

RESEARCH ARTICLE

Open Access



# Linkage mapping and QTL analysis of flowering time using ddRAD sequencing with genotype error correction in *Brassica napus*

Armin Scheben<sup>1,2</sup>, Anita A. Severn-Ellis<sup>1</sup>, Dhvani Patel<sup>1</sup>, Aneeta Pradhan<sup>1</sup>, Stephen J. Rae<sup>3</sup>, Jacqueline Batley<sup>1</sup> and David Edwards<sup>1\*</sup> 

## Abstract

**Background:** *Brassica napus* is an important oilseed crop cultivated worldwide. During domestication and breeding of *B. napus*, flowering time has been a target of selection because of its substantial impact on yield. Here we use double digest restriction-site associated DNA sequencing (ddRAD) to investigate the genetic basis of flowering in *B. napus*. An F<sub>2</sub> mapping population was derived from a cross between an early-flowering spring type and a late-flowering winter type.

**Results:** Flowering time in the mapping population differed by up to 25 days between individuals. High genotype error rates persisted after initial quality controls, as suggested by a genotype discordance of ~ 12% between biological sequencing replicates. After genotype error correction, a linkage map spanning 3981.31 cM and comprising 14,630 single nucleotide polymorphisms (SNPs) was constructed. A quantitative trait locus (QTL) on chromosome C2 was detected, covering eight flowering time genes including *FLC*.

**Conclusions:** These findings demonstrate the effectiveness of the ddRAD approach to sample the *B. napus* genome. Our results also suggest that ddRAD genotype error rates can be higher than expected in F<sub>2</sub> populations. Quality filtering and genotype correction and imputation can substantially reduce these error rates and allow effective linkage mapping and QTL analysis.

**Keywords:** *Brassica napus*, Double digest restriction-site associated DNA sequencing, Flowering, QTL, SNP

## Background

Genotyping-by-sequencing (GBS) is a powerful tool for high-throughput discovery of genetic polymorphisms in crops [1–5]. GBS comprises a range of library preparation and sequencing approaches that differ in their costs, methodical biases and the type and amount of data produced [1, 6]. Restriction site-associated DNA

sequencing (RAD) is a GBS method that can be used to cost-effectively calibrate the number and coverage of genotyped loci and single nucleotide polymorphisms (SNPs) by varying the enzymes used and the sequencing depth. A recent comparative analysis of single enzyme RAD and two enzyme double digest RAD (ddRAD) used a range of enzyme combinations in different plants and suggested that the enzyme combination of HinfI and HpyCH4IV was promising for maximising genome coverage breadth across a range of species [7]. Like other GBS approaches, ddRAD is prone to missing data and

\* Correspondence: [dave.edwards@uwa.edu.au](mailto:dave.edwards@uwa.edu.au)

<sup>1</sup>School of Biological Sciences and Institute of Agriculture, The University of Western Australia, Perth, WA, Australia

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

undercalling of heterozygous genotypes [5], but imputation and correction methods can help produce high quality genotypes. Imputation and correction approaches used in crops include the hidden Markov model based LB-Impute [8] and FSFHap [9], the sliding window based Genotype-Corrector [10] and simple heuristic approaches [11].

GBS has been used for marker discovery, linkage mapping and QTL analysis in a range of crops [12–14], including the important oilseed crop *Brassica napus*. Over 20 high density linkage maps have been generated for *B. napus* using RNA sequencing [15], the Brassica 60 K genotyping array [16, 17] and ddRAD sequencing [18]. Combined with phenotypic data, these linkage maps provide a powerful basis for identification of genes underlying agronomic traits, which can then be introduced into crop germplasm [19, 20]. Crop yield in *B. napus* is strongly dependent on flowering time, making this trait a key breeding target. Flowering time genetic pathways have been elucidated in Arabidopsis and most flowering time genes are known to be conserved between Arabidopsis and *B. napus* [21–23]. On this basis, many QTL and associated SNPs for flowering time have been detected in *B. napus* [24–31]. However, despite progress in understanding the genetic underpinnings of *B. napus* flowering time, a substantial proportion of flowering time variation remains to be explained.

There are three *B. napus* oilseed rape (OSR) growth types with considerable variation in flowering time: spring, semi-winter and winter. Spring OSR and semi-winter OSR have a low requirement for vernalization to flower and are early-flowering, whilst winter OSR has a strong vernalization requirement and is late-flowering. In *B. napus* breeding, the flowering traits of spring OSR decrease its generation time compared to winter OSR, allowing more rapid breeding cycles. Reducing vernalization requirements in winter OSR by introducing spring OSR alleles would be one approach to allow breeders to accelerate winter OSR breeding. In addition, *B. napus* hybrids are generally higher yielding than open pollinated varieties due to heterosis [32, 33]. If flowering time can be efficiently managed, heterosis could be exploited from hybridization of spring OSR and winter OSR. Identifying flowering time loci that distinguish spring OSR and winter OSR therefore has important breeding applications. Here, to identify these loci, we crossed a spring OSR and a winter OSR to generate an F<sub>2</sub> mapping population. We genotyped the progeny and parental lines using ddRAD sequencing. Finally, we constructed a high-density linkage map and carried out QTL analysis of flowering time and the related trait budding time. We present candidate regions for flowering time and budding time and discuss the use of error-prone ddRAD genotyping in heterozygous breeding populations.

## Results

### Pre-processing and aligning sequencing reads

A single individual (sample ID: 146) was excluded from further analysis as it had fewer than one million reads after trimming. In the remaining 206 samples (consisting of 199 F<sub>2</sub> progenies, 4 replicates of BnSOSR and 3 replicates of BnWOSR), a mean of 13.14 million raw paired sequences were generated per sample. A mean of 56.2% of reads were uniquely aligned with high quality. The mean coverage depth at covered bases was 9.41 × and the mean coverage breadth of the genome was 18.03%.

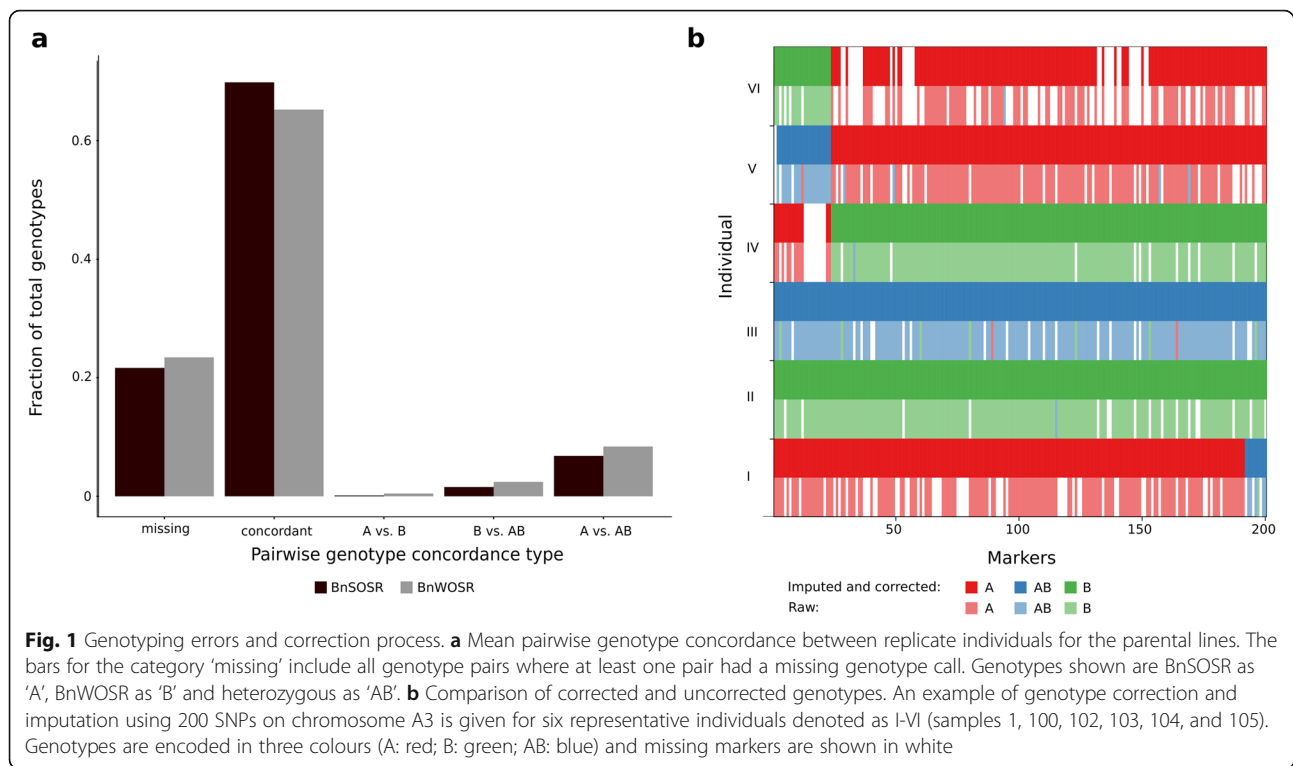
### SNP filtering and genotype correction

A total of 4,841,931 biallelic SNPs were identified in the mapping population and parental individuals. For further analysis, the seven parental individuals were excluded from the SNP set. Filtering by individual missingness, genotype depth, minor allele frequency (MAF), and genotype missingness reduced the number of SNPs to 124,804. Of the 199 progeny, 192 were retained after filtering individuals with high genotype missingness. Of the 124,804 SNPs, 50,856 did not have a heterozygous genotype in any parental individual. The SNPs with heterozygous genotypes in the parental individuals may be caused by mismapping or remaining heterozygosity in the parental genomes and were therefore excluded. Next, removing 16,647 SNPs that were monomorphic between parents and 5957 that showed segregation distortion ( $p < 0.01$ ), generated a set of 28,252 SNPs. Segregation distorted SNPs were distributed relatively evenly across chromosomes, with noticeable hotspots at the ends of chromosomes A1 and C5 (Fig. S1). Genotype-Corrector quality control removed 13,509 further SNPs after filtering homozygous SNPs located within heterozygous regions. A total of 4.94% of genotypes were corrected using Genotype-Corrector and 94.76% of missing genotypes were imputed (Fig. 1). The most frequent genotype corrections were B to AB (29.56%) and A to AB (23.48%).

