

SCIENTIFIC REPORTS



OPEN

Two QTLs controlling Clubroot resistance identified from Bulked Segregant Sequencing in Pakchoi (*Brassica campestris* ssp. *chinensis* Makino)

Hongfang Zhu, Wen Zhai, Xiaofeng Li & Yuying Zhu

Clubroot, caused by *Plasmodiophora Brassicae*, is a serious soil-borne disease in worldwide. In recent years, progression of clubroot is rapid and serious in Shanghai, China. In this study, The inheritance of clubroot resistance (CR) were determined in pakchoi using F_2 segregation population that were developed by crossing highly resistant line 'CR38' and susceptible line 'CS22'. Two novel QTLs, *qBrCR38-1* and *qBrCR38-2*, was identified by BSA-seq (Bulked Segregant Sequencing) resistant to *P. brassicae* physiological race 7. Two significant peak *qBrCR38-1* and *qBrCR38-2* were observed by three statistical methods between interval of 19.7–20.6 Mb in chromosome A07 and 20.0–20.6 Mb in chromosome A08, respectively. In addition, Polymorphic SNPs identified within target regions were converted to competitive allele-specific PCR (KASP) assays. In target regions of *qBrCR38-1* and *qBrCR38-2*, there were twenty SNP sites identified, eleven KASP markers of which are significantly associated to CR ($P < 0.05$). Seven candidate genes were identified and found to be involved in disease resistance (TIR-NBS-LRR proteins), defense responses of bacterium and fungi and biotic/abiotic stress response in the target regions harboring the two QTLs. Two novel QTLs and candidate genes identified from the present study provide insights into the genetic mechanism of CR in *B. rapa*, and the associated SNPs can be effectively used for marker-assisted breeding.

Clubroot, caused by the obligate endoparasite *Plasmodiophora Brassicae*, is recognized as a serious soil-borne disease. It infects cruciferous Brassica oil crops and vegetables and is specially associated with considerable yield losses^{1–3}. The pathogen causes abnormal growth of the plant and eventually leads to a massive gall formation on the root⁴. Clubroot prevents the transportation of water and nutrients, causing the plant to wilt and finally dies off. In recent years, the incidence of clubroot in the suburbs of Shanghai, China, has gradually become increasingly more serious. Until 2017, 39 counties and 9 towns of Shanghai had a breakout of clubroot disease and the affected area had reached 2500 hm². Pathotypes of *P. brassicae* was physiological race 7 in Qingpu district of Shanghai. Clubroot disease is difficult to completely prevent and control by cultural practices or chemical treatments such as anti-microbial compounds or crop rotations^{5,6}. Therefore, development of resistant cultivars is the most effective approach to control clubroot disease.

In *Brassica Rapa*, genetic analysis and quantitative trait locus (QTL) mapping studies have identified at least eight race-specific clubroot resistance (CR) loci. *Crr1*, *Crr2*, and *Crr4* CR genes from European turnip cv 'Siloga', which were mapped onto chromosomes A08, A01, and A06, respectively^{7,8}. The resistance genes *CRa* from turnip line 'ECD02'⁹, *CRb* from turnip line 'Gelria R'^{10,11} and *Crr3* from turnip cv. "Milan White"^{12,13} were both located on the chromosome A03. *CRk* and *CRc* were identified from turnip cv. "Debra" line ECD01 and ECD02, mapped onto A03 and A02, respectively^{14,15}. In *B. rapa* canola, *Rcr4*, *Rcr8*, and *Rcr9* breeding line T19 were mapped to chromosomes A03, A02, and A08, respectively through genotyping by sequencing¹⁶. The resistance gene *Rcr1* from pak choy cv. "Flower Nabana" was also mapped to A03 by transcriptome analysis, and two candidate genes

Shanghai Key Lab of Protected Horticultural Technology, Horticultural Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai, 201106, China. Hongfang Zhu and Wen Zhai contributed equally. Correspondence and requests for materials should be addressed to W.Z. (email: zhaiwen1105@163.com) or Y.Z. (email: yy5@saas.sh.cn)

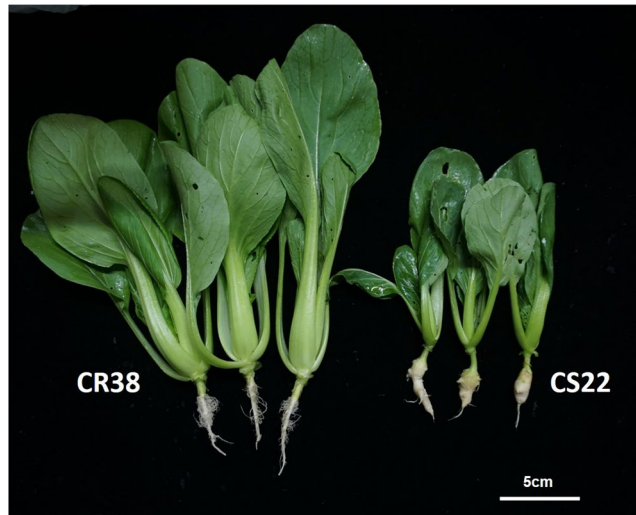


Figure 1. phenotype of the parents: CR38 (Clubroot resistance) and CS22 (Clubroot susceptible). Plants were inoculated with 7th physiology race of *P. brassicae*.

Bra019409 and *Bra019410* were screened for correlation with clubroot-resistance¹⁷. Meanwhile, clubroot-resistant gene *Rcr2* was also mapped on chromosome 3 in chinese cabbage ‘Jazz’. Five SNP markers co-segregated with *Rcr2* were developed between 22 to 26 Mb, and *Bra019410* and *Bra019413* are most likely candidates of *Rcr2*, with conserved domain of TIR-NBS-LRR resistance protein¹⁸.

