International Scholars Journals

International Journal of Plant Breeding and Genetics ISSN 2756-3847 Vol. 10 (6), pp. 001-009, June, 2023. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

Marker Development and Fine-Mapping of Downy Mildew (*Plasmopara halstedii*) (DM) Resistance *PlArg* in Sunflower (*Helianthus annuus*) using nested-association-mapping (NAM) Populations

Zining Wang (✉), Wayne (Hui) Shen, Géza Mészáros, Kathy Smith, David Pang, Steve Calhoun

Zining Wang(✉),Wayne Shen, David Pang, S & W Seed Company, TTU Innovation Hub at Research Park, 3911 4th St, M.S. 3005, Lubbock TX 79415, USA.
Géza Mészáros, S & W Seed Company, 11 Új Élet utca, Szeged, H-6710, Hungary.
Kathy Smith, S&W Seed Company, Nampa Laboratory, 4819 Lewis Lane, NAMPA, Idaho 83686, USA.
Steve Calhoun, S&W Seed Company, 6615 Winfield St, Lubbock TX 79407, USA.

Abstract

Accepted 15 March, 2023

Downy Mildew (*Plasmopara halstedii*) in Sunflower (*Helianthus annuus*) is a serious disease worldwide. The resistance gene *PlArg* from *Helianthus argophyllus* has been confirmed to be the most effective gene to date against DM races. Previously reported molecular markers for *PlArg* are of limited value for marker assisted selection (MAS) across diverse germplasm. In this study, we used the published markers, NAM populations, two consensus genetic maps and three versions of genome assemblies to test the flanking markers D2384 and D2395. We also developed a new SSR marker D2564, co-segregating with *PlArg* but in a genomic region not previously reported. After further screening new SSR markers in this region using the small population, D2317 was the closest marker to *PlArg*. Further, we tested D2317 with a larger NAM population of 256 families and a $F_{2:3}$ population of 126 families. It co-segregated with *PlArg*. D2317 has four bands in the genome and lands in two resistance-like genes, HanXRQChr01g0018861 and HanXRQChr01g0018891 separately. In this study, we used comparative genomic tools and NAM populations to develop the causal marker D2317 and find candidate genes for *PlArg*. This marker will have broad applications for MAS and will aid the cloning of *PlArg* resistance genes.

Keywords: Sunflower, marker development, fine mapping, *Plarg*, DM resistance, recombination depression, candidate genes, MAS.

INTRODUCTION

Sunflower is a very important oil and confection crop in the world. It ranks the third in production after soybean and rapeseed (Shahbandeh, 2020). Sunflower is a crop with many diseases and pests that cause great yield loss. Among the diseases, DM, caused by the biotrophic Oomycete *Plasmopara halstedii* (Berlese et de Toni), has long been considered one of the most devastating sunflower diseases worldwide (Viranyi 2015). Before 1980, there were only a few

Corresponding Author Email: ziningwang@swseedco.com
Tel: 806-646-0976

races detected in the sunflower farming regions. In recent decades, DM has spread and mutated quickly in the sunflower farming area in Europe and America. Viranyi (2015) listed 24 *P. halstedii* races in Europe and 40 races in the Americas. Thirty-six resistance genes to DM races and their differential lines have been identified (Ma et al., 2019; Spring 2019). For commercial breeding, it is very important to develop sunflower lines with broad resistance to most or all DM races.

The resistance loci *PlArg* from *H. argophyllus* has been confirmed to be resistant to all the DM races found so far (GASCUEL et al., 2015; Gilley et al., 2020). Breeders have developed sunflower lines with *PlArg* resistance to DM. The first line, ARG 1575-2, is a bulk of 41 self-pollinated F5 plants derived from the cross CMS HA89/accession *Helianthus argophyllus* Raf.-1575 (PI 468651) and was released by the USDA and the North Dakota Agricultural Experiment Station in 1989 (Seiler, 1991). RHA 419 and RHA 420 are F4-derived F6 fertility restorer lines advanced by pedigree selection from the cross RHA 373/ ARG 1575-2. RHA 373 is a restorer line released by the USDA and the North Dakota Agricultural Experiment Station in 1990. RHA 443 and RHA 464 with *PlArg* were introduced from selections within RHA 419 Gulya and Seiler. 2002; Hulke et al. 2010). As a broad-spectrum resistance gene, *PlArg* is effective against all known races of *P. halstedii* (Gascuel et al. 2015; Gilley et al. 2016; Gilley et al., 2020) and will have a broad application in breeding.

Genetic maps for *PlArg* have been published since 2004 (Table 1). Dußle et al. (2004) first mapped *PlArg* on LG1 of the sunflower genome using SSR markers. Wieckhorst et al. (2010) identified two markers co-segregating with *PlArg*. Livaja et al. (2013) mapped two single nucleotide polymorphism (SNP)-derived cleaved amplified polymorphic sequence (CAPS) markers in the *PlArg* region. Other markers were also mapped to the *PlArg* gene region (Qi et al., 2017; KÖSOĞLU et al., 2017; Pecrix et al., 2018; Şahin et al., 2018; Ramazanova et al., 2021). Although DNA markers have been discovered co-segregating with or closely linked to *PlArg*, most were developed from bi-parental populations which potentially limits their value in other populations. In addition, the region containing *PlArg* is reported to show suppressed recombination (Qi et al., 2017; Qi et al., 2020) which can limit the development of causal makers and identification of candidate genes for *PlArg*. Therefore, development of a causal marker for *Plarg* is essential for MAS.

