Aus dem Institut für Pflanzenbau und Pflanzenzüchtung I,

Justus-Liebig-Universität, Professur für Pflanzenzüchtung

The role of genome structure variation in the evolution and adaptation of life cycle traits in *Brassica* species

Habilitationsschrift

zur Erlangung des akademischen Grades der Dr. habil

und der *venia legendi*, verliehen durch den Fachbereich 09 (Agrarwissenschaften, Ökotrophologie und Umweltmanagment) der Justus-Liebig-Universität

für das Lehrgebiet „Pflanzengenetik“

eingereicht von

Dr. Sarah Schießl-Weidenweber

Dezember 2019

Table of contents

Rationale

Humanity strongly depends on successful plant production. Recent decades have seen tremendous increases in production efficiency based on agronomic, technical and plant breeding improvements. The upcoming decades, however, also pose tremendous challenges for plant production due to three major reasons: first, the global human population is still rising. Second, climate change threatens local climate stability, changes water availability and shifts climate zones. Third, the need for a climate protective agriculture limits the use of common agronomic and technical solutions. Altogether, those challenges increase the importance of fast plant breeding solutions. Plant breeders therefore need to be able to adapt new varieties more quickly to changing conditions to allow stable and high yields. Modern breeding strategies rely on informed decisions about the allelic content of a variety, which demands reliable genetic markers for stress tolerance and climatic adaptation, including life cycle adaptation. We therefore need to understand why, how and with which measures successful adaptation works on a genetic and genomic level. In that respect, we hope to learn more from evolutionary processes that were successful in adapting species to the conditions that are now in place. In the last decades, we have learnt that plant genomes are more dynamic than we expected. Allelic variance does not only affect small differences in DNA sequence, but also large structural variation. However, many important crops are polyploid and therefore have complex genomes that are hard to dissect. Therefore, structural genome variation in polyploid crops is not well understood, although it may possess a high adaptive potential. The best crop model to study effects of genome structure variation on climatic adaption is the genus *Brassica*. Its close relationship to the model plant *Arabidopsis thaliana*, the well-known relationships between the diploid and the polyploid *Brassica* species, the ease of resynthesizing polyploids from diploids and the wide use as oil crops, vegetables, fodder, condiments and ornamentals make it an excellent choice as a study system. Among the six *Brassica* species, *Brassica napus* in the form of oilseed rape has the highest economic impact and therefore the highest demand to be improved and adapted. However, genetic bottlenecks in the breeding history of oilseed rape limited genetic diversity, which is a major obstacle to further genetic gains. Introgressions from the first and second gene pool are therefore recommended, but they also affect adaptation traits. One of the most critical adaptation traits affected is synchronization of the life cycle with conditions in place. This affects two developmental transitions: germination, which is the embryo – juvenile transition, and flowering, which is the juvenile – adult transition. Both impact yield, but flowering time seems to have a stronger impact.

Here, I studied the impact of structural genome variation within the flowering time gene network in *Brassica* species with two main objectives: First, I want to understand how structural variation affects single *Brassica* flowering time gene copies to be supportive of suitable marker development. Second,

I want to understand the evolutionary processes shaping this system to inform future adaptation breeding. Chapter I presents four published papers dealing with the effects of structural variation in *Brassica* crops. Chapter II presents three published papers that studied the differential impacts for gene copies of the same flowering time gene, indicating that evolutionary processes have changed their functions. Chapter III discusses the genetic and physiological effects of abiotic stress during flowering, illuminating the need for a more holistic approach in climate change sensitive breeding.

Publications included in this thesis

Chapter I:

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

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

Schiessl SV, Huettel B, Kuehn D, Reinhardt R, Snowdon RJ (2017c) Flowering Time Gene Variation in *Brassica* Species Shows Evolutionary Principles. Front. Plant Sci. 8:183. doi: 10.3389/fpls.2017.01742

Schiessl SV, Katche E, Ihien E, Chawla HS, Mason AS (2018) The role of genomic structural variation in the genetic improvement of polyploid crops. The Crop Journal 7:127–140. doi: 10.1016/j.cj.2018.07.006

Chapter II:

Snowdon RJ, **Schiessl S** (2019) Illuminating Crop Adaptation Using Population Genomics. Molecular Plant 12:27–29. doi: 10.1016/j.molp.2018.12.014

Schiessl S, Williams N, Specht P, Staiger D, Johansson M (2019) Different copies of SENSITIVITY TO RED LIGHT REDUCED 1 show strong subfunctionalization in *Brassica napus*. BMC Plant Biol 19:372. doi: 10.1186/s12870-019-1973-x

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

Chapter III

Schiessl SV, Orantes-Bonilla M, Quezada-Martinez D, Snowdon RJ: Room for improvement: transcriptomics reveal high genetic diversity of drought resistance strategies in winter oilseed rape (unpublished manuscript)

Hatzig SV, Nuppenau J-N, Snowdon RJ, **Schiessl SV** (2018) Drought stress has transgenerational effects on seeds and seedlings in winter oilseed rape (*Brassica napus* L.). BMC Plant Biol 18:297. doi: 10.1186/s12870-018-1531-y

Introduction

Genome structure

Evolution happens at the DNA. Understanding evolution and adaptation consequently needs a thorough understanding of the structure of DNA. The most basic level of DNA structure is its sequence, and the second and most popular is the complementary double helix (Watson and Crick 1953). Eukaryotes store their DNA in the nucleus to better protect and regulate the DNA. Therein, eukaryotic DNA is organized in chromosomes, which are long linear sequences of complementary DNA strands, which can contain up to 150 000 Million base pairs (Mbp) in some fern species (Hidalgo et al. 2017). The DNA is wound around globular proteins called histones (Vergara and Gutierrez 2017) and brought into a precisely organized 3D structure by the help of additional structural proteins (Ea et al. 2015). This superstructure from DNA, histones and organizing proteins is called the chromatin, and its conformation decides whether the DNA is ready to allow transcription, can be replicated, may be damaged or repaired, or may also be recombined with other DNA (Kumar 2018; Tock and Henderson 2018; Vergara and Gutierrez 2017; Waterworth et al. 2011). Understanding the structure and organization of chromatin is therefore fundamental to our understanding of life. The following paragraphs will give a short overview over the different levels of chromatin organization.

Histones

Histones are small globular proteins with a high percentage of basic amino acids, allowing them to bind to the negatively charged DNA backbone (Cutter and Hayes 2015; Vergara and Gutierrez 2017). Different histones have been identified (H1, H2A, H2B, H3, H4) which themselves have different isoforms (Cutter and Hayes 2015). Two histones of each H2A, H2B, H3, H4 form heterooctamers around which the DNA molecule can wrap around twice over a length of 147 bp (Cutter and Hayes 2015; Vergara and Gutierrez 2017), forming a nucleosome. Nucleosomes are linked by unwound DNA, which is protected by histone H1 (Cutter and Hayes 2015; Vergara and Gutierrez 2017). The conformation and therefore the accessibility of the DNA can be regulated by the strength of the binding between histones and DNA (Cutter and Hayes 2015). The looser the binding, the more accessible, but the more unstable is the chromatin (Cutter and Hayes 2015). The regulation can be large-scale, like the compaction of the chromatin in mitosis and meiosis guided by histone H1 (Golloshi et al. 2017), medium-scale by the exchange of histone isoforms in the nucleosomes, and small-scale by chemical modifications of the histones (e.g. methylations, acetylations, phosphorylation etc.) or the DNA (methylation of cytosine) (Kumar 2018; Vergara and Gutierrez 2017). Therefore, histones play a central role in regulation of transcriptional activity and in restructuring the chromatin in cell division. In plants, the most famous example of histone-driven gene regulation is the vernalisation system, which blocks flowering before winter (He 2009). Here,

the activation of chromatin remodeling factors by cold leads to the exchange of histone H2A with another variant in the chromatin of the flowering repressor *FLOWERING LOCUS C* (*FLC*). At the same time, repressive marks are removed from histone H3, while activating marks are added (He 2009). This blocks *FLC* transcription and gives way to positive flowering regulation (He 2009). Other examples for histone-driven transcription regulation in plants are temperature sensing and stress response (Kumar 2018).

Centromeres and telomeres

During the reproductive process, plants produce haploid gametes. Each gamete needs to carry a full, but unique set of chromosomes, therefore, homologous chromosomes need to be faithfully separated. Faithful separation is organized at the kinetochore, a structure keeping homologous chromosomes attached to each other in their most condensed state until the spindle apparatus has fully formed (Scelfo and Fachinetti 2019). The kinetochore does not form randomly, but at a site characterized by highly repetitive DNA called the centromere, as it is often (but by far not always) found quite to the center of the chromosome (Scelfo and Fachinetti 2019). If a chromosome lacks a functional centromere, it has a high risk not to be passed onto the next generation (Wang and Dawe 2018).

Another feature of eukaryotic chromosomes tightly linked to the cell cycle are telomeres. Telomeres are also composed of repetitive DNA bound to special scaffolding proteins to protect the ends of the chromosome from degradation (Victorelli and Passos 2017). Those ends should normally get shorter with each round of replication, as DNA polymerases are unable to fully synthesize the lagging strand to the end (Victorelli and Passos 2017). In some cell types, however, the enzyme telomerase can resynthesize shortening telomeres (Victorelli and Passos 2017). This is necessary to maintain the capacity to divide, as cells with too short telomeres are arrested (Victorelli and Passos 2017).

Higher order structures

When cells do not undergo mitosis or meiosis and are arrested in interphase, the chromosomes are not just randomly floating around in the nucleus, but instead have a precisely defined 3D architecture in which the chromosomes occupy certain areas which are called nuclear territories (Ea et al. 2015; Golloshi et al. 2017). This architecture is lost during condensation in mitosis or meiosis, but exactly rebuild afterwards (Golloshi et al. 2017). It is still unclear how this memory process works, but it is likely to be governed by epigenetic marks (Golloshi et al. 2017). The substructures of these nuclear territories are obviously important for transcriptional activity, as GC poor territories on the outside seem to have lower transcriptional activity than those in the inside (Golloshi et al. 2017). This finding is striking in respect to structural variation, as strongly increased or decreased chromosome length might have an impact on transcriptional activity of genes otherwise non-affected

genes. Thinking of the genome as a 3D structure changes our linear way of interpreting genetic information dramatically and will possibly solve some of the obstacles which are unresolved up to now. However, when we speak of structural variation of the DNA, we normally do not think of changes in histone composition, centromere or telomere length, nor of 3D structure variation. This has mostly technical reasons: the methods to study most of those phenomena are very recent and not yet accessible to high-throughput studies. Most of the time, structural variants rather summarize large rearrangements within the DNA part of the chromatin. Nevertheless, knowledge about the respective principles of structure allows us to better understand the patterns in genomic rearrangements.

Polyploidy

Mechanisms of genome doubling

Almost all flowering plants show signatures of ancestral genome doubling events at some point in their evolution (Soltis and Soltis 2016). Individuals or species carrying more than one set of chromosomes per nucleus in all their cells as a result of genome doubling are called polyploids. The set number, also called the ploidy, can only increase if the cell cycle, either in mitosis or meiosis, is interrupted after DNA replication and before cell division. If this happens in somatic cells, this is called endoreduplication and may lead to polyploid tissue parts within an organism (Chevalier et al. 2014). This is extremely common in organ development across kingdoms, it affects humans (heart, liver), insects (gut), or everyday vegetables (tomato fruit flesh) (Orr-Weaver 2015). If somatic genome doubling affects the reproductive tissue, this may lead to the formation of a fully polyploid offspring in selfing species, as plants do not show a strictly separated germline (Walbot and Evans 2003). The resulting offspring of such an event is called autopolyploid, as the extra set of chromosomes comes from the same species. However, most polyploid events are believed to be the result of fertilization with unreduced gametes (Mason and Pires 2015; Szadkowski et al. 2011). Unreduced gametes arise naturally at low frequencies around 0.5% depending on the species (Otto 2007), but their abundance rises under stress conditions like drought and heat (Mason and Pires 2015). Fertilization with unreduced gametes may lead to triploid plants (reduced + unreduced gamete) or, at much lower probability, to tetraploid plants (unreduced + unreduced gamete) (Arrigo and Barker 2012). Triploid species, however, can either be sterile and therefore go extinct immediately (Arrigo and Barker 2012), or they can produce unreduced gametes at higher frequencies, increasing the likelihood of tetraploid offspring (Mason and Pires 2015). This mechanism is called the triploid bridge. Fertilization with unreduced gametes can happen within a species and give rise to autopolyploids, but it also may happen between closely related species sharing the same

habitat and lead to allopolyploids. Here, the extra set of chromosomes comes from a different species.

In both cases, the new polyploid species has to solve several new problems. The first and major problem is correct sorting of chromosomes during meiosis to avoid aneuploidy in the next generation (Pelé et al. 2018). Moreover, the merging of two genomes seems to cause immediate problems for chromatin organization, a phenomenon described as the 'genomic shock' (Bashir et al. 2018; McClintock 1984). This term summarizes different observations of immediate changes in the nucleus, like changes in the methylation landscape, activation of transposable elements, genomic rearrangements, changes in small RNA metabolism and alternative splicing (Diez et al. 2014; Gaeta et al. 2007; Lisch 2012; Lukens 2005; Otto 2007). Depending on the species, it has been controversially discussed if the genomic shock is real (Göbel et al. 2018; Sarilar et al. 2013), and it may be concluded that the effect size of the genomic shock depends on the species. However, the genomic shock is only observed in the first generations after whole-genome duplication, and control over the disturbed processes is regained quickly (An et al. 2019; Skalická et al. 2003).

Influence of polyploidy on genome structure variation

Having two sets of similar or even identical chromosomes has both positive and negative consequences for the plant. In the first place, polyploidy leads to higher redundancy within the genome, masking effects of negative mutations (Otto 2007; Renny-Byfield and Wendel 2014). Genetic redundancy increases the degrees of freedom in mutation and selection and therefore allows a faster evolution at gene scale (Otto 2007; Renny-Byfield and Wendel 2014). Moreover, polyploid plants can inherit variance between the subgenomes stably to the next generation, because alleles are located at homeologous, but not homologous chromosomes. This is different than for heterozygous diploids, where heterozygosity will segregate in the next generation. This phenomenon is termed fixed heterosis (Chen 2010) and is assumed to be one of the reasons for higher growth rates and better performance of some polyploids (Miller et al. 2012; Song and Chen 2015).

Redundancy, however, may also cause problems. The first major problem is chromosome segregation. In meiosis I, homologous chromosomes line up to be separated into different daughter nuclei (Mercier et al. 2015). If sequence similarity between homeologous chromosomes is comparable to sequence similarity between homologous chromosomes, this process can also happen between homeologous chromosomes, and this may have outcomes of varying severity (Gaeta and Pires 2010). This is the reason why first generation polyploids very often have reduced fertility and have a high risk of immediate extinction (Arrigo and Barker 2012; Gaeta and Pires 2010). In the worst case, all four chromosomes (homologs and homeologs) align, and the daughter nuclei receive too

less or too many chromosomes, resulting in aneuploidy (Gaeta and Pires 2010). In another scenario, each daughter cell only receives the two homologs of the same homeolog, resulting in the duplication of this chromosome, but the deletion of its homeolog (Gaeta and Pires 2010). Both patterns are very abundant in the first generations of polyploidy (Gaeta et al. 2007; Pires et al. 2004; Samans et al. 2017; Stein et al. 2017; Szadkowski et al. 2011), but normally less observed in established polyploids (Samans et al. 2017; Szadkowski et al. 2011), indicating they have evolved some mechanism to control chromosome distribution. Some authors propose that large scale inversions within homeologous chromosomes block them to fully align (Faria and Navarro 2010; Rodgers-Melnick et al. 2015; Ziolkowski and Henderson 2017), while others claim a genetic control mechanism (Jenczewski et al. 2003; Liu et al. 2006). Independently of the mechanism, some newly formed polyploids manage to overcome this barrier and restore normal fertility. This does however not always mean that subgenomes stop to interact, leading to the second major problem. Homeologous chromosomes still align in some established polyploids (Chalhoub et al. 2014; Lashermes et al. 2016), although complete loss or duplication of chromosomes is rarely observed. Instead, like homologous chromosomes also do, they exchange genetic material in crossing overs, either as a balanced exchange, or in overwriting one part of the homeolog with the respective part from the other one (Żmieńko et al. 2013). This means that large parts of the DNA sequence are either missing, copied, inverted or shifted to another place, and this is then summarized as structural variation. While different mechanisms are responsible for the arousal of structural variation, they have one thing in common: they all start with a double-strand break (DSB) that somehow has to be repaired by the cell machinery.

Double-strand breaks

Under normal circumstances, double strand breaks within the DNA are a cellular emergency case– if not repaired immediately, they result in one chromosome with no centromere and two with a missing telomere, effectively abolishing cell division capacity. DSBs can occur due to environmental stresses like UV-B radiation, reactive oxygen species as side products of photosynthesis, desiccation and pollutants like heavy metals, but also due to endogenous processes, like stalked replication forks and steric stress during DNA unwinding (Waterworth et al. 2011). Moreover, DSBs can also be introduced by transposable elements (TEs). Transposable elements are mobile DNA sequences which may excise or copy themselves from and into the genome (Lisch 2012; Vicient and Casacuberta 2017). Depending on their origin and mode of action, they can either just translocate in the genome or amplify therein (Lisch 2012). As this poses a veritable threat to the genome, the cell machinery normally tries to silence active TEs epigenetically, but a general activity is normally upheld (Lisch 2012; Vicient and Casacuberta 2017). Under some circumstances like cellular stress or hybridization,

those epigenetic marks are lost and TEs regain activity, destabilizing the genome (Lisch 2012; Vicient and Casacuberta 2017).

However, there is also a scenario where double strand breaks are even introduced deliberately into the DNA by the cell machinery itself, and this is during homologous recombination within meiosis I. The DNA is cut by an enzyme called Spo11, a transesterase (Mercier et al. 2015; Serra et al. 2018). In *A. thaliana*, it introduces 100-200 DSBs per meiosis (Serra et al. 2018). Most of those DSBs, however, are repaired as non-crossovers, and most chromosomes receive less than three crossovers per generation (Mercier et al. 2015). Crossovers do not occur randomly, instead, the number and distribution of crossovers is tightly controlled and concentrated into some areas called hotspots of meiotic recombination (Mercier et al. 2015; Tock and Henderson 2018). Those hotspots are obviously determined by chromatin structure, as, for example, nucleosome-depleted regions within promotors are enriched in meiotic hotspots (Mercier et al. 2015; Tock and Henderson 2018). Moreover, the chromatin structure during Prophase I seems to have a strong impact on DSB distribution (Tock and Henderson 2018). At this time, the chromatin is wound around a protein axis in loops bound by the protein cohesin (Tock and Henderson 2018). In the tethered loop/axis model, Spo11-1 seems to localize first at the axis, but is then translocated outside to the loop where it cuts and allows for recombination (Tock and Henderson 2018). At the axis, other proteins like Rec8 seem to protect the chromatin form DSBs (Tock and Henderson 2018).

Double-strand break repair

In general, there are two major mechanisms to repair DSBs. Outside meiosis, DSBs are mostly repaired via non-homologous end-joining (NHEJ), while the major mechanism during meiosis is homologous recombination (HR) (Puchta 2005; Waterworth et al. 2011). In somatic cells, HR is a minor pathway, and repair via NHEJ dominates (Puchta 2005). In NHEJ, open ends are linked to one other based on microsynteny of a few base pairs (Chang et al. 2017; Waterworth et al. 2011). Recent studies have linked somatic DSBs and their repair to long term genome shrinkage, although species seem to be differentially affected by such mechanisms (Vu et al. 2017). Moreover, they can lead to small insertions and deletions at the break site, as sometimes synthesis-dependent DNA repair seems to be involved (Waterworth et al. 2011).

Homologous recombination, on the other hand, involves different pathways, but all of them depend on homology (Puchta 2005). For lesions located in an area with repetitive sequences on both sites, a third of the lesions are repaired with single-strand annealing (SSA) (Daley et al. 2013; Waterworth et al. 2011). SSA leads to the deletion of one of the repetitive elements (Waterworth et al. 2011). For DSBs in non-repetitive regions, DSBs are immediately recognized by nucleases digesting one strand to create single stranded 3'-overhang ends of 100-1000 bp length (Lawrence et al. 2017; Serra et al.

2018; Waterworth et al. 2011). Those overhangs are then bound by scaffolding proteins to allow the broken strand to invade its homologous double strand to form the so-called displacement loop (Lawrence et al. 2017; Pyatnitskaya et al. 2019; Serra et al. 2018). During the process, mismatches originating from allelic variance are repaired by the mismatch repair system (Waterworth et al. 2011), effectively reducing diversity at this site. The D-loop can be dissolved in different ways: in synthesis-dependent strand annealing (SDSA), the displacement loop dissociates after the missing information is copied from a suitable homolog and then annealed to the 3'end of the broken strand, leading to repair without recombination or non-crossover (Pyatnitskaya et al. 2019; Waterworth et al. 2011). Alternatively, the D-loop region can extend and capture the second end of the broken strand (second-end capture) (Pyatnitskaya et al. 2019). Ligation of the respective ends then leads to the formation of a double Holliday junction (dHJ), which can be resolved with different outcomes, either as crossover or non-crossover (Pyatnitskaya et al. 2019; Waterworth et al. 2011). The stabilization of the dHJ is promoted by the ZMM proteins (Pyatnitskaya et al. 2019). It has been found that polymorphisms reduce the number of recombination events, possibly due to mismatch recognition promoting dissociation after the second-end capture (Serra et al. 2018). At the same time, recombination is likely to be mutagenic by process errors and is therefore also a source of genetic variation (Serra et al. 2018). In summary, most DSBs are repaired in a non-recombinant way, but they influence allelic diversity and local mutation rates in all cases.

Evolutionary effects of recombination between homeologs

In allopolyploid species, homeologous chromosomes rather than homologous chromosomes sometimes pair in meiosis (Gaeta and Pires 2010; Mercier et al. 2015). In meiosis I, this may lead to the loss of one homeolog coupled to the duplication of another when both homologs of a homeolog are pulled towards the same daughter cell (Chen et al. 2018; Mwathi et al. 2019). If the homeologous chromosomes have recombined during prophase I, this may be true for parts of those chromosomes (Chen et al. 2018; Gaeta et al. 2007; Mwathi et al. 2019; Szadkowski et al. 2011). As the segregation pattern is random and the following choice of fertilization partner is random, too, many different outcome patterns are possible, including patterns without any change (Gaeta and Pires 2010). With continued selfing, those patterns are quickly fixed, leading for example to patterns of homeologous reciprocal translocation (HRT), deletion of one homeolog coupled to duplication of the other, also called homeologous non-reciprocal translocation (HNRT), or simple deletions or duplications (Gaeta and Pires 2010). It is important to note that all those patterns may trace back to the same recombination mechanisms, but occur due to different segregation and mating events. In any case, they can affect large regions of several thousands to million base pairs and therefore very likely affect genes. Polyploid genomes are normally expected to have more than one copy of a gene, but deletions and duplications may change this copy number. This phenomenon is summarized as copy

number variation (CNV) (Żmieńko et al. 2013). In a strict sense, it only terms deletions and duplications, but as it was found that not all copies show the same functionality, HRTs and HNRTs are also included in this term. Polyploids can tolerate deletions more easily due to their higher redundancy, and can therefore accumulate more such copy number variants (CNVs) within their genome (Otto 2007). This adds another level of complexity to the polyploid genome, as different individuals of the same polyploid species may not only show sequence variation, but also a high level of copy number variation, eventually also affecting traits (Chang et al. 2015; Fopa Fomeju et al. 2014; Hu et al. 2018; Lee et al. 2016; Maron et al. 2013; Qian et al. 2016; Schiessl et al. 2017a).

Diploidization

In the long run, gene copy numbers decrease considerably, eventually until the point of non-redundancy (Renny-Byfield and Wendel 2014; Sankoff et al. 2010). At any point of this process of genome reduction, a new round of polyploidy may occur (Adams and Wendel 2005; Alix et al. 2017; Mason and Batley 2015). This makes it harder to precisely define a polyploid. Polyploids are therefore classified into ancient polyploids or paleopolyploids, which are effectively diploid, but still carry detectable traces of polyploidy, mesopolyploids, which still show genome-wide redundancy, but the original chromosome structure is already disorganized, and young or recent polyploids where subgenome structure is overall conserved (Kagale et al. 2014; Soltis and Soltis 2009). A thorough examination of the genomic resources we have up to date has shown that there is almost no plant species not carrying any traces of polyploidy (Soltis and Soltis 2009). Polyploidy is therefore an extremely common and important process in plant evolution, but in itself unstable (Otto 2007; Renny-Byfield and Wendel 2014; Sankoff et al. 2010). Obviously, although polyploidy has some advantages in terms of faster adaptation (Kagale et al. 2014; Selmecki et al. 2015), large genomes seem to be a burden to the cell (Hidalgo et al. 2017). Over time, deletions accumulate, and genome size decreases. However, this process is obviously not completely random, as some genes show amplification or at least copy number conservation against the trend (Wendel et al. 2018). The key to understand those exceptions is possibly the observation that genes with related, but not identical functions often share the same ancestor (Soltis and Soltis 2016). For example, it has been shown that genome duplication played an important role in the evolution of the AGAMOUS subfamily involved in flowering (Dreni and Kater 2014). The other way round, gene copies which have been retained mostly show differential expression patterns (Renny-Byfield and Wendel 2014). This indicates that gene duplication has a diversifying function: due to its redundancy, mutations within one copy are less severely perceived by evolution if another copy can uphold the function (Otto 2007). This copy may therefore accumulate mutations more easily and has a higher chance to achieve a new or altered function (Otto 2007). At the same time, the chance to loose function is also higher. A process altering the function of a gene is called subfunctionalisation, a process leading to a new function is

called neofunctionalisation. Processes leading to lost functions are called pseudogenisation, as their products are pseudogenes. Together, the recurring processes of whole-genome duplication and refunctionalisation drove major innovations in plant evolution, like the complex structure of flowers (Soltis and Soltis 2016).

Brassica

Model crops for evolution and adaptation

Established polyploids can have many advantages over diploids. They tend to have higher growth rates (Miller et al. 2012; Song and Chen 2015), they may show fixed heterosis (Chen 2010) and might therefore be more flexible in extreme conditions (Kagale et al. 2014), and there is potential to increase diversity and introduce new or improved traits via cross-species introgression (Mason and Batley 2015). Moreover, whole-genome duplication is one of the major driving forces of evolution (Soltis and Soltis 2016) and studying polyploids can therefore help revealing general mechanisms of adaptation and speciation. At the same time, research on polyploid plants faces many challenges: the complexity of polyploid genomes is a strong obstacle to genomic and bioinformatics studies due to its high redundancy and the unknown degree of sub- and neofunctionalization (Kyriakidou et al. 2018). Moreover, many polyploids arose long time ago, and their exact progenitors are not known. This makes it often impossible to study processes within the first generations of polyploidy.

The genus *Brassica* is excellently suited to meet most of these challenges (Snowdon 2007). Three of its diploid species, *B. rapa*, *B. nigra* and *B. oleracea*, can still be used to resynthesize three of its allotetraploid species, *B. juncea*, *B. napus* and *B. carinata*. Those relationships were recognized very early and summarized as the triangle of U (Morinaga 1934; Nagaharu 1935). All 6 U triangle species are used as crops: *B. rapa*, the A subgenome donor, is used as leafy vegetable (Pak Choi, Chinese cabbage) or beet (turnips) (Qi et al. 2017). *B. nigra*, the B subgenome donor, is known as black mustard. *B. oleracea*, the C subgenome donor, comprises a large variety of leafy vegetables (white, red and savoy cabbage, kale, Brussels sprout), flower vegetables (broccoli, cauliflower) and beets (kohlrabi) (Rakow 2004). *B. juncea* (genome AABB) and *B. napus* (genome AACC) are both mainly grown as oilseeds, while the latter has much higher economic impact on a global level. *B. carinata* (genome BBCC) is also called Ethiopian mustard and is used as condiment and leafy vegetable.

Oilseed rape *(B. napus)* is the crop with the highest economic impact among the six U triangle species and at the same time the youngest of the three allotetraploid species. It is thought to have occurred around 7500 years before now (Chalhoub et al. 2014). The subgenomes are therefore still basically intact (Parkin et al. 1995) as compared with the progenitor genomes. This creates an optimal situation to be used as a polyploid model: the interspecific hybridization can still be repeated to

study the first generations after polyploidisation (Morinaga 1934; Nagaharu 1935), while natural *B. napus* exists in a meiotically stable, but not yet fractionated form (Chalhoub et al. 2014). Both the economic impact and the close relationship to the model plant *Arabidopsis thaliana* facilitated the assembly and annotation of several reference genomes and pangenomes for *Brassica* species up to date (Bayer et al. 2017; Chalhoub et al. 2014; Golicz et al. 2016; Hurgobin et al. 2017; Liu et al. 2014; Parkin et al. 2014; Wang et al. 2011b; Yang et al. 2016). Moreover, any progress in understanding traits and processes can be more or less directly put into practice (Snowdon and Iniguez Luy 2012). *Brassica* is therefore a perfect model crop system for studying evolution and adaptation of polyploid crops.

Brassica genome structure

Polyploidy is a recurring process in plant evolution. Therefore, species considered diploid often carry traces of former whole-genome duplications within their genomes (Wendel 2015). In the diploid *Brassica* species, the genome structure shows strong evidence of a former hexaploidization step (Lagercrantz et al. 1996; Lagercrantz 1998; Lukens et al. 2003; Lysak 2005) which has been dated differently by different authors, varying between 23 Mya (Beilstein et al. 2010; Parkin et al. 2014), 19-8 Mya (Lysak 2005), and 9-5 Mya ago (Wang et al. 2011b). The allotetraploid species do therefore not only carry two copies of each locus, but rather vary between 1 and 12 (Schiessl et al. 2014) with a mean copy number of 4.4 (Parkin et al. 2010), as the expected copy number of 6 was reduced or increased by ongoing diploidization processes during lineage speciation. The original hexaploidization obviously happened in two steps: a tetraploidization followed by the addition of another subgenome (Parkin et al. 2014; Wang et al. 2011b). Likewise, the genomes of the diploid Brassicas contain three parts: the two most fractionated parts (MF1 and MF2) and the least fractionated part (LF) (Cheng et al. 2012; Wang et al. 2011b). The hexaploid ancestor gave rise to the three lineages A, B and C (corresponding to the subgenome) (Li et al. 2017). Lineage B diverged earlier and is therefore more distinct from A/C than A and C from each other (Li et al. 2017). This has consequences for the interaction of the subgenomes within the allotetraploid species. Resynthesized *B. napus* with its AACC genome configuration shows more interactions between the subgenomes (Gaeta et al. 2007; Nicolas et al. 2012; Samans et al. 2017; Szadkowski et al. 2011) than *B. juncea* (Axelsson et al. 2000) which reflects the greater distance of the B subgenome from AC. In the first generations after allopolyploidization, the subgenomes in resynthesized *B. napus* show large genomic rearrangements between the each other, in particular for those chromosomes with complete synteny (Gaeta et al. 2007; Nicolas et al. 2012; Samans et al. 2017; Szadkowski et al. 2011). The size and number of such events seems to depend on the mode of formation: lines derived from colchicine treatment and tissue culture show larger and fewer rearrangements, while lines derived from unreduced gametes show smaller and more (Szadkowski et al. 2011). Traces of rearrangements in established *B. napus*

show that although this process is slowed down, it is still ongoing (Samans et al. 2017; Schiessl et al. 2017b). The occurrence of genomic exchanges between subgenomes depends on the distance from the centromeric regions, the degree of collinearity and the subgenome, as A tends to loose less content than C (Nicolas et al. 2012), a phenomenon described as biased fractionation. The reason for the preferential loss of C versus A loci is unclear, although different explanatory models are underway. It has been proposed that it is a matter of sheer size, as the C subgenome is larger than the A subgenome, and smaller genomes would be more efficient (Samans et al. 2017). The higher content in transposable elements (TEs) is another possible explanation for its preferential loss (Wendel et al. 2018). In this model, genes on C are less expressed due to the proximity of silenced TEs and would therefore contribute less to gene expression (Wendel et al. 2018). In consequence, they would be more easily selected against. However, a correlation between gene expression patterns and gene retention in the different subgenomes was not found in the C subgenome itself (Parkin et al. 2014) and was also not found in other *Brassicas* to date.

The high prevalence of genomic exchanges in the *B. napus* genome makes a complex genome also highly dynamic. Genomic rearrangement can lead to deletions, duplications, translocations and inversions, and they all may encompass coding sequences and genes. Therefore, gene copy numbers do not only vary strongly between genes, but also between individuals, which may strongly affect trait variation. Copy number variants (CNVs) are therefore an important factor to consider in genomics of the allotetraploid *Brassica* species.

The use of genome structure information

Structural variants are particularly interesting because their putative phenotypic effect is expected to be strong compared to most sequence variants (Żmieńko et al. 2013). This is evident for deletions, which can be regarded as knockout mutants, while duplications can in some cases be regarded as overexpressors. The latter, however, depends on the regulatory mechanism of the gene in question. Translocations and inversions may in the first place not change anything. But as our understanding of chromatin regulation increases, genomic context will be more and more considered in predicting gene expression and phenotypic impact.

But there is more about the effect of structural variants and CNVs in polyploids than just expression level. Polyploids carry several copies of a gene, and the gene copies may have varying degrees of subfunctionalisation (Kagale et al. 2014). Therefore, the impact of a CNV in a polyploid depends strongly on the regulation mechanism, the degree of functional redundancy and on gene dosage (Miller et al. 2012; Soltis et al. 2016; Veitia and Potier 2015). Gene dosage is the dependency of a gene product's effect on its interaction partners, for example for proteins acting in a complex (Soltis et al. 2016; Veitia and Potier 2015). Higher copy numbers would therefore only have a measurable

effect if the copy numbers of the interaction partners would also increase (Veitia and Potier 2015). On the other hand, a lower copy number of one component would be limiting for the total complex (Veitia and Potier 2015). Some authors think gene dosage can partly explain heterosis, the performance advantage of heterozygous offspring compared to its homozygous parents (Schnable and Springer 2013). In autopolyploids, gene dosage might be overall conserved, while allopolyploids have a high chance of altered gene dosage. Moreover, paralogs may have different binding affinities with the interaction partners from the same subgenome than with those encoded by the other subgenome.

Modern plant breeding aims to find optimal gene configurations for better plant performance. The use of sequence variants for trait improvement has already increased the quality of respective approaches like genomic prediction (Andorf et al. 2019; Werner and Snowdon 2018). However, structural variants were mostly missed by such approaches (Gabur et al. 2019), and prediction therefore remained limited to a certain level. The reason for this failure is mainly the difficulty of detecting structural variants with high throughput techniques. Structural variants can be detected via three main principles: hybridization, PCR and sequencing (Gabur et al. 2019; Zhao et al. 2013). Hybridization assays comprise both very tedious and expensive approaches like Fluorescence in-situ hybridization (FISH) (Stein et al. 2017), but also high-throughput SNP arrays (Grandke et al. 2017). The latter are cheap and easy to use (Clarke et al. 2016; Mason et al. 2017), but they are mostly designed to call allelic variants, and are only suited to call large structural variants (Gabur et al. 2019). PCR approaches are more suited for smaller CNVs, but they are comparably expensive, have a low throughput and are therefore only a good option for confirming a specific CNV for a medium number of samples (Lee et al. 2016). Sequencing approaches do the best job among those techniques (Evans et al. 2015), but they also face drawbacks. CNV calling based on short read sequencing, for example, depends on the quality of the reference genome, and library preparation is still expensive for large populations (Gabur et al. 2019). Moreover, not all platforms are equally suited for each type of structural variant (Alkan et al. 2011). Long read sequencing technologies can overcome this and are cheaper in data production (Rang et al. 2018). However, the respective data output is huge and needs high computational power and considerable expertise to operate (Rang et al. 2018). Moreover, error rates are still comparably high, limiting the use of the reads for parallel SNP calling (Coster and van Broeckhoven 2019). To date, there is no technology offering a single solution, so combinations of those techniques are recommended for use in research and trait prediction (Alkan et al. 2011; Coster and van Broeckhoven 2019).

Flowering time regulation

In contrast to animals, plants to not bear reproductive organs in their juvenile stage, but need to form them in advance of reproduction (Walbot and Evans 2003). The phase change from the juvenile, vegetative form to the adult, reproductive form is irreversible and also marks the transition between biomass formation and biomass translation into seeds (Jaeger et al. 2006). The formation of flowers therefore needs strict genetic control to synchronize it correctly with various environmental stimuli, but also with its own age and with possible pollen donors (Srikanth and Schmid 2011). Flowering time regulation has been studied intensely in the model plant *A. thaliana*, but recent years have also seen more and more research on several *Brassica* species (Blümel et al. 2015). This regulatory system is mainly based on transcriptional regulation of the central flowering regulator *FLOWERING LOCUS T* (*FT*), which itself is regulated by at least six identified regulatory input pathways, being the vernalisation and the autonomous pathway, the photoperiod and the temperature pathway as well as the gibberellin and the age pathway (Andrés and Coupland 2012; Srikanth and Schmid 2011). Upregulation of FT protein activates downstream transcription factors governing flower formation (Alvarez-Buylla et al. 2010; Srikanth and Schmid 2011). Flower formation is then governed by the activation of the ABCE genes, a set of transcription factors orchestrating the necessary cell division and differentiation processes (Alvarez-Buylla et al. 2010). The vernalisation and the photoperiod pathway have overriding importance in regulating FT expression (Srikanth and Schmid 2011), while the other pathways have a modulating role and may only be decisive in extreme conditions. The main gene in vernalisation is the transcriptional inhibitor *FLOWERING LOCUC C (FLC),* which is highly expressed during the juvenile phase of the plant (Kim et al. 2009). Its gene product binds to the FT promotor to inhibit transcriptional activator from binding (Kim et al. 2009). Its high transcriptional activity is governed by its active chromatin state: the FRIGIDA complex (FRI-C) places active H2A.Z variants into the FLC chromatin (Choi et al. 2011; He 2009) and the PAF1-like complex adds activating methylation marks to histone H3 (Cao et al. 2015; He et al. 2004; Oh et al. 2008). During cold, the gene *VERNALISATION INSENSITIVE 3 (VIN3*) is upregulated and recruits the PRC1 complex with the help of long non-coding RNAs expressed from the *FLC* transcript itself (Turck and Coupland 2011; Wood et al. 2006), adding repressive marks to histone H3, and *FRI* is downregulated and therefore stops placing H2A.Z (Hu et al. 2014). This leads to a transcriptionally inactive chromatin state and gives way to positive regulation from the photoperiod pathway. The photoperiod pathway activates FT via the transcriptional activator *CONSTANS (CO),* which itself is mainly translationally controlled via the circadian clock and several photoreceptors (Johansson and Staiger 2015). CO protein is constantly produced, but is unstable in short day conditions and therefore unable to activate *FT* (Johansson and Staiger 2015). When a threshold day length is met, FT protein is produced in the leaves and transported to the apex, where it dimerizes with FLOWERING LOCUS D (FD) and activates

effector genes like *SUPPRESSOR OF CONSTANS 1 (SOC1), APETALA 1 (AP1), CAULIFLOWER (CAL)* and *LEAFY (LFY),* which themselves would activate the ABCE genes and induce flower formation (Alvarez-Buylla et al. 2010; Srikanth and Schmid 2011).

The functionality of several *Brassica FT* copies has been confirmed (Wang et al. 2009; Wang et al. 2012; Zhang et al. 2015), as well as several other *Brassica* homologs upstream or downstream of *FT,* like *FLC, FRI* or *CO* (Hou et al. 2012; Irwin et al. 2012; Irwin et al. 2016; Osborn et al. 1997; Ridge et al. 2015; Robert et al. 1998; Schranz et al. 2002; Tadege et al. 2001; Wang et al. 2011a; Wu et al. 2012; Xiao et al. 2013; Yuan et al. 2009; Zou et al. 2012), indicating that the flowering time gene network has been basically conserved within the *Brassicaceae*. However, depending on gene and ploidy level, most *Brassica* flowering time genes come in several copies. *FT*, for example, was found to have six copies within *B. napus*, while the major photoperiod pathway gene *CONSTANS* (*CO*) has four and the major vernalisation pathway gene *FLOWERING LOCUS C* (*FLC*) has even ten (Schiessl et al. 2017b). To understand and improve flowering time-related gene regulation in *Brassica* species, it is therefore crucial to understand which degree of neo-and subfunctionalisation can be observed between different copies of the major flowering time genes. Only on this basis, it is possible to understand the effect of sequence and structural variation of flowering time genes on seed yield building. In that respect, particular interest was raised for the vernalisation system. Oilseed rape is grown in three different life cycle variants: winter, semi-winter and spring. While winter types have a fully functional vernalisation system which normally needs at least eight weeks of cold exposure to allow flowering, semi-winter types would also flower without vernalisation, although later, and spring types normally do not respond to vernalisation and also lack the respective winter hardiness. These life cycle traits limit introgression between gene pools, therefore, suitable genetic markers would be highly welcome to speed up the process.

Chapters

Chapter 1: Copy number variation in *Brassica* flowering time genes

Copy number variation (CNV) is expected to be widespread in the *Brassica napus* genome and is therefore assumed to also influence flowering time genes. To allow reliable testing of phenotypic effects of CNVs in flowering time genes, it is necessary to call CNVs accurately in a large number of accessions. One of the most efficient methods to screen large populations for deletions and duplications is targeted deep sequencing. Targeted deep sequencing is a short read based sequencing technology with the aim to sequence only genetic regions of interest, but at higher sequencing depths. This technology can be both applied in diploids and polyploids, whereas it is particularly appealing in polyploids because it can capture all copies of a gene without prior knowledge of copy number. Using this technology, we studied flowering time gene variation with a focus on CNVs in two individuals of each progenitor species of *B. napus* as well as in a large diversity set of 280 accessions. The results, presented here in three experimental papers and one review, show the high prevalence of CNVs in flowering time genes in *B. napus* and indicate that they play a role in the formation of subspecies.

Chapter 1.1: Targeted deep sequencing of flowering regulators in *Brassica napus* reveals extensive copy number variation

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

SCIENTIFIC DATA

OPEN

Data Descriptor: Targeted deep sequencing of flowering regulators in *Brassica napus* reveals extensive copy number variation

Received: 28 October 2016
Accepted: 5 January 2017
Published: 14 March 2017

Sarah Schiessl[1], Bruno Huettel[2], Diana Kuehn[2], Richard Reinhardt[2] & Rod J. Snowdon[1]

Gene copy number variation (CNV) is increasingly implicated in control of complex trait networks, particularly in polyploid plants like rapeseed (*Brassica napus* L.) with an evolutionary history of genome restructuring. Here we performed sequence capture to assay nucleotide variation and CNV in a panel of central flowering time regulatory genes across a species-wide diversity set of 280 *B. napus* accessions. The genes were chosen based on prior knowledge from *Arabidopsis thaliana* and related *Brassica* species. Target enrichment was performed using the Agilent SureSelect technology, followed by Illumina sequencing. A bait (probe) pool was developed based on results of a preliminary experiment with representatives from different *B. napus* morphotypes. A very high mean target coverage of ~670x allowed reliable calling of CNV, single nucleotide polymorphisms (SNPs) and insertion-deletion (InDel) polymorphisms. No accession exhibited no CNV, and at least one homolog of every gene we investigated showed CNV in some accessions. Some CNV appear more often in specific morphotypes, indicating a role in diversification.

Design Type	individual genetic characteristics comparison design
Measurement Type(s)	copy number variation profiling
Technology Type(s)	DNA sequencing
Factor Type(s)	gene list • selectively maintained organism
Sample Characteristic(s)	*Brassica napus* • leaf

[1]Department of Plant Breeding, Justus Liebig University, IFZ Research Centre for Biosystems, Land Use and Nutrition, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany. [2]Max Planck Institute for Breeding Research, Carl-von-Linné-Weg 10, 50829 Cologne, Germany. Correspondence and requests for materials should be addressed to S.S. (email: sarah-veronica.schiessl@agrar.uni-giessen.de).

Background & Summary

Polyploid genomes present major challenges for DNA sequence analysis due to their high redundancy. Moreover, inter-subgenomic homology is a driving force for genomic rearrangements[1–3], including translocations, inversions, duplications, deletions or homeologous non-reciprocal translocations (HNRTs)[1]. Translocations are events where genomic blocks are transferred to another chromosomal location in the same orientation, whereas inversions switch the orientation. Duplications are events where genetic regions are copied to another locus, meaning that the affected region increases in copy number, whereas deletions involve a loss of genetic regions from the genome and therefore decrease the copy number of the genes therein. HNRTs are coupled duplication-deletion events, where one region of the genome replaces a respective homeologous genome region. Changes in the frequency of genes or other genomic loci due to duplications, deletions and HNRTs are collectively described as copy number variation (CNV).

Brassica napus (oilseed rape, canola) has become a model plant for studies of polyploidy effects[4], having been studied for genomic rearrangements over decades[5,6]. The strongest effects of genomic rearrangements, with regard to both size and abundance, were observed in synthetic, or resynthesized *B. napus*[1,2,7,8]. This indicates that substantial genomic rearrangements occur in the first generations after polyploidisation and are thereafter subjected to selection and fixation[2]. All the same, CNV can arise at any time and is therefore frequently observed[3]. With increasing reports associating CNV to major traits in different crop species, like wheat, maize and potato[9–11], their detection is growing in importance for genomics-assisted breeding approaches to improve crop adaptation and yield.

Flowering time is an extremely important adaptive trait, with wide-ranging implications for the breeding process in crops like *B. napus* that are grown in different ecogeographical regions. Flowering time determines the lifecycle to be either annual or biannual and is also important for local adaptation. Biannual and annual *B. napus* forms have divergent lifecycles, and hence must be bred in separate breeding programs, which in turn lowers diversity in these two genetic pools. While extremely important, flowering time regulation is also highly complex, with studies in model species[12,13] implicating over 200 different genes and non-coding RNAs. Polyploidy complicates this network not only by multiplying the gene copy number, but also by invoking potential for gene subfunctionalisation[14]. Knowledge of the underlying natural variation among *B. napus* flowering time regulators can help to understand which gene copies are decisive for restoring life cycle traits after introgression from a different gene pool. CNVs can play a central role in such variation[15].

Different methodologies have been proposed for the detection of CNVs, either based on hybridization arrays or on next-generation sequencing (NGS) technologies[3]. As the latter deliver additional information like SNPs and InDels, they are better suited to unravel causal variants for trait variation[16]. Although different approaches exist to calculate CNVs from NGS data, the simplest and most abundant approach is the approximation by calculation read depth or sequence coverage. The read depth approach is based on the assumption that sequencing reads distribute equally over the sequenced space, so that deviations in coverage can be interpreted to indicate deletions or duplications. However, for cost-effective application of DNA resequencing on large panels of genotypes, the sequencing power should be focused on regions of interest, like exomes or groups of genes. This can be done by prior enrichment for loci of interest by hybridization with probes or baits, and different commercial solutions are available for this[17]. Here, we used the Agilent SureSelect technology (Agilent Inc., Santa Clara, CA, USA) for target enrichment. The resulting sequence capture libraries were subjected to Illumina sequencing in single-end mode. Baits were designed for homologs of 35 *Arabidopsis thaliana* flowering genes, based on sequences from *B. napus* along with its diploid progenitors *B. rapa and B. oleracea*, and used for production of 120mer RNA oligonucleotides for hybridization. Sequencing was performed for 280 diverse inbred lines as a part of the ERANET-ASSYST *B. napus* diversity panel[18], containing annual and biannual oilseed, fodder, vegetable and rutabaga forms, along with exotic accessions. The results show a vast variance of SNPs, InDels and CNVs in the different breeding pools. An accompanying publication describes the biological implications of the data[15].

Methods

Plant Material and phenotyping

A panel of 280 genetically diverse *B. napus* inbred lines (self-pollinated for 5 or more generations) was grown in in Giessen, Germany (50° 35′ N, 8° 40′ E) in 2012. The plant material was part of the ERANET-ASSYST *B. napus* diversity set, described in detail by[18–20]. Biannuals were grown in autumn-sown trials, whereas annual accessions were grown in spring-sown trials. Plots were sown in a completely randomized block design with a harvest plot size of 3×1.25 m in 1 replicate (containing around 200 plants).

DNA isolation

Leaf material for genomic DNA extraction was harvested in spring 2012 from the field trial in Giessen, Germany. Pooled leaf samples were taken from at least 5 different plants per genotype, immediately shock-frozen in liquid nitrogen and kept at −20 °C until extraction. Leaf material was ground in liquid nitrogen with a mortar and pestle. DNA was extracted using a common CTAB protocol modified from Doyle and Doyle (1990) as described earlier[16]. DNA concentration was determined using a Qubit fluorometer and the Qubit dsDNA BR assay kit (Life Technologies, Darmstadt, Germany) according to

the manufacturer's protocol. DNA quantity and purity was further checked on 0.5% agarose gel (3 V/cm, 0.5xTBE, 120 min).

Selection of target genes

As described in ref. 16, a set of 29 *A. thaliana* flowering time genes was selected to cover the entire genetic network controlling flowering time, including circadian clock regulators *(CYCLING DOF FACTOR 1 (CDF1), EARLY FLOWERING 3 (ELF3), GIGANTEA (GI)* and *ZEITLUPE (ZTL)*), the input pathways for vernalisation (*EARLY FLOWERING 7 (ELF7), EARLY FLOWERING IN SHORT DAYS (EFS), FLOWERING LOCUS C (FLC), FRIGIDA (FRI), SHORT VEGTATIVE PHASE (SVP), SUPPRESSOR OF FRIGIDA 4 (SUF4), TERMINAL FLOWER 2 (TFL2), VERNALISATION 2 (VRN2), VERNALISATION INSENSITIVE 3 (VIN3)),* photoperiod sensitivity (*CONSTANS (CO), CRYPTOCHROME 2 (CRY2), PHYTOCHROME A (PHYA), PHYTOCHROME B (PHYB))* and gibberellin (*GIBBERELLIN-3-OXIDASE 1 (GA3ox1)*), along with downstream signal transducers (*AGAMOUS-LIKE 24 (AGL24), APETALA 1 (AP1), CAULIFLOWER (CAL), FLOWERING LOCUS D (FD), FLOWERING LOCUS T (FT), FRUITFUL (FUL), LEAFY (LFY), SQUAMOSA PROMOTOR PROTEIN LIKE 3 (SPL3), SUPPRESSOR OF CONSTANS 1 (SOC1), TEMPRANILLO 1 (TEM1), TERMINAL FLOWER 1 (TFL1)).* On top, we also included 6 further genes: *CIRCADIAN CLOCK ASSISTED 1 (CCA1), FLAGELLIN-SENSITIVE 2 (FLS2), GLYCIN-RICH PROTEIN 7 (GRP7), GLYCIN-RICH PROTEIN 8 (GRP8), GORDITA (GORD)* and *SENSITIVITY TO RED LIGHT REDUCED 1 (SRR1),* giving a total of 35 genes. The respective Arabidopsis identifiers are given in Supplementary Table S1.

Bait development

Before resequencing the total set of 280 accessions, 4 accessions representing divergent morphotypes were resequenced in a preliminary experiment in order to refine the bait pool (described in ref. 16). At the time, no reference sequence for *B. napus* was available. Baits were therefore produced on sequences of *B. rapa* and *B. oleracea*, using the program eArrayXD. The only exception was the target gene *FT*, for which promotor and gene sequences from *B. napus* were kindly made available by Carlos Molina, Christian Albrecht University of Kiel, Germany. A pre-publication draft (version 4.0) of the *B. napus* 'Darmor-Bzh' reference genome sequence assembly became available prior to public release, by generosity of Boulos Chalhoub, INRA, France, Unité de Recherche en Génomique Végétale.

Based on the mapping results of the preliminary experiment, the bait pool was modified in order to improve specificity. Enriched regions found in ref. 16 were classified into target regions and non-target regions by BLAST against the nucleotide database in NCBI. The bait pool was blasted against target and non-target regions (E-value cut-off e^{-10}). Baits which had excessive non-target hits were manually removed. This was the case for bait groups on the target genes *FT, FUL* and *PHYA*. For some bait groups (*AP1, CO, SOC1*), too many baits (>30%) were deleted. In these cases, baits groups (120mer oligonucleotide sequences) were created using the *B. napus* pre-publication draft reference genome sequence assembly, with the Agilent Genomic Workbench program SureDesign (Agilent Inc., Santa Clara, CA, USA). These replaced the corresponding bait groups developed previously using *B. rapa* or *B. oleracea*. Bait groups were created using the 'Bait Tiling' tool. The parameters were set as follows: Sequencing Technology: 'Illumina', Sequencing Protocol: 'Paired-End long Read (75 bp+)', 'Use Optimized Parameters (Bait length 120, Tiling Frequency 1x)', Avoid Overlap: '20', 'User defined genome', 'Avoid Standard Repeat Masked Regions'. Baits for genes on the minus-strand were developed in sense, while baits on the plus-strand were developed in antisense.

In total, 63 bait groups were created for *B. rapa* homologs of the target genes, 71 bait groups for *B. oleracea* homologs and 24 bait groups for *B. napus* homologs.

Sequence capture and sequencing

Custom bait production was carried out by Agilent Technologies (Agilent Inc., Santa Clara, CA, USA) using the output oligonucleotide sequences from SureDesign. Sequence capture was performed using the SureSelectXT 1 kb–499 kb Custom Kit (Agilent Inc., Santa Clara, CA, USA) according to the manufacturer's instructions. The resulting TruSeq DNA library (Illumina Inc., San Diego, CA, USA) was sequenced on an Illumina HiSeq 2500 sequencer at the Max Planck Institute for Breeding Research (Cologne, Germany) in 100 bp single read mode.

Sequence data analysis

Mapping. Quality control of the raw sequencing data was performed using FASTQC. Reads were mapped onto version 4.1 of the *B. napus* 'Darmor-Bzh' reference genome sequence assembly[21]. Mapping was performed using the software SOAP2 (http://soap.genomics.org.cn/soapaligner.html), with default settings and alternatively using the option r = 0 to extract uniquely aligned reads. Removal of duplicates, sorting and indexing was carried out with *samtools* version 0.1.19 (http://samtools.sourceforge.net/). Alignments were visualised using the IGV browser version 2.3.12 (http://www.broadinstitute.org/igv/).

The mean read number was 5.8 M reads, with a s.d. of 1.6 M reads. Only 3% of the sequenced samples had less than 4.2 M reads (mean–s.d.), 11% of the samples had more than 7.3 M reads (mean+s.d.). The mean alignment rate was 87%, while the mean unique alignment rate was 75%, indicating that most of

the reads could be mapped specifically (see Data Citation 1 for.bam files). The exact values, along with values for the different subsets, can be found in Table 1.

Specificity and sensitivity

In total, the captured reads aligned to 1184 distinct regions of the *B. napus* 'Darmor-Bzh' v4.1 reference genome with a mean coverage of >10. Of these, a total of 637 regions were annotated as genes, and 184 corresponded to the intended target genes. Two target gene copies (*Bna.VIN3.A01* and *Bna.VIN3.C01*) had insufficient coverage, but were nevertheless included in the target list. A total of 33 regions were identified as targets giving a BLAST hit to the *FT* promotor. A further 12 regions were identified as pseudogenes of the target genes by blasting the gene sequences to the genome. Therefore, the target included 231 regions. The average target sensitivity[17], interpreted as the percentage of target bases covered by sequence reads, was 85.6%. The average target specifity[17], or the percentage of bases mapping to the intended target, was 68.0%, and therewith indeed better than in the foregoing experiment[16]. The target regions had a mean coverage of 670x and a mean normalized coverage of 533x. The mean target coverage over the mean genome-wide coverage, also called the enrichment factor, was about 1206, and therefore also increased compared to the preliminary experiment[16] (for an overview of values see Table 2).

Detection of SNPs and InDels

Calling of single nucleotide polymorphisms (SNPs) was performed with the algorithm mpileup in the *samtools* toolkit. SNPs were filtered for a minimum mapping quality of 50 and a read depth of ≥10, using vcftools (https://vcftools.github.io/man_latest.html).

For InDel calling, a separate mapping was performed using Bowtie2 (http://sourceforge.net/projects/bowtie-bio/files/bowtie2/), as described in ref. 22. Removal of duplicates, sorting and indexing was carried out with *samtools* version 0.1.19 (http://samtools.sourceforge.net/). An initial InDel calling was performed using *samtools* mpileup, and realignment of reads around InDels was performed using GATK (https://www.broadinstitute.org/gatk/). A final InDel calling was then performed as described above. InDels were filtered for a minimum mapping quality of 30 and a read depth of 10 or more using vcftools (https://vcftools.github.io/man_latest.html). SNP and InDel annotation was performed using CooVar[23].

	total reads	alignment rate [%]	unique alignment rate [%]
mean	5768702	87.00	75.11
winter	5346938	89.21	77.04
semi	7023642	82.56	71.30
spring	6096897	84.82	73.25
swede	7038926	83.84	71.85
min	3681313	70.50	58.97
winter	4133157	74.85	63.33
semi	4298795	70.50	61.49
spring	3681313	72.18	58.97
swede	4691991	81.73	69.25
max	13869129	92.51	81.52
winter	9010533	92.51	81.52
semi	13839526	88.23	76.42
spring	13869129	91.37	81.37
swede	13724355	87.69	75.37
std dev	1581627	3.38	3.48
winter	910522	2.03	2.55
semi	3153296	5.16	4.28
spring	1807938	2.72	3.08
swede	2355237	1.73	1.85

Table 1. Sequencing and alignment values for total read number, alignment rate and unique alignment rate. The table lists the arithmetic mean, the minimal and maximal values as well as the s.d. for the total set (first line) and the respective subsets of winter type rapeseed, semi-winter rapeseed, spring type rapeseed and swedes.

	Total	winter	semi	spring	swede
Mean genome-wide coverage	0.6	0.5	0.6	0.6	0.7
Mean target coverage	668.3	641.5	754.7	688.0	756.3
Enrichment factor	1205.9	1212.7	1199.0	1202.2	1158.9
Normalized mean target coverage	532.7	541.6	522.4	524.8	507.0
Fraction of target covered (%)	85.6	86.2	84.9	84.9	85.0
Reads covering target (%)	68.0	69.2	66.4	67.2	64.8
Genome fraction covered by >10 reads (%)	0.2	0.2	0.2	0.2	0.3
Fraction of target covered by > 10 reads (%)	79.0	79.8	78.0	78.1	78.7

Table 2. Quality measures for the target enrichment for the total population and its subsets. The table lists mean values for genome-wide and target coverage, the enrichment factor calculated from that, the normalized mean target coverage, the fractions of the genome and the target covered by any read and by at least 10 reads.

We called and annotated 13053 SNPs among the 1184 captured regions, of which 5,216 were located in the target regions. Of those, 56 SNPs were either radical mutations, splice variants or stop codon mutations (gain or loss). InDel calling revealed a total of 1894 InDels, with 569 in the target regions. Only 25 InDels were frameshifts, amino acid insertions or splice variants. Potential functional variation was revealed in all homolog groups, although 7 homologs were completely conserved, while 16 carried only silent or synonymous variation. The distribution of variants on target and non-target regions is shown in Figure 1.

CNV detection

Enriched regions and coverage differences were calculated using the *bedtools* software with multiBamCov (http://bedtools.readthedocs.org/en/latest/). Read coverage for each enriched region was normalised as follows:

coveragenorm = (number of aligned reads per region*total length of genome)/(total number of aligned reads*region length).

Copy number variation (CNV) in a given region was assumed if the ratio of normalised coverage(genotype)/mean normalised coverage (all genotypes) was smaller than 0.5 or higher than 1.5, respectively.

No homolog group was found without CNV, and only two accessions were found which did not carry any CNV among the homologous copies of the target genes. Whereas it is therefore very unlikely to find two genotypes with exactly the same copy numbers, the frequency of a specific CNV among the population is generally low; about 87% of the CNVs had an abundance across the whole population of $< 10\%$. Deletions were generally more abundant than duplications (Figure 2) both for the A and C subgenomes, regarding all regions. This is expected, as genomes are known to reduce their gene space after polyploidization[24], and corresponds to similar findings from whole-genome resequencing in a diverse panel of synthetic and natural *B. napus* (Samans *et al.*, submitted). Regarding only genes, duplication and deletion events were almost twice as frequent among C subgenome than A subgenome homologs. This appears surprising, as although the C subgenome is significantly larger[25,26], its gene content is comparable to that of the A subgenome. A possible explanation for this apparent genome bias might be that transposons are more active in the C subgenome[26], increasing the frequency of small-scale homology and double-strand breaks. Considering only the target genes from the flowering-time network, the ratio was strikingly different, with more duplications of A-subgenome genes than C-subgenome genes, but more deletions of C-subgenome genes than A-subgenome target genes. This indicates that A-subgenome copies had been selected over C-subgenome copies, while the other direction is less frequent. A corresponding bias was also observed by Samans *et al.* (submitted), who suggested that the size difference between homeologous segments from the A and C subgenomes is the driving force for the directional bias, with HNRT tending to replace larger C-subgenome segments by their smaller homoeologous segments from the A-subgenome. Another possible cause could be that gene copies which co-evolved in the progenitor genomes interact more effectively with each other than they do with copies from the other progenitor, a mechanism already assumed by other authors[27]. C-subgenome copies might have been affected more often due to the high transposon content, invoking a selective advantage for homeologous exchanges that replaced C-subgenome homologs with A-subgenome copies. As flowering is a major lifecycle axis, flowering genes are under strong selection and should show such an effect more clearly than random genes.

Homeologous gene replacements associated with putative HNRTs were observed frequently among the dataset. Excluding gene copies on non-assembled reference chromosomes ('x_random'), we found

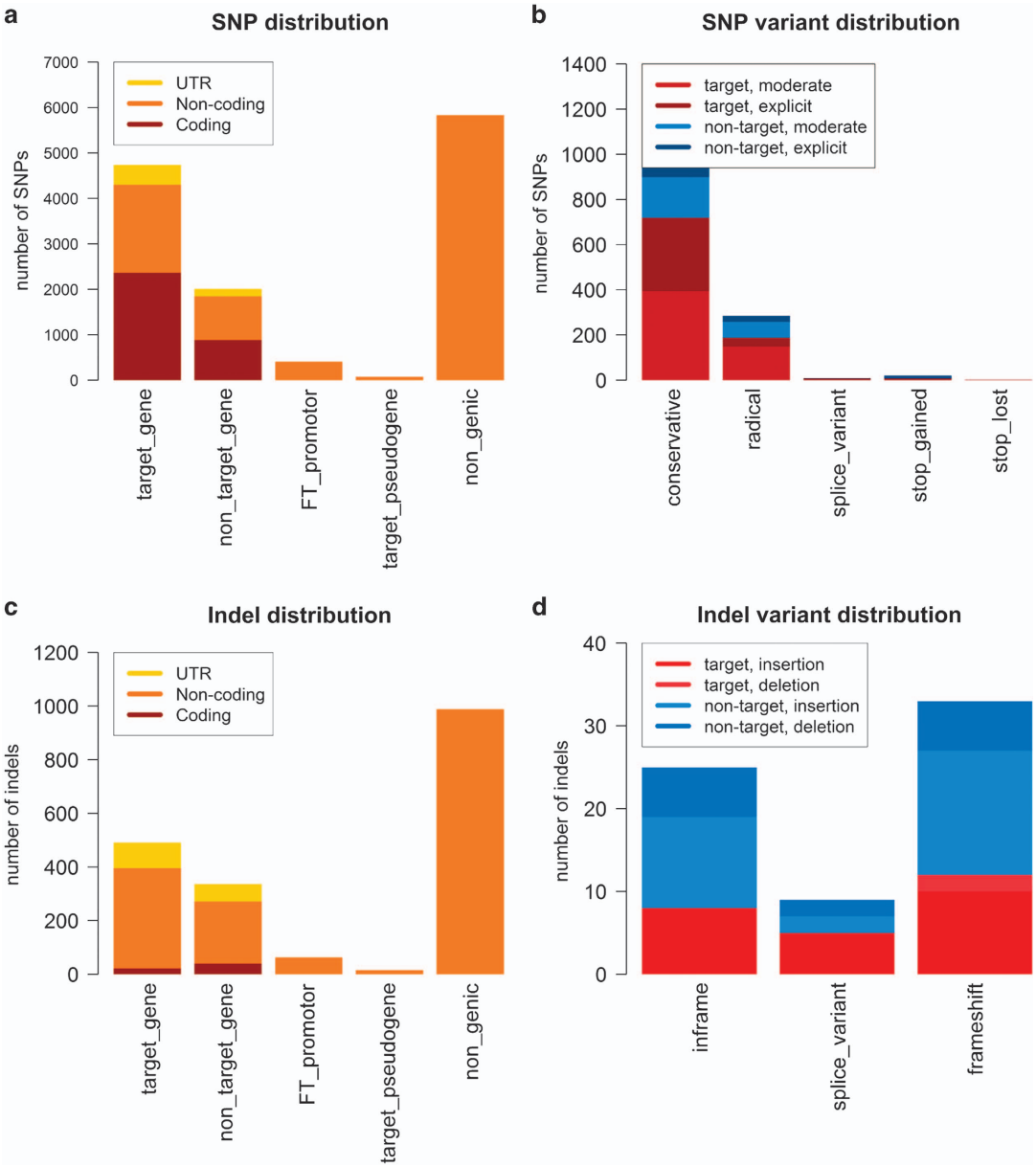


Figure 1. SNP and InDel distribution for all called SNPs and InDels meeting minimum quality requirements in the 1184 regions captured by deep sequencing. (**a**,**c**) Distribution to all analyzed regions and respective genic regions for SNPs (**a**) and InDels (**c**). (**b**,**d**): Distribution of nonsynonymous SNPs and InDels to different annotation classes regions for SNPs (**b**) and InDels (**d**).

201 putative HNRT events between homeologous target gene copies, including 165 from A-subgenome to C-subgenome homeologs and 36 from C-subgenome to A-subgenome homeologs. In contrast, we found 448 simple duplications for the same gene copies (259 on A-subgenome and 189 on C-subgenome homeologs) and 490 simple deletions (139 on A-subgenome and 351 on C-subgenome homeologs). Although there is a low but significant correlation ($r = 0.68^{***}$) between the number of simple deletions and duplications on the C subgenome, such a correlation was not observed for the A subgenome. The highest rate of HNRTs was found between the strongly homeologous chromosomes A02/C02 and A03/C03, a finding in line with other authors[1]. On the other side of the spectrum, gene copies on chromosome A08 could not be associated with HNRTs to any homeologous copies, and only one putative HNRT each was observed for chromosomes A05 and A06, respectively. High HNRT frequencies were only observed at the very ends of chromosomes, whereas gene copies located closer to the center of the chromosomes were generally not involved in exchanges (Figure 3). Similar patterns have been frequently observed in *Brassicaceae*[1,28].

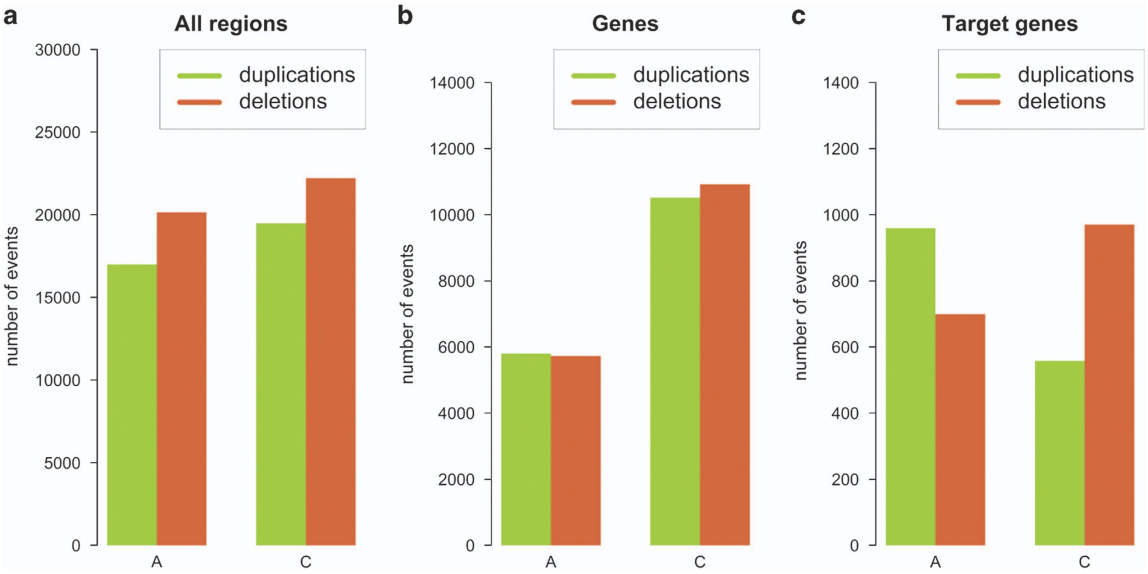


Figure 2. **Distribution of deletions and duplications across the A and C subgenomes, respectively, among the 280 genotypes of the *B. napus* sequence capture diversity panel for (a) all captured regions, (b) all captured genes, and (c) all target genes.**

Code availability

All custom codes are supplied at https://github.com/ 'targeted-deep-sequencing'.

Data Records

The aligned sequence data (.bam files) and metadata are stored at the NCBI Sequence Read Archive database, SRP-Study accession SRP087610 (Data Citation 1).

Technical Validation

All biallelic SNP, InDel and CNV data were recoded and included in a population structure analysis together with data from the Brassica 60 K Illumina Infinium SNP array (Clarke *et al.* 2016, accepted, Mason *et al.* 2016, accepted). Altogether these include 28,698 markers from the 60k SNP array and 12776 SNPs, 1894 InDels and 366 CNV markers from the present deep sequencing dataset, making a total of 43,733 markers. After pre-processing to filter for non-missing marker values >0.9, minor allele frequency >0.01 and non-missing individual markers >0.8, a total of 33,944 markers and 271 individuals were left, with 30.1% of the markers coming from the deep sequencing data. The subsequent principal component analysis for population structure revealed three main clusters. One cluster contained 137 winter-type *B. napus* accessions, one contained 93 spring-types and one semi-winter type accession, and the last cluster contained 40 genotypes of mixed origin, comprising mostly beet-forming and semi-winter types. We conclude i) that the variants called from the present deep sequencing dataset represent the species-wide diversity present in *B. napus*, in accordance with genotyping data from the SNP genotyping array, and ii) that the data correctly describe the study population.

In cooperation with Eleri Tudor from the John Innes Centre, Norwich, the data were compared to three Sanger-sequenced copies of *Bna.FRI* for 13 (A03 copy), 10 (A10 copy) and 9 (C03 copy) of our genotypes (unpublished data). The overall concordance rate between the data sets was 86%. 5 SNPs on the A03 copy had extremely low concordance rates, possibly due to a mapping problem. Removing those lead to an overall concordance rate of 90%.

We also checked for co-localizing SNPs between the datasets from the SNP genotyping array and from the deep sequencing. A total of 48 suitable SNP pairs were identified which showed apparently corresponding chromosome positions according to the BLAST search. Among this set, 3 SNP pairs had absolutely no correlation, presumably due to incorrect BLAST positions of the array SNP. For the remaining 45 SNPs, we observed a mean concordance of 91.3%, ranging from 83.0 to 97.7% with a standard deviation of 3.5%. These minor discrepancies may arise from the use of different selfing generations of some individuals for the array genotyping and the deep-sequencing. On the other hand, we noticed that concordance was lowest for SNPs discriminating between C and T, potentially suggesting a differential methylation of cytosine nucleotides.

Usage Notes

The plant material described in this paper is publically available and can be made accessible upon request.

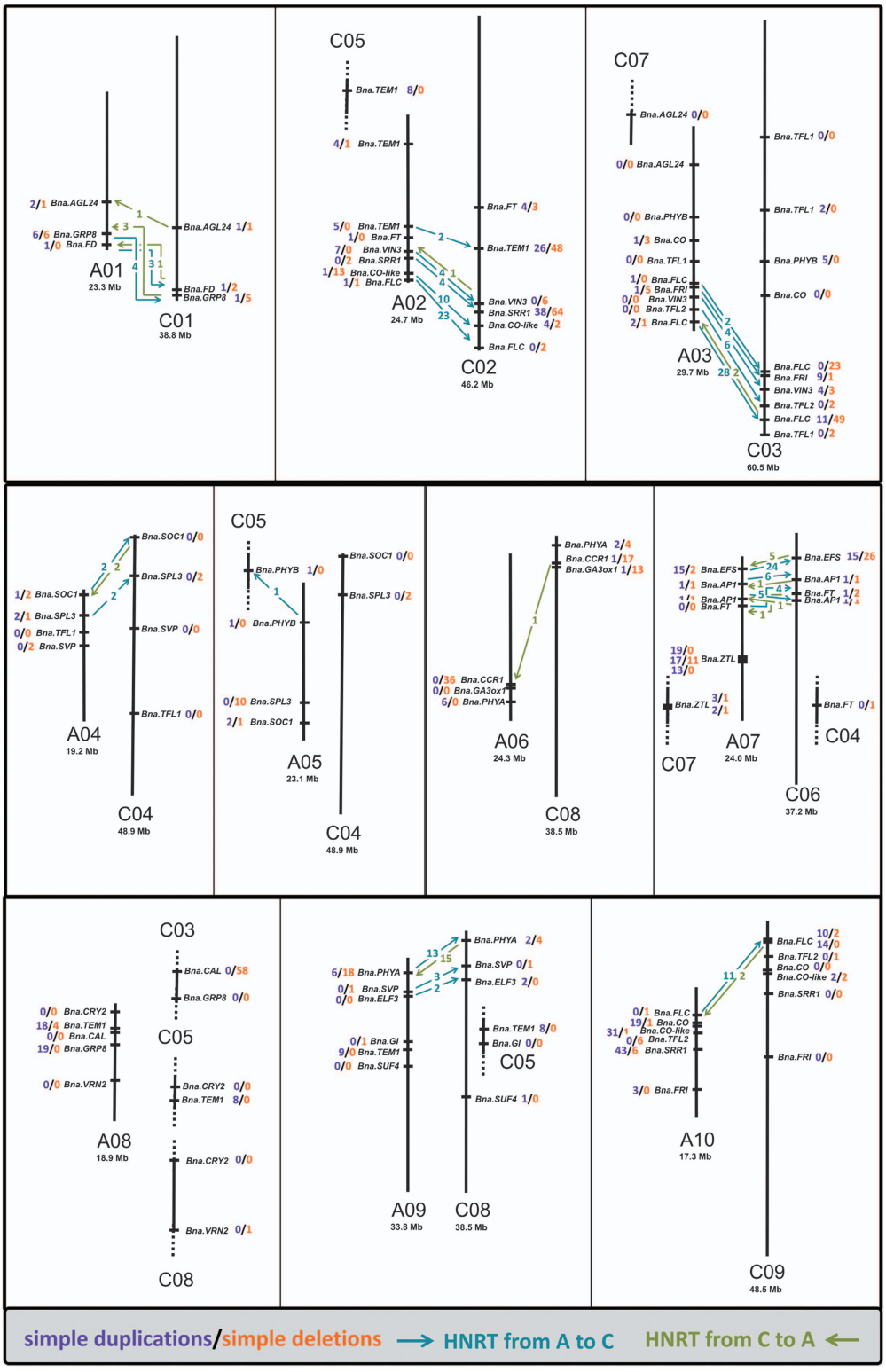


Figure 3. Landscape of homeologous target gene copies (excluding copies on x_random chromosomes). Chromosomes are shown as black vertical lines. Partial chromosomes are indicated by dotted lines at their ends. For chromosomes shown in full length, the total length is given below the chromosome name. Genetic positions are relative approximations. Arrows indicate HNRTs between homeologous copies, either from A-subgenome to C-subgenome homeologs (blue) or from C-subgenome to A-subgenome homeologs (green). The numbers behind the gene names indicate the number of simple duplication or deletion events (not involved in HNRTs) in purple and orange, respectively.

References

1. Gaeta, R. T. & Pires, J. C. Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytol* **186,** 18–28 (2010).
2. Gaeta, R. T., Pires, J. C., Iniguez-Luy, F., Leon, E. & Osborn, T. C. Genomic Changes in Resynthesized *Brassica napus* and Their Effect on Gene Expression and Phenotype. *The Plant Cell* **19,** 3403–3417 (2007).
3. Żmieńko, A., Samelak, A., Kozłowski, P. & Figlerowicz, M. Copy number polymorphism in plant genomes. *Theor Appl Genet* 1–18 (2013).
4. Mason, A. S. & Snowdon, R. J. Oilseed rape: learning about ancient and recent polyploid evolution from a recent crop species. *Plant biology (Stuttgart, Germany)* **18,** 883–892 (2016).
5. Lukens, L., Zou, F., Lydiate, D., Parkin, I. A. & Osborn, T. C. Comparison of a *Brassica oleracea* Genetic Map With the Genome of *Arabidopsis thaliana*. *Genetics* **164,** 359–372 (2003).
6. Nicolas, S. D., Monod, H., Eber, F., Chevre, A.-M. & Jenczewski, E. Non-random distribution of extensive chromosome rearrangements in *Brassica napus* depends on genome organization. *Plant J* **70,** 691–703 (2012).
7. Pires, J. C. *et al.* Flowering time divergence and genomic rearrangements in resynthesized Brassica polyploids (Brassicaceae). *Biol J Linn Soc* **82,** 675–688 (2004).
8. Szadkowski, E. *et al.* Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. *New Phytol* **191,** 884–894 (2011).
9. Díaz, A. *et al.* Copy number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*). *PLoS ONE* **7,** 1–11 (2012).
10. Springer, N. M. *et al.* Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet* **5,** e1000734 (2009).
11. Iovene, M., Zhang, T., Lou, Q., Buell, C. R. & Jiang, J. Copy number variation in potato - an asexually propagated autotetraploid species. *Plant J* **75,** 80–89 (2013).
12. Andrés, F. & Coupland, G. The genetic basis of flowering responses to seasonal cues. *Nat Rev Genet* **13,** 627–639 (2012).
13. Blümel, M., Dally, N. & Jung, C. Flowering time regulation in crops-what did we learn from Arabidopsis? *Current opinion in biotechnology* **32,** 121–129 (2015).
14. Renny-Byfield, S. & Wendel, J. F. Doubling down on genomes: polyploidy and crop plants. *American Journal of Botany* **101,** 1711–1725 (2014).
15. Schiessl, S., Hüttel, B., Kühn, D., Reinhardt, R. & Snowdon, R. J. Post-polyploidisation morphotype diversification associates with gene copy number variation. *Sci Rep.* **7,** 41845 (2017).
16. Schiessl, S., Samans, B., Hüttel, B. & Reinhardt, R. & Snowdon, R. J. Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.* **5,** 404 (2014).
17. Mamanova, L. *et al.* Target-enrichment strategies for next-generation sequencing. *Nat Meth* **7,** 111–118 (2010).
18. Bus, A., Korber, N., Snowdon, R. J. & Stich, B. Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. *Theor Appl Genet* **123,** 1413–1423 (2011).
19. Bus, A., Hecht, J., Huettel, B., Reinhardt, R. & Stich, B. High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. *BMC Genomics*, doi: 10.1186/1471-2164-13-281 (2012).
20. Körber, N. *et al.* Seedling development in a *Brassica napus* diversity set and its relationship to agronomic performance. *Theor Appl Genet* **125,** 1275–1287 (2012).
21. Chalhoub, B. *et al.* Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345,** 950–953 (2014).
22. Schmutzer, T. *et al.* Species-wide genome sequence and nucleotide polymorphisms from the model allopolyploid plant *Brassica napus*. *Scientific data* **2,** 150072 (2015).
23. Vergara, I. A., Frech, C. & Chen, N. CooVar: co-occurring variant analyzer. *BMC Research Notes* **5,** 615 (2012).
24. Sankoff, D., Zheng, C. & Zhu, Q. The collapse of gene complement following whole genome duplication. *BMC Genomics* **11,** 313 (2010).
25. Wang, X. *et al.* The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* **43,** 1035–1039 (2011).
26. Parkin, Isobel A. P. *et al.* Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biol* **15,** doi: 10.1186/gb-2014-15-6-r77 (2014).
27. Schnable, P. S. & Springer, N. M. Progress Toward Understanding Heterosis in Crop Plants. *Annu. Rev. Plant Biol.* **64,** 71–88 (2013).
28. Drouaud, J. *et al.* Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination 'hot spots'. *Genome Research* **16,** 106–114 (2006).

Data Citations

1. *NCBI Sequence Read Archive* SRP087610 (2017).

Acknowledgements

This work was financed by grant SN 14/13-1 from the German Research Foundation (DFG) within the priority programme 'Flowering Time Control: From Natural Variation to Crop Improvement.'

Author Contributions

S.S. and R.J.S. planned the experiments and wrote the manuscript. S.S. performed DNA extraction and data analysis. B.H., D.K. and R.R. performed library preparation and sequencing. We thank Eleri Tudor from the John Innes Centre, Norwich, for her courtesy to assist in validating SNP data for copies of *Bna.FRI*.

Additional Information

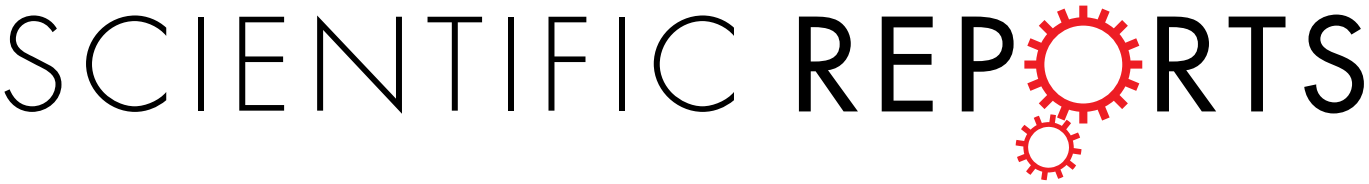
Supplementary information accompanies this paper at http://www.nature.com/sdata

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Schiessl, S. *et al.* Targeted deep sequencing of flowering regulators in *Brassica napus* reveals extensive copy number variation. *Sci. Data* 4:170013 doi: 10.1038/sdata.2017.13 (2017).

Chapter 1.2: Post-polyploidisation morphotype diversification associates with gene copy number variation

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



OPEN

Post-polyploidisation morphotype diversification associates with gene copy number variation

Received: 30 September 2016
Accepted: 03 January 2017
Published: 06 February 2017

Sarah Schiessl[1], Bruno Huettel[2], Diana Kuehn[2], Richard Reinhardt[2] & Rod Snowdon[1]

Genetic models for polyploid crop adaptation provide important information relevant for future breeding prospects. A well-suited model is *Brassica napus*, a recent allopolyploid closely related to *Arabidopsis thaliana*. Flowering time is a major adaptation trait determining life cycle synchronization with the environment. Here we unravel natural genetic variation in *B. napus* flowering time regulators and investigate associations with evolutionary diversification into different life cycle morphotypes. Deep sequencing of 35 flowering regulators was performed in 280 diverse *B. napus* genotypes. High sequencing depth enabled high-quality calling of single-nucleotide polymorphisms (SNPs), insertion-deletions (InDels) and copy number variants (CNVs). By combining these data with genotyping data from the Brassica 60 K Illumina® Infinium SNP array, we performed a genome-wide marker distribution analysis across the 4 ecogeographical morphotypes. Twelve haplotypes, including *Bna.FLC.A10*, *Bna.VIN3.A02* and the *Bna.FT* promoter on C02_random, were diagnostic for the diversification of winter and spring types. The subspecies split between oilseed/kale (*B. napus* ssp. *napus*) and swedes/rutabagas (*B. napus* ssp. *napobrassica*) was defined by 13 haplotypes, including genomic rearrangements encompassing copies of *Bna.FLC*, *Bna.PHYA* and *Bna.GA3ox1*. *De novo* variation in copies of important flowering-time genes in *B. napus* arose during allopolyploidisation, enabling sub-functionalisation that allowed different morphotypes to appropriately fine-tune their lifecycle.

Polyploid crops like wheat, potato, oats and rapeseed have been enormously successful as field crops because of their huge adaptation potential. Indeed, the fact that all flowering plants derive from ancient or recent polyploidisation events[1,2] points to an enormous evolutionary advantage associated with polyploidy. On the other hand, most polyploid events do not lead to a successful establishment of a new species[3]. Understanding how polyploids achieve adaptive potential has important implication for breeding in the context of environmental change.

On the other hand, the complexity of polyploid genomes has considerably restricted large-scale genetic studies of polyploid species[4–6], so broad conclusions are often drawn based on diploid model plants like *Arabidopsis thaliana*. The polyploid crop most closely related to *A. thaliana* is rapeseed (*Brassica napus*), making it an excellent system to transfer information from the model to the crop. Despite its very recent origin and strong allopolyploidisation bottleneck[6], rapeseed can be grown from boreal to subtropical and semi-arid areas, a result of strong differentiation into distinctly different morphotypes[7].

The morphotype with highest seed yields is the biannual winter oilseed type[8]. The prerequisites for this lifecycle are winter hardiness for winter survival, along with vernalisation requirement to avoid pre-winter flowering[7]. In subtropical areas, cultivation of semi-winter types that can be vernalised in warmer temperatures is possible[7]. Boreal or semi-arid regions have periods of low plant survival rates, either due to strong winter freezing or extreme heat stress. In these regions, annual spring types are prominent. These are neither winter-hardy nor vernalisation-dependent, and the short growing season strongly limits yield potential. *B. napus* can also be grown as beet-like forms, known as swedes or rutabagas, which form a different subspecies (ssp. *napobrassica*)[7]. Swedes are generally of winter type, however have limited winter-hardiness and require extended vernalisation to flower (Fig. 1). No wild-types of *B. napus* are known, hence the species is assumed to have arisen in cultivation[7], with at least one origin believed to be as recent as a few hundred years ago[9]. The different cultivated forms are bred in separate breeding pools, with introgression between morphotypes only in cases of extreme introgression benefit.