In the parental replicate individuals, analysis of pairwise genotype concordance identified a mean genotype discordance of 12.28% (Fig. 1). Discordance between homozygous genotypes (A vs. B) was rare, with conflicts between homozygous and heterozygous genotypes (A vs. AB, B vs. AB) making up 97.51% of genotype discordance.

### Linkage mapping

A linkage map spanning 3981.31 cM and comprising 14,630 markers was constructed using ASMap with the corrected and imputed markers (Fig. 2 and Table S1). The A genome map was 2147.15 cM with 8587 markers and the C genome map was 1834.16 cM with 6043 markers. The highest mean marker density was found on



chromosome A10, with 48.0 markers per Mb. Mean marker density per Mb was higher on the A genome (28.33) than the C genome (12.19). A supplementary map was constructed using uncorrected markers, which showed a high inflation of genetic distances with a total map length over 30,000 cM (Fig. S2). Compared to six published genetic maps that were generated with different approaches, the corrected genetic map still showed some indications of inflation (Table S2).

The correlation between genetic and physical map order provides information about the consistency between the genetic map and the reference genome. The mean Spearman's rank correlation for marker order per chromosome was 1.0 (Table 1). Several minor inconsistencies in marker order were observed (Fig. S3). All chromosomes showed mean correlations over 0.98. Mean individual crossover frequency per chromosome was 2.79.

### QTL mapping

BnSOSR flowered 20 days earlier (range: 10 to 28) and went to bud 17 days earlier (range: 12 to 20) on average than winter type BnWOSR (Table S3). In the  $F_2$  progeny, flowering times were distributed within the parental range (Fig. 3). A single significant ( $p < 0.05$ ) overlapping QTL region for the traits flowering time and budding time was detected on chromosome C2 (Fig. 3). The physical region of the QTL spanned 20.57 Mb for flowering time and 0.77 Mb for time to bud (Table S4).

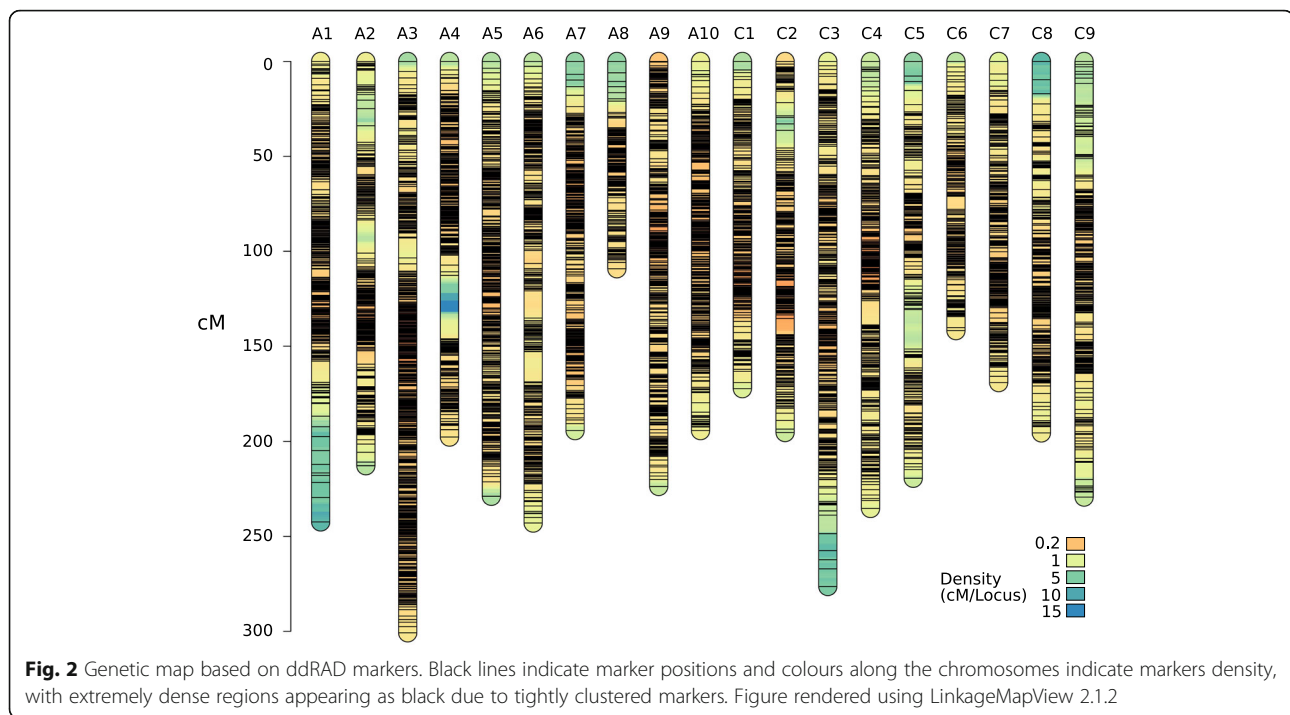
The flowering time QTL contained 8 flowering time homologs including *FLC* (Table S5), and the budding time QTL contained no known flowering time homologs. Carrying the BnWOSR allele at the QTL peak SNP led to an increase in the days to bud and flower (Fig. S4). The percentage of phenotypic variance explained for the identified QTL was 9.08% for flowering time and 8.08% for budding time. Suggestive LOD peaks are also noticeable on A2, A3 and C9.

Low coverage whole genome sequencing of 16 early-flowering and 19 late-flowering  $F_2$  individuals detected 137,696 variants on chromosome C2. The mean coverage of candidate gene exons was  $1.20 \times$  (early-flowering) and  $1.04 \times$  (late-flowering) (Table S6). No segregating non-synonymous substitutions in candidate gene coding sequences were observed (Table S5), but three candidate segregating intergenic variants within 1 kb of a candidate gene were found (Table S7).

### Discussion

#### Optimising genotyping-by-sequencing strategies

Genotyping with ddRAD was effective at generating a set of 14,630 high-quality SNPs for linkage and QTL mapping. The findings of this study can help calibrate the number of sequencing reads and genomic loci required for a range of research goals in *B. napus* and related species. For GBS, researchers often aim to optimize the genome coverage by controlling the ratio of reads sequenced to the number of loci generated. Here,



the expected maximum genome coverage breadth based on *in silico* digestion with the enzymes HpyCH4IV and HinfI was 24.6%. However, the observed mean genome coverage breadth was lower at 18.03%. An even greater inconsistency was reported when using HpyCH4IV and HinfI in *Arabidopsis* and *Glycine max*, where the expected genome coverage breadth was 29.4 and 23.1% but the observed experimental values were 4.45 and 3.33% respectively [7]. The inconsistency has been explained as a product of fragment size selection bias, redundant *in silico* loci, and insufficient sequencing reads [7]. In the *B. napus* population used here, the most important factor determining coverage was the amount of sequencing reads available for each sample. Indeed, 24 samples with high sequencing effort showed genome coverage breadth greater than the expected 24.6% up to a maximum value of 36.68%. This is particularly surprising, as coverage was calculated based only on reads aligned with high quality, which is expected to substantially reduce coverage breadth. These findings suggest that, at least in *B. napus*, igCoverage can underestimate the maximum achievable genome coverage breadth.

The high genome coverage breadth achieved using HpyCH4IV and HinfI indicates that these enzymes are well suited for high-density sampling of genome-wide diversity in *B. napus*. However, when sequencing effort is uneven between samples, high genome coverage breadth can increase genotype missingness through allele dropout. If a locus is not sequenced in enough individual samples (here the cut off was 50%) at sufficient depth, it is removed during SNP

calling or SNP filtering and becomes a missing genotype call. High levels of missingness are a common characteristic of reduced representation sequencing [3] and can limit the usefulness of genotype data in studies where genotype imputation is not possible [34]. Nevertheless, as shown in this study, by combining imputation with a high-density sampling of the genome, the limitations of genotype missingness in a mapping population can be overcome.

