Analysis of genome-wide variants through bulked segregant RNA sequencing reveals a major gene for resistance to *Plasmodiophora brassicae* in *Brassica oleracea*

Abdulsalam Dakouri¹, Xingguo Zhang¹,³, Gary Peng¹, Kevin C. Falk¹, Bruce D. Gossen¹, Stephen E. Strelkov² & Fengqun Yu¹

Two cabbage (*Brassica oleracea*) cultivars 'Tekila' and 'Kilaheerb' were identified as resistant to several pathotypes of *Plasmodiophora brassicae*. In this study, we identified a clubroot resistance gene (*Rcr7*) in 'Tekila' for resistance to pathotype 3 of *P. brassicae* from a segregating population derived from 'Tekila' crossed with the susceptible line T010000DH3. Genetic mapping was performed by identifying the percentage of polymorphic variants (PPV), a new method proposed in this study, through bulked segregant RNA sequencing. Chromosome C7 carried the highest PPV (42%) compared to the 30–34% in the remaining chromosomes. A peak with PPV (56–73%) was found within the physical interval 41–44 Mb, which indicated that *Rcr7* might be located in this region. Kompetitive Allele-Specific PCR was used to confirm the association of *Rcr7* with SNPs in the region. *Rcr7* was flanked by two SNP markers and co-segregated with three SNP markers in the segregating population of 465 plants. Seven genes encoding TIR-NBS-LRR disease resistance proteins were identified in the target region, but only two genes, *Bo7g108760* and *Bo7g109000*, were expressed. Resistance to pathotype 5X was also mapped to the same region as *Rcr7*. *B. oleracea* lines including 'Kilaheerb' were tested with five SNP markers for *Rcr7* and for resistance to pathotype 3; 11 of 25 lines were resistant, but 'Kilaheerb' was the only line that carried the SNP alleles associated with *Rcr7*. The presence of *Rcr7* in 'Kilaheerb' for resistance to both pathotypes 3 and 5X was confirmed through linkage analysis.

Clubroot, caused by the obligate pathogen *Plasmodiophora brassicae* Woronin, attacks several economically important members of the family Brassicaceae, including canola/oilseed rape (*Brassica napus* L.), broccoli and cabbage (*B. oleracea* L.).¹ The disease has been reported in more than 60 countries, with an estimated annual yield loss globally of 15%.² The pathogen induces the development of characteristic clubbing symptoms on the roots of susceptible plants, which results in wilting, premature yellowing and reduced quality and yield. Clubroot was first reported on canola on the Canadian prairies in 2003³ and is spreading across the prairie region⁴. Crop rotation, increasing soil pH, and fungicide application⁵ have been recommended to manage clubroot on *Brassica* vegetables, but the efficacy of these methods is limited. Genetic resistance remains the most effective strategy for clubroot management⁶. Several pathotypes of *P. brassicae* are present in Canada (pathotypes 2, 3, 5, 5X, 6 and 8), with pathotype 3 being the most prevalent and pathotype 5X being one of the new pathotypes that can overcome the resistance in the current cohort of clubroot resistant cultivars⁸–¹¹.

*B. napus* was originated from the interspecific hybridization of *B. rapa* and *B. oleracea*. The majority of genes for clubroot resistance that have been identified are from *B. rapa* subsp. *rapifera*.¹² Resistance in *B. rapa* is conferred by major and minor resistance genes, and resistance from *B. rapa* has been utilized widely in breeding for resistance to clubroot in *B. rapa* and *B. napus*.¹²,¹³,¹⁴ Several clubroot resistance genes have been identified and

