Justus Liebig University, Giessen
Institute of Agronomy and Plant Breeding I
Department of Plant Breeding

Implications of structural genomic variants on agronomical traits in elite winter oilseed rape (*Brassica napus* L.)

Inaugural Dissertation for a Doctorate Degree in Agricultural Sciences
in the Faculty of Agricultural Sciences, Nutritional Sciences and
Environmental Management

Examiners
Prof. Dr. Rod J. Snowdon
Prof. Dr. Matthias Frisch

Submitted by
Paul Vollrath
aus Spesenroth (Hunsrück)

Giessen 2021

“I don’t know where I’m goin’

But I sure know where I’ve been”

Whitesnake

Table of Contents

1 General Introduction

1.1 Winter oilseed rape: Origin and genomic complexity

Oilseed rape (*Brassica napus* L., 2n = 4x = 38, AACC) is an allotetraploid interspecific hybrid between the two diploid ancestors turnip rape (*Brassica rapa* L., 2n = 2x = 20, AA) and cabbage (*Brassica oleracea* L., 2n = 2x = 18, CC) (Fig. 1) (U, 1935; Parkin et al., 1995; Chalhoub et al., 2014). Comparisons among orthologous genes revealed that this spontaneous hybridization event most probably took place no more than 7,000 years ago, making *B. napus* a very young species from an evolutionary perspective (Chalhoub et al., 2014).

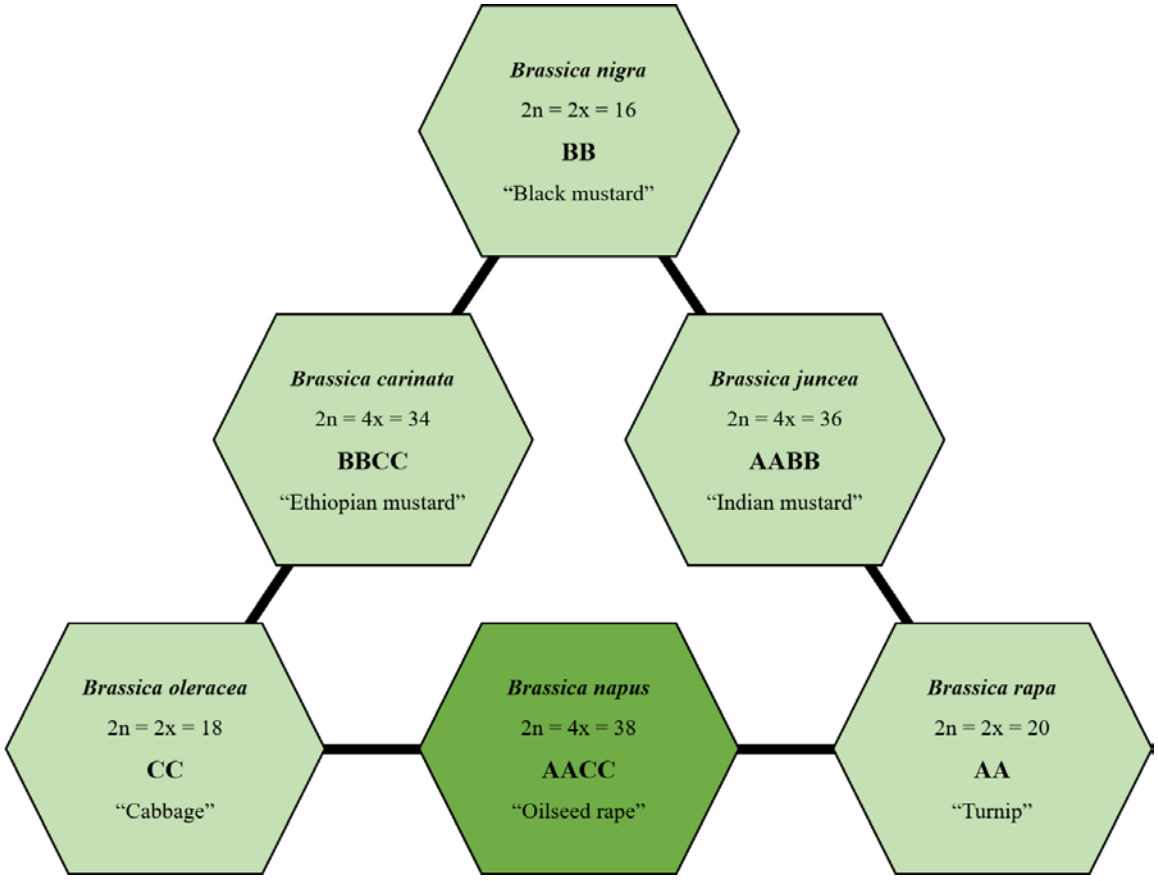

Figure 1. Cytogenetic relationship of diploid and tetraploid brassicas (according to U, 1935).

The initial allopolyploidization most likely occurred after geographically close co-cultivation of the two diploid progenitors in the Mediterranean region (Friedt and

Snowdon, 2010). The study of Chalhoub et al. (2014) also showed that the genome of *B. napus* has undergone up to 72-fold genome multiplications and thus harbors genes with high redundancy. In addition, high similarities between the A and C subgenomes resulted in high frequencies of homoeologous exchanges (Szadkowski et al., 2010; Samans et al., 2017; Hurgobin et al., 2018; Samans et al., 2018).

In 1974, the first nutritionally valuable, 00-quality (seed with zero erucic acid and low glucosinolate content) spring-type oilseed rape was released to market. First, a spontaneous mutation identified by Canadian scientists in the German spring-type cultivar 'Liho' led to seed erucic acid (C22:1) contents below 1% and built the basis of breeding for 0-quality oilseed rape (Downey and Harvey, 1963; Stefansson and Hougen, 1964). The first commercial 0-quality variety was released in Canada in the early 1970s. Secondly, the Polish cultivar 'Bronowski' was found to contain considerably reduced seed glucosinolate contents and was subsequently used in international breeding programs to introduce this quantitative trait into elite erucic-acid free breeding material (Josefsson and Appelqvist, 1968). Introgression of these two traits formed the basis for 00-quality oilseed rape cultivars. At the same time, these introgression represent two major bottlenecks in the history of oilseed rape breeding, causing a strong reduction in genetic diversity within the gene pool of adapted, high-quality oilseed rape accessions.

Oilseed rape was first cultivated as an annual, autumn-sown winter-type crop in Europe. In the beginning of the twentieth century, spring-type and semi-winter ecotypes were introduced into North America, Australia and China. Hence, *B. napus* developed adaptation

to extremely different climates and environments. Winter-, semi-winter- and spring-type oilseed rape/canola ecotypes of *B. napus* are distinguished according to their vernalization requirement, winter hardiness and flowering response to different photoperiods (Snowdon et al., 2006).

1.2 Flowering time – a crucial growth stage

As in other crops and model plants, flowering time is an important trait that clusters *B. napus* ecotypes into distinct forms. The present thesis is exclusively focused on biennial winter-type oilseed rape, which is characterized by its winter-hardiness and vernalization requirement. A second distinct ecotype of *B. napus* is spring-type oilseed rape that generally cannot withstand frost, but has no need for vernalization and is thus grown annually. Finally, an intermediate form, semi-winter-type oilseed rape, is a spring-type with intermediate winter-hardiness that is commonly grown in China and Australia (Werner and Snowdon, 2018). Due to their ecogeographical adaptations in terms of winter-hardiness and photosensitivity, completely distinct and ecotype-dependent breeding programs are essential for commercial oilseed rape breeding. In Arabidopsis it has been shown that *FLOWERING LOCUS C* (*FLC*), a key gene in the vernalization pathway, is primarily responsible for the separation between the different ecotypes. *FLC* negatively regulates one of the central flowering genes, *FLOWERING LOCUS T* (*FT*), until a prolonged period of cold leads to lower FLC levels and stops this repression. This is where the transcriptional activator *CONSTANS* (*CO*) comes into play. *CO*, the major gene in the photoperiod pathway, activates *FT* and thus induces the onset of flowering (Srikanth and Schmid, 2011;

Blümel et al., 2015). However, these genes represent only a few major players in the highly complex flowering pathway of plants.

As for most plants, flowering is a highly sensitive and perhaps the most crucial developmental phase with respect to yield formation in oilseed rape. Besides the serious risk of Sclerotinia infection during flowering, which can cause tremendous yield losses, there is a significant nutrient requirement, especially for nitrogen, during this critical growth phase (Berry et al., 2010; Stahl et al., 2019). However, weather conditions strongly impact the timing of different flowering processes via their effects on the key regulatory factors. With respect to changing climatic conditions, early summers are more frequently affected by long dry periods, resulting in drought stress during flower development and yield formation. Due to changing climatic conditions, long periods of drought have been observed more frequently in spring and early summer within the last decade in Central Europe (Lu et al., 2019). Drought stress during this crucial phase of yield development has been shown to be notably relevant to yield formation (Hohmann, 2017; Schiessl et al., 2020). Thus, it is important to adapt future varieties to the requirements that come along with ongoing climate change, including increasing periods of drought during the main period of flowering (Long et al., 2007; Borghi et al., 2019; Schiessl et al., 2020).

1.3 Blackleg – a major fungal disease of Brassicas

Another extremely important climate-dependent trait in crop plants is disease resistance. One of the most important fungal diseases of winter oilseed rape is blackleg, also known as stem canker. Blackleg is caused by the hemibiotrophic fungal pathogen *Leptosphaeria*

maculans (Desm.) Ces. & de Not. (anamorph: *Phoma lingam* (Tode ex. Fr.) Desm.). After primary infection via airborne ascospores in autumn, typical leaf spots become visible on young leaves (Hammond et al., 1985). Subsequently, the fungus grows systematically and symptomless through the lamina and petiole into the stem tissue. Typical stem canker symptoms can be observed in spring and, in worst case scenarios, lead to the collapse of the plant (West et al., 2001; Howlett et al., 2015). Blackleg is the economically most important disease of oilseed rape in all major production areas worldwide. Severe yield losses have been documented in Europe, North America and Australia (Fitt et al., 2006). For instance, yield losses greater than 50% in the 1980s in Canada and total crop failure in 1972 in Australia have been reported (Juska et al., 1997).

The most effective and sustainable strategy to counteract blackleg is the use of resistant varieties (Kutcher et al., 2013). Thus, commercial plant breeding and breeding research play a crucial role in securing the potential yield of oilseed rape. In general, two types of resistance can be distinguished. Resistance can either be based on monogenic resistance, where a single gene mediates qualitative resistance, or polygenic/ quantitative resistance, where many genes provide partial resistance. Several qualitative resistance genes against *L. maculans* are used in commercial varieties and have been described in numerous genetic studies (Balesdent et al., 2002; Delourme et al., 2006; Rimmer, 2006; van de Wouw et al., 2016). Race-specific, monogenic resistance is based on gene-for-gene interactions of plant resistance (*R*) genes and corresponding avirulence genes (*AvrLm*) of the pathogen. Race-specific resistance is usually expressed at seedling stage and because it confers complete resistance it thus causes a bimodal segregation in mapping populations (Pilet-

Nayel et al., 2017). However, rapid evolution of pathogen populations can cause a loss of function, leading to susceptibility of the plant to pathogens displaying avirulence (Sprague et al., 2006; Brun et al., 2010). Therefore, quantitative, race-non-specific resistance is a valuable tool to counteract blackleg disease.

1.4 Quantitative disease resistances: Genetics and breeding

Quantitative disease resistance is imparted by interactions of multiple genes with small to moderate effects (Poland et al., 2009; St Clair, 2010). Polygenic resistance, partial resistance, adult plant resistance, horizontal resistance and non-specific resistance are often used synonymously for quantitative resistance, although these terms do not have exactly the same meaning (Niks et al., 2015). However, all these terms contain the basic idea of quantitative resistance described above. In contrast to qualitative disease resistance that leads to complete resistance, quantitative resistance is characterized by a continuous distribution between highly susceptible and resistant phenotypes. The specific distribution is the result of different loci contributing to resistance with variable effects and segregating alleles that do not fit Mendelian segregation ratios (French et al., 2016; Pilet-Nayel et al., 2017). However, it has also been shown that a combination of three quantitative trait loci (QTL) conferred almost complete resistance against bacterial spot in tomato (*Solanum lycopersicum*) (Stall et al., 2009). Although epistatic effects of QTL were shown before, QTL are mostly additive. Thus, pyramiding of QTL was suggested to reach high levels of resistance and at the same time increase the durability of resistance (Pilet-Nayel et al., 2017). Even though quantitative resistance is not permanent, pathogens adapt significantly

slower to quantitative resistance in comparison to monogenic, qualitative resistance (Cowger and Brown, 2019). One reason for the extended durability of quantitative resistance might be that overcoming one or a few quantitative resistance genes is only of small advantage for the pathogen population, will not greatly accelerate its propagation and thus will not drastically increase the frequency of the mutation (Niks et al., 2015). Molecular mechanisms of quantitative resistance remain poorly understood, however the following hypotheses were suggested and supported by several studies. Firstly, it is highly probable that genes which regulate plant development have pleiotropic effects on disease resistance. Secondly, it has been shown that quantitative resistance genes are involved in defense response signal transduction. Lastly, weak forms of *R* genes might be involved in quantitative resistance, possibly due to mutations (Poland et al., 2009). Variants in genes responsible for quantitative resistance may lead to altered gene expression levels or more effective timing of gene expression (Niks et al., 2015).

There are several challenges in terms of breeding for quantitative disease resistance. Depending on the germplasm in a particular breeding scheme, the phenotypic expression of quantitative resistance often overlaps with *R*-gene mediated qualitative resistance and thus might be masked by a potentially strong effect of a single *R* gene. Thus, phenotypic assessment of quantitative resistance is impeded. In order to avoid this problem, a resistance breeding program needs to be designed appropriately. In addition, QTL identification highly depends on environmental conditions and can also be specific for an individual mapping population. Environment-specific or population-specific QTL were often observed and have been shown in numerous studies (Huang et al., 2016; Larkan et al., 2016;

Kumar et al., 2018). These interactions present another challenge for commercial breeding programs. Suitable selection of germplasm and trial locations in accordance with the actual breeding goal determine prospects of success at the very beginning of a long and cost-intensive process. Since stacking of major *R* genes against *L. maculans* turned out to promote the selection of specific isolates carrying all corresponding virulence alleles, a particular objective of commercial breeding nowadays is the combination of qualitative and quantitative resistance (Marcroft et al., 2012). The idea is to ensure a high degree of resistance conferred by specific *R* genes and to prolong the effectiveness by a broad quantitative resistance background. The effectiveness of this approach has been demonstrated before when *Rlm6*, an *R* gene against *L. maculans*, was overcome within three consecutive seasons after being introgressed into a highly susceptible genotype. In comparison, the combination of *Rlm6* with strong quantitative resistance prevented the *Rlm6*-mediated resistance from being overcome for several years (Brun et al., 2010). Classical QTL detection and marker-assisted selection is generally focused on genes with major effects and thus seems unsuitable to target quantitative resistance, since several hundreds of genes can potentially be involved (Corwin and Kliebenstein, 2017). Thus, genomic selection (GS) is proposed as an appropriate method to improve the level of quantitative resistance by whole-genome prediction including the consideration of small effects (Poland and Rutkoski, 2016). However, GS largely depends on the population used to train the models. Generally, biparental populations tend to be less suitable for GS in broader breeding panels because they only represent narrow diversity for quantitative traits.

For GS in elite breeding materials it is essential to use training populations which are closely related to the test population.

1.5 Use of multiparental populations for quantitative trait mapping

Multiparental mapping is a compromise that combines the advantages of biparental QTL mapping and genome-wide association mapping. One type of a multiparental mapping population is a set of biparental mapping populations connected through a common parental line. Biparental crosses have two major disadvantages: (1) As described above, they contain low genetic diversity with only two founder lines, and (2) typically early generation crossings are used, leading to a limited number of recombination events and thus to a relatively low mapping resolution (Hall et al., 2010). GWAS, first applied in human genetics (Hirschhorn and Daly, 2005), overcame these limitations of biparental QTL mapping but introduces other disadvantages at the same time. The use of large panels of diverse genotypes significantly increases resolution due to smaller linkage blocks, but does not allow the detection of rare alleles. Given a large population of several hundreds of individuals, the power of GWAS is too low to identify rare alleles, with low effects, present in only very few genotypes (Gibson, 2012). The power of detection of associations can be enhanced by increasing the sample size. The greater number of smaller LD blocks requires a higher number of molecular markers. However, the number of markers no longer poses a problem nowadays with the widespread usage of single-nucleotide polymorphism (SNP) marker arrays and state-of-the-art sequencing technologies. Another main issue of QTL identification using GWAS in a diversity panel is the degree of relatedness of individuals

within the panel, which can lead to complex population structures. However, mixed models were developed as a powerful statistical approach to account for genetic relatedness and handle population structure, hence reducing the number of false positive associations (Korte and Farlow, 2013).

In summary, multiparental mapping populations combine the advantages of biparental QTL mapping populations and diversity panels used for GWAS, and simultaneously reduce their disadvantages. They provide high power for detection of QTL and high resolution of QTL mapping. In addition, the use of multiparental mapping populations increases genetic diversity by selecting more than two founder parents for the initial crossings. The advantage of using multiparental mapping populations has been successfully demonstrated in several mapping studies for flowering time (Higgins et al., 2014; Maurer et al., 2015; Sannemann et al., 2015; Vollrath et al., 2021b) as well as blackleg resistance (Jestin et al., 2015; Gabur et al., 2018; Kumar et al., 2018; Vollrath et al., 2021a), for instance.

1.6 Recent developments in genome sequencing technologies

DNA sequencing became standard practice in molecular genetics after 1977 when sequencing by synthesis was developed by Frederick Sanger and colleagues (Sanger et al., 1977). In short, the Sanger sequencing method was based on DNA elongation using polymerase enzymes and the incorporation of 2', 3'-dideoxynucleotides (ddNTPs) that caused chain termination of DNA synthesis due to the lack of a 3'-hydroxyl group. Four distinct fluorescent dyes attached to the four different nucleotides (A, T, C, G) allowed differentiation of nucleotides. Electrophoresis further displayed a pattern of bands which

revealed the sequence of the DNA template. Sanger sequencing provided high single-base accuracy together with a considerable read length of up to 1000 bp. However, the limiting factor was the low throughput and consequent high cost.

The issue of low throughput was later addressed when second generation sequencing, which was also referred to as next-generation sequencing (NGS), was introduced. The so called second generation can be classified into two groups: sequencing by synthesis and sequencing by hybridization. The major improvement was an immense reduction of per-base sequencing costs via high-throughput, massively-parallel sequencing. Reads obtained from these technologies were generally shorter (100 to 500 bp) and less accurate in comparison to Sanger sequencing, however, massive throughput yielded enormous amounts of sequencing data that was used to build consensus sequences and thus high accuracies were achieved (Slatko et al., 2018). In general, short-read sequencing is cost-efficient and highly accurate, however, suffers from its read length in certain applications (Amarasinghe et al., 2020).

Third generation sequencing is characterized by significantly longer reads obtained from single molecule sequencing. A widely used measurement for read length distribution is the N50 value, representing the length of the shortest read among all reads representing 50% of all sequenced nucleotides in a sequencing run. In simple terms N50 is similar to a median, but more weight is given to longer reads (Lander et al., 2001). One notable advance of long-read sequencing was to make DNA amplification redundant. There are two prominent technologies mainly applied today. Pacific Biosciences (PacBio) Single Molecule Real Time sequencing was commercially released in 2011 and is today characterized by long

reads (N50 of 10 to 30 kb) together with remarkable accuracies (<1% error rate in hi-fidelity reads) on the basic principle of sequencing by synthesis and the real-time detection of incorporated nucleotides. In comparison, nanopore sequencing, developed by Oxford Nanopore Technologies (ONT) and commercially released in 2015, achieves even longer reads (N50 of up to 100 kb), with the drawback of lower read accuracies (presently up to 5% error rate) (Jain et al., 2018; Amarasinghe et al., 2020). Nanopore sequencing distinguishes individual nucleotides on the basis of electrical current fluctuations caused when single-strand DNA passes synthetic pores in a biological membrane. Both of the third generation technologies lack in terms of single base accuracy when compared to short-read sequencing technologies (Heather and Chain, 2016). In particular, nanopore sequencing faces problems when homopolymers occur. However, the big advantage is the remarkable improvement of read lengths, which make long-read technologies notably suitable for *de novo* genome assemblies and for the identification of genome-wide structural variants (SV) that were undetectable before using only short-reads (Loman et al., 2015; Chawla et al., 2020).

1.7 Association of structural genome variations with important traits

Within the last decade, there have been numerous studies addressing SV, particularly presence-absence variation (PAV), using genotyping arrays or short-read sequence data. Large-scale homoeologous exchanges and rearrangements, the co-localization with genes and the association of SV with important agronomical traits have been demonstrated (Chalhoub et al., 2014; Saxena et al., 2014; Samans et al., 2017; Stein et al., 2017; Gabur

et al., 2018; Higgins et al., 2018; Hurgobin et al., 2018). However, mostly due to above-mentioned technical limitations of short-read sequencing, small- to mid-scale structural genome variations were overseen or neglected for a long time, even though they represent a promising source of genetic diversity (Chawla et al., 2020). The implementation of long-read sequencing overcame these limitations and allowed the detection of SV, since single reads spanned repetitive regions and covered entire variants such as deletions or insertions (Mahmoud et al., 2019). No matter which sequencing technology is applied, SV detection obviously depends largely on the quality of available reference genomes. The number of available reference genome assemblies is currently strongly increasing for many crops due to decreasing costs and increasing accuracy of third generation sequencing technologies. Completeness and correctness of the assemblies have been enhanced, enabling more accurate detection of SV. This progress has also promoted the development of crop pan-genomes that can accurately display genomic rearrangements and further allow the differentiation between core and dispensable genes by comparison of individual *de novo* assemblies. These advances will promote the exploration of novel SV and their impact on agronomical traits. In the future, this approach can be expected to provide accessible and affordable data resources for practical breeding programs.

1.8 Future challenges and demands for oilseed rape breeding

Plant breeding has changed notably within the last decades. Traditional selection was steadily replaced by modern technologies like molecular markers or *de novo* sequencing. These technologies allowed the selection based on genetics rather than phenotypic

selection. Latest applications enable the prediction of phenotypes on their genomic profile using sophisticated algorithms or even machine learning approaches. In particular, breeding for quantitative traits relies on genomic prediction and selection rather than marker-assisted or even phenotypic selection. However, breeding targets in general remain similar in terms of high-yielding cultivars with maximum yield stability. Thus, disease resistance was, still is and will continue to be a major and constantly evolving breeding target. Due to growing public attention towards the environmental impact of agriculture, the focus of plant breeding and especially winter oilseed rape breeding is shifting towards further improved nutrient efficiency and disease/pest resistances. Another current subject that has been affecting plant breeding for several years now is the impact of climate change. In addition to drought, oilseed rape now requires adaptations to increasingly extreme weather conditions in order to secure yield potential. All of these topics need to be addressed in modern oilseed rape breeding to preserve oilseed rape as a highly important and valuable component of agricultural crop rotations.

1.9 Aims and objectives

Disease resistance of crop plants is one of the major challenges of plant breeding. Considering the decreasing number of available fungicidal active ingredients due to bans or restriction on their use, more innovation is required in plant breeding to improve disease resistance. Thus, latest technologies need to be accessible and affordable for commercial breeding to provide growers with highly resilient crops. The overarching goal of this thesis was the combined use of GWAS and long-read sequencing in a multiparental population

(see 1.4, 1.5 and 1.6) for accurate detection of SV associated with important agronomical traits (see 1.7), and use of this new information to implement a methodical pipeline that contributes to solve future challenges displayed in subchapter 1.8.

Many recent publications have demonstrated that SV associate with important agronomical traits in major crops. Chapters 2 and 3 of the present thesis represent current examples which investigate the implications of SV on key traits. A first example described in Chapter 2 presents several examples of deletions and insertions associated with flowering time modulation within a multiparental double haploid (DH) population derived from elite European winter oilseed rape cultivars. In particular, this study focused on an intragenic deletion within *FLOWERING LOCUS T* (*FT*), a major flowering time regulation gene, which was associated to earlier onset of flowering. A second example described in Chapter 3 demonstrates the association of SV with quantitative blackleg resistance. A major finding of this study was the identification of a large, novel insertion within the exome of a major-effect resistance gene, *Rlm9*, whose impact on qualitative resistance has previously been described in great detail. In the elite multiparental population, *Rlm9* was found within a QTL region that has only a minor effect on quantitative blackleg resistance, raising an unexpected question about the potential involvement of major-effect resistance genes in quantitative disease resistance in *B. napus*. The presented results underpin the prominent role of SV in determining important agronomical traits and its high potential value for future practical breeding programs of oilseed rape and other crops.

2 A novel deletion in *FLOWERING LOCUS T* modulates flowering time in winter oilseed rape

Vollrath, P., Chawla, H. S., Schiessl, S. V., Gabur, I., Lee, H. T., Snowdon, R. J. and Obermeier C.

Theoretical and Applied Genetics 134, 1217-1231 (2021)

doi: 10.1007/s00122-021-03768-4

Theoretical and Applied Genetics (2021) 134:1217–1231
https://doi.org/10.1007/s00122-021-03768-4

ORIGINAL ARTICLE

A novel deletion in *FLOWERING LOCUS T* modulates flowering time in winter oilseed rape

Paul Vollrath[1] · Harmeet S. Chawla[1] · Sarah V. Schiessl[1] · Iulian Gabur[1] · HueyTyng Lee[1] · Rod J. Snowdon[1] · Christian Obermeier[1]

Received: 11 August 2020 / Accepted: 6 January 2021 / Published online: 20 January 2021

Abstract

Key message **A novel structural variant was discovered in the FLOWERING LOCUS T orthologue BnaFT.A02 by long-read sequencing. Nested association mapping in an elite winter oilseed rape population revealed that this 288 bp deletion associates with early flowering, putatively by modification of binding-sites for important flowering regulation genes.**

Abstract Perfect timing of flowering is crucial for optimal pollination and high seed yield. Extensive previous studies of flowering behavior in *Brassica napus* (canola, rapeseed) identified mutations in key flowering regulators which differentiate winter, semi-winter and spring ecotypes. However, because these are generally fixed in locally adapted genotypes, they have only limited relevance for fine adjustment of flowering time in elite cultivar gene pools. In crosses between ecotypes, the ecotype-specific major-effect mutations mask minor-effect loci of interest for breeding. Here, we investigated flowering time in a multiparental mapping population derived from seven elite winter oilseed rape cultivars which are fixed for major-effect mutations separating winter-type rapeseed from other ecotypes. Association mapping revealed eight genomic regions on chromosomes A02, C02 and C03 associating with fine modulation of flowering time. Long-read genomic resequencing of the seven parental lines identified seven structural variants coinciding with candidate genes for flowering time within chromosome regions associated with flowering time. Segregation patterns for these variants in the elite multiparental population and a diversity set of winter types using locus-specific assays revealed significant associations with flowering time for three deletions on chromosome A02. One of these was a previously undescribed 288 bp deletion within the second intron of *FLOWERING LOCUS T* on chromosome A02, emphasizing the advantage of long-read sequencing for detection of structural variants in this size range. Detailed analysis revealed the impact of this specific deletion on flowering-time modulation under extreme environments and varying day lengths in elite, winter-type oilseed rape.

