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

Genome structural variation associates with fungal quantitative disease resistance in 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
1. Prof. Dr. Rod J. Snowdon
2. Prof. Dr. Matthias Frisch

Submitted by
Iulian Gabur

Giessen 2018

“If we knew what it was we were doing, it would not be called research, would it?”

Albert Einstein

Printed and published with the support of the German Academic Exchange Service.

Table of Content

1 General Introduction

1.1 Winter oilseed rape: history and economic importance

Brassica vegetables and oilseeds were among the earliest plants widely cultivated as early as 10,000 years ago. In India, archaeological findings indicate that oilseed Brassicas (probably *B. rapa*) were used as early as 4000 B.C. They spread into China and Japan approximately 2,000 years ago, while swedes (*B. napus* ssp. *napobrassica*) were used in Europe at the time of the Roman Empire. *B. rapa* was utilised for oil purposes in northern Europe around the 13th century and by the 16th century, oilseed rape was the major source of lamp oil. In Europe, systematic increase of cultivated area with oilseed rape did not happened until the 18th century (Kimber and McGregor 1995). Oilseed rape (*Brassica napus* L, 2n=4x=38, AACC; also known as rapeseed or canola) evolved in different European climate zones and hence displays variation in vegetative growth and winter hardiness. It has a strong geographical differentiation into spring or winter rapeseed genotypes, differentiation under genetic control of vernalisation requirement and flowering time. Today, oilseed is the most heavily cultivated oilseed crop and a major source of biodiesel in Europe and is present in spring forms that do not require vernalisation in North America (particularly Canada) and some parts of China. Intermediary types of oilseed rape are suitable for Asia, the Indian subcontinent and Australia (Bus et al., 2011; Snowdon et al., 2006). Worldwide, oilseed rape is the third most important oilseed crop after soybean and oil palm that have greater oil productions (FAOSTAT data, 2015: http://faostat.fao.org/). Soybean produced, on a ten years average, 236 M metric tons (2004-2013), while oilseed rape recorded in the same timeframe a production of 57.6 M metric tons. During 2004-2013, oilseed rape was cultivated on approximately 31 M ha worldwide, mainly in Canada, China, the European Union, Australia and USA. France and Germany are the biggest oilseed rape producers in Europe, with cultivated surfaces of 1.43M ha and 1.4M ha, respectively.

Oilseed rape has an amphitetraploid genome and is the result of a recent spontaneous interspecific hybridisation between *Brassica rapa* (2n=2x=20, AA) and *Brassica oleracea* (2n=2x=18, CC) (U 1935). Oilseed rape contains the full intact chromosome complements of *B. rapa* and *B. oleracea* (Parkin et al., 1995; Axelsson et al., 2000). Friedt and Snowdon (2010) suggest that the initial allopolyploidization occurred as a result of co-cultivation of the diploid progenitors in close geographical proximity, in the Mediterranean region. Recently, Chalhoub et al., (2014) published the *B. napus* reference sequence and revealed that the oilseed rape genome has undergone 72-fold multiplications and has a high genetic redundancy. When comparing orthologous genes with the diploid progenitors, *B. napus* seems to not be older than 7,500 to 12,500 years, thus, a plant species with a relatively short domestication history.

Oilseed rape brings high value in cereal crop rotations, by providing positive influence on yields of subsequent crops of wheat (Christen et al., 1992) and barley (Christen and Sieling, 1993). Moreover, in many regions of central and northern Europe, Canada and Australia, oilseed rape has an essential role in soil rejuvenation and management of monocotyledonous cereal diseases and pests (Ryan et al., 2006).

Within the last four decades, oilseed rape production increased strongly due to improvements in oil and meal quality through plant breeding, and substitution of fossil energy by clean, eco-friendly, renewable energy resources. Seed quality breeding improved substantially in the last 30 years, as oilseed rape started to be used for human consumption. However, breeding for so-called “double-low” (00) quality seeds with zero erucic acid and low glucosinolate content created a strong selection bottleneck and limited the genetic diversity in modern oilseed rape cultivars. Moreover, current winter oilseed rape cultivars are susceptible to attacks from a range of pathogens and an improvement of broad spectrum disease resistances is a continuing challenge for breeders. For example, rapid increase in the area cultivated has allowed severe

infections by soil-borne pathogens such as *Verticillium longisporum* (Dunker et al., 2008), *Sclerotinia sclerotium* (Wei et al., 2014) and *Leptospheria maculans* (Delourme et al., 2011) in major growing areas.

1.2 Major fungal pathogens economic importance and lifecycles

Major diseases of *Brassica* oilseeds are caused mainly by aerial and soil-borne root pathogens. Their economic importance varies according to geographical regions, as some diseases, for example Verticillium stem striping (*Verticillium longisporum*), are more problematic for winter oilseed rape in Europe, whereas others are widespread and affect various crops, for example Sclerotinia stem rot (*Sclerotinia sclerotium*), clubroot (*Plasmodiophora brassicae*) or blackleg stem canker (*Leptosphaeria maculans;* anamorph *Phoma lingam*).

Delourme et al. (2011) reviewed the most important diseases that cause important economic impacts on oilseed brassicas worldwide. Among them, four cause serious disease in winter oilseed rape in Europe. Blackleg is a worldwide disease that affects not only spring and winter forms of *B. napus*, but also *B. juncea* and *B. rapa*. Serious epidemics associated with blackleg have been spreading globally over the last 30 years in Australia, America and Europe (Howlett et al., 2015). Nowadays, blackleg also threatens oilseed rape production in China, where only *L. biglobosa (*anamorph *P. biglobosus)* is present and causes less damage (Fitt et al., 2006, 2008; Liu et al., 2014). Economically it was estimated that blackleg is most important disease of oilseed rape, causing yield losses in Europe, North America (especially Canada) and Australia of more than $900 M / £1000M per cropping season (Fitt et al., 2008; Fitt et al., 2011). In Canada, blackleg caused up to 50% yield losses in fields during the 1980s, when a susceptible genotype was intensively cultivated by farmers (Juska et al., 1997). In the early 2000s, blackleg disease in Canada was more successfully controlled by improved

disease management, including use of resistant varieties and extended crop rotations (Kutcher et al., 2013). However, due to favourable economic returns, 2-year rotations systems that again led to the erosion of blackleg resistance have recently ganed in popularity again (Hwang et al., 2016). Furthermore, severe epidemics are predicted to increase in the context of global warming (Evans et al., 2008; Butterworth et al., 2010).

Sclerotinia sclerotiorum is a necrotrophic pathogen that affects numerous agricultural crops around the world, most commonly in temperate regions (Willetts et al., 1980). It infects more than 400 different plant species, including soybean (*Glycine max*), chickpea (*Cicer arietinum*), sunflower (*Hellianthus annuus*) and oilseed rape (Bolton et al., 2006). Sclerotinia stem rot is reported to cause up to 80% yield losses on *Brassica* species in China (Mei et al., 2011). Elsewhere, major losses in quality and yield reductions up to 50% have been reported (Uloth et al., 2013, 2014). Estimated costs of sclerotinia damage to the canola industry in Western Australia amount to more than 23 million $AUS per annum (Neik et al., 2017).

Verticillium species also have a wide host range spanning more than 200 plant species, including the *Solanaceae*, for example tomato (*S. lycopersicum*), potato (*S. tuberosum*) and eggplant (*S. melongena*), the *Malvaceae*, for example cotton (*Gossipium arboreum*), the *Brassicaceae*, for example oilseed rape (*B. napus*), along with important tree crops such as olive (*O. europaea*), pistachio (*P. vera*) and avocado (*Persea americana*) (Fravel and Larkin, 1997). The species *V. longisporum* differs from other *Verticillium* species by being strongly restricted to *Brassica* species and by inducing premature ripening without classical wilt symptoms (Knufer, 2011). *V. longisporum* is a recent problem of *B. napus* cultivation, have become particularly widespread in cooler parts of Europe during recent decades and affecting winter oilseed rape (Evans et al., 2009) with yield losses up to 50 % (Dunker et al., 2008).

In contrast to *S. sclerotiorum* and *L. maculans*, which mainly cause infection of plant tissues above the soil surface, *V. longisporum* infects the oilseed rape via roots. Hence it may be referred to as a soil-borne, vascular plant pathogen.

1.2.1. Blackleg/phoma stem canker

In agro-ecosystems, epidemics may result from direct human intervention in decisions on succession of cropping seasons and increased adaptation in pathogen populations to host crop resistance mechanisms (Bousset et al., 2018). *Leptosphaeria maculans* is considered to be a monocyclic disease and an important cause of stem canker in oilseed rape (West et al., 2001). Blackleg epidemics are initiated in autumn, when classical leaf spots can be observed on oilseed rape leaves, persisting until early spring. Stem cankers develop later and can be observed from spring until harvest time. In Canada and Australia, secondary cycles of infection on leaf spots have been reported by Barbetti (1976) and Hall et al. (1996), however until today this has not been documented in European countries. Systematic growth of fungal hyphae lead to the development of cankers from the lead spots to petiole through xylem vessels and later to the plant stem base (Hammond et al., 1985; Travadon et al., 2009). Fungal hyphae can survive for long periods of time in stubble where fungicides treatments are not always efficient. Two fruiting bodies are known, namely pycnidia and pseudothecia. Asexual multiplication of *L. maculans* pycnidia and pseudothecia may be transmitted at short distances through rain splashes, for conidia (Pycnidiospores) or through ascospores transported by wind (Marcroft et al., 2004; Bousset et al., 2015).

Infected stubbles are the major source of inoculum and contribute directly at transmission of fungal hyphae from one harvest to another, if crop management requirements are not met (Lo-Pelzer et al., 2009; Bousset et al., 2018).

1.2.2. Sclerotinia stem rot

Sclerotinia stem rot, or white mould is caused by the fungal pathogen *Sclerotinia sclerotiorum* and has a very wide host range (Zhao et al., 2004). Infection occurs through mycelium or ascospore contact with senescent plant tissues. Generally, at flowering, ascospores colonise senescing flower petals, old or injured plant tissue, or petioles adjacent to the stem (Garg et al., 2010). The mycelium develops and spreads throughout infected tissues by producing enzymes that degrade cell wall components, such as keratin (Bolton et al., 2006). Disease symptoms become visible depending on the host species, the type of affected tissue and climatic conditions. In *Brassica* crops, infections cause pale or dark lesions on leaves that are irregular and necrotic. Lesions on stems are white to greyish in colour and develop close to the ground, eventually leading to premature wilting, stem breakage and collapse of plants.

1.2.3. Verticillium stem striping

In comparison with other *Verticillium* species, a major characteristic of *V. longisporum* is that it is able to change its lifestyle from a dormant phase to a parasitic or saprophytic phase (Eynck, 2009; Knufer, 2011). Microsclerotia are produced in the dormant phase and can survive for more than 10 years in soil (Heale and Karapapa, 1999). Microsclerotia present in soils start to geminate when they are stimulated by root exudates from host or non-host plants (Schnathorst, 1981). *V. longisporum* hyphae subsequently migrate towards the roots and the fungus penetrates directly into the root epidermal cells (Zhou et al., 2006; Eynck et al., 2007). Later, the fungus crosses into root cortex intra- and intercellularly before entering xylem vessels (Eynck et al., 2008). Here it uses compounds present in the xylem sap as a nutrient source, being well adapted to a nutrient-poor environment (Pegg and Brady, 2002). At plant senescence, the pathogen enters the parenchymatic cell. In later stages of its lifecycle the pathogen enters in a saprophytic phase and microsclerotia are formed in plant stems.

Afterwards, microsclerotia present in stem are released into the soil, where they represent a primary source of disease spread in upcoming years (Schnathorst, 1981; Neumann & Dobinson, 2003). First visual symptoms of Verticillium stem striping (VSS) occur at the beginning of plant ripening in field, when microsclerotia become visible beneath the epidermis and in the root tissue. Among the earlier symptoms, a discoloration and one-sided brownish stripes along the stem might be observed. A successful infection with *V. longisporum* restricts water and nutrient transportation by clogging the xylem vessels and causing premature ripening, even though wilt symptoms are not visible (Knufer, 2011).

1.3 Breeding oilseed rape for disease resistance: genetics and genomics of disease resistance

The increased demand for oil and vegetable production has intensified *B. napus* cultivation worldwide, but consequently also increased the incidence and severity of disease caused by various pathogens (Sanogo et al., 2015; Van de Wouw et al., 2016). Disease management options, as use of fungicides or cultural control measures, often offer inconsistent and expensive alternatives, especially for low input and low return crops. In recent years, many efforts have aimed at identifying resistance sources for major pathogens and studying the genetic determinism and mechanisms of the resistances towards breeding varieties with improved disease resistance (Delourme et al., 2011). Efficient host resistance represents a cost-effective and reliable method of disease control and resistance. Major fungal pathogens do not restrict themselves only to oilseed rape, and related species can also serve as hosts or resistance sources (Ren et al., 2016). Genetic diversity within and between *Brassica* crop species therefore represents an important tool for breeders. Moreover, other species within the *Brassicaceae* may be used as potential resource for qualitative resistance genes (R-genes) or quantitative resistance (resistance QTL) against different fungal pathogens. Breeding resistant

cultivars for blackleg, Sclerotinia stem rot and Verticillium stem striping represents a major component in disease resistance management.

In general, for fungal pathogens two types of resistances are available: monogenic and polygenic resistance, which operate during different plant development stages. Monogenic or qualitative resistance is expressed in plant seedlings up to adult plant stage and is considered to be controlled by R-genes or single-gene, race specific resistance. Polygenic or quantitative adult-plant resistance is considered as partial resistance that is mediated by many resistance factors of quantitative trait loci (QTL). Infected plants cope with pathogens either through large effect resistance genes (R-genes), moderate effect QTL, or a mixture of both categories.

Genetic studies in oilseed rape have identified specific major genes involved in disease resistance for blackleg (Delourme et al., 2006; Rimmer, 2006), but also QTL regions with low to moderate effects (Raman et al., 2012, 2013; Kumar et al., 2018). Qualitative resistance in normally expressed starting with the seedling stage and is based on gene-for-gene interactions. Until now, more than fifteen race-specific resistance genes, known as *Rlm*-genes, have been mapped in *Brassicas* on different chromosomes, or introgressed in *B. napus* from related *Brassica* species (Delourme et al., 2006; Balesdent et al., 2013; Yu et al., 2013). Most of these genes (i.e. *Rlm1-Rlm11* and *LepR1-LepR4*) are organized in R-gene clusters on different chromosomes of oilseed rape (Delourme et al., 2004). Although monogenic resistance is very efficient for specific fungal races, it exerts strong selection pressure and forces pathogens to rapidly evolve and adapt. Therefore, cultivars that rely solely on R-gene resistances can lose their effectiveness in few years if repeatedly cultivated on extensive growing areas (Rouxel et al., 2003).

Polygenic (quantitative) resistance is considered as non-race specific and more durable than qualitative resistance (Poland et al., 2009). QTL effectiveness is known to vary between seasons due to high environmental constraints. Brun et al. (2010) suggested that the

combination of qualitative and quantitative resistances in cultivars is an effective method to improve durability of blackleg disease resistance in oilseed rape. In addition, combined *Rlm6* and quantitative resistance provided effective disease resistance to blackleg for seven years in a susceptible background (Delourme et al., 2014). In adult plants, resistance QTL limit the development of necrosis at stem basis (Delourme et al., 2006; Rimmer, 2006; Raman et al., 2013; Larkan et al., 2016). Jestin et al. (2015) used a *B. napus* multi-cross design to identify diversity of quantitative blackleg resistance among different- and specific- genetic backgrounds. Larkan et al. (2016) carried out a multi-environment study and identified cysteine-rich protein kinase genes, but also four significant QTL in spring *B. napus*. Although quantitative resistance could be mapped only in the absence of major R-genes, quantitative resistance in elite varieties could represent a long-term solution for effective disease control (Pilet-Nayel et al., 2017).

The French winter oilseed rape cultivar 'Darmor' has been used as a source of resistance to blackleg in different genetic backgrounds (Pilet et al., 2001; Jestin et al., 2011). Results showed that not only the environment but also the genetic make-up of genotypes has an impact on QTL detection (Pilet et al., 2001). In literature, it was reported that some environmental conditions, especially temperature, are influencing monogenic- and polygenic-mediated resistance mechanisms against *L. maculans* (Huang et al., 2006, 2009). In Australia, QTL analysis also showed the strong influence of environment on resistance QTL mapping (Raman et al., 2012). Genotype × environment interactions represent an important limiting factor for the use of resistance QTL in breeding (Poland et al., 2009). Therefore, it is essential for commercial breeding purposes to identify environmental stable QTL and develop oilseed rape cultivars with effective resistance in diverse environments.

Sclerotinia stem rot has a wide host range, including species outside the *Brassicaceae* family (Zhao et al., 2004). Monogenic disease resistance control by single dominant genes against

S. sclerotiorum has been found in common bean (*Phaseolus vulgaris*; Schwartz et al., 2006; Viteri & Singh, 2015). Although it is possible that monogenic resistance is present in *Brassica* species, to date only hypersensitive responses were observed on stem tests (Uloth et al., 2013; You et al., 2016) or cotyledons (Garg et al., 2008; Uloth et al., 2014; Ge et al., 2015). Identification of genetic regions involved in disease control has been hampered by many inconsistencies, such as time period and type of inoculation methods, use of various isolates and other. Li et al. (2015) reported that expression of stem resistance is independent of flowering time, whereas the type and time of inoculation are major factors contributing to delayed plant-pathogen antagonistic interactions. Nevertheless, cultivars known as highly resistant to Sclerotinia stem rot in one experimental design can be susceptible in another (Neik et al., 2017).

Only a very low level of quantitative resistance to Sclerotinia stem rot is present in modern European oilseed rape varieties. Therefore, management strategies for disease control currently rely on fungicide application. So far, no major monogenic resistance (R-genes) have been identified for this disease in oilseed rape (Wu et al., 2016). However, some authors identified potential NBS-LRR class candidate genes on chromosomes A06 and C07 of the *B. napus* reference genome (Li et al., 2015), while Mei et al. (2013) identified other candidate R genes belonging to the CC-NBS-LRR class on chromosome A09. Interestingly, the majority of Sclerotinia stem rot resistance QTL have been identified in the C sub-genome, indicating that *B. oleracea* and other C-subgenome donor species may be a good source of resistance for *S. sclerotiorum*. This was recently reported for *B. oleracea* var. *capitata* genotypes from China with high level of stem and leaf resistances (You et al., 2016). Other QTL regions have been identified in Chinese semi-winter varieties (Zhao et al., 2006). However, in studies that use diverse *Brassica* populations from different geographical areas, no correlation could be seen between leaf and stem disease resistance scoring (Uloth et al., 2013; You et al., 2016). Instead, Sclerotinia stem rot resistance screening in stem and leaf tests is strongly influenced

by host subpopulations. In oilseed rape, a QTL region on chromosome C2 responsible for stem resistance was found to be homoeologous with the corresponding region on chromosome A2 (Wei et al., 2014). The existence of conserved syntenic regions involved in Sclerotionia stem rot control suggests that resistance genes could be inherited from a common ancestral species.

Within the Brassicas, the B genome of *B. nigra)* was reported to carry promising Sclerotinia stem rot resistance loci (Navabi et al., 2010). Across many Australian *B. napus* varieties that carry *B. carinata* introgressions, a high level of resistance loci were identified (Uloth et al., 2013). Garg et al. (2010) showed that introgressions from wild weedy *Brassicaceae,* especially *Erucastrum cardaminoides* or *Diplotaxis tenuisiliqua,* into *B. napus* contributed high levels of disease resistance to *S. sclerotium.*

As a soilborne vascular disease, Verticillium stem striping on oilseed rape cannot be controlled using conventional management or fungicides. Hence, long-term control can only be achieved with cultivars carrying effective quantitative disease resistance. Breeding for resistance is a major challenge and, until now, European cultivars carry only low or moderate levels of resistance against. For example, the old oilseed rape cultivar Express is considered moderately *V. longisporum*-tolerant. Within the Brassica genus, potentially useful disease resistance sources were identified in the diploid *B. oleracea* and *B. rapa* species (Eynck et al., 2009). Generally, *B. oleracea* genotypes that have high levels of erucic acid and glucosinolates in the seed were described as carriers of various levels of resistance to *V. longisporum* (Rygulla et al., 2007). However, use of such materials in breeding is complicated because of its negative influence on seed quality and flowering/adaptation traits.

Rygulla et al. (2007) produced resynthesized forms of *B. napus* by crossing *B. rapa* (syn. *campestris*; genome AA, 2n=20) with different *B. oleracea* (genome CC, 2n=18) subspecies exhibiting partial resistance to Verticillium stem striping combined with zero erucic acid and

variable glucosinolates content in the seeds. Using a segregating *B. napus* population of 163 doubled haploid (DH) lines, two major QTL regions with significant effects in multiple environments on resistance to *V. longisporum* were mapped on C-genome linkage groups N14 and N15 (since renamed as chromosomes C04 and C05, respectively), along with minor QTL on N6 (A06) and N18 (C09), respectively (Rygulla et al., 2008). In a different DH population, Express 617 x R53, genetic analysis identified a major QTL for Verticillium stem striping on chromosome C05 and another minor QTL on chromosome C01 (Obermeier et al., 2013). The resistance alleles detected on chromosome C05 originated from the resynthesized *B. napus* parent R53, while the allele at the minor QTL on chromosome C01 originated from the commercial cultivar, Express 617. Hossain (2014) demonstrated that flanking markers derived from the major *V. longisporum* resistance QTL on chromosome C05 are identical among the two previous studies. This indicates that the resistance allele is present and stable in very different C genome genetic backgrounds, allowing breeders to use it in marker-assisted breeding for transfer of resistance to Verticillium stem striping. By combining minor and major resistance QTL from different genetic origins, breeders have the potential to use marker-assisted pyramiding strategies for an effective combination of resistance alleles from synthetic donors and elite breeding lines. These results highlight the importance of quantitative disease resistance studies in order to identify stable resistance for breeding and improve understanding of the disease resistance mechanisms.

Molecular markers derived from resistance QTL with high effects are a valuable resource for breeding of new *B. napus* cultivars with effective resistance against monogenic or/and polygenic diseases. Utilisation of marker information, careful selection of pathotypes, inoculation methods and new technologies such as next-generation sequencing, gene expression profiling and others could increase our understanding of interaction between the host and pathogen and help to identify resistance loci.

Until now, genome-wide association studies (GWAS) have been used to identify resistance alleles in many agronomical important crops. Key examples include resistances against stem rust (Zhang et al., 2014) or tan spot (Kollers et al., 2014) and spot blotch in wheat (Ayana et al., 2018), southern leaf blight (Kump et al., 2011) or head smut in maize (Wang et al., 2012), blast disease in rice (Wang et al., 2014), along with blackleg disease (Fopa Fomeju et al., 2014; Luo et al., 2017) and Sclerotinia stem rot (Wei et al., 2016; Luo et al., 2017) in oilseed rape.

1.4 Structural genome diversity and disease resistance

There is growing evidence that small, medium and long range structural genome variation plays a major role in determining important agronomical traits (Saxena et al., 2014; Żmieńko et al., 2014; Schiessl et al., 2017a). Oilseed rape has become an important model crop for studies of the influence of structural genome rearrangements arising during polyploidisation on agronomic traits (Chalhoub et al., 2014, Mason & Snowdon 2016). Samans et al. (2017) observed in 32 natural and 20 synthetic allotetraploid oilseed rape accessions that structural variation plays a key role in plant adaptation and affects, in particular, genes responsible from stress responses. In both types of material a selection bias towards homeologous chromosome exchanges among the large genomic segments from the C-subgenome with smaller A-subgenome segments was observed that seems to be a source of novel genetic diversity. Within a region frequently affected by structural variations, 8 to 14 toll interleukin-1 receptor-class nucleotide binding site–leucine-rich repeat (TIR-NBS-LRR) immune receptor genes were deleted in natural *B. napus*. Also, Higgins et al. (2018) suggested that in extreme cases, *B. napus* could be affected by rearrangements ranging up to 40 Mb. Such events might have shaped the evolution of the modern oilseed rape cultivars, with zero erucic acid and low glucosinolate content in seeds. The major loci controlling low glucosinolates in modern

cultivars underwent a loss of the C-subgenome copy of a MYB 28 transcription factor and replacement with a non-functional copy. Similarly, on the A02/C02 chromosomes homoeologous genomic exchanges determinate the generation of CNV/PAV for the flowering time genes (FLC) (Schiessl et al., 2017a,b) and Cinnamoyl-CoA-reductase (CCR1) genes involved in seed fibre and seed colour (Stein et al., 2017). Hurgobin et al. (2018) constructed a *B. napus* pangenome using the 32 natural and 20 synthetic genotypes and showed that gene presence/absence variation (PAV) is an important genetic contributor to novel diversity in oilseed rape. More than 38% of genes were estimated in this study to be affected by presence-absence variation. In the same study, a total of 307 typical R-genes were identified and it was observed that in natural *B. napus* genotypes on average 126 genes were lost while in synthetic material 107 genes were lost, respectively. Gene ontology (GO) enrichment analysis of all predicted PAV genes suggested that genome duplications or deletions are affecting, in particular, genes involved in plant-type hypersensitive responses and produce novel phenotypic diversity. Nowadays, identification of extent and impact of these structural variants in crop genomes has become feasible with ongoing advances in genome sequencing technologies and simple, cost-effective, high-throughput genotyping tools. Methods, tools and examples are described in detail in Chapter 2.

1.5 Aim and objectives of this study

Improvement of disease resistance in oilseed rape should rely on the exploitation of state-of-the-art technologies and the wide genetic diversity existing in different *Brassica* species. The simultaneous investigation of resistance factors and structural genetic diversity in oilseed rape and its progenitors for multiple diseases is crucial to increase our knowledge on plant-pathogen interactions and identification of resistance alleles. From previous results, it can be deduced that many linkage groups carry genes or QTL for resistance and that some genomic

regions seem to be involved in resistance to different pathogens (Delourme et al., 2011). However, the precise comparison of the localization of resistance factors was difficult so far, due to the low number of common molecular markers between the different maps deriving from various materials. The overriding aim of this thesis was to develop and test new methods for discovery of structural genome variants associated with quantitative disease resistances in oilseed rape.

Chapter 2 discusses the origins of structural genome variation in crops from ancient and recent polyploidization and genome duplication events, and reviews modern methods to assay such variants in crop populations in order to find associations with phenotypic traits. Recent studies showed that gene presence-absence and copy number variation are at the heart of adaptive variation of crops to counter biotic stress factors. Chapter 3 describes a novel method that allows recovery of valuable information from single nucleotide absence polymorphism (SNaP) markers by population-based quality filtering of SNP hybridization data. Including SNaP markers in genetic analyses elucidated segregation of small to large-scale structural variants in a *B. napus* nested association mapping (NAM) population. Numerous new resistance QTL were identified for two major fungal diseases of oilseed rape, Sclerotinia stem rot and blackleg disease. This strategy might also be applicable for genetic mapping in many other agronomical important crop species. Although in the literature there are many reports on the prevalence of copy number variation (CNV) and presence-absence variation (PAV) in crop species, until now their influence on quantitatively inherited agronomic traits is still largely unclear. Chapter 4 analyses and debates the association of gene PAV with resistance of oilseed rape (*Brassica napus*) against the important fungal pathogen *V. longisporum*, as an example for a complex, quantitative disease resistance in the strongly rearranged genome of a recent allopolyploid crop species. Overall, the results demonstrate the prominent role of PAV in determining agronomic traits, suggesting that this important class of polymorphism should be exploited more systematically in future breeding.

2 Connecting genome structural variation with complex traits in crop plants

Gabur I, Chawla SC, Snowdon RS, Parkin IAP

Theoretical and Applied Genetics (in press)

Theoretical and Applied Genetics
https://doi.org/10.1007/s00122-018-3233-0

ORIGINAL ARTICLE

Connecting genome structural variation with complex traits in crop plants

Iulian Gabur[1] · Harmeet Singh Chawla[1] · Rod J. Snowdon[1] · Isobel A. P. Parkin[2]

Received: 15 August 2018 / Accepted: 7 November 2018

Abstract

Key message **Structural genome variation is a major determinant of useful trait diversity. We describe how genome analysis methods are enabling discovery of trait-associated structural variants and their potential impact on breeding.**

Abstract As our understanding of complex crop genomes continues to grow, there is growing evidence that structural genome variation plays a major role in determining traits important for breeding and agriculture. Identifying the extent and impact of structural variants in crop genomes is becoming increasingly feasible with ongoing advances in the sophistication of genome sequencing technologies, particularly as it becomes easier to generate accurate long sequence reads on a genome-wide scale. In this article, we discuss the origins of structural genome variation in crops from ancient and recent genome duplication and polyploidization events and review high-throughput methods to assay such variants in crop populations in order to find associations with phenotypic traits. There is increasing evidence from such studies that gene presence–absence and copy number variation resulting from segmental chromosome exchanges may be at the heart of adaptive variation of crops to counter abiotic and biotic stress factors. We present examples from major crops that demonstrate the potential of pangenomic diversity as a key resource for future plant breeding for resilience and sustainability.

Introduction: the discovery of structural variation

With rapidly increasing sophistication in genome analysis technologies, there is growing evidence that genome-wide structural variation (SV) is a major factor underlining observed phenotypic variation in eukaryotic organisms. The first report of genic SV affecting a phenotype dates back more than 80 years, when Bridges (1936) discovered that a duplication of the *Bar* gene is associated with small eyes in the fruit fly, *Drosophila*. Genomic rearrangements have been studied extensively in humans due to their association with a range of diseases. Particularly copy number variation (CNV), an important class of structural variation, has been discovered to be causal for various autoimmune disorders (Mamtani et al. 2010), including susceptibility to human immunodeficiency virus (HIV) infection (Gonzalez 2005), Parkinson's disease (Singleton 2003) and Alzheimer's disease (Rovelet-Lecrux et al. 2006; Escaramís et al. 2015). Mounting evidence supporting the importance of SV in human genetics led to the study of the same phenomena in animal species, where numerous examples have been discovered for a role of SV in important traits, for example in mice (Keane et al. 2014), cattle (Fadista et al. 2010), pigs (Esteve-Codina et al. 2013), sheep (Liu et al. 2013) and horses (Ghosh et al. 2014; Wang et al. 2014). Structural variations were initially thought to be rare in plants, but this perspective changed dramatically with the realization that almost all flowering plants derived from multiple rounds of ancient or recent polyploidization (Vialette-Guiraud et al. 2011; Van de Peer et al. 2009; Alix et al. 2017). The ability to generate reference genome sequences even for complex crop plant genomes (Edwards et al. 2013) combined with decreased costs associated with de novo genome assembly and resequencing have accelerated the study of SV (Voss-Fels and Snowdon 2016). Numerous recent reports have clearly demonstrated that both small and large genomic

Communicated by Rajeev K. Varshney.

Iulian Gabur and Harmeet Singh Chawla contributed equally to this work.

✉ Rod J. Snowdon
rod.snowdon@agrar.uni-giessen.de

[1] Department of Plant Breeding, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

[2] Agriculture and Agri-Food Canada, 107 Science Place, Saskatoon, SK S7N OX2, Canada

Published online: 17 November 2018

rearrangements can cause major phenotypic variance affecting an array of important traits in crops (Saxena et al. 2014; Neik et al. 2017; Żmieńko et al. 2014; Schiessl et al. 2017a).

Diversity of structural variants

Genome structural variants occur in diverse forms including translocations, inversions, insertion/deletion polymorphisms (InDels), copy number variation (CNV), or simple variation in microsatellite repeat number. Traditionally, InDels have been defined as short presence/absence nucleotide polymorphism ranging from 1 to 50 bp in length, whereas a variable number of copies for larger DNA segments, ranging from a few hundred bp to several kb, is generally referred to as CNV. Gene CNV represents the most intensively studied class of SV associated directly with trait variation, whereby variants affecting intergenic regions, splicing variants and/or regulatory factors could also infer SV–trait associations. This can be mainly attributed to their ease of detection using simple molecular biology methods. Presence–absence variation (PAV) represents an extreme form of CNV where whole genomic segments are deleted from individuals within a population (Saxena et al. 2014). Different kinds of SV can occur independently or simultaneously, resulting in complex genome alterations. Many important crop genomes arose from multiple polyploidy events, in some cases involving widespread recombination among homoeologous (related but non-homologous) chromosomes. Such exchanges can result in both reciprocal or non-reciprocal exchanges. The latter, often referred to as homoeologous non-reciprocal transpositions, or HNRT (Parkin et al. 1995; Pires et al. 2004; Gaeta and Chris Pires 2010), can lead to loss or gain of DNA fragments on related chromosome homoeologues and consequently to PAV and CNV. As described in more detail later in this review, examples in recent allopolyploids like *Brassica napus* have demonstrated that this kind of exchange during early rounds of polyploidization can be a key driver of modern crop genome diversity and phenotypic plasticity (Chalhoub et al. 2014; Samans et al. 2017; Hurgobin et al. 2017).

Origins of SV

Various cellular mechanisms can trigger generation of SV during meiotic or mitotic cell division. SV events are caused by recombination errors, like non-allelic homologous recombination (NAHR) (Lupski 1998), DNA break repair errors, such as non-homologous end joining (NHEJ) (Moore and Haber 1996), or replication errors, including fork stalling and template switching (FoSTeS) (Lee et al. 2007) and microhomology-mediated break-induced replication (MMBIR) (Hastings et al. 2009). NHEJ can be triggered by misguided fusion of double-strand breaks in DNA, often resulting in insertions and/or deletions; however, in rare cases NHEJ might also generate translocations (McVey and Lee 2008). FoSTeS/MMBIR is another cellular mechanism causing major structural variations (for example large rearrangements, inversions, duplications and translocations) ranging in size from a few kb to several Mb and involves fork stalling and polymerase switching at a nearby single-stranded DNA (Stankiewicz and Lupski 2010). The most likely cause of much of the CNVs observed in plants is NAHR, which is largely the result of misalignment in genomic regions housing highly identical sequences, such as repetitive DNA, leading to duplication or deletion of genomic segments and thus copy number variants. Segmental duplications appear when highly homologous genomic regions (more than 95%) are physically positioned at distances from a few kb to some Mb from one another. Furthermore, depending on the orientation of the homology, NAHR could also cause deletions (upstream orientation on the same chromosome), inversion (downstream orientation on the same chromosome) and translocation (located on different chromosomes) (Sharp et al. 2006). The abundance of repetitive sequences in plant genomes varies widely, with published frequencies ranging from around 10% in Arabidopsis (The Arabidopsis Genome Initiative 2000) to more than 85% (in wheat) (Appels et al. 2018). The prevalence of repetitive DNA, in particular in larger crop genomes, could promote the generation of dosage effects for particular sets of genes, partly explaining the large adaptive phenotypic variation existing within the plant kingdom.

Changes in ploidy can also lead to generation of SV in plants. The majority of angiosperms studied to date show evidence of polyploidization and/or whole-genome duplication in their evolutionary history, and most modern crop species have undergone recent genome duplication events that are now known to have played a significant role in dictating their path to adaptation (Fig. 1). Some major crops contain multiple copies of entire chromosomes from spontaneous genome duplication of the same species, for example autopolyploid potato (*Solanum tuberosum*; $2n = 4x = 48$), while others arose from interspecific hybridization of subgenomes among distinct, yet related species, for example allohexaploid wheat (*Triticum aestivum*; $2n = 6x = 42$) or allotetraploid oilseed rape/canola (*Brassica napus*; $2n = 4x = 38$). Many older crop species arose by ancient duplications and paleopolyploidization. For example, the diploid cabbage species *Brassica oleracea* ($2n = 2x = 18$) and *Brassica rapa* ($2n = 2x = 20$) represent paleohexaploids which have returned to a diploid state by genome fractionation (Lagercrantz et al. 1996; Tang et al. 2012; Parkin et al. 2014).