¹Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, Canada. ²Department of Agricultural, Food and Nutritional Science, University of Alberta, Alberta, Canada. ³The college of Agronomy, Henan Agricultural University, Nanyang, China. Correspondence and requests for materials should be addressed to F.Y. (email: fengqun.yu@agr.gc.ca)
mapped to chromosomes of \( B. \) \textit{rapa}; \( CRt \) and \( Rcr8 \) on chromosome \( A02 \), \( Crt16 \) and \( Rcr9 \) on chromosome \( A08 \), \( Crt2 \) on \( A01 \), \( Crt4 \) on \( A06 \), \( Crt37 \), \( CR18 \), \( CRt \), \( CRt \) and \( CRb \) on \( A03 \), \( Crt23 \), \( Rcr2 \) and \( Rcr4 \) on chromosome \( A03 \). \( CRa \), \( CRt \) and \( CRb \) and \( Crt \) have been cloned from \( B. \) \textit{rapa} \( 9-26 \). They encode toll-interleukin-1 receptor nucleotide binding site/leucine-rich repeat (TIR-NBS-LRR or TNL) proteins. In contrast to \( B. \) \textit{rapa}, no major clubroot resistance genes or lines with strong resistance have been identified in \( B. \) \textit{oleracea} \( 27-29 \).

Molecular mapping of genes in plants is based primarily on polymorphisms in genomic DNA sequences. Transcriptome analysis of short sequences generated from RNA sequencing (RNA-Seq) platforms has been used extensively for gene expression profiling and detection variants in many plant systems \( 28-30 \).

RNA-Seq is being widely used to identify single nucleotide polymorphisms (SNPs) that can be used as molecular markers \( 31,32 \). Also, bulk segregant analysis (BSA) has been applied for detecting molecular markers linked to traits of interest, such as disease resistance \( 33 \). In BSA, bulks of plants with contrasting phenotype are generated. Recent studies have used bulked segregant RNA-Seq (BSR-Seq), a combination of BSA and RNA-Seq approaches \( 23,34-36 \) to map genes of interest. Recently, BSR-Seq of genome-wide DNA variants in a \( B. \) \textit{rapa} population were used to map clubroot resistance gene \( Rcr1 \) onto chromosome \( A03 \). More than 70% of the variants between resistant (R) and susceptible (S) bulks were monomorphic in each chromosome except \( A03 \), where a significantly higher percentage of polymorphic variants (PPV) were present \( 28 \). We therefore hypothesized that a gene could be genetically mapped by identifying the PPV in a genome through BSR-Seq.

A previous study identified two cabbage (\( B. \) \textit{oleracea} ) cultivars, ’Tekila’ and ‘Kilaherb’, with strong resistance to the major strains of \( P. \) \textit{brassicae} in western Canada \( 37 \). The objectives of the current study were: (i) to identify resistance gene in the cabbage cultivars; (ii) to test if identification of PPV could be used for gene mapping; (iii) to characterize transcriptome-wide variation and map the resistance gene; (iv) to develop SNP markers tightly linked to the resistance gene; and (v) to examine DNA variation in the target region and identify the most probable candidates for the gene.

Results
Resistance to clubroot in cultivar ‘Tekila’ is controlled by one dominant gene. The resistant hybrid ‘Tekila’ was crossed with a susceptible double-haploid line T010000DH3 to develop an F\(_1\) mapping population, which is equivalent to a BC\(_1\) population, because the resistant parent was a hybrid. Five weeks after inoculation with pathotype 3 of \( P. \) \textit{brassicae}, the plants were rated for clubroot symptoms. Of the 465 \( F_1\) plant tested, 232 plants were resistant (R) and 233 were susceptible (S). The segregation ratio was 1R: 1S (\( \chi^2 = 0.7, P = 0.40 \)), which indicated that resistance to pathotype 3 in this cultivar was likely controlled by a single dominant gene, designated as \( Rcr7 \) (Resistance to clubroot 7).

Sequence alignment and read mapping. Six bulks were generated; three from the R plants and three from the S plants. RNA from each individual bulk was sequenced, resulted in six RNA sequence files. The total sequence counts were over 44 million (M) for R bulks, and 41 M for the S bulks. Approximately 80% and 83% of the total counts were assembled to reference genome in R and S pools, respectively. The total length of coding sequences assigned to chromosomes in the reference genome was 62.6 Mb. The total accumulated length of coding sequences aligned to chromosomes was ~447 MB for each of the R and S pools. This gives an estimated 7-fold depth of coverage of the gene coding portion of the reference genome \( 38 \).