Bulked segregant analysis by sequencing (BSA-seq), is an effective technique used to identify quantitative trait loci (QTLs)^{19,20}. It is a genomics tool used for genetic mapping which takes advantage of bulked-segregant analysis and high-throughput genotyping using next-generation sequencing (NGS)^{21,22}. BSA-Seq has been applied to mapping agronomically important loci in Arabidopsis, rice and wheat^{23–25}. Candidate genes for disease resistance have been identified successfully by this approach, such as a broad-spectrum resistance gene Pi65 (t) in rice²⁶, downy mildew and powdery mildew resistance QTLs in cucumber^{27,28}. BSA-seq technologies, have proven successful for rapidly establishing the association of agronomic traits with molecular markers and had a major impact on crop breeding²⁹. Meanwhile, different statistical methods for pooled QTL mapping have been proposed, including MutMap²³, G’ value³⁰ and ED³¹, that would have noise reduction and highlighting effect of QTLs. Moreover, it requires an efficient platform for applying molecular markers to marker-assisted selection (MAS) in breeding. Kompetitive Allele Specific PCR (KASP) is one of the high-throughput SNP genotyping technologies is a cost-effective, low genotyping error rates and flexible system which is widely used for genetic mapping, trait-specific markers development, germplasm characterization (genetic diversity, genetic relationship, and population structure), and quality control (QC) analysis (genetic identity, genetic purity, and parentage verification)^{32–34}.

The Pakchoi (*Brassica campestris* ssp. *chinensis* Makino), also called non-heading Chinese cabbage, is one of the most important Brassica vegetable crop in China and East Asia³⁵. Most Pakchoi cultivars are highly susceptible to the *P. brassicae*. In this study, BSA-seq was applied to determine the inheritance of clubroot resistance (CR) in Pakchoi inbred line ‘CR38’ and F₂ population. Then, the underlying QTLs were mapped by three statistical methods and examined sequence variation in the target region to identify the most probable candidate gene associated with CR.

Results

Clubroot Resistance evaluations in CR38 × CS22. ‘CR38’ were highly resistant (DSI = 3.33 in 2016, DSI = 4.76 in 2017) to the 7 physiology race of *P. brassicae* in contrast to the ‘CS22’ which was more susceptible (DSI = 100 in 2016, DSI = 94.44 in 2017). ‘CR38’ and ‘CS22’ (Fig. 1a), were crossed to develop segregating populations for QTL analysis of CR. Clubroot symptoms of 294 F₂ individuals were identified in autumn 2017. F₂ population exhibited a continuous frequency distribution with a range of 0–3 and DSI of 34.24, suggesting polygenic control of CR in this population (Table 1).

Whole genome sequencing analysis. BSA-seq were used to identify the loci controlling CR in our F₂ population, from which 22 highly resistant and 22 highly susceptible individuals were selected and two DNA pool were created. Illumina high-throughput sequencing generated 51,292,360 and 59,493,747 short reads (150 bp in length) from the CR-pool and the CS-pool, with a coverage of 98.11% (39-fold genome coverage) and 98.46% (45-fold genome coverage), respectively. Over 97% of the reads in both pools were mapped to the Chiifu-401 reference genome. The Q20% was 96.12% and 95.91% of the CR -pool and CS-pool. The effective sequencing depths for CR38 and CS22 were 20-fold and 25-fold genome coverage, respectively, which guaranteed accuracy of the subsequent analysis. The results from sequencing are presented in Table 2.

Parents	Type	Year	Disease rating					DSI
			0	1	2	3	Total	
CR38	R	2016	27	3	0	0	30	3.33
		2017	18	2	0	0	20	4.76
CS22	S	2016	0	0	0	15	15	100.00
		2017	0	0	4	18	22	94.44
CR38 × CS22	F2	2017	81	145	43	23	292	34.24

Table 1. Disease evaluations of CR in CR38 × CS22. $DSI(\%) = \frac{\sum(\text{rating class}) \times (\# \text{ plant in rating class})}{(\text{total } \# \text{ plants in treatment}) \times 3} \times 100$.

Sample Name	Clean Reads	Clean Base(G)	Clean Q20 (%)	Clean GC Content (%)	coverage rate (%)	Map reads rate (%)	Effective Rate (%)	Sequencing depth
CR-pool	51,292,360	15.39	96.12	39.62	98.11%	97.03%	99.35	39×
CS-pool	59,493,747	17.85	95.91	39.16	98.46%	97.07%	99.56	45×
CR38	27,262,388	8.18	97.75	39.91	91.89%	97.11%	98.99	20×
CS22	32,499,790	9.75	96.08	39.8	94.09%	97.08%	99.26	25×

Table 2. Summary of BSA sequencing data for each sample.

Mapping of Clubroot resistance. Sequence data were trimmed and filtered prior to analysis. Compared to the reference genome, 1,489,940 SNPs were identified between the CR38 and CS22 parents. Based on the uniquely mapped reads, association analysis between the two bulks was performed on 1,079,828 SNPs. To identify genomic regions associated with CR, three methods were used to mapping the QTLs. Firstly, Scanning each SNP in the whole genome, the Ratio of SNP-ratio in CR-pool and CS-pool was calculated³⁶ (Fig. 2A). Two significant peak *qBrCR38-1* and *qBrCR38-2* was observed between 19.7–20.6 Mb of chromosome A07 and 20.0–20.6 Mb of chromosome A08 (Fig. 2a). Next, Allelic segregation between CR and CS pools by Euclidean Distance (ED)³¹ were measured. ED⁴ was raising ED to the fourth power to decrease noise. Local polynomial regression methods (LOESS fit) of ED⁴ calculated shown in Fig. 2b. The identified peaks that have an ED above a threshold and a high allele frequency in the CR-pool. G value at each SNP were calculated, and G is a smooth version of G³⁰. The statistics result demonstrated that there were two QTLs at the same peaks position about 20.1 Mb in chromosome A07 and 20.2 Mb in chromosome A08 (Fig. 2b). We found 5441 and 1887 SNPs between the two parents in the target regions of chromosome A07 and chromosome A08.