[1]Department of Plant Breeding, Justus Liebig University, IFZ Research Centre for Biosystems, Land Use and Nutrition, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany. [2]Max Planck Institute for Breeding Research, Carl-von-Linné-Weg 10, 50829 Cologne, Germany. Correspondence and requests for materials should be addressed to S.S. (email: sarah-veronica.schiessl@agrar.uni-giessen.de)

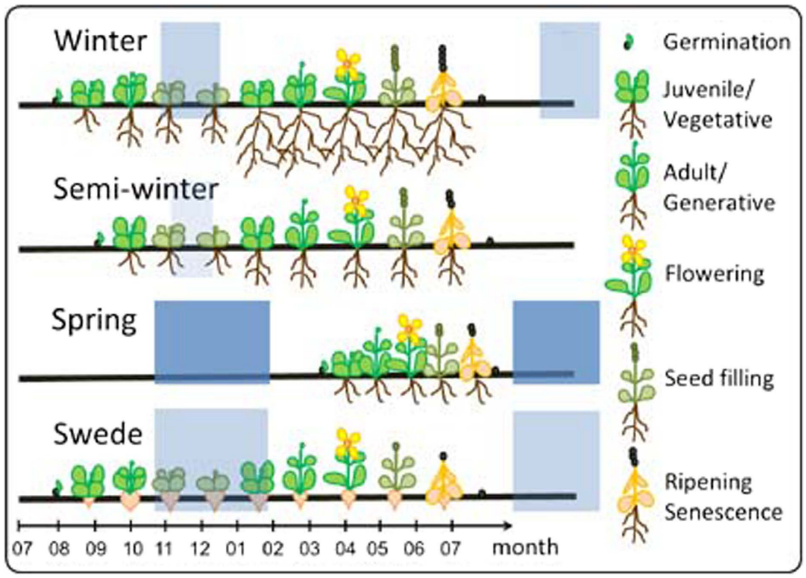


Figure 1. Schematic representation of the life cycles of the four different *Brassica napus* morphotypes. Periods of cold required for vernalisation in the respective morphotypes are indicated by blue boxes. Relative seed production is indicated by the number of grains.

However, this necessitates tedious backcrossing programs to restore the required ecogeographic adaptation characters[10]. Knowledge of the factors determining lifecycle traits like vernalisation requirement and flowering time is crucial for successful exchange of genetic material between *B. napus* gene pools[10].

Although the mechanisms of vernalisation have been studied in depth in Arabidopsis, specific winter or spring alleles were not yet defined for *B. napus*. The predominant assumption is that the underlying genetic mechanisms are identical or very similar across crucifer species. The allopolyploid *B. napus* carries two almost intact subgenomes from the ancestors *Brassica rapa* (A subgenome donor) and *Brassica oleracea* (C subgenome donor). Both ancestral subgenomes arose from a common, hexaploid ancestor, raising the theoretical copy number of Arabidopsis gene homologs to six. Due to post-polyploidisation genome reduction, the average gene copy number is 4.4[11], whereby considerable variation has been observed among different gene families, with copy number ranging from 1 to 12[12]. Homology-driven chromosome rearrangements during allopolyploidisation are a key driver of such variation[6,12]. Copy number variations (CNVs) have been found to impart large phenotypic influence in several plant species like Arabidopsis[13], wheat[14], potato[15] and maize[16], but also in domestic animals[17] and humans[18].

In Arabidopsis, *FLOWERING LOCUS C* (*FLC*) is the major repressor for the activity of the central flowering transcription factor *FLOWERING LOCUS T* (*FT*)[19]. This gene cannot be expressed before FLC protein levels drop[19], however when this occurs *FT* can be activated by the photoperiod pathway via the transcription factor *CONSTANS* (*CO*)[20]. Downregulation of *FLC* takes place at the transcriptional level. The FLC chromatin is modified and rearranged in order to stabilize a new inactive form[21,22]. Different mechanisms are involved in the structural regulation of *FLC* gene activity, including both autonomous regulators and the vernalisation pathway[22,23]. Three different mechanisms may exist for the breakdown of vernalisation requirement: (i) alteration of *FLC* regulating factors like *FRIGIDA* (*FRI*); (ii) alteration of *FLC* gene sequence or activity; (iii) alteration of *FLC* binding sites or *FT* promoter sequences. Arabidopsis annuals and biannuals have been found to differ either in *FRI* or in *FLC*[24], indicating that the Arabidopsis winter-spring split is governed by the first two levels of regulation. As a consequence, research on *B. napus* vernalisation has been heavily focused on investigating *FLC* homologs[25–27]. Indeed, a number of QTL studies in different mapping populations have suggested *FLC* loci as candidates for flowering time in *B. napus*, including populations without vernalisation requirement[28–32]. Moreover, it has been reported that a transposon insertion in the first intron of *Bna.FLC.A10* is associated with the vernalisation requirement of winter-type rapeseed[27].

The aim of the present work was therefore the definition of morphotype-specific alleles or haplotypes that might further our understanding of vernalisation control in a complex allopolyploid, and simultaneously allow breeders to successfully select for desirable lifecycle traits. By comparing results of vernalisation experiments with data from genome-wide marker distribution analysis, targeted deep-sequencing of essential flowering time regulators and the *FT* promoter, and coverage analysis to estimate CNV, we provide novel insights that reveal the complexity of post-polyploidisation morphological diversification in an important crop species.

Material and methods

Plant material and phenotyping. A panel of 280 genetically diverse *B. napus* inbred lines (selfed for 5 or more generations) was grown in Giessen, Germany (50° 35′ N, 8° 40′ E) in 2012. The plant material was part of the ERANET-ASSYST *B. napus* diversity set that has been described previously[33,34]. Winter-type rapeseed and swede

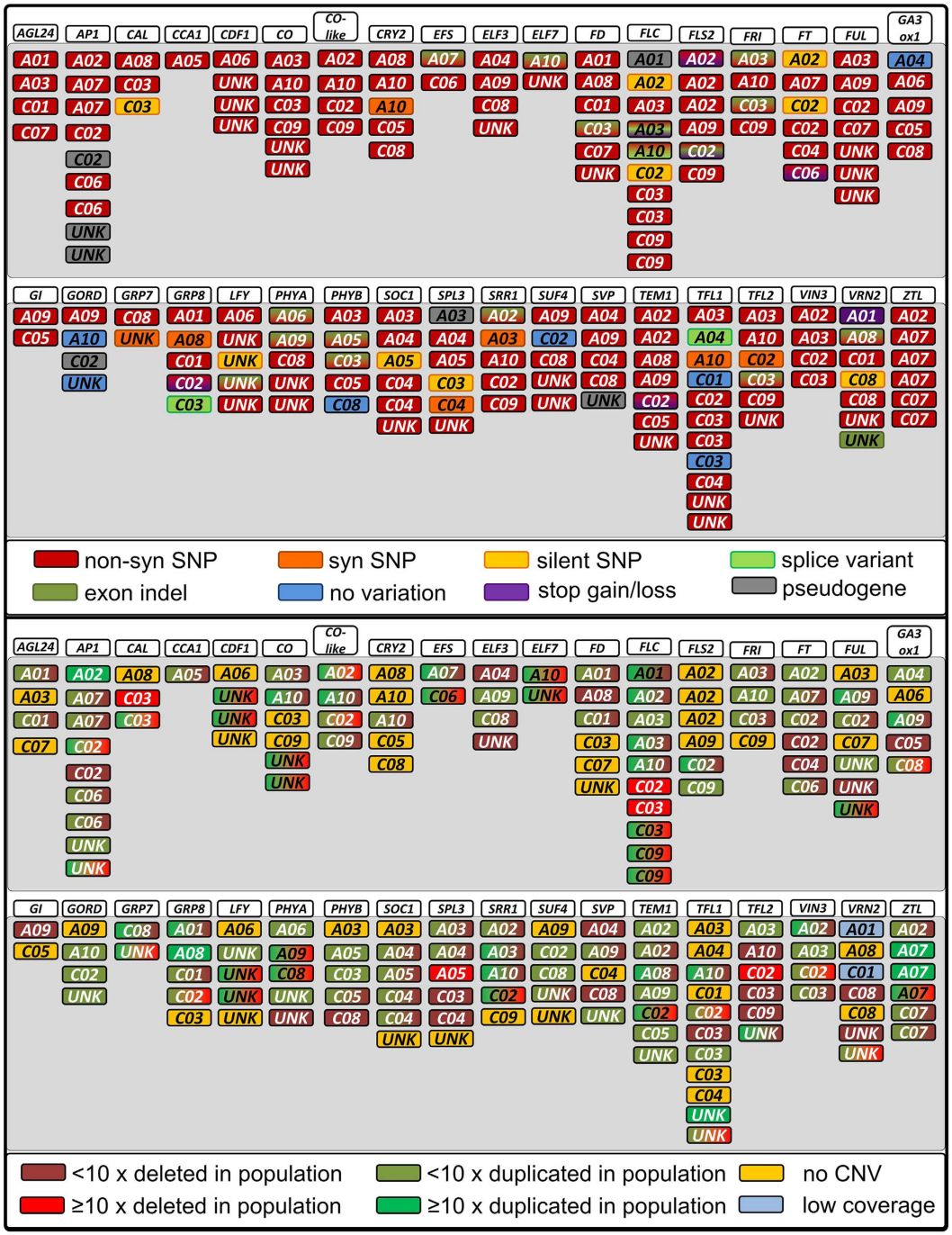


Figure 2. Distribution of SNPs/InDels (above) and CNV events (below) over all target gene copies. The chromosomal locations of the copies are given below the common Arabidopsis gene name (white background), with colors representing the respective type of sequence variation observed (see color code below each diagram). Upper panel: Silent SNPs are not indicated if synonymous or non-synonymous SNPs are present in the same copy, and synonymous SNPs are not indicated if non-synonymous SNPs are present in the same copy. Lower panel: Gene copies showing two different colors are deleted in some lines and duplicated in some others.

accessions were grown in autumn-sown trials, whereas spring-type and semi-winter accessions were grown in spring-sown trials. Plots were sown in a completely randomized block design with a harvest plot size of 3 × 1.25 m in a single replicate (containing around 200 plants).

In a separate experiment, a selection of 33 genotypes from the same set was grown in the greenhouse under semi-controlled conditions (20 °C). These genotypes were selected to represent spring, winter and swede material with different CNV patterns for *Bna.FLC*. Twenty seeds were sown in vermiculite, before being transplanted after one week into plates in soil, with 5 replicates per treatment. Four weeks after planting, these plants were either transferred to a climate chamber for vernalisation at 4 °C and short-day conditions for 6 weeks (mild

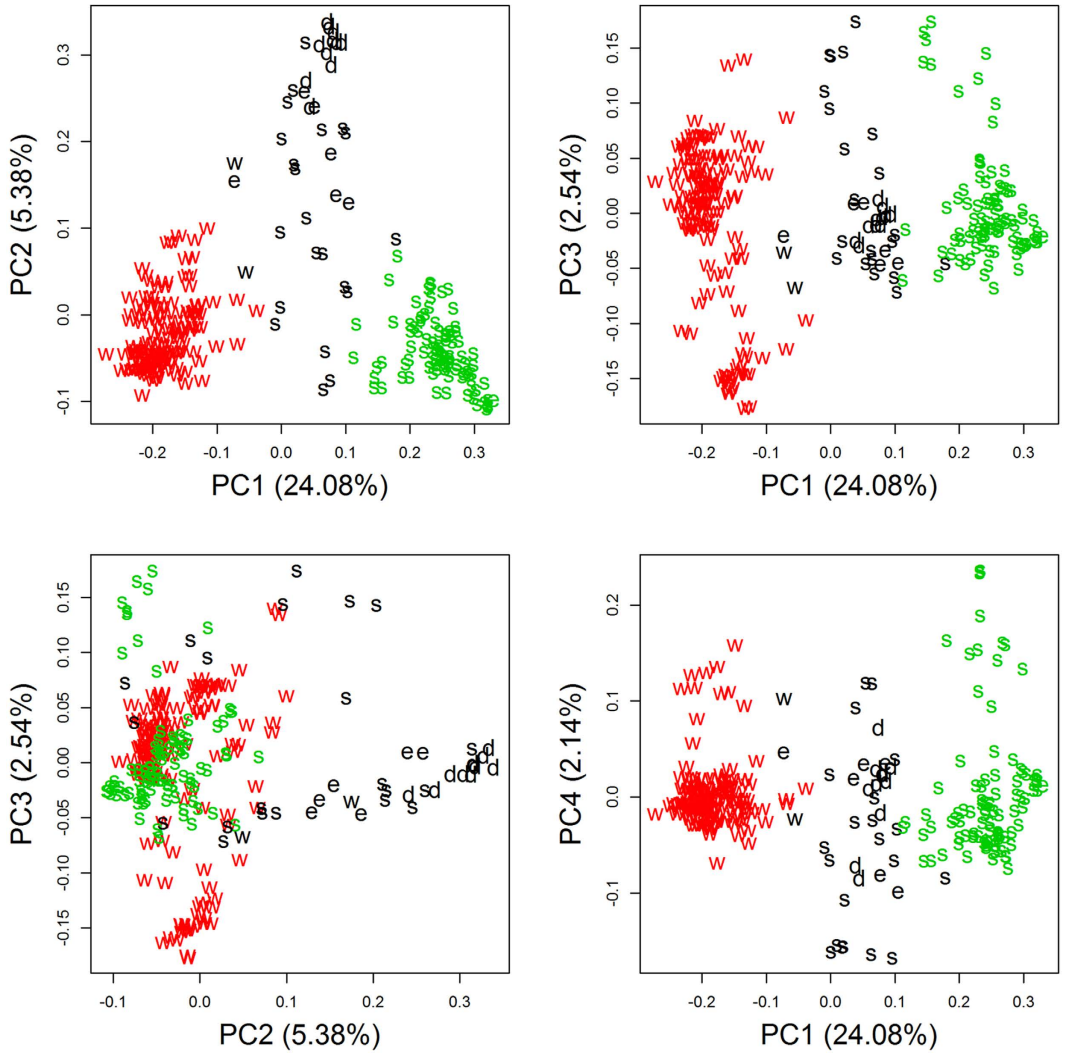


Figure 3. 2D plots of PCA for the total population. The explained variance is given in brackets. Colors indicate the cluster. Cluster 1 is shown in red, cluster 2 in green and cluster 3 in black. Letters indicate the morphotype: w for winter, s for spring, e for semi-winter and d for swedes.

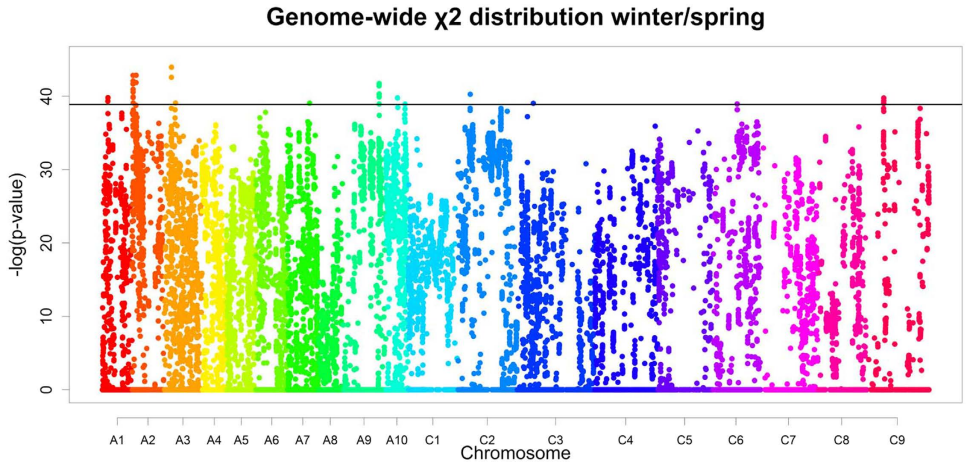


Figure 4. Genome-wide distribution of χ^2 p-values tested against equal distribution in winter and spring material. The chromosomes are coloured differently. The solid lines indicate the marker cut-off threshold of 0.1%.

Marker name	Chromosome	Position	winter allele		spring allele			located in gene	clear winter	clear spring	mixed	deletions
			winter pop	spring pop	winter pop	spring pop	−log(p)					
			split markers						ww	ss	ws	00
Bn-A01-p4641747	chrA01	4261277	132	13	4	101	39.3	*BnaA01g08820D*				
Bn-A01-p4641802	chrA01	4261332	132	13	4	101	39.3					
Bn-A01-p4803773	**chrA01**	**4413051**	**128**	**9**	**6**	**105**	**39.8**	*BnaA01g08820D*	136	102	12	3
Bn-A02-p3207085	**chrA02**	**695308**	**131**	**7**	**4**	**106**	**42.8**					
Bn-A02-p3208275	chrA02	700390	129	7	5	106	41.7					
Bn-A02-p3295898	chrA02	786193	130	8	5	101	39.9	*BnaA02g01700D*				
Bn-A02-p3297592	chrA02	787627	128	6	6	104	40.6					
Bn-A02-p3299206	chrA02	789246	130	6	5	105	42.1	*BnaA02g01710D*				
Bn-A02-p3300731	chrA02	790766	128	6	6	105	40.9	*BnaA02g01710D*	129	104	12	8
Bn-A02-p3302725	chrA02	792753	126	6	7	105	39.8					
Bn-A02-p3361391	chrA02	849106	126	6	8	104	39.1	*BnaA02g01860D*				
Bn-A02-p5907701	**chrA02**	**3096806**	**136**	**12**	**1**	**99**	**41.9**					
Bn-A02-p5917045	chrA02	3104382	137	11	2	101	42.9		148	101	2	2
Bn-A03-p6576575	chrA03	5874703	126	3	9	111	42.6	*BnaA03g12910D*	129	113	9	1
Bn-A03-p6636780	**chrA03**	**5928259**	**131**	**5**	**6**	**109**	**44.0**					
Bn-A03-p9836757	**chrA03**	**9057095**	**128**	**8**	**7**	**103**	**39.1**		136	110	na	7
Bn-A07-p15352802	**chrA07**	**17269795**	**126**	**7**	**9**	**106**	**39.0**	***BnaA07g22720D***				
Bn-A09-p30805314	chrA09	28557636	125	4	9	107	40.2	*BnaA09g40670D*	133	115	na	5
Bn-A09-p30805387	chrA09	28557709	125	4	9	110	41.4	*BnaA09g40670D*				
Bn-A09-p30887157	chrA09	28628531	131	11	5	103	39.9					
Bn-A09-p30909393	**chrA09**	**28655613**	**131**	**9**	**4**	**105**	**41.7**	***BnaA09g40920D***				
Bn-A09-p30918224	chrA09	28662308	131	13	4	101	38.9		129	101	19	4
Bn-A09-p30921980	chrA09	28664405	131	11	4	103	40.3	*BnaA09g40940D*				
Bn-A10-p7357442	**chrA10**	**9020292**	**128**	**7**	**9**	**106**	**39.8**	***BnaA10g10600D***				
Bn-A10-p7357555	chrA10	9020402	127	6	9	106	39.8	*BnaA10g10600D*	135	115	0	3
Bn-scaff_17109_2-p79906	**chrA10**	**14916811**	**121**	**1**	**17**	**111**	**38.9**		122	128	na	3
Bn-scaff_16002_1-p1767743	**chrC03**	**12604057**	**127**	**7**	**9**	**105**	**39.0**	***BnaA10g21860D***	134	114	na	5
Bn-scaff_18206_3-p62755	**chrC06**	**18959652**	**131**	**12**	**4**	**100**	**38.9**					
Bn-scaff_16912_1-p190291	chrC09	12697195	129	9	8	104	39.0		143	104	na	6
Bn-scaff_20836_1-p198809	chrC09	12804839	129	8	9	105	39.4					
Bn-scaff_20836_1-p198391	chrC09	12805246	129	8	9	105	39.4	*BnaC09g15770D*				
Bn-scaff_20836_1-p197940	**chrC09**	**12805697**	**130**	**8**	**9**	**105**	**39.8**		137	112	2	2
Bn-scaff_20836_1-p197387	chrC09	12806250	129	8	9	105	39.4					
Bn-scaff_20836_1-p196601	chrC09	12807036	129	8	9	104	39.0					
Regions from deep sequencing												
chrA02_3321143	chrA02	3321143	127	9	12	105	37.2	*Bna.SRR1.A02*				
chrA02_3862842	chrA02	3862842	120	3	19	111	37.1	*Bna.VIN3.A02*				
chrA03_5891342	chrA03	5891342	126	6	13	107	38.3	protein agamous-like 71				
chrA09_random_3749261	chrA09_random	3749261	126	12	13	102	34.4	*Bna.CCR1.A09_random*				
chrA10_14998726	chrA10	14998726	127	10	12	104	36.5	*Bna.FLC.A10*				
chrC02_random_990005	chrC02_random	990005	125	11	14	103	34.4	*Bna.FT.C02_random promoter*				

Table 1. Marker distributions of SNP markers associated with the winter-spring morphotype split in *B. napus*, along with the most closely associated markers from deep sequencing for the winter-spring split. The table shows the marker name, chromosomal position and the number of lines carrying either a winter or a spring allele in the respective winter-type and spring-type populations. The table also gives the −log(p-value) used to determine the split markers, along with the gene ID where the marker is located. If empty, the marker is non-genic. The markers with the highest −log(p-value) in each split region are shown in bold letters. The last four columns of the table show how many clear winter or spring haplotypes were counted, along with the number of mixed haplotypes and deletions. For regions only containing one marker, mixed haplotypes do not apply (na).

vernalisation) or 12 weeks (strong vernalisation), or kept in the greenhouse (no vernalisation). Begin of flowering (BBCH 61) was tracked daily for every single plant.

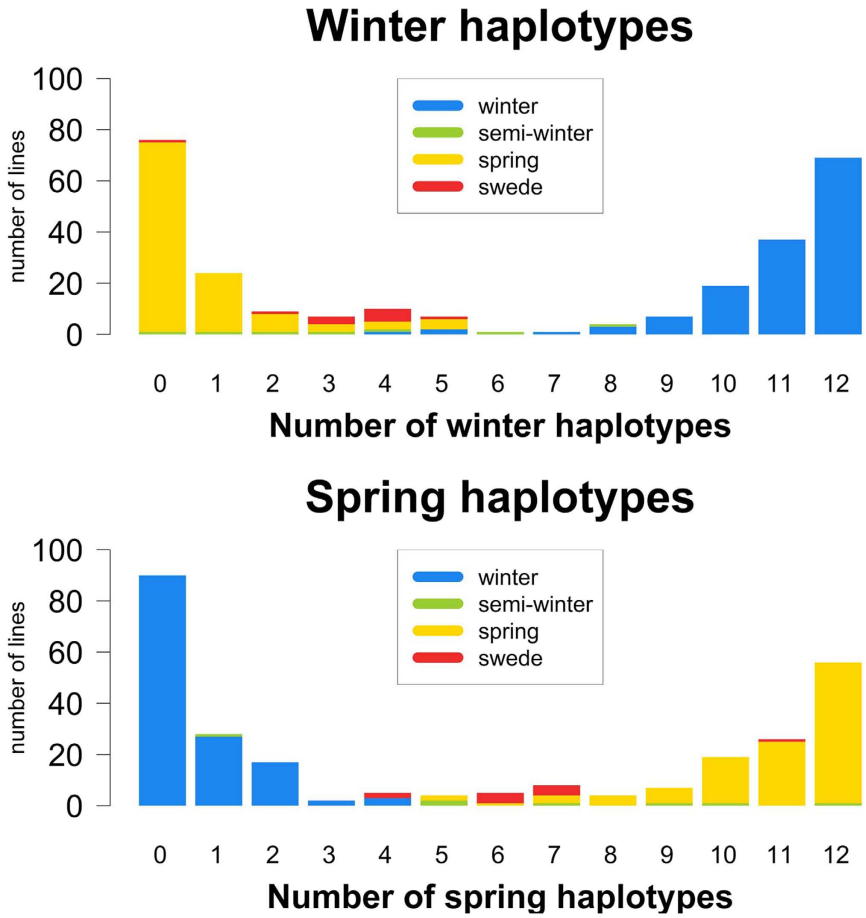


Figure 5. Distribution of clear winter haplotypes (above) and clear spring haplotypes (below) in the total population for all identified split regions. Mixed haplotypes were not counted. The distribution on morphotypes is colour-coded.

DNA isolation. Leaf material for genomic DNA extraction was harvested in spring 2012 from the field trial in Giessen, Germany. Pooled leaf samples were taken from at least 5 different plants per genotype, immediately shock-frozen in liquid nitrogen and kept at −20 °C until extraction. Leaf material was ground in liquid nitrogen with a mortar and pestle. DNA was extracted using a common CTAB protocol modified from Doyle and Doyle (1990) as described earlier[12]. DNA concentration was determined using a Qubit fluorometer and the Qubit dsDNA BR assay kit (Life Technologies, Darmstadt, Germany) according to the manufacturer's protocol. DNA quantity and purity was further checked on 0.5% agarose gel (3 V/cm, 0.5xTBE, 120 min).

Selection of target genes. As described previously[12], a set of 29 *A. thaliana* flowering time genes was selected to cover the entire genetic network controlling flowering time, including circadian clock regulators (*CYCLING DOF FACTOR 1* (*CDF1*), *EARLY FLOWERING 3* (*ELF3*), *GIGANTEA* (*GI*) and *ZEITLUPE* (*ZTL*)), the input pathways for vernalisation (*EARLY FLOWERING 7* (*ELF7*), *EARLY FLOWERING IN SHORT DAYS* (*EFS*), *FLOWERING LOCUS C* (*FLC*), *FRIGIDA* (*FRI*), *SHORT VEGTATIVE PHASE* (*SVP*), *SUPPRESSOR OF FRIGIDA 4* (*SUF4*), *TERMINAL FLOWER 2* (*TFL2*), *VERNALISATION 2* (*VRN2*), *VERNALISATION INSENSITIVE 3* (*VIN3*)), photoperiod sensitivity (*CONSTANS* (*CO*), *CRYPTOCHROME 2* (*CRY2*), *PHYTOCHROME A* (*PHYA*), *PHYTOCHROME B* (*PHYB*)) and gibberellin (*GIBBERELLIN-3-OXIDASE 1* (*GA3ox1*)), along with downstream signal transducers (*AGAMOUS-LIKE 24* (*AGL24*), *APETALA 1* (*AP1*), *CAULIFLOWER* (*CAL*), *FLOWERING LOCUS D* (*FD*), *FLOWERING LOCUS T* (*FT*), *FRUITFUL* (*FUL*), *LEAFY* (*LFY*), *SQUAMOSA PROMOTOR PROTEIN LIKE 3* (*SPL3*), *SUPPRESSOR OF CONSTANS 1* (*SOC1*), *TEMPRANILLO 1* (*TEM1*), *TERMINAL FLOWER 1* (*TFL1*)). On top, we also included *CIRCADIAN CLOCK ASSISTED 1* (*CCA1*), *FLAGELLIN-SENSITIVE 2* (*FLS2*), *GLYCIN-RICH PROTEIN 7* (*GRP7*), *GLYCIN-RICH PROTEIN 8* (*GRP8*), *GORDITA* (*GORD*) and *SENSITIVITY TO RED LIGHT REDUCED 1* (*SRR1*).

A full list of gene names and putative functions is provided in Supplementary Table 1.

Bait development. In order to perform target enrichment, complementary sequences of 120 nt length were first developed for each target region. A group of 120mer oligonucleotide sequences covering a certain target region is hereinafter referred to as a bait group for that target region, while collectively all bait groups are referred to as the bait group pool. In the present study the bait group pool for the sequence capture, developed mainly

Marker name	Chromosome	Position	non-swede allele		swede allele			located in gene	clear non-swede	clear swede	mixed	deletions
			non-swede pop	swede pop	non-swede pop	swede pop	−log(p)					
			split markers						nn	ss	sn	00
chrA03_4639027	**chrA03**	**4639027**	**260**	**0**	**0**	**11**	**57.7**	***Bna.VIN3.A03***	260	11	na	0
chrA04_12696607	**chrA04**	**12696607**	**0**	**8**	**260**	**3**	**55.8**		263	8	na	0
chrA06_5607262	chrA06	5607262	2	11	258	0	56.0	*Bna.CCR1.A06*	257	6	8	0
chrA06_5607744	chrA06	5607744	0	9	260	2	56.3	*Bna.CCR1.A06*				
chrA06_5608016	**chrA06**	**5608016**	**1**	**11**	**259**	**0**	**56.9**	***Bna.CCR1.A06***				
chrA06_5608089	**chrA06**	**5608089**	**1**	**11**	**259**	**0**	**56.9**	***Bna.CCR1.A06***				
chrA06_5614815	chrA06	5614815	0	8	260	3	55.8					
chrA08_14983629	**chrA08**	**14983629**	**260**	**3**	**0**	**8**	**55.8**	***Bna.TEM1.A08***	263	8	na	0
chrA09_11993194	**chrA09**	**11993194**	**0**	**8**	**260**	**3**	**55.8**	***BnaA09g19070D***	263	8	0	0
chrA09_11993662	**chrA09**	**11993662**	**0**	**8**	**260**	**3**	**55.8**					
chrA09_11995810	**chrA09**	**11995810**	**260**	**3**	**0**	**8**	**55.8**	***BnaA09g19070D***				
Bn-A09-p21922383	**chrA09**	**19312044**	**0**	**8**	**260**	**3**	**55.8**		263	8	na	0
chrA09_32435440	chrA09	32435440	2	11	258	0	56.0	*Bna.PHYA.A09*	258	9	4	0
chrA09_32435455	chrA09	32435455	2	11	258	0	56.0	*Bna.PHYA.A09*				
chrA09_32437048	**chrA09**	**32437048**	**0**	**10**	**260**	**1**	**56.9**	***Bna.PHYA.A09***				
chrA09_32441986	chrA09	32441986	260	2	0	9	56.3	*BnaA09g48430D*				
chrA10_17106726	**chrA10**	**17106726**	**0**	**8**	**260**	**3**	**55.8**	***Bna.ELF7.A10***	263	8	0	0
chrA10_17106744	**chrA10**	**17106744**	**0**	**8**	**260**	**3**	**55.8**	***Bna.ELF7.A10***				
chrA10_17108533	**chrA10**	**17108533**	**260**	**3**	**0**	**8**	**55.8**	***Bna.ELF7.A10***				
chrC01_1447013	**chrC01**	**1447013**	**0**	**8**	**260**	**3**	**55.8**		262	8	1	0
chrC01_1447235	**chrC01**	**1447235**	**0**	**9**	**260**	**2**	**56.3**	***Bna.FD.C01***				
chrC01_1447273	**chrC01**	**1447273**	**260**	**2**	**0**	**9**	**56.3**	***Bna.FD.C01***				
chrC01_1447516	**chrC01**	**1447516**	**0**	**9**	**260**	**2**	**56.3**	***Bna.FD.C01***				
chrC01_1447693	**chrC01**	**1447693**	**260**	**2**	**0**	**9**	**56.3**	***Bna.FD.C01***				
chrC01_1447972	**chrC01**	**1447972**	**0**	**9**	**260**	**2**	**56.3**	***Bna.FD.C01***				
Bn-scaff_16770_1-p1357882	**chrC08**	**24087909**	**0**	**8**	**260**	**3**	**55.8**		263	8	na	0
chrC08_36752901	**chrC08**	**36752901**	**260**	**1**	**0**	**10**	**56.9**	***BnaC08g42670D***	259	9	3	0
chrC08_36752954	**chrC08**	**36752954**	**0**	**10**	**260**	**1**	**56.9**	***BnaC08g42670D***				
chrC08_36753586	chrC08	36753586	1	10	259	1	56.1					
chrC08_36754224	chrC08	36754224	0	9	260	2	56.3	*BnaC08g42680D*				
chrC08_36755379	chrC08	36755379	1	10	259	1	56.1					
chrC08_36755562	chrC08	36755562	0	9	260	2	56.3					
Bn-scaff_16389_1-p12505	**chrC08**	**38113567**	**0**	**8**	**260**	**0**	**56.8**		260	8	na	3
chrC09_43739821	**chrC09**	**43739821**	**0**	**8**	**260**	**3**	**55.8**	***Bna.CO-li.C09***	263	8	na	0
Regions from deep sequencing												
chrA01_random_477115	chrA01_random	477115	5	8	255	3	51.6	mads-box protein				
chrA03_6053137	chrA03	6053137	8	9	252	2	49.6	*Bna.FRI.A03*				
chrA03_6243410	chrA03	6243410	12	8	248	2	46.2	*Bna.FLC.A03*				
chrA04_12695445	chrA04	12695445	3	8	257	3	53.3	*Bna.ELF3.A04*				
chrA05_5425314	chrA05	5425314	1	8	259	3	55.0	*Bna.SPL3.A05*				
chrA05_9211460	chrA05	9211460	4	9	256	2	52.9	ubiquitin-conjugating enzyme family protein				
chrA07_23775522	chrA07	23775522	3	8	257	3	53.3	cinnamoyl- reductase 2-2				
chrA10_1357187	chrA10	1357187	1	8	259	3	55.0	*Bna.CRY2.A10*				
chrA10_13359226	chrA10	13359226	1	9	259	2	55.4	*Bna.CO.A10*				
chrA10_14998679	chrA10	14998679	3	10	257	1	54.4	*Bna.FLC.A10*				
chrAnn_random_610372	chrAnn_random	610372	10	11	250	0	49.4	*Bna.TFL1.Ann.random*				
chrAnn_random_20504534	chrAnn_random	20504534	253	3	7	8	49.9	*Bna.VRN2.Ann.random*				
chrC03_8403949	chrC03	8403949	12	8	248	3	45.9	*Bna.FLC.C03*				
chrC03_random_5400150	chrC03_random	5400150	259	2	1	9	55.4	*Bna.FD.C03.random*				

Table 2. Marker distributions of SNP markers associated with the swede vs. non-swede morphotype split in *B. napus*, along with the most closely associated markers from deep sequencing for the swede vs. non-swede split. The table shows the marker name, chromosomal position and the number of lines carrying either a non-swede or a swede allele in the respective non-swede and swede populations. The table also gives the −log(p-value) which used to determine the split markers, along with the gene ID or the name of the gene where the marker is located. If empty, the marker is non-genic. The markers with the highest −log(p-value) in each split region are shown in bold letters. The last four columns of the table show how many clear non-swede or swede haplotypes were counted, along with the number of mixed haplotypes and deletions. For regions only containing one marker, mixed haplotypes do not apply (na).

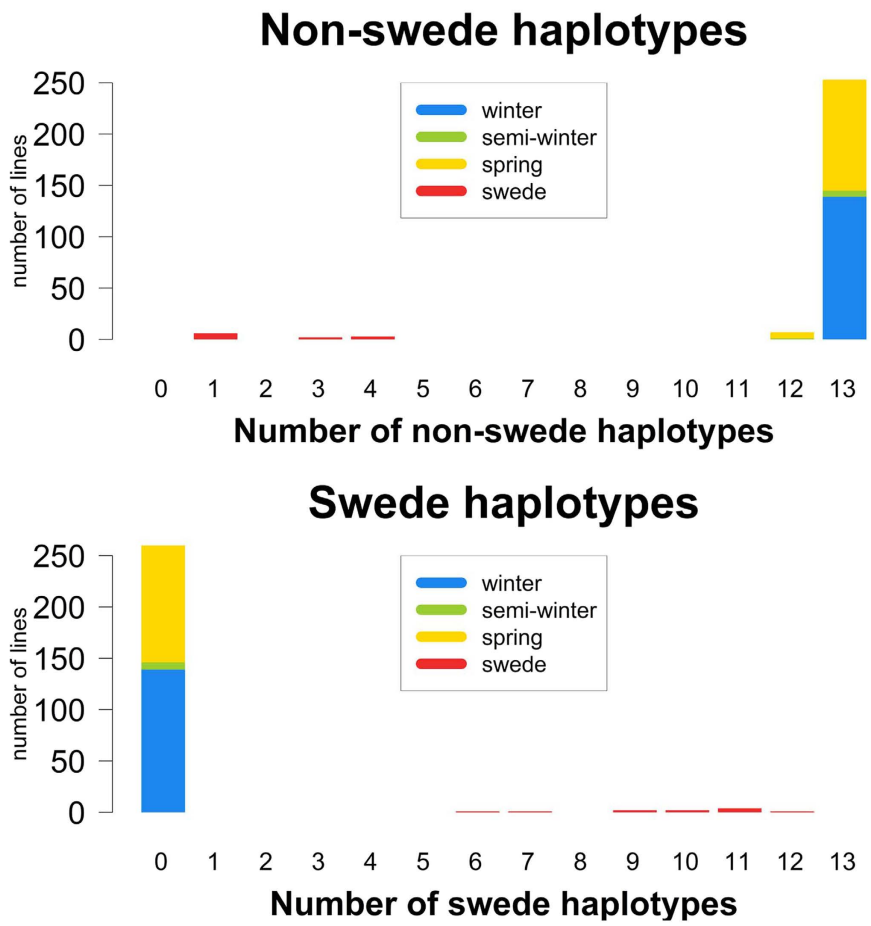


Figure 6. Distribution of clear non-swede haplotypes (above) and clear swede haplotypes (below) in the total population for all 13 identified split regions. Mixed haplotypes were not counted. The distribution on morphotypes is colour-coded.

using gene sequences from *B. rapa* or *B. oleracea*[12], was modified in order to improve specificity. Enriched regions captured in our previous study[12] were classified into target regions and non-target regions. The bait pool was then blasted against target and non-target regions with an E-value cut-off of 10^{-10}. Baits which had excessive non-target hits were manually removed. This was the case for bait groups on *FT*, *FUL* and *PHYA*. For some bait groups (*AP1*, *CO*, *SOC1*), too many baits (>30%) were deleted. In these cases, bait groups were created using a pre-publication draft (version 4.0) of the *B. napus* 'Darmor-Bzh' reference genome sequence assembly, which was kindly made available prior to public release by INRA, France, Unité de Recherche en Génomique Végétale[6], using the Agilent Genomic Workbench program SureDesign (Agilent Inc., Santa Clara, CA, USA). These replaced the corresponding bait groups developed previously using *B. rapa* or *B. oleracea*. Bait groups were created using the 'Bait Tiling' tool. The parameters were set as follows: Sequencing Technology: 'Illumina', Sequencing Protocol: 'Paired-End long Read (75 bp+)', 'Use Optimized Parameters (Bait length 120, Tiling Frequency 1x)', Avoid Overlap: '20', 'User defined genome', 'Avoid Standard Repeat Masked Regions'. Baits for genes on the minus-strand were developed in sense, while baits on the plus-strand were developed in antisense.

In total, 63 bait groups were created for *B. rapa* copies of the target genes, 71 bait groups for *B. oleracea* copies and 24 bait groups for *B. napus* copies.

Sequence capture and sequencing. Custom bait production was carried out by Agilent Technologies (Agilent Inc., Santa Clara, CA, USA) using the output oligonucleotide sequences from eArrayXD. Sequence capture was performed using the SureSelectXT 1 kb–499 kb Custom Kit (Agilent Inc., Santa Clara, CA, USA) according to the manufacturer's instructions. The resulting TruSeq DNA library (Illumina Inc., San Diego, CA, USA) was sequenced on an Illumina HiSeq 2500 sequencer at the Max Planck Institute for Breeding Research (Cologne, Germany) in 100 bp single-read mode.

Sequence data analysis. Quality control of the raw sequencing data was performed using FASTQC. Reads were mapped onto version 4.1 of the *B. napus* 'Darmor-Bzh' reference genome sequence assembly[6]. Mapping was performed using the SOAPaligner algorithm[35], with default settings and the option r = 0 to extract uniquely aligned reads. Removal of duplicates, sorting and indexing was carried out with *samtools* version 0.1.19[36]. Alignments were visualised using the IGV browser version 2.3.12[37]. Enriched regions and coverage

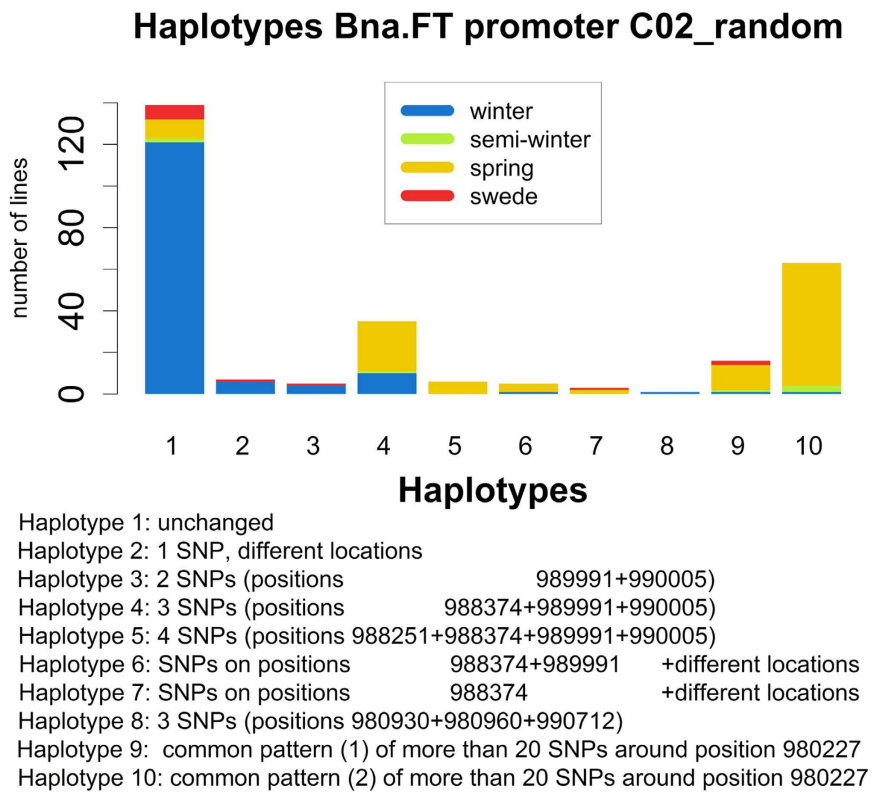


Figure 7. Haplotype distribution for the Bna.FT promoter on C02_random. The distribution on morphotypes is colour-coded. The haplotype patterns are provided in the text.

differences were calculated using the *bedtools* software with multiBamCov[38]. Calling of single nucleotide polymorphisms (SNPs) was performed with the algorithm mpileup in the *samtools* toolkit. SNP and InDel annotation was performed using CooVar[39]. Target regions were defined using the gene annotation list from the *B. napus* 'Darmor-bzh' v4.1 reference genome[6] and BLAST position results of the bait pool (E-value cut-off 10^{-100}) on the mapping reference, and used to calculate the fraction of target covered. For InDel calling, a separate mapping using Bowtie2[40] was performed, as described previously[41]. Removal of duplicates, sorting and indexing was carried out with samtools version 0.1.19. An initial InDel calling was performed using samtools mpileup, and realignment of reads around InDels was performed using GATK RealignerTargetCreator, version 3.1.1[42]. A final InDel calling was then performed as described above. InDels were filtered for a minimum mapping quality of 30 and a read depth of 10 or more using vcftools[43].

Read coverage for each captured region was normalised as follows: normalised coverage = (number of reads per region*total length of genome)/(total number of aligned reads per genotype*average read length). Copy number variation (CNV) in a given target region was assumed if the ratio of normalised coverage(genotype)/normalised coverage(all genotypes) was smaller than 0.5 or higher than 1.5, respectively.

Sequencing data for 3 genotypes from a former experiment (Silona, Campino, Magres Pajberg)[12] were analysed separately with the same pipeline to allow inclusion in the marker distribution analysis.

SNP genotyping and pre-processing. The 283 accessions were genotyped using the Brassica 60 K Illumina® Infinium SNP array by TraitGenetics GmbH (Gatersleben, Germany). We used the SNP positions as published in[44]. Heterozygous calls were treated as missing values. Moreover, we used the deep sequencing data to include all confidently called SNPs in biallelic state which lay in the analysed regions. Confidently called InDels were included by coding reference alleles as AA, insertions as CC, deletions as TT and heterozygous calls as missing values. The SNP matrix from the SNP array and the SNP and InDel data from deep sequencing were combined to one single marker file and sorted by position. The subsequent marker set contained 43733 markers. After pre-processing the marker set for non-missing marker values >0.9, minor allele frequency >0.01 and individuals (genotypes) with non-missing individual markers >0.8, we retained 33944 unique SNP markers and a population of 271 individuals for marker distribution analysis. Data pre-processing was performed with R (version 3.1.0) using the package GenABEL[45].

Population structure. Population structure analysis and visualization were performed in R (version 3.1.0) using the package SelectionTools (http://fb09-pg-s207.agrar.uni-giessen.de/~frisch-m/), which applies principal component analysis based on genetic distances calculated according to the euclidean modified Rodger's distance method. The most likely number of population subclusters was determined to be 3 by plotting the within-cluster

Gene ID	Chromosome	start	stop	deletions		duplications		mean coverage	*Gene name*
				nonswede population	swede population	nonswede population	swede population		
BnaA08g15780D	chrA08	13097823	13098361	11	9	0	0	1641.5	*no annotation*
BnaA09g48410D	chrA09	32434233	32438771	31	0	10	8	1136.2	*Bna.PHYA.chrA09*
BnaA09g57140D	chrA09_random	4043861	4045464	1	0	4	8	1675.6	*Bna.GA3ox.chrA09.random*
BnaA10g22080D	chrA10	14998617	15003197	2	0	2	9	1321.5	*Bna.FLC.chrA10*
BnaC08g38580D	chrC08	34776298	34779240	9	8	1	0	1226.1	*Bna.CCR1.chrC08*
BnaC08g38810D	chrC08	34907098	34908735	4	9	0	0	1581.1	*Bna.GA3ox.chrC08*
BnaC08g42660D	chrC08	36746642	36751390	6	10	15	0	1537.7	*Bna.PHYA.chrC08*
BnaC08g42670D	chrC08	36752307	36753108	7	9	14	0	1660.2	*germin like protein*
BnaC09g46500D	chrC09	46345350	46350092	2	9	13	0	1096.9	*Bna.FLC.chrC09*
BnaC09g46540D	chrC09	46366645	46371180	3	9	11	0	1031.5	*Bna.FLC.chrC09*

Table 3. Distribution of deletion and duplication events in the non-swede and swede populations. The table shows the gene ID, chromosomal position and the number of lines in the respective non-swede and swede populations which carry either a deletion or duplication. The table also gives the mean coverage of the respective gene and the common gene name.

sum of squares against the possible number of clusters, ranging from 1 to 15. K-means clustering was then performed in R using SelectionTools.

Marker distribution analysis. For every marker, we counted the allele frequency of the alternative allele in each morphotype pool. The ratio between the frequency of the allele in the winter pool (winter + swedes) and the spring pool (semi-winter + spring) was used to assign the allele as a winter or spring allele. If the ratio was <1, the alternative allele was denoted spring (s), if it was >1, the alternative allele was denoted winter (w). We then first tested if the marker would be suitable to explain a morphotype split, by comparing the observed distribution of w alleles in the winter pool (without swedes) and s alleles in the spring pool (without semi-winter) with the expected distribution (139/114), using a χ^2 test. Only markers which did not show significant deviation from this distribution (p-value > 0.1) were considered in the next step. In the next step, we tested the distribution against random distribution between the pools, by comparing the observed distribution of w alleles in winter/s alleles in spring/s alleles in winter/w alleles in spring against the expected random distribution of 69.5/57/69.5/57. We then considered the top 0.1% of $-\log$(p-value) as split markers. The same was done for the swede and non-swede material.

Results

Deep sequencing and variant calling. We defined regions as genetic regions which were covered with a mean coverage in the population of at least 10. In total, we analyzed 1184 regions, of which 637 regions were annotated as genes. Of these, 184 corresponded to the intended target genes. Two target genes copies for *VERNALISATION INSENSITIVE 3* (*VIN3*) (*Bna.VIN3.A01* and *Bna.VIN3.C01*) had insufficient coverage for this analysis and were not considered. Among the non-genic regions, we found 33 regions giving a BLAST hit to the *FLOWERING LOCUS T* (*FT*) promoter. A further 12 regions were identified as pseudogenes of the target genes. Those regions which were assigned to one of those classes (target genes, target pseudogenes and FT promoter) were summarized as target regions. A gene group is defined as all copies of a specific gene.

We called and annotated 13053 SNPs, of which 4806 were located in the target regions. InDel calling revealed a total of 1894 InDels, with 506 in the target regions. Only 25 InDels were frameshifts, amino acid insertions or splice variants. All gene groups showed potentially functional variation, i.e. at least one copy of the gene group carried either a non-synonymous SNP, stop codon mutation, amino acid insertion, splice variant or frameshift InDel. Altogether, only 7 copies were completely conserved, while 16 copies carried only silent or synonymous variation. Interestingly, no functional variation was observed in two copies of *Bna.FLOWERING LOCUS C* (*FLC*) (on chromosomes A02/C02) and two copies of *Bna.FT* (also on chromosomes A02/C02), respectively. On the other hand, other copies of *Bna.FLC* (A03, A10) and *Bna.FT* (C06) carried a surprisingly large range of variation. Among the genes with frameshift variants were copies of *Bna.FRIGIDA* (*FRI*), *Bna.PHYTOCHOME A* (*PHYA*), *Bna.EARLY FLOWERING IN SHORT DAYS* (*EFS*), *Bna.EARLY FLOWERING 7* (*ELF7*), *Bna.PHYTOCHROM B* (*PHYB*), *Bna.VERNALISATION 2* (*VRN2*) and *Bna.LEAFY* (*LFY*) (Fig. 2). We also calculated copy number variation (CNV) based on read depth. No gene group was found without CNV, and only two lines were found which did not carry any CNV among the target copies. The distribution of SNPs, InDels and CNVs is shown in Fig. 2.

Population structure. Among the analysed population of 271 accessions, we had 139 winter type accessions, 7 semi-winter type accessions, 114 spring type accessions and 11 swedes. Analyzing this population with a Principal Component Analysis (PCA) showed a strong population substructure, as the first principal component explained 24.1% of the variation, while further components explained 5.4, 2.5 and 2.1%, respectively (Fig. 3). The population falls into three main clusters: the first cluster contained 137 winter type accessions, the second one 93 spring type accessions and a semi-winter type accession, and the third and most diverse cluster contained 11 swedes, 6 semi-winter type accessions, 21 spring type accessions and 2 winter type accessions. We concluded that the winter material was genetically least diverse, while spring material was more diverse,

Chromosome	Marker with highest p-value	Candidate	Distance from closest split marker [kbp]
chrA01	Bn-A01-p4803773	dna topoisomerase 2	502.7
		pseudo-response regulator 2	*57.0*
		agamous-like protein 1	386.3
		probable lysine-specific demethylase jmj14-like	877.5
		small rna 2 -o-methyltransferase	990.8
chrA02	Bn-A02-p3207085	flowering locus c	557.1
		transcription factor hy5	342.1
		embryonic flower 1	281.6
		nuclear transcription factor y subunit a-1	127.5
		flowering time control protein fy	*8.1*
		zinc finger protein knuckles	44.4
		dna (cytosine-5)-methyltransferase drm2	121.7
chrA02	Bn-A02-p5917045	histone deacetylase	403.8
		auxin response factor 4	81.4
		e3 sumo-protein ligase siz1	53.5
		ap2-like ethylene-responsive transcription factor toe2	*1.9*
		pseudo-response regulator 3	8.2
		sensitivity to red light reduced protein	215.9
		multicopy suppressor of ira1	580.9
		vernalization insensitive 3	757.8
chrA03	Bn-A03-p6636780	protein agamous-like 71	19.3
		protein agamous-like 42	22.4
		protein phosphatase	8.2
		polycomb group protein embryonic flower 2	70.5
		frigida	124.9
		flowering locus c	311.8
chrA03	Bn-A03-p9836757	histone h2a	539.8
		chromatin structure-remodeling complex protein syd	445.8
		protein early flowering 4-like	164.3
		btb poz domain-containing protein	48.3
chrA07	Bn-A07-p15352802	*two-component response regulator arr5*	*352.9*
		sin3 histone deacetylase complex	654.6
		protein argonaute 7-like	894.1
		floral homeotic protein apetala 1	964.5
chrA09	Bn-A09-p30909393	*protein early flowering 3-like*	17.9
		e3 ubiquitin-protein ligase orthrus 2	67.9
		dna methyltransferase	727.3
chrA10	Bn-A10-p7357555	topoisomerase i	651.5
		swinger	285.1
		histone acetyltransferase type b catalytic subunit-like	187.2
chrA10	Bn-scaff_17109_2-p79906	transcription factor hy5	261.7
		flowering-promoting factor 1-like	99.2
		flowering locus c	81.8
chrC03	Bn-scaff_16002_1-p1767743	*btb poz domain-containing protein*	67.5
		two-component response regulator arr16	*12.8*
		phy rapidly regulated 1	743.0
chrC06	Bn-scaff_18206_3-p62755	histone z	939.5
		squamosa-promoter binding protein	*213.7*
		shatterproof1	749.7
chrC09	Bn-scaff_20836_1-p197940	*set domain isoform 1*	845.4

Table 4. Selected candidate genes for all 12 regions associated with the split between winter-type and spring-type *B. napus* accessions. The table lists the chromosome and the marker with the most significant deviation from the expected random distribution, together with candidate genes selected based on gene ontology and literature. The last column specifies the distance of the gene to the closest marker within the split-associated region. The candidate gene closest to the split region is shown in italics.

Chromosome	Marker with highest p-value	Candidate	Distance from closest split marker [kbp]
chrA03	chrA03_4639027	*vernalization insensitive 3*	*0.0*
		histone acetyltransferase type b catalytic subunit	163.4
chrA04	chrA04_12696607	protein ovule abortion 4	45.0
		protein early flowering 3-like	*0.1*
		terminal flowering 1 protein 1	413.7
chrA06	chrA06_5608089	gibberellin 3-oxidase	181.8
		cinnamoyl- reductase	*0.0*
		transcriptional factor b3 family protein	248.7
		histone acetyltransferase hac12	263.1
		protein elf4-like 4	473.9
chrA08	chrA08_14983629	*ap2-erebp rave subfamily protein rav2*	*0.0*
		cycling dof factor 2	170.7
		cullin 3	187.3
chrA09	chrA09_11993662	cullin 4	982.5
		della protein	*348.6*
		agamous-like mads-box protein agl3	367.7
chrA09	Bn-A09-p21922383	mads-box protein gordita	875.9
		protein suppressor of fri 4	759.6
		nuclear transcription factor y subunit a-7	*151.7*
chrA09	chrA09_32437048	*phytochrome a*	*0.0*
		phytochrome interacting factor 3	11.9
		histone-lysine n-methyltransferase atx2	810.9
chrA10	chrA10_17106744	probable lysine-specific demethylase elf6-like	457.6
		protein lhy cca1-like 1	51.2
		pseudo-response regulator 7	41.0
		protein early flowering 7	*0.0*
chrC01	chrC01_1447516	*bzip transcription factor*	*0.0*
chrC08	Bn-scaff_16770_1-p1357882	*dek domain-containing chromatin associated protein*	*942.4*
chrC08	chrC08_36752954	*phytochrome a*	*1.5*
		phytochrome interacting factor 3	10.5
		medea	551.6
chrC08	Bn-scaff_16389_1-p12505	histone-lysine n-methyltransferase atx2	434.0
		dna helicase	405.8
		tata-box-binding protein 2	*107.5*
chrC09	chrC09_43739821	chromo domain-containing protein lhp1-like	985.9
		della protein	866.0
		col1 protein	*0.0*
		coa	5.9
		sepallata2	47.5

Table 5. Selected candidate genes for all defined regions associated with the split between swede and non-swede *B. napus* morphotypes. The table lists the chromosome and the marker with the most significant deviation from the expected random distribution, together with candidate genes selected based on gene ontology and literature. The last column specifies the distance of the gene to the closest marker within the split-associated region. A value of "0.0" indicates that the marker lies within the gene. The candidate gene closest to the split region is shown in italics.

followed by semi-winter and swede material. Overlap between winter and spring pools is minimal, while all other types show more overlap, although swedes are more distant from the winter and spring core clusters.

Marker distribution analysis. To analyse marker distribution on a genome-wide scale, we used SNP data from the Brassica 60 K Illumina® Infinium SNP array and combined it with data from deep sequencing (SNPs and InDels). In order to find the most indicative marker patterns for the differential flowering behaviour of winter and spring material, we analyzed the differential marker pattern between the different morphotypes using the χ^2 test. We first defined "winter" and "spring" alleles by allele frequencies in the different morphotypes and assessed their distribution in both pools. First, we excluded all markers with non-suitable allele frequencies. We regard all markers as non-suitable if their minor allele frequency was too low to explain a population split. This was tested in a foregoing χ^2 test (see Methods). The remaining markers were tested against random distribution in the

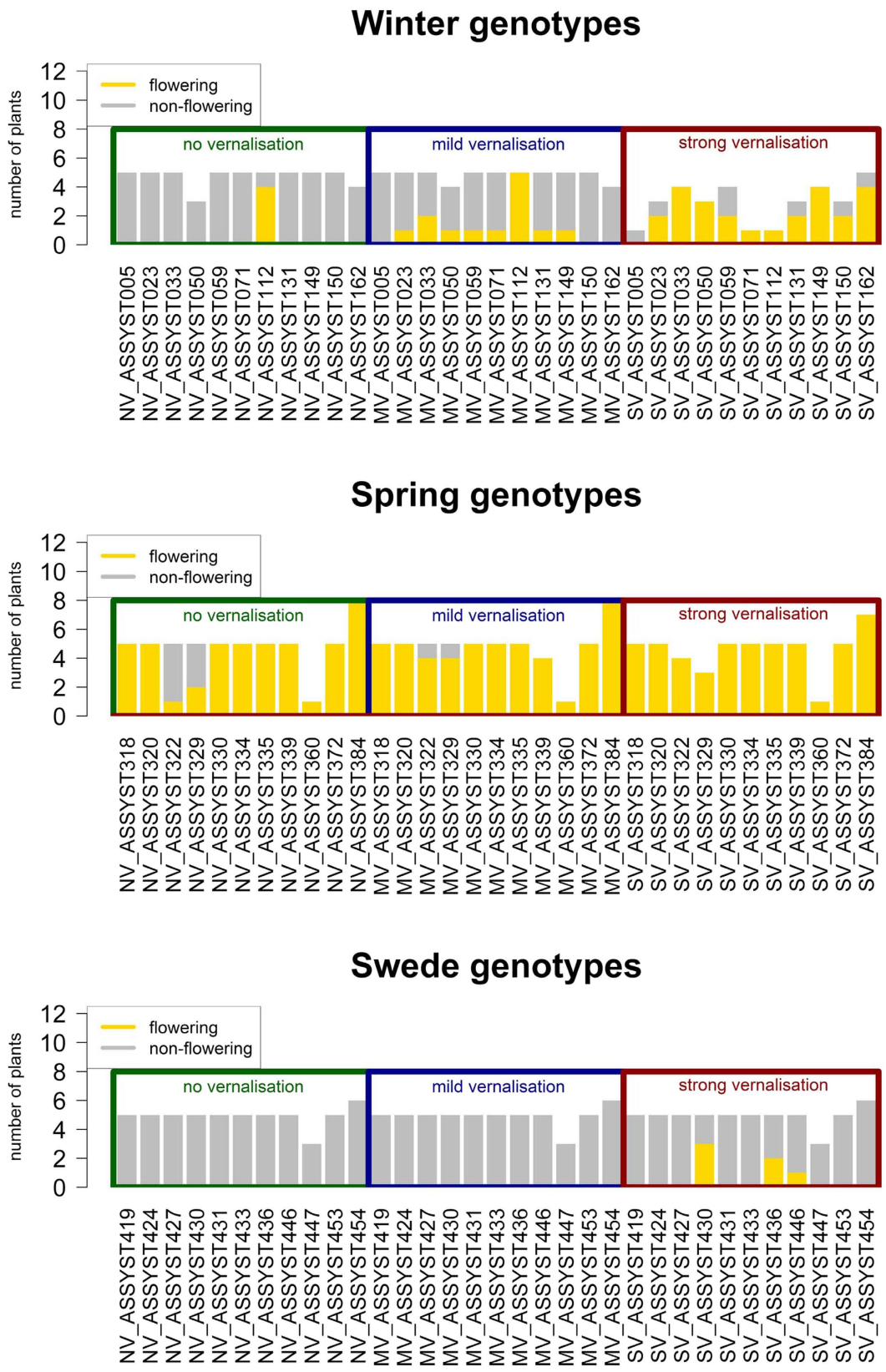


Figure 8. Distribution of flowering plants in the vernalisation trial. The height of the bars represents the number of replications which could be phenotyped. Yellow indicates flowering plants, grey indicates non-flowering plants. The different treatments are framed in different colors, with green indicating no vernalisation, blue mild vernalisation (6 weeks) and red strong vernalisation (12 weeks).

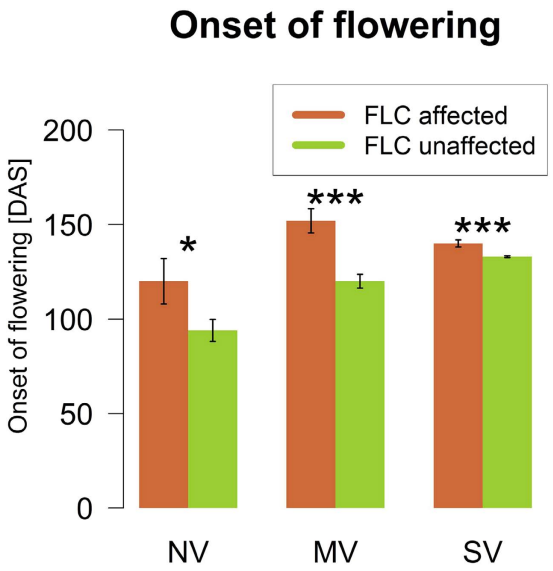


Figure 9. Barplots of flowering time recorded in days after sowing (DAS) for FLC-affected and non-affected plants among the spring genotypes in the vernalisation trial. Whiskers show standard errors. The asterisks denote the level of significance for Student's t-test (*p-value < 0.05, ***p-value > 0.001). NV: No vernalisation, MV: mild vernalisation (6 weeks), SV: strong vernalisation (12 weeks).

respective morphotypes. The same was done for swedes against non-swede accessions. Choosing a cut-off which considers the top 0.1% of markers (-log[p-value] = 38.9 for winter/spring and 55.8 for swedes), we detected 12 regions on chromosomes A01, A02, A03, A07, A09, A10, C03, C06 and C09 for the winter-spring split (Fig. 4), and 13 regions on chromosomes A03, A04, A06, A09, C01, C08 and C09 for the swede split.

Analysis of split regions. We subsequently counted how many of the 12 winter-spring split regions have a clear winter or spring pattern in each genotype, i.e. the number of cases where every split marker in the haplotype corresponded to the winter or spring state (Table 1). The distribution of lines carrying clear winter and spring haplotypes is shown in Fig. 5. Mixed haplotypes were excluded here, as they account for less than 5% of the haplotypes. From this distribution, we concluded that characterizing these regions for their haplotype pattern is sufficient to distinguish winter from spring morphotypes, but not to distinguish semi-winter or swede morphotypes. The same analysis on 13 split regions identified for the swede vs. non-swede split revealed a more explicit distribution (Table, 2 and Fig. 6). Genotyping these loci is therefore sufficient to distinguish swede morphotypes from non-swedes.

In order to exclude candidates for the respective morphotype split, we specifically looked at the variant distributions from deep sequencing in our marker set. Because these derived from sequence data, a poorly fitting distribution excludes the sequence from being a major cause for this morphotype, as sequencing covers the total variation of a gene. This is not the case for data from the SNP array, as even genic SNPs are not always completely predictive for their neighbor SNP. For the winter-spring split, only 6 sequenced regions with a distribution comparable to the detected split markers could be found, among them *Bna.VIN3.A02*, *Bna.FLC.A10* and the *Bna.FT* promoter on the non-assembled scaffolds of C02_random (Table 1). With the exception of *Bna.FLC.A10* (R10P mutation), all those SNPs are either synonymous or located in an intron. For the non-swede vs. swede split, we found 17 sequenced regions carrying variants with an acceptable distribution, for example three copies of *Bna.FLC*, two copies of *Bna.CO* and a further copy of *Bna.FLOWERING LOCUS D* (*FD*) (Table 2).

Upstream of the gene *Bna.FT* on C02_random we found two regions, spanning 4622 and 4904 bp, respectively, which retrieved BLAST hits to the A02 or C02 copies of the *Bna.FT* promoter listed by NCBI. We therefore identified these sequences as the promoter of *Bna.FT* on C02_random. Both sequences contain a CArG box core motif, whereby the first sequence also contains 3 additional FLC binding sites known from *A. thaliana*[46] and the second sequence contains 2 such FLC binding sites. No SNP is located in those motifs. We found that most (143 of 145) winter types are unchanged in both sequences or carry only minor changes, whereas most (71 of 116) of the spring population carried one of two distinctive haplotype patterns involving a SNP at position C02_random:980227 (Fig. 7). These patterns were shared by only two putative winter-type accessions, one of which is an exotic accession that may not need vernalisation, whereas the other is an accession which the vernalisation experiments revealed to have vernalisation-independent flowering (see below).

We furthermore compared the numbers of deletion and duplication events in the different morphotype pools. For the winter-spring split, we found no specific pattern for the total population. In contrast, we found several patterns of deletions and duplications which were almost exclusive to swedes, concerning split regions on A08, A09, A10, C08 and C09 (Table 3). The regions on A09/C08 (containing copies of *Bna.PHYA* and *Bna.GIBBERELLIN 3 OXIDASE 1* (*GA3ox1*) and A10/C09 (containing copies of *Bna.FLC*) are homeologous to each other. Some of these regions, particularly those on C08, appear to involve larger homoeologous exchanges that probably affect not only the detected genes.

For the winter-spring split, we found 234 genes with flowering-related gene ontology terms within 1 Mb of one of the diagnostic split markers. Examples are shown in Table 4. In the 13 regions showing a split between swedes and non-swedes, we found 260 candidate genes within 1 Mb. In this analysis, several split markers lay directly in candidate genes covered by deep sequencing, for example *Bna.VIN3*, *Bna.PHYA* (2 copies), *Bna.TEMPRANILLO1* (*TEM1*), *Bna.ELF7*, *Bna.CONSTANS* (*CO*), *Bna.CO-like* or *Bna.CINNAMOYL COA REDUCTASE 1* (*CCR1*) (Table 5). Some of those markers were non-synonymous SNPs (in *Bna.CCR1*, *Bna.TEM1*, *Bna.CO*, *Bna.CO-like*), whereas others were either synonymous or located in introns or untranslated regions (UTRs).

Vernalisation trials. In order to test if the swede-specific pattern of *Bna.FLC* deletions and duplications would affect vernalisation dependency, we conducted a vernalisation trial with a reduced set of lines (11 lines each from the winter, spring and swede panels). These were selected to represent either lines without CNV in *Bna.FLC* (as a control), or lines which have alternative patterns of deletion and duplication in *Bna.FLC*. The plants were subjected to either 6 or 12 weeks of vernalisation, or were not vernalised. We then scored the time until opening of the first flower. We found that all spring lines and one winter line were vernalisation independent. The vernalisation-independent winter line was one of the two genotypes carrying a strongly divergent *Bna.FT* promoter on C02_random. All swedes and three winter lines were found to be strongly vernalisation dependent, meaning that no plants flowered after mild vernalisation. At the end of the experiment, one winter line and 8 swede lines did not flower at all, meaning that 12 weeks of vernalisation were not sufficient to induce flowering (Fig. 8).

For the spring types, we found that lines with altered *Bna.FLC* patterns flower significantly later than lines without such changes (Fig. 9). All the same, spring lines carrying a swede pattern in *Bna.FLC* were not vernalisation dependent, indicating that this pattern is not sufficient to induce vernalisation.

Discussion

Our study aimed at identifying genetic variants which are responsible for the separation of the different morphotype pools in *B. napus*. According to population structure analyses performed in this and other studies[33,34,47], winter-type *B. napus* accessions tend to separate almost completely from other accessions, while some spring types along with semi-winter and swede material are more diverse. Here, we defined a total of 12 variant haplotypes which are diagnostic for the winter-spring split, and 13 variant haplotypes for the swede-non-swede split. Moreover, we found one winter type without vernalisation requirement that was nevertheless winter-hardy, and spring types with some degree of vernalisation responsiveness. Swedes were found to be extremely vernalisation dependent, with some variation among the accessions, presumably because swede forms have been bred to maintain their vegetative state for as long as possible. Vernalisation is a quantitative process[48], so there is natural variation in responsive temperature range and vernalisation duration[21,49–51]. Markers for such life cycle traits are extremely important in order to introgress desirable traits between ecogeographical or morphotype gene pools, for example seed quality traits from spring to winter oilseed forms[52] or resistance traits from swede to non-swede material[53].

An R10P mutation in the MADS box domain of *Bna.FLC.A10* was revealed as one candidate for the winter-spring split in *B. napus*, however our data shows that this mutation is neither the only candidate nor the best one. Neither the other *Bna.FLC* homologues, nor the detected copies of *Bna.FRI*, showed an appropriate variant distribution to explain the winter-spring split. Similar results were found for natural variation in flowering time for *A. thaliana*[54,55]. This excludes the possibility that genetic variation within *Bna.FLC* gene sequences, besides *Bna.FLC.A10*, are causal for vernalisation requirement, thus indicating that other *Bna.FLC* copies are either not responsible for vernalisation or there is variation in *cis*-regulatory elements. As shown by the vernalisation trials, spring-type plants without CNV in *Bna.FLC* copies flower earlier, indicating that *Bna.FLC* still plays a role in modulating flowering time in the absence of vernalisation requirement. Spring genotypes with a high *Bna.FLC* copy number showed accelerated flowering under vernalisation, indicating that they established weak vernalisation responsiveness. *FLC* is known to bind many other genes in *A. thaliana*[46], and it also regulates other developmental processes like germination[56], hence additional copies might be assumed to underlie strong selection. The differential degree of conservation between the copies suggests sub-functionalisation, whereby conserved *Bna.FLC.A02* and *Bna.FLC.C02* presumably retain more general roles, whereas *Bna.FLC.A10* might be more specialized towards flowering regulation. Sub-functionalisation events are characteristic for the evolution of MADS box transcription factors[57,58].

On the other hand, *Bna.SRR1.A02*, *Bna.VIN3.A02*, *Bna.AGL71.A03*, *Bna.CCR1.A09_random* and the *Bna.FT* promoter on C02_random represent further candidates for the morphotype split. *SRR1* is a clock-associated gene found to regulate *CO*, *FT* and *CYCLING DOF FACTOR 1* (*CDF1*) in *A. thaliana*[59]. Moreover, *A. thaliana srr1* mutants have reduced levels of *FLC* and respond only weakly to vernalisation[59]. Similarly, *Bna.VIN3.A02* represents a copy of another vernalisation candidate upstream of *Bna.FLC*[54]. In Arabidopsis, *VIN3* is expressed during cold and associates to the PRC2 complex to downregulate *FLC* gene activity[22,60]. *AGAMOUS-LIKE 71* (*AGL71*) is closely related to the flowering integrator *SUPPRESSOR OF CONSTANS 1* (*SOC1*) and seems to be involved in gibberellin-dependent flowering pathways[61]. Its promoter contains a CArG box for FLC binding in *A. thaliana*[62]. *CCR1* is a biosynthetic enzyme in lignin production and leaf development, which is known to regulate the concentration of the antioxidative compound ferulic acid[63]. This could be related to cold perception, as cold is partly perceived via the redox state[64]. Although all observed candidate SNPs are synonymous or silent, they may still have strong potential consequences for *cis*-regulatory elements, methylation, small RNA regulation and chromatin structure, and associated changes in the promoter. Moreover, alternative splicing was found to occur abundantly in resynthesized *B. napus*, also for copies of *Bna.FLC*[65]. On the other hand, we also cannot fully exclude that the observed effects are caused by linkage to additional genes in the neighbourhood of the investigated flowering-time regulators.

We also found patterns on the *Bna.FT* promoter on C02_random associating with the winter-spring split. Haplotype analysis showed that only two winter lines showed a strongly varying pattern, resembling most of the spring morphotypes. One of these was an exotic line, while the other was found to be vernalisation independent. This indicates that a functional promoter sequence for this gene is necessary to build up vernalisation requirement. A change in vernalisation requirement through variation in an *FT* promoter was already found in different *Brassicas*[66], narrow-leafed lupine[67] and litchi[68]. In *A. thaliana*, *FLC* was shown to bind to a CArG box located in the first intron of *FT*[69], although in cereals the binding site lies in the promoter[70]. Indeed, the *Bna.FT* copy on C02_random contains a CArG box motif. It is possible that both the promoter and the intron are responsible for *Bna.FLC* binding. All the same, in both winter and spring types it has been reported[66] that the A02 copy in *B. napus* is constitutively expressed and the C02 copy is completely silenced, whereas the copies on A07 and C06 appear to be specifically silenced in winter morphotypes but transcribed in spring morphotypes[66]. As our *Bna.FT* copy on C02_random corresponds to the C02 copy reported in the aforementioned study[66], we assume that either there was a problem with the RT-PCR due to allelic variation, or the regulatory mechanism is more complex. However, in an independent transcriptome study we were unable to detect the constitutive expression of *Bna.FT.A02*, nor of any other *Bna.FT* copy, in winter-type *B. napus* before vernalisation (C. Obermeier, unpublished data).

Genomic rearrangements are common in *B. napus*[6,41,71–75]. They are particularly predominant in the first generations after allopolyploidisation[72], but the process is ongoing and believed to have an important role in speciation[76]. Different studies found indications for genomic rearrangements between *B. napus* morphotypes[6,12,29]. In the present study we found CNVs concerning copies of *Bna.FLC*, *Bna.PHYA* and *Bna.GA3ox1* to involve duplications in the A subgenome and corresponding homoeologous deletions in the C subgenome. This indicates replacement of the C-subgenome regions by the respective A-subgenome regions, a process known as homeologous non-reciprocal translocation (HNRT)[77]. A *de novo* HNRT will erase any sub-functionalisation which may have occurred prior to the rearrangement. Our data concur with the hypothesis[27] that the *Bna.FLC.A10* copy is most specifically involved in flowering regulation. A duplication in *Bna.FLC.A10* would therefore increase vernalisation requirement. This hypothesis fits with the strong vernalisation requirement we observed in lines carrying this duplication. Differential expression of *Bna.FLC* in highly rearranged, resynthesized rapeseed was observed before[26]. All the same, this pattern can only be effective when the vernalisation system is functional, as two spring lines with the same pattern are not vernalisation dependent. One of these presumably has a defective *Bna.FT* promoter on C02_random, whereas neither of them carries the swede-specific *Bna.VIN3.A03* marker.

Other genes affected by such HNRTs are *Bna.PHYA* and *Bna.GA3ox1*. PHYA is a red/far-red perceiving photoreceptor which has a stabilising role for CO under long days[78]. This might represent a necessary co-adaptation of the photoperiodic pathway due to the strong vernalisation requirement, as later flowering means that the day length is longer at the time of flowering. This assumption is underlined by the finding that a D94G mutation in a copy of *Bna.CO-like* is a candidate for the swede split. *GA3ox1*, a biosynthetic key gene involved in GA production, is regulated by PHYB and by feedback mechanisms of downstream pathways[79]. GA also affects other developmental processes like seed germination, hypocotyl elongation and fruit set[79,80]. *Bna.GA3ox1* is therefore also a candidate for the swede morphotype, which is characterized by an enlarged hypocotyl and low seed-set. This might also apply to *Bna.CCR1* as a candidate for the swede split. In Arabidopsis *CCR1* is involved in lignin biosynthesis, leaf development regulation and regulation of the redox state[63] (see above).

All swede lines share a silent mutation in *Bna.VIN3.A03*, which is not shared by any other line. Although the consequence of this mutation is unclear, *Bna.VIN3* is a strong candidate for vernalisation requirement, particularly because another copy is a candidate for the winter-spring split (see above). Copies of *VIN3* have previously been named as candidates for vernalisation requirement and flowering time in *A. thaliana* and *B. napus*[54,81]. In *B. oleracea*, it was found that *BoVIN3* was upregulated much faster than *A. thaliana VIN3*[82], indicating that the expression is more sensitive to cold. Another upstream candidate for the strong vernalisation requirement is *Bna.ELF7*. *ELF7* is involved in chromatin remodeling of *FLC* during vernalisation[83]. A further gene variant which was not found outside the swede population lay in *Bna.TEM1*. *TEM1* encodes another repressor of *FT*, which competes with *CO* for the same genetic region to fulfill their function[84]. The variant is a non-synonymous T167R mutation that potentially affects binding to the UTR of *FT*. A further candidate, *Bna.FD*, possibly modulates FT protein effectiveness, as FD is a direct and essential interaction partner of FT in the shoot apex[85].

These results represent an excellent base for further experiments to transfer morphotype features between *B. napus* genetic pools. Moreover, they also shed light on the evolution of major flowering time genes in the aftermath of allopolyploidisation, and their role in morphotype diversification and ecogeographical adaptation. We clearly demonstrate that different copies of important flowering regulators play different regulatory roles across the vernalisation and flowering pathways. The scarcity of non-synonymous mutations, along with the observed variation in the *Bna.FT* promoter, underline the importance of *cis*-regulatory mechanisms in flowering time regulation.

References

1. Adams, K. L. & Wendel, J. F. Polyploidy and genome evolution in plants. *Curr Opin Plant Bio* **8,** 135–141 (2005).
2. Soltis, P. S. & Soltis, D. E. The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* **60,** 561–588 (2009).
3. Arrigo, N. & Barker, M. S. Rarely successful polyploids and their legacy in plant genomes. *Curr Opin Plant Bio* **15,** 140–146 (2012).
4. Brenchley, R. *et al.* Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491,** 705–710 (2012).
5. The Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature* **475,** 189–195 (2011).
6. Chalhoub, B. *et al.* Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science* **345,** 950–953 (2014).
7. Edwards, Dave, Batley, Jacqueline, Parkin, Isobel & Chittaranjan, Kole (ed.) *Genetics*, *Genomics and Breeding of Oilseed Brassicas* (CRC Press, 2011).
8. Bundessortenamt. Beschreibende Sortenliste 2016. Getreide, Mais Öl- und Faserpflanzen Leguminosen Rüben Zwischenfrüchte (2016).
9. Allender, C. J. & King, G. J. Origins of the amphiploid species Brassica napus L. investigated by chloroplast and nuclear molecular markers. *BMC Plant Biol* **10,** 54 (2010).

10. Sharpe, A. G. & Lydiate, D. J. Mapping the mosaic of ancestral genotypes in a cultivar of oilseed rape (*Brassica napus*) selected via pedigree breeding. *Genome* **46** (2003).
11. Parkin, I. A. P. Segmental Structure of the Brassica napus Genome Based on Comparative Analysis With Arabidopsis thaliana. *Genetics* **171,** 765–781 (2005).
12. Schiessl, S., Samans, B., Hüttel, B., Reinhardt, R. & Snowdon, R. J.Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species Brassica napus. *Front. Plant Sci.* **5,** 404 (2014).
13. Cao, J. *et al.* Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* **43,** 956–963 (2011).
14. Díaz, A. *et al.* Copy number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*). *PLoS ONE* **7,** 1–11 (2012).
15. Iovene, M., Zhang, T., Lou, Q., Buell, C. R. & Jiang, J. Copy number variation in potato - an asexually propagated autotetraploid species. *Plant J* **75,** 80–89 (2013).
16. Springer, N. M. *et al.* Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet* **5,** e1000734 (2009).
17. Clop, A., Vidal, O. & Amills, M. Copy number variation in the genomes of domestic animals. *Anim Genet* **43,** 503–517 (2012).
18. Żmieńko, A., Samelak, A., Kozłowski, P. & Figlerowicz, M. Copy number polymorphism in plant genomes. *Theor Appl Genet.* 1–18 (2013).
19. Srikanth, A. & Schmid, M. Regulation of flowering time: all roads lead to Rome. *Cell. Mol. Life Sci.* **68,** 2013–2037 (2011).
20. Kardailsky, I. *et al.* Activation tagging of the floral inducer FT. *Science* **286** (1999).
21. Wollenberg, A. C. & Amasino, R. M. Natural variation in the temperature range permissive for vernalization in accessions of Arabidopsis thaliana. *Plant, Cell & Env* **35,** 2181–2191 (2012).
22. He, Y. Control of the transition to flowering by chromatin modifications. *Mol. Plant* **2,** 554–564 (2009).
23. Turck, F. & Coupland, G. When vernalization makes sense. *Science* **331,** 36–37 (2011).
24. Choi, K. *et al.* The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors. *The Plant Cell* **23,** 289–303 (2011).
25. Tadege, M. *et al.* Control of flowering time by FLC orthologues in *Brassica napus*. *Plant J.* **28** (2001).
26. Pires, J. C. *et al.* Flowering time divergence and genomic rearrangements in resynthesized Brassica polyploids (Brassicaceae). *Biol J Linn Soc* **82,** 675–688 (2004).
27. Hou, J. *et al.* A Tourist-like MITE insertion in the upstream region of the BnFLC.A10 gene is associated with vernalization requirement in rapeseed (*Brassica napus* L.) (2012).
28. Udall, J. A., Quijada, P. A., Lambert, B. & Osborn, T. C. Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 2. Identification of alleles from unadapted germplasm. *Theor Appl Genet* **113,** 597–609 (2006).
29. Quijada, P. A., Udall, J. A., Lambert, B. & Osborn, T. C. Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (Brassica napus L.): 1. Identification of genomic regions from winter germplasm. *Theor Appl Genet* **113,** 549–561 (2006).
30. Nelson, M. N. *et al.* Quantitative trait loci for thermal time to flowering and photoperiod responsiveness discovered in summer annual-type *Brassica napus* L. *PLoS ONE* **9,** e102611 (2014).
31. Raman, H. *et al.* Genetic and physical mapping of flowering time loci in canola (*Brassica napus* L.). *Theor Appl Genet* **126,** 119–132 (2013).
32. Fletcher, R. S., Mullen, J. L., Heiliger, A. & McKay, J. K. QTL analysis of root morphology, flowering time, and yield reveals trade-offs in response to drought in *Brassica napus*. *Ex Bot J* (2014).
33. Bus, A., Hecht, J., Huettel, B., Reinhardt, R. & Stich, B. High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. *BMC Genomics* (2012).
34. Körber, N. *et al.* Seedling development in a *Brassica napus* diversity set and its relationship to agronomic performance. *Theor Appl Genet* **125,** 1275–1287 (2012).
35. Li, R. *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25,** 1966–1967 (2009).
36. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
37. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29,** 24–26 (2011).
38. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr prot bioinf* **47,** 11.12.1–34 (2014).
39. Vergara, I. A., Frech, C. & Chen, N. CooVar: Co-occurring variant analyzer. *BMC Research Notes* (2012).
40. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9,** 357–359 (2012).
41. Schmutzer, T. *et al.* Species-wide genome sequence and nucleotide polymorphisms from the model allopolyploid plant *Brassica napus*. *Scientific data* **2,** 150072 (2015).
42. McKenna, A. *et al.* The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20,** 1297–1303 (2010).
43. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27,** 2156–2158 (2011).
44. Schiessl, S., Iniguez-Luy, F., Qian, W. & Snowdon, R. J. Diverse regulatory factors associate with flowering time and yield responses in winter-type *Brassica napus*. *BMC Genomics* **16,** 737 (2015).
45. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23,** 1294–1296 (2007).
46. Deng, W. *et al.* FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of *Arabidopsis*. *PNAS* **108,** 6680–6685 (2011).
47. Bus, A., Korber, N., Snowdon, R. J. & Stich, B. Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. *Theor Appl Genet* **123,** 1413–1423 (2011).
48. Song, J., Irwin, J. & Dean, C. Remembering the prolonged cold of winter. *Current Biology* **23,** R807–R811 (2013).
49. Duncan, S. *et al.* Seasonal shift in timing of vernalization as an adaptation to extreme winter. *eLife* (2015).
50. Coustham, V. *et al.* Quantitative modulation of polycomb silencing underlies natural variation in vernalization. *Science* **337,** 584–587 (2012).
51. Meyer, S. E., Nelson, D. L. & Carlson, S. L. Ecological genetics of vernalization response in *Bromus tectorum* L. (*Poaceae*). *Ann Bot* **93,** 653–663 (2004).
52. Wu, G., Wu, Y., Xiao, L., Li, X. & Lu, C. Zero erucic acid trait of rapeseed (*Brassica napus* L.) results from a deletion of four base pairs in the *fatty acid elongase 1* gene. *Theor Appl Genet* **116,** 491–499 (2008).
53. Hasan, M. J. & Rahman, H. Genetics and molecular mapping of resistance to *Plasmodiophora brassicae* pathotypes 2, 3, 5, 6, and 8 in rutabaga (*Brassica napus* var. *napobrassica*). *Genome* **59,** 805–815 (2016).
54. Dittmar, E. L., Oakley, C. G., Agren, J. & Schemske, D. W. Flowering time QTL in natural populations of *Arabidopsis thaliana* and implications for their adaptive value. *Mol Ecology* **23,** 4291–4303 (2014).
55. Grillo, M. A., Li, C., Hammond, M., Wang, L. & Schemske, D. W. Genetic architecture of flowering time differentiation between locally adapted populations of *Arabidopsis thaliana*. *New Phytol* **197,** 1321–1331 (2013).
56. Chianga, G. C. K., Baruaa, D., Kramera, E. M., Amasinob, R. M. & Donohuea, K. Major flowering time gene, *FLOWERING LOCUS C*, regulates seed germination in *Arabidopsis thaliana*. *PNAS* **106,** 11661–11666 (2009).
57. Becker, A. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogen Evol* **29,** 464–489 (2003).
58. Dreni, L. & Kater, M. M. MADS reloaded: evolution of the AGAMOUS subfamily genes. *New Phytol* **201,** 717–732 (2014).
59. Johansson, M. & Staiger, D. SRR1 is essential to repress flowering in non-inductive conditions in *Arabidopsis thaliana*. *Ex Bot J* **65,** 5811–5822 (2014).