Transcriptome haplotyping. A haplotype profile was established based on the alignment between the coding sequences of the reference genome and RNA-Seq from the R and S pools. In total, 89 haplotypes were identified, with 48 biallelic and 41 triallelic types (Supplementary Table S1). Biallelic haplotypes were the most frequent within the coding region, with a genome-wide frequency (GWF) of 145,151. In contrast, triallelic types were rare, with a GWF of 170 (Supplementary Table S1). Also, 34% of variants at triallelic sites were non-synonymous and 66% were synonymous. Chromosome C3 contained the highest number of biallelic haplotypes, with chromosome-wide haplotype frequency (CWF) of 23,906. Chromosome C6 had the lowest number, with a CWF of 1,1956. Chromosome C7 had the highest number of triallelic haplotypes, with a CWF of 32, while C6 had the lowest number, with a CWF of 7 (Supplementary Table S1).

Analysis of variants and mapping of \( Rcr7 \) through analysis of PPV. Based on alignment with the reference genome, about 156 K variants were detected in either R or S pools, with 154 K SNPs and 2.6 K InDels in each pool (Supplementary Table S2). Also, 36% of variants per pool were nonsynonymous and 64% were synonymous (Supplementary Table S2). Overall, 36.5% of the SNPs and 51.4% of the InDels were polymorphic (Supplementary Fig. S1).

Chromosome C7 carried the highest PPV (42%) compared to the other eight chromosomes (30–34%) (Fig. 1a). This indicated that \( Rcr7 \) was likely located on C7, based on PPV. In addition, a PPV peak (56–73%) was found within the physical interval 41–44 Mb (Fig. 1b), which indicated that \( Rcr7 \) likely resided in this region of chromosome C7.

Confirmation of \( Rcr7 \) location and fine mapping. There were 1668 SNP sites identified between physical positions 41–44 Mb of chromosome C7 where \( Rcr7 \) was mapped using PPV (Supplementary Table S3). To fine map the gene, 465 \( F_1\) plants including 90R and 90S plants for RNA-Seq in the \( F_1\) population were analyzed with eight SNP markers at the physical position 41–44 Mb of chromosome C7 using Kompetitive Allele-Specific PCR (KASP) (Table 1, Supplementary Table S3). \( Rcr7 \) was flanked by SNP_C7_44 and SNP_C7_56 at 0.4 and 1.1 centi-Morgan (cM) respectively, in an interval of 1.5 cM. Three SNP markers (SNP_C7_34, SNP_C7_43 and SNP_C7_68) co-segregated with \( Rcr7 \) (Fig. 2), confirming that \( Rcr7 \) is located at the physical position 41–44 Mb of chromosome C7.
Identification of variants in the target region. SNP_C7_44 was located in gene Bo7g108740 encoding an Aquaporin protein at site 42,863,773 and SNP_C7_56 in gene Bo7g109090 encoding a Receptor-like protein kinase 1-like protein at site 43,094,738 causing synonymous mutations (Fig. 3). The physical distance between these two markers was 230,966 bases. There were 36 genes, including 6 genes that encoded disease resistance proteins and 1 gene that encoded disease resistance-responsive protein in this region, based on Blast2Go search (Supplementary Table S4). Blast search at http://www.arabidopsis.org/wublast/index2.jsp indicated that all seven disease resistance genes encoded TNL proteins. The number of polymorphic variants (SNP and InDel) uniquely identified from the R bulk was further assessed in the seven genes. Five TNL genes (Bo7g108830, Bo7g108840, Bo7g108850, Bo7g108870 and Bo7g109070) in the region did not show any expression and no short reads were assembled into the reference genome, so no polymorphic variants could be identified in the genes (Table 2). Short reads were identified in two TNL genes, Bo7g108760 and Bo7g109000, in both the R and S bulks (Table 2). The coding region in Bo7g108760 is 2724 bp in length (Table 2). One polymorphic InDel was identified (Table 2, Fig. 3). The coding region in Bo7g109000 is 564 bp in length and no polymorphic SNPs or InDels between the R and S bulks were identified (Table 2).