### Linkage mapping

The correlation of physical and genetic maps was high, indicating that the map is accurate and collinear with the reference genome. Similarly, collinear maps with only minor inconsistencies were also found by earlier linkage mapping studies in *B. napus* [18, 35]. The linkage map constructed here was on average 2 × larger in cM compared to six published *B. napus* linkage maps generated using different approaches with relatively similar marker densities [15, 17, 18, 35–37]. Although our mapping population is derived from two highly divergent parental lines and may enable us to sample more crossovers than other studies, some residual genotype errors or segregation distortion is expected to lead to some map inflation. In contrast to our study, which relied on an F<sub>2</sub> population, five of the compared studies used recombinant inbred lines or doubled haploid populations, which will likely suffer fewer genotyping errors due to their inherent lack of heterozygous alleles. Using a genotyping array, which is less error-prone than ddRAD genotyping, in a *B. napus* F<sub>2</sub> population also led



**Table 1** Summary of the genetic map. Spearman's rank correlation ( $\rho$ ) was calculated for the genetic marker positions and the physical marker positions on the reference genome

Chr	Length (Mb)	Length (cM)	Number of markers	$\rho$
A1	31.16	242.40	740	0.98
A2	31.34	212.85	630	1.00
A3	39.49	300.82	1581	1.00
A4	23.31	197.72	748	1.00
A5	28.6	228.92	892	1.00
A6	31.9	243.03	729	1.00
A7	28.9	194.30	954	1.00
A8	21.74	109.17	365	1.00
A9	46.72	223.68	990	1.00
A10	19.96	194.26	958	0.99
C1	47.95	172.18	739	1.00
C2	58.66	195.33	832	0.99
C3	71.85	276.37	900	1.00
4	61.04	235.26	861	0.99
C5	52.72	219.41	483	1.00
C6	44.61	141.83	504	1.00
C7	52.5	168.96	612	1.00
C8	46.29	195.51	569	1.00
C9	60.21	229.31	543	1.00
Total	798.95	3981.31	14,630	–

to a smaller genetic map of roughly half the size [17]. This suggests that ddRAD genotyping errors, and not the population type, are the main reason for the genetic map inflation. Nevertheless, the high-density and accuracy of the linkage map presented here suggest that it is useful for localizing QTL.

#### Flowering time QTL on chromosome C2

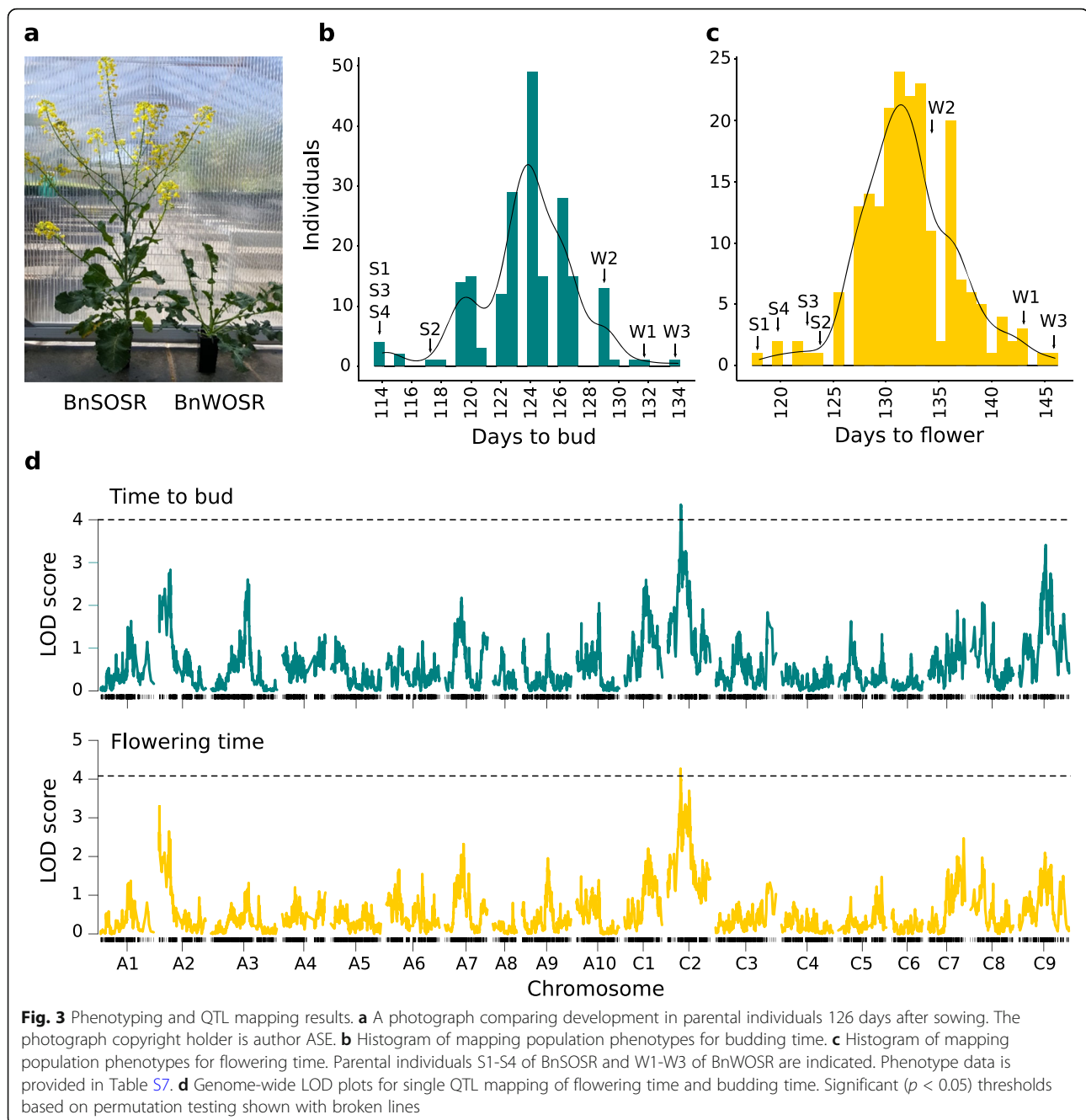
Flowering time QTL in *B. napus* are mostly found in regions syntenic with Arabidopsis chromosome 5 on *B. napus* chromosomes A2, A3, A9, A10, C2 and C3 [22]. Here, we identified a single significant locus for flowering time and budding time on C2. These two phenotypes were significantly correlated and could be linked to a likely shared QTL that explained ~ 9% of variance. This modest amount of variance explained is typical for traits that are controlled by many loci spread over multiple chromosomes, with each making minor contributions to the phenotypic effect. Additionally, the size of the QTL region (LOD confidence interval) differed for flowering time and budding time. The size of the QTL region is important, as it reflects the level of mapping resolution that was achieved. In  $F_2$  mapping populations, the confidence interval of a QTL can be large (> 1 Mb), and these populations often represent the starting point for fine-

mapping of candidate genes. In this study we used the recommended LOD threshold of 1.5 units for 95% coverage of the confidence interval [38]. However, the width of the confidence interval depends on how steep the QTL peak is, which can depend on a range of factors including marker density [39, 40].

The QTL is physically close to a locus identified in an earlier study, which found that the 60 K array SNP Bn-scaff\_18507\_1-p889927 was associated with a QTL on C2 at position 33,936,984 on v81 (position on Darmor v41: 26,548,393) explaining 6.36% of flowering time variation [28]. However, the QTL LOD peaks identified here are distant from this location. Among the known flowering time genes on C2, *FT* [41], *FLC* [42] and *FY* [43] may have particularly substantial effects. *FT* has been implicated in *B. napus* flowering time divergence [30, 44] and *FLC* has been found to explain ~ 23% of flowering time variation in *B. napus* [45]. *FY* is a suppressor of the transcription factor *FLC* [43, 46], but is distant from the QTL region identified in this study. The *FT* homologs on C2 are 1-3 Mb outside of the QTL region and are expressed at low levels in all ecotypes tested [47], which suggests *FT* may also not be the candidate gene. None of the candidate flowering time genes showed nonsynonymous or synonymous substitutions segregating with flowering time, indicating that regulatory changes may underlie the detected QTL. *FLC* and *SRR1* are located within the QTL region and were found to have variants within 1 kb of their coding sequence that may segregate between early-flowering and late-flowering  $F_2$  individuals. By further studying differential expression of the candidate genes, it may be possible to determine whether *FLC*, *SRR1* or another gene is driving differences in flowering time. Overall, the  $F_2$  experimental design presented here is only the start of the discovery process because lack of recombination between closely linked regions can hinder high-resolution mapping. Our results show one significant QTL and additional suggestive regions of interest. It is likely that several of these loci would need to be transferred into the desired genetic background to exploit heterosis between spring and winter varieties.