60. Wood, C. C. *et al.* The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *PNAS* **103,** 14631–14636 (2006).
61. Dorca-Fornell, C. *et al.* The Arabidopsis SOC1-like genes AGL42, AGL71 and AGL72 promote flowering in the shoot apical and axillary meristems. *Plant J* **67,** 1006–1017 (2011).
62. Helliwell, C. A., Wood, C. C., Robertson, M., James Peacock, W. & Dennis, E. S. The Arabidopsis FLC protein interacts directly *in vivo* with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. *Plant J* **46,** 183–192 (2006).
63. Xue, J. *et al.* CCR1, an enzyme required for lignin biosynthesis in Arabidopsis, mediates cell proliferation exit for leaf development. *Plant J* **83,** 375–387 (2015).
64. Kovi, M. R., Ergon, A. & Rognli, O. A. Freezing tolerance revisited-effects of variable temperatures on gene regulation in temperate grasses and legumes. *Curr Opin Plant Bio* **33,** 140–146 (2016).
65. Zhou, R., Moshgabadi, N. & Adams, K. L. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *PNAS* **108,** 16122–16127 (2011).
66. Wang, J. *et al.* Promoter Variation and Transcript Divergence in Brassicaceae Lineages of FLOWERING LOCUS T. *PLoS ONE* **7,** e47127 (2012).
67. Nelson, M. N. *et al.* The loss of vernalization requirement in narrow-leafed lupin is associated with a deletion in the promoter and de-repressed expression of a *Flowering Locus T* (*FT*) homologue. *New Phytol* (2016).
68. Ding, F. *et al.* Promoter difference of *LcFT1* is a leading cause of natural variation of flowering timing in different litchi cultivars (*Litchi chinensis* Sonn.). *Plant science* **241,** 128–137 (2015).
69. Searle, I. *et al.* The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling In *Arabidopsis*. *Genes & Dev* **20,** 898–912 (2006).
70. Deng, W. *et al.* Direct links between the vernalization response and other key traits of cereal crops. *Nat Comms* **6,** 5882 (2015).
71. 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).
72. Gaeta, R. T., Pires, J. C., Iniguez-Luy, F., Leon, E. & Osborn, T. C. Genomic Changes in Resynthesized *Brassica napus* and Their Effect on Gene Expression and Phenotype. *The Plant Cell* **19,** 3403–3417 (2007).
73. Nicolas, S. D., Monod, H., Eber, F., Chevre, A.-M. & Jenczewski, E. Non-random distribution of extensive chromosome rearrangements in *Brassica napus* depends on genome organization. *Plant J* **70,** 691–703 (2012).
74. Udall, J. A., Quijada, P. A. & Osborn, T. C. Detection of Chromosomal Rearrangements Derived From Homeologous Recombination in Four Mapping Populations of *Brassica napus* L. *Genetics* **169,** 967–979 (2005).
75. Szadkowski, E. *et al.* Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. *New Phytol* **191,** 884–894 (2011).
76. Faria, R. & Navarro, A. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends ecol evol* **25,** 660–669 (2010).
77. Gaeta, R. T. & Pires, J. C. Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytol* **186,** 18–28 (2010).
78. Valverde, F. *et al.* Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science* **303,** 1003–1006 (2004).
79. Gabriele, S. *et al.* The Dof protein DAG1 mediates PIL5 activity on seed germination by negatively regulating GA biosynthetic gene *AtGA3ox1*. *Plant J* **61,** 312–323 (2010).
80. Gallego-Giraldo, C. *et al.* Role of the gibberellin receptors GID1 during fruit-set in Arabidopsis. *Plant J* **79,** 1020–1032 (2014).
81. Shi, J. *et al.* Unraveling the Complex Trait of Crop Yield With Quantitative Trait Loci Mapping in *Brassica napus*. *Genetics* **182,** 851–861 (2009).
82. Ridge, S., Brown, P. H., Hecht, V., Driessen, R. G. & Weller, J. L. The role of BoFLC2 in cauliflower (*Brassica oleracea* var. *botrytis* L.) reproductive development. *Ex Bot J* **66,** 125–135 (2015).
83. He, Y., Doyle, M. R. & Amasino, R. M. PAF1-complex-mediated histone methylation of *FLOWERING LOCUS C* chromatin is required for the vernalization-responsive, winter-annual habit in *Arabidopsis*. *Genes & Dev* **18,** 2774–2784 (2004).
84. Castillejo, C. & Pelaz, S. The Balance between CONSTANS and TEMPRANILLO Activities Determines *FT* Expression to Trigger Flowering. *Current Biology* **18,** 1338–1343 (2008).
85. Abe, M. *et al.* FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex. *Science* **309,** 1052–1056 (2005).

Acknowledgements

This work was financed by grant SN 14/13-1 from the German Research Foundation (DFG) within the priority programme "Flowering Time Control: From Natural Variation to Crop Improvement."

Author Contributions

S.S. and R.S. planned the experiments and wrote the manuscript. S.S. performed DNA extraction, data analysis and phenotyping. B.H., D.K. and R.R. performed library preparation and sequencing. SS and RS prepared the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Schiessl, S. *et al.* Post-polyploidisation morphotype diversification associates with gene copy number variation. *Sci. Rep.* **7**, 41845; doi: 10.1038/srep41845 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter 1.3: Flowering Time Gene Variation in *Brassica* Species Shows Evolutionary Principles

Schiessl SV, Huettel B, Kuehn D, Reinhardt R, Snowdon RJ (2017c) Flowering Time Gene Variation in *Brassica* Species Shows Evolutionary Principles. Front. Plant Sci. 8:183. doi: 10.3389/fpls.2017.01742

frontiers in Plant Science

ORIGINAL RESEARCH
published: 17 October 2017
doi: 10.3389/fpls.2017.01742

Flowering Time Gene Variation in *Brassica* Species Shows Evolutionary Principles

Sarah V. Schiessl[1]*, *Bruno Huettel*[2], *Diana Kuehn*[2], *Richard Reinhardt*[2] and *Rod J. Snowdon*[1]

[1] Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Giessen, Germany, [2] Max Planck Institute for Breeding Research, Cologne, Germany

Flowering time genes have a strong influence on successful reproduction and life cycle adaptation. However, their regulation is highly complex and only well understood in diploid model systems. For crops with a polyploid background from the genus *Brassica*, data on flowering time gene variation are scarce, although indispensable for modern breeding techniques like marker-assisted breeding. We have deep-sequenced all paralogs of 35 *Arabidopsis thaliana* flowering regulators using Sequence Capture followed by Illumina sequencing in two selected accessions of the vegetable species *Brassica rapa* and *Brassica oleracea*, respectively. Using these data, we were able to call SNPs, InDels and copy number variations (CNVs) for genes from the total flowering time network including central flowering regulators, but also genes from the vernalisation pathway, the photoperiod pathway, temperature regulation, the circadian clock and the downstream effectors. Comparing the results to a complementary data set from the allotetraploid species *Brassica napus*, we detected rearrangements in *B. napus* which probably occurred early after the allopolyploidisation event. Those data are both a valuable resource for flowering time research in those vegetable species, as well as a contribution to speciation genetics.

Keywords: sequence capture, natural variation, polyploidy, speciation, copy number variation

OPEN ACCESS

Edited by:
Amy Litt,
University of California, Riverside, United States

Reviewed by:
J. Chris Pires,
University of Missouri, United States
Tatiana Arias,
Corporation for Biological Research, Colombia

***Correspondence:**
Sarah V. Schiessl
sarah-veronica.schiessl@agrar.uni-giessen.de

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

Received: *28 July 2017*
Accepted: *25 September 2017*
Published: *17 October 2017*

Citation:
Schiessl SV, Huettel B, Kuehn D, Reinhardt R and Snowdon RJ (2017) Flowering Time Gene Variation in Brassica Species Shows Evolutionary Principles. Front. Plant Sci. 8:1742. doi: 10.3389/fpls.2017.01742

INTRODUCTION

The genus *Brassica* is highly diverse. It contains many phenotypically extremely different vegetable, turnip and oil crops, among them garden turnip, Chinese cabbage and Pak Choi (*Brassica rapa*) and cabbage, broccoli, cauliflower, Brussels sprouts, kale, kohlrabi and savoy (*Brassica oleracea*) (Paterson et al., 2001). Both species are also diploid progenitors of *B. napus*, which comprises rapeseed/ canola and rutabagas (Chalhoub et al., 2014). Adequate regulation of flowering and flowering time is crucial for crop production especially for leafy vegetable crops as in *B. rapa* and *B. oleracea*. Early bolting limits vegetable growth and can therefore severely decrease yield. On the other hand, complete inhibition of flowering interferes with seed production. Knowledge about the impact of flowering time gene variation is therefore crucial for successful vegetable breeding. In the model plant *Arabidopsis thaliana,* flowering time is set by expression of the gene *FLOWERING LOCUS T (FT)* (Srikanth and Schmid, 2011; Blümel et al., 2015). The expression of *FT* is negatively regulated by the transcriptional repressor *FLOWERING LOCUS C (FLC)* in interaction with other genetic factors from the vernalisation pathway, and positively regulated via the transcriptional

activator *CONSTANS (CO)* in interaction with genetic factors from the photoperiod pathway and the circadian clock (Srikanth and Schmid, 2011; Blümel et al., 2015). Other pathways like the ambient temperature pathway, the age pathway, the sugar signaling pathway and the stress pathway are able to modulate the flowering response (Srikanth and Schmid, 2011; Blümel et al., 2015). All the same, vernalisation and day length have major effects on flowering time. Although *B. rapa* and *B. oleracea* are closely related to *A. thaliana*, their reaction to vernalisation is different. Whereas *A. thaliana* and *B. rapa* respond to seed vernalisation, *B. oleracea* requires plant vernalisation (Lin, 2005; Zhang et al., 2015) and is not responsive at early seedling stages. As in *A. thaliana*, annual (vernalisation-independent) and biennual (vernalisation-dependent) forms exist within both *Brassica* species (Camargo and Osborn, 1996). Foregoing studies have identified different orthologous copies of *FLC* and *FT* as strong candidates for flowering time regulation in *B. rapa* and *B. oleracea*, both in the absence and in the presence of vernalisation (Pires et al., 2004; Lin, 2005; Razi et al., 2008; Zhao et al., 2010; Li et al., 2013; Zhang et al., 2015). *Brassica rapa* has been found to carry 2 copies of *Bra.FT* (Zhang et al., 2015) (A02 and A07), 4 copies of *Bra.FLC* (Schranz et al., 2002) (A02, two copies on A03, A10) and 3 copies of *Bra.CO* (A01, A03, A10). In contrast, *B. oleracea* seems to carry 4 copies of *Bol.FT* (two copies on C02, C04, and C06), 5 copies of *Bol.FLC* (Razi et al., 2008)(one copy on C02, two copies on C03, two copies on C09) and 3 copies of *Bol.CO* (C01, C03, and C09). *Bra.FT.A07*, often referred to as *BrFT2*, was found to underlie a strong QTL for flowering time, possibly due to a transposon insertion in the mapping parent R-o-18 (Zhang et al., 2015). *Bra.FLC.A02*, also referred to as *BrFLC2*, was found to underlie QTLs for flowering time in different studies (Zhao et al., 2010; Xiao et al., 2013; Zhang et al., 2015), possibly due to a 57 bp InDel in the fourth exon and the forth intron of the gene, leading to a non-functional allele (Wu et al., 2012). Another *FLC* copy, *Bra.FLC.A10*, also referred to as BrFLC1, was associated to flowering time due to alternative splicing via variation in intron 6 (Yuan et al., 2009; Wu et al., 2012). A CO-like copy on A02 co-localized with a flowering QTL in a DH population derived from a Chinese cabbage and a rapid cycling line (Li et al., 2013). Different patterns of functional polymorphisms, including premature stop codons, non-synonymous SNPs and differential promotor structure have been found for *Bol.FLC* copies (Okazaki et al., 2007; Razi et al., 2008; Irwin et al., 2016). Both copies on C03 (formerly referred to as BoFLC3 and BoFLC5) as well as one copy on C09 (referred to as BoFLC1) were found to co-localize with flowering time QTL (Razi et al., 2008). A further copy, referred to as BoFLC2 or BoFLC4, was assumed to be a pseudogene located on C02 (Razi et al., 2008), but was found to underlie a QTL in a different study due to a 1 bp deletion (Okazaki et al., 2007) Copies of *Bol.CO* were also suggested as candidate genes for QTL in *B. oleracea*, for example, *Bol.CO.C09* (Okazaki et al., 2007). Most previous research has therefore focused on the central flowering regulators *FLC, FT,* and *CO*, whereas other genes which might modulate the flowering response have been largely ignored. In order to provide a more complete description of genetic variation in central flowering time genes, we deep-sequenced representatives of two *B. rapa* subspecies (L58, ssp. *parachinensis*, R-o-18, ssp. *tricolaris*) along with two different genotypes of *B. oleracea* ssp. *capitata* (BRA1398, Kashirka) for a set of flowering time genes, using a sequence capture approach followed by Illumina sequencing. The data allowed estimation of copy number and sequence variation including SNPs and InDels. All those sequence variants are potentially influential on the phenotype and therefore an interesting resource to vegetable breeders. Comparison to previous data from the same genes in the allopolyploid hybrid species *B. napus* (Schiessl et al., 2014, 2017a,b) provide new insight into the genetic history of *B. napus* and are discussed along with the sequence data from its diploid progenitors.

MATERIAL AND METHODS

Plant Material and DNA Extraction

Two inbred *B. rapa* lines and two *B. oleracea* genotypes were used for the present study. The two *B. rapa* lines, both annuals, were L58, a caixin line (ssp. *parachinensis*) and R-o-18, a yellow sarson line (ssp. *tricolaris*). Both had been used as parents for DH populations before (Bagheri et al., 2012; Zhang et al., 2015). The two *B. oleracea* genotypes were the annual BRA1398 (*ssp. capitata convar. botrytis var. botrytis L.*) and the biennial Kashirka (ssp. *capitata*), a late flowering Siberian kale.

Leaf material from 4 week old plants grown in pots in the greenhouse was collected and immediately shock-frozen in liquid nitrogen. DNA was then extracted from grinded leaf material using a common CTAB protocol modified from Doyle (1990) as described before (Schiessl et al., 2014). DNA concentration was measured using a Qubit fluorometer (Qubit dsDNA BR assay kit, Life Technologies, Darmstadt, Germany) according to the manufacturer's protocol. DNA quantity and purity was further checked on 0.5% agarose gel (3 V/cm, 0.5xTBE, 120 min) stained with ethidium bromide.

Target Genes

The four samples were re-sequenced using targeted deep sequencing along with 280 *B. napus* genotypes as described elsewhere (Schiessl et al., 2017a,b). In brief, flowering time genes involving the most important flowering regulation pathways as known from *Arabidopsis thaliana* were checked for *Brassica* orthologs. Those included genes from the circadian clock *(CYCLING DOF FACTOR 1 (CDF1), EARLY FLOWERING 3 (ELF3), GIGANTEA (GI),* and *ZEITLUPE (ZTL)*), the vernalisation pathway (*EARLY FLOWERING 7 (ELF7), EARLY FLOWERING IN SHORT DAYS (EFS), FLOWERING LOCUS C (FLC), FRIGIDA (FRI), SHORT VEGTATIVE PHASE (SVP), SUPPRESSOR OF FRIGIDA 4 (SUF4), TERMINAL FLOWER 2 (TFL2), VERNALISATION 2 (VRN2), VERNALISATION INSENSITIVE 3 (VIN3)*), the photoperiod pathway (*CONSTANS (CO), CRYPTO-CHROME 2 (CRY2), PHYTOCHROME A (PHYA), PHYTOCHROME B (PHYB)*) and gibberellin signaling (*GIBBERELLIN-3-OXIDASE 1 (GA3ox1)*), along with downstream signal transducers (*AGAMOUS-LIKE 24 (AGL24), APETALA 1 (AP1), CAULIFLOWER (CAL), FLOWERING LOCUS D (FD), FLOWERING LOCUS T (FT), FRUITFUL*

(FUL), LEAFY (LFY), SQUAMOSA PROMOTOR PROTEIN LIKE 3 (SPL3), SUPPRESSOR OF CONSTANS 1 (SOC1), TEMPRANILLO 1 (TEM1), TERMINAL FLOWER 1 (TFL1)). On top, we also included 6 further genes: *CIRCADIAN CLOCK ASSISTED 1 (CCA1), FLAGELLIN-SENSITIVE 2 (FLS2), GLYCIN-RICH PROTEIN 7 (GRP7), GLYCIN-RICH PROTEIN 8 (GRP8), GORDITA (GORD)* and *SENSITIVITY TO RED LIGHT REDUCED 1 (SRR1)*, giving a total of 35 genes.

Bait Development

In order to enrich for the respective target regions, a bait pool was constructed based on selected sequences from *B. rapa, B. oleracea,* and *B. napus*. A detailed description of the bait pool development can be found in (Schiessl et al., 2017a). The bait pool consisted of 178 bait groups, 63 bait groups for *B. rapa* orthologs, 71 bait groups for *B. oleracea* orthologs and 24 bait groups for *B. napus* orthologs. As a short summary, baits were first developed in the program eArrayXD using sequences from *B. rapa* and *B. oleracea*. After a preliminary sequencing test with four diverse *B. napus* genotypes (Schiessl et al., 2014), the bait pool was refined and some sequences were replaced by *B. napus* sequences using the Agilent Genomic Workbench program SureDesign (Agilent Inc., Santa Clara, CA, USA). This improved the specificity of the bait pool (Schiessl et al., 2017a). Bait groups were created using the "Bait Tiling" tool. The parameters were set as follows: Sequencing Technology: "Illumina," Sequencing Protocol: "Paired-End long Read (75 bp+)," "Use Optimized Parameters (Bait length 120, Tiling Frequency 1x)," Avoid Overlap: "20," "User defined genome," "Avoid Standard Repeat Masked Regions."

Library Preparation and Sequencing

Custom bait production was carried out by Agilent Technologies (Agilent Inc., Santa Clara, CA, USA) using the output oligonucleotide sequences from SureDesign. Sequence capture was performed at the GenomeCenter at the Max Planck Institute for Breeding Research (Cologne, Germany) using the SureSelectXT 1–499 kb Custom Kit (Agilent Inc., Santa Clara, CA, USA) according to the manufacturer's instructions. The resulting TruSeq DNA library (Illumina Inc., San Diego, CA, USA) was sequenced on an Illumina HiSeq 2500 sequencer at the Max Planck Institute for Breeding Research (Cologne, Germany) in 100 bp single read mode.

Sequence Analysis

Alignment

Quality control of the raw sequencing data was performed using FASTQC. Reads were mapped both onto version 4.1 of the *B. napus* "Darmor-Bzh" reference genome sequence assembly (Chalhoub et al., 2014) and either onto version 1.5 of the *B. rapa* "Chiifu-401-42" reference genome (Wang et al., 2011) for *B. rapa* reads or onto version 2.1 "TO1000" of the *B. oleracea* reference genome (Parkin et al., 2014) for *B. oleracea* reads. Mapping was performed using the SOAPaligner algorithm (Li R. et al., 2009), with default settings. Removal of duplicates, sorting and indexing was carried out with samtools version 0.1.19 (Li H. et al., 2009). Alignments were visualized using the IGV browser version 2.3.12 (Robinson et al., 2011). For InDel calling, a separate mapping using Bowtie2 (Langmead and Salzberg, 2012) was performed, as described in (Schmutzer et al., 2015), on the reference genome of *B. rapa* and *B. oleracea*. Removal of duplicates, sorting and indexing was again carried out with samtools version 0.1.19.

As it turned out that *Bol.FLC.C02* is likely to be misassembled in the reference genome, we cut out all *Bol.FLC* copies from the reference genome except *Bol.FLC.C02*, which we replaced by *Bol.FLC.C2.E9* (GenBank accession KU521323.1 Irwin et al., 2016). The resulting fasta file was used as artificial genome and the mapping was performed accordingly.

CNV Calling

We first defined regions with sufficient coverage (normalized mean coverage at least 10) for *B. rapa* and *B. oleracea* mapped on their respective reference genomes. A region is defined as being covered by at least two overlapping reads. The coverage was calculated using the bedtools software with multiBamCov (Quinlan, 2014) and normalized to region length, genotype read number and genome size. Those regions were subjected to BLAST against the target regions found for *B. napus* in (Schiessl et al., 2017a) using a *e*-value cut-off of e^{-50} using the program BioEdit version 7.2.0. Moreover, a bed file with the positions of those regions was compared to the gene positions of the respective reference genomes. All regions which either overlapped with an annotated gene or alternatively had a BLAST hit to a *B. napus* target region were analyzed for CNVs. In order to have comparable coverages, we used the gene positions wherever possible, and calculated the normalized coverages on those positions.

In a second step, we compared the coverage ratio between both genotypes of a species. If one genotype had less than 50% coverage than the other, we assumed an unbalanced coverage ratio, indicating a CNV. In case that the coverage of one of the genotypes was less than 30% of the other, we assumed a deletion. For all other cases, we compared the coverage of this region to the respective coverage obtained for the orthologous region in a population of 280 *B. napus* genotypes (Schiessl et al., 2017a). If the coverage ratio was less than 30%, we assumed a deletion. In all other cases, we assumed a duplication for the other genotype.

SNP and InDel Calling and Annotation

Calling of single nucleotide polymorphisms (SNPs) was performed with the algorithm mpileup in the samtools toolkit (Li H. et al., 2009). Calling of InDels was performed based on a separate alignment using Bowtie2. An initial InDel calling was first performed using samtools mpileup, and realignment of reads around InDels was then performed using GATK RealignerTargetCreator, version 3.1.1 (McKenna et al., 2010). A final InDel calling was then performed as described above. SNPs were filtered for a minimum mapping quality of 50 and a read depth of at least 10, and InDels were filtered for a minimum mapping quality of 30 and a read depth of at least 10 using vcftools (Danecek et al., 2011). SNP and InDel annotation was performed using CooVar (Vergara et al., 2012).

As InDel calling with this read length and mapping parameters is limited to InDels of a length of 18 bp, we conducted another approach where we searched for regions of

zero coverage in a 19 bp window which were strongly covered in the respective other genotype using the SOAP2 mapping, while having a low coverage using the realigned Bowtie2 mapping in the same genotype. This approach ensures for the detection of larger deletions, which are not due to reference mapping problems.

Data and Seed Availability

The aligned bamfiles are available via NCBI SRA (https://www.ncbi.nlm.nih.gov/sra/), study accession SRP119226 (Flowering time genes in *B. rapa* and *B. oleracea*).

The two accessions of *B. oleracea* BRA1398 and Kashirka (BRA1506) are available via the Genbank at IPK Gatersleben. The two accessions of *B. rapa* L58 and R-o-18 are available from the authors upon request.

RESULTS

Gene Copies in *B. rapa, B. oleracea,* and *B. napus*

In the two *B. rapa* accessions we found 1405 regions with a mean normalized coverage of at least 10, among them 222 regions co-localizing with an annotated *B. rapa* gene. Of those, 105 regions had a BLAST hit to a target gene of *B. napus* as analyzed in (Schiessl et al., 2017b), while 95 regions of those with a BLAST hit co-localized with an annotated *B. rapa* gene. We therefore analyzed 228 regions for *B. rapa,* which either co-localized with a *B. rapa* gene or had a BLAST hit to a target *B. napus* gene or both, excluding non-genic regions from the analysis. For the two *B. oleracea* accessions we found 3010 regions with a normalized coverage of at least 10, with 365 regions co-localizing with an annotated *B. oleracea* gene and 111 regions showing a BLAST hit to a target *B. napus* region. In total, we analyzed 384 regions for *B. oleracea*.

When mapping the sequencing reads from *B. rapa* and *B. oleracea* onto the *B. napus* reference genome, we found that *BnaA02g16710D* (*Bna.ZTL.A02*) had a strongly reduced normalized coverage in both *B. rapa* lines (2 and 13% of the mean *B. napus* coverage), while both *B.oleracea* lines had significant coverage at this locus (169 and 159% of the mean *B. napus* coverage). The raw read depth landscape for this locus is shown in **Figure 1**. The *B. rapa* genome also did not carry a respectively annotated gene on A02, while the *B. oleracea* genome carries an additional copy on a non-localized scaffold (**Figure 2**).

In contrast, *BnaCnng78500D* (*Bna.LFY.Cnn*) showed a much higher coverage in both *B. rapa* lines (270 and 256% of the mean *B. napus* coverage) than in both *B. oleracea* (41 and 45% of the mean *B. napus* coverage). Because the total size of the sequenced gene space was about half the size of the allotetraploid while the read number was comparable, a normalized coverage increase of around 200% would be expected for the diploid species.

In contrast to *B. napus*, we found no orthologs to the four further gene copies *BnaA03g24400D* (*Bna.SRR1.A03*), *BnaA10g27730D* (*Bna.CRY2.A10b*), *BnaAnng24480D* (*Bna.CDF1.Ann*) and *BnaAnng38870D* (*Bna.CO.Annb*) in *B. rapa*. All the same, when mapping the *B. rapa* reads onto the *B. napus* genome, we found significant coverage on those loci (176, 123, 207, and 87% of the *B. napus* coverage, respectively). This might either indicate mismapping or missing information in the reference genome. Doing so for *B. oleracea*, we did not find orthologs to *BnaC08g10770D* (*Bna.VRN2.C08*) and *BnaCnng50250D* (*Bna.LFY.Cnna*). The last one could point to a stable exchange between the A and C subgenomes similar to *BnaCnng78500D* (*Bna.LFY.Cnnb*), because it had only 51% coverage with *B. oleracea* reads, while having 136% using *B. rapa* reads. This is further supported when looking at the neighbor-joining tree for LFY (**Figure 3A**). On the other hand, we also noticed that the *B. napus* sequence of *BnaCnng50250D (Bna.LFY.Cnna)* contained patches of NNN (unknown sequence), so is possibly an artifact of the *B. napus* reference genome and might also interfere with the mapping for *BnaCnng78500D* (*Bna.LFY.Cnnb*). *Bna.VRN2.C08* had 173% using *B. oleracea* reads. For *Bol.CDF1.C02*, we found a strongly covered region without any *B. napus* ortholog, which could indicate a rearrangement from *Bol.CDF1.C02* to *Bna.CDF1.Ann* (see also **Figure 3B**). An overview of all putative gains and losses can be found in **Figure 2**. *BnaC02g00490D* (*Bna.FLC.C02*), showing 192% coverage using *B. oleracea* reads was not found to be covered at the respective locus *Bo2g166560 (Bol.FLC.C02)* in *B. oleracea* itself in both lines. We assume that this is due to a misassembled copy in the reference genome in *B. oleracea*, due to phylogenetic analysis and comparisons to sequences for *Bol.FLC.C2* published by (Irwin et al., 2016). When we performed a separate alignment only to the *Bol.FLC* copies replacing *Bol.FLC.C02* by the published version in (Irwin et al., 2016), we got a coverage in the expected range on *Bol.FLC.C2.E9*. Our data suggest that there is a misassembly between *Bol.FLC.C02* and *Bol.FLC.C03a*.

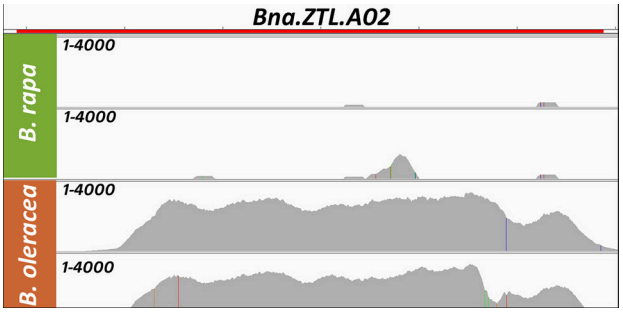


FIGURE 1 | Coverage (Raw read depth) for both genotypes of *B. rapa* (green, A subgenome) showing low coverage and both genotypes of *B. oleracea* (orange, C subgenome) showing high coverage when mapped on the *B. napus* genome for a gene located on the A subgenome in *B. napus*: *Bna.ZTL.A02* (*BnaA02g16710D)* as an example for early rearrangements in the *B. napus* genome. The height of the gray area in each lane marks the raw read depth and is shown from 1 to 4000 as read count per base as non-dimensional number. The red bar on top marks the gene extension.

Copy Number Variation

For all analyzed regions which co-localized with a respectively annotated gene, we analyzed the coverage for the annotated gene

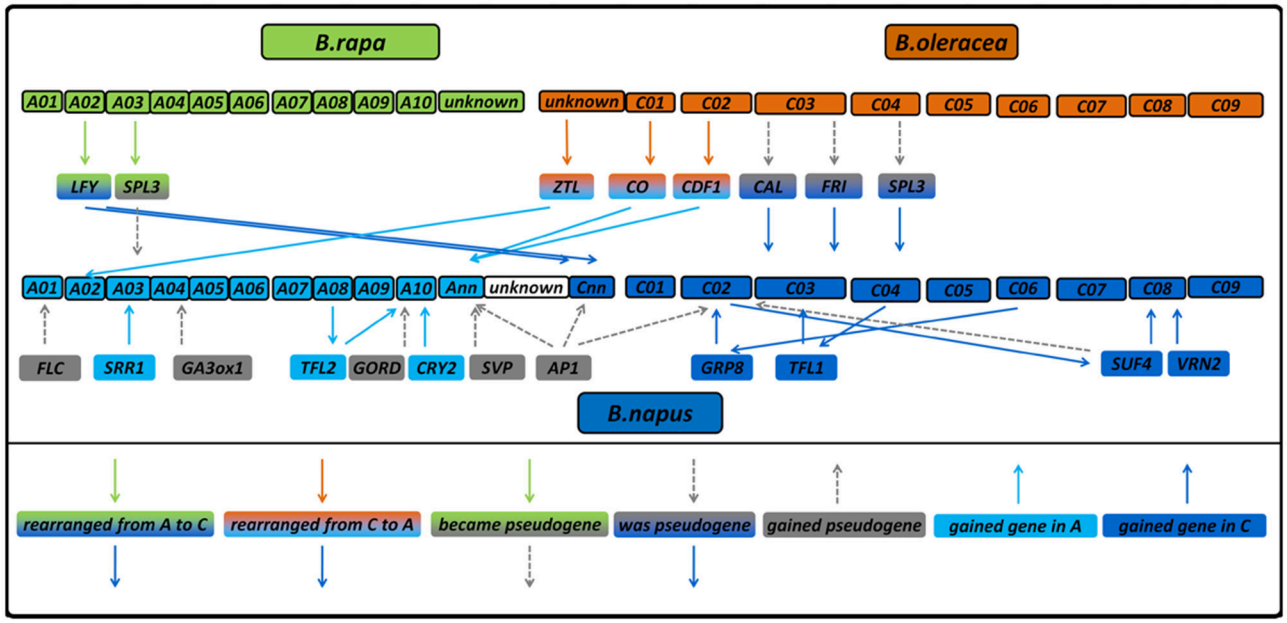


FIGURE 2 | Diagram showing gene copies which were lost, gained or rearranged in *B. napus* compared to its diploid progenitors. The diagram also shows genetic regions which were considered to be pseudogenes in *B. napus* before (Schiessl et al., 2017a) without having a homologous region in the diploids, those which have a homologous region and those which were not annotated as a gene in the diploids. Gene copies without changes are not displayed.

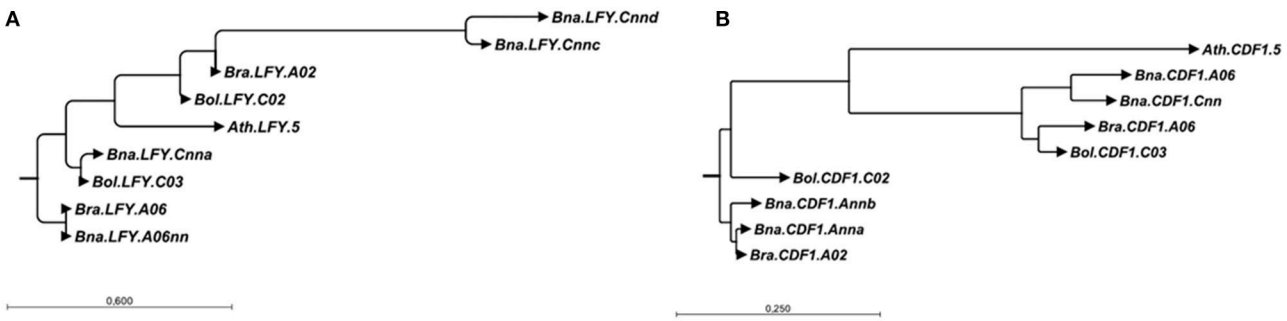


FIGURE 3 | Neighbor-joining trees for all copies of *LFY* **(A)** and *CDF1* **(B)** in *B. rapa*, *B. oleracea* and *B. napus* using the *A. thaliana* gene as outgroup. The trees were constructed based on aligments made in CLC SequenceViewer version 7.8.1 with bootstrap analysis (100 replicates).

positions. For those which only had a BLAST hit to a target *B. napus* gene, we analyzed the coverage for the total length of the covered region.

For *B. rapa*, we found 61 regions with an unbalanced coverage ratio, meaning that the coverage for one of the genotypes was at least 50% higher than for the other. Of those, 55 regions were genic in *B. rapa*. While 13 genic regions showed clear deletions (one genotype had less than 30% normalized coverage compared to the other), the coverage patterns of the other regions were less obvious. Therefore, in order to distinguish between a duplication in one genotype and a deletion in the other, we compared the normalized coverages of the *B. rapa* region to the respective normalized mean coverage of the corresponding region in a population of 280 genotypes of *B. napus* (Schiessl et al., 2017b). All regions with less than 30% coverage compared to *B. napus* were considered to be deleted. All other unbalanced coverage ratios were assigned as duplication to the respective genotype. According to this definition, we found 15 genic deletions and 25 genic duplications in the genotype L58, and 9 genic deletions and 10 duplications in the genotype R-o-18. The CNVs concerning target genes are summarized in **Table 1**.

For *B. oleracea*, we found 118 regions with an unbalanced coverage ratio, with 110 regions being genic in *B. oleracea*. Applying the same thresholds as for *B. rapa*, we found 8 genic deletions and 38 genic duplications for genotype BRA1398, as well as 11 genic deletions and 46 duplications for genotype Kashirka. The respectively concerned target genes are summarized in **Table 2**.

TABLE 1 | Gene names, Gene IDs, chromosomal location and normalized coverages for target flowering genes for the two sequenced accessions of *B. rapa*.

Target gene	Gene ID *B. napus*	Gene ID *B. rapa*	Chromosome *B. rapa*	Mean normalized coverage *B. napus* population	Mean normalized coverage *B. rapa* on *B. napus*	Mean normalized coverage *B. rapa* on *B. rapa*	Normalized coverage genotype L58 on *B. napus*	Normalized coverage genotype L58 on *B. rapa*	Assumed nature of the CNV
Bna.VIN3.A02	*BnaA02g08140D*	*Bra020445*	A02	1,561.9	4,982.2	1,441.3	6,330.0	1,833.2	Duplication
Bna.TEM1.A02	*BnaA02g14040D*	*Bra038346*	A02	1,281.5	2,111.5	702.6	2,637.1	884.7	Duplication
Bna.PHYB.A03	*BnaA03g34390D*	*Bra001650*	A03	1,637.4	3,159.4	1,076.7	3,986.9	659.3	Deletion
Bna.SVP.A04	*BnaA04g12990D*	*Bra030228*	A04	1,196.7	2,485.7	960.0	3,246.9	1,238.4	Duplication
Bna.CRY2.A08	*BnaA08g27870D*	*Bra030568*	A08	1,195.9	1,769.3	950.9	1,788.6	712.5	Deletion
Bna.SUF4.A09	*BnaA09g25530D*	*Bra023153*	A09	1,467.6	2,513.5	1,249.9	2,506.9	1,512.6	Duplication
Bna.GA3ox.A09	*BnaA09g57140D*	*Bra026757*	A09	1,675.6	1,114.2	594.5	720.1	425.4	Deletion
Bna.CO-li.A10	*BnaA10g18420D*	*Bra008668*	A10	2,248.3	3,936.3	1,398.6	4,418.0	994.0	Deletion
Bna.SUF4.Ann	*BnaAnng11220D*	*Bra039880*	A08	1,050.3	2,338.2	1,042.0	1,600.3	1,261.8	Duplication
Bna.CO.Ann	*BnaAnng39160D*	*Bra021464*	A01	795.4	2,212.1	1,009.1	3201.7	1,304.3	Duplication
Target gene	**Gene ID *B. napus***	**Gene ID *B. rapa***	**Chromosome *B. rapa***	**Mean normalized coverage *B. napus* population**	**Mean normalized coverage *B. rapa* on *B. napus***	**Mean normalized coverage *B. rapa* on *B. rapa***	**Normalized coverage genotype R-0-18 on *B. napus***	**Normalized coverage genotype R-0-18 on *B. rapa***	**Assumed nature of the CNV**
Bna.PHYA.A06	*BnaA06g05470D*	*Bra020013*	A06	1,104.8	1,423.0	826.4	1,521.4	1,077.1	Duplication
Bna.ZTL.A07	*BnaA07g01230D*	*Bra038830*	A07	1,265.0	3,121.8	1,343.2	2,839.7	1,627.4	Duplication
Bna.EFS.A07	*BnaA07g33460D*	*Bra015678*	A07	1,171.4	2,417.1	945.9	3,089.6	1,179.9	Duplication
Bna.EFS.A07	*BnaA07g33460D*	*Bra015678*	A07	1,171.4	2,417.1	945.9	3,089.6	1,179.9	Duplication
Bna.ELF7.A10	*BnaA10g27050D*	*Bra009582*	A10	1,094.2	2,348.5	1,120.3	2,609.5	1,386.0	Duplication
Bna.TFL1.Ann	*BnaAnng00810D*	*Bra028815*	A02	1,576.4	2,931.8	1,266.9	3,435.0	1,542.4	Duplication
Bna.GRP7.Ann		*Bra031210*	A09	322.3	670.1	510.3	594.5	700.5	Duplication

TABLE 2 | Gene names, Gene IDs, chromosomal location and normalized coverages for target flowering genes for the two sequenced lines of *B. oleracea*.

Target gene	Gene ID *B. napus*	Gene ID *B.oleracea*	Chromosome *B.oleracea*	Mean normalized coverage *B. napus* population	Mean normalized coverage *B. oleracea* on *B. napus*	Mean normalized coverage *B. oleracea* on *B. oleracea*	Normalized coverage genotype BRA1398 on *B. napus*	Normalized coverage genotype BRA1398 on *B.oleracea*	Assumed nature of the CNV
Bna.FT.C02	*BnaC02g23820D*	*Bo01129s030*	Scaffold01129	95.2	69.3	51.4	0.5	0.6	Deletion
Bna.FRI.C03	*BnaC03g16130D*	*NA*	C3	1,092.3	1,965.8	NA	1,721.2	NA	Duplication
Bna.FT.C06	*BnaC06g27090D*	*Bo6g099320*	C6	490.3	888.6	1209.0	627.5	1,514.5	Duplication
Bna.EFS.C06	*BnaC06g38010D*	*Bo6g121240*	C6	1,379.9	2,242.1	1,481.9	3,276.3	2,014.1	Duplication
Bna.SUF4.C08	*BnaC08g09340D*	*Bo2g121020*	C2	1,321.7	4,003.0	2,995.4	5,437.6	4,023.6	Duplication
Bna.PHYB.C08	*BnaC08g10540D*	*Bo8g043460*	C8	76.8	45.9	27.6	0.3	0.2	Deletion
Target gene	**Gene ID *B. napus***	**Gene ID *B.oleracea***	**Chromosome *B.oleracea***	**Mean normalized coverage *B. napus* population**	**Mean normalized coverage *B. oleracea* on *B. napus***	**Mean normalized coverage *B. oleracea* on *B. oleracea***	**Normalized coverage genotype Kashirka on *B. napus***	**Normalized coverage genotype Kashirka on *B. oleracea***	**Assumed nature of the CNV**
Bna.AP1.C02	*BnaC02g44500D*	*Bo2g062650*	C2	1,635.1	1,891.5	2,082.4	3,040.8	3,126.3	Duplication
Bna.TFL1.C03	*BnaC03g47080D*	*Bo3g040460*	C3	904.5	1,028.9	752.8	1,522.7	1,055.7	Duplication
Bna.CAL.C03	*BnaC03g56640D*	*Bo00825s090*	Scaffold00825	822.7	842.2	1,237.0	1,206.9	1,718.7	Duplication
Bna.TFL1.C04	*BnaC04g16750D*	*Bo4g074330*	C4	1,290.8	2,202.5	1,403.5	1,850.2	1,742.2	Duplication
Bna.SRR1.C09	*BnaC09g34850D*	*Bo9g139490*	C9	1,087.5	1,649.7	1,324.3	1,243.6	1,669.0	Duplication
Bna.CO.C09	*BnaC09g41990D*	*Bo9g163730*	C9	1,466.0	1,103.9	1,004.9	3.2	2.7	Deletion
Bna.FLC.C09	*BnaC09g46540D*	*Bo9g173400*	C9	1,031.5	4,037.8	1,523.2	4,302.0	1,885.6	Duplication
Bna.VRN2.Cnn	*BnaCnng45490D*	*Bo5g078770*	C5	207.0	253.1	155.1	337.3	227.1	Duplication

SNPs and Indels

We called a total of 4409 SNPs and 1048 short InDels for *B. rapa* and 6743 SNPs and 1092 short InDels for *B. oleracea*. For *B. rapa*, 11.1% of the total SNPs and 4.3% of the total InDels were heterozygous, while 27.6% of the total SNPs and 11.4% of the total InDels were heterozygous for *B. oleracea*. As shown in **Tables S1**, **S2**, 1436 SNPs and 22 InDels were target variants for *B. rapa*, while 1179 SNPs and 22 InDels were target variants for *B. oleracea*. Taking SNPs and InDels together, the heterozygosity in *B. rapa* was 0.8% in the target regions (1.0% for L58 and 0.6% for R-o-18) and 15.7% for *B. oleracea* (5.2% for BRA1398 and 20.2% for Kashirka). The higher level of heterozygosity in *B. oleracea* was expected, as the species has a high level of self-incompatibility. The heterozygosity between both lines also varied more for *B. oleracea* than for *B. rapa* (**Figure 4**). Only homozygous variants were considered as true variants and used for further analysis. The variant distribution is shown in **Figure 5** (R-o-18) and in **Figure 6** (Kashirka) (see **Figures S1**, **S2** for L58 and BRA1398). Almost all gene copy groups showed considerable variation in all genotypes. For L58, gene copy groups without putative functional variation were *Bra.SUF4, Bra.ELF7,* and *Bra.SVP* from the vernalisation pathway and *Bra.AP1, Bra.CAL, Bra.LFY,* and *Bra.SOC1* from the effector network. For R-o-18, gene copy groups without putative functional variation were *Bra.ELF7* from the vernalisation pathway, and *Bra.CAL* and *Bra.SOC1* from the effector network. Concerning *B. oleracea*, genotype BRA1398, only *Bol.AP1* and *Bol.SPL3* from the effector network remained without putative functional variation, while Kashirka did not show putative functional variation for *Bol.SUF4* and *Bol.ELF7* from the vernalisation pathway, for *Bol.GI* and *Bol.CDF1* from the photoperiodic pathway, for *Bol.GA3ox1* from GA signaling and for *Bol.SPL3* from the effector network.

As InDels larger than 18 bp were not detected with our read length and mapping parameters, we used an additional approach to detect larger deletions (see section Materials and Methods). For *B. rapa*, we found larger deletions for one gene copy in L58 (*Bra.PHYA.A06*) and for 5 copies in R-o-18 (among them *Bra.FLC.A02, Bra.TEM1.A02, Bra.SUF4.A08, Bra.CRY2.A10*). For *B. oleracea*, we did not find larger deletions in the annotated target genes.

Variation in Central Flowering Regulators

In *B. rapa*, the two copies of the central flowering regulator *FT* showed a low SNP variation. There was only one conservative L49I mutation in *Bra.FT.A02* in L58. However, there were 6 InDels in *Bra.FT.A02* (2 deletions, 4 insertions) for L58 and 2 deletions in the same copy in R-o-18 and an additional insertion into *Bra.FT.A07*. In contrast, there were 4 non-synonymous SNPs for *B. oleracea FT* copies. One of them was a radical W170C mutation in *Bol.FT.C04* in BRA1398, another a moderately radical H81Y mutation in the same copy in both genotypes. This copy also carried an insertion in both genotypes. The other two mutations, both found for *Bol.FT.C06* in Kashirka, were conservative (R21Q and E59D).

For *FLC* orthologs, there was more variation in *B. oleracea* than in *B. rapa*. There was a moderately radical R193P mutation and a deletion in *Bra.FLC.A02* in R-o-18 and a conservative T20P mutation along with a splice donor variant in *Bra.FLC.A10* for both genotypes, while the two copies on A03 remain almost conserved, with one insertion into *Bra.FLC.A03 (Bra022771*) shared by both genotypes. In contrast, the *B. oleracea* genotype BRA1398 shows variation in all *FLC* orthologs. Those were a moderately radical G110V mutation and a moderately conservative K79N mutation in *Bol.FLC.C03 (Bo3g005470*), a conservative I173V mutation shared with Kashirka in *Bol.FLC.C03 (Bo3g024250),* a moderately radical T176N mutation also shared with Kashirka in *Bol.FLC.C09 (Bo9g173370*) and a conservative S168N mutation in *Bol.FLC.C09 (Bo9g173400)*. Both genotypes share an insertion in *Bol.FLC.C09 (Bo9g173370*). Kashirka also carries a conservative R24Q mutation in *Bol.FLC.C03 (Bo3g024250*) and putatively has *Bol.FLC.C09 (Bo9g173400*) duplicated. As *Bol.FLC.C09a* (*Bo9g173370*) seems to have an improper stop codon, producing a distinctly longer peptide, we assume that this copy is non-functional. As we found that *Bol.FLC.C02* appears to be misassembled in the *B. oleracea* genome, we called SNPs separately compared to *Bol.FLC.C2.E9*, a sequenced copy from genotype E9 published in (Irwin et al., 2016). We called one moderately radical A75D mutation in Kashirka compared to E9.

Orthologs of the key photoperiodic transcription factor *CO* show more variation in *B. rapa* than in *B. oleracea*. *Bra.CO.A01* shows two moderately radical (A60E, C237S), two moderately conservative (D16G, P130Q) and two conservative mutations (H167Q, Q181E) in R-o-18. In L58, this copy does not carry non-synonymous SNPs, but appears to be duplicated. *Bra.CO.A03* carries a radical F146S mutation, two moderately radical (A20D, Q92L) mutations, one moderately conservative E33G mutation and one conservative I192V mutation in both genotypes. Both *B. oleracea* genotypes carry a moderately radical Y100S mutation in *Bol.CO.C01*, while BRA1398 also carries two moderately radical mutations (K145I, G223R) and one conservative E71Q mutation on *Bol.CO.C03*. The copy *Bol.CO.C09* appears to be deleted in Kashirka.

DISCUSSION

Network Variation in Vegetable Species

Our sequencing data provide considerable novel data on variation among numerous flowering time regulatory genes in *B. rapa* and *B. oleracea*. We confirmed the functional conservation of *BrFLC2 (Bra.FLC.A02)* in the *B. rapa* genotype L58 (Wu et al., 2012) and a previously identified SNP resulting in a splice variant in *BrFLC1 (Bra.FLC.A10)* (Yuan et al., 2009; Wu et al., 2012). A larger deletion in *Bra.FLC.A02* in genotype R-o-18 colocalizes with a 57 bp deletion at the same position in intron 4 and exon 4. This deletion was previously found to underlie a flowering time QTL in a DH population deriving from L58 and R-o-18 (Zhang et al., 2015). We moreover observed an InDel in *BrFT2 (Bra.FT.A07)* in R-o-18, which we assume is caused by a larger structural variant underlying another flowering time QTL in the same population (Zhang et al., 2015). For *B. oleracea*, the variation detected in the central flowering time regulators is expected to significantly influence flowering time and related processes, considering the large genetic variation

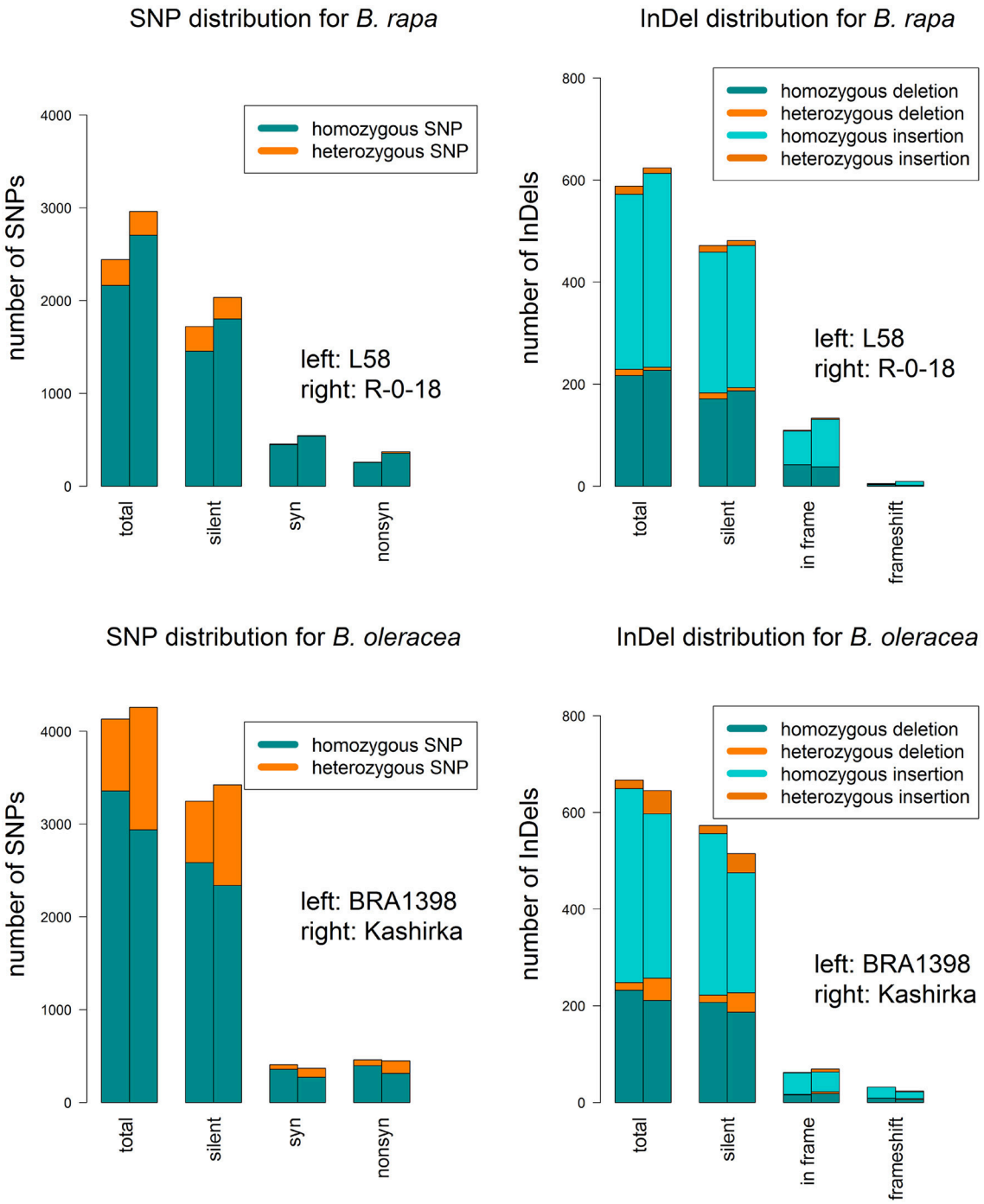


FIGURE 4 | SNP and InDel distribution for both sequenced genotypes for each *B. rapa* and *B.oleracea*. For SNPs, the barplots show the total number of SNPs (total) and their distribution on silent SNPs (silent), synonymous SNPs (syn) and non-synonymous SNPs (non-syn). For InDels, they show the total number of InDels (total) and their distribution on silent InDels (silent), in-frame variants (in frame) and frameshift InDels (frameshift). Splice variants were not displayed, as they were rare. Homozygous and heterozygous SNPs are color-coded.

including radical SNP mutations, InDels and CNVs. QTL studies for flowering time in *B. oleracea* found central regulator copies in different populations (Okazaki et al., 2007; Razi et al., 2008; Irwin et al., 2016). Both *Bol.FLC.C03* and one *Bol.FLC.C09* copy were found to underlie flowering time variation, and *Bol.FLC.C02* variation has been found to have a large influence on heading date in purple sprouting broccoli (Irwin et al., 2016) and cauliflower due to a 1 bp InDel (Okazaki et al., 2007). Furthermore, a copy

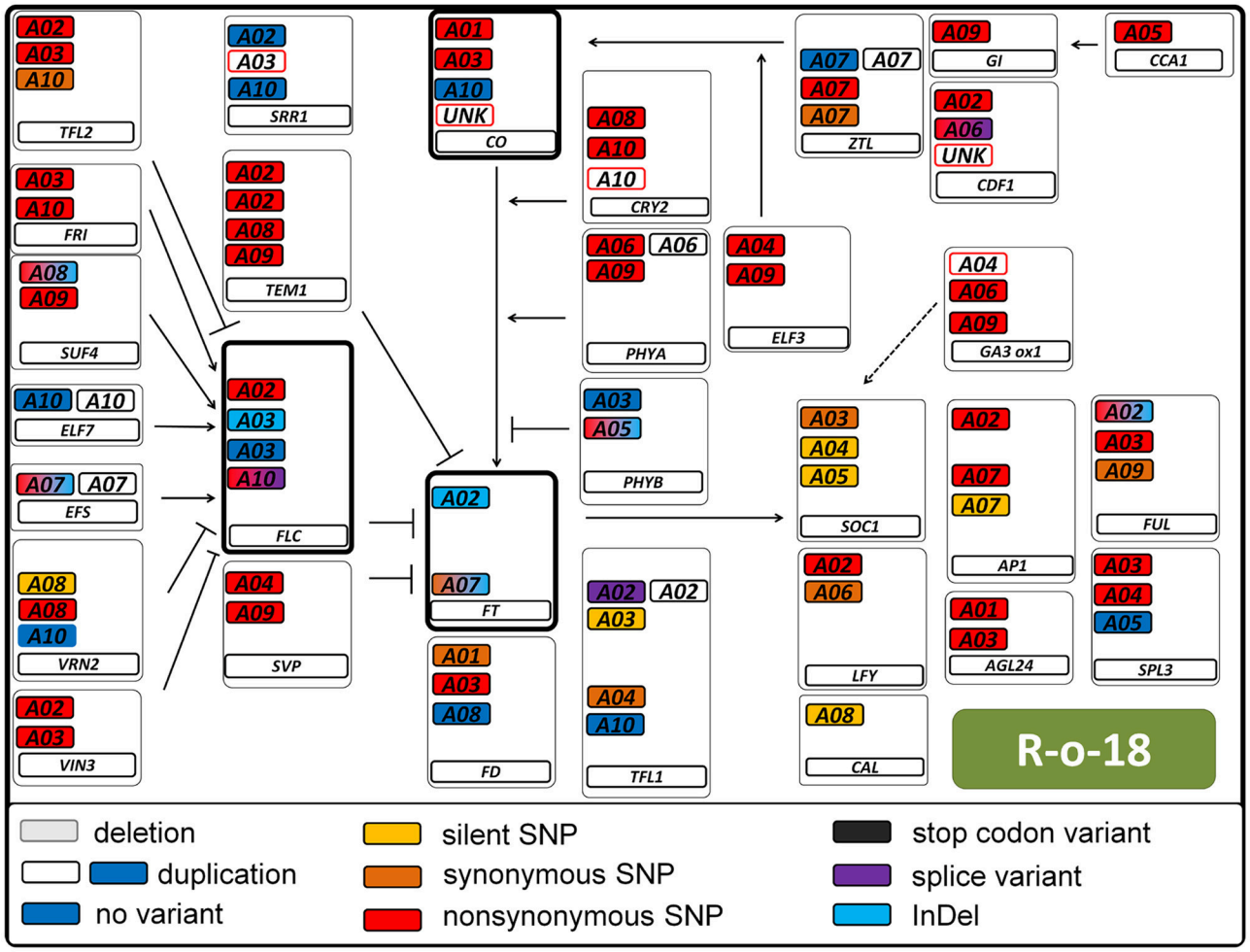


FIGURE 5 | Pathway variation for flowering time genes in the *B. rapa* genotype R-o-18. Interactions are displayed as known from *Arabidopsis thaliana*. Arrows indicate positive regulation, blunt ends indicate negative regulation. All copies of a gene are displayed in a box. The main flowering regulators *FT, FLC,* and *CO* are indicated in bold boxes. The type of variant is color coded (see legend). Duplications are indicated by a second colorless copy box. SNP colors are hierarchic, meaning for example that synonymous SNPs are not displayed if non-synonymous SNPs are present. Boxes framed in red indicate known copies from *B. napus* which were not found in *B. rapa*. Boxes framed in blue were found, but not annotated as gene.

of *Bol.CO* was found in a small-effect QTL in the same study (Okazaki et al., 2007).

Here we provide a variant framework for flowering gene variation not only for the central flowering regulators, but for the total flowering network in both vegetable species. For example, our data could help to find functional variance for QTL in *B. rapa,* not explained by *Bra.FLC* or *Bra.FT,* in a DH population derived from the reference genotype Chiifu-401 and a rapid cycling line (Li et al., 2013). The QTL in that study on A02 and A06 may correspond to candidate genes *Bra.COL.A02* and *Bra.LFY.A06*. Although we did not detect functional variation in *Bra.LFY.A06* sequences, we identified 3 non-synonymous SNPs in *Bra.COL.A02*.

Both L58 and R-o-18 are annuals, while L58 has an early-flowering phenotype (Bagheri et al., 2012). The annual behavior has been attributed to the shared splice variant in *Bra.FLC.A10* (Yuan et al., 2009), whereas the late flowering habitus of R-o-18 could be caused by the structural variation in *Bra.FT.A07*, which seems to overlay the effect of the *Bra.FLC.A02* deletion (Zhao et al., 2010; Wu et al., 2012). A further explanation could be the duplication of *Bra.CO.A01* in L58, which is likely to increase the expression of the flowering activator CO. In contrast, the late flowering Siberian Kale Kashirka is biannual, which could either be attributed to the putative duplication of *Bol.FLC.C09b (Bo9g173400)* or to the R24Q mutation in *Bol.FLC.C03 (Bo3g024250)* and the A75D mutation in *Bol.FLC.C02* (independent mapping to *Bol.FLC.C2.E9*). All other functional variants are either shared with or unique to BRA1398, which itself is annual. *Bol.FLC.C09a* (*Bo9g173370*) seems to be a pseudogene.

All the same, the DH populations studied so far all showed transgressive segregation (Zhao et al., 2010; Bagheri et al.,

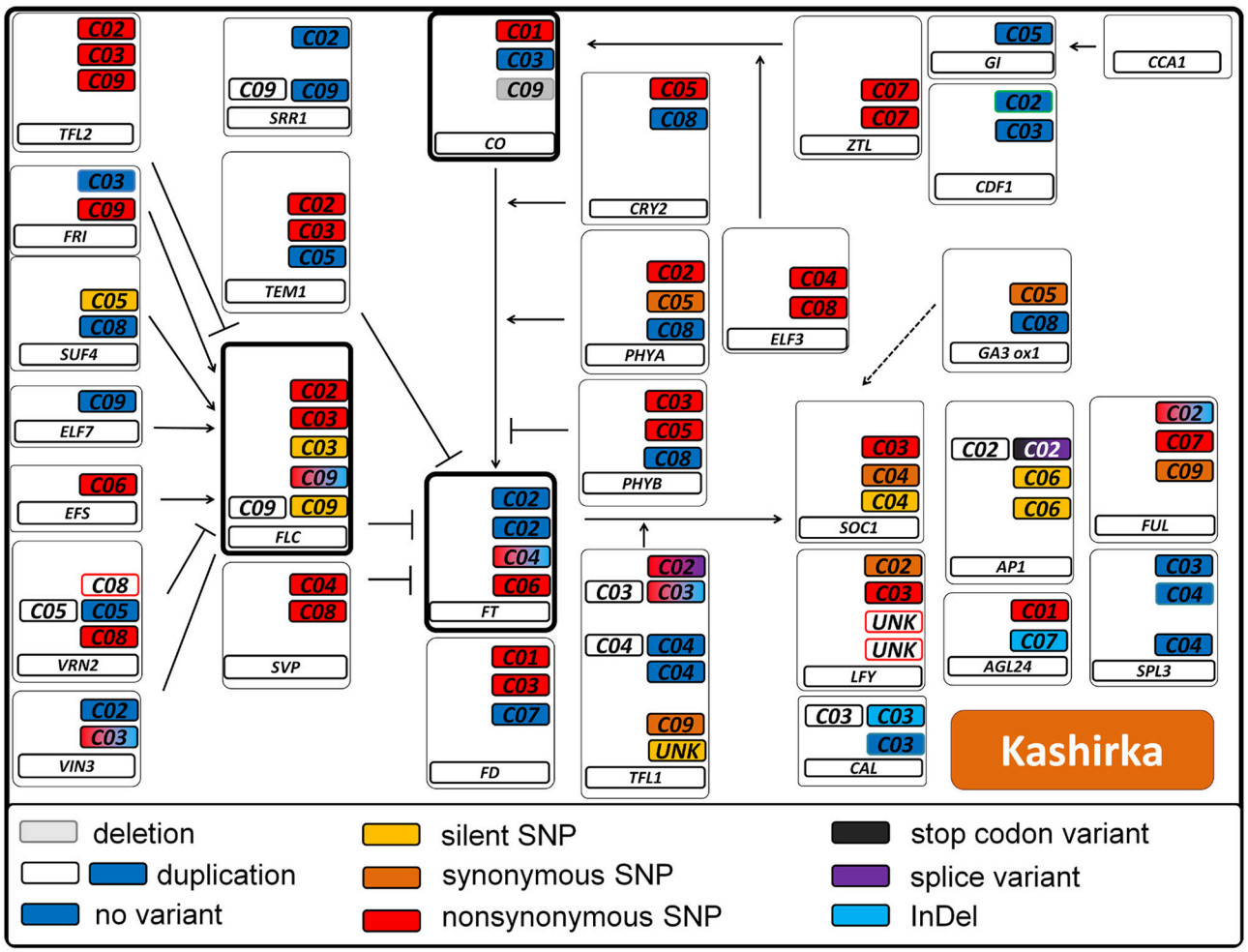


FIGURE 6 | Pathway variation for flowering time genes in the *B. oleracea* genotype Kashirka. Interactions are displayed as known from *Arabidopsis thaliana*. Arrows indicate positive regulation, blunt ends indicate negative regulation. All copies of a gene are displayed in a box. The main flowering regulators *FT, FLC,* and *CO* are indicated in bold boxes. The type of variant is color coded (see legend). Duplications are indicated by a second colorless copy box. SNP colors are hierarchic, meaning for example that synonymous SNPs are not displayed if non-synonymous SNPs are present. Boxes framed in red indicate known copies from *B. napus* which were not found in *B. oleracea*. Boxes framed in green indicate that this copy was not known from *B. napus*. Boxes framed in blue were found, but not annotated as gene.

2012; Xiao et al., 2013; Zhang et al., 2015), indicating polygenic regulation, so in order to identify and quantify the contributions from small effect genes, genome-wide association studies would have to complement the QTL studies performed so far. Our data represent a valuable resource for the development of suitable marker systems or for respective mutation studies.

Rearrangements in *Brassica napus*

Most flowering time genes in *B. napus* were found to be collinear and syntenic with their orthologues in both sequenced accessions of the progenitors *B. rapa* and *B. oleracea*, in accordance to previous findings that the donor subgenomes remain basically unaltered, although local rearrangements took place (Rana et al., 2004; Parkin, 2005; Bancroft et al., 2011; Chalhoub et al., 2014). Although, the number of sequenced accessions was limited and only covered a small part of the intraspecific variation (two subspecies for *B. rapa* and one subspecies for *B. oleracea*), we do all the same believe that they give important insights into Brassica genomics. Our sequence data indicate that some copies obviously were lost from the *B. napus* genome after polyploidisation (for example, *Bol.CDF1.C02*), while others were gained by duplication (for example, *Bna.SRR1.A03*) or rearranged to another chromosome (*Bol.TFL1.C04* to *Bna.TFL1.C03*). For two gene copies among the set (*BnaCnng78500D* (*Bna.LFY.Cnn*), *BnaA02g16710D* (*Bna.ZTL.A02*), we observed a stable exchange between the subgenomes A and C in *Brassica napus*. Using resynthesized *B. napus* as a model for polyploidization, such rearrangements were observed frequently (Gaeta et al., 2007; Szadkowski et al., 2011; Schmutzer et al., 2015), and were found to occur mostly in the first meiotic cycles after hybridization (Gaeta et al., 2007). Many times, those rearrangements in

resynthesized *B. napus* enclosed larger parts of a chromosome (Samans et al., 2017), whereas in our case, no consecutive patterns of fixed deletions or duplications were found, indicating small scale changes. As was found in Szadkowski et al. (2011), interspecific hybridization via unification of unreduced gametes causes more frequent, but smaller translocations than somatic doubling of allohaploids in *B. napus*. While the latter is mostly used for experimental hybrids, the first is more likely to occur under natural conditions. Small and stable homoeologous exchanges are therefore widespread in the *B. napus* genome and played a major role in *B. napus* speciation (Szadkowski et al., 2011; Chalhoub et al., 2014; Samans et al., 2017). Rearrangements (apart from CNVs) can change the regulatory context of a gene, change its mutation frequency and therefore contribute to speciation (Faria and Navarro, 2010). The occurrence of new pseudogenes both in *B. oleracea* and *B. napus* may be an indicator of beginning (*B. napus*) and ongoing (*B. oleracea*) diploidization after the interspecific hybridization. Pseudogenization and gene loss are general principles of genome evolution after whole genome duplication events (Sankoff et al., 2010). We expect that the total variation in each species will reveal even more such rearrangements. The sequence capture bait design used in the present study is therefore a valuable resource for further assessment of intra-and interspecific variation in *Brassica* flowering time genes.

CONCLUSIONS

Flowering time control is of major importance in crop adaptation. Knowledge about flowering time genes is crucial for improving important *Brassica* vegetable crops. Our study provides sequence variation data for all orthologous copies of 35 flowering-time regulatory genes in two accessions each of *B. rapa* and *B. oleracea*, respectively. The data confirm earlier findings on variation in central flowering time regulators, but also provide comprehensive novel data spanning numerous other genes involved in the flowering network. Rearrangement patterns compared to the allotetraploid *B. napus* revealed only small and local changes, implicating that allopolyploidisation in *B. napus* occured via unreduced gametes with small-scale homoeologous exchanges.

AUTHOR CONTRIBUTIONS

SS performed DNA extraction, bait development and data analysis. BH, DK, and RR performed library preparation and sequencing. SS and RS wrote the manuscript. All authors have read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

This work was financed by grant SN 14/13-1 from the German Research Foundation (DFG) within the priority program "Flowering Time Control: From Natural Variation to Crop Improvement."

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2017.01742/full#supplementary-material

Figure S1 | Pathway variation for flowering time genes in the *B. rapa* genotype L58. Interactions are displayed as known from *Arabidopsis thaliana*. Arrows indicate positive regulation, blunt ends indicate negative regulation. All copies of a gene are displayed in a box. The main flowering regulators FT, FLC, and CO are indicated in bold boxes. The type of variant is color coded (see legend). Duplications are indicated by a second colorless copy box. SNP colors are hierarchic, meaning for example that synonymous SNPs are not displayed if non-synonymous SNPs are present. Boxes framed in red indicate known copies from *B. napus* which were not found in *B. rapa*. Boxes framed in blue were found, but not annotated as gene.

Figure S2 | Pathway variation for flowering time genes in the *B. oleracea* genotype BRA1398. Interactions are displayed as known from *Arabidopsis thaliana*. Arrows indicate positive regulation, blunt ends indicate negative regulation. All copies of a gene are displayed in a box. The main flowering regulators FT, FLC, and CO are indicated in bold boxes. The type of variant is color coded (see legend). Duplications are indicated by a second colorless copy box. SNP colors are hierarchic, meaning for example that synonymous SNPs are not displayed if non-synonymous SNPs are present. Boxes framed in red indicate known copies from *B. napus* which were not found in *B. oleracea*. Boxes framed in green indicate that this copy was not known from *B. napus*. Boxes framed in blue were found, but not annotated as gene.

Table S1 | Total number of target SNPs and number of homozygous target SNPs for both sequenced species along with the number of target SNPs and homozygous target SNPs for each sequenced genotype.

Table S2 | Total number of target InDels and number of homozygous target InDels for both sequenced species along with the number of target InDels and homozygous target InDels for each sequenced genotype.

REFERENCES

Bagheri, H., El-Soda, M., van Oorschot, I., Hanhart, C., Bonnema, G., Jansen-van den Bosch, T., et al. (2012). Genetic analysis of morphological traits in a new, versatile, rapid-cycling *Brassica rapa* recombinant inbred line population. *Front. Plant Sci.* 3:183. doi: 10.3389/fpls.2012.00183

Bancroft, I., Morgan, C., Fraser, F., Higgins, J., Wells, R., Clissold, L., et al. (2011). Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. *Nat. Biotechnol.* 29, 762–766. doi: 10.1038/nbt.1926

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

Camargo, L. E. A., and Osborn, T. C. (1996). Mapping loci controlling flowering time in *Brassica oleracea*. *TAG Theor. Appl. Genet.* 92, 610–616. doi: 10.1007/BF00224565

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. doi: 10.1126/science.1253435

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Doyle, J. J.(1990). Isolation of plant DNA from fresh tissues. *Focus* 12, 13–15.

Faria, R., and Navarro, A. (2010). Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol. (Amst).* 25, 660–669. doi: 10.1016/j.tree.2010.07.008

Gaeta, R. T., Pires, J. C., Iniguez-Luy, F., Leon, E., and Osborn, T. C. (2007). Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* 19, 3403–3417. doi: 10.1105/tpc.107.054346

Irwin, J. A., Soumpourou, E., Lister, C., Ligthart, J.-D., Kennedy, S., and Dean, C. (2016). Nucleotide polymorphism affecting FLC expression underpins heading date variation in horticultural brassicas. *Plant J.* 87, 597–605. doi: 10.1111/tpj.13221

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* 9, 357–359. doi: 10.1038/nmeth.1923

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

Li, R., Yu, C., Li, Y., Lam, T.-W., Yiu, S.-M., Kristiansen, K., et al. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25, 1966–1967. doi: 10.1093/bioinformatics/btp336

Li, X., Ramchiary, N., Dhandapani, V., Choi, S. R., Hur, Y., Nou, I.-S., et al. (2013). Quantitative trait loci mapping in *Brassica rapa* revealed the structural and functional conservation of genetic loci governing morphological and yield component Traits in the A, B, and C subgenomes of Brassica species. *DNA Res.* 20, 1–16. doi: 10.1093/dnares/dss029

Lin, S.-I. (2005). Differential regulation of flowering LOCUS C expression by vernalization in cabbage and Arabidopsis. *Plant Physiol.* 137, 1037–1048. doi: 10.1104/pp.104.058974

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Okazaki, K., Sakamoto, K., Kikuchi, R., Saito, A., Togashi, E., Kuginuki, Y., et al. (2007). Mapping and characterization of FLC homologs and QTL analysis of flowering time in *Brassica oleracea*. TAG. Theoretical and applied genetics. *Theoretische und angewandte Genetik* 114, 595–608. doi: 10.1007/s00122-006-0460-6

Parkin, I. A. P. (2005). Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171, 2765–2781. doi: 10.1534/genetics.105.042093

Parkin, I. A., Koh, C., Tang, H., Robinson, S. J., Kagale, S., Clarke, W. E, et al. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biol.* 15:R77. doi: 10.1186/gb-2014-15-6-r77

Paterson, A. H., Lan, T.-H., Amasino, R., Osborn, T. C., and Quiros, Carlos (2001). Brassica genomics: a complement to, and early beneficiary of, the Arabidopsis sequence. *Genome Biol*. 2, reviews1011.1–reviews1011.4.

Pires, J. C., Zhao, J. W., Schranz, M. E., Leon, E. J., Quijada, P. A., Lukens, L. N., et al. (2004). Flowering time divergence and genomic rearrangements in resynthesized *Brassica polyploids* (Brassicaceae). *Biol. J. Linn. Soc.* 82, 675–688. doi: 10.1111/j.1095-8312.2004.00350.x

Quinlan, A. R. (2014). BEDTools: the Swiss-army tool for Genome feature analysis. *Curr. Prot. Bioinf.* 47, 1–34. doi: 10.1002/0471250953.bi1112s47

Rana, D., Boogaart, T., O'Neill, C. M., Hynes, L., Bent, E., Macpherson, L., et al. (2004). Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. *Plant J.* 40, 725–733. doi: 10.1111/j.1365-313X.2004.02244.x

Razi, H., Howell, E. C., Newbury, H. J., and Kearsey, M. J. (2008). Does sequence polymorphism of *FLC* paralogues underlie flowering time QTL in *Brassica oleracea*? *Theor. Appl. Genet.* 116, 179–192. doi: 10.1007/s00122-007-0657-3

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754

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

Sankoff, D., Zheng, C., and Zhu, Q. (2010). The collapse of gene complement following whole genome duplication. *BMC Genomics* 11:313. doi: 10.1186/1471-2164-11-313

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

Schiessl, S., Hüttel, B., Kuehn, D., Reinhardt, R., and Snowdon, R. J. (2017b). Post-polyploidisation morphotype diversification associates with gene copy number variation. *Sci. Rep.* 7:41845. doi: 10.1038/srep 41845

Schiessl, S., Samans, B., Hüttel, B., Reinhardt, R., and Snowdon, R. J. (2014). Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.* 5:404. doi: 10.3389/fpls.2014.00404

Schmutzer, T., Samans, B., Dyrszka, E., Ulpinnis, C., Weise, S., Stengel, D., et al. (2015). Species-wide genome sequence and nucleotide polymorphisms from the model allopolyploid plant *Brassica napus*. *Sci. Data* 2:150072. doi: 10.1038/sdata.2015.72

Schranz, M. E., Quijada, P., Sung, S.-B., Lukens, L., Amasino, R., and Osborn, T. C. (2002). Characterization and effects of the replicated flowering time gene FLC in *Brassica rapa*. *Genetics*. 162, 1457–1468.

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

Szadkowski, E., Eber, F., Huteau, V., Lode, M., Coriton, O., Jenczewski, E., et al. (2011). Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. *New Phytol.* 191, 884–894. doi: 10.1111/j.1469-8137.2011.03729.x

Vergara, I. A., Frech, C., and Chen, N. (2012). CooVar: co-occurring variant analyzer. *BMC Res. Notes*. 5:610. doi: 10.1186/1756-0500-5-615

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43, 1035–1039. doi: 10.1038/ng.919

Wu, J., Wei, K., Cheng, F., Li, S., Wang, Q., Zhao, J., et al. (2012). A naturally occurring InDel variation in BraA.FLC.b (BrFLC2) associated with flowering time variation in *Brassica rapa*. *BMC Plant Biol.* 12:151. doi: 10.1186/1471-2229-12-151

Xiao, D., Zhao, J. J., Hou, X. L., Basnet, R. K., Carpio, D. P. D., Zhang, N. W., et al. (2013). The *Brassica rapa* FLC homologue FLC2 is a key regulator of flowering time, identified through transcriptional co-expression networks. *J. Exp. Bot.* 64, 4503–4516. doi: 10.1093/jxb/ert264

Yuan, Y.-X., Wu, J., Sun, R.-F., Zhang, X.-W., Xu, D.-H., Bonnema, G., et al. (2009). A naturally occurring splicing site mutation in the *Brassica rapa* FLC1 gene is associated with variation in flowering time. *J. Exp. Bot.* 60, 1299–1308. doi: 10.1093/jxb/erp010

Zhang, X., Meng, L., Liu, B., Hu, Y., Cheng, F., Liang, J., et al. (2015). A transposon insertion in FLOWERING LOCUS T is associated with delayed flowering in *Brassica rapa*. *Plant Sci.* 241, 211–220. doi: 10.1016/j.plantsci.2015.10.007

Zhao, J., Kulkarni, V., Liu, N., Del Carpio, D. P., Bucher, J., and Bonnema, G. (2010). BrFLC2 (FLOWERING LOCUS C) as a candidate gene for a vernalization response QTL in *Brassica rapa*. *J. Exp. Bot.* 61, 1817–1825. doi: 10.1093/jxb/erq048

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

Chapter 1.4: The role of genomic structural variation in the genetic improvement of polyploid crops

Schiessl SV, Katche E, Ihien E, Chawla HS, Mason AS (2018) The role of genomic structural variation in the genetic improvement of polyploid crops. The Crop Journal 7:127–140. doi: 10.1016/j.cj.2018.07.006

Available online at www.sciencedirect.com

ScienceDirect

The role of genomic structural variation in the genetic improvement of polyploid crops

*Sarah-Veronica Schiessl, Elvis Katche, Elizabeth Ihien, Harmeet Singh Chawla, Annaliese S. Mason**

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

ARTICLE INFO

Article history:
Received 30 May 2018
Received in revised form 10 July 2018
Accepted 24 August 2018
Available online 28 September 2018

Keywords:
Presence–absence variation
Copy-number variation
Homeologous exchanges
Genome structure
Pan-genome

ABSTRACT

Many of our major crop species are polyploids, containing more than one genome or set of chromosomes. Polyploid crops present unique challenges, including difficulties in genome assembly, in discriminating between multiple gene and sequence copies, and in genetic mapping, hindering use of genomic data for genetics and breeding. Polyploid genomes may also be more prone to containing structural variation, such as loss of gene copies or sequences (presence–absence variation) and the presence of genes or sequences in multiple copies (copy-number variation). Although the two main types of genomic structural variation commonly identified are presence–absence variation and copy-number variation, we propose that homeologous exchanges constitute a third major form of genomic structural variation in polyploids. Homeologous exchanges involve the replacement of one genomic segment by a similar copy from another genome or ancestrally duplicated region, and are known to be extremely common in polyploids. Detecting all kinds of genomic structural variation is challenging, but recent advances such as optical mapping and long-read sequencing offer potential strategies to help identify structural variants even in complex polyploid genomes. All three major types of genomic structural variation (presence–absence, copy-number, and homeologous exchange) are now known to influence phenotypes in crop plants, with examples of flowering time, frost tolerance, and adaptive and agronomic traits. In this review, we summarize the challenges of genome analysis in polyploid crops, describe the various types of genomic structural variation and the genomics technologies and data that can be used to detect them, and collate information produced to date related to the impact of genomic structural variation on crop phenotypes. We highlight the importance of genomic structural variation for the future genetic improvement of polyploid crops.

Contents

* *Corresponding author.*
E-mail address: annaliese.mason@agrar.uni-giessen.de (A.S. Mason).
Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

https://doi.org/10.1016/j.cj.2018.07.006

1. Introduction

Polyploidy refers to either the duplication of a single genome (autopolyploidy) or to the combination of two or more different genomes (allopolyploidy) to make a new species [1,2]. Many important domesticated crops have been classified as allopolyploids, such as wheat (*Triticum aestivum*), tobacco (*Nicotiana tabacum*), peanut (*Arachis hypogaea*), and cotton (*Gossypium hirsutum*) [1]. Another popular example is rapeseed (allotetraploid *Brassica napus*, 2*n* = 4x = 38, AACC) which was formed by hybridization between *B. rapa* (2*n* = 2x = 20) and *B. oleracea* (2*n* = 2x = 18) [3]. By contrast, autopolyploids, such as seedless watermelon (*Citrullus lanatus*), banana (*Musa acuminata*), potato (*Solanum tuberosum*), and alfalfa (*Medicago sativa*) [1], arise within a single species by genome doubling [4].

In flowering plants in particular, polyploidy and interspecific hybridization have played a major and pervasive role in shaping plant genomes [5]. The initial merger of two genomes is often accompanied by dramatic events such as transposable-element activation (movement and replication of mobile DNA), homeologous exchanges (swapping of DNA between ancestrally related chromosomes), and DNA methylation (which may cause changes in gene expression), while the subsequent path to diploidization involves the loss, retention, or maintenance of duplicate genes with possible neo- and subfunctionalization (respectively the arising of novel gene functions and the sharing of gene functionality between duplicates) [6–8]. Together with the strongly increased sequence similarity in polyploid genomes, these processes drastically increase the likelihood of occurrence of genomic structural variants in polyploids.

Polyploids harbor great potential for crop improvement. The presence of extra gene copies and alleles can boost allelic heterosis (which confers hybrid vigor) as well as provide gene redundancy [6]. However, the multiple subgenomes and larger genome size in polyploid than in diploid crops pose some challenges to polyploid crop improvement. These challenges include decreased selection efficiency due to the contribution of multiple genes and alleles to each trait, and increased difficulties in obtaining accurate genomic and genotypic data. The latter challenge is particularly relevant for genomic structural variants, which are now known to heavily influence traits. Addressing some of these challenges requires a deep functional and structural understanding of crop genomes [9]. In this review, we present an overview of the types of genomic structural variation present in polyploids and how they can be detected, as well as the documented influence of genomic structural variation on traits in polyploids. We highlight the challenges and opportunities in exploiting the special genomic structure of polyploid crops for evolutionary and breeding research.

2. Genomic structural variation

2.1. Types of genomic structural variation in polyploids

Genomic structural variation includes all variants of the DNA sequence in which sequence blocks larger than 1 kb are transferred to a different genomic context. These transfers can have different outcomes: the sequence block to be transferred can be moved to a new locus (translocation), it can be flipped from a 5′- to -3′ to a 3′- to -5′ orientation in the same location (inversion), it can be lost (deletion), and it can be copied to a new locus (duplication). Although translocations and inversions change only the genomic context and do not affect the number of copies of a sequence present in the genome, deletions and duplications can change the copy number of the genes contained in the affected sequence block. This change can lead to individual variation in the number of copies of a gene, which is called copy-number variation (CNV). If a gene or region is simply missing in some individuals relative to others, we call these presence–absence variants (PAVs).

Polyploid genomes are, owing to their intragenomic homology, prone to so-called homeologous exchanges (HEs), in which homeologous chromosomes (ancestrally related chromosomes from different subgenomes) exchange genetic material. Particularly common in polyploids, these exchanges between homeologous chromosomes during meiosis can result in the appearance of CNVs and PAVs as well as reciprocal translocation events (Fig. 1). Although homeologous exchanges are often classed as either PAVs or CNVs, these differ from the conventional definition of PAVs and CNVS in that one part of the genome is replaced with a copy from another part of the genome, usually a homeologous region, generally conserving gene content (Fig. 1). Although transposable elements have been reported to be a major cause of structural variation [10,11], and CNV and PAV are considered to have greater effects on plant phenotype [12], HEs play a major role in generating genomic structural variation in polyploids. Investigating, capturing, and utilizing the genetic differences arising from this variation will promote the genetic improvement of polyploid crops. In the following sections we will briefly discuss these three main types of structural variation (PAVs, CNVs, and HEs), focusing on the context of polyploid crops.

2.2. Copy-number variation

Copy-number variation refers to the presence of DNA sequences (usually larger than 1 kb) in copies whose number varies between individuals or populations of the same species. Smaller elements are known as insertion/deletions [13–15]. In humans, besides the known association with sporadic and Mendelian diseases, CNV has also been associated with complex traits in humans such as autism, susceptibility to HIV, and schizophrenia. However, not only do CNVs play a role in disease and susceptibility to disease, but CNVs may also result in the emergence of advantageous traits, and thus be subject to evolutionary pressures such as selection and drift [16,17]. Several mechanisms have been proposed to explain how CNV arises, including non-allelic homologous recombination (NAHR) and non-homologous end joining (NHEJ), which are recombination-based mechanisms, and retrotransposition, which is the activation and insertion of retrotransposons. A novel replication-based mechanism known as fork stalling and template switching (FoSTeS) has been proposed to account for complex rearrangements that cannot be explained by the above mechanisms [16,18]. NAHR occurs between DNA segments of high similarity that are not alleles or homologous sequences, and usually involves low-copy-number repeats (LCR), which are DNA segments larger than 1 kb probably generated during genome duplication events [19], a major feature of polyploids.

CNV has long been known to contribute to phenotypic diversity in humans. More recently, evidence from an increasing number of studies has shown that CNV is prevalent and plays an important role in phenotypic diversity in plants. Studies in different plant species such as maize (*Zea mays*) [20], *Arabidopsis thaliana* [21], rice (*Oryza sativa*) [22], rapeseed [23,24], and wheat [25] all testify to the prevalence of CNV in plants. Interestingly, CNV in plants is generally calculated differently from that in animals. In animals, copy number is calculated as the number of copies per haploid genome, whereas in plants copy number generally refers to the number of copies per diploid genome [18,25,26]. Using

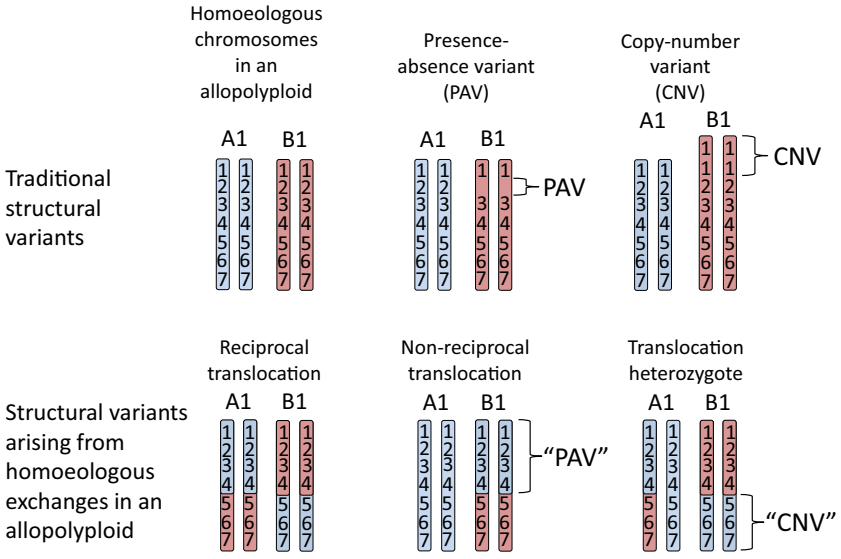


Fig. 1 – Examples of genomic structural variation that can occur in polyploids. Two pairs of homeologous chromosomes A1 and B1 with identical gene order (numbers 1–7) are presented as an example, showing presence–absence and copy-number variation as traditionally defined (respectively, loss and duplication of genes) and some of the variants that may arise from homeologous exchanges: a reciprocal translocation, which is a form of chromosome rearrangement without loss or multiplication of sequences; a non-reciprocal translocation, resulting in a "PAV-like" region with the absence of the B1 homeolog and duplication of the A1 (technically also a CNV, but these are more difficult to detect); and a translocation heterozygote such as may arise by hybridization between an individual with a fixed reciprocal translocation event and an individual without this translocation event, resulting in a 3:1 ratio of A1:B1 chromosome segments over a "CNV-like" region.

comparative genome hybridization, Springer et al. [20] identified 400 CNVs between the maize inbred lines Mo17 and B73. These CNVs were distributed across all maize chromosomes, although several conserved regions (located mostly around the centromeres) showed few or no CNVs. By sequencing 80 *Arabidopsis thaliana* accessions collected from diverse environments, Cao et al. [21] discovered 1029 CNVs, some of which overlapped with gene coding regions and thus might have an effect on phenotype. Copy-number variation was found to be particularly prevalent in polyploids and polyploid crops, such as allopolyploid wheat [25,27–29] and autotetraploid potato [13], despite the added challenge of discriminating between the multiple gene copies already present as a result of whole-genome duplication and hybridization events. Several authors have also reported CNVs affecting important adaptive and agronomic characteristics such as grain yield, frost tolerance, and flowering time [30–33]; a comprehensive overview of the effect of structural variation on phenotypic traits in polyploids appears in Section 4 of this review.