Corresponding sequences in flanking 50 bp region were used to develop allele-specific KASP primers which were genotype of the parental lines and F₂ population. There were twenty SNP sites identified at physical position 18.4 Mb – 20.8 Mb and 19.4 Mb – 21.8 Mb around regions of *qBrCR38-1* and *qBrCR38-2* (Table 3). *qBrCR38-1* was mapped in the interval of 0.35 cM flanked by Br_K_07106 and Br_K_07107. Br_K_080103 and Br_K_080107 are linkage to *qBrCR38-2* within the genetic distance of 0.33 cM. To evaluate the effect of *qBrCR38-1* and *qBrCR38-2*, F₂ population were classified into two groups (AA and aa) based on the genotype of the most tightly linked markers. Lines carrying the CR38 alleles (AA) are significantly ($P < 0.05$) clubroot resistant than those carrying the CS22 alleles (aa) in eleven SNP markers (Table 3). Close-up view of QTLs and KASP markers link to the target regions was shown in Fig. 3.

Candidate gene identification. A total of 188 and 116 annotated genes encoded sequences encompassed the target regions of *qBrCR38-1* and *qBrCR38-2*, respectively. The functional annotation of genes in the target region indicates their potential involvement in fungal disease resistance assessed for further analysis. In chromosome A07, candidate gene *BraA07002412*, homologous to *AT1G80460*, encodes protein *NHO1* (*nonhost resistant gene 1*) similar to glycerol kinase and performs a rate-limiting step in glycerol metabolism. In addition, candidate gene *BraA08002455* in chromosome A08, homologous to *AT1G65850*, encodes Toll-Interleukin-1 receptor/nucleotide binding site/leucine-rich repeat (TIR-NBS-LRR class) disease resistance protein. Moreover, six genes with F-box domain (*BraA07002249*, and *BraA07002494* in A07, *BraA08002451* and *BraA08002452* in A08) and a WRKY transcription factor (*BraA08002471*) were considered as the functional genes associated with CR (Table 4).

Discussion

The traditional fine mapping method is time-consuming and is rather laborious to require the identification of genome-wide polymorphic DNA markers for linkage analysis. Furthermore, it needs to use the genetic mapping and physical mapping to locate the target gene in the specific location of chromosomes. With the rapid development of sequencing technology and genomics, using BSA-seq of extreme mixed pools, a large number of traits can be located quickly and effectively. Compared to traditional QTL mapping, BAS-seq has the advantages of simple operation and short test period, and does not require specific near-isogenic lines to be constructed. Through high-throughput sequencing, thousands of SNPs can be captured at one time, and this variation information can be used to identify trait-related genes or QTLs as well as to develop molecular markers. The accuracy of BSA-seq results is very important to identify the resistance loci of CR, and there are many influencing factors. With improvement of reference genome in Chinese cabbage using NGS and Pacbio SMRT sequencing, it is possible

