fulton DL, Li YY, Laird MR, Horsman BG, Roche FM, Brinkman FS: Improving the specificity of high-throughput ortholog prediction. *BMC Bioinforma* 2006, **7**:270.
44. Koski LB, Golding GB: The closest BLAST hit is often not the nearest neighbor. *J Mol Evol* 2001, **52**:540–542.
45. Ehrenreich I, Hanzawa Y, Chou L, Roe J, Kover P, Purugganan M: Candidate gene association mapping of Arabidopsis flowering time. *Genetics* 2009, **183**(1):325–335.
46. Moreno MV: Genetic diversity in sunflower: analysis of a local germplasm collection as a tool to assist management and breeding programs. In *PhD thesis: University of Buenos Aires, Faculty of Exact and Natural Sciences*; 2011.
47. Stich B, Mohring J, Piepho HP, Heckenberger M, Buckler ES, Melchinger AE: Comparison of mixed-model approaches for association mapping. *Genetics* 2008, **178**:1745–1754.
48. Gu Y, Fu Y, Dowd P, Li S, Vernoud V, Gilroy S, Yang Z: A Rho family GTPase controls actin dynamics and tip growth via two counteracting downstream pathways in pollen tubes. *J Cell Biol* 2005, **169**:127.
49. Wu G, Gu Y, Li S, Yang Z: Genome-Wide Analysis of Arabidopsis Rop-Interactive CRIB Motif-Containing Proteins That Act as Rop GTPase Targets. *Plant Cell* 2001, **13**:2841–2856.
50. Ingvarsson PK, Street NR: Association genetics of complex traits in plants. *New Phytol* 2011, **189**:909–922.
51. Lee YJ, Szumlanski A, Nielsen E, Yang Z: Rho-GTPase-dependent filamentous actin dynamics coordinate vesicle targeting and exocytosis during tip growth. *J Cell Biol* 2008, **181**:1155.
52. Bashi ZD, Hegedus DD, Buchwaldt L, Rimmer R, Borhan MH: Expression and regulation of *Sclerotinia sclerotiorum* necrosis and ethylene-inducing peptides (NEPs). *Mol Plant Pathol* 2010, **11**:43–53.
53. Rafalski JA: Association genetics in crop improvement. *Curr Opin Plant Biol* 2010, **13**:174–180.
54. Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashv S, Bruggemann E, et al: Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci U S A* 2007, **104**:11376–11381.
55. Maringolo CA: *Regiones Cromosómicas Asociadas a Resistencia a Podredumbre Húmeda del Capítulo de Girasol (Sclerotinia sclerotiorum (Lib.) de Bary)*. FCA: UNMdP, Posgrado en Producción Vegetal; 2007.
56. Schneiter AA, Miller JF: Description of sunflower growth stages. *Crop Sci* 1981, **21**:901–903.
57. Escande AR, Laich FS, Pedraza MV: Field testing of honeybee-dispersed *Trichoderma* spp. to manage sunflower head rot (*Sclerotinia sclerotiorum*). *Plant Pathol* 2002, **51**:346–351.
58. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 1990, **215**:403–410.
59. Katoh K, Misawa K, Kuma K, Miyata T: MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl Acids Res* 2002, **30**:3059–3066.
60. Goloboff PA, Farris JS, Nixon KC: TNT, a free program for phylogenetic analysis. *Cladistics* 2008, **24**:774–786.
61. Rozen S, Skaletsky H: Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Edited by Krawetz S, Misener S. Totowa, NJ: Humana Press; 2000:365–386. <http://frodo.wi.mit.edu/>
62. Boyce R, Chilana P, Rose TM: iCODEHOP: a new interactive program for designing CONsensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences. *Nucl Acids Res* 2009, **37**:W222–228.
63. Fusari C, Lia V, Nishinakamasu V, Zubrzycki J, Puebla A, Maligne A, Hopp H, Heinz R, Paniego N: Single nucleotide polymorphism genotyping by heteroduplex analysis in sunflower (*Helianthus annuus* L.). *Mol Breeding* 2011, **28**:73–89.
64. Poormohammad Kiani S, Talia P, Grieu P, Maury P, Hewezi T, Gentzbittel L, Sarrafi A: Genetic analysis of plant water status and osmotic adjustment in recombinant inbred lines of sunflower under two water treatments. *Plant Sci* 2007, **178**:773–787.
65. Waits LP, Luikart G, Taberlet P: Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Mol Ecol* 2001, **10**:249–256.
66. Peakall R, Smouse PE: Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 2006, **6**: 288–295.
67. Rogers JS: *Measures of genetic similarity and distance*. Texas: Studies Genetics VII University Texas Publ 1972, **7213**:145–153.
68. Di Rienzo JA, Casanoves F, Balzarini MG, Gonzalez L, Tablada M, Robledo CW: *InfoStat versión 2009*. Argentina: Grupo InfoStat, FCA, Universidad Nacional de Córdoba; 2009.
69. Pinheiro JC, Bates DM: *Mixed-effects models in S and S-PLUS*. NewYork: Springer; 2009.

doi:10.1186/1471-2229-12-93

Cite this article as: Fusari et al.: Association mapping in sunflower for sclerotinia head rot resistance. *BMC Plant Biology* 2012 **12**:93.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