2.3. Presence–absence variation

Definitions of PAV vary. Whereas Ding et al. [34] define PAV as the presence or absence of genes within a genome or the presence of genes located in different genomic regions between genomes, the more common definition is the presence or absence of a gene in some but not all individuals of a species [20,35]. PAV has also been considered as an extreme form of CNV [14]. With the cost of genome sequencing currently decreasing, the ability to sequence many genomes simultaneously is on the rise, permitting alignment of the genomes of many individuals of the same species for comparison. One common outcome of this comparison is the identification of presence–absence variation (PAV). The extent to which two genomes can vary in terms of PAV has been demonstrated in maize, in which a comparison was made between inbred lines B73 and Mo17. On average, only 50% of sequences were shared in common between the two lines, while 25% of sequences in homologous locations were present in one inbred line but were absent in the other [36]. This result, coupled with results from other plant species, prompted an extension of the "pan-genome" concept (originally proposed in bacteria) to plants. The pan-genome is composed of the "core" genome (genes or genomic regions present in all individuals of a species) and the "dispensable" genome (present in some individuals of a species) [10,12]. Following the introduction of this concept to plants, the pan-genomes of plant species such as rice, barley, maize, and soybean have been analyzed, and the dispensable genome fraction has been shown to play an important role in evolution as well as in the complex interplay between plant species and the environment [18,37]. As a result, some authors have asked whether the dispensable genome fraction really is dispensable [10,12], or should perhaps instead be thought of as another form of adaptive variation within species.

Polyploidization and subsequent diploidization processes in plants are accompanied by subgenome fractionation, gene loss, and transposable element activation [38], processes that can increase the frequency of presence–absence variation. Despite the inherent difficulty of evaluating PAV in polyploids, several studies have made progress in this direction. Montenegro et al. [35] produced a wheat pan-genome by sequencing 18 elite cultivars and comparing them to an elite spring cultivar, and found that each cultivar had an average of 128,656 genes, with 64.3% of genes shared by all 19 cultivars. The total pan-genome content was 140,500 ± 102 genes, with 39 unique genes per individual. Sequencing the genome of the autotetraploid potato *Solanum tuberosum* ($2n = 4x = 48$) and comparing it with a heterozygous diploid genome revealed PAV in 275 genes, with 246 genes specific to the diploid [39]. The assembled pan-genome of *Brassica napus* from 53 synthetic and non-synthetic lines revealed that 38% of genes showed PAV, many of which PAVs were putatively associated with important agronomic traits such as flowering time, disease resistance, and glucosinolate content [23]. These examples highlight the importance of PAV and its potential for the genetic improvement of polyploid crops.

2.4. Homeologous exchanges

In allopolyploids, homeologous chromosomes come together in a single genome. Disomic inheritance, which is the result of strict pairing between homologous chromosomes, is sometimes enforced by pairing regulators, such as *Ph1* in allopolyploid wheat [40]. However, this process can occasionally fail even in stable allopolyploids, such that homeologous chromosomes pair and exchange genetic information [40–43], undergoing HEs (Fig. 1). Genomic variation resulting from HEs has been reported in rapeseed [44], wheat [18,45], sunflower (*Helianthus annuus*) [46], and *Tragopogon* [47,48]. HEs have also been demonstrated to affect phenotypic traits. Although Udall et al. [49] observed no marked effect of HEs in four rapeseed mapping populations, Osborn et al. [41] reported that a homeologous non-reciprocal translocation between chromosomes A7 and C6 in a rapeseed mapping population had a significant effect on seed yield. Other studies have linked HEs to other important traits such as seed quality, flowering time and fertility [24,50]. HEs have also been reported to be the major cause of gene PAV in *B. napus* amphiploids [23,51]. Hurgobin et al. [23] assembled the pan-genome of *B. napus* and reported two types of PAV: non-HE-related PAV and HE-related PAV, the latter referring to the loss of genes by replacement with their corresponding genomic segments from homeologous regions. Of the 53 accessions used to assemble the pan-genome, 30 showed HE-related PAVs, and functional annotation of these HE-related PAVs pointed to their involvement in stress, defense, and auxin pathways [23]. Lloyd et al. [51] also validated the effects on gene expression of 21 HEs between *B. napus* accessions, demonstrating major effects of some of these HEs on homeologous gene pairs. As HEs have been reported to have both adaptive and agronomic importance, cataloguing these genomic changes could play an important role in the breeding of allopolyploid crops.

3. Techniques for uncovering genomic structural variation in polyploid crops

3.1. Physical and genetic maps

Polyploid genomes are larger and more complex than their related diploid genomes. Larger genomes are more expensive to sequence, and polyploid genomes usually require more

bioinformatics expertise than diploid genomes [52]. The assembly of complex and polyploid genomes is still quite challenging, as next-generation sequencing relies on assembling short sequences that are usually much smaller than the size of genomic rearrangements [53], making it difficult to identify these events. To date, numerous genome misassemblies in polyploids have been found to result from structural variation, such as translocations and inversions between the A and D subgenomes in tetraploid cotton [53]. Chromosome rearrangements can also interfere with the construction of genetic maps, mapping of quantitative trait loci (QTL), and marker-assisted selection [54,55]. These rearrangements may also affect the accurate positioning of sequences in polyploid genomes when sequences are aligned to the reference genome of their diploid progenitors [56], complicating sequence-based genotyping approaches. However, the use of genetic mapping populations has shown great potential for addressing sequence-assembly problems in polyploids [57], and genetic and physical maps can be integrated for QTL mapping [58] or even combined to identify the effect of homeologous exchanges on phenotypes [50]. Building on these strategies, new technological developments also show promise in facilitating the identification of structural variants in complex polyploid genomes for use in crop improvement.

3.2. Discriminating between homeologous loci

A major challenge in polyploid crop improvement is in discriminating between homeologous alleles; that is, alleles present at homeologous loci (in different genomic locations), rather than homologous loci (alleles present at the same locus on two homologous chromosomes). Alleles from different homeologous genomic locations are difficult to discriminate from alleles at a single homologous locus, leading to false identification of marker (e.g. single nucleotide polymorphisms (SNP)) polymorphism [59–61]. This confusion also creates difficulty in the development of homeolog-specific and allele-specific markers, complicating the design of primers to amplify specific target regions and not corresponding homeologous regions [54]. In allopolyploids such as peanut in which the two subgenomes are highly similar (with 96% median sequence identity), distinguishing homeologous from allelic SNPs is complicated [62]. Thus, increased genetic similarity between subgenomes within polyploid species exacerbates these challenges. Consequently, the rate of SNP marker development and its application in molecular breeding is slowed in polyploid crops [63]. Polyploid crops often require higher numbers of markers than diploids, ultimately increasing cost [64]. Amplification of homeologous alleles is also an obstacle to the use of SNP arrays for studying structural variation in polyploid crops. SNP arrays can be a valuable tool for detecting structural variation: in general, they can detect PAVs in the form of segregating marker “fails” (failure to amplify an allele by multiple markers physically located contiguously on a chromosome) in a population according to the expected allele ratios [44,65,66]. However, when they are used with polyploid genomes, short oligonucleotide probes anchored onto these arrays often bind to closely related parts of the subgenomes, leading to false SNP calling [66]. Other problems caused by homeologous sequences in polyploid crop improvement include difficulty in genome-wide quantification of homeologous gene expression due to high sequence similarity between homeologous gene pairs, and the detection of more minor than major QTL in QTL mapping of polyploids [67–69]. Despite these challenges, approaches such as SNP array genotyping of doubled-haploid [70] and testcross [71] mapping populations have been highly successful in identifying chromosomal structural rearrangements and the effects of these rearrangements on phenotype [50] in complex polyploids.

With the advent of third-generation genomic technologies (first generation: sequencing of single short read sequences; e.g. Sanger sequencing; second generation: high-throughput multiplex sequencing of short (<150 bp) reads; e.g. Illumina sequencing (https://www.illumina.com/)), it is also now possible to detect structural variants in polyploid genomes with great precision, assisting in the discrimination of homeologous alleles. Unlike their predecessors, third-generation technologies rely mainly on capturing long-range genomic information, and can be broadly classified into two categories: mapping- and sequencing-based [72]. Third-generation mapping technologies such as Bionano Genomics optical mapping (https://bionanogenomics.com/) provide long-range genome structure information in the form of ordered genomic markers (restriction or marker sites) without sequencing every single nucleotide. However, third-generation sequencing technologies such as Pacific Biosciences (PacBio) Single Molecule Real Time (SMRT) sequencing (https://www.pacb.com/) and the Oxford Nanopore Technologies (https://nanoporetech.com/) sequencing platform provide actual base-pair information for ultra-long DNA molecules.

3.3. Third-generation mapping technologies

Bionano Genomics optical mapping using nano-channel arrays relies on capturing long-range genomic information in the form of restriction sites for the detection of structural variants. Introduced in 2010, it 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 can then be aligned to an in silico-generated optical map of a reference assembly. Insertions and deletions can be detected if the analyzed genotypes have additional or missing restriction sites compared to a reference assembly, although detection may be confounded by the presence of mutations in the restriction enzyme binding sites. Optical mapping has been used in genome assembly and structural-variation detection approaches in many plant species, such as wheat [73], maize [74], *Arabidopsis* [75,76], and clover (*Trifolium subterraneum* L.) [77]. One of the key factors distinguishing this approach from other technologies is that the DNA molecules are not broken into small fragments during the entire process, thus enabling the capture of long-range genomic information stretching up to several hundred kilobases. However, this approach relies, for structural-variation detection, on the availability of a high-quality reference assembly, which is often not yet available for non-model crop species. Furthermore, lack of actual nucleotide information makes it computationally challenging to isolate actual structural variant calls from the noise generated during imaging of the DNA fragments.

3.4. Third-generation sequencing technologies

Two major third generation sequencing technologies with excellent potential for detecting structural variation in complex polyploid genomes are PacBio SMRT technology and Oxford Nanopore sequencing. Unlike Bionano Optical mapping, both techniques generate actual nucleotide sequences rather than only restriction-site information. PacBio SMRT works on the principle of sequencing-by-synthesis. A single-stranded circularized DNA template is fed into a sequencing well. As the DNA polymerase synthesizes the complimentary strand base by base, a distinct fluorescent signal is generated for each base. In its current form, PacBio reads range from 2 kb to 100 kb in length. Ultra-long reads can span insertions or deletions of several thousand base pairs, thereby enabling the precise detection and mapping of variation breakpoints. A very recent example underlining the efficacy of PacBio sequencing in assembling complex polyploid genomes was the whole-genome assembly of bread wheat [78]. However, one of the possible reasons preventing more frequent use of this technology to date for detection of structural variation is the currently prohibitive financial costs associated with it. For example, a de novo assembly of a single rapeseed (*Brassica napus*, 2*n* = 4x, ~1130 Mb) genome with 80× coverage would currently cost approximately US$25,000 with a service provider (excluding the bioinformatics analysis) on the Sequel System, whereas the system itself costs US$350,000.

Oxford Nanopore sequencing provides a relatively cheap alternative to PacBio SMRT technology for detection of structural variants in polyploid genomes. The largest difference separating this technology from others is that the cost of the sequencer itself has been reduced to practically zero. This makes it easy for smaller labs to access this technology. One of the very popular sequencing platforms from the Oxford Nanopore technologies is known as the MinION. It is a small handheld device capable of sequencing DNA by measuring the minute disruptions in electric current as a DNA molecule traverses a nanopore. It is capable of delivering a read length similar to that of the PacBio SMRT technology. There are a few examples of this technology being applied to decipher structural variation in plant genomes, such as *Solanum pennellii*, for which Nanopore reads were used exclusively [79], and *Arabidopsis thaliana*, for which structural variants associated with plant growth were identified using a single Nanopore flow cell [75]. Although promising, this technology is still in its developmental phase, and high error rates and erratic data yields from Nanopore flow cells continue to be an issue. In fact, high error rates (~15%; compared to <0.5% for Illumina short reads [80]) are a major bottleneck for both the PacBio SMRT and Oxford Nanopore sequencing technologies [81]. However, thanks to the random distribution of errors across the entire length of a read, the error problem can be overcome by increasing the depth of sequencing, albeit making the entire process more expensive.

3.5. A combined strategy for the detection of structural variation in polyploids

As discussed, every individual technology has its own limitations. Accordingly, for complex genomes it would be advantageous to adopt a hybrid approach using a combination of different technologies. A very commonly used strategy for detection of structural variation in large genomes is to pursue low coverage with longer reads using either PacBio or Nanopore sequencing and to combine these data with a high number of Illumina short reads. Second-generation sequencing technologies such as Illumina together with classical genetic mapping can also be used to detect long-range genomic rearrangements. For example, using a combination of short Illumina reads with SNP data from a segregating nested association mapping (NAM) population in *Brassica napus* revealed genomic deletions associated with disease resistance [65].

As mentioned previously, pan-genomics is another popular choice for the detection of structural variation, in particular PAV. Many studies have shown the power of pan-genomics in detecting PAV, for example in wheat [35], rapeseed [23], cabbage (*Brassica oleracea*) [82], and rice [83]. However, it is important to note that the amount of information a pan-genome can provide about a species is highly dependent on the number and diversity of the individuals sequenced to create the pan-genome, and sequencing a large number of individuals inevitably incurs high costs. Declining costs combined with increasing throughput of next-generation sequencing technologies during the last decade has already led to the availability of many crop plant reference genomes: for example rice [84], maize [85], sorghum (*Sorghum bicolor*) [86], cucumber (*Cucumis sativus*) [87], soybean (*Glycine* max) [88], potato (*Solanum tuberosum*) [89], barley (*Hordeum vulgare*) [90], cotton (*Gossypium hirsutum*) [91], chickpea (*Cicer arietinum*) [92], rapeseed (*Brassica napus*) [3], bread wheat [93], common bean (*Phaseolus vulgaris*) [94], and pearl millet (*Pennisetum glaucum*) [95]. However, creating multiple reference assemblies for a single species remains challenging, owing to financial and technological constraints. In addition to large overhead expenses in creating multiple assemblies, a majority of these sequencing technologies can read only a short stretch of DNA per read. This limitation might not be a problem in dealing with diploid genomes, but it presents an inherent problem for identifying genome structural variation in polyploid crops. Allopolyploid genomes usually comprise two or more very closely related genomes, making the alignment of short DNA reads to a reference genome extremely challenging. Sequence alignment is at the heart of any structural variation detection pipeline, and a spurious alignment may lead to the wrong biological conclusions. However, with declining sequencing costs it is just a matter of time before the term “reference genome” becomes obsolete, and thousands of genome assemblies become available for every plant species, including the complex polyploid crops.

4. The influence of structural variation on traits

4.1. The challenge of linking structural variation to phenotype

Copy-number variation can also be understood as an extreme form of sequence variation. Whereas deletions abolish gene copy function, duplications can lead to variation in expression level [13] and thereby affect gene dosage. Duplications thus

have a higher risk of affecting traits than point mutations or InDels, and should accordingly be under strong selection pressure. The adaptive value of gene duplicates is illustrated by the observation that polyploid genomes tend to lose redundant gene copies [52], whereas gene copies from adaptive pathways are retained or even duplicated. For example, in the mustard family, genes from the glucosinolate pathway were highly retained after whole-genome duplications, with a retention rate of 95% compared to 45% across all genes [96]. Despite the strong phenotypic effects expected from CNV, findings linking genomic variation with phenotypes are scarce. This scarcity is due partly to the comparatively large effort needed to confirm CNVs in plant genomes. CNVs can be detected via microscopy-based approaches, e.g., fluorescent in situ hybridization (FISH); quantitative PCR (qPCR); probe-hybridization-based assays; SNP array-based methods; or next-generation sequencing. Statistically linking CNVs to phenotypes requires analysis of medium-to-large-sized populations, and these must undergo CNV detection as well as phenotyping, which can be expensive in cost and time. Moreover, some of these methods have specific drawbacks. For example, PCR-based approaches are not robust against sequence variation within primer binding sites, SNP arrays cannot easily detect duplications, and NGS approaches depend on the quality of the reference genome. These difficulties partly explain the lack of data. Another reason may be that both scientists and breeders underestimate the extent of structural variation in polyploid crops and focus rather on classical sequence variants. Recent findings of crop genomic structural variation, however, show that the assumption of a stable genome is questionable. This genomic instability is also shown by several pan-genomic approaches. For example, sequencing 10 different accessions of cabbage, a diploid with multiple polyploid events in its lineage (mesohexaploid), revealed that 18.7% of the gene copies were affected by PAV in at least one of the accessions [82]. Another pan-genome study in the related recent allotetraploid rapeseed, using 53 diverse accessions including resynthesized lines, found that 38.0% of the genes were affected by PAV [23], similar to the value of 35.7% obtained from 18 hexaploid wheat cultivars [35]. These figures illustrate the importance of structural variation in polyploid crops and the pressing need to investigate the influence of structural variants on adaptive and agronomic traits. Although transposable elements are the major source of structural variation in diploids, polyploids present even more possibilities of genomic rearrangement. Homeologous exchanges in meiosis can lead to deletions and duplications in homeologous chromosomes, depending on the degree of sequence similarity and distance from the centromeres [97]. A summary of research studies that have confirmed the phenotypic influence of CNVs in polyploid plants is presented in Table 1.

4.2. The adaptive value of structural variation

The high degree of structural variation observed within species is indicative of its adaptive value. CNVs are raw material for evolutionary adaptation and have been found to underlie several adaptive traits in crops, among them flowering time [27], cold tolerance [103], and boron tolerance [104]. They also seem to be an important mechanism for the development of herbicide resistance in weeds [105], demonstrating the power of structural variation in facilitating plant evolution even on very short time scales. Gaines et al. [105] identified the acquisition of glyphosate resistance in Palmer amaranth, a weedy species in southeast USA, by repeated duplication of the gene *5-enolpyruvylshikimate-3-phosphate synthase* (*EPSPS*). The enzyme EPSPS is the molecular target of glyphosate, and resistant populations show a 40–100 fold increase in *EPSPS* copy number along with increased protein expression and enzymatic activity [105]. Sequencing of these amplified regions revealed that they were in close vicinity to miniature inverted-repeat transposable elements (MITEs), suggesting that these MITEs have played a mechanistic role in gene amplification [106]. FISH experiments also confirmed that the *EPSPS* cassette indeed amplified as an extrachromosomal circular DNA (eccDNA) from one of two genomic copies of *EPSPS*, a finding that may explain how the species managed to evolve widespread resistance to glyphosate in less than five years [107].

Owing to the technical constraints discussed above, most studies linking CNV and phenotype focus on selected genes or sets of genes. Techniques such as exome capture can overcome this limitation for genome-wide analysis. However, few genome-wide studies have been performed to link structural variants to adaptive traits genome-wide in polyploids, owing to the high costs involved. In one example, 537 diverse accessions of tetraploid and octoploid switchgrass (*Panicum virgatum*) were sequenced, revealing 9979 genes affected by CNV or PAV [100]. Some of the CNVs could be assigned to specific ecotypes: for example, 62 deletions were specific to upland switchgrass, one of the major ecotypes of the species [100]. Some of the affected genes were also involved in photo-inhibition protection, indicative of a genome-wide adaptive response in either upland or lowland switchgrass [100].

4.3. Effects of structural variation on flowering time

A major component of climatic adaptation is flowering-time regulation, which synchronizes plant development with climatic conditions. In narrow-leafed lupin (*Lupinus angustifolius*), breeder selection for two major early flowering-time loci (*Ku* and *Julius*) allowed expansion of this crop into shorter-season environments: very recently, both loci were found to result from deletions in regulatory regions of a flowering-time gene (*LanFTc1*, a *FLOWERING LOCUS T* (*FT*) homolog) [108]. In wheat, an increase in the copy number of *Ppd-B1* conferred day-neutral flowering, whereas accessions with unaltered copy numbers were found to be photoperiod-sensitive [25]. Increased copy numbers of *Vrn-A1* were correlated with an increased vernalization requirement. In a large landmark study [27] using a global panel of 1110 diverse wheat cultivars, copy-number variants of these two loci were found to be involved in wheat adaptation worldwide. Based on the phylogeny of the population, the authors concluded that the observed gene duplications arose independently several times, underlining the importance and prevalence of the CNV adaptation mechanism. *Ppd-B1* duplications affected 56% of the Chinese subpopulation but only 10% of the total population, suggesting a role of

Table 1 – Studies reliably linking structural variation to trait variation in polyploid crops.

Species	Ploidy	Trait	Method of detection	Number and type of accessions	Reference
Brassica napus	4x	Flowering time	Sequence capture	280 diverse accessions	Schiessl et al. [98]
Brassica napus	4x	Seed quality	SNP array, FISH, NGS	3 different doubled haploid populations	Stein et al. [50]
Brassica napus	4x	Chlorophyll content	SNP array	203 diverse semi-winter accessions	Qian et al. [99]
Panicum virgatum	4x + 8x	Ecotype	Exome capture	537 diverse accessions	Evans et al. [100]
Solanum tuberosum	4x	Gene expression	FISH	16 diverse accessions	Iovene et al. [13]
Triticum aestivum	6x	Flowering time	*Taq*Man	9 diverse accessions, 77 doubled haploid lines	Díaz et al. [25]
Triticum aestivum	6x	Flowering time	*Taq*Man	1110 diverse accessions	Würschum et al. [27]
Triticum aestivum	6x	Frost tolerance	*Taq*Man	65 diverse winter accessions, 81 diverse spring accessions	Zhu et al. [101]
Triticum aestivum	6x	Frost tolerance	*Taq*Man	407 diverse winter accessions	Würschum et al. [102]
Triticum aestivum	6x	Chlorophyll content, grain size	PCR	169 recombinant inbred lines	Chang et al. [30]

photoperiod adaptation to different daylength regimes. In addition, copy numbers of *Vrn-A1* increased from southern to northern Europe. Both CNVs together explained between 3% and 30% of the phenotypic variance depending on environment. Other polyploid crops also show widespread CNVs in flowering-time genes. Using a targeted-sequencing approach in a population of 280 diverse accessions of rapeseed, a study [24] of gene copy-number variation showed that CNVs are highly abundant in all 35 flowering-time gene copy groups assessed. Similar results were found for the progenitor species *B. rapa* and *B. oleracea* [32]. In particular, two structural rearrangements (HE duplication–deletion events) affecting copies of *Bna.PHYA* (duplicated from A09, deleted on C8) and *Bna.FLC* (duplicated from A10, deleted on C09) predominated in swedes (*B. napus* ssp. *napobrassica*; [24]). *PHYA* is a photoreceptor involved in photoperiod-dependent regulation of flowering time, whereas *FLC* is the main regulator of vernalization in dicots [109]. The subspecies *napobrassica*, as a root vegetable form of *B. napus*, is extremely vernalization-dependent and flowers only after very long periods of cold, presumably because it has been selected for bolting resistance [24]. These observed HEs and CNVs in wheat and rapeseed were selected unconsciously by early breeders, highlighting the possible gains that could be made by targeted application of knowledge of structural genome variation in modern breeding.

4.4. *Effects of structural variation on frost tolerance*

Another trait tightly linked to climatic adaptation is frost tolerance. Biannual plants overwinter in the vegetative state and therefore have to withstand temperatures well below 0 °C. This requirement can limit introgression of genetic variance from an annual to a biannual background. In wheat, two major loci controlling this trait have been shown to be linked to CNVs: *FR1*, containing the previously named vernalization regulator *Vrn-A1*, and *FR2*, where three *C-REPEAT BINDING FACTOR* (*CBF*) gene copies, *CBF-A12*, *CBF-A14*, and *CBF-A15*, are located [101]. Copy numbers of *CBF-A12* and *CBF-A14* in 65 diverse winter and 81 diverse spring cultivars were significantly correlated with frost tolerance in winter accessions, but not in spring accessions [101]. Furthermore, the phenotypic effect of the CNV depended on the haplotype of the other locus: an increased copy number of *Vrn-A1* led to higher frost tolerance in accessions carrying the FR2-A2-T haplotype [101]. Later, Würschum et al. [102] showed that the *CBF-A14* CNV indeed accounts for 24.3% of the phenotypic variance of the trait in winter wheat. This example shows that both CNVs and sequence variants should be carefully integrated into the genetic models underlying modern breeding programs.

4.5. *Effects of structural variation on other agronomic traits*

Although several clear examples of CNVs affecting important agronomic traits, such as disease resistance [72,110] and metal ion tolerance [104,111], are known in diploids, results showing the effect of CNVs on agronomic traits in polyploids are surprisingly scarce. However, the rare examples we have indicate that more research in this field is sorely needed. In allotetraploid rapeseed, Qian et al. [99] used data from a genome-wide SNP array to identify a deletion encompassing a copy of *NON-YELLOWING 1* (*NYE1*), which is involved in chlorophyll degradation during senescence. The deletion was significantly associated with chlorophyll content at two different plant developmental stages in a population of 203 diverse Chinese semi-winter accessions [99]. Haplotype analysis revealed that seven accessions carried the deletion, all of which showed significantly increased chlorophyll content at both seedling and bolting stages [99]. Phylogenetic analysis suggested that the locus had been introgressed into *B. napus* from the progenitor species *B. rapa* [99]. Similarly, a deletion in a copy of *cytokinin oxidase* (*CKX*), a gene involved in photosynthesis regulation, reduced both chlorophyll content and grain size in a recombinant inbred line (RIL) population derived from a cross between a winter and a spring cultivar of wheat [30]. Stein et al. [50] identified several genomic rearrangements affecting seed quality and other agronomic traits in three different doubled-haploid (DH) populations of *B. napus* using a variety of different methods to validate these results: FISH, whole genome resequencing, genetic mapping using SNP array data, and sequence capture [50]. Different quantitative trait loci (QTL) for seed fiber content, number of seeds per silique, flowering time, and glucosinolate content could all be traced back to deletions or homeologous exchanges in this study [50]. For example, a QTL for seed color, lignin, and fiber content was associated with a duplicated fragment on chromosome C08, with a corresponding loss of a (homeologous) chromosome A09 fragment in the resynthesized mapping parent [50]. The 173-kb QTL interval

was found to harbor three important candidate genes for those traits. A CNV for one of these gene copies was additionally confirmed by sequence capture, indicating that the structural rearrangement indeed affected the genes within the interval [50].

5. Conclusions and perspectives

Today, we know that allelic variants of genes are far from explaining the totality of crop phenotypes. There is an increasing realization that SNPs do not represent all existing genetic variation within a species, and also an increasing appreciation of the role of genome structural variation in phenotypic expression [14]. This understanding has led to large-scale characterization of multiple genomes per species at a population level, resulting in the discovery of a large array of structural variants [10]. Changes in gene copy number affect the expression levels of genes, which may in turn alter phenotypes. Owing to their large size relative to single nucleotide polymorphisms (SNPs), structural variants most likely account for more heritable differences in phenotype than SNPs [112]. However, until recently, the complexity of genome structural variation was relatively unknown, particularly in plants. Genome structure is still more complex in polyploids, which have multiple gene copies and are often also the product of ancestral hybridization events between species. Polyploids carry all the same types of genomic structural variation as diploids, but taken to the extreme. Owing to the buffering effect of multiple genomes and hence redundant gene copies, CNV and PAV are hypothetically better tolerated in polyploids than in diploids [7]. Sequence homology shared between two different parts of the genome initiates non-homologous recombination events in all species, regardless of polyploid status [54]. However, polyploids, with their high proportion of sequence homology (homeology) between different subgenomes, are far more likely to undergo chromosome rearrangements [113]. These rearrangements can be classified into non-reciprocal and reciprocal translocation events, and then into duplications and deletions (Fig. 1), but all arise from exchanges between non-homologous (usually homeologous) chromosomes during meiosis [114]. These homeologous exchanges differ from PAV and CNV as traditionally defined in that a "dosage compensation" effect is usually achieved, such that gene content is relatively conserved [115,116]. For this reason, we propose that homeologous exchanges should be considered as a major, distinct category of genome structural variation, particularly in polyploids.

All three major types of genome structural variation (CNV, PAV, and HE) have been linked to important agronomic traits in polyploid crops (Table 1), despite the fact that genomic resources are still sparse for many of these species. The production of multiple reference genomes per species and the subsequent construction of "pan-genomes" that include all (or almost all) DNA present in all individuals within a species have been under way for only a short time, and most of our important crop species are still lacking pan-genome resources (Table 2). Fortunately, these resources are becoming increasingly available with the advent of new sequencing technologies, and with the corresponding drop in price and time required for genome sequencing. In the short term, detection of genome structural variation at lower costs can be optimized by combining multiple technologies, such as short-read sequencing with longer reads or optical mapping, or by the use of older methods such as molecular cytogenetics or genotyping arrays [50,66].

n.a., data not available/presented.

In this review, we have highlighted the importance of genomic structural variation for polyploid crop improvement. Genomic structural variation in polyploids affects major agronomic traits such as flowering time, frost tolerance, and seed quality traits, as well as adaptation to the environment (Table 1). This effect of structural variation on phenotype emphasizes the necessity of investing in pan-genomics, as emerging research has also pointed to the "dispensable" genome as playing an important role in adaptation within a species to different environmental conditions [10,12] and CNV and PAV have been specifically linked to traits inadvertently selected for by human agriculture in polyploid crops such as wheat [27] and rapeseed [98]. In future, efficient identification and functional validation of genomic structural variation facilitated by new technologies and pan-genome resources should play a major role in genetic improvement of polyploid crops.

Table 2 – List of species with assembled pan-genomes.

Species	Ploidy level	Represented diversity	% core	% dispensable	Reference
Arabidopsis thaliana	Diploid	80 accessions	n.a.	n.a.	Gao et al. [21]
Arabidopsis thaliana	Diploid	19 accessions	69.7	30.3	Contreras-Moreira et al. [37]
Brassica napus	Allotetraploid	33 varieties, 20 synthetics	62	38	Hurgobin et al. [23]
Brassica oleracea	Mesohexaploid	9 varieties, 1 wild relative	81.3	18.7	Golicz et al. [82]
Brassica rapa	Mesohexaploid	3 cultivars	91.2	8.8	Lin et al. [117]
Glycine soja	Diploid	7 accessions	80	20	Li et al. [118]
Hordeum vulgare	Diploid	16 accessions	n.a.	n.a.	Contreras-Moreira et al. [37]
Medicago truncatula	Diploid	15 accessions	58	42	Zhou et al. [119]
Oryza sativa	Diploid	1483 cultivars	n.a.	n.a.	Yao et al. [83]
Oryza sativa	Diploid	3010 cultivars	56.7	43.4	1001 Genomes Consortium [120]
Populus	Diploid	4 *P. nigra*, 2 *P. deltoides*, 1 *P. trichocharpa*	80.7	19.3	Pinosio et al. [121]
Solanum tuberosum	Diploid/autotetraploid	20 diploids, 47 tetraploid accessions	46–65	35–54	Hardigan et al. [39]
Triticum aestivum	Allohexaploid	18 cultivars	64.3	35.7	Montenegro et al. [35]
Zea mays	Paleopolyploid	14,129 inbred lines	60	40	Lu et al. [122]

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (MA6473/1-1, MA6473/2-1).

REFERENCES

[1] J.A. Udall, J.F. Wendel, Polyploidy and crop improvement, Crop Sci. 46 (2006) S3–S14.
[2] J. Ramsey, D.W. Schemske, Pathways, mechanisms, and rates of polyploid formation in flowering plants, Annu. Rev. Ecol. Syst. 29 (1998) 467–501.
[3] B. Chalhoub, F. Denoeud, S. Liu, I.A.P. Parkin, H. Tang, X. Wang, J. Chiquet, H. Belcram, C. Tong, B. Samans, M. Correa, C. Da Silva, J. Just, C. Falentin, C.S. Koh, I. Le Clainche, M. Bernard, P. Bento, B. Noel, K. Labadie, A. Alberti, M. Charles, D. Arnaud, H. Guo, C. Daviaud, S. Alamery, K. Jabbari, M. Zhao, P.P. Edger, H. Chelaifa, D. Tack, G. Lassalle, I. Mestiri, N. Schnel, M.C. Le Paslier, G. Fan, V. Renault, P.E. Bayer, A.A. Golicz, S. Manoli, T.H. Lee, V.H.D. Thi, S. Chalabi, Q. Hu, C. Fan, R. Tollenaere, Y. Lu, C. Battail, J. Shen, C.H.D. Sidebottom, A. Canaguier, A. Chauveau, A. Berard, G. Deniot, M. Guan, Z. Liu, F. Sun, Y.P. Lim, E. Lyons, C.D. Town, I. Bancroft, J. Meng, J. Ma, J.C. Pires, G.J. King, D. Brunel, R. Delourme, M. Renard, J.M. Aury, K.L. Adams, J. Batley, R.J. Snowdon, J. Tost, D. Edwards, Y. Zhou, W. Hua, A.G. Sharpe, A.H. Paterson, C. Guan, P. Wincker, Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome, Science 345 (2014) 950–953.
[4] C. Parisod, K. Alix, J. Just, M. Petit, V. Sarilar, C. Mhiri, M. Ainouche, B. Chalhoub, M.A. Grandbastien, Impact of transposable elements on the organization and function of allopolyploid genomes, New Phytol. 186 (2010) 37–45.
[5] P.S. Soltis, D.B. Marchant, Y. Van de Peer, D.E. Soltis, Polyploidy and genome evolution in plants, Curr. Opin. Genet. Dev. 35 (2015) 119–125.
[6] L. Comai, The advantages and disadvantages of being polyploid, Nat. Rev. Genet. 6 (2005) 836–846.
[7] A.R. Leitch, I.J. Leitch, Genomic plasticity and the diversity of polyploid plants, Science 320 (2008) 481–483.
[8] K.L. Adams, J.F. Wendel, Polyploidy and genome evolution in plants, Curr. Opin. Plant Biol. 8 (2005) 135–141.
[9] L. Chaney, A.R. Sharp, C.R. Evans, J.A. Udall, Genome mapping in plant comparative genomics, Trends Plant Sci. 21 (2016) 770–780.
[10] M. Morgante, E. De Paoli, S. Radovic, Transposable elements and the plant pan-genomes, Curr. Opin. Plant Biol. 10 (2007) 149–155.
[11] B. Piegu, R. Guyot, N. Picault, A. Roulin, A. Saniyal, H. Kim, K. Collura, D.S. Brar, S. Jackson, R.A. Wing, O. Panaud, Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice, Genome Res. 16 (2006) 1262–1269.
[12] F. Marroni, S. Pinosio, M. Morgante, Structural variation and genome complexity: is dispensable really dispensable? Curr. Opin. Plant Biol. 18 (2014) 31–36.
[13] M. Iovene, T. Zhang, Q.F. Lou, C.R. Buell, J.M. Jiang, Copy number variation in potato-an asexually propagated autotetraploid species, Plant J. 75 (2013) 80–89.
[14] R.K. Saxena, D. Edwards, R.K. Varshney, Structural variations in plant genomes, Brief. Funct. Genomics 13 (2014) 296–307.
[15] M. Zarrei, J.R. MacDonald, D. Merico, S.W. Scherer, A copy number variation map of the human genome, Nat. Rev. Genet. 16 (2015) 172–183.
[16] F. Zhang, W.L. Gu, M.E. Hurles, J.R. Lupski, Copy number variation in human health, disease, and evolution, Annu. Rev. Genomics Hum. Genet. 10 (2009) 451–481.
[17] D.F. Conrad, D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y.J. Zhang, J. Aerts, T.D. Andrews, C. Barnes, P. Campbell, T. Fitzgerald, M. Hu, C.H. Ihm, K. Kristiansson, D.G. MacArthur, J.R. MacDonald, I. Onyiah, A.W.C. Pang, S. Robson, K. Stirrups, A. Valsesia, K. Walter, J. Wei, C. Tyler-Smith, N.P. Carter, C. Lee, S.W. Scherer, M.E. Hurles, W.T.C. Control, Origins and functional impact of copy number variation in the human genome, Nature 464 (2010) 704–712.
[18] A. Żmieńko, A. Samelak, P. Kozłowski, M. Figlerowicz, Copy number polymorphism in plant genomes, Theor. Appl. Genet. 127 (2014) 1–18.
[19] W. Gu, F. Zhang, J.R. Lupski, Mechanisms for human genomic rearrangements, PathoGenetics 1 (2008) 4.
[20] N.M. Springer, K. Ying, Y. Fu, T.M. Ji, C.T. Yeh, Y. Jia, W. Wu, T. Richmond, J. Kitzman, H. Rosenbaum, A.L. Iniguez, W.B. Barbazuk, J.A. Jeddeloh, D. Nettleton, P.S. Schnable, Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content, PLoS Genet. 5 (2009), e1000734.
[21] J. Cao, K. Schneeberger, S. Ossowski, T. Gunther, S. Bender, J. Fitz, D. Koenig, C. Lanz, O. Stegle, C. Lippert, X. Wang, F. Ott, J. Muller, C. Alonso-Blanco, K. Borgwardt, K.J. Schmid, D. Weigel, Whole-genome sequencing of multiple *Arabidopsis thaliana* populations, Nat. Genet. 43 (2011) 956–963.
[22] P. Yu, C.H. Wang, Q. Xu, Y. Feng, X.P. Yuan, H.Y. Yu, Y.P. Wang, S.X. Tang, X.H. Wei, Detection of copy number variations in rice using array-based comparative genomic hybridization, BMC Genomics 12 (2011) 372.
[23] B. Hurgobin, A. Golicz Agnieszka, E. Bayer Philipp, K. Chan Chon-Kit, S. Tirnaz, A. Dolatabadian, V. Schiessl Sarah, B. Samans, D. Montenegro Juan, A.P. Parkin Isobel, J.C. Pires, B. Chalhoub, J. King Graham, R. Snowdon, J. Batley, D. Edwards, Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*, Plant Biotechnol. J. 16 (2018) 1265–1274.
[24] S.V. Schiessl, B. Huettel, D. Kuehn, R. Reinhardt, R.J. Snowdon, Post-polyploidisation morphotype diversification associates with gene copy number variation, Sci. Rep. 7 (2017), 41845.
[25] A. Díaz, M. Zikhali, A.S. Turner, P. Isaac, D.A. Laurie, Copy number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*), PLoS One 7 (2012), e33234.
[26] D.E. Cook, T.G. Lee, X.L. Guo, S. Melito, K. Wang, A.M. Bayless, J.P. Wang, T.J. Hughes, D.K. Willis, T.E. Clemente, B. W. Diers, J.M. Jiang, M.E. Hudson, A.F. Bent, Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean, Science 338 (2012) 1206–1209.
[27] T. Würschum, P.H.G. Boeven, S.M. Langer, C.F.H. Longin, W. L. Leiser, Multiply to conquer: copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat, BMC Genet. 16 (2015) 96.
[28] S.M. Langer, C.F.H. Longinand, T. Wurschum, Flowering time control in European winter wheat, Front. Plant Sci. 5 (2014) 537.
[29] C. Saintenac, D.Y. Jiang, S.C. Wang, E. Akhunov, Sequence-based mapping of the polyploid wheat genome, Genes Genome Genet. 3 (2013) 1105–1114.
[30] C. Chang, J. Lu, H.P. Zhang, C.X. Ma, G.L. Sun, Copy number variation of cytokinin oxidase gene *tackx4* associated with grain weight and chlorophyll content of flag leaf in common wheat, PLoS One 10 (2015), e0145970.
[31] A.N. Sieber, C.F.H. Longin, W.L. Leiser, T. Würschum, Copy number variation of *CBF-A14* at the *Fr-A2* locus determines frost tolerance in winter durum wheat, Theor. Appl. Genet. 129 (2016) 1087–1097.

[32] S.V. Schiessl, B. Huettel, D. Kuehn, R. Reinhardt, R.J. Snowdon, Flowering time gene variation in *Brassica* species shows evolutionary principles, Front. Plant Sci. 8 (2017) 1742.
[33] S. Förster, E. Schumann, M. Baumann, W.E. Weber, K. Pillen, Copy number variation of chromosome 5A and its association with Q gene expression, morphological aberrations, and agronomic performance of winter wheat cultivars, Theor. Appl. Genet. 126 (2013) 3049–3063.
[34] J. Ding, H. Araki, Q. Wang, P. Zhang, S. Yang, J.Q. Chen, D. Tian, Highly asymmetric rice genomes, BMC Genomics 8 (2007) 154.
[35] J.D. Montenegro, A.A. Golicz, P.E. Bayer, B. Hurgobin, H. Lee, C.-K.K. Chan, P. Visendi, K. Lai, J. Doležel, J. Batley, D. Edwards, The pangenome of hexaploid bread wheat, Plant J. 90 (2017) 1007–1013.
[36] S. Brunner, K. Fengler, M. Morgante, S. Tingey, A. Rafalski, Evolution of DNA sequence nonhomologies among maize inbreds, Plant Cell 17 (2005) 343–360.
[37] B. Contreras-Moreira, C.P. Cantalapiedra, M.J. Garcia-Pereira, S.P. Gordon, J.P. Vogel, E. Igartua, A.M. Casas, P. Vinuesa, Analysis of plant pan-genomes and transcriptomes with GET_HOMOLOGUES-EST, a clustering solution for sequences of the same species, Front. Plant Sci. 8 (2017) 184.
[38] M.X. Zhao, B.A. Zhang, D. Lisch, J.X. Ma, Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants, Plant Cell 29 (2017) 2974–2994.
[39] M.A. Hardigan, F.P.E. Laimbeer, L. Newton, E. Crisovan, J.P. Hamilton, B. Vaillancourt, K. Wiegert-Rininger, J.C. Wood, D. S. Douches, E.M. Farre, R.E. Veilleux, C.R. Buell, Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) E9999–E10008.
[40] S. Griffiths, R. Sharp, T.N. Foote, I. Bertin, M. Wanous, S. Reader, I. Colas, G. Moore, Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat, Nature 439 (2006) 749–752.
[41] T.C. Osborn, D.V. Butrulle, A.G. Sharpe, K.J. Pickering, I.A. Parkin, J.S. Parker, D.J. Lydiate, Detection and effects of a homeologous reciprocal transposition in *Brassica napus*, Genetics 165 (2003) 1569–1577.
[42] S.D. Nicolas, G. Le Mignon, F. Eber, O. Coriton, H. Monod, V. Clouet, V. Huteau, A. Lostanlen, R. Delourme, B. Chalhoub, C. D. Ryder, A.M. Chèvre, E. Jenczewski, Homeologous recombination plays a major role in chromosome rearrangements that occur during meiosis of *Brassica napus* haploids, Genetics 175 (2007) 487–503.
[43] R.T. Gaeta, J.C. Pires, Homoelogous recombination in allopolyploids: the polyploid ratchet, New Phytol. 186 (2010) 18–28.
[44] A.S. Mason, P. Chauhan, S. Banga, S.S. Banga, P. Salisbury, M. J. Barbetti, J. Batley, Agricultural selection and presence-absence variation in spring-type canola germplasm, Crop Pasture Sci. 69 (2018) 55–64.
[45] R. Cronn, R.L. Small, T. Haselkorn, J.F. Wendel, Cryptic repeated genomic recombination during speciation in *Gossypium gossypioides*, Evolution 57 (2003) 2475–2489.
[46] L.H. Rieseberg, S.C. Kim, R.A. Randell, K.D. Whitney, B.L. Gross, C. Lexer, K. Clay, Hybridization and the colonization of novel habitats by annual sunflowers, Genetica 129 (2007) 149–165.
[47] K.Y. Lim, D.E. Soltis, P.S. Soltis, J. Tate, R. Matyasek, H. Srubarova, A. Kovarik, J.C. Pires, Z.Y. Xiong, A.R. Leitch, Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (Asteraceae), PLoS One 3 (2008), e3353.
[48] R.J.A. Buggs, A.N. Doust, J.A. Tate, J. Koh, K. Soltis, F.A. Feltus, A.H. Paterson, P.S. Soltis, D.E. Soltis, Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids, Heredity 103 (2009) 73–81.
[49] J.A. Udall, P.A. Quijada, T.C. Osborn, Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L, Genetics 169 (2005) 967–979.
[50] A. Stein, O. Coriton, M. Rousseau-Gueutin, B. Samans, S.V. Schiessl, C. Obermeier, I.A.P. Parkin, A.M. Chevre, R.J. Snowdon, Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*, Plant Biotechnol. J. 15 (2017) 1478–1489.
[51] A. Lloyd, A. Blary, D. Charif, C. Charpentier, J. Tran, S. Balzergue, E. Delannoy, G. Rigaill, E. Jenczewski, Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop, New Phytol. 217 (2018) 367–377.
[52] S. Renny-Byfield, J.F. Wendel, Doubling down on genomes: polyploidy and crop plants, Am. J. Bot. 101 (2014) 1711–1725.
[53] S. Wang, J.D. Chen, W.P. Zhang, Y. Hu, L.J. Chang, L. Fang, Q. Wang, F.N. Lv, H.T. Wu, Z.F. Si, S.Q. Chen, C.P. Cai, X.F. Zhu, B. L. Zhou, W.Z. Guo, T.Z. Zhang, Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes, Genome Biol. 16 (2015) 108.
[54] D. Fu, A.S. Mason, M. Xiao, H. Yan, Effects of genome structure variation, homeologous genes and repetitive DNA on polyploid crop research in the age of genomics, Plant Sci. 242 (2016) 37–46.
[55] A.H. Paterson, J.E. Bowers, M.D. Burow, X. Draye, C.G. Elsik, C.-X. Chiang, C.S. Katsar, T.H. Lan, Y.-R. Lin, R. Ming, R.J. Wright, Comparative genomics of plant chromosomes, Plant Cell 12 (2000) 1523–1539.
[56] D. Edwards, J. Batley, R.J. Snowdon, Accessing complex crop genomes with next-generation sequencing, Theor. Appl. Genet. 126 (2013) 1–11.
[57] P.E. Bayer, P. Ruperao, A.S. Mason, J. Stiller, C.K.K. Chan, S. Hayashi, Y. Long, J. Meng, T. Sutton, P. Visendi, R.K. Varshney, J. Batley, D. Edwards, High-resolution skim genotyping by sequencing reveals the distribution of crossovers and gene conversions in *Cicer arietinum* and *Brassica napus*, Theor. Appl. Genet. 128 (2015) 1039–1047.
[58] Q.-H. Zhou, D.-H. Fu, A.S. Mason, Y.-J. Zeng, C.-X. Zhao, Y.-J. Huang, *In silico* integration of quantitative trait loci for seed yield and yield-related traits in *Brassica napus*, Mol. Breed. 33 (2014) 881–894.
[59] M. Imelfort, C. Duran, J. Batley, D. Edwards, Discovering genetic polymorphisms in next-generation sequencing data, Plant Biotechnol. J. 7 (2009) 312–317.
[60] R.E. Oliver, G.R. Lazo, J.D. Lutz, M.J. Rubenfield, N.A. Tinker, J. M. Anderson, N.H.W. Morehead, D. Adhikary, E.N. Jellen, P.J. Maughan, G.L.B. Guedira, S.M. Chao, A.D. Beattie, M.L. Carson, H.W. Rines, D.E. Obert, J.M. Bonman, E.W. Jackson, Model SNP development for complex genomes based on hexaploid oat using high-throughput 454 sequencing technology, BMC Genomics 12 (2011) 77.
[61] K.T. Lai, C. Duran, P.J. Berkman, M.T. Lorenc, J. Stiller, S. Manoli, M.J. Hayden, K.L. Forrest, D. Fleury, U. Baumann, M. Zander, A.S. Mason, J. Batley, D. Edwards, Single nucleotide polymorphism discovery from wheat next-generation sequence data, Plant Biotechnol. J. 10 (2012) 743–749.
[62] D.J. Bertioli, P. Ozias-Akins, Y. Chu, K.M. Dantas, S.P. Santos, E. Gouvea, P.M. Guimaraes, S.C.M. Leal-Bertioli, S.J. Knapp, M.C. Moretzsohn, The use of SNP markers for linkage mapping in diploid and tetraploid peanuts, Genes Genome Genet. 4 (2014) 89–96.
[63] J. Clevenger, C. Chavarro, S.A. Pearl, P. Ozias-Akins, S.A. Jackson, Single nucleotide polymorphism identification in polyploids: a review, example, and recommendations, Mol. Plant 8 (2015) 831–846.
[64] A. Rasheed, Y.F. Hao, X.C. Xia, A. Khan, Y.B. Xu, R.K. Varshney, Z.H. He, Crop breeding chips and genotyping

platforms: progress, challenges, and perspectives, Mol. Plant 10 (2017) 1047–1064.
[65] I. Gabur, H.S. Chawla, X. Liu, V. Kumar, S. Faure, A. von Tiedemann, C. Jestin, E. Dryzska, S. Volkmann, F. Breuer, R. Delourme, R. Snowdon, C. Obermeier, Finding invisible quantitative trait loci with missing data, Plant Biotechnol. J. (2018)https://doi.org/10.1111/pbi.12942.
[66] A.S. Mason, E.E. Higgins, R.J. Snowdon, J. Batley, A. Stein, C. Werner, I.A.P. Parkin, A user guide to the *Brassica* 60K Illumina Infinium (TM) SNP genotyping array, Theor. Appl. Genet. 130 (2017) 621–633.
[67] M. Lynch, A. Force, The probability of duplicate gene preservation by subfunctionalization, Genetics 154 (2000) 459–473.
[68] B.C.Y. Collard, M.Z.Z. Jahufer, J.B. Brouwer, E.C.K. Pang, An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts, Euphytica 142 (2005) 169–196.
[69] S. Akama, R. Shimizu-Inatsugi, K.K. Shimizu, J. Sese, Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid *Arabidopsis*, Nucleic Acids Res. 42 (2014) e46.
[70] A.S. Mason, J. Takahira, C. Atri, B. Samans, A. Hayward, W.A. Cowling, J. Batley, M.N. Nelson, Microspore culture reveals complex meiotic behaviour in a trigenomic *Brassica* hybrid, BMC Plant Biol. 15 (2015) 173.
[71] A.S. Mason, J. Batley, P.E. Bayer, A. Hayward, W.A. Cowling, M. N. Nelson, High-resolution molecular karyotyping uncovers pairing between ancestrally related *Brassica* chromosomes, New Phytol. 202 (2014) 964–974.
[72] T.G. Lee, B.W. Diers, M.E. Hudson, An efficient method for measuring copy number variation applied to improvement of nematode resistance in soybean, Plant J. 88 (2016) 143–153.
[73] H. Staňková, A.R. Hastie, S. Chan, J. Vrána, Z. Tulpová, M. Kubaláková, P. Visendi, S. Hayashi, M. Luo, J. Batley, D. Edwards, J. Doležel, H. Šimková, BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes, Plant Biotechnol. J. 14 (2016) 1523–1531.
[74] Y. Jiao, P. Peluso, J. Shi, T. Liang, M.C. Stitzer, B. Wang, M.S. Campbell, J.C. Stein, X. Wei, C.-S. Chin, K. Guill, M. Regulski, S. Kumari, A. Olson, J. Gent, K.L. Schneider, T.K. Wolfgruber, M.R. May, N.M. Springer, E. Antoniou, W.R. McCombie, G.G. Presting, M. McMullen, J. Ross-Ibarra, R.K. Dawe, A. Hastie, D.R. Rank, D. Ware, Improved maize reference genome with single-molecule technologies, Nature 546 (2017) 524–527.
[75] T.P. Michael, F. Jupe, F. Bemm, S.T. Motley, J.P. Sandoval, C. Lanz, O. Loudet, D. Weigel, J.R. Ecker, High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell, Nat. Commun. 9 (2018) 541.
[76] F. Jupe, A.C. Rivkin, T.P. Michael, M. Zander, T.S. Motley, J.P. Sandoval, K.R. Slotkin, H. Chen, R. Castagnon, J.R. Nery, J.R. Ecker, The Complex Architecture of Plant Transgene Insertions, bioRxiv, 2018 https://doi.org/10.1101/282772.
[77] Y. Yuan, Z. Milec, P.E. Bayer, J. Vrana, J. Dolezel, D. Edwards, W. Erskine, P. Kaur, Large-Scale Structural Variation Detection in Subterranean Clover Subtypes Using Optical Mapping Validated at Nucleotide Level, bioRxiv, 2017 https://doi.org/10.1101/232132.
[78] A.V. Zimin, D. Puiu, R. Hall, S. Kingan, B.J. Clavijo, S.L. Salzberg, The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*, GigaScience 6 (2017) 1–7.
[79] M.H.W. Schmidt, A. Vogel, A.K. Denton, B. Istace, A. Wormit, H. van de Geest, M.E. Bolger, S. Alseekh, J. Maß, C. Pfaff, U. Schurr, R. Chetelat, F. Maumus, J.-M. Aury, S. Koren, A.R. Fernie, D. Zamir, A.M. Bolger, B. Usadel, De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing, Plant Cell 29 (2017) 2336–2348.
[80] M. Schirmer, R. D'Amore, U.Z. Ijaz, N. Hall, C. Quince, Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data, BMC Bioinf. 17 (2016) 125.
[81] F.W. Li, A. Harkess, A guide to sequence your favorite plant genomes, Appl. Plant Sci. 6 (2018), e1030.
[82] A.A. Golicz, P.E. Bayer, G.C. Barker, P.P. Edger, H. Kim, P.A. Martinez, C.K.K. Chan, A. Severn-Ellis, W.R. McCombie, I.A.P. Parkin, A.H. Paterson, J.C. Pires, A.G. Sharpe, H. Tang, G.R. Teakle, C.D. Town, J. Batley, D. Edwards, The pangenome of an agronomically important crop plant *Brassica oleracea*, Nat. Commun. 7 (2016), 13390.
[83] W. Yao, G. Li, H. Zhao, G. Wang, X. Lian, W. Xie, Exploring the rice dispensable genome using a metagenome-like assembly strategy, Genome Biol. 16 (2015) 187.
[84] S.A. Goff, A draft sequence of the rice genome (*Oryza sativa* L. *ssp. japonica*), Science 296 (2002) 92–100.
[85] P.S. Schnable, D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T.A. Graves, P. Minx, A.D. Reily, L. Courtney, S.S. Kruchowski, C. Tomlinson, C. Strong, K. Delehaunty, C. Fronick, B. Courtney, S.M. Rock, E. Belter, F. Du, K. Kim, R.M. Abbott, M. Cotton, A. Levy, P. Marchetto, K. Ochoa, S.M. Jackson, B. Gillam, W. Chen, L. Yan, J. Higginbotham, M. Cardenas, J. Waligorski, E. Applebaum, L. Phelps, J. Falcone, K. Kanchi, T. Thane, A. Scimone, N. Thane, J. Henke, T. Wang, J. Ruppert, N. Shah, K. Rotter, J. Hodges, E. Ingenthron, M. Cordes, S. Kohlberg, J. Sgro, B. Delgado, K. Mead, A. Chinwalla, S. Leonard, K. Crouse, K. Collura, D. Kudrna, J. Currie, R. He, A. Angelova, S. Rajasekar, T. Mueller, R. Lomeli, G. Scara, A. Ko, K. Delaney, M. Wissotski, G. Lopez, D. Campos, M. Braidotti, E. Ashley, W. Golser, H. Kim, S. Lee, J. Lin, Z. Dujmic, W. Kim, J. Talag, A. Zuccolo, C. Fan, A. Sebastian, M. Kramer, L. Spiegel, L. Nascimento, T. Zutavern, B. Miller, C. Ambroise, S. Muller, W. Spooner, A. Narechania, L. Ren, S. Wei, S. Kumari, B. Faga, M.J. Levy, L. McMahan, P. van Buren, M.W. Vaughn, K. Ying, C.T. Yeh, S.J. Emrich, Y. Jia, A. Kalyanaraman, A.P. Hsia, W.B. Barbazuk, R.S. Baucom, T.P. Brutnell, N.C. Carpita, C. Chaparro, J.M. Chia, J.M. Deragon, J.C. Estill, Y. Fu, J.A. Jeddeloh, Y. Han, H. Lee, P. Li, D.R. Lisch, S. Liu, Z. Liu, D.H. Nagel, M.C. McCann, P. SanMiguel, A.M. Myers, D. Nettleton, J. Nguyen, B.W. Penning, L. Ponnala, K.L. Schneider, D.C. Schwartz, A. Sharma, C. Soderlund, N.M. Springer, Q. Sun, H. Wang, M. Waterman, R. Westerman, T.K. Wolfgruber, L. Yang, Y. Yu, L. Zhang, S. Zhou, Q. Zhu, J.L. Bennetzen, R.K. Dawe, J. Jiang, N. Jiang, G.G. Presting, S.R. Wessler, S. Aluru, R.A. Martienssen, S.W. Clifton, W.R. McCombie, R.A. Wing, R.K. Wilson, The B73 maize genome: complexity, diversity, and dynamics, Science 326 (2009) 1112–1115.
[86] A.H. Paterson, J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, J. Schmutz, M. Spannagl, H.B. Tang, X.Y. Wang, T. Wicker, A.K. Bharti, J. Chapman, F.A. Feltus, U. Gowik, I.V. Grigoriev, E. Lyons, C.A. Maher, M. Martis, A. Narechania, R. P. Otillar, B.W. Penning, A.A. Salamov, Y. Wang, L.F. Zhang, N.C. Carpita, M. Freeling, A.R. Gingle, C.T. Hash, B. Keller, P. Klein, S. Kresovich, M.C. McCann, R. Ming, D.G. Peterson, D. Ware Mehboob-ur-Rahman, P. Westhoff, K.F.X. Mayer, J. Messing, D.S. Rokhsar, The *Sorghum bicolor* genome and the diversification of grasses, Nature 457 (2009) 551–556.
[87] S. Huang, R. Li, Z. Zhang, L. Li, X. Gu, W. Fan, W.J. Lucas, X. Wang, B. Xie, P. Ni, Y. Ren, H. Zhu, J. Li, K. Lin, W. Jin, Z. Fei, G. Li, J. Staub, A. Kilian, E.A.G. van der Vossen, Y. Wu, J. Guo, J. He, Z. Jia, Y. Ren, G. Tian, Y. Lu, J. Ruan, W. Qian, M. Wang, Q. Huang, B. Li, Z. Xuan, J. Cao, Z. Wu Asan, J. Zhang, Q. Cai, Y. Bai, B. Zhao, Y. Han, Y. Li, X. Li, S. Wang, Q. Shi, S. Liu, W. K. Cho, J.-Y. Kim, Y. Xu, K. Heller-Uszynska, H. Miao, Z. Cheng, S. Zhang, J. Wu, Y. Yang, H. Kang, M. Li, H. Liang, X.

Ren, Z. Shi, M. Wen, M. Jian, H. Yang, G. Zhang, Z. Yang, R. Chen, S. Liu, J. Li, L. Ma, H. Liu, Y. Zhou, J. Zhao, X. Fang, G. Li, L. Fang, Y. Li, D. Liu, H. Zheng, Y. Zhang, N. Qin, Z. Li, G. Yang, S. Yang, L. Bolund, K. Kristiansen, H. Zheng, S. Li, X. Zhang, H. Yang, J. Wang, R. Sun, B. Zhang, S. Jiang, J. Wang, Y. Du, S. Li, The genome of the cucumber, *Cucumis sativus* L, Nat. Genet. 41 (2009) 1275–1281.

[88] J. Schmutz, S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng, D. Xu, U. Hellsten, G.D. May, Y. Yu, T. Sakurai, T. Umezawa, M.K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X.C. Zhang, K. Shinozaki, H.T. Nguyen, R.A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R.C. Shoemaker, S.A. Jackson, Genome sequence of the palaeopolyploid soybean, Nature 463 (2010) 178–183.

[89] X. Xu, S.K. Pan, S.F. Cheng, B. Zhang, D.S. Mu, P.X. Ni, G.Y. Zhang, S. Yang, R.Q. Li, J. Wang, G. Orjeda, F. Guzman, M. Torres, R. Lozano, O. Ponce, D. Martinez, G. De la Cruz, S.K. Chakrabarti, V.U. Patil, K.G. Skryabin, B.B. Kuznetsov, N.V. Ravin, T.V. Kolganova, A.V. Beletsky, A.V. Mardanov, A. Di Genova, D.M. Bolser, D.M.A. Martin, G.C. Li, Y. Yang, H.H. Kuang, Q. Hu, X.Y. Xiong, G.J. Bishop, B. Sagredo, N. Mejia, W. Zagorski, R. Gromadka, J. Gawor, P. Szczesny, S.W. Huang, Z.H. Zhang, C.B. Liang, J. He, Y. Li, Y. He, J.F. Xu, Y.J. Zhang, B.Y. Xie, Y.C. Du, D.Y. Qu, M. Bonierbale, M. Ghislain, M.D. Herrera, G. Giuliano, M. Pietrella, G. Perrotta, P. Facella, K. O'Brien, S.E. Feingold, L.E. Barreiro, G.A. Massa, L. Diambra, B.R. Whitty, B. Vaillancourt, H.N. Lin, A. Massa, M. Geoffroy, S. Lundback, D. DellaPenna, C.R. Buell, S.K. Sharma, D.F. Marshall, R. Waugh, G.J. Bryan, M. Destefanis, I. Nagy, D. Milbourne, S.J. Thomson, M. Fiers, J.M.E. Jacobs, K.L. Nielsen, M. Sonderkaer, M. Iovene, G.A. Torres, J.M. Jiang, R. E. Veilleux, C.W.B. Bachem, J. de Boer, T. Borm, B. Kloosterman, H. van Eck, E. Datema, B.T.L. Hekkert, A. Goverse, R.C.H.J. van Ham, R.G.F. Visser, P.G.S. Consortiu, Genome sequence and analysis of the tuber crop potato, Nature 475 (2011) 189–195.

[90] K.F.X. Mayer, R. Waugh, P. Langridge, T.J. Close, R.P. Wise, A. Graner, T. Matsumoto, K. Sato, A. Schulman, G.J. Muehlbauer, N. Stein, R. Ariyadasa, D. Schulte, N. Poursarebani, R. Zhou, B. Steuernagel, M. Mascher, U. Scholz, B. Shi, K. Madishetty, J.T. Svensson, P. Bhat, M. Moscou, J. Resnik, P. Hedley, H. Liu, J. Morris, Z. Frenkel, A. Korol, H. Bergès, S. Taudien, M. Felder, M. Groth, M. Platzer, A. Himmelbach, S. Lonardi, D. Duma, M. Alpert, F. Cordero, M. Beccuti, G. Ciardo, Y. Ma, S. Wanamaker, F. Cattonaro, V. Vendramin, S. Scalabrin, S. Radovic, R. Wing, M. Morgante, T. Nussbaumer, H. Gundlach, M. Martis, J. Poland, M. Spannagl, M. Pfeifer, C. Moisy, J. Tanskanen, A. Zuccolo, J. Russell, A. Druka, D. Marshall, M. Bayer, D. Swarbreck, D. Sampath, S. Ayling, M. Febrer, M. Caccamo, T. Tanaka, S. Wannamaker, T. Schmutzer, J.W.S. Brown, G.B. Fincher, A physical, genetic and functional sequence assembly of the barley genome, Nature 491 (2012) 711–716.

[91] F.G. Li, G.Y. Fan, C.R. Lu, G.H. Xiao, C.S. Zou, R.J. Kohel, Z.Y. Ma, H.H. Shang, X.F. Ma, J.Y. Wu, X.M. Liang, G. Huang, R.G. Percy, K. Liu, W.H. Yang, W.B. Chen, X.M. Du, C.C. Shi, Y.L. Yuan, W.W. Ye, X. Liu, X.Y. Zhang, W.Q. Liu, H.L. Wei, S.J. Wei, G.D. Huang, X.L. Zhang, S.J. Zhu, H. Zhang, F.M. Sun, X. F. Wang, J. Liang, J.H. Wang, Q. He, L.H. Huang, J. Wang, J.J. Cui, G.L. Song, K.B. Wang, X. Xu, J.Z. Yu, Y.X. Zhu, S.X. Yu, Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution, Nat. Biotechnol. 33 (2015) 524–530.

[92] R.K. Varshney, C. Song, R.K. Saxena, S. Azam, S. Yu, A.G. Sharpe, S. Cannon, J. Baek, B.D. Rosen, B. Tar'an, T. Millan, X. Zhang, L.D. Ramsay, A. Iwata, Y. Wang, W. Nelson, A.D. Farmer, P.M. Gaur, C. Soderlund, R.V. Penmetsa, C. Xu, A.K. Bharti, W. He, P. Winter, S. Zhao, J.K. Hane, N. Carrasquilla-Garcia, J.A. Condie, H.D. Upadhyaya, M.-C. Luo, M. Thudi, C. L.L. Gowda, N.P. Singh, J. Lichtenzveig, K.K. Gali, J. Rubio, N. Nadarajan, J. Dolezel, K.C. Bansal, X. Xu, D. Edwards, G. Zhang, G. Kahl, J. Gil, K.B. Singh, S.K. Datta, S.A. Jackson, J. Wang, D.R. Cook, Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement, Nat. Biotechnol. 31 (2013) 240–246.

[93] K.F.X. Mayer, J. Rogers, J. Dolezel, C. Pozniak, K. Eversole, C. Feuillet, B. Gill, B. Friebe, A.J. Lukaszewski, P. Sourdille, T.R. Endo, J. Dolezel, M. Kubalakova, J. Cihalikova, Z. Dubska, J. Vrana, R. Sperkova, H. Simkova, J. Rogers, M. Febrer, L. Clissold, K. McLay, K. Singh, P. Chhuneja, N.K. Singh, J. Khurana, E. Akhunov, F. Choulet, P. Sourdille, C. Feuillet, A. Alberti, V. Barbe, P. Wincker, H. Kanamori, F. Kobayashi, T. Itoh, T. Matsumoto, H. Sakai, T. Tanaka, J.Z. Wu, Y. Ogihara, H. Handa, C. Pozniak, P.R. Maclachlan, A. Sharpe, D. Klassen, D. Edwards, J. Batley, O.A. Olsen, S.R. Sandve, S. Lien, B. Steuernagel, B. Wulff, M. Caccamo, S. Ayling, R.H. Ramirez-Gonzalez, B.J. Clavijo, B. Steuernagel, J. Wright, M. Pfeifer, M. Spannagl, K.F.X. Mayer, M.M. Martis, E. Akhunov, F. Choulet, K.F.X. Mayer, M. Mascher, J. Chapman, J.A. Poland, U. Scholz, K. Barry, R. Waugh, D.S. Rokhsar, G.J. Muehlbauer, N. Stein, H. Gundlach, M. Zytnicki, V. Jamilloux, H. Quesneville, T. Wicker, K.F.X. Mayer, P. Faccioli, M. Colaiacovo, M. Pfeifer, A. M. Stanca, H. Budak, L. Cattivelli, N. Glover, M.M. Martis, F. Choulet, C. Feuillet, K.F.X. Mayer, M. Pfeifer, L. Pingault, K.F. X. Mayer, E. Paux, M. Spannagl, S. Sharma, K.F.X. Mayer, C. Pozniak, R. Appels, M. Bellgard, B. Chapman, M. Pfeifer, M. Pfeifer, S.R. Sandve, T. Nussbaumer, K.C. Bader, F. Choulet, C. Feuillet, K.F.X. Mayer, E. Akhunov, E. Paux, H. Rimbert, S. C. Wang, J.A. Poland, R. Knox, A. Kilian, C. Pozniak, M. Alaux, F. Alfama, L. Couderc, V. Jamilloux, N. Guilhot, C. Viseux, M. Loaec, H. Quesneville, J. Rogers, J. Dolezel, K. Eversole, C. Feuillet, B. Keller, K.F.X. Mayer, O.A. Olsen, S. Praud, A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome, Science 345 (2014) 1251788.

[94] J. Schmutz, P.E. McClean, S. Mamidi, G.A. Wu, S.B. Cannon, J. Grimwood, J. Jenkins, S. Shu, Q. Song, C. Chavarro, M. Torres-Torres, V. Geffroy, S.M. Moghaddam, D. Gao, B. Abernathy, K. Barry, M. Blair, M.A. Brick, M. Chovatia, P. Gepts, D.M. Goodstein, M. Gonzales, U. Hellsten, D.L. Hyten, G. Jia, J.D. Kelly, D. Kudrna, R. Lee, M.M.S. Richard, P.N. Miklas, J.M. Osorno, J. Rodrigues, V. Thareau, C.A. Urrea, M. Wang, Y. Yu, M. Zhang, R.A. Wing, P.B. Cregan, D.S. Rokhsar, S.A. Jackson, A reference genome for common bean and genome-wide analysis of dual domestications, Nat. Genet. 46 (2014) 707–713.

[95] R.K. Varshney, C. Shi, M. Thudi, C. Mariac, J. Wallace, P. Qi, H. Zhang, Y. Zhao, X. Wang, A. Rathore, R.K. Srivastava, A. Chitikineni, G. Fan, P. Bajaj, S. Punnuri, S.K. Gupta, H. Wang, Y. Jiang, M. Couderc, M.A.V.S.K. Katta, D.R. Paudel, K.D. Mungra, W. Chen, K.R. Harris-Shultz, V. Garg, N. Desai, D. Doddamani, N.A. Kane, J.A. Conner, A. Ghatak, P. Chaturvedi, S. Subramaniam, O.P. Yadav, C. Berthouly-Salazar, F. Hamidou, J. Wang, X. Liang, J. Clotault, H.D. Upadhyaya, P. Cubry, B. Rhoné, M.C. Gueye, R. Sunkar, C. Dupuy, F. Sparvoli, S. Cheng, R.S. Mahala, B. Singh, R.S. Yadav, E. Lyons, S.K. Datta, C.T. Hash, K.M. Devos, E. Buckler, J.L. Bennetzen, A.H. Paterson, P. Ozias-Akins, S. Grando, J. Wang, T. Mohapatra, W. Weckwerth, J.C. Reif, X. Liu, Y. Vigouroux, X. Xu, Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments, Nat. Biotechnol. 35 (2017) 969–976.

[96] J.A. Hofberger, E. Lyons, P.P. Edger, J.C. Pires, M.E. Schranz, Whole genome and tandem duplicate retention facilitated

glucosinolate pathway diversification in the mustard family, Genome Biol. Evol. 5 (2013) 2155–2173.
[97] S.D. Nicolas, H. Monod, F. Eber, A.M. Chèvre, E. Jenczewski, Non-random distribution of extensive chromosome rearrangements in *Brassica napus* depends on genome organization, Plant J. 70 (2012) 691–703.
[98] S.V. Schiessl, B. Huettel, D. Kuehn, R. Reinhardt, R.J. Snowdon, Targeted deep sequencing of flowering regulators in *Brassica napus* reveals extensive copy number variation, Sci. Data 4 (2017), 170013.
[99] L.W. Qian, K. Voss-Fels, Y.X. Cui, H.U. Jan, B. Samans, C. Obermeier, W. Qian, R.J. Snowdon, Deletion of a stay-green gene associates with adaptive selection in *Brassica napus*, Mol. Plant 9 (2016) 1559–1569.
[100] J. Evans, E. Crisovan, K. Barry, C. Daum, J. Jenkins, G. Kunde-Ramamoorthy, A. Nandety, C.Y. Ngan, B. Vaillancourt, C.L. Wei, J. Schmutz, S.M. Kaeppler, M.D. Casler, C.R. Buell, Diversity and population structure of northern switchgrass as revealed through exome capture sequencing, Plant J. 84 (2015) 800–815.
[101] J. Zhu, S. Pearce, A. Burke, D.R. See, D.Z. Skinner, J. Dubcovsky, K. Garland-Campbell, Copy number and haplotype variation at the VRN-A1 and central FR-A2 loci are associated with frost tolerance in hexaploid wheat, Theor. Appl. Genet. 127 (2014) 1183–1197.
[102] T. Würschum, C.F.H. Longin, V. Hahn, M.R. Tucker, W.L. Leiser, Copy number variations of CBF genes at the Fr-A2 locus are essential components of winter hardiness in wheat, Plant J. 89 (2017) 764–773.
[103] E. Francia, C. Morcia, M. Pasquariello, V. Mazzamurro, J.A. Milc, F. Rizza, V. Terzi, N. Pecchioni, Copy number variation at the HvCBF4-HvCBF2 genomic segment is a major component of frost resistance in barley, Plant Mol. Biol. 92 (2016) 161–175.
[104] T. Sutton, U. Baumann, J. Hayes, N.C. Collins, B.J. Shi, T. Schnurbusch, A. Hay, G. Mayo, M. Pallotta, M. Tester, P. Langridge, Boron-toxicity tolerance in barley arising from efflux transporter amplification, Science 318 (2007) 1446–1449.
[105] T.A. Gaines, D.L. Shaner, S.M. Ward, J.E. Leach, C. Preston, P. Westra, Mechanism of resistance of evolved glyphosate-resistant palmer amaranth (*Amaranthus palmeri*), J. Agric. Food Chem. 59 (2011) 5886–5889.
[106] T.A. Gaines, A.A. Wright, W.T. Molin, L. Lorentz, C.W. Riggins, P.J. Tranel, R. Beffa, P. Westra, S.B. Powles, Identification of genetic elements associated with EPSPS gene amplification, PLoS One 8 (2013), e65819.
[107] D.H. Koo, W.T. Molin, C.A. Saski, J. Jiang, K. Putta, M. Jugulam, B. Friebe, B.S. Gill, Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed *Amaranthus palmeri*, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) 3332–3337.
[108] C.M. Taylor, L.G. Kamphuis, W. Zhang, G. Garg, J.D. Berger, M. Mousavi-Derazmahalleh, P.E. Bayer, D. Edwards, K.B. Singh, W.A. Cowling, M.N. Nelson, INDEL variation in the regulatory region of the major flowering time gene *LanFTc1* is associated with vernalisation response and flowering time in narrow-leafed lupin (*Lupinus angustifolius* L.), Plant Cell Environ. (2018) https://doi.org/10.1111/pce.13320.
[109] A. Srikanth, M. Schmid, Regulation of flowering time: all roads lead to Rome, Cell. Mol. Life Sci. 68 (2011) 2013–2037.
[110] Y. Hu, J. Ren, Z. Peng, A.A. Umana, H. Le, T. Danilova, J.J. Fu, H.Y. Wang, A. Robertson, S.H. Hulbert, F.F. White, S.Z. Liu, Analysis of extreme phenotype bulk copy number variation (XP-CNV) identified the association of *rp1* with resistance to Goss's wilt of maize, Front. Plant Sci. 9 (2018) 110.
[111] L.G. Maron, C.T. Guimaraes, M. Kirst, P.S. Albert, J.A. Birchler, P.J. Bradbury, E.S. Buckler, A.E. Coluccio, T.V. Danilova, D. Kudrna, J.V. Magalhaes, M.A. Pineros, M.C. Schatz, R.A. Wing, L.V. Kochian, Aluminum tolerance in maize is associated with higher MATE1 gene copy number, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 5241–5246.
[112] N.P. Carter, Methods and strategies for analyzing copy number variation using DNA microarrays, Nat. Genet. 39 (2007) S16–S21.
[113] M. Cifuentes, L. Grandont, G. Moore, A.M. Chèvre, E. Jenczewski, Genetic regulation of meiosis in polyploid species: new insights into an old question, New Phytol. 186 (2010) 29–36.
[114] I. Schubert, M.A. Lysak, Interpretation of karyotype evolution should consider chromosome structural constraints, Trends Genet. 27 (2011) 207–216.
[115] Z.Y. Xiong, R.T. Gaeta, J.C. Pires, Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 7908–7913.
[116] M. Chester, M.J. Lipman, J.P. Gallagher, P.S. Soltis, D.E. Soltis, An assessment of karyotype restructuring in the neoallotetraploid *Tragopogon miscellus* (Asteraceae), Chromosom. Res. 21 (2013) 75–85.
[117] K. Lin, N.W. Zhang, E.I. Severing, H. Nijveen, F. Cheng, R.G.F. Visser, X.W. Wang, D. de Ridder, G. Bonnema, Beyond genomic variation - comparison and functional annotation of three *Brassica rapa* genomes: a turnip, a rapid cycling and a Chinese cabbage, BMC Genomics 15 (2014) 250.
[118] Y.H. Li, G.Y. Zhou, J.X. Ma, W.K. Jiang, L.G. Jin, Z.H. Zhang, Y. Guo, J.B. Zhang, Y. Sui, L.T. Zheng, S.S. Zhang, Q.Y. Zuo, X.H. Shi, Y.F. Li, W.K. Zhang, Y.Y. Hu, G.Y. Kong, H.L. Hong, B. Tan, J. Song, Z.X. Liu, Y.S. Wang, H. Ruan, C.K.L. Yeung, J. Liu, H.L. Wang, L.J. Zhang, R.X. Guan, K.J. Wang, W.B. Li, S.Y. Chen, R.Z. Chang, Z. Jiang, S.A. Jackson, R.Q. Li, L.J. Qiu, *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits, Nat. Biotechnol. 32 (2014) 1045–1052.
[119] P. Zhou, K.A.T. Silverstein, T. Ramaraj, J. Guhlin, R. Denny, J. Q. Liu, A.D. Farmer, K.P. Steele, R.M. Stupar, J.R. Miller, P. Tiffin, J. Mudge, N.D. Young, Exploring structural variation and gene family architecture with de novo assemblies of 15 *Medicago* genomes, BMC Genomics 18 (2017) 261.
[120] G. Consortium, C. Alonso-Blanco, J. Andrade, C. Becker, F. Bemm, J. Bergelson, K.M. Borgwardt, J. Cao, E. Chae, T.M. Dezwaan, W. Ding, J.R. Ecker, M. Exposito-Alonso, A. Farlow, J. Fitz, X.C. Gan, D.G. Grimm, A.M. Hancock, S.R. Henz, S. Holm, M. Horton, M. Jarsulic, R.A. Kerstetter, A. Korte, P. Korte, C. Lanz, C.R. Lee, D.Z. Meng, T.P. Michael, R. Mott, N. W. Muliyati, T. Nagele, M. Nagler, V. Nizhynska, M. Nordborg, P.Y. Novikova, F.X. Pico, A. Platzer, F.A. Rabanal, A. Rodriguez, B.A. Rowan, P.A. Salome, K.J. Schmid, R.J. Schmitz, U. Seren, F.G. Sperone, M. Sudkamp, H. Svardal, M. M. Tanzer, D. Todd, S.L. Volchenboum, C.M. Wang, G. Wang, X. Wang, W. Weckwerth, D. Weigel, X.F. Zhou, 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*, Cell 166 (2016) 481–491.
[121] S. Pinosio, S. Giacomello, P. Faivre-Rampant, G. Taylor, V. Jorge, M.C. Le Paslier, G. Zaina, C. Bastien, F. Cattonaro, F. Marroni, M. Morgante, Characterization of the poplar pan-genome by genome-wide identification of structural variation, Mol. Biol. Evol. 33 (2016) 2706–2719.
[122] F. Lu, M.C. Romay, J.C. Glaubitz, P.J. Bradbury, R.J. Elshire, T. Y. Wang, Y. Li, Y.X. Li, K. Semagn, X.C. Zhang, A.G. Hernandez, M.A. Mikel, I. Soifer, O. Barad, E.S. Buckler, High-resolution genetic mapping of maize pan-genome sequence anchors, Nat. Commun. 6 (2015) 6914.

Chapter 2: Differential evolution of *Brassica* flowering time genes

Many gene families seem to have a common ancestor, so obviously this ancestor has duplicated and diversified afterwards. Those processes of sub- and neofunctionalisation are central to gene evolution, but generally hard to observe due to the long periods involved. *Brassica* species are particularly well suited to study those processes, as they contain both mesohexaploid species and recent allotetraploids. It is therefore possible to study both long and short-term refunctionalisation. This does not only have theoretical value, but is also important for understanding selective responses in breeding: Breeders need to know which copies they need to check for allelic variance and which copies have lower or no importance. The papers presented in this chapter show studies of increasing evidence for subfunctionalisation across the flowering time genes *Bna.FT*, *Bna.FLC* and *Bna.SRR1*. The invited Spotlight paper in Chapter 2.1 highlights that allelic variance in only one copy of each *Bna.FT* and *Bna.FLC* is responsible for ecotype differences, while the other copies show no relevant allelic variance. The experimental paper in Chapter 2.2 shows that all nine annotated *Bna.FLC* copies have different expression patterns. Three of them do not react to cold anymore, so did loose part of their original function. Chapter 2.3 is an experimental paper showing that different copies of *Bna.SRR1* have different abilities to restore *srr1* mutants in *A. thaliana* and also identifies possible protein domains responsible for this change.

Chapter 2.1: Illuminating Crop Adaptation Using Population Genomics

Chapter II:

Snowdon RJ, **Schiessl S** (2019) Illuminating Crop Adaptation Using Population Genomics. Molecular Plant 12:27–29. doi: 10.1016/j.molp.2018.12.014

Molecular Plant
Spotlight

CellPress
PARTNER JOURNAL

Illuminating Crop Adaptation Using Population Genomics

In this issue, Wu et al. (2019) describe the largest whole-genome resequencing dataset published to date for rapeseed (*Brassica napus*), an allopolyploid crop species that originated just a few thousand years ago under anthropogenic influence and rapidly evolved into one of the world's most important oilseed crops. In almost 1000 accessions spanning species-wide germplasm for oilseed rape, a comprehensive analysis of sequence diversity related to flowering-related traits uncovered selective sweeps associated with eco-geographic adaptation and human selection, attributable particularly to divergence among homoeologs of key flowering-time regulation and ethylene synthesis/signaling genes. The authors further extend their analysis to retrospective, genome-based documentation of diversity footprints that trace the global expansion of the species across a century of breeding. The study contributes not only a rich catalog of genome-wide diversity for genetic analysis and future breeding of important agro-economic traits but also a unique conceptional framework for ongoing selective adaptation of oilseed rape crops to emerging challenges presented by changing climatic conditions in key production areas. Selective sweeps were found to contain genes involved in stress adaptation and development, especially flowering time, indicating that climatic adaptation both in terms of local stress factors and of life-cycle adaptation was the major factor underlying agronomic selection for improved seed yield.

Modulation of flowering behavior and morphotype evolution via post-polyploidization genome restructuring and homoeologous expression changes has previously been identified as a decisive factor in the success of *B. napus* as a diverse, globally adapted crop (Samans et al., 2017; Schiessl et al., 2017). This new study expands that knowledge by combining information on genome-wide SNP diversity and linkage disequilibrium (LD) with genome-wide association studies (GWAS) and gene expression data. The resulting picture underlines the pivotal role of mutations creating novel flowering and stress response phenotypes as a driver of geographic expansion. Moreover, the data show that this selection in *B. napus* was very specific to distinct gene copies and did not act in parallel on copies of the same gene. Interestingly, functional variation apparently not only arose by mutations within coding sequences but rather via mutations in promoter regions. The two most prominent flowering-time genes, *FLOWERING LOCUS T (FT)* and *FLOWERING LOCUS C* (*FLC*), were strong candidates for selection. Indeed, the expression patterns of *B. napus FT* and *FLC* homoeologs vary in expression between different ecotypes (Figure 1), indicating that promoter variation in floral regulatory genes is an essential factor in the creation of adaptive variation and needs to be considered in knowledge-based breeding approaches. The phenomenon of promoter-driven adaptation may also hold true for other closely related crops: For example, selection against premature bolting in spring-type *Brassica rapa* ecotypes was shown to be highly associated with promoter variation in a copy of *VERNALIZATION INSENSITIVE 3 (VIN3)* (Su et al., 2018). Interestingly, similar effects are observed in cereals, where mutations in promoter sequences of the cereal gene *VERNALIZATION 1* (*VRN1*) also associate with eco-geographic adaptation (Deng et al., 2015). Moreover, potential variation in promoter sequences was recently postulated as a possible driver of pleiotropic effects that appear to simultaneously modulate flowering responses and root architecture in wheat and barley (Voss-Fels et al., 2018). Such findings underline the importance of detailed knowledge about the genetic intricacies of flowering-time regulation and abiotic stress responses at the DNA sequence and regulatory expression levels. To provide the knowledge needed for targeted breeding of climate-adapted crops to cope with future challenges, we urgently need more and better resources for illuminating genome-wide and genome-deep data.

For a long time, polyploidy was a major obstacle to providing such resources. The first genome assembly for *B. napus* (Chalhoub et al., 2014) represented one of the most highly duplicated and structurally rearranged polyploid plant genomes to be completed at the time. It was also one of the last crop genomes to be assembled on a backbone of DNA reads generated from tiled BAC clones using Titanium Roche 454 and Sanger sequencing technologies. Unimaginable just a decade ago for a crop species with such a complex genome, the dimensions of this new resequencing study underline the enormous power of large-scale genome interrogation. However, as sequencing power and cost no longer present significant barriers to mining for useful genetic diversity, even in complex crop genomes, new questions arise with regard to the maximum exploitation of large-scale sequence datasets. One key concern for breeders and researchers is the need for effective, standardized, and integrated (e.g., across species) platforms to manage, mine, and utilize genomic data from extensive crop plant collections, coupled with methods and standards to unambiguously link specific genotypes and seed lots to corresponding genomic and phenotypic data. Given that data handling, storage, and downstream analysis represent potentially greater future challenges than the actual generation of large-scale genomic sequence datasets, public bioinformatics infrastructure for effective sharing and exploitation of published genome datasets are essential prerequisites to maximize the added value of genome sequences for crop genetics and breeding.

