Identification of new QTLs for resistance to Plasmodiophora brassicae in Brassica napus using genome wide association mapping

CURRENT STATUS: POSTED

Abdulsalam Dakouri
Saskatoon Research and Development Centre

Mebarek Lamara
Saskatoon Research and Development Centre

Md. Masud Karim
Saskatoon Research and Development Centre

Jinghe Wang
Saskatoon Research and Development Centre

Qilin Chen
Saskatoon Research and Development Centre

Stephen E. Strelkov
University of Alberta

Sheau-Fang Hwang
University of Alberta

Bruce D. Gossen
Saskatoon Research and Development Centre

Gary Peng
Saskatoon Research and Development Centre

Fengqun Yu
Agriculture and Agri-Food Canada

Corresponding Author
ORCiD: https://orcid.org/0000-0002-7957-7632

DOI:
SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
Brassica napus, Plasmodiophora brassicae, genetic resistance, genome-wide association mapping, resistance
Abstract
Background Clubroot of canola (Brassica napus), caused by the obligate pathogen Plasmodiophora brassicae Woronin, is a major disease worldwide. Genetic resistance remains the best strategy to manage this disease. The objective of the study was to identify and map new sources of resistance to clubroot in B. napus using genome-wide association mapping. The reaction of a collection of 177 accessions to four highly virulent pathotypes of P. brassicae was assessed. These pathotypes were selected because they were most recently identified and showed different virulence patterns on the Canadian clubroot differential (CCD) lines. The collection was then genotyped using genotyping by sequencing (GBS) method. Multi-locus mixed linear model (MMLM) was used to perform the association analysis.

Results The majority of accessions were highly susceptible (70 -100 DSI), while few individual accessions showed strong resistance (0-20 DSI) to 5X (2 accessions), 2B (7 accessions), 3A (8 accessions) and 3D (15 accessions). In total, 301,753 SNPs were mapped to 19 chromosomes. Population structure analysis indicated that the 177 accessions belong to two major populations. SNPs were associated with resistance to each pathotype using MLMM. In total, 23 significant SNP loci were identified, with 14 SNPs mapped to the A-genome and 9 to the C-genome. The SNPs were associated with resistance to pathotypes 5X (4 SNPs), 2B (9), 3A (5) and 3D (5). A blast search of 2 Mb upstream and downstream identified 61 disease resistance genes, of which 24 belonged to TIR-NBS-LRR proteins and 20 belonged to CC-NBS-LRR proteins. The distance between a SNP locus and the nearest resistance genes ranged from 0.11–1.66 Mb. This indicated that NBS-LRR gene family might have an important role in clubroot resistance in B. napus.

Conclusion The resistant B. napus lines and the SNP markers identified in this study can be used for breeding for resistance to clubroot and contribute to understanding the genetic mechanism of resistance to clubroot.

Background Canola (Brassica napus L.), also known as oilseed rape, is grown around the globe [1]. Canola is the largest crop in Canada by acreage (22,535,000 acres) and farm gate value ($16.7 billion Cdn).
Demand for a healthy oil for human consumption, biofuel production, and use of meal as a high quality feed for livestock, have produced strong prices and steadily increasing canola production (Canola Council of Canada 2019).

*Brassica napus* (AACC, 2n = 38), is a natural amphidiploid species resulting from several hybridization events between its progenitors; *B. rapa* L. (AA genome, 2n = 20) and *B. oleracea* L. (CC genome, 2n = 18) [2]. The A and C genomes remained distinct, so recombination events are rare and there is no chromosomal rearrangement [3,4]. There are two forms of *B. napus*; biannual vegetables such as rutabaga and swede, and annual oilseed or fodder subspecies that contains many economically important oil, leafy and root vegetables, and fodder crops [5].

Clubroot, caused by *Plasmodiophora brassicae* Woronin, is an important disease of canola/oilseed rape and other brassica species worldwide [6,7]. Management based on genetic resistance has been effective [8,9] but is generally not durable. The clubroot resistance (CR) genes identified have been almost exclusively from *B. rapa* [10,11]. Nonetheless, a few resistant *B. napus* lines have been used as sources of CR genes [12,13].

Linkage mapping has been used extensively to study important qualitative and quantitative traits, but it is limited to detecting pairs of alleles representing the two parents of the mapping population [14]. Association mapping of population based on linkage disequilibrium (LD) between alleles within diverse populations can be used to detect potential association between markers and traits of interest [15]. The result of advances in next-generation sequencing technologies is that analysis of genotype by sequencing has become feasible for plant species with large genome size [16]. Association mapping has been used in various host-pathogen systems including wheat [17], tomato [18], maize [19] and canola [13,20]. Genome-wide association mapping in *B. napus–P. brassicae* has been limited to one study using already established SNP array and a single pathotype [13].

In the current study, a world collection of 177 *B. napus* germplasm was genotyped by GBS approach and tested for their reaction to four very recently identified pathotypes of *P. brassicae*. These pathotypes exhibited various virulence patterns on the Canadian clubroot differential (CCD) lines. The objective was to identify novel sources of resistance to clubroot from this large collection of *B. napus*
accessions by (1) screening the collection under controlled environment, (2) assessing genetic
diversity and structure analysis of the core collection, and (3) conducting association mapping of
resistance to clubroot in this collection.

Full Text

Identification of new QTLs for resistance to Plasmodiophora brassicae in Brassica napus using genome
wide association mapping

Abdulsalam Dakouri¹, Mebarek Lamara¹,², Md. Masud Karim¹, Jinghe Wang¹, Qilin Chen¹, Stephen E.
Strelkov³, Sheau-Fang Hwang³, Bruce D. Gossen¹, Gary Peng¹, Fengqun Yu¹

¹ Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, SK
S7N 0X2, Canada.

² Institut de Recherche sur les Forêts, Université du Québec en Abitibi-Témiscamingue, Rouyn-
Noranda, QC J9X 5E4, Canada.

³ Department of Agricultural, Food and Nutritional Science, University of Alberta, Alberta, Canada.

*Corresponding author: Fengqun Yu. Email: fengqun.yu@canada.ca

Emails address of authors
Abdulsalam Dakouri : abdulsalam.dakouri@canada.ca

Mebarek Lamara: mebarek.lamara@uqat.ca

Md. Masud Karim: masud.karim@canada.ca

Jinghe Wang: jinghe.wang@canada.ca

Qilin Chen: qilin.chen@canada.ca

Bruce Gossen: bruce.gossen@canada.ca

Stephen E. Strelkov: strelkov@ualberta.ca

Sheau-Fang Hwang: sh20@ualberta.ca

Gary Peng: gary.peng@canada.ca

Fengqun Yu: fengqun.yu@canada.ca

Abstract

Background
Clubroot of canola (Brassica napus), caused by the obligate pathogen Plasmodiophora brassicae Woronin, is a major disease worldwide. Genetic resistance remains the best strategy to manage this disease. The objective of the study was to identify and map new sources of resistance to clubroot in B. napus using genome-wide association mapping. The reaction of a collection of 177 accessions to four highly virulent pathotypes of P. brassicae was assessed. These pathotypes were selected because they were most recently identified and showed different virulence patterns on the Canadian clubroot differential (CCD) lines. The collection was then genotyped using genotyping by sequencing (GBS) method. Multi-locus mixed linear model (MMLM) was used to perform the association analysis.

Results
The majority of accessions were highly susceptible (70 –100 DSI), while few individual accessions showed strong resistance (0–20 DSI) to 5X (2 accessions), 2B (7 accessions), 3A (8 accessions) and 3D (15 accessions). In total, 301,753 SNPs were mapped to 19 chromosomes. Population structure analysis indicated that the 177 accessions belong to two major populations. SNPs were associated with resistance to each pathotype using MLMM. In total, 23 significant SNP loci were identified, with 14 SNPs mapped to the A-genome and 9 to the C-genome. The SNPs were associated with resistance to pathotypes 5X (4 SNPs), 2B (9), 3A (5) and 3D (5). A blast search of 2 Mb upstream and downstream identified 61 disease resistance genes, of which 24 belonged to TIR-NBS-LRR proteins and 20 belonged to CC-NBS-LRR proteins. The distance between a SNP locus and the nearest resistance genes ranged from 0.11–1.66 Mb. This indicated that NBS-LRR gene family might have an important role in clubroot resistance in B. napus.

Conclusion
The resistant B. napus lines and the SNP markers identified in this study can be used for breeding for resistance to clubroot and contribute to understanding the genetic mechanism of resistance to clubroot.

Keywords: Brassica napus, Plasmodiophora brassicae, genetic resistance, genome-wide association mapping, resistance.

Background
Canola (Brassica napus L.), also known as oilseed rape, is grown around the globe [1]. Canola is the
largest crop in Canada by acreage (22,535,000 acres) and farm gate value ($16.7 billion Cdn). Demand for a healthy oil for human consumption, biofuel production, and use of meal as a high quality feed for livestock, have produced strong prices and steadily increasing canola production (Canola Council of Canada 2019).

*Brassica napus* (AACC, 2n = 38), is a natural amphidiploid species resulting from several hybridization events between its progenitors; *B. rapa* L. (AA genome, 2n = 20) and *B. oleracea* L. (CC genome, 2n = 18) [2]. The A and C genomes remained distinct, so recombination events are rare and there is no chromosomal rearrangement [3,4]. There are two forms of *B. napus*; biannual vegetables such as rutabaga and swede, and annual oilseed or fodder subspecies that contains many economically important oil, leafy and root vegetables, and fodder crops [5].

Clubroot, caused by *Plasmodiophora brassicae* Woronin, is an important disease of canola/oilseed rape and other brassica species worldwide [6,7]. Management based on genetic resistance has been effective [8,9] but is generally not durable. The clubroot resistance (CR) genes identified have been almost exclusively from *B. rapa* [10,11]. Nonetheless, a few resistant *B. napus* lines have been used as sources of CR genes [12,13].

Linkage mapping has been used extensively to study important qualitative and quantitative traits, but it is limited to detecting pairs of alleles representing the two parents of the mapping population [14]. Association mapping of population based on linkage disequilibrium (LD) between alleles within diverse populations can be used to detect potential association between markers and traits of interest [15]. The result of advances in next-generation sequencing technologies is that analysis of genotype by sequencing has become feasible for plant species with large genome size [16]. Association mapping has been used in various host-pathogen systems including wheat [17], tomato [18], maize [19] and canola [13,20]. Genome-wide association mapping in *B. napus-P. brassicae* has been limited to one study using already established SNP array and a single pathotype [13].

In the current study, a world collection of 177 *B. napus* germplasm was genotyped by GBS approach and tested for their reaction to four very recently identified pathotypes of *P. brassicae*. These pathotypes exhibited various virulence patterns on the Canadian clubroot differential (CCD) lines. The
objective was to identify novel sources of resistance to clubroot from this large collection of *B. napus* accessions by (1) screening the collection under controlled environment, (2) assessing genetic diversity and structure analysis of the core collection, and (3) conducting association mapping of resistance to clubroot in this collection.

**Results**

**Evaluation of clubroot reaction**

At 6 weeks after seeding, the germplasm was evaluated for resistance to four *P. brassicae* pathotypes; 5X, 2B, 3A and 3D. The disease severity index (DSI) ranged from 0 to 100. The majority of accessions were highly to completely susceptible (70–100 DSI), but several were highly resistant (0–20 DSI) to pathotypes 5X (21 accessions), 2B (7 accessions), 3A (8 accessions), and 3D (15 accessions) (Fig.1, Table S1). The correlation coefficient of severity among the four pathotypes was strongest between 2B and 3A ($r^2 = 0.77$) and weakest between 5X and 3D ($r^2 = 0.27$, Table S2).

Phenotypic data were transformed using rank-based inverse normal transformation to make the DSI values nearly fit the normal distribution required for parametric model-based association analysis (Figure S1).

**Sequence analysis and SNP discovery**

Genotyping by sequencing (GBS) data analysis was performed for the 177 *B. napus* accessions. A total of ~1.2 billion reads and ~633 million good barcoded reads were generated and split into three FASTQ fills. On average, there were 3.3 M read counts per sample (range ~1.8 to 7.7 M) and 3.1 M read counts mapped (range 76 to 96%). Sequence tags from each file were captured and merged to produce a master tag file of 4,253,499 sequence tags. The tags were then aligned to *B. napus* reference genome v4.1, using the TASSEL-GBS pipeline. A total of 2,217,292 (52.1%) tags were uniquely aligned to the reference, 1,220,090 (28.7%) aligned to multiple positions and 816,117 (19.2%) were not aligned. Uniquely mapped tags were used to calculate the tag density distribution at each site in the *B. napus* genome and for SNP calling.