The homologous regions of the A-genome of B. rapa corresponding to the seven TNL genes were searched. Two expressed TNL genes, Bo7g108760 and Bo7g109000, as well as two unexpressed TNL genes, Bo7g108870 and Bo7g109070, were homologous to the B. rapa genes Bra019305, Bra019277, Bra019297 and Bra019273 respectively. All four B. rapa genes reside in the 25 Mb region of chromosome A03. The unexpressed gene, Bo7g108840, was homologous to the B. rapa gene Bra001161 in the 15 Mb region of A03; Bo7g108830 and Bo7g108850 were homologous to the B. rapa genes Bra013698 and Bra027889, which are located at 7 Mb of A01 and 10 Mb of A09 respectively (Table 2).

Genetic mapping of clubroot resistance to pathotype 3 of P. brassicae in ‘Kilaherb’. The two flanking markers and three co-segregating markers (Table 1) were assessed on additional 25 accessions of B. oleracea, including ‘Kilaherb’ (Table 3). Eleven accessions were resistant to pathotype 3 of P. brassicae, with 0 DSI, but only ‘Kilaherb’ carried the SNP alleles associated with Rcr7.
To determine if the clubroot resistance in 'Kilaherb' was associated with Rcr7, a segregating population consisting of 50 F1 plants from the cross 'Kilaherb' × T010000DH3 was assessed with the eight SNP markers identified previously. The population segregated in a 1:1 ratio (25R and 25S) to pathotype 3 of P. brassicae, indicating that 'Kilaherb' carried a single dominant clubroot resistance gene. All eight SNP markers associated with Rcr7 in 'Tekila' were also associated with resistance in 'Kilaherb' (Fig. 4). Three SNP markers (SNP_C7_34, SNP_C7_43 and SNP_C7_68) that co-segregated with Rcr7 also co-segregated with resistance in 'Kilaherb', confirming that 'Kilaherb' carries Rcr7.

**Table 1.** List of KASP primers, together with their physical locations and flanking sequences.

| Marker ID | Physical location | Flanking sequence |
|-----------|-------------------|------------------|
| SNP_C7_20 | 42111835          | CCATGGAGAAGCTTTGAGAATGCGCTCAAACAGGTGATCCCTTGTGGGTTTCAAGCGATA[G/A] |
| SNP_C7_34 | 42889307          | CCGGGAAGTATCGGAAGCTTTTCTGAGGAAAGTATTTTGGGCTCTAGGAAAGGAAAGGAA |
| SNP_C7_42 | 42705812          | CTTTGGCAACATCAACTCTCAGGGAGGAGACATAACCTACCCGGCC[A/G] |
| SNP_C7_43 | 42887446          | TTTTGGGACTTGGAGGAGACATTTTGCTGAGGAAAGTATTTTGGGCTCTAGGAAAGGAAAGGAA |
| SNP_C7_44 | 42863773          | CATGTTCTGGACACTCGTCACTGACTGAAAGAGCTTTTCTTGGACACTCGTCACTGACTGAAAGGAAAGGAA |
| SNP_C7_45 | 42581612          | GATGGAACTCACAATCTCTCAGCCTGATGATTGGCATCTTCATCTCCGTGACGGCGATCT[T/C] |
| SNP_C7_56 | 43097438          | GCTTTGTTACTTGGAGGAGACATTTTGCTGAGGAAAGTATTTTGGGCTCTAGGAAAGGAAAGGAA |
| SNP_C7_68 | 42888001          | CAAGGTTGAAATATTGGCGACGCATTCAGGAGGCTCAATGAGTTTGGAGGCGATCT[T/C] |

**Figure 2.** Mapping of Rcr7 in the F1 population derived from the cross 'Tekila' × T010000DH3 with SNP markers using the KASP method. R for resistant, S for susceptible on the right, SNP markers on the top. PCR products amplified from R alleles are denoted in black and those from S alleles in hatched. The recombinants were identified from the F1 population consisting of 465 plants tested with pathotype 3 of P. brassicae.