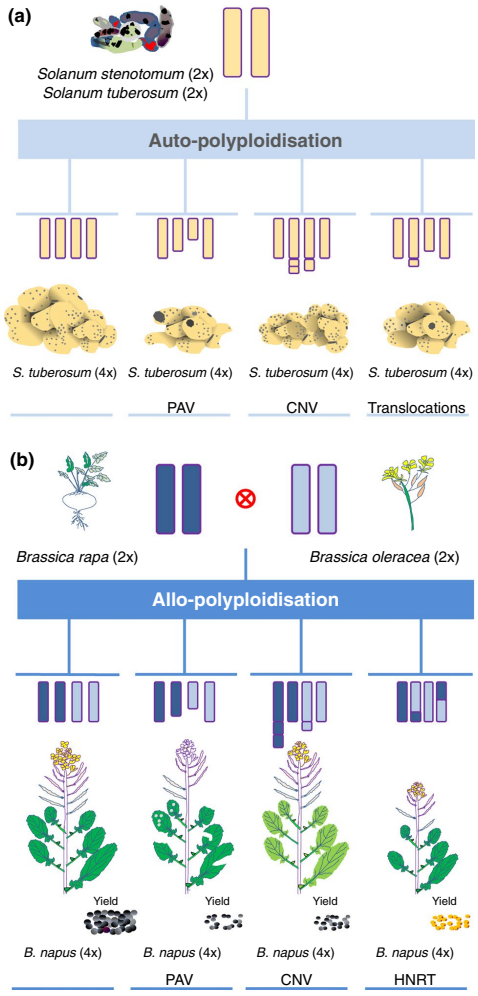


Fig. 1 Origins of different kinds of structural variants in autopolyploid and allopolyploid crops from segmental chromosome rearrangements, illustrated by the coloured bars with examples from **a** autopolyploid potato (*Solanum tuberosum*, 2n = 4x = 48) and **b** allopolyploid rapeseed/canola (*Brassica napus*, 2n = 4x = 38). Autotetraploid *S. tuberosum* arose from a spontaneous genome duplication (auto-polyploidization) of the diploid progenitor *S. stenotomum*/*S. tuberosum* (2n = 2x = 24), while *B. napus* arose from interspecific hybridization between the diploid progenitor species *B. oleracea* (2n = 2x = 18) and *B. rapa* (2n = 2x = 20). SV linked to adaptive and agronomic diversity is represented as presence–absence variation (PAV), copy number variation (CNV), translocations and homoeologous non-reciprocal transpositions (HNRT)

Visualization of large-scale SV

Classical cytology first identified evidence of large-scale chromosomal aberrations in cereals (e.g. Sears 1939), which were later confirmed as translocations using early molecular marker technologies (e.g. Gale and Devos 1998). Comparative genomic hybridization (CGH) was one of the very first methods to visualize large-scale SVs. With CGH it is possible to detect and map relative DNA sequence copy number between genomes, by hybridizing fluorescently labelled DNA from each source genome to metaphase chromosome spreads or genome-wide sequence arrays. An increase or decrease in copy number of genomic DNA (corresponding to segmental SV) can then be detected by measuring the florescence ratios between the two coloured fluorophores (Kallioniemi et al. 1992). The resolution of CGH via in situ hybridization is relatively low, with segmental SV events only visible at megabase scale, whereas array-based CGH (aCGH) can resolve smaller SV events down to a few kilobases in size. For example, Yu et al. (2011) were able to detect 641 CNVs ranging from 1.1 to 180.7 kb between two rice cultivars using aCGH. Large SV events can also be visualized directly at the chromosome level using molecular cytogenetic techniques such as fluorescence in situ hybridization (FISH) and genomic in situ hybridization (GISH) (Xiong et al. 2011; Chester et al. 2012; Horn et al. 2002; Snowdon 2007). These techniques allow physical analysis of chromosomes using chromosome arm ratios, mapping of heterochromatic regions, bacterial artificial chromosome probes (BAC-FISH) containing specific repeat sequences or molecular karyotyping of chromosome-specific probes (Xiong and Pires 2011). FISH has been used successfully in maize to analyse B chromosome non-disjunction due to *r-X1* deficiency (Tseng et al. 2017). Furthermore, the technique has been used to map homoeologous exchanges associated with agronomic traits in polyploid crop genomes. Stein et al. (2017) used BAC-FISH to identify homoeologous exchanges between two *B. napus* chromosomes associated with a QTL for seed fibre content. In potato, FISH was used to identify CNV associated with plant growth and developmental traits (Iovene et al. 2013), while Ali et al. (2016) used FISH in wheat to validate introgression of alien DNA segments that led to mosaic virus resistance. Before the availability of cheap, high-throughput genome sequencing, hybridization methods provided a relatively simple and low cost option for visualizing large SV events at the chromosomal level; however, finer resolution is required for detection of smaller SV events. With sufficient sequence depth, approaches based on next-generation sequencing (NGS) technologies provide an ideal solution.

Sequencing-based SV detection

Next-generation sequencing (NGS) approaches have accelerated the process of assembling plant reference genomes to a speed and accuracy that was unimaginable a decade ago. Furthermore, availability of methods to detect single nucleotide differences between genomes using whole-genome sequencing data, high-coverage exome sequence data or sequence capture data has been a major breakthrough in deciphering complex SV (Chen et al. 2008; Schiessl et al. 2017b). One of the key advantages of NGS-based methods for SV detection is the resolution that can be achieved by using such approaches. To date, sequenced reference genomes of varying quality are available for over 200 plant species, including most major crops (see http://www.plabipd.de/timeline_view.ep for an up-to-date overview of published plant reference genomes). As the quality of reference genome assemblies for more complex genomes continues to improve, for example by utilization of new methods like chromatin conformation associations (e.g. Mascher et al. 2017 or long-read single-molecule sequencing (Jiao et al. 2017), our ability to utilize genome-wide or targeted resequencing techniques for SV analysis in large populations will become even more powerful. Early whole-genome resequencing studies in major crops with relatively simple genomes, like sorghum, used reference-based read-mapping approaches to identify genome-wide SNP and small-scale SV. For example, Mace et al. (2013) identified 1.9 million InDels, including specific gene PAV associated with domestication and breeding, in high-coverage resequencing data from 44 genetically and geographically diverse *Sorghum bicolor* accessions. Different approaches have been developed for characterization of SVs from NGS reads, including combinations of read depth (RD), paired read (PR) and split read (SR) analysis along with de novo sequence assemblies in order to address more complex genomic re-arrangements.

Algorithms for RD analysis rely on the density of sequenced reads aligned to a locus in a reference genome for CNV identification (Alkan et al. 2009; Li and Olivier 2013). In crops, RD approaches focused mainly on calling of large SV, for example in resequencing data from sorghum (Zheng et al. 2011) or rapeseed (Samans et al. 2017). RD-based methods can detect deletions and duplications very effectively. However, limitations of the read length and the quality and coverage of the available reference genomes reduce the efficacy of this approach for detecting insertions or translocations. Furthermore, it should be noted that RD-based approaches are highly sensitive to library preparation methods. For example, PCR amplification during the sequencing library preparation can lead to either over-representation or complete absence of certain genomic regions that are difficult to amplify, which would be interpreted as duplication and deletion events, respectively, using an RD sensing algorithm. Therefore, stringent normalization of read depth is required to remove any kind of bias introduced by the library preparation. These limitations can be somewhat overcome by paired-end sequencing of single DNA fragments (paired reads). Since the sequencing library is enriched for a particular fragment size, the distance between the paired reads is pre-defined. Any insertion or deletion between the paired reads would result in a deviation from the expected mapping distance; hence, abnormally mapped read pairs might indicate the presence of SV (Korbel et al. 2007). Different types of SV can be mapped using paired reads, for example deletions or insertions (when paired reads align further apart or closer than expected), inversions (when the orientation of paired reads is inverted) or translocations (when each of the paired reads maps to different chromosomes). This approach is still highly dependent on the read coverage, size and number of repetitive elements in the genome and the quality of the reference genome, and paired read methods are best suited for detection of medium-sized insertions and deletions. However, they might not be the best choice for identification of small insertion or deletion events, due to the difficulty in distinguishing small deviations in read–pair distance from technical errors. Split read algorithms provide an alternative which also makes use of paired-end sequencing. Split read algorithms depend on accurate mapping of one of the reads from a pair, while the other read maps only partially to the reference genome (Ye et al. 2009). When reads align right across a SV break point, precise calling of breakpoints can be achieved. However, with short read NGS technologies this type of approach is only useful for detecting small-sized SV (Ye et al. 2009; Schröder et al. 2014). New opportunities to overcome these problems using long-read sequencing are described later in this review.

One major bottleneck of the methods described above is the availability of high-quality reference genomes. De novo genome sequence assembly provides the optimal method for fine-scale SV detection, but until now assembly based pangenome approaches have been largely prevented by high cost and time constraints (Hajirasouliha et al. 2010). However, costs can be significantly reduced using reduced-representation sequencing approaches which only address part of the whole genome. Reduced-representation sequencing can be achieved either by selection of restriction fragments for sequencing or by designing baits to capture certain interesting regions of the genome. Whole-exome sequencing is an example of such an approach which reduces computing and sequencing costs by focusing only on protein-coding regions. This reduces the capacity to detect large SV, but can potentially identify causal CNV when sequencing coverage

is sufficient. Exome capture has not yet been used extensively in crops, but recently a capture array was developed for barley to assay species-wide sequence diversity and SV (Mascher et al. 2013). Alternatively, targeted gene sequencing provides opportunities to capture sequence variants for specific panels of target genes, for example for QTL regions (e.g. Clarke et al. 2013) or specific biological pathways (Schiessl et al. 2017a). However, sequence capture does rely on hybridization capture and amplification steps which raise costs of library preparation and can also lead to normalization problems which must be dealt with during data analysis.

Because each method has limitations, a pragmatic approach is to use a combination of SV detection methods (Escaramís et al. 2015; Alkan et al. 2011). However, accurate and unique alignment of short sequence reads to a reference assembly is the foundation of almost every SV detection pipeline. This is extremely challenging in the case of polyploids due to the high homology between their subgenomes. The majority of the crop species reference genomes published to date are themselves based on short read sequencing, containing in some cases thousands of contigs and scaffolds that are not assembled to chromosome level due to the repetitive and complex nature of most crop genomes. The development of third-generation sequencing technologies which generate long-range sequences and enable longer, contiguous scaffolds provide new opportunities for reliable, cost-effective *de novo* assembly at whole-chromosome level (Jiao and Schneeberger 2017). In various research applications, long-read sequencing has become an efficient alternative for SV mapping and phasing. The long-read sequencing platforms from Pacific BioSciences (Menlo Park, CA, USA) and Oxford Nanopore Technologies (Oxford, UK) can provide read lengths ranging from 10 to 150 kb (Schmidt et al. 2017) depending on the DNA library quality. The sequencing error rate for both these platforms is higher compared to short read methods like Illumina sequencing. However, because the sequencing errors are randomly distributed this limitation can be overcome by increasing the depth of sequencing (Schiessl et al. 2018). By spanning rearrangement endpoints and providing more accurate reference assemblies, both of these sequencing platforms enable discovery of complex SV events which were extremely challenging to detect using only short-read methods (English et al. 2015; Chaisson et al. 2014).

Alternative technologies such as optical mapping (Bionano Genomics, San Diego, CA USA) (Lam et al. 2012) or linked-read technologies (10x Genomics, Pleasanton, CA, USA) (Mostovoy et al. 2016), which allow long distances to be effectively spanned in complex genomes, have also contributed substantially to SV detection. Even in challenging polyploid crop genomes, combinations of these different approaches can provide base pair resolution to the study of SV. Unfortunately, these high-resolution techniques are still relatively expensive, meaning that high-throughput analysis in large populations is still prohibited by cost. Until this changes, cheaper high-throughput methods like comparative genomic hybridization (CGH), single nucleotide polymorphism (SNP) arrays or real-time multiplex PCR may be viable alternatives in order to study trait-associated variants in large populations.

Analysing SV using SNP genotyping arrays

High-density SNP arrays provide a popular and cost-effective solution to analyse genetic differences among many individuals within a species. Presently the most popular platform for SNP genotyping is the Infinium™ assay from Illumina (San Diego, CA USA), which relies on hybridization of genomic fragments to probe sequences anchored in flowcells on a glass surface, with probes designed to capture approximately 50 nucleotides of unique, non-polymorphic sequence adjacent to a pre-determined SNP site. This is followed by a single-base extension using hapten-labelled nucleotides and generation of fluorescence signal by adding fluorescently labelled antibodies (Mason et al. 2017). Development of algorithms that can detect SV by quantifying the relative light intensities generated during a SNP call has been an area of research for many years. In human genetics, algorithms like QuantiSNP (Colella et al. 2007) and Birdsuite (Korn et al. 2008) use the fluorescent signal intensity of one allele relative to the other to infer a duplication or deletion event. For polyploid crop plants, the R package "gsrc" (for "genome-wide structural rearrangement calling") was developed to call rearrangements using SNP intensity information (Grandke et al. 2016). A wide range of SNP arrays have been developed in multiple crops for use in crop breeding and genetic research (Voss-Fels and Snowdon 2016).

Despite the widespread use of SNP arrays, there are some inherent problems associated with them when it comes to SV detection. The greatest problem is the limited power of detection of small SVs, due to poor resolution and ascertainment bias due to the pre-determined design of the arrays. PCR-based methods provide a simple and cheap alternative to SNP arrays, especially for detection of SVs at ultra-high resolution; however, prior knowledge of regions of interest is required and throughput is limited. Quantitative real-time PCR (qRT-PCR) and digital PCR (dPCR) are further methods capable of efficiently identifying small SV (InDels) but also translocations, inversions and CNV (Schiessl et al. 2017a; Qian et al. 2016; Ma and Chung 2001). Genes for different traits affected by PAV, CNV and InDel polymorphisms have been validated in a number of major crops using PCR, for example flowering time genes (Schiessl et al. 2017a), lignin biosynthesis genes (Stein et al. 2017)

and a stay-green gene (Qian et al. 2016) in rapeseed, copies of the boron toxicity tolerance gene *Bot1* in barley (Sutton et al. 2007), the aluminium tolerance gene *MATE1* in maize genotypes (Maron et al. 2013) and InDels in the wheat photoperiodicity genes *Ppd-A1a Ppd-B1a* (Nishida et al. 2013).

Crop pangenomes as a future reference paradigm

The unprecedented low cost and high throughput of DNA sequencing today makes it possible to generate genome sequence data for hundreds or thousands of individuals within a species. This provides a new wealth of data to discover genomic re-arrangements in crop genomes in the form of CNVs and PAVs. Insights into genomic SV have conclusively established that a single reference assembly cannot reflect the entire diversity within a species. This gave rise to the concept of pangenomes, which ideally represent all structural genome diversity present in a species. Originally coined for analysis of bacterial genomes (Tettelin et al. 2005), the pangenome concept was first adapted to plants after comparative sequencing of grass genomes revealed widespread structural variation on a previously unknown scale (Morgante et al. 2007). Since the first crop genomes became available, the pangenome concept has been investigated at many different levels especially in maize (Morgante et al. 2007; Springer et al. 2009; Lai et al. 2010; Chia et al. 2012; Hirsch et al. 2014). Most pangenome analyses so far have focused primarily on differences in gene content between individuals in a species; however, as de novo genome assemblies become more feasible in more complex organisms there is growing scope for assembly scale pangenome analysis.

A pangenome for any species is considered to comprise a so-called "core" genome, comprising all genes common to all individuals within the species, along with a "dispensable" genome consisting of partially shared genic regions that are present in some individuals but absent from others (Tettelin et al. 2005). To maximize discovery and coverage of the dispensable genome component, a pangenome should ideally include data from a broad range of individuals representative of all diversity present in the species. Pangenomic description of SV is best achieved by assembly based approaches, but due to their prohibitive cost for large and complex genomes the early pangenomes for most crop species have been generated by genomic resequencing (or in some cases skim sequencing) of representative diversity and analysis by techniques to detect and place SV in existing reference assemblies. This type of approach generally comprises three major steps: resequencing reads are mapped to a high-quality reference assembly, unmapped reads are independently assembled into additional contigs and these are inserted at the appropriate positions in the original assembly using end alignments and/or genetic mapping data. Although faster and cheaper than a de novo assembled pangenome, this method relies strongly on a high-quality reference assembly and can only capture SV in contigs that are able to be successfully assembled and placed from unmapped reads. Nevertheless, such approaches can provide cost-effective opportunities to efficiently capture genic CNV and PAV (Golicz et al. 2016a; Montenegro et al. 2017; Zhou et al. 2017).

To date there are only a handful of studies in which crop pangenomes have been created by de novo assembly of diverse individuals. The most comprehensive study so far was a pangenomic analysis of genomic variation in cultivated and wild rice (Zhao et al. 2018) in which whole-genome de novo assemblies were generated for 66 diverse genotypes chosen to represent a panel of 1529 accessions across the *Oryza sativa–Oryza rufipogon* species complex. The resulting rice pangenome identified 26,372 core genes and 16,208 dispensable genes, enabling associations of SV signatures across the pangenome to domestication sweeps and other signals of natural and artificial selection. Interestingly, several important known genes which were not observed in the original Nipponbare reference genome sequence, including the submergence tolerance1 genes *Sub1A* (Xu et al. 2006) *SNORKEL1* and *SNORKEL2* (Hattori et al. 2009), and the phosphorus-deficiency tolerance gene *Pstol* (Gamuyao et al. 2012) were discovered in the pangenome sequence (Zhao et al. 2018).

These findings reflect observations from Samans et al. (2017) in allotetraploid *B. napus* that genes involved in stress adaptation responses are particularly prevalent among genome structural variants resulting in CNV and PAV, underlining the key role of SV in crop adaption and breeding selection. Similarly, in hexaploid wheat, for which the first high-quality whole-genome reference assembly was recently published (The International Wheat Genome Sequencing Consortium 2014), a resequencing-based pangenome study including 18 wheat cultivars revealed an average of 128,656 genes per cultivar, of which 64% were found to be present in all cultivars and 49,952 genes were dispensable (Appels et al. 2018). Again, annotation of the variable set of genes revealed an enrichment for genes involved in environmental stress and defence response. Assembly based approaches to pangenome analysis will further refine these initial studies as they become more feasible with new assembly strategies and long-read sequencing capabilities. In the foreseeable future, assembly based pangenome analysis is likely to become the method of choice for generating and analysing reference genome data, even in crops with large, complex genomes like those of barley (Stein and Mascher 2019) or wheat. In a pangenome analysis based on *de novo* assemblies for wild relatives of soybean (*Glycine soja*), Li et al. (2014) found high variation in a dispensable genome

comprising around 20% of all assembled sequences, with CNV and mutations in dispensable genes showing evidence for positive selection and a strong influence on important agronomical traits. McHale et al. (2012) found previously that CNV and PAV between wild and domesticated soybean affect over 800 genes involved in biotic stress resistance, and detailed assemblies of wild vs. cultivated forms can deliver important sequence information with regard to potentially important genes that may be absent from reference cultivars.

In contrast to pangenome assembly approaches, which can miss genes not picked up by algorithms for prediction of open reading frames (ORF), some authors advocate the use of transcriptomics-based approaches as a cost-effective way to circumvent this problem. For example, He et al. (2015) introduced the concept of an ordered transcriptome for the allopolyploid *B. napus* based on gene models from its diploid progenitors *B. rapa* and *B. oleracea*, and the homoeologous diploid pan-transcriptomes as a reference to visualize SV in genetically diverse *B. napus* accessions using mRNAseq data (He et al. 2016). Such approaches provide a clear visual impression of the high degree of SV in recent polyploid crop genomes (Fig. 2). Hirsch et al. (2014) took a transcriptomics-based approach to assemble a pangenome for maize. Using this approach, they succeeded in identifying 8681 representative transcript assemblies (longest transcripts within the respective loci) which did not map to the B73 reference, 83% of which mapped only in subsets of 503 investigated lines and can be considered as dispensable genes.

Lu et al. (2015) used an alternative approach for an assembly independent pangenome analysis in maize, using linkage information to map 26 million sequence tags generated by reduced-representation sequencing of 14,129 maize inbred lines. A total of 4.4 million tags with high-confidence map

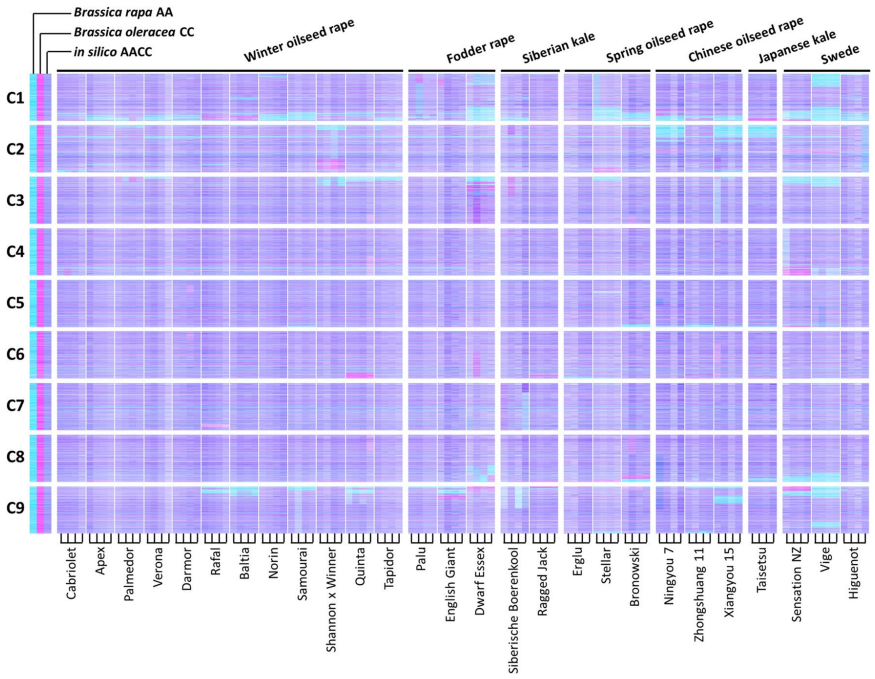


Fig. 2 Visualization of extensive structural variation (SV) caused by homoeologous genome exchanges between the A and C subgenomes of the allopolyploid crop species *Brassica napus* (oilseed rape), using Transcriptome Display Tile Plots derived from mRNAseq reads mapped to an ordered pan-transcriptome. The relative transcript abundance of homoeologous gene pairs is represented in CMYK colour space, with cyan component representing transcript abundance of the A-subgenome copy and magenta component representing transcript abundance of the C-subgenome copy. The pairs are plotted in *Brassica* C genome order (chromosomes denoted C1 to C9) for four biological replicates of each of 27 accessions of *B. napus* and controls comprising parental species and their in silico combination. Image reproduced from He et al. (2015; https://doi.org/10.1111/pbi.12657) under the terms of the Creative Commons Attribution licence 2.0

positions were selected as anchors for a high-density pangenome map. One quarter of these anchors represented PAV and showed enriched associations with phenotypic traits, providing a basis to discover genes where SV is involved in maize adaptation and agronomy. This example shows the power of combinatory approaches involving low-cost, high-throughput sequencing and population genetic analysis to define and analyse SV. Such techniques can potentially also be applied in species without extensive genomics resources.

The Brassicaceae (Cruciferae) family represents an important crop model for studying polyploidy and genomic structural re-arrangements (Mason and Snowdon 2016). Present-day allopolyploid *Brassica* crops originated by interspecific hybridization between different diploid progenitors, for example *B. napus* was formed by hybridization between *B. oleracea* and *B. rapa*. Because the diploids are themselves closely related paleopolyploids with high homoeology between their genomes, synthetic hybrids among them undergo extensive genome restructuring due to inter-homoeologue pairing during the early rounds of meiosis, leading to extensive SV (e.g. Samans et al. 2017; Zou et al. 2018). It might be reasonable to hypothesize that corresponding processes during ancient polyploidization had a similar influence on genome-wide SV and adaptive diversity in *Brassica* spp., giving rise to substantial PAV and CNV observed in present-day diploid cabbage species: such events have been found to have particularly profound effects on genes involved in biotic stress responses in *B. oleracea* (Golicz et al. 2016b) or phenylpropanoid biosynthesis in *B. rapa* (Lin et al. 2014).

Structural variation and trait diversity in major crops: key examples

Table 1 provides examples for demonstrated associations of SV to plant phenotypes in crop species. Wheat is one of the most complex plant genomes due to its large size and polyploid nature. As a result, there has been considerable delay in the detailed genomic analysis of wheat. However, early genetic mapping studies already showed that rearrangements on a number of chromosomes impact numerous important genes for resistance and adaptation traits (Nelson et al. 1995). It was also known for some time that some genes duplicated via polyploidy have remained unaltered, whereas others have been deleted or rendered non-functional by transposon insertions or premature stop codons (Gu et al. 2004). Major translocations in wheat have been associated with specific geographical regions (Riley et al. 1967; Belay and Merker 2004, 2006; Ma et al. 2015) and associated with adaptive and biotic resistance traits (Liu et al. 2016; Law and Worland 2006). With growing access to gene and sequence data, the influence of

Table 1 Examples for structural variations with demonstrated effects on agronomic traits in different crop species

Species	Type of variant	Traits associated	Reference
Barley (*Hordeum vulgare*)	CNV	Boron toxicity tolerance	Sutton et al. (2007)
	CNV	Disease resistance	Muñoz-Amatriaín et al. (2013)
Maize (*Zea mays*)	PAV, CNV	Domestication	Springer et al. (2009)
	CNV	Disease response, heterosis	Beló et al. (2010)
	CNV	–	Swanson-Wagner et al. (2010)
	CNV	Breeding selection	Jiao et al. (2012)
	CNV	Aluminium tolerance	Maron et al. (2013)
Rice (*Oryza sativa*)	PAV, CNV	Grain size, disease resistance	Xu et al. (2012)
	CNV	Disease resistance	Yang et al. (2013), Yu et al. (2013)
	InDel	Root system architecture	Uga et al. (2013)
Soybean (*Glycine max*)	PAV, CNV	Stress responses	Haun et al. (2010), McHale et al. (2012)
	CNV	Disease resistance	Lee et al. (2015)
Sorghum (*Sorghum bicolor*)	PAV, CNV	Disease resistance	Zheng et al. (2011), Mace et al. (2014)
Wheat (*Triticum aestivum*)	CNV	Vernalization, flowering time	Díaz et al. (2012), Würschum et al. (2015)
	CNV	Plant height	Li et al. (2012)
	PAV	Heading date	Nishida et al. (2013)
	CNV	Frost tolerance	Sieber et al. (2016)
	CNV	Winter hardiness	Würschum et al. (2016)
Oilseed rape (*Brassica napus*)	PAV, CNV	Flowering time	Schiessl et al. (2017b)
	HE	Seed fibre	Stein et al. (2017)
	PAV	Stay-green	Qian et al. (2016)
	PAV	Disease resistance	Gabur et al. (2018)

CNV due to polyploidization on adaptive traits like flowering time has been elucidated in more detail. For example, Díaz et al. (2012) found that variation in flowering behaviour in commercial wheat cultivars resulted from CNV for the photoperiodicity gene *Ppd-B1* and the vernalization gene *Vrn-A1*, rather than direct DNA mutations. An increase in the copy number of *Ppd-B1* was found to be associated with an early-flowering, day-length neutral phenotype, whereas plants with a higher *Vrn-A1* copy number exhibited an increased vernalization requirement. In another example, Würschum et al. (2016) reported that copy number variation of *C-repeat Binding Factor* (CBF) genes at the *Fr-A2* locus was the pivotal component for winter hardiness in a panel of 407 European winter wheat cultivars.

In addition to inter-homoeologue chromosome exchanges, interspecific hybrids of wheat with related grasses have led to rich cytogenetic stocks with segmental chromosome insertions or translocations, with a particular focus on resistance traits (Friebe et al. 1996; Wulff and Moscou 2014). One of the most well-known events is the 1BS/1RS translocation from rye, which increased drought adaptation and promoted yield performance of spring wheat in dryland production systems (Villareal et al. 1995; Reeves et al. 1999). However, the exact molecular basis of this improvement is still elusive. On the other hand, gene CNV has also been shown to shape other important phenotypic traits such as plant height in wheat. In cultivars carrying the semi-dwarfing genes *Reduced height* (*Rht*)*-B1b* and *Rht-D1b*, previously uncharacterized CNV polymorphisms of *Rht-D1* were reported to be causal for extreme dwarf phenotype, while a 90 bp insertion in *Rht-B*1 also contributed to severe dwarfism (Pearce et al. 2011). Another critical factor affecting wheat yield is photosynthetic activity determined by chlorophyll content. CNV in the wheat cytokinin oxidase gene *Tackx4*, which influences chlorophyll content and chloroplast stability via modulation of cytokinin concentration, was found to be associated with the chlorophyll content after anthesis as well as grain weight in 102 wheat varieties (Chang et al. 2016).

In barley, several studies have described gene copy number polymorphisms associated with environmental adaptation. As in wheat, CNV in the *H. vulgare CBF* orthologue at the *Frost Resistant-2 locus* (*FR-2*) was found to confer frost tolerance, with an increase in *CBF* coding sequences in winter barley compared to spring forms (Knox et al. 2010; Francia et al. 2016). Similarly, dosage effects from an increase in the number of *H. vulgare* boron transporter (*Bot1*) gene copies were found to confer boron toxicity tolerance (Sutton et al. 2007). Muñoz-Amatriaín et al. (2013) found that CNV between the barley cultivars Barke and Morex was particularly prominent for disease resistance proteins and protein kinases, while increased levels of CNVs were observed for wild accessions in comparison with cultivated barley. As for the examples mentioned above, these studies suggest a key role of SV in conferring the genome plasticity needed for adaption of barley to diverse environmental conditions.

In oilseed rape/canola, anomalies in marker segregation in mapping populations displaying otherwise normal patterns of inheritance (Parkin et al. 1995; Sharpe et al. 1995; Udall et al. 2005) provided the first evidence for exchange of genetic material between homoeologous chromosomes. Detailed elucidation revealed that in the most extreme cases such chromosomal rearrangements can range up to ~ 40 Mb in length, effectively involving whole chromosomes (Higgins et al. 2018). Interestingly, all evidence thus far shows subgenomic bias in direction of exchanges, with loss of the C genome and concomitant gain of the A genome being far more prevalent (Samans et al. 2017; Higgins et al. 2018). Early studies already suggested an important adaptive role, with a well-documented exchange between *B. napus* chromosomes A07 and C06 being associated with higher seed yields (Osborn et al. 2003). The ubiquity of such events, which have shaped the modern *B. napus* genome, was confirmed with the sequencing of the genome reference. Fixed homoeologous exchange events were found to be shared among cultivars due to intentional or inadvertent selection during allopolyploidization and/or breeding, and they thus underlie loci for a number of important traits (Chalhoub et al. 2014). One event led to loss of the C genome copy of a MYB28 transcription factor on chromosome C02 that was replaced with a non-functional A-genome copy of the same gene; this SV defined one of the strongest loci controlling the low glucosinolate phenotype that has underpinned the global success of canola as a major crop. A similar event involving the same *B. napus* chromosomes (A02/C02) created CNV for the flowering time gene *FLC* which leads to accelerated flowering in annual types. Methods to catalogue homoeologous exchanges in *B. napus* from read depth data (Samans et al. 2017) also led to the discovery of PAV underlying a QTL for seed fibre in *B. napus* (Stein et al. 2017).

A comprehensive study of natural genetic variation in homologues of 35 flowering time regulation genes in diverse *B. napus* morphotypes identified an extensive range of structural variation and potential associations to phenotypes related to flowering and secondary processes (Schiessl et al. 2017b). Different homoeologues of the vernalization response gene *Flowering Locus C* (*FLC*), the photosynthetic regulator *Phytochrome A* (*PHYA*) and the hormone *Gibberellic Acid 3-oxidase 1* (*GA3ox1*) all showed CNV and PAV associated with the derivation of *B. napus* morphotypes, again demonstrating the importance of SV on genes involved in human agricultural selection. The use of genome-wide SNP arrays to catalogue SV in multiparental segregating *B. napus* populations enables the inclusion of SV polymorphisms in genome-wide association studies. Gabur et al. (2018) utilized segregating PAV identified by

"single nucleotide absence polymorphism" (SNAP) markers (or "missing" SNP data) to discover a strong involvement of SV in the quantitative control of disease resistance in *B. napus*. Further, in combination with genome sequencing data from mapping parents the QTL could be delineated to small PAV spanning just one or a few potential candidate genes. The success of this study and the continued discovery of SVs as determinants in the control of key agronomic traits suggests that the discovery of SV should become a standard tool in future genetic analyses of crop traits.

In soybean, the world's primary pulse crop, self-pollination together with genetic bottlenecks during domestication have eroded the genetic diversity within the species (Hyten et al. 2006), with sequence polymorphism among soybean accessions typically as low as one SNP per 1000 bases. Therefore, it might also be reasonable to assume a low level of genomic SV. However, this assumption was shattered by Anderson et al. (2014) who found that a panel of 41 soybean accessions contained almost 1528 genes affected by SV. Interestingly, genes exhibiting CNVs were enriched for resistance genes with nucleotide-binding site (NBS) or NBS-leucine-rich repeat (LRR) domains, suggesting involvement of CNVs in interactions with plant pathogens. A well-known example was described by Cook et al. (2012), who found that a 31 kb sequence fragment containing an amino acid transporter, an α-SNAP protein and a WI12 (wound-inducible domain) protein that each contribute to soybean cyst nematode (SCN), one of the most devastating pathogens of soybean, was present in ten tandem copies in resistant cultivars but only a single copy in susceptible cultivars. A similar gene CNV was also reported by Liu et al. (2017) who also showed that CNV of multiple genes present in a single DNA fragment contribute towards SCN resistance.

There have been numerous examples of genome structural variations underlying commercially important traits in many vegetable species. Hardigan et al. (2016) studied genome-wide SV in homozygous clones of diploid potato (*S. tuberosum*), finding that almost 30% of the genes were tolerant to deletion or duplication, with an impact of SV on performance. As in other crop species, there was evidence that PAV and CNV impacted gene clusters in potato involved in environmental stress responses. The authors concluded that CNV may drive adaptation of potato through evolution of important pathways involved in stress responses. SVs have also been reported to play an important role in controlling several traits in tomato. Tranchida-Lombardo et al. (2018) reported over 200 deletions by resequencing and assembling of two tomato landraces. Many of these deletions were found to be localized in the genes annotated for ripening, shelf life and quality of the fruit. In cucumber (*Cucumis sativus*), a model system for sex determination studies in plants, Zhang et al. (2015) constructed a nucleotide-resolution SV map which revealed SVs in their coding regions of over 1600 genes. Using this SV map, they were able to prove that the sex determination in cucumber is controlled by CNV in four genes at the *Female* (*F*) locus.

Approximately 85% of the maize genome is composed of transposable elements (Schnable et al. 2009), which contribute significantly to genomic re-arrangements and gene PAV. In a recent effort to create a newer reference genome assembly for maize, more than 3000 SVs were detected by comparing optical maps for two inbred lines Ki11 and W22 to the B73 reference. The individual SV events ranged from 1 kb to over 1 Mb in length, with an average length of about 20 kb (Jiao et al. 2017). Because this phenomenon has been studied extensively in maize, it is not surprising that a plethora of agronomically interesting traits have been found be to be controlled by PAV in maize, ranging from abiotic and biotic stress responses to plant architecture and heterosis. For example, Wang et al. (2016) reported that an insertion in the promoter region of the *ZmVPP1* gene induces drought-dependent expression of *ZmVPP1* in drought-tolerant genotypes. The PAV in *ZmVPP1* was also associated with enhancement of photosynthetic efficiency and root development under both stress and non-stress conditions. CNV events also play a role in stress resistance responses in maize. For example, the *rp1* locus, responsible for race-specific resistance to the common rust fungus, is a hotspot for unequal crossovers leading to gain, loss or duplications in this NBS-LRR gene family. This creates a diverse haplotype makeup at the *rp1* locus, translating to variable resistance responses to various rust races (Richter et al. 1995). The same authors found a similar example for a CNV in a wall-associated kinase (*ZmWAK*) gene, which was found to confer resistance to head smut in maize. Interestingly, the responsible WAK gene was absent in many modern maize lines but present in wild relatives.

Although there is considerable indirect evidence for a role of PAV in stress responses, there is still only a handful of examples for their importance in adaptive traits in maize. Maron et al. (2013) reported association of a rare CNV in the multidrug and toxin exudation gene *MATE1* in maize to aluminium tolerance. Strikingly, the geographical origin of maize lines carrying three copies of *MATE1* coincided with highly acidic soil, implicating this CNV as an important SV conferring maize adaptation to a specific environment.

A DNA segment present or absent exclusively in germplasm adapted to a particular type of environmental cue might be indicative of the fact that genic PAV in this segment play a crucial role in adaptation. An array of INDELs in the 5′ regulatory region of the *FLOWERING LOCUS T* homologue, *LanFTc1* was found to be associated with differential vernalization response, flowering time, in narrow-leafed lupin (Taylor et al. 2018). Genotypes with no deletion exhibited an early flowering behaviour and a reduced or no response to vernalization. Such a catalogue

of structural variations could serve as the basis for the necessary environmental plasticity needed for designing the future crops adapted to wide range of environments. Darracq et al. (2018) identified in total 88 Mb of DNA in a French maize inbred line that was absent in an American inbred line, and contained 395 putative coding genes. Evidence was also observed for de novo SV in European maize alongside ancient SV, demonstrating ongoing adaptive genome evolutionary dynamics. Annotation of the novel genes revealed putative roles in biotic and biotic stress responses, in biosynthetic processes, in development, in protein synthesis and in chromatin remodelling. Intriguingly, expression of most of the novel genes was restricted to particular conditions or tissues, leading to a conclusion that at least some of the genes from the dispensable part of the genome might be involved in environmental adaptation.

The realization that heterotic pools in maize breeding programs can have vastly different gene content gave new insight into the impact of SV on heterosis. Springer et al. (2009) compared the genome structures of two maize inbred lines by comparative genomic hybridization and found that a copious amount of genomic sequences exhibited copy number differences between the two genomes. Sun et al. (2018) underlined the extent of genome-wide SV in maize by assembling the genome of Mo17 and comparing it to the B73 reference assembly. This confirmed that almost 10% of the annotated genes were exclusive to one or the other accession, while more than 20% were found to show substantial structural variation. It was also hypothesized that these SVs might be involved in heterosis and genome evolution. Furthermore, many sequences annotated as single-copy genes were found to be present in one genome but completely absent from the other genome. Although the contributing mechanisms for heterosis are still not completely elucidated and may differ from crop to crop, there is good reason to believe that fixation of complementary PAV in different heterotic pools can play an important role in exploitation of additive heterosis in hybrid breeding.