Enhancing the opportunities to mine genome sequence data for previously invisible diversity (Gabur et al., 2018a) is particularly relevant for species like *B. napus* with dynamic, allopolyploid genomes where homeologous chromosome exchanges are

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

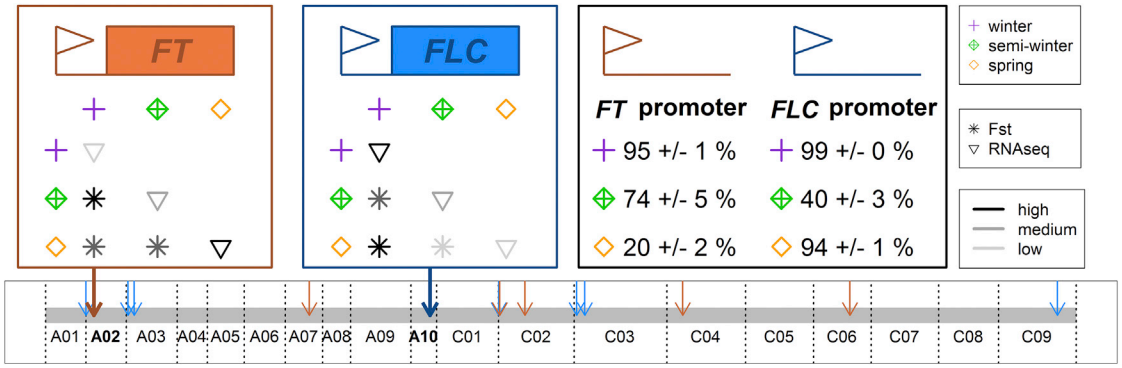


Figure 1. Selective Variants of Regulatory Flowering-Time Genes in Different *Brassica napus* ecotypes.
Eco-geographic selection acting on the flowering-time gene copies BnaA02g12130D (FT copy on chromosome A02, orange box) and BnaA10g22080D (FLC copy on chromosome A10, blue box) in *Brassica napus*. The thin orange arrows indicate the positions of the remaining FT copies in the *B. napus* genome, while the thin blue arrows indicate the positions of the remaining FLC copies. Ecotypes are represented by colored symbols (see legend). The genetic fixation index (F_{st}) between ecotypes is indicated by the tone of gray in the star symbols, with darker tones indicating higher values and therefore stronger selection. The average expression of each gene in young leaves of non-vernalized plants is indicated with a triangle, with darker triangles indicating higher expression. The table in the black box shows the degree of conservation (in %, ±SE) of decisive sequence polymorphisms within the promotor regions of the two gene copies for each ecotype.

prolific (Hurgobin et al., 2017; Gabur et al., 2018b). For example, there is growing evidence that novel genome structural variants in *B. napus* are created even during self-pollination of homozygous cultivars or during generation of doubled haploid lines (He et al., 2017). In this context, genome-wide sequences can potentially reveal a scale of structural genome diversity within "fixed" genotypes that was hitherto undiscoverable. The closer we look, the clearer it becomes that small-scale deletions affecting duplicates of *B. napus* genes or their promoter regions, many of which may be remnants of structural variants caused by illicit pairing of homoeologous chromosomes, are an important source of new adaptive and/or agronomic trait variation (Harper et al., 2012; Qian et al., 2016; Schiessl et al., 2017) with enormous potential value for breeding.

With this in mind, a further key issue of particular relevance for crops with highly duplicated and strongly restructured genomes is the quality and choice of reference genome assemblies for read mapping from short-read sequence data. The potential for ascertainment bias caused by an inability to interrogate chromosome regions not represented in specific accessions due to genome structural variation is rarely considered in resequencing studies. This is because the huge advantages of comprehensive genome sequence data outweigh any potential shortfalls in comparison with less comprehensive genotyping platforms based on fixed SNP panels that were identified in specific germplasm collections. Nevertheless, the value of enormous genome datasets such as the one described by Wu et al. (2019) will increase even further when gold-standard pan-genomes representing the full extent of available morphotype diversity become available. Here, short reads were mapped to assemblies of the winter oilseed rape cultivars Darmor-*bzh* (Chalhoub et al., 2014) and Tapidor (Bayer et al., 2017). The differences in SNP and InDel calling were small and led to the same conclusions. However, structural variants within selective sweeps were not evaluated, so the full extent of genetic diversification for crop evolution and adaptation resulting from genome restructuring has yet to be discovered. Further analysis of such expansive genome sequencing datasets in the context of pan-genomic references thus has considerable potential to deliver key information to future breeders. Powerful new methods to generate platinum-quality assemblies of complex *Brassica* genomes using long-read sequence data (Belser et al., 2018), in combination with population-scale resequencing using inexpensive short-read technologies, can help provide an ordered catalog of pan-genomic sequence diversity with huge relevance to breeders.

ACKNOWLEDGMENTS

The authors confirm that they have no conflict of interest.

Received: December 17, 2018
Revised: December 17, 2018
Accepted: December 18, 2018
Published: December 21, 2018

Rod J. Snowdon and Sarah Schiessl*

Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition Justus Liebig University Giessen, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

***Correspondence: Rod J. Snowdon (rod.snowdon@agrar.uni-giessen.de)**
https://doi.org/10.1016/j.molp.2018.12.014

REFERENCES

Bayer, P.E., Hurgobin, B., Golicz, A.A., Chan, C.K., Yuan, Y., Lee, H., Renton, M., Meng, J., Li, R., Long, Y., et al. (2017). Assembly and comparison of two closely related *Brassica napus* genomes. Plant Biotechnol. J. **15**:1602–1610.

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.C., Falentin, C., Genete, M., Berrabah, W., Chevre, A.M., Delourme, R., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat. Plants **4**:879–887.

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

Deng, W.W., Casao, M.C., Wang, P.H., Sato, K., Hayes, P.M., Finnegan, E.J., and Trevaskis, B. (2015). Direct links between the

vernalization response and other key traits of cereal crops. Nat. Commun. **6**:5882.

Gabur, I., Chawla, H.S., Liu, X.W., Kumar, V., Faure, S., von Tiedemann, A., Jestin, C., Dryzska, E., Volkmann, S., Breuer, F., et al. (2018a). Finding invisible quantitative trait loci with missing data. Plant Biotechnol. J. **16**:2102–2112.

Gabur, I., Chawla, H.S., Snowdon, R.J., and Parkin, I.A.P. (2018b). Connecting genome structural variation with complex traits in crop plants. Theor. Appl. Genet. https://doi.org/10.1007/s00122-018-3233-0.

Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P., and Bancroft, I. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. Nat. Biotechnol. **30**:798–802.

He, Z., Wang, L., Harper, A.L., Havlickova, L., Pradhan, A.K., Parkin, I.A.P., and Bancroft, I. (2017). Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. Plant Biotechnol. J. **15**:594–604.

Hurgobin, B., Golicz, A.A., Bayer, P.E., Chan, C.K., Tirnaz, S., Dolatabadian, A., Schiessl, S.V., Samans, B., Montenegro, J.D., Parkin, I.A.P., et al. (2017). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. Plant Biotechnol. J. **16**:1265–1274.

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

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**. https://doi.org/10.3835/plantgenome2017.02.0013.

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

Su, T.B., Wang, W.H., Li, P.R., Zhang, B., Li, P., Xin, X.Y., Sun, H.H., Yu, Y.J., Zhang, D.S., Zhao, X.Y., et al. (2018). A genomic variation map provides insights into the genetic basis of spring Chinese cabbage (*Brassica rapa* ssp *pekinensis*) selection. Mol. Plant **11**:1360–1376.

Voss-Fels, K.P., Robinson, H., Mudge, S.R., Richard, C., Newman, S., Wittkop, B., Stahl, A., Friedt, W., Frisch, M., Gabur, I., et al. (2018). VERNALIZATION1 modulates root system architecture in wheat and barley. Mol. Plant **11**:226–229.

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., Zhou, G., Lohwasser, U., Hua, S., Wang, H., et al. (2019). Whole-genome resequencing of a world-wide collection of rapeseed accessions reveals genetic basis of their ecotype divergence. Mol. Plant https://doi.org/10.1016/j.molp.2018.11.007.

Chapter 2.2: The vernalisation regulator *FLOWERING LOCUS C* is differentially expressed in biennial and annual *Brassica napus*

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

SCIENTIFIC REPORTS
nature research

OPEN

The vernalisation regulator *FLOWERING LOCUS C* is differentially expressed in biennial and annual *Brassica napus*

Sarah V. Schiessl[1,3*], Daniela Quezada-Martinez[1,3], Ellen Tebartz[1], Rod J. Snowdon[1] & Lunwen Qian[2*]

Plants in temperate areas evolved vernalisation requirement to avoid pre-winter flowering. In *Brassicaceae*, a period of extended cold reduces the expression of the flowering inhibitor *FLOWERING LOCUS C* (*FLC*) and paves the way for the expression of downstream flowering regulators. As with all polyploid species of the *Brassicaceae*, the model allotetraploid *Brassica napus* (rapeseed, canola) is highly duplicated and carries 9 annotated copies of *Bna.FLC*. To investigate whether these multiple homeologs and paralogs have retained their original function in vernalisation or undergone subfunctionalisation, we compared the expression patterns of all 9 copies between vernalisation-dependent (biennial, winter type) and vernalisation-independent (annual, spring type) accessions, using RT-qPCR with copy-specific primers and RNAseq data from a diversity set. Our results show that only 3 copies – *Bna.FLC.A03b*, *Bna.FLC.A10* and to some extent *Bna.FLC.C02* – are differentially expressed between the two growth types, showing that expression of the other 6 copies does not correlate with growth type. One of those 6 copies, *Bna.FLC.C03b*, was not expressed at all, indicating a pseudogene, while three further copies, *Bna.FLC.C03a* and *Bna.FLC.C09a*b, did not respond to cold treatment. Sequence variation at the *COOLAIR* binding site of *Bna.FLC.A10* was found to explain most of the variation in gene expression. However, we also found that *Bna.FLC.A10* expression is not fully predictive of growth type.

Synchronization of reproduction with favorable environmental conditions is crucial for a species' survival. For plants, timing of reproduction is mainly governed by flowering time regulation. In temperate climates, the two most important cues to induce flowering are a period of extended cold (the vernalisation pathway) and day length (the photoperiod pathway)[1]. In contrast to biennial or winter forms, annual or spring forms are lacking a functional vernalisation pathway and therefore flower without a period of extended cold.

The vernalisation pathway was found to be a complex regulatory system including transcriptional and epigenetic regulation mechanisms[2]. In *Arabidopsis thaliana*, the model *Brassicaceae*, a single copy of the transcriptional repressor *FLOWERING LOCUS C* (*FLC*) inhibits expression of the central flowering regulator *FLOWERING LOCUS T* (*FT*)[3]. In this state, *FT* repression via *FLC* overrides activating effects from the photoperiod pathway and other activating input signals[4]. FLC also represses the expression of further important flowering genes by binding to the promoter, like *FLOWERING LOCUS D* (*FD*), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), and *TEMPRANILLO 1* (*TEM1*)[5,6]. While the regulation of many *FLC* targets depends on complex formation with SHORT VEGETATIVE PHASE (SVP), *FT* and *SOC1* expression can be independently regulated by either FLC or SVP[7]. The activating function of FT depends on dimer formation with FD[8]: *FLC* itself is regulated via the vernalisation pathway[2] and via the autonomous pathway[9]. There is also cross-talk with the photoperiod pathway via the protein SENSITIVITY TO RED LIGHT REDUCED 1 (SRR1)[10]. *FLC* expression is regulated via chromatin conformation, which itself is governed via histone modifications[11]. Genetic and

[1]Department of Plant Breeding, Justus Liebig University, IFZ Research Centre for Biosystems, Land Use and Nutrition, Heinrich-Buff-Ring 26-32, 35392, Giessen, Germany. [2]Collaborative Innovation Center of Grain and Oil Crops in South China, Hunan Agricultural University, Changsha, 410128, China. [3]These authors contributed equally: Sarah V. Schiessl and Daniela Quezada-Martinez. *email: sarah-veronica.schiessl@agrar.uni-giessen.de; qianlunwen@163.com

epigenetic regulation processes induced by extended cold stabilize a chromatin state which does not allow transcription of *FLC* mRNA[12]. This inactive state is mitotically, but not meiotically stable; therefore, the next generation starts in a non-vernalized state[3,13]. In the non-vernalized state, *FLC* expression is promoted by the FRI-C, a large protein complex built up by the scaffold protein FRIGIDA (FRI) in interaction with SUPPRESSOR OF FRIGIDA 1 (SUF4)[14]. FRI-C acts to recruit general transcription factors and specific chromatin modification factors to the *FLC* chromatin[14]. Cold induces the transcription of a noncoding antisense transcript from *FLC* called *COOLAIR*, which starts to decrease *FLC* mRNA levels[12,15]. Independently from that, but later, alternative splicing produces the noncoding sense RNA *COLDAIR* from the first intron of *FLC* itself[12,15]. *COLDAIR* recruits Polycomb Group proteins (PcG) proteins forming the PRC2-like complex to the *FLC* chromatin[15], where they remove activating marks and add repressive marks to the histones in the *FLC* chromatin in interaction with the protein VERNALIZATION INSENSITIVE 3 (VIN3)[16]. This is further supported by the action of the Paf1 complex[17]. Those repressive marks are recognized and bound by the PRC2 complex containing the proteins VERNALISATION 2 (VRN2) and LIKE HETEROCHROMATIN 1 (LHP1) (also known as TFL2), keeping *FLC* repression stable until the next generation[18–20].

In *A. thaliana*, spring and winter forms have been shown to carry different allelic variants of *FLC, FRI* and *VIN3*[14,21–23]. As the main vernalisation components were found to be conserved in the *Brassicaceae*[24–31], their paralogs are also candidates for the difference between spring and winter forms in *Brassica* crops. Variation in paralogs of *FLC* was found to be associated to a change in flowering time for *B. rapa*[32–35], *B. oleracea*[25,36–38] and *B. napus*[26,28,39–41]. However, these crops have a paleopolyploid origin and were shown to carry several paralogs of *FLC*. The reference genome of *B. napus* carries nine annotated copies of *Bna.FLC*[42]. Early transformation studies revealed that at least five of them are able to delay flowering in *A. thaliana*, but to a varying extent[28]. The strongest effect was observed in plants transformed with *BnFLC1* (*Bna.FLC.A10*)[28]. Later it was discovered that a transposon insertion in the upstream region of *Bna.FLC.A10* is strongly associated to the vernalisation-dependent phenotype[39]. Depending on the genetic background investigated, different copies of *Bna.FLC* were repeatedly found within the confidence intervals of quantitative trait loci (QTL) for flowering time in *B. napus*[43–48]. Previously we found that a *R10P* mutation in the *Bna.FLC.A10* (BnaA10g22080D) copy is a strong candidate for the winter-spring split in a diverse set of *B. napus* accessions[41]. More recently, it was found that the strongest selection signal separating spring and winter type populations was also found at *Bna.FLC.A10*, correlating with differential gene expression[49]. Although different data sets indicate that *Bna.FLC.A10* might be the most decisive copy for vernalisation requirement[28,39,41], the roles of the other copies remain unclear, mostly because specific primers for some copies could not be developed[29,46]. For example, the most extensive expression study on *Bna.FLC* copies to date was not able to resolve four of the nine copies specifically, and the study was only conducted in a winter and a semi-winter type, but not in a spring type[29]. In the present study, we conducted Reverse Transcription Quantitative Real-Time PCR (RT-qPCR) to measure the specific individual expression levels of all *Bna.FLC* copies. Comparing two winter-type and two spring-type *B. napus* accessions, along with a winter-hardy but not strictly vernalisation-dependent winter accession, we measured *Bna.FLC* expression levels with and without vernalisation, and in different tissues prior to vernalisation. Our results indicate that only *Bna.FLC.A03b, Bna.FLC.A10*, and *Bna.FLC.C02* are differentially expressed between winter and spring type accessions, although tissue-specific differences exist. Comparisons with RNAseq data for a diversity set show that this differential expression is a general difference between winter and spring accessions before vernalisation. We then screened a diverse population of 53 winter and 48 spring accessions for *Bna.FLC.A10* and *Bna.FLC.C02* expression with RT-qPCR and found that only *Bna.FLC.A10* was differentially expressed between winter and spring accessions. There were, however, exceptions in both winter and spring accession, and sequence analysis revealed that this was strongly associated to a sequence variant at the *COOLAIR* binding site and a broken reading frame in exon 1. We conclude that this sequence variant contributed to the difference between biennial and annual *B. napus* forms, whereas their remaining homologs appear to have undergone subfunctionalisation and pseudogenisation after polyploidisation.

Material and Methods

Sequence analysis. Genomic sequences were extracted from the respective reference genomes for *B. napus* (the *Darmor-bzh* reference genome, version 4.1[42]), *B. rapa* (the 'Chiifu-401-42' reference genome, version 1.5[50]) and *B.oleracea* (the 'TO1000' reference genome, version 2.1[51]) using the respective annotation. A further copy of *Bna.FLC* was identified in[41] and also included. Moreover, we identified four gene copies from an unpublished version of *B. nigra* genome which was made available by the courtesy of I. Parkin. The gene IDs for the single copies can be found in Table S1. The sequence from *A. thaliana* was retrieved from The Arabidopsis Information Resource (TAIR), www.arabidopsis.org, for gene model AT5G10140.1.

Sequence alignments were performed using CLUSTAL multiple sequence alignment by MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/, version 3.8) with Default parameters. Based on this alignment, we constructed a Maximum likelihood tree and a neighbor joining tree using bootstrap analysis (100 replicates) using MEGA version 10.0.5. Exon-intron structure was determined by aligning the genomic sequences with the respective cDNA sequences from[42]. SNP variation for the five accessions also used in RT-qPCR was taken from[41].

Promoter region analysis was performed using MEME (http://meme-suite.org/tools/meme) and JASPAR (http://jaspar.genereg.net/) using motifs for *A. thaliana*[52].

Plant material, cultivation and sampling. Two *B. napus* winter accessions (Manitoba, Lisabeth) and two spring accessions (Girasol, Korall) were chosen along with a winter-hardy, vernalization-independent winter accession (Mansholt). Mansholt carries duplications for *Bna.FLC.C09a* and *Bna.FLC.C09b*, while the other accessions did not show copy number variations[53]. The plants were sown in 7 × 7 cm pots in 3 biological replicates per treatment and transplanted to 12 × 12 cm pots 4 weeks after sowing. Cultivation was performed in a greenhouse

using a 16/8 h day/night rhythm with 20/17 °C. The first sampling was performed seven weeks after sowing when the plants reach BBCH stage 14 at midday (ZT 7) by cutting off the youngest fully developed leaf. Leaves were frozen in liquid nitrogen and stored at −80 °C until RNA extraction. After the first sampling, one set of plants was brought to the cool room for eight weeks using an 8/16 h day/night rhythm at constant 5 °C. Another set of plants remained in the greenhouse as control plants. After eight weeks of vernalisation, both sets were sampled again at BBCH20.

For the analysis of tissue-specific expression, Manitoba and Korall were grown in three replicates in the greenhouse under the same conditions. Ten weeks after sowing when the plants reach BBCH15, we sampled petioles, developed and emerging leaves separately and kept the samples at −80 °C until RNA extraction.

For the analysis of the time series, Manitoba was grown in 27 replicates in the greenhouse under the same conditions. From an age of 3 weeks up to 8 weeks, we took leaf samples from 3 biological replicates every week. The plants were then subjected to a cold treatment as described above, and we sampled again after 4, 6 and 8 weeks of cold treatment.

For the population screening, 3 biological replicates of 101 accessions (see Supplementary Table 2 for a list of accessions) were sown in quickpot plates and grown for 7 weeks until sampling.

Primer design, cDNA synthesis and quantitative PCR. Total RNA was extracted using the NucleoSpin miRNA kit (Macherey-Nagel) following manufacturer's instructions, quantified using Qubit RNA Broad Range on a Qubit fluorimeter and stored at −80 °C until use.

cDNA synthesis was performed using the RevertAid cDNA synthesis kit (ThermoFisher) using 1 µg of total RNA and Oligo-DT primers. cDNA was quantified using the Qubit DNA High Sensitivity kit on a Qubit fluorimeter. Quantitative Real-time PCR was performed on a Real-Time PCR System ViiA7 cycler (Applied Biosystems) in 384-well plates. The reaction mix containing specific primers, template cDNA and FastStart Universal SYBR Green Master mix containing Rox (Roche) was pipetted by a robot (Biomek 4000, Beckman Coulter). As endogenous control, we used ubiquitin. The PCR program was as follows: initial denaturation (94 °C for 2 min), amplification and quantification (40 cycles, 95 °C for 20 sec, 60 °C for 30 sec, 72 °C for 30 sec), and final extension (72 °C for 5 min). A final melting curve was recorded between 55 and 95 °C. PCR efficiency was measured using a pool of all samples in a dilution series of 6 points. All samples were measured in 3 technical replicates. The normalized expression level was determined using the ΔCt method[54]. The primer sequences are shown in Supplementary Table 3. For comparisons between winter and spring, a Student's t-test was performed, for comparisons of more than one factor, we calculated least significant distances in R using the package agricolae.

RNAseq analysis. For validation of the results from RT-qPCR, we downloaded a data set from NCBI Sequence Read Archive (SRA, SRP069066) published in[55]. These represent RNAseq data from Illumina HiSeq2500 in single-end mode for a publically available *B. napus* diversity set (ASSYST set[56,57]) for 3 weeks old leaf samples without vernalization. Unfortunately, no biological replicates are given, so the data can only be used for comparison or pooled data analysis. We selected (1) the same 5 accessions which were used in our RT-qPCR experiment for comparison (2) 30 randomly selected spring accessions and 30 randomly selected winter accessions for pooled data analysis (see Supplementary Table 4).

Another publically available data set was downloaded to analyze the time course in gene expression for several vernalisation genes before, during and after cold treatment in the spring cultivar Westar from the NCBI SRA (SRP132445). Each sample was a pool of 6–10 individual plants. We only analyzed time points with two replicates, which were 22 days old plants (before vernalisation), 43 and 64 old plants (during vernalization), and 67 and 72 days old plants (after vernalization). For each time point, both leaves and shoot apex were sampled.

Quality control was performed using FastQC, version 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Accordingly, adapter removal and trimming was performed using Trimmomatic version 0.38[58] by first removing TruSeq adapters followed by head cropping the first 9 bases. Clean reads were mapped onto version 4.1 of the *B. napus* '*Darmor-Bzh*' reference genome using Bowtie2[59], alignment mode "very-sensitive". Removal of duplicates, sorting and indexing was carried out with *samtools* version 0.1.19[60]. We then calculated transcripts per million (TPM) using *bedtools* software with multiBamCov[61] for raw coverage calling, followed by normalization in R (version 3.1.2) according to[62] for the sum of all exons of every gene. From this, we selected a list of important vernalisation gene copies, and calculated $|\log_2(\text{mean(TPM(winter))}/\text{mean(TPM(spring))})|$. Besides *FLC*, we included other important vernalisation regulators (*FRI, SUF4, TFL2, VIN3, VRN2, SVP, SRR1*) and genes directly regulated by vernalisation (*SOC1, FD, FT, TEM1*).

Results

Phylogenetic analysis. The coding sequences of the respective homeologous copies are more related to each other than to their inter-subgenome paralogs (Fig. 1A, Supplementary Fig. 1). This is also evident from the gene structure. Most *Bna.FLC* copies have 7 exons, while the truncated pseudogene on A01 only carries exons 4–7. *Bna.FLC.C03b* and *Bna.FLC.C09b* both carry an additional exon between exons 1/2 and 6/7, respectively (Fig. 1B,C). With the exception of *Bna.FLC.A01*, similar relationships were reported elsewhere[28,29,63]. When comparing all *Bna.FLC* copies with their progenitors from *B. rapa* and *B. oleracea*, there is mostly low divergence between the progenitor and the polyploid (Fig. 1A). The largest distance was found between *Bra.FLC.A02* and *Bna.FLC.A02*, followed by *Bol.FLC.C09b* and *Bna.FLC.C09b*. The only copy without a respective homolog in the progenitors was *Bna.FLC.A01*, suggesting that it has been duplicated after the recent interspecific hybridization leading to *B. napus* (<7500 years ago[42]), most likely from *Bna.FLC.A02* (Supplementary Fig. 2). In contrast, both *Bna.FLC.C09* copies have a closely related homolog in the progenitor species *B. oleracea*, indicating that they have been duplicated earlier. The related *Brassica* species *B. nigra* was found to carry 4 *FLC* copies, however, two of those copies (*Bni.FLC.B02* and *Bni.FLC.B08b*) were truncated in our version of the *B. nigra* reference genome,

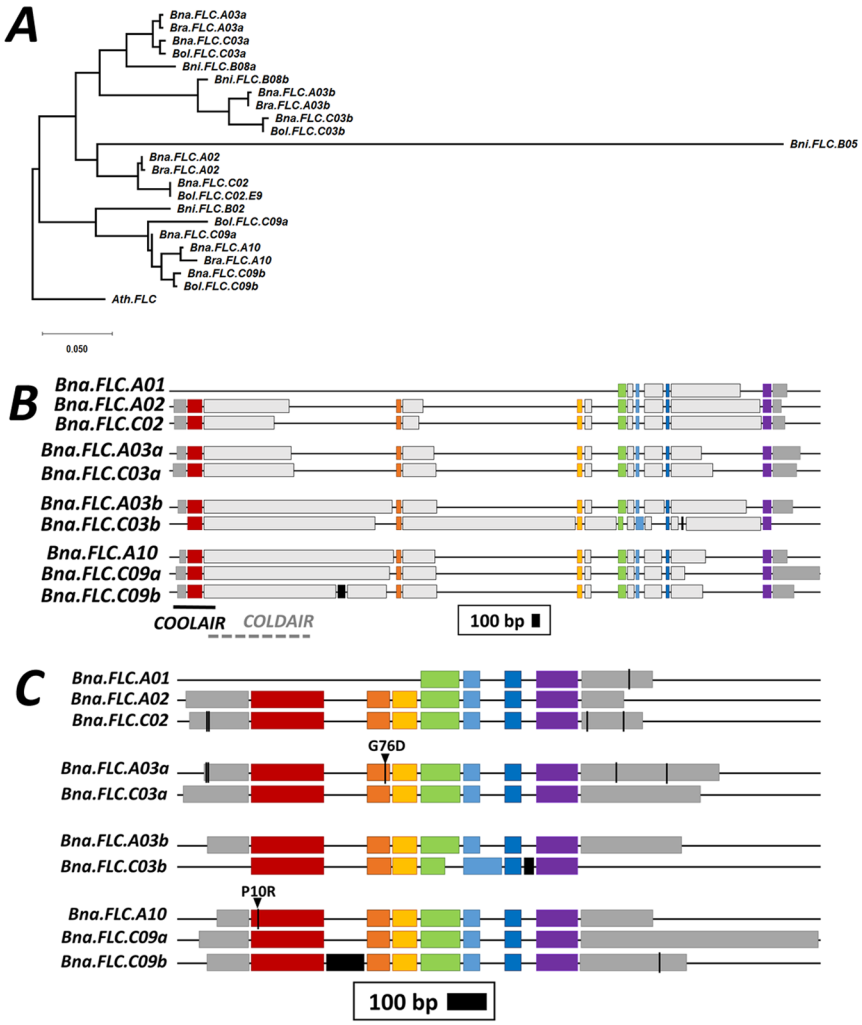


Figure 1. (**A**) Phylogenetic tree constructed with Maximum Likelihood of *Brassica FLC* from three diploid (*B. rapa, B. nigra, B. oleracea)* and one polyploid (*B. napus*) species with *Ath.FLC* as outgroup. cDNA sequences were extracted from the respective reference genomes, the sequence from *A. thaliana* was retrieved from TAIR. Sequence alignment was performed using CLUSTAL multiple sequence alignment by MUSCLE with Default parameters. Based on this alignment, a Maximum Likelihood tree was constructed using MEGA version 10.0.5. (**B**) Full gene structure including UTR (dark grey boxes), introns (light grey boxes) and exons (colored boxes). Black boxes represent extra exons. Box length is proportional to length in bp, the legend shows the length of 100 bp. The solid line in black shows the localization of *COOLAIR*, the dotted line in grey represents the approximate position of *COLDAIR*. (**C**) Exon structure including the position of sequence variants within the 5 accessions used for the initial RT-qPCR. Dark grey boxes represent UTRs, colored boxes represent exons in the same color code than for B, black boxes represent extra exons. Black lines represent a variant, black lines with a black triangle represent a non-synonymous SNP, annotated with the respective amino acid mutation.

which possibly is due to incomplete annotation, but might also indicate pseudogenization. The respective copies cluster with the same main four branches than the AC species copies, but more distantly, reflecting the higher divergence of the B genome from the AC genomes. *Bni.FLC.B05* clusters distantly with the A02/C02 clade, *Bni. FLC.B08ab* cluster with the respective A03ab/C03ab clades, and *Bni.FLC.B02* cluster with the A10/C09 clade (Fig. 1A).

The *Brassica* A and C (sub)genomes can be subdivided into a least fractionated (LF) and two more fractionated (MF1, MF2) partitions[64]. Using synteny analysis[64], we found that *Bna.FLC.A10 (BnaA10g22080D)* and *Bna.FLC.C09a (BnaC09g46500D)* are located in LF, *Bna.FLC.A03a (BnaA03g02820D)* and *Bna.FLC.C03a (BnaC03g04170D)* are located in MF1, and *Bna.FLC.A02 (BnaA02g00370D)* and *Bna.FLC.C02 (BnaC02g00490D)* are located in MF2 (Fig. 2). *Bna.FLC.A03b (BnaA03g13630D), Bna.FLC.C03b (BnaC03g16530D)* and *Bna.FLC. C09b (BnaC09g46540D)* are non-syntenic copies, indicating that they arose after the hexaploidization of the common *Brassica* ancestor.

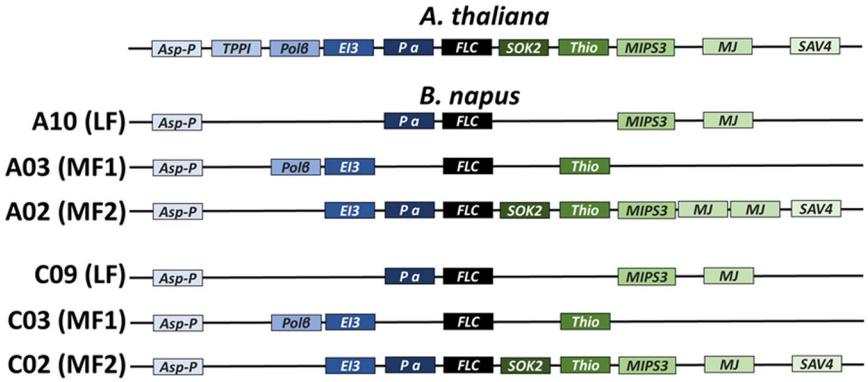


Figure 2. Graphical representation of local synteny for the syntenic *Bna.FLC* copies. The figure shows the gene order in the vicinity of *Ath.FLC, Bna.FLC.A10, Bna.FLC.A03a, Bna.FLC.A02, Bna.FLC.C09a, Bna.FLC.C03a* and *Bna.FLC.C02*, respectively. LF, MF1 and MF2 stand for least fractionated, more fractionated 1 and more fractionated 2, respectively. Synteny was derived from BRAD, gene names were derived from TAIR: Asp-P: aspartyl protease family protein, TPPI: TREHALOSE-6-PHOSPHATE PHOSPHATASE I, Polβ: DNA-directed RNA polymerase subunit beta, EI3: Ethylene insensitive 3 family protein, P a: Pollen Ole e 1 allergen and extensin family protein, SOK2 (no synonym given), Thio: Thioesterase superfamily protein, MIPS3: MYO-INOSITOL-1-PHOSPHATE SYNTHASE 3, MJ: Major facilitator superfamily protein, SAV4: SHADE AVOIDANCE 4.

Analysis of regulatory regions. We compared the sequences of three different regulatory regions using[42]: the promoter regions, intron 1 known to contain the long non-coding RNA *COLDAIR*, and the regions containing the antisense RNA *COOLAIR*.

The promoter regions (the upstream region of the gene extending until the next annotated gene) ranged from 736 bp (*Bna.FLC.C03a*) to 10,928 bp (*Bna.FLC.A03a*). The *Bna.FLC.C09b* promoter region was excluded from the analysis due to a significant amount of missing sequence data. Only a 26 bp motif (consensus **C**M**TGCG**RYR**CAC**R**TGGC**WR**TC**Y**TS**T**M**), located shortly before the 5′UTR of the gene, was highly conserved between the promoter regions of all 8 analysed copies, annotated as MA1359.1, the binding site for the transcription factor *BIGPETAL* (BPEp).The motif also overlapped with the *COOLAIR* binding site. Apart from this, motif search using MEME did not reveal a conclusive pattern. While the promoter regions of *Bna.FLC.A02* and *Bna.FLC.C02* were highly conserved both in length and motif content, the promoter regions of *Bna.FLC.A03a* and *Bna.FLC.A10* were remarkably longer than their C subgenome counterparts and motif composition was different. The promoter regions of *Bna.FLC.A03b* and *Bna.FLC.C03b* were comparable in length, but different in motif composition (Fig. 3). The promoter region of *Bna.FLC.A03b* carries a highly repetitive 29 bp motif (consensus **A**Y**TCGGACGAC**K**TATATTT**H**AGTCGT**Y**TG**) which is specific among all copies. No other motif was specific to a single promoter region. We have also retrieved the respective promoter regions from the diploid *Brassica* species *B. rapa, B. nigra* and *B.oleracea* as well as the promoter region of *A. thaliana*. The alignment shows that conservation is generally low among all 5 species, also given the strong variation in length (736 bp in *Bna.FLC.C03a* – 16946 bp in *Bni.FLC.B08b*). However, some few AT rich regions seem to be fairly conserved (Supplementary Fig. 3). When we aligned subgroups of promoter regions to the *A. thaliana* promoter region, we found that different parts of those regions were conserved within those subgroups (Supplementary Fig. 4). While the A02/C02 subgroup is more conserved towards the beginning and the end of the promoter region, the subgroup A03a/C03a is most conserved in the central part, and the subgroups A03b/C03b and A10/C09a towards the end of the region. The promoter regions from *B. nigra* showed the lowest conservation with *A. thaliana*, while the regions from the A10/C09a group showed the highest conservation with *A. thaliana*. Within *Bna.FLC.A10*, a region with a BLAST hit to the MITE transposon in front of Tapidor-*Bna.FLC.A10* reported by[39] was found to be unique among the promoter regions. When comparing motif composition in the orthologous copies using MEME excluding *B. nigra*, but including *Bna.FLC.C09b*, we found that motif composition is generally conserved between the orthologous promoter regions. Motif composition was fully conserved for the A02, the A03b, C03b and A10 copies, although distances and order were slightly varying. The A03a, C03a, C09a and C09b promoter regions varied slightly in motif composition, for example, the repetitive sequence found in *Bna.FLC.A03a* was not contained in *Bra.FLC.A03a*, indicating a later origin. Moreover, divergence between homeologs was also absent (A03b/C03b) or low (A02/C02, A03a/C03a), but high for A10/C09ab, indicating stronger divergence in regulatory motifs between the A10 and the C09 copies. The 26 bp motif overlapping with *COOLAIR* was also conserved among almost all species and copies, with the exception of the C03b promoter regions and *Bol.FLC.C02*. (Supplementary Fig. 5). The lengths of intron 1 ranged from 912 bp (*Bna.FLC.C02*) to 2,470 bp (*Bna.FLC.A10*) being considerably longer on A10/C09a and A03b/C03b, but intermediate on C09b. On A02/C02 and A03a/C03a, respectively, the first intron was missing large fragments of 1,000–1,500 bp at the 3′ end. When we aligned those to *A. thaliana COLDAIR* (HG975388.1), alignment was generally poor, making a direct comparison difficult.

Finally, we aligned the sequence of the long non-coding antisense transcript *COOLAIR* (GQ352646.1) to *Bna. FLC* and compared the overlapping regions. *COOLAIR* is supposed to cover parts of the 5′upstream region, exon 1, parts of intron 1, parts of intron 6, exon 7 and parts of the 3′downstream region. However, we were only able

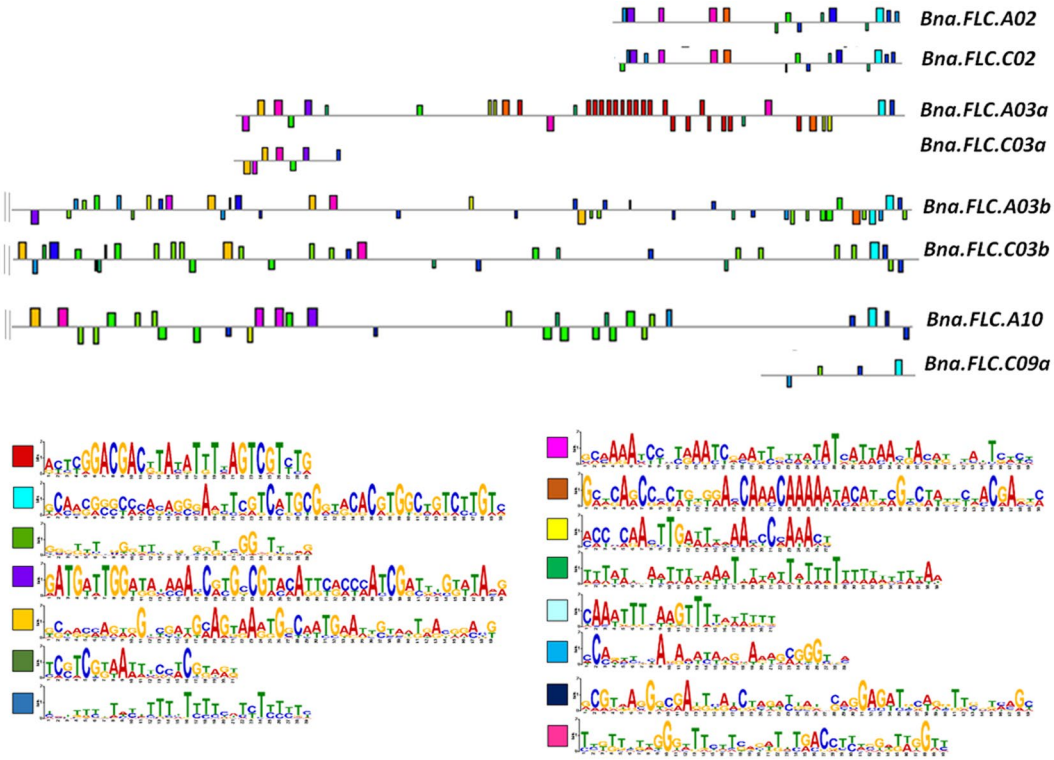


Figure 3. Motif composition of the upstream (promoter) regions of 8 analysed *Bna.FLC* gene copies done by MEME. The promoter region for *Bna.FLC.C09b* was excluded due to missing sequence information. Every motif is represented by a distinct color (see legend).

to match the 5′upstream region, exon 1 and parts of intron 1 to our sequences. Considerable variation was found at the 5′ end (3′ end of the antisense RNA) with *Bna.FLC.C02* showing an insert of 43 bp, while *Bna.FLC.A10* showed more sequence variation at this site. Both copies on C09, in contrast, were approximately 75 bp shorter at this end (Fig. 1B). While RNAseq data (see below) before vernalisation did not show clear evidence for *COOLAIR* expression, there is a pattern of directional reads in the time series of the spring type at *Bna.FLC.A10*, but coverage was too low for quantification.

***Bna.FLC* expression levels differentiate before vernalization.** We then studied the gene expression patterns of all *Bna.FLC* copies in five different accessions (3 winter and 2 spring cultivars) and found that only eight copies were expressed in leaves seven weeks after sowing (BBCH 14). *Bna.FLC.C03b* was never expressed. The highest expression levels were detected for *Bna.FLC.A10* and *Bna.FLC.C02* in winter accessions. Both copies showed a differential expression pattern between winter and spring accessions, whereby the average ratio of winter/spring expression level was 19.0 for *Bna.FLC.A10* and 2.8 for *Bna.FLC.C02*. The copy *Bna.FLC.A03b* showed no expression in the spring accessions, whereas *Bna.FLC.A02* showed considerably higher expression in Mansholt, a winter type without strict vernalisation requirement (Fig. 4).

We assessed *Bna.FLC* expression in petioles, emerging and developed leaves of 10 week old plants (BBCH 15) before vernalisation in a winter (Manitoba) and a spring accession (Korall) to detect tissue-specific patterns (Supplementary Fig. 6). *Bna.FLC.A03a* and *Bna.FLC.A10* were differentially expressed between winter and spring accessions in all tissues. *Bna.FLC.C02* expression was only significantly different in emerging leaves ($p = 0.021$) and petioles ($p = 0.008$), indicating differential expression dynamics in winter and spring material. Additionally, in petioles, *Bna.FLC.A03a* showed a significantly higher expression in Korall ($p = 0.037$), while *Bna.FLC.C09b* showed a significantly lower expression in Korall ($p = 0.019$).

For validation, we downloaded available RNAseq data published in[55]. Three week old leaves were sampled before vernalisation for the same accessions we have used here. Only *Bna.FLC.A03b* and *Bna.FLC.A10* were differentially expressed between both growth types. *Bna.FLC.C03b* expression was extremely low. In contrast to RT-qPCR, the expression of *Bna.FLC.A02, Bna.FLC.A03a, Bna.FLC.C02, Bna.FLC.C03a* and *Bna.FLC.C09b* was comparable to *Bna.FLC.A10*, but did not show differential expression between both growth types (Supplementary Fig. 7). Only *Bna.FLC.C09b* had a lower expression level. Those differences may partly be due to the different developmental stage and due to the high sequence similarity of the copies which may confound mapping.

We then analyzed a larger data set from the same study containing 30 spring and 30 winter accessions, using all winter accessions as biological replicates for "winter" and respectively so for "spring". We included other important vernalisation regulators (*FRI, SUF4, TFL2, VIN3, VRN2, SVP, SRR1*) and genes directly regulated by vernalisation (*SOC1, FD, FT, TEM1*). We found 3 gene copies which were strongly ($|\log_2(FC)| > 1$) and significantly ($p < 0.05$) upregulated in winter accessions. These were *Bna.FLC.A03b, Bna.FLC.A10* and *Bna.VRN2.Ann*. Three

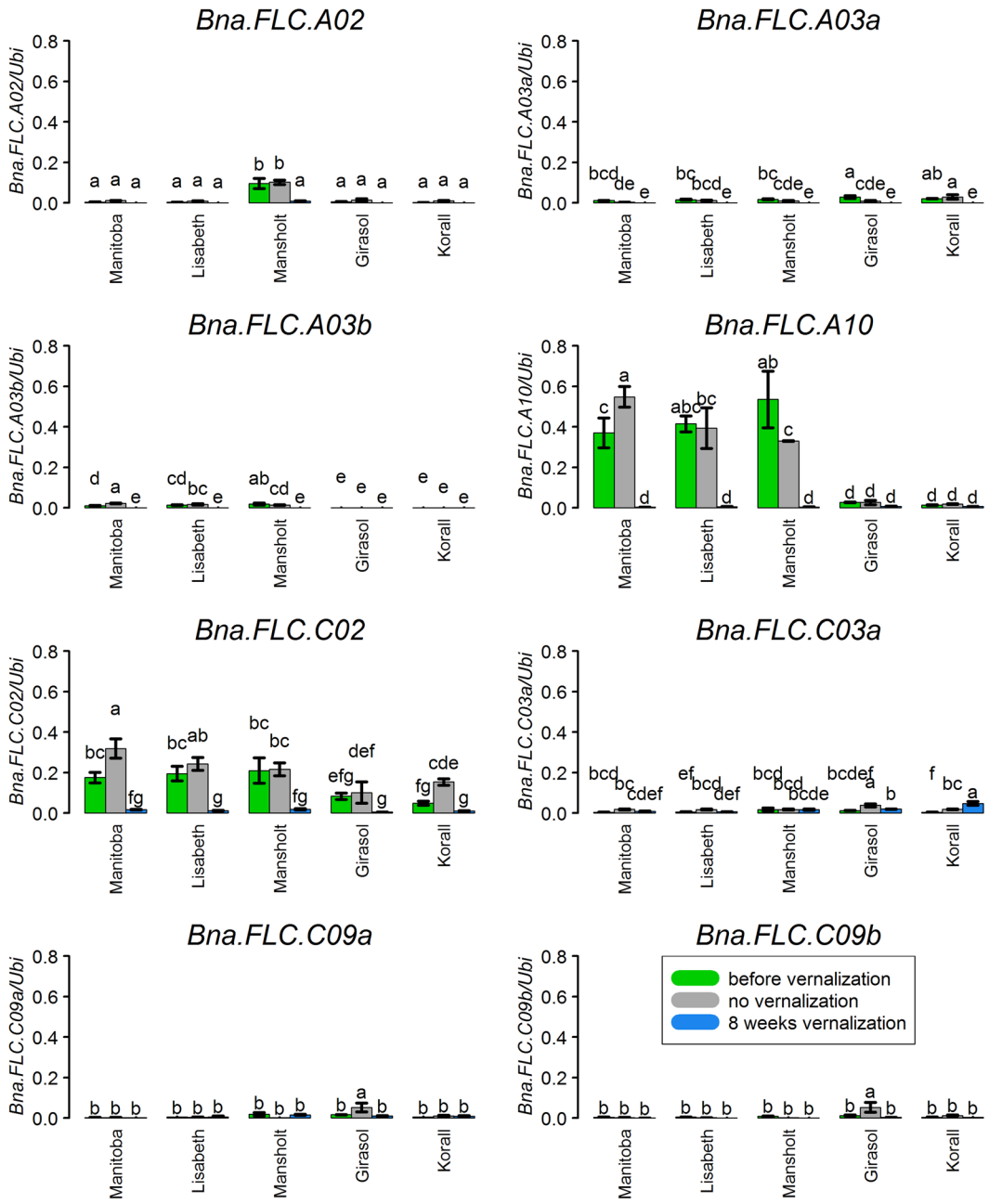


Figure 4. Relative gene expression of all eight expressed *Bna.FLC* copies in leaves before vernalisation 7 weeks after sowing (BBCH 14), and without and with vernalisation 15 weeks after sowing in the accessions Manitoba, Lisabeth (winter-type), Mansholt (winter-hardy, but not vernalisation-dependent), Girasol and Korall (both spring-type). The values were calculated from RT-qPCR using the ΔCt method and represent 3 biological replicates. Whiskers show SEM. "NE" stands for "not expressed". Small letters represent group by LSD test. The scale is the same for all 3 plots.

copies of the downstream gene *Bna.SOC1* (on A03, A05 and C04) and two copies of the central flowering time gene *Bna.FT* (on A02 and C02) were strongly and significantly downregulated in winter accessions compared to spring accessions. *Bna.VIN3, Bna.FT* and *Bna.FD* were expressed at an extremely low level, as expected (Fig. 5).

Most *Bna.FLC* copies are downregulated by cold. Without cold treatment, 15 weeks after sowing (BBCH20), *Bna.FLC* expression patterns did not show a clear pattern. Expression remained mostly constant, although single accessions show slightly decreased expression (*Bna.FLC.A03a, Bna.FLC.A03b, Bna.FLC.A10)* or slightly increased expression *(Bna.FLC.A03b, Bna.FLC.A10, Bna.FLC.C02, Bna.FLC.C03a)*. However, the changes did not affect the differential expression pattern between winter and spring types for *Bna.FLC.A10* (Fig. 4). We also used our primer set to study a time course in the winter type Manitoba for the two copies *Bna.FLC.A10* and *Bna.FLC.C02*. We have sampled independent plants of the same genotype every week from an age of 3 weeks to

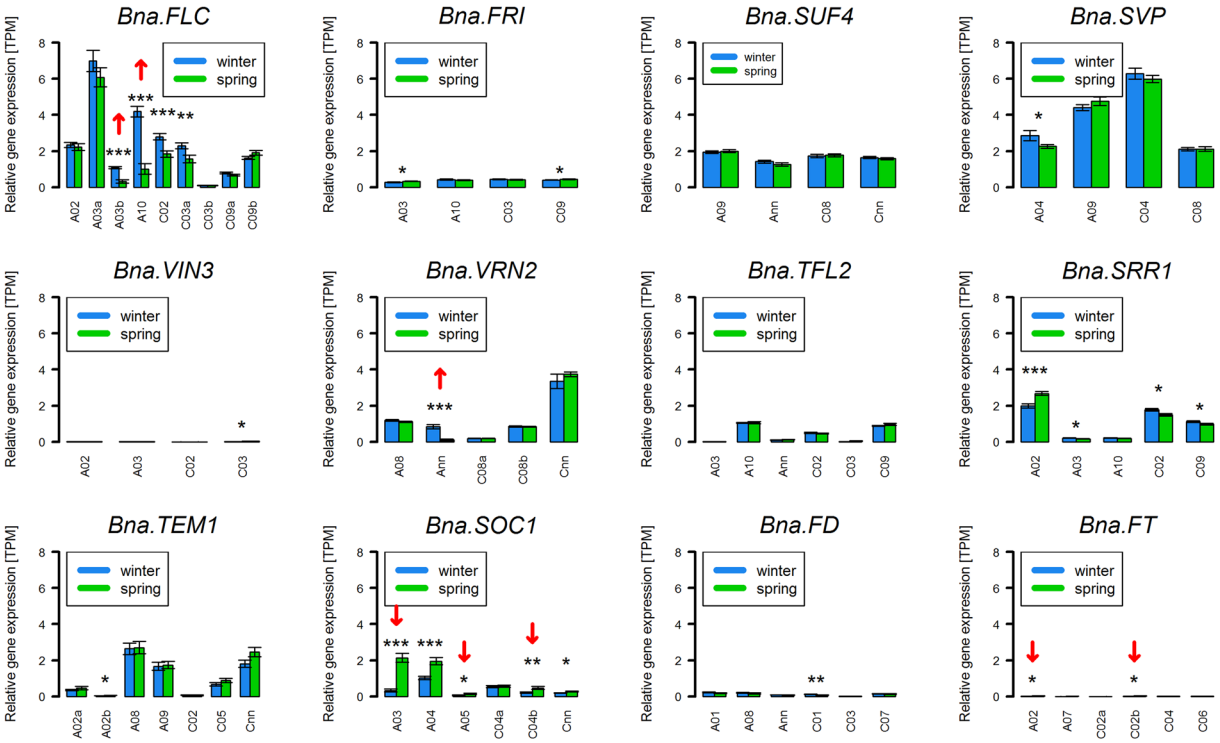


Figure 5. Normalized gene expressed levels (TPM) from RNAseq data for 12 different gene copy groups for genes involved in vernalisation (*FLC, FRI, SUF4, SVP, VIN3, VRN2, TFL2*), affecting *FLC* expression (*SRR1*) or affected by vernalisation (*TEM1, SOC1, FD, FT*) in a diverse set of 30 winter and 30 spring accessions. Whiskers represent standard errors of the mean (SEM). Asterisks show the level of significance based on the Student's t-test (*p-value < 0.05, **p-value < 0.01, ***p-value < 0.001). Red arrows indicate copies which were strongly ($|\log_2(FC)| > 1$) up or downregulated comparing winter and spring accessions. Arrows pointing upwards indicate augmented expression in winter accessions, arrows pointing downwards indicate lower expression in winter accessions.

8 weeks, when they were subjected to vernalisation. During vernalisation, we sampled at 4, 6 and 8 weeks of vernalisation. The results show clearly that the cold treatment is the only factor influencing gene expression for those copies, while the effect of age can be neglected (Fig. 6A). Longer vernalization seems to further decrease *Bna.FLC* levels, although those effects are only significant for *Bna.FLC.C02*.

After eight weeks of cold treatment (BBCH 20), most *Bna.FLC* copies were downregulated. Exceptions were *Bna.FLC.C03a*, which either did not change or even increased, and both *C09* copies, where downregulation was not significant. Downregulation was generally stronger in winter accessions than in spring accessions, partly due to the higher expression level before vernalisation (Fig. 4).

We also determined *Bna.FLC* and other vernalisation regulators' expression patterns in a time series in both leaf and apex in a publically available data set from the spring cultivar Westar. For each tissue, we analyzed five time points: 22 days (before vernalization), 43 and 64 days (during vernalization), and 67 and 72 days (after vernalization). The first sampling point was comparable to the sampling point of the other RNAseq data set, and the expression of all copies lay in the same range than the spring accessions from this diversity set. With the exception of *Bna.FLC.C02*, the expression patterns over time were similar in both tissue types. All copies with the exception of *Bna.FLC.C09a* were downregulated during cold. After cold, *Bna.FLC.A10* expression increased again to reach its original state, while both C09 copies even exceeded their original expression (Fig. 6B).

Analyzing the expression patterns of other vernalization regulators in this spring type cultivar, we found that all four copies of *Bna.FRI* got slightly upregulated and all four copies of *Bna.VIN3* got strongly upregulated during cold. *Bna.VIN3* levels returned to their original state after cold, while *Bna.FRI* levels remained almost constant. This indicates that *Bna.VIN3* expression reacts normally in spring types. Interestingly, four out of six *Bna.FT* copies were already fairly well expressed before vernalization, reflecting the spring type behavior, but got completely downregulated during cold treatment and strongly upregulated afterwards. This could possibly be due to a short day length regime in the cold room (Supplementary Figs 8–10).

Population-wide patterns for *Bna.FLC.A10* and *Bna.FLC.C02*. We then wanted to assess if the two most expressed copies (*Bna.FLC.A10* and *Bna.FLC.C02*) could be associated to growth type on a population-wide scale. For this, we have screened a large set of 101 accessions (53 winter and 48 spring accessions) for gene expression with RT-qPCR. We found that *Bna.FLC.A10* gene expression was highly significantly different between winter and spring accessions ($p < 1.8 \times 10^{-7}$), but *Bna.FLC.C02* gene expression was not. For *Bna.FLC.A10*, although the overall pattern was clear, some exceptions were observed in both winter and spring accessions (Fig. 7A). We

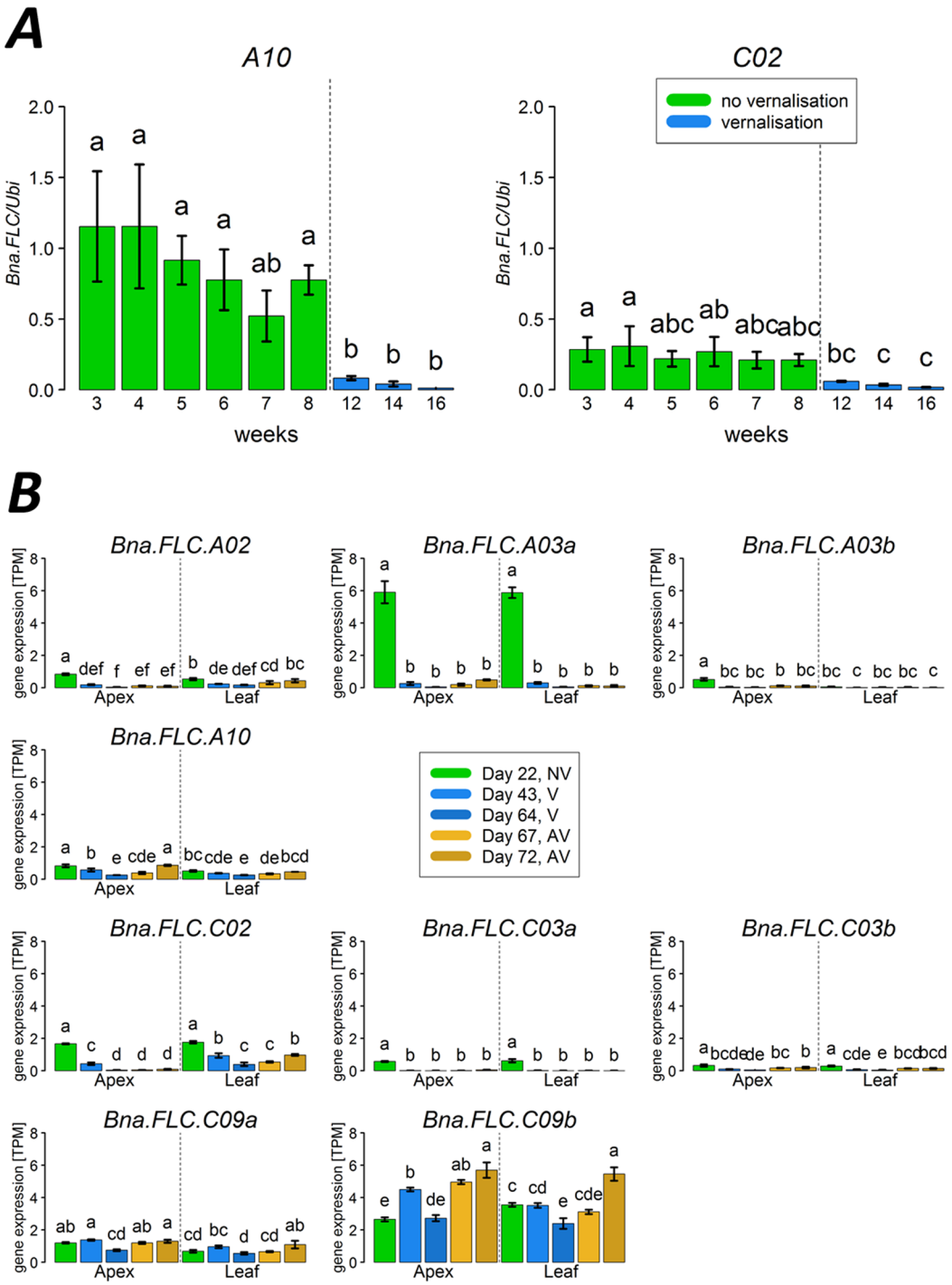


Figure 6. Time series of *Bna.FLC* expression. (**A**) Relative gene expression for *Bna.FLC.A10* and *Bna.FLC.C02* for a time series of 3–16 weeks in the winter cultivar Manitoba. (**B**) Normalized gene expressed levels (TPM) values from RNAseq from a publically available data set for the spring cultivar Westar. Whiskers show SEM. Small letters represent group by LSD test.

found four spring accessions with a high *Bna.FLC.A10* expression, while we found ten winter accessions with low *Bna.FLC.A10* expression. Although the population has been selected to reflect both early and late flowering accessions, there is no bias towards *Bna.FLC.A10* expression in early or late flowering accessions. To understand how vernalisation may be upheld or circumvent, we checked the RNAseq diversity set and found four accessions with unusual *FLC* pattern in both the spring and the winter pool. We found that in a winter background, accessions with lower *Bna.FLC.A10* have increased *Bna.SOC1* expression for a copy on A03 (still lower than spring types) and a copy on A05 (higher than spring types) and increased expression of *Bna.FT.A02* (but still considerably lower than spring types). At the same time, they have less *Bna.VIN3.A02*, even lower than spring accessions.

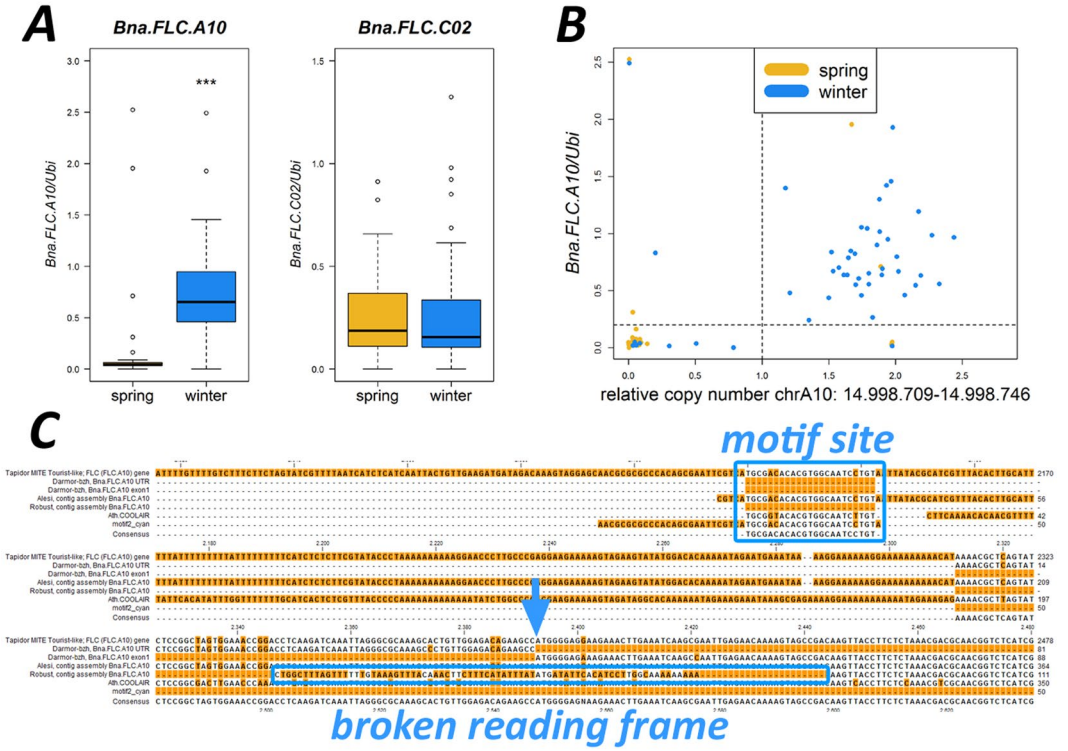


Figure 7. Relationships between the broken reading frame found to be correlating with *Bna.FLC.A10* expression. (**A**) Boxplots for relative gene expression of *Bna.FLC.A10* and *Bna.FLC.C02* in 48 spring and 53 winter accessions in leaves of 7 week old plants. (**B**) Relative copy number (RCN) of the region containing the sequence variant as a proxy for the presence of the sequence variant plotted against *Bna.FLC.A10* relative gene expression. The dotted lines separate "low" and "high" RCN/ gene expression. (**C**) Sequence alignment of the unaffected (accession Alesi) and affected (accession Robust) contig assembly with the Darmor-bzh-5′UTR, the Darmor-bzh exon 1, the Tapidor full length sequence from[39], the *A. thaliana COOLAIR* sequence and the sequence of the conserved 26 bp motif (indicated by blue frame). The site of the broken reading frame is also indicated. The blue arrow shows the ATG start site.

In a spring background, we found increased expression of *Bna.TEM1.C02* (higher than winter), but decreased expression of *Bna.FT.A02* (even lower than in winter), *Bna.VIN3.C02* (to 0) and *Bna.TFL2.C03* (approximately winter level).

To understand which causes could be responsible for this pattern, we went back to the resequencing data published in[41]. Here, we realized that a short fragment encompassing the R10P SNP was only covered by non-uniquely mapping reads in certain accessions, while perfectly covered with uniquely mapping reads in others. This points to a previously undiscovered short InDel. InDels of a size between 20–50 bp are hard to detect with short read sequencing, because the respective reads will be too divergent to be mapped. Therefore, we have extracted the reads from all *Bna.FLC* copies and reassembled them for both an affected and a non-affected accession. Aligning the resulting contigs with both the Tapidor and the Darmor-bzh version reveals that the start of the coding sequence is severely distorted in the affected accession, effectively breaking the open reading frame. The start codon is still present, but both the 5′UTR and the first exon are completely divergent from the non-affected sequence. This also affects the presence of the conserved 26 bp motif in front of the 5′UTR, which is missing in the affected accession, and the *COOLAIR* binding site. The first exon misses a fragment of 23 bp, representing a frameshift deletion (Fig. 7C). To assess this pattern for the total population, we took normalized coverage in the respective region (38 bp) as a proxy for this pattern and found that this is almost fully (two exceptions) explaining the observed expression pattern, indicating that either promoter variation or a broken reading frame of *Bna.FLC.A10* contributes to the winter-spring split (Fig. 7B).

Discussion

Regulatory diversification among *Bna.FLC* copies predates interspecific hybridization. *FLC* is a MADS-box transcription factor. The MADS-box genes represent a family of transcription factors regulating different developmental transitions, including flowering and fruit development[65]. Duplication and diversification played an important role in the evolution of the different clades in the MADS-box family[65,66]. One of the most important sources of duplication is polyploidy. Here, we analysed the expression patterns of all *FLC* homologs in *B. napus*, a recent allotetraploid which arose from two mesohexaploid ancestors. Due to the complex genetic history of *B. napus* including polyploidy and gene loss, a copy number between 4 and 5 is expected for single-copy *A. thaliana*[27,67]. *Bna.FLC*, however, was not affected by gene loss, but has rather undergone copy

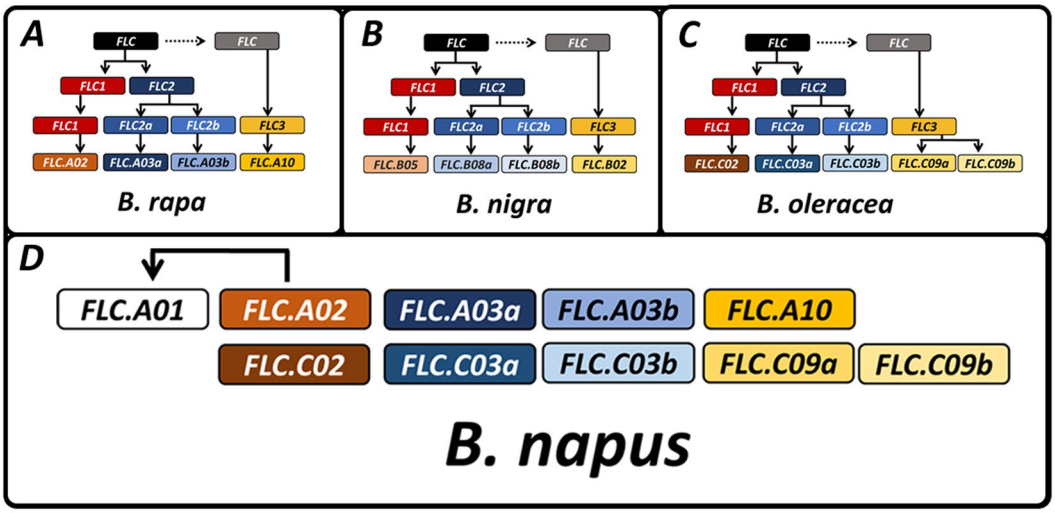


Figure 8. Proposed evolutionary path for *Bna.FLC*. The model was build based on synteny analysis and phylogenetic analysis. Pseudogenes were identified using protein sequence prediction and gene expression analysis. Gene copies are represented as colored rectangles. White color indicates a putative pseudogene. Arrows indicate ancestry. (**A–C**) Proposed evolutionary paths for the diploid *Brassica* species *B. rapa*, *B. nigra* and *B. oleracea*, respectively. (**D**) Copy configuration for the allotetraploid *B. napus*.

number amplification, as 10 loci could be identified as homologs of *Ath.FLC*. Based on our analysis, we now propose an evolutionary path for *Bna.FLC* (Fig. 8). Therein, we conclude that the duplication of the A03 and C03 copies predates the differentiation of *B. rapa, B.nigra* and *B. oleracea*, while the duplication of the C09 copies occurred after the AC split during *B. oleracea* speciation. *Bna.FLC.C03b* is considered to be a pseudogene, as we and others[29] never observed expression. The predicted exon structure and protein of *BnaFLC.C03b* differs from the other expressed copies (Fig. 1B,C), and its promoter misses conserved motifs which seem to be important for expression. We hypothesise that *BnaFLC.A01* occurred after the allopolyploidization, probably from the closest homolog *Bna.FLC.A02*, as no close homolog exists in any of the progenitors. The copies on A10 and C09 appear to be most conserved with *A. thaliana*, we therefore propose that the ancestor of this group is less recent and came from the second polyploidization step of the *Brassica* ancestor at the transition from the tetraploid to the hexaploid stage. Indeed, *Bna.FLC.A10* and *Bna.FLC.C09a* are both located within the least fractionated parts of the subgenomes[64], the youngest part of each subgenome. Given the observation that the copies which were most expressed in each subgenome are not homeologous to each other, subfunctionalisation is likely to have happened during the progenitors' evolution. This is supported by the high conservation of motif composition in the promoter regions of orthologous copies, indicating that evolution of regulatory elements predates the interspecific hybridization.

Different cold responsiveness points to subfunctionalisation of different *Bna.FLC* copies. *B. napus* is grown in diverse climates and therefore has developed different growth types. Due to the low genetic diversity within the growth type gene pools[68], suitable markers for growth type would help to speed up introgression breeding and allow for better and more effective adaptation. We therefore aimed to study the gene expression patterns of the most prominent candidate gene for vernalisation. A functional vernalisation system prevents pre-winter flowering in winter rapeseed, but is generally missing in spring rapeseed. *FLC* copies have been proposed as candidate genes for flowering time in previous QTL studies and GWAS, both in *B. napus* and in the progenitor species *B. oleracea* and *B. rapa*. In the progenitor species, *Bra.FLC.A10* and *Bol.FLC.C02* have been found to underlie QTL for flowering time[25,32,33,36]. In *B. napus*, *Bna.FLC.A10* has been proposed by different authors to be the most important vernalisation gene copy[28,39,41,49]. Others also found associations on C02 and A03 in different material[29,43,46]. Recently, it was shown that promoter variation for *Bna.FLC.A10* was linked to strong selection signatures between winter and spring material[49], but their data also show high variation in *Bna.FLC.A10* expression especially in winter accessions. *B. napus*, however, has 9 annotated gene copies and therefore underwent gene amplification rather than gene loss. We were therefore interested in the gene expression patterns of all other copies in *B. napus*. Studies on gene expression of other *Bna.FLC* copies have been carried out by other authors. Tadege *et al.*[28] was the first to qualitatively study *Bna.FLC* expression. They found that *At.FLC* delayed flowering in spring rapeseed, and that some copies of *Bna.FLC* also delayed flowering in *A. thaliana*, with *Bna.FLC.A10* having the strongest effect. They also showed a strong difference between total *Bna.FLC* level between two spring and two winter accessions by Northern blotting, all aligning well with our findings. Moreover, they found *Bna.FLC.A10* being more expressed in stems than in leaves, a finding we could not confirm. Hou *et al.*[39] studied a large BC_5F_2 population derived from a cross between Tapidor (a winter type) and Ningyou7 (a semi-winter type). They could show that *Bna.FLC.A10* downregulation was much quicker in semi-winter than in winter type, but original *Bna.FLC.A10* expression was high in both accessions, indicating that the mechanism separating semi-winter and winter might be different from the mechanism separating winter and spring, as we have found the *Bna.*

FLC:A10 expression is constantly low in spring types even before vernalization (Figs 4, 6). The authors found a 621 bp MITE insertion upstream of *Bna.FLC.A10* being associated to this lower responsiveness. They did not find allelic variance between the winter and the semi-winter type and concluded that the difference between winter and semi-winter was entirely due to the expression difference. Raman 2016 studied *Bna.FLC* expression unspecifically for the pairs *FLC1* (*Bna.FLC.A10/C09ab*), *FLC2* (*Bna.FLC.A02/C02*) and *FLC3* (*Bna.FLC.A03a/C03a*) in Australian and Japanese material[46] and found that *FLC2* expression was correlating to vernalization response, while *FLC1* and *FLC3* did not. However, we could not find a study showing expression patterns for all *Bna.FLC* copies in both winter and spring accessions to allow conclusions on possible subfunctionalisation. Therefore, we designed specific primers for all *Bna.FLC* copies and use them to test gene expression both in winter and spring accessions. We found strongly varying expression patterns in respect to expression level, response to cold and specificity to growth type, indicating a high degree of subfunctionalisation. The differential expression of *Bna.FLC.A10* between winter and spring as found by others[49] has been confirmed. At the same time, we also found *Bna.FLC.A03b* differentially expressed between winter and spring material. *Bna.FLC.C02*, however, was found to be differentially expressed in emerging leaves, but not in developed leaves, and population screening showed either a small (RNAseq data) or no difference (RT-qPCR). Moreover, the copy was shown to be deleted both in winter and spring accessions[53], which in most cases also correlated with the expression pattern (data not shown), but had obviously no effect on life cycle characteristics. Both copies on C09 and partly *Bna.FLC.C03a* were found not to respond to cold, and *Bna.FLC.C03b* seems to be pseudogene as reported by others[29]. *Bna.FLC.A02, Bna.FLC.A03a* and *partly Bna.FLC.C03a* are expressed and react to cold, but do not show differences in expression between winter and spring. This patterns indicates that *Bna.FLC.A10* is indeed the major source of variation in vernalisation requirement, while *Bna.FLC.A03b* and *Bna.FLC.C02* might still be supportive.

Accessions with low *Bna.FLC.A10* expression have a broken reading frame. When screening a population of 101 accessions for *Bna.FLC.A10*, we could show that winter accessions express *Bna.FLC.A10* significantly stronger than spring accessions. However, our data along with others[39,49] show that there are winter accessions with a very low *Bna.FLC.A10* level, which still need vernalisation to flower, and spring accessions with high *Bna.FLC.A10* levels which can flower without. Reassessment of deep-sequencing data on this copy allowed us to detect a sequence variant at the junction of the 5′UTR and exon 1 which was almost fully explanatory for *Bna.FLC.A10* level. The variant overlapped with the binding site of the regulatory antisense RNA and also with a highly conserved 26 bp promoter motif shortly before the 5′UTR. We are therefore now able to describe the nature of the promoter variation which has been proposed by others before[39,49]. This also aligns with other findings, for example, the shorter 3′ end of *COOLAIR* in both C09 copies which do not react to cold. A similar mechanism has been found in *A. thaliana*, where variation in intron 1 caused a different splicing behavior of *COOLAIR*, increasing *FLC* expression[69]. While we now are able to explain the expression patterns in *Bna.FLC.A10*, accessions with unusual expression patterns indicate that *Bna.FLC.A10* might be important, but not necessary to define the life cycle characteristics of winter and spring rapeseed. In winter accessions with low *Bna.FLC.A10*, we see upregulation of some downstream targets like *Bna.SOC1* and *Bna.FT*. In spring accessions with high *Bna.FLC.A10*, we also see a decrease in downstream targets. This could indicate that those exceptions might also have a respective phenotype, but this is not the case, as both early and late flowering accessions are among those exceptional accessions. Therefore, some other genes seem to compensate the effect, but obviously not a different *Bna.FLC* copy. In spring accessions, *Bna.TEM1.C02* is possibly serving this task, as it is known to be an early suppressor of *FT* in *A. thaliana*[70]. Apart from those exceptions, the analysis of the RNAseq data has shown that *Bna.VRN2.Ann* is also downregulated in spring accessions. Other vernalisation genes are not or not strongly differentially expressed at this time point. At the same time, copies of important downstream targets like *Bna.FT* (A02, C02) and *Bna.SOC1* (A03, A05, C04) are significantly downregulated in winter accessions, indicating that the vernalisation system indeed is active, although *Bna.FT* expression levels remain low in spring accessions, probably due to non-inductive photoperiod.

Conclusions

Our data support earlier findings showing that *Bna.FLC.A10* is the most important copy of *FLC* regulating vernalisation in winter-type *B. napus*. Moreover, we were able to show that this expression pattern is linked to a sequence variant at the *COOLAIR* binding site. However, *Bna.FLC.A10* expression level is not fully predictive of growth type, and more research needs to be done on other factors involved in vernalisation. Accessions showing exceptional patterns of *Bna.FLC.A10* might be useful for such an approach, as they represent functional mutants in the respective background. Here, those mutants have shown that other *Bna.FLC* copies do not compensate lack or overexpression of *Bna.FLC.A10*. The question remains why *Bna.FLC* copy number increased so strongly if additional homologs are not needed for vernalisation. The answer may lie in the assumption of different regulatory roles besides vernalisation. Our finding that some copies are obviously no longer cold-responsive suggests this to be the case, and is reflected by the large variation in putative regulatory regions between the copies. Together these findings indicate beginning gene copy subfunctionalisation, implying ongoing functional diversification in this recent allopolyploid.

Data availability

All primer sequences are available from Table S3.

Received: 15 January 2019; Accepted: 26 September 2019;

Published online: 17 October 2019

References

1. Blümel, M., Dally, N. & Jung, C. Flowering time regulation in crops-what did we learn from *Arabidopsis*? *Curr Opin Biotechnol* **32**, 121–129, https://doi.org/10.1016/j.copbio.2014.11.023 (2015).
2. Song, J., Irwin, J. & Dean, C. Remembering the prolonged cold of winter. *Curr Biol* **23**, R807–R811, https://doi.org/10.1016/j.cub.2013.07.027 (2013).
3. Sheldon, C. C., Rouse, D. T., Finnegan, E. J., Peacock, W. J. & Dennis, E. S. The molecular basis of vernalization: the central role of *FLOWERING LOCUS C* (*FLC*). *PNAS* **97**, 3753–3758, https://doi.org/10.1073/pnas.060023597 (2000).
4. Srikanth, A. & Schmid, M. Regulation of flowering time: all roads lead to Rome. *Cell Mol Life Sci* **68**, 2013–2037, https://doi.org/10.1007/s00018-011-0673-y (2011).
5. Deng, W. *et al*. *FLOWERING LOCUS C* (*FLC*) regulates development pathways throughout the life cycle of *Arabidopsis*. *PNAS* **108**, 6680–6685, https://doi.org/10.1073/pnas.1103175108 (2011).
6. Searle, I. *et al*. The transcription factor *FLC* confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. *Genes Dev* **20**, 898–912 (2006).
7. Mateos, J. L. *et al*. Combinatorial activities of *SHORT VEGETATIVE PHASE* and *FLOWERING LOCUS C* define distinct modes of flowering regulation in *Arabidopsis*. *Genome Biol* **16**, 31, https://doi.org/10.1186/s13059-015-0597-1 (2015).
8. Abe, M. *et al*. *FD*, a bZIP Protein Mediating Signals from the Floral Pathway Integrator *FT* at the Shoot Apex. *Science* **309**, 1052–1056, https://doi.org/10.1126/science.1113095 (2005).
9. Simpson, G. G. The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. *Curr Opin Plant Biol* **7**, 570–574, https://doi.org/10.1016/j.pbi.2004.07.002 (2004).
10. Johansson, M. & Staiger, D. *SRR1* is essential to repress flowering in non-inductive conditions in *Arabidopsis thaliana*. *Ex Bot J* **65**, 5811–5822, https://doi.org/10.1093/jxb/eru317 (2014).
11. He, Y. Control of the transition to flowering by chromatin modifications. *Mol Plant* **2**, 554–564, https://doi.org/10.1093/mp/ssp005 (2009).
12. Turck, F. & Coupland, G. When vernalization makes sense. *Science* **331**, 36–37, https://doi.org/10.1126/science.1200786 (2011).
13. Crevillén, P. *et al*. Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. *Nature*; https://doi.org/10.1038/nature13722 (2014).
14. Choi, K. *et al*. The *FRIGIDA* complex activates transcription of *FLC*, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors. *The Plant Cell* **23**, 289–303, https://doi.org/10.1105/tpc.110.075911 (2011).
15. Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331**, 76–79 (2011).
16. Wood, C. C. *et al*. The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes *VERNALIZATION INSENSITIVE 3*. *PNAS* **103**, 14631–14636 (2006).
17. Oh, S., Park, S. & van Nocker, S. Genic and global functions for *Paf1C* in chromatin modification and gene expression in *Arabidopsis*. *PLoS genet* **4**, e1000077, https://doi.org/10.1371/journal.pgen.1000077 (2008).
18. Exner, V. *et al*. The Chromodomain of *LIKE HETEROCHROMATIN PROTEIN 1* Is Essential for H3K27me3 Binding and Function during *Arabidopsis* Development. *PLoS ONE* **4**, e5335, https://doi.org/10.1371/journal.pone.0005335 (2009).
19. Mylne, J. S. *et al*. *LHP1*, the *Arabidopsis* homologue of *HETEROCHROMATIN PROTEIN1*, is required for epigenetic silencing of *FLC*. *PNAS* **103**, 5012–5017, https://doi.org/10.1073/pnas.0507427103 (2006).
20. Mozgova, I., Köhler, C. & Hennig, L. Keeping the gate closed: functions of the polycomb repressive complex *PRC2* in development. *PlantJ* **83**, 121–132, https://doi.org/10.1111/tpj.12828 (2015).
21. Dittmar, E. L., Oakley, C. G., Agren, J. & Schemske, D. W. Flowering time QTL in natural populations of *Arabidopsis thaliana* and implications for their adaptive value. *Mol Ecol* **23**, 4291–4303, https://doi.org/10.1111/mec.12857 (2014).
22. Gazzani, S., Gendall, A. R., Lister, C. & Dean, C. Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiol* **132**, 1107–1114, https://doi.org/10.1104/pp.103.021212 (2003).
23. Mendez-Vigo, B., Pico, F. X., Ramiro, M., Martinez-Zapater, J. M. & Alonso-Blanco, C. Altitudinal and Climatic Adaptation Is Mediated by Flowering Traits and *FRI*, *FLC*, and *PHYC* Genes in *Arabidopsis*. *Plant Physiol* **157**, 1942–1955, https://doi.org/10.1104/pp.111.183426 (2011).
24. Camargo, L. E. A. & Osborn, T. C. Mapping loci controlling flowering time in *Brassica oleracea*. *Theor Appl Genet* **92**, 610–616, https://doi.org/10.1007/s001220050170 (1996).
25. Okazaki, K. *et al*. Mapping and characterization of *FLC* homologs and QTL analysis of flowering time in *Brassica oleracea*. *Theor Appl Genet* **114**, 595–608, https://doi.org/10.1007/s00122-006-0460-6 (2007).
26. Osborn, T. C. *et al*. Comparison of Flowering Time Genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. *Genetics* 1123–1129 (1997).
27. Schiessl, S., Samans, B., Hüttel, B., Reinhardt, R. & Snowdon, R. J. Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. *FrontPlant Sci* **5**, 404, https://doi.org/10.3389/fpls.2014.00404 (2014).
28. Tadege, M. *et al*. Control of flowering time by *FLC* orthologues in *Brassica napus*. *Plant J* **28** (2001).
29. Zou, X. *et al*. Comparative Analysis of *FLC* Homologues in *Brassicaceae* Provides Insight into Their Role in the Evolution of Oilseed Rape. *PLoS ONE* **7**, e45751, https://doi.org/10.1371/journal.pone.0045751 (2012).
30. Irwin, J. A. *et al*. Functional alleles of the flowering time regulator *FRIGIDA* in the *Brassica oleracea* genome. *BMC plant biol* **12**, 21, https://doi.org/10.1186/1471-2229-12-21 (2012).
31. Wang, N. *et al*. Flowering time variation in oilseed rape (*Brassica napus* L.) is associated with allelic variation in the *FRIGIDA* homologue *BnaA.FRI.a*. *Ex Bot J* **62**, 5641–5658, https://doi.org/10.1093/jxb/err249 (2011).
32. Wu, J. *et al*. A naturally occurring InDel variation in *BraA.FLC.b* (*BrFLC2*) associated with flowering time variation in *Brassica rapa*. *BMC plant biol* **12**, 151, https://doi.org/10.1186/1471-2229-12-151 (2012).
33. Yuan, Y.-X. *et al*. A naturally occurring splicing site mutation in the *Brassica rapa FLC1* gene is associated with variation in flowering time. *Ex Bot J* **60**, 1299–1308, https://doi.org/10.1093/jxb/erp010 (2009).
34. Zhao, J. *et al*. *BrFLC2* (*FLOWERING LOCUS C*) as a candidate gene for a vernalization response QTL in *Brassica rapa*. *Ex Bot J* **61**, 1817–1825, https://doi.org/10.1093/jxb/erq048 (2010).
35. Xiao, D. *et al*. The *Brassica rapa FLC* homologue *FLC2* is a key regulator of flowering time, identified through transcriptional co-expression networks. *Ex Bot J* **64**, 4503–4516, https://doi.org/10.1093/jxb/ert264 (2013).
36. Irwin, J. A. *et al*. Nucleotide polymorphism affecting *FLC* expression underpins heading date variation in horticultural brassicas. *Plant J* **87**, 597–605, https://doi.org/10.1111/tpj.13221 (2016).
37. Razi, H., Howell, E. C., Newbury, H. J. & Kearsey, M. J. Does sequence polymorphism of *FLC* paralogues underlie flowering time QTL in *Brassica oleracea*? *Theor Appl Genet* **116**, 179–192, https://doi.org/10.1007/s00122-007-0657-3 (2008).
38. Ridge, S., Brown, P. H., Hecht, V., Driessen, R. G. & Weller, J. L. The role of *BoFLC2* in cauliflower (*Brassica oleracea* var. *botrytis* L.) reproductive development. *Ex Bot J* **66**, 125–135, https://doi.org/10.1093/jxb/eru408 (2015).
39. Hou, J. *et al*. A Tourist-like MITE insertion in the upstream region of the *BnFLC.A10* gene is associated with vernalization requirement in rapeseed (*Brassica napus* L.). *BMC Plant Biol* **12**, 238, https://doi.org/10.1186/1471-2229-12-238 (2012).
40. Pires, J. C. *et al*. Flowering time divergence and genomic rearrangements in resynthesized Brassica polyploids (Brassicaceae). *Biol J Linn Soc* **82**, 675–688 (2004).

