Local Duplication of TIR-NBS-LRR Gene Marks Clubroot Resistance in *Brassica napus* cv. Tosca

Piotr M. Kopec[1], Katarzyna Mikolajczyk[2], Ewa Jajor[3], Agnieszka Perek[3], Joanna Nowakowska[2], Christian Obermeier[4], Harmeet Singh Chawla[4,5], Marek Korbas[3], Iwona Bartkowiak-Broda[2] and Wojciech M. Karlowski[1]*

[1] Department of Computational Biology, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University Poznan, Poznan, Poland, [2] Department of Genetics and Breeding of Oilseed Crops, Plant Breeding and Acclimatization Institute-National Research Institute, Poznan, Poland, [3] Institute of Plant Protection - National Research Institute, Poznan, Poland, [4] Department of Plant Breeding, Justus-Liebig-Universitaet Giessen, Giessen, Germany, [5] Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

Clubroot, caused by *Plasmodiophora brassicae* infection, is a disease of growing importance in cruciferous crops, including oilseed rape (*Brassica napus*). The affected plants exhibit prominent galling of the roots that impairs their capacity for water and nutrient uptake, which leads to growth retardation, wilting, premature ripening, or death. Due to the scarcity of effective means of protection against the pathogen, breeding of resistant varieties remains a crucial component of disease control measures. The key aspect of the breeding process is the identification of genetic factors associated with variable response to the pathogen exposure. Although numerous clubroot resistance loci have been described in *Brassica* crops, continuous updates on the sources of resistance are necessary. Many of the resistance genes are pathotype-specific, moreover, resistance breakdowns have been reported. In this study, we characterize the clubroot resistance locus in the winter oilseed rape cultivar "Tosca." In a series of greenhouse experiments, we evaluate the disease severity of *P. brassicae*-challenged "Tosca"-derived population of doubled haploids, which we genotype with Brassica 60 K array and a selection of SSR/SCAR markers. We then construct a genetic map and narrow down the resistance locus to the 0.4 cM fragment on the A03 chromosome, corresponding to the region previously described as *Crr3*. Using Oxford Nanopore long-read genome resequencing and RNA-seq we review the composition of the locus and describe a duplication of TIR-NBS-LRR gene. Further, we explore the transcriptomic differences of the local genes between the clubroot resistant and susceptible, inoculated and control DH lines. We conclude that the duplicated TNL gene is a promising candidate for the resistance factor. This study provides valuable resources for clubroot resistance breeding programs and lays a foundation for further functional studies on clubroot resistance.

Keywords: *Brassica napus*, *Plasmodiophora brassicae*, Oxford Nanopore, TNL, RNA-Seq, QTL, resistance, duplication

INTRODUCTION

Plasmodiophora brassicae Wor., an obligate, soil-borne parasite of crucifers (Brassicaceae), is an agent responsible for clubroot disease. During the two-stage infection (Kageyama and Asano, 2009; Liu et al., 2020), the pathogen hijacks multiple nodes of host metabolism and induces hyperplasia and hypertrophy of the underground organs leading to a prominent galling. The galls act as a major physiological sink that supports the proliferation and development of the pathogen while reducing the fitness of the host (Malinowski et al., 2019). Deformations of the root system impair the plant's capacity for water and nutrient uptake, leading to growth retardation, wilting, and premature, non-optimal flowering (Korbas et al., 2009). Many important crops cultivated worldwide, including oilseed rape (*Brassica napus*), belong to the Brassicaceae (Dixon, 2007). Clubroot disease has been becoming a global problem of increasing economic impact in cruciferous crops and has been ranked under the top 10 most significant worldwide threats to oilseed rape production (Dixon, 2009; Zheng et al., 2020). An infection of oilseed rape was shown to cause up to 60% loss of yield at relatively low spore densities, and total yield failure at a higher pathogen pressure (Strehlow et al., 2015). Once introduced, *P. brassicae* is hard to eradicate. Resting spores can live in the soil for up to 20 years (Wallenhammar, 1996), and spread easily via, for example, dirt on farm equipment (Cao et al., 2009). Many protective measures against the pathogen, e.g., crop rotation or chemical control, are of limited efficiency (Hwang et al., 2014). Therefore, the breeding of resistant plant varieties remains a crucial component of clubroot control efforts.

The key aspect of the breeding process is the identification of genetic features associated with plant response to pathogen exposure. Numerous clubroot resistance loci were described in *Brassica* crops (Landry et al., 1992; Figdore et al., 1993; Voorrips and Visser, 1993; Grandclément and Thomas, 1996; Voorrips et al., 1997; Matsumoto et al., 1998; Moriguchi et al., 1999; Suwabe et al., 2003, 2006; Hirai et al., 2004; Laurens and Thomas, 2004; Piao et al., 2004; Rocherieux et al., 2004; Nomura et al., 2005; Saito et al., 2006; Sakamoto et al., 2008; Nagaoka et al., 2010; Kato et al., 2012; Chen et al., 2013; Chu et al., 2013, 2014; Hatakeyama et al., 2013, 2017; Pang et al., 2014, 2018; Zhang et al., 2014; Fredua-Agyeman and Rahman, 2016; Lee et al., 2016; Huang et al., 2017; Yu et al., 2017; Dakouri et al., 2018; Hirani et al., 2018; Nguyen et al., 2018; Peng et al., 2018; Laila et al., 2019; Choi et al., 2020; Farid et al., 2020; Karim et al., 2020) and are reviewed in (Neik et al., 2017; Lv et al., 2020).

Several genetic studies were performed in *B. napus*. Manzanares-Dauleux et al. (2000) described a major resistance gene *Pb-Bn1* on the A04 chromosome and two quantitative loci on A04 and C05 chromosomes. The resistance derived from the DH ECD-04 line (selected from *Brassica rapa* subsp. rapifera), which was utilized in many breeding programs for the development of clubroot-resistant cultivars, including winter oilseed rape "Mendel," was mapped to the *CRa/CRb* region on the A03 chromosome (Diederichsen and Sacristan, 1996; Diederichsen et al., 2006; Fredua-Agyeman and Rahman, 2016; Zhang et al., 2016). Werner et al. (2008) mapped 19 QTLs spread

across 8 chromosomes. In addition, a couple of association studies were conducted on the *B. napus/P. brassicae* pathogenic model. Li et al. (2016) identified 9 loci, 7 of which were not described previously. Hejna et al. (2019) identified 2 major and 7 minor loci, with the most prominent peak overlapping the *CRa* region. Fredua-Agyeman et al. (2020) identified three genomic hotspots corresponding to *Crr3/CRk/CRd* and *CRa/CRb/CRb^Kato* regions on A03 and *Crr1* region on A08 in a GWAS study of 124 rutabaga accessions from Nordic countries.

Additionally, two resistance genes were cloned thus far: *CRa* (Ueno et al., 2012) and *Crr1* (Hatakeyama et al., 2013). Both genes belong to the TIR-NBS-LRR (TNL; Toll/interleukin-1 receptor-like – nucleotide-binding site – leucine-rich repeat) protein domain family, reported as a key component of effector-triggered immunity (DeYoung and Innes, 2006; McHale et al., 2006).

Despite a seemingly ample collection of resistance loci, continuous updates on the sources of resistance are necessary. *P. brassicae* shows pathogenic specialization, and the host's resistance genes often confer immunity to only subsets of pathotypes. Moreover, the breakdown of clubroot resistance in the case of some *P. brassicae* pathotypes has been repeatedly reported (Diederichsen et al., 2014; Strelkov et al., 2016).

In this study, we map the resistance locus of the Swedish resynthesis-derived winter-type oilseed rape cultivar "Tosca" (Happstadius et al., 2003; Diederichsen et al., 2009) to a small region on the A03 chromosome. Using the long-read Oxford Nanopore (ON) sequencing technology, we review the genomic structure of the locus in "Tosca" as well as in susceptible "BRH-1" breeding line. In addition, we perform an RNA-seq experiment to identify infection-induced differentially expressed genes. These data are subsequently linked to the genic composition of the resistance locus. Based on the results, we attribute the "Tosca" resistance phenotype to a locus constitutively expressing a duplicated TNL gene, located within the *Crr3* (Hirai et al., 2004) region, directly upstream of the region homologous to the *CRd* (Pang et al., 2018). This study provides valuable resources for clubroot-resistant rapeseed breeding programs and lays a foundation for further functional studies on clubroot resistance.

MATERIALS AND METHODS

Plant Material

A doubled haploid (DH) segregating population of 250 DH lines was developed by Plant Breeding Strzelce Ltd. (IHAR-PIB Group; division in Borowo) from a cross of a winter oilseed rape (*B. napus*) clubroot resistant cultivar "Tosca" and a susceptible BRH-1 breeding line, using isolated microspore culture technique as described in (Cegielska-Taras et al., 2002; Szała et al., 2020).

Pathogen Source, Preparation, and Plant Inoculation

Samples of *B. napus* root galls induced by *P. brassicae* were collected from infested oilseed rape fields in Lower Silesian Province in Poland. The inoculum for the greenhouse

experiments was prepared by isolating resting spores from the galls. The galls were blended, and the homogenate was filtered through a layer of gauze and centrifuged for 5 min at 3,500 rpm to obtain a clear suspension. Spore density was measured using a 0.1 mm deep, improved Neubauer counting chamber (Marienfeld-Superior) and a bright field microscope (Olympus BX 50). The density was adjusted to 1×10^8 spores/ml. For inoculation, each experimental pot containing five 1-week-old seedlings was watered with the spore suspension. The same batch of inoculum was used in all experiments. Additionally, to assess the *P. brassicae* pathotype, galls from 25 DH lines were collected and individually processed into a set of spore suspensions. *P. brassicae* pathotype of every suspension was classified using the Somé system (Some et al., 1996).

Experimental Design and Conditions

The greenhouse experiments were performed between April 2018 and August 2019 in the Research Centre of Quarantine, Invasive and Genetically Modified Organisms – Institute of Plant Protection National Research Institute. The experiment followed the principle of augmented design. The plants were grown in a series of 6 temporally successive blocks (batches). Each of the batches included around 60 test DH lines, augmented with 8 reference lines ("checks") – 6 phenotypically extreme DH genotypes that were selected from the first experiment and parental lines.

For every line, 15 plants were grown in 3 pots: 2 pots for treated (inoculated) and 1 pot for untreated control, 5 plants each. Pots were randomly distributed in 4 trays for treated and 2 trays for untreated control plants. Separate, fixed trays were used for treated and untreated plants to avoid water or soil-borne contamination. The soil pH value was 6.0. The temperature ($\pm 0.5°C$) was set to 18°C/16°C day/night regime for the first 2 weeks of cultivation, and then elevated to 20°C/18°C. The photoperiod was set to a 14 h/10 h light/darkness scheme. The air humidity ($\pm 3\%$) was 60%. Soil humidity was kept in the range between 60 and 70%.