Two genomic tools have been developed for sunflower genetic research. First, two high-density consensus SNP maps of sunflower have been constructed with over 15,000 SNPs markers (Bowers et al. 2012; Talukder et al. 2014; Hulke et al. 2015). Second, three versions of genome assemblies have been completed, namely Helianthus annuus r1.2, HA412HO bronze assembly and HA412HOv2.0 assembly (Badouin et al., 2017; http://sunflowergenome.org). These whole genome genetic maps and genome sequences will greatly facilitate the marker development and map-based cloning of sunflower genes.

Nested association mapping (NAM) populations are another tool for marker development and gene identification. Buckler et al., (2009) first used NAM to study the genetic architecture of maize flowering time. Compared with the biparental populations, NAM populations have more diversified germplasm in the populations and increased their diversity of genetic background (Sallam et al., 2020). The recombination depression can be decreased using NAM populations due to increased polymorphisms in the target region. Another advantage is that co-segregating markers found from NAM populations are more likely causal markers that will work for more diverse germplasm in MAS.

In this study, we used the modified NAM populations, genome sequences and published maps to finely map the *PlArg* resistance gene, developed causal markers for MAS and identified the possible candidate genes. The causal marker found will have a broader application in the sunflower breeding programs.

MATERIALS AND METHODS

Two NAM populations from the breeding materials were constructed for mapping purposes. The crosses were made between the common parent RHA 419that carries the *PlArg* resistance allele and other parental lines, namely ST 4391, STR 4409, STR 4365, STR 4415, STR 4395-3, STR 4293HO, STR 4023OR, STR 4027HO and STR 4023, by the contractor (Héctor J. Martinuzzi) of S & W Seed Company.

The first NAM population (Table 2) was designed for screening the published markers and the initial fine mapping of *PlArg*. The F_3 families were inoculated with DM race 714. Race 714 was chosen because it shows good phenotypes and it can overcome most of the DM resistance genes except a few resistance genes including *Plarg*. Thirteen segregating families were chosen to make the NAM populations. From each F_3, two resistant plants and two susceptible plants were chosen to make up the populations with 52 plants.

The second NAM populations were designed to finely map the *PlArg* gene with 256 families from 10 populations (Table 3). The F_2 and F_3 populations were planted in Lubbock, Texas for DNA extraction in 2020. The harvested seeds of nine $F_{3:4}$ families and one $F_{2:3}$ family were shipped to Nampa, Idaho for phenotyping. The 256 families were used for the final mapping.

The phenotyping was done in Nampa Idaho in 2020. Briefly, 40 seeds were surface sterilized and placed on damp germ paper in petri plates. After 72 hours, germinated seeds were immersed in 20 ml spore solution (30-50,000 spores/ml) of race 714 and incubated at room temperature for 3 - 5 hours. Seedlings were transferred to flats with soil and placed in the greenhouse at 17 - 24°C with 12-16 hours of light. Lumens averaged 11,000 - 16,000 lux. After 7 days, flats were covered with domes (100% rel. hum.) and placed in a dark growth room at 16°C for 40 hours. This environment induces sporulation. Any sporulation on leaves or stems was considered

susceptible phenotype.

For DNA extraction, CTAB DNA extraction method was used (Xin and Chen, 2012). DNA concentration was quantified using Nano Drop 2000 (Thermo Fisher Scientific, Dallas, Texas). For the SSR genotyping, the PCR components consist of 5 µl 2 × JumpStart, 0.6 µl primer mix, 50 ng DNA and 4.1 µl dH2O. For PCR reactions, the initial cycle was 94°C for 3'; then 35 cycles of 94°C for 30'', 60°C for 30'', 72°C for 30''; and 72°C for 5' as the final extension followed by 4°C for overnight. The PCR was performed on T-100 Thermo Cycler (Bio-Rad Laboratories Inc, California, USA), run on a 3% Metaphor gel and visualized on Azure Biosystem 300 (Azure Biosystems, Inc, California, USA). For STS markers from the publication, the PCR components and the genotyping process were the same as SSR genotyping except that the recommended Tm was used and the PCR was run and visualized in 1.2% agarose gel.

For the SNP genotyping, the PCR components followed the PACE-2.0-User-Guide (3CR Bioscience). For PCR reactions, the initial cycle was 95°C for 15'; then 10 cycles of 95°C for 20'', 65-57°C touch-down for 60''each cycle dropped 0.8°C; and 34 cycles of 95°C for 20'', 57°C for 60''; PCR stopped at 30°C at the final step.

For the data analyses, Chi-squared analyses were carried out on the $F_{3:4}$ population segregating for the *PlArg* gene to detect deviations from the expected Mendelian ratios for co-dominant (1:2:1) or dominant (3:1) markers. Genetic map was constructed using Join Map Version 5.1 (Van Ooijen, 2011) using an LOD score of 7.0 under default settings. Recombination fractions were converted to centimorgans (cM) using the mapping function of Kosambi (Kosambi, 1994).