The raw sequence data for SNP calling were also analysed using the TASSEL-GBS pipeline. A total of 399,234 unfiltered SNPs and 355,680 filtered SNPs were called for the 177 accessions, with a mean of individual depth of 8.5 ± 2 SD and mean site depth of 6.7 ± 11.4 SD. Of the 355,680 filtered SNPs,
301,753 SNPs were mapped to the 19 chromosomes; the remaining SNPs were randomly distributed without specific chromosome assignment. Only variants mapped to chromosomes were kept for further analyses.

**Variant analysis and annotation**

There were more SNPs in the C-genome (160,174 SNPs) than the A-genome (141,579 SNPs). Chromosome A03 had the highest number of SNPs within the A-genome, while C03 contained the highest number of SNPs in the C-genome (Table 1). The mean density per Kb was 2.12 SNP / Kb across the 19 chromosomes. In general, SNP density was higher in the C-genome (2.55 SNPs / Kb) than the A-genome (1.70 SNPs / Kb). C07 had the highest number of SNPs per Kb (2.88) and A10 had the lowest (1.43) (Table 1). The vast majority of SNPs were bi-allelic (90%), and only 10% were multi-allelic (Figure S2). There was a positive correlation \( (r^2 = 0.80) \) between chromosome length and the number of SNPs, but only a weak correlation \( (r^2 = 0.3) \) between the number of SNPs and the number of SNPs per Kb.

The SNPs were annotated using the VariantAnnotation package of R. About 37% of SNPs were annotated within coding regions, 22% within introns, 31% within promoter regions, 0.3% within splice sites, and 9.7% mapped to other genetic regions (Figure S3). A more detailed SNP annotation was performed using the Variant Effect Predictor (Figure S3). For SNPs within coding regions, 17% were non-synonymous, 18% were upstream-gene variants, 9% were downstream-gene variants, 23% were synonymous variants, 14% were intron variants, 15% intergenic variants, and 4% were located in the splice site regions and 5’ and 3’ UTRs (Figure S2C). Overall, more SNPs were annotated to the A-genome than the C-genome (Figure S3).

**Genetic diversity and population structure**

For genetic diversity analysis, the SNP markers were filtered at a minor allele frequency (MAF) of 0.05 and minimum sample count of 80%, which resulted in 140,195 good quality SNPs. The mean MAF was the same for the A- and C-genomes (MAF = 0.14). Chromosome C01 had the highest MAF (0.16), followed by C03 and A07 (0.15), and lowest in chromosomes A09 and C09 (0.12) (Table 1). The mean marker heterozygosity \( (H_e) \) was 0.06 and the mean accession heterozygosity was 0.14. The average
polymorphic information content (PIC) was the same for A and C-genomes (0.26). PIC was highest in chromosome C01 (0.27) and lowest (0.24) in A09 (Table 1). The ratio of transitions (changes from A <-> G and C <-> T) to transversions (changes from A <-> C, A <-> T, G <-> C or G <-> T) was 3.22. Population structure analysis indicated the existence of two major group populations, and analysis using the Evanno criterion supported this result (Fig. 2). Population 1 contained 63 accessions (35.6%) representing all continents, while population 2 contained 114 accessions (64.4%), mainly from Europe. A phylogenetic tree using the neighbour-joining algorithm produced two major clusters and six subclusters (Figure S4).

Analysis of molecular variance
Analysis of molecular variance on the six subclusters (SCA-I, II, III,SCB-I, II and III) identified significant genetic differences between major clusters, among subclusters, and among individuals within subclusters ($p < 0.001$). Variance within subclusters accounted for 87.7% of the total variance, with only 7.5% among sub-cluster and 4.7% among major clusters. The fixation index ($F_{st}$) value was 0.21, which indicated that the accessions belonged to two closely related groups (Table S3). Sub-cluster pairwise $F_{st}$ values ranged from 0.03 between SCB-I and SCB-II to 0.16 between SCA-I and SCB-II (Table S4).

Linkage disequilibrium analysis
Linkage disequilibrium in the association panel was calculated using Pearson’s $r^2$ statistic on pairwise combinations of SNPs present across the 19 chromosomes of $B.\ napus$ (Figure S5). The average LD ($r^2$) across the genome was 0.15. The mean LD was 0.10 in the A-genome and 0.19 in the C-genome. LD values ranged from 0.01 in A09 to 0.19 in C01 (Table 1). Across the genome, LD decayed very rapidly ($r^2 = 0.20$) within 300 Kb (Figure S5).

Association analysis
Genome-wide association analysis for clubroot severity was conducted using the following models: general linear model (GLM), mixed linear model (MLM), compressed mixed linear model (CMLM), enriched compressed mixed linear model (ECMLM), and multi-locus mixed model (MLMM). The quantile-quantile (Q-Q) plots, from all models revealed that, save for significant SNPs, the distribution
of observed -log10(p) was closest to the expected distribution in the MLMM compared to other models, therefore associations were identified using this model. A significance threshold of $P < 0.5/N$ (N: number of SNPs) was used for detecting significant SNPs. The MLMM-genome-wide association study (GWAS) detected 23 SNPs associated with resistance to the four *P. brassicae* pathotypes including four SNPs associated with resistance to 5X, nine SNPs to 2B, five to 3A and five to 3D. The name, physical position, $P$ value and -log($P$ value) are presented in Table 2. Across genome, the A-genome carried 14 SNP loci and the C-genome carried 11 loci (Table 2, Fig. 3).

**Candidate resistance genes**

A Blast search identified 61 nucleotide binding site/leucine-rich repeat (NBS-LRR) resistance proteins and non-NBS-LRR resistance genes within the 2 Mb sequence upstream and downstream of 19 out of 23 significant SNP loci detected in our study (Table 3). The majority of resistance genes appeared as clusters of 2 to 10 genes, while they appeared as a single gene in other cases. On A01, one resistance gene (*BnaA01g28560D*) was found at ~0.5 Mb from the A01_19406286 locus associated with resistance to pathotype 5X. On A03, one Enhanced Disease Resistance 2-like (*BnaA03g03110D*) gene and two TIR-NBS-LRR resistance (*BnaA03g03260D, BnaA03g03270D*) genes were detected at 0.11–0.2 Mb distance from A03_651104541 locus associated with resistance to pathotype 2B. Additionally, a cluster of two TIR-NBS-LRR resistance (*BnaA03g44070D, BnaA03g44080D*) genes and a Disease Resistance RRS1-like isoform X1 were detected at 0.73 Mb distance from A03_68863700 locus associated with resistance to pathotype 2B (Table 3). A cluster of six TIR-NBS-LRR resistance (*BnaA03g45000D, BnaA03g45010D, BnaA03g45020D, BnaA03g45040D, BnaA03g45050D*) genes were identified on A03 at 0.11–0.13 Mb from the A03_71057307 locus associated with resistance to pathotype 3D (Table 3). On A04, a gene (*BnaA04g06780D*) encoding a Disease Resistance-Responsive (dirigent-like protein) protein family and four putative disease resistance genes (*BnaA04g06520D, BnaA04g06530D, BnaA04g06550D, BnaA04g06580D*) were identified at 0.63–0.92 Mb from A04_83864566 locus associated with resistance to 2B. On A05, two CC-NBS-LRR (*BnaA05g24990D, BnaA05g25000D*) genes were identified at ~0.2 Mb from A05_115772286 locus associated with resistance to pathotype 2B. On A08, one disease resistance gene (*BnaA08g02210D*) was detected at
GWAS has been widely used to identify and map QTLs for quantitatively inherited traits in a wide
range of plant species. In the current study, GWAS was used to identify and map new sources of resistance to four highly aggressive pathotypes (5X, 2B, 3A, 3D) of P. brassicae in 177 accessions of B. napus. The majority of the accessions were highly susceptible to all four pathotypes (80–100 DSI), while ~10% showed high levels of resistance (0–25 DSI). This supported previous reports that sources of high levels of resistance to clubroot were much less common in B. napus than in B. rapa [11,12]. In total, 23 SNPs were identified: 14 SNPs on the A-genome and 9 on the C-genome. This indicated that the A-genome (from B. rapa) carried more QTLs for clubroot resistance, but the C-genome (from B. oleracea) could be a potential source for clubroot resistance improvement [13].

One of the major factors that may affect the accuracy of GWAS analysis is the existence of population structure within the population used for GWAS. The analysis confirmed that the core collection of accessions represented two different populations. A multi-locus mixed linear model (MMLM) was used to analysis the association between the phenotypes and the SNP markers because it provided the best fit in Q-Q plots between SNP markers and the DSI for the four pathotypes for the models assessed.

QTLs for clubroot resistance have been identified previously in B. napus [21,22] and several have been mapped to chromosomes C03, C06, and C09 [13]. We believe that all 23 of the QTLs identified in the current study are novel because they were located at different physical locations on the chromosomes from QTLs identified previously and were associated with resistance to different pathotypes.

The majority of plant disease resistance genes identified to date have been classified as toll-interleukin-1 receptor/nucleotide binding site/leucine-rich repeat (TIR-NBS-LRR or TNL) proteins or coiled coil/nucleotide binding site/leucine-rich repeat (CC-NBS-LRR or CNL) proteins. The ratio of TNLs to CNLs differs among plant species, likely because their R genes are adapted to different pathogens [23,24]. About 70% of NBS-LRR genes in Brassicaceae family belongs to TNLs [25,26,27].

In the current study, 61 resistance genes were identified within 2 Mb upstream and 2 Mb downstream of the SNPs associated with resistance to the four pathotypes. The resistance genes belonged mainly to the TNL family (24 genes) or the CNL family (20 genes). The frequency of TNLs and CNLs was
highest on A09 (11 genes) followed in decreasing order by A03, A08, C09, C01 and C03. The uneven
distribution of TNLs and CNLs is not uncommon in other plant species [30,31,32,33]. The vast majority
of TNLs and CNLs appeared in clusters of 2 to 10 TNLs and CNLs, which is similar to the results of
previous studies in B. napus [29], Arabidopsis, Medicago truncatula and Solanum tuberosum
[30,34,35].
The remaining resistance genes were non-TNL genes (nTNL), comprised of four enhanced disease
resistance-like genes, two disease resistance-responsive (dirigent-like protein) family genes, and
seven putative disease resistance proteins. A set of nTNLs with RPP13 domain (called RNLs) was also
detected. A group of nTNL genes with RPW8 domain (RNL) had been identified in previous studies
[27,36,37], but was not observed in the current study.
A previous phylogenetic analysis of nTNLs and CNLs from five Brassicaceae species indicated that
RNLs are likely derived from the CNL lineage [27,29]. The function of RNLs is yet to be determined,
but they have no direct response to the pathogen and may have not the same duplication rates as
TNLs and CNLs, which explains their lower abundance in the genome [27,29]. They may have a role in
defence-signal transduction [38] or as helpers of other NBS genes [38]. The role of other nTNLs is also
unknown.
Conclusion
The current study identified several accessions of B. napus with high levels of resistance to four
pathotypes of P. brassicae. Genome-wide association mapping analysis detected and mapped 23 SNP
loci associated with resistance to the four pathotypes. This information will be used in subsequent
genetic analysis of bi-parental populations to verify the SNPs and fine map the functional genes
responsible for resistance to each pathotype and for marker-assisted breeding of resistance to
clubroot in canola.
Materials and methods
Plant and pathogen materials
Germplasm of Brassica napus consisting of 177 accessions from 32 countries, provided by three gene
banks (Plant Genetic Resources of Canada (PGRC), Centre for Genetic Resources of the Netherlands
and Agricultural Research Service, USDA, USA), was selected for study (Table S1). These accessions
represented collections from Europe (123 accessions), Asia (29), North America (20), Oceania (2), South America (1), Africa (1), and one accession of unknown origin (Table S1). The accessions were oilseed rape (146 accessions), fodder rapa (21), Swede rape (7), rutabaga (2) and turnip (1). The growth habit was predominantly winter type (129), with some spring type (48) accessions (Table S1). Plants for GBS analysis were grown in a growth chamber up to the 3–4 leaf stage. A total of 100 mg of leaf tissue was collected from each accession, immediately frozen in liquid nitrogen and then lyophilized in a freeze dryer for approximately 48 h. The freeze-dried tissues were ground to a fine powder using a tissue lyser (Qiagen, Newtown City, USA).

