



# Fine mapping and candidate gene screening of the downy mildew resistance gene *RPF1* in Spinach

Hongbing She<sup>1</sup> · Wei Qian<sup>1</sup> · Helong Zhang<sup>1</sup> · Zhiyuan Liu<sup>1</sup> · Xiaowu Wang<sup>1</sup> · Jian Wu<sup>1</sup> · Chunda Feng<sup>2</sup> · James C. Correll<sup>2</sup> · Zhaosheng Xu<sup>1</sup>

Received: 11 June 2018 / Accepted: 17 August 2018 / Published online: 22 September 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

**Key message** A SLAF-BSA approach was used to locate the *RPF1* locus. The three most likely candidate genes were identified which provide a basis for cloning the resistance gene at the *RPF1* locus.

**Abstract** Spinach downy mildew is a globally devastating oomycete disease. The use of downy mildew resistance genes constitutes the most effective approach for disease management. Hence, the objective of the present study was to fine map the first-reported resistance locus *RPF1*. The resistance allele at this resistance locus was effective against races 1–7, 9, 11, 13, and 15 of *Peronospora farinosa* f. sp. *spinaciae* (*P. effusa*). The approach fine mapped *RPF1* using specific-locus amplified fragment sequencing (SLAF-Seq) technology combined with bulked segregant analysis. A 1.72 Mb region localized on chromosome 3 was found to contain *RPF1* based on association analysis. After screening recombinants with the SLAF markers within the region, the region was narrowed down to 0.89 Mb. Within this region, 14 *R* genes were identified based on the annotation information. To identify the genes involved in resistance, resequencing of two resistant inbred lines (12S2 and 12S3) and three susceptible inbred lines (12S1, 12S4, and 10S2) was performed. The three most likely candidate genes were identified via amino acid sequence analysis and conserved domain analysis between resistant and susceptible inbred lines. These included *Spo12729*, encoding a receptor-like protein, and *Spo12784* and *Spo12903*, encoding a nucleotide-binding site and leucine-rich repeat domains. Additionally, based on the sequence variation in the three genes between the resistant and susceptible lines, molecular markers were developed for marker-assisted selection. The results could be valuable in cloning the *RPF1* alleles and improving our understanding of the interaction between the host and pathogen.

---

Communicated by Richard G.F. Visser.

---

Hongbing She and Wei Qian have contributed equally to this work.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00122-018-3169-4>) contains supplementary material, which is available to authorized users.

---

✉ James C. Correll  
jcorrell@uark.edu

✉ Zhaosheng Xu  
xuzhaosheng@caas.cn

<sup>1</sup> Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China

<sup>2</sup> University of Arkansas, Fayetteville, AR, USA

## Introduction

Spinach downy mildew, caused by *Peronospora farinosa* f. sp. *spinaciae* (*Pfs*) Byford (= *P. effusa*), is one of the most destructive diseases of spinach worldwide (Qian et al. 2016). A major outbreak of downy mildew can significantly reduce spinach yield and quality, thus posing a serious threat to production. Developing new resistance cultivars is the most successful approach for managing spinach downy mildew (Correll et al. 1994).

A number of downy mildew resistance genes have been identified in other plants. A total of 28 downy mildew resistance genes (designated RPP for Recognition *Peronospora Parasitica*) have been reported in *Arabidopsis*, among which six genes (*RPP1*, *RPP4*, *RPP5*, *RPP8*, *RPP13*, and *RPP27*) have been cloned (Slusarenko and Schlaich 2003; Wang and Cao 2006; Nemri et al. 2010). In addition, the resistance genes *DM3*, *PI8*, and *MrRPP1* were cloned and characterized in lettuce, sunflower, and grapevine, respectively

(Meyers et al. 1998; Radwan et al. 2005; Feechan et al. 2013). To date, all the cloned downy mildew resistance genes are members of known plant disease resistance (*R*) gene classes. *R* genes encode proteins (*R* proteins) that play an essential role in defense against many pathogens and pests, including bacteria, viruses, fungi, and oomycetes, as well as nematodes and insects (Dangl and Jones 2001; Meyers et al. 2005). *R* genes are grouped into different classes according to their protein structure. Most *R* genes encode proteins with an extracellular nucleotide-binding site (NBS) and leucine-rich repeats (LRRs). Based on the structure of the N-terminal, these NBS-LRR-type proteins have been divided into two classes: One is called TIR-NBS-LRR or TNL proteins which have a N-terminal TIR (toll/mammalian interleukin 1 receptor) domain, and the other is known as CC-NBS-LRR or CNL proteins which encode a CC (coiled-coiled) motif (Dangl and Jones 2001; Meyers et al. 2005). The second class encodes Ser/Thr kinases (S/TK) such as the *Pto* gene from tomato, which confers resistance to *Pseudomonas syringae* pv tomato. (Martin et al. 1993). The third class contains receptor-like kinases (RLKs) with an extracellular LRR domain followed by a transmembrane region and a cytoplasmic Ser/Thr kinase domain, such as the rice *Xa21* gene, which confers resistance to *Xanthomonas oryzae* pv. *oryzae* (Song et al. 1995). Other *R* genes encode receptor-like proteins (RLPs), which are similar to RLK genes. Compared with RLKs, RLPs lack a cytoplasmic Ser/Thr kinase domain. The class contains the tomato *Cf-2*, *Cf-4*, *Cf-5*, and *Cf-9* genes that confer resistance to *Cladosporium fulvum* (Tör et al. 2004). The fifth class of *R* genes only encodes a possible coiled-coil domain. Members of this class encompass genes such as the *RPW8* gene, which is a broad-spectrum downy mildew resistance gene in *Arabidopsis thaliana* (Dangl and Jones 2001; Xiao et al. 2001). To date, all of the cloned genes conferring resistance to the downy mildew pathogen belong to the NBS-LRR class, except for *RPP27*, which was the first gene that was found to encode RLPs in *Arabidopsis*. A total of 139 NBS-LRR genes were identified in the spinach genome, and CNL and NL constitute the most abundant classes of NBS-LRR genes, representing 40 and 36% of the NBS-LRR genes, respectively (Xu et al. 2017).

A total of six spinach downy mildew resistance genes/alleles (*RPF1-RPF6*, **Resistance** against *Peronospora Farinose*) have been identified to date. However, none of these resistance genes have been cloned (Correll et al. 2011). Sixteen races have been reported since 1824, 12 of which have been identified in the past 20 years (Correll et al. 2011; Plantum 2016; Kubota et al. 2017). Race identification depends on the disease reactions on a set of diverse host differentials including open-pollinated, commercial hybrids, and breeding lines. The host differentials contain six resistance genes that confer resistance to

multiple races of *Pfs*. For instance, *RPF1* provides resistance to race 1–7, 9, 11, 13, and 15, whereas the *RPF2* gene provides resistance against race 1–10 (Irish et al. 2003, 2008; Feng et al. 2014). The codominant molecular marker DM-1, developed from an amplified fragment length polymorphism (AFLP) marker, is closely linked ( $\approx 1.7$  cM) to the *RPF1* locus (Irish et al. 2008). In addition, the marker RGA 5B14r, which was identified from a spinach bacterial artificial chromosome (BAC) library, is closely linked to the *RPF1* locus with a genetic distance of 1.7 cM and co-segregated with the marker DM-1 (Feng et al. 2015). Although the DM-1 marker has been widely used for marker-assisted selection (MAS), the distance of the marker from the suspected resistance gene limits its full utility. Currently, a draft genome of spinach has been published, and five genes close to DM-1 markers were deemed to be potential candidate genes of *RPF1* (Xu et al. 2017).