The huge diversity in rice, combined with well-defined phylogeny available for the genus *Oryza*, the small size of the genome and vast genome sequence resources make it an ideal candidate for studying effects of genomic SV on traits. Bai et al. (2016) generated a CNV map, at single nucleotide resolution for 50 rice accessions, comprising 9196 deletions compared to the Nipponbare reference genome. More than 2000 annotated genes were reported to be affected by CNV. Uga et al. (2013) identified the gene *Deep Rooting 1* (*DRO1*) as a key regulator of root system architecture with a profound effect on yield under different water regimes in rice. A japonica upland rice (Kinandang Patong) containing a full-length *DRO1* copy was found to have a deeper root system architecture, whereas the indica lowland rice cultivar IR64 carried a truncated copy due to a 1 bp InDel. This results in shallower roots due to the introduction of a premature stop codon in *DRO1* (Uga et al. 2013). Yu et al. (2011) and Yao et al. (2015) both found enrichment for disease resistance or defence response genes among dispensable genes in large rice resequencing studies. Wang et al. (2015) reported CNV at the *Grain Length on Chromosome 7* (GL7) locus associated with regulation of grain dimensions in rice. A tandem duplication of a DNA fragment within the *GL7* locus lead was found to cause upregulation of *GL7* expression and suppression of its negative regulator, thereby resulting in an increased grain length and improved grain appearance.

Outlook

As more and more genome data become available for major crops, our insight into the profound importance of SV on trait diversity continues to grow. Understanding the contribution of gene copy number and presence–absence variation to important traits will be an important factor in improving the accuracy and efficacy of many new genetic technologies in plant breeding, from genomic selection to genome editing strategies. Just a decade ago, the notion that full, high-quality reference genome assemblies for any major crop could be generated reasonably simply and quickly, at low cost, was barely conceivable. Today this is (almost) a reality, and a new era of high-throughput pangenomic analyses is set to dominate crop genetic studies in the immediate future. Although current costs of third-generation DNA sequencing technologies and chromatin conformation technologies for scaffold improvement are still high, and computational bottlenecks associated with creation of reference assemblies need to be overcome, the plummeting cost of long-read sequencing and improvement in computational algorithms and hardware could make de novo genome assembly more routine in the foreseeable future, even for complex polyploid crop genomes. One aspect of interest for breeders in a changing world is associations of SV with ecogeographical adaptations, abiotic stress adaptation and biotic stress responses. This knowledge opens fascinating new opportunities to learn from adaptive evolution of polyploid crop species in order to improve crop resilience against biotic and abiotic stress constraints in the face of climate change. From a broader perspective, studies of SV in model and crop plants derived from recent polyploidy have revealed an involvement of gene dosage and/or PAV in a wide number of different traits under natural and human (breeding) selection and showed how genome rearrangements resulting from de novo polyploidization might even be used to generate new variation for breeding. Such examples underline the role of SV as a key driver of genetic diversity for future breeding of sustainable, resilient and healthy crops. Novel methods to detect,

assay, harness and select for useful SV events will therefore be a valuable future resource for crop breeding.

Author contribution statement IG and HSC drafted the manuscript. IAP and RJS contributed text sections and edited the content.

Funding Deutscher Akademischer Austauschdienst and Deutsche Forschungsgemeinschaft (Grant No. SN14/17-1).

Compliance with ethical standards

Conflict of interest All authors jointly state that there is no conflict of interest. IAPP and RJS serve on the editorial board for this journal, but this is not considered to constitute a conflict of interest.

References

Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JSP (2017) Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. Ann Bot 120(2):183–194. https://doi.org/10.1093/aob/mcx079

Ali N, Heslop-Harrison JP, Ahmad H, Graybosch RA, Hein GL, Schwarzacher T (2016) Introgression of chromosome segments from multiple alien species in wheat breeding lines with wheat streak mosaic virus resistance. Heredity 117:114–123. https://doi.org/10.1038/hdy.2016.36

Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, Sahinalp SC, Gibbs RA, Eichler EE (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet 41(10):1061–1067. https://doi.org/10.1038/ng.437

Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. Nat Rev Genet 12(5):363–376. https://doi.org/10.1038/nrg2958

Anderson JE, Kantar MB, Kono TY, Fu F, Stec AO, Song Q, Cregan PB, Specht JE, Diers BW, Cannon SB, McHale LK, Stupar RM (2014) A roadmap for functional structural variants in the soybean genome. G3 (Bethesda) 4(7):1307–1318. https://doi.org/10.1534/g3.114.011551

Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak CJ, Choulet F, Distelfeld A, Poland J, Ronen G, Sharpe AG, Pozniak C, Barad O, Baruch K, Keeble-Gagnère G, Mascher M, Ben-Zvi G, Josselin A-A, Himmelbach A, Balfourier F, Gutierrez-Gonzalez J, Hayden M, Koh C, Muehlbauer G, Pasam RK, Paux E, Rigault P, Tibbits J, Tiwari V, Spannagl M, Lang D, Gundlach H, Haberer G, Mayer KFX, Ormanbekova D, Prade V, Šimková H, Wicker T, Swarbreck D, Rimbert H, Felder M, Guilhot N, Kaithakottil G, Keilwagen J, Leroy P, Lux T, Twardziok S, Venturini L, Juhász A, Abrouk M, Fischer I, Uauy C, Borrill P, Ramirez-Gonzalez RH, Arnaud D, Chalabi S, Chalhoub B, Cory A, Datla R, Davey MW, Jacobs J, Robinson SJ, Steuernagel B, van Ex F, Wulff BBH, Benhamed M, Bendahmane A, Concia L, Latrasse D, Alaux M, Bartoš J, Bellec A, Berges H, Doležel J, Frenkel Z, Gill B, Korol A, Letellier T, Olsen O-A, Singh K, Valárik M, van der Vossen E, Vautrin S, Weining S, Fahima T, Glikson V, Raats D, Číhalíková J, Toegelová H, Vrána J, Sourdille P, Darrier B, Barabaschi D, Cattivelli L, Hernandez P, Galvez S, Budak H, Jones JDG, Witek K, Yu G, Small I, Melonek J, Zhou R, Belova T, Kanyuka K, King R, Nilsen K, Walkowiak S, Cuthbert R, Knox R, Wiebe K, Xiang D, Rohde A, Golds T, Čížková J, Akpinar BA, Biyiklioglu S, Gao L, N'Daiye A, Kubaláková M, Šafář J, Alfama F, Adam-Blondon A-F, Flores R, Guerche C, Loaec M, Quesneville H, Condie J, Ens J, Maclachlan R, Tan Y, Alberti A, Aury J-M, Barbe V, Couloux A, Cruaud C, Labadie K, Mangenot S, Wincker P, Kaur G, Luo M, Sehgal S, Chhuneja P, Gupta OP, Jindal S, Kaur P, Malik P, Sharma P, Yadav B, Singh NK, Khurana J, Chaudhary C, Khurana P, Kumar V, Mahato A, Mathur S, Sevanthi A, Sharma N, Tomar RS, Holušová K, Plíhal O, Clark MD, Heavens D, Kettleborough G, Wright J, Balcárková B, Hu Y, Salina E, Ravin N, Skryabin K, Beletsky A, Kadnikov V, Mardanov A, Nesterov M, Rakitin A, Sergeeva E, Handa H, Kanamori H, Katagiri S, Kobayashi F, Nasuda S, Tanaka T, Wu J, Cattonaro F, Jiumeng M, Kugler K, Pfeifer M, Sandve S, Xun X, Zhan B, Batley J, Bayer PE, Edwards D, Hayashi S, Tulpová Z, Visendi P, Cui L, Du X, Feng K, Nie X, Tong W, Wang L (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science. https://doi.org/10.1126/science.aar7191

Bai Z, Chen J, Liao Y, Wang M, Liu R, Ge S, Wing RA, Chen M (2016) The impact and origin of copy number variations in the *Oryza* species. BMC Genomics 17:261. https://doi.org/10.1186/s12864-016-2589-2

Belay G, Merker A (2004) Cytogenetic studies in Ethiopian landraces of tetraploid wheat (*Tviticum Turgidum* L.). II. Spontaneous chromosome translocations and fertility. Hereditas 126(1):35–43. https://doi.org/10.1111/j.1601-5223.1997.00035.x

Belay G, Merker A (2006) Cytogenetic analysis of a spontaneous 5B/6B translocation in tetraploid wheat landraces from Ethiopia, and implications for breeding. Plant Breed 117(6):537–542. https://doi.org/10.1111/j.1439-0523.1998.tb02203.x

Beló A, Beatty MK, Hondred D, Fengler KA, Li B, Rafalski A (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theor Appl Genet 120(2):355–367. https://doi.org/10.1007/s00122-009-1128-9

Bridges CB (1936) The BAR "gene" a duplication. Science 83(2148):210–211. https://doi.org/10.1126/science.83.2148.210

Chaisson MJP, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M, Landolin JM, Stamatoyannopoulos JA, Hunkapiller MW, Korlach J, Eichler EE (2014) Resolving the complexity of the human genome using single-molecule sequencing. Nature 517(7536):608–611. https://doi.org/10.1038/nature13907

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Correa 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, Canaguier A, Chauveau A, Berard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, 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) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345(6199):950–953. https://doi.org/10.1126/science.1253435

Chang C, Lu J, Zhang H-P, Ma C-X, Sun G (2016) Copy number variation of cytokinin oxidase gene *Tackx4* associated with grain weight and chlorophyll content of flag leaf in common wheat. PLoS ONE 10(12):e0145970. https://doi.org/10.1371/journal.pone.0145970

Chen W, Kalscheuer V, Tzschach A, Menzel C, Ullmann R, Schulz MH, Erdogan F, Li N, Kijas Z, Arkesteijn G, Pajares IL, Goetz-Sothmann M, Heinrich U, Rost I, Dufke A, Grasshoff U, Glaeser B, Vingron M, Ropers HH (2008) Mapping translocation breakpoints by next-generation sequencing. Genome Res 18(7):1143–1149. https://doi.org/10.1101/gr.076166.108

Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE (2012) Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). Proc Natl Acad Sci USA 109(4):1176–1181. https://doi.org/10.1073/pnas.1112041109

Chia J-M, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai J, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhäjärvi T, Rong T, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang Z, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang G, Xu Y, Ware D (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44:803. https://doi.org/10.1038/ng.2313

Clarke WE, Parkin IA, Gajardo HA, Gerhardt DJ, Higgins E, Sidebottom C, Sharpe AG, Snowdon RJ, Federico ML, Iniguez-Luy FL (2013) Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover sequence nucleotide polymorphisms (SNP) in *Brassica napus* L. PLoS ONE 8(12):e81992. https://doi.org/10.1371/journal.pone.0081992

Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J (2007) QuantiSNP: an objective Bayes hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res 35(6):2013–2025. https://doi.org/10.1093/nar/gkm076

Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE, Diers BW, Jiang J, Hudson ME, Bent AF (2012) Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. Science 338(6111):1206–1209. https://doi.org/10.1126/science.1228746

Darracq A, Vitte C, Nicolas S, Duarte J, Pichon J-P, Mary-Huard T, Chevalier C, Bérard A, Le Paslier M-C, Rogowsky P, Charcosset A, Joets J (2018) Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. BMC Genomics 19(1):119. https://doi.org/10.1186/s12864-018-4490-7

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(3):e33234. https://doi.org/10.1371/journal.pone.0033234

Edwards D, Batley J, Snowdon RJ (2013) Accessing complex crop genomes with next-generation sequencing. Theor Appl Genet 126(1):1–11. https://doi.org/10.1007/s00122-012-1964-x

English AC, Salerno WJ, Hampton OA, Gonzaga-Jauregui C, Ambreth S, Ritter DI, Beck CR, Davis CF, Dahdouli M, Ma S, Carroll A, Veeraraghavan N, Bruestle J, Drees B, Hastie A, Lam ET, White S, Mishra P, Wang M, Han Y, Zhang F, Stankiewicz P, Wheeler DA, Reid JG, Muzny DM, Rogers J, Sabo A, Worley KC, Lupski JR, Boerwinkle E, Gibbs RA (2015) Assessing structural variation in a personal genome—towards a human reference diploid genome. BMC Genomics 16(1):332. https://doi.org/10.1186/s12864-015-1479-3

Escaramís G, Docampo E, Rabionet R (2015) A decade of structural variants: description, history and methods to detect structural variation. Brief Funct Genomics 14(5):305–314. https://doi.org/10.1093/bfgp/elv014

Esteve-Codina A, Paudel Y, Ferretti L, Raineri E, Megens H-J, Silió L, Rodríguez MC, Am Groenen M, Ramos-Onsins SE, Pérez-Enciso M (2013) Dissecting structural and nucleotide genome-wide variation in inbred Iberian pigs. BMC Genomics 14(1):148. https://doi.org/10.1186/1471-2164-14-148

Fadista J, Thomsen B, Holm L-E, Bendixen C (2010) Copy number variation in the bovine genome. BMC Genomics 11(1):284. https://doi.org/10.1186/1471-2164-11-284

Francia E, Morcia C, Pasquariello M, Mazzamurro V, Milc JA, Rizza F, Terzi V, Pecchioni N (2016) Copy number variation at the *HvCBF4–HvCBF2* genomic segment is a major component of frost resistance in barley. Plant Mol Biol 92(1):161–175. https://doi.org/10.1007/s11103-016-0505-4

Friebe B, Jiang J, Raupp WJ, McIntosh RA, Gill BS (1996) Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. Euphytica 91(1):59–87. https://doi.org/10.1007/BF00035277

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. https://doi.org/10.1111/pbi.12942

Gaeta RT, Chris Pires J (2010) Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol 186(1):18–28. https://doi.org/10.1111/j.1469-8137.2009.03089.x

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95(5):1971. https://doi.org/10.1073/pnas.95.5.1971

Gamuyao R, Chin JH, Pariasca-Tanaka J, Pesaresi P, Catausan S, Dalid C, Slamet-Loedin I, Tecson-Mendoza EM, Wissuwa M, Heuer S (2012) The protein kinase *Pstol1* from traditional rice confers tolerance of phosphorus deficiency. Nature 488(7412):535–539. https://doi.org/10.1038/nature11346

Ghosh S, Qu Z, Das PJ, Fang E, Juras R, Cothran EG, McDonell S, Kenney DG, Lear TL, Adelson DL, Chowdhary BP, Raudsepp T (2014) Copy number variation in the horse genome. PLoS Genet 10(10):e1004712. https://doi.org/10.1371/journal.pgen.1004712

Golicz AA, Batley J, Edwards D (2016a) Towards plant pangenomics. Plant Biotechnol J 14(4):1099–1105. https://doi.org/10.1111/pbi.12499

Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, Chan CKK, Severn-Ellis A, McCombie WR, Parkin IAP, Paterson AH, Pires JC, Sharpe AG, Tang H, Teakle GR, Town CD, Batley J, Edwards D (2016b) The pangenome of an agronomically important crop plant *Brassica oleracea*. Nat Commun 7:13390. https://doi.org/10.1038/ncomms13390

Gonzalez E (2005) The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307(5714):1434–1440. https://doi.org/10.1126/science.1101160

Grandke F, Snowdon R, Samans B (2016) gsrc: an R package for genome structure rearrangement calling. Bioinformatics 3:545–546. https://doi.org/10.1093/bioinformatics/btw648

Gu YQ, Coleman-Derr D, Kong X, Anderson OD (2004) Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four *Triticeae* Genomes. Plant Physiol 135(1):459. https://doi.org/10.1104/pp.103.038083

Hajirasouliha I, Hormozdiari F, Alkan C, Kidd JM, Birol I, Eichler EE, Sahinalp SC (2010) Detection and characterization of novel sequence insertions using paired-end next-generation sequencing. Bioinformatics 26(10):1277–1283. https://doi.org/10.1093/bioinformatics/btq152

Hardigan MA, Crisovan E, Hamilton JP, Kim J, Laimbeer P, Leisner CP, Manrique-Carpintero NC, Newton L, Pham GM, Vaillancourt B, Yang X, Zeng Z, Douches DS, Jiang J, Veilleux RE, Buell CR (2016) Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. Plant Cell 28(2):388–405. https://doi.org/10.1105/tpc.15.00538

Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in gene copy number. Nat Rev Genet 10(8):551–564. https://doi.org/10.1038/nrg2593

Hattori Y, Nagai K, Furukawa S, Song X-J, Kawano R, Sakakibara H, Wu J, Matsumoto T, Yoshimura A, Kitano H, Matsuoka M, Mori H, Ashikari M (2009) The ethylene response factors *SNORKEL1* and *SNORKEL2* allow rice to adapt to deep water. Nature 460(7258):1026–1030. https://doi.org/10.1038/nature08258

Haun WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, Jeddeloh JA, Jia G, Springer NM, Vance CP, Stupar RM (2010) The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. Plant Physiol 155(2):645–655. https://doi.org/10.1104/pp.110.166736

He Z, Cheng F, Li Y, Wang X, Parkin IAP, Chalhoub B, Liu S, Bancroft I (2015) Construction of *Brassica* A and C genome-based ordered pan-transcriptomes for use in rapeseed genomic research. Data Brief 4:357–362. https://doi.org/10.1016/j.dib.2015.06.016

He Z, Wang L, Harper AL, Havlickova L, Pradhan AK, Parkin IAP, Bancroft I (2016) Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. Plant Biotech J 15:594–604. https://doi.org/10.1111/pbi.12657

Higgins EE, Clarke WE, Howell EC, Armstrong SJ, Parkin IAP (2018) Detecting *de Novo* homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array. G3 (Bethesda) 8(8):2673–2683. https://doi.org/10.1534/g3.118.200118

Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Penagaricano F, Lindquist E, Pedraza MA, Barry K, de Leon N, Kaeppler SM, Buell CR (2014) Insights into the maize pan-genome and pan-transcriptome. Plant Cell 26(1):121–135. https://doi.org/10.1105/tpc.113.119982

Horn R, Snowdon R, Kusterer B (2002) Structural genome analysis using molecular cytogenetic techniques. In: Esser K, Lüttge U, Beyschlag W, Hellwig F (eds) Progress in botany: genetics, physiology, ecology. Springer, Berlin, pp 55–79

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 (2017) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. Plant Biotechnol J 16(7):1265–1274. https://doi.org/10.1111/pbi.12867

Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA 103(45):16666. https://doi.org/10.1073/pnas.0604379103

Iovene M, Zhang T, Lou Q, Buell CR, Jiang J (2013) Copy number variation in potato—an asexually propagated autotetraploid species. Plant J 75(1):80–89. https://doi.org/10.1111/tpj.12200

Jiao W-B, Schneeberger K (2017) The impact of third generation genomic technologies on plant genome assembly. Curr Opin Plant Biol 36:64–70. https://doi.org/10.1016/j.pbi.2017.02.002

Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, Zhang M, Xie S, Lai J (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44(7):812–815. https://doi.org/10.1038/ng.2312

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin C-S, Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR, Springer NM, Antoniou E, McCombie WR, Presting GG, McMullen M, Ross-Ibarra J, Dawe RK, Hastie A, Rank DR, Ware D (2017) Improved maize reference genome with single-molecule technologies. Nature 546:524. https://doi.org/10.1038/nature22971

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258(5083):818. https://doi.org/10.1126/science.1359641

Keane TM, Wong K, Adams DJ, Flint J, Reymond A, Yalcin B (2014) Identification of structural variation in mouse genomes. Front Genet 5:1061. https://doi.org/10.3389/fgene.2014.00192

Knox AK, Dhillon T, Cheng H, Tondelli A, Pecchioni N, Stockinger EJ (2010) CBF gene copy number variation at *Frost Resistance-2* is associated with levels of freezing tolerance in temperate-climate cereals. Theor Appl Genet 121(1):21–35. https://doi.org/10.1007/s00122-010-1288-7

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders ACE, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science 318(5849):420–426. https://doi.org/10.1126/science.1149504

Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D (2008) Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet 40(10):1253–1260. https://doi.org/10.1038/ng.237

Lagercrantz U, Putterill J, Coupland G, Lydiate D (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. Plant J 9(1):13–20

Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M, Jiao Y, Ni P, Zhang J, Li D, Guo X, Ye K, Jian M, Wang B, Zheng H, Liang H, Zhang X, Wang S, Chen S, Li J, Fu Y, Springer NM, Yang H, Wang J, Dai J, Schnable PS, Wang J (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42(11):1027–1030. https://doi.org/10.1038/ng.684

Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, Deshpande P, Cao H, Nagarajan N, Xiao M, Kwok P-Y (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 30(8):771–776. https://doi.org/10.1038/nbt.2303

Law CN, Worland AJ (2006) The control of adult-plant resistance to yellow rust by the translocated chromosome 5BS–7BS of bread wheat. Plant Breed 116(1):59–63. https://doi.org/10.1111/j.1439-0523.1997.tb00975.x

Lee JA, Carvalho CMB, Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131(7):1235–1247. https://doi.org/10.1016/j.cell.2007.11.037

Lee TG, Kumar I, Diers BW, Hudson ME (2015) Evolution and selection of *Rhg1*, a copy-number variant nematode-resistance locus. Mol Ecol 24(8):1774–1791. https://doi.org/10.1111/mec.13138

Li W, Olivier M (2013) Current analysis platforms and methods for detecting copy number variation. Physiol Genom 45(1):1–16. https://doi.org/10.1152/physiolgenomics.00082.2012

Li Y, Xiao J, Wu J, Duan J, Liu Y, Ye X, Zhang X, Guo X, Gu Y, Zhang L, Jia J, Kong X (2012) A tandem segmental duplication (TSD) in green revolution gene *Rht-D1b* region underlies plant height variation. New Phytol 196(1):282–291. https://doi.org/10.1111/j.1469-8137.2012.04243.x

Li Y, Zhou G, Ma J, Jiang W, Jin L, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang S, Zuo Q, Shi X, Li Y, Zhang W, Hu Y, Kong G, H-l H, Tan B, Song J, Liu Z, Wang Y, Ruan H, Yeung CKL, Liu J, Wang H, Zhang L, Guan R, Wang K, Li W, Chen S, Chang R, Jiang Z, Jackson SA, Li R, Qiu L (2014) *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and

agronomic traits. Nat Biotechnol 32(10):1045–1052. https://doi.org/10.1038/nbt.2979
Lin K, Zhang N, Severing EI, Nijveen H, Cheng F, Visser RGF, Wang X, de Ridder D, Bonnema G (2014) Beyond genomic variation—comparison and functional annotation of three *Brassica* rapagenomes: a turnip, a rapid cycling and a Chinese cabbage. BMC Genomics 15(1):250. https://doi.org/10.1186/1471-2164-15-250
Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F, Du L (2013) Analysis of copy number variations in the sheep genome using 50 K SNP BeadChip array. BMC Genomics 14(1):229. https://doi.org/10.1186/1471-2164-14-229
Liu M, Stiller J, Holušová K, Vrána J, Liu D, Doležel J, Liu C (2016) Chromosome-specific sequencing reveals an extensive dispensable genome component in wheat. Sci Rep. https://doi.org/10.1038/srep36398
Liu S, Kandoth PK, Lakhssassi N, Kang J, Colantonio V, Heinz R, Yeckel G, Zhou Z, Bekal S, Dapprich J, Rotter B, Cianzio S, Mitchum MG, Meksem K (2017) The soybean *GmSNAP18* gene underlies two types of resistance to soybean cyst nematode. Nat Commun 8:14822. https://doi.org/10.1038/ncomms14822
Lu F, Romay MC, Glaubitz JC, Bradbury PJ, Elshire RJ, Wang T, Li Y, Li Y, Semagn K, Zhang X, Hernandez AG, Mikel MA, Soifer I, Barad O, Buckler ES (2015) High-resolution genetic mapping of maize pan-genome sequence anchors. Nat Commun. https://doi.org/10.1038/ncomms7914
Lupski JR (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 14(10):417–422. https://doi.org/10.1016/S0168-9525(98)01555-8
Ma L, Chung WK (2001) Quantitative analysis of copy number variants based on real-time lightcycler PCR. In: Haines JL, Korf BR, Morton CC, Seidman CE, Seidman JG, Smith DR (eds) Current protocols in human genetics, vol 107. Wiley, Hoboken, pp 7.21.1–7.21.8
Ma J, Stiller J, Zheng Z, Wei Y, Zheng Y-L, Yan G, Doležel J, Liu C (2015) Putative interchromosomal rearrangements in the hexaploid wheat (*Triticum aestivum* L.) genotype 'Chinese Spring' revealed by gene locations on homoeologous chromosomes. BMC Evol Biol 15:37. https://doi.org/10.1186/s12862-015-0313-5
Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, Campbell BC, Hu W, Innes DJ, Han X, Cruickshank A, Dai C, Frère C, Zhang H, Hunt CH, Wang X, Shatte T, Wang M, Su Z, Li J, Lin X, Godwin ID, Jordan DR, Wang J (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun 4:2320. https://doi.org/10.1038/ncomms3320
Mace E, Tai S, Innes D, Godwin I, Hu W, Campbell B, Gilding E, Cruickshank A, Prentis P, Wang J, Jordan D (2014) The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes. BMC Plant Biol. https://doi.org/10.1186/s12870-014-0253-z
Mamtani M, Anaya J-M, He W, Ahuja SK (2010) Association of copy number variation in the *FCGR3B* gene with risk of autoimmune diseases. Genes Immun 11(2):155–160. https://doi.org/10.1038/gene.2009.71
Maron LG, Guimaraes CT, Kirst M, Albert PS, Birchler JA, Bradbury PJ, Buckler ES, Coluccio AE, Danilova TV, Kudrna D, Magalhaes JV, Pineros MA, Schatz MC, Wing RA, Kochian LV (2013) Aluminum tolerance in maize is associated with higher *MATE1* gene copy number. Proc Natl Acad Sci USA 110(13):5241–5246. https://doi.org/10.1073/pnas.1220766110
Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Ayling S, Steuernagel B, Pfeifer M, D'Ascenzo M, Akhunov ED, Hedley PE, Gonzales AM, Morrell PL, Kilian B, Blattner FR, Scholz U, Mayer KFX, Flavell AJ, Muehlbauer GJ, Waugh R, Jeddeloh JA, Stein N (2013) Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. Plant J 76(3):494–505. https://doi.org/10.1111/tpj.12294
Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, Bayer M, Ramsay L, Liu H, Haberer G, Zhang X-Q, Zhang Q, Barrero RA, Li L, Taudien S, Groth M, Felder M, Hastie A, Šimková H, Staňková H, Vrána J, Chan S, Muñoz-Amatriaín M, Ounit R, Wanamaker S, Bolser D, Colmsee C, Schmutzer T, Aliyeva-Schnorr L, Grasso S, Tanskanen J, Chailyan A, Sampath D, Heavens D, Clissold L, Cao S, Chapman B, Dai F, Han Y, Li H, Li X, Lin C, McCooke JK, Tan C, Wang P, Wang S, Yin S, Zhou G, Poland JA, Bellgard MI, Borisjuk L, Houben A, Doležel J, Ayling S, Lonardi S, Kersey P, Langridge P, Muehlbauer GJ, Clark MD, Caccamo M, Schulman AH, Mayer KFX, Platzer M, Close TJ, Scholz U, Hansson M, Zhang G, Braumann I, Spannagl M, Li C, Waugh R, Stein N (2017) A chromosome conformation capture ordered sequence of the barley genome. Nature 544(7651):427–433. https://doi.org/10.1038/nature22043
Mason AS, Snowdon RJ (2016) Oilseed rape: learning about ancient and recent polyploid evolution from a recent crop species. Plant Biol (Stuttg) 18(6):883–892. https://doi.org/10.1111/plb.12462
Mason AS, Higgins EE, Snowdon RJ, Batley J, Stein A, Werner C, Parkin IAP (2017) A user guide to the Brassica 60 K Illumina Infinium™ SNP genotyping array. Theor Appl Genet 130(4):621–633. https://doi.org/10.1007/s00122-016-2849-1
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(4):1295–1308. https://doi.org/10.1104/pp.112.194605
McVey M, Lee SE (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet 24(11):529–538. https://doi.org/10.1016/j.tig.2008.08.007
Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan C-KK, Visendi P, Lai K, Doležel J, Batley J, Edwards D (2017) The pangenome of hexaploid bread wheat. Plant J 90(5):1007–1013. https://doi.org/10.1111/tpj.13515
Moore JK, Haber JE (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. Mol Cell Biol 16(5):2164–2173. https://doi.org/10.1128/MCB.16.5.2164
Morgante M, de Paoli E, Radovic S (2007) Transposable elements and the plant pan-genomes. Curr Opin Plant Biol 10(2):149–155. https://doi.org/10.1016/j.pbi.2007.02.001
Mostovoy Y, Levy-Sakin M, Lam J, Lam ET, Hastie AR, Marks P, Lee J, Chu C, Lin C, Džakula Ž, Cao H, Schlebusch SA, Giorda K, Schnall-Levin M, Wall JD, Kwok P-Y (2016) A hybrid approach for *de novo* human genome sequence assembly and phasing. Nat Methods 13(7):587–590. https://doi.org/10.1038/nmeth.3865
Muñoz-Amatriaín M, Eichten SR, Wicker T, Richmond TA, Mascher M, Steuernagel B, Scholz U, Ariyadasa R, Spannagl M, Nussbaumer T, Mayer KFX, Taudien S, Platzer M, Jeddeloh JA, Springer NM, Muehlbauer GJ, Stein N (2013) Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. Genome Biol 14(6):203. https://doi.org/10.1186/gb-2013-14-6-r58
Neik TX, Barbetti MJ, Batley J (2017) Current status and challenges in identifying disease resistance genes in *Brassica napus*. Front Plant Sci 8:1788. https://doi.org/10.3389/fpls.2017.01788
Nelson JC, Sorrells ME, Van-Deynze AE, Lu YH, Atkinson M, Bernard M, Leroy P, Faris JD, Anderson JA (1995) Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5, and 7. Genetics 141(2):721–731

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 Breeding 31(1):27–37. https://doi.org/10.1007/s11032-012-9765-0

Osborn TC, Butrulle DV, Sharpe AG, Pickering KJ, Parkin IAP, Parker JS, Lydiate DJ (2003) Detection and effects of a homeologous reciprocal transposition in *Brassica napus*. Genetics 165(3):1569–1577

Parkin IAP, Sharpe AG, Keith DJ, Lydiate DJ (1995) Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). Genome 38(6):1122–1131. https://doi.org/10.1139/g95-149

Parkin IAP, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, Denoeud F, Belcram H, Links MG, Just J, Clarke C, Bender T, Huebert T, Mason AS, Pires JC, Barker G, Moore J, Walley PG, Manoli S, Batley J, Edwards D, Nelson MN, Wang X, Paterson AH, King G, Bancroft I, Chalhoub B, Sharpe AG (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome Biol 15(6):R77. https://doi.org/10.1186/gb-2014-15-6-r77

Pearce S, Saville R, Vaughan SP, Chandler PM, Wilhelm EP, Sparks CA, Al-Kaff N, Korolev A, Boulton MI, Phillips AL, Hedden P, Nicholson P, Thomas SG (2011) Molecular characterization of *Rht-1 d*warfing genes in hexaploid wheat. Plant Physiol 157(4):1820. https://doi.org/10.1104/pp.111.183657

Pires JC, Zhao JW, Schranz ME, Leon EJ, Quijada PA, Lukens LN, Osborn TC (2004) Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (*Brassicaceae*). Biol J Linn Soc Lond 82(4):675–688. https://doi.org/10.1111/j.1095-8312.2004.00350.x

Qian L, Voss-Fels K, Cui Y, Jan HU, Samans B, Obermeier C, Qian W, Snowdon RJ (2016) Deletion of a stay-green gene associates with adaptive selection in *Brassica napus*. Mol Plant 9(12):1559–1569. https://doi.org/10.1016/j.molp.2016.10.017

Reeves TG, Rajaram S, van Ginkel M, Trethowan R, Braun H, Cassaday K (1999) New wheats for a secure, sustainable future. CIMMYT, Mexico

Richter TE, Pryor TJ, Bennetzen JL, Hulbert SH (1995) New rust resistance specificities associated with recombination in the *Rp1* complex in maize. Genetics 141(1):373–381

Riley R, Coucoli H, Chapman V (1967) Chromosomal interchanges and the phylogeny of wheat. Heredity 22:233. https://doi.org/10.1038/hdy.1967.29

Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerrière A, Vital A, Dumanchin C, Feuillette S, Brice A, Vercelletto M, Dubas F, Frebourg T, Campion D (2006) *APP* locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 38(1):24–26. https://doi.org/10.1038/ng1718

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

Saxena RK, Edwards D, Varshney RK (2014) Structural variations in plant genomes. Brief Funct Genomics 13(4):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. https://doi.org/10.1038/sdata.2017.13

Schiessl S-V, Katche E, Ihien E, Chawla HS, Mason AS (2018) The role of genomic structural variation in the genetic improvement of polyploid crops. Crop J. https://doi.org/10.1016/j.cj.2018.07.006

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(10):2336–2348. https://doi.org/10.1105/tpc.17.00521

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, van Buren P, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956):1112–1115. https://doi.org/10.1126/science.1178534

Schröder J, Hsu A, Boyle SE, Macintyre G, Cmero M, Tothill RW, Johnstone RW, Shackleton M, Papenfuss AT (2014) Socrates: identification of genomic rearrangements in tumour genomes by re-aligning soft clipped reads. Bioinformatics 30(8):1064–1072. https://doi.org/10.1093/bioinformatics/btt767

Sears ER (1939) Cytogenetic studies with polyploid species of wheat. I. chromosomal aberrations in the progeny of a haploid of *Triticum Vulgare*. Genetics 24(4):509–523

Sharp AJ, Cheng Z, Eichler EE (2006) Structural variation of the human genome. Annu Rev Genomics Hum Genet 7:407–442. https://doi.org/10.1146/annurev.genom.7.080505.115618

Sharpe AG, Parkin IA, Keith DJ, Lydiate DJ (1995) Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome 38(6):1112–1121

Sieber A-N, Longin CFH, Leiser WL, Würschum T (2016) Copy number variation of *CBF-A14* at the *Fr-A2* locus determines frost tolerance in winter durum wheat. Theor Appl Genet 129(6):1087–1097. https://doi.org/10.1007/s00122-016-2685-3

Singleton AB (2003) Synuclein locus triplication causes Parkinson's disease. Science 302(5646):841. https://doi.org/10.1126/science.1090278

Snowdon RJ (2007) Cytogenetics and genome analysis in *Brassica* crops. Chromosome Res 15(1):85–95. https://doi.org/10.1007/s10577-006-1105-y

Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL, Barbazuk WB, Jeddeloh JA, Nettleton D, Schnable PS, Ecker JR (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet 5(11):e1000734. https://doi.org/10.1371/journal.pgen.1000734

Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. Annu Rev Med 61(1):437–455. https://doi.org/10.1146/annurev-med-100708-204735

Stein N, Mascher M (2019) Capturing pangenome diversity for breeding: a case study in barley. Theor Appl Genet **(to be published in same special issue)**

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(11):1478–1489. https://doi.org/10.1111/pbi.12732

Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H, Song W, Zhang M, Cui Y, Dong X, Liu H, Ma X, Jiao Y, Wang B, Wei X, Stein JC, Glaubitz JC, Lu F, Yu G, Liang C, Fengler K, Li B, Rafalski A, Schnable PS, Ware DH, Buckler ES, Lai J (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. Nat Genet 50(9):1289–1295. https://doi.org/10.1038/s41588-018-0182-0

Sutton T, Baumann U, Hayes J, Collins NC, Shi B-J, Schnurbusch T, Hay A, Mayo G, Pallotta M, Tester M, Langridge P (2007) Boron-toxicity tolerance in barley arising from efflux transporter amplification. Science 318(5855):1446–1449. https://doi.org/10.1126/science.1146853

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, Ware D, Springer NM (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res 20(12):1689–1699. https://doi.org/10.1101/gr.109165.110

Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, Wang X, Freeling M, Pires JC (2012) Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. Genetics 190(4):1563–1574. https://doi.org/10.1534/genetics.111.137349

Taylor CM, Kamphuis LG, Zhang W, Garg G, Berger JD, Mousavi-Derazmahalleh M, Bayer PE, Edwards D, Singh KB, Cowling WA, Nelson MN (2018) INDEL variation in the regulatory region of the major flowering time gene *LanFTc1* is associated with vernalization response and flowering time in narrow-leafed lupin (*Lupinus angustifolius* L.). Plant Cell Environ. https://doi.org/10.1111/pce.13320

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, DeBoy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc Natl Acad Sci USA 102(39):13950. https://doi.org/10.1073/pnas.0506758102

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796. https://doi.org/10.1038/35048692

The International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345(6194):1251788. https://doi.org/10.1126/science.1251788

Tranchida-Lombardo V, Aiese Cigliano R, Anzar I, Landi S, Palombieri S, Colantuono C, Bostan H, Termolino P, Aversano R, Batelli G, Cammareri M, Carputo D, Chiusano ML, Conicella C, Consiglio F, D'Agostino N, de Palma M, Di Matteo A, Grandillo S, Sanseverino W, Tucci M, Grillo S (2018) Whole-genome re-sequencing of two Italian tomato landraces reveals sequence variations in genes associated with stress tolerance, fruit quality and long shelf-life traits. DNA Res 25(2):149–160. https://doi.org/10.1093/dnares/dsx045