Introduction

Structural genome variation (SV) is widespread in crop genomes and frequently associated with important traits (Gabur et al. 2019). Recent improvements in genome sequencing enable detection of SV larger than 30 base pairs (Chawla et al. 2020). Different types of SV, including deletions, insertions, duplications, translocation and inversions (Feuk et al. 2006), can have a substantial effect on gene expression and accordingly on the phenotypes of plants (Alonge et al. 2020; Saxena et al. 2014; Wang et al. 2016). Examples include determination of different agronomic traits such as flowering time (Díaz et al. 2012; Schiessl et al. 2017b; Song et al. 2020), disease resistance (Gabur et al. 2018; Hurgobin et al. 2018), heading date (Nishida et al. 2013), seed weight (Song et al. 2020) and biotic stress response (McHale et al. 2012). Due to their cumulative size, more nucleotides tend to be affected by SV than by SNP variants on a genome-wide scale (De Coster et al. 2019).

Brassica napus ($2n = 4x = 38$, AACC) is an allotetraploid interspecific hybrid between *Brassica rapa* ($2n = 2x = 20$, AA) and *Brassica oleracea* ($2n = 2x = 18$, CC) (Chalhoub

Communicated by Maria Laura Federico.

Supplementary information The online version contains supplementary material available at (https://doi.org/10.1007/s00122-021-03768-4).

✉ Christian Obermeier
Christian.obermeier@agrar.uni-giessen.de

[1] Department of Plant Breeding, Justus Liebig University, Giessen, Germany

et al. 2014; Parkin et al. 1995; U 1935). Due to the high similarity between the A and the C subgenomes, homoeologous exchanges occur very frequently (Chalhoub et al. 2014; Hurgobin et al. 2018; Stein et al. 2017; Szadkowski et al. 2010). Many studies to date have focused on SV in resynthesized oilseed rape, which is known to be affected by numerous and large variation. SV has been visualized using genetic mapping (Song et al. 1995; Gabur et al. 2020), molecular cytogenetics (Xiong et al. 2011), optical mapping (Gabur et al. 2020) and genome-wide short-read sequencing (Chalhoub et al. 2014; Samans et al. 2017), but all of these methods are only capable of accurately detecting large-scale SV (> 100,000 bp; Samans et al. 2017). However, using long-read sequencing, we recently demonstrated that small-scale SV events (30–10,000 bp) are also unexpectedly widespread in natural *B. napus* (Chawla et al. 2020), with considerable implications for gene function and diversity. Sequence-capture experiments demonstrated a major impact of SV on flowering-time regulatory pathway genes (Schiessl et al. 2017a, b), which play a key role in crop adaptation and yield performance. This study aims to capture the relevance of genome-wide SV in elite European winter oilseed rape and investigate the specific impact of SV on flowering time in a narrow breeding gene pool.

Facing climate change and the trend toward increasingly warmer winters in Europe, in conjunction with an earlier onset of vegetative growth after winter, flowering time is a crucial factor for plant development, seed formation and ultimately yield production. However, flowering time is determined by many genes, which complicates its examination (Blümel et al. 2015). Previous studies on flowering time in *B. napus* focused primarily on identification of major genetic factors responsible for differentiation of winter, semi-winter and spring morphotypes (Long et al. 2007; Raman et al. 2019; Schiessl et al. 2017b, 2019; Song et al. 2020; Wu et al. 2019; Xu et al. 2016). Due to the strong effects of major mutations in key flowering-time regulators, which are jointly responsible for ecogeographical differentiation, minor effects of allelic differentiation tend to be obscured in such studies. For breeding, however, where fine-tuning of flowering within ecotypes is critical, dissection of minor-effect loci in elite breeding gene pools can be more relevant than major-effect loci which disturb local adaptation. Therefore, in this study we deliberately carried out an analysis of flowering-time variation in elite winter cultivars, aiming to find novel genomic regions for fine adjustment of flowering modulation. We first investigated SV events associated with flowering-time variation in a double haploid (DH) multiparental association mapping population derived from crossing six inbred lines from elite commercial cultivars to a common parent (also a modern elite cultivar). Single-locus assays of interesting SV were further assessed for their impact on flowering time in a winter-type *B. napus* diversity set. The results revealed associations of SV with the onset of flowering in a number of flowering-time genes within QTL regions in elite oilseed rape. The findings can help to breed for adapted oilseed rape cultivars that fulfill current and future requirements in the course of climatic changes.

Material and methods

Plant material

A multiparental population comprising 354 DH lines was used in this study (Supplementary Table S1 and Supplementary Fig. 1). The panel consists of six subpopulations derived by crossing six elite founder lines ('Adriana', 'Alpaga', 'DK Cabernet', 'Galileo', 'King 10′ and the DH line 'JN') to the common elite parent 'Lorenz'. DH families from the crosses King 10 × Lorenz, Adriana × Lorenz, and JN × Lorenz were produced by NPZ Innovation GmbH (Holtsee, Germany). DH families from the crosses Lorenz × Alpaga and Lorenz × DK Cabernet were produced by Syngenta Seeds GmbH (Bad Salzuflen, Germany), while the family from the cross Lorenz × Galileo was produced by KWS SAAT SE (Einbeck, Germany). All six subpopulations comprised 60 DH lines except for the cross Lorenz × Galileo, which comprised 54 DH lines. In addition, we also investigated a set of 140 genetically diverse inbred lines of winter-type oilseed rape selected from the ERANET-ASSYST *B. napus* diversity set described in Körber et al. (2012) (Supplementary Table S2).

Field trials and phenotyping data

Field trials were conducted in an alpha lattice design with two replications at five independent locations across Germany in the 2017/2018 and 2018/2019 growing seasons. Plants were grown in plots of 6 m^2 (4 × 1.5 m) with a sowing density of 45 seeds per m^2. In 2017/18, the trials were located in Hadmersleben (Syngenta Seeds GmbH) and Nienstädt (Monsanto Agrar Deutschland GmbH), while in 2018/19 trials were carried out in Hovedissen (W. von Borries-Eckendorf GmbH & Co. KG), Leutewitz (Deutsche Saatveredelung AG) and Soerup (Bayer CropScience AG). In addition, we accessed flowering-time data from Schiessl et al. (2015) for 140 winter-type lines of the ERANET-ASSYST diversity set assessed over a total of 13 environments from 2010 to 2013, including flowering-time data from Giessen (Germany) in 2011 and 2012, Gross Gerau (Germany) in 2010, 2011 and 2012, Beibei (China) and Temuco (Chile) in 2011 and 2012, and Rauischholzhausen (Germany) in 2010, 2011, 2012 and 2013. We further assessed the diversity set for flowering time in a randomized complete block design with a plot size of 12.5 m^2 (10 × 1.25 m) in one replicate in

Rauischholzhausen (Germany) in 2018. Start of flowering was defined as BBCH 61 (10% of flowers on main raceme open) (Lancashire et al. 1991) and is reported in all trials as the number of days after sowing.

SNP genotyping and linkage analysis

The entire population including the seven parents was genotyped using the *Brassica* 60 k Illumina Infinium™ SNP array. By allowing zero mismatches and gaps, a stringent alignment using BLASTN (Altschul et al. 1990) was conducted and we were able to anchor 27,832 SNP uniquely to a single position of the publicly available *B. napus* genome assembly Darmor-*bzh* v4.1 (Chalhoub et al. 2014). Failed and nonspecific SNP were excluded from further analysis. Using the software Haploview, we defined linkage disequilibrium (LD) blocks based on their confidence intervals according to Gabriel et al. (2002).

Genome-wide association studies (GWAS)

The R package GenABEL (Aulchenko et al. 2007) was used to conduct a genome-wide association study (GWAS). Markers with more than 10% missing data or a minor allele frequency lower than 5% were excluded, as were genotypes with more than 10% missing data. The kinship and principal component analysis were included in the model to adjust for population stratification. We applied a false discovery rate (FDR) of ≤ 0.1 (Benjamini and Hochberg 1995) to call a marker–trait association significant. In addition, to reduce the type II error rate, a threshold of $-\log_{10}(p \text{ value}) \geq 3.0$ for marker–trait associations was defined to consider a marker as a putative SNP. Phenotypic variance (R^2) was estimated using the formula $R^2 = \frac{\chi}{n-2+\chi}$. As shown by Gabur et al. (2018), calling of single nucleotide absence polymorphisms (SNaP) by segregation patterns can provide additional markers and reveal previously undetected QTL in *B. napus*. Using this procedure, we called SNaP markers for loci that failed the threshold of < 10% failed calls but showed the expected 1:1 segregation for presence–absence polymorphisms in the respective subfamilies.

Whole genome sequencing and variant calling

Long-read sequencing on the MinION device from Oxford Nanopore Technologies (ONT) (Oxford, UK) was used to sequence genomic DNA from each parental line to a genome coverage of at least 20×. For this purpose, high molecular weight (HMW) DNA isolation was performed using a protocol modified from Mayjonade et al. (2016), as described in Chawla et al. (2020). In order to achieve a longer average read length, we conducted a size selection step using the Circulomics Short Read Eliminator Kit (https://www.circulomics.com/sre) prior to the sequencing library preparation. The kit consists of a buffer that separates small DNA fragments by centrifugation. The size-selected HMW-DNA was subsequently used to prepare the sequencing library. We utilized the SQK-LSK109 sequencing kit provided by Oxford Nanopore Technologies using their recommended protocol. Library preparation consisted of the following two steps: end repair and adapter ligation, each followed by a cleaning step with magnetic beads. Finally, a library of 500–1000 ng of HMW-DNA was transferred into the MinION flow cell for sequencing. MinION raw signal data were processed using the base-caller Guppy version 3.2.1 (https://community.nanoporetech.com/downloads/guppy/release_notes). Read quality was evaluated with the tool MinIONQC (Lanfear et al. 2019), and raw reads with a minimum Q score of 7 were aligned to the *B. napus* reference genome Darmor-*bzh* v4.1 (Chalhoub et al. 2014) using the NGMLR long-read mapper version 0.2.7 (Sedlazeck et al. 2018). Subsequently, the alignment file in SAM format was converted to a sorted BAM file using SAMtools version 1.9 (Li et al. 2009) and genome-wide SV was called using Sniffles version 1.0.10 (Sedlazeck et al. 2018). Settings and subsequent filtering of SV followed Chawla et al. (2020) with a focus only on deletions and insertions. Bedtools version 2.29.2 was used to detect SV overlapping with genes (Quinlan et al. 2010). Using the Integrative Genomics Viewer version 2.6.1, we produced coverage plots and illustrated SV and genes (Robinson et al. 2011). The segregation ratio of SV was tested conducting a χ^2 test. The online database JASPAR was used to screen for transcription factor binding sites (http://jaspar.genereg.net/) (Khan et al. 2018).

SV validation

To validate interesting SV events detected by the variant calling, we used a standard PCR approach using primers flanking the deletions/insertions. Primer design was performed using Primer3Plus (Untergasser et al. 2007) and BLAT (Kent 2002) via the *Brassica napus* Genome Browser (http://www.genoscope.cns.fr/brassicanapus). All primer sequences can be found in Supplementary Table S3. Sanger sequencing of PCR products was conducted to confirm unique primer binding. Besides PCR, we confirmed the 288 bp deletion in *FLOWERING LOCUS T* on chromosome A02 (*BnaFT.A02*; *BnaA02g12130D*) using 100 bp single-end Illumina short read sequencing data from 140 inbred winter-type lines of the diversity set (Schiessl et al. 2017b). Short reads from genotypes carrying the deletion fail to map accurately to *BnaFT.A02* in the Darmor-*bzh* version v4.1 reference genome due to the deleted 288 bp and 92 unresolved nucleotides (Ns) within the reference. Therefore, we modified the reference by adding a corrected version of

BnaFT.A02 using long read data from cultivar Express 617 (Lee et al. 2020), once with and once without the 288 bp sequence. Alignment, removal of duplicates, sorting, indexing and normalized mean coverage (NMC) calling for the 288 bp position were performed according to Schiessl et al. (2017c). An NMC value below 0.5 indicated association with a confirmed deletion.

Results

GWAS for identification of flowering-time QTL

A total of 13,746 high-quality polymorphic SNP markers were identified in the mapping population. Additionally, 578 SNaP markers were called by analyzing the segregation patterns of raw array data in the subpopulations, as described in Gabur et al. (2018). Using all 14,324 markers, eight genomic regions associated with flowering time in winter oilseed rape were detected in at least two of five field environments (Fig. 1 and Supplementary Table S4). Four of the detected QTL were located on chromosome A02, in LD blocks with sizes of 352 kb, 498 kb, 451 kb and 383 kb. Three QTL were identified on chromosome C02 in LD blocks of 331 kb, 860 kb and 360 kb, respectively, while a 914 kb QTL interval was detected at the distal end of chromosome C03. Broad sense heritability (Falconer and Mackay 1996) for the onset of flowering was high ($h^2 = 0.83$).

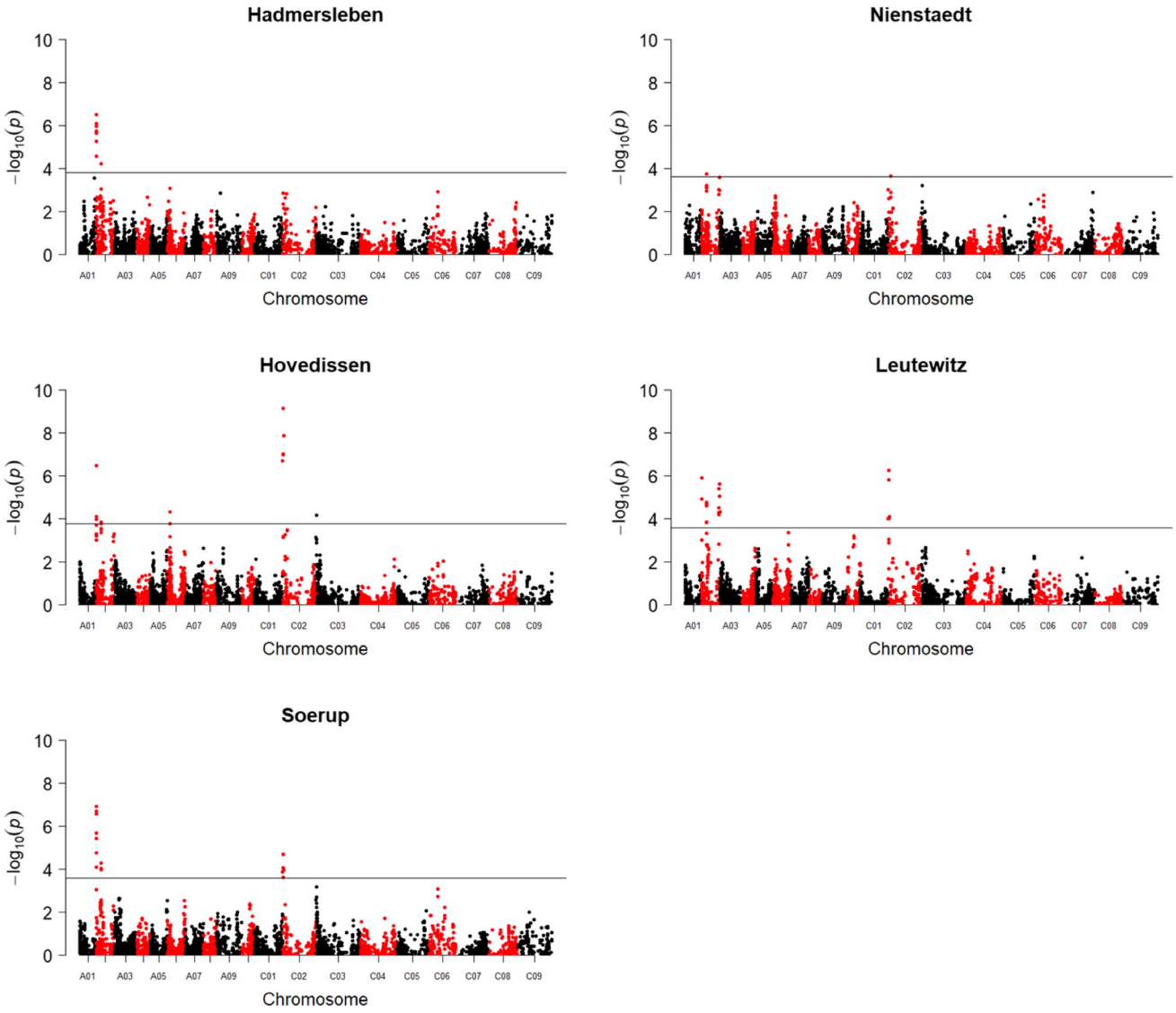

Fig. 1 Manhattan plots showing marker–trait associations for flowering time (days after sowing) in the multiparental population in five independent environments across Germany ($n = 352$). The solid line represents the threshold for significant SNP markers (FDR ≤ 0.1)

Genome-wide structural variant detection in seven elite parental lines using long-read sequencing

A total of 250 Gb of raw sequencing data was generated from 29 MinION flowcells, with an average yield of 35.7 Gb per genotype. A genome coverage of 23 × to 43 × was achieved for the seven parents of the mapping population. The first 20 flowcell runs delivered average N50 values of 21,967 bp. After using the Circulomics Short Read Eliminator Kit for removing short reads from the library in the following nine sequencing runs, the N50 increased to an average of 30,526 bp, so that the final N50 of the cumulative data from all flowcells ranged from 20,418 to 29,260 bp for all seven genotypes. A summary of sequencing results including read numbers, read lengths and the quality of the reads can be found in Supplementary Table S5. Mapping rates using the Darmor-*bzh* v4.1 reference genome ranged from 63 to 72% for the seven parental lines (Supplementary Table S6). Variant calling revealed a total of 50,762 unique SV across the seven parental lines compared to Darmor-*bzh*, of which 13,374 were polymorphic across the seven parental lines. Conversely, 37,388 SV calls were consistent for all seven parental lines and therefore monomorphic in the population, but different to the Darmor-*bzh* reference; these calls likely include assembly errors of Darmor-*bzh*. Of the polymorphic SV events we detected, 6,990 were deletions and 6,384 were insertions, with sizes ranging from 30 bp to 26 kb. The majority (47.5%) had a size between 100 and 1,000 bp (Fig. 2). Median sizes for deletions and insertions were 429 bp and 263 bp, respectively. All detected polymorphic deletions and insertions were included in subsequent analyses.

Genes affected by SV within flowering-time QTL

To explore the potential impact of the genome-wide SV on protein coding genes, we firstly compared the positions of the detected SV with the gene regions for 101,040 genes annotated in the Darmor-*bzh* reference genome (Chalhoub et al. 2014). In total, 4,840 (4.8%) of all annotated genes

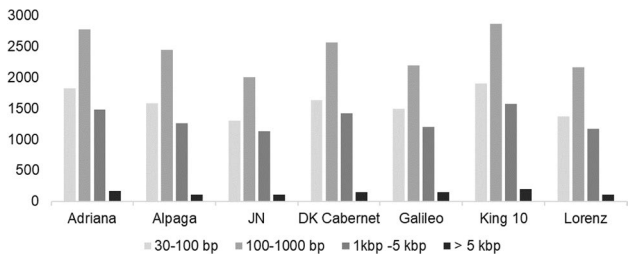

Fig. 2 Total numbers of polymorphic SV detected in seven parental lines by comparison to the *B. napus* reference Darmor-*bzh* v4.1 grouped in four different size classes

in the genome coincided with polymorphic SV events in the parental lines of the mapping population. A similar frequency of genes coinciding with SV events was observed within the eight flowering-time QTL (6.7%, 52 of 771). To investigate whether flowering time might be affected by intragenic SV, we analyzed the genes within QTL regions in more detail. In total, we found 771 annotated genes contained within the eight QTL regions, spanning 1.684 Mb on chromosome A02, 1.551 Mb on chromosome C02 and 914 kb on chromosome C03. By filtering gene ontology (GO) terms of the *B. napus* gene annotations on these 771 genes for terms that can be linked to flowering time, we identified 71 genes annotated to be involved in flowering-time regulation, among them copies of the vernalization gene *FLOWERING LOCUS C* (*FLC*) and the circadian rhythm gene *LATE ELONGATED HYPOCOTYL* (*LHY*) (Table 1 and Supplementary Table S7). Six out of these 71 genes exhibit SV. Of these six genes, five were detected in the QTL regions on chromosome A02 and one in the QTL region on chromosome C02. The affected gene on chromosome C02 is not homoeologous to any of the five flowering-related genes on chromosome A02, and the QTL regions on A02 and C02 do not represent homoeologous regions based on comparison of harbored genes as defined by Chalhoub et al. (2014). By analyzing regions up to 10 kb upstream of the start codons of the flowering-time genes, we found a large 1,361 bp deletion in the parental line JN in the promoter region of *BnaFT.A02* (*BnaA02g12130D*) (Table 1). In three of the seven parental lines, the same gene also displayed a 288 bp deletion in the second intron.

Association of intragenic deletions with flowering time in the multiparental population

To evaluate trait associations with deletions in flowering-time genes, we developed specific PCR assays for all seven SV events in the six QTL-linked flowering-time candidate genes and used these to analyze the segregation of size polymorphisms identified in the parental lines in relevant subpopulations (Fig. 3, Supplementary Table S3). The seven PCR assays were first validated in the seven parental lines of the multiparental population. In 47 of these 49 PCR assays, the alleles called by ONT data from the parental lines were reconfirmed, whereas a 38 bp insertion, designated SV7, could not be reconfirmed in parental lines King 10 and Galileo and was excluded from further analyses. The PCR assays were then applied to the respective subpopulations which showed the corresponding polymorphism and tested for the expected 1:1 segregation. In most cases, the observed segregation in the subfamily did not differ significantly from the expected 1:1 segregation ratio (Table 2). However, three SV did not fit the expected segregation in 1 (from 3), 2 (from 5) and 2 (from 5) polymorphic subpopulations, respectively,

Table 1 Flowering-time candidate genes and coinciding SV identified between 7 elite parental lines within QTL for flowering time analyzed in a *B. napus* multiparental elite population ($n = 354$)

QTL name	Size of LD block in kb	Start–end position	Number of genes/ flowering-time genes in Darmor-*bzh*	Number of genes/ flowering-time genes with structural variants*	Gene IDs	Blast2GO annotation of gene	Name: size and type of SV (relative to Darmor-*bzh*)	PCR assay name: product size polymorphism (see Fig. 3)
A02_1	352	12,297–364,239	92/10 (10.9%)	14 / 2 (14.2%)	*BnaA02g00160D*; *BnaA02g00910D*	Pollen development; regulation of flower development	SV1: 373 bp deletion; SV2: 64 bp deletion	P133: 1130 bp/757 bp; P255: 419 bp/362 bp
A02_2	498	6,311,854–6,810,319	59/8 (13.6%)	6/1 (16.7%)	Promoter of *BnaFT.A02*; *BnaA02g12130D* (*BnaFT.A02*)	-; photoperiodism, flowering, positive regulation of flower development	SV3: 1,361 bp deletion; SV4: 288 bp deletion	P311: 1924 bp/563 bp; P416: 797 bp/519 bp
A02_3	451	22,569,685–23,021,163	73/3 (4.1%)	2/0 (0%)	–	–	–	–
A02_4	383	24,112,932–24,495,807	72/11 (15.3%)	5 / 2 (40%)	*BnaA02g33650D*; *BnaA02g33940D*	Circadian rhythm; entrainment of circadian clock	SV5: 1,313 bp deletion; SV6: 35 bp deletion	P513: 819 bp/no product; P632: 667 bp/632 bp
C02_1	331	512,169–842,796	70/0 (0%)	2/0 (0%)	–	–	–	–
C02_2	860	1,028,042–1,888,045	153/11 (7.2%)	2/1 (50%)	*BnaC02g03640D*	Regulation of flower development	SV7: 38 bp insertion	P732: 262 bp/300 bp
C02_3	360	3,168,965–3,528,564	62/5 (8.1%)	10/0 (0%)	–	–	–	–
C03_1	914	2,434–915,986	190/23(12.1%)	11/0 (0%)	–	–	–	–
Sum	4,149	-	771/71 (9.2%)	52/6 (11.5%)	–	–	–	–

Flowering-time genes were defined according to Schiessl et al. (2015) as genes associated by Blast2GO with at least one of the following gene ontology terms: flower, vernalization, photoperiod, circadian, floral, vegetative to reproductive, vegetative phase change, pollen carpel, sepal, petal

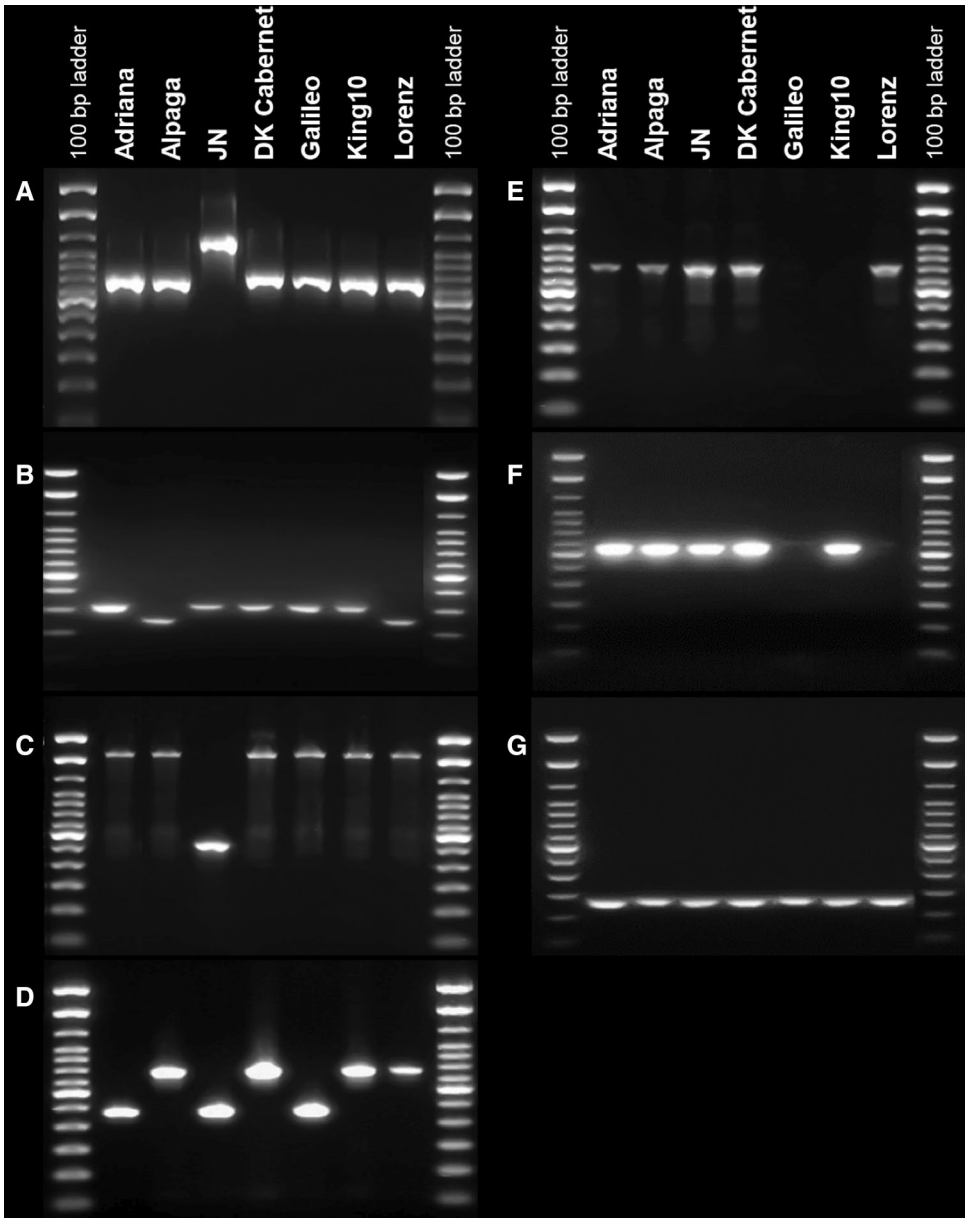

Fig. 3 Size polymorphism within six flowering-time genes from three QTL on A02 and one QTL on C02 detected by PCR in seven parents of the *B. napus* multiparental population with the common parent Lorenz. **a** SV1: 373 bp deletion, **b** SV2: 64 bp deletion, **c** SV3: 1,361 bp deletion, **d** SV4: 288 bp deletion, **e** SV5: 1,313 bp deletion, **f** SV6: 35 bp deletion, **g** SV7: 38 bp insertion (also see Tables 1 and 2 for more details)