Bulked segregate analysis (BSA) is a method used for rapidly identifying DNA markers that are tightly linked to regions of interest. Two gene pools are generated via the selection of a pair of bulked DNA samples of individuals from a segregating population with extreme phenotypes (e.g., resistant and susceptible to a specific disease). Thus, the two bulked DNA samples differ genetically in the region of interest but otherwise have the same genetic background (Michelmore et al. 1991). BSA has been extensively used for mapping significant traits in many crops over the past two decades, such as rice (Venuprasad et al. 2009), maize (Quarrie et al. 1999), and barley (Poulsen et al. 1995). In recent years, the advent of advanced next-generation sequencing (NGS) technologies has led to the establishment of a large number of methods for marker development, including restriction site-associated sequencing (RAD-Seq) (Baird et al. 2008), whole-genome resequencing (WGRs) (Xie et al. 2010), genotyping by sequencing (GBS) (Elshire et al. 2011), and specific-locus amplified fragment sequencing (SLAF-Seq) (Sun et al. 2013); the latter being used in the current study. The combination of BSA and SLAF-Seq technologies is regarded as an efficient and powerful approach to map important traits in many plants such as rice (Zheng et al. 2016), cucumber (Liang et al. 2016), barley (Qin et al. 2015), maize (Xia et al. 2015), and sweet cherry (Jin et al. 2016).

In this study, the resistant inbred line 12S3, a female and recurrent parent containing the resistance gene at the *RPF1* locus, and the susceptible inbred line 12S4, a male parent lacking the resistance gene at the *RPF1* locus, were used as parents. The objectives of the present study were to fine map the *RPF1* locus and identify the candidate genes via BSA-SLAF-Seq technology in the BC<sub>1</sub> and F<sub>2</sub> populations derived from the crosses of F<sub>1</sub>/12S3 and 12S3/12S4, respectively. The results of the current study will play a crucial role in

the cloning of the resistance gene and developing disease-resistant cultivars.

## Materials and methods

### Plant materials

The  $F_1$  population was developed from a cross between the inbred lines 12S3/12S4. 12S3, a female and recurrent parent with the *RPF1* locus, is resistant to 11 races of *Pfs* (*Pfs*1–7, 9, 11, 13, 15); however, 12S4, a male parent, is susceptible to these races. A mapping population of the 148  $BC_1$  and 200  $F_2$  lines was generated by the backcrossing of  $F_1$ /12S3 and self-pollination of  $F_1$  plants, respectively.  $BC_1$  population were used for preliminary mapping of *RPF1*, and  $F_2$  population were inoculated to genetic analysis and fine mapping. All of these materials were obtained from the Spinach Research Group, Institute of Vegetables and Flowers (IVF), Chinese Academy of Agricultural Sciences (CAAS).

### Inoculation and genetic analysis

The  $F_2$  populations were planted in commercial potting mix in 28 cm × 42 cm flats, and eight 12S4 plants were grown per flat as a control group. All of these plants were cultivated in a greenhouse in 2015. Seedlings with two true leaves (25–30 day-old plants) were spray-inoculated with a previously reported sporangial suspension ( $2.5 \times 10^5$  sporangial/mL) of *Pfs* race 9 (Qian et al. 2016). The inoculum was applied using a sprayer (Fig. S6a). After being inoculated, the plants were placed in a dew chamber maintained at 18–20 °C with 100% relative humidity for 24 h in dark and then transferred to a growth chamber maintained at 18–20 °C with a 12 h photoperiod for 6 days. After the 6 days, the plants were returned to the dew chamber for 24 h in dark, and then seedlings were scored as resistant or susceptible as previously described by Correll et al. (2010) (Fig. S6b).

The segregation ratios of the  $F_2$  populations were analyzed using a Chi-square test ( $\chi^2$ ) with SAS software (version 9.2; SAS Institute Inc., Cary, NC, US).

### DNA extraction and SLAF library construction for high-throughput sequencing

The fresh leaves from the two parents (12S3 and 12S4) and the  $BC_1$  individuals were collected and then frozen in liquid nitrogen prior to DNA extraction. Total genomic DNA was extracted using the cetyltrimethyl ammonium bromide (CTAB) method reported by Murray and Thompson (1980). The DNA quality and concentration were assessed by electrophoresis on 1.0% agarose gels and an ND-2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE,

USA). DNA samples from the two parents and 148  $BC_1$  individuals were used for SLAF library construction and sequencing.

Initially, SLAF-pre-design experiment was performed to avoid repeated SLAFs and obtain an even distribution of SLAFs for maximum SLAF-Seq efficiency. The SLAF library was then constructed using the pre-design scheme. The *RsaI* and *HaeIII* (New England Biolabs, NEB) restriction enzymes were used to completely digest the genomic DNA from each sample separately. After digestion, adenine and duplex tag-labeled sequencing adapters were added to the DNA fragments. Following this, 20 polymerase chain reaction (PCR) cycles were conducted to ensure sufficient concentration of the DNA fragments to meet the sequencing requirements. The PCR products were then purified and pooled, and the sample was assessed using 2% agarose gel electrophoresis. Following gel purification, DNA fragments of 364–414 bp were excised and diluted for paired-end sequencing. Finally, SLAF-Seq was performed using the Illumina HiSeq 2500 platform (Illumina, Inc; San Diego, CA, USA), and then single-nucleotide polymorphism (SNP) genotyping and evaluation were performed (Sun et al. 2013).

### Analysis of SLAF-Seq data and association analysis

The SLAF-Seq data was analyzed using the procedures detailed in Sun et al. (2013). All SLAF paired-end reads with clear index information were clustered based on sequence similarity, which was detected using one-to-one alignment in BLAST (Kent 2002). Sequences with over 95% identity were grouped into one SLAF locus. SLAFs with two, three, or four tags were identified as polymorphic SLAFs.

Based on the genotypes of the  $BC_1$  population as assessed using marker DM-1, the sequencing data of 50 homozygous resistant (RR) individuals and 50 heterozygous resistant (Rr) individuals were selected as the sequencing results of the RR pool and Rr pool, respectively. The association analysis was subsequently performed by Euclidean distance (ED) followed by Loess regression analysis (Loess fit with a polynomial exponent of 1) using the Akaike Information Criterion (AICc), which identifies the regions containing the mutation and generates a list of trait-related candidate regions. ED association analysis was calculated as follows:

$$ED = \sqrt{(A_{RR} - A_{Rr})^2 + (T_{RR} - T_{Rr})^2 + (C_{RR} - C_{Rr})^2 + (G_{RR} - G_{Rr})^2}$$

In this study,  $A_{RR}$ ,  $G_{RR}$ ,  $C_{RR}$ ,  $T_{RR}$  indicate the depth of four types of base in the RR pool while  $A_{Rr}$ ,  $G_{Rr}$ ,  $C_{Rr}$ ,  $T_{Rr}$  represent the depth of four types of base in the Rr pool. In theory, the higher the ED value, the closer the object site (Geng et al. 2016).