Tseng S-H, Peng S-F, Cheng Y-M (2017) Analysis of B chromosome nondisjunction induced by the *r-X1* deficiency in maize. Chromosome Res 42(2):223. https://doi.org/10.1007/s10577-017-9567-7

Udall JA, Quijada PA, Osborn TC (2005) Detection of chromosomal rearrangements derived from homologous recombination in four mapping populations of *Brassica napus* L. Genetics 169(2):967–979. https://doi.org/10.1534/genetics.104.033209

Uga Y, Sugimoto K, Ogawa S, Rane J, Ishitani M, Hara N, Kitomi Y, Inukai Y, Ono K, Kanno N, Inoue H, Takehisa H, Motoyama R, Nagamura Y, Wu J, Matsumoto T, Takai T, Okuno K, Yano M (2013) Control of root system architecture by *DEEPER ROOTING 1* increases rice yield under drought conditions. Nat Genet 45(9):1097–1102. https://doi.org/10.1038/ng.2725

Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K (2009) The flowering world: a tale of duplications. Trends Plant Sci 14(12):680–688. https://doi.org/10.1016/j.tplants.2009.09.001

Vialette-Guiraud ACM, Adam H, Finet C, Jasinski S, Jouannic S, Scutt CP (2011) Insights from ANA-grade angiosperms into the early evolution of *CUP-SHAPED COTYLEDON* genes. Ann Bot 107(9):1511–1519. https://doi.org/10.1093/aob/mcr024

Villareal RL, Toro E, Mujeeb-Kazi A, Rajaram S (1995) The 1BL/1RS chromosome translocation effect on yield characteristics in a *Triticum aestivum* L. cross. Plant Breed 114(6):497–500. https://doi.org/10.1111/j.1439-0523.1995.tb00843.x

Voss-Fels K, Snowdon RJ (2016) Understanding and utilizing crop genome diversity via high-resolution genotyping. Plant Biotechnol J 14(4):1086–1094. https://doi.org/10.1111/pbi.12456

Wang W, Wang S, Hou C, Xing Y, Cao J, Wu K, Liu C, Zhang D, Zhang L, Zhang Y, Zhou H (2014) Genome-wide detection of copy number variations among diverse horse breeds by array CGH. PLoS ONE 9(1):e86860. https://doi.org/10.1371/journal.pone.0086860

Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, Fang Y, Zeng L, Xu E, Xu J, Ye W, Meng X, Liu R, Chen H, Jing Y, Wang Y, Zhu X, Li J, Qian Q (2015) Copy number variation at the *GL7* locus contributes to grain size diversity in rice. Nat Genet 47:944. https://doi.org/10.1038/ng.3346

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(10):1233–1241. https://doi.org/10.1038/ng.3636

Wulff BBH, Moscou MJ (2014) Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. Front Plant Sci 5:692. https://doi.org/10.3389/fpls.2014.00692

Würschum T, Boeven PHG, Langer SM, Longin CFH, Leiser WL (2015) Multiply to conquer: copy number variations at *Ppd-B1* and *Vrn-A1* facilitate global adaptation in wheat. BMC Genet 16(1):949. https://doi.org/10.1186/s12863-015-0258-0

Würschum T, Longin CFH, Hahn V, Tucker MR, Leiser WL (2016) Copy number variations of CBF genes at the *Fr-A2* locus are essential components of winter hardiness in wheat. Plant J 89(4):764–773. https://doi.org/10.1111/tpj.13424

Xiong Z, Pires JC (2011) Karyotype and identification of all homoeologous chromosomes of allopolyploid *Brassica napus* and its diploid progenitors. Genetics 187(1):37–49. https://doi.org/10.1534/genetics.110.122473

Xiong Z, Gaeta RT, Pires JC (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc Natl Acad Sci USA 108(19):7908–7913. https://doi.org/10.1073/pnas.1014138108

Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ (2006) *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature 442(7103):705–708. https://doi.org/10.1038/nature04920

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, Wang W (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30(1):105–111. https://doi.org/10.1038/nbt.2050

Yao W, Li G, Zhao H, Wang G, Lian X, Xie W (2015) Exploring the rice dispensable genome using a metagenome-like assembly strategy. Genome Biol 16:187. https://doi.org/10.1186/s13059-015-0757-3

Yang S, Li J, Zhang X, Zhang Q, Huang J, Chen J-Q, Hartl DL, Tian D (2013) Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. Proc Natl Acad Sci USA 110(46):18572–18577. https://doi.org/10.1073/pnas.1318211110

Ye K, Schulz MH, Long Q, Apweiler R, Ning Z (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25(21):2865–2871. https://doi.org/10.1093/bioinformatics/btp394

Yu P, Wang C, Xu Q, Feng Y, Yuan X, Yu H, Wang Y, Tang S, Wei X (2011) Detection of copy number variations in rice using array-based comparative genomic hybridization. BMC Genomics 12(1):372. https://doi.org/10.1186/1471-2164-12-372

Yu P, Wang C-H, Xu Q, Feng Y, Yuan X-P, Yu H-Y, Wang Y-P, Tang S-X, Wei X-H (2013) Genome-wide copy number variations in *Oryza sativa* L. BMC Genomics 14(1):649. https://doi.org/10.1186/1471-2164-14-649

Zhang Z, Mao L, Chen H, Bu F, Li G, Sun J, Li S, Sun H, Jiao C, Blakely R, Pan J, Cai R, Luo R, Van de Peer Y, Jacobsen E, Fei Z, Huang S (2015) Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. Plant Cell 27(6):1595–1604. https://doi.org/10.1105/tpc.114.135848

Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Zhang L, Huang T, Wang Y, Fan D, Zhao Y, Wang Z, Zhou C, Chen J, Zhu C, Li W, Weng Q, Xu Q, Wang Z-X, Wei X, Han B, Huang X (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat Genet 50(2):278–284. https://doi.org/10.1038/s41588-018-0041-z

Zheng L-Y, Guo X-S, He B, Sun L-J, Peng Y, Dong S-S, Liu T-F, Jiang S, Ramachandran S, Liu C-M, Jing H-C (2011) Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). Genome Biol 12(11):R114. https://doi.org/10.1186/gb-2011-12-11-r114

Zhou P, Silverstein KAT, Ramaraj T, Guhlin J, Denny R, Liu J, Farmer AD, Steele KP, Stupar RM, Miller JR, Tiffin P, Mudge J, Young ND (2017) Exploring structural variation and gene family architecture with *de novo* assemblies of 15 *Medicago* genomes. BMC Genomics 18(1):261. https://doi.org/10.1186/s12864-017-3654-1

Żmieńko A, Samelak A, Kozłowski P, Figlerowicz M (2014) Copy number polymorphism in plant genomes. Theor Appl Genet 127(1):1–18. https://doi.org/10.1007/s00122-013-2177-7

Zou J, Hu D, Mason AS, Shen X, Wang X, Wang N, Grandke F, Wang M, Chang S, Snowdon RJ, Meng J (2018) Genetic changes in a novel breeding population of *Brassica napus* synthesized from hundreds of crosses between *B. rapa* and *B. carinata*. Plant Biotechnol J 16(2):507–519. https://doi.org/10.1111/pbi.12791

3 Finding invisible quantitative trait loci with missing data

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

Plant Biotechnology Journal (2018), vol 16:2102-2112

doi: 10.1111/pbi.12942

Plant Biotechnology Journal

Plant Biotechnology Journal (2018) **16**, pp. 2102–2112 doi: 10.1111/pbi.12942

Finding invisible quantitative trait loci with missing data

Iulian Gabur[1,#], Harmeet S. Chawla[1,#], Xiwei Liu[1], Vinod Kumar[2], Sébastien Faure[3], Andreas von Tiedemann[4], Christophe Jestin[5], Emmanuelle Dryzska[6], Susann Volkmann[7], Frank Breuer[7], Régine Delourme[2], Rod Snowdon[1] and Christian Obermeier[1,*]

[1]*Department of Plant Breeding, Justus Liebig University, Giessen, Germany*
[2]*IGEPP, INRA, AGROCAMPUS OUEST, Univ Rennes, Le Rheu, France*
[3]*Biogemma, Mondonville, France*
[4]*Section of General Plant Pathology and Crop Protection, Georg August University, Göttingen, Germany*
[5]*Terres Inovia, Thiverval-Grignon, France*
[6]*Syngenta Seeds France SAS, Saint-Sauveur, France*
[7]*KWS SAAT SE, Einbeck, Germany*

Received 23 February 2018;
revised 26 April 2018;
accepted 28 April 2018.
**Correspondence* (Tel +49 641 9937426; fax +49 641 9937429; email christian.obermeier@agrar.uni-giessen.de)
[#]Both contributed equally to this article.

Keywords: *Brassica napus*, quantitative resistance, presence–absence variation, single nucleotide absence polymorphism, SNaP.

Summary

Evolutionary processes during plant polyploidization and speciation have led to extensive presence–absence variation (PAV) in crop genomes, and there is increasing evidence that PAV associates with important traits. Today, high-resolution genetic analysis in major crops frequently implements simple, cost-effective, high-throughput genotyping from single nucleotide polymorphism (SNP) hybridization arrays; however, these are normally not designed to distinguish PAV from failed SNP calls caused by hybridization artefacts. Here, we describe a strategy to recover valuable information from single nucleotide absence polymorphisms (SNaPs) by population-based quality filtering of SNP hybridization data to distinguish patterns associated with genuine deletions from those caused by technical failures. We reveal that including SNaPs in genetic analyses elucidate segregation of small to large-scale structural variants in nested association mapping populations of oilseed rape (*Brassica napus*), a recent polyploid crop with widespread structural variation. Including SNaP markers in genomewide association studies identified numerous quantitative trait loci, invisible using SNP markers alone, for resistance to two major fungal diseases of oilseed rape, *Sclerotinia* stem rot and blackleg disease. Our results indicate that PAV has a strong influence on quantitative disease resistance in *B. napus* and that SNaP analysis using cost-effective SNP array data can provide extensive added value from 'missing data'. This strategy might also be applicable for improving the precision of genetic mapping in many important crop species.

Introduction

Structural variation in genomes of humans, animals and plants is an essential class of genetic polymorphism that is today commonly used for genomic analysis (Dolatabadian *et al.*, 2017). Common forms of short and longer range structural variation include small insertions and deletions (InDels), copy number variation (CNV) and presence–absence variation (PAV). Traditionally, InDels have been defined as short presence–absence polymorphisms spanning from 1 to 50 bp, whereas CNV results from gain or losses of larger DNA segments ranging from a few nucleotides to several kb of DNA in the size range of genes (reviewed in Saxena *et al.*, 2014; Żmieńko *et al.*, 2014). An extreme form of CNV is characterized by deletions of DNA sequences in one or more individuals of a population, which is also termed PAV (Saxena *et al.*, 2014).

In recent years genetic diversity for structural genome variation in the form of InDels, CNV and PAV have been investigated widely in humans (Iafrate *et al.*, 2004), bacteria (Arrach *et al.*, 2008), animals (Graubert *et al.*, 2007; Guryev *et al.*, 2008; Snijders *et al.*, 2005; Wilson *et al.*, 2006) and plants (Batley *et al.*, 2003; Hurgobin *et al.*, 2017; Qian *et al.*, 2016; Schiessl *et al.*, 2017; Springer *et al.*, 2009; Stein *et al.*, 2017). In crops, PAV has been attributed to evolutionary processes associated with natural selection and breeding (Hurgobin *et al.*, 2017; Żmieńko *et al.*, 2014). Completion of reference genomes for most major crops and rapidly decreasing prices for genotyping-by-sequencing (GBS) have facilitated identification of PAV on a whole-genome level. On the other hand, detection of PAV in GBS and skim-sequencing datasets can be complicated by difficulties in distinguishing genuine deletions from regions with insufficient sequence coverage, along with bioinformatic challenges associated with haplotype imputation.

Gene CNV has been implicated in the control of many agronomic traits in different crop species, for example flowering time and plant height in *Brassica napus*, oilseed rape (Schiessl *et al.*, 2017), ancestral evolution events and domestication in maize (Springer *et al.*, 2009), or vernalization and winter hardiness in wheat (Würschum *et al.*, 2015, 2017). CNV is also involved in resistance against pathogens, with nucleotide-binding leucine-rich repeat (NB-LRR), thaumatin-like protein (TLP) and receptor-like kinase (RLK) genes being commonly involved in local gene duplications leading to variable copy number (Saxena *et al.*, 2014).

Brassica napus (oilseed rape, canola, kale, rutabaga/swede) is a recent allopolyploid crop species that arose from interspecific hybridization between two diploid progenitors, *B. oleracea* and *B. rapa* (Snowdon, 2007), and rapidly acquired vast ecogeographic and agronomic diversity that ultimately led to its establishment as a globally important crop. Post polyploidization

homoeologous exchanges during meiosis between chromosomes of the A and C subgenomes have been identified as a major driver of genome diversity in *B. napus* (Chalhoub *et al.*, 2014; Hurgobin *et al.*, 2017; Samans *et al.*, 2017). Homoeologous rearrangements, including gene conversions (Chalhoub *et al.*, 2014), CNV (Schiessl *et al.*, 2017), PAV and segmental deletions (Hurgobin *et al.*, 2017; Samans *et al.*, 2017), underlie widespread structural and functional genome variation in both natural *B. napus* forms and in *de novo*, synthetic *B. napus* accessions. All these recent studies demonstrated that in *B. napus* gene terms associated with plant resistance and stress responses are strongly enriched among genes affected by deletions due to homoeologous exchanges, suggesting that PAV might be an important mechanism in crop disease resistance.

Major diseases of oilseed rape are caused by fungal pathogens transmitted by airborne or soil-borne spores. In comparison with cereal crops, where gene-for-gene resistance interactions with important fungal virulence genes (e.g. for rust and mildew diseases) play an important role in breeding for crop protection, the most effective adult-plant resistances to fungal diseases in oilseed rape are quantitative in nature. Vastly increased production, shorter crop rotations and global warming have led to particularly strong disease pressure in major growing areas for the fungal pathogens *Verticillium longisporum*, *Sclerotinia sclerotiorum* and *Leptosphaeria maculans* (Barbetti *et al.*, 2012; Siebold and von Tiedemann, 2012). Among these, blackleg disease (Phoma stem canker) caused by *L. maculans* (anamorph *Phoma lingam*) is the most economically important disease of oilseed rape in Europe, North America and Australia, while Sclerotinia stem rot caused by the necrotrophic pathogen *S. sclerotiorum* causes substantial yield losses in all major growing areas throughout Europe, Australia, North America and China (Delourme *et al.*, 2011). Major genes for resistance to blackleg disease (reviewed by Delourme *et al.*, 2011) are effective at seedling stage but more durable when used in association with quantitative, adult-plant resistance. No major-gene resistance to *S. sclerotiorum* is available in *B. napus* so that identification and combination of quantitative genetic resistance factors are essential for breeding.

Resources for high-throughput genomics are today broadly implemented for Brassica crops. Reference genome sequences for *B. rapa* (The *Brassica rapa* Genome Sequencing Project Consortium, 2011); *B. oleracea* (Liu *et al.*, 2014) and *B. napus* (Chalhoub *et al.*, 2014) have been supplemented by large-scale resequencing (Schmutzer *et al.*, 2015) or transcriptome datasets (He *et al.*, 2015). As in many other major crops, one of the most broadly used tools for genetic analysis in *Brassica* crops is a community-designed, high-density single nucleotide (SNP) genotyping array (Clarke *et al.*, 2016; Mason *et al.*, 2017). It has been extensively applied for high-density genetic mapping and QTL analysis (e.g. Liu *et al.*, 2013; Luo *et al.*, 2017; Wang *et al.*, 2015), genomewide association studies for a wide range of traits (e.g. Hatzig *et al.*, 2015; Li *et al.*, 2014, 2016; Schiessl *et al.*, 2015; Sun *et al.*, 2016; Wan *et al.*, 2017; Xu *et al.*, 2016) and genomic selection (Jan *et al.*, 2016; Zou *et al.*, 2016).

One objective of this study was to improve the resolution of QTL mapping for fungal disease resistance in *B. napus* using high-density SNP array data. We also elucidate the relevance and role of small-scale and large-scale PAV in the *B. napus* genome in relation to disease resistance. Data filtering approaches were designed to identify presence–absence variants in high-density SNP array data and include these 'missing' data as an additional dimension in genomewide association studies.

Results

From failed SNP calls to single nucleotide absence polymorphism markers

Failed SNP calls are commonly observed in genotyping experiments applying chip hybridization technologies. Thus, markers which show excessive frequencies of failed calls are often removed from genotyping matrices for downstream analyses (Mason *et al.*, 2017). Although failed SNP calls are expected to be due to technical artefacts, in case of genuine deletions they may also represent biologically useful information due to potential association with gene PAV. We distinguished failed SNP calls representing single nucleotide absence polymorphisms (SNaP) from random technical failures in raw SNP chip data by filtering for segregation patterns in a multiparental homozygous mapping population. In an inbred or homozygous biparental mapping population, a segregation allele frequency of 50% is expected for a SNaP (e.g. A:failed, or C:failed in Figure 1a for SNP30 and SNP31), because the presence-allele will be amplified only from one parental line and is absent in the other parental line. In our multiparental population, five diverse parents were crossed with one common parent so that a SNaP has an expected frequency of failed calls within each subpopulation of 50%, whereas the expected frequency of failed calls across the total population will be 10% if the absence derives from (only) one of the diverse parents. A threshold of 10% is usually used in the standard filtering approach as failed SNP call frequency to exclude markers from further analyses, thus eliminating potential SNaP markers.

According to the standard filtering approach as described in experimental procedures, 18 068 polymorphic SNP markers (63.5% of a total of 28 073 anchored markers) were selected for SNP-based genomewide association studies (GWAS). In a refined filtering approach, a three-step method was used to select SNaP markers that could indicate structural presence–absence polymorphisms.

Many of the 36.5% SNP markers removed during this standard filtering procedure showed a low frequency of failed calls in the 5 nested association mapping (NAM) subpopulations. However, separate investigation of each subpopulation revealed SNaP frequencies that were frequently close to the expected segregation pattern of a biparental homozygous population (50%). Figure 1b shows an example of different observed SNP marker patterns within and across subpopulations (SP1 = subpopulation 1, SP10 = subpopulation 10), which would be excluded from analysis using common filtering approaches. For example, markers SNP30 and SNP31 show a dominant A/failed or C/failed allele pattern with ~50% frequency across all subfamilies, because the absence derives from the recurrent parent, whereas SNP2500 and SNP2501 show ~50% A/failed or C/failed in only one subfamily because the absence derives from only one of the 5 diverse parents. All 10 005 excluded SNPs were reanalysed for specific segregation patterns within the five subpopulations, using an allele frequency threshold of 15%–85% between failed/present calls and considering segregation distortion. SNaP markers identified by segregation patterns were validated by physical positioning in the reference genome and identification of neighbouring SNP loci which showed corresponding SNaP patterns in the same genotypes. Using this approach, a number

of 3627 SNaPs were identified, recovered and included in subsequent GWAS.

The mean genomic distance between polymorphic markers improved from ~37.6 kb when using only the 18 068 filtered SNPs to ~32.2 kb when including the genomewide SNaP markers (Table S1), and chromosome regions with low SNP density were found to be more evenly covered when SNaPs were added. Chromosomes C01 and C02, two of the *B. napus* chromosomes most substantially affected by structural rearrangements (e.g. Xiong and Pires, 2011), showed the strongest representation of SNaPs, with 1021 on C01 and 629 on C02 and an increase in haplotype blocks of 59% and 54%, respectively (Table S1).

Genomic regions displaying SNaPs show segmental deletions in the parental lines

SNaPs in a number of selected regions were validated by comparing SNP genotyping data with Illumina short-read genomic sequence coverage from the six parental lines in the corresponding chromosome segments (genomic resequencing data from Schmutzer *et al.*, 2015). The size of putatively deleted regions was manually verified and compared to genome positions of flanking sequences of the SNP assays corresponding to the respective SNaPs. Particular focus was placed on identification of small and medium scale structural variation across all chromosomes (e.g. consecutive SNaPs implying potential gene PAV). Figure 2 shows examples for deletions on chromosomes A03 and A07, respectively. Four consecutive SNaPs were detected within a 5-kb interval on chromosome A03 from position 10 075 405 to 10 080 123 bp. Physical anchoring of these markers to the reference genome showed consecutive failed calls, and the deletion in parental lines H165, RS13/6, CRY1 and MOY4 is confirmed by the read mapping data from whole-genome sequencing (Figure 2a). Similarly, two consecutive SNaPs on chromosome A07 exhibit failed calls in the parental lines CRY1 and MOY4, consistent with sequencing reads aligned to the reference genome (Figure 2b). From 3627 SNaPs selected using the customized filtering approach on all chromosomes, 3405 (89%) showed the corresponding SNaP between the respective parental lines. Furthermore, from 100 randomly selected SNaP markers spanning all chromosomes, we confirmed 95% putative deletions using genomewide resequencing data from the six parents.

Another 35-kb deletion was detected by SNaPs in the NAM panel which was localized at position 21 935–21 965 kb on chromosome A03 in the recurrent parental line DH5ON. The SNaPs present within this region were associated with disease resistance (see below). The deletion was further validated by PCR using specific primer pairs targeting the corresponding chromosome region. As expected, DH5ON showed no amplification of specific primer pairs for four consecutive regions (containing six genes) in the estimated deleted interval, whereas the expected PCR products that indicate presence were visible in the other parental lines and positive controls (Figure 3). Based on the SNP genotyping data, this deletion was carried by 128 offspring lines of the 200 investigated NAM lines (64% frequency). A further 12-kb deletion detected by SNaPs associated with blackleg resistance on chromosome C04 was also validated using the same approach and including Sanger sequencing of PCR products (Figure S1). The results confirmed deletion of three genes in parent MOY4 and in the respective

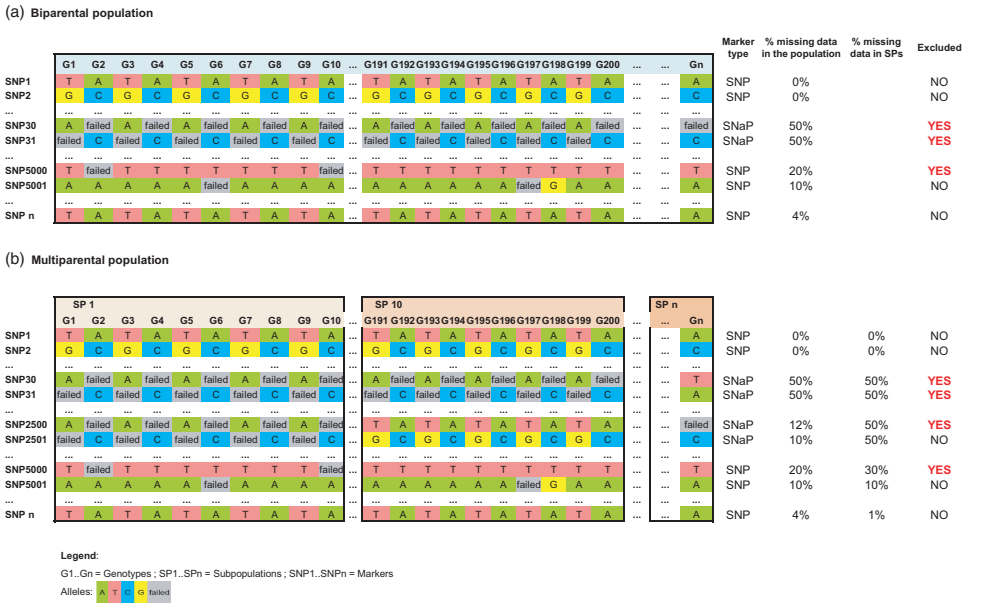


Figure 1 Schematic representation of allele segregation patterns and frequencies within and across subpopulations for different single nucleotide polymorphism (SNP) types observed in homozygous biparental (a) or multiparental mapping populations (b) and SNP probes excluded from analysis applying standard filtering procedures.

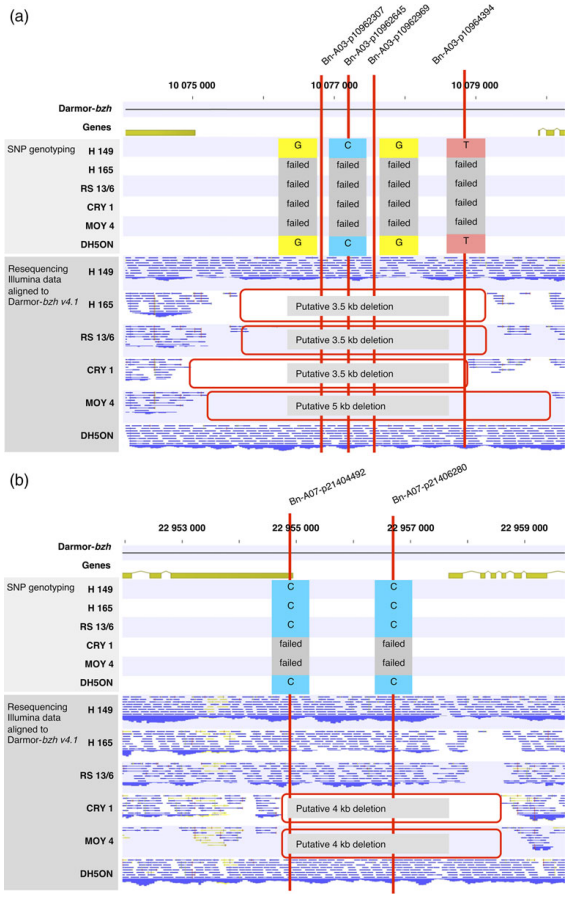


Figure 2 Physical anchoring of single nucleotide absence polymorphism (SNaP) markers (red lines) and Illumina resequencing data for six nested association mapping (NAM) parental lines to Darmor-*bzh* and comparison with SNP segregation patterns in 200 NAM lines for (a) four consecutive SNaP markers on chromosome A03, Bn-A03-p10962307 (10 076 697 bp), Bn-A03-p10962645 (10 077 034 bp), Bn-A03-p10962969 (10 077 358 bp), Bn-A03-p10964394 (10 078 777 bp) and two consecutive SNaP markers on chromosome A07, Bn-A07-p21404492 (22 954 748 bp) and Bn-A07-p21406280 (22 956 683 bp).

subfamily (3.5% frequency in all 200 lines, 17.5% frequency in the subfamily).

Deletion sizes and segregation in NAM subfamilies

Physical location of SNP probes corresponding to SNaP markers confirmed small, medium and large deletions up to chromosome-size in the five segregating subpopulations (Figure 4). On chromosome C02, a large range deletion was detected by consecutive SNaP calls in the resynthesized *B. napus* parent H165, consistent with whole-genome resequencing data and read coverage analysis. This deletion segregates in the DH5ON x H165 subpopulation in our study. Chromosome C01 also shows large deletions in the parents H149 and MOY4. In general, the natural *B. napus* parent DH5ON exhibits the lowest frequency of genome restructuring and SNaPs, but these segregate in all subfamilies.

SNaP markers reveal hidden QTL for disease resistance

Genomewide SNP-trait association analyses were first conducted using the 18 076 quality-filtered SNP markers that were polymorphic in the NAM subpopulations. In order to reduce the rate of false-positive marker–trait associations, a mixed linear model was applied that controls population substructure using the first two principle components and the kinship matrix. Phenotype data from blackleg disease screening in field trials in France revealed 12 significant SNP-trait associations (higher than the arbitrary selected significance threshold of $-\log10_{(P\text{-value})} \geq 3$) with p-values between 7.88E-05 and 8.53E-04 (Figure S2a). Blackleg resistance from the German field screening identified 52 significant SNP-trait associations with p-values higher than the arbitrary selected significance threshold ($-\log10_{(P\text{-value})} \geq 3$). After correction for false discovery rate (FDR, see experimental procedures), however, no SNPs remained in either trial with confirmed associations to blackleg disease resistance.

Repeating the GWAS including 3627 SNaPs together with the 18 076 SNPs identified a total of 38 resistance-associated markers on the French trial data (a 3.2-fold increase, Figure S2b, Table 1). In contrast to the SNP analysis, six associations involving SNaP markers on chromosome C04 were also

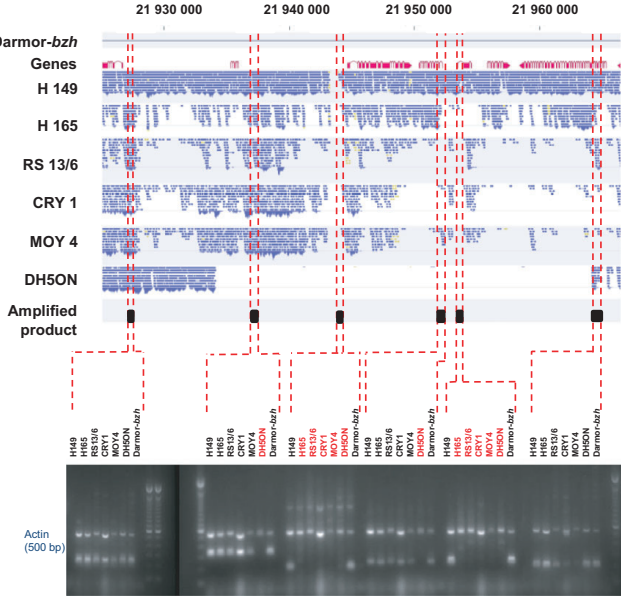


Figure 3 Alignment of Illumina resequencing data for six nested association mapping parents to the Darmor-*bzh* reference and comparison with PCR amplification results for six genes contained within a 30-kb deletion (position 21 934 109 to 21 964 245 bp on chromosome A03) in the common parent DH5ON (Actin gene and genotype Darmor-*bzh* are used as controls, and genotypes with no amplified products are shown in red font).

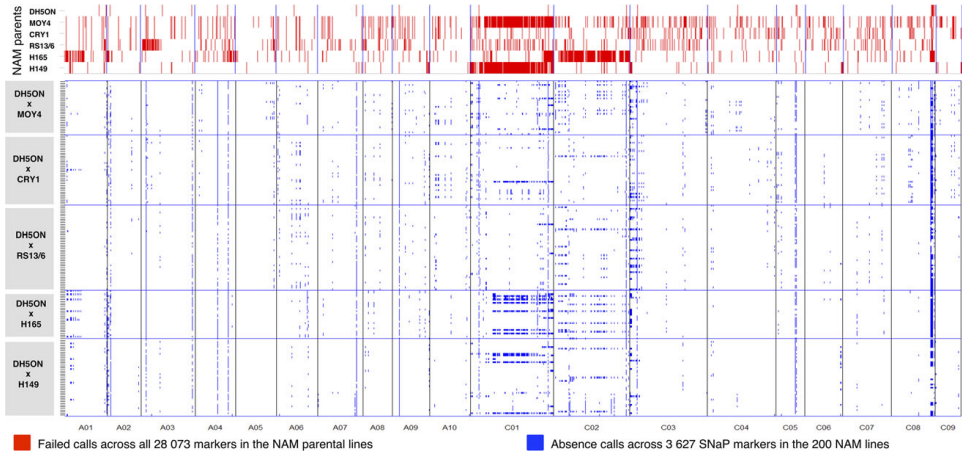


Figure 4 Genomewide deletion patterns visualized by single nucleotide polymorphism marker distribution with failed calls (red) in six nested association mapping (NAM) parental lines compared to single nucleotide absence polymorphism (SNaP) marker segregation patterns in the 200 *B. napus* nested association mapping (BnNAM) lines and subpopulations (blue).

significant at FDR ≤ 0.1. These identified two new QTL located in strongly conserved LD blocks from 0–40 and 200–400 kb at the proximal end of C04 (Figure S3). GWAS for blackleg resistance screening data from the German field trial revealed 115 significant associations (a 2.2-fold increase compared to the SNP markers alone), of which 41 were also significant at FDR ≤ 0.1 (Table S3).

Similarly, using only SNP markers obtained with the commonly used filtering criteria, we identified 47 significant SNP-trait associations for *Sclerotinia* stem rot resistance with p-values over the arbitrary significance threshold. Adding the SNaP markers, a total of 65 markers (Table S2) associated with *Sclerotinia* stem rot resistance using the arbitrary selected significance threshold (a 1.4-fold increase). SNaP markers revealing deletions inherited

from the common elite parent DH5ON were associated with resistance to *Sclerotinia* stem rot resistance for the QTL present on A03 at position 21 Mb. The absence of these alleles in DH5ON was reconfirmed by SNP analysis in all parents and the BnNAM population by resequencing of parents and by PCR analyses (see details above, Figure 3). In contrast to the analysis with SNPs alone, a number of SNaP markers showed significant associations (exceeding the arbitrary selected threshold) to resistances against both pathogens.

Significant SNP-trait associations for blackleg disease and *Sclerotinia* stem rot resistance were compared with previous studies performed in other *B. napus* mapping panels (Table S4).

Deletions associate with both susceptibility and resistance

We observed opposing scenarios of PAV and its association with disease resistance. For example, an ~25 kb deletion on chromosome A01 in the parental line H165 associated with *Sclerotinia* stem rot susceptibility in the subpopulation derived from this parent (Figure 5a). In contrast, an ~30 kb deletion on chromosome A03 associated with resistance to *Sclerotinia* stem rot in all five subpopulations (Figure 5b). In general, for both diseases, absence alleles were more frequently associated with susceptibility. For blackleg, all SNaP alleles associated with susceptibility were absence alleles, while in two cases out of 30 marker–trait associations for Sclerotinia resistance a SNaP absence allele was associated with resistance.

Discussion

SNP hybridization arrays are nowadays commonly used in genetic plant analyses (reviewed by Voss-Fels and Snowdon, 2016). Here, we demonstrate that standard data quality filtering approaches can remove large numbers of potentially useful marker information that can mask QTL caused by PAV. We also show that these SNaP markers are detecting deletions ranging from small (1 bp) to large (chromosome-wide) size in segregating *B. napus* populations. This has been described before using whole-genome sequencing data for single *B. napus* genotypes (Chalhoub *et al.*, 2014; Hurgobin *et al.*, 2017; Samans *et al.*, 2017). In the allopolyploid genome of *B. napus*, high homoeology between the diploid progenitor genomes (A and C subgenomes) led to extensive structural genome variation a widespread phenomenon in the pangenomes of many crop plants (for a review see Dolatabadian *et al.*, 2017). Thus, the refined SNP filtering approach might be applicable to many agronomical important diploid and polyploidy crop genomes such as maize, sorghum, cotton wheat and others.

Using standard filtering quality procedures to select 60K Brassica Illumina SNP array markers facilitated detection of a number of regions corresponding to previously identified loci conferring disease resistance. For blackleg, studies using biparental crosses and association mapping on diverse germplasm sets have identified a small number of major loci with monogenic inheritance (Delourme *et al.*, 2014; Larkan *et al.*, 2014; Raman *et al.*, 2012a,b) as well as multiple quantitative trait loci (QTL) involved in disease resistance (e.g. Delourme *et al.*, 2006; Fomeju *et al.*, 2015; Fopa Fomeju *et al.*, 2014; Jestin *et al.*, 2011; Kaur *et al.*, 2009; Larkan *et al.*, 2016). Comparison of the location of these QTL with the QTL discovered in this study is difficult as different marker types were used for genetic mapping by most authors. From the 50 QTL discovered in this study using SNP and SNaP markers, only 3 coincide with previously mapped major *Rlm* and quantitative resistance loci using SNP probe and SSR sequence anchoring to Darmor-*bzh* (Larkan *et al.*, 2014, 2016; Raman *et al.*, 2012a,b, 2016). Partial resistance to *Sclerotinia* stem rot has been found in some Chinese semi-winter *B. napus* genotypes (Zhao *et al.*, 2006) and spring-type oilseed rape lines from China and Australia (Li *et al.*, 2009). QTL mapping in various Chinese DH populations has identified many loci responsible for partial resistance to *Sclerotinia* stem rot in RV298 × P1804 (Zhao *et al.*, 2006), Huazhuang 5 × J7005 (Wu *et al.*, 2013), Express × SWU 7 (Wei *et al.*, 2014) and in natural populations by Wei *et al.* (2016) and Wu *et al.* (2016). From 37 QTL discovered in this study using SNP and SNaP markers, only seven were reported in the literature before. After adding SNaP markers, we observed a 1.6 -to 3.5-fold increase of QTL regions associated with blackleg and *Sclerotinia* stem rot resistance (in total 57 new QTL were found).