Table 2 Segregation ratios of six SV events coinciding with five flowering-time candidate genes within the investigated QTL regions of the multiparental population ($n = 354$)

SV type		SV size [in bp]	Lorenz x					
			Adriana	Alpaga	JN	DK Cabernet	Galileo	King 10
SV1	Deletion	373	m	m	29:31 (0.796)	m	m	m
SV2	Deletion	64	32:28 (0.606)	m	5:54 (0.000)***	28:32 (0.606)	36:18 (0.014)*	29:31 (0.796)
SV3	Deletion	1361	m	m	31:29 (0.796)	m	m	m
SV4	Deletion	288	43:17 (0.001)***	m	31:29 (0.796)	m	29:25 (0.586)	m
SV5	Deletion	1313	m	m	m	m	34:20 (0.057)	34:26 (0.302)
SV6	Deletion	35	29:31 (0.796)	39:21 (0.020)*	42:17 (0.001)**	30:30 (1.000)	m	28:32 (0.606)

Detailed descriptions of the SV events can be found in Table 3. Numbers in brackets are the probabilities ($P[\chi^2]$) that the observed data fit the expected 1:1 segregation pattern. Asterisks show level of significance based on χ^2 test (*p value < 0.05, **p value < 0.01, ***p value < 0.001). m: monomorphic, not tested in this subpopulation

but did fit the expected segregation in the remaining subpopulations. To understand the segregation distortions that were observed for several SV, we sequenced the PCR products. Sanger sequencing resulted in clear chromatograms, excluding the possibility of nonspecific primer binding. Particularly for allotetraploid *B. napus*, it is well known that segregation distortion in certain genome regions can occur within the process of DH production (Pilet et al. 2001; Zhao et al. 2005), either due to linkage with loci impacting DH regeneration or due to homoeologous exchanges in the F1 parent. Linked SNP markers showed the same patterns of segregation distortion, which additionally confirms a biological cause and excludes the possibility of technical issues leading to the observed distortion. However, the 60 DH lines of the subpopulations were produced from up to five single F1 plants. Thus, it is possible that the segregation distortion could be due to de novo SV in individual F1 plants. The SV calls were included in GWAS and allelic effects were calculated (Table 3). For SV4, we detected a significant association with flowering time in the multiparental population in all five locations. Effects varied between 0.38 and 1.42 days earlier onset of flowering for genotypes carrying the deletion. Interestingly, a significant effect of the SV1 deletion on flowering time (0.84–1.62 days earlier flowering) was detected in only three out of the five locations (Hadmersleben, Hovedissen and Soerup), whereas the deletion allele of SV5 showed a significant effect (0.52–1.91 days earlier flowering) only in the other two locations (Leutewitz and Nienstaedt).

A novel 288 bp deletion in *BnaFT.A02* associates with flowering time

The 288 bp deletion within the second intron of *BnaFT.A02* (SV4) was investigated in further detail because *FT* is a well described key regulator in the flowering-time pathway (Guo et al. 2014; Helliwell et al. 2006; Schiessl et al. 2017c, b; Srikanth et al. 2011) and this specific deletion has not been described before. Validation by Sanger sequencing confirmed that three of the seven parental inbred lines (Adriana, JN, Galileo) carry the 288 bp deletion, whereas four, including the common parent Lorenz, are identical to the reference genome in this gene region. A PCR marker spanning the deletion was used to genotype the entire multiparental population for the SV polymorphism (Supplementary Table S8). Adding this PCR marker as an additional marker to the GWAS revealed a significant association (LOD score = 4.57) explaining 5.64% (R^2) of the phenotypic variation. On average, we observed a slight, but significant earlier onset of flowering (effect of deletion = -−0.88 days) for genotypes having this deletion (Fig. 4). However, at location Leutewitz a stronger effect (− 1.42 days) was observed (Table 3). We screened the 288 bp sequence for transcription factor binding sites using JASPAR (Khan et al. 2018).

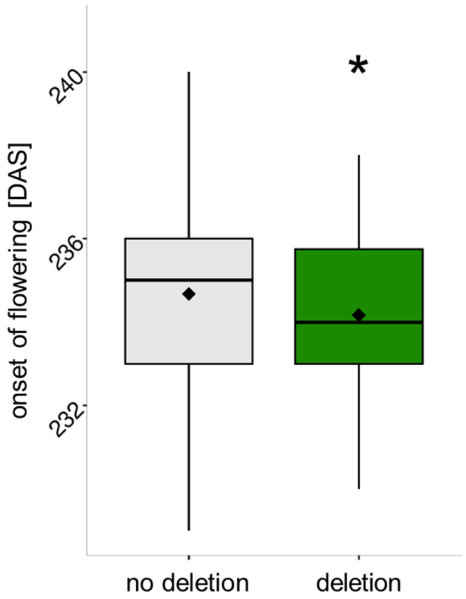

Fig. 4 Two boxplots visualize the difference for onset of flowering (days after sowing) depending on the presence/absence of the 288 bp sequence in intron 2 of the gene *BnaFT.A02*. Lines carrying the deletion show an on average earlier onset of flowering of about one day (effect of deletion = − 0.88) in the *B. napus* multiparental population ($n = 354$). Significant difference is highlighted with green color and asterisks (*p value < 0.05, **p value < 0.01, ***p value < 0.001)

Table 3 Allelic effects of the six investigated SV events coinciding with five flowering-time candidate genes from the QTL in the multiparental population ($n = 354$)

SV	SV type	Hadmersleben		Hovedissen		Leutewitz		Nienstaedt		Soerup	
		Effect	R^2	Effect	R^2	Effect	R^2	Effect	R^2	Effect	R^2
1	Deletion	−0.84	7.44	−1.62	4.85	ns	–	ns	–	−1.47	8.20
2	Deletion	ns	–	ns	–	ns	–	ns	–	ns	–
3	Deletion	ns	–	ns	–	ns	–	–	–	ns	–
4	Deletion	−0.38	3.63	−1.01	4.06	−1.42	3.99	−0.51	3.48	−0.80	4.84
5	Deletion	ns	–	ns	–	−1.91	6.25	−0.52	3.55	ns	–
6	Deletion	ns	–	ns	–	ns	–	ns	–	ns	–

Effect numbers describe the change in number of days to onset of flowering. R^2: phenotypic variation (%) explained by the respective SV. ns: nonsignificant

The analysis revealed binding sites for the *CIRCADIAN CLOCK-ASSOCIATED 1* (*CCA1*) and *LHY* along with several members of the REVEILLE family of transcription factors (Supplementary Table S9), suggesting that this region in the second intron of *BnaFT.A02* may have functional relevance for flowering-time modulation. None of the five other known copies of *BnaFT* in the *B. napus* genome associated to flowering-time QTL, suggesting that they are not responsible for flowering-time variation in this population (Supplementary Table S10).

Distribution of the 288 bp deletion in genetically diverse winter-type *B. napus*

The relevance of the 288 bp deletion in *BnaFT.A02* for flowering-time modulation in winter-type oilseed rape was investigated in a selection of accessions from the ERANET-ASSYST *B. napus* diversity set for which comprehensive flowering-time data were already available (Schiessl et al. 2014, 2015, 2017b). We re-analyzed Illumina short-read data produced by Schiessl et al. (2015, 2017b) from sequence capture of flowering-pathway genes, because the 288 bp deletion was not detectable in the previous analysis, where the Darmor-*bzh* v4.1 reference genome was used for alignment. However, the modified reference sequence without these 288 bp in *BnaFT.A02* allowed us to accurately map the deletion in the diversity set using the short-read sequence capture data (Supplementary Table S11). This approach confirmed the authenticity of the deletion and revealed a widespread distribution in winter-type oilseed *B. napus* cultivars. In addition, the use of short reads enabled us to confirm the borders of the deletion. Of 140 winter accessions used in this study, 34 were found to carry the deletion (NMC < 0.5), whereas 91 were similar to the reference genome (NMC > 1.5). For 15 accessions, analysis of short-read sequencing data coverage and relative copy number calculation did not reveal a clear result (NMC 0.5–1.5). We validated the approach based on short-read sequencing coverage analysis by conducting a PCR on all accessions using the same primers as described above (assay P416, Supplementary Table S3). We confirmed 32 of the 34 NMC deletion calls, whereas two accessions showed contradictory results. Out of 91 accessions without NMC deletion calls, 85 were confirmed. Two showed contradictory patterns, and four accessions were found to be heterozygous. The remaining 15 accessions include three carrying the deletion, eight without the deletion and four which were revealed to be heterozygous. This shows that relative coverage analysis based on Illumina short read data fails to accurately call a 288 bp deletion using 100 bp single-read data. For each of the 14 independent environments, a Student's *t* test was performed to calculate the phenotypic effect of the deletion assessed using the PCR assay. In this germplasm collection, a significant association of the deletion with flowering time was found only for the locations in China and Chile in 2012, but not for the German locations. However, a trend toward early flowering associated with the deletion was observable in almost all environments (Fig. 5).

Discussion

Flowering time is an important agronomical trait that has been investigated intensively using QTL analyses and GWAS in oilseed rape. However, most studies on flowering time in *B. napus* investigated germplasm sets with broad genetic diversity including spring, semi-winter and winter-type accessions and focused on major effect QTL. Multiple flowering-time QTL with major effects and high heritabilities have been mapped, and key genes from the flowering and vernalization pathways were repeatedly implicated (Long et al. 2007; Raman et al. 2019; Song et al. 2020; Wu et al. 2019; Xu et al. 2016). In contrast, this study focused on identification of minor QTL which have not already been fixed in adapted elite breeding pools. This was achieved using a multiparental mapping population whose founders were exclusively modern, elite, European winter-type *B. napus* accessions. The aim was to identify small-effect polymorphisms relevant for modulation of flowering time in elite, adapted, winter oilseed rape breeding material. Given this fundamental difference in approach, it is not surprising that our QTL results barely overlap with previous findings. Major differences separating the three major ecotypes (spring, semi-winter and winter types) tend to mask smaller effects responsible for intra-ecotype variation. Some of the significant marker–trait associations for flowering time in the present work were stable only over a limited number of environments, suggesting that QTL effects are minor and may interact with specific environmental factors. Using 158 winter-type accessions, Schiessl et al. (2015) also revealed flowering-time QTL that are highly sensitive to local environments. In that study, none of the central flowering regulators (e.g., *Bna.FT*, *CONSTANS*, *GIGANTEA*) or central vernalization genes (e.g., *Bna.FLC*) was located within QTL regions for the onset of flowering in winter-type *B. napus*. The QTL found by Schiessl et al. (2015) also did not overlap with the QTL we found in our elite winter-type population. This indicates that fine regulation of the onset of flowering in winter-type *B. napus* is environmentally dependent and potentially controlled by multiple different genome regions with small effects.

In contrast to a previous report for disease resistance (Gabur et al. 2018, 2020), inclusion of SNaP markers (array-based presence/absence markers) in the QTL analysis did not result in detection of additional QTL associated with flowering time. This might suggest that SV may not

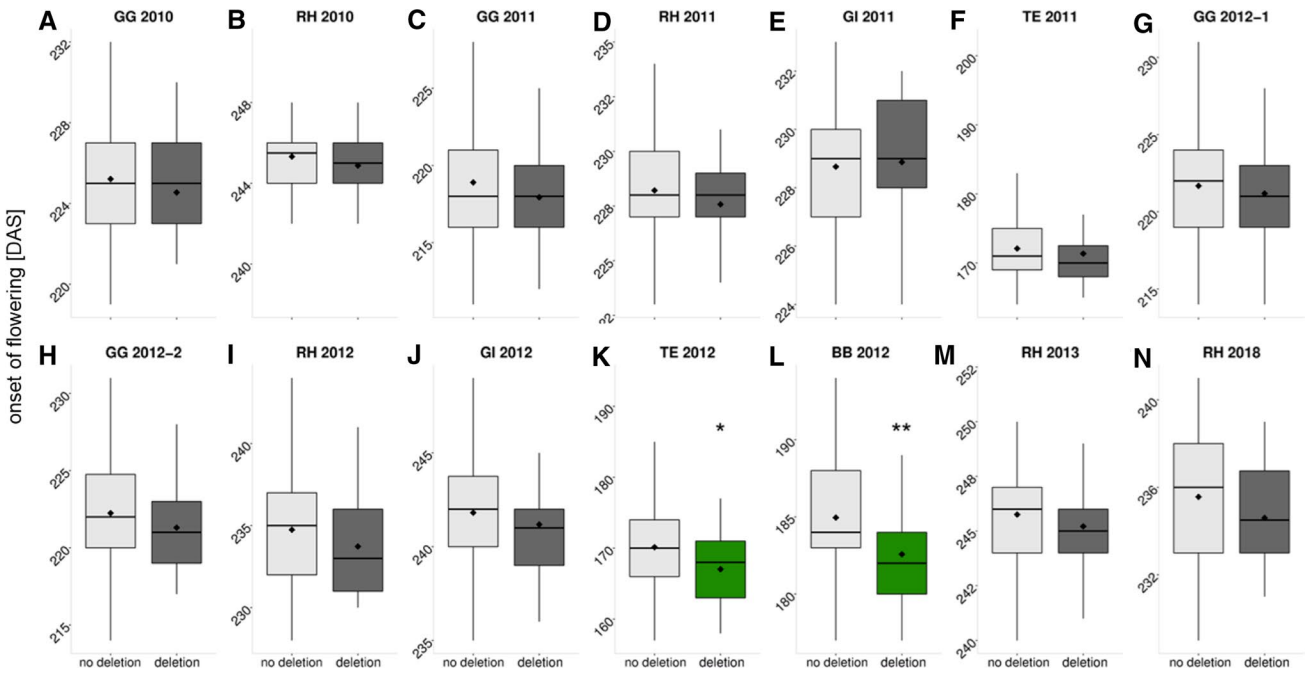

Fig. 5 The boxplots visualize the difference in the onset of flowering (days after sowing) depending on the presence/absence of the 288 bp sequence in intron 2 of the gene *BnaFT.A02* in 14 environments for the ERANET-ASSYST diversity set. The environments are abbreviated as follows: Germany: Giessen (GI), Gross Gerau (GG), Rauischholzhausen (RH); Chile: Temuco (TE); China: Beibei, Chongqing (BB), with their respective year of harvest. Black diamonds represent the mean values. Significant differences tested with Student's *t* test are highlighted with green color and asterisks (**p* value < 0.05, ***p* value < 0.01, ****p* value < 0.001) ($n = 140$)

strongly impact flowering-time modulation in elite winter oilseed rape cultivars. However, since we showed associations between SV and flowering-time modulation, it appears more likely that SV events associated with flowering traits are in LD to neighboring SNP markers from the 60 k *Brassica* SNP array in these elite mapping parents. The number of polymorphic SV events we detected among the seven elite parental lines (13,374) is similar to the number of polymorphic SNP detected between these lines using the 60 k *Brassica* SNP array (13,746), suggesting that structural variation is a significant source of sequence polymorphism in elite cultivars with a potentially very important functional role. We found that around 5% of genes are affected by small to mid-scale SV events in each of the seven elite parents, corresponding to the frequency of intragenic SV reported by Chawla et al. (2020) for older winter oilseed rape cultivars. This underlines the importance of functional gene modification during post-polyploidization genome restructuring in *B. napus* (Chawla et al. 2020; Song et al. 2020). The high level of functionally relevant SV polymorphism in modern elite cultivars underlines the high relevance of this kind of variant for selection and breeding and the need to develop a better platform than the 60 k *Brassica* SNP array for population-wide high-throughput genotyping and identification of SV.

Most of the deletions/insertions we identified in candidate genes within flowering-time QTL in elite cultivars reside in introns. With respect to *BnaFT.A02* (*BnaA02g12130D*), one deletion lies in the second intron and another in the putative promoter region. The 1.3 kb deletion in the promoter region of *BnaFT.A02* is associated with the different *B. napus* ecotypes (Chawla et al. 2020), but does not completely explain the differentiation between ecotypes (Schiessl et al. 2019). Because the populations we investigated are exclusively vernalization-dependent, this deletion did not show significant associations with flowering time. In contrast, SV1 (a 373 bp deletion in *BnaA02g00160D*), SV5 (a 1,313 bp deletion in *BnaA02g33650D*) and SV4 (a 288 bp deletion in *BnaFT.A02*) showed significant associations with flowering time. Whereas a direct functional relevance of *BnaFT.A02* is already well established, further functional studies are required to elucidate the potential role of the *B. napus* genes *BnaA02g001160D* and *BnaA02g33650D* in coordination of flowering time. *BnaA02g00160D* is a homolog of *A. thaliana SETDOMAIN GROUP 15* (*SDG15*; *AT5G09790*), a H3K27 methyltransferase which was previously described to be involved in the control of transposal activity and genome stability (Ma et al. 2018). Other methyltransferases have described roles in flowering-time regulation, like *SDG7* (Lee et al. 2015), while *SDG15* was only implicated in pollen formation so far (Raynaud et al. 2006) and is therefore considered less likely to directly

influence flowering time here. *BnaA02g33650D* is a copy of *BTB AND TAZ DOMAIN PROTEIN 1* (*BT1*), which is known to plan a role in both male and female gametophyte development (Robert et al. 2009). Hence, this gene is expected to be involved in flower development, but not necessarily in flowering-time regulation, which takes place sometime prior to flower development.

Schiessl et al. (2015) did not detect the QTL harboring *BnaFT.A02* by GWAS using markers from the 60 k *Brassica* SNP array. This might be due to a strong LD decay around SV events or the lower number of accessions and the population structure of the diversity set used in this study. Moreover, the failure to detect a 288 bp deletion using a sequence capture strategy based on short single-read 100 bp Illumina sequencing (Schiessl et al. 2017b) clearly reveals the advantage of long-read sequencing to detect polymorphic small- to mid-size deletions. The 288 bp deletion in *BnaFT.A02* is associated with flowering time at all five field locations in Germany in both the 2017/2018 and 2018/2019 growing seasons. Both years were characterized by extremely warm and dry summers in Germany, in strong contrast to the environmental conditions in the years 2010–2013 during the study of Schiessl et al. (2015). This may suggest a QTL-by-environment interaction that is more active under extreme environments, which in turn might explain why day length is only crucial in the diversity set.

FT is one of the key regulators of flowering time in the well-investigated flowering-time pathway (Helliwell et al. 2006; Srikanth et al. 2011). Several studies, particularly in Arabidopsis, elucidated interactions of *FT* with other major genes in this pathway (Faure et al. 2007; Turck et al. 2008). For instance, the single copy of *FT* in Arabidopsis is suppressed by *FLC*, which is downregulated during extended periods of cold temperature to unlock the repression (Helliwell et al. 2006; Searle et al. 2006). However, the allopolyploid *B. napus* genome possesses multiple *FT* orthologs and paralogs with incompletely elucidated function. For example, *BnaFT.A02* is most likely not regulated by *Bna.FLC*. Instead, Wang et al. (2009) showed that the characteristic CArG box to which *FLC* normally binds is missing in this particular *FT* copy, and Guo et al. (2014) demonstrated that *BnaFT.A02* is expressed independently of vernalization. On the other hand, *FT* expression is activated by *CONSTANS* (*CO*), whose activity is promoted by an increase in photoperiod (Kobayashi et al. 1999; Srikanth et al. 2011; Turck et al. 2008; Wigge et al. 2005). In agreement with Schiessl et al. (2015), who demonstrated the day length dependency of *BnaFT.A02*, we found that a deletion polymorphism in *BnaFT.A02* in winter-type *B. napus* only associated with flowering time during trials in Chile and China, where day lengths differed significantly from the German trial locations. This supports a vernalization-independent, day-length-dependent activation of *BnaFT.A02* (Corbesier et al. 2007; Guo et al. 2014; Schiessl et al. 2015).

Expression of *FT* can be detected in leaves and apical meristems (Wigge et al. 2005). Corbesier et al. (2007) proposed the *FT* protein moves from the leaves through the phloem to the meristem. Changes in *FT* expression have been shown to significantly change flowering time in Arabidopsis (Kobayashi et al. 1999) as well as in major crops like barley (Faure et al. 2007), wheat (Yan et al. 2006), maize (Lazakis et al. 2011; Meng et al. 2011) and potato (Navarro et al. 2011). In winter-type *B. napus, Bna.FT* expression changes caused by allelic variants were associated with altered flowering time (Tudor et al. 2020), while high *FT* expression is known to correlate with earlier flowering (Raman et al. 2019). To investigate potential mechanisms by which the 288 bp deletion within the second intron of *BnaFT.A02* might regulate expression of this *FT* ortholog, we searched the sequence spanning the SV for putative regulatory motifs known to be involved in controlling gene expression. Besides the main '*FT*-interactors' *FLC* and *CO*, several other genes are already known to interact with *FT*. For example, *SCHLAFMÜTZE* (Mathieu et al. 2009), *TEMPRANILLO 1* (Castillejo et al. 2008) and *SHORT VEGETATIVE PHASE* (Kobayashi et al. 2007) repress *BnaFT.A02* expression by binding to it. Screening for transcription factor binding sites within the deleted 288 bp sequence revealed binding sites for the well-known circadian clock genes *CCA1* and *LHY*. In addition, *LHY* was found to be located in QTL C03_1. We also noted two unannotated genes with similarity to known flowering-time genes, *GA20-oxidase* (*BnaC02g01710D*) and *CONSTANS-LIKE* (*BnaC02g06280D*), within QTL C02_1 and C02_3, respectively. This suggests that this SV polymorphism may modulate *FT* expression by altering the binding characteristics to one or both of these genes, which are known to act together by the formation of dimers (Lu et al. 2009; Seo et al. 2012) and which have known involvement in photoperiod responses (Fujiwara et al. 2008; Li et al. 2008).

In the past, the prevalence of SV within genes has been mostly ignored in most crop species due to the lack of resolution of short-read resequencing. Recent breakthroughs in long-range sequencing technologies have led to increased application of these tools in recent years, particularly for whole-genome assembly (Michael et al. 2018; Schmidt et al. 2017). To a limited extent, long-read resequencing has also been applied to explore genome-wide SV patterns; however, the focus to date was generally on broad comparisons between small numbers of very diverse accessions (e.g., Chawla et al. 2020; Todesco et al. 2020), or strongly restructured genomes of mapping parents with synthetic background (Gabur et al. 2019). Associations of SV with flowering time, disease resistance and eco-geographical differentiation in diverse oilseed rape *B. napus* accessions (Chawla et al. 2020) revealed a general relevance of SV for

breeding. However, the extent, relevance and role of SV for important traits in modern elite cultivars have still not been addressed or documented in detail. Here, we show that SV in key regulator genes of flowering time is associated with flowering-time modulation in elite winter oilseed rape cultivars. The present study stresses the potential value of SV impacting important agronomic traits within a set of well-adapted and strongly selected varieties. Our results suggest that the utilization of markers developed from SV can add additional benefit to breeding programs. Our findings can be used to fine-adjust commercial varieties in terms of flowering time, an increasingly important adaptation trait in the face of climate change.

Acknowledgements We thank Regina Illgner for excellent technical help. We thank all German breeding companies organized in the GFPi (Gemeinschaft zur Förderung von Pflanzeninnovation e. V.) for production of the elite multiparental population and for performing field trials and providing phenotyping data.

Author Contribution statement CO and RS conceived the idea. PV, HSC and CO developed the methodology. CO and RS sourced the funding. SVS provided phenotype and reanalyzed Illumina sequencing data. PV performed the laboratory analysis. PV and CO carried out data curation. PV, HSC, SVS, IG, HTL and CO analyzed the data. PV and CO drafted the manuscript. PV, HSC, SVS, IG, HTL, RS and CO revised and gave further inputs to this manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. PV and CO received funding from the German Federal Ministry of Education and Research (BMBF) within the project PhomaDur (Grant Number BLE: 2818205115). This work was also supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, 031A532B).