## Fine mapping of the *RPF1* locus

To narrow down the location of the *RPF1* region, polymorphic SLAF markers in the predicted region were designed using Kompetitive Allele Specific PCR (KASP) assay primers to screen recombinants from the susceptible individuals of the F<sub>2</sub> population. Additionally, the codominant marker DM-1 was also designed using the KASP assay primer (DM-1-43), based on the sequence information of the dominant band and recessive band. All these KASP primers were designed by the LGC company (Shanghai, China). The primer sequences of the 22 markers are shown in Table S1. KASP was performed in a total reaction volume of 5  $\mu$ L containing 2.5  $\mu$ L 2 $\times$  KASP Master mix, 0.07  $\mu$ L KASP Assay mix, and 2.5  $\mu$ L genomic DNA diluted to 20–30 ng/ $\mu$ L. The reactions took place in 384-well PCR plates, and two no-template controls (NTCs) were included on each plate. All amplifications were performed on a Veriti 384 Well Thermal Cycler (Applied Biosystems, Foster City, CA, US) under the following conditions: 15 min at 94 °C; followed by ten cycles of 20 s at 94 °C and 60 s at 61 °C (0.6 °C drop per cycle), achieving a final annealing temperature of 55 °C; followed by a further 26 cycles of 20 s at 94 °C and 60 s at 55 °C. All plates should be read below 40 °C in a 7900 HT Fast Real-Time PCR System (Applied Biosystems), and the data were analyzed using SDS2.3 software (supplied by Applied Biosystems). If sufficiently defined genotype clusters are not obtained after the initial KASP thermal cycles, an additional thermal cycle is required, namely three cycles of 94 °C for 20 s, and 57 °C for 60 s, after which the data are assessed again. The additional thermal cycling and plate reading were carried out until defined genotype clusters were obtained.

## Analysis of candidate genes

Based on the spinach genome annotations in SpinachBase (<http://www.spinachbase.org/cgi-bin/spinach/tool/download.cgi>), all *R* genes in the final mapping regions were selected as candidate genes.

Resequencing data were obtained from five spinach lines, including the susceptible (10S2, 12S1, and 12S4) and resistant (12S2 and 12S3) individuals. Prior to mapping, low-quality reads were filtered according to the method of Cheng et al. (2016). The high-quality reads were used for mapping, which was conducted using the Burrows-Wheeler alignment (BWA) software (version 0.7.15; parameter “mem”). SNPs and InDels were detected in the genomes of the five lines using SAMtools software (version 1.3.1) using the following parameters: (1) mpileup -q 20 -Q 30 -ugf ref.fa and (2) bcftools view -p 0.9 -cg. Based on the variations (SNPs and InDels), 14 candidate genes in the resistant and

susceptible lines were translated into proteins using Open Reading Frame Finder (ORF Finder, parameter “default”) (<https://www.ncbi.nlm.nih.gov/orffinder/>). The amino acid sequences and conserved domains of the 14 candidate genes in the resistant and susceptible lines were then compared. Finally, the gene exhibiting conserved domain variation was considered as the best candidate gene for the *RPF1* locus. Alignment of amino acid sequences and conserved domain analysis were performed using MEGA7 software (parameter “ClustaW”) and the online tool SMART (<http://smart.embl-heidelberg.de/>) (parameter “default”), respectively.

## Expression analysis

Relative expression of the candidate genes was quantified in a quantitative real-time (qRT)-PCR assay, using a QuantStudio™ 12 K Flex Real-Time PCR System (Applied Biosystems). Leaf material was harvested from two resistant (12S2 and 12S3) and three susceptible (12S1, 12S4, and 10S2) lines inoculated with *Pfs9* at the end of inoculation test when all of the susceptible plants have susceptible symptoms. The RNA extraction was performed using a TRIzol reagent (Thermo Fisher Scientific Inc.), and cDNA was synthesized using a TranScript One-Step gDNA Removal and cDNA Synthesis Kit (TransGen Biotech, Beijing, China). The primer sequences used in qRT-PCR analysis are shown in Table S1. Relative expression was calculated by normalization using the reference genes *G6PDH* (glucose-6-phosphate dehydrogenase). qRT-PCR was performed in a 10- $\mu$ L reaction volume containing 70 ng template cDNA, 0.2  $\mu$ M of each primer, 0.2  $\mu$ M ROX Reference Dye II, 3.4  $\mu$ L ddH<sub>2</sub>O, and 5  $\mu$ L 2 $\times$  SYBR Fast qPCR Mix (TaKaRa, [www.takara-bio.com/](http://www.takara-bio.com/)). The reactions were performed in a 384-well plate, and all genes expression analyses were conducted with three independent biological replicates and three independent technical replicates for each sample. The thermal cycling condition is as follows: 50 °C for 2 min, 95 °C for 10 min, followed by 35 cycles at 95 °C for 15 s and 60 °C for 1 min. The relative genes expression levels were determined using the  $2^{-\Delta\Delta C_T}$  method (Livak and Schmittgen 2001).

## Results

### Phenotypic and genetic analysis

To assess the inheritance of resistance to *Pfs* in 12S3, crosses were made between 12S3 and 12S4. All the F<sub>1</sub> plants were resistant to downy mildew. In the F<sub>2</sub> population, 153 of 200 individuals constituted resistant plants and 47 constituted susceptible plants, with the Chi-square test confirming the segregation ratio to be 3:1 (Table 1). The results indicated

**Table 1** Genetic analysis of resistance to spinach downy mildew in the  $F_2$  population

Population	Total plants number	Resistant plant	Susceptible plant	$\chi^2_{3:1}$	$\chi^2_{0.05}$
$F_2$	200	153	47	0.167	3.841

$\chi^2_{3:1} < \chi^2_{0.05} = 3.841$ , indicating that the resistance to *Pfs* in 12S3 was controlled by a single dominant gene

that the resistance to *Pfs* in 12S3 was controlled by a single dominant gene, designated the *RPF1* locus in a previous report (Irish et al. 2008).

### Fine mapping of *RPF1*

A total of 4080 high-quality polymorphic SLAFs had been obtained to construct a high-density genetic map of spinach (Qian et al. 2017). Association analysis was then performed by Euclidean distance using the sequencing data of two pools (Rr and RR). The association threshold was 0.01, and 21 SLAF markers on LG1, from 0 to 10.8 cM, were significantly correlated with the target gene. After aligning sequences of the 21 SLAF markers and DM-1 to the spinach genome, these SLAF markers and DM-1 were mapped on chromosome 3 from 0.05 to 1.77 Mb. The results of the mapping and association analysis are shown in Table 2 and Fig. 1a, respectively.

To further refine the mapping region, the 22 markers were used to screen recombinants from the susceptible individuals of  $F_2$  population. Of these 22 markers, six markers were not available due to lack of segregation (KM3020126, KM1091570, and KM3712903), distorted segregation (KM2531431), and a lack of signal (KM1813691 and KM1591384). Eventually, 12 recombinants were obtained and the *RPF1* candidate region was reduced to a 0.89 Mb interval flanked by marker KM256645 (0.34 Mb) and KM1179193 (1.23 Mb) (Fig. 1b).

### Analysis of candidate genes in the candidate region

A total of 177 genes were found within the 0.89 Mb interval on chromosome 3 (Table S2). Based on the annotation of the spinach reference genome database in SpinachBase (<http://www.spinachbase.org/cgi-bin/spinach/tool/download.cgi>), 14 *R* genes were identified as the candidate genes at the *RPF1* locus (Table 3).