To determine if the clubroot resistance in 'Kilaherb' was associated with Rcr7, a segregating population consisting of 50 F1 plants from the cross 'Kilaherb' × T010000DH3 was assessed with the eight SNP markers identified previously. The population segregated in a 1:1 ratio (25R and 25S) to pathotype 3 of P. brassicae, indicating that 'Kilaherb' carried a single dominant clubroot resistance gene. All eight SNP markers associated with Rcr7 in 'Tekila' were also associated with resistance in 'Kilaherb' (Fig. 4). Three SNP markers (SNP_C7_34, SNP_C7_43 and SNP_C7_68) that co-segregated with Rcr7 also co-segregated with resistance in 'Kilaherb', confirming that 'Kilaherb' carries Rcr7.

**Genetic mapping of resistance to pathotype 5X of P. brassicae in 'Tekila' and 'Kilaherb'.** 'Tekila' and 'Kilaherb' were resistant to pathotype 5X of P. brassicae. To test if the resistance to 5X was associated with Rcr7, a subset of the 'Tekila' and 'Kilaherb' segregating populations comprising 88 F1 plants each were inoculated with pathotype 5X. There were 50 R and 38 S plants in the 'Tekila' population and 43 R and 55 S plants in the 'Kilaherb' population. Segregation for R and S in both populations was consistent with an expected ratio of 1:1 ($\chi^2 = 1.64$ and $P = 0.20$ in the 'Tekila' population and $\chi^2 = 1.47$ and $P = 0.23$ in the 'Kilaherb' population). When both populations were tested with the eight SNP markers linked to Rcr7, all of the markers co-segregating with resistance to pathotype 5X in 'Tekila' and the genetic map for resistance to pathotype 5X for 'Kilaherb' was exactly the same as that for resistance to pathotype 3 (Fig. 4).
Discussion

Until recently, genetic mapping to identify genetic loci governing traits of interest was time-consuming and laborious. Development of high throughput sequencing (HTS) technology has made sequencing of whole genomes comparatively quick and affordable. One application of HTS is mapping by sequencing (MBS). MBS has been used to map causal genes in the genomes of organisms such as *Arabidopsis thaliana*[^3],[^39],[^40], *Caenorhabditis elegans*[^41], wheat[^35] and maize[^36]. A previous study identified a high proportion of PPV on chromosome A03 of *B. rapa* adjacent to the clubroot resistance gene *Rcr1*[^28]. The current study used NGS and identification of PPV to map a major clubroot resistance gene from *B. oleracea*, which was designated as *Rcr7*. A higher PPV was identified within the physical interval 41–44 Mb of chromosome C7 in *B. oleracea*, which indicated that *Rcr7* was likely located in this region. This was confirmed via conventional mapping using linkage analysis. Taken together, these results supported the proposal that identification of PPV could be used for genetic mapping of genes of interest. However, its application and efficacy comparing with other MBS methods still needs to be determined.

[^3]: Arabidopsis thaliana
[^39]: Arabidopsis thaliana
[^40]: Arabidopsis thaliana
[^41]: Caenorhabditis elegans
[^35]: wheat
[^36]: maize
[^28]: Rcr1
[^42]: Previous QTL studies
[^43]: Rcr1, Rcr2 and Rcr4
[^44]: Rcr8 and Rcr9
[^45]: Pathotype 5X

---

**Table 2.** List of genes encoding disease resistance related proteins in the *Rcr7* C-genome target region and homologous regions in A-genome of *Brassica rapa*. Each of these genes were in the TIR-NBS-LRR class, based in gene sequence blasts at [http://www.arabidopsis.org/wublast/index2.jsp](http://www.arabidopsis.org/wublast/index2.jsp). Gene sequence was blasted at [http://brassicadb.org/brad/blastPage.php](http://brassicadb.org/brad/blastPage.php).
to pathotypes 2, 3, 5, 6 and 8