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, Levy Y, Harel TH, Shalev-Schlosser G, Amsellem Z, Razifard H, Caicedo AL, Tieman DM, Klee H, Kirsche M, Aganezov S, Ranallo-Benavidez TR, Lemmon ZH, Kim J, Robitaille G, Kramer M, Goodwin S, McCombie WR, Hutton S, van Eck J, Gillis J, Eshed Y, Sedlazeck FJ, van der Knaap E, Schatz MC, Lippman ZB (2020) Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell 182:146–161. https://doi.org/10.1016/j.cell.2020.05.021

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aulchenko YS, de Koning D-J, Haley C (2007) Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics 177:577–585. https://doi.org/10.1534/genetics.107.075614

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300. https://doi.org/10.2307/2346101

Blümel M, Dally N, Jung C (2015) Flowering time regulation in crops—what did we learn from *Arabidopsis*? Curr Opin Biotechnol 32:121–129. https://doi.org/10.1016/j.copbio.2014.11.023

Castillejo C, Pelaz S (2008) The balance between CONSTANS and TEMPRANILLO activities determines FT expression to trigger flowering. Curr Biol 18:1338–1343. https://doi.org/10.1016/j.cub.2008.07.075

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Wang X, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker P (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345:950–953. https://doi.org/10.1126/science.1253435

Chawla HS, Lee H, Gabur I, Vollrath P, Tamilselvan-Nattar-Amutha S, Obermeier C, Schiessl SV, Song J-M, Liu K, Guo L, Parkin IAP, Snowdon RJ (2020) Long-read sequencing reveals widespread intragenic structural variants in a recent allopolyploid crop plant. Plant Biotechnol J. https://doi.org/10.1111/pbi.13456

Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, Coupland G (2007) FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. Science 316:1030–1033. https://doi.org/10.1126/science.1141752

De Coster W, De Rijk P, De Roeck A, De Pooter T, D'Hert S, Strazisar M, Kristel S, van Broeckhoven C (2019) Structural variants identified by oxford nanopore promethion sequencing of the human genome. Genome Res 29:1178–1187. https://doi.org/10.1101/gr.244939.118

Díaz A, Zikhali M, Turner AS, Isaac P, Laurie DA (2012) Copy number variation affecting the Photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*). PLoS ONE 7:e33234. https://doi.org/10.1371/journal.pone.0033234

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Prentice Hall, London

Faure S, Higgins J, Turner A, Laurie DA (2007) The *FLOWERING LOCUS T*-like Gene Family in Barley (*Hordeum*

vulgare). Genetics 176:599–609. https://doi.org/10.1534/genetics.106.069500
Feuk L, Marshall CR, Wintle RF, Scherer SW (2006) Structural variants: changing the landscape of chromosomes and design of disease studies. Hum Mol Genet 15:R57-66. https://doi.org/10.1093/hmg/ddl057
Fujiwara S, Oda A, Yoshida R, Niinuma K, Miyata K, Tomozoe Y, Tajima T, Nakagawa M, Hayashi K, Coupland G, Mizoguchi T (2008) Circadian clock proteins *LHY* and *CCA1* regulate SVP protein accumulation to control flowering in *Arabidopsis*. Plant Cell 20:2960–2971. https://doi.org/10.1105/tpc.108.061531
Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229. https://doi.org/10.1126/science.1069424
Gabur I, Chawla HS, Liu X, Kumar V, Faure S, von Tiedemann A, Jestin C, Dryzska E, Volkmann S, Breuer F, Delourme R, Snowdon R, Obermeier C (2018) Finding invisible quantitative trait loci with missing data. Plant Biotechnol J 16:2102–2112. https://doi.org/10.1111/pbi.12942
Gabur I, Chawla HS, Lopisso DT, von Tiedemann A, Snowdon RJ, Obermeier C (2020) Gene presence-absence variation associates with quantitative *Verticillium longisporum* disease resistance in *Brassica napus*. Sci Rep 10:4131. https://doi.org/10.1038/s41598-020-61228-3
Gabur I, Chawla HS, Snowdon RJ, Parkin IAP (2019) Connecting genome structural variation with complex traits in crop plants. Theor Appl Genet 132:733–750. https://doi.org/10.1007/s00122-018-3233-0
Guo Y, Harloff HJ, Jung C, Molina CM (2014) Mutations in single *FT*- and *TFL1*-paralogs of rapeseed (*Brassica napus* L.) and their impact on flowering time and yield components. Front Plant Sci 5:282. https://doi.org/10.3389/fpls.2014.00282
Helliwell CA, Wood CC, Robertson M, James Peacock W, Dennis ES (2006) The arabidopsis FLC protein interacts directly in vivo with *SOC1* and *FT* chromatin and is part of a high-molecular-weight protein complex. Plant J 46:183–192. https://doi.org/10.1111/j.1365-313X.2006.02686.x
Hurgobin B, Golicz AA, Bayer PE, Chan C-KK, Tirnaz S, Dolatabadian A, Schiessl SV, Samans B, Montenegro JD, Parkin IAP, Pires JC, Chalhoub B, King GJ, Snowdon R, Batley J, Edwards D (2018) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. Plant Biotechnol J 16:1265–1274. https://doi.org/10.1111/pbi.12867
Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12:656–664. https://doi.org/10.1101/gr.229202
Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Chèneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, Mathelier A (2018) JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res 46:D1284. https://doi.org/10.1093/nar/gkx1188
Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T (1999) A pair of related genes with antagonistic roles in mediating flowering signals. Science 286:1960–1962. https://doi.org/10.1126/science.286.5446.1960
Kobayashi Y, Weigel D (2007) Move on up, it's time for change—mobile signals controlling photoperiod-dependent flowering. Genes Dev 21:2371–2384. https://doi.org/10.1101/gad.1589007
Körber N, Wittkop B, Bus A, Friedt W, Snowdon RJ, Stich B (2012) Seedling development in a *Brassica napus* diversity set and its relationship to agronomic performance. Theor Appl Genet 125:1275–1287. https://doi.org/10.1007/s00122-012-1912-9
Lancashire PD, Bleiholder H, van den Boom T, Langelüddeke P, Stauss R, Weber E, Witzenberger A (1991) A uniform decimal code for growth stages of crops and weeds. Annal Appl Biol 119:561–601. https://doi.org/10.1111/j.1744-7348.1991.tb04895.x
Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B (2019) MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics 35:523–525. https://doi.org/10.1093/bioinformatics/bty654
Lazakis CM, Coneva V, Colasanti J (2011) *ZCN8* encodes a potential orthologue of *Arabidopsis FT* florigen that integrates both endogenous and photoperiod flowering signals in maize. J Exp Bot 62:4833–4842. https://doi.org/10.1093/jxb/err129
Lee H, Chawla HS, Obermeier C, Dreyer F, Abbadi A, Snowdon R (2020) Chromosome-scale assembly of winter oilseed rape *Brassica napus*. Front Plant Sci 496:1–13. https://doi.org/10.3389/fpls.2020.00496
Lee J, Yun J-Y, Zhao W, Shen W-H, Amasino RM (2015) *SDG7* and proper timing of vernalization. PNAS 112:2269–2274. https://doi.org/10.1073/pnas.1423585112
Li D, Liu C, Shen L, Wu Y, Chen H, Robertson M, Helliwell CA, Ito T, Meyerowitz E, Yu H (2008) A repressor complex governs the integration of flowering signals in *Arabidopsis*. Dev Cell 15:110–120. https://doi.org/10.1016/j.devcel.2008.05.002
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Long Y, Shi J, Qiu D, Li R, Zhang C, Wang J, Hou J, Zhao J, Shi L, Park B-S, Choi SR, Lim YP, Meng J (2007) Flowering time quantitative trait loci analysis of oilseed *Brassica* in multiple environments and genomewide alignment with *Arabidopsis*. Genetics 177:2433–2444. https://doi.org/10.1534/genetics.107.080705
Lu SX, Knowles SM, Andronis C, Ong MS, Tobin EM (2009) *CIRCADIAN CLOCK ASSOCIATED1* and *LATE ELONGATED HYPOCOTYL* function synergistically in the circadian clock of *Arabidopsis*. Plant Physiol 150:834–843. https://doi.org/10.1104/pp.108.133272
Ma Z, Castillo-González C, Wang Z, Sun D, Hu X, Shen X, Potok ME, Zhang X (2018) Arabidopsis serrate coordinates histone methyltransferases *ATXR5/6* and RNA processing factor *RDR6* to regulate transposon expression. Dev Cell 45:769–784. https://doi.org/10.1016/j.devcel.2018.05.023
Mathieu J, Yant LJ, Mürdter F, Küttner F, Schmid M (2009) Repression of flowering by the miR172 target *SMZ*. PLoS Biol 7:e1000148. https://doi.org/10.1371/journal.pbio.1000148
Mayjonade B, Gouzy J, Donnadieu C, Pouilly N, Marande W, Callot C, Langlade N, Muños S (2016) Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. Biotechniques 61:203–205. https://doi.org/10.2144/000114460
McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddeloh JA, Stupar RM (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. Plant Physiol 159:1295–1308. https://doi.org/10.1104/pp.112.194605
Meng X, Muszynski MG, Danilevskaya ON (2011) The *FT*-like *ZCN8* Gene functions as a floral activator and is involved in photoperiod sensitivity in maize. Plant Cell 23:942–960. https://doi.org/10.1105/tpc.110.081406
Michael TP, Jupe F, Bemm F, Motley ST, Sandoval JP, Lanz C, Loudet O, Weigel D, Ecker JR (2018) High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. Nat Commun 9:541. https://doi.org/10.1038/s41467-018-03016-2

Navarro C, Abelenda JA, Cruz-Oró E, Cuéllar CA, Tamaki S, Silva J, Shimamoto K, Prat S (2011) Control of flowering and storage organ formation in potato by *FLOWERING LOCUS T*. Nature 478:119–122. https://doi.org/10.1038/nature10431

Nishida H, Yoshida T, Kawakami K, Fujita M, Long B, Akashi Y, Laurie DA, Kato K (2013) Structural variation in the 5′ upstream region of photoperiod-insensitive alleles *Ppd-A1a* and *Ppd-B1a* identified in hexaploid wheat (*Triticum aestivum* L.), and their effect on heading time. Mol Breed 31:27–37. https://doi.org/10.1007/s11032-012-9765-0

Parkin IA, Sharpe AG, Keith DJ, Lydiate DJ (1995) Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). Genome 38:1122–1131. https://doi.org/10.1139/g95-149

Pilet ML, Duplan G, Archipiano M, Barret P, Baron C, Horvais R, Tanguy X, Lucas MO, Renard M, Delourme R (2001) Stability of QTL for field resistance to blackleg across two genetic backgrounds in oilseed rape. Crop Sci 41:197–205. https://doi.org/10.2135/cropsci2001.411197x

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033

Raman H, Raman R, Qiu Y, Yadav AS, Sureshkumar S, Borg L, Rohan M, Wheeler D, Owen O, Menz I, Balasubramanian S (2019) GWAS hints at pleiotropic roles for *FLOWERING LOCUS T* in flowering time and yield-related traits in canola. BMC Genom 20:636. https://doi.org/10.1186/s12864-019-5964-y

Raynaud C, Sozzani R, Glab N, Domenichini S, Perennes C, Cella R, Kondorosi E, Bergounioux C (2006) Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in Arabidopsis. Plant J 47:395–407. https://doi.org/10.1111/j.1365-313X.2006.02799.x

Robert HS, Quint A, Brand D, Vivian-Smith A, Offringa R (2009) BTB and TAZ domain scaffold proteins perform a crucial function in *Arabidopsis* development. Plant J 58:109–121. https://doi.org/10.1111/j.1365-313X.2008.03764.x

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29:24–26. https://doi.org/10.1038/nbt.1754

Samans B, Chalhoub B, Snowdon RJ (2017) Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species *Brassica napus*. Plant Genome 10:1–15. https://doi.org/10.3835/plantgenome2017.02.0013

Saxena RK, Edwards D, Varshney RK (2014) Structural variations in plant genomes. Brief Funct Genom 13:296–307. https://doi.org/10.1093/bfgp/elu016

Schiessl S, Huettel B, Kuehn D, Reinhardt R, Snowdon R (2017a) Post-polyploidisation morphotype diversification associates with gene copy number variation. Sci Rep 7:41845. https://doi.org/10.1038/srep41845

Schiessl S, Huettel B, Kuehn D, Reinhardt R, Snowdon RJ (2017b) Targeted deep sequencing of flowering regulators in *Brassica napus* reveals extensive copy number variation. Sci Data 4:170013. https://doi.org/10.1038/sdata.2017.13

Schiessl S, Iniguez-Luy F, Qian W, Snowdon RJ (2015) Diverse regulatory factors associate with flowering time and yield responses in winter-type *Brassica napus*. BMC Genom 16:737. https://doi.org/10.1186/s12864-015-1950-1

Schiessl S, Samans B, Hüttel B, Reinhard R, Snowdon RJ (2014) Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. Front Plant Sci 2014:1–14. https://doi.org/10.3389/fpls.2014.00404

Schiessl SV, Huettel B, Kuehn D, Reinhardt R, Snowdon RJ (2017c) Flowering time gene variation in *Brassica* Species shows evolutionary principles. Front Plant Sci 2017:1–13. https://doi.org/10.3389/fpls.2017.01742

Schiessl SV, Quezada-Martinez D, Tebartz E, Snowdon RJ, Qian L (2019) The vernalisation regulator *FLOWERING LOCUS C* is differentially expressed in biennial and annual *Brassica napus*. Sci Rep 9:14911. https://doi.org/10.1038/s41598-019-51212-x

Schmidt MH-W, Vogel A, Denton AK, Istace B, Wormit A, van de Geest H, Bolger ME, Alseekh S, Maß J, Pfaff C, Schurr U, Chetelat R, Maumus F, Aury J-M, Koren S, Fernie AR, Zamir D, Bolger AM, Usadel B (2017) De Novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. Plant Cell 29:2336–2348. https://doi.org/10.1105/tpc.17.00521

Searle I, He Y, Turck F, Vincent C, Fornara F, Kröber S, Amasino RA, Coupland G (2006) The transcription factor *FLC* confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. Genes Dev 20:898–912. https://doi.org/10.1101/gad.373506

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC (2018) Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods 15:461–468. https://doi.org/10.1038/s41592-018-0001-7

Seo PJ, Park M-J, Lim M-H, Kim S-G, Lee M, Baldwin IT, Park C-M (2012) A Self-Regulatory Circuit of *CIRCADIAN CLOCK-ASSOCIATED1* underlies the circadian clock regulation of temperature responses in *Arabidopsis*. Plant Cell 24:2427–2442. https://doi.org/10.1105/tpc.112.098723

Song J-M, Guan Z, Hu J, Guo C, Yang Z, Wang S, Liu D, Wang B, Lu S, Zhou R, Xie W-Z, Cheng Y, Zhang Y, Liu K, Yang Q-Y, Chen L-L, Guo L (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. Nat Plants 6:34–45. https://doi.org/10.1038/s41477-019-0577-7

Song K, Lu P, Tang K, Osborn TC (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc Natl Acad Sci U S A 92:7719–7723. https://doi.org/10.1073/pnas.92.17.7719

Srikanth A, Schmid M (2011) Regulation of flowering time: all roads lead to Rome. Cell Mol Life Sci 68:2013–2037. https://doi.org/10.1007/s00018-011-0673-y

Stein A, Coriton O, Rousseau-Gueutin M, Samans B, Schiessl SV, Obermeier C, Parkin IAP, Chèvre A-M, Snowdon RJ (2017) Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. Plant Biotechnol J 15:1478–1489. https://doi.org/10.1111/pbi.12732

Szadkowski E, Eber F, Huteau V, Lodé M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ, Chalhoub B, Jenczewski E, Chèvre A-M (2010) The first meiosis of resynthesized *Brassica napus*, a genome blender. New Phytol 186:102–112. https://doi.org/10.1111/j.1469-8137.2010.03182.x

Todesco M, Owens GL, Bercovich N, Légaré J-S, Soudi S, Burge DO, Huang K, Ostevik KL, Drummond EBM, Imerovski I, Lande K, Pascual-Robles MA, Nanavati M, Jahani M, Cheung W, Staton SE, Muños S, Nielsen R, Donovan LA, Burke JM, Yeaman S, Rieseberg LH (2020) Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature. https://doi.org/10.1038/s41586-020-2467-6

Tudor EH, Jones DM, He Z, Bancroft I, Trick M, Wells R, Irwin JA, Dean C (2020) QTL-seq identifies *BnaFTA.02* and *BnaFLCA.02* as candidates for variation in vernalisation requirement and response in winter oilseed rape (*Brassica napus*). Plant Biotechnol J. https://doi.org/10.1111/pbi.13421

Turck F, Fornara F, Coupland G (2008) Regulation and identity of florigen: *FLOWERING LOCUS T* moves center stage. Annu Rev Plant Biol 59:573–594. https://doi.org/10.1146/annurev.arplant.59.032607.092755

U N, (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *Brassica napus* and peculiar mode of fertilization. Jap J Bot 7:389–453

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35:W71–W74. https://doi.org/10.1093/nar/gkm306

Wang J, Long Y, Wu B, Liu J, Jiang C, Shi L, Zhao J, King GJ, Meng J (2009) The evolution of *Brassica napus FLOWERING LOCUS T* paralogues in the context of inverted chromosomal duplication blocks. BMC Evol Biol 9:271. https://doi.org/10.1186/1471-2148-9-271

Wang X, Wang H, Liu S, Ferjani A, Li J, Yan J, Yang X, Qin F (2016) Genetic variation in *ZmVPP1* contributes to drought tolerance in maize seedlings. Nat Genet 48:1233–1241. https://doi.org/10.1038/ng.3636

Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, Lohmann JU, Weigel D (2005) Integretation of spatial and temporal information during floral induction in *Arabidopsis*. Science 309:1056–1059. https://doi.org/10.1126/science.1115983

Wu D, Liang Z, Yan T, Xu Y, Xuan L, Tang J, Zhou G, Lohwasser U, Hua S, Wang H, Chen X, Wang Q, Zhu Le, Maodzeka A, Hussain N, Li Z, Li X, Shamsi IH, Jilani G, Wu L, Zheng H, Zhang G, Chalhoub B, Shen L, Yu H, Jiang L (2019) Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. Mol Plant 12:30–43. https://doi.org/10.1016/j.molp.2018.11.007

Xiong Z, Gaeta RT, Pires JC (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc Natl Acad Sci U S A 108:7908–7913. https://doi.org/10.1073/pnas.1014138108

Xu L, Hu K, Zhang Z, Guan C, Chen S, Hua W, Li J, Wen J, Yi B, Shen J, Ma C, Tu J, Fu T (2016) Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.). DNA Res 23:43–52. https://doi.org/10.1093/dnares/dsv035

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J (2006) The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. Proc Natl Acad Sci U S A 103:19581–19586. https://doi.org/10.1073/pnas.0607142103

Zhao J, Becker HC, Zhang D, Zhang Y, Ecke W (2005) Oil content in a European × Chinese rapeseed population: QTL with additive and epistatic effects and their genotype-environment interactions. Crop Sci 45:51–59

3 Dissection of quantitative blackleg resistance reveals novel variants of resistance gene *Rlm9* in elite *Brassica napus*

Vollrath, P., Chawla, H. S., Alnajar, D., Gabur, I., Weber, S., Ehrig, L., Lee, H. T., Koopmann, B., Snowdon, R. J. and Obermeier C.

Frontiers in Plant Science 12:749491 (2021)

doi: 10.3389/fpls.2021.749491

frontiers in Plant Science

ORIGINAL RESEARCH
published: 18 November 2021
doi: 10.3389/fpls.2021.749491

Dissection of Quantitative Blackleg Resistance Reveals Novel Variants of Resistance Gene *Rlm9* in Elite *Brassica napus*

*Paul Vollrath[1], Harmeet S. Chawla[2], Dima Alnajar[3], Iulian Gabur[1,4], HueyTyng Lee[1], Sven Weber[1], Lennard Ehrig[1], Birger Koopmann[3], Rod J. Snowdon[1] and Christian Obermeier[1]**

[1] Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, Giessen, Germany, [2] Department of Plant Sciences, Crop Development Centre, University of Saskatchewan, Saskatoon, SK, Canada, [3] Plant Pathology and Crop Protection Division, Department of Crop Sciences, Georg August University of Göttingen, Göttingen, Germany, [4] Department of Plant Sciences, Faculty of Agriculture, Iasi University of Life Sciences, Iaşi, Romania

Blackleg is one of the major fungal diseases in oilseed rape/canola worldwide. Most commercial cultivars carry *R* gene-mediated qualitative resistances that confer a high level of race-specific protection against *Leptosphaeria maculans*, the causal fungus of blackleg disease. However, monogenic resistances of this kind can potentially be rapidly overcome by mutations in the pathogen's avirulence genes. To counteract pathogen adaptation in this evolutionary arms race, there is a tremendous demand for quantitative background resistance to enhance durability and efficacy of blackleg resistance in oilseed rape. In this study, we characterized genomic regions contributing to quantitative *L. maculans* resistance by genome-wide association studies in a multiparental mapping population derived from six parental elite varieties exhibiting quantitative resistance, which were all crossed to one common susceptible parental elite variety. Resistance was screened using a fungal isolate with no corresponding avirulence (*AvrLm*) to major *R* genes present in the parents of the mapping population. Genome-wide association studies revealed eight significantly associated quantitative trait loci (QTL) on chromosomes A07 and A09, with small effects explaining 3–6% of the phenotypic variance. Unexpectedly, the qualitative blackleg resistance gene *Rlm9* was found to be located within a resistance-associated haploblock on chromosome A07. Furthermore, long-range sequence data spanning this haploblock revealed high levels of single-nucleotide and structural variants within the *Rlm9* coding sequence among the parents of the mapping population. The results suggest that novel variants of *Rlm9* could play a previously unknown role in expression of quantitative disease resistance in oilseed rape.

Keywords: ONT, structural variation, blackleg, *Brassica napus*, long-read sequencing, Rlm9

OPEN ACCESS

Edited by:
Jacqueline Batley, University of Western Australia, Australia

Reviewed by:
Angela Van De Wouw, The University of Melbourne, Australia
Francis Chuks Ogbonnaya, Grains Research and Development Corporation, Australia

***Correspondence:**
Christian Obermeier
christian.obermeier@agrar.uni-giessen.de

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *29 July 2021*
Accepted: *29 October 2021*
Published: *18 November 2021*

Citation:
Vollrath P, Chawla HS, Alnajar D, Gabur I, Lee H, Weber S, Ehrig L, Koopmann B, Snowdon RJ and Obermeier C (2021) Dissection of Quantitative Blackleg Resistance Reveals Novel Variants of Resistance Gene Rlm9 in Elite Brassica napus. Front. Plant Sci. 12:749491. doi: 10.3389/fpls.2021.749491

INTRODUCTION

Oilseed rape/canola (*Brassica napus* L.) is one of the most important vegetable oil crops. As a recent allotetraploid crop, originating from an interspecific hybridization event between its two diploid ancestors *Brassica rapa* (2n = 2x = 20, AA) and *Brassica oleracea* (2n = 2x = 18, CC) (Nagaharu, 1935), *B. napus* (2n = 4x = 38, AACC) carries a highly complex and dynamic genome which is

affected by many small-scale and large-scale structural variations (Parkin et al., 1995; Chalhoub et al., 2014; Stein et al., 2017; Hurgobin et al., 2018; Chawla et al., 2020). Many studies revealed high frequencies of homoeologous exchanges between the highly similar A and C subgenomes (Szadkowski et al., 2010; Chalhoub et al., 2014; Samans et al., 2017).

The hemibiotrophic fungal pathogen *Leptosphaeria maculans* (Desm.) Ces. & de Not. [anamorph: *Phoma lingam* (Tode ex. Fr.) Desm.] causes stem canker in *B. napus*. This disease, also known as blackleg, is a major problem in almost all oilseed rape and canola-growing regions around the globe. Substantial yield losses have been reported in Australia, North America and several European countries (Fitt et al., 2006). The primary infection of winter oilseed rape takes place in autumn via airborne ascospores. Additionally, secondary infections are likely through pycnidiospores formed within the asexual pycnidia. The spores penetrate the host tissue via stomata or wounds and colonize intercellular spaces of the mesophyll. From there, the fungus starts its symptomless, biotrophic growth systematically through the petiole into the stem. Here the pathogen kills the cells, leading to girdling and rotting of the stem base. As a result, the plant tends to ripen prematurely and severe infections can lead to serious lodging and death (West et al., 2001). Resistance breeding is the most sustainable and effective method to counteract *L. maculans*. Resistance of *B. napus* against *L. maculans* is often divided into two classes of resistance mechanisms. A distinction is made between race-specific, qualitative resistance determined by major genes, and non-race-specific quantitative resistance provided by numerous minor effect genes. Qualitative resistance against *L. maculans*, considered as complete resistance, has been investigated in considerable detail and used extensively in commercial breeding programs of *B. napus*, due to the high efficacy and convenient assessment at the cotyledon stage (Rimmer and van den Berg, 1992; Balesdent et al., 2005; Delourme et al., 2006; Elliott et al., 2016). However, it has been observed that rapid adaptation of the pathogen populations can overcome *R* gene-mediated resistance in the field within a few seasons (Rouxel et al., 2003; Sprague et al., 2006; Brun et al., 2010; Zhang et al., 2016; van de Wouw et al., 2017). Thus, commercial breeders place special focus on quantitative disease resistance. Quantitative resistance is influenced by multiple genes, and the incomplete nature of the resistance decreases the selection pressure on the pathogen population and consequently increases the durability of the resistance (St Clair, 2010; Delourme et al., 2014). In contrast to qualitative resistance, the assessment of quantitative resistance is more challenging as it is mainly expressed at adult plant stages and highly influenced by environmental conditions (Fitt et al., 2006; Huang et al., 2009). A main target of breeding is therefore the combination of highly effective *R* gene-mediated resistance with a broad and durable quantitative resistance (Brun et al., 2010; Pilet-Nayel et al., 2017). However, this clear distinction between the two types of resistance has recently been questioned in several studies (Thomma et al., 2011; Delplace et al., 2020). Also, in some cases the expression of partial resistance in adult plants imparted by major resistance genes has been demonstrated (Chantret et al., 1999; Raman et al., 2018). The complexity of the *B. napus–L. maculans* pathosystem and possible corresponding genes often makes the assessment of quantitative resistance difficult. Therefore, the objective of the present study was to identify genomic regions involved in quantitative blackleg resistance by excluding the possibility of effective *R* gene-*Avr* gene interactions. A multiparental mapping population was tested under controlled greenhouse conditions with a selected highly virulent fungal isolate. In addition, whole-genome long-read re-sequencing using Oxford Nanopore Technology (ONT) was conducted to reveal the implications of single nucleotide variants (SNV) and structural genome variations (SV) on an agronomical highly important trait within the narrow genepool of European elite winter oilseed rape. Recently, Chawla et al. (2020) demonstrated a previously unknown extent of genome-wide, small to medium-sized SV events within *B. napus* genes using ONT long-read sequencing technology. In contrast, SNV have to date been largely ignored in quantitative trait analysis of important crop traits, due to the difficulty of assaying SV on a genome-wide scale in complex crop genomes. In the past few years, however, rapidly decreasing costs and increasing accuracy of long-read sequencing from the ONT or Pacific Biosciences platforms has opened the way to include genome-wide SV data from long-read sequences in QTL analysis and interpretation. Here we successfully called SNV in ONT data from the seven elite winter oilseed rape parents of the multiparental mapping population, enabling us to associate genome-wide SV with single nucleotide polymorphism (SNP) haplotypes carrying blackleg resistance QTL.

MATERIALS AND METHODS

Plant Material

A *B. napus* multiparental population comprising 354 double haploid (DH) lines derived from seven European elite winter oilseed rape varieties was tested for quantitative blackleg resistance in multiple greenhouse screenings. The mapping population consisted of six subfamilies derived from crosses of the elite parent "Lorenz" to six elite founder lines ("Adriana," "Alpaga," "DK Cabernet," "Galileo," "King 10," and the DH line "JN"). Each of the subfamilies comprised 60 DH lines except for the cross Lorenz × Galileo, which comprised 54 DH lines. The common parent Lorenz was previously classified as highly susceptible to blackleg disease, whereas the other six founder lines were all known to carry quantitative blackleg resistance (unpublished data, breeding companies). The German breeding companies NPZ Innovation GmbH (Holtsee, Germany), Syngenta Seeds GmbH (Bad Salzuflen, Germany), and KWS SAAT SE & Co. KGaA (Einbeck, Germany) produced and provided the DH families. In parallel, another panel of 256 diverse winter oilseed rape inbred lines was tested for blackleg resistance in a 2-year field trial with one plot per field. These accessions were part of the ERANET-ASSYST *B. napus* diversity set, previously described by Bus et al. (2011).