Resequencing of 12S2, 12S3, 10S2, 12S1, and 12S4 was performed by Biomarker Technologies (Beijing, China). A total of 151 Mb raw paired reads were obtained, and the low-quality reads were filtered out, resulting in 136 Mb clean reads (Table S3). The total SNPs and InDels of the 14 candidate genes were 17,521 and 96, respectively. The number

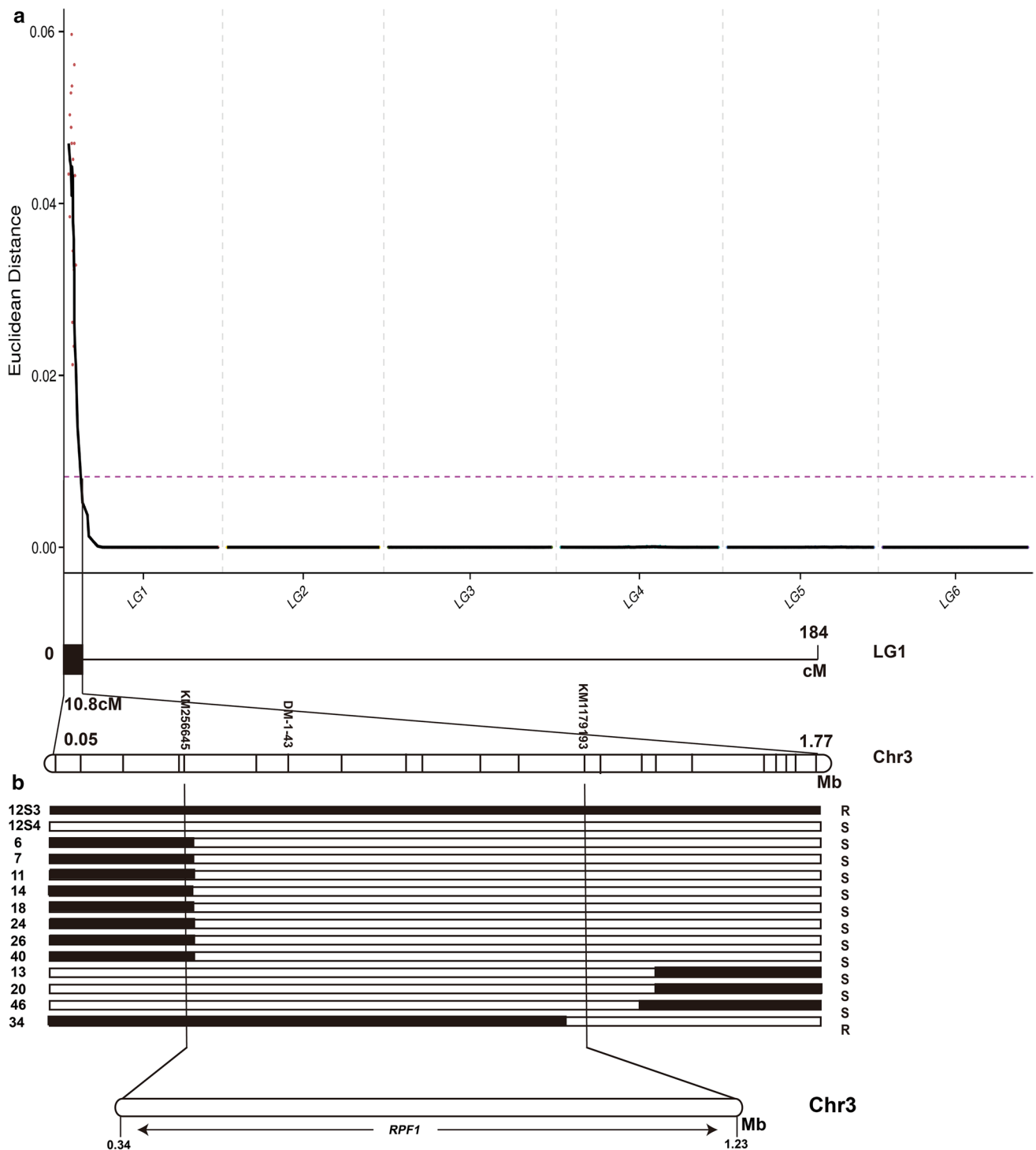
**Table 2** Position of 22 markers closely associated with the *RPF1* locus on the genome

Chr ID	Marker	Start (bp)	End (bp)
Chr3	KM1813691	59,999	60,098
	KM1091570	116,023	115,823
	KM2799188	210,684	210,884
	KM256247	335,283	335,083
	KM256645	346,823	346,623
	KM255893	507,514	507,314
	DM-1-43	578,831	578,985
	KM256436	697,817	697,620
	KM3019892	841,444	841,642
	KM3020126	877,739	877,939
	KM1179494	1,007,056	1,007,257
	KM1178722	1,092,561	1,092,735
	KM1179193	1,239,297	1,239,490
	KM1179370	1,257,213	1,257,408
	KM2544246	1,366,932	1,367,132
	KM1591079	1,398,464	1,398,264
	KM1591384	1,479,150	1,478,953
	KM2532196	1,687,745	1,687,545
	KM2532038	1,696,819	1,696,619
	KM2531431	1,698,987	1,698,811
	KM3712903	1,703,078	1,702,878
	KM1000661	1,773,768	1,773,568

of variations in the exons and introns of 14 candidate genes is shown in Table 4.

### Screening of candidate genes

After comparative analysis of the amino acid sequences and conserved domains of the 14 candidate genes in the resistant and susceptible lines, three significant genes, namely *Spo12784*, *Spo12903*, and *Spo12729*, were obtained. Both *Spo12784* and *Spo12903* encode an NBS-LRR domain, and *Spo12729* encodes a receptor-like protein. The three genes differ majorly between the resistant and susceptible lines (Fig. 2). Compared with the conserved domain of *Spo12784* in the resistant lines, the susceptible lines lacked the LRR\_8 domain, which was ascribed to a change in codon ACA into AAG at position –1700 in the resistant lines, resulting in amino acid substitutions (from C into L) at position +561 (Fig. S7). Conversely, compared with the domain of *Spo12903* in the susceptible lines, the resistant lines lacked the LRR\_8 domain. This was attributed to one nucleotide substitution (from A into G) at position –720 in the resistant lines, resulting in threonine being substituted for isoleucine at position +1081 (Fig. S8). As for *Spo12729*, a signal peptide domain was present in the resistant lines, whereas a transmembrane region was present in the susceptible lines.



**Fig. 1** Mapping analysis of *RPF1*. **a** Association analysis to locate the *RPF1* locus to a 10.8 cM region on LG1. X-axis=linkage group; Y-axis=ED value; red line indicates the threshold value. **b** Fine-mapping analysis delimited *RPF1* to a 0.89 Mb interval flanked by the SNP markers KM256645 and KM1179193. Twelve recombinants

are shown. The black segments represent the resistant 12S3 genotype, while the white segments indicate the susceptible 12S4 genotype. Phenotype scores (R=resistant, S=susceptible) are indicated on the right (color figure online)

This was due to codon AAT changing into GGT at position +529 in the resistant lines, which led to asparagines substituting glycine at position +177 (Fig. S9). Additionally, the

number of LRR domains differs between the resistant and susceptible lines, but it remains unclear which SNP determines the variation.