Resting spores of field collections of strains L-G02, F.183–14, F.3–14 and F.1–14 representing pathotypes 5X, 2B, 3A and 3D respectively of *P. brassicae* (Canadian Clubroot Differential) system, [39]were increased on canola and stored as frozen clubbed roots at −20°C until needed. Resting spores were extracted from the frozen clubs as described by [40], and adjusted to a concentration of $1 \times 10^7$ resting spores/mL. Spores of each pathotype were applied separately to the host entries.

**Evaluation of clubroot reaction**

Seed of each host genotype was pre-germinated on moistened filter paper in a Petri dishes. One-week-old seedlings of each host line and pathotype were inoculated by dipping the entire root system in the resting spore suspension for 10 s. The inoculated seedlings were then immediately planted in 6 × 6 × 6 cm plastic pots filled with Sunshine LA4 potting mixture, with one seedling per pot. The pots were thoroughly watered and transferred to a greenhouse at 21°C ± 2°C with a 16 h photoperiod. The potting mixture was kept saturated with tap water at pH 6.5 for the first week after inoculation and then watered and fertilized as required.

Six weeks after inoculation, the seedlings were gently removed from the potting mix, the roots of each plant were washed with tap water, and each root was rated for clubroot symptom development on a 0 to 3 scale [41], where: 0 = no clubs, 1 = a few small clubs on less than one-third of the roots, 2 = moderate clubs (small to medium-sized clubs on 1/3 to 2/3 of the roots), and 3 = severe clubs (medium to large-sized clubs on > 2/3 of the roots). A DSI was then calculated using the formula of [42] as modified by [41]:

15
DSI % = \[\Sigma(n\times0+n\times1+n\times2+n\times3)N\times3\times100\]

Where \(n\) is the number of plants in a class; \(N\) is the total number of plants in an experimental unit; and 0, 1, 2 and 3 are the symptom severity classes.

**Sequence analysis and SNP discovery**

The accession sequences were analyzed using GBS. In brief, GBS involves four major steps: DNA sample preparation, library construction, library sequencing and SNP calling. DNA extraction was performed using the DNeasy 96 plant kit as per the manufacturer’s instruction (Qiagen). To reduce the genome complexity, DNA was digested with ApeKI, a methylation-sensitive restriction enzyme. The fragments produced by digestion were directly ligated to enzyme-specific adapters followed by PCR amplification. The samples divided into two pools of 96 samples each followed by two runs of Illumina HiSeq 2500 (Illumina Inc., USA). DNA alignment was generated with BWA software version 0.7.8-r455. The GBS-TASSEL pipeline [43] was used for SNP calling, and VCF and HapMap genotype files were generated. Initial SNP filtration was performed with the following settings: MAF > 0.01 and missing data per site < 90%. Accessions with too much missing data were removed. Depth, missingness and heterozygosity were calculated using VCFtools V.0.1.12 [44]. Genotyping and SNP calling was performed at the Genomic Diversity Facility, Cornell University (http://www.bio-tech.cornell.edu/brc/brc/services).

**Variant annotation**

Variants were annotated to regions of the *B. napus* reference genome using R, implemented using “VariantAnnotation” [45], and Variant Effect Predictor (VEP, [46]), and variant locations were characterised as coding, intron, splice site, promoter and intergenic regions.

**Genetic diversity and population structure**

Population-based genetic diversity, including allele frequencies, MAF, and average heterozygosity, were computed using TASSEL 5.2.18 software [47]. Polymorphic information content (PIC) values [48] was calculated for SNP markers using the formula \((PIC = 1- (maf^2+(1-maf)^2))-(2maf^2(1-maf)^2))\). The ratio of transitions to transversions was calculated using the [49] 2-parameter model, implemented in MEGA7 [50].

Structure analysis of the accessions was conducted using STRUCTURE software v2.2 [51]. A subset of
10,094 SNPs was selected that was evenly distributed across the genome with one SNP per 100 Kb. The admixture model and correlated allele frequency were applied with a burn-in period of 50,000 iterations and 100,000 replications of Markov Chain Monte Carlo (MCMC). Five runs were performed to calculate the mean likelihood for the number of populations K, ranging from 1 to 10, and the mean of the log-likelihood estimates LnP(D) for each K. The ad-hoc statistic ΔK was used to determine optimal number groups [52]. Structure output was visualized using STRUCTURE HARVESTER web-based software [53]).

Analysis of molecular variance
Analysis of molecular variance (AMOVA) was conducted using Arlequin v.3.5 software [54] to estimate the genetic variance among clusters and sub-clusters of the A and C genome haplotypes. In this analysis, the distance matrix among samples was computed to estimate the genetic structure of the haplotypes. Genetic variance components were estimated, and the total variance was partitioned among major clusters, among sub-clusters within major clusters, and within subclusters. The significance of the variance components was tested using 1,000 permutations. The fixation index \( F_{st} \), an estimation of population differentiation and genetic distance based on genetic polymorphism data, was calculated.

Linkage disequilibrium (LD) analysis
LD decay across the B. napus genome was measured and a correlation matrix of \( r^2 \) values was computed between all pairs of polymorphic SNPs with MAF ≥ 5% using the GAPIT V2 package [55].

Association analysis
Data for the disease DSI were transformed using rank-based inverse normal transformation implemented as the rnttransform function in the GenABEL R [56]. Association was analyzed for a subset of 10,094 SNP markers with MAF ≥ 5% using the following models: general linear model (GLM), mixed linear model (MLM), compressed mixed linear model (CMLM), enriched compressed mixed linear model (ECMLM), and multi-locus mixed model (MLMM) implemented in the GAPIT V2 package of R [55]. A kinship matrix of the accessions was calculated and principle components analysis was used to account for population structure and accessions relatedness.

Candidate resistance genes
Using Blast2Go software [57], the sequence region neighboring (2 Mb upstream and downstream) of the significant SNPs were searched for candidate genes encoding disease resistance proteins potentially responsible for resistance to each pathotype of *P. brassicae*.

**List of abbreviations**

AMOVA: analysis of molecular variance; CC: coiled-coil; CCD: Canadian clubroot differential; CMLM: compressed mixed linear model; CNL: CC-NBS-LRR; CR: clubroot resistance; DSI: disease severity index; ECMLM: enriched compressed mixed linear model; GLM: general linear model; GWAS: genome wide association analysis; GBS: genotyping by sequencing; LD: linkage disequilibrium; LRR: leucine-rich repeat; MAF: minor allele frequency; MCMC: markov chain monte carlo; MLM: mixed linear model; MMLM: multilocus mixed linear model; NBS: nucleotide-binding site; nTNL: non-TIR-NBS-LRR; PGRC: plant genetic resource of Canada; PIC: Polymorphic information content; Q-Q: quantile-quantile; QTL: quantitative trait locus; SNP: single nucleotide polymorphism; TIR: Toll-interleukin–1 receptor; TNL: TIR-NBS-LRR; USDA: United States department of Agriculture; UTR: untranslated region

**Acknowledgments**

The authors are grateful to Melissa Kehler, Victor Manolii, Md Mizanur Rahaman and the summer students Yasmina Bekkaoui and Kurtis Flavel for their technical support.

**Declarations**

**Ethics approval and consent to participate**

Not applicable

**Consent to publish**

Not applicable

**Availability of data and materials**

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

**Competing interests**

The authors declare that they have no competing interest.

**Funding**

This work was funded by a competitive grant from SaskCanola under Canola Agronomic Research Program. The funding body played no role in the design of the study and collection, analysis, and
interpretation of data and in writing the manuscript.

Author contributions
FY and AD conceived of and designed the study; AD and JW conducted the experiments; AD, ML, MMK and QC analyzed data; SES, SFH, BDG and GP provided important resources and facilities. AD drafted the manuscript. All authors reviewed the manuscript and approved the final draft.

Reference
Nagaharu U. Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. Japanese Journal of Botany. 1935; 7: 389–452.
Allender CJ, King GJ. Origins of the amphiploid species Brassica napus L. investigated by chloroplast and nuclear molecular markers. BMC Plant Biol. 2010; 10:54. doi:10.1186/1471-2229-10-54.
Howell EC, Kearsey MJ, Jones GH, King GJ, Armstrong SJ. A and C genome distinction and chromosome identification in Brassica napus by sequential fluorescence in situ hybridization and genomic in situ hybridization. Genetics. 2008;180(4):1849–1857. doi:10.1534/genetics.108.095893
Parkin, IAP, and Lydiate, DJ. Conserved patterns of chromosome pairing and recombination in Brassica napus crosses. Genome. 1997; 40: 496-504.
McNaughton IH. Swedes and rapes—Brassica napus (Cruciferae). In: Smartt J, Simmonds NW (eds) Evolution of crop plants, 2nd edn. Longman Scientific & Technical, London, UK. 1995; 68-75.
Dixon, G. R. The occurrence and economic impact of Plasmodiophora brassicae and clubroot disease. J. Plant Growth Regul. 2009; 28:194–202.
Karling JS. The Plasmodiophorales 2nd ed. Hafner Publishing Company, Inc., New York (1968).
Voorrips, RE.Plasmodiophora brassicae: Aspects of pathogenesis and resistance in Brassica oleracea. Euphytica. 1995; 83:139 -146 (1995).
Diederichsen E, Frauen M, Linders E, Hatakeyama K, Hirai M. Status and perspectives of clubroot resistance breeding in crucifer crops. J. Plant Growth Regul. 2009; 28:265–281.
Buczacki ST, Toxopeus H, Mattusch P, Johnston TD, Dixon GR, Hobolth LA. Study of physiologic specialization in Plasmodiophora brassicae: proposals for attempted rationalization through an international approach. Transaction of British Mycological Society. 1975; 65:295-303.
Piao Z, Ramchiary N, Lim YP. Genetics of clubroot resistance in Brassica species. J. Plant Growth Regul. 2009; 28: 252–264.
Rahman H, Shakir A, Hasan MJ. Breeding for clubroot resistant spring canola (Brassica napus L.) for the Canadian prairies: Can the European winter canola cv. Mendel be used as a source of resistance? Can. J. Plant Sci. 2011; 91: 447–458.
Li L, Luo Y, Chen B, et al. A Genome-Wide Association Study Reveals New Loci for Resistance to Clubroot Disease in Brassica napus. Front Plant Sci. 2016;7:1483. doi:10.3389/fpls.2016.01483
Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, et al. Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. 2005; 44: 1054-1064
Gupta PK, Rustgi S, Kulwal PL. Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol. Biol. 2005; 57: 461–485.
Elshire RJ, Glaubitz JC, Sun Q, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011; 6:e19379. doi:10.1371/journal.pone.0019379.
Crossa J, Burgueno J, Dreisigacker S, et al. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics. 2007; 177:1889-1913.
Ranc N, Muños S, Xu J, et al. Genome-wide association mapping in tomato (Solanum lycopersicum) is
possible using genome admixture of Solanum lycopersicum var. cerasiforme. G3. 2012; 2: 853–864. doi:10.1534/g3.112.002667.

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, et al. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. 2011; 43: 163–168.

Wang N, Chen B, Xu K, et al. Association Mapping of Flowering Time QTLs and Insight into Their Contributions to Rapseed Growth Habits. Front Plant Sci. 2016;7:338. doi:10.3389/fpls.2016.00338.

Manzanares-Dauleux MJ, Delourme R, Baron F, Thomas G. Mapping of one major gene and of QTLs involved in resistance to clubroot in Brassica napus. Theor. Appl. Genet. 2000; 101: 885–891. 10.1007/s001220051557

Werner S, Diederichsen E, Frauen M, Schondelmaier J, Jung C. Genetic mapping of clubroot resistance genes in oilseed rape. Theor. Appl. Genet. 2008; 116: 363–372. 10.1007/s00122-007-0674-2.

Leister D. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends In Genetics: TIG. 2004; 20: 116–122. doi:10.1016/j.tig.2004.01.007.

Lozano R, Ponce O, Ramirez M, Mostajo N, Orjeda G. Genome-wide identification and mapping of NBS-encoding resistance genes in Solanum tuberosum Group Phureja. PLoS One. 2012; 7: e34775. doi:10.1371/journal.pone.0034775

Mun J, Yu H, Park S, Park B. Genome-wide identification of NBS encoding resistance genes in Brassica rapa. Molecular Genetics and Genomics. 2009; 282: 617–631. doi:10.1007/s00438–009–0492–0.

Yu J, Tehrim S, Zhang F, Tong C, Huang J, Cheng X, Dong C, Zhou Y, Qin R, Hua W, Liu S. Genome-wide comparative analysis of NBSencoding genes between Brassica species and Arabidopsis thaliana. BMC Genomics. 2014; 15: 3. doi:10.1186/1471-2164-15-3.