Despite the controlled experimental conditions, we observed a significant batch effect – seasonal phenotypic variability among the analyzed DH lines. Therefore, an additional experiment was carried out including more lines in one, common batch (242, including the checks) at the expense of the number of tested plants per line (5 instead of 10). Additionally, to promote the infection by *P. brassicae*, the temperature was elevated to 20°C for the first 2 weeks and 24°C/20°C for the next five.

Phenotyping and Phenotypic Data Analysis

After 7 weeks of growth (42 days after inoculation), the plants were phenotyped for classical underground morphological symptoms of clubroot disease. Each plant was removed from the ground and washed with water. The degree of infection (DOI) was evaluated on a 4-degree scale (Vigier et al., 1989), where 0 indicates a healthy root system, 1 refers to 1–10% of root system altered (small galls on lateral roots), 2 denotes 11–50% root system altered, and 3 describes 51–100% root system altered. The

disease index (DI) for each genotype by batch was then calculated by obtaining the arithmetic mean of the DOI and rescaling it to the percent scale.

To obtain the DI over the entire experiment, adjusted for the batch effect (phenotypic variability between the greenhouse runs), the DOI data were fit into a linear mixed model using the lme4 library (Bates et al., 2015) for R (R Core Team, 2020):

$$P_{ij} = \mu + g_i + B_j + (gb)_{ij} + e_{ij} \tag{1}$$

where P_{ij} stands for the phenotype of the *i*th genotype in the *j*th batch, μ is the general mean of the experiment, g_i is the random effect of the *i*th genotype, B_j is the fixed effect of the *j*th batch, $(gb)_{ij}$ is the random effect of the interaction between the *i*th genotype and *j*th batch, and e_{ij} is the error term. Next, the conditional mode (Best Linear Unbiased Prediction; BLUP) of the genotype was obtained. The BLUP-DI values were used in a subsequent QTL mapping.

Heritability Estimation

Broad sense heritability (H^2) of the DOI was estimated after (Stahl et al., 2019) following the concept of (Piepho and Möhring, 2007), with the equation:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + SE^2} \tag{2}$$

where σ_G^2 is the genetic variance, derived from a full random model (Eq. 3) and SE^2 is the squared standard error of the difference between the means, derived from a mixed model (Eq. 4). The analysis was conducted using the R packages lmerTest (Kuznetsova et al., 2016), lsmeans (Lenth, 2016), and lme4 (Bates et al., 2015).

$$P_{ij} = \mu + g_i + b_j + (gb)_{ij} + e_{ij} \tag{3}$$

where P_{ij} stands for the phenotype of the *i*th genotype in the *j*th batch, μ is the general mean of the experiment, g_i is the random effect of the *i*th genotype, b_j is the random effect of the *j*th batch, $(gb)_{ij}$ is the random effect of the interaction between the *i*th genotype and *j*th batch, and e_{ij} is the error term.

$$P_{ij} = \mu + G_i + B_j + (gb)_{ij} + e_{ij} \tag{4}$$

where P_{ij} stands for the phenotype of the *i*th genotype in the *j*th batch, μ is the general mean of the experiment, G_i is the fixed effect of the *i*th genotype, B_j is the fixed effect of the *j*th batch, $(gb)_{ij}$ is the random effect of the interaction between the *i*th genotype and *j*th batch, and e_{ij} is the error term.

To investigate the reliability of each of the batches, their influence on the H^2 was assessed by recalculating the H^2 with a leave-one-out approach.

Genotyping

The plants were genotyped using The *Brassica* 60 K *Illumina Infinium*[TM] SNP array (Clarke et al., 2016) and a set of SSR and SCAR markers of known clubroot resistance loci (**Supplementary Table 1**). For *Brassica* 60 K genotyping, the plant

material collected from young leaves was sent to the commercial service provider TraitGenetics in Gatersleben (Germany) for DNA isolation and further processing. For SSR/SCAR analysis, the DNA was extracted from young leaves using a modified CTAB method (Doyle and Doyle, 1990). PCR amplification products were visualized on a 1.5% agarose gel (SCAR) and using the ABI PRISM 3130 OXL capillary electrophoresis (SSR).

Filtering of Genotyping Data

To check for duplicates, the lines were clustered with complete linkage based on Jaccard's distance. Lines with <0.05 distance were regarded as duplicates, and only one of them (randomly selected) was used in further analyses. Lines with more than 0.02% of heterozygous calls were discarded. Homomorphic (>95%) markers and markers with distorted segregation patterns (1:3) were also removed from further analyses. Redundant markers were binned.

Genetic Map Construction and QTL Mapping

A genetic map was constructed using the R/qtl package (Broman et al., 2003). For ordering the markers, the R/TSPmap program was used (Monroe et al., 2017). QTL Mapping was conducted with Haley-Knott regression implemented in the scan1 function of the R/qtl2 (Broman et al., 2019) package. $\log 10(p)$ significance cutoff was determined using a permutation test with $n = 1,000$.

Genome Sequencing of Parental Lines

Genomic DNA from parental lines was sequenced using ON technology. DNA from young leaves was extracted following a protocol described by Chawla et al. (2020). The libraries were prepared using the SQK-LSK109 kit, following the manufacturer's recommendations, and sequenced on R9.4.1 Flow Cells.

Genome Sequencing Data Analysis

The *B. napus* reference genomes used for the study were: Darmor-*bzh* v4.1 (Chalhoub et al., 2014), deposited on EnsemblPlants as AST_PRJEB5043_v1; Express 617 assembly v1 (Lee et al., 2020); reference pan-genome v0 (Song et al., 2020). Darmor-*bzh* genes within the mapped resistance locus were functionally classified using Pannzer2 (Törönen et al., 2018).

Base calling from ON signals was performed using Guppy, and raw reads were mapped to the reference genomes with minimap2 (Li, 2018) with -x map-ont parameters, and filtered for uniquely mapping reads with samtools (Li et al., 2009) using -q 60 option. Local SNV calling was performed using longshot (Edge and Bansal, 2019), with default parameters. SV calling was executed using sniffles (Sedlazeck et al., 2017), with –min_support 5 option. The potential effect of the variants differing parental accessions was determined with the SnpEff (Cingolani et al., 2012).

The reads overlapping the TNL gene on the Express 617 *B. napus* genome assembly were extracted from the raw sequence file, assembled using Redbean (wtdbg2; Ruan and Li, 2020), and polished once using the wtpoa-cns tool.

Transcriptome Sequencing

For transcriptomic experiments, one resistant and one susceptible DH line were selected. The roots of two biological replicates per line of infected and control plants were harvested on the day of phenotyping (7 weeks after inoculation), immediately frozen in liquid nitrogen, and stored at $-80°C$. The tissue was blended in liquid nitrogen using a mortar and pestle. The total RNA was extracted with the Qiagen Plant RNeasy kit. TruSeq mRNA strand-specific libraries were prepared and sequenced on Illumina NovaSeq600 in a 2×150 bp paired-end layout. Library preparation and sequencing were conducted by Macrogen.

Reconstruction of the TNL Genes-Encoded Transcripts Using RNA-Seq Reads

RNA-seq reads were pooled by DH line and mapped to the Express 617 genomic sequence assembly supplemented with the fragment containing the duplication as a pseudochromosome using STAR (Dobin et al., 2013) with the following parameters: –outFilterMismatchNoverLmax 0.1 –outFilterMismatchNoverReadLmax 0.1 –alignIntronMax 2000 –alignIntronMin 15 –outSAMprimaryFlag AllBestScore. Reads mapping to the pseudochromosome were then assembled using Trinity (Grabherr et al., 2011) with -genome_guided_bam –genome_guided_max_intron 2000 options. Assembled transcripts were re-mapped to the Express 617 reference sequence supplemented with a pseudochromosome with a minimap2 -x splice for validation. ORFs were predicted and translated using NCBI's ORFfinder[1]. For sequence comparison, the CDS and protein sequences were aligned with EMBL-EBI's Clustal Omega and EMBOSS Needle (Madeira et al., 2019). The sequence-based prediction of protein domains was carried out with InterProScan (Jones et al., 2014).

Analysis of Differential Gene Expression

Raw RNA-seq reads were trimmed using Trimmomatic (Bolger et al., 2014) with default options and mapped to the reference genome using STAR with the following parameters: –outFilterMismatchNoverLmax 0.1 –outFilterMismatchNoverReadLmax 0.1 –alignIntronMax 2000 –alignIntronMin 15. The fragments were counted using the featureCounts (Liao et al., 2014) program with -s 2 -p -M flags. The differential expression analysis was performed using limma (Ritchie et al., 2015; Law et al., 2016)/edgeR (Robinson et al., 2010) R packages, following the procedure described by Law et al. (2016). Raw counts were normalized via TMM, and log-CPM values were used for the DE analysis. The fit was processed with limma's treat() with lfc = 1 parameter, thus genes with fold-change significantly larger than 2 were deemed as differentially expressed. Gene Ontology enrichment analysis was conducted with g:Profiler (Raudvere et al., 2019). To assess the expression of the "Tosca" TNL copies, the reads were mapped to the reference with the fragment containing the duplication

---

[1] https://www.ncbi.nlm.nih.gov/orffinder/

attached as a pseudochromosome, and TPM values of TMM normalized counts were calculated.

Creation of Figures

All plots were generated with r/ggplot2 (Wickham, 2009). Figures were assembled with Inkscape[2].

RESULTS

Phenotyping of "Tosca" × "BRH-1" DH Population

To identify the locus harboring resistance to clubroot disease in the "Tosca" winter oilseed rape cultivar, a mapping population of 250 DH lines was developed in a cross with a susceptible, double-low line BRH-1. For phenotyping experiments, the lines were divided into 7 groups and tested separately in controlled environmental conditions. In addition to the tested DH lines, each experimental batch contained a set of referential "checks" – both parents and 6 DH lines used in all experiments (**Figure 1**, colored lines). These 6 lines were identified in the first experiment as showing contrasting phenotypes (resistant or susceptible) and no visible developmental abnormalities. In each of the experiments, 1-week-old seedlings were inoculated with a suspension of *P. brassicae* spores prepared from the pathotype-P3-dominant environmental sample. After a total of 7 weeks of growth, the plants were examined for infection-induced morphological pathologies of roots and evaluated on a 4-step severity-dependent scale. For every line, a percent scale DI was calculated.

Analysis of the first 6 batches clearly demonstrated that despite random grouping of the lines the DI distribution was not equal between the batches (Kruskal–Wallis test, $p < 10^{-6}$), with mean DI ranging from 13.13 to 46.86% (**Figure 1A** and **Table 1**). Phenotypes of the 8 checks were shifting according to the batches' trend, showing that the observed differences were independent

[2]Harrington et al. http://www.inkscape.org/

TABLE 1 | Summary statistics of the DI of the seven batches and BLUP-DI.