**Table 3** Candidate genes within the region on chromosome 3 from 0.34 to 1.23 Mb

Gene ID	Physical position (Mb)	Annotation
<i>Spo12762</i>	0.374	Serine/threonine/dual-specificity protein kinase
<i>Spo12719</i>	0.386	Serine/threonine/tyrosine-protein kinase catalytic domain
<i>Spo12765</i>	0.416	Serine/threonine/dual-specificity protein kinase
<i>Spo12773</i>	0.502	Leucine-rich repeat (LRR)
<i>Spo12729</i>	0.547	Leucine-rich repeat (LRR)
<i>Spo12730</i>	0.549	Leucine-rich repeat (LRR)
<i>Spo12736</i>	0.667	NB-ARC; leucine-rich repeat (LRR)
<i>Spo12784</i>	0.686	NB-ARC; leucine-rich repeat (LRR)
<i>Spo12793</i>	0.861	Serine/threonine-protein kinase
<i>Spo12903</i>	0.966	Leucine-rich repeat (LRR); NB-ARC
<i>Spo12905</i>	0.977	NB-ARC; leucine-rich repeat (LRR)
<i>Spo12908</i>	1.037	Disease resistance protein
<i>Spo12916</i>	1.147	Leucine-rich repeat (LRR)
<i>Spo12821</i>	1.212	CC-NBS-LRR disease resistance protein

**Table 4** SNP and InDel variation in the exons and introns of the 14 candidate genes

Genes ID	SNPs		InDels	
	Exon	Intron	Exon	Intron
<i>Spo12762</i>	2	7	0	1
<i>Spo12719</i>	2	2	1	0
<i>Spo12765</i>	0	11	0	1
<i>Spo12773</i>	70	38	7	11
<i>Spo12729</i>	115	0	0	0
<i>Spo12730</i>	27	0	1	0
<i>Spo12736</i>	201	16	2	0
<i>Spo12784</i>	105	0	1	0
<i>Spo12793</i>	62	264	3	31
<i>Spo12903</i>	120	1	0	0
<i>Spo12905</i>	133	3	1	1
<i>Spo12908</i>	32	91	0	5
<i>Spo12916</i>	82	105	8	9
<i>Spo12821</i>	146	117	5	8

The remaining candidate genes were excluded using the same approach. The three candidate genes *Spo12762*, *Spo12719*, and *Spo12765* possess identical amino acid sequences between the resistant and susceptible lines (Fig. S10, Fig. S11, and Fig. S12). Thus, the three candidate genes constituted highly conserved proteins and were not the best candidate genes at the *RPF1* locus.

The candidate gene *Spo12916* encoded RLKs with extracellular receptors, a transmembrane domain, and a S/TK domain. Several amino acid mutations were observed at the amino acid level, but the conserved domain did not differ between the resistant and susceptible lines (Fig. S13). Furthermore, the RLKs can be further divided into RD and non-RD kinases (R and D indicate arginine and aspartic

acid, respectively), based on the presence or absence of a conserved arginine (R) residue preceding the DxxxxN motif (de Oliveira et al. 2014). To date, most non-RD kinases are involved in plant immunity recognition, whereas RD kinases are thought to play an important role in other processes, such as development (Dardick and Ronald 2006; Hurni et al. 2015). As the candidate gene *Spo12916* is an RD kinase, it was thus eliminated from the analysis.

The candidate gene *Spo12821* encodes a 1275-aa-predicted CC-NBS-LRR protein with homology to a *Beta vulgaris* putative disease resistance protein RGA4 (GenBank accession No. XP\_010690548). However, the resequencing data showed that the candidate gene was a heterozygote in the susceptible inbred line 10S2, suggesting that the gene was not the candidate resistance gene at the *RPF1* locus.

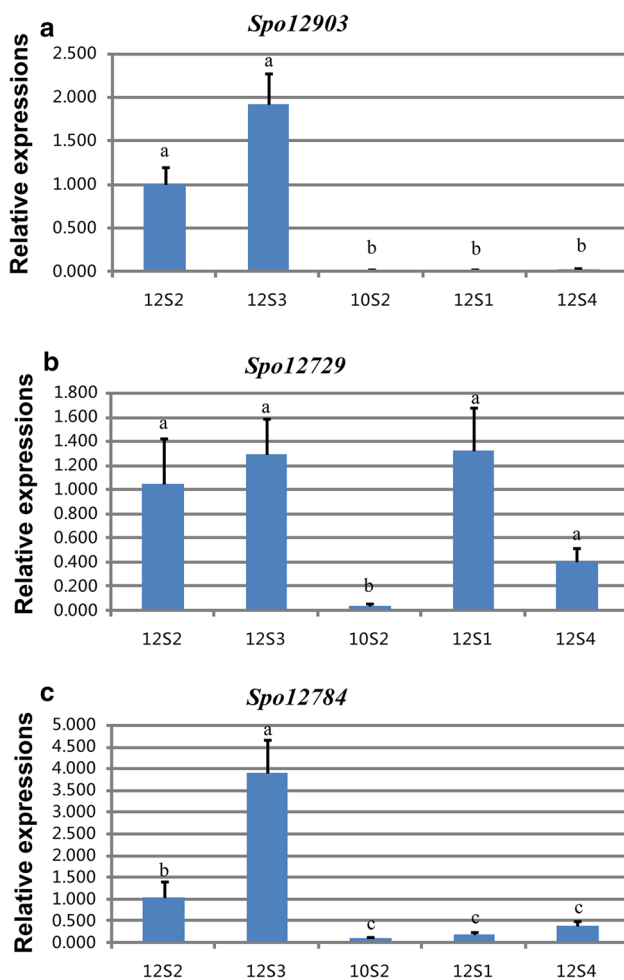
The remaining six genes *Spo12773*, *Spo12730*, *Spo12736*, *Spo12793*, *Spo12905*, and *Spo12908* encoded the RLK, RLP, NBS-LRR, S/TK, CC-NBS-LRR, d1m0za\_ domains, respectively. All of these genes possessed several mutations in the coding region, while the conserved domain did not differ majorly between the resistant and susceptible lines (Fig. S13). Therefore, all of the above genes do not constitute the best candidate resistance genes at the *RPF1* locus.

### Expression analysis of the *RPF1* candidate genes

To clarify the transcription patterns of the three candidate genes in the inoculated resistant and susceptible lines, qRT-PCR analysis was performed to analyze the transcript levels in the two resistant and three susceptible lines. The results are summarized in Fig. 3. The *Spo12903* was significantly expressed in the two resistant lines, 12S2, and 12S3, compared with those in the three susceptible lines, 12S1, 12S4, and 10S2 (Fig. 3a). Similarly, the *Spo12784* relative expression







**Fig. 3** Relative expressions levels of the three candidate genes in the inoculated resistant and susceptible lines. **a–c** Relative expression of *Spo12903*, *Spo12729*, and *Spo12784*, respectively. The error was calculated from the standard deviation of three biological replicates, and different letters on the top of the bars indicate significant differences (Student's test,  $p < 0.01$ ). Formula  $2^{-\Delta\Delta C_T}$  is used to calculate relative expression

for the further mapping of *RPF1*. However, two major deficiencies of the approach have been noted. Firstly, fine mapping of *RPF1* was not performed using any mapping population. Secondly, only NBS-LRR genes were selected as the candidate genes. There are five classes of *R* genes, and only two classes (NBS-LRR and RLP) of *R* genes have been reported to have resistance against downy mildew. Although the majority of downy mildew resistance genes encode NBS-LRR, there is the possibility that other classes of *R* genes may be involved in downy mildew resistance. Thus, the other *R* gene classes may have been overlooked.