#### Genotype errors and correction

We detected high pairwise genotype discordance within the duplicate parental samples. Because genotype errors in either of the compared duplicate samples can lead to discordance, the genotype error rate can be roughly estimated as half of the discordance (~ 6%). In line with GBS results in a rice  $F_2$  population [11], most errors can be attributed to undercalling of heterozygous alleles (97.51%). The parental lines are homozygous, though residual heterozygosity and mismapping can lead to heterozygous allele calls. The true error rate in the progeny



is therefore likely higher than in the parents, because  $F_2$  populations contain an expected 50% heterozygous alleles.

Calling heterozygous SNPs accurately requires depths substantially higher than those required for calling homozygous SNPs [48, 49]. The moderate sequencing depths used in this study (9.41 $\times$ ) may thus lead to inflation of sequencing noise and insufficient allele sampling, which can result in undercalling of heterozygous alleles [50]. The percentage of errors attributed to undercalling of heterozygotes may even be an underestimate, as

errors between apparently homozygous alleles may be caused by conflicting erroneous genotype calls of a heterozygous allele.

The genotype error rates found here are higher than error rates reported in the literature, even for heterozygous populations. For example, Malmberg et al. (2018a) analysed a heterozygous *B. napus* panel with different skim sequencing coverages and filtered genotype calls using a minimum read depth of 5. The authors found error rates of 2.1% error at 2 $\times$  sequencing coverage and 4.2% error at 1 $\times$  sequencing coverage [50]. Similarly, an

error rate of 3% was found using GBS in a bovine population with a minimum read depth of 5 [51]. In a ddRAD genotyping study using a mapping population of cichlid fishes, an investigation of genotype errors found error rates of 4.41% at 8× coverage [52]. This suggests that GBS can lead to higher genotype error rates than expected in samples such as  $F_2$  mapping populations.

A substantial effect of genotype errors on linkage mapping was found. Cumulative inflation of linkage map length is often caused by genotype errors that introduce spurious double recombination events into the map [53]. In addition, high levels of missing data and markers with segregation distortion may also affect the mapping distance as these alter the calculated recombination rate [54]. It has been estimated that every 1% error rate in a marker adds approximately two cM to the linkage map [55]. Linkage map inflation has been previously reported for GBS data in wheat [53, 54] and rice [11]. In one of the studies on wheat, errors inflated the linkage map by up to 11 times [53]. In a linkage and QTL mapping study of *B. rapa* based on SNPs derived from GBS, high error rates (19.6%) were found and the resulting A genome linkage map was inflated, spanning 4802.52 cM [56].

Linkage and QTL mapping in major crops are commonly carried out using highly accurate commercial genotyping arrays such as the Illumina Infinium Brassica 60 K array [57]. Genotyping arrays may introduce sampling bias because they only genotype previously known SNPs. An advantage of GBS over genotyping arrays is that regions missing from the reference genome can be genotyped and used for linkage mapping. However, genotyping arrays have the important advantages of more accurate calling of heterozygous genotypes and low missing data. Although genotype arrays also produce errors that frequently involve heterozygous sites, the error rates are likely lower at 1–2% [58, 59]. Our results suggest that in heterozygous populations, genotyping arrays will generate markers with substantially higher accuracy than GBS. Here, to increase genotype accuracy for GBS data, genotype correction was applied. We find that genotype correction substantially decreased genetic map inflation, underlining the value of a correction step in heterozygous populations analysed using GBS at low to moderate sequencing coverage.

## Conclusions

We report a QTL on chromosome C2 for flowering time and budding time in a *B. napus* winter type × spring type cross. This QTL and the additional suggestive loci can be fine-mapped and backcrossed into the parental varieties to facilitate flowering time control in hybrid spring type × winter type varieties. An optimised combination of enzymes was also identified using in silico analysis,

and the resulting number of empirical ddRAD loci and SNPs demonstrate the effectiveness of the enzyme combination HinfI and HpyCH4IV. In addition, we show that ddRAD generates high levels of genotype errors that can impact linkage map construction. By filtering SNPs and applying genotype imputation and correction an accurate map could be constructed, allowing effective QTL analysis of flowering time. Further investigation of the loci controlling flowering time and maturation will allow *B. napus* breeders to better exploit variation in winter and spring types.

## Methods

### Plant material and phenotyping

Plant materials were provided by BASF (Ghent). No formal identification was required for this cultivated plant material. The mapping population resulted from a cross between an early-flowering spring line (BnSOSR) and a late-flowering winter line (BnWOSR) carried out by BASF. The  $F_2$  population consisting of 200 individuals as well as 4 BnSOSR and 3 BnWOSR parental replicates was sown directly into 10×10×15 cm pots in a phytotron at the University of Western Australia, Perth, in 2017. The  $F_2$  population was investigated based on single plants. Temperature was maintained between 18 and 22 °C. To ensure flowering occurred, vernalization was initiated 63 days after sowing and plants were moved to a controlled environment chamber with a constant temperature of 4 °C and an 8-h light period. Plants were watered twice a week by hand. After approximately 6 weeks, and a total of 108 days after sowing, the plants were returned to the phytotron. This 6 week vernalization period is sufficiently long to ensure that variation in flowering time in the  $F_2$  population is not driven by variation in vernalization response [60, 61]. Each pot was provided with a dripper and connected to the irrigation system. Plants were watered twice a day for 1 min. A total of 125 ml of fertilizer with micro minerals was provided by hand every 2 weeks. The time of first floral buds appearing and the date of first flower opening were recorded. All plants were grown until seed set. To investigate whether phenotype data were normally distributed, Shapiro-Wilke's tests were carried and histograms plotted using ggpubr 0.2.1 [62].

### Restriction enzyme selection and digestion

DNA fragmentation was carried out by simultaneous digestion using two restriction enzymes. Suitable restriction enzyme pairs, that created sticky or overhanging ends, were selected based on their reaction buffer and incubation temperature compatibility to allow simultaneous digestion. The software IgCoverage 1.0 [7] was used to carry out in silico digestion of the *B. napus* genome (Darmor-bzh v8.1 [63]); using these enzyme pairs. The

number of fragments within the 100–600 bp size range with different ends (LengthDeFrag100–600), as well as the expected percentage of genome coverage breadth generated by the selected 26 restriction enzyme pairs, were then compared. Out of the 26 suitable enzyme pairs selected, 18 pairs showed coverage breadth > 20% (Table S8). The enzyme pair *HinfI* and *HpyCH4IV* (New England Biolabs, Ipswich, USA) was selected based on the number of fragments, genome coverage, availability and cost per sample. This pair was predicted to generate 840,663 fragments with different ends within the 100–600 bp range, which covered 24.6% of the genome. The suitability of the selected restriction enzyme pair was confirmed by digesting 400 ng of genomic DNA using 5 Units of each restriction enzyme and NEB CutSmart™ buffer (10×) (New England Biolabs (NEB), Ipswich, USA). The reaction was incubated for 4 h at 37 °C and the results were visualised using the LabChip GX Touch 24 (PerkinElmer, Waltham, USA).

#### Adapter design

Adapters for the ddRAD protocol were designed based on the adapters and indexed primers used by Peterson et al. (2012). Barcoded adapters were modified to create a complementary overhang for the *HpyCH4IV* restriction enzyme, while the common adapter was altered to create a complementary overhang for the frequent-cutter *HinfI*. The adapters were assembled by annealing 10 μM forward and reverse strand oligos as described in Peterson et al. (2012). The adapter concentrations to be used in the ligation step for the barcoded and common adapters were determined using the molarity calculator described by Peterson et al. (2012). The average distance between the restriction sites required for the calculation was calculated using the estimated *in silico* digestion results obtained using the IgCoverage package.

#### Sequencing library preparation

Genomic DNA was extracted from leaf material using the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. DNA concentrations were quantified using the broad range Qubit 3.0 Fluorometric assay (Invitrogen, Carlsbad, USA), while DNA quality was assessed with the LabChip GX Touch 24 (PerkinElmer, Waltham, USA). Modified versions of the Peterson et al. (2012) and Clark et al. (2014) protocols were used to construct the ddRAD libraries. The extracted gDNA was normalised at 50 ng/μL and 200 ng of DNA of each sample was digested in a 20 μL reaction volume containing restriction enzyme/s and recommended buffer. Digestion for the preparation of the ddRAD libraries was carried out using *HpyCH4IV* (5 U) and *HinfI* (5 U) in NEB CutSmart™ buffer. The reaction was incubated at 37 °C for 4 h.