Batch	Min	Max	Mean	Median	Q_1	Q_3	SD	H^2_{LOO}	ΔH^2
1	0.00	90.48	30.77	36.67	0.00	55.56	28.87	0.70	0.03
2	0.00	100.00	46.84	57.143	0.00	83.33	40.02	0.65	0.08
3	0.00	50.00	13.11	6.70	0.00	26.70	15.29	0.79	−0.03
4	0.00	62.96	18.89	16.67	5.36	30.56	14.79	0.77	−0.04
5	0.00	92.59	44.23	58.33	0.00	66.67	32.29	0.73	0.00
6	0.00	100.00	38.59	40.00	0.00	73.30	36.02	0.77	−0.02
7	0.00	100.00	63.84	100.00	13.30	100.00	43.97	0.52	0.21
BLUP-DI	−6.45	77.22	37.15	46.87	11.71	56.92	24.32		

H^2_{LOO} is the H^2 without given batch; $\Delta H^2 = H^2 - H^2_{LOO}$.

of the selection of the lines in each of the subsets (**Figure 1A**). The differences are well explained by seasonal changes, with a higher incidence observed in batches carried out during spring and summer, despite that the plants were grown in controlled greenhouse conditions.

Since the main goal of this experiment was to identify DH lines with the strongest resistant phenotype, we decided to repeat the tests using growing conditions that promote *P. brassicae* infection. Therefore, in the last 7th batch (**Figure 1B**), we have included most (240 out of 250) of the DH lines and increased the temperature by 2°C (for details see section "Materials and Methods"). In all 7 batches, the DI values followed a clear bimodal distribution, suggesting that the majority of the phenotypic effect is linked to a single locus. To adjust the DI for the batch effect for QTL mapping, the phenotyping data were fit to a linear mixed model and the BLUP of the genotypic effect (BLUP-DI) was obtained (**Figure 1C**). The estimated broad-sense heritability of the disease severity for the entire experiment was $H^2 = 0.729$.

Identification of the Resistance Locus by Genetic Mapping

The mapping population was genotyped using the *Brassica* 60 k SNP array (Clarke et al., 2016) and a set of SSR and SCAR
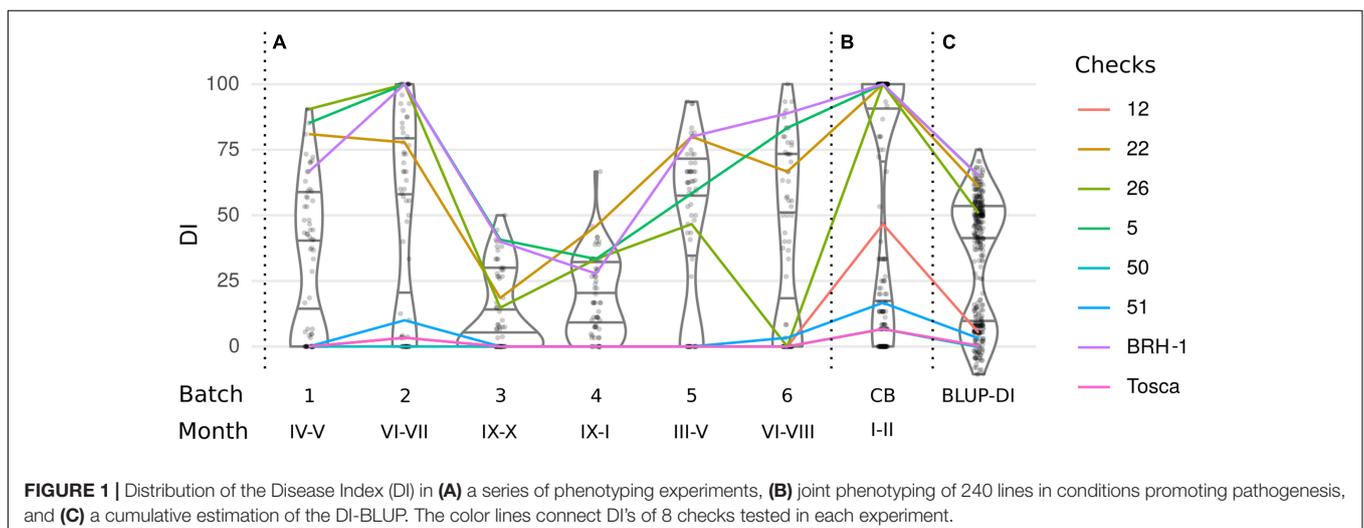
FIGURE 1 | Distribution of the Disease Index (DI) in **(A)** a series of phenotyping experiments, **(B)** joint phenotyping of 240 lines in conditions promoting pathogenesis, and **(C)** a cumulative estimation of the DI-BLUP. The color lines connect DI's of 8 checks tested in each experiment.

markers linked to various clubroot resistance *loci*. Markers showing segregation distortion or high heterozygosity were discarded from further analysis. Segregating "Failed" SNP calls were regarded as potential presence-absence variants (Gabur et al., 2018) and kept in the analysis. The constructed genetic map consisted of 1,406 bins of cosegregating markers distributed among 19 linkage groups corresponding to the 19 chromosomes of *B. napus*. The total length of the map was 1866.1 cM with an average spacing of 1.3 cM and a max spacing of 37.2 cM (**Supplementary Table 4**).

QTL mapping on BLUP-DI data revealed a single locus on the A03 chromosome (**Supplementary Table 5**). Bayes Credible Interval (BCI) for the QTL spanned 0.4 cM between 11.980 and 12.378 cM on the genetic map (**Figure 2A** and **Supplementary Table 5**). The same region, although with a larger BCI span, was detected in individual QTL mappings for every phenotyping batch (**Supplementary Table 6**). No evidence suggested the involvement of other loci affecting the trait. The locus exhibited a large effect, with a 45.65 difference between mean values of BLUP-DI for the bin of markers exhibiting the strongest correlation with the phenotype (**Figures 2B,C** and

Supplementary Table 5). Recombination events in the proximity of the QTL were identified and compared with the phenotypes. This analysis revealed that the state of a single bin of markers, cosegregating with the representative Bn-A03-p15102212 marker at 11.980 cM was sufficient to explain the phenotype (**Figure 2D**).

To physically anchor the resistance locus, either probe or primer sequences (depending on the marker type) from within the bin with the strongest correlation with the trait were aligned to the Darmor-*bzh* 4.1 reference genome. The closest markers flanking the bin spanned a region of 91,088 bp on supercontig LK031800. However, three of the markers cosegregating with the peak marker mapped to a supercontig LK033659, suggesting that the reference genome has been misassembled.

To resolve this discrepancy, we have mapped both contigs (LK031800 and LK033659) to the recently published *B. napus* reference pangenome (Song et al., 2020) and Express 617 assemblies (Lee et al., 2020). As shown in **Figure 3**, both contigs map to the same region in the newer long-read-based genomic sequences. The LK031800 (2,713,116 bp) contig maps to two distinct distant parts of the reference sequence that are separated by region corresponding to LK033659 (51,625 bp) and a small

FIGURE 2 | Results of genetic mapping. **(A)** LOD score for QTL presence along the A03 chromosome genetic map. Red lines span Bayes Credible Intervals. **(B)** Effect of "BRH-1" (yellow) and "Tosca" (green) alleles along the A03 chromosome genetic map. **(C)** Phenotype (BLUP-DI) distribution of DH lines carrying "BRH-1" yellow and "Tosca" green alleles at the peak marker. **(D)** Analysis of recombinants. The plot shows a genotype (white: "BRH-1"-inherited, gray: "Tosca"-inherited) at markers surrounding the mapped locus. The phenotype for 18 DH lines recombining in the proximity of the locus is shown as a DI from one of the 1–6 batches, and, if available, 7th, common batch, as well as BLUP-DI. A marker solely explaining the phenotype is highlighted in teal.

FIGURE 3 | Physical localization of the resistance locus (black line) in different assemblies of the *Brassica napus* genome defined by the bin of cosegregating markers at 11.98 cM. Colored bars represent Darmor-*bzh* 4.1 contigs. Gray color depicts sequences in the pangenome and Express 617 not covered by the Darmor-*bzh* 4.1 contigs.

contig LK038676 of 4,535 bp in size. For further verification of the Darmor-*bzh* being misassembled, a new set of 6 SCAR markers was designed upstream, within, and downstream of the LK033659 (Tsc, **Supplementary Table 2**). The PCR results confirmed the observed segregation pattern and were in perfect agreement with the primers' physical location. Summarizing, the locus defined by the bin of markers (with the representative marker Bn-A03-p15102212) is covered by both new reference *B. napus* assemblies as well as two contigs from Darmor-*bzh* 4.1, but in the latter case, one of the contigs (LK031800) is misassembled (**Figure 3**).

Concluding, the genetic factor of resistance to *P. brassicae* infection is located in the region covering 124,463 bp on the reference pangenome, 97,783 bp on the Express 617 assembly, and 118,111 bp on the Darmor-*bzh* 4.1 assembly. Genetic and physical evidence suggests that the mapped resistance locus falls within the region homologous to the *Crr3* locus (Hirai et al., 2004), directly upstream of the region homologous to the *CRd* (Pang et al., 2018; **Figure 4**). Accordingly, the resistance locus identified in this study is hereafter referred to as *Crr3^{Tsc}*. The *Crr3^{Tsc}* contains 25 annotated protein-coding genes. The full list with functional descriptions is included in **Supplementary Table 7**.

Structural Variation Within the Resistance Locus Between "Tosca" and "BRH-1"

To explore in detail the properties of the region covering resistance in the "Tosca" genetic background, we have sequenced

the genomes of the parental lines with ON technology. The reads mapped to the Express 617 reference genomic sequence consistently overlapped the resistance locus genomic region (97,783 bp) with an average read coverage of 16.24 for "Tosca" and 28.50 for "BRH-1," with 93.6 and 91.9% of positions covered with at least 5 reads, respectively (**Figure 5A** and **Supplementary Figure 1**).

The long-read mapping results showed differences between the "Tosca" and reference genome assembly. Interestingly, the read coverage depth of the fragment overlapping a TNL gene (BnaA03g29300D) was approximately doubled in comparison to the surrounding sequences in "Tosca," but not in "BRH-1" (**Figure 5A** and **Supplementary Figure 1**). Moreover, neither of the reads spanned the entire gene and both of its flanking regions, and many of them mapped twice to the gene. To further investigate these observations, we have *de novo* reassembled this fragment using exclusively long reads mapping to the gene. The new assembly, supported by 20× average coverage and multiple span-through reads, revealed a 7 kb duplication in the "Tosca", but not "BRH-1" genome (including a full copy of the TNL gene – described below; yellow box in **Figure 5B** and **Supplementary Figure 2**).

The duplication identified in "Tosca" was further confirmed using a pair of primers flanking the polymorphic site (TD1_F1/TD1_R, **Supplementary Table 2**) that generate different product lengths for "BRH-1," both duplicated "Tosca" TNL paralogs and their homeolog from the C genome in "Tosca" and "BRH-1" (C genome homeologs are nearly identical in both lines; **Supplementary Figure 3**). PCR reactions performed

FIGURE 4 | Schematic location of the *Crr3^Tsc* region in the context of other resistance loci and markers located on Express 617 chromosome A03. **(A)** A general overview of the location of *Rcr1/Cra/Crb/Crb^Kato* and *Crr3/CRd/Crk* regions on the A03 chromosome. **(B)** Zoomed region from the *Crr3/CRd/Crk* fragment (marked with a black box on **A**). Markers labeled with the same color cosegregate in the mapping DH population. No genetic data were obtained for markers indicated with black. The precise start of the *CRd* locus could not be physically mapped onto Express 617 assembly.

on the parents and 11 DH lines with varying degrees of infection resistance revealed the expected pattern of bands, with both "Tosca"-specific alleles segregating with the resistance phenotype (**Figure 6**).

The duplicated region covers the entire TNL gene and a fragment homologous to STP6 **Figure 5B** and **Supplementary Figure 2**. In Darmor-*bzh,* the STP6 fragment is annotated as a distinct gene (BnaA03g29290D), while in Express 617 assembly, the TNL and STP genes are merged into a single entity (A03p030030.1_BnaEXP). Neither the TNL nor the STP fragment is annotated in the scaffoldA03 of the pangenome.

Our RNA-seq data analysis suggests that in both copies the TNL and STP fragment form a single transcription unit; however, the CDS terminates before reaching the STP fragment. The duplicates are separated by a 12 kbp spacer, containing two ~1.2 kbp transcribed, spliced regions. Blast search of genomic and transcriptomic sequences of both transcribed fragments did not provide any conclusive results. Transcript variants of the

transcribed regions contained fragmented ORFs (up to 117 aa in length) with limited similarity to known proteins.

Additionally, the duplicated genes (including the STP fragment) differ in their splicing structure, producing transcripts with 8 and 9 exons, respectively. They share more than 90% identity on the genomic sequence level, with most differences situated downstream of the TNL encoding ORF. On the transcript level, the similarity is 90.9% (4529/4982). The protein sequences are identical in 94.9% (1051/1108) and similar in 96.3 (1067/1108). The alignment has 13 gaps (1.2%), with 10 located at the very end.

Sequence Variation Within Resistance Locus

The location and molecular effect of variants differing between "Tosca" and "BRH-1" were assessed using SNPeff with reference to the Express 617 assembly. The region covering resistance

FIGURE 5 | (A) Simplified plot showing resistance locus coverage of ONT long reads aligned to the Express 617 reference genome. A sharp, flat peak (indicated by an arrow) in "Tosca" genomic data marks the location of the fragment containing the TNL gene (yellow). Darker colors on the coverage plot represent insertions larger than 10 bp. **(B)** Schematic overview of (i) depth of coverage of ON reads (DoC) and (ii) within the duplicated "Tosca" region. Yellow – a fragment corresponding to the BnaA03g29300D gene; green – homologous fragments of STP6 gene. **(C)** Pairwise alignment-based comparison of CDS sequence variability between TNL homologs in "Tosca" (T1,T2) and "BRH-1." Synonymous substitutions are marked as empty circles below the axis, non-synonymous are indicated by filled circles above.

to *P. brassicae* infection contains 1521 polymorphic sites: 1327 SNPs, 61 indels, 1 small duplication, and 132 mixed-type variants. The large, duplicated region, covering the TNL gene found in the "Tosca" genome, was omitted from the SNPeff analysis and evaluated separately.

Most of the detected variants are located within intergenic regions. Exonic and intronic polymorphisms account for 6.15 and 7% of the total sequence variability, respectively. The non-synonymous/synonymous substitution ratio is 0.51. The molecular effect, as defined by SNPeff, was high for 0.25%, moderate for 2.07%, and low for 4.62% of the variants. The remaining differences were classified as modifiers. Both genotypes showed the presence of insertions and deletions, located mainly in introns and intergenic regions. Sequences of insertions larger than 500 bp did not show similarity to any annotated genes. A more detailed, gene-oriented analysis of the SNP effect revealed that among protein-coding genes with detectable expression in at least one of the studied lines,

only two genes, namely A03p030010.1_BnaEXP (Darmor-*bzh* BnaA03g29270D) and A03p030120.1_BnaEXP/A03p030120.2_ BnaEXP (Darmor-*bzh* BnaA03g57410D), contain high-effect variants. A03p030120.1_BnaEXP/A03p030120.2_BnaEXP encodes a metallochaperone and carries a splice donor variant in the "Tosca" cultivar. Besides, this gene harbors the highest variability, with 29 missense and 18 synonymous differences between the lines. A03p030010.1_BnaEXP encodes a chaperonin and contains a premature stop codon in the "BRH-1" (**Supplementary Table 8**).

To reduce the effect of potential ambiguous mapping, the coding sequences of the duplicated "Tosca" TNL genes were compared based on a transcript assembly. The predicted CDS contained a large number of variants between the "BRH-1" gene and both copies from "Tosca" (**Figure 5C**). Interestingly, as noted before, the "Tosca" paralogs differ considerably at the sequence and gene structure levels (**Figure 5C**, T1-T2; **Supplementary Figure 2**).

FIGURE 6 | Allele-discriminating PCR. The gel electropherogram **(B)** shows a result of PCR reaction designed as depicted on the schematic **(A)**. B, T1, and T2 stand for BnaA03g29300D copies from "BRH-1" and "Tosca," respectively. C03 represents a homoeologous region on the C03 chromosome, identical between the parental lines. Expected products are color-coded according to the schematic, and juxtaposed with the gel. Gel lanes: M – size marker, T – "Tosca," B – "BRH-1," W – water-containing control reaction, 266-56 – selected lines, recombining in the proximity of the resistance locus. Susceptible lines are underlined with red, resistant with teal. Both "Tosca" alleles (fragments 224 bp and 577 bp) segregated with the resistance.

The *C*-terminal coding fragments have different lengths (60 bp and 30 bp). The lack of 30 bp in the T1 gene results in a frameshift(s) and, consequently, in a different amino acid sequence at the *C*-end of the encoded protein. Apart from the *C*-terminal variance, the sequences differ with regard to 6 synonymous, 36 non-synonymous substitutions, and 9 in-frame deletions (6 and 3 bp long). 35 of the missense variants cluster within and in close vicinity to the LRR domain coding sequence, with the rest of the protein sequence differing only at one position near the *N*-terminus (**Figure 5C**).

The "BRH-1" TNL homologous gene has the same *C*-terminal composition as the "Tosca" paralog T2 (**Figure 5C**, T2-BRH). The BRH and T2 genes differ with regard to 1 in-frame deletion (3 bp), 12 synonymous, and 63 non-synonymous substitutions. On the other hand, the BRH and T1 genes, besides the *C*-terminus variance, differ with regard to 15 synonymous and 59 non-synonymous substitutions (**Figure 5C**, T1-BRH). Similar to the T1–T2 comparison, nearly all differences between "BRH-1" and "Tosca" gene copies are localized near the *C*-terminus and in the LRR domain coding sequence. The cDNA, CDS, and protein sequences of the "Tosca" and "BRH-1" genes are available in the **Supplementary Material**.

Differential Gene Expression Analysis of Resistant and Susceptible DH Lines

To further characterize the 25 genes located within the locus harboring resistance to clubroot disease, we have examined the differences in transcript levels between inoculated and non-inoculated control roots from resistant and susceptible DH lines in the context of the global pattern of differentially expressed genes.

RNA-seq comparison of transcript accumulation between non-inoculated control resistant versus susceptible plants showed a differential signal for 1,247 genes, with only one located within the resistance locus – the BnaA03g29270 gene (**Supplementary Tables 9,10**). This gene shows a significantly higher (logFC = 1.9, adjusted *p*-value $\leq$ 0.0017) expression level in the resistant line. The gene encodes for a homolog of an *Arabidopsis thaliana* chaperone protein (CCT3).

Subsequently, we have explored differentially expressed genes in roots 46 days after inoculation (DAI) with *P. brassicae*, using a resistant and susceptible line from the mapping population.

In the case of the resistant line, we have identified 111 genes that showed differential transcript accumulation after inoculation (91 up- and 20 down-regulated genes; **Supplementary Tables 9, 10**). 53 of these genes were differentially expressed only in the resistant line. Most of them fall into three general Gene Ontology classes: chitin metabolism, regulation of growth, and defense response (**Supplementary Table 11**). None of the differentially expressed genes identified in the resistant line was located within the resistance locus. The nearest gene showing a differential expression pattern – BnaA03g28780D (encoding Hevein-like preprotein, reported to be involved in the defense response against fungi and bacteria) – is located 200 kbp upstream from the locus.

Analysis of the inoculated susceptible line revealed a much high number (6821) of differentially expressed genes (**Supplementary Tables 9, 10**). Among them, 2,778 were up- and 4,043 were down-regulated. The Gene Ontology-based assignment showed a much more diverse spectrum of molecular functions, among others: oxidative stress response, carbohydrate metabolism, lignin metabolism, chitin metabolism, defense response, auxin signaling (**Supplementary Table 12**). The KEGG pathway enrichment analysis yielded significant hits for phenylpropanoid biosynthesis, biosynthesis of secondary metabolites, glutathione metabolism, stilbenoid, diarylheptanoid and gingerol biosynthesis, and flavonoid biosynthesis.

Four of the differentially expressed genes were located within the resistance locus: down-regulated germin-like protein (BnaA03g29240D – ortholog of AtGLP8, AT3G05930), up-regulated Sugar Transporter Protein (BnaA03g29310D – ortholog of AtSTP6, AT3G05960), down-regulated Fantastic Four protein (BnaA03g57340D – ortholog of AtFAF4, AT3G06020), up-regulated protein trichome birefringence-like (BnaA03g57390D – ortholog of AtTBL10, AT3G06080). None of them, however, were differentially expressed in the resistant line. Moreover, their expression levels were similar in resistant and susceptible control, non-inoculated plants.

The transcript levels of the TNL gene BnaA03g29300D remained unchanged for "BRH-1" and both "Tosca" copies in both lines after inoculation; however, the "BRH-1" and "Tosca" copies were expressed at relatively high levels in the control and the inoculated plants. The transcript levels of the two "Tosca" copies detected in the resistant line added up to twice the amount of the transcript level of one copy expressed in the susceptible line (**Supplementary Table 13**).

DISCUSSION

Based on genetic mapping of a population of 250 DH plants, we were able to identify a single locus conferring resistance to clubroot disease in the winter oilseed rape cultivar "Tosca." The "Tosca" resistance has a different background than the widely utilized "ECD-04," introgressed into the "Mendel" cultivar (Diederichsen and Sacristan, 1996; Diederichsen et al., 2006; Fredua-Agyeman and Rahman, 2016), which makes the source relevant in the *B. napus* breeding efforts.

The identified "Tosca" resistance locus, designated as $Crr3^{Tsc}$, in *B. napus* is located on the A03 chromosome within a previously described *Crr3* locus described in *B. rapa* (Hirai et al., 2004; Saito et al., 2006), which together with *CRk* (Sakamoto et al., 2008; Matsumoto et al., 2012), and *CRd* (Pang et al., 2018) forms a larger cluster of clubroot resistance genetic factors. This cluster has been recently spotted in a GWA study in a panel of *B. napus* ssp. *napobrassica* (rutabaga), which, like "Tosca", are of Nordic origin (Fredua-Agyeman et al., 2020). The region of ∼750 kbp identified in this GWA study was associated with resistance to *P. brassicae* pathotypes 2B and 8P (classified according to Canadian Clubroot Differential Set), which are subsets of Some's P2 pathotype (Strelkov et al., 2018). Here, we show that a region of

∼120 kbp of the $Crr3^{Tsc}$ locus explains the resistance to field isolates consisting of a mixture of pathotypes with the highest prevalence of P3.

The locus was anchored based on a single bin of 11 marker sequences to a region of 97,783 bp of the *B. napus* Express 617 genome assembly. Within the locus, we have identified 25 protein-coding genes. 13 of them were found to be constitutively expressed at the late stage of infection and 4 were found to be differentially expressed between contrasting susceptible and resistant lines of the mapping population. Some of these genes show a functional annotation that makes them interesting candidates to be involved in various stages of *P. brassicae* infection.

Resistance might be expressed constitutively or induced after the initial infection with the pathogen. The non-inoculated control plants showed significant differences in constitutive gene expression patterns, but only 1 out of 1247 was located within the mapped resistance locus. The differentially expressed gene BnaA03g29270D, a homolog of *Arabidopsis thaliana* chaperone protein CCT3, does not offer a direct and evident connection to the mechanism of plant resistance and is unlikely to be involved in resistance expression. The analysis of gene expression affected by the interaction with the pathogen at the later stage of infection provided more interesting candidates. For example, BnaA03g29310D gene which is a homolog of AtSTP6. STPs are monosaccharide/H + symporters that mediate the transport of monosaccharides from the apoplast into the cells (Büttner, 2010). We have observed a significant upregulation of STP6 in infected roots of susceptible, but not of resistant plants. STP family genes, namely STP8 and STP13, have previously been reported to be up-regulated upon clubroot infection in *A. thaliana* (Walerowski et al., 2018), while STP4, STP12, STP1 showed a differential expression pattern in an infected, clubroot-susceptible *Brassica oleracea* cultivar CS-JF1 (Zhang et al., 2019). However, as the expression was affected only in the susceptible line, this gene might have been up-regulated in response to a successful transformation of plant metabolism by the clubroot pathogen during the invasion of the roots. Thus, a potential resistance effect would have to be driven by the inhibition of the expression induction. A similar phenomenon might be responsible for the expression of the gene BnaA03g57390D, harbored within the resistance locus, encoding a homolog of Trichome Birefringence Like 10 (TBL10) protein. This gene was up-regulated in the roots of susceptible, infected plants, but remained constant in the resistant line. TBL proteins are modifiers of the cell wall (Bischoff et al., 2010; Yuan et al., 2016; Gao et al., 2017) and AtTBL10 was found to be involved in O-acetylation of pectin (Stranne et al., 2018). In many reports, pectin hypoacetylation has been linked to increased disease resistance (reviewed in Pauly and Ramírez, 2018), which may explain the lack of TBL10 induction in clubroot defense reaction. Another down-regulated gene from the resistance locus, BnaA03g57340D, is a member of the FANTASTIC FOUR (FAF) protein family. Its expression was down-regulated in susceptible, but not in resistant plants. Overexpression of FAF members was shown to inhibit root growth, which could be rescued by exogenous sucrose (Wahl et al., 2010). Thus,

FAF might perform a role in integrating auxin and sugar signaling during infection progression, allowing the pathogen to manipulate the physiological processes of susceptible plants for more efficient infection progression. Another differentially expressed gene from the resistance locus which has been described to be involved in disease resistance expression is BnaA03g29240D – a homolog of Germin-like protein 8 (GLP8). GLPs are well established as an important component of the biotic stress response (Dunwell et al., 2008; Ilyas et al., 2016). In *B. napus,* GLPs are involved in oxidative burst initiation during *Sclerotinia sclerotiorum* infection (Rietz et al., 2012). GLP5 was also found to have higher expression in a line of *B. rapa* carrying the *Rcr1* clubroot resistance gene (Song et al., 2016). In our study, GLP8 expression was highly reduced in the roots of susceptible plants, though it remained constant in the resistant plants.

Because the expression analysis did not reveal a clear, dominant candidate gene located in the resistance locus that could be responsible for the resistance in "Tosca" we have further explored the properties of the mapped genomic fragment by sequencing the parental cultivars with Oxford Nanopore technology. Detailed analysis of the sequencing data showed a high level of polymorphism on a single nucleotide as well as a larger scale. Despite a large number of differences between the parental lines, most of them were located within non-coding (genic and non-genic) regions. Additionally, the biological impact of the majority of the polymorphisms was predicted to be low in most of the 25 genes. However, the mapping of the ON reads revealed one striking difference – a large duplication in "Tosca" that covered a full copy of the BnaA03g29300D gene. The duplicated gene contains TIR, NB-ARC, and LRR domains, thus belonging to the TNL subclass of NLR genes. Many of these genes are known to be involved directly or indirectly in the recognition of pathogen effector molecules and initiation of downstream defense responses (reviewed in: Dubey and Singh, 2018; de Araújo et al., 2019). As these genes are known to be involved in the very early stages of signaling cascades, they potentially could be differentially expressed in the early stages of the infection process. The identification of a recent copy of the TNL genes to some extent conforms with this presumption. Assuming that recently duplicated genes retain their original function, we may speculate that the effect of enhanced resistance to pathogen infection in "Tosca" is linked with cumulatively elevated expression (2 times) of two copies of the TNL genes. We cannot, however, exclude an alternative possibility that the new copy of the gene acquired new specificity toward the particular *P. brassicae* pathotypes or that both copies are involved in a more complex resistance initiation (see later). So far, two clubroot resistance genes have been cloned, *Crr1a* (Hatakeyama et al., 2013) and *Cra/CRb* (Ueno et al., 2012; Hatakeyama et al., 2017), both encoding TNL proteins.

The identification of the genomic fragment corresponding to the region defined by the peak marker for the resistance locus was not straightforward using the Darmor-*bzh* 4.1 genome assembly. The coverage of the region was not complete and one of the three contigs mapping to this fragment was misassembled.

The locus, however, was correctly placed on the long-read-based reference pangenome and Express 617 assemblies. The Darmor-*bzh* 4.1 reference assembly was constructed prior to the advance of long-read sequencing technologies mainly based on Illumina short-read sequencing and is thus highly fragmented (Lee et al., 2020). We have to note that between the submission and publication of this article, an upgraded, Oxford Nanopore-based version of Darmor-bzh genome (v10) was published, in which the region in question is assembled in agreement with the results of our study and other long-read assemblies (Rousseau-Gueutin et al., 2020). Moreover, it has been shown that in *B. napus* more than 50% of known RGA copies do not occur in a single reference genotype assembly but require analysis of pangenome assemblies for detection (Dolatabadian et al., 2020). Thus, our analysis represents a typical example of how the use of long-read sequencing technology and pangenome sequence assemblies allows more efficient dissection of plant disease resistance loci.

Frequent duplications and clusterization of resistance-related NLR genes are a well-established phenomenon (reviewed in de Araújo et al., 2019; van Wersch and Li, 2019). BnaA03g29300D is flanked by homologous STP6 sequences. This configuration may have served as a foundation for homology-dependent duplication events, for example, by unequal crossing-over. Remarkably, both TNL gene copies seem to be functional, i.e., neither underwent pseudogenization. Both are constitutively expressed in 7-week-old plants and the transcripts contain full-length ORF's. Importantly, the copies harbor a large proportion of polymorphic, non-synonymous sites observed between "Tosca" and "BRH-1," nearly all of which lay within the pattern-recognizing LRR domain, implicating a strong positive selection acting on this domain. Positive selection promoting rapid sequence changes in the NLR genes, especially within LRR domains has already been frequently reported (Bergelson et al., 2001; Mondragón-Palomino et al., 2002; Yang et al., 2013; Karasov et al., 2014; Gao et al., 2018; Han, 2019). LRR domains are reported to be the major factors determining recognition specificity (reviewed in Padmanabhan et al., 2009); therefore, the apparent differences in the amino acid sequence domain may be responsible for the capability of the "Tosca" to sense the *P. brassicae* elicitors and induce the downstream defense response. Remarkably, the two tandem paralogs in the "Tosca" genome themselves differ significantly in the amino acid sequence of their LRR domains. The differences may allow for a broader range of elicitor recognition or the coordination of a more complex response to infection. The TNL proteins are known to engage in functional homo- and heterodimerization (Williams et al., 2014) and various modes of entanglement in the defense response initiation, which often involves clustered genes (reviewed in de Araújo et al., 2019; van Wersch and Li, 2019). Nonetheless, in the case of previously described clubroot resistance locus *Cra/Crb/CRb*^{kato}*,* which consists of at least six tandemly repeated NLR genes, a majority of the resistance effect is attributed to a single gene, with residual, unexplained effect (Hatakeyama et al., 2017).

Furthermore, the BnaA03g29300D gene is a homolog of *B. rapa* Bra001175, which has been shown to have a higher

level of expression in the clubroot-resistant, compared to the susceptible genotype of the *CRd*-carrying line during the early stages of the infection process, at day 13 after inoculation (Pang et al., 2018). In our study, using RNA-seq data, when each of the duplicated genes was tested separately, we could not find statistical differences in BnaA03g29300D expression between resistant and non-resistant inoculated lines and no evidence of induction after inoculation. However, both duplicated genes are expressed at a similar, relatively high level in the roots of "Tosca"-background plants, showing a cumulative 2-times higher transcript accumulation before infection compared to the "BRH-1"-derived, single copy line.

In summary, the search for the genetic background of resistance to *P. brassicae* infection in *B. napus* cv. Tosca revealed a complex picture of genomic and transcriptomic changes. Based on genetic mapping, structural genomics, expression analyses, and functional annotation, we conclude that the TNL gene (BnaA03g29300D) duplication is most likely to be involved in the resistance. Certainly, further experimental tests, including a gene knockout and functional complementation must be conducted to confirm the role of this gene, and/or its duplication, in the resistance against *P. brassicae*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI SRA, BioProject number: PRJNA685314.

AUTHOR CONTRIBUTIONS

KM, CO, and WK conceived the research. EJ and AP performed and MK supervised the phenotyping experiments. KM and JN performed and IB-B supervised the SSR and SCAR genotyping experiments and production of the DH line seeds. PK collected, processed, and analyzed the phenotyping, genotyping, and sequencing data and performed association analyses. HSC generated a part of Oxford Nanopore sequencing data. WK supervised and coordinated the whole project. PK and WK wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.639631/full#supplementary-material

REFERENCES

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bergelson, J., Kreitman, M., Stahl, E. A., and Tian, D. (2001). Evolutionary dynamics of plant R-genes. *Science* 292, 2281–2285. doi: 10.1126/science.1061337

Bischoff, V., Nita, S., Neumetzler, L., Schindelasch, D., Urbain, A., Eshed, R., et al. (2010). TRICHOME BIREFRINGENCE and its homolog AT5G01360 encode plant-specific DUF231 proteins required for cellulose biosynthesis in *Arabidopsis*. *Plant Physiol.* 153, 590–602. doi: 10.1104/pp.110.153320

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Broman, K. W., Gatti, D. M., Simecek, P., Furlotte, N. A., Prins, P., Sen, Ś, et al. (2019). R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics* 211, 495–502. doi: 10.1534/genetics.118.301595

Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. doi: 10.1093/bioinformatics/btg112

Büttner, M. (2010). The *Arabidopsis* sugar transporter (AtSTP) family: an update. *Plant Biol.* 12(Suppl. 1), 35–41. doi: 10.1111/j.1438-8677.2010.00383.x

Cao, T., Manolii, V. P., Strelkov, S. E., Hwang, S.-F., and Howard, R. J. (2009). Virulence and spread of *Plasmodiophora brassicae* [clubroot] in Alberta, Canada. *Can. J. Plant Pathol.* 31, 321–329. doi: 10.1080/07060660909507606

Cegielska-Taras, T., Tykarska, T., Szała, L., Kuraś, L., and Krzymański, J. (2002). Direct plant development from microspore-derived embryos of winter oilseed rape *Brassica napus* L. ssp. Oleifera (DC.) Metzger. *Euphytica* 124, 341–347.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953.

Chawla, H. S., Lee, H., Gabur, I., Vollrath, P., Tamilselvan-Nattar-Amutha, S., Obermeier, C., et al. (2020). Long-read sequencing reveals widespread intragenic structural variants in a recent allopolyploid crop plant. *Plant Biotechnol. J.* 19, 240–250. doi: 10.1111/pbi.13456

Chen, J., Jing, J., Zhan, Z., Zhang, T., Zhang, C., and Piao, Z. (2013). Identification of novel QTLs for isolate-specific partial resistance to *Plasmodiophora brassicae* in *Brassica rapa*. *PLoS One* 8:e85307. doi: 10.1371/journal.pone.0085307

Choi, S. R., Oh, S. H., Chhapekar, S. S., Dhandapani, V., Lee, C. Y., Rameneni, J. J., et al. (2020). Quantitative trait locus mapping of clubroot resistance and *Plasmodiophora brassicae* pathotype banglim-specific marker development in *Brassica rapa*. *Int. J. Mol. Sci.* 21:4157. doi: 10.3390/ijms21114157

Chu, M., Song, T., Falk, K. C., Zhang, X., Liu, X., Chang, A., et al. (2014). Fine mapping of Rcr1 and analyses of its effect on transcriptome patterns during infection by *Plasmodiophora brassicae*. *BMC Genomics* 15:1166. doi: 10.1186/1471-2164-15-1166

Chu, M., Yu, F., Falk, K. C., Liu, X., Zhang, X., Chang, A., et al. (2013). Identification of the Clubroot Resistance Gene Rpb1 and introgression of the resistance into canola breeding lines using a marker-assisted approach. *Acta Hortic.* 2013, 599–605. doi: 10.17660/actahortic.2013.1005.74

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Clarke, W. E., Higgins, E. E., Plieske, J., Wieseke, R., Sidebottom, C., Khedikar, Y., et al. (2016). A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus

markers in the allotetraploid genome. *Theor. Appl. Genet.* 129, 1887–1899. doi: 10.1007/s00122-016-2746-7

Dakouri, A., Zhang, X., Peng, G., Falk, K. C., Gossen, B. D., Strelkov, S. E., et al. (2018). Analysis of genome-wide variants through bulked segregant RNA sequencing reveals a major gene for resistance to *Plasmodiophora brassicae* in *Brassica oleracea. Sci. Rep.* 8:17657. doi: 10.1038/s41598-018-36 187-5

de Araújo, A. C., Fonseca, F. C. A., Cotta, M. G., Alves, G. S. C., and Miller, R. N. G. (2019). Plant NLR receptor proteins and their potential in the development of durable genetic resistance to biotic stresses. *Biotechnol. Res. Innov.* 3, 80–94.

DeYoung, B. J., and Innes, R. W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat. Immunol.* 7, 1243–1249. doi: 10.1038/ni1410

Diederichsen, E., Beckmann, J., Schondelmeier, J., and Dreyer, F. (2006). Genetics of clubroot resistance in *Brassica napus* 'Mendel'. *Acta Hortic.* 706, 307–312. doi: 10.17660/actahortic.2006.706.35

Diederichsen, E., Frauen, M., Linders, E. G. A., Hatakeyama, K., and Hirai, M. (2009). Status and perspectives of clubroot resistance breeding in crucifer crops. *J. Plant Growth Regul.* 28, 265–281. doi: 10.1007/s00344-009-9100-0

Diederichsen, E., Frauen, M., and Ludwig-Müller, J. (2014). Clubroot disease management challenges from a German perspective. *Can. J. Plant Pathol.* 36, 85–98. doi: 10.1080/07060661.2013.861871

Diederichsen, E., and Sacristan, M. D. (1996). Disease response of resynthesized *Brassica napus* L. lines carrying different combinations of resistance to Plasmodiophora brassicae Wor. *Plant Breed.* 115, 5–10. doi: 10.1111/j.1439-0523.1996.tb00862.x

Dixon, G. R. (2007). *Vegetable Brassicas and Related Crucifers.* Wallingford: CABI.

Dixon, G. R. (2009). The occurrence and economic impact of *Plasmodiophora brassicae* and clubroot disease. *J. Plant Growth Regul.* 28, 194–202. doi: 10.1007/s00344-009-9090-y

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Dolatabadian, A., Bayer, P. E., Tirnaz, S., Hurgobin, B., Edwards, D., and Batley, J. (2020). Characterization of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant Biotechnol. J.* 18, 969–982. doi: 10.1111/pbi.13262

Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fesh tissue. *Focus* 12:1053.

Dubey, N., and Singh, K. (2018). "Role of NBS-LRR proteins in plant defense," in *Molecular Aspects of Plant-Pathogen Interaction*, eds A. Singh and I. Singh (Singapore: Springer), 115–138. doi: 10.1007/978-981-10-7371-7_5

Dunwell, J. M., George Gibbings, J., Mahmood, T., and Saqlan Naqvi, S. M. (2008). Germin and germin-like proteins: evolution, structure, and function. *Crit. Rev. Plant Sci.* 27, 342–375. doi: 10.1080/07352680802333938

Edge, P., and Bansal, V. (2019). Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nat. Commun.* 10:4660.

Farid, M., Yang, R.-C., Kebede, B., and Rahman, H. (2020). Evaluation of *Brassica oleracea* accessions for resistance to *Plasmodiophora brassicae* and identification of genomic regions associated with resistance. *Genome* 63, 91–101. doi: 10.1139/gen-2019-0098

Figdore, S. S., Ferreira, M. E., Slocum, M. K., and Williams, P. H. (1993). Association of RFLP markers with trait loci affecting clubroot resistance and morphological characters in *Brassica oleracea* L. *Euphytica* 69, 33–44. doi: 10.1007/bf00021723

Fredua-Agyeman, R., and Rahman, H. (2016). Mapping of the clubroot disease resistance in spring *Brassica napus* canola introgressed from European winter canola cv. "Mendel.". *Euphytica* 211, 201–213. doi: 10.1007/s10681-016-1730-2

Fredua-Agyeman, R., Yu, Z., Hwang, S.-F., and Strelkov, S. E. (2020). Genome-wide mapping of loci associated with resistance to clubroot in *Brassica napus* ssp. napobrassica (Rutabaga) Accessions From Nordic Countries. *Front. Plant Sci.* 11:742. doi: 10.3389/fpls.2020.00742

Gabur, I., Chawla, H. S., Liu, X., Kumar, V., Faure, S., von Tiedemann, A., et al. (2018). Finding invisible quantitative trait loci with missing data. *Plant Biotechnol. J.* 16, 2102–2112. doi: 10.1111/pbi.12942

Gao, Y., He, C., Zhang, D., Liu, X., Xu, Z., Tian, Y., et al. (2017). Two trichome birefringence-like proteins mediate xylan acetylation, which is essential for leaf blight resistance in rice. *Plant Physiol.* 173, 470–481. doi: 10.1104/pp.16.01618

Gao, Y., Wang, W., Zhang, T., Gong, Z., Zhao, H., and Han, G.-Z. (2018). Out of water: the origin and early diversification of plant R-genes. *Plant Physiol.* 177, 82–89. doi: 10.1104/pp.18.00185

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Grandclément, C., and Thomas, G. (1996). Detection and analysis of QTLs based on RAPD markers for polygenic resistance to *Plasmodiophora brassicae* woron in *Brassica oleracea* L. *Theor. Appl. Genet.* 93, 86–90. doi: 10.1007/s001220050251

Han, G.-Z. (2019). Origin and evolution of the plant immune system. *New Phytol.* 222, 70–83. doi: 10.1111/nph.15596

Happstadius, I., Ljungberg, A., Kristiansson, B., and Dixelius, C. (2003). Identification of *Brassica oleracea* germplasm with improved resistance to *Verticillium wilt. Plant Breed.* 122, 30–34. doi: 10.1046/j.1439-0523.2003.00 774.x

Hatakeyama, K., Niwa, T., Kato, T., Ohara, T., Kakizaki, T., and Matsumoto, S. (2017). The tandem repeated organization of NB-LRR genes in the clubroot-resistant CRb locus in *Brassica rapa* L. *Mol. Genet. Genomics* 292, 397–405. doi: 10.1007/s00438-016-1281-1

Hatakeyama, K., Suwabe, K., Tomita, R. N., Kato, T., Nunome, T., Fukuoka, H., et al. (2013). Identification and characterization of Crr1a, a gene for resistance to clubroot disease (*Plasmodiophora brassicae* Woronin) in *Brassica rapa* L. *PLoS One* 8:e54745. doi: 10.1371/journal.pone.0054745

Hejna, O., Havlickova, L., He, Z., Bancroft, I., and Curn, V. (2019). Analysing the genetic architecture of clubroot resistance variation in *Brassica napus* by associative transcriptomics. *Mol. Breed.* 39:112.

Hirai, M., Harada, T., Kubo, N., Tsukada, M., Suwabe, K., and Matsumoto, S. (2004). A novel locus for clubroot resistance in *Brassica rapa* and its linkage markers. *Theor. Appl. Genet.* 108, 639–643. doi: 10.1007/s00122-003-1475-x

Hirani, A. H., Gao, F., Liu, J., Fu, G., Wu, C., McVetty, P. B. E., et al. (2018). Combinations of independent dominant loci conferring clubroot resistance in all four turnip accessions (*Brassica rapa*) from the european clubroot differential set. *Front. Plant Sci.* 9:1628. doi: 10.3389/fpls.2018.01628

Huang, Z., Peng, G., Liu, X., Deora, A., Falk, K. C., Gossen, B. D., et al. (2017). Fine mapping of a clubroot resistance gene in chinese cabbage using SNP markers identified from bulked segregant RNA sequencing. *Front. Plant Sci.* 8:1448. doi: 10.3389/fpls.2017.01448

Hwang, S.-F., Howard, R. J., Strelkov, S. E., Gossen, B. D., and Peng, G. (2014). Management of clubroot (*Plasmodiophora brassicae*) on canola (*Brassica napus*) in western Canada. *Can. J. Plant Pathol.* 36, 49–65. doi: 10.1080/07060661.2013. 863806

Ilyas, M., Rasheed, A., and Mahmood, T. (2016). Functional characterization of germin and germin-like protein genes in various plant species using transgenic approaches. *Biotechnol. Lett.* 38, 1405–1421. doi: 10.1007/s10529-016-2129-9

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kageyama, K., and Asano, T. (2009). Life Cycle of *Plasmodiophora brassicae. J. Plant Growth Regul.* 28, 203–211. doi: 10.1007/s00344-009-9101-z

Karasov, T. L., Horton, M. W., and Bergelson, J. (2014). Genomic variability as a driver of plant–pathogen coevolution? *Curr. Opin. Plant Biol.* 18, 24–30. doi: 10.1016/j.pbi.2013.12.003

Karim, M. M., Dakouri, A., Zhang, Y., Chen, Q., Peng, G., Strelkov, S. E., et al. (2020). Two clubroot-resistance genes, Rcr3 and Rcr9wa, mapped in *Brassica rapa* using bulk segregant RNA sequencing. *Int. J. Mol. Sci.* 21:5033. doi: 10. 3390/ijms21145033

Kato, T., Hatakeyama, K., Fukino, N., and Matsumoto, S. (2012). Identificaiton of a clubroot resistance locus conferring resistance to a *Plasmodiophora brassicae* classified into pathotype group 3 in Chinese cabbage (*Brassica rapa* L.). *Breed. Sci.* 62, 282–287. doi: 10.1270/jsbbs.62.282

Korbas, M., Jajor, E., and Budka, A. (2009). Clubroot (*Plasmodiophora brassicae*) - a threat for oilseed rape. *J. Plant Protect. Res.* 49, 446–451. doi: 10.2478/v10045-009-0071-78

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2016). *lmerTest: Tests in Linear Mixed Effects Models. R Package version 2.0–30.*

Laila, R., Park, J.-I., Robin, A. H. K., Natarajan, S., Vijayakumar, H., Shirasawa, K., et al. (2019). Mapping of a novel clubroot resistance QTL using ddRAD-seq in

Chinese cabbage (*Brassica rapa* L.). *BMC Plant Biol.* 19:13. doi: 10.1186/s12870-018-1615-8

Landry, B. S., Hubert, N., Crete, R., Chang, M. S., Lincoln, S. E., and Etoh, T. (1992). A genetic map for *Brassica oleracea* based on RFLP markers detected with expressed DNA sequences and mapping of resistance genes to race 2 of *Plasmodiophora brassicae* (Woronin). *Genome* 35, 409–420. doi: 10.1139/g92-061

Laurens, F., and Thomas, G. (2004). Inheritance of resistance to clubroot (*Plasmodiophora brassicae* Wor.) in Kale (*Brassica oleracea* ssp. Acephala). *Hereditas* 119, 253–262. doi: 10.1111/j.1601-5223.1993.00253.x

Law, C. W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G. K., et al. (2016). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research* 5:ISCB Comm J-1408. doi: 10.12688/f1000research.9005.3

Lee, H., Chawla, H. S., Obermeier, C., Dreyer, F., Abbadi, A., and Snowdon, R. (2020). Chromosome-scale assembly of winter oilseed rape *Brassica napus*. *Front. Plant Sci.* 11:496. doi: 10.3389/fpls.2020.00496

Lee, J., Izzah, N. K., Choi, B.-S., Joh, H. J., Lee, S.-C., Perumal, S., et al. (2016). Genotyping-by-sequencing map permits identification of clubroot resistance QTLs and revision of the reference genome assembly in cabbage (*Brassica oleracea* L.). *DNA Res.* 23, 29–41.

Lenth, V. R. (2016). Least-squares means: the R package lsmeans. *J. Statist. Softw.* 69, 1–33.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). 1000 genome project data processing subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, L., Luo, Y., Chen, B., Xu, K., Zhang, F., Li, H., et al. (2016). A genome-wide association study reveals new loci for resistance to clubroot disease in *Brassica napus*. *Front. Plant Sci.* 7:1483. doi: 10.3389/fpls.2016.01483

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Liu, L., Qin, L., Zhou, Z., Wilhelmina, G. H., Liu, S., and Wei, Y. D. (2020). Refining the life cycle of *Plasmodiophora brassicae*. *Phytopathology* 110, 1704–1712. doi: 10.1094/phyto-02-20-0029-r

Lv, H., Fang, Z., Yang, L., Zhang, Y., and Wang, Y. (2020). An update on the arsenal: mining resistance genes for disease management of *Brassica* crops in the genomic era. *Hortic. Res.* 7:34. doi: 10.1038/s41438-020-0257-9

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641.

Malinowski, R., Truman, W., and Blicharz, S. (2019). Genius architect or clever thief—how *Plasmodiophora brassicae* reprograms host development to establish a pathogen-oriented physiological sink. *Mol. Plant Microbe Interact.* 32, 1259–1266. doi: 10.1094/mpmi-03-19-0069-cr

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

Matsumoto, E., Ueno, H., Aruga, D., Sakamoto, K., and Hayashida, N. (2012). Accumulation of three clubroot resistance genes through marker-assisted selection in chinese cabbage (*Brassica rapa* ssp. pekinensis). *J. Jpn. Soc. Hortic. Sci.* 81, 184–190. doi: 10.2503/jjshs1.81.184

Matsumoto, E., Yasui, C., Ohi, M., and Tsukada, M. (1998). Linkage analysis of RFLP markers for clubroot resistance and pigmentation in Chinese cabbage (*Brassica rapa* ssp. pekinensis). *Euphytica* 104:79. doi: 10.1023/A:1018370418201

McHale, L., Tan, X., Koehl, P., and Michelmore, R. W. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 7:212.

Mondragón-Palomino, M., Meyers, B. C., Michelmore, R. W., and Gaut, B. S. (2002). Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res.* 12, 1305–1315. doi: 10.1101/gr.159402

Monroe, J. G., Grey Monroe, J., Allen, Z. A., Tanger, P., Mullen, J. L., Lovell, J. T., et al. (2017). TSPmap, a tool making use of traveling salesperson problem solvers in the efficient and accurate construction of high-density genetic linkage maps. *BioData Mining* 10:38. doi: 10.1186/s13040-017-0158-0

Moriguchi, K., Kimizuka-Takagi, C., Ishii, K., and Nomura, K. (1999). A genetic map based on RAPD, RFLP, isozyme, morphological markers and QTL analysis for clubroot resistance in *Brassica oleracea*. *Breed. Sci.* 49, 257–265. doi: 10.1270/jsbbs.49.257

Nagaoka, T., Doullah, M. A. U., Matsumoto, S., Kawasaki, S., Ishikawa, T., Hori, H., et al. (2010). Identification of QTLs that control clubroot resistance in *Brassica oleracea* and comparative analysis of clubroot resistance genes between *B. rapa* and *B. oleracea*. *Theor. Appl. Genet.* 120, 1335–1346. doi: 10.1007/s00122-010-1259-z

Neik, T. X., Barbetti, M. J., and Batley, J. (2017). Current status and challenges in identifying disease resistance genes in *Brassica napus*. *Front. Plant Sci.* 8:1788. doi: 10.3389/fpls.2017.01788

Nguyen, M. L., Monakhos, G. F., Komakhin, R. A., and Monakhos, S. G. (2018). The new clubroot resistance locus is located on chromosome A05 in Chinese Cabbage (*Brassica rapa* L.). *Russ. J. Genet.* 54, 296–304. doi: 10.1134/s1022795418030080

Nomura, K., Minegishi, Y., Kimizuka-Takagi, C., Fujioka, T., Moriguchi, K., Shishido, R., et al. (2005). Evaluation of F2 and F3 plants introgressed with QTLs for clubroot resistance in cabbage developed by using SCAR markers. *Plant Breed.* 124, 371–375. doi: 10.1111/j.1439-0523.2005.01105.x

Padmanabhan, M., Cournoyer, P., and Dinesh-Kumar, S. P. (2009). The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. *Cell. Microbiol.* 11, 191–198. doi: 10.1111/j.1462-5822.2008.01260.x

Pang, W., Fu, P., Li, X., Zhan, Z., Yu, S., and Piao, Z. (2018). Identification and mapping of the clubroot resistance gene CRd in chinese cabbage (*Brassica rapa* ssp. pekinensis). *Front. Plant Sci.* 9:653. doi: 10.3389/fpls.2018.00653

Pang, W., Liang, S., Li, X., Li, P., Yu, S., Lim, Y. P., et al. (2014). Genetic detection of clubroot resistance loci in a new population of *Brassica rapa*. *Hortic. Environ. Biotechnol.* 55, 540–547. doi: 10.1007/s13580-014-0079-5

Pauly, M., and Ramírez, V. (2018). New insights into wall polysaccharide O-Acetylation. *Front. Plant Sci.* 9:1210. doi: 10.3389/fpls.2018.01210

Peng, L., Zhou, L., Li, Q., Wei, D., Ren, X., Song, H., et al. (2018). Identification of quantitative trait loci for clubroot resistance in *Brassica oleracea* with the use of *Brassica* SNP microarray. *Front. Plant Sci.* 9:822. doi: 10.3389/fpls.2018.00822

Piao, Z. Y., Deng, Y. Q., Choi, S. R., Park, Y. J., and Lim, Y. P. (2004). SCAR and CAPS mapping of CRb, a gene conferring resistance to *Plasmodiophora brassicae* in Chinese cabbage (*Brassica rapa* ssp. pekinensis). *Theor. Appl. Genet.* 108, 1458–1465. doi: 10.1007/s00122-003-1577-5

Piepho, H. P., and Möhring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177, 1881–1888. doi: 10.1534/genetics.107.074229

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198.

Rietz, S., Bernsdorff, F. E. M., and Cai, D. (2012). Members of the germin-like protein family in *Brassica napus* are candidates for the initiation of an oxidative burst that impedes pathogenesis of *Sclerotinia sclerotiorum*. *J. Exp. Bot.* 63, 5507–5519. doi: 10.1093/jxb/ers203

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Rocherieux, J., Glory, P., Giboulot, A., Boury, S., Barbeyron, G., Thomas, G., et al. (2004). Isolate-specific and broad-spectrum QTLs are involved in the control of clubroot in *Brassica oleracea*. *Theor. Appl. Genet.* 108, 1555–1563. doi: 10.1007/s00122-003-1580-x

Rousseau-Gueutin, M., Belser, C., Da Silva, C., Richard, G., Istace, B., Cruaud, C., et al. (2020). Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience* 9:giaa137. doi: 10.1093/gigascience/giaa137

Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* 17, 155–158. doi: 10.1038/s41592-019-0669-3

Saito, M., Kubo, N., Matsumoto, S., Suwabe, K., Tsukada, M., and Hirai, M. (2006). Fine mapping of the clubroot resistance gene, Crr3, in *Brassica rapa*. *Theor. Appl. Genet.* 114, 81–91. doi: 10.1007/s00122-006-0412-1

Sakamoto, K., Saito, A., Hayashida, N., Taguchi, G., and Matsumoto, E. (2008). Mapping of isolate-specific QTLs for clubroot resistance in Chinese cabbage (*Brassica rapa* L. ssp. pekinensis). *Theor. Appl. Genet.* 117, 759–767. doi: 10.1007/s00122-008-0817-0

Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., et al. (2017). Accurate detection of complex structural variations using single molecule sequencing. *Biorxiv* [Preprint]. doi: 10.1101/169557

Some, A., Manzanares, M. J., Laurens, F., Baron, F., Thomas, G., and Rouxel, F. (1996). Variation for virulence on *Brassica napus* L. amongst Plasmodiophora brassicae collections from France and derived single-spore isolates. *Plant Pathol.* 45, 432–439. doi: 10.1046/j.1365-3059.1996.d01-155.x

Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* 6, 34–45. doi: 10.1038/s41477-019-0577-7

Song, T., Chu, M., Lahlali, R., Yu, F., and Peng, G. (2016). Shotgun label-free proteomic analysis of clubroot (*Plasmodiophora brassicae*) resistance conferred by the gene Rcr1 in *Brassica rapa*. *Front. Plant Sci.* 7:1013. doi: 10.3389/fpls.2016.01013

Stahl, A., Vollrath, P., Samans, B., Frisch, M., Wittkop, B., and Snowdon, R. J. (2019). Effect of breeding on nitrogen use efficiency-associated traits in oilseed rape. *J. Exp. Bot.* 70, 1969–1986. doi: 10.1093/jxb/erz044

Stranne, M., Ren, Y., Fimognari, L., Birdseye, D., Yan, J., Bardor, M., et al. (2018). TBL10 is required for O -acetylation of pectic rhamnogalacturonan-I in *Arabidopsis thaliana*. *Plant J.* 96, 772–785. doi: 10.1111/tpj.14067

Strehlow, B., de Mol, F., and Struck, C. (2015). Risk potential of clubroot disease on winter oilseed rape. *Plant Dis.* 99, 667–675. doi: 10.1094/pdis-05-14-0482-re

Strelkov, S. E., Hwang, S.-F., Manolii, V. P., Cao, T., and Feindel, D. (2016). Emergence of new virulence phenotypes of *Plasmodiophora brassicae* on canola (*Brassica napus*) in Alberta, Canada. *Eur. J. Plant Pathol.* 145, 517–529. doi: 10.1007/s10658-016-0888-8

Strelkov, S. E., Hwang, S.-F., Manolii, V. P., Cao, T., Fredua-Agyeman, R., Harding, M. W., et al. (2018). Virulence and pathotype classification of *Plasmodiophora brassicae* populations collected from clubroot resistant canola (*Brassica napus*) in Canada. *Can. J. Plant Pathol.* 40, 284–298. doi: 10.1080/07060661.2018.1459851

Suwabe, K., Tsukazaki, H., Iketani, H., Hatakeyama, K., Fujimura, M., Nunome, T., et al. (2003). Identification of two loci for resistance to clubroot (*Plasmodiophora brassicae* Woronin) in *Brassica rapa* L. *Theor. Appl. Genet.* 107, 997–1002. doi: 10.1007/s00122-003-1309-x

Suwabe, K., Tsukazaki, H., Iketani, H., Hatakeyama, K., Kondo, M., Fujimura, M., et al. (2006). Simple sequence repeat-based comparative genomics between *Brassica rapa* and *Arabidopsis thaliana*: the genetic origin of clubroot resistance. *Genetics* 173, 309–319. doi: 10.1534/genetics.104.038968

Szała, L., Sosnowska, K., and Cegielska-Taras, T. (2020). Induced chromosome doubling in microspores and regenerated haploid plants of *Brassica napus*. *Acta Biol. Cracoviensia Ser. Bot.* 62, 23–31.

Törönen, P., Medlar, A., and Holm, L. (2018). PANNZER2: a rapid functional annotation web server. *Nucleic Acids Res.* 46, W84–W88. doi: 10.1093/nar/gky350

Ueno, H., Matsumoto, E., Aruga, D., Kitagawa, S., Matsumura, H., and Hayashida, N. (2012). Molecular characterization of the CRa gene conferring clubroot resistance in *Brassica rapa*. *Plant Mol. Biol.* 80, 621–629. doi: 10.1007/s11103-012-9971-5

van Wersch, S., and Li, X. (2019). Stronger when together: clustering of Plant NLR Disease resistance Genes. *Trends Plant Sci.* 24, 688–699. doi: 10.1016/j.tplants.2019.05.005

Vigier, B., Chiang, M. S., and Hume, D. J. (1989). Source of resistance to clubroot (*Plasmodiophora brassicae* Wor.) in triazine-resistant spring canola (rapeseed). *Can. Plant Dis. Surv.* 69, 113–115.

Voorrips, R. E., Jongerius, M. C., and Kanne, H. J. (1997). Mapping of two genes for resistance to clubroot (*Plasmodiophora brassicae*) in a population of doubled

haploid lines of *Brassica oleracea* by means of RFLP and AFLP markers. *Theor. Appl. Genet.* 94, 75–82. doi: 10.1007/s001220050384

Voorrips, R. E., and Visser, D. L. (1993). Examination of resistance to clubroot in accessions of *Brassica oleracea* using a glasshouse seedling test. *Netherlands J. Plant Pathol.* 99, 269–276. doi: 10.1007/bf01974308

Wahl, V., Brand, L. H., Guo, Y.-L., and Schmid, M. (2010). The FANTASTIC FOUR proteins influence shoot meristem size in *Arabidopsis thaliana*. *BMC Plant Biol.* 10:285. doi: 10.1186/1471-2229-10-285

Walerowski, P., Gündel, A., Yahaya, N., Truman, W., Sobczak, M., Olszak, M., et al. (2018). Clubroot disease stimulates early steps of phloem differentiation and recruits sweet sucrose transporters within developing galls. *Plant Cell* 30, 3058–3073. doi: 10.1105/tpc.18.00283

Wallenhammar, A. C. (1996). Prevalence of *Plasmodiophora brassicae* in a spring oilseed rape growing area in central Sweden and factors influencing soil infestation levels. *Plant Pathol.* 45, 710–719. doi: 10.1046/j.1365-3059.1996.d01-173.x

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

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Berlin: Springer Science & Business Media.

Williams, S. J., Sohn, K. H., Wan, L., Bernoux, M., Sarris, P. F., Segonzac, C., et al. (2014). Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science* 344, 299–303. doi: 10.1126/science.1247357

Yang, S., Li, J., Zhang, X., Zhang, Q., Huang, J., Chen, J.-Q., et al. (2013). Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18572–18577. doi: 10.1073/pnas.1318211110

Yu, F., Zhang, X., Peng, G., Falk, K. C., Strelkov, S. E., and Gossen, B. D. (2017). Genotyping-by-sequencing reveals three QTL for clubroot resistance to six pathotypes of *Plasmodiophora brassicae* in *Brassica rapa*. *Sci. Rep.* 7:4516.

Yuan, Y., Teng, Q., Zhong, R., Haghighat, M., Richardson, E. A., and Ye, Z.-H. (2016). Mutations of *Arabidopsis* TBL32 and TBL33 affect xylan acetylation and secondary wall deposition. *PLoS One* 11:e0146460. doi: 10.1371/journal.pone.0146460

Zhang, H., Feng, J., Hwang, S.-F., Strelkov, S. E., Falak, I., Huang, X., et al. (2016). Mapping of clubroot (*Plasmodiophora brassicae*) resistance in canola (*Brassica napus*). *Plant Pathol.* 65, 435–440. doi: 10.1111/ppa.12422

Zhang, T., Zhao, Z., Zhang, C., Pang, W., Choi, S. R., Lim, Y. P., et al. (2014). Fine genetic and physical mapping of the CRb gene conferring resistance to clubroot disease in *Brassica rapa*. *Mol. Breed.* 34, 1173–1183. doi: 10.1007/s11032-014-0108-1

Zhang, W., Wang, S., Yu, F., Tang, J., Yu, L., Wang, H., et al. (2019). Genome-wide identification and expression profiling of sugar transporter protein (STP) family genes in cabbage (*Brassica oleracea* var. capitata L.) reveals their involvement in clubroot disease responses. *Genes* 10:71. doi: 10.3390/genes10010071

Zheng, X., Koopmann, B., Ulber, B., and von Tiedemann, A. (2020). A global survey on diseases and pests in oilseed rape—current challenges and innovative strategies of control. *Front. Agron.* 2:908. doi: 10.3389/fagro.2020.590908