Most of these previous QTL analyses included other types of markers than SNP chip markers, including anonymous amplified fragment length polymorphism or other PCR marker systems (Delourme *et al.*, 2008b; Huang *et al.*, 2016) that also partially address presence–absence polymorphism. Combining different marker types or marker systems has been shown to increase the power of genetic mapping because different markers access different genome features (García-Lor *et al.*, 2012; Larkan *et al.*, 2014; Raman *et al.*, 2012a,b). The advantages of SNP array

Table 1 Summary of significant marker–trait associations and QTL regions by applying GWAS for SNP markers only and for SNP plus SNaP markers

	SNP-trait associations				QTL regions				
	SNP	SNP and SNaP	Overlapping	New	SNP	SNP and SNaP	Overlapping	New	Fold increase
Blackleg France field	12	38	12	26	7	22	6	18	3.14
Blackleg German field	52	120	55	65	14	28	13	15	2.00
SSR_AUDPC	15	16	15	1	5	8	4	4	1.60
SSR_7dai	16	17	15	2	7	11	5	6	1.57
SSR_14dai	9	12	9	3	4	11	2	9	2.75
SSR_21dai	7	20	7	13	2	7	2	5	3.50

AUDPC, area under the disease progress curve; SSR, *Sclerotinia* stem rot; dai, dai after inoculation; QTL, quantitative trait loci; SNP, single nucleotide polymorphism; SNaP, single nucleotide absence polymorphism; GWAS, genomewide association studies.

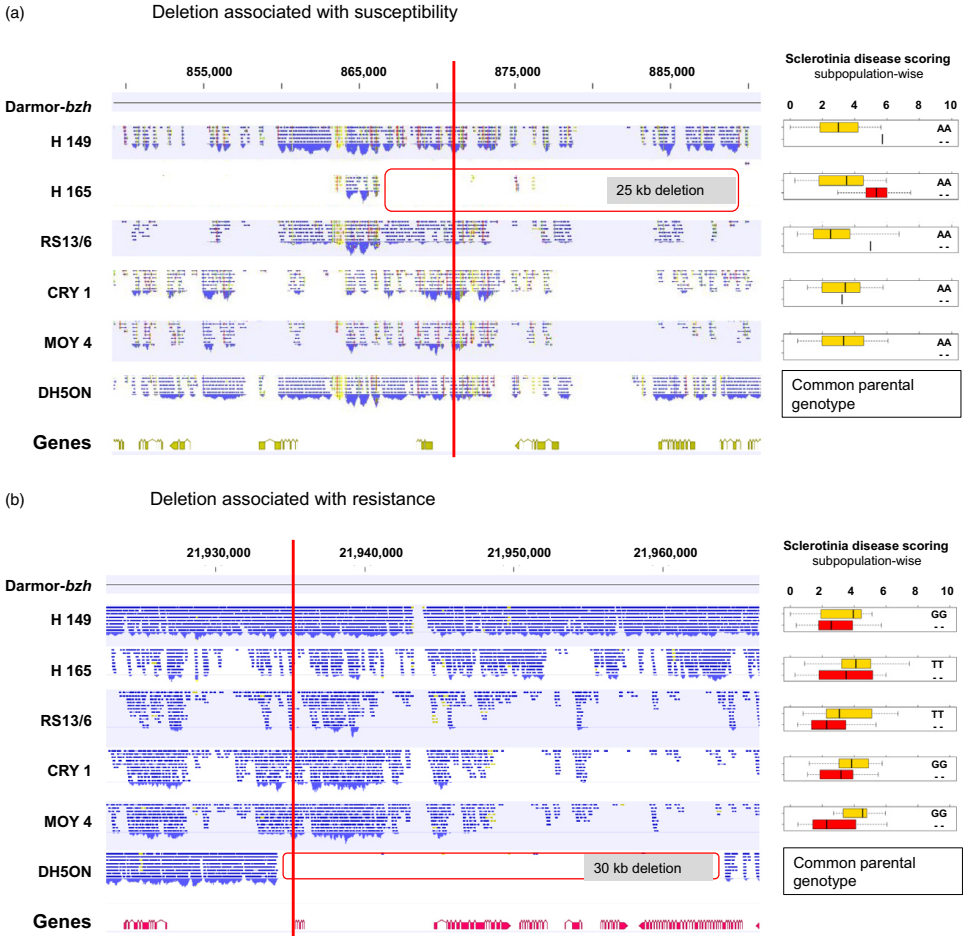


Figure 5 Alignment of resequencing data in six parents of the nested association mapping population showing deletions in gene range size (left, in red) compared to single nucleotide absence polymorphism marker–trait association for Sclerotinia stem rot disease resistance in five subpopulations (right). Effect of an absence allele on phenotype in the respective subfamilies showing an association with (a) susceptibility on chromosome A01 at position 873 225 bp and an association with (b) resistance on chromosome A03 at position 21 934 764 bp.

screening, which provide robust, low-cost, high-resolution data for genetic mapping and GWAS, may therefore be further boosted by addition of methods to assess PAV via SNaP scoring.

Including SNaP markers in GWAS analysis for quantitative resistance to blackleg disease and *Sclerotinia* stem rot markedly increased significant marker–trait associations. The frequent localization of new QTL in regions affected by PAV confirms the hypothesis that PAV has particular relevance for disease resistance (Hurgobin *et al.*, 2017). Offspring of resynthesized *B. napus* with high rates of presence–absence and other structural variations may therefore have an increased potential for use in resistance breeding of *B. napus*.

It is known that PAV and other structural variation in plants affect stress response genes, particularly genes involved in disease resistance (McHale *et al.*, 2012; Shen *et al.*, 2006; Tan *et al.*, 2012). Examples for mutations that lead to loss of functions are numerous and have been described for many traits including disease resistance (Dolatabadian *et al.*, 2017). In case of R genes, plant disease resistance is determined by complementary pairs of resistance (R) genes from the plant and avirulence (Avr) genes from the invading pathogen. In this gene-for-gene interaction, an avirulence protein binds to the corresponding resistance protein triggering plant defence responses. This interaction can be disrupted by a mutation in the plant R gene or in the fungal Avr gene (Bonas and Lahaye, 2002). Examples for deletions leading to a loss of function are less frequent, but have also been described. For example, presence–absence polymorphisms associated with race-specific R genes are a common phenomenon estimated to affect

about 20% of R genes in Arabidopsis and rice (Shen *et al.*, 2006). Most R genes act in a dominant manner and null alleles are consequently associated with susceptibility. When grown in the absence of targeting pathogens, plants carrying specific R genes were found to be up to 10% less fit than plants lacking the R gene (MacQueen and Bergelson, 2016; Tian *et al.*, 2003). Thus, deletion of R genes can be beneficial and may be a common feature of crop resistance. Association of deletions resulting in a loss of function with susceptibility is consistent with our finding that in the majority of cases the null alleles of SNaP markers were associated with *Sclerotinia* stem rot and blackleg disease susceptibility, not with resistance. However, in 6% of cases, we found a homozygous null allele of a SNaP marker associated with resistance against *Sclerotinia* stem rot. This may indicate that a plant factor allowing the pathogen to invade more efficiently is deleted or mutated. For example, Uppalapati *et al.* (2012) described a mutation of the *irg1* gene (*inhibitor of rust germ tube differentiation 1*), which is involved in wax accumulation on the leaf surface. A homozygous mutation was found to hinder the germ tubes of the Asian soybean rust and two other fungal pathogens, preventing them from developing successfully on the leaves of *Medicago truncatula*. Association of mutations or deletions with resistance in natural populations is a rare phenomenon, as loss of function is typically recessive and in heterozygous genotypes resistance can be masked by the dominant allele or by a dosage effect.

The detection of resistance-associated deletions and a number of new genetic loci for blackleg and *Sclerotinia* stem rot resistance in this study demonstrate the usefulness of using missing data to map invisible QTL. Analyses of genes in deleted segments associated with resistance QTL is a promising new approach to deciphering the genetic basis of quantitative resistances in oilseed rape and other crop species. The described strategy for genetic mapping using SNaP markers will also be useful for dissection of major agronomical traits in molecular plant breeding of polyploid crops.

Experimental procedures

Plant material

A subset of a *B. napus* nested association mapping (BnNAM) population was used in this study. The BnNAM population consists of 50 genetically diverse winter *B. napus* accessions (20 exotic *B. napus*, 30 resynthesized *B. napus*) crossed with an elite doubled haploid winter-type line (DH5ON). Each of the 50 subpopulations is composed of ≥50 doubled haploid lines per cross (where both parents are natural *B. napus*) or ≥50 single-backcross recombinant inbred lines (BC1-RILs) for crosses with one resynthesized *B. napus* parent (Snowdon *et al.*, 2015). The present study used five BnNAM subpopulations with a total of 200 BC1-RILs (Table 2) derived from synthetic *B. napus* founders carrying multiple quantitative resistances.

Phenotypic analysis of traits

Blackleg resistance testing of the 200 BnNAM accessions was conducted in a field screening by Syngenta (Toulouse, France) and KWS SAAT SE (Einbeck, Germany) in 2015/2016. The area of necrosis at the plants base stem was evaluated for 30 plants per genotype using a 1–6 scale at crop maturity in late June. This procedure is also known as the G2 index, where a score of 1 corresponds to complete absence of affected tissue, while a score of 6 corresponds to 100% area affected, a broken stem or a dead plant (Delourme *et al.*, 2008a,b; Huang *et al.*, 2016). *Sclerotinia* stem rot resistance tests were conducted in a field screening at KWS SAAT SE, Einbeck, Germany, in 2015/2016. Resistance was assessed using a toothpick stem inoculation method similar to the method described by Zhao and Meng (2003). Plants in plots were inoculated after flowering with toothpicks that were previously soaked with potato dextrose broth and overgrown with *S. sclerotiorum* mycelia. The toothpicks were inserted into the centre of the main stem and the lengths of necrotic surface were measured at 7, 14 and 21 days after inoculation (dai). Approximately 25 plants were scored for each of the 200 tested BnNAM lines. Using the lesion lengths recorded at the three dates, an area under the disease progress curve (AUDPC) was calculated according to Obermeier *et al.* (2013). Mean values for each of the three time points (7, 14, 21 dai) and the AUDPC values were used for GWAS.

SNP genotyping and quality control

The entire BnNAM panel was genotyped with the 60K Illumina Infinium Brassica SNP array containing 52 158 SNP probes. Using the Darmor-*bzh* reference v4.1 (Chalhoub *et al.*, 2014), we anchored 28 073 SNP marker using BLASTN as described by Qian *et al.* (2014). Initially, all markers that exhibited a minor allele frequency (MAF) <0.05 and a failed call frequency >90% were removed from the SNP data set. Subsequently, SNP markers that were previously anchored to the Darmor-*bzh* reference v4.1 were included in a customized pipeline, regardless of whether they were polymorphic or monomorphic for expected SNP alleles, to evaluate whether they show segregation patterns consistent with a presence–absence polymorphism

Table 2 Parents, genetic origin and composition of nested association mapping subpopulations used for blackleg and Sclerotinia stem rot resistance evaluation

Parental lines	Type	Accession name	Number of RILs	Mother	Variety/type	Father	Variety/type
PBY033	Synthetic	H149	48	*Brassica oleracea* conv. *capitata* var. *medullosa*	'Cavalier rouge'	*B. rapa* ssp. *chinensis*	Pak Choi
PBY034	Synthetic	H165	28	*B. oleracea* conv. *capitata* var. *sabauda*	Wirsing	*B. rapa* ssp. *chinensis*	Pak Choi
PBY040	Synthetic	RS13/6	53	*B. rapa* ssp. *chinensis*	Pak Choi	*B. oleracea* conv. *botrytis* var. *alboglabra*	Broccoli
PBY050	Synthetic	CRY1	41	*B. rapa* spp. *trilocularis*	Yellow Sarson	*B. cretica*	–
PBY052	Synthetic	MOY4	31	*B. rapa* spp. *trilocularis*	Yellow Sarson	*B. montana*	–
PBY061	Elite	DH5ON	–	*B. napus* ssp. *napus*	Oase	*B. napus* ssp. *napus*	Nugget

(SNaPs), indicating a possible deletion. If a SNP exhibited a presence–absence segregation pattern ratio in at least one of the subfamilies, the presence SNP allele was recoded as AA and the failed SNP call as BB to enable inclusion in the marker matrix for the GWAS.

Genomewide association studies, linkage disequilibrium analysis and haplotype construction

Association analyses were conducted using the R package GenABEL (Aulchenko *et al.*, 2007). A mixed linear model approach that increases detection power (Stich *et al.*, 2008) was adjusted for population stratification by including the kinship matrix and the first two principal components as covariates (Price *et al.*, 2006). For determining significant SNP-trait association, we applied a FDR of ≤0.1 (Benjamini and Hochberg, 1995). Additionally, a significance cut-off value was set at $-\log_{10}(1/n)$, where n represents the number of SNP markers. To reduce the type II error rate, we also captured the SNP-trait associations for disease resistance using an arbitrary threshold of $-\log_{10(P\text{-value})} \geq 3$ as previously done by Hatzig *et al.* (2015) and Raman *et al.* (2016). Whole-genome linkage disequilibrium was calculated using the squared allele frequency correlations (r^2) between pairs of SNPs. Only markers with MAF ≥ 0.05 were included in the analysis. Haplotype patterns were assessed for SNPs and SNaPs that showed significant marker trait association. Haplotype blocks were defined using the confidence interval method described by Gabriel *et al.* (2002) in Haploview version 4.2 (Barrett *et al.*, 2005) and the R package GenABEL (Aulchenko *et al.*, 2007).

PCR validation of PAV

Specific primers were designed for all genes present in QTL intervals that showed significant associations between SNaP markers and traits on chromosomes A03 and C04. Additionally, primers specific for an *Actin* gene copy located on C04 (BnaC04g27010D) were designed and included in a multiplex PCR to ensure that no technical errors occurred during the tests. Additionally, we included for PCR the *B. napus* reference genotype Darmor-*bzh* as a positive control for PCR amplification. PCR primer information, reaction setup and conditions are listed in Table S5.

Comparative sequence analysis of QTL regions from six NAM parents

The publically available *B. napus* Darmor-*bzh* reference genome assembly v. 4.1 (Chalhoub *et al.*, 2014) and the resequencing data sets of the six BnNAM parents with a 12 to 15× coverage of Illumina 100-bp paired-end-sequencing (Schmutzer *et al.*, 2015) were used for comparative analysis of read coverage and PAV in selected QTL regions. Illumina sequence reads of the six parental lines were aligned to the reference genome with CLC Genomics Workbench v.9.0 software (Qiagen Bioinformatics, Aarhus, Denmark). Putative PAVs were visually inspected using CLC Genomics Workbench. To remove putative false-positive aligned short Illumina reads, a minimum threshold of five reads aligned to a physical position in the reference was set, similar to the approach described by Schmutzer *et al.* (2015).

Acknowledgements

We acknowledge Matthias Frisch (Justus Liebig University Giessen) for advice in statistical analyses of GWAS data. Funding was received within the European PLANT-KBBE-IV research funding programme for the French-German consortium GeWiDis (Exploiting Genome Wide Diversity for disease resistance improvement in oilseed rape) in Germany by the federal Ministry of Education and Research (BMBF, grant 031A325) and in France by the L'Agence nationale de la recherché (ANR, grant ANR-13-KBBE-0004-01), respectively. Iulian Gabur acknowledges funding through a scholarship from the German Academic Exchange Service (DAAD). SNP genotyping data were kindly provided by the PRE-BREED-YIELD consortium funded by BMBF (grant no. 0315964).

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

IG, RS and CO designed the concept and wrote the manuscript. IG and CO and analysed the data. IG, CO, XL and HSC performed the PCR analyses. VK, RD, AvT, SF, CJ, ED, SV and FB contributed to phenotype data acquisition. All authors read and approved the manuscript.

References

Arrach, N., Porwollik, S., Cheng, P., Cho, A., Long, F., Choi, S.H. and McClelland, M. (2008) Salmonella serovar identification using PCR-based detection of gene presence and absence. *J. Clin. Microbiol.* **46**, 2581–2589.

Aulchenko, Y.S., Koning, D.J. and Haley, C. (2007) Genome wide rapid association using mixed model and regression: a fast and simple method for genome wide pedigree-based quantitative trait loci association analysis. *Genetics*, **177**, 577–585.

Barbetti, M.J., Banga, S.S. and Salisbury, P.A. (2012) Challenges for crop production and management from pathogen biodiversity and diseases under current and future climate scenarios—case study with oilseed Brassicas. *Field. Crop. Res.* **127**, 225–240.

Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

Batley, J., Barker, G., O'Sullivan, H., Edwards, K.J. and Edwards, D. (2003) Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiol.* **132**, 84–91.

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

Bonas, U. and Lahaye, T. (2002) Plant disease resistance triggered by pathogen-derived molecules: refined models of specific recognition. *Curr. Opin. Microbiol.* **5**, 44–45.

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

Clarke, W.E., Higgins, E., Plieske, J., Wieseke, R., Sidebottom, C., Khedikar, Y., Batley, J. *et al.* (2016) A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theor. Appl. Genet.* **129**, 1887–1899.

Delourme, R., Chèvre, A.M., Brun, H., Rouxel, T., Balesdent, M.H., Dias, J.S., Salisbury, P. *et al.* (2006) Major gene and polygenic resistance to *Leptosphaeria maculans* in oilseed rape (*Brassica napus*). *Eur. J. Plant Pathol.* **114**, 41–52.

Delourme, R., Brun, H., Ermel, M., Lucas, M.O., Vallee, P., Domin, C., Walton, G. *et al.* (2008a) Expression of resistance to *Leptosphaeria maculans* in *Brassica napus* double haploid lines in France and Australia is influenced by location. *Ann. Appl. Biol.* **153**, 259–269.

Delourme, R., Piel, N., Horvais, R., Pouilly, N., Domin, C., Vallée, P., Falentin, C. *et al.* (2008b) Molecular and phenotypic characterization of near isogenic lines at QTL for quantitative resistance to *Leptosphaeria maculans* in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* **117**, 1055–1067.

Delourme, R., Barbetti, M., Snowdon, R., Zhao, J. and Manazanares-Dauleux, M.J. (2011) Genetics and genomics of disease resistance. In *Genetics, Genomics and Breeding of Oilseed Brassicas* (Edwards, D., Batley, J., Parkin, I.A.P. and Kole, C., eds), pp. 276–303. Boca Raton, FL: Science Publishers.

Delourme, R., Bousset, L., Ermel, M., Duffé, P., Besnard, A.L., Marquer, B., Fudal, I. *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.

Dolatabadian, A., Patel, D.A., Edwards, D. and Batley, J. (2017) Copy number variation and disease resistance in plants. *Theor. Appl. Genet.* **130**, 2479–2490.

Fomeju, B.F., Falentin, C., Lassalle, G., Manzanares-Dauleux, M.J. and Delourme, R. (2015) Comparative genomic analysis of duplicated homoeologous regions involved in the resistance of *Brassica napus* to stem canker. *Front. Plant Sci.* **6**, 772.

Fopa Fomeju, B., Falentin, C., Lassalle, G., Manzanares-Dauleux, M.J. and Delourme, R. (2014) Homoeologous duplicated regions are involved in quantitative resistance of *Brassica napus* to stem canker. *BMC Genom.* **15**, 498.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.

García-Lor, A., Luro, F., Navarro, L. and Ollitrault, P. (2012) Comparative use of InDel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity: a perspective for genetic association studies. *Mol. Genet. Genomics*, **287**, 77–94.

Graubert, T.A., Cahan, P., Edwin, D., Selzer, R.R., Richmond, T.A., Eis, P.S., Shannon, W.D. *et al.* (2007) A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet.* **3**, e3.

Guryev, V., Saar, K., Adamovic, T., Verheul, M., van Heesch, S.A., Cook, S., Pravenec, M. *et al.* (2008) Distribution and functional impact of DNA copy number variation in the rat. *Nat. Genet.* **40**, 538–545.

Hatzig, S.V., Frisch, M., Breuer, F., Nesi, N., Ducournau, S., Wagner, M.H., Leckband, G. *et al.* (2015) Genome-wide association mapping unravels the genetic control of seed germination and vigor in B*rassica napus*. *Front. Plant Sci.* **6**, 221.

He, Z., Cheng, F., Li, Y., Wang, X., Parkin, I.A., Chalhoub, B., Liu, S. *et al.* (2015) Construction of Brassica A and C genome-based ordered pan-transcriptomes for use in rapeseed genomic research. *Data Brief*, **4**, 357–362.

Huang, Y.J., Jestin, C., Welham, S.J., King, G.J., Manzanares-Dauleux, M.J., Fitt, B.D.L. and Delourme, R. (2016) Identification of environmentally stable QTL for resistance against *Leptosphaeria maculans* in oilseed rape (*Brassica napus*). *Theor. Appl. Genet.* **129**, 169–180.

Hurgobin, B., Golicz, A.A., Bayer, P.E., Chan, C.-K.K., Tirnaz, S., Dolatabadian, A., Schiessl, S.V. *et al.* (2017) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol. J.* https://doi.org/10.1111/pbi.12867.

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W. *et al.* (2004) Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951.

Jan, H.U., Abbadi, A., Lücke, S., Nichols, R.A. and Snowdon, R.J. (2016) Genomic prediction of testcross performance in canola (*Brassica napus*). *PLoS ONE*, **11**, e0147769.

Jestin, C., Lodé, M., Vallée, P., Domin, C., Falentin, C., Horvais, R., Coedel, S. *et al.* (2011) Association mapping of quantitative resistance for *Leptosphaeria maculans* in oilseed rape (*Brassica napus* L.). *Mol. Breed.* **27**, 271–287.

Kaur, S., Cogan, N., Ye, G., Baillie, R., Hand, M., Ling, A., McGearey, A. *et al.* (2009) Genetic map construction and QTL mapping of resistance to blackleg (*Leptosphaeria maculans*) disease in Australian canola (*Brassica napus* L.) cultivars. *Theor. Appl. Genet.* **120**, 71–83.

Larkan, N.J., Lydiate, D.J., Yu, F., Rimmer, S.R. and Borhan, M.H. (2014) Co-localisation of the blackleg resistance genes *Rlm2* and *LepR3* on *Brassica napus* chromosome A10. *BMC Plant Biol.* **14**, 387.

Larkan, N.J., Raman, H., Lydiate, D.J., Robinson, S.J., Yu, F., Barbulescu, D.M., Raman, R. *et al.* (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 Biol.* **16**, 183.

Li, C., Liu, S., Sivasithamparam, K. and Barbetti, M. (2009) New sources of resistance to Sclerotinia stem rot caused by *Sclerotinia sclerotiorum* in Chinese and Australian *Brassica napus* and *B. juncea* germplasm screened under Western Australian conditions. *Australas. Plant Pathol.* **38**, 149–152.

Li, F., Chen, B., Xu, K., Wu, J., Song, W., Bancroft, I., Harper, A.L. *et al.* (2014) Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). *DNA Res.* **21**, 355–367.

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

Liu, L., Qu, C., Wittkop, B., Yi, B., Xiao, Y., He, Y., Snowdon, R.J. *et al.* (2013) A high-density SNP map for accurate mapping of seed fibre QTL in *Brassica napus* L. *PLoS ONE*, **8**, e83052.

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A.P., Zhao, M. *et al.* (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* **5**, 3930.

Luo, Z., Wang, M., Long, Y., Huang, Y., Shi, L., Zhang, C., Liu, X. *et al.* (2017) Incorporating pleiotropic quantitative trait loci in dissection of complex traits: seed yield in rapeseed as an example. *Theor. Appl. Genet.* **130**, 1569–1585.

MacQueen, A. and Bergelson, J. (2016) Modulation of R-gene expression across environments. *J. Exp. Bot.* **67**, 2093–2105.

Mason, A.S., Higgins, E., Snowdon, R.J., Batley, J., Stein, A., Werner, C. and Parkin, I.A.P. (2017) A user guide to the Brassica 60K Illumina Infinium™ SNP genotyping array. *Theor. Appl. Genet.* **130**, 621–633.

McHale, L.K., Haun, W.J., Xu, W.W., Bashkar, P.B., Anderson, J.E., Hyten, D.L., Gerhardt, D.J. *et al.* (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiol.* **159**, 1295–1308.

Obermeier, C., Hossain, M.A., Snowdon, R.J., Knüfer, J., 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.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909.

Qian, L., Qian, W. and Snowdon, R.J. (2014) Sub-genomic variation patterns as a signature of breeding in rapeseed. *BMC Genom.* **15**, 1170.

Qian, L., Voss-Fels, K., Cui, Y., Jan, H.U., Samans, B., Obermeier, C., Qian, W. *et al.* (2016) Deletion of a stay-green gene associates with adaptive selection in *Brassica napus*. *Mol. Plant*, **9**, 1559–1569.

Raman, R., Taylor, B., Marcroft, S., Stiller, J., Eckermann, P., Coombes, N., Rehman, A. *et al.* (2012a) 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.

Raman, R., Taylor, B., Lindbeck, K., Coombes, N., Barbulescu, D., Salisbury, P. and Raman, H. (2012b) Molecular mapping and validation of *Rlm*1 genes for resistance to *Leptosphaeria maculans* in canola (*Brassica napus* L.). *Crop Pasture Sci.* **63**, 1007–1017.

Raman, H., Raman, R., Coombes, N., Song, J., Diffey, S., Kilian, A., Lindbeck, K. *et al.* (2016) Genome-wide association study identifies new loci for resistance to *Leptosphaeria maculans* in canola. *Front. Plant Sci.* **7**, 1513.

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.

Saxena, R.K., Edwards, D. and Varshney, R.K. (2014) Structural variations in plant genomes. *Brief Funct. Genomics*, **13**, 296–307.

Schiessl, S., Iniguez-Luy, F., Qian, W. and Snowdon, R.J. (2015) Diverse regulatory factors associate with flowering time and yield responses in winter-type *Brassica napus*. *BMC Genom.* **16**, 737.

Schiessl, S., Huettel, B., Kuehn, D., Reinhard, R. and Snowdon, R.J. (2017) Post-polyploidisation morphotype diversification associates with gene copy number variation. *Sci. Rep.* **7**, 41845.

Schmutzer, T., Samans, B., Dyrska, E., Lespinasse, D., Micic, Z., Abel, S., Duchscherer, P. *et al.* (2015) Species-wide genome sequence and nucleotide polymorphism datasets from the model allopolyploid plant *Brassica napus*. *Sci. Data*, **2**, 150072.

Shen, J.D., Araki, H., Chen, L.L., Chen, J.Q. and Tian, D.C. (2006) Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. *Genetics*, **172**, 1243–1250.

Siebold, M. and von Tiedemann, A. (2012) Potential effects of global warming on oilseed rape pathogens in Northern Germany. *Fungal Ecol.* **5**, 62–72.
Snijders, A.M., Nowak, N.J., Huey, B., Fridlyand, J., Law, S., Conroy, J., Tokuyasu, T. *et al.* (2005) Mapping segmental and sequence variations among laboratory mice using BAC array CGH. *Genome Res.* **15**, 302–311.
Snowdon, R.J. (2007) Cytogenetics and genome analysis in Brassica crops. *Chromosome Res.* **15**, 85–95.
Snowdon, R.J., Abbadi, A., Kox, T., Schmutzer, T. and Leckband, G. (2015) Heterotic haplotype capture: precision breeding for hybrid performance. *Trends Plant Sci.* **20**, 410–413.
Springer, N.M., Ying, K., Fu, Y., Ji, T., Yeh, C.T., Jia, Y., Wu, W. *et al.* (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet.* **5**, e1000734.
Stein, A., Coriton, O., Rousseau-Gueutin, M., Samans, B., Schiessl, S.V., Obermeier, C., Parkin, I.A. *et al.* (2017) Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. *Plant Biotechnol. J.* **15**, 1478–1489.
Stich, B., Möhring, J., Piepho, H.P., Heckenberger, M., Buckler, E.S. and Melchinger, A.E. (2008) Comparison of mixed-model approaches for association mapping. *Genetics*, **178**, 1745–1754.
Sun, C., Wang, B., Wang, X., Hu, K., Li, K., Li, Z., Li, S. *et al.* (2016). Genome-wide association study dissecting the genetic architecture underlying the branch angle trait in rapeseed (*Brassica napus* L.). *Sci. Rep.* **6**, 33673.
Tan, S., Zhong, Y., Hou, H., Yang, S. and Tian, D. (2012) Variation of presence/absence genes among Arabidopsis populations. *BMC Evol. Biol.* **12**, 86.
The *Brassica rapa* Genome Sequencing Project Consortium. (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **43**, 1035–1040.
Tian, L., Wang, J., Fong, M.P., Chen, M., Cao, H., Gelvin, S.B. and Chen, Z.J. (2003) Genetic control of developmental changes induced by disruption of Arabidopsis histone deacetylase 1 (AtHD1) expression. *Genetics*, **165**, 399–409.
Uppalapati, S.R., Ishiga, Y., Doraiswamy, V., Bedair, M., Mittal, S., Chen, J., Nakashima, J. *et al.* (2012) Loss of abaxial leaf epicuticular wax in *Medicago truncatula irg1/palm1* mutants results in reduced spore differentiation of anthracnose and nonhost rust pathogens. *Plant Cell*, **24**, 353–370.
Voss-Fels, K. and Snowdon, R.J. (2016) Understanding and utilizing crop genome diversity via high-resolution genotyping. *Plant Biotechnol. J.* **14**, 1086–1094.
Wan, H., Chen, L., Guo, J., Li, Q., Wen, J., Yi, B., Ma, C. *et al.* (2017) Genome-wide association study reveals the genetic architecture underlying salt tolerance-related traits in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* **8**, 593.
Wang, X., Yu, K., Li, H., Peng, Q., Chen, F., Zhang, W., Chen, S. *et al.* (2015) High-density SNP map construction and QTL identification for the apetalous character in *Brassica napus* L. *Front. Plant Sci.* **6**, 1164.
Wei, D., Mei, J., Fu, Y., Disi, J.O., Li, J. and Qian, W. (2014) Quantitative trait loci analyses for resistance to *Sclerotinia sclerotiorum* and flowering time in *Brassica napus*. *Mol. Breed.* **34**, 1797–1804.
Wei, L., Jian, H., Lu, K., Filardo, F., Yin, N., Liu, L., Qu, C. *et al.* (2016) Genome-wide association analysis and differential expression analysis of resistance to *Sclerotinia* stem rot in *Brassica napus*. *Plant Biotechnol. J.* **14**, 1368–1380.
Wilson, G.M., Flibotte, S., Missirlis, P.I., Marra, M.A., Jones, S., Thornton, K., Clark, A.G. *et al.* (2006) Identification by full-coverage array CGH of human DNA copy number increases relative to chimpanzee and gorilla. *Genome Res.* **16**, 173–181.
Wu, J., Cai, G., Tu, J., Li, L., Liu, S., Luo, X., Zhou, L. *et al.* (2013) Identification of QTLs for resistance to Sclerotinia stem rot and BnaC.IGMT5.a as a candidate gene of the major resistant QTL *SRC6* in *Brassica napus*. *PLoS ONE*, **8**, e67740.
Wu, J., Zhao, Q., Yang, Q., Liu, H., Li, Q., Yi, X., Cheng, Y. *et al.* (2016) Comparative transcriptomic analysis uncovers the complex genetic network for resistance to *Sclerotinia sclerotiorum* in *Brassica napus*. *Sci. Rep.* **6**, 19007.
Würschum, T., Boeven, P.H.G., Langer, S.M., Longin, C.F.H. and Leiser, W.L. (2015) Multiply to conquer: copy number variations at *Ppd-B1* and *Vrn-A1* facilitate global adaptation in wheat. *BMC Genet.* **16**, 96.
Würschum, T., Longin, F.H., Hahn, V., Tucker, M.R. and Leiser, W.L. (2017) Copy number variations of CBF genes at the Rr-A2 locus are essential components of winter hardiness in wheat. *Plant J.* **89**, 764–773.
Xiong, Z. and Pires, J.C. (2011) Karyotype and identification of all homoeologous chromosomes of allopolyploid *Brassica napus* and its diploid progenitors. *Genetics*, **187**, 37–49.
Xu, L., Hu, K., Zhang, Z., Guan, C., Chen, S., Hua, W., Li, J. *et al.* (2016) Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.). *DNA Res.* **23**, 43–52.
Zhao, J. and Meng, J. (2003) Genetic analysis of loci associated with partial resistance to *Sclerotinia sclerotiorum* in rape-seed (*Brassica napus* L.). *Theor. Appl. Genet.* **106**, 759–764.
Zhao, J., Udall, J.A., Quijada, P.A., Grau, C.R., Meng, J. and Osborn, T.C. (2006) Quantitative trait loci for resistance to *Sclerotinia sclerotiorum* and its association with a homeologous non-reciprocal transposition in *Brassica napus* L. *Theor. Appl. Genet.* **112**, 509–516.
Żmieńko, A., Samelak, A., Kozłowski, P. and Figlerowicz, M. (2014) Copy number polymorphism in plant genomes. *Theor. Appl. Genet.* **127**, 1–18.
Zou, J., Zhao, Y., Liu, P., Shi, L., Wang, X., Wang, M., Meng, J. *et al.* (2016) Seed quality traits can be predicted with high accuracy in *Brassica napus* using genomic data. *PLoS ONE*, **11**, e0166624.

Supporting information

Additional Supporting Information may be found online in the supporting information section at the end of the article.

Figure S1. Sequence analyses for a QTL detected for blackleg disease resistance in DH line MOY4, covering a 12 kb region on chromosome C04 (C04_QTL1). (a) Anchoring of consensus Sanger reads to the reference genome Darmor-*bzh*; (b) anchoring of Sanger reads to individual NAM parents of five targeted genes; (c) PCR amplification of genes.
Figure S2. Manhattan plots resulting from genome-wide association analysis (GWAS) for blackleg resistance in the NAM panel using (a) SNP markers, (b) SNP and SNaP markers. The *x*-axis represents the marker positions along each chromosome anchored to the Darmor-*bzh* reference; the *y*-axis shows the $-\log10_{(P\text{-value})}$ for the trait–marker association. The solid horizontal line indicates the arbitrary selected threshold at $-\log10_{(P\text{-value})} \geq 3$ and the dashed line indicates the significance threshold $-\log10_{(P\text{-value})} \geq 4.33$ or FDR <0.10.
Figure S3. Detection of a QTL for blackleg disease resistance on chromosome C04 using GWAS with (a1) only SNP markers, and (a2) SNP plus SNaP markers. Haplotype patterns reveal two blocks at the beginning of the chromosome, one (BnPAV_C04_1) spanning 40k and harbouring 3 SNaP markers (b1), and 13 genes (c1), and a second (BnPAV_C04_2) spanning 200k and harbouring 2 SNaP markers (b2), and 19 genes (c2), respectively.
Table S1. Mean genomic distances and haploblock numbers between SNP markers and between SNP and SNaP markers, respectively.
Table S2. Summary of SNP-trait associations with a $-\log10_{(P\text{-value})} \geq 3$.
Table S3. Summary of SNP- and SNaP-trait associations with a $-\log10_{(P\text{-value})} \geq 3$.
Table S4. Common QTL regions with literature.
Table S5. Information on PCR primer sequences, reaction setups and cycling conditions.

4 Mapping the presence of absence: Gene presence-absence associated with quantitative disease resistance in *Brassica napus*

Gabur I, Chawla HS, von Tiedemann A, Snowdon RJ, Obermeier C

Scientific Reports (under review)

Scientific Reports, Full Research Paper

Mapping the presence of absence: Gene presence-absence associated with quantitative disease resistance in *Brassica napus*

Iulian Gabur[1*], Harmeet Singh Chawla[1*], Andreas von Tiedemann[2], Rod J. Snowdon[1], Christian Obermeier[1]

[1]Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, 35392 Giessen, Germany

[2]Section of General Plant Pathology and Crop Protection, Georg August University Göttingen, 37077 Göttingen, Germany

[*] These authors contributed equally to this work.

Author for correspondence:

Christian Obermeier

Tel: +49 641 9937426

Email: christian.obermeier@agrar.uni-giessen.de

Figures: 4; all 4 in color

Tables: 1

Supplementary Figures: 3

Supplementary Tables: 10

Abstract

Although copy number variation (CNV) and presence-absence variation (PAV) have been discovered in selected gene families in most crop species, the global prevalence of these polymorphisms in most complex genomes is still unclear and their influence on quantitatively inherited agronomic traits is still largely unknown. Here we analyze the association of gene PAV with resistance of oilseed rape (*Brassica napus*) against the important fungal pathogen *Verticillium longisporum*, as an example for a complex, quantitative disease resistance in the strongly rearranged genome of a recent allopolyploid crop species. Using Single Nucleotide absence Polymorphism (SNaP) markers to efficiently trace PAV in breeding populations, we significantly increased the resolution of loci influencing *V. longisporum* resistance in biparental and multi-parental mapping populations. Gene PAV, assayed by resequencing mapping parents, was observed in 23-51% of the genes within confidence intervals of quantitative trait loci (QTL) for *V. longisporum* resistance, and high-priority candidate genes identified within QTL were all affected by PAV. The results demonstrate the prominent role of gene PAV in determining agronomic traits, suggesting that this important class of polymorphism should be exploited more systematically in future breeding.