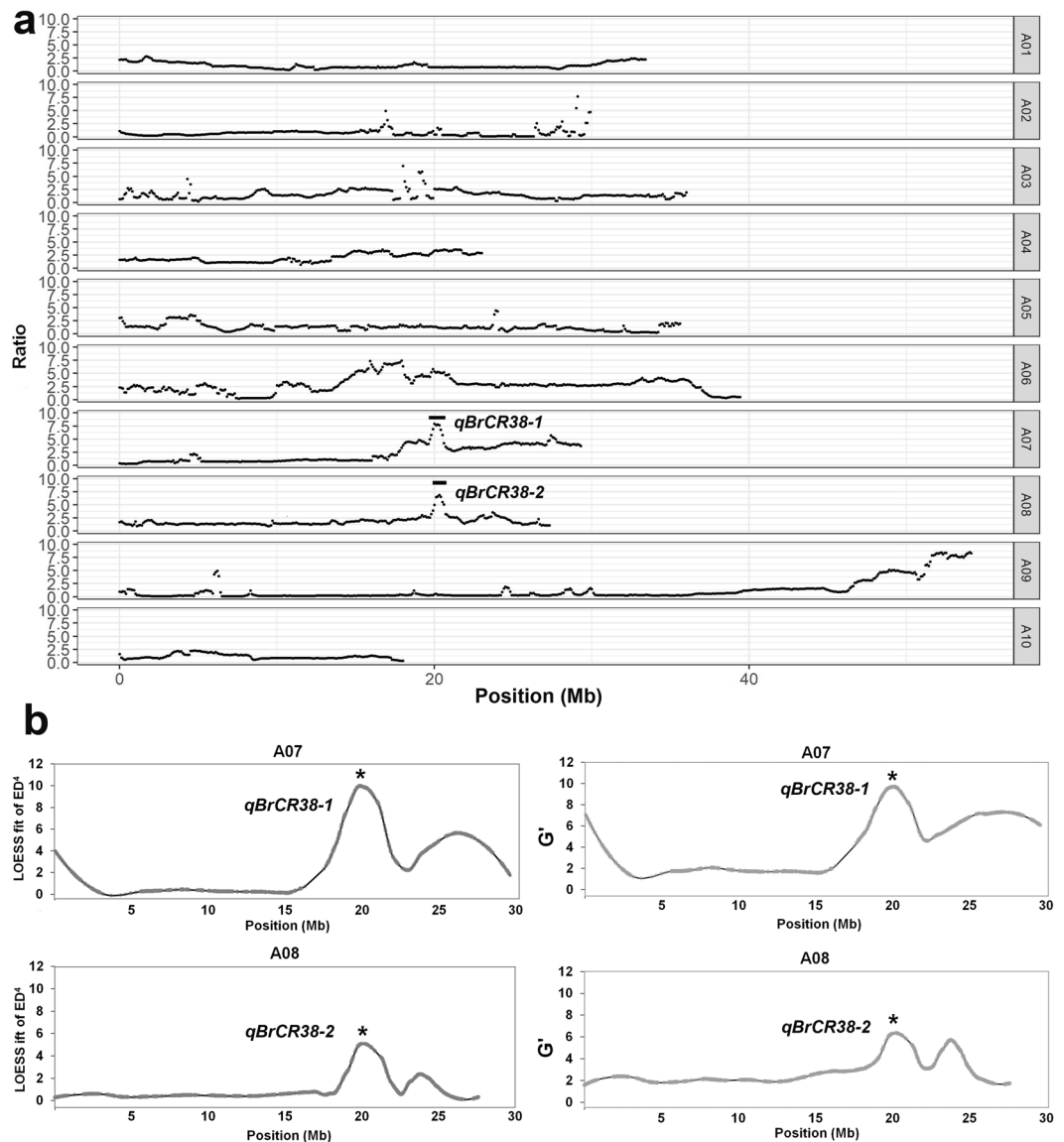


Figure 2. Genome-wide scan for clubroot resistance QTLs using BSA-seq. Two significant QTL of *qBrCR38-1* and *qBrCR38-2* detected in chromosomes A07 and A08 using three statistical approach. (a) Ratio (CR-pool/CS-pool) of the SNP-ratios (Resistant alleles/sensitive alleles) is presented. The significance peak is indicated by the horizontal dotted line. (b) Distribution of G' value and Loess fit curve calculated using ED^4 data on chromosome A07 and A08

to map genes for different traits in Chinese cabbage crops effectively and get more complete gene sequences. In order to highlight the significance of the positioning interval for better noise reduction, Three appropriate statistical approach were used to observed same two significant peak of *qBrCR38-1* and *qBrCR38-2*. SNP-ratio and ED^4 increasing CR sensitivity near the high allele frequency locus in the CR-pool. G' is expected to decrease much more rapidly around the causal site, implying narrower intervals of support around QTLs³⁰. BSA-Seq combination with KASP analysis is a powerful approach for fine mapping of causal genes. KASP is a cost-effective single-step genotyping technology, cheaper than SSRs and more flexible than genotyping by sequencing (GBS) or array-based genotyping. In our research, KASP markers linked CR were obtained in the target region of QTLs. Alignments of the physical map corresponds to the genetic map around target regions were showed good collinearity (Fig. 4). The inconsistencies such as Br_K_080103 might be caused by chromosomal rearrangement in Pakchoi. Eleven SNP sites were significantly associated to CR, but nine SNP sites have weak effective to CR in chromosome A08. Thus it is speculated that *qBrCR38-1* is a major effective locus. In general, our research provide an important way to molecular marker assisted selection for CR on backcross breeding and gene pyramiding in *B.rapa*.

Several recent studies reported that there were at least eight pathotypes of *P. brassicae* pathogen. Pathotypes 2, 3, 4, 5, 6, 7, 10, 11 represent the main races of *P. brassicae* pathogen in China as classified by Williams identification system^{37,38}. The clubroot-resistant genes *Crr1*, *Crr2* and *CRb* have resistance to physiological race 4^{7,8}.

Marker Name	QTL	Chr	Target region (Mb)	Position (bp)	P value
Br_K_070101	<i>qBrCR38-1</i>	A07	19.7–20.6	18391107	0.027*
Br_K_070113	<i>qBrCR38-1</i>	A07	19.7–20.6	19759488	0.026*
Br_K_070105	<i>qBrCR38-1</i>	A07	19.7–20.6	19995501	0.008**
Br_K_070106	<i>qBrCR38-1</i>	A07	19.7–20.6	20107291	0.003**
Br_K_070107	<i>qBrCR38-1</i>	A07	19.7–20.6	20209491	0.002**
Br_K_070115	<i>qBrCR38-1</i>	A07	19.7–20.6	20333647	0.002**
Br_K_070109	<i>qBrCR38-1</i>	A07	19.7–20.6	20389706	0.002**
Br_K_070110	<i>qBrCR38-1</i>	A07	19.7–20.6	20447844	0.003**
Br_K_070116	<i>qBrCR38-1</i>	A07	19.7–20.6	20501616	0.009**
Br_K_070103	<i>qBrCR38-1</i>	A07	19.7–20.6	20813113	0.008**
Br_K_080101	<i>qBrCR38-2</i>	A08	20.0–20.6	19389881	0.345
Br_K_080112	<i>qBrCR38-2</i>	A08	20.0–20.6	19918702	0.240
Br_K_080115	<i>qBrCR38-2</i>	A08	20.0–20.6	19936667	0.328
Br_K_080118	<i>qBrCR38-2</i>	A08	20.0–20.6	20067617	0.356
Br_K_080107	<i>qBrCR38-2</i>	A08	20.0–20.6	20254038	0.484
Br_K_080109	<i>qBrCR38-2</i>	A08	20.0–20.6	20444233	0.531
Br_K_080111	<i>qBrCR38-2</i>	A08	20.0–20.6	20608800	0.652
Br_K_080120	<i>qBrCR38-2</i>	A08	20.0–20.6	20628063	0.536
Br_K_080121	<i>qBrCR38-2</i>	A08	20.0–20.6	20638194	0.652
Br_K_080103	<i>qBrCR38-2</i>	A08	20.0–20.6	21748451	0.048*

Table 3. KASP marker in target region. * $P < 0.05$, ** $P < 0.01$.