The digested DNA was ligated respectively to the unique barcoded and common adapters using T4 DNA ligase (Thermo Invitrogen, Carlsbad, USA). An 18 μL master mix containing ligation buffer, 200 U of T4 ligase and the common adapter was prepared and added directly to the 20 μL digest reaction, after which the individual barcoded adapters were added. The reaction was incubated at 22 °C for 2 h, followed by 65 °C for 20 min, then cooled to 4 °C at a ramp rate of 2 °C per 90 s. To accommodate variation in DNA concentration and quality the samples were not pooled after ligation but individually purified and double size selected to enable enrichment of fragments between 250 and 800 bp. The total volume of the samples was adjusted to 100 μL by adding 60 μL of nuclease free water. Double size selection was carried out by adding 50 μL of a 1:4 (0.5X) mixture of AMPure XP Beads (Beckman Coulter, Brea, USA) to PEG buffer (20% PEG w/v, 2.5 M NaCl) to remove fragments > 800 bp. The supernatant was transferred to 20 μL of a 1:1 (0.7X) Ampure XP bead to PEG buffer mixture to collect fragments > 250 bp. The beads were washed using 80% ethanol and the fragments eluted in 30 μL nuclease free water.

A 10 μL aliquot of the size selected DNA was used for PCR amplification. A 40 μL master mix of Phusion Hot-Start High-Fidelity Master Mix Polymerase (Thermo Fisher Scientific, Waltham, USA) and the Forward (0.5 μM) and Reverse primers (0.5 μM) was prepared. Samples were amplified at 98 °C for 2 min, followed by 15 cycles of 98 °C for 15 s, 62 °C for 30 s, 72 °C for 30 s, and a final extension for 5 min at 72 °C. Amplified libraries were cleaned using 1.5X Ampure XP Beads to sample volume to remove primer dimers. The resulting library DNA concentrations were determined using the High Sensitivity (HS) Qubit 3.0 Fluoro metric assay. Library quality and fragment size distribution were visualised using the LabChip GX Touch 24. Equimolar amounts of the prepared libraries were pooled and loaded on a 1.5% agarose gel to enrich and select fragments between 300 and 700 bp. The DNA was recovered using the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany). The final library concentration, quality and size distribution were assessed again and adjusted to 20 nM DNA using 10 nM Tris Buffer (pH 8.5, 0.1% Tween 20, 10 nM). The final libraries were sent to the KCCG Core facility at the Garvan Institute for Medical Research for paired end sequencing on the HiSeq X Ten platform.

To help detect potentially causal variants contributing to flowering time in the  $F_2$  population, the 16 individuals that flowered in less than 127 days and the 19 individuals that flowered after over 137 days were sequenced at low coverage (Table S9). Genomic DNA for low coverage whole genome sequencing was extracted as described above. Sequencing libraries were prepared using the



Illumina TruSeq® Nano DNA Library Prep kit based on the guidelines provided by the manufacturer. Equimolar amounts of uniquely indexed libraries were pooled and sent for paired end sequencing on the HiSeq 2500 platform at the Australian Genome Research Facility.

#### Adapter trimming and quality trimming

The Illumina bcl2fastq 2.20.0.422 pipeline [64] was used to convert base call files to FASTQ format. Paired-end ddRAD sequencing reads were demultiplexed using sabre 1.0 [65] with a single mismatch allowed. Raw FASTQ files were trimmed of adapter sequences and low quality bases with Trimmomatic 0.36 [66]. For adapter trimming, a maximum mismatch score of 2 was used for the adapter sequence, together with a palindrome clip score threshold of 30 and a simple clip score threshold of 10. Low quality bases with a Phred+33 score below 3 were trimmed from the start and end of the read. Sliding window trimming was carried out using a 4-base wide window, trimming the bases when the average quality per base was below 15. The Illumina TruSeq3-PE adapter list provided with Trimmomatic was used for adapter trimming. All reads with fewer than 36 bases after trimming were discarded. Reads that were unpaired after trimming were also discarded. Following read pre-processing, the untrimmed and trimmed reads were analysed using the diagnostic tool fastQC [67]. The results of fastQC for each sample were then aggregated and summarised using multiQC [68]. The multiQC report was used to verify that adapters had been removed and read quality was high.

#### Aligning sequencing reads

Trimmed reads were mapped using BWA 0.7.17 with the BWA-MEM algorithm [69] to the *B. napus* Darmorbzh v8.1 assembly [63] using default parameters. After alignment, SAM files were converted to BAM format using SAMtools 1.8 [70]. Unmapped reads, supplementary alignments and reads with a mapping quality below 20 were discarded. This filter removes multi-mapping reads, which commonly occur in *B. napus* due to the homeologous regions of its polyploid genome. The mapping results were analysed using SAMtools stats and mosdepth 0.2.3 [71]. The number of ddRAD loci were calculated from mosdepth per-base output using BEDTools 2.26.0 [72] to merge neighboring loci within 100 bp.

#### Calling single nucleotide variants

Variants were called using GATK 3.6 [73]. First, BAM alignments were indexed using SAMtools, then HaplotypeCaller was used to call SNPs for each individual sample. Genotyping was carried out using GATK GenotypeGVCF using default setting with auto index

creation and locking when reading rods disabled. Results per chromosome were merged using GATK CatVariants. Variants were filtered using VCFtools 0.1.15 [74]. Indels and multiallelic SNPs were excluded (`--remove-indels --max-alleles 2 --min-alleles 2`). Before filtering SNPs, individuals with >0.9 missing genotypes were removed. To reduce the rate of heterozygous alleles incorrectly called as homozygous alleles due to insufficient read depth, genotypes with a depth <5 (`--minDP 5`) were set to missing. SNPs were discarded if they displayed a minor allele frequency <0.05 (`--maf 0.05`) or when genotypes were not present in >80% of all individuals (`--max-missing 0.8`). Genotype discordance was calculated with snpEff 4.3 t [75] using the duplicate samples for the parents (spring type BnSOSR with  $n = 4$ , winter type BnWOSR with  $n = 3$ ) with pairwise comparisons of genotypes for SNPs passing the above filters. Heterozygosity per individual was calculated using VCFtools.

The parentage assignment and filtering of distorted SNPs was carried out using the custom script vcf2gt.py [76], which uses cyvcf2 0.8.0 [77] to parse VCF files. A chi-square test implemented in scipy 1.2.0 [78] was carried out to identify and discard SNPs with significant segregation distortion ( $p < 0.01$ ) based on the expected  $F_2$  segregation ratio of 1:2:1. Further filtering removed SNPs that were heterozygous in at least one of the parents or that were not polymorphic between the parents. The script also converted the SNPs in VCF format to a genotype matrix in AB format (A: homozygous allele from Parent 1; B: homozygous allele from Parent 2; AB: heterozygous allele; -: missing allele).

Genotypes were imputed and corrected using Genotype-Corrector 1.0 [10], which is well-suited for  $F_2$  populations with moderate to low genotype missingness like the one analysed here. This software uses the order of SNPs on the genome reference and a sliding-window approach to impute and correct genotypes based on neighboring genotypes in  $F_2$  populations. Before correction, up to eight consecutive homozygous SNPs within 150 bp genomic intervals in heterozygous regions were binned into a single SNP with Genotype-Corrector qc\_hetero. This helps prevent miscorrection of heterozygous genotypes to homozygous genotypes when using the sliding window approach. With the 20% missing SNPs used here, the expected accuracy of Genotype-Corrector is >95%, based on empirical testing in crop mapping populations [10].

#### Linkage mapping

Linkage mapping was carried out using the MSTMap algorithm [79] implemented in the R package ASMap [80]. The sum of recombination events objective function was used to find the optimal sequence of loci, and

the  $p$ -value threshold for clustering markers into linkage groups was set to  $1e^{-23}$  based on evaluation of a range of values from  $1e^{-14}$  to  $1e^{-29}$ . The kosambi distance function was used to estimate genetic distances between SNPs, and rare recombination events were treated as errors (*detectBadData = True*). Linkage groups were assigned chromosome names based on marker positions on the reference genome. Small linkage groups that did not represent an entire chromosome, were merged with chromosomal linkage groups and genetic distances recalculated if an unambiguous assignment was possible using physical marker positions. Linkage groups with < 7 markers were discarded. The estimated pairwise recombination fractions between markers were calculated using the *rqtl* function *plotRF* [38]. Recombination fractions were used to identify outlier markers that are not in LD with neighboring markers and to manually correct marker order using physical marker positions. A total of 95 outlier markers were removed from further analysis. To ascertain the quality of the genetic map, the correlation between marker order on the genetic map and the reference genome was calculated using a Spearman's rank correlation test in R. Crossover frequency was estimated using the *rqtl* function *locateXO* and a custom python script *crossover.py* [76].

### QTL mapping

QTL mapping was conducted with *rqtl* scanone using a single QTL model and the non-parametric model for flowering and budding time because these traits did not follow a normal distribution. Genome-wide significance thresholds for logarithm of the odds scores (LOD) were estimated using a permutation test with 1000 iterations [81]. The 1.5-LOD drop interval of each QTL position was estimated using *rqtl* *lodint*. The percentage of variance explained for each QTL was calculated using *rqtl* *fitqtl* with Haley-Knott regression. The custom interactive R script used for QTL mapping in Rstudio 1.1.456 [82] was based on the *rqtl* manual [38].