**Table 3.** Clubroot severity (disease severity index, DSI) and SNP marker profiles on accessions of *Brassica oleracea*. ‘+’ The presence of SNP allele associated with *Rcr7*; ‘−’ absence of SNP allele associated with *Rcr7*. 

| Accession ID | Name            | DSI | SNP_C7 marker |
|--------------|-----------------|-----|---------------|
|              |                 |     | 45 | 34 | 68 | 44 | 56 |
| CN 35413     | White Flowered  | 78  | −  | −  | −  | −  | −  |
| Kailaan-Big Boy | Kailaan-Big Boy | 44  | −  | −  | −  | −  | −  |
| CN87022      | Polycaul       | 50  | −  | −  | −  | −  | −  |
| CGN14040     | Kailan         | 71  | −  | −  | −  | −  | −  |
| CGN14041     | Kailan         | 86  | −  | −  | −  | −  | −  |
| Chinese kale | Kailaan-Li      | 0   | −  | −  | −  | −  | −  |
| Chinese kale | Kailaan-Li      | 0   | −  | −  | −  | −  | −  |
| CK 1312      | Nobel Jade     | 38  | −  | −  | −  | −  | −  |
| CHK1058      | Green Jade     | 73  | −  | −  | −  | −  | −  |
| MU538B       | UI Lan Midwater | 0   | −  | −  | −  | −  | −  |
| MU550B       | Green Pearl    | 0   | −  | −  | −  | −  | −  |
| 422C         | Guy Lon        | 100 | −  | −  | −  | −  | −  |
| 1004         | 1004           | 0   | −  | −  | −  | −  | −  |
| 1005         | 1005           | 30  | −  | −  | −  | −  | −  |
| 1006         | 1006           | 0   | −  | −  | −  | −  | −  |
| Jl04         | Jl04           | 0   | −  | −  | −  | −  | −  |
| Jl02         | Jl02           | 24  | −  | −  | −  | −  | −  |
| H01          | H01            | 50  | −  | −  | −  | −  | −  |
| H03          | H03            | 0   | −  | −  | −  | −  | −  |
| H02          | H02            | 0   | −  | −  | −  | −  | −  |
| H01          | H01            | 33  | −  | −  | −  | −  | −  |
| B04          | B04            | 0   | −  | −  | −  | −  | −  |
| B03          | B03            | 11  | −  | −  | −  | −  | −  |
| DH3          | T010000DH3     | 100 | −  | −  | −  | −  | −  |
| Kilaherb     | Kilaherb       | 0   | +  | +  | +  | +  | +  |
| Tekila       | Tekila         | 0   | +  | +  | +  | +  | +  |

...
Only 2 of the 7 TNL genes associated with Rcr7 (Bo7g108760 and Bo7g109000) were expressed in this study. No polymorphic variants were identified in Bo7g109000, but one polymorphic InDel was found in Bo7g108760. Therefore, Bo7g108760 is possibly a candidate of Rcr7. However, the profiles of gene expression and DNA variants were not fully captured in this study due to low depth of sequencing, so other candidates may exist. The identity of the TNL gene corresponding to Rcr7 will be addressed after the gene has been cloned.

The transition forms (C → T, G → A) dominated single-base substitutions with the ratio of 1.5 (transition): 1 (transversion). This ratio was similar to the 2:1 ratio reported from the human genome. Despite the biallelic nature of most SNPs in the current analysis, some loci with three segregating nucleotides may also exist.

Figure 4. Genetic mapping of clubroot resistance in 'Kilaherb' × T010000DH3 F1 population for resistance to pathotypes 3 and 5X.
in certain populations. Triallelic SNPs were previously reported in the human genome at a very low frequency compared to biallelic variants. Human transcriptome analysis showed that triallelic sites would likely cause non-synonymous changes. In the current study, 34% of the triallelic SNPs resulted in non-synonymous variation in the B. oleracea reference genome.