41. Schiessl, Hüttel, B., Kuehn, D., Reinhardt, R. & Snowdon, R. J. Post-polyploidisation morphotype diversification associates with gene copy number variation. *Sci Rep* 41845, https://doi.org/10.1038/srep41845 (2017).
42. Chalhoub, B. *et al*. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950–953, https://doi.org/10.1126/science.1253435 (2014).
43. Fletcher, R. S., Mullen, J. L., Heiliger, A. & McKay, J. K. QTL analysis of root morphology, flowering time, and yield reveals trade-offs in response to drought in *Brassica napus*. *Ex Bot J*; https://doi.org/10.1093/jxb/eru423 (2014).
44. Nelson, M. N. *et al*. Quantitative trait loci for thermal time to flowering and photoperiod responsiveness discovered in summer annual-type *Brassica napus* L. *PLoS ONE* **9**, e102611, https://doi.org/10.1371/journal.pone.0102611 (2014).
45. Quijada, P. A., Udall, J. A., Lambert, B. & Osborn, T. C. Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 1. Identification of genomic regions from winter germplasm. *Theor Appl Genet* **113**, 549–561, https://doi.org/10.1007/s00122-006-0323-1 (2006).
46. Raman, H. *et al*. Genome-wide association analyses reveal complex genetic architecture underlying natural variation for flowering time in canola. *Plant Cell Environ* **39**, 1228–1239, https://doi.org/10.1111/pce.12644 (2016).
47. Raman, H. *et al*. Genetic and physical mapping of flowering time loci in canola (*Brassica napus* L.). *Theor Appl Genet* **126**, 119–132, https://doi.org/10.1007/s00122-012-1966-8 (2013).
48. Udall, J. A., Quijada, P. A., Lambert, B. & Osborn, T. C. Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed *(Brassica napus* L.): 2. Identification of alleles from unadapted germplasm. *Theor Appl Genet* **113**, 597–609, https://doi.org/10.1007/s00122-006-0324-0 (2006).
49. Wu, D. *et al*. Whole-genome resequencing of a world-wide collection of rapeseed accessions reveals genetic basis of their ecotype divergence. *Mol Plant*. https://doi.org/10.1016/j.molp.2018.11.007 (2018).
50. Wang, X. *et al*. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* **43**, 1035–1039, https://doi.org/10.1038/ng.919 (2011).
51. Parkin, I. A. P. *et al*. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biol* (2014).
52. Khan, A. *et al*. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Research* **46**, D260–D266, https://doi.org/10.1093/nar/gkx1126 (2018).
53. Schiessl, H. B., Kuehn, D., Reinhardt, R. & Snowdon, R. J. Targeted deep sequencing of flowering regulators in *Brassica napus* reveals extensive copy number variation. *Sci data* **4**, 170013, https://doi.org/10.1038/sdata.2017.13 (2017).
54. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408, https://doi.org/10.1006/meth.2001.1262 (2001).
55. Havlickova, L. *et al*. Validation of an updated Associative Transcriptomics platform for the polyploid crop species *Brassica napus* by dissection of the genetic architecture of erucic acid and tocopherol isoform variation in seeds. *PlantJ* **93**, 181–192, https://doi.org/10.1111/tpj.13767 (2018).
56. Bus, A., Korber, N., Snowdon, R. J. & Stich, B. Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. *Theor Appl Genet* **123**, 1413–1423, https://doi.org/10.1007/s00122-011-1676-7 (2011).
57. Körber, N. *et al*. Seedling development in a *Brassica napus* diversity set and its relationship to agronomic performance. *Theor and Appl Genet* **125**, 1275–1287, https://doi.org/10.1007/s00122-012-1912-9 (2012).
58. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, https://doi.org/10.1093/bioinformatics/btu170 (2014).
59. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9**, 357–359 (2012).
60. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, https://doi.org/10.1093/bioinformatics/btp352 (2009).
61. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr prot bioinf* **47**, 11, https://doi.org/10.1002/0471250953.bi1112s47 (2014).
62. Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. **131**, 281–285, https://doi.org/10.1007/s12064-012-0162-3 (2012).
63. Schranz, M. E. *et al*. Characterization and Effects of the Replicated Flowering Time Gene *FLC* in *Brassica rapa*. *Genetics* (2002).
64. Cheng, F., Wu, J., Fang, L. & Wang, X. Syntenic gene analysis between *Brassica rapa* and other *Brassicaceae* species. *Front plant science* **3**, 198, https://doi.org/10.3389/fpls.2012.00198 (2012).
65. Becker, A. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogen Evol* **29**, 464–489, https://doi.org/10.1016/S1055-7903(03)00207-0 (2003).
66. Dreni, L. & Kater, M. M. MADS reloaded: evolution of the AGAMOUS subfamily genes. *New Phytol* **201**, 717–732, https://doi.org/10.1111/nph.12555 (2014).
67. Parkin, I. A. P. *et al*. Towards unambiguous transcript mapping in the allotetraploid *Brassica napus*. *Genome* **53**(11), 929–38, https://doi.org/10.1139/G10-053. (2010).
68. Snowdon, R. J. & Iniguez Luy, F. L. Potential to improve oilseed rape and canola breeding in the genomics era. *Plant Breeding* **131**, 351–360, https://doi.org/10.1111/j.1439-0523.2012.01976.x (2012).
69. Li, P., Tao, Z. & Dean, C. Phenotypic evolution through variation in splicing of the noncoding RNA *COOLAIR*. *Genes Dev* **29**, 696–701, https://doi.org/10.1101/gad.258814.115 (2015).
70. Matias-Hernandez, L., Aguilar-Jaramillo, A. E., Marin-Gonzalez, E., Suarez-Lopez, P. & Pelaz, S. *RAV genes: regulation of floral induction and beyond.* **114**, 1459–1470, https://doi.org/10.1093/aob/mcu069 (2014).

Acknowledgements

The authors like to thank the Justus Liebig University for granting SVS with financial support for this project. The authors also thank Iulian Gabor for technical support in the lab and Franziska Hintermeier for sampling the Manitoba time series. We also thank Huey Tyng Lee for assembling the *Bna.FLC.A10* sequences.

Author contributions

S.V.S., L.Q. and R.J.S. conceived the project. S.V.S. conducted the bioinformatics analysis of the RNAseq data, the phylogenetic analysis and the analysis of the regulatory regions. D.Q.M. and E.T. conducted the RT-qPCR. All authors have contributed in writing the manuscript and have read and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-51212-x.

Correspondence and requests for materials should be addressed to S.V.S. or L.Q.

Chapter 2.3: Different copies of *SENSITIVITY TO RED LIGHT REDUCED 1* show strong subfunctionalization in *Brassica napus*

Schiessl S, Williams N, Specht P, Staiger D, Johansson M (2019) Different copies of *SENSITIVITY TO RED LIGHT REDUCED 1* show strong subfunctionalization in *Brassica napus*. BMC Plant Biol 19:372. doi: 10.1186/s12870-019-1973-x

Schiessl et al. BMC Plant Biology (2019) 19:372
https://doi.org/10.1186/s12870-019-1973-x

BMC Plant Biology

RESEARCH ARTICLE Open Access

Different copies of SENSITIVITY TO RED LIGHT REDUCED 1 show strong subfunctionalization in *Brassica napus*

Sarah Schiessl[2†], Natalie Williams[1†], Pascal Specht[2], Dorothee Staiger[1] and Mikael Johansson[1*]

Abstract

Background: Correct timing of flowering is critical for plants to produce enough viable offspring. In *Arabidopsis thaliana* (Arabidopsis), flowering time is regulated by an intricate network of molecular signaling pathways. Arabidopsis *srr1–1* mutants lacking SENSITIVITY TO RED LIGHT REDUCED 1 (SRR1) expression flower early, particularly under short day (SD) conditions (1). SRR1 ensures that plants do not flower prematurely in such non-inductive conditions by controlling repression of the key florigen *FT*. Here, we have examined the role of SRR1 in the closely related crop species *Brassica napus*.

Results: Arabidopsis SRR1 has five homologs in *Brassica napus*. They can be divided into two groups, where the A02 and C02 copies show high similarity to AtSRR1 on the protein level. The other group, including the A03, A10 and C09 copies all carry a larger deletion in the amino acid sequence. Three of the homologs are expressed at detectable levels: *A02*, *C02* and *C09*. Notably, the gene copies show a differential expression pattern between spring and winter type accessions of *B. napus*. When the three expressed gene copies were introduced into the *srr1–1* background, only *A02* and *C02* were able to complement the *srr1–1* early flowering phenotype, while *C09* could not. Transcriptional analysis of known SRR1 targets in *Bna.SRR1*-transformed lines showed that *CYCLING DOF FACTOR 1 (CDF1)* expression is key for flowering time control via SRR1.

Conclusions: We observed subfunctionalization of the *B. napus SRR1* gene copies, with differential expression between early and late flowering accessions of some *Bna.SRR1* copies. This suggests involvement of *Bna.SRR1* in regulation of seasonal flowering in *B. napus*. The *C09* gene copy was unable to complement *srr1–1* plants, but is highly expressed in *B. napus*, suggesting specialization of a particular function. Furthermore, the C09 protein carries a deletion which may pinpoint a key region of the SRR1 protein potentially important for its molecular function. This is important evidence of functional domain annotation in the highly conserved but unique SRR1 amino acid sequence.

Keywords: *Arabidopsis*, *Brassica napus*, Flowering, Cross-species functionality, Subfunctionalization

Background

Plants need to synchronize their reproductive activity to the optimal growth season, to ensure maximal reproductive output. Consequently, onset of flowering is tightly controlled by a network of signals originating from developmental, as well as environmental signaling pathways [2–4]. After reaching a critical developmental age, plants will respond to favorable environmental stimuli and flowering will be initiated [5]. In long day (LD) plants, flowering is promoted in spring and summer when the days are longer than the nights. Day length is measured by the inner circadian clock that maintains a ca 24-h cyclic rhythm of gene and protein expression of clock components that in turn regulate downstream processes. When light coincides with the expression of components of the so-called photoperiodic pathway of flower induction, expression of "florigen" *FLOWERING LOCUS T (FT)* is promoted in the leaves [6, 7]. FT protein then travels through the vasculature to the shoot apex where flower formation is initiated [8–10]. CONSTANS (CO) is a key signal

* Correspondence: mikael.johansson@uni-bielefeld.de
†Sarah Schiessl and Natalie Williams contributed equally to this work.
[1]RNA Biology and Molecular Physiology, Faculty for Biology, Bielefeld University, Universitaetsstrasse 25, 33615 Bielefeld, Germany
Full list of author information is available at the end of the article

integrator for photoperiodic flowering. Its transcription is controlled by the circadian clock through the GIGANTEA (GI) clock component that interacts with FLAVIN BINDING, KELCH REPEAT, F-Box 1 (FKF1) in coincidence with light. FKF1 then represses the activity of CDF transcription factors, which have a repressive role on *CO* expression [11–13]. This allows accumulation of *CO* transcript in the afternoon and CO protein expression. CO in turn promotes expression of *FT* by binding to its promoter and thus initiating flowering. Transcription of *FT* is also tightly regulated by both promotive and repressive elements that integrate signals from various environmental and developmental signaling pathways [2]. An important *FT* repressor in this transcriptional landscape is the MADS box transcription factor FLOWERING LOCUS C (FLC), which has an important role as a repressor of flowering in unfavorable conditions, as its expression level is reduced by extended periods of cold [14, 15].

The main genetic factors of the flowering time regulation network have been conserved throughout *Brassicaceae*, as revealed by genome sequencing in recent years [16–20]. This conservation indicates that their function might be similar as in the model species Arabidopsis. Additionally, many quantitative trait loci (QTL) studies and genome-wide association studies (GWAS) for flowering time have found homologs of Arabidopsis flowering time genes in the confidence intervals of associated markers [21–29]. However, the most important crop plants from the *Brassicaceae* come from the genus *Brassica*, including important vegetable species like cabbage, cauliflower (*Brassica oleracea*), Chinese cabbage (*Brassica rapa*), but also the important oilseed crop rapeseed (*Brassica napus*). *Brassica* species share a whole-genome triplication, and *B. napus* arose from a recent interspecific hybridization between *B. rapa* (A subgenome donor) and *B. oleracea* (C subgenome donor), expanding the theoretically expected copy number of Arabidopsis homologs in allotetraploid *B. napus* to 6 (Brassica triplication × 3, hybridization × 2) [30, 31]. After polyploidization, many different processes like homologous recombination and the action of transposable elements led to a strong genome reorganization. Together with selective processes, this reorganization individually changed the specific gene copy numbers, now varying between 1 and 12, and possibly varying between individuals [16, 32, 33]. In the course of evolution, single copies might evolve differently and give rise to new expression patterns or functions through a process called subfunctionalization [34]. The degree of subfunctionalization is gene specific. Subfunctionalization has played an important role in evolution of flowering time control [35, 36].

SENSITIVITY TO RED LIGHT REDUCED (SRR1) is essential for repression of flowering in non-inductive photoperiods in Arabidopsis [1]. Mutant *srr1–1* plants flower particularly early under SD conditions and show a reduced sensitivity to the lengthening of the photoperiod. SRR1 acts to promote the expression of several direct repressors of *FT*, including *CDF1*, the *TEMPRANILLO (TEM)* transcription factors that are also involved in gibberellic acid biosynthesis and *FLC*, ensuring that flowering is prevented in non-inductive conditions. In addition, SRR1 has roles in setting the correct pace of the circadian clock and in mediating red light signaling [37]. SRR1 was also found to be important for control of flowering time in natural conditions, together with many genes closely associated to the circadian clock in a combined genome-wide association (GWAS) and linkage mapping study in Arabidopsis [38]. The protein structure of SRR1 is unknown and it does not contain any known protein motifs, although it is highly conserved between species, with homologs present in yeast and mammals [37, 39]. In *Brassica rapa*, a quantitative trait loci (QTL) study combining whole genome transcript variation with flowering time QTLs, identified the *BrSRR1* ortholog as a candidate associated with flowering and the expression of *BrFT* [40]. Furthermore, the *Bna.SRR1.A02* copy has recently been identified as one of the candidates genes responsible for the morphotypic split between biannual and annual forms in *B. napus* [41]. This suggests that the role for SRR1 in flowering time control may be conserved among *Brassicaceae*.

B. napus carries 5 copies of *Bna.SRR1* located on chromosomes *A02*, *A03*, *A10*, *C02* and *C09*. It is unclear if all of them have maintained the original function or if they have undergone subfunctionalization processes. Here, we examine the functionality of the *Bna.SRR1* copies by expression analysis in *B. napus* and complementation of Arabidopsis *srr1–1* mutants. We show that two groups of different gene structures have evolved and that only some *Bna.SRR1* gene copies are functional in Arabidopsis. This indicates a strong subfunctionalization of *Bna.SRR1* and provides new information about SRR1 function.

Results

Phylogeny of SRR1 in Brassicaceae

We searched 13 sequenced *Brassicaceae* species for homologs of *A. thaliana SRR1*. Copies of *SRR1* were found in all 13 species (Fig. 1a). Most of them (8 out of 13, *A. thaliana, A. lyrata, Capsella rubella, Thelungiella salsunginea, Thelungiella halophila, Aethionema arabicum, Leavenworthia alabamicum, Schrenkiella parvula*) carried only one copy of *SRR1*, whereas *B. rapa* and *B.oleracea* each carried two copies, *Camelina sativa* carried three copies, *Sisymbrium irio* four copies and *B. napus* five copies. Thus, *B. napus* carries one copy more than expected from its progenitor species.

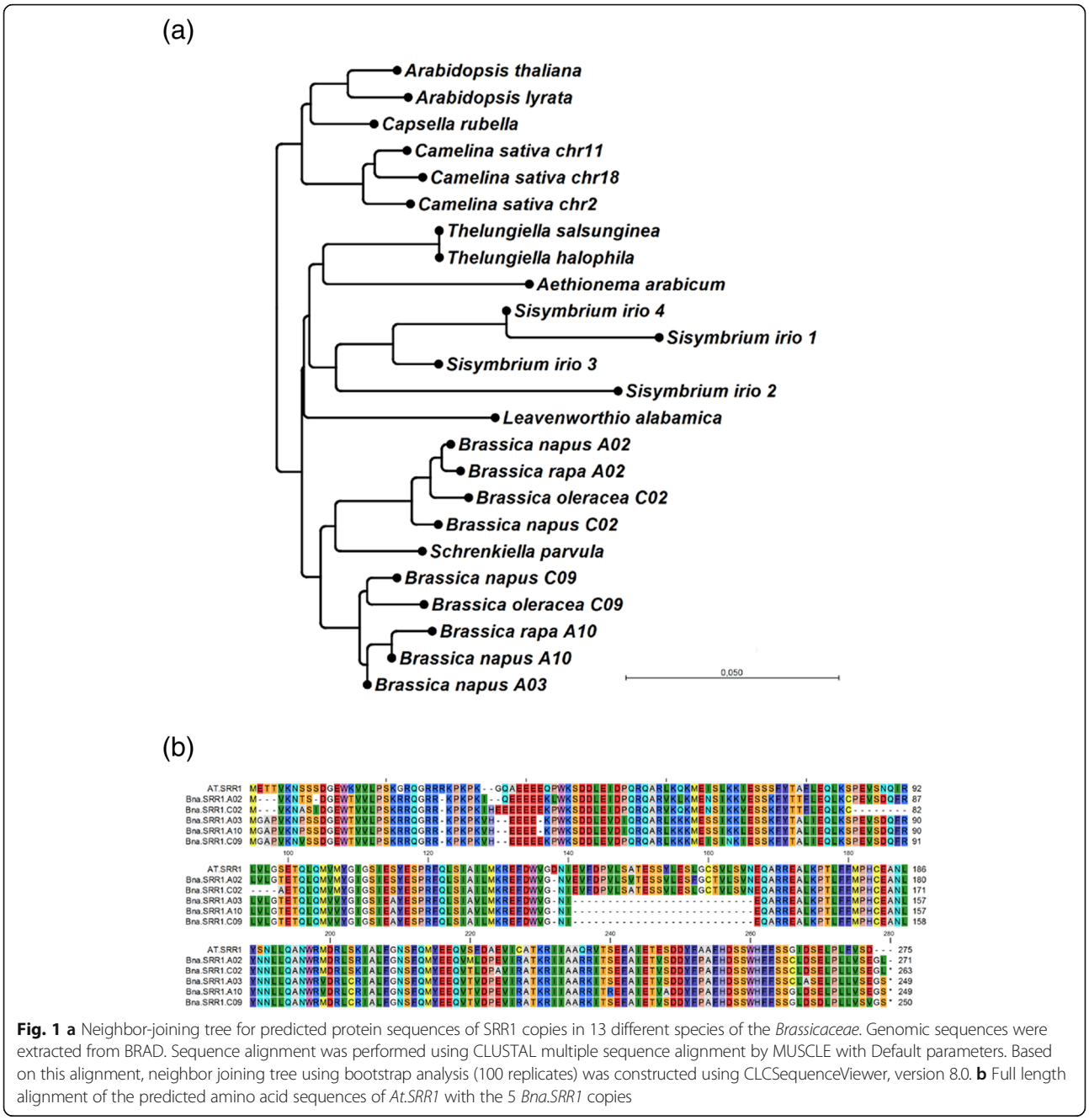


Fig. 1 **a** Neighbor-joining tree for predicted protein sequences of SRR1 copies in 13 different species of the *Brassicaceae*. Genomic sequences were extracted from BRAD. Sequence alignment was performed using CLUSTAL multiple sequence alignment by MUSCLE with Default parameters. Based on this alignment, neighbor joining tree using bootstrap analysis (100 replicates) was constructed using CLCSequenceViewer, version 8.0. **b** Full length alignment of the predicted amino acid sequences of *At.SRR1* with the 5 *Bna.SRR1* copies

Sequence comparisons indicate that the *Bna.SRR1.A03* copy arose from a duplication of the *Bna.SRR1.A10* copy (Fig. 1a).

Gene sequence analysis shows that the five *Bna.SRR1* copies can be divided into two groups, based on their predicted amino acid sequence. The first group, consisting of the *A02* and *C02* gene copies, is more similar to the AtSRR1 protein although several amino acid changes have occurred (Fig. 1b). The second group, consisting of the *A03*, *A10* and *C09* gene copies, all have a 21 amino acid deletion in their protein sequences, compared to the AtSRR1 protein and the A02 and C02 proteins (similarity to AtSRR1: A02: 83.6% and C02: 80.7% conservation vs A03: 73.4%, A10: 73.8% and C09 74.9% conservation). Only one copy in *B. rapa* and *B.oleracea* and two copies in *S. irio* showed similar deletions in this region. A 13 amino acid deletion is also found in the C02 protein, which is unique for this homolog (Fig. 1b).

Not all *Bna.SRR1* copies are expressed

By sampling the Manitoba winter type accession, requiring an extended period of cold to be able to flower, and

the Korall spring type accession, which does not, potential seasonal differences in expression were examined. For 10 week old plants, emerging leaves, developed leaves and petioles were sampled and expression levels of the different copies were tested in the sampled tissues with RT-qPCR using copy-specific primers. This revealed that only three of the five gene copies were expressed at detectable levels, namely the *A02*, *C02* and the *C09* gene copies (Fig. 2). Of these, the *C09* copy was expressed at higher levels compared to the other gene copies, accumulating to about two times the levels of the *A02* copy in all tested tissues in the Manitoba winter type and to an even higher ratio in the Korall spring type (Fig. 2). The *C02* copy was expressed at lower levels than both the *A02* and *C09*. In emerging leaves, all expressed gene copies were expressed at higher levels in the winter type Manitoba, compared to the spring type Korall (Fig. 2a). In developed leaves, expression levels were more similar between the accessions and the *C09* copy was expressed at a slightly higher, but not significantly, level in the spring type Korall compared to the winter type Manitoba (Fig. 2b). In petioles, expression of the *A02* and *C02* copies was only detectable in the Manitoba winter type while the *C09* copy was expressed at high levels in both Korall and Manitoba (Fig. 2c). Thus, there is a much more prominent difference in expression level between accessions in emerging leaves compared to developed leaves. This may suggest that the *Bna.SRR1* genes have an important regulatory role at an earlier stage of development in the Manitoba winter accession compared to the Korall spring accession. To examine whether these findings were accession-specific or dependent on the winter type vernalization requirement, nine additional winter and spring accessions of the ASSYST collection [42, 43], were sampled for emerging leaf material and the expression of *Bna.SRR1 A02, C02* and *C09* was examined. Five accessions were classified as early flowering and four as late flowering of the winter types, while four accessions were early flowering and five late flowering of the examined spring types. Analysis of these accessions revealed a large variation in expression of the *A02* gene copy between the accessions (Fig. 3a). Interestingly, the late-flowering spring lines had a statistically significant ($p > 0.01$, two-factorial ANOVA) higher expression of the *A02* copy compared to the early-flowering spring lines.

The *C02* gene copy was expressed differently between accessions, expression levels were generally higher in the winter accessions, but in several accessions, no expression at all was detected (Fig. 3b).

Expression of the *C09* copy was more stable between the different accessions and comparable to what was observed in the Korall and Manitoba accessions, suggesting presence of the *C09* gene product is important in both winter and spring types (Fig. 3c). Additionally, to examine *Bna.SRR1* expression in other tissues, roots, stems and flowers were sampled from the spring accession Ability as well as roots and stems from the winter accession Zephir. Expression of *Bna.SRR1* was subsequently tested. No *Bna.SRR1* gene copy could be detected in roots, while expression of A02, *C02* and *C09* was detected in stems (Additional file 1: Figure S1). Here, the *A02* copy was expressed at higher levels than the *C02* and *C09* copies in the winter accession, while the *C09* copy had a similar level of expression in both accessions in stems and in flowers in the spring accession. The copy on *C02* was expressed at similar levels to *C09* in stems in both accessions, but not detectable in flowers. In conclusion, the *A02* and *C09* copies were detected in stem and flower tissue, while the *C02* copy was only detected in stems, suggesting possible tissue-specific subfunctionalization between the gene copies.

Bna.SRR1 gene copies show different ability to rescue early flowering in *srr1–1*

To examine whether the *Bna.SRR1* gene copies may have a similar function in flowering as the Arabidopsis

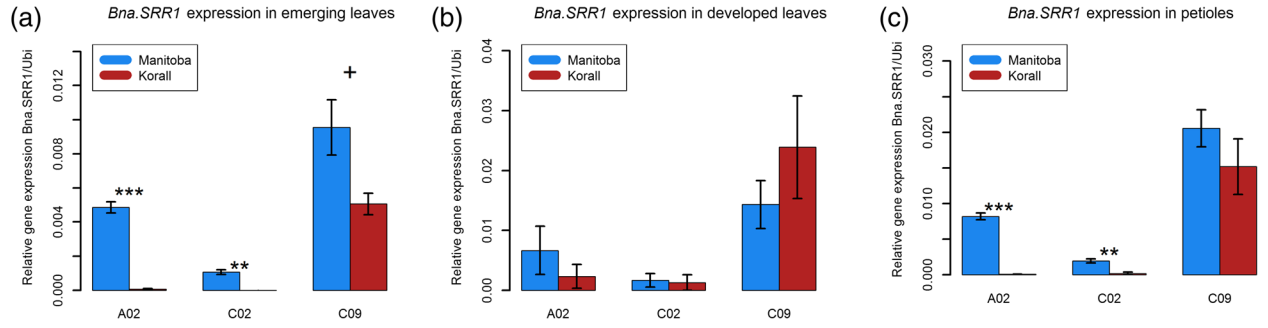


Fig. 2 Relative gene expression of the three expressed *Bna.SRR1* copies in emerging and developed leaves and petioles in the accessions Manitoba (winter-type) and Korall (spring-type) without vernalization from rosette material with approximately 5 developed leaves. (**a**) Emerging leaves, (**b**) developed leaves, (**c**) petioles. The values were calculated from RT-qPCR using the **Δ**Ct method and represent mean of 3 biological replicates. Error bars show standard error of mean. Asterisks show the level of significance based on the Student's t-test (*p-value< 0.05, **p-value< 0.01, ***p-value< 0.001)

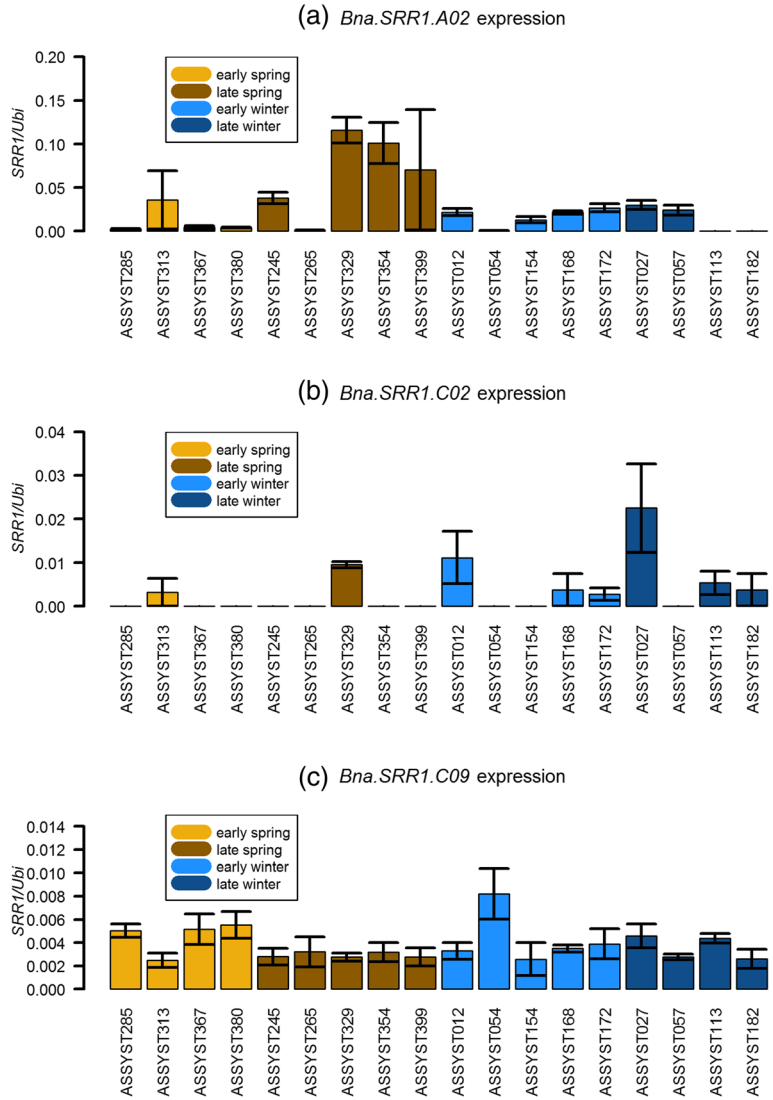


Fig. 3 Relative gene expression of *Bna.SRR1* gene copies in early- and late-flowering spring and winter accessions from the ASSYST collection. (**a**) *Bna.SRR1 A02*, (**b**) *Bna.SRR1 C02*, (**c**) *Bna.SRR1 C09*. The values were calculated from RT-qPCR using the ΔCt method and represent mean of 3 biological replicates. Error bars show standard error of mean

SRR1 gene, the three gene copies shown to be expressed in *B. napus* (*A02*, *C02* and *C09*) were introduced into *srr1–1* mutant plants. About 1500 bp of the promoter region and the genes including the 3′ untranslated region were amplified from genomic *B. napus* DNA using PCR and introduced into the HPT1 binary vector [44]. Subsequently, *srr1–1* mutant plants were transformed with these vectors to introduce the *Bna.SRR1* copies into Arabidopsis. The transformed plant lines were tested for their flowering phenotype under SDs, where *srr1–1* mutants are known to have a strong early flowering phenotype [1]. Flowering time of the transformed plant lines was then measured. The plants transformed with the *A02* gene copy as well as the *C02* copy flowered similar to Col-7 wt plants, thus fully complementing the early flowering phenotype of *srr1–1* (Fig. 4a, b). In comparison, the plants transformed with the *C09* copy flowered with the same leaf numbers as the *srr1–1* mutants (Fig. 4c). This suggests that the differences in *C09* compared to the other homologs may be critical for the proteins' ability to repress flowering in Arabidopsis. In contrast, the deletion in *C02* has no relevance for the function of the protein in regulation of flowering.

To examine how the difference in amino acid composition in C09 may alter the protein, the predicted protein structure of the different SRR1 copies was generated using the PredictProtein resource [45]. This showed that the SRR1 homologs are predicted to have a very similar structure (Additional file 2: Figure S2). The major difference in C09 compared to the other copies is that one α-

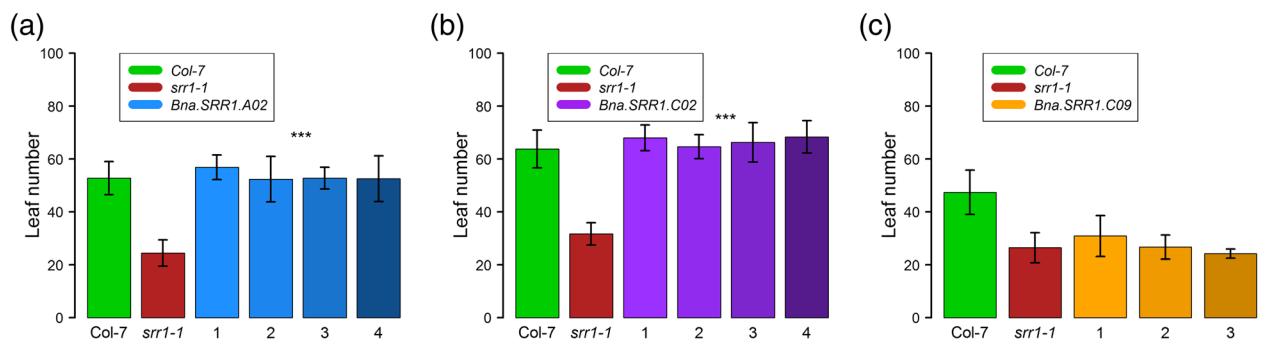


Fig. 4 Flowering time of *srr1–1* plants transformed with *Bna.SRR1* gene copies. Plants were grown in SD conditions (16 h light: 8 h dark, 20 °C). (**a**) *Bna.SRR1 A02*-transformed lines, (**b**) *Bna-SRR1 C02*-transformed lines, (**c**) *Bna.SRR1 C09*-transformed lines. Leaves of at least 15 plants per line were counted at flowering. Error bars show standard deviation. Asterisks show the level of significance based on the Student's t-test compared to not transformed *srr1–1* plants (*p-value< 0.05, **p-value< 0.01, ***p-value< 0.001)

helix, predicted to be mainly exposed, is missing through the deletion. The prediction does not suggest that the deletion renders the protein unstable.

Expression of *Bna.SRR1* gene copies in Arabidopsis

As expression levels of the different *Bna.SRR1* gene copies differed strongly in *B. napus*, the level of expression of *Bna.SRR1.A02*, which could complement flowering in Arabidopsis, and *C09*, which could not, was tested in the Arabidopsis lines transformed with the respective gene copies.

RT-qPCR analysis showed that in comparison to the endogenous *SRR1* gene copy, both *Bna.SRR1* genes introduced into the *srr1–1* background were expressed at lower levels (Fig. 5). For the *A02* copy, these low levels of expression were obviously sufficient to complement the flowering phenotype. The *C09* copy was also expressed at lower levels than *AtSRR1*, but higher than *A02* in the tested lines, reaching ca 30% of expression levels of *AtSRR1*. The level of expression of the *A02* copy does not appear to be critical for the function of SRR1, as low amounts of transcript are sufficient to fulfill its role in flowering time control. A comparison of the promoter structure between the *SRR1* gene copies using the MEME suite [46] revealed two enriched motifs common in all the gene promoters, although their distribution is somewhat different between the genes (Additional file 3: Fig. S3). The motifs, a SORLIP motif and an ARF motif, have been described to be involved in light-regulated gene expression and as an auxin response factor binding site, respectively [47, 48]. They were located close to the start of the coding sequence in

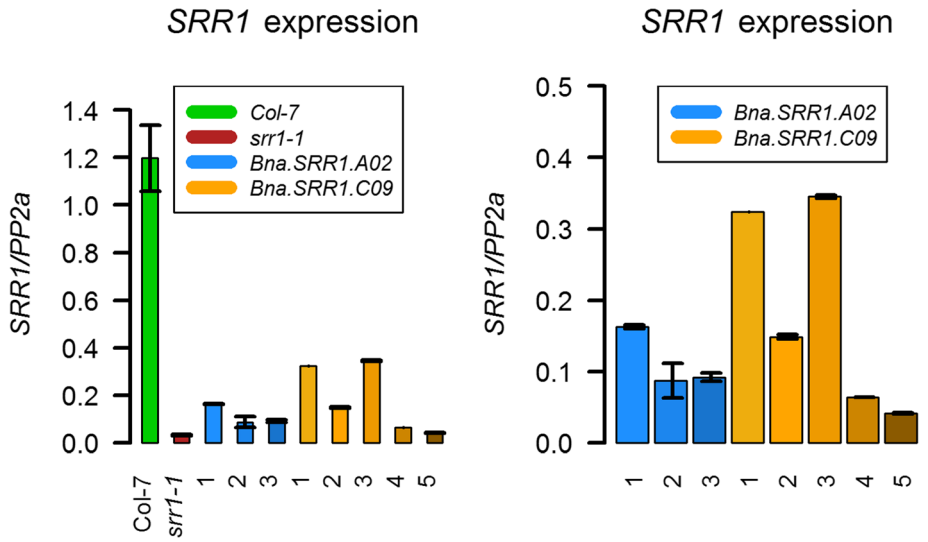


Fig. 5 Expression of *SRR1* in Col-7 wt and *srr1–1* lines transformed with the *Bna.SRR1* gene copies. Left panel: Expression in comparison with Col-7 wt of three independent transformed lines transformed with *Bna.SRR1 A02* and *Bna.SRR1 C09*, respectively. Right panel: magnified comparison of expression between the *Bna.SRR1 A02*-transformed and *Bna.SRR1 C09*-transformed lines. Error bars show standard error of mean

AtSRR1, while they were located further upstream in the *Bna.SRR1* gene promoters. Although the factors regulating SRR1 expression are unknown, this may indicate that the efficiency of transcriptional activation of the *Bna.SRR1* genes in Arabidopsis is different, which could explain the reduced expression levels of the *Bna.SRR1* copies compared to endogenous *AtSRR1*.

Expression of SRR1 targets in *Bna.SRR1* transformed lines

SRR1 acts in several pathways regulating flowering by promoting the expression of *FT* repressors [1]. To examine how the *Bna.SRR1* copies affected known targets of AtSRR1 in regulation of gene expression, their transcript levels in plants carrying the *AtSRR1*-complementing *A02* and the non-complementing *C09* gene copies were measured.

To confirm that the complemented phenotype in *BnaSRR1.A02* lines was due to restoration of the *FT* expression pattern, a time series was sampled in 3-h intervals over 24-h in SD conditions and analyzed using RT-qPCR. This revealed that in *A02*-transformed lines, *FT* was expressed at very low wt-like levels, while elevated expression was observed in *C09*-transformed lines, notably at the for flowering induction critical time point ZT9, as well as in *srr1–1* mutants (Fig. 6a, Additional file 4: Fig. S4). Furthermore, analysis of *CDF1* expression, a known repressor of *FT* and a target of SRR1 showed that *CDF1*, with an expected peak of expression in the morning, was expressed as in Col-7 in the lines transformed with the *A02* gene copy. Meanwhile, *CDF1* was expressed at reduced levels in the morning and expression peaked earlier in the *C09*-transformed lines (Fig. 6b). This was similar to the expression pattern observed in *srr1–1* mutants and thus *C09* had no complementing effect on *CDF1* expression.

TEM1 and *TEM2* transcription factors are other known targets of SRR1, which are involved in regulation of flowering through the GA biosynthesis pathway [49]. Due to their redundancy and similar expression, *TEM1* was examined to determine whether the *Bna.SRR1* copies could affect their expression. Interestingly, whereas *srr1–1* showed reduced expression, as seen previously [1], both the *A02* and *C09*-transformed lines showed an expression pattern similar or even slightly enhanced compared to Col-7 wt, suggesting complementation of *TEM1* expression (Fig. 6c). TEM1 is known to repress the gibberellic acid biosynthesis gene *GIBBERELLIN 3-OXIDASE 1* (*GA3OX1)* [49]. To confirm the rescued expression of *TEM1* in *A02* and *C09*-transformed lines, *GA3OX1* expression was measured. Transcript levels were elevated in *srr1–1* compared to wt, congruent with previous observations [1]. In the *C09*-transformed lines no significant difference was seen while expression in *A02*-transformed lines was somewhat reduced (Additional file 5: Figure S5). The results support the elevated *TEM1* levels in the transformed lines. In conclusion, this suggests that the function of SRR1 in transcriptional regulation is fully rescued by the *A02* gene copy for all tested genes, while the *C09* gene copy can only complement *TEM1* expression, which is not enough to rescue the early flowering phenotype of *srr1–1*.

Discussion

Our data show that SRR1 is highly conserved in *Brassicaceae,* suggesting an important function in growth and development within this family (Fig. 1a). However, its unique protein structure has made the prediction of key regions of the protein or a molecular mode of action difficult. Interestingly, we found that the crop species *Brassica napus* carries in total five homologs with differences in gene and protein structure between them, including a larger deletion in the A03, A10 and C09 proteins, compared to the A02 and C02 proteins and AtSRR1 (Fig. 1b). As this could suggest subfunctionalization between the different gene copies, we tested their level of expression in *B. napus* followed by a functional analysis of the expressed copies in Arabidopsis *srr1–1* background, where *AtSRR1* is not expressed.

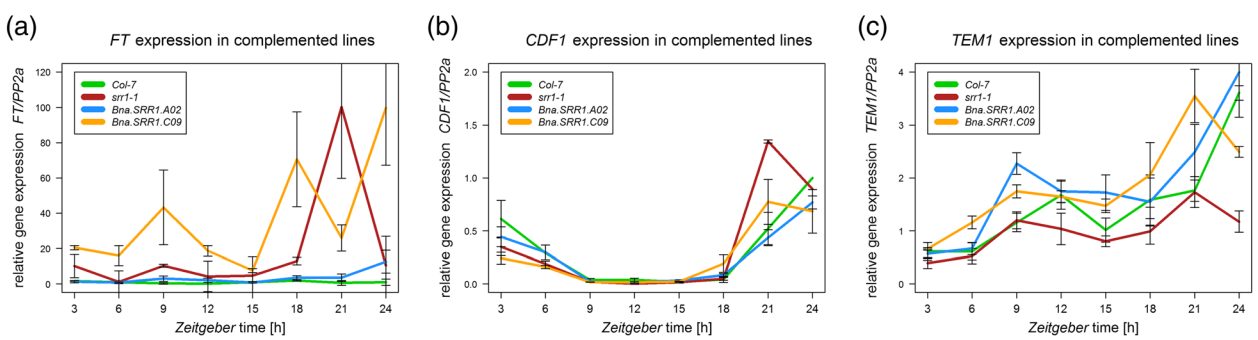


Fig. 6 Relative expression of known SRR1 targets in lines transformed with *Bna.SRR1 A02* and *Bna.SRR1.C09*. (**a**) *FT*, (**b**) *CDF1*, (**c**) *TEM1*. The values represent the average expression of three independently transformed lines. Error bars show standard error of mean

Differential gene expression suggests subfunctionalization

Initial gene expression analysis in the *B. napus* accessions Manitoba (winter type) and Korall (spring type) showed that only three of the five gene copies were expressed, *A02*, *C02* and *C09* (Fig. 2). Expression of the same copies was also detected in stems, while in flowers only *A02* and *C09* was detected (Additional file 1: Figure S1). In roots, no *Bna.SRR1* copy was detected.

Differential expression of *B. napus* flowering gene copies has been shown in several reports [25, 50–52]. Here, the *C09* copy is consistently expressed in all tested accessions and tissues, while the *A02* and C02 copies are expressed at different levels depending on accession in emerging leaves (Fig. 3).

In the winter type Manitoba and the spring type Korall, comparison of expression between developed and emerging leaves showed that differences in expression between accessions was lower in comparison to emerging leaves (Fig. 2a, b), suggesting that the *A02* and *C02* gene copies may have a repressive role on flowering at early stages of development, when highly expressed. This may suggest that they have a comparable role as SRR1 in Arabidopsis, in suppressing flowering until the conditions are more favorable. Interestingly, a similar pattern has been observed for the important flowering time regulator, *FLC*, where three out of nine copies were differentially expressed between winter and spring material (Quezada et al., submitted). One copy of *Bna.FLC* was never expressed [53], similar to what we found for *Bna.SRR1.A03* and *Bna.SRR1.A10*, indicating pseudogenization.

Thus, in Manitoba and Korall, the larger difference in *Bna.SRR1* expression in emerging leaves compared to developed leaves between the Korall spring and the Manitoba winter type may suggest that high expression early in the developmental cycle in the winter type is desirable to prevent premature flowering. This could account for a mechanism measuring the ratio of developing to differentiated leaves, allowing flowering only after a certain leaf mass has been reached. Developing leaves could likewise send a “stop” signal, which is only overridden if enough differentiated tissue has developed.

However, in the extended analysis of emerging leaves of several other accessions, *A02* expression displayed a large variation suggesting that such a mechanism may be accession-dependent. *A02* was particularly highly expressed in several late-flowering spring accessions, suggesting it may have a role in delaying flowering in these accessions (Fig. 3a). This function may be overruled by *FLC* in winter accessions with a vernalization requirement.

Expression of *C02* also varied between accessions, suggesting a possible accession-specific function, while expression of *C09* was much more stable between accessions in the extended analysis (Fig. 3b, c). In comparison, presence of the *C09* gene product seems to be of general importance in *B napus* and thus have the different gene copies subfunctionalized to perform specific roles in this species.

Complementation reveals potential key protein domain of SRR1

Flowering time experiments with the three gene copies being expressed in *B. napus* showed that only the *A02* and *C02* gene copies can complement the early flowering phenotype of *srr1–1* while the *C09* copy cannot (Fig. 4). This suggests that the differences in *C09* may be critical for the function of the SRR1 protein in Arabidopsis, at least in regard to its role in regulating flowering. The most obvious candidate region to be critical for correct function is the 21 amino acid deletion in C09. In comparison, although the C02 protein product also carries a deletion in another part of the protein, it could still complement the loss of AtSRR1 in *srr1–1* plants (Fig. 1b, Fig. 4). As the SRR1 protein sequence does not contain any known regulatory elements, this is an important finding, indicating that this region of the protein may be critical for proper function. This deletion is a highly conserved SRR1-unique sequence in *Brassicacea* and this specific deletion only occurs in *B. napus*. Taking into account that the *A02* and *C02* copies are the same ones expressed at much lower levels in the spring type compared to the winter type, this further indicates that their expression may be necessary to prevent undesirable premature flowering in the winter type, acting as a repressive signal in months preceding the cold season.

The dysfunction of the *C09* gene copy in Arabidopsis could be either due to an important function-specific binding region of the protein being excluded through the altered protein sequence, or due to direct degradation of the C09 protein product. However, the performed protein structure prediction suggests that C09 still has a similar structure to the other SRR1 copies, and only one predicted helix structure is missing (Additional file 2: Figure S2). Considering the experimental results, this deletion may be important for interactions or protein modifications necessary for the regulation of flowering time. Further biochemical studies are however necessary to confirm that the region deleted in C09 is the determining factor.

Considering that SRR1 in Arabidopsis is also involved in circadian regulation and light signaling [37], it is possible that the *A02* and *C09* copy may have a specialized functions in *B. napus,* through subfunctionalization. Being that the *C09* gene copy is unique for *B. napus* may also suggest a species-specific specialization. Its exact function requires more detailed analyses in *B. napus*.

The expression analysis in lines complemented with *B. napus* gene copies in Arabidopsis shows that the expression levels of the introduced genes were much lower than the endogenous *SRR1* in Col-7 wt plants (Fig. 5). This was, however, sufficient for *A02* to be able to complement the *srr1–1* early flowering phenotype, suggesting that low *SRR1* expression levels are enough for proper function. Expression of the *C09* copy was lower than endogenous *SRR1*, but higher than *Bna.SRR1.A02*. Thus, considering that expression of the *A02* lines was sufficient to complement the flowering phenotype of *srr1–1*, it is unlikely that the level of *C09* expression is a major factor in the inability of the *C09* gene copy to do the same (Fig. 5).

CDF1 is key to regulate flowering through SRR1

Analysis of known targets of SRR1 showed that the *A02* gene copy was able to replace AtSRR1 function in regard to its role in regulating expression of flowering time regulators, including the key florigen *FT,* the important *FT* repressor, *CDF1,* and *TEM1* (Fig. 6). In contrast, the *C09* copy was unable to rescue *SRR1* function, as the *C09*-transformed lines showed *srr1–1*-like expression patterns of *CDF1* and *FT*. Conversely, *TEM1* expression levels were rescued to WT levels by *C09*, but this seems to have a limited effect on flowering, as *C09*-transformed plants flowered like *srr1–1* mutants. In conclusion, the data suggests that the key target for floral repression by SRR1 is *CDF1,* where an altered expression is observed in *srr1–1,* as well as in *C09*-transformed lines (Fig. 6a). *TEM1* appears to be rescued by both the *A02* and *C09* gene copies (Fig. 6c), although this is not enough to rescue the early flowering phenotype in *C09*-transformed lines. This indicates that the differences in C09, most notably the deleted region, may be necessary for SRR1 control of *CDF1* expression.

Our data suggests that these gene copies may have a similar molecular mode of action in *B. napus* as in Arabidopsis and may be able to influence expression of *B. napus* homologs to other known flowering time components, which have been shown to be also present in *B. napus* [33]. Furthermore, the consistent expression levels of the *C09* copy compared to the variation in *A02* expression may suggest that the gene copies have subfunctionalized to acquire specific roles in *B. napus* that may or may not be related to the regulation of flowering. This information may help to map the signaling network controlling flowering time in *B. napus*, enabling the identification of key factors in breeding.

Conclusions

We have shown that SRR1, an important Arabidopsis flowering time regulator, has several homologs in *Brassica napus*. Their expression patterns varied and major alterations in amino acid composition were found. The differences in expression between winter and spring type accessions suggest their expression may be of importance to flowering ability.

Only two of three expressed copies could complement the early flowering *srr1–1* mutant phenotype, showing cross-species functionality. The C09 copy, with a 21 amino acid deletion compared to A02, C02, and AtSRR1, failed to complement the early flowering phenotype. *C09* is, however, consistently expressed in *B. napus*, suggesting strong subfunctionalization between the gene copies. The presented data may be used in the future for further characterization of the flowering time pathway in *B. napus* and highlights the possibility that the *B. napus* gene copies may have taken on specific functions throughout evolution.

Methods

Sequence analysis

Whole genome sequences for *A. thaliana, A. lyrata, B. napus, B. rapa, B.oleracea, Camelina sativa, Capsella rubella, Thelungiella salsunginea, Thelungiella halophila, Aethionema arabicum, Leavenworthia alabamicum, Schrenkiella parvula* and *Sisymbrium irio* were retrieved from http://brassicadb.org/brad/ftpTrans.php. The five known copies of *B. napus* were then used for a BLAST search against each of the genomes. *Bna.SRR1* copies were then selected using a cut-off value of 10^{-50} for Brassica and Arabidopsis, while using a cutoff of 10^{-20} for the remaining species. Fragments shorter than 200 bp were excluded. To avoid missing gene information, 100 bp were added to the start and stop of each BLAST position. For all species except the *Brassica* species and *Arabidopsis thaliana*, peptide sequences were predicted using GENSCAN (http://genes.mit.edu/GENSCAN.html) with "Arabidopsis" as organism. For *Brassica* and *A. thaliana*, we used the peptide sequence information from the respective peptide prediction published within their reference genomes.

Plant material and growth conditions

Arabidopsis thaliana

The T-DNA mutant *srr1–1* in the Col-7 background has been described [1, 37]. All seeds were stratified for 3 d at 4 °C before putting on soil. Seeds grown on plates were surface sterilized and stratified for 3 d at 4 °C before plating on agar-solidified half-strength MS (Murashige & Skoog) medium (Duchefa) supplemented with 0.5% sucrose and 0.5 g MES. Plants were grown in Percival incubators AR66-L3 (CLF Laboratories) in 100 μmol m^{-2} s^{-1} light intensity, with the light-dark and temperature conditions as indicated.

Brassica napus

A winter accession (Manitoba) and a spring accessions (Korall) of oilseed rape were sown in 7 × 7 cm pots in 3 biological replicates and transplanted to 12 × 12 cm pots 4 weeks after sowing. For the extended expression analysis, a diversity set consisting of 10 winter and 10 spring accessions was sown in quickplates in 3 biological replicates. Cultivation was performed in a greenhouse using a 16 h/8 h day/night rhythm with 20 °C /17 °C. For Manitoba and Korall, we sampled petioles, developed and emerging leaves separately 10 weeks after sowing. For the diversity set, we selected 9 winter and 9 spring accessions for the youngest developed leaf 8 weeks after sowing. The other two accessions were grown 3 weeks further and we sampled stems, roots and flowers separately. Tissues were frozen in liquid nitrogen and stored at − 80 °C until RNA extraction.

Flowering time experiments

Seeds were germinated as described above and grown on soil in a random fashion. Flowering time was determined by counting the rosette leaves when the bolt was > 0.5 cm tall [54].

Cloning

Genomic DNA from *Brassica napus* was amplified using Phusion Proofreading polymerase (Thermo Fischer) and primers with specific restriction sites. The amplified DNA was separated on an agarose gel and extracted using a GeneJet gel extraction kit (Thermo Fischer) and then ligated into a pJET2.1 cloning vector using the CloneJet kit (Thermo Fischer). The insert was digested and separated on an agarose gel and then cloned into a pHPT1 binary vector [44], using T4 Ligase (Thermo Fischer). The resulting construct was transformed into Agrobacterium and then into Arabidopsis *srr1–1* plants using the floral dip method.

Transcript analysis

Arabidopsis material

Total RNA was extracted using from plant material using Tri Reagent as previously described or using Universal RNA Purification Kit (Roboklon) following manufacturer's instructions.

For cDNA, 2 μg of total RNA was DNAse-treated using RQ1 RNAse-free DNAse (Promega) and reverse transcribed using AMV Reverse Transcripase (Roboklon) according to manufacturer's instructions.

qPCR was performed with iTaq Sybr Green Supermix (Bio-Rad) according to manufacturer's instructions. The normalized expression level was determined using the ΔCt method, with *PP2a* (At1g69960) as a reference gene as described [55]. The primer sequences can be found in Additional file 6: Table S1.

Brassica napus material

Total RNA was extracted using the NucleoSpin miRNA kit (Macherey-Nagel) following manufacturer's instructions. The eluted RNA was quantified using Qubit RNA Broad Range on a Qubit fluorimeter and stored at − 80 °C until use.

Primers were designed based on the Darmor-bzh reference genome, version 4.1 (Chalhoub et al. 2014). Specificity was confirmed by aligning the predicted cDNA with CLUSTAL multiple sequence alignment by MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/, version 3.8). The primer sequences can be found in Additional file 6: Table S1.

cDNA synthesis was performed using the RevertAid cDNA synthesis kit (ThermoFisher) using 1 μg of total RNA and Oligo-dT primers. The amount of cDNA was quantified using the Qubit DNA High Sensitivity kit on a Qubit fluorimeter. Quantitative Real-time PCR was performed on a Real-Time PCR System ViiA7 cycler (Applied Biosystems) in 384 well plates. The reaction mix containing specific primers, the template cDNA and FastStart Universal SYBR Green Master mix containing Rox (Roche) was pipetted by a robot (Biomek 4000, Beckman Coulter). As endogenous control, we used ubiquitin. The PCR program was as follows: initial denaturation (94 °C for 2 min), amplification and quantification (40 cycles, 95 °C for 20 s, 60 °C for 30 s, 72 °C for 30 s), and a final extension (72 °C for 5 min). At the end, a melting curve was recorded between 55 and 95 °C. PCR efficiency was measured using a pool of all samples in a dilution series of 6 points. All samples were measured in 3 technical replicates. The normalized expression level was determined using the ΔCt method.

Additional files

Additional file 1: Figure S1. Relative gene expression of *Bna.SRR1* gene copies in different tissues of the Ability spring and Zephir winter accessions. The values were calculated from RT-qPCR using the ΔCt method and represent mean of 3 biological replicates. Error bars show standard error of mean. (TIF 10547 kb)

Additional file 2: Figure S2. Protein structure predictions based on the PredictProtein server. Red squares in the first row indicate predicted alpha-helices, blue squares indicate strands. Yellow boxes in the second row indicate buried regions while blue boxes indicate exposed regions. Grey boxes in the third row indicate disordered regions. The red dotted squares highlight a predicted helix missing in Bna.C09 compared to the other predicted SRR1 copies. (TIF 517 kb)

Additional file 3: Figure S3. *AtSRR1* and *BnSRR1* promoter alignment. Two enriched motifs were discovered using MEME. Sequences from *A.thaliana* and *B.napus* 1 kb upstream from the transcriptional start site were used with a minimal motif length of 6 and maximum of 10 (Bailey and Elkan, Proc Int Conf Intell Syst Mol Biol, 2:28–36,1994). Motifs were determined to be statistically significant with an E-value lower than 0.05. SORLIP 2 binding site is associated with *PhyA* signaling, while ARF (Auxin Response Factor) binding sites are intrinsic for the auxin response. Enriched motifs are underlined and binding sites are highlighted in gray. (TIF 577 kb)

Additional file 4: Figure S4. Expression of FT at zeitgeber time 9 (9 h after lights on, ZT9) in plants grown in SDs (8 h light:16 h dark, 20 °C). The values represent biological replicates of three independently transformed lines. Error bars show standard error of mean. (TIF 5273 kb)

Additional file 5: Figure S5. Expression of TEM1 target *GA3OX1* at zeitgeber time 8 (8 h after lights on, ZT8) in plants grown in SDs (8 h light:16 h dark, 20 °C). The values represent biological replicates of three independently transformed lines. Error bars show standard error of mean. Asterisks show the level of significance based on the Student's t-test compared to Col-7 wt plants. (TIF 6591 kb)

Additional file 6: Primer sequences. (XLSX 10 kb)

Abbreviations
LD: Long day; RT-qPCR: Real Time Quantitative PCR; SD: Short day

Acknowledgements
This project was initiated in the DFG Priority Program SPP1530 on Flowering Time Control (STA 653/5).

Consent to publication
Not applicable.

Authors'contributions
MJ, SVS and DS designed the experiments. MJ, SVS, NW and PS performed experiments. MJ and SVS performed data analysis. MJ, SVS and DS wrote the manuscript with input from all authors. All authors have read and approved the final manuscript.

Funding
MJ is supported by the DFG grant JO 1252/2. SVS has been funded by grant SN14/14–2. The funding bodies had no role in the experimental design, data analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials
All data generated or analysed during this study are included in this published article and its supplementary information files.

Ethics approval and consent to participate
Not applicable.

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

Author details
[1]RNA Biology and Molecular Physiology, Faculty for Biology, Bielefeld University, Universitaetsstrasse 25, 33615 Bielefeld, Germany. [2]Department of Plant Breeding, Justus Liebig University, IFZ Research Centre for Biosystems, Land Use and Nutrition, Heinrich-Buff-Ring 26-32, 35392 Giessen, Giessen, Germany.

Received: 20 December 2018 Accepted: 13 August 2019
Published online: 22 August 2019

References
1. Johansson M, Staiger D. SRR1 is essential to repress flowering in non-inductive conditions in Arabidopsis thaliana. J Exp Bot. 2014;65(20):5811–22.
2. Johansson M, Staiger D. Time to flower: interplay between photoperiod and the circadian clock. J Exp Bot. 2015;66:719–30.
3. Jung C, Muller AE. Flowering time control and applications in plant breeding. Trends Plant Sci. 2009;14(10):563–73.
4. McClung CR, Lou P, Hermand V, Kim JA. The importance of ambient temperature to growth and the induction of flowering. Front Plant Sci. 2016;7:1266.
5. Hyun Y, Richter R, Coupland G. Competence to flower: age-controlled sensitivity to environmental cues. Plant Phys. 2017;173(1):36–46.
6. Song YH, Shim JS, Kinmonth-Schultz HA, Imaizumi T. Photoperiodic flowering: time measurement mechanisms in leaves. Annu Rev Plant Biol. 2015;66:441–64.
7. An H, Roussot C, Suárez-López P, Corbesier L, Vincent C, Piñeiro M, et al. CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis. Development. 2004;131(15):3615–26.
8. Corbesier L, Vincent C, Jang SH, Fornara F, Fan QZ, Searle I, et al. FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science. 2007;316(5827):1030–3.
9. Jaeger KE, Wigge PA. FT protein acts as a long-range signal in Arabidopsis. Curr Biol. 2007;17(12):1050–4.
10. Mathieu J, Warthmann N, Küttner F, Schmid M. Export of FT protein from phloem companion cells is sufficient for floral induction in Arabidopsis. Curr Biol. 2007;17(12):1055–60.
11. Suarez-Lopez P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G. CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature. 2001;410(6832):1116–20.
12. Sawa M, Nusinow DA, Kay SA, Imaizumi T. FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. Science. 2007;318(5848):261–5.
13. Fornara F, Panigrahi KCS, Gissot L, Sauerbrunn N, Rühl M, Jarillo JA, et al. Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. Dev Cell. 2009;17(1):75–86.
14. Michaels SD, Amasino RM. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell. 1999;11(5):949–56.
15. Searle I, He Y, Turck F, Vincent C, Fornara F, Kröber S, et al. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. Genes Dev. 2006;20(7):898–912.
16. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, et al. Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science. 2014;345(6199):950–3.
17. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, et al. The genome of the mesopolyploid crop species Brassica rapa. Nat Genet. 2011;43(10):1035–9.
18. Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA, et al. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun. 2014;5:3930.
19. Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, et al. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. Genome Biol. 2014;15(6):R77.
20. Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, et al. The genome of the extremophile crucifer Thellungiella parvula. Nat Genet. 2011;43(9):913–8.
21. Fletcher RS, Mullen JL, Heiliger A, McKay JK. QTL analysis of root morphology, flowering time, and yield reveals trade-offs in response to drought in Brassica napus. J Exp Bot. 2015;66(1):245–56.
22. Li X, Ramchiary N, Dhandapani V, Choi SR, Hur Y, Nou IS, et al. Quantitative trait loci mapping in Brassica rapa revealed the structural and functional conservation of genetic loci governing morphological and yield component traits in the a, B, and C subgenomes of Brassica species. DNA Res. 2013;20(1):1–16.
23. Nelson MN, Rajasekaran R, Smith A, Chen S, Beeck CP, Siddique KH, et al. Quantitative trait loci for thermal time to flowering and photoperiod responsiveness discovered in summer annual-type Brassica napus L. PLoS One. 2014;9(7):e102611.
24. Okazaki K, Sakamoto K, Kikuchi R, Saito A, Togashi E, Kuginuki Y, et al. Mapping and characterization of FLC homologs and QTL analysis of flowering time in Brassica oleracea. Theor Appl Genet. 2007;114(4):595–608.
25. Raman H, Raman R, Coombes N, Song J, Prangnell R, Bandaranayake C, et al. Genome-wide association analyses reveal complex genetic architecture underlying natural variation for flowering time in canola. Plant Cell Environ. 2016;39(6):1228–39.
26. Raman H, Raman R, Eckermann P, Coombes N, Manoli S, Zou X, et al. Genetic and physical mapping of flowering time loci in canola (Brassica napus L.). Theor Appl Genet. 2013;126(1):119–32.
27. Udall JA, Quijada PA, Lambert B, Osborn TC. Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (Brassica napus L.): 2. Identification of alleles from unadapted germplasm. Theor Appl Genet. 2006;113(4):597–609.
28. Zhao J, Kulkarni V, Liu N, Del Carpio DP, Bucher J, Bonnema G. BrFLC2 (FLOWERING LOCUS C) as a candidate gene for a vernalization response QTL in Brassica rapa. J Exp Bot. 2010;61(6):1817–25.

29. Schiessl S, Iniguez-Luy F, Qian W, Snowdon RJ. Diverse regulatory factors associate with flowering time and yield responses in winter-type Brassica napus. BMC Genomics. 2015;16:737.
30. Lagercrantz U, Putterill J, Coupland G, Lydiate D. Comparative mapping in Arabidopsis and Brassica, fine scale genome collinearity and congruence of genes controlling flowering time. Plant J. 1996;9(1):13–20.
31. Lysak MA, Koch MA, Pecinka A, Schubert I. Chromosome triplication found across the tribe Brassiceae. Genome Res. 2005;15(4):516–25.
32. Schiessl S, Huettel B, Kuehn D, Reinhardt R, Snowdon RJ. Targeted deep sequencing of flowering regulators in Brassica napus reveals extensive copy number variation. Sci Data. 2017;4:170013.
33. Schiessl S, Samans B, Huttel B, Reinhard R, Snowdon RJ. Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species Brassica napus. Front Plant Sci. 2014;5:404.
34. Renny-Byfield S, Wendel JF. Doubling down on genomes: polyploidy and crop plants. Am J Bot. 2014;101(10):1711–25.
35. Becker A, Theissen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Phylogenet Evol. 2003;29(3):464–89.
36. Dreni L, Kater MM. MADS reloaded: evolution of the AGAMOUS subfamily genes. New Phytol. 2014;201(3):717–32.
37. Staiger D, Allenbach L, Salathia N, Fiechter V, Davis SJ, Millar AJ, et al. The Arabidopsis SRR1 gene mediates phyB signaling and is required for normal circadian clock function. Genes Dev. 2003;17(2):256–68.
38. Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, et al. Linkage and association mapping of Arabidopsis thaliana flowering time in nature. PLoS Genet. 2010;6(5):e1000940.
39. Fiechter V, Cameroni E, Cerutti L, Virgilio C, Barral Y, Fankhauser C. The evolutionary conserved BER1 gene is involved in microtubule stability in yeast. Curr Genet. 2008;53(2):107–15.
40. Xiao D, Zhao JJ, Hou XL, Basnet RK, Carpio DPD, Zhang NW, et al. The Brassica rapa FLC homologue FLC2 is a key regulator of flowering time, identified through transcriptional co-expression networks. J Exp Bot. 2013; 64(14):4503–16.
41. Schiessl S, Huettel B, Kuehn D, Reinhardt R, Snowdon R. Post-polyploidisation morphotype diversification associates with gene copy number variation. Sci Rep. 2017;7:41845.
42. Bus A, Korber N, Snowdon RJ, Stich B. Patterns of molecular variation in a species-wide germplasm set of Brassica napus. Theor Appl Genet. 2011; 123(8):1413–23.
43. Korber N, Wittkop B, Bus A, Friedt W, Snowdon RJ, Stich B. Seedling development in a Brassica napus diversity set and its relationship to agronomic performance. Theor Appl Genet. 2012;125(6):1275–87.
44. Schöning JC, Streitner C, Meyer IM, Gao Y, Staiger D. Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in Arabidopsis. Nucleic Acids Res. 2008;36(22):6977–87.
45. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, et al. PredictProtein--an open resource for online prediction of protein structural and functional features. Nucleic Acids Res. 2014;42:W337–43.
46. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME suite: tools for motif discovery and searching. Nucleic Acids Res. 2009; 37(suppl_2):W202–W8.
47. Hudson ME, Quail PH. Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. Plant Physiol. 2003; 133(4):1605–16.
48. Ulmasov T, Hagen G, Guilfoyle TJ. Dimerization and DNA binding of auxin response factors. Plant J. 1999;19(3):309–19.
49. Osnato M, Castillejo C, Matías-Hernández L, Pelaz S. TEMPRANILLO genes link photoperiod and gibberellin pathways to control flowering in Arabidopsis. Nature Comm. 2012;3:808.
50. Tadege M, Sheldon CC, Helliwell CA, Stoutjesdijk P, Dennis ES, Peacock WJ. Control of flowering time by FLC orthologues in Brassica napus. Plant J. 2001;28(5):545–53.
51. Wang J, Hopkins CJ, Hou J, Zou X, Wang C, Long Y, et al. Promoter variation and transcript divergence in Brassicaceae lineages of FLOWERING LOCUS T. PLoS One. 2012;7(10):e47127.
52. Wang N, Qian W, Suppanz I, Wei L, Mao B, Long Y, et al. Flowering time variation in oilseed rape (Brassica napus L.) is associated with allelic variation in the FRIGIDA homologue BnaA.FRI.a. J Exp Bot. 2011;62(15):5641–58.
53. Zou X, Suppanz I, Raman H, Hou J, Wang J, Long Y, et al. Comparative analysis of FLC homologues in Brassicaceae provides insight into their role in the evolution of oilseed rape. PLoS One. 2012;7(9):e45751.
54. Steffen A, Fischer A, Staiger D. Methods Mol Biol. 2014;1158:285–95.
55. Streitner C, Hennig L, Korneli C, Staiger D. Global transcript profiling of transgenic plants constitutively overexpressing the RNA-binding protein AtGRP7. BMC Plant Biol. 2010;10(1):221.

Publisher's Note

PART 3: Flowering time and drought stress

Flowering time is not only an irreversible switch from vegetative to reproductive growth, but also the phase where plants are most sensitive to stress. Stress during flowering limits plant growth and seed yield, even if conditions are optimal afterwards. For Europe, climate change is predicted to lead to wetter winters, but drier springs, therefore largely affecting flowering time of winter oilseed rape. It is therefore necessary to understand how stress tolerance during flowering can be improved by breeding. At the same time, it is largely unknown how stress signaling and flowering time regulation interact, and which genetically accessible tolerance strategies might exist. It has been proposed that miRNAs might play a role in a matter similar to the age pathway, so *MIRNA* genes might represent a possible breeding target. To shed light into this hypothesis, we conducted a controlled drought stress experiment for winter oilseed rape. The results are presented in two chapters: chapter 3.1 is an unpublished manuscript showing how a developmental delay in drought stressed accessions is associated to miRNA and mRNA expression. Chapter 3.2 shows how drought stress during flowering time does even affect plant development in the next generation.

Chapter 3.1: Room for improvement: transcriptomics reveal high genetic diversity of drought resistance strategies in winter oilseed rape

Schiessl SV, Orantes-Bonilla M, Quezada-Martinez D, Snowdon RJ: Room for improvement: transcriptomics reveal high genetic diversity of drought resistance strategies in winter oilseed rape (unpublished manuscript)

Room for improvement: transcriptomics reveal high genetic diversity of drought resistance strategies in winter oilseed rape

Sarah Schiessl, Mauricio Orantes-Bonilla, Daniela Quezada-Martinez, Rod Snowdon

Department of Plant Breeding, Justus Liebig University, IFZ Research Centre for Biosystems, Land Use and Nutrition, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

Abstract

Spring droughts are expected to become more frequent in Central Europe as a result of climate change. This will affect biennual crops like winter oilseed rape (*Brassica napus*) during flowering, which is the most sensitive stage for yield building. Breeding for more spring drought tolerant plants is therefore a major adaptation strategy for a higher resilience of agricultural production. However, data on the diversity of genetic regulation of drought response and drought tolerance during this stage under realistic conditions are not available, therefore, no breeding targets or markers are identified so far. Here, we assessed the phenotypic plasticity of drought response for eight adverse *B. napus* accessions under semi-controlled, field-like conditions, and linked their stress strategies to gene and miRNA expression at an early and a late stress time point. We found that drought resistance strategies were highly diverse, both on the phenotypic and on the genetic level. The finding that phenotypically similar accessions have varying degrees of drought resistance due to more effective molecular protection mechanisms reveals good possibilities for genetic improvement of drought resistance in winter oilseed rape by combining positive alleles for ROS scavenging, source/sink ratio and regulation of developmental timing. Moreover, we identified putative *MIRNA* genes in the *B. napus* genome which respond to stress and may also be involved in protective mechanisms, representing possible breeding targets. Our data provide a framework for further validation studies on a single-locus level as well as a valuable resource for applied breeding.

Introduction

The Synthesis Report of the International Panel on Climate Change (IPCC) predicts a medium to high likelihood for an increase in water restrictions in Central Europe in the next decades (IPCC, 2014). This will particularly affect the period of late spring/ early summer (Lu *et al.*, 2019). One of the crops which is likely to be affected by late spring drought is oilseed rape (*Brassica napus*). Winter oilseed rape used to be well adapted to temperate climates, as the highest yields per area are reported in Central Europe and Chile (FAOSTAT 2019). Severe droughts within the last decade (2011, 2015, 2018, 2019) have shown how much changes in precipitation may affect future production of oilseed rape. Oilseed rape is not only a common source for biodiesel and edible oil, but also an important part of crop rotation regimes with cereals (Sieling and Christen, 2015). As irrigation may eventually increase the risk of soil salinity (Tal, 2016), the most important measure to reduce the risk of harvest

reduction is the development of drought tolerant crops. However, data on genetic variation associated to drought stress response in *B. napus* is still scarce, and most of the studies which were conducted on the characterization of drought stress response in *B. napus* investigated plants at the juvenile stage and/or in small pots which do not reflect field conditions (Hatzig *et al.*, 2014; Zhang *et al.*, 2015; Wang *et al.*, 2017). This is most likely not predictive of the expected drought scenario for winter oilseed rape, where drought hits the plants before or during flowering, the most sensitive stage in regards to drought stress (Borghi *et al.*, 2019). One strategy to avoid flowering during stress would therefore be breeding for earlier flowering cultivars. Drought signaling and developmental timing are connected, but this link is unfortunately not well characterized (Kazan and Lyons, 2016). For example, a well-known reaction of plants to drought is drought escape, which is accelerated reproduction due to earlier flowering (Kooyers, 2015). Others have shown that there is a positive correlation between flowering time and water use efficiency (Lovell *et al.*, 2013). Some few flowering genes have been proposed to be involved in drought escape, among them *FRIGIDA (FRI), GIGANTEA (GI), FLOWERING LOCUS T (FT)* and *TWIN SISTER OF FT (TSF*), mostly via influencing the level of abscisic acid (Lovell *et al.*, 2013; Riboni *et al.*, 2013; Kazan and Lyons, 2016). To avoid negative pleiotropic effects between developmental timing and drought resistance, it is therefore necessary to precisely understand the interactions between flowering time regulation and drought signaling. In general, both regulatory networks are well characterized, especially the flowering time network (Srikanth and Schmid, 2011; Blümel *et al.*, 2015), but links between them have been established mainly in *Arabidopsis thaliana* (Kazan and Lyons, 2016). Here, *GI* seems to play a central role to integrate different stress signals into the flowering network, together with plant hormones like gibberellins and IAA, but also micro RNAs (miRNAs) (Kazan and Lyons, 2016). Among the miRNAs involved, miR156 and miR172 obviously play the most important role in drought related signaling to the flowering network by interacting with *SQUAMOSA PROMOTOR LIKE 3* and *9* (*SPL3/9*), and *WRKY44*, respectively (Kazan and Lyons, 2016). As miRNAs are encoded in the genome as *MIRNA* genes, they are interesting targets for plant breeding, but they are still very poorly annotated in crop genomes and are therefore normally not considered as candidates in association studies. Moreover, *B. napus* is a recent allotetraploid (2n=AACC) and therefore has a highly complex genome, which complicates direct transfer of knowledge both for protein-coding genes and *MIRNA* genes.

We therefore wanted to characterize the drought stress response of winter oilseed rape during flowering time in realistic testing conditions, including more genotypic diversity than before. Moreover, we wanted to test if *B. napus* homologs of the same genes would react to drought stress as known from model species, and if so, which copies would do so. Finally, we aimed at identifying respective *MIRNA* genes in the *B. napus* genome which would potentially also be useful breeding targets. To this purpose, we selected eight contrasting winter oilseed rape accessions and grew them in large containers in a system shown to provide good estimates of field performance in both irrigated and non-irrigated conditions (Hohmann *et al.*, 2016). Shortly before flowering, we subjected the plants to a 3 weeks water withdrawal period, reflecting a spring without rainfall. We phenotyped and sampled the plants throughout their life cycle and performed RNAseq and small RNAseq at an early (8 days) and a late (21 days) stress time point. The results show that there is a large phenotypic plasticity in drought resistance strategies, indicating a broad genetic variability which could be combined to improve drought resistance in oilseed rape.

Material and Methods

Plant material, growth conditions and stress treatment

We selected eight contrasting *B. napus* winter accessions from the ERANET-ASSYST diversity set (Bus *et al.*, 2011; Körber *et al.*, 2012) in respect to flowering time and drought tolerance based on earlier field data. We classified Pollen and Musette as early-flowering and drought tolerant, Hokkai3Go and Liporta as early-flowering and drought sensitive, Zephir and NKNemax as late-flowering and drought tolerant and Alaska and Campari as late flowering and drought sensitive. Those accessions were sown in large containers containing 75 % sand and 25 % soil (Hohmann *et al.*, 2016) with 9 plants per container in a water-controlled environment in October 2014 at the field station in Rauischholzhausen, Germany. The containers were left under a transparent roof without temperature regulation, allowing vernalisation during winter. Water demand was determined by regular weighing of the containers with a specialized hydraulic scale, and water was supplied to a water capacity of 60 %. The plants were grown with optimal water supply until the first plants reached BBCH50 (April 2015). The stress containers were not watered subsequently for 3 weeks, while control containers were watered normally. Afterwards, the stress containers were again watered to approximately 60 % water capacity, and all plants were grown until ripening.

Sampling and phenotyping

The plants were sampled at different developmental stages for different tissues. The tissue was cut with a razorblade, put into a 50 ml plastic tube, immediately frozen in liquid nitrogen and subsequently stored at -80°C. We sampled the plants at BBCH18 (rosette stage before vernalisation), BBCH20 (rosette stage after vernalisation), BBCH50 (shooting before flowering and stress application), BBCH60 (opening of the first flowers, 8 days of stress treatment) and BBCH65 (full flowering, 21 days of stress treatment). We used different plants from the same container for each sampling, starting from the left plant in the back. We therefore sampled BBCH18 leaves (left, back), BBCH20 leaves (middle, back), BBCH 50 leaves and buds (right, back), BBCH 60 leaves and buds/ flowers (left, middle) and BBCH 65 leaves and flowers (right, middle) We took the last developed true leaf. At 2 time points, we cut the above-ground part of one plant of each container out in order to determine fresh and dry matter as well as water content. This was the left middle plants directly after stress and the middle front plant at harvest. At harvest, we moreover separated the middle plant into the main raceme and the side branches and counted the number of side branches as well as the number of siliques. We determined beginning (at least 1 plant with open flowers in the container) and end of flowering (no open flowers in the container). Moreover, we regularly counted living primary side branches and measured plant height. In total, we collected data on plant height (PH), number of side branches (NSB), fresh matter (FM), dry matter (DM), water content (WC), number of siliques on the main raceme (SMR), number of siliques on the side branches (SSB), total number of siliques (TNS), number of grains per plant (GPP), numbers of grains per silique (GPS), thousand grain weight (TGW), yield per plant (YPP), yield per container (YPC), seed water content (SWC), beginning of flowering (BOF) and end of flowering (EOF).

Small RNA sequencing

RNA extraction and sequencing

Frozen plant tissue was ground to fine powder using a mortar and a pestle. 50 mg of this powder was used to extract total RNA using the NucleoSpin miRNA extraction kit (Macherey-Nagel) according to the manufacturer's instructions. Small RNA samples (control and stress, leaves) from BBCH60 (early

stress) and BBCH65 (full stress) as well as leaf samples BBCH50 (control and stress) and control samples for Campari for BBCH18, BBCH20 were sent to BGI Genomics Ltd. (Hong Kong, China). Quality control was performed using an Agilent BioAnalyzer 2100 (Agilent). Small RNA sequencing was then performed using the BGISEQ-500 as described in (Fehlmann *et al.*, 2016), and the resulting raw read files were filtered to exclude low quality reads (Q20<90 %, adaptor and null rate>10 %, small RNA tag rate >20 %), reads with 5'-primer contaminants, reads without 3' primers, reads without insert, reads with polyA and reads shorter than 18 nt.

Alignment and coverage calling

Clean reads were mapped onto the *Darmor-bzh* v4.1 reference genome (Chalhoub *et al.*, 2014) using Bowtie2 (Langmead and Salzberg, 2012), version 2.1.0 using the following options: mode --very-sensitive, -q, -p 40, -L 10. A bed file containing all regions covered with at least 9 reads in at least 1 sample was created using custom scripts involving bedtools (Quinlan, 2014) version 2.17.0 and R, version 3.1.2. The raw coverage of all pooled samples on those regions was called using bedtools multiCov, and the coverage was normalized to the number of reads per sample and to the region length. Regions with a mean normalized coverage >10 were selected for further analysis.

MIRNA gene prediction and annotation

The selected regions were analysed using NovoMIR version 1.11 (Teune and Steger, 2010), and the regions predicted to be a MIRNA gene were selected for further analysis. The reads mapping to those regions were extracted using bedtools, and the frequency of each putative miRNA was counted in R for every sample. For every sample, the putative miRNA list was filtered for sequences having a count>10. All sample lists were merged by sequence, and the mean frequency per sequence was determined. This matrix was then filtered for sequences with a mean abundance >5 to avoid analysis of low expression miRNAs or degradation products. Likewise, we found 2649 putative miRNA sequences which were expressed in at least one sample in any condition and time point. Those sequences were blasted against different databases to attain a suitable annotation. Those databases contained (1) the sequences of both the mature and the precursor miRNAs of all available *Brassicaceae* from mirbase (Griffiths-Jones *et al.*, 2008), (2) the sequences of both the mature and the precursor miRNAs of all other plant species, (3) the small RNA sequences from Rfam database (Kalvari *et al.*, 2018), (4) the sequences from the TIGR Repeats database (Ouyang and Buell, 2004), representing repetitive elements and (5) the *B. napus* chloroplast genome. All sequences were then hierarchically annotated, meaning that sequences which could be annotated with (1) were excluded from the annotation process with (2). We then extracted only those 2067 small RNA sequences which were perfectly matching the respective sequence in a predicted *MIRNA* gene, excluding all small RNA sequences with gaps and mismatches to avoid including sequencing errors. We noticed that many of those are mere length variants, indicating beginning degradation, partly specific to certain samples (Campari at early stress in particular), hampering direct comparisons of small RNA sequence abundance between samples. Instead of analyzing small RNA sequence expression on their own, we therefore analyzed *MIRNA* gene expression by adding up expression values of all small RNA sequences perfectly matching a putative *MIRNA* gene. It has to be noted that 794 small RNA sequences matched more than one single *MIRNA* gene and were included into all matching *MIRNA* gene expression patterns, as distinction was not possible. At the same time, those *MIRNA* gene regions were annotated using the respective annotations from the small RNA sequences. In annotation conflicts, we combined annotations, allowing ambiguity to avoid loss of information.

Prediction of target genes

The list of filtered putative miRNAs was used as input using the web-based program psTargetRNA (http://plantgrn.noble.org/psRNATarget/analysis), with *Brassica napus* unigene transcript v5 as database and Default settings for release 2017, but Expectation cutoff lowered to 3. The output was merged with the results of a BLAST2GO analysis for the total genome in order to provide the best available target gene annotation.

RNAseq

RNA extraction and sequencing

For RNAseq, we sequenced all leaf samples from BBCH60 and BBCH65 for both control and stress. Five samples were out of stock and could not be sequenced. These were for BBCH 60 (early stress): 1 biological replicate of Musette in stress conditions, and for BBCH65 (late stress): 1 biological replicate of Zephir in control conditions and each one biological replicate of Campari, Liporta and Musette under stress conditions. Therefore, each accession and treatment was still replicated. The leaf samples were sent to Novogene Corporation (Hong Kong, China) for RNA extraction and RNA sequencing.

Alignment and TPM calculation

Quality control of the downloaded reads was performed using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Accordingly, adapter removal and trimming was performed using Trimmomatic version 0.38 (Bolger *et al.*, 2014) by first removing TruSeq adapters followed by head cropping the first 9 bases. Clean reads were mapped onto version 4.1 of the *B. napus* 'Darmor-Bzh' reference genome sequence assembly (Chalhoub *et al.*, 2014) using Bowtie2 (Langmead and Salzberg, 2012), alignment mode "very-sensitive". Removal of duplicates, sorting and indexing was carried out with *samtools* version 0.1.19 (Li *et al.*, 2009). We then calculated TPM *bedtools* software with multiBamCov (Quinlan, 2014) for raw coverage calling, followed by normalization in R (version 3.1.2) according to (Wagner *et al.*, 2012) for the sum of all exons of every gene. From this, we calculated $|\log_2(\text{mean(TPM(control))/mean(TPM(stress))}|$ for each accession at both stress time points, and selected genes with more than half of the accessions having a value >1.

WGCNA

Weighted Gene Co-Expression Network Analysis was performed in RStudio/Bioconductor using the libraries "AnnotationDbi", "impute", "GO.db", "preprocessCore" and the package "WGCNA". As genotypic input, we used the TPM values of all genes with a mean TPM value >2, separately so for each stress time point. As phenotypic input, we used the mean phenotypic value for each of the 36 traits we have measured (see above). Outliers were excluded based on genetic distance tree plots, and soft-thresholding power was chosen manually for each data set to optimize mean connectivity.

Validation of gene expression by RT-qPCR

To confirm the changes in gene expression due to stress in some of the predicted candidate genes, we selected two copies of *SPL* genes (*Bna.SPL3.A04, Bna.SPL4.C03*) for RT-qPCR validation. Moreover, we also confirmed gene expression changes for a copy of *FT*, *Bna.FT.A02*, to confirm genetic regulation of the main flowering regulator. RT-qPCR was performed for leaf samples taken from the accession Campari at BBCH18 (before vernalisation), BBCH20 (after vernalisation), BBCH50 (before flowering and stress), BBCH60 (beginning of flowering, early stress) and BBCH65 (full flower, late stress) for Campari, control plants, and BBCH50, BBCH60 and BBCH65 for Campari, stress plants. RNA was extracted as described above (see Small RNA sequencing). cDNA synthesis was performed using

the RevertAid cDNA synthesis kit (ThermoFisher) using 1 µg of total RNA and random primers. The amount of cDNA was quantified using the Qubit DNA High Sensitivity kit on a Qubit fluorimeter. Quantitative Real-time PCR was performed on a Real-Time PCR System ViiA7 cycler (Applied Biosystems) in 384 well plates. The reaction mix containing specific primers, the template cDNA and FastStart Universal SYBR Green Master mix containing Rox (Roche) was pipetted by a robot (Biomek 4000, Beckman Coulter). As endogenous control, we used ubiquitin. PCR efficiency was measured using a pool of all samples in a dilution series of 6 points. All samples were measured in 3 technical replicates. The normalized expression level was determined using the ΔCt method (Livak and Schmittgen, 2001). The primer sequences can be found in ***Table 1***.

Table 1: Primer sequences as used for RT-qPCR validation.

Gene copy	**Gene ID**	**Primer sequence (5'-3')**
Bna.SPL3.A04	*BnaA04g19840D*	CACCAAGAGGAGTTGCAGGAGA
		CAGAAGAGAGCAAGCAAAGGCTAA
Bna.SPL4.C03	*BnaC03g18800D*	GCCAAAGCTCCTGTTGTTCG
		ACGGAGGTCATGAAACCTGCT
Bna.FT.A02	*BnaA02g12130D*	CTTACGAGAGTCCAAGGCCC
		ACGGGAAGGCCGAGATTGTA
	Ubiquitin	CTTCTTCGGCCTCAACTGGTT
		GAAGATGATCTGCCGCAAGTGT

Results

Drought stress delayed plant development

Plants suffering from drought stress showed a significant developmental delay in all accessions, although there were differences in the severity of the delay. Only one accession (NKNemax) started flowering significantly later under stress, but all accessions ended flowering significantly later under stress (Figure 1a, 1b). This delay was still visible at harvest, where the seed water content of stressed plants was still significantly higher in Alaska, Campari, Musette and NKNemax, indicating delayed ripening (Figure 1e, 1f). Stress significantly and strongly reduced plant height in all accessions at all time points until harvest (Figure 1c, 1d). In contrast, while fresh matter was significantly reduced after three weeks of drought stress in all accessions, only Campari showed a reduced fresh matter at harvest. Dry matter was significantly reduced in Campari, Hokkai3Go, Musette, NKNemax and Pollen directly after stress and in Alaska, Campari, Hokkai3Go and Musette at harvest (Supplementary Figure S1). The number of side branches showed altered dynamics for some stressed and control accessions (Figure 1i-l). Generally, the number of side branches first increased slightly until flowering, but then started to decrease as side branches died off. During stress and directly after stress, NKNemax showed a significantly reduced number of side branches, but this difference was overcome 6 weeks after stress and at harvest. Campari showed a significantly reduced number of side branches after stress, but an increased number of side branches 6 weeks after stress. Pollen showed a significantly higher number of side branches 6 weeks after stress, but this difference was lost until harvest. Alaska and Zephir increased their number of side branches significantly 6 weeks after stress and kept this increased number until harvest. Plant water content was significantly decreased after stress, but recovered until harvest and was not significantly different in any accession at harvest. Five accessions had significantly reduced average plant yield at harvest (Alaska, Campari, Hokkai3Go, Musette and Zephir), while the yield of the middle plants was only significantly reduced in Zephir (Figure 1g, 1h). The number of siliques did not show a significant reduction in any genotype, neither for the main raceme, the side branches or the total plant, although NKNemax showed a strong trend for an increase under stress (Supplementary Figure S2). Hokkai3Go and Pollen did not show any change in hundred grain weight, while Alaska, Campari, Musette and NKNemax reduced it and Liporta and Zephir increased it under stress (Supplementary Figure S4). We conclude that the stress level applied was severe, but as Liporta, Pollen and NKNemax did not reduce yield significantly, they must have been able to cope with water scarcity. In the present stress scenario, they can therefore be regarded as drought tolerant, while Alaska, Campari, Hokkai3Go, Musette and Zephir can be regarded as drought sensitive. Based on those results, we grouped the eight accessions into four different ideotypes: Alaska/Musette (reduced grain weight), Campari/NKNemax (less grains per silique, reduced grain weight), Hokkai-3-go/Pollen (no yield component reduced) and Liporta/Zephir (no biomass reduction, higher grain weight) (***Figure 2***).