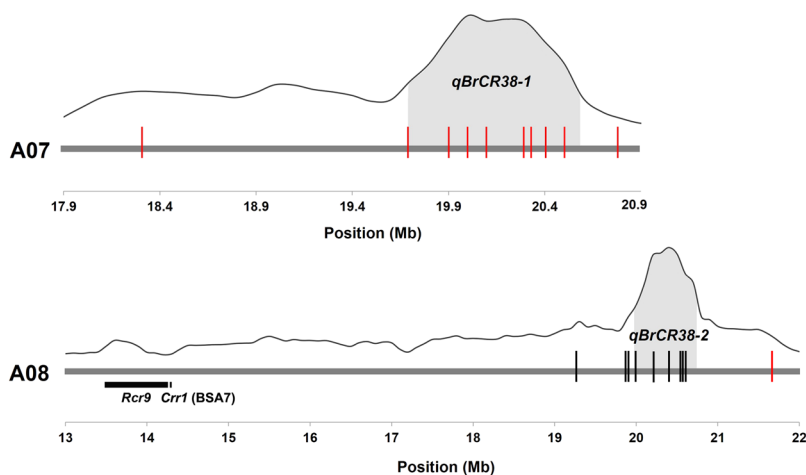


Figure 3. Close-up view of *qBrCR38-1* and *qBrCR38-2* and KASP markers link to the target regions of CR. Vertical lines represent the position of KASP markers, where red lines were significant ($P < 0.05$) associated with CR. Gray areas is the target regions.

Gene ID	Chr	Position	<i>A. thaliana</i> homologs	Annotations
BraA07002249	A07	19177733–19181091	AT4G02310	F-box family protein
BraA07002412	A07	19975263–19977367	AT1G80460	NHO1 (nonhost resistance gene.)
BraA07002494	A07	20387426–20389115	AT1G78750	F-box family protein
BraA08002451	A08	19715762–19716810	AT4G39756	F-box family protein
BraA08002452	A08	19719445–19720378	AT4G39753	F-box family protein
BraA08002455	A08	19759513–19760286	AT1G65850	disease resistance protein (TIR-NBS-LRR class)
BraA08002471	A08	19872938–19875271	AT4G39410	WRKY transcription factor

Table 4. Candidate gene annotation.

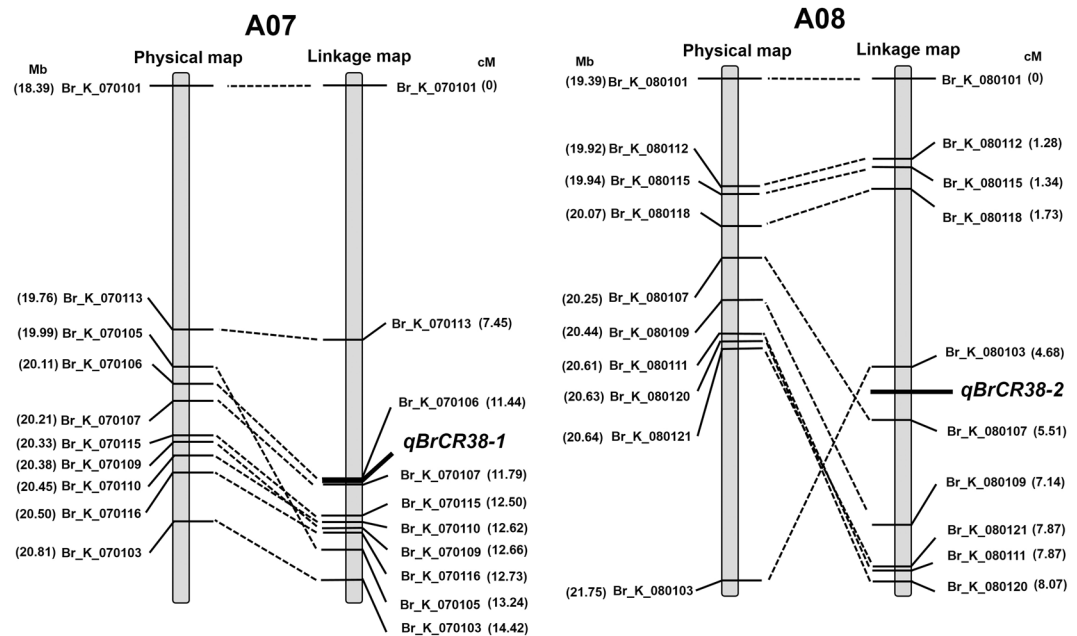


Figure 4. The physical map corresponds to the genetic map around region of *qBrCR38-1* and *qBrCR38-2*. Broken lines drawn regions defined by KASP markers on chromosome A07 and A08. Physical locations are on the left, and the genetic distance is shown on the right.

Rcr1-Rcr9 have high resistance to pathotypes 2, 3, 5, 6 and 8 in Canada^{17,18,39}. In the present study, the resistant parent “CR38” has high resistance to *P. brassicae* of physiological race 7 which is widespread in Shanghai, China. There were no clubroot resistant genes, even no resistant cultivars. Therefore, the development of cultivars and the mapping genes resistant to physiological race 7 in Pakchoi will provide a basis for disease resistance breeding of Chinese cabbage.

Previous researches have applied several strategies to identify and map genes and QTL involved in CR in *B.rapa*. *Crr1* and *Rcr9*¹⁶ were both located in chromosomes A08, while no genes or QTLs were found in chromosomes A07. In the present study, clubroot resistance in “CR38” was identified as quantitative inheritance and was regulated by multiple genes. In addition, *qBrCR38-1* and *qBrCR38-2* in chromosomes A07 and A08 associated with CR in Pakchoi were detected by BSA-seq. We blasted the sequence from linkage marker (BSA7) of *Crr1* and *Rcr9* target region to *B.rapa* reference genome V2.5. *Crr1* and *Rcr9* were located in 14.27 Mb and 13.5–14.2 Mb respectively, thus showing that these two genes were not in the same chromosomal position as *qBrCR38-2* (Fig. 3).

Annotation information and homologous sequence alignment of candidate genes revealed new disease-resistant genes involved in CR and further studies are required. NOH1 was reported to be related to both general and specific resistance against bacterium and fungi⁴⁰. TIR-NBS-LRR resistant protein plays an important role in regulating the activity of plant disease resistance proteins⁴¹. Recent studies suggested that F-box proteins may up-regulate defense-related gene expression in rice⁴², and play a role in cell death and defense responses activated during pathogen recognition in tobacco and tomato⁴³. WRKY transcription factors family is known to be involved in biotic/abiotic stress response⁴⁴. Overall, Genomic intervals of QTL loci still need to be narrowed down to determine candidate genes. Candidate genes identified from the present study provide insights into the genetic mechanism of clubroot resistant in *B.rapa*.