In the present study, we mapped *RPF1* into an interval of 0.89 Mb on chromosome 3 using 148 BC<sub>1</sub> and 200 F<sub>2</sub> populations. A total of 14 candidate genes were identified

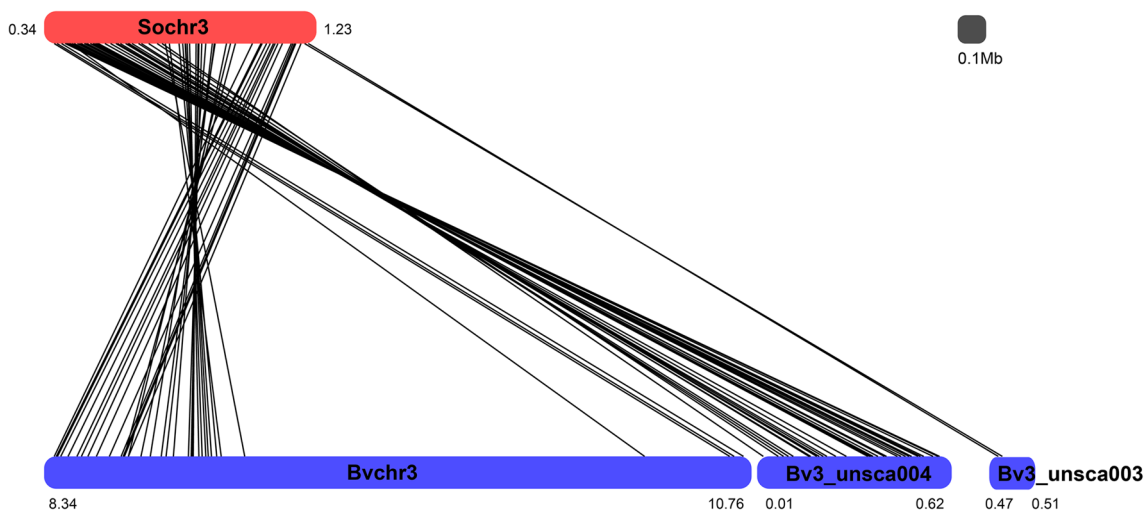
within the 0.89 Mb region of the spinach genome. To avoid missing the candidate genes, chromosomal synteny analysis of the region of the *RPF1* locus in spinach and sugar beet was performed using MCScanX (Wang et al. 2012). The results showed that the region of *RPF1* in spinach shared a striking colinearity to Bvchr3 of sugar beet (Fig. 4). A total of 10 *R* genes were identified in the syntenic region of sugar beet, eight of which shared colinearity with some of the 14 candidate genes in spinach. The results of the syntenic analysis are shown in Table S4, and the 10 *R* genes in syntenic region of sugar beet are shown in Table S5.

Furthermore, conserved domain analysis of the 14 candidate genes indicated that among 14 candidate genes, five genes (*Spo12736*, *Spo12784*, *Spo12903*, *Spo12905*, and *Spo12821*) contained the NBS-LRR domain; *Spo12729* and *Spo12730* belonged to RLP; another two genes (*Spo12773* and *Spo12916*) belonged to RLK; and *Spo12762*, *Spo12719*, *Spo12765*, and *Spo12793* encoded a Ser/Thr kinase domain (Table 3). As most downy mildew resistance genes encode NBS-LRR and RLP, the candidate genes possession of the NBS-LRR and RLP domain were regarded as the crucial genes. To differentiate these resistance genes, a comparison of the amino acid sequences and conserved domain of the 14 candidate genes in the resistant and susceptible lines revealed that the gene exhibiting conserved domain variation constituted an important candidate gene. Ultimately, *Spo12729*, *Spo12784*, and *Spo12903* were selected as the best candidate resistance genes at the *RPF1* locus.

Additionally, expression levels of the three genes in two inoculated resistant and two susceptible lines were analyzed by qRT-PCR. The result indicated that *Spo12784* and *Spo12903* of the inoculated resistant lines exhibited the higher expression than the susceptible lines. There were no obvious differential transcript levels in *Spo12729* (Fig. 3). However, further analysis is required to confirm the candidate genes, including transforming the three genes into susceptible plants.

### Homology analysis of the best candidate genes and known downy mildew resistance genes

The genes *Spo12784* and *Spo12903* encoded a 1126-aa and 1206-aa NBS-LRR domain, respectively. *Spo12729* encoded a 634-aa RLP. In this study, eight known downy mildew resistance genes, namely *RPP1*, *RPP4*, *RPP5*, *RPP8*, *RPP13*, and *RPP27* from *Arabidopsis*, and *Dm3* and *PI8* from lettuce and sunflower, respectively, were obtained from Plant Resistance Genes database (<http://prgdb.org/old.php>). To determine the homology of the three candidate genes and other downy mildew resistance genes, homology analysis of the amino acid sequences of the three candidate



**Fig. 4** Synteny analysis of the candidate region of *RPF1* in spinach and sugar beet. The red box annotated with “Sochr3” indicates chromosome 3 of spinach. The blue box annotated with “Bvchr3” represents chromosome 3, while the blue boxes annotated with

“Bv3\_unsca004” and “Bv3\_unsca003” inside represent chromosome 3 unlocalized genomic scaffolds 004 and 003 of sugar beet, respectively. The numbers on both sides of the chromosomes represent the physical location (color figure online)

genes and the eight known downy mildew resistance genes was performed using online tool BLASTP ([https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)). The alignment results showed that among the eight known downy mildew resistance genes, only *RPP8*, *RPP13*, and *PI8* shared high homology with *Spo12784* and *Spo12903*. Specifically, *Spo12784* shared 28, 27, and 25% identity with *RPP8*, *RPP13*, and *PI8*, respectively, and *Spo12784* shared 28, 33, and 28% identity with *RPP8*, *RPP13*, and *PI8*, respectively. However, only *RPP27* shared high homology (27% identity) with *Spo12729*. Additionally, the candidate genes *Spo12903* and *Spo12784* and *Spo12729* and *RPP27* each were grouped together in the phylogenetic tree (Fig. 5). Based on comparative analysis of the conserved domain in Fig. 2, it is evident that *Spo12784*, *Spo12903*, and *PI8*, *RPP8*, and *RPP13* possess a similar structure containing the NBS-LRR domain, and *Spo12729* and *RPP27* have a similar structure with a signal peptide and LRR domain.