Zhang YM, Shao ZQ, Wang Q, Hang YY, Xue JY, Wang B, Chen JQ. Uncovering the dynamic evolution of nucleotide-binding site-leucinerich repeat (NBS-LRR) genes in Brassicaceae. Journal of Integrative Plant Biology. 2016; 58:165–177. doi:10.1111/jipb.12365.

Dakouri A, Zhang X, Peng G et al. Analysis of genome-wide variants through bulked segregant RNA sequencing reveals a major gene for resistance to Plasmodiophora brassicae in Brassica oleracea. Scientific Reports. 2018; 8:17657.

Alamery S, Tinnaz S, Bayer P, et al. Genome-wide identification and comparative analysis of NBS-LRR resistance genes in Brassica napus. Crop Pasture Sci. 2017; 69: 79–94. doi: 10.1071/CP17214.

Meyers B, Kozik A, Griego A, Kuang H, Michelmore R. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. The Plant Cell. 2003; 15: 809–834. doi:10.1105/tpc.009308.

Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Molecular Genetics and Genomics. 2004; 271: 402–415. doi:10.1007/s00438004–0990-z.

Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers B, Boerjan W, Martin F. Genome-wide identification of NBS resistance genes in Populus trichocarpa. Plant Molecular Biology. 2008; 66: 619–636. doi:10.1007/s11103-008–9293-9.

Yang S, Zhang X, Yue J-X, Tian D, Chen J-Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. Molecular Genetics and Genomics. 2008; 280: 187–198. doi:10.1007/s00438008–0355–0.

Ameline-Torregrosa C, Wang B-B, O’Blenness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant Medicago truncatula. Plant Physiology. 2008; 146: 5–21. doi:10.1104/pp.107.104588.

Jupe F, Pritchard L, Etherington G, MacKenzie K, Cock P, Wright F, Sharma SK, Bolser D, Bryan G, Jones J, Hein I. Identification and localisation of the NB-LRR gene family within the potato genome. BMC Genomics. 2012; 13: 75. doi:10.1186/1471-2164-13-75.
Bonardi V, Tang S, Stallmann A, Roberts M, Cherkis K, Dangl JL. Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108: 16463–16468. doi:10.1073/pnas.1113726108

Collier SM, Hamel L-P, Moffett P. Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. Molecular Plant-Microbe Interactions. 2011; 24: 918–931. doi:10.1094/MPMI–03–11–0050.

Shao Z-Q, Zhang Y-M, Hang Y-Y, Xue J-Y, Zhou G-C, Wu P, Wu X-Y, Wu X-Z, Wang Q, Wang B. Long-term evolution of nucleotide binding site-leucine-rich repeat genes: understanding gained from and beyond the legume family. Plant Physiology. 2014; 166: 217–234. doi:10.1104/pp.114.243626.

Strelkov, S. E., Hwang, S. F., Manolii, V. P., Cao, T., Fredua-Agyeman, R., Harding, M. W., Peng, G., Gossen, B. D., McDonald, M. R., and Feindel, D. Virulence and pathotype classification of Plasmodiophora brassicae populations collected from clubroot resistant canola (Brassica napus) in Canada. Can. J. Plant Pathol. 20185; 40: 284–298. DOI: 10.1080/07060661.2018.1459851.

Strelkov SE, Tewari JP, Smith‐Degenhardt E. Characterization of Plasmodiophora brassicae populations from Alberta, Canada. Canadian Journal of Plant Pathology. 2006; 28: 467–474.

Kuginuki Y, Yoshikawa H, Hirai M. Eur. J. Plant Pathol. 1999; 105: 327–332.

Horiuchi S, and Hori M. Bull. Chugoku Natl. Agric. Exp. Stn. Ser. E. (Environ. Div.), 1980; 17:33–55.

Glaubitz JC., Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS ONE. 2014; 9:E90346. doi:10.1371/journal.pone.0090346

Roussel V, Koenig J, Beckert M, Balfourier F. Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. Theor Appl Genet. 2004; 108:920–930

Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution. 1980; 16:111–120.

Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016; 33:1870–1874.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155: 945–959.

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 2005; 4: 2611–2620. doi:10.1111/j.1365-294X.2005.02553.

Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2011; 4: 359–361.

Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010; 10: 564–567.

Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, et al. GAPIT Version 2: An enhanced integrated tool for genomic association and prediction. Plant Genome. 2018; 9: 2–9. doi:
Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007; 23: 1294-1296.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids research. 2008; 36: 3420-3435. doi:10.1093/nar/gkn176.

Table 1 Genome wide distribution of SNPs, minor allele frequency (MAF), Heterozygosity, polymorphic information content (PIC) and average Linkage disequilibrium (LD)

| Chromosome | Start | End | Total No. seq | SNP | SNP/Kb | MAF | Heterozygosity | PIC | Average LD |
|------------|-------|-----|---------------|-----|--------|-----|----------------|-----|------------|
| A01        | 2024  | 23251220 | 23250         | 13062| 1.78   | 0.14| 0.08           | 0.24| 0.090      |
| A02        |       |      |               |     |        |     |                |     |            |
919
24785167
24784
12455
1.99
0.13
0.08
0.23
0.080
A03
808
29746073
29745
20541
1.45
0.14
0.07
0.24
0.060
A04
1717
19141470
19140
10562
1.81
0.14
0.07
|    |    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|----|
| A05 | 2697 | 23052978 | 23050 | 14917 | 1.55 | 0.14 | 0.06 | 0.24 | 0.076 |    |    |    |
| A06 | 2120 | 24372251 | 24370 | 14696 | 1.66 | 0.13 | 0.06 | 0.22 | 0.075 |    |    |    |
| A07 | 10938 | 24000655 | 23990 | 14232 |    |    |    |    |    |    |    |    |
1.69
0.15
0.07
0.24
0.070
A08
1729
18958296
18957
10281
1.84
0.13
0.07
0.22
0.084
A09
1327
33857792
33857
18702
1.81
0.12
0.07
0.21
0.010
A10
4083
|        |        |        |        |        |
|--------|--------|--------|--------|--------|
| 17366872 | 17363  | 12131  | 1.43   | 0.14   |
|        | 0.07   | 0.23   | 0.080  |        |
|        |        |        |        | Average (A-subgenome) |
| 23851  | 14158  | 1.70   | 0.14   | 0.07   |
|        | 0.23   | 0.072  | 0.190  |        |
| C01    | 8039   | 38812658 | 38805  |
|        | 17087  | 2.27   | 0.16   | 0.08   |
|        | 0.27   | 0.190  |        |        |
| C02    |        |        |        |        |
| Chromosome | Start   | End     | Total No. seq | SNP | SNP/Kb | MAF | Heterozygosity |
|------------|---------|---------|---------------|-----|--------|-----|----------------|
|            | 1607    | 46186975|               |     |        |     |                |
|            | 46185   | 17662   |               |     |        |     |                |
|            | 2.61    | 0.14    |               |     |        |     |                |
|            | 0.09    | 0.24    |               |     |        |     |                |
|            | 0.146   |         |               |     |        |     |                |
|            | C03     | 760     |               |     |        |     |                |
|            | 60565276| 60565   |               |     |        |     |                |
|            | 25136   | 2.41    |               |     |        |     |                |
|            | 0.15    | 0.09    |               |     |        |     |                |
|            | 0.24    |         |               |     |        |     |                |
|            | 0.073   |         |               |     |        |     |                |
| PIC | Average LD | C04 | 1773 | 48929072 | 48927 | 19053 | 2.57 | 0.14 | 0.08 | 0.24 | 0.140 |
|-----|------------|-----|------|---------|-------|-------|------|------|------|------|-------|
| C05 | 3386       | 43172068 | 43169 | 16540 | 2.61 | 0.13 | 0.10 | 0.22 | 0.074 |
| C06 | 1745       | 37224854 | 37223 | 14761 |
48501448
48500
18295
2.65
0.12
0.09
0.21
0.075

Average (C-subgenome)
45178
17797
2.55
0.14
0.09
0.23
0.107

Average (genome)
34514
15978
2.12
0.14
0.08
0.23
0.088

Table 2 List of significant SNPs, chromosomes, physical location and P values

Pathotype

SNP locus
| Chromosome | Position  | P.value     | -log(P values) |
|------------|-----------|-------------|----------------|
| A01        | 19406286  | 2.49E-07    | 6.60           |
| A08        | 2719211   | 5.39E-06    | 5.27           |
| A09        | 8083774   | 1.11E-05    | 4.96           |
| C08        | 20763231  | 4.88E-05    | 4.31           |
| A07        | 147954943 |             |                |
3509616
9.95E-10
9.00
A04_83864566
A04
6035183
1.15E-07
6.94
A03_651104541
A03
1393088
5.23E-07
6.28
A08_186203638
A08
17751690
3.94E-06
5.40
2B
C01_276968702
C01
38290946
5.56E-06
5.26
C03_688995474
C03
6293335
| Value  | Label  |
|--------|--------|
| 7.75E-06 | 5.11   |
| A05_115772286 | A05    |
| 18791143   | 9.44E-06 |
| 5.02       | A09_215839211 |
| A09        | 28425222   |
| 2.18E-05   | 4.66     |
| A03_68863700 | A03     |
| 20801907   | 3.63E-05  |
| 4.44       | C01_341096049 |
| C01        | 17366972   |
| 1.95E-08   | 7.71      |
| A10_222415846 | A10    |
| A10        | 1136417   |
| 5.56E-08   | 7.25      |

33
3A
C04_403341747
C04
19039176
3.50E-07
6.46
C09_608174205
C09
11282154
6.73E-07
6.17
C09_634081519
C09
37189468
2.21E-05
4.66
A04_90743077
A04
12913694
7.21E-07
6.14
C09_638459286
C09
41567235
2.76E-06
5.56
3D
| Pathotype | SNP          | Chromosome | Gene ID                  | Candidate gene region (Mb) | Distance to SNP (Mb) | Description |
|----------|--------------|------------|--------------------------|----------------------------|----------------------|-------------|
| 5X       | A03_71057307 | A03        | 22995514                 | 3.13E-06                   | 5.50                 | BnaA01g28560D |
|          | A01_17862282 | A01        | 17862282                 | 3.31E-06                   | 5.48                 |             |
|          | C03_685453245| C03        | 2751106                  | 4.46E-05                   | 4.35                 |             |
| Gene Accession | Description                  | Relative Expression |
|---------------|------------------------------|---------------------|
| BnaA03g03110D | Disease resistance           | 0.47                |
| A03_651104541 | ENHANCED Disease RESISTANCE 2-like | 0.11               |
| BnaA03g03260D | TIR-NBS-LRR class gene       | 0.19                |
| BnaA03g44070D | ENHANCED Disease RESISTANCE 2-like isoform X1 | 1.30               |
20.5–23.5  
0.73  
TIR-NBS-LRR class gene  
A03  
BnaA03g44080D  

20.5–23.5  
0.73  
TIR-NBS-LRR class gene  
A03  
BnaA03g44090D  

Disease resistance RRS1-like isoform X1  
2B  
A03_68863700  
A03  
BnaA03g45000D  

20.5–23.5  
0.13  
TIR-NBS-LRR class gene  
A03  
BnaA03g45010D  

20.5–23.5  
0.12  
TIR-NBS-LRR class gene  
A03  
BnaA03g45020D
20.5–23.5
0.12
TIR-NBS-LRR class gene
3D
A03_71057307
A03
BnaA03g45040D
20.5–23.5
0.11
TIR-NBS-LRR class gene
A03
BnaA03g45050D
20.5–23.5
0.11
TIR-NBS-LRR class gene
A03
BnaA03g45970D
20.5–23.5
0.46
TIR-NBS-LRR class gene
A04
BnaA04g06520D
5.5–6.5
0.92
Putative Disease resistance protein
A04
BnaA04g06530D
5.5–6.5  
0.92  
Putative Disease resistance protein  
2B  
A04_83864566  
A04  
BnaA04g06550D  
5.5–6.5  
0.91  
Putative Disease resistance protein  
A04  
BnaA04g06580D  
5.5–6.5  
0.82  
Putative Disease resistance protein  
A04  
BnaA04g06780D  
5.5–6.5  
0.63  
Disease resistance-responsive (dirigent-like protein)  
2B  
A05_115772286  
A05  
BnaA05g24990D  
18–19  
0.22  
CC-NBS-LRR class gene
A05
BnaA05g25000D
18–19
0.21
CC-NBS-LRR class gene
5X
A08_171171159
A08
BnaA08g02210D
1–3
0.94
CC-NBS-LRR class gene
A08
BnaA08g24820D
17–18
0.52
CC-NBS-LRR class gene
A08
BnaA08g24860D
17–18
0.49
CC-NBS-LRR class gene
2B
A08_186203638
A08
BnaA08g26370D
17–18
0.12
CC-NBS-LRR class gene
A08
BnaA08g26380D
17-18
0.13
CC-NBS-LRR class gene
A08
BnaA08g26400D
17-18
0.14
CC-NBS-LRR class gene
A08_186203638
A08
BnaA08g26410D
17-18
0.14
CC-NBS-LRR class gene
A09
BnaA09g13280D
7-9
0.76
TIR-NBS-LRR class gene
A09
BnaA09g13850D
7-9
0.20
TIR-NBS-LRR class gene
A09
BnaA09g13890D
7–9
0.16