**Materials and Methods**

Schematic flowchart of the experimental procedure is shown in Supplementary Fig. S2.

**Plant mapping population.** The study assessed F1 populations derived from crosses of two clubroot-resistant hybrid cabbage cultivars, 'Tekila' and 'Kilaherb' (Syngenta Canada), with a clubroot-susceptible doubled haploid (DH) line 'T010000DH3' derived from Chinese kale (B. oleracea) cultivar 'TO1434' (Saskatoon Research and Development Centre, Saskatoon, SK, Canada). 'Tekila' and 'Kilaherb' were completely resistant to pathotypes 2, 3, 5, 6 and 8 of *P. brassicae* in an initial screening and strains of pathotype 5X were used for inoculation in this study. Clubroot reaction system (Applied Biosystem, Mississauga, ON). Linkage analysis was performed using JoinMap V.4.54.

www.kbioscience.co.uk/download/KASP.swf. The assays were performed using a StepOne Plus real time qPCR instrument (Applied Biosystems, USA). The KASP assay and linkage analysis.

**Evaluation of plants for resistance to clubroot.** A field population of *P. brassicae* identified as pathotype 322 and a field population L-G02 of pathotype 5X were used for inoculation in this study. Clubroot reaction was assessed in controlled environment studies, as described by Chu, et al. Briefly, the plants were assessed for clubroot symptoms at 5 weeks after seeding using a 0 to 3 scale, where 0 = no symptoms and 3 = large galls. A clubroot rating of 0 was defined as R, 1–3 as S. A disease severity index (DSI) was computed based on 7-14 plants of each of the *B. oleracea* accessions.

**RNA sequencing, read mapping, variant analysis and gene annotation.** At 12 days after seeding, leaf tissue was collected from each F1 plant for RNA extraction. The leaf tissue of each individual plant was stored at −80°C until after its clubroot reaction had been assessed. Then the F1 plants were separated into resistant and susceptible pools. From each pool, three bulks were randomly generated, with each bulk containing 30 plants. RNA was extracted from each bulk using Qiagen RNA extraction kits as per manufacturer's instructions (Qiagen, Toronto, ON). A cDNA library for each bulk was constructed following Illumina TruSeq RNA Sample Preparation v2 Guide (RS-122-9001DOC, Illumina Inc.; San Diego, CA). Sequencing samples were prepared using MiSeq Reagent Kit V3, as per the manufacturer's instruction (MS-102-3001, Illumina Inc.). Sequencing was performed at the University of Saskatchewan, Saskatoon, SK using an Illumina MiSeq® System (Illumina Inc.) with 75 cycles, which resulted in 75 bp pair-end reads. The short sequence reads of the pooled resistance bulks and pooled susceptible bulks were mapped to the nine chromosomes (C1 to C9) of *B. oleracea* reference genome v2.1.38 following the pooled sample assembly method described by28 using SeqMan NextGen in DNASTAR.12 software (Lasergene Inc., Madison, WI). For variant analysis, pooled sample assembly (PSA) was selected because it tends to produce a more complete sequence coverage compared to single-sample assembly.28

Variants (SNPs and Indels) analysis was performed using SeqMan Pro 12 in DNASTAR.12 software, utilizing the following parameters: SNP% ≥ 15%, P not ref ≥50%, Q-call ≥15 and depth ≥5.

Gene annotation was analyzed with Blast2GO. Further confirmation of the genes with TNL domains was performed with Arabidopsis thaliana WU-BLAST2 Search at http://www.arabidopsis.org/wublast/index2.jsp.