### Identification of flowering time genes

A total of 306 *Arabidopsis* flowering time (FT) genes from the FLOWERING Interactive Database [83] was downloaded from The Arabidopsis Information Resource [84]. These genes include homologs of the known *B. napus* flowering time genes. BLAST+ 2.2.29 [85, 86] analysis of the FT genes against the reference genome was carried out to find gene homologs using a cut-off of  $1e^{-6}$  (following [87]). Overlapping hits were merged using BEDtools. The gene names in the v81 annotation [63] were identified using BEDtools to obtain gene annotations overlapping the BLAST alignments.

### Variant analysis in early-flowering and late-flowering F<sub>2</sub> individuals with low coverage whole genome sequencing

Reads were aligned and variants called using the approach described above for the ddRAD sequencing data. To help detect potentially causal variants, less stringent VCFtools filters were applied to exclude only variants with high missingness or low minor allele counts (*--max-missing 0.25 --mac 5*). To detect candidate variants that were segregating between the late-flowering and the early-flowering samples, we used a simple set of thresholds. First, we required four or more genotype calls for each of the two groups. Secondly, we required over 70% of the genotype calls within each group to be consistent. Finally, the consensus genotype calls had to differ between groups.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-020-02756-y>.

**Additional file 1: Table S1.** Linkage map constructed using ASMap with corrected and imputed ddRAD markers derived from an BnSOSR x BnWOSR F<sub>2</sub> population. **Table S2.** Mean markers and linkage map sizes across maps generated in this study. Abbreviations: doubled haploid (DH), recombinant inbred lines (RIL), amplified fragment length polymorphisms (AFLP), Brassica 60 K genotyping array (Brassica 60 k). **Table S3.** Phenotypic variation in flowering time and time to bud in the F<sub>2</sub> mapping population and the parental lines. Phenotypes of spring type BnSOSR and winter type BnWOSR parental lines are shown as means of all replicates of each type ( $n = 4$ ,  $n = 3$ , respectively). **Table S4.** QTL identified using genome-wide single QTL scans for budding time (B) and flowering time (FT). For each QTL, the phenotypic variance explained (PVE) and additive effect (AE) are shown. **Table S5.** Summary of variants impacting candidate flowering time genes on chromosome C2. The QTL regions ranges from 4,345,729 to 24,916,709. **Table S6.** Exon coverage of candidate genes on chromosome 2 generated by whole genome skim sequencing of 16 early-flowering and 19 late-flowering F<sub>2</sub> individuals. **Table S7.** Summary of intergenic variants impacting candidate genes and segregating between early-flowering and late-flowering F<sub>2</sub> individuals. **Table S8.** In silico double enzyme digest analysis. Frag: Total number of fragments; FragDe: total number of different end fragments; LengthDeFrag: total length in bases of different end fragments; De-Frag100–600: total number of different end fragments between 100 bases and 600 bases; LengthDeFrag100–600: total length in bases of different end fragments between 100 bases and 600 bases; % Coverage: percentage of the reference genome covered by fragments between 100 bases and 600 bases. **Table S9.** Phenotype data for flowering time and budding time in the mapping population and parents. **Table S10.** Sample list for the 194 progeny and 7 parental individuals used in this study.

**Additional file 2: Fig. S1.** Segregation distorted loci across all chromosomes. The significance threshold ( $p > 0.01$ ) is shown as a red dotted line. **Fig. S2.** Comparison of linkage group sizes in an uncorrected genetic map and a corrected genetic map. Corrected linkage groups are aligned centrally to uncorrected groups. **Fig. S3.** Physical (x-axis) and genetic (y-axis) marker positions on all chromosomes in Mb and cM respectively. Spearman's rank correlation test result shown in the top left corner of each plot. **Fig. S4.** Effect plot for the budding time (left) and flowering time (right) QTL on C2 at positions 4,673,904 and 4,655,461 respectively. The 'AA' genotype is BnSOSR and the 'BB' genotype is BnWOSR.

### Abbreviations

RAD: Restriction-site associated DNA sequencing; ddRAD: Double digest RAD; GBS: Genotyping-by-sequencing; LOD: Logarithm of the odds; VCF: Variant

call format; QTL: Quantitative trait locus; SNP: Single nucleotide polymorphism; MAF: Minor allele frequency; OSR: Oilseed rape; WOSR: Winter OSR; SOSR: Summer OSR

#### Authors' contributions

DE, JB and SR conceived and supervised the project. ASE grew the plants and carried out phenotyping. ASE, AP and DP prepared the sequencing libraries. ASE wrote the phenotyping and library preparation sections of the Materials and Methods. AS carried out the analyses and drafted the manuscript. AS, DE, JB and SR wrote the final manuscript and all authors provided comments and edits. All authors read and approved the final manuscript.

#### Funding

This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. A.S. was supported by an IPRS awarded by the Australian government. This work was supported by the Australia Research Council (Projects LP160100030, LP140100537 and LP130100925) and BASF. Bayer transferred this work with the CropScience business to BASF in 2018. Funding bodies were not involved in the design of the study, collection of data, analysis, interpretation of data or in writing the manuscript.

#### Availability of data and materials

All sequences have been deposited in SRA (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA640838>) with individual accessions listed in Table S10. Scripts used to for this study are available at [https://github.com/ascheben/bn\\_gbs/](https://github.com/ascheben/bn_gbs/). Genotype data are available via figshare (<https://doi.org/10.6084/m9.figshare.13139378.v1>). The plant material used in this study is available from BASF but restrictions apply to its availability and so it is not publicly available.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

BASF supported this work and employs SR.

#### Author details

<sup>1</sup>School of Biological Sciences and Institute of Agriculture, The University of Western Australia, Perth, WA, Australia. <sup>2</sup>Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY 11724, USA. <sup>3</sup>BASF Agricultural Solutions Belgium NV, BASF Innovation Center Gent, Technologiepark-Zwijnaarde 101, 9052 Ghent, Belgium.