Resistance Screenings and Data Analysis

The multiparental mapping population was screened for adult-plant blackleg resistance under controlled conditions in the greenhouse of Georg August University of Göttingen in 11 independent screening experiments, each individual screening included all 354 genotypes with two plant individuals (replicates) per genotype in a completely randomized design. In total 22 plant individuals were tested per genotype across 11 screening experiments (22 replicates per genotype). Manual infection was carried out at developmental stage BBCH 13–14 (3–4 true leaves, leaf pairs, or whorls unfolded) (Lancashire et al., 1991). *L. maculans* was propagated on oatmeal agar medium 2 weeks before infection. A mycelial agar plug was then placed at the stem base, slightly above the axil of the first true leaf, after wounding of the infection site using a needle. Subsequently, plants were grown under foil tunnels for 72 h to ensure appropriate humidity and temperature for a successful infection. At 49 days post infection (dpi), a cross section was cut at the stem base to estimate the length (L), girdling (G), and penetration depth (P) of the blackleg lesions. L was measured in mm whereas G and P were visually scored as percentage of the total circumference and diameter, respectively. Next, these scores were converted into individual 0–9 scales for each score. Using L, G, and P scoring values at 49 dpi, the Volume of Diseased Tissue (VDT) value was calculated using a formula modified from Kutcher et al. (1993):

$$VDT = \left(1 - (\frac{1-P}{9})^2\right) * \frac{G}{9} * L$$

Also, single screening means of the two replicates per screening, adjusted means across all the eleven screenings, were calculated using the R packages lmerTest version 3.1-2 (Kuznetsova et al., 2017) and lsmeans version 2.30-0 (Lenth, 2016). This approach allowed the assessment of QTL stability across different screenings.

The diversity set was grown in 2016/2017 and 2017/2018 in field trials in Rauischholzhausen, Germany. It was grown under normal farming practices with no use of fungicides. The fields were chosen based on close crop rotation and known high natural blackleg infection pressure. Single plots per genotype were sown and analyzed in a randomized complete block design, with plot sizes of 12.5 m^2 (10 m × 1.25 m). At developmental stage BBCH 83–85 (30–50% of pods ripe, seeds black, and hard), 20 plants from the middle row of each plot were uprooted and cut at the stem base. Visual scores from 1 to 6 for blackleg infestation at the resulting cross section were used to calculate the G2 index for blackleg adult plant stem infection (Pilet et al., 1998; Aubertot et al., 2004) using the formula:

$$G2\ index = [(N1x0) + (N2x1) + (N3x3) + (N4x5) + (N5x7) + (N6x9)]\ /Nt,$$

where N1, N2, N3, N4, N5, and N6 are the number of stems with scores 1, 2, 3, 4, 5, and 6, respectively, and Nt is the total number of stems assessed.

Characterization of Fungal Isolates

Leaf samples with characteristic lesions of *L. maculans* were collected in the field trials and dried. Leave segments were incubated in humid chambers to allow spore release from pycnidia. Single pycnidium isolates were prepared by plating spores on SNA medium amended with 200 ppm streptomycin for 6 days. Petri dishes were incubated under UV light at 20°C. Subsequently, a mycelial plug was transferred to V8 medium supplemented with 200 ppm streptomycin and incubated for 2 weeks under the same conditions. Spore suspensions were prepared by adjustment to 1×10^7 spores/ml using a hemocytometer. To characterize *L. maculans* isolates, cotyledon tests were applied using a differential set of *B. napus* harboring major resistance genes (**Supplementary Table 3**; Balesdent et al., 2002, 2006; Marcroft et al., 2012). Shortly, 10 µl spore suspension was applied on each lobe of cotyledon after injuring it with a needle. Eight replications were used. Symptoms were evaluated 14 days post inoculation according to the IMASCORE rating scale, where class 1 shows typical hypersensitive reactions and class 6 reflects tissue collapse with sporulation. Classes 1–3 were considered as resistance reactions, while classes 4–6 were noted as susceptible ones (Balesdent et al., 2001).

Single Nucleotide Polymorphism Genotyping and Analysis of Linkage Disequilibrium

All the investigated *B. napus* accessions were genotyped using the *Brassica* 60k Illumina Infinium™ SNP array. The *B. napus* reference genome assembly Darmor-*bzh* version 10 (Rousseau-Gueutin et al., 2020) was used to anchor 34,079 markers uniquely to a single position of the genome (**Supplementary Table 1**). Markers mapping unspecifically to multiple positions were excluded from further analyses. Finally, single hits were filtered for a cut-off *e*-value of $1e^{-15}$. Heterozygote SNP calls were considered as missing data since we should not expect heterozygote calls for DH or inbred lines, hence it can be assumed that these calls are mostly due to technical artifacts. Genome-wide linkage disequilibrium (LD) was calculated using the R package SelectionTools version 19.4[1]. Prior to LD analysis, markers were filtered for minor allele frequency (MAF) ≥0.05 and a maximum of 10% missing data per marker and DH line. A tolerance threshold of $r^2 > 0.4$ was set to assign markers in strong LD to respective LD blocks.

Genome-Wide Association Studies

Genome-wide association studies were conducted using the R package GenABEL version 1.8-0 (Aulchenko et al., 2007). Markers for the multiparental population were filtered as described above for LD analysis. This approach led to a set of 17,869 polymorphic and unique markers, including 16,400 SNP markers and 1,469 Single Nucleotide absence Polymorphism (SNaP) markers called as described by Gabur et al. (2018). The linear mixed model was adjusted for population structure by consideration of identity-by-state estimates and the first

[1]http://population-genetics.uni-giessen.de/~software/

two principal components (PC) as covariates. To reduce false positive rates, LD blocks were determined as suggestive QTL when a minimum of two markers per block showed trait associations in at least two individual greenhouse screening rounds. A LOD score of −log10 (p-value) ≥3.0 was applied as threshold for suggestive marker-trait associations. Finally, QTL were determined after correction for false discovery rate (FDR ≤ 0.1) when performing GWAS with the adjusted means across all greenhouse screening rounds.

In order to validate QTL discovered in the multiparental population using the greenhouse data, we also performed an independent GWAS in the diversity panel using phenotypic data from the field trials. Similar filtering steps led to 23,603 polymorphic SNP markers. Here, LD-based QTL were defined by considering kinship and PC and applying a LOD score of −log10 (p-value) ≥3.0 as an arbitrary threshold for putative marker-trait associations.

Functional Annotation of Darmor-*bzh* Genes

Functional annotation data for Darmor-*bzh* v4.1 produced by Gabur et al. (2020) were used together with genome-wide functional annotation of Darmor-*bzh* v10 performed using the "Automatic assignment of Human Readable Descriptions" (AHRD)[2] package (**Supplementary Table 2**). AHRD obtains the functional annotations for gene models by blasting them to various publicly available protein databases such as Swiss-Prot, TAIR or trEMBL. Two hundred best scoring blast results (based on e-value) were chosen from each of the above-mentioned databases. Description for all the resulting blast hits was then assigned a score using a multi-step approach. In the first step every description line was subjected to a couple of regular expression filters, removing descriptions such as "Whole genome shotgun sequence" and other vague terms like "OS = Arabidopsis thaliana." In the subsequent step the description lines were broken down into single tokens. These tokens were then pushed through a blacklist filter, thereby discarding all the tokens present in the blacklist. Every token was then assigned an overlap score based on the bit score, the database score, and the overlap score of the blast match. In the last step the token score was divided by a correction factor to remove any bias toward longer or shorter description lines. For the exact database and software versions please refer to the "Darmor10_input_go_prediction.yaml" file in the **Supplementary Material**.

Whole Genome Long-Read Resequencing and Variant Calling

Long-read sequencing was performed for all the seven parental lines of the mapping population using Oxford Nanopore Technologies (ONT). DNA extraction was conducted as described in Chawla et al. (2020). In addition, DNA size selection was performed using the Circulomics Short Read Eliminator Kit (Circulomics Inc.). The recommended kit LSK-109 from ONT was used for DNA library preparation. Long-read sequencing was performed using the MinION device from ONT. Subsequently, basecalling was executed with Guppy version 4.0.14 and reads were aligned to the *B. napus* reference assembly Darmor-*bzh* v10 using the long-read mapper NGMLR version 0.2.7 (Sedlazeck et al., 2018). BAM files were created using samtools version 1.9 (Li et al., 2009). For genome-wide SV detection the variant caller Sniffles version 1.0.12 was used with default settings (Sedlazeck et al., 2018). Deletions and insertions with a minimum size of 30 bp were classified as SV. In addition, SNV calling was conducted using the deep neural network based variant caller Clair version 2 (Luo et al., 2020). The Clair module callVarBam, with a model for ONT data, was used to call SNV from BAM files. To reduce false positive calls, SNV calls were filtered for a minimum quality using a shell script kindly provided by Fritz Sedlazeck and Medhat Mahmoud, Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX, United States. The script calculates the most appropriate cut-off value for SNV quality filtering according to the recommendations of the authors of Clair. Luo et al. (2020) observed that quality scores of variants derived from ONT data are usually bimodally distributed, so that high-quality base calls can be extracted by setting a quality cut-off at a value corresponding to the bottom of the valley between the two peaks plus 50. In addition to quality filtering, a stringent filtering for only homozygous calls and a minimum allele frequency (AF) of 0.5 for the variant were applied. Then, the files of the single genotypes were merged using bcftools version 1.10.2 and were subsequently used as an input to invoke the force-calling parameter of Clair. Next, all samples were run again to force-call the SNV provided within the merged file. This again was followed by filtering for quality, homozygous calls and a minimum AF of 0.1. A lower threshold for AF was chosen after force-calling, since previous SNV calling already indicates the presence of these variants. Subsequently, CLC Genomics Workbench (v9.0, QIAGEN Digital Insights, Aarhus, Denmark) was used to align the sequences of the parental lines and to predict the impacts of amino acid changes caused by SNV.

Single Nucleotide Variants Validation

Due to the high number of SNV, the sequence of the gene A07p27010.1_BnaDAR was selected to validate the SNV calling method. For this purpose, sets of primers were designed to amplify the selected region and Sanger Sequencing was performed in the mapping parents. Primers were designed using the online tool Primer3Plus (Untergasser et al., 2007).

RESULTS

Identification of a *L. maculans* Isolate Without *R* Gene Interaction for Use in Quantitative Resistance Screening

From the 644 *L. maculans* isolates collected in Northern Germany and characterized by Winter and Koopmann (2016), one isolate was selected (isolate 1.4.1.15). This isolate has been shown in cotyledon test with a *B. napus* differential set to harbor virulence alleles against *R* genes *Rlm1*, *Rlm2*, *Rlm4*, *Rlm7*, *Rlm9, LepR2,* and *LepR3*. The isolate was tested on the cotyledons of the

[2]https://github.com/groupschoof/AHRD

seven elite oilseed rape parents of the multiparental mapping population. It showed a susceptible interaction in all cotyledon tests indicating that none of the parents exhibit any qualitative major resistance against this fungal isolate based on plant *R* gene and fungal avirulence gene interaction (**Supplementary Table 3**). Thus, this virulent isolate 1.4.1.15 recovered from a field in Peine (Germany) in 2013 was used subsequently for the greenhouse screenings for quantitative resistance, in order to avoid interaction of major monogenic resistance genes with fungal avirulence genes that could putatively mask the effects of minor quantitative resistance loci in the mapping population.

European Elite *B. napus* Accessions Show Genetic Variation for Quantitative Blackleg Resistance

Mixed linear models (MLM) demonstrated significant genotypic variation within the multiparental population ($p < 0.001$). VDT values from greenhouse screenings showed a normal distribution within the six subfamilies as well as in the entire mapping population, confirming the quantitative inheritance of blackleg resistance in this population (**Supplementary Figure 1**). LD analysis resulted in an average number of 33 LD blocks per chromosome. Single greenhouse screening rounds using mean VDT values from 2 replicates per genotype revealed together up to 326 marker-trait associations in GWAS. GWAS using mean VDT values composed of 22 replicates for each genotype merged from a total number of 11 screening rounds identified 84 significant marker-trait associations (**Supplementary Table 4**). After FDR correction a total of eight QTL regions were identified (**Table 1**), seven on chromosome A09 and one on chromosome A07. No QTL were detected on C-subgenome chromosomes. As expected for quantitative resistance, phenotypic variation explained by individual SNP markers was low and was ranging from 3.2 to 5.6% (**Table 1**). QTL stacking in the multiparental population revealed that the allele combination present in the common parent Lorenz led to the highest susceptibility compared to alleles from the other parental lines. This confirmed the initial assumption and the choice of Lorenz as the susceptible common parent. Almost all resistance alleles were derived from the six founder lines, whereas only one resistance allele was derived from the common parent Lorenz. The ideal allele combination of the eight QTL resulted in an estimated effect on resistance of 35.5% (−2.06 on the VDT scale). This beneficial combination was already present in the two founder lines DK Cabernet and JN (**Supplementary Table 5**).

Candidate Gene Analysis for Genes Involved in Quantitative Resistance

Based on two approaches together we identified 128 genes across the eight QTL regions that are associated to defense response and/or resistance in at least one reference genome (**Supplementary Table 6**). The QTL on chromosome A07 (haploblock A07.b304) was detected in the highest number of individual screening rounds (**Table 1**). The haploblock had a size of 888 kb in the investigated multiparental mapping population and contained 53 SNP markers and 193 genes in

TABLE 1 | Blackleg resistance QTL (VDT trait) identified in a *B. napus* multiparental elite mapping population in 11 independent greenhouse screening rounds ($n = 354$) and overlapping resistance QTL (G2 or stem lesion) in the ERANET-ASSYST diversity set in two field trials ($n = 256$).

		Multiparental elite mapping population					ERANET-ASSYST diversity set				
QTL ID	**Chromo-some**	**Start-end position of LD block/QTL**	**Size of QTL**	**LOD#**	**R^2#**	**Detection in screening rounds***	**Start-end position of LD block/QTL**	**Size of QTL**	**LOD#**	**R^2#**	**Detection in field trial**
A07.b304	A07	20,107,508–20,995,393	888 kb	3.35	3.40%	T, 3, 6, 7, 9, 10	20,079,904–20,378,231	298 kb	3.57	7.79%	2017/18
A09.b366	A09	13,704,742–14,702,671	998 kb	5.17	5.48%	T, 1, 3, 7	12,981,315–14,141,992	1,160 kb	3.46	6.82%	2016/17
A09.b369	A09	15,852,358–16,237,668	385 kb	5.22	5.53%	T, 1, 3, 7	ns	ns	ns	ns	–
A09.b370	A09	16,917,123–16,939,597	22 kb	3.17	3.19%	T, 1, 3	ns	ns	ns	ns	–
A09.b374	A09	18,825,078–19,494,241	669 kb	5.07	5.36%	T, 2, 3, 11	ns	ns	ns	ns	–
A09.b380	A09	33,761,934–33,966,136	204 kb	5.02	5.32%	T, 1, 2, 3, 11	ns	ns	ns	ns	–
A09.b381	A09	34,149,762–38,832,384	4,683 kb	3.50	3.58%	T, 1, 3	ns	ns	ns	ns	–
A09.b382	A09	38,844,647–41,182,649	2,338 kb	3.88	4.09%	T, 1, 3	41,110,836–42,454,513	1,343 kb	3.07	5.53%	2016/17

*#LOD and R^2 values for peak markers; *T = data from 11 screening rounds combined (adjusted mean values), numbers indicate different screening rounds with 2 plant individuals per genotype; ns, not significant.*

Darmor-*bzh* v10 (**Supplementary Table 7**). However, some genes in the interval revealed no annotation in Darmor-*bzh* v4.1 because no homologous genes exist between Darmor-*bzh* v4.1 and v10 or because Blast2GO revealed no annotation for the Darmor-*bzh* v4.1 genes (60 of 193). Thus, the protein sequences of the 193 *B. napus* Darmor-*bzh* v10 genes from the QTL interval were additionally aligned to the Arabidopsis reference genome Araport11 (Cheng et al., 2017) and literature links were evaluated exhibiting some additional, more detailed functional annotations. Out of 193 genes, 17 were associated to defense response and/or resistance (**Supplementary Table 7**). In particular, we found substantiated evidence in the literature that three of these 17 genes have a function related to fungal plant resistance in *A. thaliana* and *B. napus* in interaction with common *B. napus* fungal pathogens (A07p26890.1_BnaDAR, A07p28430.1_BnaDAR, A07p27010.1_BnaDAR). Most interestingly, the gene A07p27010.1_BnaDAR has been shown to be the major resistance gene *Rlm9*, which imparts qualitative resistance against *L. maculans* in *B. napus* (Larkan et al., 2020). To evaluate for the presence of resistance gene analogs (RGAs) in the QTL region of A07.b304 and their polymorphisms between the parents of the multiparental mapping population, we also located the QTL interval for Darmor-*bzh* versions 4.1 (Chalhoub et al., 2014) and 8.1 (Bayer et al., 2017) based on the flanking SNP markers (**Supplementary Table 8**). However, two out of five putative RGAs identified in this interval showed no polymorphism and none of them had been annotated to be involved in plant resistance.

Identification of Major Resistance Gene *Rlm9* as Candidate Gene for Quantitative Blackleg Resistance

The detected quantitative resistance region has been mapped before for the qualitative resistance gene *Rlm9* in young seedlings (Larkan et al., 2016b; Raman H. et al., 2020). However, testing of virulence complexity proved that the isolate used for greenhouse screening for quantitative resistance in our experiments shows no gene-for-gene interaction with *Rlm1*, *Rlm2*, *Rlm4*, *Rlm7*, *Rlm9, LepR2,* and *LepR3* (see above). On the other hand, variant calling using ONT long-read sequencing data revealed that the sequence diversity in the gene A07p27010.1_BnaDAR (*Rlm9*) was considerably higher than in all other genes within the QTL interval and all putative resistance genes in any of the other identified QTL (**Supplementary Table 7**). Within the QTL interval, the peak SNP marker lies just 35 kb away from the *Rlm9* gene. Within the sequence of *Rlm9*, we found 142 polymorphic SNV calls and a 6 kb insertion in four of the seven parental lines (Adriana, Alpaga, Galileo, and King 10). Parental lines Lorenz (susceptible), DK Cabernet and JN displayed an identical haplotype, comprising 24 SNPs in LD to the peak SNP marker, whereas the resistant genotypes harboring the insertion within the sequence of *Rlm9* (Adriana, Alpaga, Galileo, King 10) showed a deviating haplotype (**Figure 1**). PCR and Sanger sequencing proved both, the authenticity of the 6 kb insertion and the correctness of the 142 intragenic SNV calls in A07p27010.1_BnaDAR. Each of the 142 SNV calls from ONT data was confirmed to be correct, whereby the Sanger sequencing also revealed a further 60 SNV within this particular gene. These 60 false negatives can be explained by the strict filtering process applied to eliminate false positive calls in the SNV calling approach. Nevertheless, the results indicate that SNV calling from "noisy" long reads produced with ONT provides reliable variant calls for genetic analysis.

Rlm9 codes for a wall-associated kinase like (WAKL) protein. A search for motifs using the Conserved Domain Search tool implemented in NCBI along with the Pfam database revealed three conserved domains, two within exon 1 and one within exon 3. These comprise an extracellular galacturonan-binding domain (GUB_WAK), a C-terminal wall-associated kinase (WAK) and an intracellular Serine/Threonine protein kinase domain (Ser/Thr_kinase). In addition, an EGF-like domain is located in exon 2 (Larkan et al., 2020). Based on the Sanger sequencing we found that three of the seven parental genotypes (Lorenz, DK Cabernet, JN) carry an identical *Rlm9* allele to Darmor-*bzh*, whereas the other four harbor a 6 kb insertion within the second exon along with 202 SNV throughout the entire gene (Adriana, Alpaga, Galileo, King 10). These SNV caused in total 92 non-synonymous amino acid changes. We observed 19 amino acid changes within the GUB_WAK domain (84.2% identity to Darmor-*bzh*), 15 amino acid changes within the WAK domain (86.7% identity to Darmor-*bzh*) and 23 amino acid changes as well as a stop codon within the Ser/Thr_kinase domain (91.3% identity to Darmor-*bzh*). The EGF-like domain in exon 2 was disrupted by the 6 kb insertion in genotypes Adriana, Alpaga, Galileo, and King 10 (**Figure 2**).

Candidate Genes in Co-localizing Quantitative Trait Loci From Greenhouse and Field Trials

The QTL regions and candidate genes where initially identified in the elite population under controlled conditions excluding *R* gene interaction. To putatively support these candidate genes, we also identified QTL for adult plant resistance under field conditions in a diversity set. The field trials using the ERANET-ASSYST diversity set relied on natural infection. Characterization of field isolates sampled at the field site in Rauischholzhausen (Germany) revealed fungal isolates harboring the avirulence alleles *AvrLm3* or *AvrLm7* in combination with virulence alleles *avrLm1, avrLm2, avrLm4,* and *avrLm9*. This situation represents the typical German field situation as reported previously (Winter and Koopmann, 2016). This implies that monogenic *Rlm3* and *Rlm7*, also located on chromosome A07 (Larkan et al., 2016b; Raman R. et al., 2020) and potentially harbored by some of the genotypes of the diversity set, could mask quantitative resistance associated with the investigated genomic region on chromosome A07. However, monogenic *Rlm9* effects should not mask quantitative resistance effects associated with *Rlm9* in this field situation. Three of the eight QTL found in the analysis with the multiparental mapping population were also found in field trials for blackleg stem lesions in the ERANET-ASSYST diversity set (QTL A07.b304, A09.b366 and A09.b382, **Table 1** and **Supplementary Table 9**). The LD blocks for these three QTL

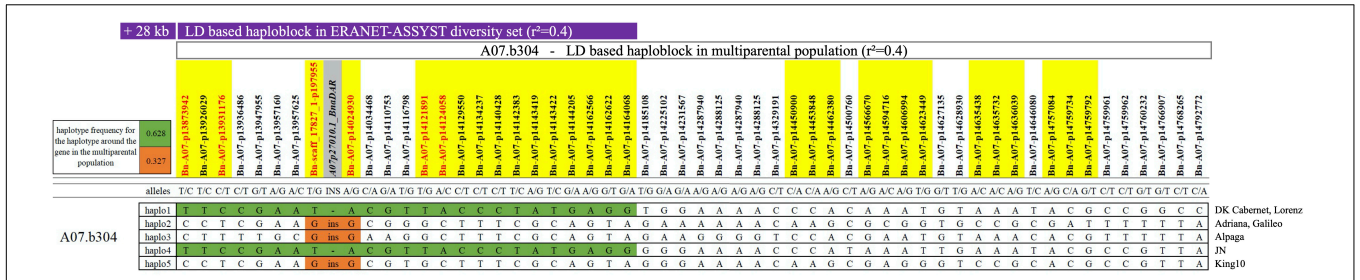

FIGURE 1 | LD based haploblock A07.b304 (r^2 = 0.4) containing 54 single nucleotide polymorphism (SNP) markers and *Rlm9* (A07p27010.1_BnaDAR). Green color indicates the two haplotypes with 24 identical SNP markers between the three parental lines Lorenz, DK Cabernet and JN. Haplotypes 2, 3, and 5 are harboring the 6 kb insertion in the sequence of *Rlm9* (orange). Purple bar highlights the overlap with the LD based QTL identified in the diversity set in field trials. Gray highlight indicates Rlm9 gene, yellow highlight indicates SNP markers identified in one individual greenhouse screening round, red color indicates SNP markers identified in at least two individual greenhouse screening rounds.

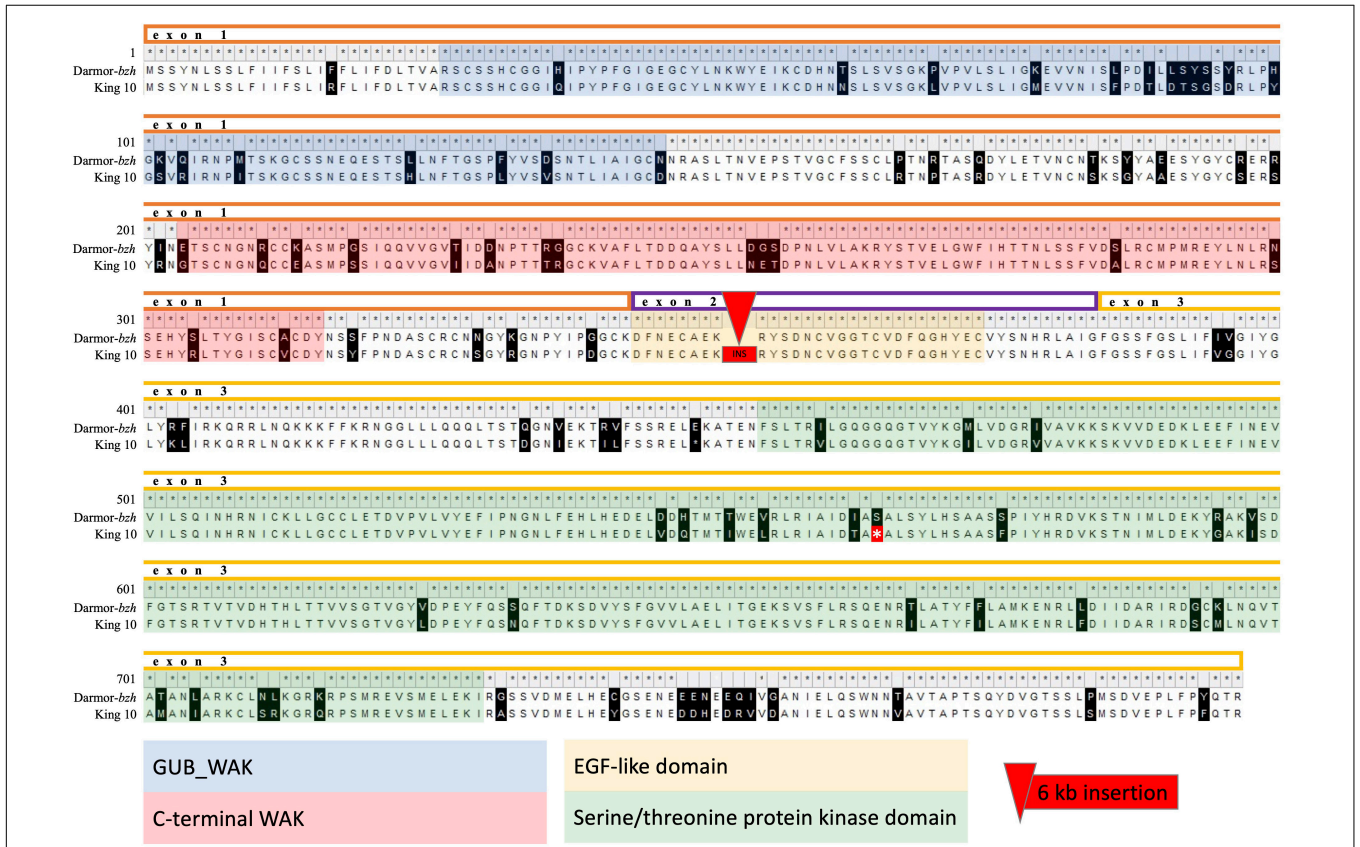

FIGURE 2 | Alignment of the protein sequence of King 10 against the reference Darmor-*bzh* v10. King 10 is representative for the four parental lines harboring a 6 kb insertion and 202 SNV leading to 92 non-synonymous amino acid changes. Different colors indicate exons and conserved domains. The asterisk indicates a stop codon.

partly overlap in both populations. Assuming that the same genes are involved in quantitative resistance expression in both winter oilseed rape populations, the overlap in the QTL interval might be useful to narrow the search for potential candidate genes. By considering overlapping QTL regions as high-confidence QTL intervals, we reduced the areas of interest from 888 kb to 271 kb, 998 kb to 473 kb, and 2,338 kb to 72 kb, respectively. In total, these three intervals contained 133 genes (68, 54, and 11). The number of candidate genes was reduced from 68 to 5, 54 to 2, and 11 to 1 for the three co-localizing QTL intervals for QTL A07.b304, A09.b366, and A09.b382 by filtering for GO terms associated to defense response and/or resistance (**Table 2** and **Supplementary Table 10**). *Rlm9* is localized in the overlapping QTL region of both populations, the multiparental mapping population and the diversity set tested under controlled conditions in the greenhouse and in the field, respectively.