Key words: breeding, genome-wide gene variation, presence-absence variation, polyploidy, rapeseed

INTRODUCTION

Duplication of genes followed by diversification is a common process shaping the evolution of plant species by natural and artificial (breeding) selection [1]. Genes can be duplicated by different mechanisms, including tandem duplication, transposon-mediated duplication, segmental duplication, or in the most extreme form by whole-genome duplication (WGD) or polyploidization. WGD is common in the evolutionary history of many wild and cultivated plant species. Different terms have been used frequently to describe short- and long-range genomic duplication and genome structural variation (SV), a term originally defined in reference to insertions, deletions and inversions greater than 1 kb in size [2,3,4]. In contrast to small-scale insertion-deletion (InDel) polymorphisms, which are generally defined as small insertions or deletions of a few nucleotides (up to 50 bp), SV in the size range of genes (up to a few kb) can give rise to copy number variation (CNV) or presence/absence variation (PAV). The latter is an extreme form of CNV where fragments in the size range of genes are missing from the genomes of some investigated genotypes.

Genes affected by CNV in diploid plant species have been linked to local adaptation of wild populations [5] and to important agronomical traits in crops, including abiotic and biotic stress tolerance [1,3,6,7, 8]. However, the strong impact of CNV and other forms of SV in polyploid crop genomes on evolution and trait selection was not recognized until the last few years, when recognition of their relevance was facilitated by large-scale genotyping and genome sequencing in large breeding populations of numerous crops [9]. One recent genotyping technology which is particularly suitable for detection of long-range SV is Bionano optical genome mapping using nano-channel arrays [10]. This method involves imaging of high-molecular weight, fluorescently-labeled DNA molecules and creation of large restriction maps represented as stretches of light and dark regions (resembling a barcode), which then can be aligned to an *in silico* generated optical map of a reference genome assembly. A key factor distinguishing this approach from previous technologies for SV analysis is that the DNA molecules are not sheared, thus enabling the capture of long-range genomic information stretching up to several hundred kilobases. In combination with accurate genome assemblies even for strongly complex polyploid crop genomes, optical mapping opens new avenues for dissection of genomic rearrangements associated with traits relevant for commercial plant breeding.

In polyploid crop species, duplications, InDels and PAV have been found to control a number of important agronomical traits, for example flowering time and vernalization requirement in oilseed rape [11,12] and wheat [13], abiotic stress tolerance in wheat [14,15] and biotic stress tolerance in tobacco [16]. Oilseed rape (*Brassica napus*) is a recent polyploid crop species originating from the inter-specific hybridization between the two diploid progenitor species, *B. oleracea* and *B. rapa*. Due to high levels of homoeology between the two progenitor subgenomes, widespread structural rearrangements are a common phenomenon within the rapeseed genome [17,18,19], while its ancestral hexaploid progenitor

genomes already carried intensive structural and functional modifications through long-term genome fractionation and evolution [20,21,22,23].

Previously, was demonstrated that biotic stress tolerance is commonly associated with inheritable single nucleotide presence/absence polymorphism in segregating breeding populations [24]. Here we focus on gene PAV and evaluate the biological relevance and prevalence of gene PAV associated with resistance to a common fungal disease of oilseed rape, Verticillium stem striping caused by *Verticillium longisporum* [24]. By genetic mapping of genome-wide SNP/SNaP markers, along with short-read Illumina sequencing in combination with long-range optical mapping, we demonstrate that inclusion of presence-absence polymorphisms in quantitative trait locus (QTL) mapping strategies enables also reliable identification of gene PAV with a putative role in *V. longisporum* resistance. The results provide a valuable example for the importance of pangenomic gene variation for breeding of a key trait in a major polyploid crop.

RESULTS

SNaP markers increase QTL detection power

Analysis of raw genotype data from the Brassica 60k SNP array [26,27] for a doubled haploid population ExR53-DH, showed a high percentage of failed calls for some individual SNP probes across all analyzed 244 genotypes. Commonly, SNP probes with a high percentage of failed calls across all genotypes (>10%) are considered technical artefacts and these SNP probes are usually completely removed from genetic analyses. Here instead of being discarded due to excessive missing data, these markers could be recovered from the dataset as Single Nucleotide absence Polymorphism (SNaP) markers [24] and subsequently implemented for biparental genetic mapping. Two genetic maps were produced, one using only SNP markers (the "SNP map") and one using SNP plus SNaP markers (the "SNaP map"), respectively. Comparison of the SNP and SNaP maps revealed that large chromosomal regions were not covered in the SNP map (e.g. for chromosome A03 compare Figure 1a and 1b). Surprisingly, QTL mapping using the SNaP map increased the number of detectable QTL from five to 17 (Supplementary Table S1), with substantially increased LOD scores also indicating a dramatic increase in QTL detection power when including SNaP marker data. Furthermore, the map resolution and precision across QTL intervals were considerably increased by inclusion of SNaP markers. Interestingly, some QTL detected only in the SNaP map contained only SNP markers within the QTL confidence interval (e.g. q23k-BP-A1-1, q23k-BP-A1-2, q23k-BP-A3-2), while other QTL spanned intervals containing only SNaP markers (e.g. q23k-BP-A3-1). Many SNaP markers clustered in groups, spanning large regions up to chromosome scale, while other SNaP markers were located within blocks of SNP markers (Supplementary Table S1).

Additionally, based on segregation of parental lines for *V. longisporum* resistance, a subset of subpopulations from crosses of a common elite oilseed rape parent with five synthetic *B. napus* parents was selected from the *B. napus* nested association mapping (NAM) panel described by Snowdon et al. [28]. GWAS using only SNP markers revealed a total of 18 marker-trait associations with p-values greater than -log(p-value) ≥ 3, on chromosomes A02, A07, C04, C05 and C06, however none of the associated SNP markers exceeded the false discovery rate (FDR) threshold (Supplementary Table S2, Supplementary Figure S2). In contrast, GWAS including both SNP and SNaP markers increased the number of markers associated with resistance from 18 to 41 (Supplementary Figure S2) and simultaneously increased the QTL detection power in the NAM panel (Supplementary Table S2). Most of the additional detected QTL contained only SNaP markers. A total of three markers, localized on chromosomes A07, C04 and C08 showed low p-values which overcame the FDR threshold.

Common resistance QTL colocalize with PAV in diverse genetic backgrounds

Sizes of QTL intervals for *V. longisporum* resistance showed differences between the biparental QTL mapping and the multi-parental GWAS (Supplementary Table S1 and S2). Generally, smaller QTL intervals are expected in a NAM-GWAS approach, as higher numbers of recombinations are expected from crosses involving multiple non-related parents. On the other hand, QTL intervals in a GWAS mapping approach can only be measured for markers which can be positionally anchored, whereas biparental mapping can also consider marker loci which can be genetically but not physically anchored. This led to some common resistance QTL with smaller confidence intervals observed in the biparental mapping compared to the NAM-GWAS (e.g. QTL on chromosomes A05 and C05).

Using only SNP markers, comparison of QTL detected by biparental QTL mapping and by GWAS revealed one common QTL on chromosome C05, designated q17k-BP-C5/q18k-NAM-C5-2 (Figure 2, Supplementary Tables S1 and S2), which had the highest coefficient of determination of all QTL in the biparental mapping and a medium coefficient of determination in the NAM-GWAS. In contrast, adding SNaP markers to the QTL analyses revealed 5 common overlapping QTL (Figure 2, Table 1) harbouring 2 to 90 genes per confidence interval (Supplementary Table S3).

Gene ontology and enrichment analysis for genes underlying resistance QTL

Gene ontology enrichment analysis of the biparental population revealed five enriched GO terms in 17 QTL regions, with highest significance (topgoFisher score) attributable to the terms ‘chitin catabolic process’ and ‘response to biotic stimulus’ (Supplementary Table S4). In the NAM population eight enriched terms for genes harboured in 28 QTL regions were mainly related to cell growth. In the co-localizing QTL sections six enriched GO terms were identified. A total of 144 genes are harboured

within the confidence intervals of the 5 co-localizing resistance QTL (Supplementary Table S5). The GO terms for the 144 genes harboured in the five consensus QTL intervals reflect annotations from just 8 genes (labelled in green in Supplementary Table S5) related to cell-wall modification (expansins) and pathogen defence (defensins) on chromosome A05 (CoLOC2) and selenium binding on chromosome C08 (CoLOC5). One gene each returned annotation terms containing 'response to stress' and 'systemic acquired resistance', respectively. For four genes an annotation term 'defense response to fungus' or 'response to symbiotic fungus' was assigned, mostly based on plant defensin genes, whereas seven genes returned annotations containing the term 'cell wall', mostly based on expansin genes and pectin esterase-like protein genes.

Both long and short-range PAV influence *V. longisporum* resistance

To validate if SNaP markers can help to reliably detect otherwise invisible QTL, resistance-associated QTL detected with SNP and SNaP markers in the ExR53-DH population were physically located in the *B. napus* Darmor-*bzh* reference genome and the corresponding sequences were compared with optical genome maps. *De novo* assembly for Express 617 and R53 was performed using 300 Gb of BioNano data (~250x coverage) from Express 617 and 140 Gb (~116x coverage) from R53, respectively. DNA molecules from both genotypes exhibited a very high N50 > 180 kbp, ensuring that long-range genomic information was covered (Supplementary Table S6). One nick label was detected for every 10,000 bp of the molecules, also indicating a uniform coverage and a high SV detection power. The final assembly comprised 1,368 and 1,331 optical maps with N50 values of 234 kb and 235 kb, respectively. The total lengths of the optical mapping assemblies were 978 Mb for Express 617 and 874 Mb for R53. The high molecule sizes, along with the total size of the assemblies close to the predicted genome size for *B. napus*, indicate a good assembly quality suitable for reliable detection of long-range SV. Optical maps enabled accurate detection of small to medium-size deletions and insertions in the size range of genes (from 3 to 5 kb). Large-scale deletions detected with optical mapping data were also consistently detected by consecutively anchored SNaP markers in the ExR53-DH genetic map (Supplementary Figure S1).

The 17 detected QTL regions harboured between 3 and 21 SNP and/or SNaP markers. In total, 122 markers were contained within QTL regions and 72 could be anchored to the Darmor-*bzh* reference genome (Supplementary Table S7). All regions were investigated for long-range structural variants within QTL regions in parental genomes by analyzing the optical map data and comparing the data with SNP/SNaP marker patterns in the segregating population. Medium to long-range structural variation (>3 kb deletions) were confirmed for 16 out of 18 (89%) SNaP marker positions in the two parents (Supplementary Table S7), suggesting that genetically anchored consecutive SNaP markers

can reliably detect medium to long-range presence/absence polymorphism associated with *V. longisporum* resistance.

Figure 1 shows an example for comparison of genetic mapping from the DH population, reference anchoring of SNP and SNaP markers and optical mapping data of two parents for chromosome A03, which harbours two additional QTL not detectable using the SNP map. One of the QTL detectable only with the SNaP map, q23k-BP-A3-1, harboured 4 SNaP markers, whereas another, QTL q23k-BP-A3-2, harboured 14 SNP markers within the confidence interval (Supplementary Table S7). The QTL q23-BP-A3-1 is localized in a region with long consecutively ordered stretches of SNaP markers. For the parental line Express 617, both QTL regions were covered in the optical maps (Figure 1c), whereas for the synthetic *B. napus* parent R53 about 6 Mb overlapping the QTL region q-23k-BP-A3-1 was not covered. The region not covered in R53 corresponds to a 13.8 cM interval on the genetic SNaP map (3.456 Mb), with the resistance allele contributed by Express 617 (Figure 1a and b). This lack of optical map alignments in QTL q23k-BP-A3-1, combined with segregating, consecutively anchored SNaP markers in the segregating DH population, confirms the deletion of this large chromosomal region in parental line R53 and shows that this deletion is involved in resistance expression. In contrast, the second QTL, q23k-BP-A3-2, is located in a region on chromosome A03 with long consecutively mapped stretches of SNP markers. Flanking the QTL region, only isolated SNaP markers were mapped. The isolated SNaP markers detected close to the QTL thus probably represent short-range PAV, potentially down to even single-nucleotide level. Nevertheless, saturating the genetic map on chromosome A03 by adding SNaP markers facilitated the detection of QTL q23k-BP-A3-2, whereas no QTL could be mapped in this region using only the SNP map (Figure 1a, Supplementary Table S1). This example demonstrates that the addition of SNaP markers for genetic mapping was causal for a higher detection power of *V. longisporum* resistance QTL associated with both presence/presence as well as presence/absence polymorphisms.

Gene PAV is a key determinant of *V. longisporum* resistance

In order to identify putatively absent genes from regions associated with *V. longisporum* disease resistance in the NAM panel, we combined Illumina 60K SNP chip array genotyping and Illumina resequencing data for the six NAM parental lines with GWAS data from the segregating NAM population. Illumina resequencing confirmed medium to long-range presence/absence variation in the respective parents for 17 QTL regions (Supplementary Table S6). 60% of the 28 detected resistance QTL were affected by medium to long-range PAV in the NAM population.

Figure 3 shows an example of the analyses for the SNaP marker Bn-A03-p10964394, which is associated with *V. longisporum* resistance within the QTL interval q22k-NAM-A3-2 (Supplementary Table S8). This region CoLOC1 (Supplementary Table S3) also overlaps with QTL q23k-BP-A3-1

detected in the biparental cross (Figure 3a). Comparison of the marker-trait segregation in the NAM panel with marker and resequencing data for the six parental genotypes confirmed the expected pattern and revealed a putatively deleted interval within a part (8 %) of the QTL region of the susceptible parent. This deletion corresponds to a 30 kb region containing 10 protein-coding genes (Figure 3c) within the entire QTL interval of 382 kb containing a total of 90 genes (Figure 3b). However, from 5 markers (3 SNaPs, 2 SNPs) within the LD block of the common QTL interval CoLOC1, only the SNaP marker Bn-A03-p10964394 within the deleted region in R53 is significantly associated with *V. longisporum* resistance, suggesting that presence/absence polymorphism is involved in resistance. Based on the resequencing data of the parental lines, two to four genes are affected by deletions within the QTL interval (Figure 3c), namely BnaA03g21190D (coding for an uncharacterized protein), BnaA03g21200D (coding for a skp1-like protein involved in ubiquitin-dependent protein catabolic processes), BnaA03g21210D (coding for aquaporin pip1-2 protein involved in transporter activities), BnaA03g21220D (coding for an uncharacterized protein) and BnaA03g21230D (coding for an ATP-dependent helicase brm-like protein involved in DNA binding).

SNP/SNaP marker patterns within *V. longisporum* QTL regions were also compared with parental whole-genome sequence coverage in the two parents of the biparental population ExR53-DH. Complete or partial presence/absence events were reconfirmed by resequencing coverage analysis of the two parents for 12 out of 18 reference-anchored SNaP markers (67%) and for 5 out of 6 QTL regions harbouring one or several reference-anchored SNaP markers (83%, Supplementary Table S9). The 14 cM QTL interval q23k-BP-A3-1 spans 3,45 Mb on the Darmor-*bzh* reference, containing 683 genes (Supplementary Table S9). Part of this region is shown in detail in Figure 1 as co-localizing QTL region CoLOC1. Within CoLOC1 a total of 382 kb, containing 90 genes, overlaps in the biparental and multi-parental QTL analyses (Supplementary Table S3 and Figure 3).

Identification of genes affected by PAV by genome resequencing

Using genomic resequencing data from 52 accessions [29], including the six NAM parents investigated in our study, we tested the software package SGSGeneLoss [30] for gene loss calling and classification of genes as PAV genes or non-PAV genes. Parameters of 5% minimum gene size coverage (lostCutoff = 0.05) and minimum coverage of two reads for a gene (minCov = 2) was used by Hurgobin et al. [19] to call a gene as present in a genotype. However, using these parameters for the 6 NAM parents used in our study and comparing gene loss calling with PCR amplification data for 11 gene fragments from two genomic regions on chromosome A03 and C04 (Gabur et al.,[24], Figure 3 and Supplementary Figure S3) revealed an inaccurate gene fragment loss classification in 9 of 66 cases (86 % accuracy) and no PAV was called for the chromosome C04 regions. This suggests that the chosen parameters were not stringent enough in the six genotypes to distinguish genuine gene loss from data processing

noise due to misalignments. Hence, gene loss was accordingly underestimated using these SGSGeneLoss program parameters (and using Darmor-*bzh* v.4.1 as a reference). We therefore developed a customized pipeline which calibrated gene loss calling parameters based on PCR wet lab data and also included the gene size in the calculations to reduce further alignment biases. This procedure increased accuracy of gene loss calling to 96%. Both methods were then used for gene loss calling in the six NAM parents, Express617 and R53, firstly across the whole genome and then within identified QTL regions. Using our modified approach, 49 % of total genes (49,701 of 101,039) were estimated to be affected by gene loss and 51 % of genes within the 28 QTL regions (1,334 of 2,601) were estimated to be affected by gene loss (PAV genes) between the 6 NAM parents. Between Express 617 and R53, 23 % of total genes (23,772 of 101,039) and 23 % of genes within 17 QTL regions (590 of 2,646) were estimated to be affected by gene loss. A total of 144 genes were harboured in the 5 QTL intervals which co-localized between the biparental and multi-parental QTL mapping. Of these, 74 (52 %) in the NAM panel and 37 (26 %) in the ExR53-DH panel were estimated to be affected by PAVs. The genes estimated to be affected by PAV overlap in 20% of cases within the 5 co-localizing QTL regions between the two populations (Supplementary Table S10).

Identification of candidate genes associated with PAV

Within the 17 detected *V. longisporum* resistance QTL regions in the ExR53-DH population 590 (23%) from a total of 2,646 genes are affected by PAV based on coverage analysis of the parental resequencing data. In contrast, within the 28 detected QTL in the NAM population 1,334 genes from 2,601 (51%) are affected by PAV. Within the 5 QTL regions that co-localize between the ExR53-DH and NAM populations, 37% (37/144) of genes are affected by PAV in ExR53-DH and 52% (74/144) are affected by PAV in the NAM population. In total, 28 of 144 genes affected by PAV (20%) are common between the biparental and NAM populations. Analysis of genes from QTL regions affected by presence/absence polymorphism in both populations revealed that 43% of these genes (12/28) have no GO terms assigned.

DISCUSSION

The reanalysis and integration of SNP array data, short-range Illumina sequencing data and long-range Bionano optical mapping data with QTL data provided new insights into the importance of gene PAV for disease resistance expression against Verticillium stem striping in oilseed rape. This indicates that some *V. longisporum* resistance QTL are located in chromosome regions undetectable by SNP markers alone and will thus remain undiscovered in traditional SNP array-based analyses. Corresponding results were obtained for the multi-parental population using reference-anchored markers. These

findings indicate that association of PAV with disease resistance in allopolyploid *B. napus* is a common and prevalent phenomenon. Hence, the use of SNaP markers for genetic mapping, QTL analyses and breeding for disease resistance in oilseed rape could prove to be of great interest for scientists and oilseed rape breeders. The recent allopolyploid crop species *B. napus* shows a strong abundance of SV and genomic rearrangements [18,31], and to date it is unknown how widespread this phenomenon is in other, older polyploid crop species. However, the ability to extract SNaP marker calls from commonly used, cost-effective genotyping array data [32] makes this technique interesting for high-throughput discovery of trait-associated PAV in other polyploid crops like wheat, which is known to exhibit PAV and CNV for a number of important agricultural traits [13,15,33,34].

Although intra-species PAV has been reported in many plant species, it is still unclear for most crops how many genes, what kind of genes and what traits are predominantly affected. Especially in complex polyploid crops, considerably higher numbers of diverse, high-quality reference assemblies are needed before reliable pangenomic analysis of genome-wide gene PAV becomes possible. In maize, Beló et al. [35] used comparative genomic hybridization arrays for detailed genome-wide analysis of SV, and a more recent study discovered some degree of CNV between 100 analyzed lines across more than 90% of the maize genome [36]. Hurgobin et al. [19] classified 38% of all *B. napus* genes in a diversity set of 53 genotypes as PAV-affected, including genes involved in important agronomical traits. However, bioinformatic inference of PAV from high-throughput sequencing data without experimental validation can lead to false classification of gene PAV. Here we demonstrate that using complementary PCR data to calibrate bioinformatics tools the proportion of gene PAV may be even higher in selected breeding materials.

For nucleotide binding site-leucine-rich repeat (NLR) resistance genes, which are often organized in clusters or tandem repeats in a number of plant species and crops, numerous studies have shown that fitness costs can lead to multiplication and deletion of gene family members, particularly in *A. thaliana* [37,38,39] but also in *B. napus* [17]. Samans et al. [18] described two gene clusters of 14 and 8 TIR-NBS-LRR genes, respectively, on *B. napus* chromosome C09, which are frequently deleted in natural *B. napus* compared to synthetic *B. napus* accessions. Part of one of these regions, containing a total of 7 TIR-NBS-LRR genes affected by PAV, was also identified in in our study to harbour a QTL for *V. longisporum* resistance. However, the results of our study also clearly show that gene PAVs within resistance QTL regions are not limited to clusters of typical NLR genes, but widespread and affecting functionally very diverse genes and families. In fact, only 54 of 2601 genes we detected within the 28 resistance QTL (2%) were typical NLR genes.

Different mixtures of *V. longisporum* lineages/pathotypes were used for disease resistance screening in the two different *B. napus* mapping panels. Comparison of QTL regions detected by including SNaP markers between the biparental and the multi-parental populations revealed a considerable number of common QTL. This indicates broad-spectrum, lineage/pathotype-independent resistance reactions in

genetically diverse germplasm which are of great interest for commercial resistance breeding. Thus we focused for a more detailed analysis on these 5 common QTL regions, which harbour 2 to 90 genes each within the co-localizing section of the QTL intervals.

Based on the quantitative genetic nature of the disease resistance, we assumed that single specialized genes involved in very different biological functions from each QTL interval will not contribute to *V. longisporum* resistance, but rather a number of genes from common biosynthesis pathways or with common biological functions from one or all QTL intervals. Thus, to prioritize candidate PAV genes putatively involved in broad-range resistance expression we performed a gene ontology enrichment analysis for all QTL-associated genes, separately for both populations, and also for the genes from the 5 common QTL intervals. The enriched GO terms for the genes from the QTL regions were quite different between the two populations, being mainly related to chitin and cell-wall metabolism for the biparental ExR53-DH population and mainly related to cell growth for the multi-parental NAM population. This difference is not unexpected, as the disease resistance screenings are known to be highly susceptible to the environment and slightly different mixes of *V. longisporum* lineages/pathotypes were used for inoculation of the two populations. However, the result might also indicate that common as well as different resistance mechanisms are activated by different pathotypes. Within the 144 genes from the co-localizing QTL putatively involved in pathotype-independent resistance, a number of genes coding for selenium-binding proteins, for plant defensin proteins and for expansin proteins have been found to be affected by PAV. In *A. thaliana,* expression of selenium binding proteins is tightly linked to detoxification processes related to oxidative stress [40]. Plant defensins are major components of the innate immune system of plants, are involved in the cell wall integrity signaling pathway and often show a potent, broad-spectrum antifungal activity [41]. The antifungal protein RsAFP2 from *Raphanus sativus*, a close relative of *B. napus*, has been described to exhibit antifungal activitiy against the fungus *V. dahliae*, which is closely related to *V. longisporum* [42]. Furthermore, a synthetic defensin expressed in *A. thaliana* has also been shown to exhibit antifungal activity against *V. dahliae* [43]. Expansins mediate cell wall-loosening and down-regulation of an expansin-like protein in *A. thaliana* that has been shown to increase resistance against nectotrophic fungi [44].

The prioritization of these candidate PAV genes mainly involved in cell wall integrity, growth and modification is consistent with our earlier findings that QTL for the concentration of soluble simple phenylpropanoids, which are putative precursors and degradation products of cell-wall modifications and cross-linking, are co-localizing with major resistance QTL in the biparental population ExR53-DH. In addition, the concentrations of some of these cell wall-associated compounds are significantly correlated with *V. longisporum* resistance [45].

MATERIALS AND METHODS

Phenotyping for *Verticillium longisporum* resistance in *B. napus*

Resistance phenotyping was conducted in the greenhouse at Georg August University Göttingen, Germany. In order to represent a broad range of pathogenicity traits occurring in oilseed rape fields, a spore suspension mixture of *V. longisporum* isolates VL43 (lineage A1/D1, North Germany), VLS3 (lineage A1/D1, Sweden) and PD589 (lineage A1/D3, Japan) [46,47] with a density of 1 x 10^6 spores/ml concentrations each was used to inoculate the 200 NAM lines applying the root-dipping method [48] in four experiments. Each experiment included 20 inoculated and 20 control plants for each tested genotype. Rating of symptoms was done weekly over a 4-week period, using the 1-9 disease scoring scale described by Eynck et al. [48]. Resistance screening for the biparental population ExR53-DH was performed similarly using a mixture of isolates VL40 and VL43 (both lineage A1/D1) with 202 DH lines described in detail in Obermeier et al. [45].

High molecular weight DNA isolation for optical mapping

High molecular weight (HMW) DNA isolation was carried out for Express 617 and R53 according to the IrysPrep™ Plant Tissue-Nuclei protocol provided by Bionano Genomics. Young leaves (approximately 2 grams) were harvested from dark-treated rapeseed plants. The harvested leaves were immediately fixed with 2% formaldehyde followed by homogenization in isolation buffer containing PVP-10, BME and Triton X-100. The isolated nuclei were then purified on Percoll cushions. Purified nuclei were further embedded in an agarose matrix. Agarose plugs were further subjected to proteinase K treatment followed by rigorous washings steps. Finally, HMW DNA was recovered by melting the plugs using GELase™ (Epicentre) treatment. An additional drop dialysis step was performed to ensure ultra-clean DNA. High molecular weight DNA was further subjected to sequence-specific nick-labeling using the IrysPrep™ Labeling-NLRS protocol provided by Bionano Genomics. HMW DNA was subjected to digestion by the single-stranded nicking endonuclease *Nt.BspQI* (recognition site GCTCTTC). The nicks created by *Nt.BspQI* were then repaired using fluorophore-labeled nucleotides. Nicked and labeled single DNA molecules were subsequently loaded onto an IrysChip for imaging on the Bionano Genomics Irys system.

Bio-informatics analysis for Bionano optical mapping data

DNA molecule images generated from the Irys system were computationally translated into single-molecule optical maps. These single molecules were then assembled into consensus maps using the dedicated IrysSolve pipeline (v5134) provided by Bionano Genomics. An *in silico* optical map was

generated for the Darmor-*bzh* v. 4.1 reference [17] reference using Knickers v1.5.5 and was used to calculate noise parameters for the final assembly. Optical map assemblies from Express 617 and R53 were finally aligned to the Darmor-*bzh* reference using the runCharacterize script provided by Bionano Genomics, with the settings published by Kawakatsu et al. [49]. The alignment was imported into Bionano IrysView (v2.5.1.29842) software for visualizing and detecting structural variations between Express 617, R53 and the Darmor-*bzh B. napus* reference genome.

Genetic mapping and QTL analysis

244 DH lines from the F1 of the cross Express 617 x R53 and the two parents were analyzed with the Brassica 60k Illumina Infinium array [26]. DNA was extracted from leaves using the CTAB method [50] and array genotyping assays were outsourced to TraitGenetics (Seeland, Germany). SNP calls with >85 % failed calls across all 244 genotypes were removed from further analyses. Also, SNP probe calls with >90 % or < 10% of a single allele across the population were removed, leaving a total of 22,064 SNP markers. From these, 4,654 SNP probes (21.1%) showed a segregation pattern with one allele displaying a failed call and 17,410 SNP probes showed a normal biallelic segregation. A "SNaP map" was created from all quality-filtered 22,064 markers (2,714 marker bins), including biallelic and presence/absence polymorphisms, while a "SNP map" was created using only the 17,410 markers showing presence/presence polymorphism (2,176 marker bins). Genetic maps were created using the software MSTMap [51] applying the kosambi distance function and a cut-off *p* value of 10^{-30}. QTL analyses were performed using the software QGene 4.3.9 and 4.4.0 [52] applying composite interval mapping with a scan interval of 1 milliMorgan and automatic forward cofactor selection. Mean normalized AUDPC values from the four *V. longisporum* greenhouse resistance screenings were used as trait input data [45].

Genome-wide association studies

The NAM panel was also genotyped using the 60K Illumina Infinium Brassica SNP array as described above and data was filtered according to Gabur et al. [24]. Using the Darmor-*bzh* reference v4.1 we anchored 28,073 SNP markers by BLASTn using CLC Genomics Workbench v. 9.0 (Qiagen Bioinformatics). The SNP map contained 18,068 markers, the SNaP map contained 21,695 markers. Association analyses were conducted using the R package GenABEL [53]. A mixed linear model approach that increases detection power [54] was adjusted for population stratification by including the kinship matrix and the first two principal components as covariates [55]. Stringent significance cutoff values were set at a false discovery rate (FDR) correction of 10% [56]. To reduce the type II error rate,

we also captured the SNP-trait associations for disease resistance using an arbitrary threshold of $-\log10(p) \geq 3$.

Linkage disequilibrium (LD) analysis and haplotype construction

Whole genome linkage disequilibrium (LD) was calculated using the squared allele-frequency correlations (r^2) between pairs of SNPs. Only markers with a maximum of 10% missing data and MAF ≥ 0.05 were included in the analysis. Haplotype patterns were assessed for SNP and SNaP markers that showed significant marker trait association at the adjusted Bonferroni threshold of $-\log10(\text{P-value}) \geq 4.33$. Haplotype blocks were defined with the confidence interval method described by Gabriel et al. [57] in Haploview version 4.2 [58] and visualized with the R package LDheatmap [59].

Resequencing and coverage analysis

The Illumina 250 bp paired-end raw sequencing data for the two parents of the biparental population, Express 617 and R53, was described previously by Stein et al. [60], while Illumina 100 bp paired-end raw sequencing data for the 6 NAM parents was described by Schmutzer et al. [29]. Sequences were aligned to the Darmor-*bzh* v. 4.1 reference using CLC Genomics Workbench v. 9.0 (Qiagen Bioinformatics). Genes from QTL regions were classified as affected or not affected by presence/absence variation based on coverage analysis of the WGS data from the parental lines. Coverage differences were calculated using the bedtools software package v. 2.27.0 with multiBamCov. A minimum cutout threshold of 1.5 aligned reads was used to differentiate between gene presence and absence. The threshold was selected using PCR data for calibration available for the six NAM parental lines from Gabur et al. [24]. Additionally, we used the SGSGeneLoss v0.1 described by Golicz et al. [30]. The visualization in Figure 1 was performed using the software MapChart v.2.3, while Figure 3 was generated using CLC Genomics Workbench 9 track lists. Identification of homoeologous exchanges (HE) was performed using the method described by Samans et al. [18]. The visualization in Figure 4 was generated partly by using the R package 'gsrc' [32].

Gene ontology enrichment analyses

101,039 Darmor-*bzh* peptide sequences (Brassica_napus.annotation_v5.pep.fa.gz) downloaded from http://www.genoscope.cns.fr/brassicanapus/data/ were used as input for Blast2Go v. 4.1.9 to produce gene ontology information for all *B. napus* genes. The R package topGO v1.0 [61] was used for gene ontology enrichment analysis.

ACKNOWLEDGEMENTS

We thank Jutta Schaper (Georg August University Goettingen) for excellent technical assistance in resistance phenotyping, Matthias Frisch (Justus Liebig University Giessen) for advice in statistical analyses, Philip Howard and Andrew Leitch (Queen Mary University of London) for processing HMW DNA samples and analysis on a Bionano Genomics Irys System for optical mapping. Iulian Gabur acknowledges funding through a scholarship from the DAAD. Funding was provided by the German Federal Ministry of Food and Agriculture (BMEL grant 28-1-45.051-10) and from the German Federal Ministry of Education and Research (BMBF grant 031A325 for the German-French Plant-KBBE consortium GEWIDIS and grant 0315964 for the national consortium PreBreedYield). The authors declare no conflict of interest.

AUTHOR CONTRIBUTION

CO and RS designed the research. AvT produced the phenotyping data. HSC performed the Optical mapping and bioinformatics analysis. IG performed genetic and other bioinformatics analyses. IG, HSC and CO analyzed the data. IG and CO wrote the original draft; all authors discussed and approved the final manuscript.

Competing Interests: The authors declare no competing interests.

REFERENCES

1 Panchy, N., Lehti-Shiu, M., & Shiu, S.-H. Evolution of Gene Duplication in Plants. Plant Physiol. **171**, 2294–2316 (2016).

2 Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. Nature Reviews Genetics **12**, 363–376 (2011).

3 Saxena, R. K., Edwards, D. & Varshney, R. K. Structural variations in plant genomes. Brief. Funct. Genomics **13**, 296–307 (2014).

4 Żmieńko, A., Samelak, A., Kozłowski, P., & Figlerowicz, M. Copy number polymorphism in plant genomes. Theor. Appl. Genet. **127**, 1–18 (2014).

5 Xu, X. et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat. Biotechnol **30**, 105–111 (2012).

6 Yang, S. et al. Rapidly evolving R genes in diverse grass species confer resistance to rice blast

disease. Proc Natl Acad Sci **110**, 18572–18577 (2013).

7 Renny-Byfield, S. & Wendel, J. F. Doubling down on genomes: polyploidy and crop plants. American Journal of Botany **101**, 1711–1725 (2014).

8 Lee, T. G., Kumar, I., Diers, B. W. & Hudson, M. E. Evolution and selection of *Rhg1*, a copy-number variant nematode-resistance locus. Molecular Ecology **24**, 1774–1791 (2015).

9 Bevan, M. W., Uauy, C., Wulff, B. B. H., Zhou, J., Krasileva, K. & Clark, M. D. Genomic innovation for crop improvement. Nature **543**, 346–354 (2017).

10 Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. Nature Reviews Genetics **7**, 85–97 (2006).

11 Schiessl, S., Hüttel, B., Kühn, D., Reinhardt, R. & Snowdon, R. J. Post-polyploidisation morphotype diversification associates with gene copy-number variation. Scientific Reports **7**, 41845 (2017a).

12 Schiessl, S., Hüttel, B., Kühn, D., Reinhardt, R. & Snowdon, R. J. Targeted deep sequencing of flowering regulators in *Brassica napus* reveals extensive copy number variation. Sci. Data **4**, 170013 (2017b).

13 Würschum, T., Boeven, P. H. G., Langer, S. M., Longin, C. F. H. & Leiser, W. L. Multiply to conquer: copy number variations at *Ppd-B1* and *Vrn-A1* facilitate global adaptation in wheat. BMC Genet. **16**, 96 (2015).

14 Sieber, A.-N., Longin, C. F. H., Leiser, W. L. & Würschum, T. Copy number variation of *CBF-A14* at the Fr-A2 locus determines frost tolerance in winter durum wheat. Theoretical and Applied Genetics **129**, 1087–1097 (2016).

15 Würschum, T., Longin, C. Friedrich H., Hahn, V., Tucker, M. R. & Leiser, W. L. Copy number variations of *CBF* genes at the Fr-A2 locus are essential components of winter hardiness in wheat. Plant J. **89**, 764–773 (2017).

16 Chen, S. et al. Unstable allotetraploid tobacco genome due to frequent homeologous recombination, segmental deletion, and chromosome loss. Molecular Plant **11**, 914–927 (2018).

17 Chalhoub, B. et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science **345**, 950-953 (2014).

18 Samans, B., Chalhoub, B. & Snowdon, R. J. Surviving a genome collision: Genomic signatures of allopolyploidiation in the recent crop species *Brassica napus*. Plant Genome **10**, 1-15 (2017).

19 Hurgobin, B.et al. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. Plant Biotechnol. J. **16**, 1265–1274 (2018).

20 Wang, X. et al.; *Brassica rapa* Genome Sequencing Project Consortium. The genome of the mesopolyploid crop species *Brassica rapa*. Nat. Genet. 43, 1035–1039 (2011).

21 Cheng. F., Mandakova, T., Wu, J., Xie, Q., Lysak, M. A. & Wang, X. Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. Plant Cell **25**, 1541–1554 (2013).

22 Liu, S. et al. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun. **5**, 3930 (2014).

23 Parkin, I. A. et al. Transcriptome and methylome profiling reveals relics of genome dominance

in the mesopolyploid *Brassica oleracea*. Genome Biol. **15**, R77 (2014).

24 Gabur, I. et al. Finding invisible quantitative trait loci with missing data. Plant Biotech. J.; 10.1111/pbi.12942 (2018).

25 Depotter, J. R. et al. *Verticillium longisporum*, the invisible threat to oilseed rape and other brassicaceous plant hosts. Mol. Plant Pathol. **17**, 1004-1016 (2016).

26 Clarke, W. E. et al. A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. Theor. Appl. Genet. **129**, 1887-1899 (2016).

27 Mason, A. S., Higgins, E. E., Stein, A., Werner, C., Batley, J., Parkin, I. A. P. & Snowdon, R. J. A user guide to the Brassica 60K Illumina Infinium™ SNP genotyping array. Theor. Appl. Genet. **130**, 621-633 (2017).

28 Snowdon, R. J., Abbadi, A., Kox, T., Schmutzer, T. & Leckband, G. Heterotic Haplotype Capture: precision breeding for hybrid performance. Trends Plant Sci. **20**, 410-413 (2015).

29 Schmutzer, T. et al. Species-wide genome sequence and nucleotide polymorphisms from the model allopolyploid plant *Brassica napus*. Scientific Data **2**, 150072 (2015).

30 Golicz, A. A. et al. Gene loss in the fungal canola pathogen *Leptosphaeria maculans*. Funct. Integr. Genomics **15**, 189–196 (2015).