In conclusion, BSA-seq Combine with KASP analysis was carried out to conduct the mapping of CR in Pakchoi. As a result, two novel QTLs regions on chromosome A07 and A08 respectively were detected, and seven genes in target chromosome region associated with the disease resistance were considered as candidates for CR. These genes need to be further studied, and the associated SNPs should be used for marker-assisted breeding of CR in *B.rapa*.

Materials and Methods

Plant materials and phenotypic data collection. Two Pakchoi inbred line, CR38 (Clubroot resistance) and CS22 (Clubroot susceptible), were used as parents to generate 292 F₂ individuals for studies on inheritance and genetic mapping. CR38 is cold tolerant type and highly resistant to the 7th physiology race of *Plasmodiophora Brassicae* by using the inoculation method of Williams (1966)⁴⁵. The pathogen was propagated on Pakchoi, and the clubs in infected roots were stored at -20°C until required. All CR test plants were sown in a pot containing 5×10^6 spores per gram of dry soil, and cultivation was carried out in a growth chamber at 25/20 $^{\circ}\text{C}$ (day/night) with a photo-period of 14 h. At 6 weeks after sowing, the root symptoms of each plant were evaluated. A disease severity index (DSI) of two parents was calculated as according to Suwabe (2003)⁸. Each test included control-resistant and susceptible cultivars. All plant materials examined in this study were obtained from Shanghai academy of agricultural science.

Library construction and Whole genome sequencing. Total genomic DNA from the two parental lines, CR38 and CS22, was isolated from young leaves according to the CTAB method⁴⁶. Equal amounts of genomic DNA were sampled from individuals with leaves (22 DNA samples for each extreme trait) and bulked to generate the Clubroot Resistance pool (CR-pool) and Clubroot Sensitive pool (CS-pool). Pair-end sequencing libraries with a read length of 150 bp and insert sizes of approximately 350 bp were subjected to whole-genome re-sequencing with Illumina HiSeq 2500. Short reads obtained from both parents and two DNA-bulks were aligned against the *B. rapa* Chiifu-401 reference genome sequence v2.5 (<http://brassicadb.org/brad/>) to obtain the consensus sequence using BWA software⁴⁷. Reads of CR-pool and CS-pool were separately aligned to CR38 and CS22 consensus sequence reads to call SNPs with GATK tools software⁴⁸. Heterozygote alleles in two parents were filtered out during the process. Raw read data are archived at the NCBI Sequence Read Archive (SRA) under Accession PRJNA497673.

QTL mapping and target region annotation. Assuming that “A” is a SNP identified between CR38 and CS22, the CR38 and CS22 genotypes at this site are “AA” and “aa”, respectively. For the CR-pool, nRA and nRa are the numbers of reads containing “A” and “a”, respectively. Three methods were used to mapping QTLs. Firstly, SNP-ratio (Resistant alleles/sensitive alleles) of CR-pool and CS-pool were calculated. Then, CR-pool SNP-ratio was divided by CS-pool SNP-ratio and plotted across the genomic regions that showed ratio peaks, which indicate the possible existence of a QTL³⁶

$$\text{Ratio} = (\text{nRA}/\text{nRa})/(\text{nSA}/\text{nSa})$$

Read depth for each allele at segregating allelic SNPs in 500 kb sliding windows was summed using a 100 kb step increment.

The ED for each SNP was calculated according to a previous report³¹. ED is the sum of 100 ED-SNP values within a window of 100 consecutive SNPs. ED⁴ was then calculated by raising ED to the fourth power. Loess fit curves for SNP allele frequency ED⁴. G, the G value averaged across neighboring SNPs, was calculated as described by Magwene *et al.* (2011)³⁰. These methods for mapping common target regions were considered as significant loci.

A list of genes within the target regions was generated using the *B. rapa* genome annotation data (<http://brassicadb.org/brad/>), and these genes were used to query the GO annotation on the website of gene ontology to identify putative resistance related genes for further analysis.

KASP Marker development and genotyping in target region. Polymorphic SNPs identified within the target regions for clubroot resistance were converted into KASP markers (Supplementary Table S1). Each SNP site were parsed to retrieve the flanking sequences 50 bp for KASP markers in the target region. The criteria for selection were that the flanking sequences (a) did not contain any other SNP and InDels, (b) had no more than four consecutive repeats, (c) was close to the ratio peaks on the physical location and (d) Primer sequences were relatively unique in the genome.

For each SNP, two allele-specific forward primers and one common reverse primer were designed using the Primer3 software. Target region SNPs were converted into KASP primers and were used to test the entire F2 population. Reactions were performed in 384 well plates, with a final reaction volume of 3 μ l, which contained 1.48 μ l of KASP 2X reaction mix, 50 ng of template DNA, 0.17 μ M Hex forward primer, 0.17 μ M FAM forward primer and 0.42 μ M universal reverse primer. The following cycling conditions were used: 15 min at 94 °C followed by 10 touchdown cycles of 20 s at 95 °C and 60 s at 65 °C (dropping 0.8 °C per cycle); after the final annealing temperature of 56 °C was achieved, there were 26 cycles of 20 s at 94 °C and 60 s at 57 °C. Thermocycling and fluorescence readings were performed on Hydrocycler and PHERAstar of LGC SNPline platform. Genotyping data viewed as a cluster plot by SNPviewer software supported from LGC Genomics (<http://www.lgcgenomics.com>). The significance of the correlation coefficients between genotype and phenotype was determined with t-tests. Linkage groups were performed with JoinMap 4.1 (Van Ooijen and Voorrips, 2001).