TIR-NBS-LRR class gene
A09
BnaA09g13900D
7–9
0.15

TIR-NBS-LRR class gene
5X
A09_195497763
A09
BnaA09g14320D
7–9
0.11

TIR-NBS-LRR class gene
A09
BnaA09g14420D
7–9
8.26

CC-NBS-LRR class gene
A09
BnaA09g14550D
7–9
0.30
CC-NBS-LRR class gene
A09
BnaA09g14560D
7–9
0.30
CC-NBS-LRR class gene
A09
BnaA09g14570D
7–9
0.31
CC-NBS-LRR class gene
A09
BnaA09g14680D
7–9
0.37
CC-NBS-LRR class gene
A09
BnaA09g42680D
27.1–30.1
1.28
CC-NBS-LRR class gene
A09
BnaA09g43420D
27.1–30.1
1.66
Disease resistance-responsive (dirigent-like protein) family
2B
Disease resistance-responsive (dirigent-like protein) family

ENHANCED Disease RESISTANCE 4-like

ENHANCED Disease RESISTANCE 4-like

ENHANCED DISEASE RESISTANCE-like protein (DUF1336)
CC-NBS-LRR class gene
C01
BnaC01g39050D
37–39
0.46
Disease resistance protein RPS6 isoform X1
C01
BnaC01g40270D
37–39
0.36
TIR-NBS-LRR class gene
3A
C01
BnaC01g40280D
37–39
0.36
TIR-NBS-LRR class gene
C01_276968702
C01
BnaC01g40300D
37–39
0.37
TIR-NBS-LRR class gene
C01
BnaC01g40310D
37–39
0.38
TIR-NBS-LRR class gene
C01
BnaC01g40460D
37–39
0.44
Putative Disease resistance protein At4g11170
2B
C03_685453245
C03
BnaC03g04690D
2–3
0.48
TIR-NBS-LRR class gene
3D
C03_688995474
C03
BnaC03g05380D
2–3
0.16
TIR-NBS-LRR class gene
3A
C04_403341747
C04
BnaC04g18730D
17–19
0.48
CC-NBS-LRR class gene
5X
C08_579178095
C08
BnaC08g17450D
20.7–21.1
0.31
TIR-NBS-LRR class gene
C09
BnaC09g14400D
11–12
0.26
TIR-NBS-LRR class gene
C09
BnaC09g14870D
11–12
0.12
TIR-NBS-LRR class gene
3A
C09_608174205
C09
BnaC09g15010D
11–12
0.24
CC-NBS-LRR class gene
C09
BnaC09g15020D
11–12
0.24  
CC-NBS-LRR class gene  
C09  
BnaC09g15110D  
11-12  
0.39  
CC-NBS-LRR class gene  
3D  
A09_638459286  
C09  
BnaC09g38250D  
41-42  
0.38  
Disease resistance  

Figure legends  
**Fig 1.** Frequency distribution of accessions plotted against clubroot severity (disease severity index, DSI) for four pathotypes 5X, 2B, 3A, and 3D indicated in the figure.  
**Fig 2.** Population structure analysis of the 177 accessions based on A. model-based Bayesian clustering using STRUCTURE for K = 2 groups, and B. estimation of the number of sub-populations for K values of 1 to 10.  
**Fig 3.** Manhattan plots of association analysis using the multilocus mixed linear model (MMLM) model P+K for pathotypes A. 5X, B. 2B, C. 3A and D. 3D. The horizontal line represents the threshold of significance (-log_{10}0.5/10094 = 4.30).  

Additional files  
*Table S1* List of accessions, their growth habits, type, origin and DSIs.  
*Table S2* Pearson correlation coefficient ($r^2$) between DSIs of the four pathotypes.  
*Table S3* Analysis of molecular variance (AMOVA) design and results.
Table S4 Distance method: Sub-Cluster Pairwise FST differences.

Figure S1 Frequency distribution of rank-transformed disease severity index for pathotypes (a) 5X-LG2, (b) 2B, (c) 3A, and (d) 3D.

Figure S2 Relative frequency distribution (%) of Bi-/Multi-allelic SNPs

Figure S3 Variant annotation results, A. the distribution of SNPs within genic regions; coding region, introns and promoter region, B. Derailed SNP annotation based on Variant Effect Predictor software, C. Distribution of annotated SNP across B. napus chromosomes.

Figure S4 Phylogenetic analysis of the 177 B. napus accessions based SNP markers A. Major cluster A and subclusters SCA-I, SCA-II and SCA-III, B. Major cluster B and subcluster SCB-I, SCB-II, SCB-III. The geographical distribution at the continent level is also illustrated.

Figure S5 Genome wide linkage disequilibrium (LD) decay plot as a function of physical distance (bp). LD decay assessed in a B. napus collection of 177 accessions LD estimates are reported as squared correlations of allele frequencies (r2).

Results
Evaluation of clubroot reaction
At 6 weeks after seeding, the germplasm was evaluated for resistance to four P. brassicae pathotypes; 5X, 2B, 3A and 3D. The disease severity index (DSI) ranged from 0 to 100. The majority of accessions were highly to completely susceptible (70–100 DSI), but several were highly resistant (0–20 DSI) to pathotypes 5X (21 accessions), 2B (7 accessions), 3A (8 accessions), and 3D (15 accessions) (Fig.1, Table S1). The correlation coefficient of severity among the four pathotypes was strongest between 2B and 3A ($r^2 = 0.77$) and weakest between 5X and 3D ($r^2 = 0.27$, Table S2).

Phenotypic data were transformed using rank-based inverse normal transformation to make the DSI values nearly fit the normal distribution required for parametric model-based association analysis (Figure S1).

Sequence analysis and SNP discovery
Genotyping by sequencing (GBS) data analysis was performed for the 177 B. napus accessions. A total of ~1.2 billion reads and ~633 million good barcoded reads were generated and split into three FASTQ fills. On average, there were 3.3 M read counts per sample (range ~1.8 to 7.7 M) and 3.1 M
read counts mapped (range 76 to 96%). Sequence tags from each file were captured and merged to produce a master tag file of 4,253,499 sequence tags. The tags were then aligned to *B. napus* reference genome v4.1, using the TASSEL-GBS pipeline. A total of 2,217,292 (52.1%) tags were uniquely aligned to the reference, 1,220,090 (28.7%) aligned to multiple positions and 816,117 (19.2%) were not aligned. Uniquely mapped tags were used to calculate the tag density distribution at each site in the *B. napus* genome and for SNP calling.

The raw sequence data for SNP calling were also analysed using the TASSEL-GBS pipeline. A total of 399,234 unfiltered SNPs and 355,680 filtered SNPs were called for the 177 accessions, with a mean of individual depth of 8.5 ± 2 SD and mean site depth of 6.7 ± 11.4 SD. Of the 355,680 filtered SNPs, 301,753 SNPs were mapped to the 19 chromosomes; the remaining SNPs were randomly distributed without specific chromosome assignment. Only variants mapped to chromosomes were kept for further analyses.

**Variant analysis and annotation**

There were more SNPs in the C-genome (160,174 SNPs) than the A-genome (141,579 SNPs). Chromosome A03 had the highest number of SNPs within the A-genome, while C03 contained the highest number of SNPs in the C-genome (Table 1). The mean density per Kb was 2.12 SNP / Kb across the 19 chromosomes. In general, SNP density was higher in the C-genome (2.55 SNPs / Kb) than the A-genome (1.70 SNPs / Kb). C07 had the highest number of SNPs per Kb (2.88) and A10 had the lowest (1.43) (Table 1). The vast majority of SNPs were bi-allelic (90%), and only 10% were multi-allelic (Figure S2). There was a positive correlation ($r^2 = 0.80$) between chromosome length and the number of SNPs, but only a weak correlation ($r^2 = 0.3$) between the number of SNPs and the number of SNPs per Kb.

The SNPs were annotated using the VariantAnnotation package of R. About 37% of SNPs were annotated within coding regions, 22% within introns, 31% within promoter regions, 0.3% within splice sites, and 9.7% mapped to other genetic regions (Figure S3). A more detailed SNP annotation was performed using the Variant Effect Predictor (Figure S3). For SNPs within coding regions, 17% were non-synonymous, 18% were upstream-gene variants, 9% were downstream-gene variants, 23% were
synonymous variants, 14% were intron variants, 15% intergenic variants, and 4% were located in the splice site regions and 5’ and 3’ UTRs (Figure S2C). Overall, more SNPs were annotated to the A-genome than the C-genome (Figure S3).

**Genetic diversity and population structure**

For genetic diversity analysis, the SNP markers were filtered at a minor allele frequency (MAF) of 0.05 and minimum sample count of 80%, which resulted in 140,195 good quality SNPs. The mean MAF was the same for the A- and C-genomes (MAF = 0.14). Chromosome C01 had the highest MAF (0.16), followed by C03 and A07 (0.15), and lowest in chromosomes A09 and C09 (0.12) (Table 1). The mean marker heterozygosity ($H_e$) was 0.06 and the mean accession heterozygosity was 0.14. The average polymorphic information content (PIC) was the same for A and C-genomes (0.26). PIC was highest in chromosome C01 (0.27) and lowest (0.24) in A09 (Table 1). The ratio of transitions (changes from A <-> G and C <-> T) to transversions (changes from A <-> C, A <-> T, G <-> C or G <-> T) was 3.22.

Population structure analysis indicated the existence of two major group populations, and analysis using the Evanno criterion supported this result (Fig. 2). Population 1 contained 63 accessions (35.6%) representing all continents, while population 2 contained 114 accessions (64.4%), mainly from Europe. A phylogenetic tree using the neighbour-joining algorithm produced two major clusters and six subclusters (Figure S4).

**Analysis of molecular variance**

Analysis of molecular variance on the six subclusters (SCA-I, II, III,SCB-I, II and III) identified significant genetic differences between major clusters, among subclusters, and among individuals within subclusters ($p < 0.001$). Variance within subclusters accounted for 87.7% of the total variance, with only 7.5% among sub-cluster and 4.7% among major clusters. The fixation index ($F_{st}$) value was 0.21, which indicated that the accessions belonged to two closely related groups (Table S3). Sub-cluster pairwise $F_{st}$ values ranged from 0.03 between SCB-I and SCB-II to 0.16 between SCA-I and SCB-II (Table S4).

**Linkage disequilibrium analysis**

Linkage disequilibrium in the association panel was calculated using Pearson’s $r^2$ statistic on pairwise
combinations of SNPs present across the 19 chromosomes of *B. napus* (Figure S5). The average LD ($r^2$) across the genome was 0.15. The mean LD was 0.10 in the A-genome and 0.19 in the C-genome. LD values ranged from 0.01 in A09 to 0.19 in C01 (Table 1). Across the genome, LD decayed very rapidly ($r^2 = 0.20$) within 300 Kb (Figure S5).

**Association analysis**

Genome-wide association analysis for clubroot severity was conducted using the following models: general linear model (GLM), mixed linear model (MLM), compressed mixed linear model (CMLM), enriched compressed mixed linear model (ECMLM), and multi-locus mixed model (MLMM). The quantile-quantile (Q-Q) plots, from all models revealed that, save for significant SNPs, the distribution of observed $-\log10(p)$ was closest to the expected distribution in the MLMM compared to other models, therefore associations were identified using this model. A significance threshold of $P < 0.5/N$ ($N$: number of SNPs) was used for detecting significant SNPs. The MLMM-genome-wide association study (GWAS) detected 23 SNPs associated with resistance to the four *P. brassicae* pathotypes including four SNPs associated with resistance to 5X, nine SNPs to 2B, five to 3A and five to 3D. The name, physical position, $P$ value and $-\log(P$ value) are presented in Table 2. Across genome, the A-genome carried 14 SNP loci and the C-genome carried 11 loci (Table 2, Fig. 3).

**Candidate resistance genes**

A Blast search identified 61 nucleotide binding site/leucine-rich repeat (NBS-LRR) resistance proteins and non-NBS-LRR resistance genes within the 2 Mb sequence upstream and downstream of 19 out of 23 significant SNP loci detected in our study (Table 3). The majority of resistance genes appeared as clusters of 2 to 10 genes, while they appeared as a single gene in other cases. On A01, one resistance gene (*BnaA01g28560D*) was found at ~0.5 Mb from the A01_19406286 locus associated with resistance to pathotype 5X. On A03, one Enhanced Disease Resistance 2-like (*BnaA03g03110D*) gene and two TIR-NBS-LRR resistance (*BnaA03g03260D, BnaA03g03270D*) genes were detected at 0.11–0.2 Mb distance from A03_651104541 locus associated with resistance to pathotype 2B. Additionally, a cluster of two TIR-NBS-LRR resistance (*BnaA03g44070D, BnaA03g44080D*) genes and a Disease Resistance RRS1-like isoform X1 were detected at 0.73 Mb distance from A03_68863700 locus.
associated with resistance to pathotype 2B (Table 3). A cluster of six TIR-NBS-LRR resistance
\(BnaA03g45000D, BnaA03g45010D, BnaA03g45020D, BnaA03g45040D, BnaA03g45050D\) genes
were identified on A03 at 0.11–0.13 Mb from the A03_71057307 locus associated with resistance to
pathotype 3D (Table 3). On A04, a gene \(BnaA04g06780D\) encoding a Disease Resistance-Responsive
(dirigent-like protein) protein family and four putative disease resistance genes \(BnaA04g06520D,
BnaA04g06530D, BnaA04g06550D, BnaA04g06580D\) were identified at 0.63–0.92 Mb from
A04_83864566 locus associated with resistance to 2B. On A05, two CC-NBS-LRR \(BnaA05g24990D,
BnaA05g25000D\) genes were identified at ~0.2 Mb from A05_115772286 locus associated with
resistance to pathotype 2B. On A06, one disease resistance gene \(BnaA08g02210D\) was detected at
0.94 Mb from A08_171171159 locus associated with resistance to pathotype 5X. Also, a cluster of six
CC-NBS-LRR genes were found at 0.12–0.52 Mb from A08_186203638 associated with resistance to
2B. On A09, a cluster of five TIR-NBS-LRR resistance genes \(BnaA09g13280D, BnaA09g13850D,
BnaA09g13890D, BnaA09g13900D, BnaA09g14320D\) and five CC-NBS-LRR genes \(BnaA09g14420D,
BnaA09g14550D, BnaA09g14560D, BnaA09g14570D, BnaA09g14580D\) were detected at 0.11–0.76
Mb from A09_195497763 locus associated with resistance to 5X. In addition, one CC-NBS-LRR gene
\(BnaA09g42680D\), two Disease Resistance-Responsive genes (dirigent-like protein) and two
Enhanced Disease Resistance 4-like genes were found at 0.50–1.67 Mb from A09_215839211 locus
associated with resistance to pathotype 2B. On A10, one enhanced disease resistance-like gene
\(BnaA10g04000D\), and one CC-NBS-LRR gene \(BnaA10g05000D\) were detected at 0.98–1.50 Mb
from the A10_222415846 locus associated with resistance to pathotype 3A.

On C01, four TIR-NBS-LRR resistance genes \(BnaC01g40270D, BnaC01g40280D, BnaC01g40300D,
BnaC01g40310D\) and two non-NBS-LRR disease resistance genes \(BnaC01g39050D,
BnaC01g40460D\) were identified at 0.36–0.46 Mb from the locus at C01_276968702 associated with
resistance to pathotype 3A (Table 3). On C03, two TIR-NBS-LRR resistance genes \(BnaC03g05380D,
BnaC03g04690D\) located at 0.16 Mb and 0.48 Mb from C03_68899547 and C03_685453245 loci
associated with resistance to pathotype 3D were detected. On C04, one TIR-NBS-LRR resistance gene
\(BnaC08g17450D\) and one CC-NBS-LRR resistance gene \(BnaC04g18730D\) mapped at 0.31 Mb and
0.48 Mb, respectively, from the C04_403341747 locus associated with resistance to pathotype 3A (Table 3). On C08, one TIR-NBS-LRR resistance (BnaC08g17450D) gene was identified at 0.31 Mb from C08_579178095 locus associated with resistance to 5X. On C09, two TIR-NBS-LRR resistance genes (BnaC09g14400D, BnaC09g14870D) and three CC-NBS-LRR resistance genes (BnaC09g15010D, BnaC09g15020D, BnaC09g15110D) located at 0.12–0.4 Mb from C09_608174205 locus associated with resistance to 3A. In addition, a non-NBS-LRR disease resistance gene (BnaC09g38250D) located at 0.38 Mb from C09_638459286 locus associated with resistance to 3D was detected (Table 3).

Discussion

GWAS has been widely used to identify and map QTLs for quantitatively inherited traits in a wide range of plant species. In the current study, GWAS was used to identify and map new sources of resistance to four highly aggressive pathotypes (5X, 2B, 3A, 3D) of *P. brassicae* in 177 accessions of *B. napus*. The majority of the accessions were highly susceptible to all four pathotypes (80–100 DSI), while ~10% showed high levels of resistance (0–25 DSI). This supported previous reports that sources of high levels of resistance to clubroot were much less common in *B. napus* than in *B. rapa* [11,12]. In total, 23 SNPs were identified: 14 SNPs on the A-genome and 9 on the C-genome. This indicated that the A-genome (from *B. rapa*) carried more QTLs for clubroot resistance, but the C-genome (from *B. oleracea*) could be a potential source for clubroot resistance improvement [13].

One of the major factors that may affect the accuracy of GWAS analysis is the existence of population structure within the population used for GWAS. The analysis confirmed that the core collection of accessions represented two different populations. A multi-locus mixed linear model (MMLM) was used to analysis the association between the phenotypes and the SNP markers because it provided the best fit in Q-Q plots between SNP markers and the DSI for the four pathotypes for the models assessed.

QTLs for clubroot resistance have been identified previously in *B. napus* [21,22] and several have been mapped to chromosomes C03, C06, and C09 [13]. We believe that all 23 of the QTLs identified in the current study are novel because they were located at different physical locations on the chromosomes from QTLs identified previously and were associated with resistance to different
pathotypes.

The majority of plant disease resistance genes identified to date have been classified as toll-interleukin-1 receptor/nucleotide binding site/leucine-rich repeat (TIR-NBS-LRR or TNL) proteins or coiled coil /nucleotide binding site/leucine-rich repeat (CC-NBS-LRR or CNL) proteins. The ratio of TNLs to CNLs differs among plant species, likely because their R genes are adapted to different pathogens [23,24]. About 70% of NBS-LRR genes in Brassicaceae family belongs to TNLs [25,26,27].

In the current study, 61 resistance genes were identified within 2 Mb upstream and 2 Mb downstream of the SNPs associated with resistance to the four pathotypes. The resistance genes belonged mainly to the TNL family (24 genes) or the CNL family (20 genes). The frequency of TNLs and CNLs was highest on A09 (11 genes) followed in decreasing order by A03, A08, C09, C01 and C03. The uneven distribution of TNLs and CNLs is not uncommon in other plant species [30,31,32,33]. The vast majority of TNLs and CNLs appeared in clusters of 2 to 10 TNLs and CNLs, which is similar to the results of previous studies in B. napus [29], Arabidopsis, Medicago truncatula and Solanum tuberosum [30,34,35].

The remaining resistance genes were non-TNL genes (nTNL), comprised of four enhanced disease resistance-like genes, two disease resistance-responsive (dirigent-like protein) family genes, and seven putative disease resistance proteins. A set of nTNLs with RPP13 domain (called RNLs) was also detected. A group of nTNL genes with RPW8 domain (RNL) had been identified in previous studies [27,36,37], but was not observed in the current study.

A previous phylogenetic analysis of nTNLs and CNLs from five Brassicaceae species indicated that RNLs are likely derived from the CNL lineage [27,29]. The function of RNLs is yet to be determined, but they have no direct response to the pathogen and may have not the same duplication rates as TNLs and CNLs, which explains their lower abundance in the genome [27,29]. They may have a role in defence-signal transduction [38] or as helpers of other NBS genes [38]. The role of other nTNLs is also unknown.

Conclusion

The current study identified several accessions of B. napus with high levels of resistance to four
Pathotypes of *P. brassicae*. Genome-wide association mapping analysis detected and mapped 23 SNP loci associated with resistance to the four pathotypes. This information will be used in subsequent genetic analysis of bi-parental populations to verify the SNPs and fine map the functional genes responsible for resistance to each pathotype and for marker-assisted breeding of resistance to clubroot in canola.

**Materials And Methods**

**Plant and pathogen materials**

Germplasm of *Brassica napus* consisting of 177 accessions from 32 countries, provided by three gene banks (Plant Genetic Resources of Canada (PGRC), Centre for Genetic Resources of the Netherlands and Agricultural Research Service, USDA, USA), was selected for study (Table S1). These accessions represented collections from Europe (123 accessions), Asia (29), North America (20), Oceania (2), South America (1), Africa (1), and one accession of unknown origin (Table S1). The accessions were oilseed rape (146 accessions), fodder rapa (21), Swede rape (7), rutabaga (2) and turnip (1). The growth habit was predominantly winter type (129), with some spring type (48) accessions (Table S1). Plants for GBS analysis were grown in a growth chamber up to the 3–4 leaf stage. A total of 100 mg of leaf tissue was collected from each accession, immediately frozen in liquid nitrogen and then lyophilized in a freeze dryer for approximately 48 h. The freeze-dried tissues were ground to a fine powder using a tissue lyser (Qiagen, Newtown City, USA).

Resting spores of field collections of strains L-G02, F.183–14, F.3–14 and F.1–14 representing pathotypes 5X, 2B, 3A and 3D respectively of *P. brassicae* (Canadian Clubroot Differential) system, [39] were increased on canola and stored as frozen clubbed roots at −20°C until needed. Resting spores were extracted from the frozen clubs as described by [40], and adjusted to a concentration of $1 \times 10^7$ resting spores/mL. Spores of each pathotype were applied separately to the host entries.

**Evaluation of clubroot reaction**

Seed of each host genotype was pre-germinated on moistened filter paper in a Petri dishes. One-week-old seedlings of each host line and pathotype were inoculated by dipping the entire root system in the resting spore suspension for 10 s. The inoculated seedlings were then immediately planted in 6 × 6 × 6 cm plastic pots filled with Sunshine LA4 potting mixture, with one seedling per pot. The pots
were thoroughly watered and transferred to a greenhouse at 21°C ± 2°C with a 16 h photoperiod. The potting mixture was kept saturated with tap water at pH 6.5 for the first week after inoculation and then watered and fertilized as required.

Six weeks after inoculation, the seedlings were gently removed from the potting mix, the roots of each plant were washed with tap water, and each root was rated for clubroot symptom development on a 0 to 3 scale [41], where: 0 = no clubs, 1 = a few small clubs on less than one-third of the roots, 2 = moderate clubs (small to medium-sized clubs on 1/3 to 2/3 of the roots), and 3 = severe clubs (medium to large-sized clubs on > 2/3 of the roots). A DSI was then calculated using the formula of [42] as modified by [41]:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

Where \( n \) is the number of plants in a class; \( N \) is the total number of plants in an experimental unit; and 0, 1, 2 and 3 are the symptom severity classes.

Sequence analysis and SNP discovery
The accession sequences were analyzed using GBS. In brief, GBS involves four major steps: DNA sample preparation, library construction, library sequencing and SNP calling. DNA extraction was performed using the DNeasy 96 plant kit as per the manufacturer’s instruction (Qiagen). To reduce the genome complexity, DNA was digested with ApeKI, a methylation-sensitive restriction enzyme. The fragments produced by digestion were directly ligated to enzyme-specific adapters followed by PCR amplification. The samples divided into two pools of 96 samples each followed by two runs of Illumina HiSeq 2500 (Illumina Inc., USA). DNA alignment was generated with BWA software version 0.7.8-r455. The GBS-TASSEL pipeline [43] was used for SNP calling, and VCF and HapMap genotype files were generated. Initial SNP filtration was performed with the following settings: MAF > 0.01 and missing data per site < 90%. Accessions with too much missing data were removed. Depth, missingness and heterozygosity were calculated using VCFtools V.0.1.12 [44]. Genotyping and SNP calling was performed at the Genomic Diversity Facility, Cornell University (http://www.bio-tech.cornell.edu/brc/brc/services).
Variant annotation
Variants were annotated to regions of the *B. napus* reference genome using R, implemented using “VariantAnnotation” [45], and Variant Effect Predictor (VEP, [46]), and variant locations were characterised as coding, intron, splice site, promoter and intergenic regions.

Genetic diversity and population structure
Population-based genetic diversity, including allele frequencies, MAF, and average heterozygosity, were computed using TASSEL 5.2.18 software [47]. Polymorphic information content (PIC) values [48] was calculated for SNP markers using the formula (PIC = 1- (maf^2+(1-maf)^2)-(2maf^2(1-maf)^2)). The ratio of transitions to transversions was calculated using the [49] 2-parameter model, implemented in MEGA7 [50].

Structure analysis of the accessions was conducted using STRUCTURE software v2.2 [51]. A subset of 10,094 SNPs was selected that was evenly distributed across the genome with one SNP per 100 Kb. The admixture model and correlated allele frequency were applied with a burn-in period of 50,000 iterations and 100,000 replications of Markov Chain Monte Carlo (MCMC). Five runs were performed to calculate the mean likelihood for the number of populations K, ranging from 1 to 10, and the mean of the log-likelihood estimates LnP(D) for each K. The ad-hoc statistic ΔK was used to determine optimal number groups [52]. Structure output was visualized using STRUCTURE HARVESTER web-based software [53]).

Analysis of molecular variance
Analysis of molecular variance (AMOVA) was conducted using Arlequin v.3.5 software [54] to estimate the genetic variance among clusters and sub-clusters of the A and C genome haplotypes. In this analysis, the distance matrix among samples was computed to estimate the genetic structure of the haplotypes. Genetic variance components were estimated, and the total variance was partitioned among major clusters, among sub-clusters within major clusters, and within subclusters. The significance of the variance components was tested using 1,000 permutations. The fixation index (Fst), an estimation of population differentiation and genetic distance based on genetic polymorphism data, was calculated.

Linkage disequilibrium (LD) analysis
LD decay across the *B. napus* genome was measured and a correlation matrix of $r^2$ values was computed between all pairs of polymorphic SNPs with MAF $\geq$ 5% using the GAPIT V2 package [55].

**Association analysis**

Data for the disease DSI were transformed using rank-based inverse normal transformation implemented as the `rntransform` function in the GenABEL R [56]. Association was analyzed for a subset of 10,094 SNP markers with MAF $\geq$ 5% using the following models: general linear model (GLM), mixed linear model (MLM), compressed mixed linear model (CMLM), enriched compressed mixed linear model (ECMLM), and multi-locus mixed model (MLMM) implemented in the GAPIT V2 package of R [55]. A kinship matrix of the accessions was calculated and principle components analysis was used to account for population structure and accessions relatedness.

**Candidate resistance genes**

Using Blast2Go software [57], the sequence region neighboring (2 Mb upstream and downstream) of the significant SNPs were searched for candidate genes encoding disease resistance proteins potentially responsible for resistance to each pathotype of *P. brassicae*.

**List Of Abbreviations**

AMOVA: analysis of molecular variance; CC: coiled-coil; CCD: Canadian clubroot differential; CMLM: compressed mixed linear model; CNL: CC-NBS-LRR; CR: clubroot resistance; DSI: disease severity index; ECMLM: enriched compressed mixed linear model; GLM: general linear model; GWAS: genome wide association analysis; GBS: genotyping by sequencing; LD: linkage disequilibrium; LRR: leucine-rich repeat; MAF: minor allele frequency; MCMC: markov chain monte carlo; MLM: mixed linear model; MMLM: multilocus mixed linear model; NBS: nucleotide-binding site; nTNL: non-TIR-NBS-LRR; PGRC: plant genetic resource of Canada; PIC: Polymorphic information content; Q-Q: quantile-quantile; QTL: quantitative trait locus; SNP: single nucleotide polymorphism; TIR: Toll-interleukin–1 receptor; TNL: TIR-NBS-LRR; USDA: United States department of Agriculture; UTR: untranslated region.

**Declarations**

**Acknowledgments**

The authors are grateful to Melissa Kehler, Victor Manolii, Md Mizanur Rahaman and the summer students Yasmina Bekkaoui and Kurtis Flavel for their technical support.

**Ethics approval and consent to participate**
Not applicable

Consent to publish
Not applicable

Availability of data and materials
The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interest.

Funding
This work was funded by a competitive grant from SaskCanola under Canola Agronomic Research Program. The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Author contributions
FY and AD conceived of and designed the study; AD and JW conducted the experiments; AD, ML, MMK and QC analyzed data; SES, SFH, BDG and GP provided important resources and facilities. AD drafted the manuscript. All authors reviewed the manuscript and approved the final draft.

Reference
1. Nagaharu U. Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. Japanese Journal of Botany. 1935; 7: 389-452.

2. Allender CJ, King GJ. Origins of the amphiploid species Brassica napus L. investigated by chloroplast and nuclear molecular markers. BMC Plant Biol. 2010; 10:54. doi:10.1186/1471-2229-10-54.

3. Howell EC, Kearsey MJ, Jones GH, King GJ, Armstrong SJ. A and C genome distinction and chromosome identification in Brassica napus by sequential fluorescence in situ hybridization and genomic in situ hybridization. Genetics. 2008;180(4):1849-1857. doi:10.1534/genetics.108.095893

4. Parkin, IAP, and Lydiate, DJ. Conserved patterns of chromosome pairing and
recombination in Brassica napus crosses. Genome. 1997; 40: 496–504.

5. McNaughton IH. Swedes and rapes—Brassica napus (Cruciferae). In: Smartt J, Simmonds NW (eds) Evolution of crop plants, 2nd edn. Longman Scientific & Technical, London, UK. 1995; 68–75.

6. Dixon, G. R. The occurrence and economic impact of Plasmodiophora brassicae and clubroot disease. J. Plant Growth Regul. 2009; 28:194–202.

7. Karling JS. The Plasmodiophorales 2nd ed. Hafner Publishing Company, Inc., New York (1968).

8. Voorrips, RE. Plasmodiophora brassicae: Aspects of pathogenesis and resistance in Brassica oleracea. Euphytica. 1995; 83:139–146 (1995).

9. Diederichsen E, Frauen M, Linders E, Hatakeyama K, Hirai M. Status and perspectives of clubroot resistance breeding in crucifer crops. J. Plant Growth Regul. 2009; 28:265–281.

10. Buczacki ST, Toxopeus H, Mattusch P, Johnston TD, Dixon GR, Hobolth LA. Study of physiologic specialization in Plasmodiophora brassicae: proposals for attempted rationalization through an international approach. Transaction of British Mycological Society. 1975; 65:295–303.

11. Piao Z, Ramchiary N, Lim YP. Genetics of clubroot resistance in Brassica species. J. Plant Growth Regul. 2009; 28: 252–264.

12. Rahman H, Shakir A, Hasan MJ. Breeding for clubroot resistant spring canola (Brassica napus L.) for the Canadian prairies: Can the European winter canola cv. Mendel be used as a source of resistance? Can. J. Plant Sci. 2011; 91: 447–458.

13. Li L, Luo Y, Chen B, et al. A Genome-Wide Association Study Reveals New Loci for Resistance to Clubroot Disease in Brassica napus. Front Plant Sci. 2016; 7:1483. doi:10.3389/fpls.2016.01483
14. Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, et al. Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. 2005; 44: 1054-1064

15. Gupta PK, Rustgi S, Kulwal PL. Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol. Biol. 2005; 57: 461-485.

16. Elshire RJ, Glaubitz JC, Sun Q, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011; 6:e19379. doi:10.1371/journal.pone.0019379.

17. Crossa J, Burgueno J, Dreisigacker S, et al. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics. 2007; 177:1889-1913.

18. Ranc N, Muños S, Xu J, et al. Genome-wide association mapping in tomato (Solanum lycopersicum) is possible using genome admixture of Solanum lycopersicum var. cerasiforme. G3. 2012; 2: 853-864. doi:10.1534/g3.112.002667.

19. Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, et al. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. 2011; 43: 163-168.

20. Wang N, Chen B, Xu K, et al. Association Mapping of Flowering Time QTLs and Insight into Their Contributions to Rapeseed Growth Habits. Front Plant Sci. 2016;7:338. doi:10.3389/fpls.2016.00338.

21. Manzanares-Dauleux MJ, Delourme R, Baron F, Thomas G. Mapping of one major gene and of QTLs involved in resistance to clubroot in Brassica napus. Theor. Appl. Genet. 2000; 101: 885-891. 10.1007/s001220051557

22. Werner S, Diederichsen E, Frauen M, Schondelmaier J, Jung C. Genetic mapping of
clubroot resistance genes in oilseed rape. Theor. Appl. Genet. 2008; 116: 363–372. 10.1007/s00122-007-0674-2.

23. Leister D. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends In Genetics: TIG. 2004; 20: 116–122. doi:10.1016/j.tig.2004.01.007.

24. Lozano R, Ponce O, Ramirez M, Mostajo N, Orjeda G. Genome-wide identification and mapping of NBS-encoding resistance genes in Solanum tuberosum Group Phureja. PLoS One. 2012; 7: e34775. doi:10.1371/journal.pone.0034775

25. Mun J, Yu H, Park S, Park B. Genome-wide identification of NBS encoding resistance genes in Brassica rapa. Molecular Genetics and Genomics. 2009; 282: 617–631. doi:10.1007/s00438-009-0492-0.

26. Yu J, Tehrim S, Zhang F, Tong C, Huang J, Cheng X, Dong C, Zhou Y, Qin R, Hua W, Liu S. Genome-wide comparative analysis of NBS encoding genes between Brassica species and Arabidopsis thaliana. BMC Genomics. 2014; 15: 3. doi:10.1186/1471-2164-15-3.

27. Zhang YM, Shao ZQ, Wang Q, Hang YY, Xue JY, Wang B, Chen JQ. Uncovering the dynamic evolution of nucleotide-binding site-leucinerich repeat (NBS-LRR) genes in Brassicaceae. Journal of Integrative Plant Biology. 2016; 58:165–177. doi:10.1111/jipb.12365.

28. Dakouri A, Zhang X, Peng G et al. Analysis of genome-wide variants through bulked segregant RNA sequencing reveals a major gene for resistance to Plasmodiophora brassicae in Brassica oleracea. Scientific Reports. 2018; 8:17657.

29. Alamery S, Tirnaz S, Bayer P, et al. Genome-wide identification and comparative analysis of NBS-LRR resistance genes in Brassica napus. Crop Pasture Sci. 2017; 69: 79–94. doi: 10.1071/CP17214.
30. Meyers B, Kozik A, Griego A, Kuang H, Michelmore R. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. The Plant Cell. 2003; 15: 809–834. doi:10.1105/tpc.009308.

31. Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Molecular Genetics and Genomics. 2004; 271: 402–415. doi:10.1007/s00438004–0990-z.

32. Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers B, Boerjan W, Martin F. Genome-wide identification of NBS resistance genes in Populus trichocarpa. Plant Molecular Biology. 2008; 66: 619–636. doi:10.1007/s11103–008–9293–9.

33. Yang S, Zhang X, Yue J-X, Tian D, Chen J-Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. Molecular Genetics and Genomics. 2008; 280: 187–198. doi:10.1007/s00438008–0355–0.

34. Ameline-Torregrosa C, Wang B-B, O’Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant Medicago truncatula. Plant Physiology. 2008; 146: 5–21. doi:10.1104/pp.107.104588.

35. Jupe F, Pritchard L, Etherington G, MacKenzie K, Cock P, Wright F, Sharma SK, Bolser D, Bryan G, Jones J, Hein I. Identification and localisation of the NB-LRR gene family within the potato genome. BMC Genomics. 2012; 13: 75. doi:10.1186/1471–2164–13–75.

36. Bonardi V, Tang S, Stallmann A, Roberts M, Cherkis K, Dangl JL. Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. Proceedings of the National Academy of Sciences of the United
37. Collier SM, Hamel L-P, Moffett P. Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. Molecular Plant-Microbe Interactions. 2011; 24: 918-931. doi:10.1094/MPMI-03-11-0050.

38. Shao Z-Q, Zhang Y-M, Hang Y-Y, Xue J-Y, Zhou G-C, Wu P, Wu X-Y, Wu X-Z, Wang Q, Wang B. Long-term evolution of nucleotide binding site-leucine-rich repeat genes: understanding gained from and beyond the legume family. Plant Physiology. 2014; 166: 217-234. doi:10.1104/pp.114.243626.