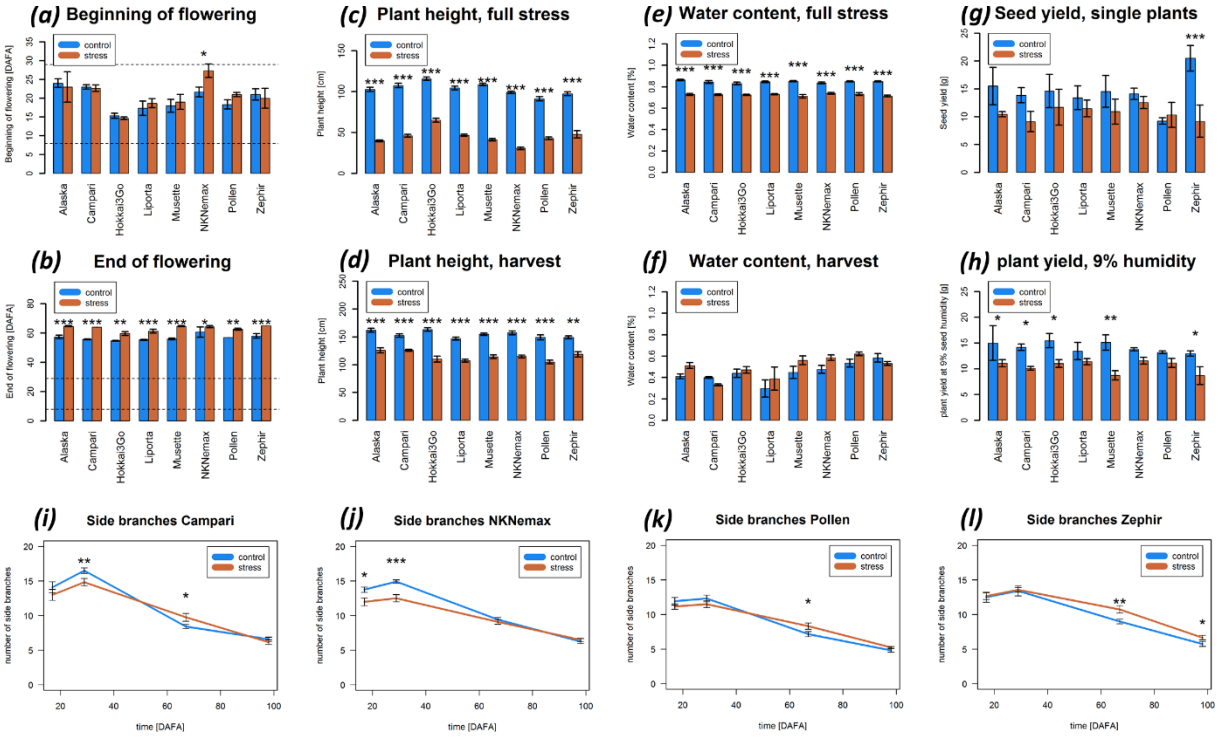


Figure 1: Barplots for (a) Beginning of flowering (b) End of flowering (c,d) plant height at full stress and at harvest (e,f) Water content at full stress and at harvest (g) Single plant yield (h) corrected average plant yield per container. (I,j,k,l) represent line plots for number of side branches for 4 different accessions: Campari, NKNemax, Pollen and Zephir, respectively. The dotted lines in (a,b) represent the stress period. All values represent 3 biological replicates. Whiskers represent standard errors of the mean (SEM). Significant differences are indicated at $p < 0.05$*, $p < 0.01$ $p < 0.001$***.**

Under well-watered conditions, seed yield at 9 % humidity correlates positively with fresh and dry matter at full bloom (full stress), with plant height before flowering, at full bloom and at harvest, and with the total number of siliques at harvest (Supplementary Figure S5). Under stress conditions, seed yield correlates negatively with fresh and dry matter at full stress and with the number of side branches at early and full stress (Supplementary Figure S6). Taking both treatments together, seed yield correlates positively to plant height at all time points after begin of stress treatment, to fresh and dry matter at full stress and at harvest and also weakly to the total number of siliques. Also, seed yield correlates negatively to end of flowering and to seed humidity at harvest (Supplementary Figure S7). Together, those results indicate that the developmental delay under stress is associated to seed yield reduction.

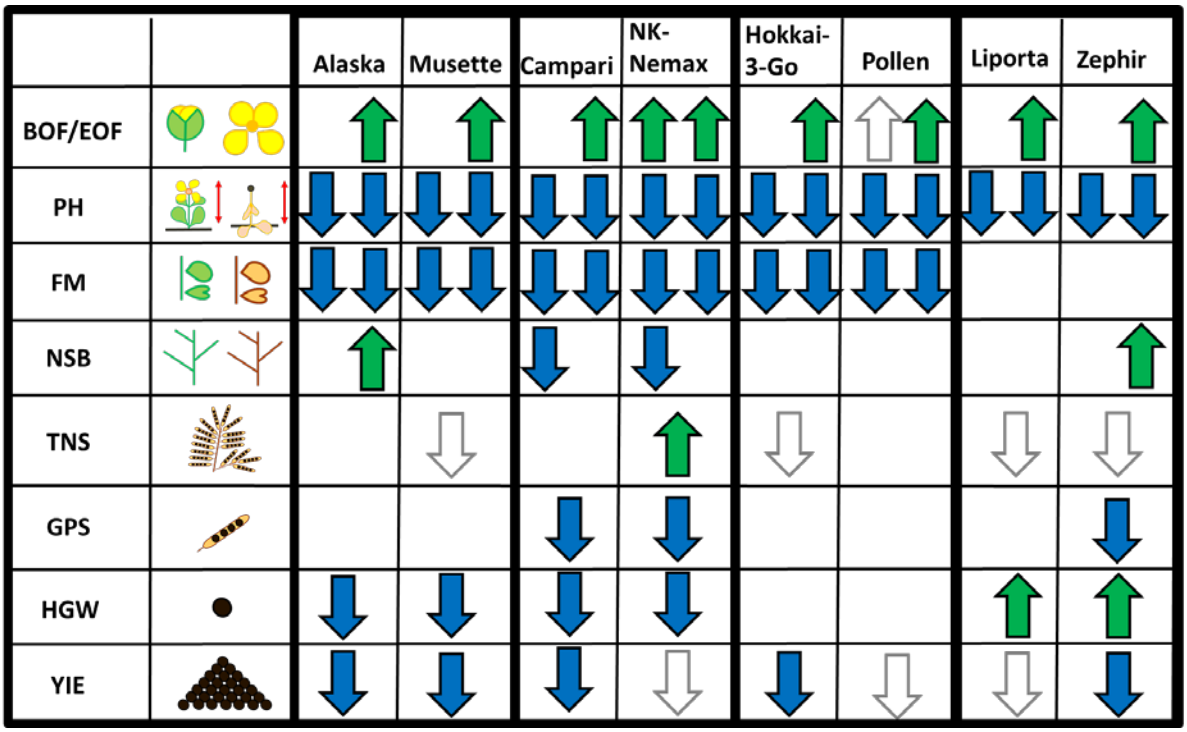


Figure 2: Graphical summary of the behavior of the most important life cycle, biomass and yield traits under stress. Accessions with similar phenotypic behavior were grouped into ideotypes (thick black boxes). Filled arrows indicate significant reductions; grey non-filled arrows indicate non-significant trends. For biomass parameters, the left arrow indicates the behavior at full stress, the right arrow indicates the behavior at harvest. BOF: Beginning of flowering, EOF: End of flowering, PH: Plant height, FM: Fresh matter, NSB: Number of side branches, TNS: Total number of siliques, GPS: grains per silique, HGW: hundred grains weight, YIE: total container yield at 9 % humidity.

Specific target gene expression patterns associate to stress reaction

To link the phenotypic variation to the respective gene expression variation, we performed a Weighted Gene Co-expression Network Analysis (WGCNA) for both early and full stress using the TPM expression values of the RNAseq analysis with a mean value >2 as genotypic input and the 36 phenotypic traits data as phenotypic input. Likewise, we found 8 modules strongly associated to main developmental and yield traits like plant height, flowering time and seed yield for the early stress time point (Figure 3), and 6 modules for the full stress time point (Supplementary Figure S8). For early stress, the modules positively associating with yield were named “lightyellow”, “magenta”, “midnightblue” and “turquoise”, the modules negatively associated to yield were named “pink”, “darkgreen”, “red” and “white”. For late stress, the modules positively associating with yield were named “salmon”, “greenyellow”, and “yellow”, the modules negatively associated to yield were named “blue”, “midnightblue”, and “turquoise”. According to the following GO term enrichment analysis, the genes in these modules were mainly associated with stress reaction and photosynthesis at early stress (***Table 2***). At full stress, the modules were associated to stress reaction, rRNA reorganisation, photosynthesis and sucrose transport (***Table 3***).

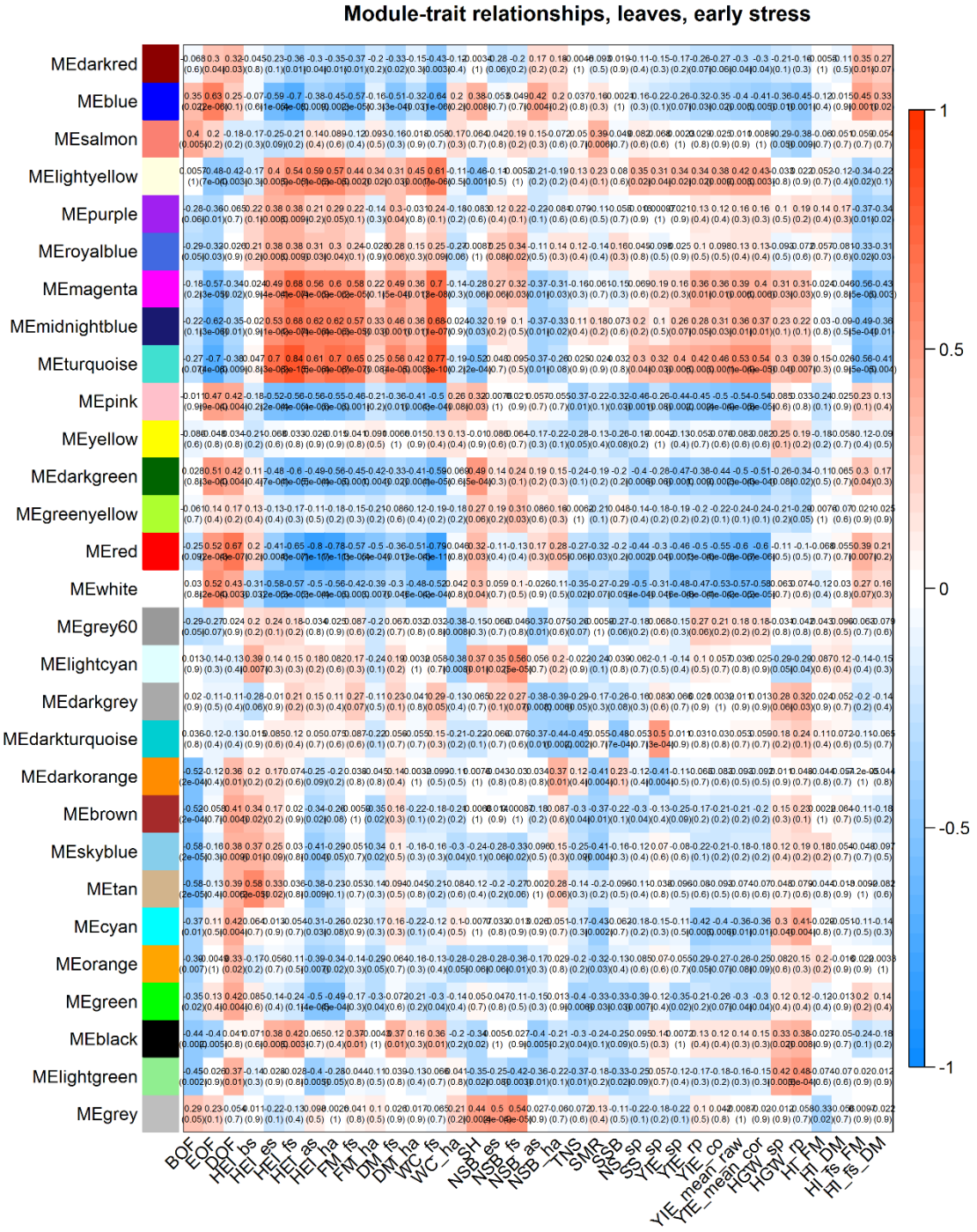


Figure 3: Heatmap of the gene module-trait correlation for 36 phenotypic traits for leaves at early stress. Each module is named with a random color. The number on top of each cell represents the correlation coefficient (r). The number below in brackets represents the respective p-value.

Table 2: Top five GO terms found for different associated modules from WGCNA based on RNAseq data for leaves, early stress, grouped by the major biological processes found. Each module was named with a random color. The table gives the p-value from the hypergeometric test indicating how significant the enrichment has been, and the OddsRatio, indicating how strongly the GO term has been enriched within this set of genes compared to a random gene set of the same size. GO terms were filtered for OddsRatio>5, and the 10 highest ranking GO terms were selected for classification. For each process, the top five GO terms were selected for presentation.

Top 5 terms associated to photosynthesis			
Pvalue	**OddsRatio**	**Term**	**module**
1.488E-160	62.40	photosynthetic electron transport in photosystem I	turquoise
6.52333E-15	60.97	photosynthesis, light harvesting in photosystem I	turquoise
2.28789E-53	59.38	carbon fixation	turquoise
2.01237E-26	51.62	energy coupled proton transport, down electrochemical gradient	magenta
2.00462E-16	50.28	photosynthesis, light harvesting in photosystem II	turquoise
Top 5 terms associated to stress reaction			
Pvalue	**OddsRatio**	**Term**	**module**
2.52286E-14	55.69	response to desiccation	lightyellow
3.08514E-26	27.53	detection of stimulus	midnightblue
2.81461E-38	25.17	systemic acquired resistance	midnightblue
2.66296E-37	22.38	defense response, incompatible interaction	midnightblue
7.33098E-12	20.50	calcium ion transport	lightyellow
Top 5 terms associated to protein turnover			
Pvalue	**OddsRatio**	**Term**	**module**
9.53666E-21	35.27	protein repair	turquoise
6.54545E-21	22.95	translational elongation	pink
1.52269E-42	21.54	proteasome core complex assembly	magenta
3.32778E-40	17.54	proteasome assembly	magenta
1.01956E-39	12.29	ribonucleoprotein complex biogenesis	pink
Top 5 terms associated to RNA regulation			
Pvalue	**OddsRatio**	**Term**	**module**
6.07872E-86	28.58	RNA methylation	pink
4.0265E-20	20.88	purine nucleoside triphosphate biosynthetic process	magenta
1.54455E-71	18.94	RNA modification	pink
2.06129E-58	12.33	macromolecule methylation	pink
8.40659E-11	7.44	pyrimidine ribonucleotide biosynthetic process	pink
Top 5 terms, non-specified			
Pvalue	**OddsRatio**	**Term**	**module**
1.43763E-64	28.90	cellular metabolic compound salvage	magenta
3.36662E-20	28.14	regulation of response to biotic stimulus	midnightblue
1.26008E-19	26.13	regulation of multi-organism process	midnightblue
3.93245E-11	25.77	multidimensional cell growth	lightyellow
3.02441E-19	24.88	detection of biotic stimulus	midnightblue

Table 3: Top five GO terms found for different associated modules from WGCNA based on RNAseq data for leaves, full stress, grouped by the major biological processes found. Each module was named with a random color. The table gives the p-value from the hypergeometric test indicating how significant the enrichment has been, and the OddsRatio, indicating how strongly the GO term has been enriched within this set of genes compared to a random gene set of the same size. GO terms were filtered for OddsRatio>5, and the 10 highest ranking GO terms were selected for classification. For each process, the top five GO terms were selected for presentation.

Top 5 terms associated to photosynthesis			
Pvalue	**OddsRatio**	**Term**	**module**
8.21338E-32	749.05	photosynthesis, light harvesting in photosystem II	greenyellow
2.94125E-80	230.83	photosynthesis, light harvesting	greenyellow
1.5595E-186	158.77	photosynthesis	greenyellow
3.0268E-172	143.38	photosynthesis, light reaction	greenyellow
2.51985E-47	85.39	photosynthetic electron transport in photosystem I	greenyellow
Top 5 terms associated to stress reaction			
Pvalue	**OddsRatio**	**Term**	**module**
5.74193E-17	10.02	response to superoxide	yellow
5.74193E-17	10.02	response to oxygen radical	yellow
8.63002E-13	8.54	response to desiccation	yellow
3.71837E-16	7.90	vacuole organization	blue
1.04078E-35	7.60	hyperosmotic salinity response	yellow
Top 5 terms associated to rRNA processing			
Pvalue	**OddsRatio**	**Term**	**module**
1.33054E-14	58.10	rRNA export from nucleus	turquoise
5.11907E-87	53.60	rRNA processing	greenyellow
1.46862E-12	41.40	ribosomal small subunit assembly	turquoise
1.04896E-19	34.01	maturation of SSU-rRNA	turquoise
4.37688E-14	27.94	cleavage involved in rRNA processing	turquoise
Top 5 terms associated to sucrose transport			
Pvalue	**OddsRatio**	**Term**	**module**
1.04862E-13	32.84	sucrose transport	blue
7.56256E-13	26.68	oligosaccharide transport	blue
Top 5 terms, non-specified			
Pvalue	**OddsRatio**	**Term**	**module**
7.10372E-50	266.95	protein-chromophore linkage	greenyellow
2.14121E-70	70.69	response to far red light	greenyellow
1.20184E-10	10.49	sphingoid biosynthetic process	blue
1.05785E-17	9.10	autophagy	yellow
2.55797E-06	7.37	cold acclimation	yellow

Resistance gene candidates

We wanted to find genetic patterns associated to a higher drought stress resistance. The phenotypic data have shown that similar strategies in yield component reduction did not result in the same level of resistance. For example, Liporta and Zephir showed a comparable strategy when comparing developmental traits and yield components, but Liporta is resistant and Zephir is not (***Figure 1***). We assumed that comparing gene expression between resistant and tolerant accessions of the same ideotype would be revealing in terms of candidate resistance genes. In detail, we assumed that there are genes which fulfill a protective function, leading to a reduced subjective stress level like reactive oxygen species (ROS) scavenging enzymes, and genes which respond to a subjective stress level, like genes related to metabolism. In this model, resistant accessions would express protective genes earlier or more strongly, allowing for a better growth and yield building (***Figure 4***). We therefore first identified differentially expressed genes (DEGs) between control and stress plants in leaf tissue at

early and late stress. Those make the totality of stress-protective and stress-responding genes. We found 6885 differentially expressed genes (DEGs) for leaves at early stress and 6066 DEGs for late stress. At early stress, there was a strong enrichment in photosynthesis-related genes and genes involved in chlorophyll synthesis for the downregulated genes, while stress-related and root development genes tended to be upregulated (***Figure 4***). At full stress, this is no longer the case, although photosynthesis DEGs are still partially downregulated compared to the control. However, we noticed that there is a general decrease in expression of photosynthesis DEGs and a general increase in expression of stress-related DEGs between the early and late sampling point in control plants, while translation-related DEG expression decreased, possibly due to leaf aging. Alaska, NKNemax, Pollen and Zephir showed a very pronounced reaction at early stress, while Campari and Liporta showed a medium response and Hokkai-3-go and Musette reacted weakly. At full stress, no accession showed such a strong response anymore, Alaska, Liporta, Musette and Zephir still showed a medium response, while the remaining reacted only weakly. In general, the stress response was much less synchronized at full stress than at early stress.

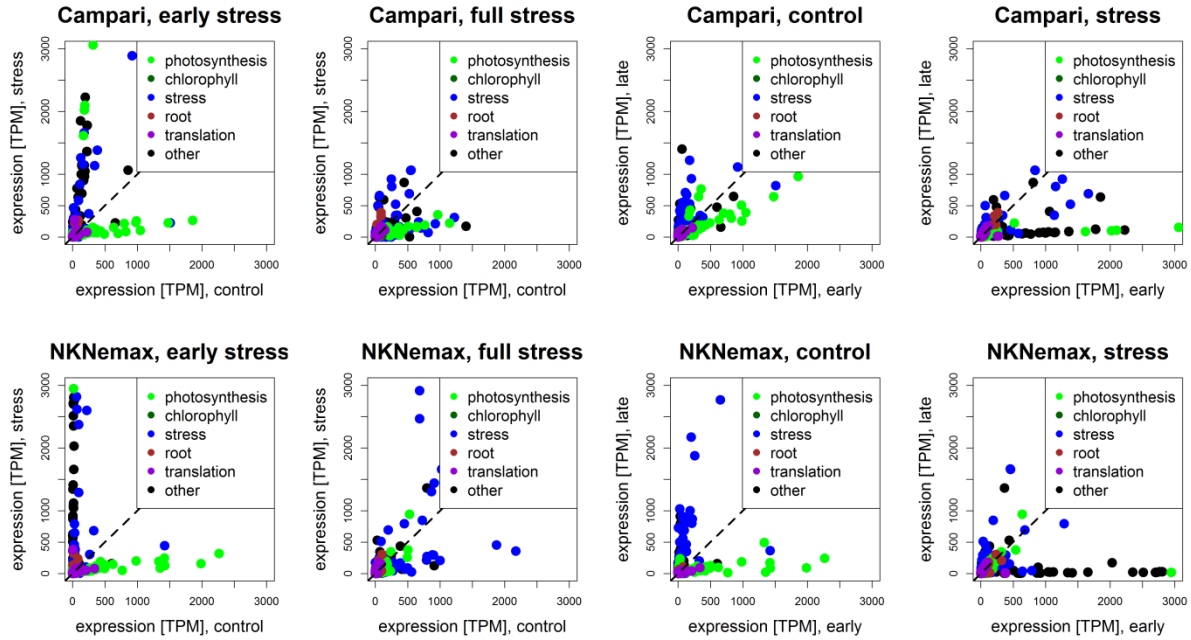


Figure 4: Scatterplots of gene expression values of all DEGs for Campari (sensitive) and NKNemax (resistant). The color code indicates the process in which the gene copy is involved. Upregulated genes show up at the left side of the plot, while downregulated genes show up at the right side of the plot.

We then compared gene expression under stress for the three resistant accessions to the respective susceptible accessions from the same ideotype (Campari/ NKNemax, Hokkai-3-go/ Pollen and Zephir/ Liporta, see also **Figure 2**). If a DEG was upregulated under stress, we assumed a protective gene if the resistant accession would express it more strongly under stress compared to the susceptible one. In contrast, if a DEG was downregulated under stress, we assumed a protective gene if the resistant accession would express it less strongly compared to the susceptible accession. On top of the three ideotype pairs, we also compared all three resistant accessions to all five susceptible accessions.

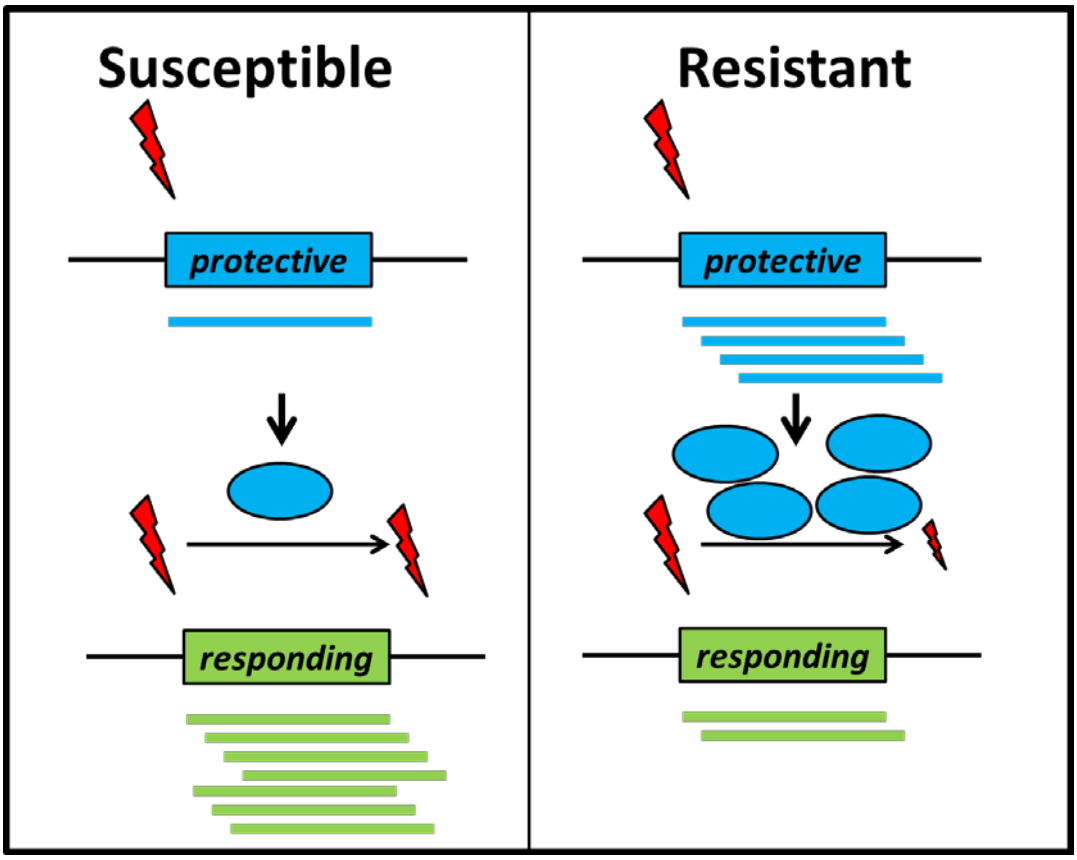


Figure 5: Graphical representation of the assumed gene expression model for detection of candidate resistance genes. In resistant accessions, protective genes would react more sensitively to stress, reducing the subjective stress level. Therefore, stress-responding genes would be expressed less strongly.

The results of this analysis are summarized in Table 4. In total, we identified 591 candidate resistance genes at early stress and 632 candidate resistance genes at late stress. At early stress, one candidate gene copy was found in all datasets (*Bna.MAP3K13.C05*), while there was no such finding at full stress. The expression patterns of *Bna.MAP3K13.C05* are shown in ***Figure 5***. We also analysed the GO term enrichment in the respective data sets. We found that at early stress, the candidate resistance genes for Campari/NKNemax were mainly involved in response to reactive oxygen species and oxidative stress as well as small molecule biosynthesis. At full stress, the respective candidate genes were mainly involved in the plant's innate immune system and in programmed cell death. For Hokkai-3-go/Pollen, the top GO terms mainly involved biosynthesis pathway, like isoprenoid biosynthesis or chlorophyll synthesis, at early stress, and terms related to the response to plant hormones like jasmonic acid and salicylic acid at full stress. The pair Zephir/Liporta mainly differed in genes involved in photosynthesis, response to carbohydrate and ion transport at early stress and in genes involved in response to light and reactive oxygen species at full stress.

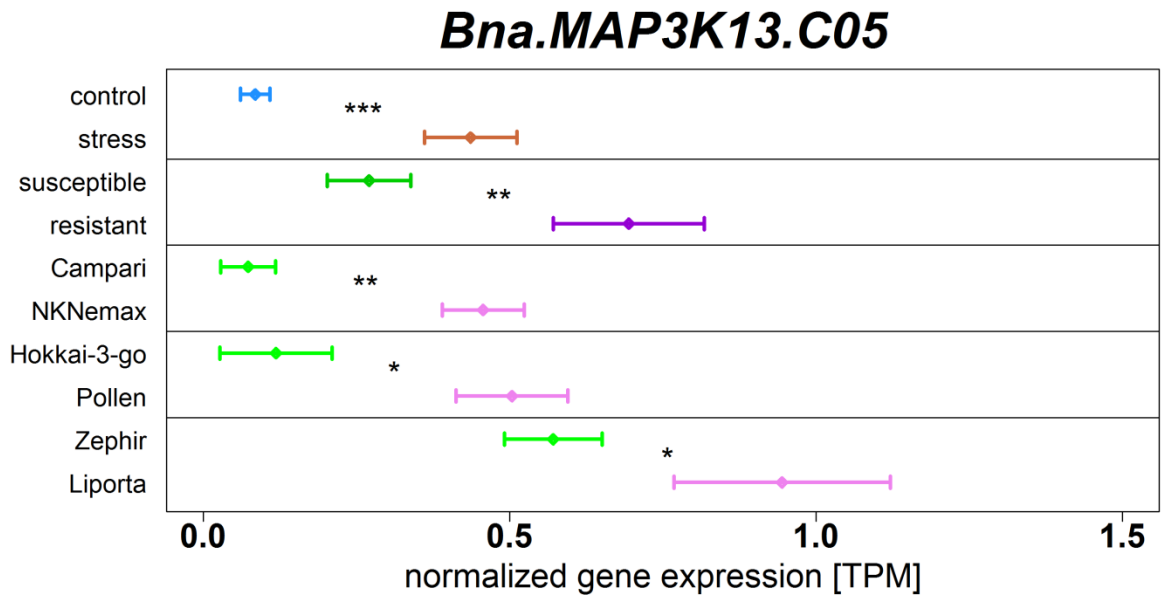


Figure 6: Normalized gene expression for the *B. napus* C05 copy of *mitogen-activated protein kinase kinase kinase 13* (*BnaC05g04950D*) at early stress. Values show mean values over all control accessions (control), all stress accessions (stress), all five susceptible accessions (susceptible) or all three resistant accessions (resistant) or mean values of three biological replicates of the respective indicated accession. Whiskers represent standard errors of the mean (SEM). Significant differences are indicated at p < 0.05*, p < 0.01 p < 0.001***.**

Some miRNA families have stress-specific variants

We then performed small RNA sequencing to identify putative miRNAs linked to stress or stress-specific gene expression patterns. We mapped them on the *Darmor-bzh* v4.1 reference genome and extracted genomic regions region of a minimum coverage predicted to be a *MIRNA* gene. We ended up with 507 putative expressed *MIRNA* genes, 196 on the A subgenome and 311 on the C subgenome.

For those *MIRNA* genes, we separately compared the cumulative normalized gene expression for all eight accessions between control and stress at early and at late stress. Differential gene expression was assumed if two-sided Student's t-test resulted in a p-value < 0.005 and the fold change between control and stress exceeded 2 in at least one accession. Likewise, we found that 65 and 64 putative *MIRNA* gene regions were differentially expressed at early stress and full stress, respectively, and were denoted as DEM (Differentially Expressed *MIRNA*). A majority of DEMs were downregulated: 44 of 64 DEMs were downregulated at early stress and 46 of 64 DEMs at full stress. About half of the DEMs were overlapping between time points, showing the same pattern in all cases. The full list of identified DEMs can be found in **Table 5**. The results show that not all *MIRNA* gene regions with the same annotation react equally to stress. For example, only 4 out of 23 members of miR156 family are upregulated at early stress, but 9 are upregulated at full stress, indicating that some members do not react to drought stress, while others have altered dynamics. This shows that some miRNA families in *B. napus* have stress-specific and stress-unspecific variants. Interestingly, more than twice of the DEMs were located in the C subgenome, exceeding the ratio between all *MIRNA* gene regions.

Table 4: List of differentially expressed MIRNA gene regions (DEMs) at early and full drought stress. *The number stands for the number of members with the respective*

annotation. ns: not significant, down/up: down/upregulated with a fold change of at least 2 at a p-value<0.005 in at least one of 8 accessions.

	early stress			**full stress**			
DEM	**ns**	**down**	**up**	**ns**	**down**	**up**	**total**
miR156	19	0	4	14	0	9	23
miR157	0	6	0	0	6	0	6
miR158	6	1	0	6	1	0	7
miR160	11	3	0	12	2	0	14
miR162/miR1512/miR9563/miR9568	26	0	1	24	0	3	27
miR164/miR3439/miR3446/miR9563	17	0	8	25	0	0	25
miR167	5	2	0	7	0	0	7
miR169	2	7	0	5	4	0	9
miR319	10	6	0	12	4	0	16
miR391	3	0	0	0	3	0	3
miR395	3	2	0	3	2	0	5
miR397	0	1	0	0	1	0	1
miR398	0	1	0	0	1	0	1
miR399	4	4	0	4	4	0	8
miR404	21	1	0	22	0	0	22
miR5718	0	2	0	0	2	0	2
miR5725	0	1	0	0	1	0	1
miR6030	3	0	0	0	3	0	3
miR7504/miR9563	15	0	0	10	5	0	15
miR827	0	1	0	0	1	0	1
miR858	1	0	1	1	0	1	2
miR952	1	0	0	0	0	1	1
miR9558	0	2	0	1	1	0	2
miR9563	15	1	2	14	3	1	18
MIRNA_pred_A02_2	0	0	1	0	0	1	1
MIRNA_pred_C01_7	0	1	0	1	0	0	1
MIRNA_pred_C03_11	1	0	0	0	1	0	1
MIRNA_pred_C04_1	0	0	1	1	0	0	1
MIRNA_pred_C07_7	1	1	0	2	0	0	2
MIRNA_pred_C08_7/C09_6/Cnn_1	1	0	2	1	0	2	3
MIRNA_pred_Cnn_4	0	0	1	1	0	0	1
MIRNA_pred_Cnn_8	1	1	0	1	1	0	2

We predicted possible target genes for the small RNA sequences using psTargetRNA and compared the expression patterns of their respective DEMs to the expression patterns of their predicted targets. With the exception of two previously not annotated *MIRNA* genes, all DEMs had predicted target genes associated to them. Consequently, we checked if the putative target gene would also be a DEG at the same time point. Four DEMs had no associated target DEG (miR395, miR404, miR827 and a previously not annotated one). In the case of further four DEMs, the target DEGs all showed an unexpected pattern, as the change in expression was directly and not indirectly proportional. We all the same have kept them in the data set to avoid assumption bias.

The detected DEM/DEG pairs point to an involvement of miRNAs in all major processes we have observed to be affected under stress with WGCNA: developmental regulation, stress response and photosynthesis. For example, we found several members of the miR156 family being strongly upregulated under stress, connected to downregulation of several copies of *SQUAMOSA PROMOTOR LIKE (SPL)* genes, namely *SPL1, SPL3, SPL4, SPL10* and *SPL11*. Other DEMs associated to DEGs involved in flowering time regulation were miR169, miR319, miR398 and miR7504/miR9563. We also found DEMs in connection with upregulation of superoxide dismutase and DREB genes (miR398), chaperones (miR157) and serine-threonine-kinases (miR952). Moreover, several DEMs were found to be associated to DEGs like PSII-subunit Q2 (miR164/miR3439/miR3446/miR9563), phosphate dikinase (MIRNA_pred_A02_2) and seduheptulose-bisphosphatase (MIRNA_pred_C08_7/ C09_6/Cnn_1).

Interestingly, differential dynamics were also observed in terms of DEM/DEG pairs. For example, we detected ten DEMs annotated with miR156. Four were upregulated at early stress; nine were upregulated at full stress, while three overlapped between time points. However, they also seemed to target different gene copies: at early stress, *Bna.SPL3.A04*, *Bna.SPL11.A06*, *Bna.SPL10.A07* and *Bna.SPL1.C04* were downregulated. At full stress, this affected *Bna.SPL4.C03*, *Bna.SPL1.C07* and *Bna.SPL10.C09*. Other copies of the same genes did not react at all, possibly indicating co-evolution of certain miRNA/ gene pairs.

We noticed that seven DEGs among those DEM/DEG pairs were also candidate resistance genes at each time point. However, at early stress, the DEM expression pattern did not allow conclusions on a respective tolerance mechanism. For late stress, miR398 targeting two copies of superoxide dismutase showed a pattern in line with a possible association to tolerance in Liporta.

Finally, we randomly selected two copies of *SPL* genes (*Bna.SPL3.A04, Bna.SPL4.C03*) and one copy of the main flowering regulator *FLOWERING LOCUS T* (*FT*), *Bna.FT.A02*, for RT-qPCR validation in one accession (Campari) over five sampling points throughout life time: before vernalisation (BBCH18), after vernalisation (BBCH20), before flowering (BBCH50), at beginning of flowering (BBCH60, early stress time point) and at full flower (BBCH65, late stress time point). For the overlapping time points, the data are in good concordance with the data retrieved from RNAseq (Figure 7). The time course shows that under control conditions, as expected, *Bna.FT.A02* is strongly increasing after vernalisation and with the onset of longer days, to reach a maximum at full flower. Under stress conditions, *Bna.FT.A02* gene expression increases less strongly until beginning of flowering and then decreases again, indicating that the regulation of onset of flowering is under stress-sensitive control. For *Bna.SPL3.A04*, there was a visible increase in gene expression until beginning of flowering, followed by sharp decrease in control conditions. In stress conditions, *Bna.SPL3.A04* does not peak at beginning of flowering, but is significantly reduced in gene expression compared to the control. For *Bna.SPL4.C03*, no differences between control and stress conditions were found in Campari.

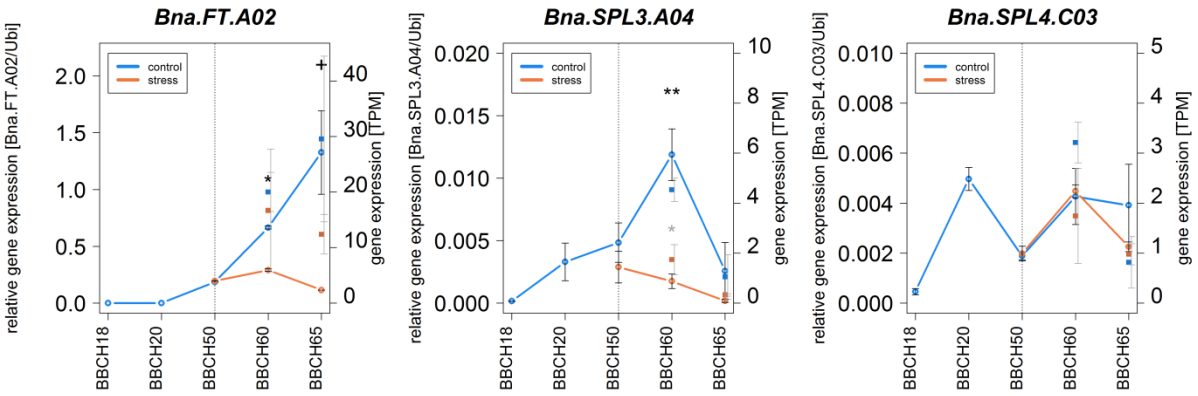


Figure 7: Gene expression in the accession Campari as measured by RT-qPCR (left axis) and by RNAseq (right axis) for *Bna.FT.A02, Bna.SPL3.A04* and *Bna.SPL4.C03* at five different time points: before vernalisation (BBCH18), after vernalisation (BBCH20), before flowering (BBCH50), and at the beginning of flowering (BBCH60, early stress time point) and at full flower (BBCH65, late stress time point). Beginning of stress treatment is indicated by the dotted line. RT-qPCR results are shown as line plot with dots and black whiskers, RNAseq results are shown as squares with grey whiskers beside to avoid overlap. Whiskers represent standard errors of the mean (SEM). Black asterisks show significant differences for RT-qPCR, grey asterisks show significant differences for RNAseq. Significant differences are indicated at $p<0.1^{+}$, $p < 0.05$*, $p < 0.01$.**

Discussion

The stress treatment we have conducted was designed to reflect a spring without rainfall, a scenario expected to occur significantly more often in Central Europe due to climate change (Lu *et al.*, 2019). Central Europe is a major growing area for oilseed rape with the highest yields per area worldwide (FAOSTAT 2019). The container system we used to simulate drought conditions allows for a close-to-normal root development and has been tested before to be well-suited to predict field performance (Hohmann *et al.*, 2016). Our results are therefore of high practical relevance for climate adapted breeding in this important oilseed crop. In the present study, three out of eight accessions accessions could keep their yield level even under stress conditions, while the other five showed varying levels of yield reduction (26-42 % reduced yield level). In an earlier study, we found that seed quality of the same plants was also reduced under stress, although this was independent of yield level (Hatzig *et al.*, 2018). Phenotypic evaluation indicates that there is no general drought tolerance or susceptibility mechanism among the accessions, but plant development was strongly delayed in all accessions under water-limited conditions, indicating a water saver strategy (Urban *et al.*, 2016). In detail, we noted that both Liporta and Zephir did not reduce their biomass under stress, and in consequence showed an increased grain weight under stress, reflecting a better source/sink ratio. Source/ sink ratio improvements have been proposed as a major mechanism to improve crop yields (Chang *et al.*, 2017). However, while this helped to compensate seed yield for Liporta, turning it resistant, it caused a decreased grain number per silique for Zephir, leaving it susceptible. Campari and NKNemax both had a reduced number of side branches under stress, but not at harvest, and both of them had reduced grain weight and grain number per silique. However, NKNemax had a higher number of siliques, compensating for grain weight and number, possibly due to its late flowering. NKNemax started to flower shortly before the end of the stress period, allowing full flower during optimal water conditions, which likely resulted in more successful pollinations. If stress occurs during gamete formation, developing pollen or ovules are undernourished and fail to finish maturation (Borghi *et al.*, 2019), therefore, it is beneficial to form flowers after the stress period. Hokkai-3-go and Pollen

did not show changes in the number of side branches, and both showed no significant reduction of any yield component alone, although Hokkai-3-go showed a non-significant trend for a decrease in number of siliques, which is therefore possibly the main cause for its overall yield reduction. This again could be linked to flowering time, as Hokkai-3-go was the earliest flowering accession among the total set (***Figure 1***) and mostly flowered during stress, likely to reduce the number of successfully pollinated flowers. Overall, this shows that the assessment of single yield components is not sufficient to characterize or predict drought resistance. It is necessary to understand the molecular basis of the differential behavior to improve breeding strategies for climate adaptation in terms of source/sink ratio and developmental timing. We therefore analyzed the transcriptional response of the plants at early and full stress. A WGCNA analysis clearly showed that the processes which are mainly affected during stress are photosynthesis and stress reaction. It is long known that photosynthesis is decreased under water shortage, as stomata close and limit CO_2 availability (Fang and Xiong, 2015). Our data show that all photosynthetic processes from plastid localization, carbon fixation, photosystem assembly and photosynthetic light reactions were downregulated on the transcriptional level. Similar results using transcriptomic data on stressed and non-stressed *B. napus* have been found before in seedlings (Zhang *et al.*, 2015; Wang *et al.*, 2017), although only one gene copy among the associated modules (*BnaA04g09660D*) was shared with the candidate gene set (169 genes) from (Wang *et al.*, 2017) and one gene copy (*BnaA07g03740D*) was shared with the candidate genes set (11 genes) from (Zhang *et al.*, 2015). Proteomics studies have shown the same processes to be involved at the protein level, although the correlation between the transcriptional and the translational level was only moderate (Koh *et al.*, 2015). The low overlap in candidate genes from other studies is not surprising given the low overlap in DEGs we have observed between early and late stress, indicating that stress response is highly dynamic. While some accessions do not greatly change their expression pattern between early and late stress (Alaska, Hokkai-3-go, Liporta, Zephir), others have their maximum stress response at early stress (Campari, NKNemax, Pollen) or also late stress (Musette) (***Figure 4***). This is agreement with the finding that resistance to drought stress was achieved by different phenotypic strategies. Diversity in drought resistance strategies in different accessions has been found before with a lower number of accessions (Urban *et al.*, 2016). This stresses the pressing need for the assessment of more genotypic variance of *B. napus* under realistic drought conditions. Respectively, our analysis of possible tolerance genes shows that there is hardly any general candidate gene for drought resistance in our data set. The putative protective genes involved in drought resistance were different, and they were also involved in different processes, with the notable exception of a copy of *MAP3K13* at early stress (***Figure 5***). *A. thaliana MAP3K13* has previously been implicated in nitrogen starvation signaling upstream of *MAPKK3* (Marchive *et al.*, 2013), which is a signal integrator for stress-signaling pathways (Colcombet *et al.*, 2016). It is possible that this copy of *B. napus MAP3K13* has acquired an additional role in drought stress signaling. Tolerant accessions express the copy more strongly, indicating that they are able to protect themselves more quickly. Earlier studies also pointed to a very quick reaction of the *B. napus* proteome under drought (Koh *et al.*, 2015). Here, at early stress, improvements were obviously reached with a better response to oxidative stress, reduced chlorophyll synthesis or a stronger downregulation of photosynthesis, for example by downregulating the key enzyme of the photosynthetic dark reaction, RuBisCo. At full stress, resistant accessions showed beneficial expression patterns for genes involved in systemic resistance, hormone response and response to reactive oxygen species. For Liporta/Zephir at early stress, resistance genes were enriched in photosynthesis genes, indicating resistance was mainly based on a higher photosynthetic capacity in Liporta, allowing to make use of the higher biomass, while at late stress, resistance seems to be

mainly due to ROS scavenging. For NKNemax/Campari at both early and late stress, resistance genes were mostly stress-response genes, indicating that besides later flowering, NKNemax might also have a more effective signaling and ROS scavenging system. For Hokkai-3-go/Pollen at early stress, Pollen as the tolerant variety seems to be able to uphold photosynthesis more strongly, as photosynthesis genes are not so strongly downregulated under stress as in the susceptible accession. At late stress, tolerance genes are enriched in stress and defense response. In total, this illustrates that there is both a huge variation in drought stress responses in *B. napus* as well as a high potential in improving drought tolerance by combining protective loci. Moreover, the analysis shows the importance of active developmental synchronization with stress period. Generally, shifting the flowering period to a later time point was beneficial in our stress treatment, especially for one accession which also started to flower later, while all accessions ended flowering later under stress. However, even in accessions which did not change onset of flowering, *Bna.FT.A02* levels were found to be reduced in reaction to stress. Expression of this copy was recently found to be a major determinant of flowering time in rapeseed (Wu *et al.*, 2018). This indicates that in accessions which did not shift onset of flowering, threshold levels of *Bna.FT.A02* to induce flowering were already reached at the onset of stress, but subsequent reduction decreased formation of more flowers to avoid fertilization during the stress period. For NKNemax, threshold levels seemed to be reached later, allowing shifting flowering to a later time point, which was obviously a successful tolerance strategy. At this point, it has to be noted that this is likely to be specific for this stress scenario, while one extra week of drought might have led to a completely different results. In general, most of the affected flowering genes were involved in the gibberellin response pathway, the age pathway and downstream signaling, indicating these are main entry points for crosstalk of stress signaling into flowering regulation. miRNAs seem to play an important supportive role in this crosstalk via the age pathway, as several differentially expressed *MIRNA* genes are predicted to target differentially expressed genes therein, mostly *SPL* genes. *SPL* genes are mostly important regulators of plant developmental timing in reaction to environmental stimuli to accelerate phase transitions (Wang and Wang, 2015). The miR156/SPL complex is one of the best characterized and conserved regulatory miRNA/gene modules (Wang and Wang, 2015) and the role of *SPL3* in flowering regulation is well documented (Wang, 2014). miR156 levels itself seem to react on sugar level, which itself depends on photosynthetic efficiency (Wang, 2014). The observed stress-related decline in photosynthetic gene expression might therefore have led to a decline in sugars, which in turn increases miR156 levels and in consequence decreases the level of *SPL* gene expression, delaying development. As miR156 is also connected to the regulation of several DELLA proteins, this may possibly explain changes in gibberellin and age pathway genes (Wang and Wang, 2015). Apart from miR156, we also detected other miRNA/gene modules involved in developmental timing, like miR319/TCP4, pointing to a more general role of miRNAs in stress-induced developmental delay. TCP factors orchestrate processes in leaf differentiation, like cell division activity, but also photosynthetic activity (Bresso *et al.*, 2018), Downregulation of photosynthesis, as observed via downregulation of important enzymes of the Calvin cycle and various subunits of photosystem I and II, was also partly connected to other mi*RNAs*, as well as upregulation of important stress-responsive genes like superoxide dismutase and *DREB* genes. Most *MIRNA* were downregulated under stress, releasing stable expression of their targets, in line with downregulation of miRNA producing enzymes. This is concordant with a proteomics study on drought stress in *B. napus* finding most of the differentially expressed proteins were upregulated (Koh *et al.*, 2015). Interestingly, miRNA expression in most cases had different or even adverse association to different copies of their respective target genes, which indicates a functional specialization of specific miRNA-target gene copy pairs. This was also true for miRNA families

themselves, which partly show differential stress responses for their different members. This points to subfunctionalisation on the level of miRNA, but also their targets, requiring more studies to precisely characterize the single interaction patterns in future.

Conclusions

Our results show that there is a high diversity in drought resistance strategies for spring drought scenarios in winter oilseed rape, both on the phenotypic and the genetic and miRNA level. This may allow combining advantageous alleles from genes in different resistance pathways like ROS scavenging, improved photosynthetic efficiency and developmental timing to produce a superior ideotype. We have also shown that *MIRNA* genes may represent a previously overlooked breeding target for better stress resistance. To implement those results into practice, more research needs to be done now to unravel the specific genetic mechanisms of the putative protectors, but also on trait associations, to finally identify usable allelic diversity for marker development for breeding programs.

References

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

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120.

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

Bresso EG, Chorostecki U, Rodriguez RE, Palatnik JF, Schommer C. 2018. Spatial Control of Gene Expression by miR319-Regulated TCP Transcription Factors in Leaf Development. *Plant Physiology* **176,** 1694–1708.

Bus A, Korber 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.

Chalhoub B, Denoeud F, Liu S *et al.* 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345,** 950–953.

Chang T-G, Zhu X-G, Raines C. 2017. Source-sink interaction: a century old concept under the light of modern molecular systems biology. *Journal of Experimental Botany* **68,** 4417–4431.

Colcombet J, Sözen C, Hirt H. 2016. Convergence of Multiple MAP3Ks on MKK3 Identifies a Set of Novel Stress MAPK Modules. *Frontiers in Plant Science* **07,** 217.

Fang Y, Xiong L. 2015. General mechanisms of drought response and their application in drought resistance improvement in plants. *Cellular and molecular life sciences CMLS* **72,** 673–689.

Fehlmann T, Reinheimer S, Geng C *et al.* 2016. cPAS-based sequencing on the BGISEQ-500 to explore small non-coding RNAs. *Clinical epigenetics* **8,** 123.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Research* **36,** D154-8.

Hatzig S, Zaharia LI, Abrams S *et al.* 2014. Early osmotic adjustment responses in drought-resistant and drought-sensitive oilseed rape. *Journal of Integrative Plant Biology* **56,** 797–809.

Hatzig SV, Nuppenau J-N, Snowdon RJ, Schießl SV. 2018. Drought stress has transgenerational effects on seeds and seedlings in winter oilseed rape (Brassica napus L.). *BMC Plant Biology* **18,** 297.

Hohmann M, Stahl A, Rudloff J, Wittkop B, Snowdon RJ. 2016. Not a load of rubbish: simulated field trials in large-scale containers. *Plant, Cell & Env* **39,** 2064–2073.

IPCC. 2014. *Climate Change 2014, Synthesis Report*.

Kalvari I, Argasinska J, Quinones-Olvera N *et al.* 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research* **46,** D335-D342.

Kazan K, Lyons R. 2016. The link between flowering time and stress tolerance. *Journal of Experimental Botany* **67,** 47–60.

Koh J, Chen G, Yoo M-J *et al.* 2015. Comparative proteomic analysis of Brassica napus in response to drought stress. *Journal of proteome research,* 3068–3081.

Kooyers NJ. 2015. The evolution of drought escape and avoidance in natural herbaceous populations. *Plant Science* **234,** 155–162.

Körber N, Wittkop B, Bus A, Friedt W, Snowdon RJ, Stich B. 2012. Seedling development in a *Brassica napus* diversity set and its relationship to agronomic performance. *Theoretical and Applied Genetics* **125,** 1275–1287.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359.

Li H, Handsaker B, Wysoker A *et al.* 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)* **25,** 402–408.

Lovell JT, Juenger TE, Michaels SD *et al.* 2013. Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation. *Proceedings of the Royal Society B: Biological Sciences* **280,** 20131043.

Lu J, Carbone GJ, Grego JM. 2019. Uncertainty and hotspots in 21st century projections of agricultural drought from CMIP5 models. *Scientific Reports* **9,** 157.

Marchive C, Roudier F, Castaings L *et al.* 2013. Nuclear retention of the transcription factor NLP7 orchestrates the early response to nitrate in plants. *Nature communications* **4,** 1713.

Ouyang S, Buell CR. 2004. The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Research* **32,** D360-3.

Quinlan AR. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr prot bioinf* **47,** 11.

Riboni M, Galbiati M, Tonelli C, Conti L. 2013. GIGANTEA Enables Drought Escape Response via Abscisic Acid-Dependent Activation of the Florigens and SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1. *Plant Physiology* **162,** 1706–1719.

Sieling K, Christen O. 2015. Crop rotation effects on yield of oilseed rape, wheat and barley and residual effects on the subsequent wheat. *Archives of Agronomy and Soil Science* **22,** 1–19.

Srikanth A, Schmid M. 2011. Regulation of flowering time: all roads lead to Rome. *Cellular and Molecular Life Sciences* **68,** 2013–2037.

Tal A. 2016. Rethinking the sustainability of Israel's irrigation practices in the Drylands. *Water research*.

Teune J-H, Steger G. 2010. NOVOMIR: De Novo Prediction of MicroRNA-Coding Regions in a Single Plant-Genome. *Journal of nucleic acids* **2010**.

Urban MO, Vasek J, Klima M *et al.* 2016. Proteomic and physiological approach reveals drought-induced changes in rapeseeds: Water-saver and water-spender strategy. *Journal of Proteomics* **152,** 188–205.

Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in biosciences = Theorie in den Biowissenschaften* **131,** 281–285.

Wang H, Wang H. 2015. The miR156/SPL Module, a Regulatory Hub and Versatile Toolbox, Gears up Crops for Enhanced Agronomic Traits. *Molecular Plant* **8,** 677–688.

Wang J-W. 2014. Regulation of flowering time by the miR156-mediated age pathway. *Journal of Experimental Botany* **65,** 4723–4730.

Wang P, Yang C, Chen H, Song C, Zhang X, Wang D. 2017. Transcriptomic basis for drought-resistance in Brassica napus L. *Scientific reports* **7,** 40532.

Wu D, Liang Z, Yan T *et al.* 2018. Whole-genome resequencing of a world-wide collection of rapeseed accessions reveals genetic basis of their ecotype divergence. *Molecular Plant*.

Zhang J, Mason AS, Wu J *et al.* 2015. Identification of Putative Candidate Genes for Water Stress Tolerance in Canola (Brassica napus). *Frontiers in Plant Science* **6,** 1–13.

Chapter 3.2: Drought stress has transgenerational effects on seeds and seedlings in winter oilseed rape (*Brassica napus* L.)

Hatzig SV, Nuppenau J-N, Snowdon RJ, **Schiessl SV** (2018) Drought stress has transgenerational effects on seeds and seedlings in winter oilseed rape (*Brassica napus* L.). BMC Plant Biol 18:297. doi: 10.1186/s12870-018-1531-y

Hatzig et al. BMC Plant Biology (2018) 18:297
https://doi.org/10.1186/s12870-018-1531-y

BMC Plant Biology

RESEARCH ARTICLE Open Access

Drought stress has transgenerational effects on seeds and seedlings in winter oilseed rape (*Brassica napus* L.)

Sarah V. Hatzig[1], Jan-Niklas Nuppenau[2], Rod J. Snowdon[1] and Sarah V. Schießl[1*]

Abstract

Background: Drought stress has a negative effect on both seed yield and seed quality in *Brassica napus* (oilseed rape, canola). Here we show that while drought impairs the maternal plant performance, it also increases the vigour of progeny of stressed maternal plants. We investigated the transgenerational influence of abiotic stress by detailed analysis of yield, seed quality, and seedling performance on a growth-related and metabolic level. Seeds of eight diverse winter oilseed rape genotypes were generated under well-watered and drought stress conditions under controlled-environment conditions in large plant containers.

Results: We found a decrease in seed quality in seeds derived from mother plants that were exposed to drought stress. At the same time, the seeds that developed under stress conditions showed higher seedling vigour compared to non-stressed controls.This effect on seed quality and seedling vigour was found to be independent of maternal plant yield performance.

Conclusions: Drought stress has a positive transgenerational effect on seedling vigour. Three potential causes for stress-induced improvement of seedling vigour are discussed: (1) Heterotic effects caused by a tendency towards a higher outcrossing rate in response to stress; (2) an altered reservoir of seed storage metabolites to which the seedling resorts during early growth, and (3) inter-generational stress memory, formed by stress-induced changes in the epigenome of the seedling.

Keywords: Canola, Rapeseed, Drought stress, Seed germination, Seedling vigour, Metabolite analysis, Seed quality, Fatty acids, Amino acids, Intergenerational stress memory

Background

Drought stress is one of the most important abiotic factors impairing seed and biomass yield in global agriculture. Forecasts predict an increasing frequency of insufficient precipitation and consequent aridity in many parts of the world [1, 2]. Therefore, mankind must develop sustainable strategies to protect crop production as well as crop quality under limiting conditions. Breeding for drought-adapted varieties is an important building block in such a strategy. Modern breeding programs require a profound understanding of the specific implications that water stress has on yield and yield quality parameters. The present study helps expand our knowledge on these implications in winter oilseed rape (*Brassica napus* L.), by providing data on maternal and transgenerational effects of drought stress on yield, seeds and seedling vigour. *B. napus* is one of the most important oil crops worldwide [3].

Elimination of seed erucic acid and reduction of seed glucosinolate content (double-low seed quality) facilitated a global boom in production of oilseed rape and canola (*B. napus*), today the second-most important oilseed crop in the world behind soybean. Besides its use as feedstock in Europe and for biofuel production, the oil from oilseed rape and canola also plays a significant role for human consumption. Due to its favourable fatty acid composition with high amounts of mono- and polyunsaturated fatty acids, the

* Correspondence: sarah-veronica.schiessl@agrar.uni-giessen.de
[1]Department of Plant Breeding, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany
Full list of author information is available at the end of the article

consumption of oilseed rape and canola oil has been described to benefit human health [4]. Although desirable seed oil qualities are genetically fixed in modern germplasm collections, moderate fluctuations in seed quality and composition can be linked to environmental influences prevailing during seed production [5–8]. Besides the need for high oil yield, high amounts of desirable fatty acids and low concentrations of undesirable seed components, such as erucic acid and glucosinolates, a fast and uniform germination and high seed vigour are essential for good crop establishment and yield stability. Moreover, enhanced seed and seedling vigour improve plant density and spatial arrangement in the field, extend the growth duration [9] and have a direct positive effect on crop formation and growth [10]. The need for optimal seed germination and vigour characteristics thus requires seed production environments that maximise vigour performance.

Here, we analysed the effect of limiting water conditions during seed production in two respects: First, we investigated the consequences of maternal drought stress on seed yield and seed quality. Subsequently, we evaluated the effect of maternal water supply on germination and seedling vigour performance of progeny from stressed and non-stressed maternal plants of genotypes with varying maternal response to drought stress. Plants from eight contrasting winter oilseed rape genotypes were grown under optimal water supply vs. water shortage during the critical phase of flowering, in a large-container growth system which accurately simulates a field growth environment with deep soil whilst allowing careful control of water supply under semi-controlled greenhouse conditions [11]. As expected, optimal water supply during seed production ensured high seed quality and composition. Unexpectedly, however, lower quality seeds from plants grown under drought stress showed favourable effects on seedling vigour potential in comparison to the seeds from non-stressed plants. This indicates that the optimal growth environment in respect to water availability may differ for farmers and seed producers. While drought can pose a significant threat to seed quality and yield for farmers, drought stress effects during seed production may be advantageous for seed producers in terms of optimising germination and vigour characteristics of commodity seed.

Results

Maternal treatments: Seed yield and seed quality

The maternal plants were subjected to a 3 weeks water withdrawal starting shortly before flowering. Total seed yield was significantly affected by this drought treatment for all accessions except *Liporta*, *NK Nemax* and *Pollen* (Fig. 1a). The strongest yield reduction under stress (42%) was found in the accession *Musette*. For accessions *Alaska*, *Campari*, *Hokkai 3-Go* and *Zephir*, we observed reductions of 26, 29, 29 and 33%, respectively. Significant reduction in thousand seed weight was observed only for *Alaska* and *NK Nemax* (Fig. 1b). The number of seeds per silique was significantly reduced in *Campari*, *NK Nemax* and *Zephir* under drought stress (Fig. 1c), however no significant change in number of siliques per plant was observed except in *NK Nemax*, which showed a significant increase in number of siliques under drought stress (Fig. 1d). Number of seeds per plant was significantly reduced in genotypes *Campari*, *Hokkai 3-Go* and *Musette* (Fig. 1e).

Seed oil and protein content along with the fatty acid composition were strongly affected by the drought stress treatment (Fig. 2). Under drought treatment all accessions reacted with a significant decrease in seed oil content (Fig. 2a) and corresponding significant increase in protein content (Fig. 2b). The strongest effect was observed in *NK Nemax*, which showed 14.4% reduced seed oil and 29.6% increased protein. The weakest effect was seen in *Campari* where oil decreased by 6.5% and protein increased by 15.6% under drought stress. In the drought stress treatment, a general tendency towards higher seed glucosinolate contents was observed, but this effect was statistically significant in *Hokkai 3-Go* only (+ 12.5%; Fig. 2c). In contrast to this trend, *Liporta* showed a significant decrease in seed glucosinolate content (− 42.6%) in the drought stress treatment.

Seeds produced under drought stress showed distinct changes in fatty acid patterns compared to seeds developed under optimal water conditions (Additional file 1). Oleic acid was significantly decreased in the genotypes *Alaska*, *Campari* and *NK Nemax,* while mean decreases for *Hokkai* 3-Go, *Musette* and *Zephir* were not statistically significant (Fig. 3a). All genotypes except *Hokkai 3-Go* and *Zephir* showed a significant increase in their proportion of linoleic acid (Fig. 3b), while significant increases in linolenic acid were observed for *Alaska*, *NK Nemax* and *Zephir* (Fig. 3c). Other fatty acids (stearic acid, linoleic acid, linolenic acid, palmitic acid, gadoleic acid, behenic acid, palmitoleic acid, lignoceric acid and myristic acid) were generally elevated in the maternal drought variant in three or more of the eight accessions. No effects of drought stress were observed for eicosenic or erucic acid.

Seed germination characteristics

The absolute germination rate (GR96) under constant watering (0 Mpa) showed a small but significant increase in seeds derived from the maternal drought treatment in *Liporta*, *Musette*, *NK Nemax* and *Zephir*, which showed GR96 values of 99, 99, 96 and 98% in seeds from the maternal control treatment but germinated to 100% in seeds from the maternal drought treatment. *Alaska*,

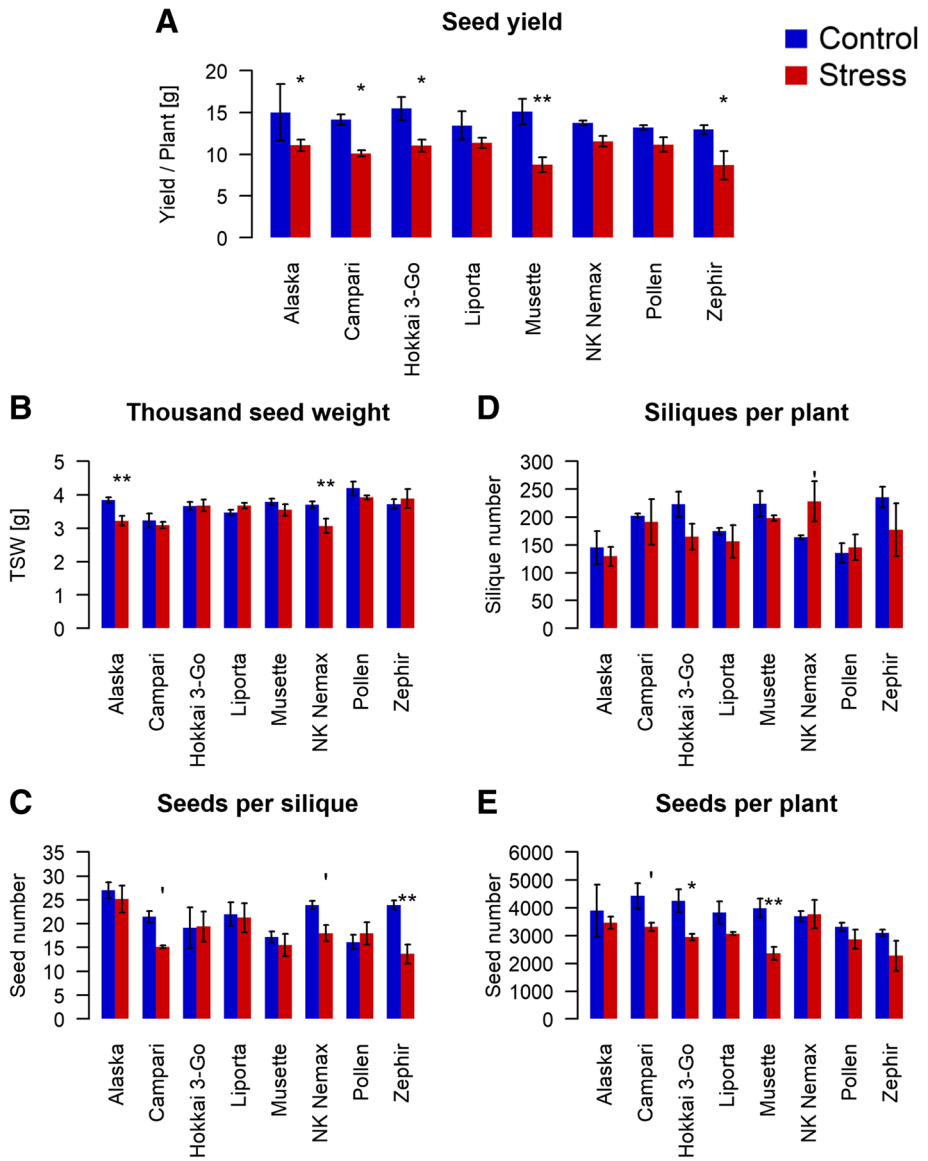


Fig. 1 Effect of drought stress on (**a**) Total seed yield, (**b**) Thousand seed weight, (**c**) Number of seeds per silique, (**d**) Number of siliques per plant and (**e**) number of seeds per plant of 8 diverse winter oilseed rape genotypes cultivated in a semi-controlled container trial. Bars are means of three replicates with standard errors. Significant differences at $p < 0.1'$, $p < 0.05$* and $p < 0.01$**

Campari, *Hokkai 3-Go* and *Pollen* showed 100% germination in both maternal treatments. Under moderate osmotic stress (− 0.5 Mpa), no significant difference in GR96 was be observed between the maternal treatments, with mean GR96 values of 98.3 and 97.5% for maternal control and drought treatment, respectively.

Accessions reacted significantly different in their mean germination time (MGT) under constant watering (Fig. 4a). Whereas *Hokkai 3-Go*, *Pollen* and *Zephir* showed a significant increase in MGT in seeds from the maternal drought treatment, *Musette* and *NK Nemax* showed higher MGT values in the seeds from the maternal control treatment. MGT was equal for both maternal treatments under moderate osmotic stress, except in *Alaska*, *NK Nemax* and *Pollen*, which showed higher MGT values in the maternal stress treatment (Fig. 4b). Comparing the two in vitro scenarios, constant watering and moderate osmotic stress, uniformity (U) in seed germination was significantly higher under constant watering (data not shown). However, no significant difference in U was observed between the two maternal treatments, neither under constant watering nor under moderate osmotic stress.

Seedling vigour performance

Significant differences in seedling vigour were observed both between accessions and the two different maternal

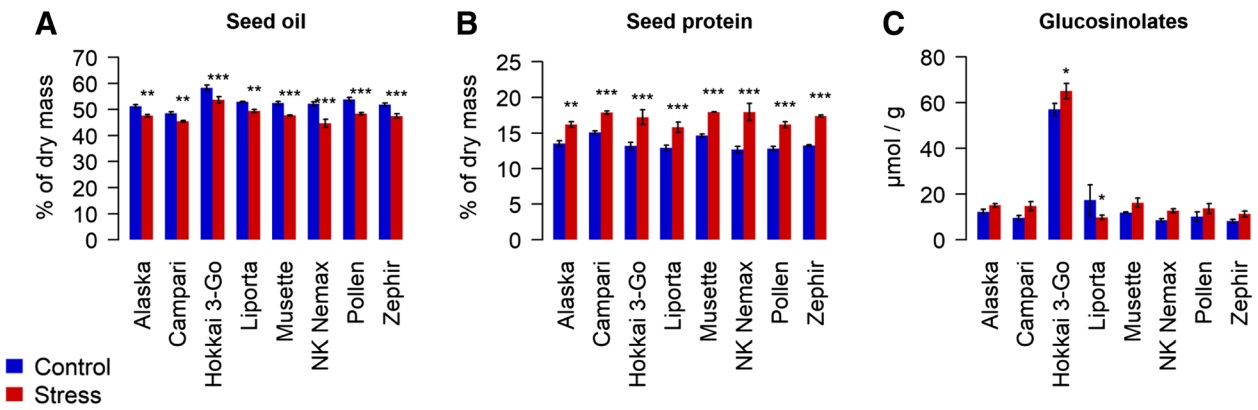


Fig. 2 Effect of drought stress on (**a**) seed oil amount, (**b**) seed protein amount and (**c**) concentrations of seed glucosinolates of 8 diverse winter oilseed rape genotypes cultivated in a semi-controlled container trial. Bars are means of three replicates with standard errors. Significant differences at $p < 0.05$*, $p < 0.01$** and at $p < 0.001$***

water treatments (Fig. 5). For all accessions except *Musette*, seedling fresh weights were found to be significantly increased by the maternal drought stress treatment. The strongest effect was found in *Zephir* which showed a mean increase of 36.7% relative to the control treatment, followed by *Hokkai 3-Go* and *Alaska* with 34.0 and 27.2% increase, respectively. The lowest significant difference was observed for *Liporta,* which showed an increase by 14.3% relative to the control. Seedling vigour was found to be uncorrelated to mean germination time for both maternal water treatments. A replication trial with 4 of those 8 accessions in the next season showed the same trend (Additional file 2: Figure S5).

Seedling metabolite patterns

In seedlings from the maternal drought treatment, concentrations of several amino acids and nitrogen compounds were significantly increased in diverse accessions compared to seedlings from the maternal control treatment (Additional file 3). From these metabolites, ammonium (NH_4^+), Histidine (His), Asparagine (Asn), S-methylcystein sulfoxide (SMCSO), Glutamine (Gln), Arginine (Arg), Glycine (Gly), Aspartate (Asp), Threonine (Thr), α-Alanine (α-Ala), γ-aminobutyric acid (GABA), Tyrosine (Tyr), Valine (Val) and Isoleucine (Ile) showed significantly higher concentrations in three or more of the eight genotypes under investigation. The concentrations of sugars and organic acids were similarly increased in the maternally drought treated seedlings of several accessions (Additional file 3). Most differences between the maternal treatments were observed for malate, fructose and sucrose (significantly different for six, six and three of the eight accessions, respectively). A significant role in the trans-generational response to drought stress was implicated for NH_4^+, SMCSO, Gln, Gly, Tyr, Val, fructose and malate. Strong correlations were found between relative concentrations of these

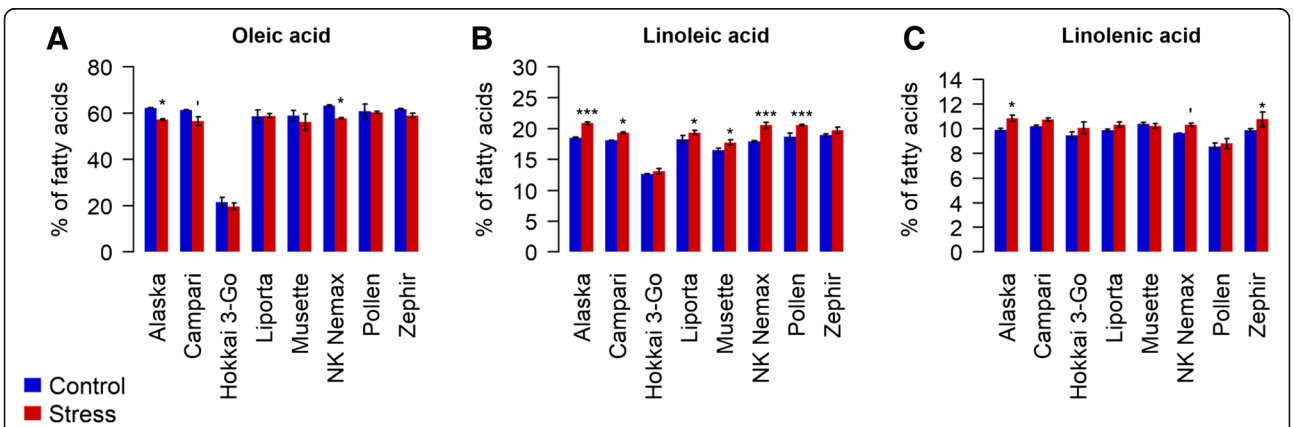


Fig. 3 Effect of drought stress on the amount of (**a**) Oleic acid, (**b**) Linoleic acid and (**c**) Linolenic acid measured in the seeds of 8 diverse winter oilseed rape genotypes cultivated in a semi-controlled container trial. Bars are means of three replicates with standard errors. Significant differences at $p < 0.1$', $p < 0.05$* and at $p < 0.001$***

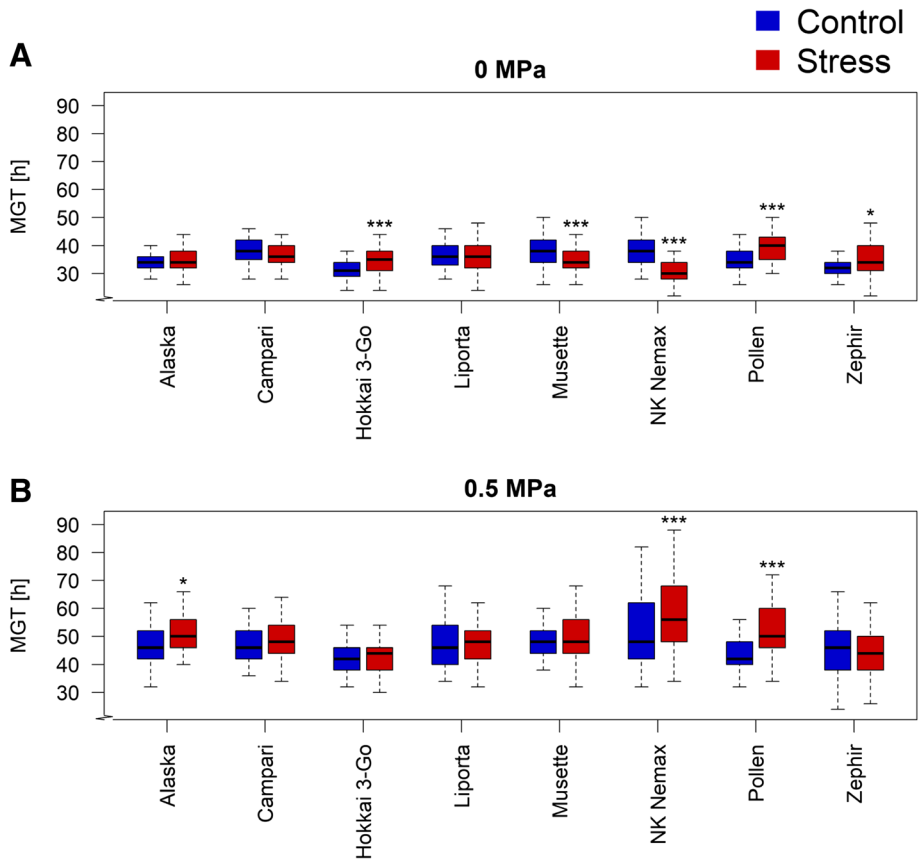


Fig. 4 Effect of maternal drought stress on mean germination time (MGT) of seeds from 8 diverse winter oilseed rape genotypes, cultivated in a semi-controlled container trial: (**a**) germination performance under 0 MPa and (**b**) germination performance under moderate osmotic stress (− 0.5 MPa). Boxplots represent performances of 100 seeds each. Significant differences at $p < 0.05$* and at $p < 0.001$***

metabolites and relative seedling FW in seeds from the maternal drought treatment compared to the control (Additional file 4, Fig. 6).

Multivariate analysis of seed quality, seedling vigour and seedling metabolic patterns

Using a principal component analysis (PCA) for all seed quality characteristics and seedling metabolite concentrations with significant differences between the maternal treatments, we found a clear spatial separation between offspring plants derived from the maternal stress and maternal control treatment (Fig. 7a). It is evident that the most important determinant is the maternal treatment, while genotypic specificities were subordinated. The accessions *Hokkai 3-Go* showed a similar separation between the maternal treatments but diverged from all other genotypes. As *Hokkai 3-Go* was the only genotype with ++ seed quality, this spatial distance may be explained by differences in its fatty acid patterns. The strongest contributions regarding the separation of quality groups could be attributed to the fatty acids oleic acid, palmitic acid, linoleic acid and palmitoleic acid (Fig. 7b). The largest separating effect would be expected from erucic acid, however this fatty acid was not considered as a variable in the PCA, as it showed no significant differences between the maternal treatments. Furthermore, Fig. 7b shows that Thr, Tyr, malate, Tyr, Asp, Gly, Gln and NH_4^+ mainly contribute to the separation between maternal control and stress treatment.

Discussion

Our data show that during the critical stage of flowering, water supply has a pronounced transgenerational effect on seedling vigour of the progeny. The strength of the drought stress effects on both maternal seed yield and transgenerational seedling performance is strongly genotype-specific. The drought-induced reduction in seed yield confirms the long-known relationship between water supply and yield development during the critical phase of flowering [12–14]. Unexpectedly, the positive transgenerational effect of drought stress on vigour is in contrast to the negative effect of maternal stress on the yield performance of the maternal plants. However, the growth-stimulating transgenerational effect in seedlings due to maternal drought stress was independent from

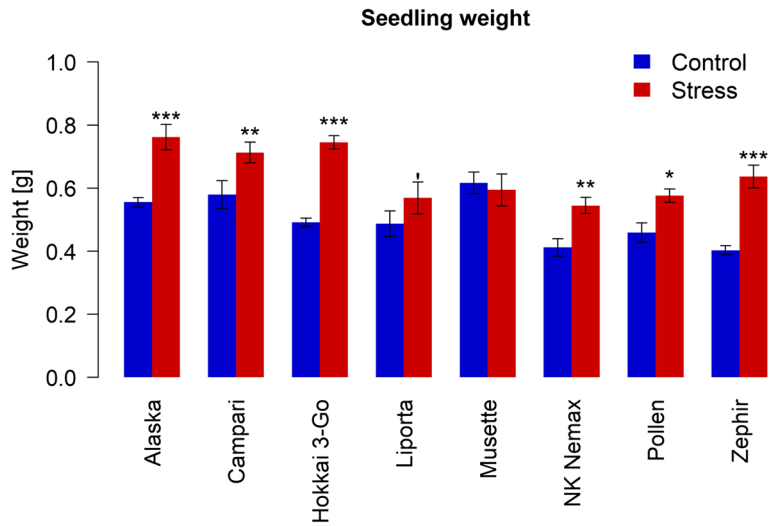


Fig. 5 Effect of maternal drought stress on seedling growth performance of seeds harvested from 8 diverse winter oilseed rape genotypes cultivated in a semi-controlled container trial. Bars are means of three replicates with standard errors. Significant differences at $p < 0.1$', $p < 0.05$*, $p < 0.01$** and $p < 0.001$***

the extent of yield reduction in the mother plant. Interestingly, maternal drought stress had differential effects on seed germination performance depending on the accession. However, the accession had no effect on seedling vigour. In fact, the lack of any significant correlation between Mean Germination time (MGT) and seedling fresh weight suggests that the seedling performance is completely decoupled from germination performance. Our experiments under two different osmotic conditions suggest that the water availability during germination has a stronger influence on germination performance than water availability during foregoing seed production. In contrast, seedling development appears to be clearly enhanced by maternal drought, even in accessions like *Hokkai 3-Go, Pollen and Zephir* which showed delayed germination in seeds of the maternal drought treatment and subsequently exhibited a higher seedling biomass than the maternal control treatment. In a smaller replication trial performed in the following season, the same enhancement effect was observed. These findings contradict the assumption that optimal seedling development presupposes an appropriate germination performance. Instead, our findings corroborate a weak link between germination and post-germination seedling growth, as already hypothesized by [15].

Three possible explanations are proposed for the observed transgenerational effects on seedling vigour: Heterosis, changes in seed quality, and intergenerational stress memory involving an alteration of growth-effective metabolic processes.

In our experiments, we used openly pollinated plants to avoid known negative effects of bagging on seed quality. In consequence, we expect a certain level of cross-pollination, and differences in vigour between progenies from different stress environments could possibly arise from differential outcrossing rates for stressed and control progenies, resulting in different levels of seedling heterosis. Assuming that drought stress negatively affects male fertility [16, 17], higher cross pollination rates might be expected for drought stressed mother plants. Moreover, it has been shown that an enhanced biomass development due to heterosis can already be observed during the early stage of seedling growth [18]. To investigate this phenomenon, we performed a separate quantification of homozygous- and heterozygous progenies by Kompetitive Allele Specific PCR (KASP) genotyping (data not shown), however this analysis did not reveal significant differences in cross pollination rates between the two maternal treatments. Hence, we believe that differential heterosis does not underlie the effects of maternal drought stress on seedling vigour of the progeny.

Changes in seed quality due to the maternal drought treatment are another possible reason for improved seedling vigour. Changes in seed composition alter the reservoir of storage metabolites on which the seedling relies during post-germination seedling growth. A marked reduction in seed oil content and an associated increase in seed protein content are common observations under drought stress [5, 13, 19, 20]. Considering the strong genotype x environment interaction for seed quality in *B. napus* [21], all genotypes would be expected to react to maternal drought with a similar shift in their seed composition patterns. Increased glucosinolate production in seeds maturing under water stress, as suggested by [5], could not be confirmed statistically by our results. In contrast, we observed a shift in fatty acid

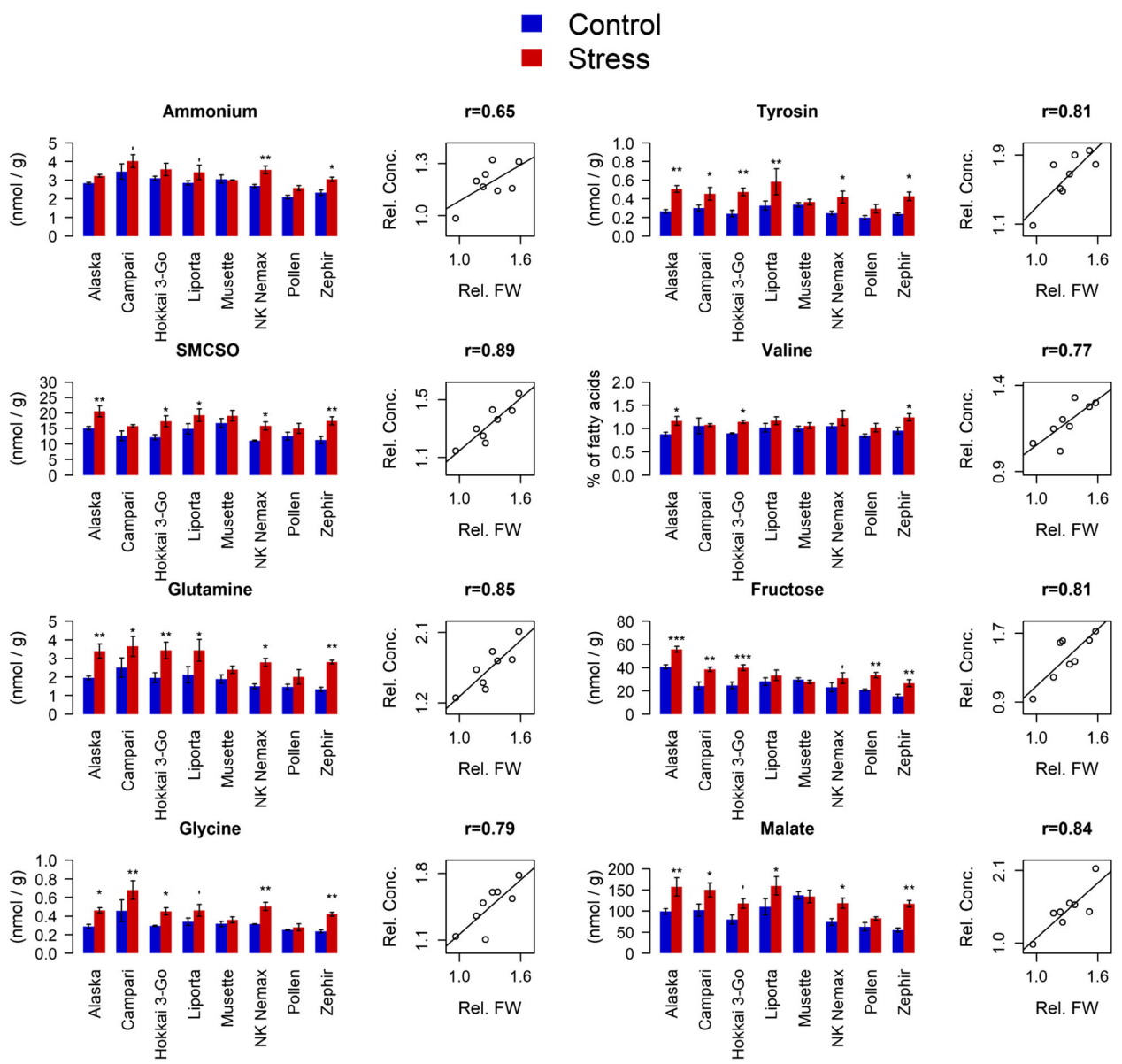


Fig. 6 Effect of maternal drought stress on seedling metabolite concentrations of the offspring derived from 8 diverse winter oilseed rape genotypes cultivated in a semi-controlled container trial. Bars are means of three replicates with standard errors. Significant differences at $p < 0.1'$, $p < 0.05*$, $p < 0.01**$ and $p < 0.001***$. Scatter diagrams show correlations between the corresponding relative metabolite concentration (Rel. Conc.) and relative seedling fresh weight (Rel. FW) under drought stress

patterns, especially at the expense of oleic acid and in favour of polyunsaturated fatty acids like linoleic and linolenic acid, due to reduced water supply [20]. The accessions *Liporta*, *NK Nemax* and *Pollen*, showing no significant yield reduction under stress, nevertheless showed similar declines in seed quality – especially in terms of seed oil content, seed protein content and fatty acid composition. This shows that seed quality is more sensitive to drought stress than the total yield levels. We assume that fatty acid modification processes towards long-chained, poly-unsaturated fatty acids proceed without restriction, whereas the delivery of corresponding precursors like stearic or oleic acid appears to have been limited by the stress (Fig. 8). However, we could not confirm that this led to a net increase in storage lipid biosynthesis beyond the extent given under non-stress conditions. Gene expression analysis has shown that expression of genes encoding fatty acid modifying enzymes like fatty acid elongase 1 (FAE1), one of the core enzymes involved in erucic acid biosynthesis in the *Brassicaceae*, peak to a later time-point than the basic fatty acid synthesis machinery [22]. Therefore, a

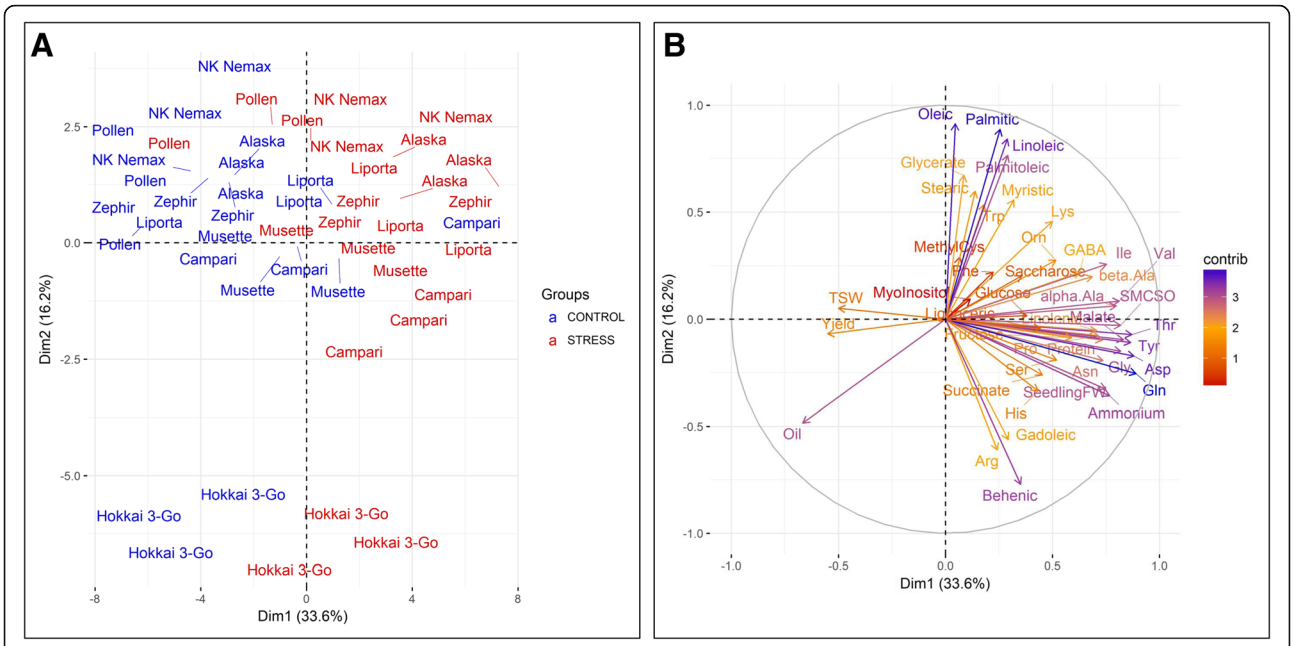


Fig. 7 Two-dimensional principal component analysis showing (**a**) the multivariate variation in terms of metabolite concentrations among seeds and seedlings derived from 8 open-pollinated winter oilseed rape genotypes cultivated under control and drought stress conditions in a semi-controlled container trial. All biological replicates are shown. Seedlings from maternal control conditions are shown in blue, seedlings from maternal stress conditions are shown in red. Proximity between samples indicates similarity in metabolite patterns. The axes show the first and second principal component along with the variation explained by them in brackets. (**b**) The same data set now represented as variable vectors indicating the strength and direction of each seed component and seedling metabolite. Vectors which are perpendicular to each other show independence of the component, vectors pointing in opposite directions show a negative correlation, while vectors with a small angle between them show a positive correlation. The color denotes the weight of the contribution to the total variation in the data set, with higher values meaning a stronger contribution. Amino acids are abbreviated as 3 letter IUPAC code. Fatty acids are named without the extension "acid". TSW: Thousand seed weight, FW: Fresh weight

higher quantity of prolonged and poly-unsaturated fatty acids in the seed might account for an earlier onset of maturity [20]. However, it is unlikely that premature maturity explains the observed shift in fatty acid patterns. Indeed, we found that the flowering period was prolonged under water shortage in our experiment (data not shown), suggesting a delayed maturity.

Besides seed oil and protein quality parameters, seedling metabolite patterns were also substantially shifted by the maternal drought treatment (Fig. 9). Higher concentrations of free amino acids could be explained by enhanced amino acid biosynthesis and/or enhanced protein degradation or, alternatively, an inhibition of amino acid degradation and/or protein synthesis. Higher monosaccharide- and disaccharide concentrations in seedling tissues might indicate an inhibition of glycolytic processes and/or polysaccharide synthesis, or an enhanced carbon assimilation and/or polysaccharide degradation. As seedling growth is generally enhanced due to maternal drought stress, it appears more likely that growth promoting processes, like carbon assimilation were stimulated, rather than an enhancement of catabolic processes. A possible explanation for an alteration of metabolic processes in the seedling which ultimately lead to enhanced seedling growth is intergenerational stress memory, defined as "a stress imprint that extends from one stressed generation of organisms to at least the first stress-free offspring generation" [23]. This model explains transgenerational effects by stress-induced changes in the epigenome of the plant, amongst them changes in DNA methylation patterns [24, 25] or histone modifications [26]. Such transgenerational effects can instil an adaptive advantage when the progeny is exposed to the maternal stress conditions [24, 27]. While we cannot exclude the possibility of epigenetic changes in our study, we can at least not confirm the adaptive value of the induced changes. However, faster growth before stress induction can improve the plants' survival chances in a new stress scenario and might therefore represent an adaptation in life cycle. Germination performance in particular showed no advantage in seeds derived from maternal drought treatment under osmotic stress conditions. If such epigenetic changes were present, our data suggest that they are mainly maternally inherited. Otherwise, we would have expected differences in seedling biomass and metabolite patterns among individuals from the same

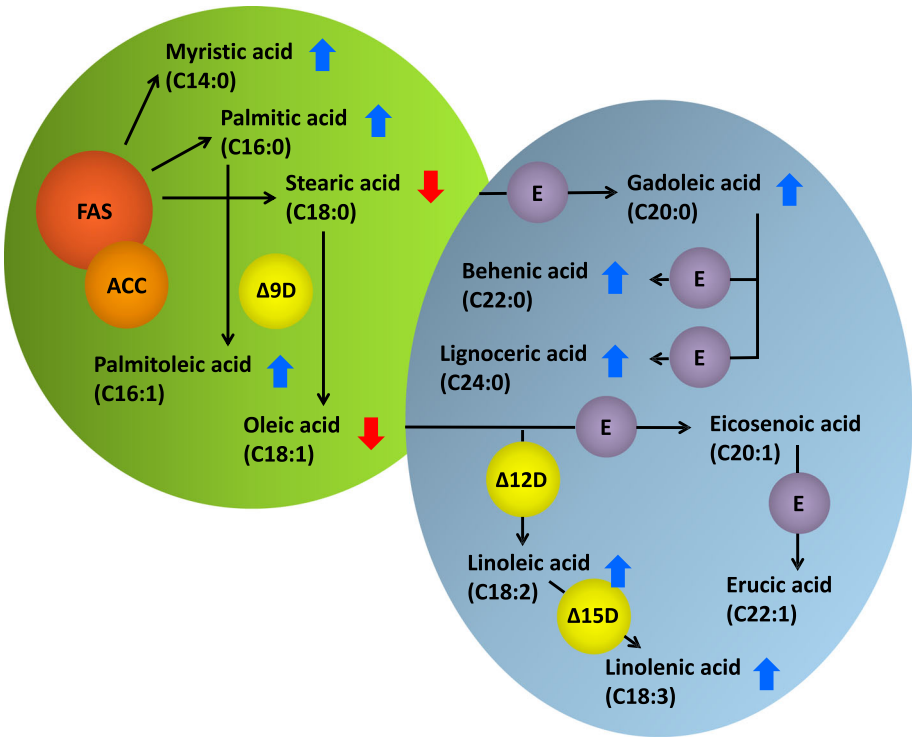


Fig. 8 Principal scheme of fatty acid biosynthesis in chloroplasts (green body) and endoplasmatic reticulum (blue body) of rapeseed. Arrows indicate, whether relative amounts of fatty acids have increased or decreased under drought stress in three or more of the 8 observed winter oilseed rape genotypes. FAS: Fatty acid synthase, ACC: Acetyl-CoA-Carboxylase, Δ9D: Δ9-Desaturase, Δ12D: Δ12-Desaturase, Δ15D: Δ15-Desaturase, E: Elongase

maternal stress treatment, as pollination was equally likely to occur from both stressed or non-stressed pollinators. This is in agreement with other studies suggesting that although both the maternal as well as the paternal environment can form specific transgenerational responses, the post-zygotic maternal effects were generally stronger [28–30].

Conclusions

The conclusions from our study have opposite consequences for farmers and seed producers. Farmers are primarily interested in achieving high seed yield levels with optimum seed quality, hence drought during seed maturation can have a negative influence. In contrast, commercial seed producers might possibly take advantage of water deficits during the generative growth phase of winter oilseed rape, as seedling vigour performance could be positively influenced by maternal drought stress, for example by generating seeds in water-limited production areas. On the other hand, a commercial benefit can only be achieved if the drought environment does not reduce seed number. This was at least not the case for most of the observed genotypes under the stress conditions applied within our study (Fig. 1e). Seed weight was not negatively correlated to seedling vigour in six out of the eight investigated genotypes in our study, meaning that a reduction in thousand seed weight may not have a negative effect on seed number. In summary, our results provide an interesting new approach for optimization of the commercial seed production process in winter oilseed rape.

Methods

Plant material and cultivation of maternal plants

We selected 8 diverse winter oilseed rape (WOSR) inbred lines from the ERANET-ASSYST diversity set [31] based on previous analyses of their flowering time and field performance under water-limiting conditions. The genotype selection comprised inbred lines of the double-low seed quality accessions *Alaska*, *Campari*, *Liporta*, *Musette*, *NK Nemax*, *Pollen* and *Zephir* along with the high erucic acid, high glucosinolate genotype *Hokkai 3-Go*. In the growing period 2014/2015, we grew all accessions under semi-controlled conditions in the large container growth system described by [11], which enables simulated field growth conditions and planting densities with a deep soil profile and exact control of soil water capacity. The containers were filled to a depth of 90 cm with a mixture of 25% soil and 75% sand. Around 30 seeds per container were sown by hand on

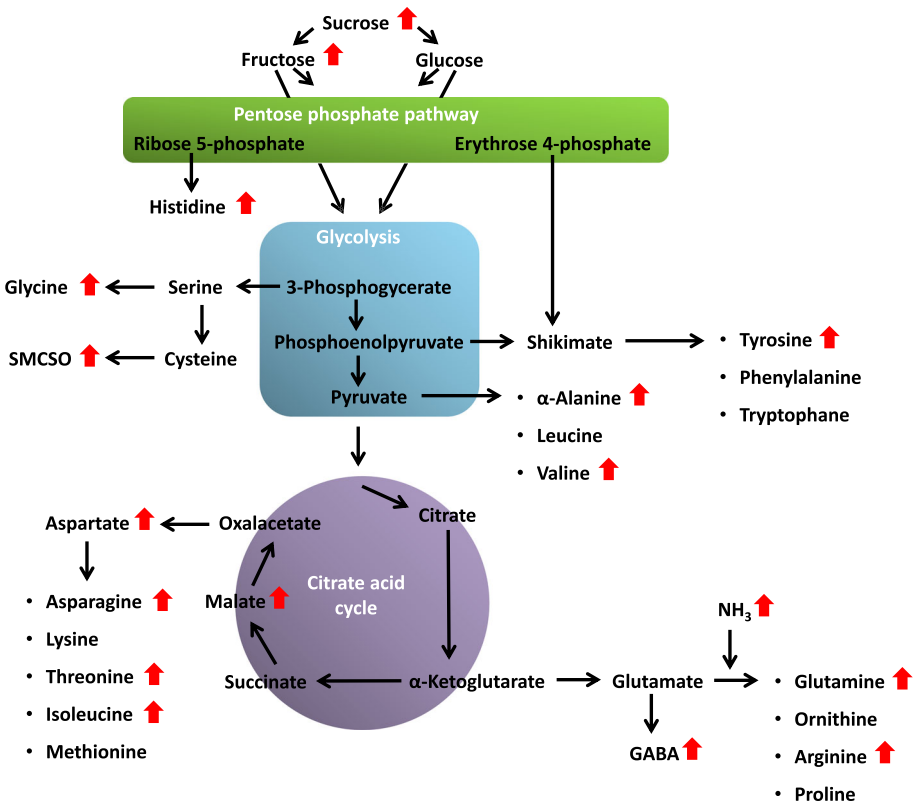


Fig. 9 Principal scheme of the biosynthesis and derivation of different amino acid and nitrogen compounds in rapeseed. Arrows indicate, whether concentrations of metabolites have increased or decreased under drought stress in three or more of the 8 observed winter oilseed rape genotypes

30.11.2014. Seedlings were thinned to 9 plants per container 25 d after sowing, with uniform spacing of the nine remaining plants. The plants were fertilized on 02.04.2015 with 1.6 g NH_4NO_3 per container. Each genotype was cultivated under well-watered conditions (control) and drought stress conditions (stress). For each accession and treatment, three containers were set up as independent biological replicates, making a total of 48 test containers. The containers were ranged in a fully randomized block design. Additional planted containers were arranged around the container block to minimize boundary effects. Optimal water supply was guaranteed by watering the plants to a minimum soil water capacity of 60%. Water demand was determined by regular weighing of the containers with a hydraulic scale hoist. In the stress treatment water supply was suspended for three weeks from begin of flowering (BBCH 50) until the onset of full flowering (BBCH65). Afterwards the plants were re-watered to the same level as the control treatment (60% water capacity). Soil water capacity was calculated based on container weights assuming a soil dry weight of 138.9 kg, a container weight of 12 kg, and a varying total plant weight. Total plant weight was estimated to be 1 kg at the start of the stress treatment. For control plants, we assumed a biomass gain of 100 g per container and day. For stress plants, we assumed a biomass gain of 100 g per container and day in the first week, 50 g per container and day in the second week, and no gain in the second week. The resulting differences in soil water content before and after watering are shown in Additional file 5: Figure S4. All plants were openly pollinated. Seeds from both the well-watered and drought stress treatment (hereinafter referred to as maternal treatments) were harvested on 13.07.2015 and stored under dry conditions.

A replicate of this experiment using 4 of those 8 accessions (Alaska, NK Nemax, Pollen, Zephir) was performed in the next season. Vernalized plants were potted into containers on 07.04.2016 with 5 plants per container. The stress treatment was induced on 23.05.2016. At the same day, the plants were bagged to ensure self-pollination. The resulting seeds were harvested on 23.08.2016 and stored under dry conditions.

Analysis of seed quality

Seed oil content, seed protein content and seed glucosinolate content was determined using near-infrared reflectance spectroscopy (NIRS, Unity SpectraStar 2500,

Brookfield, USA). NIRS measurements were performed using standard procedures as described in [32].

Fatty acid quantification was performed by gas chromatography (GC) analysis. In a test tube, 300 mg seeds were grinded together with 2 mL petroleum benzine for 2 min at maximum speed with a T25 digital Ultra Turrax sample grinder (IKA Works, Inc. Wilmington, USA). Another 2 mL petroleum benzine were added and the sample was vortexed for 10 s. After 30 min, 0.8 mL of supernatant was pipetted into a new test tube, which was placed under the laboratory fume hood until the petroleum benzine was completely evaporated. Subsequently, 2 mL sodium methylate were added to the remaining oil in the test tube and the sample was vortexed for 10 s. The sample was covered and left to settle for 30 min. Afterwards 1.7 mL iso-octane were added and the solution was shaken carefully. The sample was again covered and left to settle for 30 min. Supernatant of the upper phase was collected and used for GC analysis (TRACE GC Ultra Gas Chromatograph, Thermo Scientific, Waltham, USA). GC analysis was performed with the GC capillary column BPX70 (SGE Analytical Science, Milton Keynes, GB). As specified for rapeseed oil fatty acid composition, the fatty acid methyl ester mixture F07 (Carl Roth GmbH + Co. KG, Karlsruhe, Deutschland) was used as a standard. For analysis, the following temperature ramps were used: Start at 160 °C for 1 min, incremental increase of temperature every min by 15 °C until temperature reaches 210 °C, 210 °C for 30 s, 220 °C for 6 min. The samples were measured in two technical replicates.

Determination of seed germination characteristics

Mean germination time (MGT; in h), germination rate within 96 h (GR96; in %) and uniformity of germination, measured as the difference between the time to reach 10 and 90% of germination (U; in h), were phenotyped under in vitro conditions at 20 °C, using the automated phenotyping platform of the variety control office of the French national seed testing agency (Station Nationale d'Essais de Semences, Groupe d'Etude et de contrôle des Variétés et des Semences—GEVES, Angers, France). GR96 can be considered as the absolute germination rate, as no further increase in germination was observed 96 h after imbibition start. Germination was assayed under well-watered conditions as well as under moderate osmotic stress at an intensity of − 0.5 MPa. Seed germination analysis was carried out with 4 × 25 seeds per genotype and treatment. Detailed information about the phenotyping system is given by [33].

Analysis of seedling vigour

For analysis of seedling vigour, 12 seeds harvested from each container were sown in small plastic pots (8 × 7 × 7 cm) filled with 350 g sand. Germination and seedling growth took place in the greenhouse under controlled conditions as follows: 12 h/12 h light/dark, temperature 20–24 °C during the light period and 12 °C - 16 °C during the dark period; minimum light intensity 10 klx during the light period. The pots were watered via dish watering. Water capacity of the sand was determined as the difference in weight between sand filled pots dried for 24 h at 85 °C in a drying cabinet, and sand filled pots after watering. Water capacity was subsequently maintained up at 75% during the whole experiment. Seedlings were harvested 7 d after sowing. In each case, 12 seedlings were pooled and samples were shock-frozen immediately after fresh weight (FW) determination of each pooled sample. Consequently, three pooled samples were analyzed for each accession and maternal treatment. The shock-frozen samples were freeze-dried for 5 d and used for subsequent metabolite analysis.

Metabolite analysis

Metabolite analysis was performed on a Waters Acquity ultraperformance liquid chromatography machine with diode array detection (UPLC-DAD) using methods and software described in the Waters Corporation user manual. The manual was adapted for oilseed rape tissue by [34, 35]. The AccQtag method was used to quantify amino acids and the integration software Empower (Waters Corporation, Milford, USA) was used for analysis. Samples were resuspended in 100 mL distilled water. Subsequently, 5 mL were derivatized using AccQTag Ultra Derivatization Kit, according to the manufacturer's recommendations. An external standard of 100 mmol/L of each amino acid was run every 10 samples. Quantification of sugars was performed using a gas chromatography-flame ionization detector (GC-FID) System from Agilent Technologies (Santa Clara, CA, USA) according to [36]. The integrated Agilent software ChemStation Rev.B.04.02 was used for data analysis. Samples were resuspended in 50 mL pyridine (100%) with methoxamine hydrochloride (240 mmol/L), then derivatized with 50 mL MSTFA (N-methyl-N-(trimethylsilyl)trifluoro acetamide) (100%). An external standard containing 400 mmol/L of each sugar, sugar alcohol and organic acid was run every 10 samples.

Statistical analysis

Data analysis was carried out using R (version 3.3.3, 2017). Compliance with normal distribution and homogeneity of variances were evaluated using a Bartlett test. After global testing with ANOVA, Student's *t*-test was applied for pairwise comparisons. In terms of consistency testing, no adjustment was chosen for the statistical analysis.

Significances were tested at levels of $p < 0.1'$, $p < 0.05$*, $p < 0.01$** and $p < 0.001$***. Correlations were calculated applying the Pearson's product-moment correlation.

Additional files

Additional file 1: Seed fatty acid composition: Relative amounts of fatty acids in the seeds of 8 diverse winter oilseed rape genotypes cultivated under control and drought stress conditions. Values are means of three replicates + standard errors. Significant differences at $p < 0.1'$, $p < 0.05$*, $p < 0.01$** and $p < 0.001$***. (XLSX 13 kb)

Additional file 2: Effect of maternal drought stress on seedling fresh weight of 4 winter oilseed rape genotypes cultivated in a semi-controlled container trial after self-pollination. Bars are means of three replicates with standard errors. Significant differences at $p < 0.1'$. (TIF 4218 kb)

Additional file 3: Seedling metabolome: Concentrations of different nitrogen compounds, amino acids, sugars and organic acids in seedlings derived from 8 diverse winter oilseed rape genotypes cultivated under control and drought stress conditions. Values are means of three replicates + standard errors. Significant differences at $p < 0.1'$, $p < 0.05$*, $p < 0.01$** and $p < 0.001$***. (XLSX 18 kb)

Additional file 4: Trait correlations: Table summarizing all significant correlations ($p < 0.1$) between the two main parameters total maternal seed yield (Yield) and seedling fresh weight building of the progeny (SFW) and different single yield, seedling growth, seed quality and seedling metabolome parameters, determined in 8 diverse winter oilseed rape genotypes grown in a semi-controlled container trial under control (C) and drought stress conditions (S). S + C: Correlations were calculated among both treatments. S/C: Correlations were calculated between the relative values of each trait-trait combinations as quotient of value from stress treatment to value from control treatment. Seeds/Sil.: Number of seeds per silique, Seeds/Pl.: Number of seeds per plant. (JPG 236 kb)

Additional file 5: Soil water content during the stress trial before watering (above) and after watering (below). The numbers indicate days after stress initiation. Values are means of three replicates + standard errors. Significant differences at $p < 0.1'$, $p < 0.05$*, $p < 0.01$** and $p < 0.001$***. (TIF 4218 kb)

Abbreviations

++: high seed erucic acid and glucosinolate content; Arg: Arginine; Asn: Asparagine; Asp: Aspartate; BBCH: Scale for evalutaion of plant developmental stage developed from Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie; Cys: Cysteine; FAE1: Fatty acid elongase 1; FW: Fresh weight; GABA: γ-aminobutyric acid; Gln: Glutamine; Glu: Glutamate; Gly: Glycine; GR96: Germination rate within 96 h after imbibition start; His: Histidine; Hyp: Hydroxyproline; Ile: Isoleucine; Leu: Leucine; Lys: Lysine; Met: Methionine; Met.-Cys: Methyl-cysteine; MGT: Mean germination time; Orn: Ornithine; PCA: Principal component analysis; Phe: Phenylalanine; Pro: Proline; Ser: Serine; SMCSO: S-methylcystein sulfoxide; Thr: Threonine; Trp: Tryptophane; Tyr: Tyrosine; U: Uniformity; Val: Valine; α-Ala: α-Alanine; β-Ala: β-Alanine

Acknowledgements

The authors thank Liane Renno, Anja Pöltl, Mechthild Schwarte and Lothar Behle-Schalk for practical help and support during the container trial and Petra Degen for excellent technical assistance during NIRS and GC analysis. Further thank goes to GEVES for implementing the seed germination analysis at Phenosem platform (Anger, France) and INRA (Rennes, France) for undertaking the metabolite analysis. We also like to thank Katharina Tyson for proofreading the manuscript on English language style.

Funding

The work was performed within the framework of the transnational cooperation project SYBRACLIM with financial support from the Federal Ministry of Education and Research (BMBF) grant 031A549A and grant SN14/14–2 from the DFG (Deutsche Forschungsgemeinschaft) within the Priority Program 1530: Flowering Time Control.

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Authors' contributions

SVH, RJS and SVS designed the experiment. SVH and SVS performed the container trial and yield analysis. SVS and JNN designed and performed the greenhouse trial and sample preparation. SVH designed the seed quality, seed germination and metabolite analysis and performed the statistical analysis. SVH and SVS wrote the manuscript and all authors corrected and approved the final version.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Department of Plant Breeding, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany. [2]Department of Ecology, Environment and Plant Sciences, 106 91 Stockholm, Sweden.

Received: 26 July 2018 Accepted: 15 November 2018
Published online: 23 November 2018

References

1. Bates BC, Kundzewicz ZW, Wu S, Palutikof JP, Eds. Climate change and water. Technical Paper of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva. 2008. 210 pp. https://drive.google.com/file/d/0B1gFp6loo3akcFFFeGRRVFNYM0E/view.
2. Li Y, Ye W, Wang M, Yan X. Climate change and drought. A risk assessment of crop-yield impacts. Clim Res. 2009;39:31–46.
3. FAOSTAT Food and agriculture data. Food and Agriculture Organisation of the United Nations, Rome. 2017. http://www.fao.org/faostat/en. Accessed 8 Jan 2017.
4. Wittkop B, Snowdon RJ, Friedt W. Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. Euphytica. 2009;170:131–40.
5. Jensen CR, Mogensen VO, Mortensen G, Fieldsend JK, Milford GFJ, Andersen MN, et al. Seed glucosinolate, oil and protein contents of field-grown rape (*Brassica napus* L.) affected by soil drying and evaporative demand. Field Crop Res. 1996;47:93–105.
6. Gao J, Thelen KD, Min DH, Smith S, Hao X, Gehl R. Effects of manure and fertilizer applications on canola oil content and fatty acid composition. Agron J. 2010;102:790–7.
7. Onemli F. Fatty acid content of seed at different development stages in canola on different soil types with low organic matter. Plant Prod Sci. 2014;17:253–9.
8. Delourme R, Falentin C, Huteau V, Clouet V, Horvais R, Gandon B, et al. Genetic control of oil content in oilseed rape (*Brassica napus* L.). Theor Appl Genet. 2006;113:1331–45.
9. Ellis RH. Seed and seedling vigour in relation to crop growth and yield. Plant Growth Regul. 1992;11:249–55.
10. TeKrony DM, Egli DB. Relationship of seed vigor to crop yield: a review. Crop Sci. 1991;31:816–22.
11. Hohmann M, Stahl A, Rudloff J, Wittkop B, Snowdon RJ. Not a load of rubbish. Simulated field trials in large-scale containers. Plant Cell Environ. 2016;39:2064–73.
12. Richards RA, Thurling N. Variation between and within species of rapeseed (*Brassica campestris* and *B. napus*) in response to drought stress. I sensitivity at different stages of development. Aust J Agric Res. 1978;29:469–77.

13. Champolivier L, Merrien A. Effects of water stress applied at different growth stages to *Brassica napus* L. var. *oleifera* on yield, yield components and seed quality. Eur J Agron. 1996;5:153–60.
14. Din J, Khan SU, Ali I, Gurmani AR. Physiological and agronomic response of canola varieties to drought stress. J Anim Plant Sci. 2011;21:78–82.
15. Hatzig SV, Frisch M, Breuer F, Nesi N, Ducournau S, Wagner MH, et al. Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*. Front Plant Sci. 2015;6:221.
16. Dorion S, Lalonde S, Saini HS. Induction of male sterility in wheat by meiotic-stage water deficit is preceded by a decline in invertase activity and changes in carbohydrate metabolism in anthers. Plant Physiol. 1996;111:137–45.
17. Saini HS. Effects of water stress on male gametophyte development in plants. Sex Plant Reprod. 1997;10:67–73.
18. Meyer RC, Törjék O, Becher M, Altmann T. Heterosis of biomass production in *Arabidopsis*. Establishment during early development. Plant Physiol. 2004; 134:1813–23.
19. Bouchereau A, Clossais-Besnard N, Bensaoud A, Leport L, Renard M. Water stress effects on rapeseed quality. Eur J Agron. 1996;5:19–30.
20. Aslam MN, Nelson MN, Kailis SG, Bayliss KL, Speijers J, Cowling WA. Canola oil increases in polyunsaturated fatty acids and decreases in oleic acid in drought-stressed Mediterranean-type environments. Plant Breed. 2009;128:348–55.
21. Guo Y, Si P, Wang N, Wen J, Yi B, Ma C, et al. Genetic effects and genotype × environment interactions govern seed oil content in *Brassica napus* L. BMC Genet. 2017;18(1). https://doi.org/10.1186/s12863-016-0468-0.
22. Baud S, Lepiniec L. Regulation of *de novo* fatty acid synthesis in maturing oilseeds of *Arabidopsis*. Plant Physiol Bioch. 2009;47:448–55.
23. Lämke J, Bäurle I. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. Genome Biol. 2017;18:124.
24. Boyko A, Blevins T, Yao Y, Golubov A, Bilichak A, Ilnytskyy Y, et al. Transgenerational adaptation of *Arabidopsis* to stress requires DNA methylation and the function of dicer-like proteins. PLoS One. 2010;5:e9514.
25. Ou X, Zhang Y, Xu C, Lin X, Zang Q, Zhuang T, et al. Transgenerational inheritance of modified DNA methylation patterns and enhanced tolerance induced by heavy metal stress in rice (*Oryza sativa* L.). PLoS One. 2012;7:e41143.
26. Lang-Mladek C, Popova O, Kiok K, Berlinger M, Rakic B, Aufsatz W, et al. Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in *Arabidopsis*. Mol Plant. 2010;3:594–602.
27. Whittle CA, Otto SP, Johnston MO, Krochko JE. Adaptive epigenetic memory of ancestral temperature regime in *Arabidopsis thaliana*. Botany. 2009;87:650–7.
28. Lacey EP. Parental effects in *Plantago lanceolata* L. I.: a growth chamber experiment to examine pre- and postzygotic temperature effects. Evolution. 1995;50:865–78.
29. Galloway LF. The effect of maternal and paternal environments on seed characters in the herbaceous plant *Campanula americana* (*Campanulaceae*). Am J Bot. 2001;88:832–40.
30. Suter L, Widmer A. Environmental heat and salt stress induce transgenerational phenotypic changes in *Arabidopsis thaliana*. PLoS One. 2013;8:e60364.
31. Bus A, Körber N, Snowdon RJ, Stich B. Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. Theor Appl Genet. 2011;123:1413–23.
32. Tillmann P, Reinhardt TC, Paul C. Networking of near infrared spectroscopy instruments for rapeseed analysis: a comparison of different procedures. J Near Infrared Spectrosc. 2000;8:103–7.
33. Demilly D, Ducournau S, Wagner MH, Dürr C. Digital imaging of seed germination. In: Dutta Gupta S, Ibaraki Y, editors. Plant image analysis - fundamentals and applications. Boca Raton: CRC Press; 2014. p. 147–63.
34. Albert B, Le Cahérec F, Niogret MF, Faes P, Avice JC, Leport L, et al. Nitrogen availability impacts oilseed rape (*Brassica napus* L.) plant water status and proline production efficiency under water-limited conditions. Planta. 2012; 236:659–76.
35. Deleu C, Faes P, Niogret MF, Bouchereau A. Effects of the inhibitor of the g-aminobutyrate-transaminase, vinyl-gaminobutyrate, on development and nitrogen metabolism in *Brassica napus* seedlings. Plant Physiol Biochem. 2013;64:60–9.
36. Lugan R, Niogret MF, Kervazo L, Larher FR, Kopka J, Bouchereau A. Metabolome and water status phenotyping of *Arabidopsis* under abiotic stress cues reveals new insight into ESK1 function. Plant Cell Environ. 2009; 32:95–108.

The role of genome structure variation in the evolution and adaptation of flowering time in *Brassica* species

Successful plant breeding depends on genetic variation. For a long time, genetic variation was purely understood as sequence variation. Technological advances now have shown the abundance and importance of structural variation for trait variation, indicating that we need tools to screen large populations for this type of variant. This is even more urgent in polyploid crops like *Brassica napus*, where genomic redundancy increases the likelihood of structural variation. Aiming at a better understanding of climatic adaptation to inform breeding strategies, I have focused on the role of genome structure variation in flowering time genes in this and related species. In Chapter 1, I show the set up and results of a pipeline to reliable call copy number variants (CNVs) in large diversity sets in a group of selected target flowering time genes. Chapter 2 focuses on differential roles of different homologs of specific flowering time genes, indicating that findings from *A. thaliana* are necessary, but not sufficient to understand genetic regulation of related *Brassicaceae*. Chapter 3 elaborates on the interaction of flowering time and drought stress, showing that small RNA variation introduces a new layer of complexity in regulation of this complex trait. The following paragraphs will explain, link and discuss the main findings from the presented papers.

CNVs are expected to be widespread in the *B. napus* genome and are therefore assumed to also influence flowering time genes. We applied targeted deep sequencing to a large diversity set of 280 accessions of *B. napus* to reveal the extent and nature of the assumed CNVs in a collection of 35 flowering time related genes. Targeted deep sequencing is a short read based sequencing technology with the aim to sequence only genetic regions of interest. Here, we used the SureSelect Sequence capture methodology from Agilent technologies followed by Illumina single end sequencing. To this purpose, we developed a bait pool, which is an RNA oligonucleotide library used to hybridize the sample DNA. The oligonucleotides are complementary to the intended target sequence of the selected flowering time gene copies. As genomic resources for *B. napus* were limited at the time of bait development, baits were partly also developed using sequence information for the progenitor species *B. rapa* and *B. oleracea*. The original bait pool had been tested on four accessions in a preliminary experiment and was further improved based on the results published before (Schiessl et al. 2014) to increase specificity and enrichment factor. Likewise, we were able to sequence 184 gene copies, homologs of 35 *A. thaliana* genes, at an average coverage of 670 x. This coverage can then be used to reliably call sequence variants, but also CNVs, based on read coverage. We found 5216 target SNPs and 569 target InDels, of which 56 and 25 were predicted to have severe impacts on gene function, respectively, representing important allelic variants for future breeding approaches. Moreover, we detected 448 simple duplications, 490 simple deletions and 201 putative HNRTs across

the gene copies that could unambiguously be localized. This adds up to a slight net gene loss as expected for a recent polyploid (Renny-Byfield and Wendel 2014; Sankoff et al. 2010). At the same time, the distribution of duplications and deletions on the A and C subgenomes was different, indicating C subgenome copies tend to be lost, while A subgenome copies tend to be duplicated. Generally, including also off-target regions, we found that the C subgenome generally harbors almost twice as many events as the A subgenome, indicating that more DSBs took place, possibly due to the higher transposon content of the C subgenome (Parkin et al. 2014). HNRTs were also strongly biased between A and C in the same direction and were most frequent between chromosomes of high homeology like A02/C02 and A03/C03. This indicates beginning biased fractionation and suggests a higher number of DSBs in the C subgenome to be responsible for this bias. During homeologous recombination, the non-broken strand is normally used as template and its information is likely to be overretained both in SDSA and DSBR mechanisms (Waterworth et al. 2011). Moreover, HNRTs got much more frequent towards the telomeric ends and were almost absent from the centromeric areas, in full agreement with other authors, pointing to homeologous recombination as main repair mechanism (Nicolas et al. 2012; Samans et al. 2017). Although more than 87 % of CNVs occurred at a frequency lower than 10 %, each of the 280 accessions carried at least one CNV. The results indicate that CNVs are highly prevalent and need to be considered as important source of genetic variation. We therefore compared CNV patterns between different ecotypes of *B. napus* and found two HNRTs on chromosomes A09/C08 and A10/C09 encompassing copies of *Bna.PHYA, Bna.GA3OX1* and *Bna.FLC* being almost exclusive for swedes (ssp. *napobrassica*), indicating that these HNRTs played a role in formation of the subspecies. In case of the vernalisation regulator *Bna.FLC*, the A10 copy has replaced the C09 copies. Our own and other reasearcher's data point to a strong subfunctionalisation in *Bna.FLC* (see below), leaving the A10 copy responsible for most of the vernalisation requirement. This fits to the observation that swedes are extremely vernalisation dependent. Notably, such patterns were less obvious when comparing flowering time genes between winter and spring accessions. *Bna.PHY.A09* and *Bna.TEM1.C02* were deleted in about 24.6 % and 34.2 % of spring accessions, while only being deleted in 1.5 % and 2.9 % of winter accessions, respectively. At the same time, *Bna.SRR1.A03, Bna.SRR1.A10* and *Bna.EFS.A07* were duplicated in 23.7 %, 26.6 % and 21.5 % of winter accessions, while only being duplicated in 1.8 %, 1.8 % and 3.7 % of spring accessions, respectively, and *Bna.TFL1.A10* was duplicated in 43.9 % of spring accessions, while only being duplicated in 2.9 % of winter accessions. At the same time, sequence variants were more successful in separating winter from spring accessions, as 12 haplotypes could be defined which are sufficient to separate the subpopulations. This indicates that although some CNVs seem to be beneficial for a certain ecotype, they are far from being fixed, which might either indicate that they arose quite recently, or alternatively that we observe admixture of subpopulations with different

independent CNV events. In both cases, those CNVs could potentially have a diversifying effect. Structural variation could therefore be a mechanism to overcome reduction of genetic diversity by genetic bottlenecks like interspecific hybridization, not only by deleting or duplicating genes, but also by associated InDels at the DSB site and shifts in selection pressure on different genes. To better understand the dynamics of CNV arousal and fixation, we also sequenced two accessions of each *B. rapa* and *B. oleracea* for the same gene set. Therein, we discovered several stable translocations between the *B. napus* subgenome in comparison to the progenitors, indicating that genomic exchanges between subgenomes may indeed eventually get fixed, either due to genetic drift or due to selection. At the same time, we found a surprisingly high number of CNVs within the diploid progenitors, pointing to intrasubgenomic mechanisms of genomic rearrangements. Redundancy in the progenitor genomes is still about a factor of more than 2, which possibly might play a role in such a mechanism (Parkin et al. 2014; Wang et al. 2011b). Those findings might also be of relevance for other crops. As summarized in the presented review paper, CNVs were also found to play a role for trait variation in *Panicum virgatum*, *Solanum tuberosum* and *Triticum aestivum*, but reliable data on such traits are scarce due to the difficulty to genotype this type of variant on a large scale. Recent technologies advances like long read sequencing and reduced representation sequencing might soon overcome these barriers and genome-wide data for CNVs might become available on a population scale, allowing answering many of the leftover questions: Is the amount of CNVs we observed in flowering time genes representative for all genes? Are there CNVs explaining subspeciation events, and how quickly are CNVs purged out or fixed? How can we use structural variation to improve modern breeding strategies like genomic selection? And what can we learn from CNVs about the degree of subfunctionalisation?

Subfunctionalisation is hypothesized to be one of the major drivers of gene evolution (Renny-Byfield et al. 2014; Soltis and Soltis 2016), but is rarely observed in action due to the long periods involved. In *B. napus* and progenitors, we have an excellent model system to study such processes as they are happening, as they represent snap shots of a development at different time points. Assuming no subfunctionalisation took place, gene copies should act in an additive manner. Consequently, the same phenotypic effect should be observed for variation in each copy. However, when analyzing sequence variants called from a large resequencing panel of almost 1000 *B. napus* accessions (Wu et al. 2018), spring, semi-winter and winter ecotypes were found to mainly differ in one copy of two genes, namely *Bna.FT.A02* and *Bna.FLC.A10*. *Bna.FT* has six copies and *Bna.FLC* has nine annotated copies in *B. napus*, but the other five and eight, respectively, seem to have no or only little effect, indicating they do not act additively. Interestingly, the genetic variants within those genes are both promoter variants, indicating that gene expression was a major factor in the process of subfunctionalization. We therefore studied gene expression patterns of all nine annotated *Bna.FLC*

copies and found strong differentiation between the copies. While one copy on chromosome C03 seems to be a pseudogene, both copies on chromosome C09 and another copy on chromosome C03 did not react to cold anymore – a major feature of *FLC* gene function in *A. thaliana*. Among the remaining copies, only *Bna.FLC.A10* showed a population-wide differential expression pattern between winter and spring accessions, which was mostly, but not fully explanatory of vernalisation dependency. The expression pattern itself was almost completely associated to a small structural variant at the promoter-TSS junction where a long non-coding RNA is supposed to bind in the process of *Bna.FLC* downregulation. The non-responsive C09 copies also show structural variation at this site, indicating that structural variation at regulatory sites of flowering time genes had a decisive influence on geographic adaptation of *B. napus* and also influences the differentiation of gene copies. As pointed out, *Bna.FLC.A10* expression patterns were not solely explanatory for winter-spring-differentiation. We therefore also studied another candidate gene from our deep-sequencing trial, *Bna.SRR1*. In *A. thaliana*, this gene is involved in the photoperiod pathway and seems to interact with other pathways, mainly the vernalisation pathway. It has five copies in *B. napus*, of which two are almost fully identical in sequence (on A03 and A10). When assessing *Bna.SRR1* gene expression, we found that those identical copies are almost not expressed, while all the three remaining copies were differentially expressed between winter and spring accessions. Complementation assays into *A. thaliana* subsequently showed that only the copies on A02 and C02 were able to complement the *srr1-1* phenotype, while the C09 copy could not. Interestingly, the copies on A03, A10 and C09 are lacking a region of 21 amino acids, which is conserved between the A02/C02 copies and with *A. thaliana*. Analysis of downstream gene expression in those complemented lines has shown that the C09 copy can rescue *TEM1* expression, but not *CDF1*, indicating that this 21 aa region could be decisive for interaction with *CDF1,* which in turn seems to be a necessary step in controlling flowering time by *Bna.SRR1*. This is therefore an example how structural variants affect gene function by altering their protein interaction interface. The pattern of the deletion across the copies suggests that this structural variation predates the speciation event of *B. rapa* and *B. oleracea*, as it's conserved between the respective copies of the A and C subgenome. While the non-affected copies have obviously upheld their function in both progenitor species, the affected copies show a different pattern: while the A03/A10 copies are not expressed anymore, the C09 copy is still highly expressed and also still seems to be able to induce *TEM1* expression. This indicates different evolutionary trajectories for *Bna.SRR1.A10*, which obviously underwent pseudogenization, and *Bna.SRR1.C09*, which possibly acquired a different or modified role. It would be interesting to study *SRR1* expression in the progenitors to better understand the evolutionary consequences of these differential patterns. Understanding these differential patterns will eventually allow us to screen for useful variation in a much better focused way in future.

One of the major challenges we are facing in the context of adaptation breeding is abiotic stress during flowering. Surprisingly, not a lot is known about the interaction of flowering time regulation and drought stress, reflecting the complexity of both pathways and at the same time stressing the pressing need for more research. To understand how flowering time regulation and drought stress signaling do interact, we conducted a semi-controlled drought stress treatment in eight contrasting accessions of winter oilseed rape, using leaf samples for RNAseq and small RNA sequencing. We found that drought stress during flowering delayed plant development until harvest, reducing seed yield in five out of eight accessions. In the next generation, seeds from stressed plants showed a lower quality, but higher seedling vigor irrespective of yield level of the mother plant, indicating that some transgenerational mechanism exists to adapt the next generation to suboptimal conditions. Tolerance strategies varied strongly between accessions, indicating large variation and low selection pressure up to now. The same was true for gene expression patterns: whereas we did find general genetic responses for example on photosynthesis and stress signaling genes, specific resistance genes were mostly unique for each of the tolerant accessions. Moreover, we found hints that some of the resistance genes were targets of miRNAs, and the respective *MIRNA* genes could therefore also be breeding targets. Interestingly, we found that some of those stress-related miRNAs showed stress-specific variants, indicating that those *MIRNA* genes have a stress-responsive promoter. Flowering time was most likely affected via the gibberellin and age pathways and therefore also via small RNA signaling. Interestingly, most small RNAs were not predicted to target all copies of a gene, indicating that subfunctionalisation also affects small RNA regulation. Mutation of small RNAs or small RNA target sites is possibly one of the quickest and most effective ways to modify and modulate gene expression patterns in polyploids and may represent a first step in gene diversification and adaptation, as the example of *Bna.FLC* has shown.

In summary, the presented manuscripts show that structural variation across flowering time genes has influenced adaptation to different environments by differentially affecting copies of the same gene. Small RNA regulation and promoter variation seem to play a supportive role in subfunctionalization and therefore also shape the effect of structural variation. However, the view on this interplay is just opening, and we are far from seeing a complete model which would allow predicting adaptation responses from genetic variation. More interdisciplinary approaches are needed to fill in the knowledge gaps which remain, tearing down borders between basic and applied research.

Summary

Adaptation breeding is one of the key factors of minimizing possible negative impacts of climate change on plant production. Modern breeding approaches use genetic information to optimize the outcome of respective breeding programs. However, many crops are polyploid and therefore have complex and highly redundant genomes. The high degree of redundancy also increases the likelihood of structural genome variation. Moreover, duplicated genes may over time undergo sub-or neofunctionalisation, which complicates knowledge transfer from model systems and requires thoughtful approaches to dissect the underlying genetic regulation for adaptation traits. *Brassica napus* is the polyploid crop with the closest relationship to the model plant *Arabidopsis thaliana* and has a well-known evolutionary ancestry. In the presented works, *B. napus* was used as a model to study the influence of genomic redundancy, structural genome variation and subfunctionalisation on adaptation traits like flowering time and drought stress response. It could be shown that structural genome variation was extremely abundant in flowering time genes in *B. napus*, but also within its progenitor species *B. rapa* and *B. oleracea*. Some genomic exchange events obviously played an important role in formation of a subspecies, demonstrating the high adaptive value of structural variants. Moreover, it was found that important flowering time regulators underwent strong subfunctionalisation linked to either promoter variation or protein structure variation, indicating not all copies have the same adaptive value for breeding. When studying gene expression networks together with small RNA expression networks in reaction to drought stress at flowering, it was found that subfunctionalisation obviously also affects small RNA regulation patterns, adding a further dimension to the complexity of polyploid genomes. Obviously, different miRNA variants are upregulated under stress than under control conditions, indicating subfunctionalisation of small RNA expression assists in the fine-tuning of abiotic stress response in newly formed polyploids. In summary, it was found that understanding subfunctionalisation is key to judge the possible outcomes of both classical sequence variants and structural variants. In that respect, the variation of regulatory elements like promoters and small RNAs should attain more attention.

Zusammenfassung

Die Züchtung angepasster Sorten ist ein Schlüsselfaktor, um mögliche negative Einflüsse des Klimawandels auf die Pflanzenproduktion zu minimieren. Moderne Züchtungsmethoden nutzen genetische Informationen, um das Ergebnis entsprechender Züchtungsprogramme zu optimieren. Viele Kulturpflanzenarten sind jedoch polyploid und besitzen daher komplexe und hochgradig redundante Genome. Der hohe Grad an Redundanz erhöht auch die Wahrscheinlichkeit struktureller Genomvariation. Zudem können duplizierte Gene über längere Zeiträume eine Sub-oder Neofunktionalisierung durchlaufen, was den Wissenstransfer aus Modellsystemen erschwert und gut durchdachte Forschungsansätze zur Aufklärung genetischer Regulation von Anpassungsmerkmalen erfordert. *Brassica napus* ist die polyploide Kulturpflanze, die am nächsten mit der Modellpflanze *Arabidopsis thaliana* verwandt ist, gleichzeitig weist sie eine gut bekannte Evolutionsgeschichte auf. In den hier vorgestellten Arbeiten wurde *B. napus* als Modellsystem genutzt, um den Einfluss von genomischer Redundanz, struktureller Genomvariation und Subfunktionalisierung auf Anpassungsmerkmale wie Blühzeit und Trockenstressreaktion zu untersuchen. Es konnte gezeigt werden, dass strukturelle Genomvariation in Blühzeitgenen von *B. napus* extrem häufig ist, und auch in den Vorläuferarten *B. rapa* und *B. oleracea* häufig vorkommt. Einige der genomsichen Austauschereignisse spielten offensichtliche eine wichtige Rolle bei der Entstehung einer Unterart, was den adaptiven Wert struktureller Genomvariation unterstreicht. Darüber hinaus wurde in wichtigen Blühzeitgenen eine starke Subfunktionalisierung beobachtet, die entweder mit Promotorvariation oder Variation der Proteinstruktur einherging, was darauf hinweist, dass nicht alle Genkopien den gleichen adaptiven Wert besitzen. Durch die gemeinsame Untersuchung von Genexpressionsnetzwerken mit der Expression kleiner RNAs bei Trockenstress während der Blüte wurde deutlich, dass Subfunktionalisierung auch die Regulationsmuster von kleinen RNAs betrifft, was der Komplexität polyploider Genome eine weitere Dimension hinzufügt. Offensichtlich werden unter Stress andere miRNAs hochreguliert als unter Kontrollbedingungen, was darauf hinweist, dass die Subfunktionalisierung kleiner RNAs in neu gebildeten Polyploiden dazu beiträgt, die Stressantwort anzupassen. Insgesamt wurde festgestellt, dass das Verständnis der Subfunktionalisierung der Schlüssel ist, um den erwarteten Effekt sowohl von klassischer Sequenzvariation als auch von struktureller Genomvariation beurteilen zu können. In diesem Zusammenhang sollte die Variation regulatorischer Elemente wie Promotoren und kleiner RNAs mehr Aufmerksamkeit erfahren.

Further publications

Apart from the works presented in this thesis, I have also contributed to the following publications:

Book chapter

Schiessl SV, Mason AS: Ancient and recent polyploid evolution in *Brassica* (2019), accepted

Peer-reviewed journal articles

Hatzig SV, **Schiessl S**, Stahl A, Snowdon RJ (2015) Characterizing root response phenotypes by neural network analysis. EXBOTJ 66:5617–5624. doi: 10.1093/jxb/erv235

Shah S, Weinholdt C, Jedrusik N, Molina C, Zou J, Große I, **Schiessl S**, Jung C, Emrani N (2018) Whole-transcriptome analysis reveals genetic factors underlying flowering time regulation in rapeseed (Brassica napus L.). Plant, Cell & Environment 41:1935–1947. doi: 10.1111/pce.13353

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 Biotechnology Journal. doi: 10.1111/pbi.12867

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 Biotechnology Journal. doi: 10.1111/pbi.12732

Gaebelein R, **Schiessl SV**, Samans B, Batley J, Mason AS (2019) Inherited allelic variants and novel karyotype changes influence fertility and genome stability in Brassica allohexaploids. New Phytol. doi: 10.1111/nph.15804

Mwathi MW, **Schiessl SV**, Batley J, Mason AS (2019) "Doubled-haploid" allohexaploid Brassica lines lose fertility and viability and accumulate genetic variation due to genomic instability. Chromosoma. doi: 10.1007/s00412-019-00720-w

The following publications have been included in my PhD thesis:

Schiessl S, Samans B, Hüttel B, Reinhardt R, Snowdon RJ (2014) Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. Front. Plant Sci. 5:404. doi: 10.3389/fpls.2014.00404

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

Acknowledgements

My attitude towards achievements has been majorly tailored quite early on by the movie “Cool Runnings”. The bob slide team’s coach tells the bob driver the following two sentences: “A gold medal is a wonderful thing. But if you’re not enough without it, you’ll never be enough with it.” Having such a coach is even more wonderful than a gold medal, and I was extremely lucky to have even two of them. During my postdoc projects, Rod Snowdon and Annaliese Mason gifted me with much of their valuable time, supportive feedback, helpful advice, necessary challenges, surprising opportunities and invaluable trust. I wouldn’t be where I am now if they would not have taken the effort to support me again and again.

A good race does not only need a supportive coach, but also a powerful team. Wonderful colleagues at the institute of plant breeding at the JLU Giessen have turned science into a home, both in Rod’s and in Annaliese’s group, allowing me to stay joyful after defeats, rethink stupid ideas, overcome doubting myself and eagerly run into new battle fields. This is, in chronological order, the achievement of my office colleagues Sarah Hatzig, Anna Stein, Paul Vollrath, Manar Makhoul and Jenny Huey Tyng Lee, but also of other lunch company like Roman Gäbelein, Isabelle Deppé and Katharina Tyson. Moreover, very skillful and organized staff prevented failures and allowed me to also celebrate victories – and this is true for all levels, for the best secretary team on earth (featuring Sabine Schomber and Ulla Riedmeier), the magical greenhouse mystery team (Benjamin Wittkop, Annette Plank, Birgit Keiner, Juliette Kellermann and more), the lab wizards and witches (Liane Renno and Stavros Tzigos), the knights of the field station in Rauischholzhausen (Lothar Behle-Schalk, Mechthild Schwarte and Anja Pöltl), the bioinformatics task force (first Birgit Samans and then Jenny Huey Tyng Lee) and also to the young and motivated student helpers (Max Merlau, Natalie Skouteris) within the institute. And finally, I was warmly welcomed in the Senckenberg institute for Molecular Evolution of plants for my beginning year as a guest scientist in the lab of Georg Zizka. This was possible because Juraj Paule was extremely supportive and open for this collaboration, which turns out to be fruitful and useful for both sides.

This shows that success is a multi-location trait, and none of the projects we conducted would have been possible without the valuable cooperation with other institutes in Germany and abroad. We cooperated successfully with the GenomeCenter in Cologne (Richard Reinhardt/ Bruno Hüttel), the Plant Breeding Institute of CAU in Kiel (Christian Jung/ Smit Shah), the circadian clock lab in Bielefeld (Dorothee Staiger/ Mikael Johannson), the Plant Breeding Institute at the University of Göttingen (Antje Schierholt), the INRA in France (Anne-Marie Chèvre), the John Innes Center in the UK (Eleri Tudor), the University of Western Australia (Dave Edwards/ Bhavna Hurgobin/ Jacqueline Batley/ Margareth Mwathi), the Niigata University in Japan (Dan Shea), and with both the Southwest

University (Wei Qian) and Hunan Agricultural University (Lunwen Qian) in China. Their contributions and input have been invaluable for the advances we were able to take, and I'm very much looking forward to take new challenges together with those named.

But this work is dedicated to my students. I never enjoyed my work more than in teaching the little bit I know about science and watching it to grow bigger and bigger in their hands. It was their questions, their energy, their doubts and their courage that inspired and motivated me to stay critical, curious and still able to be amazed. I therefore like to thank Katharina Tyson, Kathrin Marquardt, Daniela Quezada-Martinez, Alban Mariette, Roberto Santamaria, Laurin Spahn, Franziska Hintermeier, Pascal Specht, Natalie Skouteris and Ellen Tebartz for their trust, joyfulness and power, turning ideas into reality. You are the reason why this present work got necessary and I hope it makes you as proud as it made me in being your supervisor.

In "Cool Runnings", the driver asks the coach afterwards: "Coach, how do I know that I'm enough?". That's indeed the main question, and I am very lucky that I found someone who has a million good answers to that. She is always there to remind me of them before I forget, carrying me over each infinite abyss when my wings are broken, healing the wounding, sending me back to the sky and waiting for me to come back, regardless if I bring star dust or broken bones again. This is a miracle I will never fully understand, but it turns me into the most grateful person every single day: I thank my wife Sina Weidenweber for all her endless care and constant support during this time. You made it worth trying.

References

Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. Curr Opin Plant Bio 8:135–141. doi: 10.1016/j.pbi.2005.01.001

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

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

Alvarez-Buylla ER, Benitez M, Corvera-Poire A, Cador AC, Folter S de, Buen AG de, Garay-Arroyo A, Garcia-Ponce B, Jaimes-Miranda F, Perez-Ruiz RV, Piñeyro-Nelson A, Sanchez-Corrales YM (2010) Flower Development. The Arabidopsis Book. doi: 10.1199/tab.0127

An H, Qi X, Gaynor ML, Hao Y, Gebken SC, Mabry ME, McAlvay AC, Teakle GR, Conant GC, Barker MS, Fu T, Yi B, Pires JC (2019) Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid *Brassica napus*. Nat Commun 10:2878. doi: 10.1038/s41467-019-10757-1

Andorf C, Beavis WD, Hufford M, Smith S, Suza WP, Wang K, Woodhouse M, Yu J, Lübberstedt T (2019) Technological advances in maize breeding: past, present and future. TAAG 132:817–849. doi: 10.1007/s00122-019-03306-3

Andrés F, Coupland G (2012) The genetic basis of flowering responses to seasonal cues. Nat Rev Genet 13:627–639. doi: 10.1038/nrg3291

Arrigo N, Barker MS (2012) Rarely successful polyploids and their legacy in plant genomes. Curr Opin Plant Bio 15:140–146. doi: 10.1016/j.pbi.2012.03.010

Axelsson T, Bowman CM, Sharpe AG, Lydiate DJ, Lagercrantz U (2000) Amphidiploid *Brassica juncea* contains conserved progenitor genomes. Genome 43:679–688. doi: 10.1139/g00-026

Bashir T, Chandra Mishra R, Hasan MM, Mohanta TK, Bae H (2018) Effect of Hybridization on Somatic Mutations and Genomic Rearrangements in Plants. Int J Mol Sci 19:3758. doi: 10.3390/ijms19123758

Bayer PE, Hurgobin B, Golicz AA, Chan C-KK, Yuan Y, Lee H, Renton M, Meng J, Li R, Long Y, Zou J, Bancroft I, Chalhoub B, King GJ, Batley J, Edwards D (2017) Assembly and comparison of two closely related *Brassica napus* genomes. Plant Biotechnol J 15:1602–1610. doi: 10.1111/pbi.12742

Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. PNAS 107:18724–18728. doi: 10.1073/pnas.0909766107

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

Cao Y, Wen L, Wang Z, Ma L (2015) SKIP Interacts with the Paf1 Complex to Regulate Flowering via the Activation of *FLC* Transcription in *Arabidopsis*. Mol Plant 8:1816–1819. doi: 10.1016/j.molp.2015.09.004

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 345:950–953. doi: 10.1126/science.1253435

Chang C, Lu J, Zhang H-P, Ma C-X, Sun G (2015) Copy Number Variation of Cytokinin Oxidase Gene *Tackx4* Associated with Grain Weight and Chlorophyll Content of Flag Leaf in Common Wheat. PLoS ONE 10:e0145970. doi: 10.1371/journal.pone.0145970

Chang HHY, Pannunzio NR, Adachi N, Lieber MR (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol 18:495–506. doi: 10.1038/nrm.2017.48

Chen S, Ren F, Zhang L, Liu Y, Chen X, Li Y, Zhang L, Zhu B, Zeng P, Li Z, Larkin RM, Kuang H (2018) Unstable Allotetraploid Tobacco Genome due to Frequent Homeologous Recombination, Segmental Deletion, and Chromosome Loss. Mol Plant 11:914–927. doi: 10.1016/j.molp.2018.04.009

Chen ZJ (2010) Molecular mechanisms of polyploidy and hybrid vigor. Trends Plant Sci 15:57–71. doi: 10.1016/j.tplants.2009.12.003

Cheng F, Wu J, Fang L, Sun S, Liu B, Lin K, Bonnema G, Wang X (2012) Biased gene fractionation and dominant gene expression among the subgenomes of Brassica rapa. PLoS ONE 7:e36442. doi: 10.1371/journal.pone.0036442

Chevalier C, Bourdon M, Pirrello J, Cheniclet C, Gévaudant F, Frangne N (2014) Endoreduplication and fruit growth in tomato: evidence in favour of the karyoplasmic ratio theory. J Exp Bot 65:2731–2746. doi: 10.1093/jxb/ert366

Choi K, Kim J, Hwang H-J, Kim S, Park C, Kim SY, Lee I (2011) The FRIGIDA complex activates transcription of *FLC*, a strong flowering repressor in *Arabidopsis*, by recruiting chromatin modification factors. Plant Cell 23:289–303. doi: 10.1105/tpc.110.075911

Clarke WE, Higgins EE, Plieske J, Wieseke R, Sidebottom C, Khedikar Y, Batley J, Edwards D, Meng J, Li R, Lawley CT, Pauquet J, Laga B, Cheung W, Iniguez-Luy F, Dyrszka E, Rae S, Stich B, Snowdon RJ, Sharpe AG, Ganal MW, Parkin IAP (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. TAAG 129:1887–1899. doi: 10.1007/s00122-016-2746-7

Coster W de, van Broeckhoven C (2019) Newest Methods for Detecting Structural Variations. Trends Biotechnol 37:973–982. doi: 10.1016/j.tibtech.2019.02.003

Cutter AR, Hayes JJ (2015) A brief review of nucleosome structure. FEBS Lett 589:2914–2922. doi: 10.1016/j.febslet.2015.05.016

Daley JM, Kwon Y, Niu H, Sung P (2013) Investigations of Homologous Recombination Pathways and Their Regulation. Yale J Biol Med:453–461

Diez CM, Roessler K, Gaut BS (2014) Epigenetics and plant genome evolution. Curr Opin Plant Bio 18:1–8. doi: 10.1016/j.pbi.2013.11.017

Dreni L, Kater MM (2014) *MADS* reloaded: evolution of the *AGAMOUS* subfamily genes. New Phytol 201:717–732. doi: 10.1111/nph.12555

Ea V, Baudement M-O, Lesne A, Forné T (2015) Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. Genes 6:734–750. doi: 10.3390/genes6030734

Evans J, Crisovan E, Barry K, Daum C, Jenkins J, Kunde-Ramamoorthy G, Nandety A, Ngan CY, Vaillancourt B, Wei C-L, Schmutz J, Kaeppler SM, Casler MD, Buell CR (2015) Diversity and population structure of northern switchgrass as revealed through exome capture sequencing. Plant J. 84:800–815. doi: 10.1111/tpj.13041

Faria R, Navarro A (2010) Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends ecol evol 25:660–669. doi: 10.1016/j.tree.2010.07.008

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 Genomics 15:498. doi: 10.1186/1471-2164-15-498

Gabur I, Chawla HS, Snowdon RJ, Parkin IAP (2019) Connecting genome structural variation with complex traits in crop plants. TAAG 132:733–750. doi: 10.1007/s00122-018-3233-0

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic Changes in Resynthesized *Brassica napus* and Their Effect on Gene Expression and Phenotype. Plant Cell 19:3403–3417. doi: 10.1105/tpc.107.054346

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

Göbel U, Arce AL, He F, Rico A, Schmitz G, Meaux J de (2018) Robustness of Transposable Element Regulation but No Genomic Shock Observed in Interspecific Arabidopsis Hybrids. Genome Biol Evol 10:1403–1415. doi: 10.1093/gbe/evy095

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 (2016) The pangenome of an agronomically important crop plant *Brassica oleracea*. Nat Commun 7:13390. doi: 10.1038/ncomms13390

Golloshi R, Sanders JT, McCord RP (2017) Genome organization during the cell cycle: unity in division. Wires Syst Biol Med 9. doi: 10.1002/wsbm.1389

Grandke F, Snowdon R, Samans B (2017) gsrc: an R package for genome structure rearrangement calling. Bioinformatics 33:545–546. doi: 10.1093/bioinformatics/btw648

He Y (2009) Control of the transition to flowering by chromatin modifications. Mol Plant 2:554–564. doi: 10.1093/mp/ssp005

He Y, Doyle MR, Amasino RM (2004) PAF1-complex-mediated histone methylation of *FLOWERING LOCUS C* chromatin is required for the vernalization-responsive, winter-annual habit in *Arabidopsis*. Gene Dev 18:2774–2784. doi: 10.1101/gad.1244504

Hidalgo O, Pellicer J, Christenhusz M, Schneider H, Leitch AR, Leitch IJ (2017) Is There an Upper Limit to Genome Size? Trends Plant Sci 22:567–573. doi: 10.1016/j.tplants.2017.04.005

Hou J, Long Y, Raman H, Zou X, Wang J, Dai S, Xiao Q, Li C, Fan L, Liu B, Meng J (2012) A Tourist-like *MITE* insertion in the upstream region of the *BnFLC.A10* gene is associated with vernalization requirement in rapeseed (*Brassica napus* L.). BMC Plant Biol 12:238. doi: 10.1186/1471-2229-12-238

Hu X, Kong X, Wang C, Ma L, Zhao J, Wei J, Zhang X, Loake GJ, Zhang T, Huang J, Yang Y (2014) Proteasome-Mediated Degradation of FRIGIDA Modulates Flowering Time in Arabidopsis during Vernalization. Plant Cell. doi: 10.1105/tpc.114.132738

Hu Y, Ren J, Peng Z, Umana AA, Le H, Danilova T, Fu J, Wang H, Robertson A, Hulbert SH, White FF, Liu S (2018) Analysis of Extreme Phenotype Bulk Copy Number Variation (XP-CNV) Identified the Association of rp1 with Resistance to Goss's Wilt of Maize. Front Plant Sci 9:110. doi: 10.3389/fpls.2018.00110

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. doi: 10.1111/pbi.12867

Irwin JA, Lister C, Soumpourou E, Zhang Y, Howell EC, Teakle G, Dean C (2012) Functional alleles of the flowering time regulator *FRIGIDA* in the *Brassica oleracea* genome. BMC Plant Biol 12:21. doi: 10.1186/1471-2229-12-21

Irwin JA, Soumpourou E, Lister C, Ligthart J-D, Kennedy S, Dean C (2016) Nucleotide polymorphism affecting FLC expression underpins heading date variation in horticultural brassicas. Plant J. 87:597–605. doi: 10.1111/tpj.13221

Jaeger KE, Graf A, Wigge PA (2006) The control of flowering in time and space. J Exp Bot 57:3415–3418

Jenczewski E, Eber F, Grimaus A, Huet S, Lucas MO, Monod H, Chèvre A-M (2003) *PrBn*, a major gene controlling homeologous pairing in oilseed rape (*Brassica napus*) haploids. Genetics

Johansson M, Staiger D (2015) Time to flower: interplay between photoperiod and the circadian clock. J Exp Bot 66:719–730. doi: 10.1093/jxb/eru441

Kagale S, Robinson SJ, Nixon J, Xiao R, Huebert T, Condie J, Kessler D, Clarke WE, Edger PP, Links MG, Sharpe AG, Parkin IAP (2014) Polyploid evolution of the *Brassicaceae* during the Cenozoic era. Plant Cell 26:2777–2791. doi: 10.1105/tpc.114.126391

Kim D-H, Doyle MR, Sung S, Amasino RM (2009) Vernalization: Winter and the Timing of Flowering in Plants. Annu. Rev. Cell Dev. Biol. 25:277–299. doi: 10.1146/annurev.cellbio.042308.113411

Kumar SV (2018) H2A.Z at the Core of Transcriptional Regulation in Plants. Mol Plant 11:1112–1114. doi: 10.1016/j.molp.2018.07.002

Kyriakidou M, Tai HH, Anglin NL, Ellis D, Strömvik MV (2018) Current Strategies of Polyploid Plant Genome Sequence Assembly. Front Plant Sci 9:1660. doi: 10.3389/fpls.2018.01660

Lagercrantz U (1998) Comparative Mapping Between *Arabidopsis thaliana* and *Brassica nigra* Indicates That Brassica Genomes Have Evolved Through Extensive Genome Replication Accompanied by Chromosome Fusions and Frequent Rearrangements. Genetics 150:1217–1228

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 9:13–20. doi: 10.1046/j.1365-313X.1996.09010013.x

Lashermes P, Hueber Y, Combes M-C, Severac D, Dereeper A (2016) Inter-genomic DNA Exchanges and Homeologous Gene Silencing Shaped the Nascent Allopolyploid Coffee Genome (*Coffea arabica* L.). G3 6:2937–2948. doi: 10.1534/g3.116.030858

Lawrence EJ, Griffin CH, Henderson IR (2017) Modification of meiotic recombination by natural variation in plants. J Exp Bot 68:5471–5483. doi: 10.1093/jxb/erx306

Lee TG, Diers BW, Hudson ME (2016) An efficient method for measuring copy number variation applied to improvement of nematode resistance in soybean. Plant J. 88:143–153. doi: 10.1111/tpj.13240

Li P, Zhang S, Li F, Zhang S, Zhang H, Wang X, Sun R, Bonnema G, Borm TJA (2017) A Phylogenetic Analysis of Chloroplast Genomes Elucidates the Relationships of the Six Economically Important *Brassica* Species Comprising the Triangle of U. Front Plant Sci 8:716. doi: 10.3389/fpls.2017.00111

Lisch D (2012) How important are transposons for plant evolution? Nat Rev Genet 14:49–61. doi: 10.1038/nrg3374

Liu Z, Adamczyk K, Manzanares-Dauleux M, Eber F, Lucas M-O, Delourme R, Chèvre AM, Jenczewski E (2006) Mapping *PrBn* and other quantitative trait loci responsible for the control of homeologous chromosome pairing in oilseed rape (*Brassica napus* L.) haploids. Genetics 174:1583–1596. doi: 10.1534/genetics.106.064071

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin, Isobel A P, Zhao M, Ma J, Yu J, Huang S, Wang X, Wang J, Lu K, Fang Z, Bancroft I, Yang T-J, Hu Q, Wang X, Yue Z, Li H, Yang L, Wu J, Zhou Q, Wang W, King GJ, Pires JC, Lu C, Wu Z, Sampath P, Wang Z, Guo H, Pan S, Yang L, Min J, Zhang D, Jin D, Li W, Belcram H, Tu J, Guan M, Qi C, Du D, Li J, Jiang L, Batley J, Sharpe AG, Park B-S, Ruperao P, Cheng F, Waminal NE, Huang Y, Dong C, Wang L, Li J, Hu Z, Zhuang M, Huang Y, Huang J, Shi J, Mei D, Liu J, Lee T-H, Wang J, Jin H, Li Z, Li X, Zhang J, Xiao L, Zhou Y, Liu Z, Liu X, Qin R, Tang X, Liu W, Wang Y, Zhang Y, Lee J, Kim HH, Denoeud F, Xu X, Liang X, Hua W, Wang X, Wang J, Chalhoub B, Paterson AH (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun 5:3930. doi: 10.1038/ncomms4930

Lukens L, Zou F, Lydiate D, Parkin IAP, Osborn TC (2003) Comparison of a *Brassica oleracea* Genetic Map With the Genome of *Arabidopsis thaliana*. Genetics:359–372

Lukens LN (2005) Patterns of Sequence Loss and Cytosine Methylation within a Population of Newly Resynthesized *Brassica napus* Allopolyploids. Plant Physiol 140:336–348. doi: 10.1104/pp.105.066308

Lysak MA (2005) Chromosome triplication found across the tribe *Brassiceae*. Genome Res 15:516–525. doi: 10.1101/gr.3531105

Maron LG, Guimarães CT, Kirst M, Albert PS, Birchler JA, Bradbury PJ, Buckler ES, Coluccio AE, Danilova TV, Kudrna D, Magalhaes JV, Piñeros MA, Schatz MC, Wing RA, Kochian LV (2013) Aluminum tolerance in maize is associated with higher MATE1 gene copy number. PNAS 110:5241–5246. doi: 10.1073/pnas.1220766110

Mason AS, Batley J (2015) Creating new interspecific hybrid and polyploid crops. Trends Biotechnol 33:436–441. doi: 10.1016/j.tibtech.2015.06.004

Mason AS, Pires JC (2015) Unreduced gametes: meiotic mishap or evolutionary mechanism? Trends Genet 31:5–10. doi: 10.1016/j.tig.2014.09.011

Mason AS, Higgins EE, Snowdon RJ, Batley J, Stein A, Werner C, Parkin IAP (2017) A user guide to the *Brassica* 60K Illumina Infinium™ SNP genotyping array. TAAG 130:621–633. doi: 10.1007/s00122-016-2849-1

McClintock B (1984) The significance of responses of the genome to challenges. Science:792–801

Mercier R, Mézard C, Jenczewski E, Macaisne N, Grelon M (2015) The molecular biology of meiosis in plants. Annu Rev Plant Biol 66:297–327. doi: 10.1146/annurev-arplant-050213-035923

Miller M, Zhang C, Chen ZJ (2012) Ploidy and Hybridity Effects on Growth Vigor and Gene Expression in *Arabidopsis thaliana* Hybrids and Their Parents. G3 2:505–513. doi: 10.1534/g3.112.002162

Morinaga T (1934) Interspecific hybridization in Brassica. Cytologia:62–67

Mwathi MW, Schiessl SV, Batley J, Mason AS (2019) "Doubled-haploid" allohexaploid *Brassica* lines lose fertility and viability and accumulate genetic variation due to genomic instability. Chromosoma. doi: 10.1007/s00412-019-00720-w

Nagaharu U (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilisation. Jap J Bot 7:389–452

Nicolas SD, Monod H, Eber F, Chevre A-M, Jenczewski E (2012) Non-random distribution of extensive chromosome rearrangements in *Brassica napus* depends on genome organization. Plant J 70:691–703. doi: 10.1111/j.1365-313X.2012.04914.x

Oh S, Park S, van Nocker S (2008) Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. PLoS Genet 4:e1000077. doi: 10.1371/journal.pgen.1000077

Orr-Weaver TL (2015) When bigger is better: the role of polyploidy in organogenesis. Trends Genet 31:307–315. doi: 10.1016/j.tig.2015.03.011

Osborn TC, Kole C, Parkin IAP, Sharpe AG, Kuiper M, Lydiate DJ, Trick M (1997) Comparison of Flowering Time Genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. Genetics:1123–1129

Otto SP (2007) The evolutionary consequences of polyploidy. Cell 131:452–462. doi: 10.1016/j.cell.2007.10.022

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. doi: 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

Parkin IAP, Clarke WE, Sidebottom C, Zhang W, Robinson SJ, Links MG, Karcz S, Higgins EE, Fobert P, Sharpe AG (2010) Towards unambiguous transcript mapping in the allotetraploid *Brassica napus*. Genome 53:929–938. doi: 10.1139/G10-053

Pelé A, Rousseau-Gueutin M, Chèvre A-M (2018) Speciation Success of Polyploid Plants Closely Relates to the Regulation of Meiotic Recombination. Front Plant Sci 9:907. doi: 10.3389/fpls.2018.00907

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 82:675–688

Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot 56:1–14. doi: 10.1093/jxb/eri025

Pyatnitskaya A, Borde V, Muyt A de (2019) Crossing and zipping: molecular duties of the ZMM proteins in meiosis. Chromosoma. doi: 10.1007/s00412-019-00714-8

Qi X, An H, Ragsdale AP, Hall TE, Gutenkunst RN, Chris Pires J, Barker MS (2017) Genomic inferences of domestication events are corroborated by written records in *Brassica rapa*. Mol Ecol 26:3373–3388. doi: 10.1111/mec.14131

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. doi: 10.1016/j.molp.2016.10.017

Rakow G (2004) Species Origin and Economic Importance of *Brassica*. Biotechnology in Agriculture and Forestry, vol 54. Springer, Berlin

Rang FJ, Kloosterman WP, Ridder J de (2018) From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biol 19:90. doi: 10.1186/s13059-018-1462-9

Renny-Byfield S, Wendel JF (2014) Doubling down on genomes: polyploidy and crop plants. Am J Bot 101:1711–1725. doi: 10.3732/ajb.1400119

Ridge S, Brown PH, Hecht V, Driessen RG, Weller JL (2015) The role of BoFLC2 in cauliflower (*Brassica oleracea* var. *botrytis* L.) reproductive development. J Exp Bot 66:125–135. doi: 10.1093/jxb/eru408

Robert LS, Robson F, Sharpe A, Lydiate D, Coupland G (1998) Conserved structure and function of the *Arabidopsis* flowering time gene *CONSTANS* in *Brassica napus*. Plant Mol Biol:763–772

Rodgers-Melnick E, Bradbury PJ, Elshire RJ, Glaubitz JC, Acharya CB, Mitchell SE, Li C, Li Y, Buckler ES (2015) Recombination in diverse maize is stable, predictable, and associated with genetic load. PNAS 112:3823–3828. doi: 10.1073/pnas.1413864112

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

Sankoff D, Zheng C, Zhu Q (2010) The collapse of gene complement following whole genome duplication. BMC Genomics 11:313. doi: 10.1186/1471-2164-11-313

Sarilar V, Palacios PM, Rousselet A, Ridel C, Falque M, Eber F, Chèvre A-M, Joets J, Brabant P, Alix K (2013) Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized *Brassica napus* allotetraploids. New Phytol 198:593–604. doi: 10.1111/nph.12156

Scelfo A, Fachinetti D (2019) Keeping the Centromere under Control: A Promising Role for DNA Methylation. Cells 8. doi: 10.3390/cells8080912

Schiessl S, Samans B, Hüttel B, Reinhardt R, Snowdon RJ (2014) Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. Front Plant Sci 5:404. doi: 10.3389/fpls.2014.00404

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

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

Schnable PS, Springer NM (2013) Progress Toward Understanding Heterosis in Crop Plants. Annu Rev Plant Biol 64:71–88. doi: 10.1146/annurev-arplant-042110-103827

Schranz ME, Quijada P, Sung S-B, Lukens L, Amasino R, Osborn TC (2002) Characterization and Effects of the Replicated Flowering Time Gene FLC in *Brassica rapa*. Genetics:1457–1468

Selmecki AM, Maruvka YE, Richmond PA, Guillet M, Shoresh N, Sorenson AL, De S, Kishony R, Michor F, Dowell R, Pellman D (2015) Polyploidy can drive rapid adaptation in yeast. Nature 519:349–352. doi: 10.1038/nature14187

Serra H, Choi K, Zhao X, Blackwell AR, Kim J, Henderson IR (2018) Interhomolog polymorphism shapes meiotic crossover within the Arabidopsis RAC1 and RPP13 disease resistance genes. PLoS Genet 14:e1007843. doi: 10.1371/journal.pgen.1007843

Skalická K, Lim KY, Matyásek R, Koukalová B, Leitch AR, Kovarík A (2003) Rapid evolution of parental rDNA in a synthetic tobacco allotetraploid line. Am J Bot:988–996

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

Snowdon RJ, Iniguez Luy FL (2012) Potential to improve oilseed rape and canola breeding in the genomics era. Plant Breeding 131:351–360. doi: 10.1111/j.1439-0523.2012.01976.x

Soltis DE, Misra BB, Shan S, Chen S, Soltis PS (2016) Polyploidy and the proteome. Biochim Biophys Acta 1864:896–907. doi: 10.1016/j.bbapap.2016.03.010

Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. Annu Rev Plant Biol 60:561–588. doi: 10.1146/annurev.arplant.043008.092039

Soltis PS, Soltis DE (2016) Ancient WGD events as drivers of key innovations in angiosperms. Curr Opin Plant Bio 30:159–165. doi: 10.1016/j.pbi.2016.03.015

Song Q, Chen ZJ (2015) Epigenetic and developmental regulation in plant polyploids. Curr Opin Plant Bio 24:101–109. doi: 10.1016/j.pbi.2015.02.007

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

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

Szadkowski E, Eber F, Huteau V, Lode M, Coriton O, Jenczewski E, Chevre AM (2011) Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. New Phytol 191:884–894. doi: 10.1111/j.1469-8137.2011.03729.x

Tadege M, Sheldon CC, Helliwell CA, Stoutjesdijk P, Dennis ES, Peacock WJ (2001) Control of flowering time by FLC orthologues in *Brassica napus*. Plant J 28:545–553

Tock AJ, Henderson IR (2018) Hotspots for Initiation of Meiotic Recombination. Front Genet 9:521. doi: 10.3389/fgene.2018.00521

Turck F, Coupland G (2011) When vernalization makes sense. Science 331:36–37. doi: 10.1126/science.1200786

Veitia RA, Potier MC (2015) Gene dosage imbalances: action, reaction, and models. Trends Biochem Sci 40:309–317. doi: 10.1016/j.tibs.2015.03.011

Vergara Z, Gutierrez C (2017) Emerging roles of chromatin in the maintenance of genome organization and function in plants. Genome Biol 18:96. doi: 10.1186/s13059-017-1236-9

Vicient CM, Casacuberta JM (2017) Impact of transposable elements on polyploid plant genomes. Ann Bot 120:195–207. doi: 10.1093/aob/mcx078

Victorelli S, Passos JF (2017) Telomeres and Cell Senescence - Size Matters Not. EBioMedicine 21:14–20. doi: 10.1016/j.ebiom.2017.03.027

Vu GTH, Cao HX, Reiss B, Schubert I (2017) Deletion-bias in DNA double-strand break repair differentially contributes to plant genome shrinkage. New Phytol 214:1712–1721. doi: 10.1111/nph.14490

Walbot V, Evans MMS (2003) Unique features of the plant life cycle and their consequences. Nat Rev Genet 4:369–379. doi: 10.1038/nrg1064

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

Wang N, Qian W, Suppanz I, Wei L, Mao B, Long Y, Meng J, Muller AE, Jung C (2011a) Flowering time variation in oilseed rape *(Brassica napus* L.) is associated with allelic variation in the *FRIGIDA* homologue *BnaA.FRI.a*. J Exp Bot 62:5641–5658. doi: 10.1093/jxb/err249

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun J-H, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park B-S, Weisshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IAP, Batley J, Kim J-S, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon S-J, Choi S-R, Lee T-H, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z (2011b) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43:1035–1039. doi: 10.1038/ng.919

Wang J, Hopkins CJ, Hou J, Zou X, Wang C, Long Y, Kurup S, King GJ, Meng J, Ahn JH (2012) Promoter Variation and Transcript Divergence in *Brassicaceae* Lineages of *FLOWERING LOCUS T*. PLoS ONE 7:e47127. doi: 10.1371/journal.pone.0047127

Wang N, Dawe RK (2018) Centromere Size and Its Relationship to Haploid Formation in Plants. Mol Plant 11:398–406. doi: 10.1016/j.molp.2017.12.009

Waterworth WM, Drury GE, Bray CM, West CE (2011) Repairing breaks in the plant genome: the importance of keeping it together. New Phytol 192:805–822. doi: 10.1111/j.1469-8137.2011.03926.x

Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171:737–738. doi: 10.1038/171737a0

Wendel JF (2015) The wondrous cycles of polyploidy in plants. Am J Bot 102:1753–1756. doi: 10.3732/ajb.1500320

Wendel JF, Lisch D, Hu G, Mason AS (2018) The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Curr Opin Genet Dev 49:1–7. doi: 10.1016/j.gde.2018.01.004

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

Wood CC, Robertson M, Tanner G, Peacock WJ, Dennis ES, Helliwell CA (2006) The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. PNAS 103:14631–14636

Wu J, Wei K, Cheng F, Li S, Wang Q, Zhao J, Bonnema G, Wang X (2012) A naturally occurring InDel variation in *BraA.FLC.b* (BrFLC2) associated with flowering time variation in *Brassica rapa*. BMC Plant Biol 12:151. doi: 10.1186/1471-2229-12-151

Wu D, Liang Z, Yan T, Xu Y, Xuan L, Tang J, Zhou G, Lohwasser U, Hua S, Wang H, Chen X, Wang Q, Le Zhu, Maodzeka A, Hussain N, Li Z, Li X, Shamsi IH, Jilani G, Wu L, Zheng H, Zhang G, Chalhoub B, Shen L, Yu H, Jiang L (2018) Whole-genome resequencing of a world-wide collection of rapeseed accessions reveals genetic basis of their ecotype divergence. Mol Plant. doi: 10.1016/j.molp.2018.11.007

Xiao D, Zhao JJ, Hou XL, Basnet RK, Carpio DPD, Zhang NW, Bucher J, Lin K, Cheng F, Wang XW, Bonnema G (2013) The *Brassica rapa FLC* homologue *FLC2* is a key regulator of flowering time, identified through transcriptional co-expression networks. J Exp Bot 64:4503–4516. doi: 10.1093/jxb/ert264

Yang J, Liu D, Wang X, Ji C, Cheng F, Liu B, Hu Z, Chen S, Pental D, Ju Y, Yao P, Li X, Xie K, Zhang J, Wang J, Liu F, Ma W, Shopan J, Zheng H, Mackenzie SA, Zhang M (2016) The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. Nat Genet 48:1225–1232. doi: 10.1038/ng.3657

Yuan Y-X, Wu J, Sun R-F, Zhang X-W, Xu D-H, Bonnema G, Wang X-W (2009) A naturally occurring splicing site mutation in the *Brassica rapa FLC1* gene is associated with variation in flowering time. J Exp Bot 60:1299–1308. doi: 10.1093/jxb/erp010

Zhang X, Meng L, Liu B, Hu Y, Cheng F, Liang J, Aarts MGM, Wang X, Wu J (2015) A transposon insertion in *FLOWERING LOCUS T* is associated with delayed flowering in *Brassica rapa*. Plant Sci 241:211–220. doi: 10.1016/j.plantsci.2015.10.007

Zhao M, Wang Q, Wang Q, Jia P, Zhao Z (2013) Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 14 Suppl 11:S1. doi: 10.1186/1471-2105-14-S11-S1

Ziolkowski PA, Henderson IR (2017) Interconnections between meiotic recombination and sequence polymorphism in plant genomes. New Phytol 213:1022–1029. doi: 10.1111/nph.14265

Żmieńko A, Samelak A, Kozłowski P, Figlerowicz M (2013) Copy number polymorphism in plant genomes. TAAG:1–18. doi: 10.1007/s00122-013-2177-7

Zou X, Suppanz I, Raman H, Hou J, Wang J, Long Y, Jung C, Meng J, Wu K (2012) Comparative Analysis of *FLC* Homologues in *Brassicaceae* Provides Insight into Their Role in the Evolution of Oilseed Rape. PLoS ONE 7:e45751. doi: 10.1371/journal.pone.0045751