31 Higgins, E. E., Clarke, W. E., Howell, E. C., Armstrong, S. J. & Parkin, I. A. P. Detecting de novo homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array. G3: Genes|Genomes|Genetics; 10.1534/g3.118.200118 (2018).

32 Grandke, F., Snowdon, R. & Samans, B. gsrc: an R package for genome structure rearrangement calling. Bioinformatics **33**, 545–546 (2017).

33 Díaz, A., Zikhali, M., Turner, A. S., Isaac, P. & Laurie, D. A. 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 (2012).

34 Nishida, H., Yoshida, T., Kawakami, K., Fujita, M., Long, B., Akashi, Y., Laurie, D. A. & Kato, K. 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. Breeding **31**, 27–37 (2013).

35 Beló, A., Beatty, M. K., Hondred, D., Fengler, K. A., Li B. & Rafalski, A. Allelic genome structural variations in maize detected by array comparative genome hybridization. Theor. Appl. Genet. **120**, 355–367 (2010).

36 Darracq, A.et al. Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. BMC Genomics **19**, 119 (2018).

37 Shen, J. D., Araki, H., Chen, L. L., Chen, J. Q. & Tian, D. C. Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. Genetics &, 1243–1250 (2006).

38 Tan, S., Zhong, Y., Hou, H., Yang, S. & Tian, D. Variation of presence/absence genes among Arabidopsis populations. BMC Evol. Biol. **12**, 86 (2012).

39 Borrelli, G.M.et al. Regulation and Evolution of NLR Genes: A Close Interconnection for Plant Immunity. Int J Mol Sci. **19**, 6 (2018).

40 Valassakis, C., Livanos, P., Minopetrou, M., Haralampidis, K. & Roussis, A. Promoter analysis and functional implications of the selenium binding protein (SBP) gene family in *Arabidopsis thaliana*. Journal of Plant Physiology **224–225**, 19-29 (2018).

41 Parisi, K., Shafee, T. M. A., Quimbar, P., van der Weerden, N. L., Bleackley, M. R. & Anderson M. A. The evolution, function and mechanisms of action for plant defensins. Semin. Cell Dev. Biol;10.1016/j.semcdb.2018.02.004 (2018).

42 De Samblanx, G. W. et al. Mutational analysis of a plant defensin from radish (*Raphanus sativus* L.) reveals two adjacent sites important for antifungal activity, J. Biol. Chem. **272**, 1171–1179 (1997).

43 Li, F., Shen, H., Wang, M., Fan, K., Bibi, N., Ni, M., Yuan, S. & Wang X. A synthetic antimicrobial peptide BTD-S expressed in *Arabidopsis thaliana* confers enhanced resistance to *Verticillium dahliae*. Molecular Genetics and Genomics **291**, 1647–1661 (2016).

44 Abuqamar, S., Ajeb, S., Sham, A., Enan, M. R. & Iratni, R. A mutation in the expansin-like A2 gene enhances resistance to necrotrophic fungi and hypersensitivity to abiotic stress in *Arabidopsis thaliana*. Mol Plant Pathol. **14**, 813-27 (2013).

45 Obermeier, C., Hossain, M. A., Snowdon, R. J., Knüfer, J., von Tiedemann, A. & Friedt, W. Genetic analysis of phenylpropanoid metabolites associated with resistance against *Verticillium longisporum* in *Brassica napus*. Mol. Breeding **31**, 347–361 (2013).

46 Zeise, K. & von Tiedemann, A. Host specialization among vegetative compatibility groups of *Verticillium dahliae* in relation to *Verticillium longisporum*. J. Phytopathology **150**, 112–119 (2002).

47 Novakazi, F., Inderbitzin, P. Sandoya, G., Hayes, R. J., von Tiedemann, A. & Subbarao, K. V. The three lineages of the diploid hybrid *Verticillium longisporum* differ in virulence and pathogenicity. Phytopathology **105**, 662–673 (2015).

48 Eynck, C., Koopmann, B., Karlovsky, P. & von Tiedemann, A. Internal resistance in winter oilseed rape inhibits systemic spread of the vascular *pathogen Verticillium longisporum*. Phytopathology **99**, 802-811 (2009).

49 Kawakatsu, T. et al. Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. Cell **166**, 492-505 (2016).

50 Doyle, J. J. & Doyle, J. L. Isolation of plant DNA from fresh tissue. Focus **12**, 13-15 (1990).

51 Wu, Y., Bhat, P. R., Close, T. J. & Lonardi, S. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet. **4**:e1000212 (2008).

52 Joehanes, R. & Nelson, J. C. QGene 4.0, an extensible Java QTL-analysis platform. Bioinformatics **24**, 2788–2789 (2008).

53 Aulchenko, Y.S., Koning, D. J. & Haley, C. Genome wide rapid association using mixed model and regression: a fast and simple method for genome wide pedigree-based quantitative trait loci association analysis. Genetics **177**, 577-585 (2007).

54 Stich, B., Möhring, J., Piepho, H. P., Heckenberger, M., Buckler, E. S. & Melchinger, A. E. Comparison of mixed-model approaches for association mapping. Genetics **178**, 1745–1754

(2008).

55 Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. **38**, 904-909 (2006).

56 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B **57**, 289–300 (1995).

57 Gabriel, S. B. et al. The structure of haplotype blocks in the human genome. Science **296**, 2225-2229 (2002).

58 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics **21**, 263-265 (2005).

59 Shin, J. H., Blay, S., McNeney, B. & Graham J. LDheatmap: An R function for graphical display of pairwise linkage disequilibria between Single Nucleotide Polymorphisms. Journal of Statistical Software **16**, 1–9 (2006).

60 Stein, A. et al. Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. Plant Biotechnol. J. **15**, 1478–1489 (2017).

61 Alexa, A., Rahnenführer, J. & Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics **22**, 1600–1607 (2006).

476

477 **Figure Legends**

478 **Figure 1.** Genetic and physical localisation of biallelic SNP and presence/absence SNaP markers in
479 relation to two *V. longisporum* resistance QTL on chromosome A03 in the population ExR53-DH and
480 its parents. (a) Genetic linkage maps showing positions of biallelic SNP (black) and SNaP (red)
481 markers. Green lines connect consensus markers between the different map versions. (b) Positions of
482 SNP probe sequences anchored by BLASTn to the Darmor-*bzh* reference sequence. (c) Optical
483 Bionano genome maps (blue) of the two parental lines Express 617 and R53 aligned to the Darmor-*bzh*
484 reference sequence (green). The pink lines connect marker positions flanking QTL regions in the
485 SNaP map, the physical map and the optical maps (c). No resistance QTL were detected using the
486 genetic SNP map.

487

488 **Figure 2.** Comparison of QTL positions (in Mbp anchored to Darmor-*bzh*) obtained by QTL mapping
489 in the biparental population ExR53-DH using maps produced using only SNP markers (light green,
490 hatched) and SNP plus SNaP markers (dark green, hatched), and by GWAS in a NAM panel with 5
491 subpopulations using maps produced using only SNP markers (light blue, solid) and SNP plus SNaP
492 markers (dark blue, solid). Red boxes indicate QTL regions overlapping in biparental QTL and GWAS
493 multi-parental mapping.

494

Figure 3. Comparison of SNP/SNaP marker polymorphism, sequence coverage and gene content at the co-localizing QTL CoLOC1 on *B. napus* chromosome A03. (a) Positions of QTL for *V. longisporum* resistance in the NAM panel (purple) and in the biparental mapping population ExR53-DH (green). (b) Chromosome interval and annotated genes in the QTL region in the *B. napus* Darmor-*bzh* v4.1 reference genome. (c) Allele patterns of reference-anchored markers in parents of six NAM subpopulations along with resequencing coverage data in the six parents. Red boxes indicate confirmed segmental deletions involving genes. Failed alleles represent SNaP (absence) alleles in susceptible parents.

Supplementary Figure Legends

Supplementary Figure S1. Comparison of genome-anchored 60k Illumina Infinium array polymorphic marker data (a) (4 and 12 replicated samples) with genome-anchored Optical Mapping Bionano data for parental genotypes Express617 and R53 for a 3.4 Mbp section from the end of chromosome C01. Consecutively anchored failed calls (grey boxes) in (a) and missing optical mapping molecule alignment (no blue bars) in (b) indicate long-range deletions (absence) in Express617 or R53.

Supplementary Figure S2. GWAS for *V. longisporum* resistance in NAM population comparing SNP map with SNaP map.

Supplementary Figure S3. Comparison of (a) failed SNaP calls in array genotyping data in red with (b) whole genome sequence coverage data analyses marking deleted regions (red), and homeologous exchanges with duplications in the A genome (light blue) and homeologous exchanges with duplications in the C genome (dark blue) for 6 NAM parental lines.

Table Legends

Table 1. Comparison of QTL locations detected by biparental QTL mapping in ExR53-DH and by GWAS in a NAM panel using SNP and SNaP markers.

Supplementary Table Legends

Supplementary Table S1. Comparison of QTL for *V. longisporum* resistance detected by using a genetic map from ExR53-DH population produced from Single Nulceotide Polymorphism (SNP) markers and a genetic map produced by using additional Single Nucleotide absence Polymorphism (SNaP) markers.

Supplementary Table S2. SNP and SNaP markers associated with *V. longisporum* resistance in GWAS for a NAM panel.

Supplementary Table S3. Co-localising QTL regions between the ExR53-DH and NAM panels.

Supplementary Table S4. Gene Ontology enrichment analysis using topGO for genes harboured in QTL regions in the biparental (ExR53-DH) and multi-parental (NAM) populations and for the overlapping QTL regions for both populations.

Supplementary Table S5. Gene annotation for QTL intervals co-localizing for biparental QTL mapping in ExR53-DH population and multi-parental GWAS QTL mapping in NAM population.

Supplementary Table S6. Raw data accquisitaion from Bionano optical genome mapping for two parents of mapping population ExR53-DH.

Supplementary Table S7. Comparison of SNP/SNaP marker patters in ExR53-DH with BioNano optical Mapping coverage in parents for *V. longisporum* QTL regions.

Supplementary Table S8. Comparison of SNP/SNaP marker patters in NAM population with whole genome sequencing (WGS) coverage in parents for *V. longisporum* QTL regions.

Supplementary Table S9. Comparison of SNP/SNaP marker patters in ExR53-DH with WGS coverage in parents for *V. longisporum* QTL regions.

Supplementary Table S10: PAV and Homoelougous exchanges (HE) analysis on WGS coverage data of parental lines for genes harboured in QTL regions in the biparental (ExR53-DH) and multiparental (NAM) populations and for the overlapping QTL regions for both populations

548 **Tables**

549

550 **Table 1.** Comparison of QTL locations detected by biparental QTL mapping in ExR53-DH and by GWAS in a NAM panel using SNP and SNaP markers.

chromosome	QTL ID in biparental population	Start position of QTL interval in ExR53-DH	Stop position of QTL in ExR53-DH	Size of QTL interval (bp)	QTL ID in NAM population	Start position of QTL LD block in NAM population	Stop position of QTL LD block in NAM population	Size of LD block (bp)
chrA03	q23k-BP-A3-1	7,963,059	11,419,476	3,456,417	q22k-NAM-A3-2	10,075,388	10,458,202	382,814
chrA05	q23k-BP-A5-2	2,357,535	2,473,365	115,830	q22k-NAM-A5	2,384,153	2,808,636	424,483
	q23k-BP-C5-1	3,670,200	3,672,842	2,642				
	and	and	and	and				
chrC05	q23k-BP-C5-2	3,688,115	3,949,617	261,502	q22k-NAM-C5-2	3,089,132	3,698,279	609,147
chrC08	q23k-BP-C8	801,925	13,488,675	12,686,750	q22k-NAM-C8	12,201,749	12,596,542	394,793

Figures

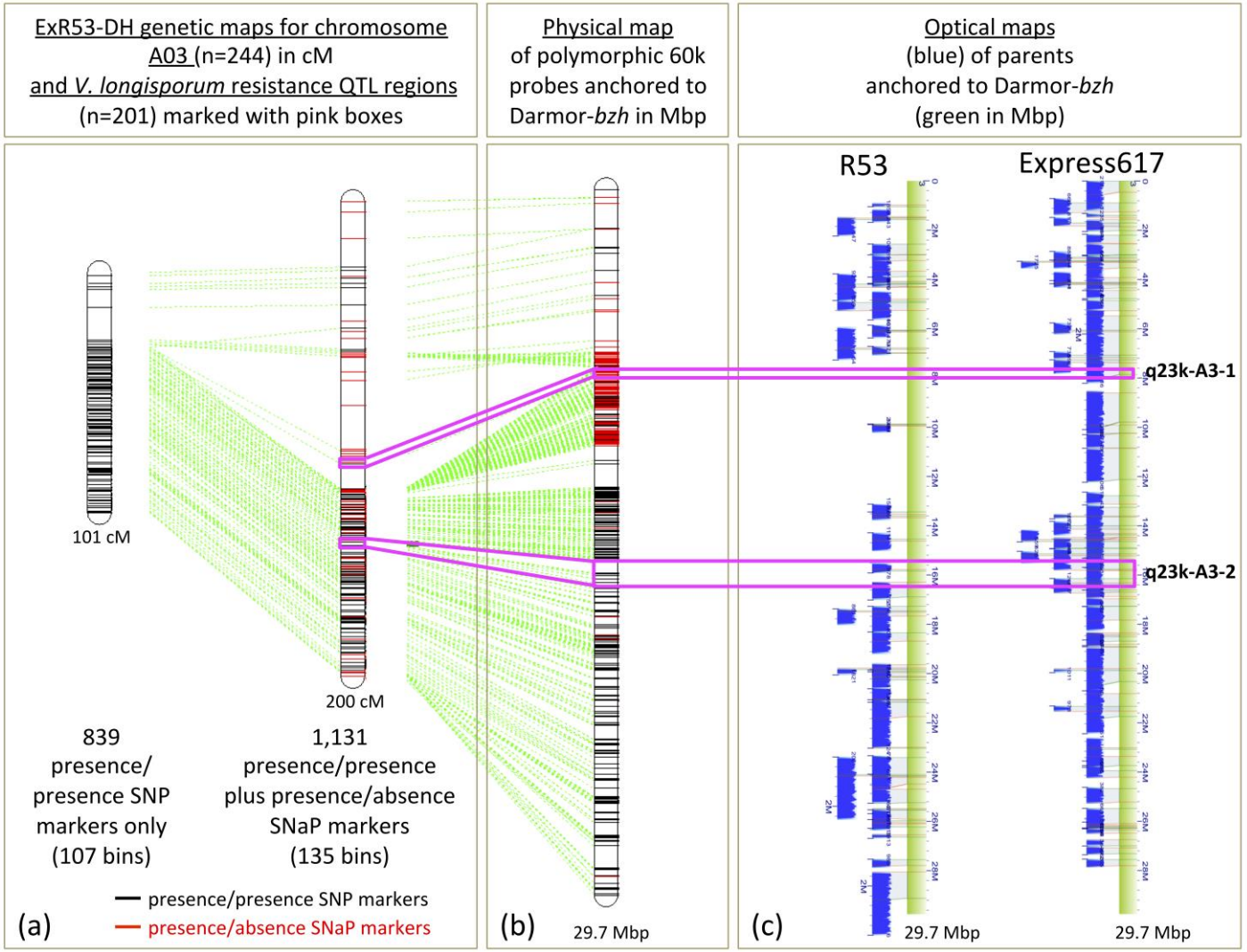

ExR53-DH genetic maps for chromosome A03 (n=244) in cM
and V. longisporum resistance QTL regions
(n=201) marked with pink boxes
Physical map
of polymorphic 60k
probes anchored to
Darmor-bzh in Mbp
Optical maps
(blue) of parents
anchored to Darmor-bzh
(green in Mbp)
R53
Express617
101 cM
200 cM
839
presence/
presence SNP
markers only
(107 bins)
1,131
presence/presence
plus presence/absence
SNaP markers
(135 bins)
presence/presence SNP markers
presence/absence SNaP markers
(a)
(b)
29.7 Mbp
(c)
29.7 Mbp
29.7 Mbp
q23k-A3-1
q23k-A3-2

Figure 1. Genetic and physical localisation of biallelic SNP and presence/absence SNaP markers in relation to two *V. longisporum* resistance QTL on chromosome A03 in the population ExR53-DH and its parents. (a) Genetic linkage maps showing positions of biallelic SNP (black) and SNaP (red) markers. Green lines connect consensus markers between the different map versions. (b) Positions of SNP probe sequences anchored by BLASTn to the Darmor-*bzh* reference sequence. (c) Optical Bionano genome maps (blue) of the two parental lines Express 617 and R53 aligned to the Darmor-*bzh* reference sequence (green). The pink lines connect marker positions flanking QTL regions in the SNaP map, the physical map and the optical maps (c). No resistance QTL were detected using the genetic SNP map.

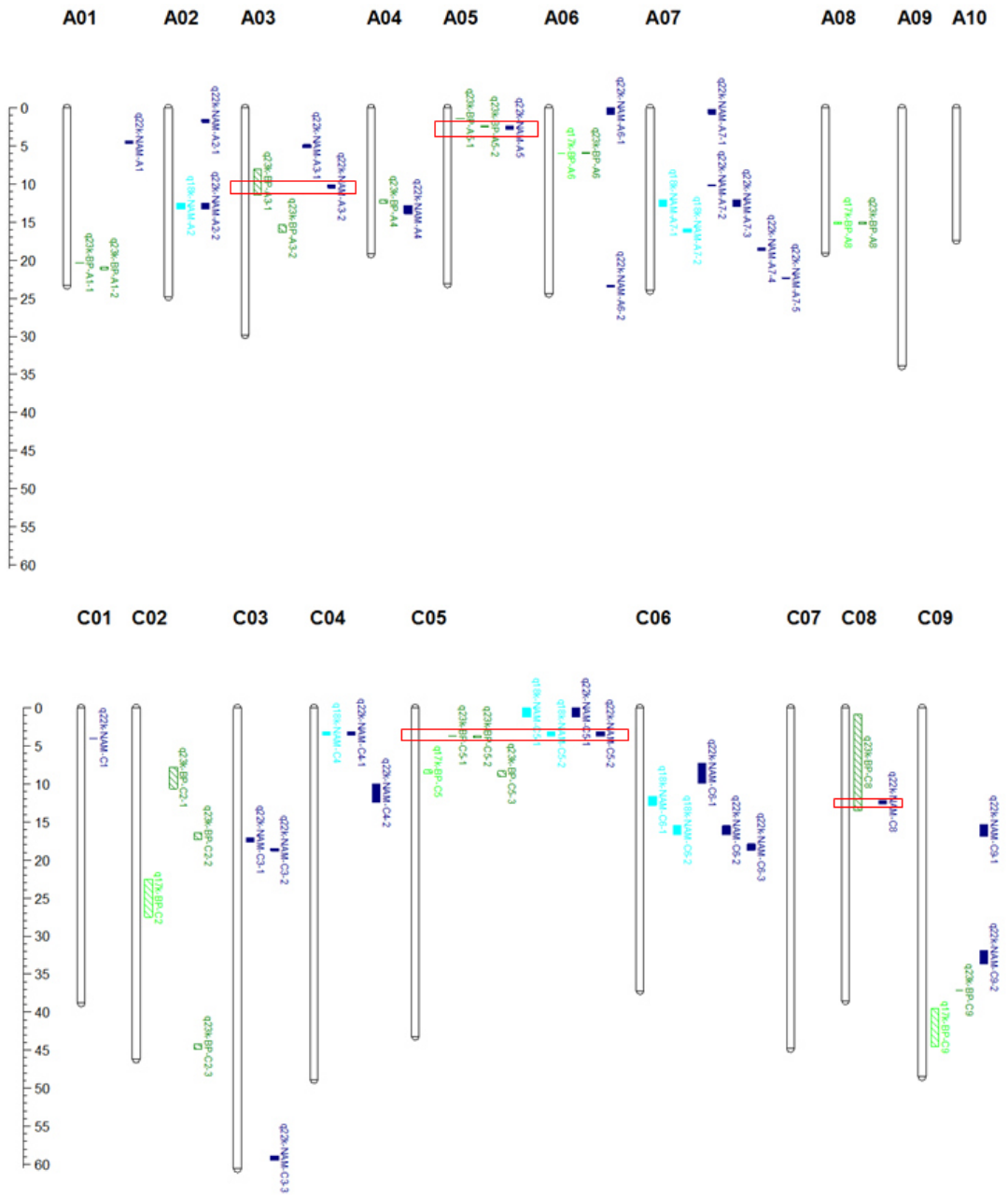
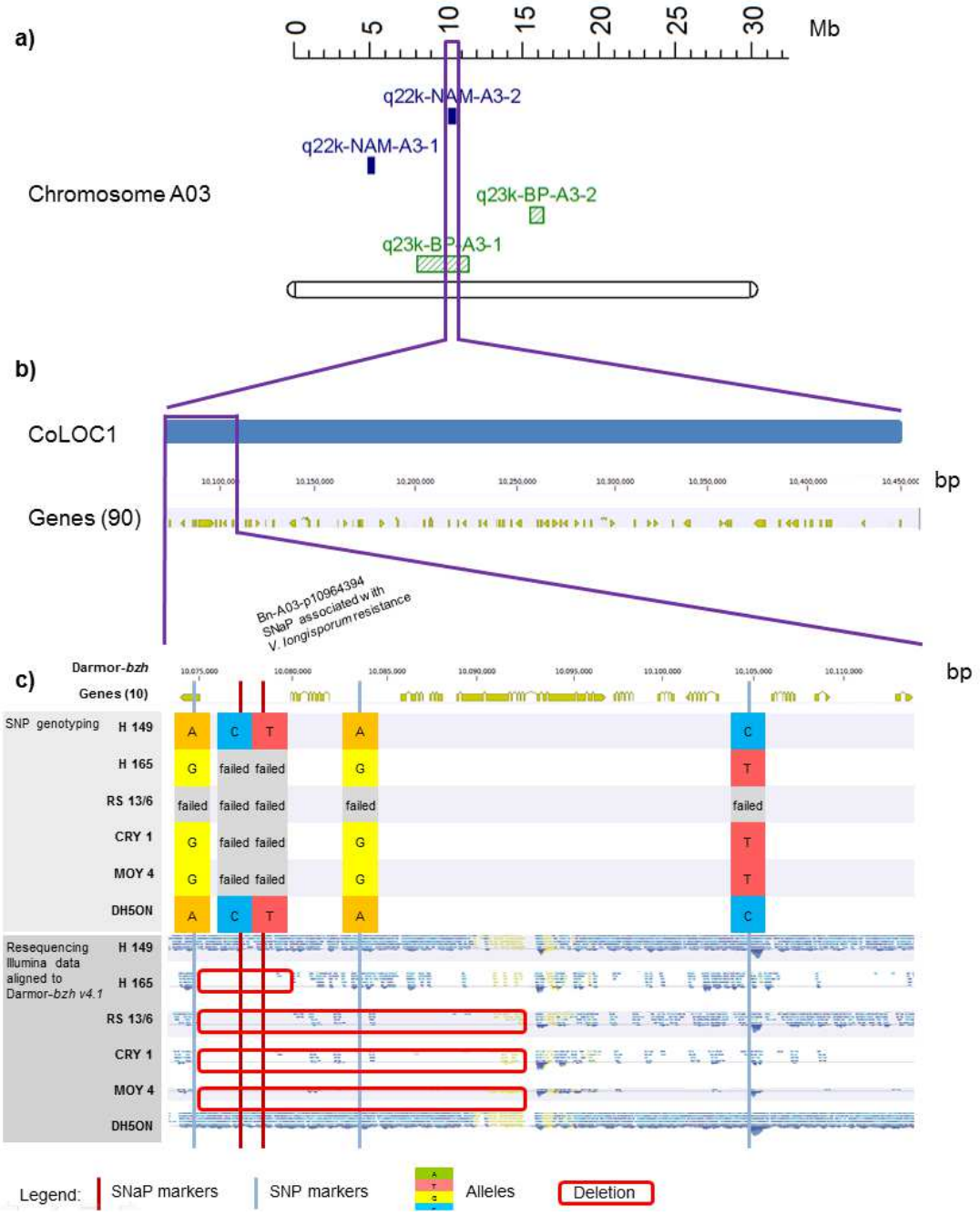


Figure 2. Comparison of QTL positions (in Mbp anchored to Darmor-*bzh*) obtained by QTL mapping in the biparental population ExR53-DH using maps produced using only SNP markers (light green, hatched) and SNP plus SNaP markers (dark green, hatched), and by GWAS in a NAM panel with 5 subpopulations using maps produced using only SNP markers (light blue, solid) and SNP plus SNaP markers (dark blue, solid). Red boxes indicate QTL regions overlapping in biparental QTL and GWAS multi-parental mapping.



564

565 **Figure 3.** Comparison of SNP/SNaP marker polymorphism, sequence coverage and gene content at
566 the co-localizing QTL CoLOC1 on *B. napus* chromosome A03. (a) Positions of QTL for *V.*
567 *longisporum* resistance in the NAM panel (purple) and in the biparental mapping population ExR53-
568 DH (green). (b) Chromosome interval and annotated genes in the QTL region in the *B. napus* Darmor-
569 *bzh* v4.1 reference genome. (c) Allele patterns of reference-anchored markers in parents of six NAM
570 subpopulations along with resequencing coverage data in the six parents. Red boxes indicate
571 confirmed segmental deletions involving genes. Failed alleles represent SNaP (absence) alleles in
572 susceptible parents

573

5 Discussion

5.1 Evaluating structural variation in oilseed rape

In the last few decades, molecular genetics and genome analysis techniques have played an important role in detection of genome structural variations and demonstrated that breeding efforts should consider these types of variations. The recent development of simple, cost-efficient, high-throughput SNP marker assays, new sequencing technologies, high-density genetic and physical mapping resources for major crop species today enables direct linkage of SV to important agricultural crop traits. More detailed studies will likely lead to the discovery of many genomic regions involved in important traits and improve our understanding of the genetics, inheritance and variation for small, medium and large-scale genome modifications in crops. Already by using these new tools, a number of recent studies have demonstrated that SV within genes are associated with selected gene families and important traits in plants, and particularly in some polyploid crops.

This thesis hypothesized that presence/absence variation (PAV) – particularly involving genes – is more widespread in the allopolyploid genome of *Brassica napus* (oilseed rape) than previously assumed (Chapter 2), and that it affects a large proportion of genes, across the whole genome, with a broad impact on agronomical important traits. As an example for an important polygenic trait complex, quantitative disease resistances against three major fungal pathogens were associated with SV in populations of structurally rearranged *B. napus* breeding lines. In chapter 3, a simple strategy was developed to extract valuable SV information from high-density SNP hybridization array data, by distinguishing failed SNP calls associated with genuine presence/absence variants from random technical failures of SNP assays and using them for QTL mapping. Moreover, to demonstrate that genome PAV commonly extends to long-range gene PAV, complementary technologies (Optical mapping

and genome-wide resequencing coverage analysis) were used for parental lines of two very diverse mapping populations (Chapter 4).

Whole-genome duplication (WGD) or polyploidization determines small, medium or long-range structural variation in plant genomes and is followed by gene diversification and specialisation. This is a process that may shape the evolutionary history of many wild and cultivated plant species which undergo artificial and/or natural selection (Panchy et al., 2016). Although it is known that polyploidization is a major driver of gene diversification, genes in plant species can also be duplicated by other mechanisms, such as tandem duplication, transposon-mediated duplication or segmental duplication. In the literature, different terms have been used frequently to define various genomic duplications and structural variations bigger that 1kb in size (Saxena et al., 2014). Extensive genome structural variation is considered in some publications as a widespread phenomenon and its full extent is not always completely revealed by pangenome studies (Dolatabadian et al., 2017). The recent allopolyploid crop species *B. napus* shows a strong abundance of SV and genomic rearrangements (Samans et al., 2017; Higgins et al., 2018). Nowadays, scientists have the ability to extract presence/absence or copy-number polymorphic marker calls using cost-effective genotyping array data (Grandke et al., 2017), or whole-genome sequencing data. Such methods were used in high-throughput discovery of trait-associated gene presence-absence variations of selected gene families in other polyploid crops like wheat, which is known to exhibit structural variations for a number of important agricultural traits (Nishida et al., 2013; Würschum et al., 2015, 2017).

The newest technological breakthroughs in whole-genome sequencing are still comparatively costly, but genome assemblies and resequencing have already contributed to substantial advances in crop genetics and breeding. The availability of reference genome sequences from

many plant species enables new insights into the genetic architecture and abundance of SV within and across species. The use of high-throughput SNP marker assays for detection of structural variations in complex plant genomes represents a cost-efficient solution to SV analysis in large populations. In this thesis, chapter 3 demonstrated that biotic stress tolerance is commonly associated with inheritable single nucleotide presence/absence polymorphism (SNaP) in segregating bi- or multi-parental breeding populations. Instead of discarding SNP array data due to excessive missing data, these markers could be recovered from the dataset using an analysis of the segregation patters of a missing allele coming from a parental line. Subsequently, the use of these markers in genetic mapping increased the detection of regions under directional selection for disease resistance. Results indicate that some disease resistance QTL are located in chromosome regions undetectable by SNP markers alone and will thus remain undiscovered in traditional SNP array-based genome wide studies. Similar results were obtained for bi- and multi-parental populations using markers anchored to the *B. napus* Darmor-*bzh* reference genome (Chalhoub et al., 2014). In general, additional detected QTL regions contained either few or only SNaP markers which were anchored in consecutive order in Darmor-*bzh*, suggesting that these QTL are genuinely in deleted genome regions affected by medium to long-range PAV and have a significant effect on quantitative disease resistance mapping (Mason et al., 2017). Whole genome sequencing data for single *B. napus* genotypes has been used to describe similar large-scale presence-absence events (Chalhoub et al., 2014; Samans et al., 2017; Hurgobin et al., 2017), which occur commonly in the allopolyploid *B. napus* genome due to a high incidence of homoeologous exchanges following allopolyploidization between the the A and C diploid progenitor subgenomes (Chalhoub et al., 2014).

Detailed analysis of raw Brassica 60K Illumina Infinium™ SNP genotyping array data for polymorphic SNP probes displaying one allele as a ‘failed’ call revealed a segregation pattern

in the ExR53-DH and the NAM mapping populations. This segregation patterns corresponds to a presence/absence polymorphism. More than 10,005 SNP markers were removed during this standard quality filtering for low frequency of failed calls and minor allele. Among the excluded SNPs, presence-absence segregation patterns were consistent from a number of 3,627 SNaPs in the NAM panel and 4,657 SNaPs in the ExR53 population. All markers identified using segregation patters were recovered and included in further analysis. For the bi-parental population, two distinct genetic maps were produced, one using only 17,626 SNP markers (2,176 bins) and one using 22,283 SNP and SNaP markers (2,714 bins), respectively. Comparative analysis of the two maps revealed that large chromosomal regions were not covered in the genetic map that was produces only with SNP markers, leading to a higher map resolution and identification of QTL not previously detected. These undiscovered chromosome regions may contain important genes with a high influence on quantitative disease resistance.

5.2 Linking structural variation with disease resistance QTL in oilseed rape

Most important agronomical traits show a complex, quantitative inheritance. This limits the effectiveness of marker-assisted selection applications in breeding programs where identification of major genes and markers involved in complex traits is crucial.

Linkage-based mapping of QTL regions and genome-wide association studies (GWAS) are complementary approaches commonly used for deciphering quantitative traits. In this study, multi-environment disease screening of three major fungal pathogens of oilseed rape was done in a bi-parental segregating population derived from the elite cultivar Express and the resynthesized genotype R53 (ExR53; Obermeier et al., 2013) and in a segregating NAM population from crosses of a common elite *B. napus* parental line, DH5ON, with five

synthetic oilseed rape recurrent parental genotypes (Snowdon et al., 2015). Available phenotyping data was used to obtain a wide overview of the genomic regions involved in qualitative and quantitative disease resistance.

In the ExR53 population, QTL mapping using genetic maps with SNaP markers increased the total number of detectable QTL from 5 to 17, indicating an improved QTL detection power. Moreover, genetic map resolution was influenced positively. Furthermore, a small number of detected QTL contained exclusively SNaP markers, while others contained only SNP markers. Interestingly, a large number of QTL intervals were spanning on large genome areas and contained marker clusters with both SNP and SNaP markers.

Genome-wide association studies for disease resistance was performed in the NAM panel using (1) only SNP markers, and (2) SNP plus SNaP markers that were physically anchored on the *B. napus* Darmor-*bzh* reference genome (Chalhoub et al., 2014; Mason et al., 2017). Comparative GWAS revealed a medium number of SNP-trait associations, whereas after addition of SNaP markers the total number of significant associations increased considerably. For example, for Verticillium stem striping the total number significant markers increased from 18 to 41, while for blackleg the number increased from 65 to 153 and for Sclerotiania stem rot the number increased from 47 to 65. Surprisingly, new QTL for disease resistance were also identified, suggesting that genomic regions influencing quantitative disease resistance are affected by SV in oilseed rape.

Short-range Illumina and long-range BioNano optical mapping sequencing data combined with Illumina Infinium hybridization array data provided new insights into the importance of SV for disease resistance. Collectively these findings indicate that SV is common and a prevalent phenomenon of qualitative and quantitative disease resistance in the recent allopolyploid *B. napus* genome.

Intra-species gene presence-absence variation been reported in many plant species, but it is still unclear what kind of genes and agronomic important traits are predominantly affected. In polyploid crop genomes, pangenomic studies in large genotype collections are an important tool for comprehensive genome-wide gene PAV analysis. Detailed reports were published both for model species (Tan et al., 2012) and for crop species with relative small genomes, such as rice (Zhao et al., 2018). In crops like rapeseed with more complex genomes, large populations were analysed for selected resistance copy-number gene variations (Dolatabadian et al., 2017) or for targeted sequence capture studies (Schiessl et al., 2017a,b), Beló et al. (2010) used comparative genomic hybridization arrays in maize for detailed genome-wide SV analysis, while Darracq et al. (2018) discovered CNV between 100 analyzed genotypes across more than 90% of the maize genome. For oilseed rape, Hurgobin et al. (2018) reported that 38% of all genes were affected by presence-absence variation in a diversity set of 53 genotypes. SV was reported to have particular relevance for disease resistance, but also in important agronomical traits. Structural variation associated with agronomical important traits of oilseed rape has been suggested previously using data from the Brassica Illumina 60K SNP array (Qian et al., 2016; Mason et al., 2017; Stein et al., 2017).

Disease resistance QTL detected in this study correspond to some previously identified and important genomic regions known to confer resistance. In literature, blackleg studies using bi-parental crosses and diverse germplasm sets have identified a major loci with qualitative inheritance (Raman et al., 2012; Delourme et al., 2014; Larkan et al., 2014; Kumar et al., 2018) as well as multiple QTL involved in disease resistance (Delourme et al., 2006; Jestin et al., 2011; Fopa Fomeju et al., 2014; Fomeju et al., 2015; Larkan et al., 2016; Kumar et al., 2018). However, direct comparison of genomic regions with the QTL identified in this study is difficult, as previously published articles used different marker types for genetic mapping. From around 50 blackleg resistance QTL reported, only 7 coincide with previously mapped

major *Rlm* and quantitative resistance loci anchoring to the reference genome Darmor-*bzh* (Larkan et al., 2014, 2016; Raman et al., 2012, 2016; Luo et al., 2017). Sclerotinia stem rot partial resistance has been mapped in very diverse Chinese semi-winter oilseed rape genotypes (Zhao et al., 2006) and spring-type oilseed rape genotypes from China and Australia (Li et al., 2015). Published QTL mapping results reported many disease resistance loci responsible for partial resistance to Sclerotinia stem rot in numerous DH populations (Zhao et al., 2006; Wu et al., 2013; Wei et al., 2014; Luo et al., 2017) and in a diversity panel (Wei et al., 2016; Wu et al., 2016). In total, only 7 QTL (out of 57 QTL) were identical between the present study and these previous mapping studies. In contrast, a major QTL on *B. napus* chromosome C05 for Verticillium stem striping resistance (Rygulla et al., 2008; Obermeier et al., 2013) could be confirmed in the present study.