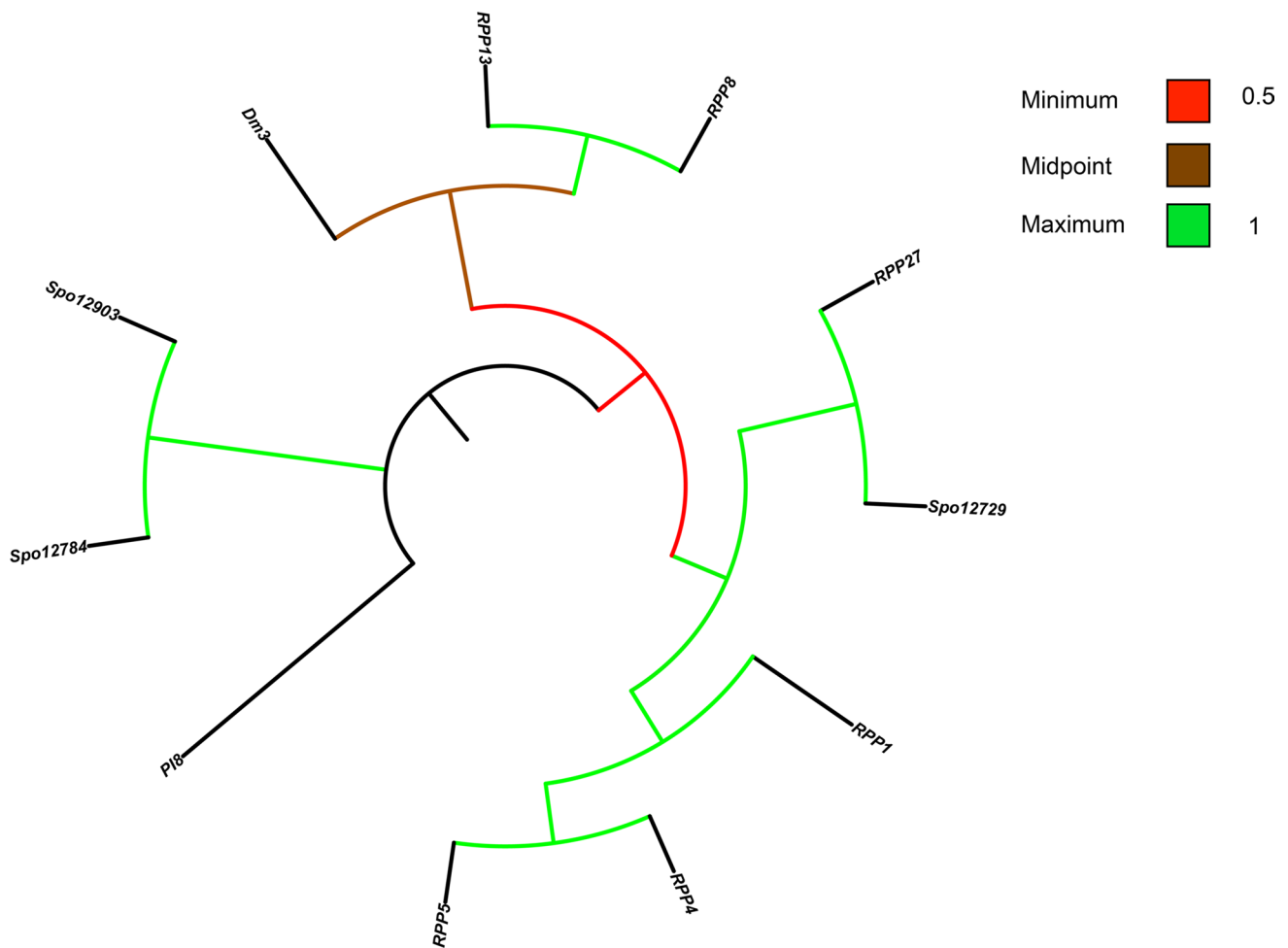
The NBS-LRR domain can be further divided into CNL and TNL classes according to the structure of the N-terminus domain. The TNL domains are only observed in dicotyledons, such as the *N* and *M* genes from tobacco and flax, respectively, while CNL domains are found in both in monocotyledons and dicotyledons, such as the *Lr1* and *Lr10* genes in monocotyledonous wheat, and *RPP8* and *RPP13* in the model eudicot *A.thaliana* (She et al. 2017). Spinach is a eudicot, and 139 NBS-LRR genes were identified in the spinach genome. The major NBS-LRR gene groups in spinach are CNL and TNL, representing 40 and 36% of the NBS-LRR genes, respectively, and only one TNL gene was observed in the spinach genome (Xu et al. 2017). *Spo12784*

and *Spo12903* encode a CNL and NL domain, respectively. *RPP8*, *RPP13*, and *PI8* also contain a CNL domain. Therefore, compared with the TNL (*RPP1*, *RPP4*, and *RPP5*) and RLP (*RPP27*) class R genes, a CNL class R gene shared the highest homology with *Spo12784* and *Spo12903*. Additionally, both *Spo12729* and *RPP27* encode an RLP domain. Thus, only *RPP27* shared the highest homology with *Spo12729*, and the two genes clustered together in the phylogenetic tree (Fig. 5).

### Based on candidate gene variation, SNP markers were developed for MAS

In recent years, molecular markers have been widely used in many crop breeding programs (Devran et al. 2015). Compared with traditional breeding methods, MAS is able to improve breeding efficiency and accelerate breeding progress. Therefore, the development of markers tightly linked to the gene of interest is significant for breeders. KASP is considered as an adaptable, efficient, and effective approach in SNP genotyping analysis and has been widely applied on many plants, including cotton, wheat, and potato (Graves et al. 2016).

In the present study, three crucial amino acid mutations of *Spo12784*, *Spo12903*, and *Spo12729* were obtained via comparative analysis of the conserved domain in the resistant and susceptible lines. Based on the amino acid mutation, three SNP variants of the candidate genes were identified that can be developed as KASP markers for MAS. This would be valuable in cloning the *RPF1* alleles and evaluating how each may function.



**Fig. 5** Cluster analysis of the three candidate genes and other downy mildew resistance genes. The color indicates the bootstraps value. Red represents a minimum bootstrap value, and green represents a maximum bootstrap value (color figure online)

## Conclusion

In this study, a SLAF-Seq approach combined with BSA was successfully used to locate the *RPF1* gene in spinach. A localized region of 10.8 cM with 21 SLAF markers on LG1 was detected to be tightly associated with the *RPF1* locus. A 1.72 Mb region on chromosome 3 was obtained by aligning the 21 SLAF markers and the DM-1 marker close to *RPF1* in the spinach genome. To narrow down the candidate region, the 22 markers were developed using KASP assay primers to screen recombinants from an  $F_2$  population. Ultimately, a 0.89-Mb region with 14 candidate *R* genes was discovered. To further evaluate the candidate genes, resequencing of two resistant inbred lines (12S2 and 12S3) and three susceptible inbred lines (12S1, 12S4, and 10S2) was performed. The three most likely candidate genes (*Spo12729*, *Spo12784*, and *Spo12903*) were identified from 14 genes via amino acid sequence

analysis and conserved domain analysis between the resistant and susceptible inbred lines. These results provide a basis for cloning the resistance gene at the *RPF1* locus and determining the functionality of the resistance gene, as well as for developing suitable markers for spinach breeding.

**Author contribution statement** QW and XZ designed the study. SH and QW conducted the experiments. SH, QW, and LZ analyzed the data. SH wrote the manuscript. QW, FC, and CJ made the revision of the manuscript. QW, ZH, and XZ prepared and collected the samples. WX and WJ helped conduct experiments.

**Acknowledgements** This work was performed at the Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture, Beijing, China, and was supported by National Natural Science Foundation of China (31401872), the National Science and

Technology Pillar Program during the Twelfth Five-Year Plan Period of China (2014BAD01B08), Beijing Scientific Program of Municipal Commission of Science and Technology (Z171100001517014), the Chinese Academy of Agricultural Sciences Innovation Project (CAAS-ASTIP-IVFCAAS), and Fundamental Research Funds for Central Non-profit Scientific Institution (1610102018004).

## Compliance with ethical standards

**Conflict of interest** We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

**Ethical standards** The authors declare that this study complies with the current laws of the countries in which the experiments were performed.