39. Strelkov, S. E., Hwang, S. F., Manolii, V. P., Cao, T., Fredua-Agyeman, R., Harding, M. W., Peng, G., Gossen, B. D., McDonald, M. R., and Feindel, D. Virulence and pathotype classification of Plasmodiophora brassicae populations collected from clubroot resistant canola (Brassica napus) in Canada. Can. J. Plant Pathol. 20185; 40: 284-298. DOI: 10.1080/07060661.2018.1459851.

40. Strelkov SE, Tewari JP, Smith-Degenhardt E. Characterization of Plasmodiophora brassicaepopulations from Alberta, Canada. Canadian Journal of Plant Pathology. 2006; 28: 467-474.

41. Kuginuki Y, Yoshikawa H, Hirai M. Eur. J. Plant Pathol. 1999; 105:327-332.

42. Horieuchi S, and Hori M. Bull. Chugoku Natl. Agric. Exp. Stn. Ser. E. (Environ. Div.), 1980; 17:33-55.

43. Glaubitz JC., Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS ONE. 2014; 9:E90346. doi:10.1371/journal.pone.0090346

44. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. Bioinformatics. 2011; 27:2156-2158. doi:10.1093/bioinformatics/btr330

45. Obenchain V, Lawrence M, Carey V, Gogarten S, Shannon P and Morgan M.
“VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants.” Bioinformatics. 2014; 30: 2076-2078. doi: 10.1093/bioinformatics/btu168.

46. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. Genome Biology 2016; 17:122. doi:10.1186/s13059-016-0974-4.

47. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics. 2007; 23: 2633–2635. doi:10.1093/bioinformatics/btm308. PMID:17586829.

48. Roussel V, Koenig J, Beckert M, Balfourier F. Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. Theor Appl Genet. 2004; 108:920–930.

49. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution. 1980; 16:111-120.

50. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016; 33:1870–1874.

51. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155: 945-959.

52. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 2005; 4: 2611-2620. doi:10.1111/j.1365-294X.2005.02553.

53. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Consr Genet Resour.
Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010; 10: 564-567.

Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, et al. GAPIT Version 2: An enhanced integrated tool for genomic association and prediction. Plant Genome. 2018; 9: 2–9. doi: 10.3835/plantgenome2015.11.0120.

Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM.. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007; 23: 1294–1296.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids research. 2008; 36: 3420–3435. doi:10.1093/nar/gkn176.

Tables

Table 1 Genome wide distribution of SNPs, minor allele frequency (MAF), Heterozygosity, polymorphic information content (PIC) and average Linkage disequilibrium (LD)

| Chromosome | Start | End       | Total No. seq | SNP   | SNP/Kb |
|------------|-------|-----------|---------------|-------|--------|
| A01        | 2024  | 23251220  | 23250         | 13062 | 1.78   |
| A02        | 919   | 24785167  | 24784         | 12455 | 1.99   |
| A03        | 808   | 29746073  | 29745         | 20541 | 1.45   |
| A04        | 1717  | 19141470  | 19140         | 10562 | 1.81   |
| A05        | 2697  | 23052978  | 23050         | 14917 | 1.55   |
| A06        | 2120  | 24372251  | 24370         | 14696 | 1.66   |
| A07        | 10938 | 24000655  | 23990         | 14232 | 1.69   |
|    | Start  | End       | Total No. seq | SNP  | SNP/Kb |
|----|--------|-----------|---------------|------|--------|
| A08| 1729   | 18958296  | 18957         | 10281|        |
| A09| 1327   | 33857792  | 33857         | 18702|        |
| A10| 4083   | 17366872  | 17363         | 12131|        |
| Average (A-subgenome) | | | 23851 | 14158 |        |
| C01| 8039   | 38812658  | 38805         | 17087|        |
| C02| 1607   | 46186975  | 46185         | 17662|        |
| C03| 760    | 60565276  | 60565         | 25136|        |
| Chromosome | Start  | End       | Total No. seq | SNP  | SNP/Kb |
| C04| 1773   | 48929072  | 48927         | 19053|        |
| C05| 3386   | 43172068  | 43169         | 16540|        |
| C06| 1745   | 37224854  | 37223         | 14761|        |
| C07| 7046   | 44766293  | 44760         | 15558|        |
| C08| 6385   | 38472912  | 38467         | 16082|        |
| C09| 1884   | 48501448  | 48500         | 18295|        |
| Average (C-subgenome) | | | 45178 | 17797 |        |
| Average (genome) | | | 34514 | 15978 |        |
Table 2 List of significant SNPs, chromosomes, physical location and P values
| Pathotype | SNP locus       | Chromosome | Position | P.value  | -log(\text{P value}) |
|-----------|----------------|------------|----------|----------|----------------------|
| 5X        | A01_19406286   | A01        | 19406286 | 2.49E-07 | ε                    |
|           | A08_171171159  | A08        | 2719211  | 5.39E-06 | ε                    |
|           | A09_195497763  | A09        | 8083774  | 1.11E-05 | 4                    |
|           | C08_579178095  | C08        | 20763231 | 4.88E-05 | 4                    |
| 2B        | A07_147954943  | A07        | 3509616  | 9.95E-10 | ε                    |
|           | A04_83864566   | A04        | 6035183  | 1.15E-07 | ε                    |
|           | A03_651104541  | A03        | 1393088  | 5.23E-07 | ε                    |
|           | A08_186203638  | A08        | 17751690 | 3.94E-06 | ε                    |
|           | C01_276968702  | C01        | 38290946 | 5.56E-06 | ε                    |
|           | C03_688995474  | C03        | 6293335  | 7.75E-06 | ε                    |
|           | A05_115772286  | A05        | 18791143 | 9.44E-06 | ε                    |
|           | A09_215839211  | A09        | 28425222 | 2.18E-05 | 4                    |
|           | A03_68863700   | A03        | 20801907 | 3.63E-05 | 4                    |
| 3A        | C01_341096049  | C01        | 17366972 | 1.95E-08 | 7                    |
|           | A10_222415846  | A10        | 1136417  | 5.56E-08 | 7                    |
|           | C04_403341747  | C04        | 19039176 | 3.50E-07 | €                    |
|           | C09_608174205  | C09        | 11282154 | 6.73E-07 | €                    |
|           | C09_634081519  | C09        | 37189468 | 2.21E-05 | 4                    |
| 3D        | A04_90743077   | A04        | 12913694 | 7.21E-07 | €                    |
|           | C09_638459286  | C09        | 41567235 | 2.76E-06 | 5                    |
|           | A03_71057307   | A03        | 22995514 | 3.13E-06 | 5                    |
|           | A01_17862282   | A01        | 17862282 | 3.31E-06 | 5                    |
|           | C03_685453245  | C03        | 2751106  | 4.46E-05 | 4                    |
| Pathotype | SNP               | Chromosome | Gene ID          | Candidate gene region (Mb) |
|-----------|-------------------|------------|-----------------|---------------------------|
| 5X        | A01_19406286      | A01        | BnaA01g28560D   | 19-20                     |
| 2B        | A03_651104541     | A03        | BnaA03g03110D   | 1-2                       |
| 2B        |                   | A03        | BnaA03g03260D   | 1-2                       |
| 2B        |                   | A03        | BnaA03g03270D   | 1-2                       |
| 2B        | A03_651104541     | A03        | BnaA03g43880D   | 20.5-23.5                 |
| 2B        |                   | A03        | BnaA03g44070D   | 20.5-23.5                 |
| 2B        |                   | A03        | BnaA03g44080D   | 20.5-23.5                 |
| 2B        |                   | A03        | BnaA03g44090D   | 20.5-23.5                 |
| 2B        |                   | A03        | BnaA03g45000D   | 20.5-23.5                 |
| 2B        |                   | A03        | BnaA03g45010D   | 20.5-23.5                 |
| 3D        | A03_68863700      | A03        | BnaA03g45020D   | 20.5-23.5                 |
| 3D        |                   | A03        | BnaA03g45040D   | 20.5-23.5                 |
| 3D        |                   | A03        | BnaA03g45050D   | 20.5-23.5                 |
| 3D        |                   | A03        | BnaA03g45970D   | 20.5-23.5                 |
| 2B        | A04_83864566      | A04        | BnaA04g06520D   | 5.5-6.5                   |
| 2B        |                   | A04        | BnaA04g06530D   | 5.5-6.5                   |
| 2B        |                   | A04        | BnaA04g06550D   | 5.5-6.5                   |
| 2B        |                   | A04        | BnaA04g06580D   | 5.5-6.5                   |
| 2B        |                   | A04        | BnaA04g06780D   | 5.5-6.5                   |
| 2B        | A05_115772286     | A05        | BnaA05g24990D   | 18-19                     |
| 2B        |                   | A05        | BnaA05g25000D   | 18-19                     |
| 5X        | A08_171171159     | A08        | BnaA08g02210D   | 1-3                       |
| 2B        | A08_186203638     | A08        | BnaA08g24820D   | 17-18                     |
| 2B        |                   | A08        | BnaA08g24860D   | 17-18                     |
| 2B        |                   | A08        | BnaA08g26370D   | 17-18                     |
| 2B        |                   | A08        | BnaA08g26380D   | 17-18                     |
| 2B        |                   | A08        | BnaA08g26400D   | 17-18                     |
| 2B        | A08_186203638     | A08        | BnaA08g26410D   | 17-18                     |
|           |                   | A09        | BnaA09g13280D   | 7.9                       |
| 5X                          | A09_195497763   | A09  | BnaA09g13850D | 7-9   |
|-----------------------------|----------------|------|---------------|-------|
|                             |                | A09  | BnaA09g13890D | 7-9   |
|                             |                | A09  | BnaA09g13900D | 7-9   |
|                             |                | A09  | BnaA09g14320D | 7-9   |
|                             |                | A09  | BnaA09g14420D | 7-9   |
|                             |                | A09  | BnaA09g14550D | 7-9   |
|                             |                | A09  | BnaA09g14560D | 7-9   |
|                             |                | A09  | BnaA09g14570D | 7-9   |
|                             |                | A09  | BnaA09g14680D | 7-9   |
| 2B                          | A09_215839211   | A09  | BnaA09g42680D | 27.1-30.1 |
|                             |                | A09  | BnaA09g43420D | 27.1-30.1 |
|                             |                | A09  | BnaA09g43430D | 27.1-30.1 |
|                             |                | A09  | BnaA09g39390D | 27.1-30.1 |
|                             |                | A09  | BnaA09g39400D | 27.1-30.1 |
| 3A                          | A10_222415846   | A10  | BnaA10g04000D | 1-3   |
|                             |                | A10  | BnaA10g05000D | 1-3   |
| 3A                          | C01_276968702   | C01  | BnaC01g39050D | 37-39 |
|                             |                | C01  | BnaC01g40270D | 37-39 |
|                             |                | C01  | BnaC01g40280D | 37-39 |
|                             |                | C01  | BnaC01g40300D | 37-39 |
|                             |                | C01  | BnaC01g40310D | 37-39 |
|                             |                | C01  | BnaC01g40460D | 37-39 |
| 2B                          | C03_685453245   | C03  | BnaC03g04690D | 2-3   |
|                             |                | C03  | BnaC03g05380D | 2-3   |
| 3D                          | C03_688995474   | C03  | BnaC03g17450D | 20.7-21.1 |
| 3A                          | C04_403341747   | C04  | BnaC04g18730D | 17-19 |
| 5X                          | C08_579178095   | C08  | BnaC08g17450D | 20.7-21.1 |
|                             |                | C09  | BnaC09g14400D | 11-12 |
|                             |                | C09  | BnaC09g144870D | 11-12 |
|                             |                | C09  | BnaC09g15010D | 11-12 |
|                             |                | C09  | BnaC09g15020D | 11-12 |
|                             |                | C09  | BnaC09g15110D | 11-12 |
| 3D                          | A09_638459286   | C09  | BnaC09g38250D | 41-42 |

Figures
Figure 1

Frequency distribution of accessions plotted against clubroot severity (disease severity index, DSI) for four pathotypes 5X, 2B, 3A, and 3D indicated in the figure.
Figure 2

Population structure analysis of the 177 accessions based on A. model-based Bayesian clustering using STRUCTURE for $K = 2$ groups, and B. estimation of the number of sub-populations for $K$ values of 1 to 10.
Figure 3

Manhattan plots of association analysis using the multilocus mixed linear model (MMLM) model P+K for pathotypes A. 5X, B. 2B, C. 3A and D. 3D. The horizontal line represents the threshold of significance (-log100.5/10094 = 4.30).

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

  eq1.jpg
  Additional files-final version.docx