**KASP assay and linkage analysis.** The sequences flanking the selected SNP sites were used for designing KASP primers (Table 1). DNA oligoicers were synthesized at Integrated DNA Technologies Inc. (Coralville, IA, USA). The primers carry standard FAM or HEX tails (FAM tail: 5′GAAGGTCGAGTCAACGGATT3′; hex TAIL 5′GAAGGTCGAGTCAACGGATT3′) with the targeted SNP at the 3′ end. KASP assays were performed following the principle and procedure available at http://www.kbioscience.co.uk/reagents/KASP_manumat.pdf and http://www.kbioscience.co.uk/download/KASP.swf. The assays were performed using a StepOne Plus real time qPCR system (Applied Biosystem, Mississauga, ON). Linkage analysis was performed using JoinMap V.4.54.

**References**

1. Karling, J. S. The Plasmodiophorales 2nd ed. Hafner Publishing Company, Inc., New York (1968).
2. Dixon, G. R. The occurrence and economic impact of *Plasmodiophora brassicae* on canola. *Can. Plant Dis. Surv.* 61, 65–70 (1981).
3. Tewari, J. P. Identification of clubroot of crucifer crops. *J. Plant Growth Regul.* 18, 117–127 (1999).
4. Voorrips, R. E. *Arabidopsis thaliana* gene expression database. (2002).
5. Donald, E. C. & Porter, I. Integrated control of clubroot. *Annu. Rev. Phytopathol.* 45, 403–413 (2007).
6. Voorrips, R. E. *Arabidopsis thaliana* gene expression database. (2002).
7. Diederichsen, E., Beckmann, J., Schonadelmeier, J. & Dreyer, F. Genetics of clubroot resistance in *Brassica napus* ‘Mendel’. *Acta Hort.* 706, 307–311 (2006).
8. Xue, C., Cao, T., Howard, R. J., Hwang, S. F. & Strelkov, S. E. Isolation and variation in virulence of single-spore isolates of *Plasmodiophora brassicae* from Canada. *Plant Dis.* 92, 456–462 (2008).
9. Strelkov, S. E. et al. Incidence of clubroot in Alberta in 2007. *Can. Plant Dis. Surv.* 88, 101–102 (2008).
10. Strelkov, S. E., Hwang, S. F., Manolidis, V. P., Cao, T. & Feindel, D. Emergence of new virulence phenotypes of *Plasmodiophora brassicae* on *Brassica napus* in Alberta, Canada. *Eur. J. Plant Pathol.* 145, 517–529 (2016).
11. Yu, F. et al. Genotyping-by-sequencing reveals three QTL for clubroot resistance to six pathotypes of *Plasmodiophora brassicae* in *Brassica rapa*. *Scientific Reports* (in press) (2017).
12. Buczacki, S. T. et al. Study of physiologic specialization in *Plasmodiophora brassicae*: Proposals for attempted rationalization through an international approach. *Trans. Br. Mycol. Soc.* 65, 295–303 (1975).
13. Gustafsson, M. & Falt, A. S. Genetic studies on resistance to clubroot in *Brassica napus*. *Annu. Appl. Biol.* 108, 409–415 (1986).
14. Diederichsen, E., Beckmann, J., Schonadelmeier, J. & Dreyer, F. Genetics of clubroot resistance in *Brassica napus* ‘Mendel’. *Acta Hort.* 706, 307–311 (2006).
A Strategy for Direct Mapping and Identification of Mutation by Whole-Genome Sequencing. This work was supported by Canola Agri-Science Cluster of Growing Forward II.

Acknowledgements

We thank Dr. I. Parkin for providing the susceptible parental line T0100000D3H, L. McGregor, M. Kehler, and Drs. T. Song, and M. Karim for technical assistance. Special thanks go to Dr. Markus Hecker for providing MiSeq facility for sequencing. This work was supported by Canola Agri-Science Cluster of Growing Forward II.
Author Contributions
F.Y., G.P. and K.C.F. conceived of the study. A.D. designed and conducted the study. A.D. and F.Y. analyzed data; X.Z. developed the mapping populations; B.D.G. and S.E.S. provided materials and helped with phenotyping; A.D. and F.Y. drafted the manuscript. All authors reviewed the manuscript and approved the final draft.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-36187-5.

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) 2018