References

- Wallenhammar, A. C. Observations On Yield Loss From Plasmodiophora Brassicae Infections in Spring Oilseed Rape. *Journal of Plant Diseases & Protection*. **105**, 1–7 (1998).
- Dixon, G. R. The Occurrence and Economic Impact of Plasmodiophora Brassicae and Clubroot Disease. *J Plant Growth Regul.* **28**, 194–202 (2009).
- Strelkov, S. E. & Hwang, S. Clubroot in the Canadian Canola Crop: 10 Years Into the Outbreak. *Can J Plant Pathol.* **36**, 27–36 (2014).
- Ingram, D. S. & Tommerup, I. C. The Life History of Plasmodiophora Brassicae Woron. *Roy Soc London Proc Ser B Biol Sci.* **180**, 103–112 (1972).
- Tsushima, S. Perspective of Integrated Pest Management: A Case Study: Clubroot Disease of Crucifers (Pesticide Science in the 21st Century). *J PESTIC SCI.* **25**, 296–299 (2000).
- Voorrips, R. E. Plasmodiophora Brassicae: Aspects of Pathogenesis and Resistance in Brassica Oleracea. *Euphytica.* **83**, 139–146 (1995).
- Suwabe, K. *et al.* Simple Sequence Repeat-Based Comparative Genomics Between Brassica Rapa and Arabidopsis Thaliana: The Genetic Origin of Clubroot Resistance. *Genetics.* **173**, 309–319 (2006).
- Suwabe, K. *et al.* Identification of Two Loci for Resistance to Clubroot (Plasmodiophora Brassicae Woronin) in Brassica Rapa L. *Theoretical & Applied Genetics.* **107**, 997–1002 (2003).
- Matsumoto, E., Yasui, C., Ohi, M. & Tsukada, M. Linkage Analysis of Rflp Markers for Clubroot Resistance and Pigmentation in Chinese Cabbage (Brassica Rapa Ssp. Pekinensis). *Euphytica.* **104**, 79 (1998).
- Piao, Z. Y., Deng, Y. Q., Choi, S. R., Park, Y. J. & Lim, Y. P. Scar and Caps Mapping of Crb, a Gene Conferring Resistance to Plasmodiophora Brassicae in Chinese Cabbage (Brassica Rapa Ssp. Pekinensis). *Tag.theoretical & Applied Genetics.theoretische Und Angewandte Genetik.* **108**, 1458–1465 (2004).