TABLE 2 | Genes within the *L. maculans* resistance QTL of the multiparental population overlapping with QTL of the diversity set based on GO terms that can be associated with plant resistance and numbers of genomic variants in the multiparental population.

Gene ID	QTL ID	Range of overlap in kb	Gene annotation	Structural variants (SV)	No. of single nucleotide variant (SNV) calls
A07p26790.1_BnaDAR	A07.b304	271 kb	Nuclear pore complex protein NUP96	–	1
A07p26870.1_BnaDAR	A07.b304	271 kb	Cinnamoyl CoA reductase	–	18
A07p26890.1_BnaDAR	A07.b304	271 kb	Transcription factor WRKY	–	17
A07p27010.1_BnaDAR	A07.b304	271 kb	Wall-associated receptor kinase-like 10	6 kb insertion	142
A07p27130.1_BnaDAR	A07.b304	271 kb	ALA-interacting subunit	–	–
A09p21820.1_BnaDAR	A09.b366	473 kb	Protein ENHANCED DISEASE RESISTANCE 2-like	–	96
A09p21880.1_BnaDAR	A09.b366	473 kb	Heat shock transcription factor	–	8
A09p44190.1_BnaDAR	A09.b382	72 kb	UPF0183 protein	–	–

DISCUSSION

Quantitative resistance against *L. maculans* exhibiting minor effects in *B. napus* is difficult to detect as it is frequently masked by qualitative resistances exhibiting major effects controlled by race-specific *R* genes in the host and pathogen. Pathogen-associated molecular patterns (PAMPs) are considered to initiate a broad-spectrum resistance against a pathogen species, termed PAMP-triggered immunity (PTI), whereas race-specific pathogen effectors contribute to pathogen virulence and can induce effector-triggered immunity (ETI) in the plant host on a gene-for-gene interaction model, also called qualitative resistance (Jones and Dangl, 2006). However, in the last decade it has become clear that this strict distinction between PTI and ETI might need reconsideration (Thomma et al., 2011). In addition, the extent to which PTI is associated with quantitative resistance expression also remains to be clarified (Delplace et al., 2020). ETI and PTI are both mainly involved in pathogen perception and signaling, whereas quantitative resistance goes beyond that and is controlled by numerous genes with diversified functions (Corwin and Kliebenstein, 2017). Qualitative and quantitative resistance cannot normally be clearly distinguished in field studies. The complexity of the fungal population in the field, harboring different avirulence genes, and the genetic composition of the plant population, harboring unknown minor and major resistance allele combinations, generally cause difficulties in association studies of quantitative resistance under field conditions. Thus, our strategy in this study was to identify and use a fungal *L. maculans* isolate that has no qualitative resistance *R* gene interaction with the parental oilseed rape cultivars of our multiparental population, and to use this isolate to map quantitative minor effect QTL by GWAS under controlled conditions.

Using this strategy, we identified a normal distribution of disease, indicating the exclusively quantitative inheritance of blackleg resistance in the mapping population. The absence of effective gene-for-gene interactions that can usually be observed in the *B. napus–L. maculans* pathosystem allowed us to identify genomic regions explaining only a small portion of phenotypic variation (<6%). In accordance with a previous study, we also observed high QTL-by-environment interactions, which is common for quantitatively inherited disease resistance even under controlled greenhouse conditions (Obermeier et al., 2013) and is even more common due to varying disease pressure throughout field trials (Huang et al., 2016; Larkan et al., 2016a; Kumar et al., 2018; Raman et al., 2018).

In most studies mapping quantitative blackleg resistance, different methods for disease scoring of cross sections at the crown or scoring of survival rate were used to estimate the disease severity (Huang et al., 2016; Larkan et al., 2016a; Gabur et al., 2018; Kumar et al., 2018; Raman et al., 2018; Raman R. et al., 2020). In the field trials of the present study, G2 index scoring of cross sections at the crown was applied due to the immense workload that needs to be achieved within a short period of time. However, the G2 index depends strongly on precise cuts to accurately assess the cross sections. This phenotyping method is time-saving compared to VDT scoring and provides useful and informative data from field evaluations of blackleg disease. In addition, to gain more detailed insights into the growth of the fungus, the VDT scoring method of Kutcher et al. (1993) was used for the assessments in the greenhouse trials. The high overlap of QTL results with previous studies underpins the relevance of the applied phenotyping method and further reveals that some of the identified QTL regions are crucial within international germplasm (Jestin et al., 2015; Raman et al., 2016, 2018; Kumar et al., 2018; Fikere et al., 2020; Raman H. et al., 2020; Raman R. et al., 2020; **Supplementary Table 11**).

We mapped quantitative resistance on chromosomes A09 and A07 in the multiparental population in the greenhouse. All QTL explained less than 5% of the phenotypic variation suggesting that genomic selection approaches are more suitable than marker-assisted selection for breeding toward quantitative *L. maculans* resistance in oilseed rape. Some of the QTL regions identified in the present study on chromosome A09 and A07 overlapped with previously described QTL (**Supplementary Table 11**). However, chromosome A07, in contrast to A09, is mainly known to harbor qualitative resistance genes providing race-specific resistances (Delourme et al., 2006). Although we ensured the specific assessment of quantitative resistance by selecting a fungal isolate with no gene-for-gene interaction with

the mapping population in our greenhouse trials, to our surprise a QTL explaining less than 5% of the phenotypic variance was identified in a genome region on chromosome A07 known to harbor a cluster of *R* genes (*Rlm3*, *Rlm4*, *Rlm7, and Rlm9*) involved in qualitative resistance expression. This suggests that this genome region is either a genomic hot spot where genes involved in qualitative as well as quantitative resistance are tightly clustered (linkage) or that some genes known to be involved in major qualitative resistance can also impart quantitative effects on resistance. Although it is possible that linkage exists between *R* gene clusters which are non-functional at the cotyledon stage in our population and genes involved explicitly in quantitative resistance expression in the LD block on chromosome A07, no clear candidate genes could be identified by GO analysis or polymorphism detection between the parental genotypes. The most striking polymorphism was detected in the well described major resistance gene *Rlm9*, which could suggest that this gene is also involved in adult plant resistance expression in the investigated multiparental population. In addition, *Rlm9* co-localized with an overlapping QTL region in a diversity set for field resistance in adult plants subjected to infection by an *L. maculans* population with an avirulent *AvrRlm9* gene composition. This observation supports the involvement of this candidate gene in quantitative resistance expression in these environments and populations.

This result is in accordance with previous reports that some major *Rlm*/*LepR* genes also have quantitative effects on adult plant resistance. For example, the fungal *AvrLmS-Lep2* gene-for-gene interaction with a *B. napus R* gene, *LepR2*, shows a qualitative intermediate resistance response at the cotyledon stage and partial resistance at the adult plant stage (Long et al., 2011; Dandena et al., 2019; Neik et al., 2020, preprint). The hypothesis that *Rlm9* is involved in quantitative resistance is also supported by Raman et al. (2018), who reported a quantitative resistance effect of the location harboring *Rlm9* in the greenhouse on adult plants, using a plant population segregating for *Rlm9* and an *L. maculans* isolate carrying a corresponding functional avirulence gene *AvrLm5-9* allele. That result also suggested that *Rlm9*-mediated resistance was expressed at the adult plant stage. The concept of *R* genes mediating quantitative resistance has been previously suggested and discussed in other studies that made similar observations (Chantret et al., 1999; Raman et al., 2018). A possible weak, constitutive expression of *R* genes at adult plant stage may lead to partial resistance with only minor effects. Our results suggest that further studies investigating *R* gene expression at adult plant stage or even accompanying an entire growth period could give deeper insights into these repeatedly observed findings.

Our detailed analysis of the molecular polymorphism for *Rlm9* in the parents of the mapping population also supports this hypothesis. In particular, non-synonymous amino acid changes, an inserted stop codon and especially the large insertion in *Rlm9* most likely alter the transcript or even interrupt the transcription of the gene in four of the seven parents of the mapping population. Usually, *R* genes encode nucleotide-binding site leucine-rich repeat (NBS-LRR) proteins (McHale et al., 2006). Wall-associated kinase (WAK) and wall-associated kinase-like (WAKL) genes are a newly discovered class of race-specific plant receptor-like kinase resistance genes involved in qualitative resistance (Larkan et al., 2020). However, although some WAKL genes have been shown to be involved in race-specific gene-for-gene interactions, for example *Rlm9* in oilseed rape and *Stb9* in wheat (Keller and Krattinger, 2018), other WAKL genes like *RFO1* (*RESISTANCE TO FUSARIUM OXYSPORUM 1*)/*WAKL22* in Arabidopsis have been found to be involved in broad-spectrum resistance against *F. oxysporum* f. sp. *matthioli* and other races (Diener and Ausubel, 2005) and against *Verticillium longisporum* (Johansson et al., 2006) in Arabidopsis, while *ZmWAK1* confers quantitative resistance to northern corn leaf blight in maize (Hurni et al., 2015). The protein structure of the WAKs and some WAKLs can be divided into an extracellular and an intracellular compartment, which are connected by a transmembrane domain. These proteins are characterized by a cytoplasmic Ser/Thr kinase domain in the cell interior and an extracellular domain that is similar to epidermal growth factor domains in vertebrates (epidermal growth factor; EGF-like domain). This occurs as a calcium-binding EGF domain (EGF-Ca^{2+}) and/or as an EGF2-like domain, although in some cases it is slightly degenerate (Verica and He, 2002). The exact function of these EGF-like domains is still largely unclear. However, previous studies demonstrated that EGF-like domains are involved in protein-protein interactions (Kuroda and Tanizawa, 1999). This extracellular domain is disrupted in four of the seven parents in our study by a 6 kb insertion, which might thus be expected to potentially interrupt (unknown) protein-protein-interactions and could consequently impart a quantitative impact on resistance activity. Other parts of the extracellular domain are bound to the pectin of the cell wall (Wagner and Kohorn, 2001), but can also serve as receptors for oligogalacturonides (OGs), which, among other things, arise from mechanical destruction of the pectin and act as DAMPs (damage-associated molecular patterns) to activate the plant immune system (Brutus et al., 2010). This part contains a conserved GUB_WAK domain (galacturonan-binding wall-associated receptor kinase). This suggests that WAKL genes might be involved in more broad-spectrum resistance in some pathogen-host interactions by sensing DAMPs. Hence, our results in oilseed rape suggest a possible dual function of *Rlm9*. On the one hand, *Rlm9* is expected to be involved in race-specific PTI in oilseed rape when challenged with *L. maculans* isolates carrying *AvrLm5-9*, which trigger a strong qualitative gene-for-gene resistance effect in plant genotypes with a functional interacting *Rlm9* gene as in Raman et al. (2018). However, in contrast to Raman et al. (2018), the results in our study suggest that the gene-for-gene interaction for *Rlm9* is not only expressed at the cotyledon stage, but also to a lesser extent in adult plants. In contrast to Raman et al. (2018), we used a *L. maculans* isolate which did not harbor a corresponding avirulence gene (*avrLm5-9*). Thus, even though no gene-for-gene interaction of *Rlm9* with its corresponding avirulence gene is expected in our experiment, still we found a weak expression of quantitative resistance of 5%. This indicates that *Rlm9* may have other additional features triggering quantitative resistance, for example by sensing DAMPs. Only if some functional domains are disrupted, like the EGF-like domain in four of the seven

genotypes in our study, might this weak quantitative resistance effect also be lost. All in all, the results of the present study confer with previous studies that observed *R* gene mediated resistance under controlled and field conditions at adult plant stages in *B. napus* (Delourme et al., 2004; Raman et al., 2012). Based on these collective observations, we hypothesize that the role of *Rlm9* in this interaction deserves more detailed functional analysis in future.

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 (accession: PRJNA751459).

AUTHOR CONTRIBUTIONS

CO and RS conceived the idea and sourced the funding. PV, HC, and CO developed the methodology. PV generated the genetic and field data. DA and BK generated the greenhouse data. PV and CO performed data curation. PV, HC, and HL analyzed the sequence and SV data. PV, SW, LE, and IG performed the quantitative genetic analysis. PV, RS, and CO drafted and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

PV and CO received funding from the German Federal Ministry of Education and Research (BMBF) within the project PhomaDur (Grant Number BLE: 2818205115). This work was also supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, and 031A532B).

ACKNOWLEDGMENTS

We thank Regina Illgner, Jutta Schaper, and Anita Popovska for excellent technical help. We also thank all German breeding companies organized in the GFPi (Gemeinschaft zur Förderung von Pflanzeninnovation e. V.) for production of the elite NAM population, for performing field trials, and for providing phenotyping data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.749491/full#supplementary-material

Supplementary Figure 1 | Phenotypic distribution of VDT values from greenhouse screenings for the entire mapping population **(A)** and the individual subfamilies: Adriana × Lorenz **(B)**, Lorenz × Alpaga **(C)**, Lorenz × DK Cabernet **(D)**, Lorenz × Galileo **(E)**, JN × Lorenz **(F)**, and King 10 × Lorenz **(G)**.

REFERENCES

Aubertot, J.-N., Schott, J.-J., Penaud, A., Brun, H., and Doré, T. (2004). Methods for sampling and assessment in relation to the spatial pattern of phoma stem canker (*Leptosphaeria maculans*) in oilseed rape. *Eur. J. Plant Pathol.* 110, 183–192. doi: 10.1023/B:EJPP.0000015359.61910.3b

Aulchenko, Y. S., Koning, D.-J. D., and Haley, C. (2007). Genomewide Rapid Association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait Loci Association analysis. *Genetics* 177, 577–585. doi: 10.1534/genetics.107.075614

Balesdent, M. H., Attard, A., Ansan-Melayah, D., Delourme, R., Renard, M., and Rouxel, T. (2001). Genetic Control and host range of avirulence toward *Brassica napus* cultivars quinta and jet neuf in *Leptosphaeria maculans*. *Phytopathology* 91, 70–76. doi: 10.1094/PHYTO.2001.91.1.70

Balesdent, M. H., Attard, A., Kühn, M. L., and Rouxel, T. (2002). New avirulence genes in the phytopathogenic fungus *Leptosphaeria maculans*. *Phytopathology* 92, 1122–1133. doi: 10.1094/PHYTO.2002.92.10.1122

Balesdent, M. H., Barbetti, M. J., Li, H., Sivasithamparam, K., Gout, L., and Rouxel, T. (2005). Analysis of *Leptosphaeria maculans* race structure in a worldwide collection of isolates. *Phytopathology* 95, 1061–1071. doi: 10.1094/PHYTO-95-1061

Balesdent, M.-H., Louvard, K., Pinochet, X., and Rouxel, T. (2006). A large-scale survey of races of *Leptosphaeria maculans* occurring on oilseed rape in France. *Eur. J. Plant Pathol.* 114, 53–65. doi: 10.1007/s10658-005-2104-0

Bayer, P. E., Hurgobin, B., Golicz, A. A., Chan, C.-K. K., Yuan, Y., Lee, H., et al. (2017). Assembly and comparison of two closely related *Brassica napus* genomes. *Plant Biotechnol. J.* 15, 1602–1610. doi: 10.1111/pbi.12742

Brun, H., Chèvre, A.-M., Fitt, B. D. L., Powers, S., Besnard, A.-L., Ermel, M., et al. (2010). Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. *New Phytol.* 185, 285–299. doi: 10.1111/j.1469-8137.2009.03049.x

Brutus, A., Sicilia, F., Macone, A., Cervone, F., and de Lorenzo, G. (2010). A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9452–9457. doi: 10.1073/pnas.1000675107

Bus, A., Körber, N., Snowdon, R. J., and Stich, B. (2011). Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. *Theor. Appl. Genet.* 123, 1413–1423. doi: 10.1007/s00122-011-1676-7

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435

Chantret, N., Pavoine, M. T., and Doussinault, G. (1999). The race-specific resistance gene to powdery mildew, MlRE, has a residual effect on adult plant resistance of winter wheat line RE714. *Phytopathology* 89, 533–539. doi: 10.1094/PHYTO.1999.89.7.533

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

Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., and Town, C. D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 89, 789–804. doi: 10.1111/tpj.13415

Corwin, J. A., and Kliebenstein, D. J. (2017). Quantitative resistance: more than just perception of a pathogen. *Plant Cell* 29, 655–665. doi: 10.1105/tpc.16.00915

Dandena, H. B., Zhang, Q., Zhou, T., Hirani, A. H., Liu, Z., Fernando, D. W. G., et al. (2019). Analysis of quantitative adult plant resistance to blackleg in *Brassica napus*. *Mol. Breed.* 39:124. doi: 10.1007/s11032-019-1035-y

Delourme, R., Bousset, L., Ermel, M., Duffé, P., Besnard, A. L., Marquer, B., et al. (2014). Quantitative resistance affects the speed of frequency increase but not

the diversity of the virulence alleles overcoming a major resistance gene to *Leptosphaeria maculans* in oilseed rape. *Infect. Genet. Evol.* 27, 490–499. doi: 10.1016/j.meegid.2013.12.019

Delourme, R., Chèvre, A. M., Brun, H., Rouxel, T., Balesdent, M. H., Dias, J. S., et al. (2006). Major gene and polygenic resistance to *Leptosphaeria maculans* in oilseed rape (*Brassica napus*). *Eur. J. Plant Pathol.* 114, 41–52. doi: 10.1007/s10658-005-2108-9

Delourme, R., Pilet-Nayel, M.-L., Archipiano, M., Horvais, R., Tanguy, X., Rouxel, T., et al. (2004). A cluster of major specific resistance genes to *Leptosphaeria maculans* in *Brassica napus*. *Phytopathology* 94, 578–583. doi: 10.1094/phyto.2004.94.6.578

Delplace, F., Huard-Chauveau, C., Dubiella, U., Khafif, M., Alvarez, E., Langin, G., et al. (2020). Robustness of plant quantitative disease resistance is provided by a decentralized immune network. *Proc. Natl. Acad. Sci. U.S.A.* 117, 18099–18109. doi: 10.1073/pnas.2000078117

Diener, A. C., and Ausubel, F. M. (2005). RESISTANCE TO FUSARIUM OXYSPORUM 1, a dominant *Arabidopsis* disease-resistance gene, is not race specific. *Genetics* 171, 305–321. doi: 10.1534/genetics.105.042218

Elliott, V. L., Marcroft, S. J., Howlett, B. J., and van de Wouw, A. P. (2016). Gene-for-gene resistance is expressed in cotyledons, leaves and pods, but not during late stages of stem colonization in the *Leptosphaeria maculans*-*Brassica napus* pathosystem. *Plant Breed.* 135, 200–207. doi: 10.1111/pbr.12343

Fikere, M., Barbulescu, D. M., Malmberg, M. M., Spangenberg, G. C., Cogan, N. O. I., and Daetwyler, H. D. (2020). Meta-analysis of GWAS in canola blackleg (*Leptosphaeria maculans*) disease traits demonstrates increased power from imputed whole-genome sequence. *Sci. Rep.* 10:14300. doi: 10.1038/s41598-020-71274-6

Fitt, B. D. L., Brun, H., Barbetti, M. J., and Rimmer, S. R. (2006). World-wide importance of phoma stem canker (*Leptosphaeria maculans* and *L. biglobosa*) on oilseed rape (*Brassica napus*). *Eur. J. Plant Pathol.* 114, 3–15. doi: 10.1007/s10658-005-2233-5

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

Gabur, I., Chawla, H. S., Lopisso, D. T., Tiedemann, A. V., Snowdon, R. J., and Obermeier, C. (2020). Gene presence-absence variation associates with quantitative *Verticillium longisporum* disease resistance in *Brassica napus*. *Sci. Rep.* 10:4131. doi: 10.1038/s41598-020-61228-3

Huang, Y. J., Jestin, C., Welham, S. J., King, G. J., Manzanares-Dauleux, M. J., Fitt, B. D. L., et al. (2016). Identification of environmentally stable QTL for resistance against *Leptosphaeria maculans* in oilseed rape (*Brassica napus*). *Theor. Appl. Genet.* 129, 169–180. doi: 10.1007/s00122-015-2620-z

Huang, Y. J., Pirie, E. J., Evans, N., Delourme, R., King, G. J., and Fitt, B. D. L. (2009). Quantitative resistance to symptomless growth of *Leptosphaeria maculans* (phoma stem canker) in *Brassica napus* (oilseed rape). *Plant Pathol.* 58, 314–323. doi: 10.1111/j.1365-3059.2008.01957.x

Hurgobin, B., Golicz, A. A., Bayer, P. E., Chan, C.-K. K., Tirnaz, S., Dolatabadian, A., et al. (2018). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol. J.* 16, 1265–1274. doi: 10.1111/pbi.12867

Hurni, S., Scheuermann, D., Krattinger, S. G., Kessel, B., Wicker, T., Herren, G., et al. (2015). The maize disease resistance gene Htn1 against northern corn leaf blight encodes a wall-associated receptor-like kinase. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8780–8785. doi: 10.1073/pnas.1502522112

Jestin, C., Bardol, N., Lodé, M., Duffé, P., Domin, C., Vallée, P., et al. (2015). Connected populations for detecting quantitative resistance factors to phoma stem canker in oilseed rape (*Brassica napus* L.). *Mol. Breed.* 35:167. doi: 10.1007/s11032-015-0356-8

Johansson, A., Staal, J., and Dixelius, C. (2006). Early responses in the *Arabidopsis*-*Verticillium longisporum* pathosystem are dependent on NDR1, JA- and ET-associated signals via cytosolic NPR1 and RFO1. *Mol. Plant Microbe Interact.* 19, 958–969. doi: 10.1094/MPMI-19-0958

Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286

Keller, B., and Krattinger, S. G. (2018). A new player in race-specific resistance. *Nat. Plants* 4, 197–198. doi: 10.1038/s41477-018-0126-9

Kumar, V., Paillard, S., Fopa Fomeju, B., Falentin, C., Deniot, G., Baron, C., et al. (2018). Multi-year linkage and association mapping confirm the high number of genomic regions involved in oilseed rape quantitative resistance to blackleg. *Theor. Appl. Genet.* 131, 1627–1643. doi: 10.1007/s00122-018-3103-9

Kuroda, S., and Tanizawa, K. (1999). Involvement of epidermal growth factor-like domain of NELL proteins in the novel protein-protein interaction with protein kinase C. *Biochem. Biophys. Res. Commun.* 265, 752–757. doi: 10.1006/bbrc.1999.1753

Kutcher, H. R., van den Berg, C. G. J., and Rimmer, S. R. (1993). Variation in pathogenicity of *Leptosphaeria maculans* on *Brassica* spp. based on cotyledon and stem reactions. *Can. J. Plant Pathol.* 15, 253–258. doi: 10.1080/07060669309501920

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmertest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lancashire, P. D., Bleiholder, H., van den Boom, T., Langelüddeke, P., Stauss, R., Elfriede, W., et al. (1991). A uniform decimal code for growth stages of crops and weeds. *Ann. Appl. Biol.* 1991, 561–601. doi: 10.1111/j.1744-7348.1991.tb04895.x

Larkan, N. J., Ma, L., Haddadi, P., Buchwaldt, M., Parkin, I. A. P., Djavaheri, M., et al. (2020). The *Brassica napus* wall-associated kinase-like (WAKL) gene Rlm9 provides race-specific blackleg resistance. *Plant J.* 104, 892–900. doi: 10.1111/tpj.14966

Larkan, N. J., Raman, H., Lydiate, D. J., Robinson, S. J., Yu, F., Barbulescu, D. M., et al. (2016a). Multi-environment QTL studies suggest a role for cysteine-rich protein kinase genes in quantitative resistance to blackleg disease in *Brassica napus*. *BMC Plant Biol.* 16:183. doi: 10.1186/s12870-016-0877-2

Larkan, N. J., Yu, F., Lydiate, D. J., Rimmer, S. R., and Borhan, M. H. (2016b). Single R gene introgression lines for accurate dissection of the *Brassica*–*Leptosphaeria* pathosystem. *Front. Plant Sci.* 7:1771. doi: 10.3389/fpls.2016.01771

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *J. Stat. Softw.* 69, 1–33. doi: 10.18637/jss.v069.i01

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

Long, Y., Wang, Z., Sun, Z., Fernando, D. W. G., McVetty, P. B. E., and Li, G. (2011). Identification of two blackleg resistance genes and fine mapping of one of these two genes in a *Brassica napus* canola cultivar 'Surpass 400'. *Theor. Appl. Genet.* 122, 1223–1231. doi: 10.1007/s00122-010-1526-z

Luo, R., Wong, C.-L., Wong, Y.-S., Tang, C.-I., Liu, C.-M., Leung, C.-M., et al. (2020). Exploring the limit of using a deep neural network on pileup data for germline variant calling. *Nat. Mach. Intell.* 2, 220–227. doi: 10.1038/s42256-020-0167-4

Marcroft, S. J., Elliott, V. L., Cozijnsen, A. J., Salisbury, P. A., Howlett, B. J., and van de Wouw, A. P. (2012). Identifying resistance genes to *Leptosphaeria maculans* in Australian *Brassica napus* cultivars based on reactions to isolates with known avirulence genotypes. *Crop Pasture Sci.* 63, 338–350. doi: 10.1071/CP11341

McHale, L., Tan, X., Koehl, P., and Michelmore, R. W. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 7:212. doi: 10.1186/gb-2006-7-4-212

Neik, T. X., Ghanbarnia, K., Ollivier, B., Scheben, A., Severn-Ellis, A., Larkan, N. J., et al. (2020). Two independent approaches converge to the cloning of a new *Leptosphaeria maculans* avirulence effector gene, AvrLmS-Lep2. *bioRxiv* [Preprint]. doi: 10.1101/2020.10.02.322479

Obermeier, C., Hossain, M. A., Snowdon, R., Knüfer, J., von Tiedemann, A., and Friedt, W. (2013). Genetic analysis of phenylpropanoid metabolites associated with resistance against *Verticillium longisporum* in *Brassica napus*. *Mol. Breed.* 31, 347–361. doi: 10.1007/s11032-012-9794-8