## References

- Baird NA, Etter PD, Atwood TS et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376
- Correll JC, Morelock TE, Black M C et al (1994) Economically important diseases of spinach. *Plant Dis* 78:653–660
- Cheng F, Sun R, Hou X et al (2016) Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. *Nat Genet* 48:1218–1224
- Correll JC, du Toit L, Koike ST et al (2010) Guidelines for spinach downy mildew: *Peronospora farinosa* f. sp. *spinaciae* (Pfs)
- Correll JC, Bluhm BH, Feng C et al (2011) Spinach: better management of downy mildew and white rust through genomics. *Eur J Plant Pathol* 129:193–205
- Dangl JL, Jones JD (2001) Plant pathogens and integrated defence responses to infection. *Nature* 411:826
- Dardick C, Ronald P (2006) Plant and animal pathogen recognition receptors signal through non-RD kinases. *PLoS Pathog* 2:e2
- Devran Z, Kahveci E, Özkaynak E et al (2015) Development of molecular markers tightly linked to *Pvr4* gene in pepper using next-generation sequencing. *Mol Breed* 35(4):101
- de Oliveira LFV, Christoff AP, de Lima JC et al (2014) The Wall-associated Kinase gene family in rice genomes. *Plant Sci* 229:181–192
- Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379
- Feechan A, Anderson C, Torregrosa L et al (2013) Genetic dissection of a TIR-NB-LRR locus from the wild North American grapevine species *Muscadinia rotundifolia* identifies paralogous genes conferring resistance to major fungal and oomycete pathogens in cultivated grapevine. *Plant J* 76:661–674
- Feng C, Correll JC, Kammeijer KE et al (2014) Identification of new races and deviating strains of the spinach downy mildew pathogen *Peronospora farinosa* f. sp. *spinaciae*. *Plant Dis* 98:145–152
- Feng C, Bluhm BH, Correll JC (2015) Construction of a spinach bacterial artificial chromosome (BAC) library as a resource for gene identification and marker development. *Plant Mol Biol Rep* 33:1996–2005
- Geng X, Jiang C, Yang J et al (2016) Rapid identification of candidate genes for seed weight using the SLAF-Seq method in *Brassica napus*. *PLoS ONE* 11:e0147580
- Graves H, Rayburn AL, Gonzalez-Hernandez JL et al (2016) Validating DNA polymorphisms using KASP assay in prairie cordgrass (*Spartina pectinata* Link) populations in the US. *Front Plant Sci* 6:1271
- Hurni S, Scheuermann D, Krattinger SG et al (2015) The maize disease resistance gene *Hm1* against northern corn leaf blight encodes a wall-associated receptor-like kinase. *Proc Natl Acad Sci* 112:8780–8785
- Irish BM, Correll JC, Koike ST et al (2003) Identification and cultivar reaction to three new races of the spinach downy mildew pathogen from the United States and Europe. *Plant Dis* 87:567–572
- Irish BM, Correll JC, Feng C et al (2008) Characterization of a resistance locus (Pfs-1) to the spinach downy mildew pathogen (*Peronospora farinosa* f. sp. *spinaciae*) and development of a molecular marker linked to Pfs-1. *Phytopathology* 98:894–900
- Jin W, Wang H, Li M et al (2016) The R2R3 MYB transcription factor *PavMYB10.1* involves in anthocyanin biosynthesis and determines fruit skin colour in sweet cherry (*Prunus avium* L.). *Plant Biotechnol J* 14:2120–2133
- Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12:656–664
- Kubota M, Tamura O, Nomura Y et al (2017) Recent races of spinach downy mildew pathogen *Peronospora farinosa* f. sp. *spinaciae* in Japan. *J Gen Plant Pathol* 83:117–120
- Liang D, Chen M, Qi X et al (2016) QTL mapping by SLAF-seq and expression analysis of candidate genes for aphid resistance in cucumber. *Front Plant Sci* 7:1000
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  methods. *Methods* 25:402–408
- Martin GB, Brommonschenkel SH, Chunwongse J et al (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* 262:1432–1436
- Meyers BC, Chin DB, Shen KA et al (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell* 10:1817–1832
- Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. *Curr Opin Plant Biol* 8:129–134
- Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci* 88:9828–9832
- Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucl Acids Res* 8:4321–4326
- Nemri A, Atwell S, Tarone AM et al (2010) Genome-wide survey of Arabidopsis natural variation in downy mildew resistance using combined association and linkage mapping. *Proc Natl Acad Sci* 107:10302–10307
- Plantum (2016) Denomination of Pfs: 16, a new race of downy mildew in spinach. <http://www.plantum.nl/321519635/Document?documentregistrationid=780926983>. Accessed 5 Sept 2016
- Poulsen DME, Henry RJ, Johnston RP et al (1995) The use of bulk segregant analysis to identify a RAPD marker linked to leaf rust resistance in barley. *Theor Appl Genet* 91:270–273
- Qian W, Feng CD, Zhang HL et al (2016) First report of race diversity of the spinach downy mildew pathogen, *Peronospora effusa*, in China. *Plant Dis* 100:1248
- Qian W, Fan G, Liu D et al (2017) Construction of a high-density genetic map and the X/Y sex-determining gene mapping in spinach based on large-scale markers developed by specific-locus amplified fragment sequencing (SLAF-seq). *BMC Genomics* 18:276
- Qin D, Dong J, Xu F et al (2015) Characterization and fine mapping of a novel barley Stage Green-Revertible Albino Gene (*HvSGRA*) by bulked segregant analysis based on SSR assay and specific length amplified fragment sequencing. *BMC Genomics* 16:838
- Quarrie SA, Lazić-Jančić V, Kovačević D et al (1999) Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. *J Exp Bot* 50:1299–1306

- Radwan O, Mouzeyar S, Nicolas P et al (2005) Induction of a sunflower CC-NBS-LRR resistance gene analogue during incompatible interaction with *Plasmopara halstedii*. *J Exp Bot* 56:567–575
- She HB, Fan GY, Zhang HL et al (2017) Cloning and analysis of NBS-LRR type disease-resistant gene analogs in *Spinach oleracea* L. *China Veg* 5:26–33 (in Chinese)
- Slusarenko AJ, Schlaich NL (2003) Downy mildew of *Arabidopsis thaliana* caused by *Hyaloperonospora parasitica* (formerly *Peronospora parasitica*). *Mol Plant Pathol* 4:159–170
- Song WY, Wang GL, Chen LL et al (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* 270:1804
- Sun X, Liu D, Zhang X et al (2013) SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE* 8:e58700
- Tör M, Brown D, Cooper A et al (2004) *Arabidopsis* downy mildew resistance gene *RPP27* encodes a receptor-like protein similar to *CLAVATA2* and tomato *Cf-9*. *Plant Physiol* 135:1100–1112
- Venuprasad R, Dalid CO, Del Valle M et al (2009) Identification and characterization of large-effect quantitative trait loci for grain yield under lowland drought stress in rice using bulk-segregant analysis. *Theor Appl Genet* 120:177–190
- Wang S, Cao J (2006) Advances in molecular mechanisms of the resistance of *Arabidopsis thaliana* to downy mildew in Chinese. *Plant Prot* 5:004
- Wang Y, Tang H, DeBarry JD et al (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucl Acids Res* 40:e49
- Xia C, Chen L, Rong T et al (2015) Identification of a new maize inflorescence meristem mutant and association analysis using SLAF-seq method. *Euphytica* 202:35–44
- Xiao S, Ellwood S, Calis O et al (2001) Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by *RPW8*. *Science* 291:118–120
- Xie W, Feng Q, Yu H et al (2010) Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. *Proc Natl Acad Sci* 107:10578–10583
- Xu C, Jiao C, Sun H et al (2017) Draft genome of spinach and transcriptome diversity of 120 *Spinacia* accessions. *Nat Commun* 8:15275
- Zheng W, Wang Y, Wang L et al (2016) Genetic mapping and molecular marker development for *Pi65 (t)*, a novel broad-spectrum resistance gene to rice blast using next-generation sequencing. *Theor Appl Genet* 129:1035–1044