11. Kato, T., Hatakeyama, K., Fukino, N. & Matsumoto, S. Fine Mapping of the Clubroot Resistance Gene *Crb* and Development of a Useful Selectable Marker in Brassica Rapa. *Breed Sci.* **63**, 116–124 (2013).
12. Saito, M. *et al.* Fine Mapping of the Clubroot Resistance Gene, *Crr3*, in Brassica Rapa. *Tag.theoretical & Applied Genetics.theoretische Und Angewandte Genetik.* **114**, 81–91 (2006).
13. Hirai, M. *et al.* A Novel Locus for Clubroot Resistance in Brassica Rapa and its Linkage Markers. *Theoretical & Applied Genetics.* **108**, 639–643 (2004).
14. Sakamoto, K., Saito, A., Hayashida, N., Taguchi, G. & Matsumoto, E. Mapping of Isolate-Specific Qtls for Clubroot Resistance in Chinese Cabbage (Brassica Rapa L. Ssp. Pekinensis). *Theor Appl Genet.* **117**, 759–767 (2008).
15. Matsumoto, E., Ueno, H. & Aruga, D. Accumulation of Three Clubroot Resistance Genes through Marker-Assisted Selection in Chinese Cabbage (Brassica Rapa Ssp. Pekinensis). *Journal of the Japanese Society for Horticultural Science.* **81**, 184–190 (2012).
16. Yu, F. *et al.* Genotyping-by-Sequencing Reveals Three Qtl for Clubroot Resistance to Six Pathotypes of Plasmodiophora Brassicae in Brassica Rapa. *Sci Rep.* **7** (2017).
17. Yu, F. *et al.* Identification of Genome-Wide Variants and Discovery of Variants Associated with Brassica Rapa Clubroot Resistance Gene *Rcr1* through Bulk Segregant Rna Sequencing. *Plos One.* **11**, e153218 (2016).
18. Huang, Z. *et al.* Fine Mapping of a Clubroot Resistance Gene in Chinese Cabbage Using Snp Markers Identified From Bulk Segregant Rna Sequencing. *Front Plant Sci.* **8**, 1448 (2017).
19. Michelmore, R. W., Paran, I. & Kesseli, R. V. Identification of Markers Linked to Disease-Resistance Genes by Bulk Segregant Analysis: A Rapid Method to Detect Markers in Specific Genomic Regions by Using Segregating Populations. *P Natl Acad Sci USA* **88**, 9828–9832 (1991).
20. Watanabe, S. *et al.* A Map-Based Cloning Strategy Employing a Residual Heterozygous Line Reveals that the Gigantea Gene is Involved in Soybean Maturity and Flowering. *Genetics.* **188**, 395–407 (2011).
21. Bolger, M. E. *et al.* Plant Genome Sequencing — Applications for Crop Improvement. *Plant Biotechnol J.* **26**, 31–37 (2014).
22. Takagi, H. *et al.* Qtl-Seq: Rapid Mapping of Quantitative Trait Loci in Rice by Whole Genome Resequencing of Dna From Two Bulk Segregant Populations. *The Plant Journal.* **74**, 174–183 (2013).
23. Abe, A. *et al.* Genome Sequencing Reveals Agronomically Important Loci in Rice Using Mutmap. *Nat Biotechnol.* **30**, 174–178 (2012).
24. Trick, M. *et al.* Combining Snp Discovery From Next-Generation Sequencing Data with Bulk Segregant Analysis (Bsa) to Fine-Map Genes in Polyploid Wheat. *Bmc Plant Biol.* **12**, 14 (2012).
25. Austin, R. S. *et al.* Next-Generation Mapping of Arabidopsis Genes. *Plant Journal for Cell & Molecular Biology.* **67**, 715 (2011).
26. Zheng, W. *et al.* Genetic Mapping and Molecular Marker Development for Pi65(T), a Novel Broad-Spectrum Resistance Gene to Rice Blast Using Next-Generation Sequencing. *Theoretical & Applied Genetics.* **129**, 1035–1044 (2016).
27. KT, W., J, V., C, Z., K, S. & S, L. Qtl Mapping for Downy Mildew Resistance in Cucumber Via Bulk Segregant Analysis Using Next-Generation Sequencing and Conventional Methods. *Theor Appl Genet.* **130**, 199–211 (2017).
28. Zhang, P., Zhu, Y., Wang, L., Chen, L. & Zhou, S. Mining Candidate Genes Associated with Powdery Mildew Resistance in Cucumber Via Super-Bsa by Specific Length Amplified Fragment (Slaf) Sequencing. *Bmc Genomics.* **16**, 1058 (2015).
29. Song, J., Li, Z., Liu, Z., Guo, Y. & Qiu, L. J. Next-Generation Sequencing From Bulk Segregant Analysis Accelerates the Simultaneous Identification of Two Qualitative Genes in Soybean. *Front Plant Sci.* **8** (2017).
30. Magwene, P. M., Willis, J. H. & Kelly, J. K. The Statistics of Bulk Segregant Analysis Using Next Generation Sequencing. *Plos Comput Biol.* **7**, e1002255 (2011).
31. Hill, J. T. *et al.* Mmapp: Mutation Mapping Analysis Pipeline for Pooled Rna-Seq. *Genome Res.* **23**, 687–697 (2013).
32. Semagn, K., Babu, R., Hearne, S. & Olsen, M. Single Nucleotide Polymorphism Genotyping Using Kompetitive Allele Specific Pcr (Kasp): Overview of the Technology and its Application in Crop Improvement. *Mol Breeding.* **33**, 1–14 (2014).
33. B T Ertiro, V. O. M. W. & Semagn, M. L. A. K. Comparison of Kompetitive Allele Specific Pcr (Kasp) and Genotyping by Sequencing (Gbs) for Quality Control Analysis in Maize. *Bmc Genomics* (2015).
34. Bourras, S. *et al.* Multiple Avirulence Loci and Allele-Specific Effector Recognition Control the Pm3 Race-Specific Resistance of Wheat to Powdery Mildew. *Plant Cell.* **27**, 2991 (2015).
35. Xi-jin, H. Advances in Breeding of Non-Heading Chinese Cabbage. *Journal of Nanjing Agricultural University* (2003).
36. Soyk, S. *et al.* Bypassing Negative Epistasis On Yield in Tomato Imposed by a Domestication Gene. *CELL.* **169**, 1142–1155 (2017).
37. Shen, X. Q., Nie, K., Wu, Q., Zhang, Y. G. & Meng, X. H. Initial Research Report On Differentiation Identification of Chinese Cabbage Clubroot Main Physiological Races. *China Vegetables* (2009).
38. Zhang, H. *et al.* Resistance to Plasmodiophora Brassicae in Brassica Rapa and Brassica Juncea Genotypes From China. *Plant Dis.* **99**, 8–14 (2015).
39. Feng, J., Jiang, J., Feindel, D., Strelkov, S. E. & Hwang, S. F. The Gene *Cr811* is Present Exclusively in Pathotype 5 and New Emerged Pathotypes of the Clubroot Pathogen Plasmodiophora Brassicae. *Eur J Plant Pathol.* **145**, 615–620 (2016).
40. Kang, L. *et al.* Interplay of the Arabidopsis Nonhost Resistance Gene *Nho1* with Bacterial Virulence. *P Natl Acad Sci USA* **100**, 3519–3524 (2003).
41. Van, O. G. *et al.* Structure-Function Analysis of the Nb-Arc Domain of Plant Disease Resistance Proteins. *J Exp Bot.* **59**, 1383 (2008).
42. Yang, Y., Zhang, L. Y., Zhu, K. & Liu, Y. L. Overexpression of a Rice Defense-Related F-Box Protein Gene *Osdrl1* in Tobacco Improves Disease Resistance through Potentiation of Defense Gene Expression & Dagger. *Physiol Plantarum.* **134**, 440 (2008).
43. Ha, V. D. B. *et al.* The F-Box Protein *Acrl89/Acif1* Regulates Cell Death and Defense Responses Activated During Pathogen Recognition in Tobacco and Tomato. *Plant Cell.* **20**, 697–719 (2008).
44. Phukan, U. J., Jeena, G. S. & Shukla, R. K. Wrky Transcription Factors: Molecular Regulation and Stress Responses in Plants. *Front Plant Sci.* **7**, 760 (2016).
45. Williams, P. H. A System for the Determination of Races of Plasmodiophora Brassicae that Infect Cabbage and Rutabaga. *Phytopathology.* **56**, 624–626 (1966).
46. Doyle, J. J. A Rapid Dna Isolation Procedure for Small Amounts of Fresh Leaf Tissue. *Phytochem Bull.* **19**, 11–15 (1987).
47. Li, H. & Durbin, R. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics.* **25**, 1754–1760 (2009).
48. McKenna, A. *et al.* The Genome Analysis Toolkit: A Mapreduce Framework for Analyzing Next-Generation Dna Sequencing Data. *Genome Res.* **20**, 1297–1303 (2010).

Acknowledgements

This work was supported by the Shanghai Municipal Agriculture Commission 7-8-1 (2014) and National Key R&D Program of China 2017YFD0101803. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We also thank Huazhi Rice Bio-tech co., LTD for providing the LGC SNline platform.

Author Contributions

H.Z. and W.Z. conceived and conducted the major part of the research including preparation of the samples, conducted bio-informatics analysis and wrote the manuscript. X.L. carried out part of the experiments. Y.Z. designed and guided the research and was involved in manuscript preparation. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-44724-z>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019