In case of gene-for-gene interaction, plant disease resistance is determined by resistance (R) genes that trigger defence responses when an avirulence (Avr) protein from the invading pathogen is detected. This interaction can be disrupted by mutations that occur either in plant R gene or in the fungal Avr gene (Bonas and Lahaye, 2002). Presence-absence polymorphisms associated with R genes are considered common and may affect about 20% of R-genes in the model (Arabidopsis) and cultivated plant species (rice) (Shen et al., 2006). Moreover, when plants that carry race-specific R-genes are grown in the absence of pathogens, a fitness penalty of up to 10% has been reported (Tian et al., 2003; MacQueen & Bergelson, 2016). Especially for nucleotide binding site-leucine-rich repeat (NLR) resistance genes, numerous studies have shown that fitness costs can lead to multiplication and deletion of gene family members, particularly in *A. thaliana* (Tan et al., 2012; Borrelli et al., 2018) but also for oilseed rape (Chalhoub et al., 2014; Samans et al., 2017). Thus, deletion of R genes can be considered as a common feature of crop disease resistance. The findings in the present studies are consistent with the results in the literature, as the majority of failed SNaP alleles

correspond to gene loss of function and susceptibility to pathogens. In contrast, however, for Sclerotinia stem rot, 6% of failed marker calls were associated with resistance to fungal attacks, indicating a deletion of a plant factor that allowed the pathogen to infect. Uppalapati et al. (2012) reported that a homozygous mutations of the *irg1* gene (*inhibitor of rust germ tube differentiation 1*) determinates a less efficient development of germ tubes of the Asian soybean rust and two other fungal pathogens on the leaves of *Medicago truncatula*. The *irg1* gene is involved in wax accumulation on the leaf surface. Generally, association of deletions with resistance is considered a rare phenomenon, while a loss of function is typically recessive and in heterozygous genotypes resistance can be masked by the dominant allele or by a dosage effect. Two clusters of TIR-NBS-LRR genes were deleted from this QTL region in natural *B. napus* compared to synthetic *B. napus* accessions (Samans et al., 2017). The same QTL region on chromosome C09 contains typical R genes affected by structural variations. Moreover, results clearly show that gene PAV events within resistance QTL regions are not limited to clusters of typical R genes, but are widespread and affect very diverse gene families. By detecting deletions associated with susceptibility or resistance for blackleg disease, Sclerotinia stem rot and Verticillium stem striping resistance, this study demonstrated the usefulness of SNaP markers to map genomic regions and QTL that were previously not detected. Analyses of gene PAV associated with resistance QTL is another promising new approach to deciphering the genetic basis of quantitative resistances in oilseed rape. Hence, inclusion of SV studies in traditional genetic mapping, QTL analyses and breeding could prove to be of great interest for oilseed rape breeders. Offspring of resynthesized *B. napus* with high incidence of SV may therefore have an increased potential for use in disease resistance breeding of modern oilseed cultivars.

5.3 Conclusions and future prospects

Overall, this thesis describes the effect of genome-wide structural variation on qualitative and quantitative disease resistance in oilseed rape. Identifying the extent and impact of SV is becoming more feasible in crop genomes with ongoing advances in high-throughput genotyping and genome sequencing technologies. Corresponding tools are today available for all major crop plants, enabling cheap and ultra-fast analysis of whole-genome SV and marker profiling in large populations and generation of accurate, long sequence reads on a genome-wide scale. Within the framework of this dissertation, it is shown that SNP genotyping data could be used efficiently to map genomic regions involved in disease resistance of oilseed rape to three major fungal pathogens. Confirmation that gene presence/absence variation strongly affects genes within confidence intervals for major fungal disease resistance in oilseed rape, and use of presence/absence markers in QTL mapping facilitated immediate resolution of important QTL to just a few candidate genes. Identified major disease resistance QTL and candidate genes can be potentially used by breeders to reinstate genetic resistance into European elite cultivars. Furthermore, loci responsible for multiple disease resistances from resynthesized *B. napus* lines could provide an opportunity to increase the diversity in the oilseed rape gene pool, while maintaining desired characteristics. Candidate genes for multi-disease resistance to major fungal pathogens represent ideal targets for further functional validations.

In the near future our insight into the prevalence of SV on important agronomic traits will continue to grow. Understanding the contribution of gene deletions/duplications to traits will be an important factor in improving the accuracy and efficacy of many new genetic technologies in plant breeding. This thesis opens new opportunities to learn from adaptive evolution of a recent polyploid crop species in order to improve crop resilience against biotic

constraints. Harnessing and exploiting SV may be a key component in future breeding of sustainable, resilient crops, and can advance our understanding of the inheritance of complex quantitative traits in polyploid crops. The thesis findings contribute unique and valuable examples of how structural variation influences agronomic traits in a major crop and how this phenomenon can be incorporated in molecular plant breeding.

6 Summary

Brassica napus L. (oilseed rape/canola) originated through spontaneous interspecific hybridisation between turnip rape (*Brassica rapa* L., syn. *campestris*; genome AA, 2n = 20) and cabbage (*Brassica oleracea* L.; genome CC, 2n = 18), resulting in an amphidiploid genome comprising the full chromosome complements of its two progenitors. Because no wild *B. napus* forms are known, it is assumed that the species arose relatively recently, when the parental species began being cultivated in geographical proximity due to anthropogenic influences. The occurrence of spontaneous chromosome doubling in crosses among closely related Brassica diploid species and the high homoeology between the diploid progenitor genomes (A and C subgenomes) have led to extensive structural genome variation within the crop genome. New genotyping methods that can generate large datasets from individual genotypes are available nowadays also for oilseed rape. Technologies like the 60k Illumina SNP array, short-read next-generation sequencing and long-read 3rd generation sequencing can provide valuable, extremely detailed insight into genome structural rearrangements generated during interspecific hybridisation between related diploid species. The impact of genome wide structural variations on plant phenotypes is of crucial importance to scientists and plant breeders in order to optimize parental crosses with maximal genetic gain.

To address this issue, the thesis describes effects of small and long range structural variation on three major fungal pathogens of oilseed rape. A diverse nested association mapping (NAM) population of 200 lines, representing crosses among five resynthesized *B. napus* accessions with a common elite *B. napus* donor, was genotyped using the 60K single-nucleotide-polymorphism (SNP) Infinium array. Phenotypic evaluations for disease resistance to Phoma stem canker/blackleg (*Leptosphaeria maculans*), Sclerotinia stem rot (*Sclerotinia sclerotiorum*) and Verticillium stem striping (*Verticillium longisporum*) were performed in greenhouse experiments and field trails across France and Germany in order to identify qualitative and quantitative resistances. After analysis on population structure, detailed

measurements of genetic diversity and linkage disequilibrium, a genome-wide association study (GWAS) was performed for each disease. Major QTL regions were identified on various chromosomes that were highly associated to disease resistance. Additionally, the presence of co-localising QTL regions among the three studied pathogens was observed, indicating putative multiple-disease resistance mechanisms.

The key finding of the dissertation is a novel strategy to recover valuable information from single nucleotide absence polymorphisms (SNaP) by population-based quality filtering of SNP hybridization data to distinguish patterns associated with genuine deletions from those caused by technical failures. Standard data quality filtering approaches can remove large numbers of potentially useful marker information that can mask QTL caused by structural variations. I reveal that including SNaP markers in genetic analyses elucidate segregation of small to large-scale structural variants in nested association mapping populations of oilseed rape. Including SNaP markers in genome wide association studies identified numerous new quantitative trait loci, invisible using SNP markers alone, for resistance to three major fungal diseases of oilseed rape. The thesis results indicate that presence-absence variation has a strong influence on quantitative disease resistance in *B. napus* and that SNaP analysis using cost-effective SNP array data can provide extensive added value from 'missing data'. For blackleg disease, 50 QTL were mapped in this study using SNP and SNaP markers. Partial resistance to Sclerotinia stem rot was found, and 37 QTL were mapped using SNP and SNaP markers. From these, only seven were previously reported in the literature. After adding SNaP markers, a 1.6 -to 3.5-fold increase in detected QTL regions associated with blackleg and Sclerotinia stem rot resistance was observed, corresponding to a total of 57 new QTL detected. Most of the additional QTL for *V. longisporum* resistance detected in the NAM panel contained only SNaP markers (18 out of 28 QTL), suggesting that regions affected by long-range presence/absence variation strongly affect quantitative disease resistance. Re-

analysis and integration of SNP array data, short-range next generation sequencing data and long-range BioNano optical mapping data with QTL data provided new insights into the importance of structural variations for disease resistance expression against major fungal pathogens of oilseed rape.

The frequent localization of new QTL in regions affected by structural genome variations confirms the hypothesis that copy-number and presence-absence variations have particular relevance for disease resistance. Offspring of resynthesized *B. napus* with high rates of presence–absence and other structural variations may therefore have an increased potential for use in resistance breeding of *B. napus*. This strategy might also be applicable for improving the precision of genetic mapping in many important crop species.

7 References

Ayana GT, Ali S, Sidhu JS, Gonzalez Hernandez JL, Turnipseed B, Sehgal SK (2018) Genome-Wide Association Study for Spot Blotch Resistance in Hard Winter Wheat. Front Plant Sci. 2018 9:926.

Axelsson T, Bowman CM, Sharpe AG, Lydiate DJ, Lagercrantz U (2000) Amphidiploid *Brassica juncea* contains conserved progenitor genomes. Genome 43: 679–688.

Balesdent MH, Fudal I, Ollivier B, Bally P, Grandaubert J, Eber F, Chèvre AM, Leflon M, Rouxel T (2013) The dispensable chromosome of *Leptosphaeria maculans* shelters an effector gene conferring avirulence towards *Brassica rapa*. New Phytol 198:887–898

Barbetti MJ (1976) The role of pycniospores of *Leptosphaeria maculans* in the spread of blackleg disease in rape. Australian Journal of Experimental Agriculture and Animal Husbandry 16: 911–914.

Beló A, Beatty MK, Hondred D, Fengler KA, Li B, Rafalski A (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theor. Appl. Genet. 120:355–367.

Bonas U, Lahaye T (2002) Plant disease resistance triggered by pathogen-derived molecules: refined models of specific recognition. Curr Opin Microbiol 5:44-45.

Bolton MD, Thomma BP, Nelson BD (2006) *Sclerotinia sclerotiorum* (Lib.) de Bary: biology and molecular traits of a cosmopolitan pathogen. Mol Plant Pathol 7:1–16.

Borrelli GM, Mazzucotelli E, Marone D, Crosatti C, Michelotti V, Valè G, Mastrangelo AM (2018) Regulation and Evolution of NLR Genes: A Close Interconnection for Plant Immunity. Int J Mol Sci. 19:6

Bousset L, Ermel M, Lebreton L (2018) The full life cycle of *Leptosphaeria maculans* completed on inoculated oilseed rape incubated under controlled conditions. Plant Pathol, 67: 1321-1328.

Bousset L, Jumel S, Garreta V, Picault H, Soubeyrand S (2015) Transmission of *Leptosphaeria maculans* from a cropping season to the following one. Annals of Applied Biology 166:530–43.

Brun H, Chèvre A-M, Fitt BDL, Powers S, Besnard A-L, Ermel M, Huteau V, Marquer B, Eber F, Renard M, Andrivon D (2010) Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. New Phytol 185:285–299

Bus A, Körber N, Snowdon RJ, Stich B (2011) Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. Theoretical and Applied Genetics 123: 1413–1423.

Butterworth MH, Semenov MA, Barnes A, Moran D, West JS, Fitt BDL (2010) North-south divide; contrasting impacts of climate change on crop yields in Scotland and England. Journal of the Royal Society Interface 7: 123–130.

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

Christen O and Sieling K (1993) The effect of different preceding crops on the development, growth and yield of winter barley. Journal of Agronomy and Crop Science 171: 114–123.

Christen O, Sieling K, Hanus H (1992) The effect of different preceding crops on the development, growth and yield of winter wheat. European Journal of Agronomy 1: 21–28.

Darracq A, Vitte C, Nicolas S, Duarte J, Pichon JP, Mary-Huard T, et al. (2018) Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. BMC Genomics 19:119.

Delourme R, Bousset L, Ermel M, Duffé P, Besnard AL, Marquer B, Fudal Linglin J, Chadoeuf J, Brun H (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

Delourme R, Chevre AM, Brun H, Rouxel T, Balesdent MH, Dias JS, et al. (2006) Major gene and polygenic resistance to *Leptosphaeria maculans* in oilseed rape (*Brassica napus*). Eur J Plant Pathol 114: 41–52

Delourme R, Pilet-Nayel ML, Archipiano M, Horvais R, Tanguy X, Rouxel T, Brun H, Renard A, Balesdent AH (2004) A cluster of major specific resistance genes to *Leptosphaeria maculans* in *Brassica napus*. Phytopathology 94: 578–583

Delourme R, Barbetti M, Snowdon R, Zhao J, Manazanares-Dauleux MJ (2011) Genetics and genomics of disease resistance. In Genetics, Genomics and Breeding of Oilseed Brassicas (Edwards D, Batley J, Parkin IAP. and Kole C, eds), pp. 276–303. Boca Raton, FL: Science Publishers.

Dolatabadian A, Patel DA, Edwards D, Batley J (2017) Copy number variation and disease resistance in plants. Theor Appl Genet 130: 2479-2490.

Dunker S, Keunecke H, Steinbach P, von Tiedemann A (2008) Impact of *Verticillium longisporum* on yield and morphology of winter oilseed rape (*Brassica napus*) in relation to systemic spread in the plant. Journal of Phytopathology 156: 698-707.

Evans N, Baierl A, Semenov MA, Gladders P, Fitt BDL (2008) Range and severity of a plant disease increased by global warming. Journal of the Royal Society Interface 5: 525–531

Evans N, Gladders P, Fitt BDL, Tiedemann AV (2009) Altered distribution and life cycles of major pathogens in Europe. In: Feldmann F, Alford DV, Furk C (eds), Crop Plant Resistance to Biotic and Abiotic Factors. Deutsche Phytomedizinische Gesellschaft Braunschweig, Germany, pp. 302-308.

Eynck C (2008) Identification of resistance sources and characterization of resistance factors in *Brassica* species to *Verticillium longisporum*. Dissertation, Universität Göttingen.

Eynck C, Koopmann B, Karlovsky P, von Tiedemann A (2009) Internal resistance in winter oilseed rape inhibits systemic spread of the vascular pathogen *Verticillium longisporum*. Phytopathology 99: 802-811.

Eynck C, Koopmann B, Grunewaldt-Stöcker G, Karlovsky P, Tiedemann Av (2007) Differential interactions of *Verticillium longisporum* and *Verticillium dahliae* with *Brassica napus* detected with molecular and histological techniques. European Journal of Plant Pathology 118: 259-274.

Fravel DR. and Larkin RP. (1997) Effect of Sublethal Stresses on Microsclerotia of *Verticillium dahliae*. In: Tjamos EC, Rowe RC, Heale JB, Fravel DR. Advances in Verticillium: Research and disease management. Proceedings of the Seventh International Verticillium Symposium, APS Press, St. Paul, Minnesota, pp. 301-306.

Fitt BDL, Brun H, Barbetti MJ, Rimmer SR (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

Fitt BDL, Hu BC, Li ZQ, Liu SY, Lange RM, Kharbanda PD, Butterworth MH, White RP (2008) Strategies to prevent spread of *Leptosphaeria maculans* (phoma stem canker) onto oilseed rape crops in China; costs and benefits. Plant Pathol. 57:652–664.

Fitt BDL, Fraaije BA, Chandramohan P, Shaw MW (2011) Impacts of changing air composition on severity of arable crop disease epidemics. Plant Pathology 60: 44–53.

Friedt W and Snowdon R (2010) Oilseed Rape. In: Oil Crops Vollmann J and Rajcan I, (eds), pp. 91–126. Springer New York, New York, NY.

Fopa Fomeju B., Falentin C, Lassalle G, Manzanares-Dauleux MJ, Delourme R (2014) Homoeologous duplicated regions are involved in quantitative resistance of *Brassica napus* to stem canker. BMC Genom. 15: 498.

Fomeju BF, Falentin C, Lassalle G, Manzanares-Dauleux M J, Delourme R (2015) Comparative genomic analysis of duplicated homoeologous regions involved in the resistance of *Brassica napus* to stem canker. Front. Plant Sci. 6:772.

Garg H, Atri C, Sandhu PS, Kaur B, Renton M, Banga SK, et al. (2010) High level of resistance to *Sclerotinia sclerotiorum* in introgression lines derived from hybridization between wild crucifers and the crop *Brassica* species *B. napus* and *B. juncea*. Field Crops Res. 117:51-58.

Garg H, Sivasithamparam K, Banga SS, Barbetti MJ (2008) Cotyledon assay as a rapid and reliable method of screening for resistance against *Sclerotinia sclerotiorum* in *Brassica napus* genotypes. Australas. Plant Pathol. 37:106–111.

Ge XT, You MP, Barbetti MJ (2015) Virulence differences among *Sclerotinia sclerotiorum* isolates determines host cotyledon resistance responses in *Brassicaceae* genotypes. Eur. J. Plant Pathol. 143:527–541.

Grandke F, Snowdon R, Samans B (2017) gsrc: an R package for genome structure rearrangement calling. Bioinformatics 33:545–546.

Hall R, Chigogora JL, Phillips LG (1996) Role of seedborne inoculum of *Leptosphaeria maculans* in development of blackleg on oilseed rape. Canadian Journal of Plant Pathology 18:35–42.

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

Heale JB and Karapapa KV (1999) The *Verticillium* threat to Canada's major oilseed crop: Canola. Canadian Journal of Plant Pathology 21:1-7.

Higgins EE, Clarke WE, Howell EC, Armstrong SJ, Parkin IAP (2018) Detecting de novo homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array. G3: Genes|Genomes|Genetics, doi:10.1534/g3.118.200118.

Hossain AM (2014) Genetic and Metabolomic Analysis of Phenolic Compounds Associated with *Verticillium longisporum* Resistance in Oilseed Rape (*Brassica napus* L.). Dissertation, Universität Giessen.

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

Hwang S-F, Strelkov SE, Peng G, Ahmed H, Zhou Q, Turnbull G (2016) Blackleg (*Leptosphaeria maculans*) severity and yield loss in Canola in Alberta, Canada. Plants 5:31.

Huang YJ, Li ZQ, Evans N, Rouxel T, Fitt BDL, Balesdent MH (2006) Fitness cost associated with loss of the *AvrLm4* avirulence function in *Leptosphaeria maculans* (phoma stem canker of oilseed rape). European Journal of Plant Pathology 114:77–89.

Huang YJ, Pirie EJ, Evans N, Delourme R, King GJ, Fitt BDL (2009) Quantitative resistance to symptomless growth of *Leptosphaeria maculans* (phoma stem canker) in *Brassica napus* (oilseed rape). Plant Pathology 58:314–23.

Hurgobin B, Golicz AA, Bayer PE, Chan, CK, Tirnaz S, Dolatabadian A, et al. (2018) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. Plant Biotech J, doi: 10.1111/pbi.12791

Jestin C, Bardol N, Lodé M, Duffé P, Domin C, Vallée P, Mangin B, Manzanares-Dauleux MJ, Delourme R (2015) Connected populations for detecting quantitative resistance factors to phoma stem canker in oilseed rape (*Brassica napus* L.). Mol Breed 35:1–16.

Jestin C, Lodé M, Vallée P, Domin C, Falentin C, Horvais R, et al. (2011) Association mapping of quantitative resistance for *Leptosphaeria maculans* in oilseed rape (*Brassica napus* L.). Mol. Breed. 27:271–287.

Juska A, Busch L, Tanaka K (1997) The blackleg epidemic in Canadian rapeseed as a 'normal agricultural accident'. Ecological Applications 7: 1350–1356.

Kimber DS, McGregor DI (1995) The species and their origin, cultivation and world production. In: Kimber D, McGregor DI (eds) *Brassica* oilseeds: production and utilization. CABI Publishing, Wallingford, pp 1–9

Kollers S, Rodemann B, Ling J, Korzun V, Ebmeyer E, Argillier O, et al. (2014) Genome-wide association mapping of tan spot resistance (*Pyrenophora tritici-repentis*) in European winter wheat. Mol. Breed. 34 363–371.

Knufer J (2011) Improvement of Winter Oilseed Rape Resistance to *Verticillium longisporum* – Assessment of Field Resistance and Characterization of Ultrastructural Plant Responses. Dissertation. Universität Göttingen.

Kumar V, Paillard S, Fopa-Fomeju B, 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.

Kutcher HR, Brandt SA, Smith EG, Ulrich D, Malhi SS, 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.

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

Larkan NJ, Raman H, Lydiate DJ, Robinson SJ, Yu F, Barbulescu DM, Raman R, Luckett DJ, Burton W, Wratten N, Salisbury PA, Rimmer SR, 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 Biol 16:183.

Larkan NJ, Lydiate DJ, Yu F, Rimmer SR, Borhan MH (2014) Co-localisation of the blackleg resistance genes Rlm2 and LepR3 on *Brassica napus* chromosome A10. BMC Plant Biol 14:387.

Li J, Zhao Z, Hayward A, Cheng H, Fu D (2015) Integration analysis of quantitative trait loci for resistance to *Sclerotinia sclerotiorum* in *Brassica napus*. Euphytica 205:483–489.

Liu Z, Latunde-Dada AO, Hall AM, Fitt BDL (2014) Phoma stem canker disease on oilseed rape (*Brassica napus*) in China is caused by *Leptosphaeria biglobosa* 'brassicae' Eur J Plant Pathol. 140: 841–857.

Lo-Pelzer E, Aubertot JN, David O, Jeuffroy MH, Bousset L (2009) Relationship between severity of blackleg (*Leptosphaeria maculans/ L. biglobosa* species complex) and subsequent primary inoculum production on oilseed rape stubble. Plant Pathology 58:61–70.

Luo Z, Wang M, Long Y, Huang Y, Shi L, Zhang C, Liu X, et al. (2017) Incorporating pleiotropic quantitative trait loci in dissection of complex traits: seed yield in rapeseed as an example. Theor. Appl. Genet. 130:1569–1585.

MacQueen A, Bergelson J (2016) Modulation of R-gene expression across environmentsJ Exp Bot 67: 2093-2105.

Marcroft SJ, Sprague SJ, Pymer SJ, Salisbury PA, Howlett BJ (2004) Crop isolation, not extended rotation length, reduces blackleg (*Leptosphaeria maculans*) severity of canola (*Brassica napus*) in south-eastern Australia. Australian Journal of Experimental Agriculture 44:601–6.

Mason AS, Higgins EE, Stein A, Werner C, Batley J, Parkin IAP, Snowdon RJ (2017) A user guide to the Brassica 60K Illumina Infinium™ SNP genotyping array. Theor. Appl. Genet. 130:621-633.

Mason A, Snowdon RJ (2016) Oilseed rape: Learning about ancient and recent polyploid evolution from a recent crop species. Plant Biol 18: 883-892

Mei J, Ding Y, Lu K, Wei D, Liu Y, Disi JO, et al. (2013) Identification of genomic regions involved in resistance against *Sclerotinia sclerotiorum* from wild *Brassica oleracea*. Theor. Appl. Genet. 126:549–

556.

Mei J, Qian L, Disi JO, Yang X, Li Q, Li J, et al. (2011) Identification of resistant sources against S*clerotinia sclerotiorum* in *Brassica* species with emphasis on *B. oleracea*. Euphytica 177:93-399.

Navabi ZK, Strelkov SE, Good AG, Thiagarajah MR, Rahman MH (2010) *Brassica* B-genome resistance to stem rot (*Sclerotinia sclerotiorum*) in a doubled haploid population of *Brassica napus* × *Brassica carinata*. Can. J. Plant Pathol. 32:237–246.

Neik TX, Barbetti MJ, Batley J (2017) Current Status and Challenges in Identifying Disease Resistance Genes in *Brassica napus*. Frontiers in Plant Science 8:1788.

Neumann MJ, Dobinson KF (2003) Sequence tag analysis of gene expression during pathogenic growth and microsclerotia development in the vascular wilt pathogen *Verticillium dahliae*. Fungal Genet Biol, 38:54-62.

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. Breeding 31:27–37.

Obermeier C, Hossain MA, Snowdon RJ, Knüfer J, von Tiedemann A, Friedt W (2013) Genetic analysis of phenylpropanoid metabolites associated with resistance against *Verticillium longisporum* in *Brassica napus*. Molecular Breeding 31:347-361.

Panchy N, Lehti-Shiu M, Shiu SH (2016) Evolution of Gene Duplication in Plants. Plant Physiol. 171: 2294–2316.

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

Pegg GF and Brady BL (2002) *Verticillium* Wilts. Wallingford, UK: CAB Publishing

Pilet ML, Duplan G, Archipiano M, Barret P, Baron C, Horvais R, et al. (2001) Stability of QTL for field resistance to blackleg across two genetic backgrounds in oilseed rape. Crop Sci. 41:197–205.

Pilet-Nayel ML, Moury B, Caffier V, Montarry J, Kerlan MC, Fournet S, Durel CE, Delourme R (2017) Quantitative resistance to plant pathogens in pyramiding strategies for durable crop protection. Front Plant Sci 8:1838

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

Qian L, Voss-Fels K, Cui Y, Jan HU, Samans B, Obermeier C, Qian W, Snowdon RJ (2016) Deletion of a stay-green gene associates with adaptive selection in Brassica napus. Mol Plant 9: 1559-1569

Ryan MH, Kirkegaard JA, Angus JF (2006) Brassica crops stimulate soil mineral N accumulation. Australian Journal of Soil Research 44: 367-377.

Raman H, Raman R, Larkan N (2013) Genetic dissection of blackleg resistance loci in rapeseed (B*rassica napus* L). In: Andersen SB (ed) Plant breeding from laboratories to fields. InTech, Luton. https://doi.org/10.5772/53611. ISBN 978-953-51-1090-3

Raman R, Taylor B, Marcroft S, Stiller J, Eckermann P, Coombes N, Rehman A, Lindbeck K, Luckett D, Wratten N, Batley J, Edwards D, Wang X, Raman H (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.

Ren L, Xu L, Liu F, Chen K, Sun C, Li J, et al. (2016) Host range of *Plasmodiophora brassicae* on cruciferous crops and weeds in China. Plant Dis. 100:933–939.

Rimmer SR (2006) Resistance genes to *Leptosphaeria maculans* in *Brassica napus*. Can J Plant Pathol 28:S288–S297.

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.

Rygulla W, Seyis F, Lühs W, Eynck C, von Tiedemann A, Friedt W, Snowdon RJ (2007) Combination of resistance to *Verticillium longisporum* from zero erucic acid *Brassica oleracea* and oilseed *Brassica rapa* genotypes in resynthesised rapeseed (*Brassica napus*) lines. Plant Breeding 126:596-602.

Rygulla W, Snowdon RJ, Friedt W, Happstadius I, Cheung WY, Chen D (2008) Identification of quantitative trait loci for resistance against *Verticillium longisporum* in oilseed rape (*Brassica napus* L.). Phytopathol 98:215-221.

Sanogo S, Lujan PA, Baucom D (2015) First report of *Sclerotinia sclerotiorum* on cabbage in New Mexico. Plant Dis. 99:891.

Samans B, Chalhoub B, Snowdon RJ (2017) Surviving a genome collision: Genomic signatures of allopolyploidization in the recent crop species Brassica napus. Plant Genome, doi: 10.3835/plantgenome2017.02.0013

Saxena RK, Edwards D, Varshney RK (2014) Structural variations in plant genomes. Brief Funct. Genomics, 13:296–307.

Schnathorst (1981) Life cycle and Epidemiology of *Verticillium.* In: Mace, M. E., Bell, A. A. and Beckman, C. H. Fungal wilt diseases of plants: 81- 111, Academic Press, London.

Schwartz HF, Otto K, Terán H, Lema M, Singh SP (2006) Inheritance of white mold resistance in P*haseolus vulgaris* × *P. coccineus* crosses. Plant Dis. 90:1167–1170.

Schiessl S, Hüttel B, Kühn D, Reinhardt R, Snowdon RJ (2017a) Targeted deep sequencing of flowering regulators in *Brassica napus* reveals extensive copy number variation. Sci Data 4: 170013

Schiessl S, Huettel B, Kuehn D, Reinhardt R, Snowdon RJ (2017b) Flowering time gene variation in *Brassica* species shows evolutionary principles. Front Plant Sci 8: 1742

Shen JD, Araki H, Chen LL, Chen JQ, Tian DC (2006) Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. Genetics 172:1243-1250.

Snowdon RJ, Lühs W, Friedt W (2006) Oilseed rape. In: Kole C. (ed) Genome Mapping and Molecular Breeding, Vol. 2: Oilseeds. Springer Verlag, Heidelberg, NewYork/Berlin/Toyko, pp.55-114.

Snowdon RJ, Abbadi A, Kox T, Schmutzer T, Leckband G (2015) Heterotic Haplotype Capture: Precision breeding for hybrid performance. Trends Plant Sci 20:410-413

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

Tan S, Zhong Y, Hou H, Yang S, Tian D (2012) Variation of presence/absence genes among Arabidopsis populations. BMC Evol. Biol. 12:86.

Tian L, Wang J, Fong MP, Chen M, Cao H, Gelvin SB, Chen ZJ (2003) Genetic control of developmental changes induced by disruption of Arabidopsis histone deacetylase 1 (AtHD1) expression. Genetics 165:399-409.

Travadon R, Marquer B, Ribulé A, et al. (2009) Systemic progress of *Leptosphaeria maculans* from cotyledons to hypocotyl in oilseed rape: influence of the number of infection sites, of intraspecific competition and of host polygenic resistance. Plant Pathology 58:461–469.

Uppalapati SR, Ishiga Y, Doraiswamy V, Bedair M, Mittal S, Chen J, Nakashima J, Tang Y, Tadege M, Ratet P, Chen R, Schultheiss H, Mysorea KS (2012) Loss of abaxial leaf epicuticular wax in *Medicago*

truncatula irg1/palm1 mutants results in reduced spore differentiation of anthracnose and nonhost rust pathogens. Plant Cell 24:353-370.

Uloth M, You MP, Finnegan PM, Banga SS, Yi H, Barbetti MJ (2014) Seedling resistance to *Sclerotinia sclerotiorum* as expressed across diverse cruciferous species. Plant Dis. 98:184–190.

Uloth MB, You MP, Finnegan PM, Banga SS, Banga SK, Sandhu PS, et al. (2013) New sources of resistance to *Sclerotinia sclerotiorum* for crucifer crops. Field Crops Res. 154:40–52.

Van de Wouw AP, Idnurm A, Davidson JA, Sprague SJ, Khangura RK, Ware AH, et al. (2016) Fungal diseases of canola in Australia: identification of trends, threats and potential therapies. Australas. Plant Pathol. 45:415–423.

Viteri DM, Singh SP (2015) Inheritance of White Mold Resistance in an Andean Common Bean A 195 and its Relationship with Andean G 122. Crop Sci. 55:44-49.

Wang M, Yan J, Zhao J, Song W, Zhang X, Xiao Y, Zheng Y (2012) Genome-wide association study (GWAS) of resistance to head smut in maize. Plant Sci. 196: 125–131.

Wang C, Yang Y, Yuan X, Xu Q, Feng Y, Yu H, Wang Y, et al. (2014) Genome-wide association study of blast resistance in indica rice. BMC Plant Biol. 14: 311.

West JS, Kharbanda PD, Barbetti MJ, 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.

Wei D, Mei J, Fu Y, Disi JO, Li J, Qian W (2014) Quantitative trait loci analyses for resistance to *Sclerotinia sclerotiorum* and flowering time in *Brassica napus*. Mol. Breed. 34:1797–1804.

Wei L, Jian H, Lu K, Filardo F, Yin N, Liu L, Qu C, et al. (2016) Genome wide association analysis and differential expression analysis of resistance to Sclerotinia stem rot in *Brassica napus*. Plant Biotechnol. J. 14:1368–1380.

Willetts HJ, Wong JAL, Kirst GD (1980) The biology of *Sclerotinia sclerotiorum, S. trifoliorum*, and *S. minor* with emphasis on specific nomenclature. Bot. Rev. 46: 101–165.

Wu J, Cai G, Tu J, Li L, Liu S, Luo X, Zhou L, Fan C, Zhou Y (2013) Identification of QTLs for resistance to Sclerotinia stem rot and BnaC.IGMT5.a as a candidate gene of the major resistant QTL SRC6 in *Brassica napus*. PLoS ONE 8:e67740.

Wu J, Zhao Q, Yang Q, Liu H, Li Q, Yi X, Cheng Y, Guo L, Fan C, Zhou Y (2016) Comparative transcriptomic analysis uncovers the complex genetic network for resistance to *Sclerotinia sclerotiorum* in *Brassica napus*. Sci Rep 6:19007.

You MP, Uloth MB, Li XX, Banga SS, Banga SK, Barbetti MJ (2016) Valuable new resistances ensure improved management of Sclerotinia Stem Rot (*Sclerotinia sclerotiorum*) in horticultural and oilseed *Brassica* species. J. Phytopathol. 164:291–299.

Würschum T, Boeven PHG, Langer SM, Longin CFH, Leiser WL (2015) Multiply to conquer: copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. BMC Genet. 16:96.

Würschum T, Longin C, Friedrich H, Hahn V, Tucker MR, Leiser WL (2017) Copy number variations of CBF genes at the Fr-A2 locus are essential components of winter hardiness in wheat. Plant J. 89:764–773

Yu F, Gugel RK, Kutcher HR, Peng G, Rimmer SR (2013) Identification and mapping of a novel blackleg resistance locus LepR4 in the progenies from *Brassica napus* × *B rapa* subsp sylvestris. Theor Appl Genet 126:307–315.

Zhang D, Bowden RL, Yu J, Carver BF, Bai G (2014) Association analysis of stem rust resistance in U.S. winter wheat. PLoS ONE, 9, e103747.

Zhao J, Peltier AJ, Meng J, Osborn TC, Grau CR (2004) Evaluation of Sclerotinia Stem Rot resistance in oilseed *Brassica napus* using a petiole inoculation technique under greenhouse conditions. Plant Dis. 88:

1033–1039.

Zhao J, Udall JA, Quijada PA, Grau CR, Meng J, Osborn TC (2006) Quantitative trait loci for resistance to *Sclerotinia sclerotiorum* and its association with a homeologous non-reciprocal transposition in *Brassica napus* L. Theor. Appl. Genet. 112: 509–516.

Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, et al. (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat. Genet. 50: 278–284.

Zhou L, Hu Q, Johansson A, Dixelius C (2006) *Verticillium longisporum* and *V. dahliae*: infection and disease in *Brassica napus*. Plant Pathology 55: 137-144.

Żmieńko A, Samelak A, Kozłowski P, Figlerowicz M (2014) Copy number polymorphism in plant genomes. Theor. Appl. Genet. 127:1–18.

List of abbreviations

ANOVA	Analysis of variance
AUDPC	Area under the disease progress curve
BP	Biparental
bp	Base pairs
cM	centi Morgan
CNV	Copy number variation
DH	Double haploid
DNA	Deoxyribonucleic acid
FAO	Food and agriculture organization of the United Nations
FDR	False discovery rate
GO	Gene ontology
GWAS	Genome-wide association studies
Ha	Hectare
HE	Homeologous exchanges
HMW	High molecular weight
InDel	Insertion and deletion polymorphisms
K	Kinship
LD	Linkage disequilibrium
M	Million
MAF	Minor allele frequency
Mbp	Million base pare
MLM	Mixed linear model
NAM	Nested association mapping
NBS-LRR	Nucleotide-binding site leucine-rich repeat
PAV	Presence-absence polymorphisms
PCA	Principal component analysis
PCR	Polymerase chain reaction
QTL	Quantitative trait locus
RIL	Recombinant inbred line
SNaP	Single Nucleotide absence Polymorphism
SNP	Single Nucleotide Polymorphism
SV	Structural variation
WGD	Whole-genome duplication
WGS	Whole-genome sequencing

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".

Acknowledgments

I am very grateful to many people who have encouraged and supported me during the time spent in Giessen and the elaboration of this thesis.

First of all, I want to express my gratitude to my first PhD supervisor, Prof. Dr. Rod J. Snowdon. He supervised my scientific development with trust, patience and outstanding scientific competence. I will always remember your very positive and friendly personality, open-mindedness and creativity. I could not have imagined having a better mentor for my thesis. Thank you, Rod, for everything!

My sincere thanks also go to my second academic supervisor, Prof. Dr. Matthias Frisch, for his many suggestions and continuous advice during the time spent in Giessen. Statistics would have been a very complex and hard to understand field without your assistance and valuable inputs

Another very important person that contributed directly at my scientific career development is Dr. Christian Obermeier. I am grateful for your co-supervision and guidance during the elaboration of this thesis. Your vast knowledge in plant breeding was of most importance for me. Our daily scientific discussions were very productive and extremely helpful. Without your precious support it would not have been possible to conduct this research. Thank you, Christian, for always been there with answers and suggestions.

Thanks to all my colleagues and friends from the Department of Plant Breeding for the nice working atmosphere and your help in all circumstances. You have contributed not only to a pleasant but also a very productive work-atmosphere. Many thanks I want to send out to Harmeet Singh Chawla, Roman Gäbelein, Subhadra Chakrabarty, Isabelle Deppe, André Schaffasz, Paul Vollrath, Daniela R. Quezada Martinez, Katharina Tyson, Christian Werner,

Kai Voss-Fels, Lunwen Qian, Timm Bernhard, Andreas Stahl, Benjamin Wittkop, Annaliese Mason, HueyTyng Lee, Fabian Grandke. Further, I would like to express my thankfulness to my favorite lab-officers and friends, Stavros Tzigos and Liane Renno. I am grateful for the stimulating scientific discussions, the 'poster sessions', conferences, meeting and all the fun we have had.

I am very grateful to the German Academic Exchange Service (DAAD) for financing my studies.

Last but not the least, many more thanks go out to my family that always supported me. First of all, my best friend, partner and wife, Diana, who was always there for me. Her unconditional love and understanding were crucial and allowed me to focus on my work. Sorry for the long working days. I'm truly thankful for everything you did. TD. My sincere thanks go to my beloved mother and father for being my model and for always encouraging me. I'm also really thankful to my brother, his wife and two beautiful daughters.

I've never taken things for granted and I know that my achievements would have not been possible without all your help. Sincerely, thank you!