Parkin, I. A., Sharpe, A. G., Keith, D. J., and Lydiate, D. J. (1995). Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). *Genome* 38, 1122–1131. doi: 10.1139/g95-149

Pilet, M. L., Delourme, R., Foisset, N., and Renard, M. (1998). Identification of QTL involved in field resistance to light leaf spot (*Pyrenopeziza brassicae*) and blackleg resistance (*Leptosphaeria maculans*) in winter rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* 97, 398–406. doi: 10.1007/s001220050909

Pilet-Nayel, M.-L., Moury, B., Caffier, V., Montarry, J., Kerlan, M.-C., Fournet, S., et al. (2017). Quantitative resistance to plant pathogens in pyramiding strategies for durable crop protection. *Front. Plant Sci.* 8:1838. doi: 10.3389/fpls.2017.01838

Raman, H., McVittie, B., Pirathiban, R., Raman, R., Zhang, Y., Barbulescu, D. M., et al. (2020). Genome-Wide Association mapping identifies novel Loci for

quantitative resistance to blackleg disease in canola. *Front. Plant Sci.* 11:1184. doi: 10.3389/fpls.2020.01184

Raman, H., Raman, R., Coombes, N., Song, J., Diffey, S., Kilian, A., et al. (2016). Genome-wide Association study identifies new loci for resistance to *Leptosphaeria maculans* in canola. *Front. Plant Sci.* 7:1513. doi: 10.3389/fpls.2016.01513

Raman, H., Raman, R., Diffey, S., Qiu, Y., McVittie, B., Barbulescu, D. M., et al. (2018). Stable quantitative resistance loci to blackleg disease in canola (*Brassica napus* L.) over continents. *Front. Plant Sci.* 9:1622. doi: 10.3389/fpls.2018.01622

Raman, R., Diffey, S., Barbulescu, D. M., Coombes, N., Luckett, D., Salisbury, P., et al. (2020). Genetic and physical mapping of loci for resistance to blackleg disease in canola (*Brassica napus* L.). *Sci. Rep.* 10:4416. doi: 10.1038/s41598-020-61211-y

Raman, R., Taylor, B., Marcroft, S., Stiller, J., Eckermann, P., Coombes, N., et al. (2012). Molecular mapping of qualitative and quantitative loci for resistance to *Leptosphaeria maculans* causing blackleg disease in canola (*Brassica napus* L.). *Theor. Appl. Genet.* 125, 405–418. doi: 10.1007/s00122-012-1842-6

Rimmer, S. R., and van den Berg, C. G. J. (1992). Resistance of oilseed *Brassica* spp. to blackleg caused by *Leptosphaeria maculans*. *Can. J. Plant Pathol.* 14, 56–66. doi: 10.1080/07060669209500906

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, 1–16. doi: 10.1093/gigascience/giaa137

Rouxel, T., Penaud, A., Pinochet, X., Brun, H., Gout, L., Delourme, R., et al. (2003). A 10-year survey of populations of *Leptosphaeria maculans* in France indicates a rapid adaptation towards the Rlm1 resistance gene of oilseed rape. *Eur. J. Plant Pathol.* 109, 871–881. doi: 10.1023/A:1026189225466

Samans, B., Chalhoub, B., and Snowdon, R. J. (2017). Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species *Brassica napus*. *Plant Genome* 10, 1–15. doi: 10.3835/plantgenome2017.02.0013

Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., Haeseler, A. V., et al. (2018). Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* 15, 461–468. doi: 10.1038/s41592-018-0001-7

Sprague, S. J., Balesdent, M.-H., Brun, H., Hayden, H. L., Marcroft, S. J., Pinochet, X., et al. (2006). Major gene resistance in *Brassica napus* (oilseed rape) is overcome by changes in virulence of populations of *Leptosphaeria maculans* in France and Australia. *Eur. J. Plant Pathol.* 114, 33–40. doi: 10.1007/s10658-005-3683-5

St Clair, D. A. (2010). Quantitative disease resistance and quantitative resistance loci in breeding. *Annu. Rev. Phytopathol.* 48, 247–268. doi: 10.1146/annurev-phyto-080508-081904

Stein, A., Coriton, O., Rousseau-Gueutin, M., Samans, B., Schiessl, S. V., Obermeier, C., et al. (2017). Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. *Plant Biotechnol. J.* 15, 1478–1489. doi: 10.1111/pbi.12732

Szadkowski, E., Eber, F., Huteau, V., Lodé, M., Huneau, C., Belcram, H., et al. (2010). The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytol.* 186, 102–112. doi: 10.1111/j.1469-8137.2010.03182.x

Thomma, B., Nürnberger, T., and Joosten, M. H. A. J. (2011). Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *Plant Cell* 23, 4–15. doi: 10.1105/tpc.110.082602

Nagaharu, U. (1935). Genome analysis in *Brassica* with special reference to the experimental formation of *Brassica napus* and peculiar mode of fertilization. *Jpn. J. Bot.* 7, 389–453.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35, W71–W74. doi: 10.1093/nar/gkm306

van de Wouw, A. P., Howlett, B. J., and Idnurm, A. (2017). Changes in allele frequencies of avirulence genes in the blackleg fungus, *Leptosphaeria maculans*, over two decades in Australia. *Crop Pasture Sci.* 69, 20. doi: 10.1071/CP16411

Verica, J. A., and He, Z. (2002). The cell Wall-Associated Kinase (WAK) and WAK-like kinase gene family. *Plant Physiol.* 129, 455–459.

Wagner, T. A., and Kohorn, B. D. (2001). Wall-Associated Ainases are expressed throughout plant development and are required for cell expansion. *Plant Cell* 13, 303–318. doi: 10.1105/tpc.13.2.303

West, J. S., Kharbanda, P. D., Barbetti, M. J., and Fitt, B. D. L. (2001). Epidemiology and management of *Leptosphaeria maculans* (phoma stem canker) on oilseed rape in Australia, Canada and Europe. *Plant Pathol.* 50, 10–27. doi: 10.1046/j.1365-3059.2001.00546.x

Winter, M., and Koopmann, B. (2016). Race spectra of *Leptosphaeria maculans*, the causal agent of blackleg disease of oilseed rape, in different geographic regions in northern Germany. *Eur. J. Plant Pathol.* 145, 629–641. doi: 10.1007/s10658-016-0932-8

Zhang, X., Peng, G., Kutcher, H. R., Balesdent, M.-H., Delourme, R., and Fernando, W. G. D. (2016). Breakdown of Rlm3 resistance in the *Brassica napus*–*Leptosphaeria maculans* pathosystem in western Canada. *Eur. J. Plant Pathol.* 145, 659–674. doi: 10.1007/s10658-015-0819-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

4 Discussion

4.1 Implementation of multiparental mapping populations

Multiparental populations are valuable plant breeding resources used to identify chromosome regions and alleles with effects on important agronomical traits in crops. The analyses presented in Chapters 2 and 3 are based on a multiparental DH population derived from crosses of six elite founder lines to one common parental line. This particular population can also be described as nested population with multiple biparental crosses sharing a common parent that links the subpopulations. The six founder lines were selected due to their known broad quantitative blackleg resistance, whereas the common parent was known to be highly susceptible towards blackleg. This particular design allowed the identification of eight QTL associated with quantitative blackleg resistance, including four novel QTL and four QTL that have been already described in previous studies. Two hundred and thirty-five SV were found to be located within these QTL regions. The association of several intragenic SV with blackleg resistance suggested the potential impact of SV in expression of disease resistance. In particular, a large novel insertion within the coding DNA sequence (CDS) of the well-characterized *R* gene *Rlm9*, which was found at the peak of a QTL for blackleg resistance, supports the hypothesis that SV can cause novel variation for quantitative resistance.

Differences regarding the onset of flowering were observed during the field trials and thereafter assessed and analyzed. Eight QTL were found to be associated with flowering time modulation. Within these genomic regions we found 52 intragenic SV, thus suggesting that SV affects flowering time modulation in oilseed rape. This corresponds to the findings of Chawla et al. (2020), who discovered that genome-wide SV clearly differentiate winter,

spring and semi-winter *B. napus* ecotypes, indicating that small to mid-size SV events in flowering pathway genes were a key factor in ecogeographical adaptation between different *B. napus* ecotypes.

This particular multiparental intercross design provides several advantages in comparison to biparental crosses and diversity panels (Huang et al., 2011). On the one hand, association mapping using multiparental mapping populations has greater statistical power compared to the use of diversity panels because allele frequencies are typically closer to the ideal 1:1 segregation. On the other hand, it provides larger genetic diversity, higher resolution and thus mapping precision due to the higher number of recombinations when compared to populations derived from biparental crosses. Usually, biparental populations have only very few recombinations per chromosome, leading to large LD blocks. Both power and resolution of the mapping approach are determined by the size of the mapping population and the number of markers, in a proportional manner. Additionally, the population structure, which is incorporated in the underlying linear models, is comparatively small for multiparental populations. This helps to prevent inflation of QTL significance (Korte and Farlow, 2013; Pascual et al., 2016; Scott et al., 2020).

The present thesis also revealed genetic diversity within the narrow gene pool of European elite winter oilseed rape germplasm. Genomic variants associated with quantitative traits derived from highly adapted cultivars demonstrate the immense potential of variant identification and utilization in commercial breeding programs. Another main advantage of the design of the mapping population was the use of well-characterized cultivars from German oilseed rape breeding companies. The selection of founder lines carrying no

effective *R* genes allowed the mapping of QTL for quantitative blackleg resistance by avoiding masking effects of major *R* genes. In contrast, large diversity panels are usually less well-characterized and potentially carry highly effective *R* genes. In such materials, major gene-for-gene interactions between *R* genes and their corresponding avirulence genes tend to mask quantitative effects. In addition, mapping QTL in populations derived from current breeding germplasm enhances the reliability and thus the applicability of discovered QTL, especially since minor-effect QTL can be unstable across different germplasm (Kang et al., 2008; Liu et al., 2010).

4.2 Implications of structural genome variation

Numerous previous studies have demonstrated SV detection in plants and their implications in important agronomical traits by application of short-read sequencing technologies. For instance, the association of presence-absence variation (PAV) with disease resistance in rice (*Oryza sativa*) (Xu et al., 2011) and stress responses in soybean (*Glycine max*) (McHale et al., 2012) have been demonstrated using Illumina short-read sequencing. Another study demonstrated the impact of a deletion in the putative promoter region of the *Ppd-1* gene that was associated with heading time in wheat (*Triticum aestivum*) (Nishida et al., 2013) by the use of Sanger sequencing. Also in oilseed rape there are several examples of PAV affecting flowering time (Schiessl et al., 2017), disease resistance (Gabur et al., 2018) or other traits like stay-green (Qian et al., 2016).

Even though short-reads are usually highly accurate, these methods lack sensitivity for detection of small to mid-scale SV due to unspecific alignments to reference genomes as a

consequence of the short read lengths (Sedlazeck et al., 2018; Chawla et al., 2020). For this reason, the re-sequencing approach of the present work for detection of deletions and insertions is based on Oxford Nanopore Technology (ONT) long-read sequencing. ONT sequencing resulted in single reads longer than 300 kb and provided an N50 of above 30 kb. Thus, depending on the coverage of a particular region, individual SV often were captured and confirmed by a number of reads spanning the entire length of the variant. The analyses in Chapter 2 are based on the widely used *B. napus* reference genome Darmor-*bzh* version 4.1 (Chalhoub et al., 2014). To further enhance the reliability of variant detection in the study in Chapter 3, we made use of one of the latest and most contiguous reference genome assemblies existing for *B. napus*. Darmor-*bzh* version 10 (Rousseau-Gueutin et al., 2020), which was assembled using long-read sequencing data, improved significantly in terms of completeness and continuity in comparison to previous versions 4.1 (Chalhoub et al., 2014) and 8.1 (Bayer et al., 2017). The use of long-read sequencing led to substantial improvements especially for the assembly of pericentromeric regions, improved gene completion and unraveled large misassembled inverted regions (Rousseau-Gueutin et al., 2020). This can be clearly seen through the great discrepancy between the SV results of the two reference genomes that were used in Chapters 2 and 3.

A comparison of these two reference genome assemblies for the total number of SV and polymorphic SV between the parents of the multiparental population disclosed great differences. Alignment to Darmor-*bzh* v4.1 revealed a total of 50,760 SV, but only 13,374 that were polymorphic between the parental lines of the mapping population (26.3%). In comparison, the alignment to Darmor-*bzh* v10 revealed in total 35,829 SV, of which 25,233

where polymorphic (70.4%). The high number of monomorphic SV when using version 4.1 led to the assumption that a certain number of SV calls might be due to assembly errors of this reference genome. This assumption was later proven by aligning long-read sequencing data from Darmor-*bzh* v10 to the reference version 4.1, which revealed 13,533 SV calls. This number is approximately the difference between the number of SV called using Darmor-*bzh* v4.1 and Darmor-*bzh* v10 (14,931). As these two assemblies were derived from the same genotype, neglecting the possibility of different seed lots contaminated during propagation and used for sequencing of the same accession and other possible pitfalls, this high number of SV can be predominantly explained by assembly errors of the older assembly (v4.1) that did not use any long-read data. More detailed numbers of SV can be taken from Chapters 2 and 3. However, referring to Darmor-*bzh* v10 we found 5,177 genes out of 108,190 predicted genes (4.8%) to harbor polymorphic SV. These numbers are in line with a comparable study by Chawla et al. (2020), who reported that approximately 5% of the genes in the winter oilseed rape accession 'Express 617' are affected by SV, though Darmor-*bzh* v4.1 was used in this study. However, the same study also reported up to 10% of genes affected by SV in other genotypes, including semi-winter, spring and synthetic accessions.

This thesis also provides an idea of the reliability of SNV detection from long-read sequencing data using a deep neural network based variant caller. ONT sequencing data has been stated to be inadequate for SNV or even small variant identification due to poor single base accuracies (Amarasinghe et al., 2020; Zhang et al., 2020). However, the software tool Clair provides a special algorithm that has been specifically developed to

overcome these limitations (Ameur et al., 2019; Luo et al., 2020). Even though single ONT reads can be inaccurate, an appropriate coverage providing a reliable consensus sequence in combination with the Clair variant caller yielded reliable SNV calls, as the results in Chapter 3 indicate. Validation of called SNV in a particular genomic region by Sanger sequencing resulted in a confirmation rate of 100%, suggesting that this approach is extremely reliable. However, the practice of strict filtering also led to a certain number of false negatives; in other words, a number of true SNV were not detected. Further studies are needed to confirm these results and to further adapt quality filtering with respect to data quality and the specific aim of a particular analysis.

Chapters 2 and 3 demonstrate several examples of intragenic deletions and insertions that are associated with flowering time modulation and quantitative blackleg resistance. Prioritization of candidate genes was performed according to the presence of variants within the CDS. Even if the associated SV were not causal for the observed phenotypic effect, they might still represent useful marker targets that can be utilized in commercial breeding.

Even though long-read sequencing is rapidly improving, it is still characterized by a higher error rate in comparison to short-read sequencing (Lu et al., 2016; Mahmoud et al., 2019). However, with further ongoing technical improvements along with declining costs, long-read sequencing may become a standard sequencing approach particularly for large and complex crop genomes. With respect to all advantages stated in this thesis, long reads enhance detection accuracy of SV in crop plants and moreover allow deeper insights towards the understanding of their implications in agricultural traits.

4.3 Future prospects in genomic research

Even though this thesis presents several examples of genetic variation associating with agronomical traits, it nevertheless illustrates drawbacks resulting from the use of a single reference genome for data analyses. Using a single linear reference always implies the bias that each SV that was identified represents only a polymorphism between the sequenced genotype and one single reference genotype. In contrast, potential variation that is equal in both genotypes is undetectable. Pan-genome assemblies represent a promising approach to overcome this problem. The basic strategy of pan-genome approaches is to compare multiple individual genomes in order to differentiate their core genome, representing the proportion present in all individuals, from their dispensable genome, which is not present in all individuals (Golicz et al., 2019). In plants, the first pan-genome was developed for *Glycine soja* (Li et al., 2014), who demonstrated that ~80% of the genome were similar among seven individual genotypes and thus represent the core genome. Obviously, the definition of the core genome depends on the germplasm used to build a particular pan-genome. In wheat, 64.3% of the genes of a pan-genome created from 18 wheat cultivars were determined as core genes, whereas that remainder were reported to be dispensable and displayed PAV within this particular set of wheat cultivars (Montenegro et al., 2017). A *B. napus* pan-genome generated using eight accessions of three different ecotypes revealed ~56% core-gene clusters, ~42% dispensable gene clusters and ~2% genotype-specific gene clusters (Song et al., 2020). The functional genetic variation that was shown in these studies would have been largely unrevealed if only a single reference genome had been used for

comparisons. These results demonstrate the additional value that pan-genomics is capable of contributing to research and breeding.

However, the application of pan-genomes still lacks standardized methodology. To date, a pan-genome can be simplified as an accumulation of various genomes. In order to make pan-genomes more accessible to genetic analyses, the construction of reference pan-genome graphs was proposed in human genetics (Li et al., 2020). A reference graph would overcome the bias of a single linear reference by representation of an assembly of multiple genomes in a specific format. According to the authors of *minigraph*, a reference pan-genome graph would serve as a valuable resource of complex genetic variation that is applicable to genetic analyses. However, there are applications where no benefit is expected from pan-genomes. Thus, the authors suggest it as an addition to commonly used reference genomes (Li et al., 2020).

Broader "pan-omic" approaches, in which pan-genomes are complemented by pan-transcriptomes and pan-proteomes, have enormous potential to enhance understanding of regulatory mechanisms in plants and the exploration of gene functions and pathways (Jaiswal et al., 2020). Considering constantly decreasing prices and improving accuracies of modern sequencing technologies, the availability of "pan-omic" data for crop plants can be expected to increase in the near future, as will the number of individual accessions contributing to a single pan-genome in any given species.

4.4 Future prospects of oilseed rape breeding

There are several challenges associated with breeding of oilseed rape in Europe. Current difficulties for growers include management of volunteer oilseed rape, providing ideal sowing conditions with special focus on weed control, water regime and crop management during the season. Unlike optimal fertilization and appropriate weed control, the control and treatment of fungal diseases represents a particular example where growers are heavily reliant on innovation in plant breeding. Increasing restrictions on active plant protection substances in European countries in recent years have already limited the number of available fungicides. In coming years these restrictions are expected to become even more stringent. Also in regard to other crop inputs, environmental impacts of modern agriculture are increasingly monitored by strict policy controls and public scrutiny, demanding efficient crops with low environmental impacts. Furthermore, the demand for organic products is likely to increase due to a positive perception in the public mind. All these factors are great motivators for plant breeders to meet growing demand for safe and healthy food by developing and providing high yielding cultivars harboring resistance or tolerance to biotic and abiotic stresses. Independent of the relevant breeding aim, modern breeding methods are essential in this regard.

Genetic variation is the primary prerequisite for plant breeding. Although this thesis only provides one example in an individual crop, it nevertheless demonstrates the high frequency and importance of SV as a valuable source of genetic variation for breeding for crop adaptation and resilience. Independent of the trait and the breeding method, SV are an additional source of genomic variation that should and can definitely be targeted by

breeders. Recent studies already demonstrated that the consideration of SV for GS led to an increase in prediction accuracy (Weisweiler et al., 2019). For adaptation of quantitative traits, which are controlled by many QTL with low effects, GS is rapidly becoming the method of choice to improve future cultivars.

In many agriculture crop production regions around the globe, oilseed rape is the only profitable leaf crop and therefore an inevitable member of common crop rotations. Plant breeding remains the most important technique towards maintaining this status by countering the multiple future challenges to which oilseed rape is exposed.

4.5 Conclusion

As sequencing prices continue to drop, increasing numbers of studies are describing genetic variation in all kinds of crop and non-crop plants. Today it is already feasible to sequence entire mapping populations or diversity collection with several hundreds or thousands of individuals, revealing increasingly deeper insights about the underlying genomic variation. Not only costs, but also methodology and sequencing accuracies, reliability and repeatability are constantly improving. The present thesis demonstrates the identification of novel genetic variants in the allopolyploid crop *B. napus*, giving examples for two economically important traits. Furthermore, results presented in Chapters 2 and 3 propose the consideration and the use of structural genome variants in commercial breeding programs. QTL for flowering time and quantitative blackleg resistance identified in elite populations can be used by breeders to improve modern varieties by combination of desirable alleles that already exist within elite breeding germplasm. However, it still will always be a main target to introgress new diversity from less adapted materials or related species and to make favorable alleles accessible to breeding programs.

The success of SV investigation applying a re-sequencing approach depends strongly on the quality of available reference genomes. Different references lead to varying results. Thus, pan-genomics will fundamentally upgrade plant sciences. The definition of core and dispensable genomes optimizes the investigation of PAV within a species. Further research regarding the identification of candidate genes and the understanding of their underlying mechanisms and gene networks they are involved in will particularly benefit from extended "pan-omics" approaches.

5 Summary

Brassica napus L. (oilseed rape) is an allotetraploid crop originated from a spontaneous hybridization event between its two diploid ancestors *Brassica rapa* L. (turnip) and *Brassica oleracea* L. (cabbage). It is assumed that this interspecific hybridization took place no more than 7,000 years ago, when turnip rape and cabbage were first cultivated in close proximity to each other. The A and C subgenomes of this relatively recent polyploid crop share highly similar homoeologous chromosomes with conserved gene order. Thus, there is a high probability of homoeologous exchanges during meiosis. Hence, highly frequent structural genome variations (SV), including deletions, insertions, inversion and translocation are expected and have already been demonstrated in several studies.

In recent years, long-read sequencing technologies have added a powerful new dimension to eukaryote genomics. Long-read sequencing technologies combine great read lengths with acceptable single base accuracies and allow the detection of small- to mid-scale SV that could not be revealed by Illumina short-read sequencing datasets. Applying long-read sequencing to investigate an interconnected multiparental population of European elite winter oilseed rape revealed that ~ 5% of the annotated genes harbored a potentially functional SV event. This relatively high number of intragenic SV justified a more detailed investigation regarding the association of SV with agricultural traits.

In the present thesis, genome-wide association studies were performed to investigate flowering time modulation and quantitative blackleg resistance within a multiparental population derived from seven European elite winter oilseed rape cultivars. Several novel QTL were identified and previously known QTL regions were confirmed. Most interestingly, this work focused on SV within these QTL regions to evaluate the

implications of SV on the investigated traits. The two major findings were a novel, previously undetected deletion within *FLOWERING LOCUS T*, a key gene of the flowering time pathway and a large, novel insertion within *Rlm9*, a well-investigated blackleg resistance gene. These SV were associated to quantitative phenotypic variation for flowering time and blackleg resistance, respectively. Furthermore, the results provided new insight into the suitability of long-read sequencing data for detection of single-nucleotide variants and underlined the relevance of reference genome quality for SV analyses.

In summary, increasing accuracies and decreasing prices of constantly evolving long-read sequencing technologies make long-read sequencing the method of choice for detection of genomic variation. The unexpectedly high prevalence and functional relevance of SV and their association with important agronomical traits allow the recommendation for intensification of their use in commercial breeding programs and further emphasize their potential value for the development of modern, high yielding and stress tolerant cultivars.

6 Zusammenfassung

Brassica napus L. (Raps) ist eine allotetraploide Kulturpflanze, die aus einer spontanen Hybridisierung seiner beiden diploide Vorfahren *Brassica rapa* L. (Rübsen) und *Brassica oleracea* L. (Kohl) entstanden ist. Diese interspezifische Hybridisierung fand vor circa 7.000 Jahren statt, als Rübsen und Kohl erstmals in unmittelbarer Nähe zueinander angebaut wurden. Die A und C Subgenome dieser relativ jungen Kulturpflanze haben sehr ähnliche homöologe Chromosomen mit konservierter Anordnung der Gene. Dadurch ergibt sich eine hohe Wahrscheinlichkeit von homöologen Rekombinationen während der Meiose. Daher werden zahlreiche strukturelle Genomvariationen (SV), wie zum Beispiel Deletionen, Insertionen, Inversionen und Translokationen erwartet und wurden bereits in einigen Studien gezeigt.

In den vergangenen Jahren haben Long-Read-Sequenzierungstechnologien der eukaryotischen Genomforschung eine neue Dimension verliehen. Long-Read-Sequenzierungstechnologien vereinen extreme Längen einzelner DNA Moleküle mit akzeptabler Genauigkeit der einzelnen Basen und ermöglichen somit die Detektion von kleinen bis mittelgroßen SV, die bei der Verwendung von Illumina Datensätzen mit sehr kurzen Reads nicht nachgewiesen werden konnten. Die Verwendung der Oxford Nanopore Sequenzierungstechnologie zur Untersuchung einer multiparentalen, europäischen Elite-Winterraps Population offenbarte, dass circa 5 % der annotierten Gene eine potentiell funktionale SV tragen. Diese relativ hohe Anzahl intragenischer SV begründete eine detaillierte Untersuchung ihrer Assoziation mit agronomischen Merkmalen.

In der vorliegenden Arbeit wurden genomweite Assoziationskartierungen zur Untersuchung von Blühzeitpunkt und quantitativer *Phoma* Resistenz durchgeführt. Hierzu wurde eine multiparentale Population, die aus sieben europäischen Elite-Winterrapssorten entwickelt wurde, verwendet. Es konnten einige neue QTL identifiziert sowie einige bereits bekannte Regionen bestätigt werden. Der Fokus dieser Arbeit lag vor allem auf den SV innerhalb dieser Regionen, um die Auswirkungen von SV auf die untersuchten Merkmale bewerten zu können. Die beiden bedeutendsten Ergebnisse waren eine zuvor unbekannte Deletion innerhalb von *FLOWERING LOCUS T*, einem Schlüssel-Gen des Gen-Netzwerks zur Regulation der Blüte sowie eine große, neuartige Insertion innerhalb von *Rlm9*, einem sehr gut untersuchten *Phoma* Resistenzgens. Diese SV waren assoziiert mit der quantitativen phänotypischen Variation des Blühzeitpunkts bzw. der *Phoma* Resistenz. Des Weiteren lieferten die Ergebnisse dieser Arbeit eine Idee zur Verwendbarkeit von Long-Read-Sequenzierungstechnologien zur Detektion von Einzelbasen-Varianten und untermalten die Wichtigkeit der Qualität des Referenzgenoms für derartige Analysen.

Zusammenfassend lässt sich festhalten, dass die zunehmende Genauigkeit sowie das sinkende Preisniveau, die Long-Read-Sequenzierungstechnologie zur Methode der Wahl für die Detektion von genomischer Variation machen. Die unerwartet hohe Verbreitung und funktionelle Bedeutung von SV und deren Assoziation mit wichtigen agronomischen Merkmalen erlaubt die Empfehlung zur intensiveren Nutzung von SV in kommerziellen Züchtungsprogrammen. Überdies betont es den potentiellen Wert für die Entwicklung von modernen, stresstoleranten Sorten mit hohem Ertragspotential.

7 References

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME and Gouil Q (2020) Opportunities and challenges in long-read sequencing data analysis. *Genome biology* 21: 30.