RESULTS

Screening of published markers in the *PlArg* region

To test the published markers in the *PlArg* region, we chose the two flanking SNPs from the publication (Qi and Ma, 2020) first, namely NSA_005063 and NSA_002851 that have been used for MAS. They were renamed as D2384 and D2395 separately in this study. After screening the small NAM population of 52 families (Table 2), markers D2384 and D2395 had two recombinants each and are not close to the resistance gene.

To better understand the *PlArg* region in sunflower genome and decide the fine-mapping region, we aligned the markers with the three versions of sunflower genome sequences and two genetic maps (Table 4), namely Helianthus annuus r1.2, HA412HO bronze assembly, HA412HOv2.0, the SNP genetic map (Talukder et al., 2014), and the SSR genetic map (Bowers et al., 2012). The versions of genome sequences are quite different from each other. The previous mapping of *Plarg* was on Chr.01. The region was flanked by D2384 and D2395 and covers 119.5-134.4 Mb of Chr.01 ofHA412HOv2.0, 123.3-137.8 Mb of Chr.01 of HA412HO bronze assembly, and possibly 115.7-126.1 Mb of Chr.01 of Helianthus annuus r1.2, in addition that marker D2395 aligned to Chr.17 instead of Chr.01. These regions on Chr.01 could not align very well with any markers of SSR map (Bowers et al., 2012). However, the region flanked by D2384 and D2359 overlapped with some SNP markers on the SNP map (Talukder et al., 2014) at 20.277 cM of Chr.01. We identified 17 SNP markers from this map (Table 3), including NSA_004291,002867, 005063, 008037,003789, 002851, 004270, 007595, 001835, 004107, 004078, 003438, 004590, 002989, 003892, 004494, and 002898. We screened the small population with the 17 SNP markers. The results showed that these markers had at least two recombinants; none of these markers were closer to the loci than D2384 and D2395.To make the mapping simpler, these markers were not included in the map (Fig.1).

Marker development in the region of *PlArg*

To find a closer marker in the region flanked by D2384 and D2395, new SSR and SNP primers were designed from the corresponding regions of the three versions of genome assembly (Table 4). They were D2513, D2582, D2564, D2574, D2627, and D2584. These markers were used to screen the first NAM population. To our surprise, marker D2564 co-segregated with *PlArg* on the first genetic map (Fig. 1) but is not flanked by D2384 and D2395. This is different from previous report (Qi and Ma 2020) and confirmed that *Plarg* is not flanked by D2384 and D2395.

To test if marker D2564 is on a resistance gene or close by, we did blast searches using the primer sequences against the three versions of genome sequences. In the Bronze version, there was no resistance-like genes in the 20 genes flanking D2564. In the HA412HOv2.0 version, Ha412HOChr01g00002354-RA was found to be one protein kinase-like geneclose to the D2564 marker. In the Helianthus annuus r1.2 version, four disease resistance-like genes were identified to flank D2564, including HanXRQChr01g0021231 (kinase), HanXRQChr01g0021241 (protein kinase-like), HanXRQChr01g0021251 (putative protein kinase) and HanXRQChr01g0021261 (protein kinase-like). Therefore, SSR primers of D2827-D2842 were designed on the five gene sequences. After screening the small population with these primer pairs, no marker was found to be co-segregating with *PlArg*. The results showed that the genome region flanking D2564 may be misassembled or D2564 happens to not be closely related to any of these five kinase-related genes because we could not find co-segregating markers in these genes. To make the genetic map simpler, these markers were not included in the map (Fig. 1).

Screening published markers and newly designed primers from the new region

In order to identify other cosegregating markers and the can-

Table 1. Mapping populations markers and races used in publications.

No.	Populations	Race	R:H:S	Distortion	Cosegregating or Flanking Markers	Marker types	References
1	cmsHA342 × Arg1575-2	730	28:82:16	Yes	ORS-662	SSR	Dußle et al., 2004
2	cmsHA342 x ARG1575-2	730	26:114:45	Yes	HT211; ORS662	SSR	Wieckhorst et al., 2011
3	cmsHA342 x ARG1575-2	330, 100, 710, 730	25:114:44	Yes	ORS662, HT211,RGC52a, RGC52b, RGC151, HT722, and ORS716	SSR, CAPs	Wieckhorst et al., 2010
4	RHA 419 x RHA-N-49	730	20:39:24	No	ORS716, ORS662 and ORS675	SSR	Imerovski et al., 2014
5	HA 89 x RHA464	734	32:76:31	No	NSA_007595 and NSA_001835	SNP	Qi et al., 2017
6	RHA-419 × Colombi				NSA002867 and NSA006138	SNP	KÖSOĞLU et al., 2017
	RHA-419 × P64LC53				NSA002867 and NSA006138	SNP	KÖSOĞLU et al., 2017
	RHA-419 × Oliva				NSA002867 and NSA006138	SNP	KÖSOĞLU et al., 2017
7	RHA419 x 9758R		63:29:10	Yes	Marker tests	SSR	Şahin et al., 2018
	RHA-419 x CL		3:17:02	Yes			Şahin et al., 2018
	RHA-419 x OL		4:16:05	Yes			Şahin et al., 2018
	RHA-419 x P64LC53		4:25:05	Yes			Şahin et al., 2018
8	HA-R3 x RHA 464	734	Selection		ORS 610, NSA_005063, NSA_002851	SSR, SNP	Qi et al., 2020

Table 2. The small NAM populations for the screening of public marker and initial mapping.

COD-1/pedigree	Pedigree	Race Testing	No. Plants
S&W-901	STR 4391/RHA 419-F3	714	4
S&W-902	STR 4391/RHA 419-F3	714	4
S&W-903	STR 4391/RHA 419-F3	714	4
S&W-908	STR 4409/RHA 419-F3	714	4
S&W-909	STR 4365/RHA 419-F3	714	4
S&W-911	STR 4365/RHA 419-F3	714	4
S&W-913	STR 4415/RHA 419-F3	714	4
S&W-916	STR 4395-3/RHA 419-F3	714	4
S&W-917	STR 4293HO/RHA 419-F3	714	4
S&W-918	STR 4023OR/RHA 419-F3	714	4
S&W-920	STR 4027HO/RHA 419-F3	714	4
S&W-921	STR 4027HO/RHA 419-F3	714	4
S&W-922	STR 4027HO/RHA 419-F3	714	4

Table 3. The largeNAM populations for the final fine mapping of *PlArg.*

Plot No.	Pedigree	Race Testing	No. Plants
901B	(STR4391/RHA419)-F4	714	12
902B	(STR4391/RHA419)-F4	714	18
909A	(STR4365/RHA419)-F4	714	10
913B	(STR4415/RHA419)-F4	714	16
916A	(STR4395-3/RHA419)-F4	714	28
918A	(STR4023OR/RHA419)-F4	714	29
918B	(STR4023OR/RHA419)-F4	714	27
921A	(STR4027HO/RHA419)-F4	714	29
921B	(STR4027HO/RHA419)-F4	714	26
95576/2	(STR4023/RHA419)-F3	714	61

didate genes, other regions outside the region flanked by D2384 and D2395 were considered. Because the marker D2564 is outside of the flanking region of D2384 and D2395in the genetic map, we picked some published SSR markers from outside region of the D2395 from the genetic maps (Bowers et al., 2012; Pecrix et al., 2018; Şahin et al., 2018; Ramazanova et al., 2021). These markers were also aligned to the three versions of genomes sequences (Table 4). SSR markers ORS606, ORS716, RGC151, HT446, ORS662, HT722 and ORS675 are linked to *PlArg* at 26.87 cM of Chr.01 of the map. After screening the small population, these markers did not co-segregate with *Plarg* (Fig. 1).These markers are germplasm related markers, but not causal markers for *Plarg*.

To develop cosegregating markers in the region at 26.87 cM of the SSR map (Bowers et al., 2012), we designed new SSR

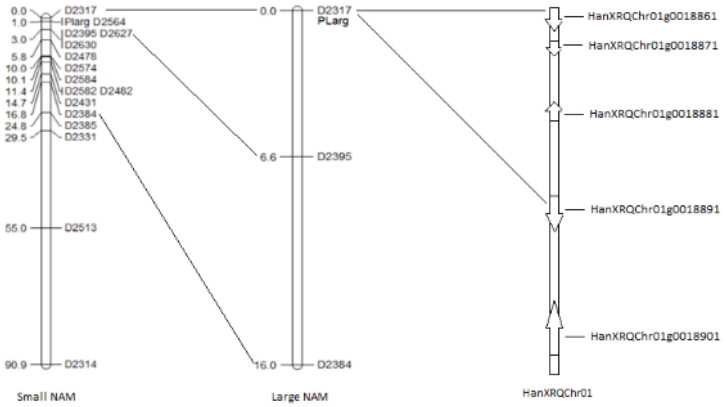

Fig. 1 Marker development, fine mapping and identification of candidate genes for *PlArg* (HanXRQ assembly is used to show the locations of the candidate genes).

primers in the region of the published sequences, identified as markers D2323, D2324, D2317, D2516, D2506, D2443, D2460, D2432 and D2428. After screening the first NAM population with these markers, D2317 had two missing data and was the closest marker to *PlArg* except D2564 (Fig.1).

To finely map the important markers, we screened the large NAM population with 256 families (Table 3) using D2317, D2384 and D2395. Marker D2317 co-segregated with the resistance gene. D2384 and D2395 were 6.6cM and 16.0cM away from the gene (Fig.1), respectively. Because the large NAM populations were used, D2317 is likely a causal marker that can work for all the germplasm we used in this study. To confirm this marker works for other germplasm, we tested a $F_{2:3}$ population of 126 families different from the NAM populations. It showed 100 match with the phenotypes. Therefore, D2317 is a causal marker developed and tested by NAM populations so far.

At the same time, to confirm that *PlArg* is controlled by one dominant gene, we did statistical analysis for the phenotypes of the larger population (Table 5) with 256 families. The ratio of resistant: segregating: susceptible plants fits a 1:2:1 model. *PlArg* is controlled by one dominant locus.

Sequence analysis and candidate gene prediction

Marker D2317 clearly showed the resistant, susceptible and heterozygous genotypes (Fig. 2). The susceptible genotype has one shortest band; the resistant genotype has four bands, the shortest band and three resistance specific bands; the heterozygous genotype has four bands but with half strengths of the bands from resistant and susceptible genotypes.

The primer sequences of D2317 were used to do blast searches against the Phytozome 12 version of genome sequence and showed 100% match to the aligning sequences in genes HanXRQChr01g0018861 and HanXRQChr01g0018891 (Fig. 1). Both genes are disease resistance related genes containing a toll/interleukin-1 receptor homology (TIR) domain, NB-ARC and leucine-rich-repeat (LRR) domain. The proposed amplicons for both genes are 240 and 239 bp separately.

Wieckhorst (2011) indicated that RGC151SW1 is a promising candidate for *Plarg*, but not yet confirmed its function. Sequence blasts showed that HanXRQChr01g0018891 has the highest similarity to RGC151SW1. Our results showed similar mapping results to the previous study. However, D2317 is the only causal marker developed so far.

DISCUSSION

NAM populations are a good tool to relieve recombination depression

In previous studies for *PlArg* mapping, biparental populations were used to map this gene (Table 1). The advantage is that it is easier to make biparental populations than to make NAM populations. These genetic maps have lower resolutions in the *PlArg* region. Markers show linkage to the gene in one population and may not work for other populations (Wieckhorst et al., 2010; Wieckhorst2011). Further, there is depressed recombination (Wieckhorst et al., 2010; Qi et al., 2017) in the *PlArg* region. Identification of closer or on-gene markers has been limited by the suppression. All these disadvantages limit the identification of candidate genes and on-gene co-segregating markers. To overcome these disadvantages, we developed NAM populations (Table 2 and 3) using RHA 419 as the common (trait donor) parent. We tested the published SSR markers and the SNP markers from all the published genetic maps for *PlArg* and both the consensus large genetic maps (Bowers et al. 2012; Talukder et al. 2014) using the small populations. We did not find a

Table 4. Alignment of markers, genome sequences and genetic maps.

Marker ID	Type	Helianthus annuus r1.2	HA412HO bronze assembly	HA412HOv2.0	Bowerset al. (2012) Map	Talukderet al. (2014) Map	Public Marker ID	References
D2502	SSR	35827585..35827792	162258968..162259111	19999140..19999283			ORS610	Qi et al., 2020
D2323	SSR			113314521..113314685				New marker
D2332, D2325	SNP	110646690..110647004	116583138..116582842	114604063..114603767			NSA_006138	KÖSOĞLU et al., 2017
D2501	SSR			116342693..116342672	26.87		ORS606..1	Bowers et al., 2012
D2484	SNP			117342352..117342674			NSA_004078	Talukder et al., 2014
D2324	SSR			118676474..118676474				New marker
D2487	SNP			117487189..117486775		20.277	NSA_004291	Talukder et al., 2014
D2315	SSR	105695755..105696124	157825291..157825610	120489332..120489651	26.87		ORS716	Bowers et al., 2012
D2317	SSR						RGC151	New marker
D2333	SNP	111757968..111850120	119724113..119879076	121802691..121802575			RGC151	Wieckhorst, 2011
				121804219..121803346	26.87		RGC151	Bowers et al., 2012
D2516	SSR	102245855..102245911	130597173..130597199	121973516..121973516				New marker
D2506	SSR	101654626..101654683	131028482..131028984	122012400..122012899			Co7_11	New marker
D2504	SSR			122012400..122012899			Co3..2	Bowers et al., 2012
D2505	SSR			122012400..122012899			Co3..4	Bowers et al., 2012
D2507	SSR			124008286..124007793	26.87		HT446	Bowers et al., 2012
D2314	SSR	106696433..106696796	129881321..130089712	124000933..124001247	26.87		ORS662	Wieckhorst, 2011
D2316	STS	106703496..106704036	129874109..129874599	124008286..124007793	26.87		HT722	Wieckhorst, 2011
D2443	SSR			124807125..124807325				New marker
D2503	SSR			129389642..129389885	26.87		ORS675	Bowers et al., 2012
D2460	SSR		123472114..123471861					New marker
D2432	SSR		123435718..123458233					New marker
D2428	SSR		123359147..123359329					New marker
D2478	SNP	113640940..113640996	129149733..129149755	130042935..130043008		20.277	NSA_002867	Talukder et al., 2014
D2331, D2324	SNP	113640925..113640981	129149733..129150065	130043123..130042791			NSA_002867	KÖSOĞLU et al., 2017
D2467			145479019..145478997	130043123..130042791				Talukder et al., 2014
D2431	SSR	115003229..115003285	123444092..123444300	126940928..126940744				New marker
D2384*	SNP	115737891..115738736	123281157..123281478	119492024..119491703		20.277	NSA_005063	Talukder et al., 2014; Qi et al., 2020
D2513	SSR	122230734..122230790	135380517..135380491	131865100..131865074				New marker
D2582	SNP	122952117..122952173	135860544..135860619	132631438..132631515				New marker
D2564	SSR	122941914..122941971	135850554..138516384	132641698..132642014				New marker
D2574	SSR	122913797..122913854	138463737..138463714	132670638..132670848				New marker
D2627	SNP		137582238..137582219	133371199..133371218				New marker
D2584	SNP	124244873..124244929	137582236..137582124	133371201..133371313				New marker
D2630	SNP	123185736..123185793	137830487..137830404	133625683..133625600			NSA_008037	Talukder et al., 2014; Qi et al., 2020
D2482	SNP		216,708,457	134212305..134212736		20.277	NSA_003789	Talukder et al. 2014
D2385, D2395*	SNP	N/A CHR17	124006084..124006420	134300280..134299940		20.277	NSA_002851	Talukder et al., 2014; Qi et al., 2020
D2486	SNP		90114947	134399689..134399294		20.277	NSA_004270	Talukder et al., 2014
D2396	SNP	126114647..126114718	140507252..140507591	136351328..136351667			NSA_007595	Qi, 2017
D2397	SNP	131048146..131048717	143344108..143343691	137814402..137814819			NSA_001835	Qi, 2017
D2485	SNP		73,098,204			20.277	NSA_004107	Talukder et al., 2014
D2483	SNP		91,790,916			20.277	NSA_004078	Talukder et al., 2014
D2481	SNP		96,021,650			20.277	NSA_003438	Talukder et al., 2014
D2489	SNP		118,946,696			20.277	NSA_004590	Talukder et al., 2014
D2480	SNP		120,929,802			20.277	NSA_002989	Talukder et al., 2014
D2482	SNP		123,609,294			20.277	NSA_003892	Talukder et al., 2014
D2488	SNP		125,266,908			20.277	NSA_004494	Talukder et al., 2014
D2479	SNP		125,830,250			20.277	NSA_002898	Talukder et al., 2014

Fig. 2. Screening of some populations using D2317. R: resistant; S: susceptible; H: heterozygous. 1-4 indicate the four bands on the metaphor gel. Band 1-4 sizes are about 254, 248, 244 and 240 bp.

Published marker co-segregating with *PlArg* in the NAM population.

We further developed new SSR markers in the region flanked by D2384 and D2395. Most markers showed different distances as expected. However, D2564 was identified to cosegregate with *PlArg* in the small NAM populations, showing that the NAM populations work better than biparental populations and can overcome the disadvantages of biparental populations and closer markers can be identified. The mapping of D2564 and D2317 showed that the *PlArg*

Table 5. Phenotypes of the populations.

No. Plants	R	H	S	Total
Observed	66	134	56	256
Expected	64	128	64	256

X^2=1.34; $\alpha_{0.01}$=9.210

gene is out of the region flanked by D2384 and D2395, further showing the mapping of this gene using NAM populations can be more accurate than using biparental populations. The published SSR and SNP markers from biparental populations are not cosegregating markers in the NAM populations and could not be good markers for MAS. Even further, the identification of cosegregating markers D2384 and D2317 showed that the recombination suppression in the *PlArg* region has been relieved by using the NAM populations. Finally, the identification of D2317 as a causal marker was a result using NAM populations that is more powerful than biparental populations.

Comparative mapping is also a good tool for marker development

We first tested the published markers in the *PlArg* region flanked by D2384 and D2395 (Table 1 and Table 4). These markers are mainly at 20.227 cM of the SNP map (Talukder et al., 2014) which corresponds to the location of 20.4 Mb on Chr.01 in the genome. We did not find any markers cosegregating with *PlArg*. To find a closer marker, we aligned this map to the three different versions of the sunflower genome sequence assembly (Table 4). Because we do not know which version is the best, we developed markers in the corresponding regions of the three versions. We found the newly designed marker D2384 from the version Phytozome 12 was cosegregating with *PlArg* in the small populations. However, the marker developed in the nearby disease resistance-like genes still showed distance to the resistance gene *PlArg*, indicating that the marker D2384 is an isolated marker misassembled in this region. Our initial genetic map indicates that the *PlArg* is outside the region flanked by D2384 and D2395. Therefore, we further aligned the published markers of Bower's map (Bowers et al., 2012) to the three versions of genome sequences (Table 4). The published markers in the Bower's map are mainly located at 28.67 cM of the map. The region covers about 15 Mb of the genome. After we tested the published SSR and CAPs markers using our small NAM populations, they were not cosegregating with *PlArg*. We further designed new primers for SSR and STS in this region. Marker D2317 cosegregated with the resistance *PlArg*. Even though we spent a lot of time and made an effort to figure out the right gene region for *PlArg*, the comparative mapping of markers, maps and genome sequences helped us find the right target region and the co-segregating markers. One possibility is that the *Plarg* loci was from a wild species and the gene arrangements may be different from the sequenced lines. Or this gene region has structural diversity of the sequenced lines. Therefore, comparative genomics is such a good tool to use in marker development.

Causal variations could be identified using NAM populations

Among the published genetic maps for *PlArg*, one map (Wieckhorst, 2011) indicated a few markers cosegregating with the gene using biparental populations. They are ORS662, ORS716, RGC151 and HT722. We tested these markers after we targeted the gene to the 28.67 cM of the Bower's map. Unfortunately, these markers did not cosegregate with *PlArg* in our NAM populations (Fig.1; Table 4). This indicated that these markers work for specific germplasm used for mapping in the publications and they may not be the causal variations around the resistance genes. In marker assisted selection for *PlArg* resistance, the application of these markers will be limited. To further develop markers cosegregating with *PlArg*, we designed primers in the region and found the cosegregating marker D2317 in both the small and large NAM populations of 50 and 256 families. To further tests if this marker works for other germplasm, we tested a $F_{2:3}$ population of 126 families from our breeding materials that were not included in the NAM populations. Genotypes of D2317 and the phenotypes of the families all matched with each other. This shows that D2317 works for the germplasm used and may be a causal marker good for MAS for *PlArg* resistance in the sunflower breeding program. D2317 is on two of the candidate resistance genes. Further function test of the D2317 is needed to verify the genetic variation is the cause of the phenotype. To find other cosegregating markers other than D2317, we designed SSR primers and SNP primers in the candidate genes (Table 4). The tests using NAM populations showed that these markers did not cosegregate with the resistance and they were not markers from causal variations. This result again confirmed that D2317 is so far the only causal marker for *PlArg* resistance.

Candidate genes have been discovered

Wieckhorst (2011) indicated that RGC151SW1 from ARG1575-2 is a promising candidate by annotation, RT-PCR and copy number analysis. To understand the relationship between our marker D2317 and the candidate genes, we aligned the marker primer sequences to the gene. The proposed marker sequence in ARG1575-2 has 253 bp and 239 bp in HA342.

The size difference is because of a 14 bp INDEL. To further understand how many copies of the marker sequences in the genome, we blast the 239 bp sequence against the Helianthus annuus r1.2 genome sequence. The sequence landed on the kinase-related genes HanXRQChr01g0018891 with 100% identity covering 239 bp; 96% identity on HanXRQChr01g0018861 with a coverage of 240 bp; the two genes have 100% identity in the primer sequences. This may indicate that the two counterpart genes in RHA 419 genome may be related to the resistance. Further in Fig. 2, we can see clearly in the resistant lines, there are four bands, namely 1-4. This suggests that maybe there are four copies of resistance genes in the RHA 419 genome. Beside the counterpart genes to HanXRQChr01g0018861 and HanXRQChr01g0018891 in RHA419 genome, there may be two more copies of genes together conferring the high level of resistance. These four possible candidate genes should come from *Helianthus argophyllus* genome. Sequencing of its genome may reveal the four candidate genes. In order to elucidate how the marker D2317 works and which genes confer the *PlArg* resistance, we are sequencing the amplicons of the marker and the candidate genes.

ACKNOWLEDGEMENTS

We are grateful to Dr. Lili Qi for her communications on her published SNP markers and related germplasm. We thank Drs. William Underwood and Christopher Misar from USDA for providing the strains of DM for phenotyping. We also thank Emilian Onisan, S&W's previous breeder for useful discussions of choosing the breeding materials for constructing the NAM populations.

Author contribution

Z.W. designed the project and performed most of the experiments; W. S. designed, led and organized the project; G. M. provided the populations; D.P. did some experiments; K. S. screened the populations; S.C. approved and led the project.

Funding

This work was supported by S & W Seed Company.

Data availability

All data supporting the results of this study are available within the article.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests

REFERENCES

Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Brière C, Owens GL, Carrère S, Mayjonade B, Legrand L, Gill N, Kane NC, Bowers JE, Hubner S, Bellec A, Bérard A, Bergès H, Blanchet N, Boniface MC, Brunel D, Catrice O, Chaidir N, Claudel C, Donnadieu C, Faraut T, Fievet G, Helmstetter N, King M, Knapp SJ, Lai Z, Le Paslier MC, Lippi Y, Lorenzon L, Mandel JR, Marage G, Marchand G, Marquand E, Bret-Mestries E, Morien E, Nambeesan S, Nguyen T, Pegot-Espagnet P, Pouilly N, Raftis F, Sallet E, Schiex T, Thomas J, Vandecasteele C, Varès D, Vear F, Vautrin S, Crespi M, Mangin B, Burke JM, Salse J, Muños S, Vincourt P, Rieseberg LH, Langlade NB (2017) The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature 546:148–152. https://doi.org/10.1038/nature22380

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez Villeda H, da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. *Science* 325: 714–718. doi:10.1126/science.1174276

Bowers JE, Bachlava E, Brunick RL, Rieseberg LH, Knapp SJ, Burke JM (2012) Development of a 10,000 Locus Genetic Map of the Sunflower Genome Based on Multiple Crosses, *G3 Genes|Genomes|Genetics*, 721–729. https://doi.org/10.1534/g3.112.002659

Dußle CM, Hahn V, Knapp SJ, Bauer E (2004) *PlArg* from *Helianthus argophyllus* is unlinked to other known downy mildew resistance genes in sunflower. Theor Appl Genet109:1083-6. doi: 10.1007/s00122-004-1722-9.

Gascuel Q, Martinez Y, Boniface MC, Vear F, Pichon M, Godiard L (2015) The sunflower downy mildew pathogen Plasmopara halstedii. Mol Plant Pathol16:109-22. doi: 10.1111/mpp.12164.

Gilley MA, Markell SG, GulyaTJ and Misar CG (2016) Prevalence and virulence of *Plasmopara halstedii* (downy mildew) in sunflowers. In: Proceedings of the 38th Sunflower Research Forum, Fargo ND. 12–13 Jan. 2016. National Sunflower Association, Bismarck, ND.

Gilley MA, Gulya TJ, Seiler GJ, Underwood W, Hulke BS, Misar CG, and Markell SG (2020). Determination of Virulence Phenotypes of *Plasmopara halstedii* in the United States. Plant Disease. 104:2823-2831. https://doi.org/10.1094/PDIS-10-19-2063-RE

Gulya T and Seiler GJ (2002). Registration of five restorer sunflower germplasm. Crop Science 42: 989-991

Hulke BS, Miller JF, Gulya TJ, Vick BA (2010). Registration of the oilseed sunflower genetic stocks HA 458, HA 459, and HA 460 possessing genes for resistance to downy mildew. J Plant Reg 4:93–97

Hulke BS, Grassa CJ, Bowers JE, Burke JM, Qi LL, Talukder ZI, Rieseberg LH (2015). A unified single nucleotide polymorphism map of sunflower (*Helianthus annuus L.*) derived from current genomic resources. Crop Sci 55:1696–1702.

Imerovski I, Aleksandra Dimitrijevic, Dragana Miladinovic, Sinisa Jocic, BoskoDedic, Sandra Cvejic, Gordana Surlan-Momirovic (2014). Identification and validation of breeder-friendly DNA markers for *PlArg* gene in sunflower. Mol Breeding 34:779–788 DOI:10.1007/s11032-014-0074-7

Kosambi D (1994). The estimation of map distances from recombination values. Ann Eugenics. 12:172–5.

Kösoğlu K, Yumuk S, Aydın Y, Evci G, Uncuoğlu A (2017). Use of SNP markers by KASP assay for MAS studies in sunflower against *Plasmopara halstedii*. Turkish Journal of Agriculture and Forestry 41: 480-489.

Livaja M, Wang Y, Wieckhorst S *et al.* (2013) BSTA: a targeted approach combines bulked segregant analysis with next-generation sequencing and *de novo* transcriptome assembly for SNP discovery in sunflower. *BMC Genomics* 14, 628. doi.org/10.1186/1471-2164-14-628

Ma G, Song Q, Underwood WR, hang ZW, Fiedler JD, Li XH, Qi LL (2019). Molecular dissection of resistance gene cluster and candidate gene identification of Pl_{17} and Pl_{19} in sunflower by whole-genome resequencing. *Sci Rep* **9**, 14974. https://doi.org/10.1038/s41598-019-50394-8

Pecrix Y, Penouilh-Suzette C, Muños S, Vear F and Godiard L (2018). Ten broad spectrum resistances to downy mildew physically mapped on the sunflower genome. Front. Plant Sci. 9:1780. doi: 10.3389/fpls.2018.01780

Qi L, Ma G (2020). Marker-assisted gene pyramiding and the reliability of using SNP markers located in the recombination suppressed regions of sunflower (*Helianthus annuus L.*). Genes 11: 10. Available from https://doi.org/10.3390/genes11010010.

Qi LL, Talukder ZI, Hulke BS, Foley ME (2017). Development and dissection of diagnostic SNP markers for the downy mildew resistance genes *PlArg* and Pl8 and maker-assisted gene pyramiding in sunflower (*Helianthus annuus L.*). Mol. Genet. Genom. 292: 551–563.

Ramazanova SA, Badyanov EV, Guchetl SZ (2021).Validation of microsatellite markers to identify Pl6, Pl8 and Plarg genes that control resistance to *Plasmopara halstedii* in sunflower. Caspian Journal of Environmental Sciences, Vol. 19 No. 5 pp. 915-920 DOI: 10.22124/CJES.2021.5266.

Şahin EÇ, Kalenderoğlu A, Aydın Y, Evci G, Uncuoğlu AA (2018).SSR Markers Suitable for Marker Assisted Selection in Sunflower for Downy Mildew Resistance. Open Life Sciences.13:319-326. DOI: 10.1515/biol-2018-0039. PMID: 33817099; PMCID: PMC7874726.

Sallam AH, Manan F, Bajgain P, Martin M, Szinyei T, Conley E, Brown-Guedira G, Muehlbauer G J, Anderson JA, & Steffenson B J (2020). Genetic architecture of agronomic and quality traits in a nested association mapping population of spring wheat. Plant Genome, 13, e20051. https://doi.org/10.1002/tpg2.20051

Seiler GJ (1991) Registration of 13 downy mildew tolerant interspecific sunflower germplasm lines derived from wild annual species. Crop Sci. 31:1714-1716

Shahbandeh M (2020). Global oilseed production 2020/21https://www.statista.com/statistics/267271/worldwide-oilseed-production-since-2008/

Spring O (2019) Spreading and global pathogenic diversity of sunflower downy mildew – Review. Plant Protect. Sci., 55: 149–158

Talukder ZI, Gong L, Hulke BS, Pegadaraju V, Song Q, Schultz Q, Qi L (2014). A high-density SNP Map of sunflower derived from RAD-sequencing facilitating fine-mapping of the rust resistance gene R12. PLoS One. 9(7):e98628. doi: 10.1371/journal.pone.0098628. PMID: 25014030; PMCID: PMC4094432.

Van Ooijen JW(2011). Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genetics Research 93:343-349

Viranyi F, Gulya TJ, Tourvieille de Labrouhe D (2015). Recent changes in the pathogenic variability of *Plasmopara halstedii* (sunflower downy mildew) populations from different continents. *Helia* 38: 149–162

Wieckhorst S, Bachlava E, Dußle CM, Tang S, Gao W, Saski C, Knapp SJ, Schön CC, Hahn V, Bauer E (2010). Fine mapping of the sunflower resistance locus *PlArg* introduced from the wild species *Helianthus argophyllus*. Theor. Appl. Genet121: 1633–1644

Wieckhorst S (2011). Characterization of the *PlArg* locus mediating resistance against *Plasmopara halstedii* in sunflower. TECHNISCHE UNIVERSITÄT MÜNCHEN.

Xin ZG and Chen JP (2012) A high throughput DNA extraction method with high yield and quality. Plant Methods 8:26