Received: 16 July 2020 Accepted: 25 November 2020

Published online: 07 December 2020

#### References

- K. R. Andrews, J. M. Good, M. R. Miller, G. Luikart, P. A. Hohenlohe (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet* 17, 81–92. <https://doi.org/10.1038/nrg.2015.28>.
- K. Voss-Fels, R. J. Snowdon (2016) Understanding and utilizing crop genome diversity via high-resolution genotyping. *Plant Biotechnol J* 14, 1086–1094. <https://doi.org/10.1111/pbi.12456>.
- J. A. Poland, T. W. Rife (2012) Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5, 92–102. <https://doi.org/10.3835/plantgenome2012.05.0005>.
- S. Deschamps, V. Llaça, G. D. May (2012) Genotyping-by-sequencing in plants. *Biology* 1, 460–483. <https://doi.org/10.3390/biology1030460>.
- A. Scheben, J. Batley, D. Edwards (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotech J* 15, 149–161. <https://doi.org/10.1111/pbi.12645>.
- X. Xu, G. Bai (2015) Whole-genome resequencing: changing the paradigms of SNP detection, molecular mapping and gene discovery. *Mol Breed* 35, 33–33. <https://doi.org/10.1007/s11032-015-0240-6>.
- Y. B. Fu, G. W. Peterson, Y. Dong (2016) Increasing genome sampling and improving SNP genotyping for Genotyping-by-sequencing with new combinations of restriction enzymes. *Genes Genom Genet* 6, 845. <https://doi.org/10.1534/g3.115.025775>.
- C. A. Frago, C. Heffelfinger, H. Y. Zhao, S. L. Dellaporta (2016) Imputing genotypes in biallelic populations from low-coverage sequence data. *Genetics* 202, 487–495. <https://doi.org/10.1534/genetics.115.182071>.
- K. Swarts, H. Li, J. A. Romero Navarro, D. An, M. C. Romay, S. Hearne, C. Acharya, J. C. Glaubitz, S. Mitchell, R. J. Elshire (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *Plant Genome* 7, 1–12. <https://doi.org/10.3835/plantgenome2014.05.0023>.
- C. Miao, J. Fang, D. Li, P. Liang, X. Zhang, J. Yang, J. C. Schnable, H. Tang (2018) Genotype-corrector: improved genotype calls for genetic mapping in F2 and RIL populations. *Sci Rep* 8, 10088. <https://doi.org/10.1038/s41598-018-28294-0>.
- T. Furuta, M. Ashikari, K. K. Jena, K. Doi, S. Reuscher (2017) Adapting genotyping-by-sequencing for rice F2 populations. *Genes Genom Genet* 7, 881–893. <https://doi.org/10.1534/g3.116.038190>.
- J. G. Uitendewilgen, A. M. Wolters, B. D'Hoop, T. J. Borm, R. G. Visser, H. J. van Eck (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS One* 8, e62355. <https://doi.org/10.1371/journal.pone.0062355>.
- H. Sonah, L. O'Donoghue, E. Cober, I. Rajcan, F. Belzile (2015) Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol J* 13, 211–221. <https://doi.org/10.1111/pbi.12249>.
- M. M. Malmberg, L. W. Pembleton, R. C. Baillie, M. C. Drayton, S. Sudheesh, S. Kaur, H. Shinokura, P. Verma, G. C. Spangenberg, H. D. Daetwyler et al. (2018) Genotyping-by-sequencing through transcriptomics: implementation in a range of crop species with varying reproductive habits and ploidy levels. *Plant Biotechnol J* 16, 877–889. <https://doi.org/10.1111/pbi.12835>.
- I. Bancroft, C. Morgan, F. Fraser, J. Higgins, R. Wells, L. Clissold, D. Baker, Y. Long, J. L. Meng, X. W. Wang et al. (2011) Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. *Nat Biotechnol* 29, 762–766. <https://doi.org/10.1038/nbt.1926>.
- B. Li, J. Gao, J. Chen, Z. Wang, W. Shen, B. Yi, J. Wen, C. Ma, J. Shen, T. Fu et al. (2019) Identification and fine mapping of a major locus controlling branching in *Brassica napus*. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-019-03506-x>.
- F. Sun, J. Liu, W. Hua, X. Sun, X. Wang, H. Wang (2016) Identification of stable QTLs for seed oil content by combined linkage and association mapping in *Brassica napus*. *Plant Sci* 252, 388–399. <https://doi.org/10.1016/j.plantsci.2016.09.001>.
- X. Chen, X. Li, B. Zhang, J. Xu, Z. Wu, B. Wang, H. Li, M. Younas, L. Huang, Y. Luo et al. (2013) Detection and genotyping of restriction fragment associated polymorphisms in polyploid crops with a pseudo-reference sequence: a case study in allotetraploid *Brassica napus*. *BMC Genomics* 14, 346–346. <https://doi.org/10.1186/1471-2164-14-346>.
- A. Abe, S. Kosugi, K. Yoshida, S. Natsume, H. Takagi, H. Kanzaki, H. Matsumura, K. Yoshida, C. Mitsuoaka, M. Tamiru et al. (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol* 30, 174–178. <https://doi.org/10.1038/nbt.2095>.
- D. Edwards, J. Batley, R. J. Snowdon (2013) Accessing complex crop genomes with next-generation sequencing. *Theor Appl Genet* 126, 1–11. <https://doi.org/10.1007/s00122-012-1964-x>.
- M. Tadege, C. C. Sheldon, C. A. Helliwell, P. Stoutjesdijk, E. S. Dennis, W. J. Peacock (2001) Control of flowering time by *FLC* orthologues in *Brassica napus*. *Plant J* 28, 545–553. <https://doi.org/10.1046/j.1365-3113X.2001.01182.x>.
- I. A. Parkin, S. M. Gulden, A. G. Sharpe, L. Lukens, M. Trick, T. C. Osborn, D. J. Lydiate (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171, 765–781. <https://doi.org/10.1534/genetics.105.042093>.
- L. S. Robert, F. Robson, A. Sharpe, D. Lydiate, G. Coupland (1998) Conserved structure and function of the *Arabidopsis* flowering time gene *CONSTANS* in *Brassica napus*. *Plant Mol Biol* 37, 763–772. <https://doi.org/10.1023/A:1006064514311>.
- H. J. Jian, A. X. Zhang, J. Q. Ma, T. Y. Wang, B. Yang, L. S. Shuang, M. Liu, J. N. Li, X. F. Xu, A. H. Paterson et al. (2019) Joint QTL mapping and transcriptome sequencing analysis reveal candidate flowering time genes in *Brassica napus* L. *BMC Genomics* 20, 21. <https://doi.org/10.1186/s12864-018-5356-8>.

25. B. J. Li, W. G. Zhao, D. R. Li, H. B. Chao, X. P. Zhao, N. Ta, Y. H. Li, Z. B. Guan, L. X. Guo, L. N. Zhang et al. (2018) Genetic dissection of the mechanism of flowering time based on an environmentally stable and specific QTL in *Brassica napus*. *Plant Sci* 277, 296–310. <https://doi.org/10.1016/j.plantsci.2018.10.005>.
26. H. Raman, R. Raman, P. Eckermann, N. Coombes, S. Manoli, X. X. Zou, D. Edwards, J. L. Meng, R. Prangnell, J. Stiller et al. (2013) 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>.
27. Y. S. Shen, Y. Xiang, E. S. Xu, X. H. Ge, Z. Y. Li (2018) Major co-localized QTL for plant height, branch initiation height, stem diameter, and flowering time in an alien introgression derived *Brassica napus* DH population. *Front Plant Sci* 9, 390. <https://doi.org/10.3389/fpls.2018.00390>.
28. L. P. Xu, K. N. Hu, Z. Q. Zhang, C. Y. Guan, S. Chen, W. Hua, J. N. Li, J. Wen, B. Yi, J. X. Shen et al. (2016) Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.). *DNA Res* 23, 43–52. <https://doi.org/10.1093/dnares/dsv035>.
29. M. N. Nelson, R. Rajasekaran, A. Smith, S. Chen, C. P. Beeck, K. H. M. Siddique, W. A. Cowling (2014) 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>.
30. H. Raman, R. Raman, Y. Qiu, A. S. Yadav, S. Sureshkumar, L. Borg, M. Rohan, D. Wheeler, O. Owen, I. Menz et al. (2019) GWAS hints at pleiotropic roles for *FLOWERING LOCUS T* in flowering time and yield-related traits in canola. *BMC Genomics* 20, 636. <https://doi.org/10.1186/s12864-019-5964-y>.
31. Osborn TC, Kole C, Parkin IA, Sharpe AG, Kuiper M, Lydiat DJ, Trick M. Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. *Genetics*. 1997;146:1123–9.
32. K. P. Starmer, J. Brown, J. B. Davis (1998) Heterosis in spring canola hybrids grown in northern Idaho. *Crop Sci* 38, 376–380. <https://doi.org/10.2135/cropsci1998.0011183X003800020018x>.
33. H. Cutforth, B. McConkey, S. Brandt, Y. Gan, G. Lafond, S. Angadi, D. Judiesch (2009) Fertilizer N response and canola yield in the semiarid Canadian prairies. *Can J Plant Sci* 89, 501–503. <https://doi.org/10.4141/Cjps08128>.
34. H. Alipour, G. H. Bai, G. R. Zhang, M. R. Bihanta, V. Mohammadi, S. A. Peyghambari (2019) Imputation accuracy of wheat genotyping-by-sequencing (GBS) data using barley and wheat genome references. *PLoS One* 14, e0208614. <https://doi.org/10.1371/journal.pone.0208614>.
35. X. D. Wang, K. J. Yu, H. G. Li, Q. Peng, F. Chen, W. Zhang, S. Chen, H. L. Maolong, J. F. Zhang (2015) High-density SNP map construction and QTL identification for the apetalous character in *Brassica napus* L. *Front Plant Sci* 6, 1164. <https://doi.org/10.3389/fpls.2015.01164>.
36. W. Ecke, A. Kampouridis, K. Ziese-Kubon, A. C. Hirsch (2015) Identification and genetic characterization by high-throughput SNP analysis of introvarietal substitution lines of rapeseed (*Brassica napus* L.) with enhanced embryogenic potential. *Theor Appl Genet* 128, 587–603. <https://doi.org/10.1007/s00122-015-2455-7>.
37. W. E. Clarke, E. E. Higgins, J. Plieske, R. Wieseke, C. Sidebottom, Y. Khedikar, J. Batley, D. Edwards, J. L. Meng, R. Y. Li et al. (2016) A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theor Appl Genet* 129, 1887–1899. <https://doi.org/10.1007/s00122-016-2746-7>.
38. K. W. Broman, H. Wu, S. Sen, G. A. Churchill (2003) R/QTL: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. <https://doi.org/10.1093/bioinformatics/btg112>.
39. Dupuis J, Siegmund D. Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics*. 1999;151:373–86.
40. Mangin B, Goffinet B, Rebai A. Constructing confidence intervals for QTL location. *Genetics*. 1994;138:1301–8.
41. F. Turck, F. Fornara, G. Coupland (2008) Regulation and identity of florigen: *FLOWERING LOCUS T* moves center stage. *Annu Rev Plant Biol* 59, 573–594. <https://doi.org/10.1146/annurev.arplant.59.032607.092755>.
42. S. D. Michaels, R. M. Amasino (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11, 949–956. <https://doi.org/10.1105/tpc.11.5.949>.
43. G. G. Simpson, P. P. Dijkwel, V. Quesada, I. Henderson, C. Dean (2003) *FY* is an RNA 3' end-processing factor that interacts with *FCA* to control the *Arabidopsis* floral transition. *Cell* 13, 777–787. [https://doi.org/10.1016/S0092-8674\(03\)00425-2](https://doi.org/10.1016/S0092-8674(03)00425-2).
44. D. Z. Wu, Z. Liang, T. Yan, Y. Xu, L. J. Xuan, J. Tang, G. Zhou, U. Lohwasser, S. J. Hua, H. Y. Wang et al. (2019) Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. *Mol Plant* 12, 30–43. <https://doi.org/10.1016/j.molp.2018.11.007>.
45. R. Raman, S. Diffey, J. Carling, R. B. Cowley, A. Kilian, D. J. Luckett, H. Raman (2016) Quantitative genetic analysis of grain yield in an Australian *Brassica napus* doubled-haploid population. *Crop Pasture Sci* 67, 298–307. <https://doi.org/10.1071/Cp15283>.
46. Feng W, Michaels SD. Dual roles for *FY* in the regulation of *FLC*. *Plant Signal Behav*. 2011;6:703–5 <https://dx.doi.org/10.4161/psb.6.5.15052>.
47. J. Wang, C. J. Hopkins, J. N. Hou, X. X. Zou, C. N. Wang, Y. Long, S. Kurup, G. J. King, J. L. Meng (2012) Promoter variation and transcript divergence in Brassicaceae lineages of *FLOWERING LOCUS T*. *PLoS One* 7, e47127. <https://doi.org/10.1371/journal.pone.0047127>.
48. D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59. <https://doi.org/10.1038/nature07517>.
49. T. Maruki, M. Lynch (2017) Genotype calling from population-genomic sequencing data. *Genes Genom Genet* 7, 1393–1404. <https://doi.org/10.1534/g3.117.039008>.
50. M. M. Malmberg, D. M. Barbulescu, M. C. Drayton, M. Shinozuka, P. Thakur, Y. O. Ogaji, G. C. Spangenberg, H. D. Daetwyler, N. O. I. Cogan (2018) Evaluation and recommendations for routine genotyping using skim whole genome re-sequencing in canola. *Front Plant Sci* 9, 1809. <https://doi.org/10.3389/fpls.2018.01809>.
51. J. S. Brouard, B. Boyle, E. M. Ibeagha-Awemu, N. Bissonnette (2017) Low-depth genotyping-by-sequencing (GBS) in a bovine population: strategies to maximize the selection of high quality genotypes and the accuracy of imputation. *BMC Genet* 18, 32. <https://doi.org/10.1186/s12863-017-0501-y>.
52. F. Henning, H. J. Lee, P. Franchini, A. Meyer (2014) Genetic mapping of horizontal stripes in Lake Victoria cichlid fishes: benefits and pitfalls of using RAD markers for dense linkage mapping. *Mol Ecol* 23, 5224–5240. <https://doi.org/10.1111/mec.12860>.
53. P. Bajgain, M. N. Rouse, J. A. Anderson (2016) Comparing genotyping-by-sequencing and single nucleotide polymorphism chip genotyping for quantitative trait loci mapping in wheat. *Crop Sci* 56, 232–248. <https://doi.org/10.2135/cropsci2015.06.0389>.
54. W. Hussain, P. S. Baenziger, V. Belamkar, M. J. Guttieri, J. P. Venegas, A. Easterly, A. Sallam, J. Poland (2017) Genotyping-by-sequencing derived high-density linkage map and its application to QTL mapping of flag leaf traits in bread wheat. *Sci Rep* 7, 16394. <https://doi.org/10.1038/s41598-017-16006-z>.
55. C. Sautenac, D. Y. Jiang, S. C. Wang, E. Akhunov (2013) Sequence-based mapping of the polyploid wheat genome. G3 - genes Genom. Genet. 3, 1105–1114. <https://doi.org/10.1534/g3.113.005819>.
56. F. Q. Yu, X. G. Zhang, G. Peng, K. C. Falk, S. E. Strelkov, B. D. Gossen (2017) Genotyping-by-sequencing reveals three QTL for clubroot resistance to six pathotypes of *Plasmodiophora brassicae* in *Brassica rapa*. *Sci Rep* 7, 4516. <https://doi.org/10.1038/s41598-017-04903-2>.
57. A. S. Mason, E. E. Higgins, R. J. Snowdon, J. Batley, A. Stein, C. Werner, I. A. Parkin (2017) A user guide to the *Brassica* 60K Illumina Infinium SNP genotyping array. *Theor Appl Genet* 130, 621–633. <https://doi.org/10.1007/s00122-016-2849-1>.
58. N. A. Tinker, S. M. Chao, G. R. Lazo, R. E. Oliver, Y. F. Huang, J. A. Poland, E. N. Jellen, P. J. Maughan, A. Kilian, E. W. Jackson (2014) A SNP genotyping Array for Hexaploid oat. *Plant Genome* 7. <https://doi.org/10.3835/plantgenome2014.03.0010>.
59. H. X. Hong, L. Xu, J. Liu, W. D. Jones, Z. Q. Su, B. T. Ning, R. Perkins, W. G. Ge, K. Miclaus, L. Zhang et al. (2012) Technical reproducibility of genotyping SNP arrays used in genome-wide association studies. *PLoS One* 7, e44483. <https://doi.org/10.1371/journal.pone.0044483>.
60. J. C. Richter, C. Mollers (2018) Genetic variation for vernalization requirement of winter oilseed rape. *Acta Hort* 1202, 87–91. <https://doi.org/10.17660/ActaHortic.2018.1202.13>.
61. M. Rapacz, A. Markowski (1999) Winter hardiness, frost resistance and vernalization requirement of European winter oilseed rape (*Brassica napus* var. *oleifera*) cultivars within the last 20 years. *J Agron Crop Sci* 183, 243–253. <https://doi.org/10.1046/j.1439-037x.1999.00346.x>.
62. A. Kassambara (2018) 'ggplot2' Based Publication Ready Plots. Accessed: 19 March 2018. <https://github.com/kassambara/ggpubr>.



63. P. E. Bayer, B. Hurgobin, A. A. Golicz, C. K. Chan, Y. Yuan, H. Lee, M. Renton, J. Meng, R. Li, Y. Long et al. (2017) Assembly and comparison of two closely related *Brassica napus* genomes. *Plant Biotechnol J* 15, 1602–1610. <https://doi.org/https://doi.org/10.1111/pbi.12742>.
64. Illumina Inc. (2017) Accessed: 7 May 2019. <http://sapac.support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html>.
65. Najoshi (2013) sabre. Accessed: 7 May 2019. <https://github.com/najoshi/sabre>.
66. A. M. Bolger, M. Lohse, B. Usadel (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
67. S. R. Andrews (2010) FastQC: a quality control tool for high throughput sequence data. Accessed: 4 September 2018. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
68. P. Ewels, M. Magnusson, S. Lundin, M. Kaller (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. <https://doi.org/https://doi.org/10.1093/bioinformatics/btw354>.
69. Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Preprint at <https://arxiv.org/abs/1303.3997>.
70. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
71. B. S. Pedersen, A. R. Quinlan (2018) Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 34, 867–868. <https://doi.org/10.1093/bioinformatics/btx699>.
72. A. R. Quinlan, I. M. Hall (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
73. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43, 491–498. <https://doi.org/10.1038/ng.806>.
74. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
75. P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Y. Lu, D. M. Ruden (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w (1118); iso-2; iso-3. *Fly* 6, 80–92. <https://doi.org/10.4161/fly.19695>.
76. A. Scheben (2019) Accessed: 7 May 2019. [https://github.com/ascheben/bn\\_gbs](https://github.com/ascheben/bn_gbs).
77. B. S. Pedersen, A. R. Quinlan (2017) cyvcf2: fast, flexible variant analysis with Python. *Bioinformatics* 33, 1867–1869. <https://doi.org/10.1093/bioinformatics/btx057>.
78. E. Jones, T. Oliphant, P. Peterson (2001) SciPy: Open Source Scientific Tools for Python. Accessed: 19 January 2019. <http://www.scipy.org/>.
79. Y. H. Wu, P. R. Bhat, T. J. Close, S. Lonardi (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet* 4, e1000212. <https://doi.org/10.1371/journal.pgen.1000212>.
80. J. Taylor, D. Butler (2017) R package ASMap: efficient genetic linkage map construction and diagnosis. *J Stat Softw* 79, 1–29. <https://doi.org/10.18637/jss.v079.i06>.
81. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics*. 1994;138:963–71.
82. R Studio Team (2015) RStudio: Integrated Development for R. Accessed: 11 November 2017. <http://www.rstudio.com/>.
83. F. Bouche, G. Lobet, P. Tocquin, C. Perilleux (2016) FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Res* 44, D1167–D1171. <https://doi.org/10.1093/nar/gkv1054>.
84. P. Lamesch, T. Z. Berardini, D. H. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D. L. Alexander, M. Garcia-Hernandez et al. (2012) The Arabidopsis information resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res* 40, D1202–D1210. <https://doi.org/10.1093/nar/gkr1090>.
85. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman (1990) Basic local alignment search tool. *J Mol Biol* 215, 403–410. [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2).
86. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden (2009) BLAST plus : architecture and applications. *BMC Bioinformatics* 10, 421. <https://doi.org/10.1186/1471-2105-10-421>.
87. J. H. Yang, K. Osman, M. Iqbal, D. J. Stekel, Z. W. Luo, S. J. Armstrong, F. C. H. Franklin (2013) Inferring the *Brassica rapa* interactome using protein-protein interaction data from *Arabidopsis thaliana*. *Front Plant Sci* 3, 297. <https://doi.org/10.3389/fpls.2012.00297>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