Ameur A, Kloosterman WP and Hestand MS (2019) Single-Molecule Sequencing: Towards Clinical Applications. *Trends in biotechnology* 37: 72–85.

Balesdent MH, Attard A, Kühn ML and Rouxel T (2002) New Avirulence Genes in the Phytopathogenic Fungus *Leptosphaeria maculans*. *Phytopathology* 92: 1122–1133.

Bayer PE, Hurgobin B, Golicz AA, Chan C-KK, Yuan Y, Lee H, Renton M, Meng J, Li R, Long Y, Zou J, Bancroft I, Chalhoub B, King GJ, Batley J and Edwards D (2017) Assembly and comparison of two closely related *Brassica napus* genomes. *Plant biotechnology journal* 15: 1602–1610.

Berry PM, Spink J, Foulkes MJ and White PJ (2010) The physiological basis of genotypic differences in nitrogen use efficiency in oilseed rape (*Brassica napus* L.). *Field Crops Research* 119: 365–373.

Blümel M, Dally N and Jung C (2015) Flowering time regulation in crops—what did we learn from Arabidopsis? *Current opinion in biotechnology* 32: 121–129.

Borghi M, Perez de Souza L, Yoshida T and Fernie AR (2019) Flowers and climate change: a metabolic perspective. *New Phytologist* 224: 1425–1441.

Brun H, Chèvre A-M, Fitt BDL, Powers S, Besnard A-L, Ermel M, Huteau V, Marquer B, Eber F, Renard M and Andrivon D (2010) Quantitative resistance increases the

durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. *The New phytologist* 185: 285–299.

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Wang X, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C and Wincker P (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345: 950–953.

Chawla HS, Lee H, Gabur I, Vollrath P, Tamilselvan-Nattar-Amutha S, Obermeier C, Schiessl SV, Song J-M, Liu K, Guo L, Parkin IAP and Snowdon RJ (2020) Long-read sequencing reveals widespread intragenic structural variants in a recent allopolyploid crop plant. *Plant biotechnology journal* 19: 240–250.

Corwin JA and Kliebenstein DJ (2017) Quantitative Resistance: More Than Just Perception of a Pathogen. *The Plant cell* 29: 655–665.

Cowger C and Brown JKM (2019) Durability of Quantitative Resistance in Crops: Greater Than We Know? *Annual review of phytopathology* 57: 253–277.

Delourme R, Chèvre AM, Brun H, Rouxel T, Balesdent MH, Dias JS, Salisbury P, Renard M and Rimmer SR (2006) Major gene and polygenic resistance to *Leptosphaeria maculans* in oilseed rape (*Brassica napus*). *European Journal of Plant Pathology* 114: 41–52.

Downey RK and Harvey BL (1963) Methods of breeding for oil quality in rape. *Canadian Journal of Plant Science* 43: 271–275.

Fitt BDL, Brun H, Barbetti MJ and Rimmer SR (2006) World-wide importance of phoma stem canker (*Leptosphaeria maculans* and *L. biglobosa*) on oilseed rape (*Brassica napus*). *European Journal of Plant Pathology* 114: 3–15.

French E, Kim B-S and Iyer-Pascuzzi AS (2016) Mechanisms of quantitative disease resistance in plants. *Seminars in cell & developmental biology* 56: 201–208.

Friedt W and Snowdon RJ (2010) "Oilseed Rape" in *Oil Crops: Handbook of Plant Breeding*, eds. Vollmann Johann, and Rajcan Istvan (Springer Verlag, Berlin/Heidelberg): 91–126.

Gabur I, Chawla HS, Liu X, Kumar V, Faure S, Tiedemann A von, Jestin C, Dryzska E, Volkmann S, Breuer F, Delourme R, Snowdon R and Obermeier C (2018) Finding invisible quantitative trait loci with missing data. *Plant biotechnology journal* 16: 2102–2112.

Gibson G (2012) Rare and Common Variants: Twenty arguments. *Nature Reviews Genetics* 13: 135–145.

Golicz AA, Bayer PE, Bhalla PL, Batley J and Edwards D (2019) Pangenomics Comes of Age: From Bacteria to Plant and Animal Applications. *Trends in Genetics* 36: 132–145.

Hall D, Tegström C and Ingvarsson PK (2010) Using association mapping to dissect the genetic basis of complex traits in plants. *Briefings in functional genomics* 9: 157–165.

Hammond KE, Lewis BG and Musa TM (1985) A systemic pathway in the infection of oilseed rape plants by *Leptosphaeria maculans*. *Plant Pathology* 34: 557–565.

Heather JM and Chain B (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics* 107: 1–8.

Higgins EE, Clarke WE, Howell EC, Armstrong SJ and Parkin IAP (2018) Detecting *de Novo* Homoeologous Recombination Events in Cultivated *Brassica napus* Using a Genome-Wide SNP Array. *G3: Genes, Genomes, Genetics* 8: 2673–2683.

Higgins RH, Thurber CS, Assaranurak I and Brown PJ (2014) Multiparental Mapping of Plant Height and Flowering Time QTL in Partially Isogenic Sorghum Families. *G3: Genes, Genomes, Genetics* 4: 1593–1602.

Hirschhorn JN and Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6: 95–108.

Hohmann MA (2017) *Untersuchung von Anzuchtsystemen und Selektionstechniken zur Züchtung von Trockenstressresistenz in Winterraps (Brassica napus L.). (In German). Dissertation, University of Giessen, Germany*.

Howlett BJ, Lowe RGT, Marcroft SJ and van de Wouw AP (2015) Evolution of virulence in fungal plant pathogens: exploiting fungal genomics to control plant disease. *Mycologia* 107: 441–451.

Huang X, Paulo M-J, Boer M, Effgen S, Keizer P, Koornneef M and van Eeuwijk FA (2011) Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population. *Proceedings of the National Academy of Sciences of the United States of America* 108: 4488–4493.

Huang YJ, Jestin C, Welham SJ, King GJ, Manzanares-Dauleux MJ, Fitt BDL and Delourme R (2016) Identification of environmentally stable QTL for resistance against *Leptosphaeria maculans* in oilseed rape (*Brassica napus*). *Theoretical and applied genetics* 129: 169–180.

Hurgobin B, Golicz AA, Bayer PE, Chan C-KK, Tirnaz S, Dolatabadian A, Schiessl SV, Samans B, Montenegro JD, Parkin IAP, Pires JC, Chalhoub B, King GJ, Snowdon R, Batley J and Edwards D (2018) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant biotechnology journal* 16: 1265–1274.

Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, Malla S, Marriott H, Nieto T, O'Grady J, Olsen HE, Pedersen BS, Rhie

A, Richardson H, Quinlan AR, Snutch TP, Tee L, Paten B, Phillippy AM, Simpson JT, Loman NJ and Loose M (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature biotechnology* 36: 338–345.

Jaiswal AK, Tiwari S, Tavares GC, da Silva WM, Castro Oliveira L de, Ibraim IC, Guimarães LC, Gomide ACP, Jamal SB, Pantoja Y, Tiwary BK, Burkovski A, Munir F, Thi HHP, Ullah N, Ali A, Giovanetti M, Alcantara LCJ, Kaur J, Dhawan D, Imchen M, Krishna Vennapu R, Kumavath R, Corredor M, Figueiredo HCP, Barh D, Azevedo V and Castro Soares S de (2020) "Pan-omics focused to Crick's central dogma" in *Pan-genomics: Applications, challenges, and future prospects*, ed. Barh, D (London: Academic Press): 1–41.

Jestin C, Bardol N, Lodé M, Duffé P, Domin C, Vallée P, Mangin B, Manzanares-Dauleux MJ and Delourme R (2015) Connected populations for detecting quantitative resistance factors to phoma stem canker in oilseed rape (*Brassica napus* L.). *Molecular Breeding* 35: 167.

Josefsson E and Appelqvist L-A (1968) Glucosinolates in seed of rape and turnip rape as affected by variety and environment. *Journal of the Science of Food and Agriculture* 19: 564–570.

Juska A, Busch L and Tanaka K (1997) The Blackleg Epidemic in Canadian Rapeseed as a "Normal Agricultural Accident". *Ecological Applications* 7: 1350–1356.

Kang S-T, Kwak M, Kim H-K, Choung M-G, Han W-Y, Baek I-Y, Kim MY, Van K and Lee S-H (2008) Population-specific QTLs and their different epistatic interactions for pod dehiscence in soybean [*Glycine max* (L.) Merr.]. *Euphytica* 166: 15–24.

Korte A and Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant methods* 9: 29.

Kumar V, Paillard S, Fopa-Fomeju B, Falentin C, Deniot G, Baron C, Vallée P, Manzanares-Dauleux MJ and Delourme R (2018) Multi-year linkage and association mapping confirm the high number of genomic regions involved in oilseed rape quantitative resistance to blackleg. *Theoretical and applied genetics* 131: 1627–1643.

Kutcher HR, Brandt SA, Smith EG, Ulrich D, Malhi SS and Johnston AM (2013) Blackleg disease of canola mitigated by resistant cultivars and four-year crop rotations in western Canada. *Canadian Journal of Plant Pathology* 35: 209–221.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH,

Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng J-F, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, La Bastide M de, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen H-C, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JGR, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AFA, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang S-P,

Yeh R-F, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Patrinos A and Morgan MJ (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.

Larkan NJ, Raman H, Lydiate DJ, Robinson SJ, Yu F, Barbulescu DM, Raman R, Luckett DJ, Burton W, Wratten N, Salisbury PA, Rimmer SR and Borhan MH (2016) Multi-environment QTL studies suggest a role for cysteine-rich protein kinase genes in quantitative resistance to blackleg disease in *Brassica napus*. *BMC plant biology* 16: 183.

Li H, Feng X and Chu C (2020) The design and construction of reference pangenome graphs with minigraph. *Genome biology* 21: 265.

Li Y-h, Zhou G, Ma J, Jiang W, Jin L-g, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang S-s, Zuo Q, Shi X-h, Li Y-f, Zhang W-k, Hu Y, Kong G, Hong H-l, Tan B, Song J, Liu Z-x, Wang Y, Ruan H, Yeung CKL, Liu J, Wang H, Zhang L-j, Guan R-x, Wang K-j, Li W-b, Chen S-y, Chang R-z, Jiang Z, Jackson SA, Li R and Qiu L-j (2014) *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature biotechnology* 32: 1045–1052.

Liu W, Kim MY, Van K, Sun S and Lee S-H (2010) Identification of Population-Specific QTLs for Flowering in Soybean. *Journal of Crop Science and Biotechnology* 13: 213–218.

Loman NJ, Quick J and Simpson JT (2015) A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nature methods* 12: 733–735.

Long Y, Shi J, Qiu D, Li R, Zhang C, Wang J, Hou J, Zhao J, Shi L, Park B-S, Choi SR, Lim YP and Meng J (2007) Flowering Time Quantitative Trait Loci Analysis of Oilseed Brassica in Multiple Environments and Genomewide Alignment with Arabidopsis. *Genetics* 177: 2433–2444.

Lu H, Giordano F and Ning Z (2016) Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* 14: 265–279.

Lu J, Carbone GJ and Grego JM (2019) Uncertainty and hotspots in 21st century projections of agricultural drought from CMIP5 models. *Scientific reports* 9: 4922.

Luo R, Wong C-L, Wong Y-S, Tang C-I, Liu C-M, Leung C-M and Lam T-W (2020) Exploring the limit of using a deep neural network on pileup data for germline variant calling. *Nature Machine Intelligence* 2: 220–227.

Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C and Sedlazeck FJ (2019) Structural variant calling: the long and the short of it. *Genome biology* 20: 246.

Marcroft SJ, van de Wouw AP, Salisbury PA, Potter TD and Howlett BJ (2012) Effect of rotation of canola (*Brassica napus*) cultivars with different complements of blackleg resistance genes on disease severity. *Plant Pathology* 61: 934–944.

Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, Schumann E, Kilian B, Reif JC and Pillen K (2015) Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC genomics* 16: 290.

McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddeloh JA and Stupar RM (2012) Structural Variants in the Soybean Genome Localize to Clusters of Biotic Stress-Response Genes. *Plant physiology* 159: 1295–1308.

Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan C-KK, Visendi P, Lai K, Doležel J, Batley J and Edwards D (2017) The pangenome of hexaploid bread wheat. *The Plant journal: for cell and molecular biology* 90: 1007–1013.

Niks RE, Qi X and Marcel TC (2015) Quantitative Resistance to Biotrophic Filamentous Plant Pathogens: Concepts, Misconceptions, and Mechanisms. *Annual review of phytopathology* 53: 445–470.

Nishida H, Yoshida T, Kawakami K, Fujita M, Long B, Akashi Y, Laurie DA and Kato K (2013) Structural variation in the 5′ upstream region of photoperiod-insensitive alleles Ppd-A1a and Ppd-B1a identified in hexaploid wheat (*Triticum aestivum* L.), and their effect on heading time. *Molecular Breeding* 31: 27–37.

Parkin IAP, Sharpe AG, Keith DJ and Lydiate DJ (1995) Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). *Genome* 38: 1122–1131.

Pascual L, Albert E, Sauvage C, Duangjit J, Bouchet J-P, Bitton F, Desplat N, Brunel D, Le Paslier M-C, Ranc N, Bruguier L, Chauchard B, Verschave P and Causse M (2016)

Dissecting quantitative trait variation in the resequencing era: complementarity of bi-parental, multi-parental and association panels. *Plant Science* 242: 120–130.

Pilet-Nayel M-L, Moury B, Caffier V, Montarry J, Kerlan M-C, Fournet S, Durel C-E and Delourme R (2017) Quantitative Resistance to Plant Pathogens in Pyramiding Strategies for Durable Crop Protection. *Frontiers in plant science* 8: 1838.

Poland J and Rutkoski J (2016) Advances and Challenges in Genomic Selection for Disease Resistance. *Annual review of phytopathology* 54: 79–98.

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC and Nelson RJ (2009) Shades of gray: the world of quantitative disease resistance. *Trends in plant science* 14: 21–29.

Qian L, Voss-Fels K, Cui Y, Jan HU, Samans B, Obermeier C, Qian W and Snowdon RJ (2016) Deletion of a Stay-Green Gene Associates with Adaptive Selection in *Brassica napus*. *Molecular plant* 9: 1559–1569.

Rimmer RS (2006) Resistance genes to *Leptosphaeria maculans* in *Brassica napus*. *Canadian Journal of Plant Pathology* 28: 288–297.

Rousseau-Gueutin M, Belser C, Da Silva C, Richard G, Istace B, Cruaud C, Falentin C, Boideau F, Boutte J, Delourme R, Deniot G, Engelen S, Carvalho JF de, Lemainque A, Maillet L, Morice J, Wincker P, Denoeud F, Chèvre A-M and Aury J-M (2020) Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience* 9: 1–16.

Samans B, Chalhoub B and Snowdon RJ (2017) Surviving a Genome Collision: Genomic Signatures of Allopolyploidization in the Recent Crop Species *Brassica napus*. *Plant Genome* 10.

Samans B, Snowdon R and Mason AS (2018) "Homoeologous Exchanges and Gene Losses Generate Diversity and Differentiate the *B. napus* Genome from that of Its Ancestors" in *The Brassica napus Genome*, eds. Liu, S, Snowdon, R, and Chalhoub, B (Cham: Springer International Publishing): 131–148.

Sanger F, Nicklen S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74: 5463–5467.

Sannemann W, Huang BE, Mathew B and Léon J (2015) Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. *Molecular Breeding* 35: 86.

Saxena RK, Edwards D and Varshney RK (2014) Structural variations in plant genomes. *Briefings in functional genomics* 13: 296–307.

Schiessl S, Huettel B, Kuehn D, Reinhardt R and Snowdon RJ (2017) Targeted deep sequencing of flowering regulators in *Brassica napus* reveals extensive copy number variation. *Scientific data* 4: 170013.

Schiessl SV, Quezada-Martinez D, Orantes-Bonilla M and Snowdon RJ (2020) Transcriptomics reveal high regulatory diversity of drought tolerance strategies in a biennial oil crop. *Plant Science* 297: 110515.

Scott MF, Ladejobi O, Amer S, Bentley AR, Biernaskie J, Boden SA, Clark M, Dell'Acqua M, Dixon LE, Filippi CV, Fradgley N, Gardner KA, Mackay IJ, O'Sullivan D, Percival-Alwyn L, Roorkiwal M, Singh RK, Thudi M, Varshney RK, Venturini L, Whan A, Cockram J and Mott R (2020) Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity* 125: 396–416.

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, Haeseler A von and Schatz MC (2018) Accurate detection of complex structural variations using single-molecule sequencing. *Nature methods* 15: 461–468.

Slatko BE, Gardner AF and Ausubel FM (2018) Overview of Next-Generation Sequencing Technologies. *Current protocols in molecular biology* 122: e59.

Snowdon RJ, Lühs W and Friedt W (2006) "Oilseed Rape" in *Genome Mapping and Molecular Breeding in Plants, Vol. 2: Oilseeds*, ed. Kole, C (Springer-Verlag, Heidelberg, New York, Berlin, Tokyo): 55–114.

Song J-M, Guan Z, Hu J, Guo C, Yang Z, Wang S, Liu D, Wang B, Lu S, Zhou R, Xie W-Z, Cheng Y, Zhang Y, Liu K, Yang Q-Y, Chen L-L and Guo L (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature Plants* 6: 34–45.

Sprague SJ, Balesdent M-H, Brun H, Hayden HL, Marcroft SJ, Pinochet X, Rouxel T and Howlett BJ (2006) Major gene resistance in *Brassica napus* (oilseed rape) is overcome by changes in virulence of populations of *Leptosphaeria maculans* in France and Australia. *European Journal of Plant Pathology* 114: 33–40.

Srikanth A and Schmid M (2011) Regulation of flowering time: all roads lead to Rome. *Cellular and molecular life sciences* 68: 2013–2037.

St Clair DA (2010) Quantitative Disease Resistance and Quantitative Resistance Loci in Breeding. *Annual review of phytopathology* 48: 247–268.

Stahl A, Vollrath P, Samans B, Frisch M, Wittkop B and Snowdon RJ (2019) Effect of breeding on nitrogen use efficiency-associated traits in oilseed rape. *Journal of experimental botany* 70: 1969–1986.

Stall RE, Jones JB and Minsavage GV (2009) Durability of Resistance in Tomato and Pepper to Xanthomonads Causing Bacterial Spot. *Annual review of phytopathology* 47: 265–284.

Stefansson BR and Hougen FW (1964) Selection of rape plants (*Brassica napus*) with seed oil practically free from erucic acid. *Canadian Journal of Plant Science* 44: 359–364.

Stein A, Coriton O, Rousseau-Gueutin M, Samans B, Schiessl SV, Obermeier C, Parkin IAP, Chèvre A-M and Snowdon RJ (2017) Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. *Plant biotechnology journal* 15: 1478–1489.

Szadkowski E, Eber F, Huteau V, Lodé M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ, Chalhoub B, Jenczewski E and Chèvre A-M (2010) The first meiosis of resynthesized *Brassica napus*, a genome blender. *The New phytologist* 186: 102–112.

U N (1935) Genome analysis in Brassica with special reference to the experimental formation of *Brassica napus* and peculiar mode of fertilization. *Journal of Japanese Botany* 7: 389–453.

Van de Wouw AP, Marcroft SJ and Howlett BJ (2016) Blackleg disease of canola in Australia. *Crop and Pasture Science* 67: 273–283.

Vollrath P, Chawla HS, Alnajar D, Gabur I, Lee H, Weber S, Ehrig L, Koopmann B, Snowdon RJ and Obermeier C (2021a) Dissection of Quantitative Blackleg Resistance Reveals Novel Variants of Resistance Gene *Rlm9* in Elite *Brassica napus*. *Frontiers in plant science* 12:749491.

Vollrath P, Chawla HS, Schiessl SV, Gabur I, Lee H, Snowdon RJ and Obermeier C (2021b) A novel deletion in *FLOWERING LOCUS T* modulates flowering time in winter oilseed rape. *Theoretical and applied genetics* 134: 1217–1231.

Weisweiler M, Montaigu Ad, Ries D, Pfeifer M and Stich B (2019) Transcriptomic and presence/absence variation in the barley genome assessed from multi-tissue mRNA sequencing and their power to predict phenotypic traits. *BMC genomics* 20: 787.

Werner C and Snowdon R (2018) "Genome-Facilitated Breeding of Oilseed Rape" in *The Brassica napus Genome*, eds. Liu, S, Snowdon, R, and Chalhoub, B (Cham: Springer International Publishing): 245–269.

West JS, Kharbanda PD, Barbetti MJ and Fitt BDL (2001) Epidemiology and management of *Leptosphaeria maculans* (phoma stem canker) on oilseed rape in Australia, Canada and Europe. *Plant Pathology* 50: 10–27.

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J and Wang W (2011) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature biotechnology* 30: 105–111.

Zhang H, Jain C and Aluru S (2020) A comprehensive evaluation of long read error correction methods. *BMC genomics* 21: 889.

List of abbreviations

A	Adenine
bp	Base pairs
C	Cytosine
CDS	Coding sequence
ddNTPs	Dideoxynucleotides
DH	Double haploid
DNA	Deoxyribonucleic acid
FLC	Flowering Locus C
FT	Flowering Locus T
G	Guanine
GS	Genomic selection
GWAS	Genome-wide association study
kb	Kilo base pairs
LD	Linkage disequilibrium
MAS	Marker-assisted selection
NGS	Next-generation sequencing
ONT	Oxford Nanopore Technologies
PAV	Presence-absence variation
QTL	Quantitative trait locus
SV	Structural genome variation
SNV	Single-nucleotide variation
T	Thymine

Declaration

I declare that the dissertation here submitted is entirely my own work, written without any illegitimate help by any third party and solely with materials as indicated in the dissertation. I have indicated in the text where I have used texts from already published sources, either word for word or in substance, and where I have made statements based on oral information given to me. At all times during the investigations carried out by me and described in the dissertation, I have followed the principles of good scientific practice as defined in the "Statutes of the Justus Liebig University Gießen for the Safeguarding of Good Scientific Practice".

Gießen, Tuesday, December 7, 2021 **Paul Vollrath**

Acknowledgements

I would like to express my particular thanks to the people who made my time in Gießen as enjoyable and unforgettable as it was, but also people who supported me during various stages in my life.

First of all, I want to express my gratitude to the head of the Department of Plant Breeding at the Justus Liebig University in Giessen and the first supervisor of my PhD, Prof. Dr. Rod J. Snowdon. Rod gave me the chance to start as a PhD student after finishing my master's at the department. Aiming for a career in crop protection, he convinced me that resistance breeding could also be considered as crop protection.

This brings me to another person who I am equally thankful to. Dr. Christian Obermeier who wrote the project *PhomaDur* and with that ensured the financing and thus my employment. As the PI of the project, co-supervisor of my PhD work and my personal contact for any questions, he was always willing to help, discuss and to give advice that substantially improved my work. Neither my publications nor the present thesis would have been possible without your suggestion, help and patience. Thank you, Christian.

I would also like to say thank you to Prof. em. Dr. Dr. h.c. Wolfgang Friedt who welcomed me warmly within my first days at the department and made me feel comfortable right from the beginning by saying: '*Endlich nochmal jemand aus dem Hunsrück*.'

I would like to come back to Rod before ending this section. Rod has the talent to motivate by giving pragmatic and incredibly straight forward advice that highly encourage you to keep working and thus reaching a certain goal, to prepare a presentation or to finish a

publication. Whenever I felt lost during my time as a PhD student, a quick chat with Rod put me back on track. Thank you, Rod.

To proceed chronologically, I would like to say thank you to Sarah Schießl-Weidenweber, Sarah Hatzig and Anna Stein. It was a pleasure to start my PhD having you as office mates. I am quite sure that I had many and most of all stupid questions at the very beginning. Nonetheless, all of you have always been helpful and friendly and thus helped me to feel comfortable within the department and in particular in our office. Thank you, Sarah, Sarah and Anna.

Another special thank you goes to Andreas Stahl and Benjamin Wittkop. Andreas already supervised me during the preparation of my master's thesis. After starting as a PhD student both of you were always available for my questions regarding field trials. Starting from planning the trials, any issues during the seasons, data analysis and interpretation. Thank you, Andy and Benny.

Going on from my practical work in the field, greenhouse and lab, I need to say a big 'thank you' to all the super helpful and friendly technicians within the department. Representative for all of you, I want to especially thank Stavros Tzigos who guided me at the very start of my 'lab career' and helped me to get a little steadier hand when using the pipette. Furthermore, I want to say a big 'thank you' to Regina Illgner. Reduced to a few simple words, Regina is one of those technicians, a PhD student can be happy to work with. Regina was highly motivated, proactive, flexible and stress resistant. Your reliability made my PhD

life considerably easier. Thank you, Regina, Stavros, Sabine, Birgit, Annette, Anja, Karl-Heinz, Swetlana, Nelly, Petra and Ingrid.

In the same breath I want to mention three more people who made my time as a PhD student less complicated. Knowing to have secretaries in the department on whom one can rely on 100 percent is a great relief for any PhD student. As personal problem solvers and reminders of anything, you have been very important to me and my progress. Thank you, Frau Schomber, Ulla and Frau Kessler-Turturici.

Many more thanks go out to my colleagues, friends and suffering comrades. I had an awesome time in Gießen during my studies. Starting with my bachelor's and master's, Gießen was an unbelievable experience to me and I made many friends. To name all of them would exceed the scope. However, I want to express my deepest gratitude for fruitful discussions, mutual support, entertaining coffee breaks, fantastic conferences, relaxing after work beers, tasty Dönerstage and crazy Friday nights. I was lucky to meet guys like you. Thank you, Harmeet, Iulian, Jenny, Hansi, Stjepan, Sven, Manar, Lennard, Kai, Christian, Timm, André, Benny, Andy, Steffen, Dani, Paula, Mauricio, Roman, Subhadra, Isabelle, Luisa and Natalja.

A very special 'thank you' goes to my 'Greek friend'. Stavros, you helped me from my very first day with basic lab work. For sure, this was not only fun for you. Nonetheless, you always gave me the feeling to be welcome to ask you anything. However, the greatest thanks are for your welcome diversion from working routine. Thanks for unforgettable,

almost five years, for nice after work drinking and eating. I am glad and proud to call you my friend. Thank you, Stavros.

Last but not least, one of my biggest thanks goes out to my beloved family. Above all, I want to express my deepest gratitude to the love of my life, my wife, Michelle. She has already been at my side before I was even thinking about going to Gießen for studying. Her infinite love and support helped me to keep following my goals and dreams. Thank you, Michelle.

No less important I want to express my deepest gratitude to my mother Andrea, my father Karl-Otto and my sister Laura. I could think of no better childhood, upbringing and support during every stage in my life. I am glad to know you to back me.

The work was funded by the German Federal Ministry of Food and Agriculture (BMEL grant number BLE: 2818205115) and the Society for Promotion of Plant Innovation (GFPi e.V., Bonn, Germany) within the cooperation project PhomaDur. Additional support was provided by all German breeding companies organized in the GFPi as well as by the Plant Pathology and Crop Protection Section, Department of Crop Science, Georg August University, Göttingen. This work was also supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI).