Check for updates

# Chromosome-Scale Assembly of Winter Oilseed Rape *Brassica napus*

HueyTyng Lee[1], Harmeet Singh Chawla[1], Christian Obermeier[1], Felix Dreyer[2], Amine Abbadi[2] and Rod Snowdon[1]*

[1] Department of Plant Breeding, Justus Liebig University Giessen, Giessen, Germany, [2] NPZ Innovation GmbH, Holtsee, Germany

Rapeseed (*Brassica napus*), the second most important oilseed crop globally, originated from an interspecific hybridization between *B. rapa* and *B. oleracea*. After this genome collision, *B. napus* underwent extensive genome restructuring, via homoeologous chromosome exchanges, resulting in widespread segmental deletions and duplications. Illicit pairing among genetically similar homoeologous chromosomes during meiosis is common in recent allopolyploids like *B. napus*, and post-polyploidization restructuring compounds the difficulties of assembling a complex polyploid plant genome. Specifically, genomic rearrangements between highly similar chromosomes are challenging to detect due to the limitation of sequencing read length and ambiguous alignment of reads. Recent advances in long read sequencing technologies provide promising new opportunities to unravel the genome complexities of *B. napus* by encompassing breakpoints of genomic rearrangements with high specificity. Moreover, recent evidence revealed ongoing genomic exchanges in natural *B. napus*, highlighting the need for multiple reference genomes to capture structural variants between accessions. Here we report the first long-read genome assembly of a winter *B. napus* cultivar. We sequenced the German winter oilseed rape accession 'Express 617' using 54.5x of long reads. Short reads, linked reads, optical map data and high-density genetic maps were used to further correct and scaffold the assembly to form pseudochromosomes. The assembled Express 617 genome provides another valuable resource for *Brassica* genomics in understanding the genetic consequences of polyploidization, crop domestication, and breeding of recently-formed crop species.

Keywords: winter oilseed rape, genome assembly, long reads, *Brassica napus*, crop genomics

## INTRODUCTION

*Brassica napus* subsp. *oleifera*, commonly known as rapeseed or canola, is the second most important oilseed crop globally (Food and Agriculture Organization of the United Nations, 2019). It originated from a natural hybridization event between *B. rapa* (AA, 2n = 2x = 20) and *B. oleracea* (CC, 2n = 2x = 18) no more than 7.5 1000 years ago (Chalhoub et al., 2014). Rapeseed was already widely cultivated in Europe from the 15th to 18th centuries for lamp fuel and soap production (Appelqvist and Ohlson, 1972). Following the introduction of double-low varieties (with low erucic acid and low glucosinolates in the seed) in the 1970s, modern rapeseed/canola varieties today deliver a high-value vegetable oil which can also be used for biodiesel production, while

the extraction meal provides a high quality, protein-rich animal feed (Friedt and Snowdon, 2009). Oilseed rape is also a major component of crop rotations in most cereal-dominated agricultural systems (Friedt et al., 2018).

*Brassica napus* has an allotetraploid genome composition (AACC, 2n = 4x = 38) (Nagaharu, 1935; Allender and King, 2010). The formation of an allotetraploid involves the challenge to combine subgenomes of distinct species, with individual evolutionary history and epigenetic patterns, into one (Osborn et al., 2003; Comai, 2005). Studies of synthetic allopolyploids and natural neo-allopolyploids showed that the hybridization process of divergent genomes causes instant and prolonged alteration of gene expression, DNA methylation patterns and transposable elements regulation (Salmon et al., 2005; Lukens et al., 2006; Buggs et al., 2010; Chelaifa et al., 2010; Coate et al., 2014; Rigal et al., 2016; Edger et al., 2017). The genomic sequences of allopolyploids are also restructured as a result of illicit pairing of non-homologous chromosomes during meiosis, which encourages homoeologous exchange (HE) events (Gaeta et al., 2007; Xiong et al., 2011). HEs result in the replacement of chromosomal segments of one subgenome with another, and is hypothesized to lead to genome diploidization through fixation of HEs with time (Lysak et al., 2007; Mandáková et al., 2010). In *B. napus*, HEs were revealed through extensive structural rearrangements when the genome of a natural line was compared to the ancestral progenitors (Chalhoub et al., 2014). When compared among seven diverse *B. napus* genotypes, both shared and specific HEs up to a few 100 kb in size were found (Chalhoub et al., 2014), suggesting that HE is an ongoing process in *B. napus*. The mixture of older, fixed HEs and newly-formed HEs explains the wide-spread variations, such as reciprocal (Lombard and Delourme, 2001; Osborn et al., 2003; Piquemal et al., 2005) and non-reciprocal (Udall et al., 2005) translocations, between genotypes. These genotype-specific HEs have been shown to give rise to novel genetic diversities related to important agronomic traits such as flowering time (Pires et al., 2004; Chalhoub et al., 2014; Schiessl et al., 2017), leaf morphology (Osborn et al., 2003; Gaeta et al., 2007), and seed content (Harper et al., 2012; Qian et al., 2016).

The motivation of producing a highly-contiguous *B. napus* genome is clear, particularly from the aspects of breeding research. Genomes of high contiguity enable accurate design of SNP markers to obtain uniquely-mapped probes for marker-assisted selection. A direct comparison of GWAS and genomic selection results between highly-fragmented and chromosomal-scale assemblies of the blueberry genome shows better predictive ability and narrowing of QTL regions (Benevenuto et al., 2019). Similarly, high-quality genomes of the bread wheat (The International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018) and its progenitors (Avni et al., 2017; Luo et al., 2017; Zhao et al., 2017; Ling et al., 2018) have enabled novel gene-to-trait discoveries such as dissection of shattering (Avni et al., 2017) and powdery mildew resistance (Ling et al., 2018). A complete genome assembly also helps fine-tune various decisions in breeding programs, such as target positions for genomic introgressions and identification of potential targets for genome editing CRISPR-Cas9 technologies (Gao, 2018). The genomic

characteristics of *B. napus* increase the complexity of studying the genome sequences in three major respects. Firstly, homoeologous regions between subgenomes hamper the genome assembly process due to low sequence specificity. Ambiguities of highly similar sequences are difficult to resolve for assembly algorithms, particularly during the read clustering process (Nagarajan and Pop, 2013). In *B. napus*, reads originating from homoeologous regions cannot always be accurately assigned to individual subgenomes, and subgenomic distinction is further blurred by ongoing HE events. Recent assemblers adopt the *k*-mer binning method to resolve haplotypes using parental genome assemblies (Koren et al., 2018). Since high-quality assemblies of the *B. napus* progenitors are available (Belser et al., 2018), this approach could be plausible however it is nevertheless unable to fully resolve HE events, which interfere with subgenomic separation of *k*-mers. Secondly, as in many other complex crop genome assemblies, the high content of repetitive sequences in *B. napus* interfere with the construction of continuous chromosomes. The two diploid progenitors *B. rapa* and *B. oleracea* are both products of multiple paleopolyploidization events, where large-scale rearrangements occur following divergence from a common ancestor (Parkin et al., 2005). As a result, *B. napus* has potentially accumulated up to 72x multiplication since the origins of angiosperms and about 34.8% of the genome are estimated to be repeats (Chalhoub et al., 2014). Thirdly, the genome assembly of any single cultivar always fails to capture the entire genomic repertoire in a species, hence the need to use a pangenome as reference is recognized in crops (Tao et al., 2019). In oilseed rape, this need is highlighted by the HE-driven variations found between cultivars. Recent evidence shows ongoing HEs even in homozygous cultivars during self-pollination (He et al., 2017), suggesting that the variations between individuals of the same cultivar could be largely underestimated.

Genome assemblies for three cultivars of *B. napus* have been published to date (Chalhoub et al., 2014; Bayer et al., 2017; Sun et al., 2017), with Darmor-*bzh* and Tapidor being the two winter-type genotype represented. The Darmor-*bzh* genome is widely used as a standard reference genome from studies ranging from gene loss (e.g., Hurgobin et al., 2018) to SNP marker-assisted analyses like genome wide association studies (GWAS) (e.g., Gabur et al., 2018). However, all three genome assemblies were constructed prior to the advance of long-read technologies, therefore these assemblies are highly fragmented. To illustrate, the N50 read length of an Oxford Nanxopore MinION single flowcell run today is about 32 kbp (**Supplementary Table S2**), approximately the same size as the N50 contig length in the Darmor-*bzh* assembly (Chalhoub et al., 2014). The long-read technologies, led today by Pacific Biosciences (Eid et al., 2009) and Oxford Nanopore Technologies (Loman et al., 2015), revolutionized genomic research by producing continuous sequences of 10s to 100s of kilobases in length. They are now used to resolve complex and repetitive regions in plant genomes (for example Schmidt et al., 2017; Belser et al., 2018). Long-reads are therefore well-suited to resolve the aforementioned complications in assembling the *B. napus* genome by encompassing HE breakpoints and transposable elements.

Here we report the sequencing and genome assembly of the German winter oilseed rape accession 'Express 617' using 54.5x coverage with Pacific Biosciences long reads. Express 617 is a natural winter oilseed rape accession widely used in many existing mapping populations for linkage analyses of traits such as seed quality (Badani et al., 2006; Stein et al., 2013, 2017), seed yield and yield architecture (Radoev et al., 2008), heterosis (Basunanda et al., 2010) and disease resistance (Obermeier et al., 2013). Short reads, optical map data and genetic maps were used to further correct and scaffold the assembly to form pseudochromosomes. The Express 617 genome is assembled to 925 Mb in size, approximate to the flow cytometry estimation of 1132 Mb (Johnston et al., 2005). The base accuracy and pseudomolecule contiguity were validated using short read libraries, SNP markers and long read alignments. The genome was annotated to contain 12.5% of coding sequences (89857 predicted genes) and 37.5% of repetitive elements. The assembly was also compared to two other published *B. napus* genomes to identify collinear regions. A total of 56 same-chromosome collinear blocks of 488 Mb in size were identified in Express 617 (53%) when compared to the Darmor-*bzh* genome. In comparison, only 230 Mb (25%) of Express 617 are collinear with the ZS11 genome. This long-read genome of *B. napus* is expected to contribute to further understanding of HE in *B. napus* and its role in generation of genetic diversity for quantitative trait expression (Gabur et al., 2019). This assembly expands the genomic repository of *B. napus*, particularly for winter-type accessions, and consequently promotes exploitation of genomics advancement in oilseed rape and canola breeding programs.

## RESULTS

### The Express 617 Genome Assembly, Gene Set, and Repetitive Elements

The total size of Express 617 genome assembly is 925 Mb, where placed pseudochromosomes are 765 and 160 Mb remained as unplaced random scaffolds (**Supplementary Table S3**). As shown in **Table 1**, this genome size is larger than three previously published assemblies, whereas the percentage of N-bases

**TABLE 2 |** Gene annotation and evaluation of the Express 617 genome.

|  | Express 617 |
| --- | --- |
| Number of genes | 89857 |
| Number of transcripts | 99481 |
| coreGF | Weighted score: 0.95 |
|  | Number of missing coreGFs: 159 |
| Number of BUSCOs found | Total complete: 4358 (94.8%) |
|  | Complete single copy: 866 |
|  | Complete duplicated: 3492 |
|  | Fragmented: 11 |
|  | Missing: 227 |
| Number of transcripts aligned to pan-transcriptome | 87012 |
| Number of proteins containing InterPro domains | 87951 |

(quantity of gaps) is lowest among all five assemblies. The high contiguity of Express 617 is also reflected in the length of N50 scaffolds (4.8 Mb) prior to pseudochromosome construction.

The genome consists of 12.5% coding sequences, 89857 genes with 99481 transcripts (**Table 2**) and 37.5% repetitive elements (**Supplementary Table S4**). The transcripts have an average length of 1924.8 bp, with an average of 5.22 exons each. Average lengths of intron and exon are 183.3 and 226.5 bp, respectively. A total of 87951 transcripts contain at least one known protein domain that can be found in curated protein databases. As observed in all other plant genomes, the majority of the Express 617 repeats are long terminal repeats (LTRs) (28.3% of all repetitive bases masked), where 22.2% are Gypsy and 16.8% Copia retrotransposons. The non-LTR subclass I and subclass II comprise 4.9 and 13.5%, respectively, while the remaining transposable elements remain uncharacterized (25.5%). Satellites, simple repeats and low complexity sequences make up another 5.2% of all repeat sequences.

Consistent with previous studies (Chalhoub et al., 2014), the chromosomes of subgenome A have higher gene density with lesser repetitive elements when compared to subgenome C (**Supplementary Figure S1**). This is explained by subgenomic dominance, a phenomenon documented in many polyploids such

**TABLE 1 |** Assembly statistics of the Express 617 genome in comparison with three previously published *B. napus* genome assemblies.

|  | Darmor-*bzh* v4.1 (Chalhoub et al., 2014) | Darmor-*bzh* v8.1 (Bayer et al., 2017) | ZS11 (Sun et al., 2017) | Tapidor (Bayer et al., 2017) | Express 617 |
| --- | --- | --- | --- | --- | --- |
| Total genome size (Mb)/percentage of Ns | 850/13.17% | 850/13.16% | 976/7.05% | 636/5.16% | 925/0.09% |
| Length of pseudochromosomes (Mb) | 645 | 798 | 854 | 627 | 765 |
| Length of unplaced scaffolds (Mb) | 204 | 51 | 120 | 8.4 | 160 |
| Number of scaffolds prior to pseudochromosome construction | 20702 | - | 3460 | 21280 | 1632 |
| Length of N50 scaffold prior to pseudochromosome construction (bases) | 763688 | - | 602220 | 197031 | 4882293 |

as cotton (Renny-Byfield et al., 2015) and maize (Schnable et al., 2011), where homoeologous copies of "dispensable" genes are preferentially silenced (Edger et al., 2018).

## Evaluation of the Assembly Quality

We took multiple steps to avoid common errors in the assembly process, and then extensively evaluated the results. The correctness of the assembly was evaluated in three ways, (1) base-level errors, (2) large-scale translocations, and (3) completeness of the gene set.

Base-level errors are single nucleotide mutations and short indels that usually arise from the sequencing process. The error rate of raw PacBio long reads was estimated to be up to 15% (Korlach, 2013). Using the alignment of two libraries of Illumina short reads, we assessed the error rate of a subset of PacBio reads (10% of total nucleotides in all reads). By allowing single-end mapping, 7% of the total nucleotides of mapped reads were mismatches, which is half of the maximum estimated error assuming that Illumina reads have near-to-zero sequencing errors (Glenn, 2011). To reduce the effect of long read sequencing errors, we used consensus long reads that were generated by self-alignment. We also incorporated high coverage of short reads during the assembly, as well as post-assembly error correction. To measure the base-level accuracy of the genome, five libraries of Illumina sequencing reads, of which four were used to construct the assembly and one was sequenced independently, were used. A total of 89% of paired-end reads aligned concordantly in the correct direction and insert size, with zero mismatches and gaps (**Supplementary Table S5**).

Large structural error is a primary concern when assembling polyploidy genomes, particularly allopolyploids with frequent HEs like *B. napus* (Samans et al., 2017). These errors could manifest through a few assembly processes, for example (1) wrongly-placed scaffolds during the construction of pseudochromosomes due to non-specific matching to genetic maps, (2) short mate-paired libraries and linked-reads could be unspecific to differentiate between regions of homoeologous chromosomes, and form wrongly-joined scaffolds, and (3) regions with high density of repetitive elements may form small scaffolds and could be wrongly-placed as described in (1). To evaluate large-scale errors, a combination of SNP markers and long read alignments was used. First, the distribution pattern of gene allelic SNPs in the AC *Brassica* genomics platform (He et al., 2015) generated by the genome-ordered graphical genotypes (GOGGs) method (He and Bancroft, 2018) was manually inspected. The correctness of the assembly was measured by low amount of alternating parental alleles in individual recombinant lines, with the assumption that allelic SNPs segregate across a mapping population while interhomeolog and interparalog SNPs do not. We detect a total of 24 regions of discording allelic patterns indicating putatively incorrect gene order, which could originate from incorrectly placed scaffolds or misjoin of scaffolds (example in **Supplementary Figure S2**). They were labeled as potentially misassembled. To confirm that they were indeed true misjoins and not inaccuracies introduced during GOGG such as ambiguously-mapped orthologs, we used alignments of long reads to the assembly. Long read alignments were generated using PacBio reads which were used to construct the assembly and additionally 17x of Nanopore ONT reads. A total of 86% (562142) of the Nanopore reads and 99% (4328786) of PacBio reads aligned to the assembly. True misjoins were identified by refining the resolution of breakpoints, which are characterized as a huge decrease of mapped reads and an enrichment of split-reads, as seen in **Supplementary Figure S3**. When supported by high coverage of mapped reads for both PacBio and Nanopore technologies, a putatively misassembled region was dismissed, as shown in **Supplementary Figure S4**. Using read coverage as supporting evidence, a total of seven regions, ranging from 123 kb to 3 Mb were identified as misassemblies. All cases of true errors have one or both breakpoints in stretches of Ns. Ns were introduced as gap-fillers during the construction of pseudochromosomes and scaffolding of using linked-reads. Regions with high frequency of Ns therefore symbolizes difficult regions where their local sequence proximity, termed "edges" in an assembly graph (Wick et al., 2015), cannot be resolved. These regions were extracted and retained as unplaced scaffolds. **Supplementary Figure S5** shows the final arrangements of scaffolds based on the genetic versus physical distance of a total of 24469 markers (17478 in Express 617 × R53; 8469 in Express 617 × V8; 12140 in Express 617 × SGDH14) in each pseudochromosome. The pseudochromosomes were also compared to the progenitors *B. rapa* and *B. oleracea* genomes to show that sequence similarities of subgenomes are as expected (**Supplementary Figure S6**).

The completeness of predicted genes were evaluated with a set of well-conserved genes across plant species using PLAZA coreGF (van Bel et al., 2012) and BUSCO (Simão et al., 2015). Out of 2928 core green plants gene families in PLAZA, 2803 were identified in the predicted gene set, therefore obtaining a weighted score of 0.95. BUSCO (v4.0.4) detected 4358 (94.8%) out of 4596 complete orthologous groups within the Brassicales lineage dataset, with 3492 being duplicated and 866 being single copy. In comparison (**Supplementary Figure S7**), version 4.1 of Darmor-*bzh* has 4378 (95.2%) complete BUSCOs, version 8.1 of Darmor-*bzh* has 4379 (95.2%), ZS11 has 4263 (92.7%) and Tapidor has 4162 (90.6%). Additionally, a publicly-available single-ended RNAseq library was used to ballpark the accuracy of annotated introns. A total of 342674 introns were predicted using RNAseq, where 229278 (67%) matched to the introns annotated, 104883 overlapped with predicted gene region, and 8513 were in intergenic regions. This indicates that 62% of introns annotated are supported by external data, which was not used in the annotation pipeline.

## Comparison Between Express 617 and Other *B. napus* Genomes

Whole genome alignment between Express 617 and Darmor-*bzh* shows high sequence similarity in all chromosomes (**Figure 1**). The secondary alignments of lower similarity between homoeologous chromosomes can also be observed.

To examine the shuffling of chromosomal segments, the gene-level collinearity between genomes, which is defined as the conservation of gene order within syntenic regions

**FIGURE 1 |** Dot plot comparison between Express 617 and Darmor-*bzh* v4.1 genome. Sequence similarity is color coded from 0 to 1.
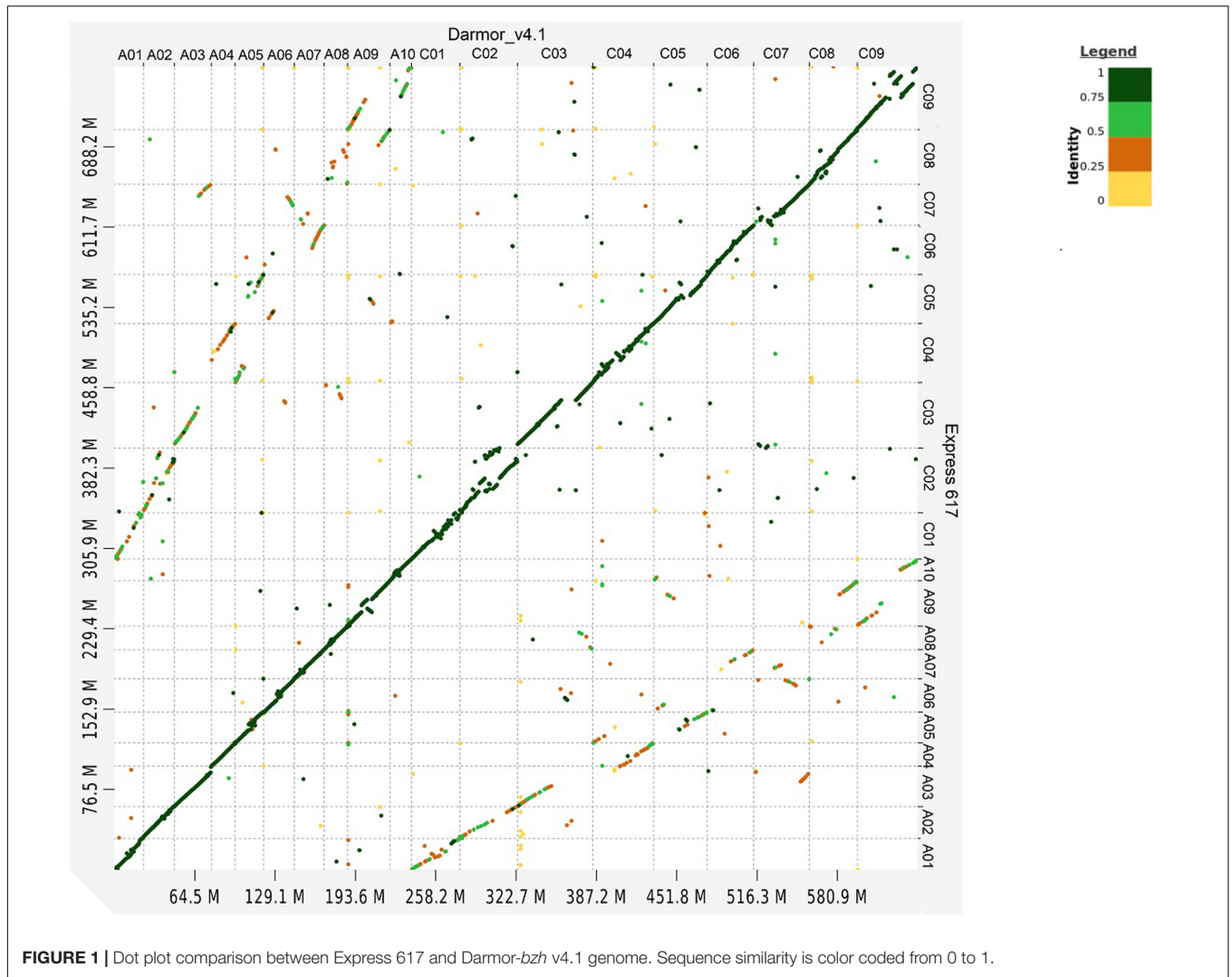
(Coghlan et al., 2005), was identified. A total of 120 collinear blocks, linked by 77840 gene pairs, were found between Express 617 and Darmor-*bzh* (**Figure 2A**). Out of 120, 56 blocks linked by 45410 gene pairs correspond to the same chromosomes. These 56 blocks made up 488 Mb (53%) of the Express 617 genome and 425 Mb (57%) of the Darmor-*bzh* genome. In comparison, Express 617 shared 100 collinear blocks (38145 gene pairs) with ZS11 (**Figure 2B**), a spring oilseed rape line, where 56 of the blocks (21982 gene pairs; 230 Mb of Express 617 and 274 Mb in ZS11) are in the same chromosomes.

## DISCUSSION

Improved contiguity of Express 617 genome in comparison to other *B. napus* assemblies is evident in the low number of scaffolds, high N50 scaffold length and the low percentage of Ns in total genome size (**Table 1**). This improvement is expected as the all other four are short read assemblies. The Tapidor assembly has the lowest contiguity as it was assembled with

about 30x of Illumina short reads and the contigs were placed using SNPs (Bayer et al., 2017). ZS11 and Darmor-*bzh* were constructed with more comprehensive data, including higher coverage of short read sequencing, long range mate-paired reads and BAC-by-BAC approach (Sanger-sequenced for Darmor-*bzh* and Illumina-sequenced for ZS11) (Chalhoub et al., 2014; Sun et al., 2017). For both assemblies the general approach was that first BAC sequences were used to form contigs, gaps were filled with short reads, and then genetic maps were used to place contigs. Since the maximum read length was 100 bp, the assemblies produced are enriched with gaps, as reflected on the percentage of Ns. Better contiguity also means that the intergenic and repeat-rich regions such as centromere are better assembled. This will enhance the development in molecular breeding such as application of transposable element markers (Bhat et al., 2020), following increasing understanding of the role of transposable elements in crops, such as in disease resistance [example in pepper (Kim et al., 2017)], domestication [example in rice (Li et al., 2017)], and adaptations [example in maize (Lai et al., 2017)].
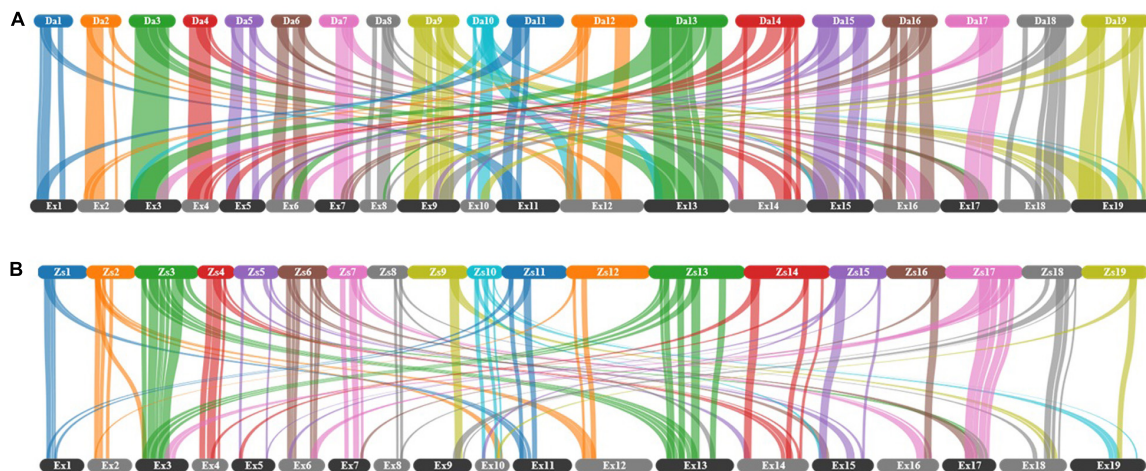
**FIGURE 2 |** Collinearity between Express 617 and two other *B. napus* assemblies for all chromosomes. **(A)** Darmor-*bzh* v4.1 versus Express 617, **(B)** ZS11 versus Express 617. Collinear blocks are indicated as connecting bars between genomes. The chromosomes were labeled with two letter indicating the cultivar followed by chromosome number, where 1 to 10 corresponds to chromosomes A1 to A10 and 11 to 19 corresponds to chromosomes C1 to C9.

A considerable amount of scaffolds (160 Mb) cannot be placed in the Express 617 pseudochromosomes. Due to the subgenomic similarities and frequency of HEs, a relatively conservative approach was taken to construct this assembly. To avoid false positives and wrong conclusions led by misassemblies, the assembled scaffolds were broken in two independent steps, therefore trading off contiguity for accuracy. Misjoins were broken first during the incorporation of optical map data, and second during GOGG evaluation. Optical maps provide independent long-range evidence for the connection of scaffolds. However, in the circumstances of conflicts between optical map and the assembled sequences, a decision has to be made to resolve conflicts. Since optical maps are not error-free (Jiao et al., 2017), the software *Chimericognizer* (Pan and Lonardi, 2019) used alignments adjacent to the conflicts to estimate the confidence of chimeric sites. A total of 92 scaffolds (out of 1547) in the assembly were identified as chimeric and broken to form 206 scaffolds. Stitching of all the scaffolds was then attempted next using *Novo&Stitch* (Pan et al., 2018). Similarly, using the GOGG approach followed by long read mapping, a total of seven regions were identified as misjoins and breakpoints were cut. Even though the correct chromosome and position of these misassembled blocks can be identified using GOGG patterns, they are of low resolution. In other words, there is no way to accurately determine a breakpoint for insertion of these blocks. These blocks were therefore retained as random scaffolds, with the putative chromosome appended to the scaffold name (**Supplementary Table S3**). Out of all unplaced scaffolds, 38 Mb were assigned to chromosomes and contain 2946 genes, whereas 122 Mb contain 2803 genes, with 34.3% of them being repetitive elements.

The completeness of gene space is one of the ways to evaluate an assembly (Veeckman et al., 2016). Coding sequences made up of 12.5% in Express 617, comparable to 11.9% in Darmor-*bzh* v4.1 (Chalhoub et al., 2014). Both coreGF and BUSCO indicate a 95% completeness of conserved orthologous groups in Express 617 genome. Based only on BUSCO results, this is comparable to Darmor-*bzh*, where 20 more orthologous groups were identified, and more superior to Tapidor and ZS11 (**Supplementary Figure S7**). We postulated that the error rate of long read sequencing affect the accuracy of gene prediction, as observed in human genome assemblies (Watson and Warr, 2019). This could possibly also be reflected on the lower number of confident genes in Express 617 when compared to Darmor-*bzh*. Using Illumina short reads as a benchmark, PacBio raw reads have an overall mapping rate of 71% and an error rate of 7%, whereas the assembly has an overall mapping rate of 99% (perfect mapping rate of 89%) and an error rate of 0.5%. This improvement is largely contributed by the pre-assembly consensus read construction and multiple rounds of short read polishing. We anticipate better polishing softwares, such as sequencing signal-based tools (such as Nanopolish[1]) to resolve the 0.5% uncorrected errors. Nevertheless, we argue that from the perspective of the amount of resources and time used, long read technology has definitely increased the efficiency to produce high quality genomes. For example, the BAC-by-BAC pooled strategy used in Darmor-*bzh* is known to be highly accurate yet expensive and include labor-intensive processes such as fingerprinting clones.

Another possibility for undetected errors to persist in this assembly is that when correlating genetic maps to physical positions, two assumptions were used (1) the genetic maps accurately represent the Express 617 genome, and (2) each marker probe mapped correctly to the chromosomal position of origin. However, even though Express 617 is the common parent of three populations used to generate genetic maps used, there are two populations with parents of synthetic backgrounds (R53 and

---

[1]https://github.com/jts/nanopolish

V8). Synthetic accessions are known to contain more HEs than non-synthetics (Sharpe et al., 1995; Liu et al., 2014; Rousseau-Gueutin et al., 2017; Hurgobin et al., 2018). For example, a large part of chromosome C02 in R53 was known to be replaced by A02 (Stein et al., 2017). To reduce the manifestations of these HEs in Express 617 pseudochromosomes, weighted priority was given to population of natural lines (Express 617 × SGDH14). The limitation of the second assumption is the specificity of probe mapping. As the length of marker probe is only 50 bp, it could map to multiple positions (8227 out of 44113 of mapped probes are non-unique). Even with uniquely mapped probes, scaffolds could still be wrongly assigned to homeologous or similar regions of non-homeologous chromosomes. Homeologous mappings can be observed in **Supplementary Figure S5**, particularly between A01 and C01, A03 and C03, A09 and C08, and A09 and C08, which are known hotspots for HE events (Chalhoub et al., 2014; Lloyd et al., 2018). Also, the density of markers are not consistent along the chromosomes. To illustrate, the first misjoin in chromosome A01 (position 1846993) detected by GOGG only have adjacent markers at a 58 kbp distance upstream and 139 kb downstream. Since only uniquely-mapped markers were used, repetitive or highly homoeologous regions contribute to large gaps between markers. This potentially explains how the misjoin was formed during the assembly, and how it was not detected with 10x linked reads and optical mapping.

*Brassica napus* morphotypes cluster into winter, semi-winter and spring growth habits based on SNPs (Diers and Osborn, 1994; Becker et al., 1995; Bus et al., 2011; Gazave et al., 2016; Delourme et al., 2018) and show sequence and copy number variation in flowering time regulatory genes (Schiessl et al., 2017). Collinearity comparisons between the two winter-type cultivars Express 617 and Darmor-*bzh,* and between Express 617 and the Chinese semi-winter cultivar ZS11 reflected this expectation of genetic diversity. However, we nevertheless cannot disregard the influence of assembly quality and completeness in collinearity studies. Regions that are not collinear could arise from true genetic diversity, assembly artifacts such as misassemblies and gap regions, or unidentified genes. Repetitive elements in the genome are likely to be the major contributor of these regions. For example, repeat-masking approach was found to be the main cause of varying number of repeat-containing disease resistance genes in four *B. napus* genomes, instead of true biological variations (Bayer et al., 2018).

# MATERIALS AND METHODS

## Whole Genome Sequencing of Plant Material

### Illumina and Pacific Biotechnologies Sequencing
Approximately 40 g of fresh leaf tissue was collected from an advanced inbred line ($>F_{11}$) of the winter type oilseed rape accession "Express 617." DNA libraries of 350 bp, 450 bp, 2 kbp, 5 kbp, and 10 kbp were constructed and subjected to paired-end sequencing on the Illumina HiSeq 2000 platform. The 20-kb SMRTbell library was prepared using SMRTbell Template

Prep Kit 1.0-SPv3, where the qualified high-molecular weight DNA were fragmented to approximately 20 kb, followed by damage repair, end repair and adapter ligation. Size selection was then performed using BluePippin$^{TM}$ Size-Selection System (Sage Science, Beverly, MA, United States). The quality of purified library was checked using Qubit (Invitrogen) and Advanced Analytical Fragment Analyzer (AATI). The SMRTbell-Polymerase Complex was prepared using Sequel$^{TM}$ Binding Kit 2.0 and sequenced on Sequel SMRT Cell 1M v2. A 6 h movie using the Sequel Sequencing kit 2.0. 10x Genomics libraries also constructed and sequenced on the Illumina platform to produce GemCode linked reads. All sequencing described above was outsourced to Novogene, Co., Ltd. (China). Raw reads obtained were deposited to the NCBI Short Read Archive (PRJNA587046). Sequencing depths obtained for each library are recorded in **Supplementary Table S1**.

### DNA Isolation for Oxford Nanopore Sequencing
The DNA isolation was carried out in accordance to high molecular weight DNA isolation protocol as described by Mayjonade et al. (2016). Approximately, 5 g of fresh leaves were harvested from rapeseed plants at 4–6 leaf stage. This frozen leaf was immediately frozen using liquid nitrogen. This frozen leaf was subjected to mechanical grinding using a mortar and pestle. 4–5 ml of pre-heated lysis buffer [1% (w/v) PVP10, 50 nM EDTA, 1.25% (w/v) SDS, 1% (w/v) Na$_2$S$_2$O$_5$, 5 mM C$_4$H$_{10}$O$_2$S$_2$, 100 mM TRIS pH 8, 500 nM NaCl, 1% (v/v) Triton X-100, 1% (w/v) PVP40] to the frozen leaf samples for disrupting the cell wall. This was followed by incubation of the lysate for 30 min at 37°C. In order to precipitates sodium dodecyl sulfate (SDS) and SDS-bound proteins, 0.3 volumes of 5 M Potassium Acetate was added to the lysate and spun at 8000 $g$ for 12 min at 4°C. Clean DNA was then recovered by fishing out the DNA using magnetic beads.

### Size Selection and Library Preparation for Oxford Nanopore Sequencing
1–3 μg of DNA was subjected to size selection using Circulomics short-read eliminator XL kit (Circulomics, Inc., Baltimore, MD, United States) according to the manufacturer's instruction. The kit uses selective precipitation to deplete DNA fragments shorter than 40 kb. The size selected DNA was then used for the preparation of the sequencing library, using SQK-LSK109 (Oxford Nanopore Technologies) kit in accordance with the manufacturer's recommendations. Following the library preparation, DNA was finally loaded onto an Oxford Nanopore MinION flow cell (version R9.4.1) for sequencing. The raw fast5 files produced by the MinION device were then base-called using *Guppy* 3.0.3 (Oxford Nanopore Technologies) with "dna_r9.4.1_450bps_hac.cfg" model using standard parameters to generate fastq file. **Supplementary Table S2** shows the statistics of reads generated.

## Optical Map Construction
DNA isolation for optical mapping was performed according to the IrysPrep$^{TM}$ Plant Tissue-Nuclei protocol provided by BioNano Genomics. Nearly 2 grams of young leaves were harvested from dark-treated rapeseed plants, immediately

followed by fixing with 2% formaldehyde. In order to isolate the intact nuclei, fixed leaf material was subjected to homogenization in an isolation buffer containing BME, Triton X-100 and PVP-10. Purified nuclei were then embedded into an agarose gel matrix. Finally, the DNA was recovered by melting the agarose plugs using GELase[TM] (Epicentre) treatment. Sequence specific nick labeling using Nt.BspQI (recognition site GCTCTTC) was performed on the isolated DNA using the IrysPrep[TM] Labeling-NLRS protocol by BioNano Genomics. Finally these single DNA molecules were loaded onto an IrysChip for imaging on the BioNano Genomics Irys platform. The DNA molecules were imaged using the BioNano Irys System and were computationally translated into single-molecule optical maps. Single optical maps were then assembled into a consensus map with IrysSolve pipeline (v5134) provided by BioNano Genomics, and deposited as **supplementary file** in NCBI BioProject PRJNA587046.

## Genetic Maps Construction

Genetic maps were constructed for the two biparental populations Express 617 × R53 (ExR53-DH) and Express 617 × V8 (ExV8-DH) using 60K Illumina Infinium Brassica SNP array, SSR and AFLP marker data obtained from 244 and 216 lines, respectively. SNP and SNaP marker data were filtered according to Gabur et al. (2018). SSR and AFLP marker data were taken from the genetic maps produced by Radoev et al. (2008) and Basunanda et al. (2010). Genetic maps were constructed using the software *MSTMap* (Wu et al., 2008) applying the Kosambi map function. The genetic linkage map for Express 617 × SGDH14 was produced using 60K Illumina Infinium Brassica SNP array marker data obtained from 139 lines using the software package *JoinMap* 4.1 (Stam, 1993; van Ooijen, 2011) applying the Kosambi map function (Behnke et al., 2018).

## Genome Assembly

To increase read accuracy, PacBio raw reads were self-aligned to generate consensus reads using *Daligner* v1.0 (Myers, 2014) using default parameters. Consensus reads were then assembled using *FALCON* (falcon-2017.11.02-16.04-py2.7) (Chin et al., 2016) to form unitigs. Unitigs were then further polished using the consensus algorithm Quiver (SMRT Link v5.0.1)[2]. Illumina short reads were used to correct small-scale errors using default parameters of *Pilon* (pilon-1.18.jar) (Walker et al., 2014). To increase contiguity, PacBio reads were used to further scaffold the unitigs (SSPACE-standard) (Boetzer et al., 2011). 10x Genomics data were first processed by trimming the first 16 bp barcode and subsequence 7 bp random primer sequence of the first mate of each pair, and then aligned to the scaffolds to form super-scaffolds using *fragScaff* (version 140324) (Adey et al., 2014). Assembly procedures described above were performed by Novogene, Co., Ltd.

The obtained super-scaffolds were corrected for large-scale chimeric regions originating from misassembly by comparing to an Express 617 BioNano optical map using *Chimericognizer* (Pan and Lonardi, 2019) where junctions of chimeric scaffolds were broken with the following parameters "-a 1.5 -b 1 -d 25 -e

---

[2]https://github.com/PacificBiosciences/GenomicConsensus

50000 -h 50000 -r 80000". Scaffolds were then corrected using *Novo&Stitch* (Pan et al., 2018) with the default strict parameters that are equivalent to "-a 3000 -b 0.1 -c 10000 -d 0.5 -e 0.9 -h 25 -r 0.2".

The corrected scaffolds were then arranged into pseudochromosmes using three high-resolution SNP-based genetic maps, including Express 617 × SGDH14, Express 617 × V8 and Express 617 × R53. Weighted priority 3, 2 and 1 were given to the listed maps respectively based on the synthetic origin of parents. Pseudochromosome construction was completed using the software *ALLMAPS* (Tang et al., 2015) with the following parameters "python –m jcvi.assembly.allmaps path –mincount = 10 –links = 25". Scaffolds that map to multiple linkage groups were identified as potentially chimeric, and the breakpoints were detected using "python –m jcvi.assembly.allmaps split" and "python –m jcvi.assembly.patch refine". Corrected pseudochoromosomes were produced by repeating the *ALLMAPS* run with the broken scaffolds.

## Gene Annotation

Repetitive sequences were identified using *RepeatModeler* (Smit and Hubley, 2008), a repeat family identification and modeling package which performs two *de novo* repeat-finding programs *RECON* (Bao, 2002) and *RepeatScout* (Price et al., 2005). The repeats identified were soft-masked in the assembly using *RepeatMasker* vopen-4.0.7 (Smit et al., 2013).

The gene prediction pipeline *BRAKER2* (Hoff et al., 2019) was used to train an Express 617-specific gene model and then predict genic regions. *BRAKER2* executes the gene predictor *Augustus* (Hoff and Stanke, 2013, 2018) internally. First, the proteomes of two species *Arabidopsis thaliana* (Proteome ID UP000006548) and *B. napus* (Proteome ID UP000028999) were aligned to the genome using *Genome Threader* (Gremme et al., 2005), and provided as evidence for model training in *Augustus*. The trained parameters were then used, together protein homology hints, to accurately predict genes. Fragmented predictions and potential pseudogenes were further filtered with (1) *Augustus* script "python Augustus/scripts/getAnnoFastaFromJoingenes.py –s TRUE" and (2) high identity to all 341468 proteins in the *Brassica* genus (Taxon identifier 3705) in UniProt release 2019_08 (The UniProt Consortium, 2019), with an alignment coverage of 80% to both target and query, a percentage identity of 80% and above, and -evalue 10e-5 using *BLASTP* (Camacho et al., 2009).

## Evaluation of Base-Level Accuracy

To estimate the error rate of PacBio reads, two Illumina sequencing libraries (SRR10382360 and SRR10382369) were aligned to the assembly using *Bowtie2* version 2.2.6 (Langmead and Salzberg, 2012) with the following parameters "bowtie2 -I 200 -X 500 –end-to-end –no-discordant". Error rate was estimated by calculating the ratio of mismatch bases ("mismatches") to total bases of mapped reads ("bases mapped") from the output of samtools stats version 1.9 (Li et al., 2009).

Five Illumina sequencing libraries (SRR10382360, SRR10382369, SRR10382370, SRR10382371, SRR1030294) were aligned to the assembly using *Bowtie2* version 2.2.6 (Langmead and Salzberg, 2012) with the following parameters

"bowtie2 –I 200 –X 500 –end-to-end –no-mixed –no-discordant". Read pairs which align perfectly were counted with the following command "cat file.sam | grep -v "ˆ@" | cut -f1,6 | uniq -c | grep 150M | grep -vc "1 "". The same five libraries were also aligned to a subset of PacBio corrected reads.

## Evaluation of Scaffolding Accuracy

Nanopore sequences (SRR10383383) were first filtered by q10 using *NanoFilt* (De Coster et al., 2018) and corrected to replace sequencing noise with consensus using *Canu* (Schmidt et al., 2017) with the following parameters "-genomeSize = 1g -correctedErrorRate = 0.105 -minReadLength = 3000 -minOverlapLength = 2000 -corOutCoverage = 200 "batOptions = -dg 3 -db 3 -dr 1 -ca 500 -cp 50" -ovlMerDistinct = 0.975". PacBio reads were also corrected using *Canu* (Schmidt et al., 2017) with the same parameters except for "-correctedErrorRate = 0.045". The corrected Nanopore and PacBio reads, were aligned to the assembly using *NGMLR* version 0.2.7 (Sedlazeck et al., 2018) with the following parameters "-x ont –no-smallinv".

The GOGG method (He and Bancroft, 2018), which uses the distribution pattern of gene allelic SNPs of 134 lines in a mapping population the AC *Brassica* genomics platform (He et al., 2015) to detect large structural misassemblies, was performed. The results were manually inspected for blocks with deviating patterns. The collinearity of putatively misassembled blocks with the AC pantranscriptome, *Arabidopsis thaliana* and *Thellungiella parvula* were used as supporting evidence for misjoins. Breakpoints of misjoins were resolved by inspecting the alignments of long reads with *IGV* (Robinson et al., 2011) in putatively misassembled regions. When supported by clear alteration of read coverage, the breakpoints were cut to isolate the misassembled blocks using *fastasubseq* under *Exonerate* suite (Slater and Birney, 2005). Gene annotation was updated to the corrected assembly using *flo* (Pracana et al., 2017) which implements the UCSC tool *liftover* (Kuhn et al., 2013).

## Evaluation of Gene Set Completeness

The completeness of genic regions were evaluated with two standard assessment pipelines, *BUSCO* v4.0.4 (Simão et al., 2015) and *PLAZA coreGF* (van Bel et al., 2012), where the presence of highly-conserved orthologs was used to score an assembly. *BUSCO* was performed on the genome assembly using the lineage dataset brassicales_odb10. Results were plotted using the generate_plot.py script of *BUSCO*. Since the coreGF python script does not set alignment threshold, predicted proteins were first aligned to PLAZA_2.5_proteome.fasta using *BLASTP* (Camacho et al., 2009), and only alignment with above 60 percentage identity and "-evalue 10e-5 –qcov_hsp_perc 60" were used to calculate for weighted score against the "greenplants" coreGFs.

Single-ended mRNA sequencing data (SRR3134083) was aligned to the assembled genome using *HISAT2* (Kim et al., 2019) and converted to intron boundaries using *bam2hints* (Augustus 3.2.1) (Hoff and Stanke, 2013, 2018). Positions were compared using *windowBed* (v2.25.0) (Quinlan and Hall, 2010).

Predicted proteins were evaluated with presence of known protein domains using *InterProScan* v5.33-72.0 (Jones et al., 2014) with the following parameters "interproscan.sh -appl TIGRFAM, PANTHER, Pfam, PrositeProfiles, PrositePatterns –iprlookup –goterms –pa".

The predicted coding sequences were also aligned to the A and C genome-based ordered pan-transcriptome (He et al., 2015) using *BLASTN* (Camacho et al., 2009) with the following parameters "-qcov_hsp_perc 60, -evalue 10e-5" and only alignments coverage of 60% to both target and query, a percentage identity of 60% and above were counted.

## Genome-to-Genome Comparison

Whole genome alignment between assemblies was performed using the graphical interface of *D-Genies* (Cabanettes and Klopp, 2018), which invokes *Minimap2* (Heng Li, 2018) internally and generates dotplots. Dotplots were displayed by applying a match size filtering, where matches that were grouped in the bins of smaller sizes (first and second out of 7 bins) were removed. Genomes used were *B. napus* Darmor-*bzh* v 4.1 (Chalhoub et al., 2014), *B. rapa* and *B. oleracea* (Belser et al., 2018).

Collinearity between genomes was identified by first obtaining orthologous genes and then these genes were used as anchor to detect synteny and collinearity using *MCScanX* (Wang et al., 2012). Since the *MCScanX* recommends around five hits per transcript, the alignment run was performed in two steps: (1) aligning all transcripts of both genomes to themselves and to each other using *BLASTN* (Camacho et al., 2009) with following parameters (2) filtering hits by alignment coverage and percentage identity of 80% and above. *MCScanX* was performed with parameters "match_score 50, match_size 10, gap_penalty -1, overlap_window 10, e_value 1e-05, max_gaps 10". The output was then plotted for visualization using *SynVisio*[3] to display blocks with final score of 10000 and above.

## DATA AVAILABILITY STATEMENT

## AUTHOR CONTRIBUTIONS

HL conducted the analysis and drafted the manuscript. HC and CO constructed the optical map and HC sequenced the Nanopore long reads. CO, FD, and AA constructed genetic maps. RS and HL conceived the study. All critically reviewed and edited the manuscript.

---

[3]https://kiranbandi.github.io/synvisio/

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00496/full#supplementary-material

## REFERENCES

Adey, A., Kitzman, J. O., Burton, J. N., Daza, R., Kumar, A., Christiansen, L., et al. (2014). In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. *Genome Res.* 24, 2041–2049. doi: 10.1101/gr.178319.114

Allender, C. J., and King, G. J. (2010). Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. *BMC Plant Biol.* 10:54. doi: 10.1186/1471-2229-10-54

Appelqvist, L.-A., and Ohlson, R. (1972). *Rapeseed : Cultivation, Composition, Processing and Utilization.* Amsterdam: Elsevier Publishing Company.

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 357, 93–97. doi: 10.1126/science.aan0032

Badani, A. G., Snowdon, R. J., Wittkop, B., Lipsa, F. D., Baetzel, R., Horn, R., et al. (2006). "Colocalization of a partially dominant gene for yellow seed colour with a major QTL influencing acid detergent fibre (ADF) content in different crosses of oilseed rape (*Brassica napus*). *Genome* 49, 1499–1509. doi: 10.1139/g06-091

Bao, Z. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* 12, 1269–1276. doi: 10.1101/gr.88502

Basunanda, P., Radoev, M., Ecke, W., Friedt, W., Becker, H. C., and Snowdon, R. J. (2010). Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* 120, 271–281. doi: 10.1007/s00122-009-1133-z

Bayer, P. E., Edwards, D., and Batley, J. (2018). Bias in resistance gene prediction due to repeat masking. *Nat. Plants* 4, 762–765. doi: 10.1038/s41477-018-0264-0

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

Becker, H. C., Engqvist, G. M., and Karlsson, B. (1995). Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. *Theor. Appl. Genet.* 91, 62–67. doi: 10.1007/BF00220859

Behnke, N., Suprianto, E., and Möllers, C. (2018). A major QTL on chromosome C05 significantly reduces acid detergent lignin (ADL) content and increases seed oil and protein content in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* 131, 2477–2492. doi: 10.1007/s00122-018-3167-6

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F. C., Falentin, C., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* 4, 879–887. doi: 10.1038/s41477-018-0289-4

Benevenuto, J., Ferrão, L. F. V., Amadeu, R. R., and Munoz, P. (2019). How can a high-quality genome assembly help plant breeders? *GigaScience* 8:giz068. doi: 10.1093/gigascience/giz068

Bhat, R. S., Shirasawa, K., Monden, Y., Yamashita, H., and Tahara, M. (2020). "Developing transposable element marker system for molecular breeding," in *Legume Genomics*, eds M. Jain, and R. Garg, (New York, NY: Springer), 233–251. doi: 10.1007/978-1-0716-0235-5_11

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27, 578–579. doi: 10.1093/bioinformatics/btq683

Buggs, R. J., Elliott, N. M., Zhang, L., Koh, J., Viccini, L. F., Soltis, D. E., et al. (2010). Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. *New Phytol.* 186, 175–183. doi: 10.1111/j.1469-8137.2010.03205.x

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

Cabanettes, F., and Klopp, C. (2018). D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6:e4958. doi: 10.7717/peerj.4958

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421

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

Chelaifa, H., Monnier, A., and Ainouche, M. (2010). Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina* × *Townsendii* and *Spartina Anglica* (Poaceae). *New Phytol.* 186, 161–174. doi: 10.1111/j.1469-8137.2010.03179.x

Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050–1054. doi: 10.1038/nmeth.4035

Coate, J. E., Bar, H., and Doyle, J. J. (2014). Extensive translational regulation of gene expression in an allopolyploid (*Glycine dolichocarpa*). *Plant Cell* 26, 136–150. doi: 10.1105/tpc.113.119966

Coghlan, A., Eichler, E. E., Oliver, S. G., Paterson, A. H., and Stein, L. (2005). Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends Genet.* 21, 673–682. doi: 10.1016/j.tig.2005.09.009

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. doi: 10.1093/bioinformatics/bty149

Delourme, R., Laperche, A., Bouchet, A.-S., Jubault, M., Paillard, S., Manzanares-Dauleux, M.-J., et al. (2018). "Genes and quantitative trait loci mapping for major agronomic traits in *Brassica napus* L," in *The Brassica Napus Genome*, eds S. Liu, R. Snowdon, and B. Chalhoub, (Cham: Springer International Publishing), 41–85. doi: 10.1007/978-3-319-43694-4_3

Diers, B. W., and Osborn, T. C. (1994). Genetic diversity of oilseed *Brassica napus* germ plasm based on restriction fragment length polymorphisms. *Theor. Appl. Genet.* 88, 662–668. doi: 10.1007/BF01253968

Edger, P. P., McKain, M. R., Bird, K. A., and van Buren, R. (2018). Subgenome assignment in allopolyploids: challenges and future directions. *Curr. Opin. Plant Biol.* 42, 76–80. doi: 10.1016/j.pbi.2018.03.006

Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29, 2150–2167. doi: 10.1105/tpc.17.00010

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi: 10.1126/science.1162986

Food and Agriculture Organization of the United Nations (2019). *FAOSTAT 2017*. Available online at: http://faostat.fao.org (accessed September, 2019).

Friedt, W., and Snowdon, R. (2009). "Oilseed rape," in *Oil Crops*, eds J. Vollmann, and I. Rajcan, (New York, NY: Springer), 91–126. doi: 10.1007/978-0-387-77594-4_4

Friedt, W., Tu, J., and Fu, T. (2018). "Academic and economic importance of *Brassica napus* rapeseed," in *The Brassica Napus Genome*, eds S. Liu, R. Snowdon, and B. Chalhoub, (Cham: Springer International Publishing), 1–20. doi: 10.1007/978-3-319-43694-4_1

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

Gabur, I., Chawla, H. S., Snowdon, R. J., and Parkin, I. A. P. (2019). Connecting genome structural variation with complex traits in crop plants. *Theor. Appl. Genet.* 132, 733–750. doi: 10.1007/s00122-018-3233-0

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

Gao, C. (2018). The future of CRISPR technologies in agriculture. *Nat. Rev. Mol. Cell Biol.* 19, 275–276. doi: 10.1038/nrm.2018.2

Gazave, E., Tassone, E. E., Ilut, D. C., Wingerson, M., Datema, E., Witsenboer, H. M., et al. (2016). Population genomic analysis reveals differential evolutionary histories and patterns of diversity across subgenomes and subpopulations of *Brassica napus* L. *Front. Plant Sci.* 7:525. doi: 10.3389/fpls.2016.00525

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* 11, 759–769. doi: 10.1111/j.1755-0998.2011.03024.x

Gremme, G., Brendel, V., Sparks, M. E., and Kurtz, S. (2005). Engineering a software tool for gene structure prediction in higher organisms. *Inform. Softw. Technol.* 47, 965–978. doi: 10.1016/j.infsof.2005.09.005

Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., et al. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat. Biotechnol.* 30, 798–802. doi: 10.1038/nbt.2302

He, Z., and Bancroft, I. (2018). Organization of the genome sequence of the polyploid crop species *Brassica juncea*. *Nat. Genet.* 50, 1496–1497. doi: 10.1038/s41588-018-0239-0

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

He, Z., Wang, L., Harper, A., Havlickova, L., Pradhan, A., Parkin, I., and Bancroft, I. (2017). Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. *Plant Biotechnol. J.* 15, 594–604. doi: 10.1111/pbi.12657

Hoff, K. J., and Stanke, M. (2013). WebAUGUSTUS–a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Res.* 41, W123–W128. doi: 10.1093/nar/gkt418

Hoff, K. J., Lomsadze, A., Borodovsky, M., and Stanke, M. (2019). "Whole-genome annotation with BRAKER," in *Gene Prediction*, ed. M. Kollmar, (New York, NY: Springer), 65–95. doi: 10.1007/978-1-4939-9173-0_5

Hoff, K. J., and Stanke, M. (2018). Predicting genes in single genomes with AUGUSTUS. *Curr. Protoc. Bioinform.* 65:e57. doi: 10.1002/cpbi.57

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

Jiao, W. B., Accinelli, G. G., Hartwig, B., Kiefer, C., Baker, D., Severing, E., et al. (2017). Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. *Genome Res.* 27, 778–786. doi: 10.1101/gr.213652.116

Johnston, J. S., Pepper, A. E., Hall, A. E., Chen, Z. J., Hodnett, G., Drabek, J., et al. (2005). Evolution of genome size in brassicaceae. *Ann. Bot.* 95, 229–235. doi: 10.1093/aob/mci016

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

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Kim, S., Park, J., Yeom, S. I., Kim, Y. M., Seo, E., Kim, K. T., et al. (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* 18:210. doi: 10.1186/s13059-017-1341-9

Koren, S., Rhie, A., Walenz, B. P., Dilthey, A. T., Bickhart, D. M., Kingan, S. B., et al. (2018). De novo assembly of haplotype-resolved genomes with trio binning. *Nat. Biotechnol.* 36, 1174–1182. doi: 10.1038/nbt.4277

Korlach, J. (2013). *Understanding Accuracy in SMRT§Sequencing: Pacific Biosciences*. Available online at: https://www.mscience.com.au/upload/pages/pacbioaccuracy/perspective-understanding-accuracy-in-smrt-sequencing.pdf

Kuhn, R. M., Haussler, D., and Kent, W. J. (2013). The UCSC genome browser and associated tools. *Brief. Bioinform.* 14, 144–161. doi: 10.1093/bib/bbs038

Lai, X., Schnable, J. C., Liao, Z., Xu, J., Zhang, G., Li, C., et al. (2017). Genome-wide characterization of non-reference transposable element insertion polymorphisms reveals genetic diversity in tropical and temperate maize. *BMC Genomics* 18:702. doi: 10.1186/s12864-017-4103-x

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 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, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, X., Guo, K., Zhu, X., Chen, P., Li, Y., Xie, G., et al. (2017). Domestication of rice has reduced the occurrence of transposable elements within gene coding regions. *BMC Genomics* 18:55. doi: 10.1186/s12864-016-3454-z

Ling, H.-Q., Ma, B., Shi, X., Liu, H., Dong, L., Sun, H., et al. (2018). Genome sequence of the progenitor of wheat a subgenome *Triticum urartu*. *Nature* 557, 424–428. doi: 10.1038/s41586-018-0108-0

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, A. P. I., et al. (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* 5:3930. doi: 10.1038/ncomms4930

Lloyd, A., Blary, A., Charif, D., Charpentier, C., Tran, J., Balzergue, S., et al. (2018). Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. *New Phytol.* 217, 367–377. doi: 10.1111/nph.14836

Loman, N. J., Quick, J., and Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* 12, 733–735. doi: 10.1038/nmeth.3444

Lombard, V., and Delourme, R. (2001). A consensus linkage map for rapeseed (*Brassica napus* L.): construction and integration of three individual maps from DH populations. *Theor. Appl. Genet.* 103, 491–507. doi: 10.1007/s001220100560

Lukens, L. N., Pires, J. C., Leon, E., Vogelzang, R., Oslach, L., and Osborn, T. (2006). 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

Luo, M. C., Gu, Y. Q., Puiu, D., Wang, H., Twardziok, S. O., Deal, K. R., et al. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* 551, 498–502. doi: 10.1038/nature24486

Lysak, M. A., Cheung, K., Kitschke, M., and Bures, P. (2007). Ancestral chromosomal blocks are triplicated in brassiceae species with varying chromosome number and genome size. *Plant Physiol.* 145, 402–410. doi: 10.1104/pp.107.104380

Mandáková, T., Joly, S., Krzywinski, M., Mummenhoff, K., and Lysak, M. A. (2010). Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. *Plant Cell* 22, 2277–2290. doi: 10.1105/tpc.110.074526

Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., et al. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *BioTechniques* 61, 203–205. doi: 10.2144/000114460

Myers, G. (2014). "Efficient local alignment discovery amongst noisy long reads," in *Algorithms in Bioinformatics*, eds D. Brown, and B. Morgenstern, (Berlin: Springer), doi: 10.1007/978-3-662-44753-6_5

Nagaharu, U. (1935). Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn. J. Bot.* 7, 389–452.

Nagarajan, N., and Pop, M. (2013). Sequence assembly demystified. *Nat. Rev. Genet.* 14, 157–167. doi: 10.1038/nrg3367

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

Osborn, T. C., Pires, J. C., Birchler, J. A., Auger, D. L., Chen, Z. J., Lee, H. S., et al. (2003). Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* 19, 141–147. doi: 10.1016/S0168-9525(03)00015-5

Pan, W., and Lonardi, S. (2019). Accurate detection of chimeric contigs via bionano optical maps. *Bioinformatics* 35, 1760–1762. doi: 10.1093/bioinformatics/bty850

Pan, W., Wanamaker, S. I., Ah-Fong, A. M. V., Judelson, H. S., and Lonardi, S. (2018). Novo&Stitch: accurate reconciliation of genome assemblies via optical maps. *Bioinformatics* 34, i43–i51. doi: 10.1093/bioinformatics/bty255

Parkin, I. A., Gulden, S. M., Sharpe, A. G., Lukens, L., Trick, M., Osborn, T. C., et al. (2005). Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171, 765–781. doi: 10.1534/genetics.105.042093

Piquemal, J., Cinquin, E., Couton, F., Rondeau, C., Seignoret, E., Doucet, I., et al. (2005). Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. *Theor. Appl. Genet.* 111, 1514–1523. doi: 10.1007/s00122-005-0080-6

Pires, J. C., Zhao, J., 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

Pracana, R., Priyam, A., Levantis, I., Nichols, R. A., and Wurm, Y. (2017). The fire ant social chromosome supergene variant SB shows low diversity but high divergence from SB. *Mol. Ecol.* 26, 2864–2879. doi: 10.1111/mec.14054

Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics* 21(Suppl. 1), i351–i358. doi: 10.1093/bioinformatics/bti1018

Qian, L., Voss-Fels, K., Cui, Y., Jan, H. U., Samans, B., Obermeier, C., et al. (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

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Radoev, M., Becker, H. C., and Ecke, W. (2008). Genetic analysis of heterosis for yield and yield components in rapeseed (*Brassica napus* L.) by quantitative trait locus mapping. *Genetics* 179, 1547–1558. doi: 10.1534/genetics.108.089680

Renny-Byfield, S., Gong, L., Gallagher, J. P., and Wendel, J. F. (2015). Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. *Mol. Biol. Evol.* 32, 1063–1071. doi: 10.1093/molbev/msv001

Rigal, M., Becker, C., Pélissier, T., Pogorelcnik, R., Devos, J., Ikeda, Y., et al. (2016). Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in *Arabidopsis thaliana* F1 epihybrids. *Proc. Natl. Acad. Sci. U.S.A.* 113, E2083–E2092. doi: 10.1073/pnas.1600672113

Robinson, J. T., Thorvaldsdóttir, 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

Rousseau-Gueutin, M., Morice, J., Coriton, O., Huteau, V., Trotoux, G., Nègre, S., et al. (2017). The impact of open pollination on the structural evolutionary dynamics, meiotic behavior, and fertility of resynthesized allotetraploid *Brassica napus* L. *G3 (Bethesda)* 7, 705–717. doi: 10.1534/g3.116.036517

Salmon, A., Ainouche, M. L., and Wendel, J. F. (2005). Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Mol. Ecol.* 14, 1163–1175. doi: 10.1111/j.1365-294X.2005.02488.x

Samans, B., Chalhoub, B., and Snowdon, R. J. (2017). Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species. *Plant Genome* 10, 1–15. doi: 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. doi: 10.1038/srep41845

Schmidt, M. H., Vogel, A., Denton, A. K., Istace, B., Wormit, A., van de Geest, H., et al. (2017). De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *Plant Cell* 29, 2336–2348. doi: 10.1105/tpc.17.00521

Schnable, J. C., Springer, N. M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4069–4074. doi: 10.1073/pnas.1101368108

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

Sharpe, A. G. I, Parkin, A. P., Keith, D. J., and Lydiate, D. J. (1995). Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). *Genome* 38, 1112–1121. doi: 10.1139/g95-148

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Slater, G., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* 6:31. doi: 10.1186/1471-2105-6-31

Smit, A. F. A., and Hubley, R. (2008). *RepeatModeler Open-1.0*. http://www.repeatmasker.org (accessed October, 2018).

Smit, A. F. A., Hubley, R., and Green, P. (2013). *RepeatMasker Open-4.0*. http://www.repeatmasker.org (accessed October, 2019).

Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: join map. *Plant J.* 3, 739–744. doi: 10.1111/j.1365-313X.1993.00739.x

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

Stein, A., Wittkop, B., Liu, L., Obermeier, C., Friedt, W., and Snowdon, R. J. (2013). "Dissection of a major QTL for seed colour and fibre content in *Brassica napus* reveals colocalization with candidate genes for phenylpropanoid biosynthesis and flavonoid deposition. *Plant Breed.* 132, 382–389. doi: 10.1111/pbr.12073

Sun, F., Fan, G., Hu, Q., Zhou, Y., Guan, M., Tong, C., et al. (2017). "The high-quality genome of *Brassica napus* cultivar 'ZS11' reveals the introgression history in semi-winter morphotype. *Plant J.* 92, 452–468. doi: 10.1111/tpj.13669

Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J. C., et al. (2015). ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biol.* 16:3. doi: 10.1186/s13059-014-0573-1

Tao, Y., Zhao, X., Mace, E., Henry, R., and Jordan, D. (2019). Exploring and exploiting pan-genomics for crop improvement. *Mol. Plant* 12, 156–169. doi: 10.1016/j.molp.2018.12.016

The International Wheat Genome Sequencing Consortium (IWGSC), Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191

UniProt Consortium, (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049

Udall, J. A., Quijada, P. A., and Osborn, T. C. (2005). Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L. *Genetics* 169, 967–979. doi: 10.1534/genetics.104.033209

van Bel, M., Proost, S., Wischnitzki, E., Movahedi, S., Scheerlinck, C., Van de Peer, Y., et al. (2012). Dissecting plant genomes with the PLAZA comparative genomics platform. *Plant Physiol.* 158, 590–600. doi: 10.1104/pp.111.189514

van Ooijen, J. W. (2011). Multipoint maximum likelihood mapping in a full-Sib family of an outbreeding species. *Genet. Res. (Camb)* 93, 343–349. doi: 10.1017/S0016672311000279

Veeckman, E., Ruttink, T., and Vandepoele, K. (2016). Are we there yet? Reliably estimating the completeness of plant genome sequences. *Plant Cell* 28, 1759–1768. doi: 10.1105/tpc.16.00349

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection

and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal. pone.0112963

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Watson, M., and Warr, A. (2019). Errors in long-read assemblies can critically affect protein prediction. *Nat. Biotechnol.* 37, 124–126. doi: 10.1038/s41587-018-0004-z

Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of de novo genome assemblies: fig. 1. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383

Wu, Y., Bhat, P. R., Close, T. J., and Lonardi, S. (2008). Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet.* 4:e1000212. doi: 10.1371/journal.pgen.1000212

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

Zhao, G., Zou, C., Li, K., Wang, K., Li, T., Gao, L., et al. (2017). The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nat. Plants* 3, 946–955. doi: 10.1038/s41477-017-0067-8

**Conflict of Interest:** AA and FD were employed by the company NPZ Innovation GmbH.

The remaining 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.