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

Armin Scheben[1a], Anita Severn-Ellis[1], Dhwani Patel[1], Aneeta Pradhan[1], Stephen Rae[2], Jacqueline Batley[1], David Edwards[1*]

[1]School of Biological Sciences and Institute of Agriculture, The University of Western Australia, Perth, WA, Australia

[2]BASF Agricultural Solutions Belgium NV, BASF Innovation Center Gent, Technologiepark-Zwijnaarde 101, 9052 Gent

[a] Current affiliation: Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724, United States of America

[*] Correspondence to: David Edwards (dave.edwards@uwa.edu.au)

Abstract

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_2 mapping population was derived from a cross between an early-flowering spring type and a late-flowering winter type. 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 3,605.70 cM and compromising 14,630 single nucleotide polymorphisms (SNPs) was constructed. A quantitative trail locus (QTL) on chromosome C2 was detected in the vicinity of flowering time genes including *FT* and *FLC*. 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_2 populations. Quality filtering and genotype correction and imputation can substantially reduce these error rates and allow effective linkage mapping and QTL analysis.

Introduction

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*).

GBS has been used for marker discovery, linkage mapping and QTL analysis in a range of crops (*8-10*), including the important oilseed crop *Brassica napus* (*11-13*). Over 20 high density linkage maps have been generated for *B. napus* using RNA sequencing (*14*), the Brassica 60K genotyping array (*15, 16*) and ddRAD sequencing (*12*). 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 (*17, 18*). 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* (*19-21*). Many QTL and associated SNPs for flowering time have been detected in *B. napus* (*22-29*). Despite this 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 (*30, 31*). 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_2 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 sample ("146") was excluded from further analysis as it had fewer than one million reads after trimming. In the remaining 206 samples, 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 5,957 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 (Figure 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 (Figure 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% (Figure 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.

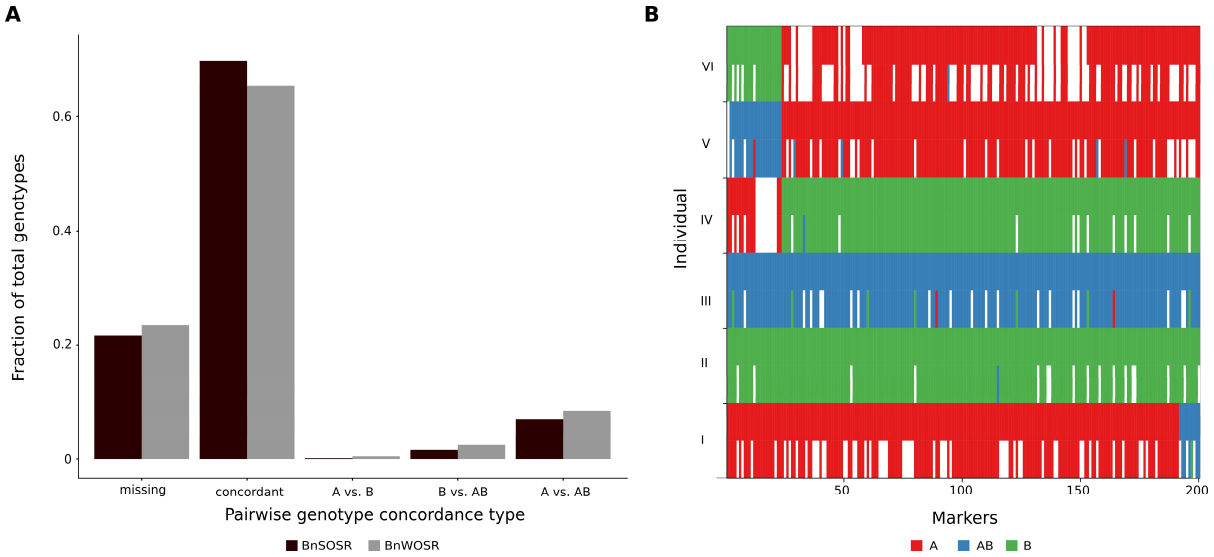

Figure 1. A) Mean pairwise genotype concordance between replicate individuals for the parental lines. The bars for the category 'missing' include all genotype pairs where at least one pair had a missing genotype call. Genotypes shown are BnSOSR as 'A', BnWOSR as 'B' and heterozygous as 'AB'. B) Comparison of corrected and uncorrected genotypes. An example of genotype correction and imputation using 200 SNPs on chromosome A3 is given for six representative individuals denoted as I-VI (samples 1, 100, 102, 103, 104, and 105). Genotypes are encoded in three colours (A: red; B: green; AB: blue) and missing markers are shown in white. For each individual, the lower row shows the genotypes before the correction and imputation step and the upper row shows the genotypes after this step.

Linkage mapping

A linkage map spanning 3,605.70 cM and compromising 14,630 markers was constructed using ASMap with the corrected and imputed markers (Figure 2 and Table S2). The A genome map was 1,771.53 cM with 8,587 markers and the C genome was 1,661.98 cM with 6,043 markers. The highest mean marker density was found on chromosome A10, with 47.90 markers per Mb.

Mean marker density per Mb was higher on the A genome (28.90) than the C genome (12.18). 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 (Figure S2). Compared to six published genetic maps that were generated with different approaches, the corrected genetic map still showed some indications of inflation (Table S3).

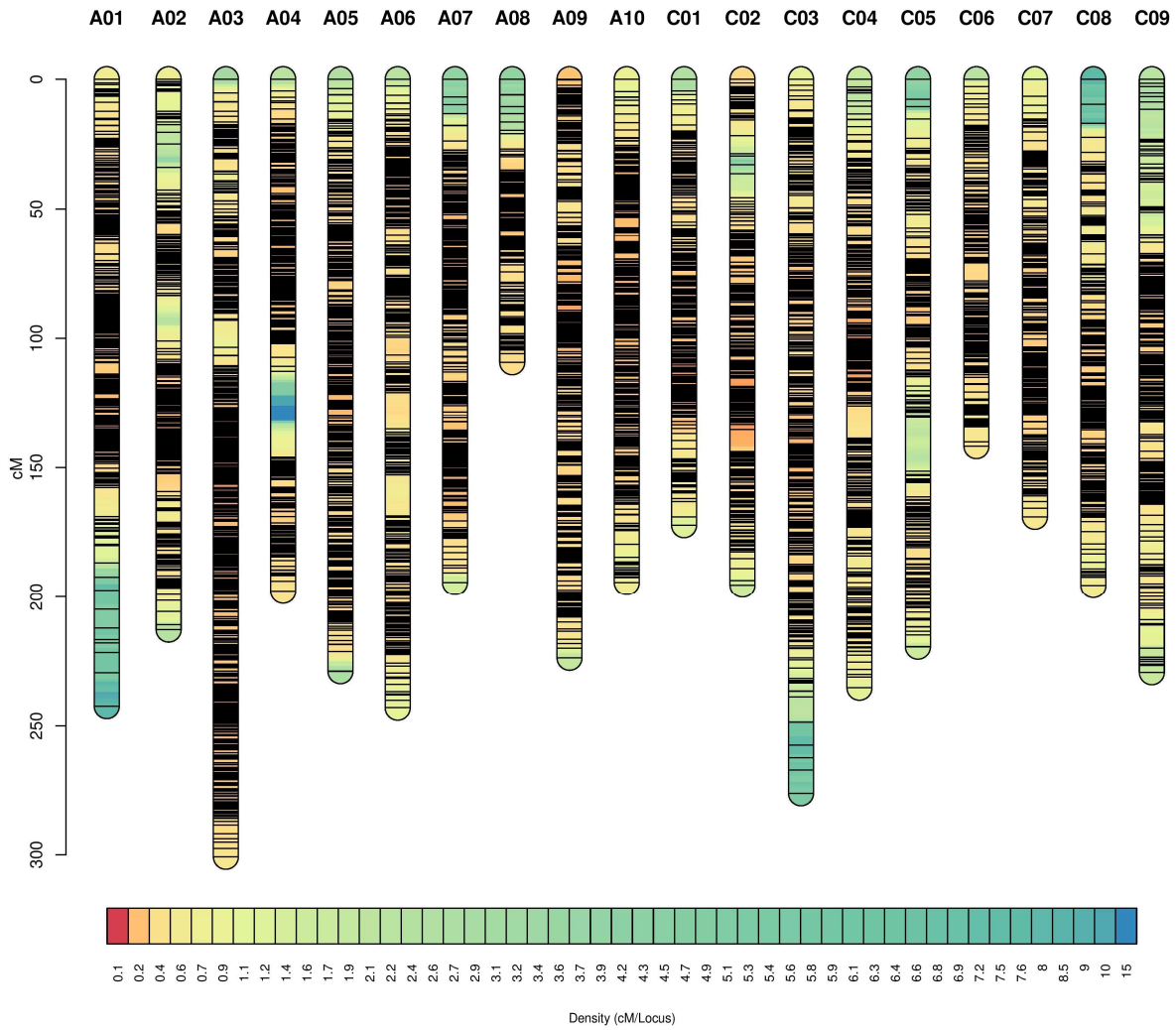

Figure 2. Genetic map based on ddRAD markers. Black lines indicate marker positions and colours along the chromosomes indicate markers density, with extremely dense regions appearing as black due to tightly clustered markers. Figure rendered using LinkageMapView 2.1.2 (*32*).

121 The correlation between genetic and physical map order provides information about the
122 consistency between the genetic map and the reference genome. The mean Spearman's rank
123 correlation for marker order per chromosome was 1.0 (Table 1). Several minor inconsistencies
124 in marker order were observed (Figure S3). All chromosomes showed mean correlations over
125 0.98. Mean individual crossover frequency per chromosome was 2.79.

126 Table 1. Summary of the genetic map. Spearman's rank correlation (rho) was calculated for the genetic
127 marker positions and the physical marker positions on the reference genomes.

Chr	Length (Mb)	Length (cM)	Number of markers	rho
A1	31.16	242.40	740	0.98
A2	31.34	185.10	630	1.00
A3	39.49	300.82	1,581	1.00
A4	23.31	58.35	748	1.00
A5	28.6	228.92	892	1.00
A6	31.9	34.53	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
A1	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
C4	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	3,605.70	14,630	-

128

129 ## QTL mapping

130 BnSOSR flowered 20 days earlier (range: 10 to 28) and went to bud 17 days earlier (range: 12
131 to 20) on average than winter type BnWOSR (Table S4). In the F_2 progeny, flowering times
132 were distributed within the parental range (Figure 3). A single significant ($p<0.05$) overlapping
133 QTL region for the traits flowering time and budding time was detected on chromosome C2
134 (Figure 3). The physical region of the QTL spanned 20.57 Mb for flowering time and 0.77 Mb
135 for time to bud (Table S5). The flowering time QTL contained 10 flowering time homologs
136 including *FT* (Table S6), and the budding time QTL contained no known flowering time
137 homologs. Carrying the BnWOSR allele at the QTL peak SNP led to an increase in the days to
138 bud and flower (Figure S4). The percentage of phenotypic variance explained for the identified
139 QTL was 9.08% for flowering time and 8.08% for budding time. Suggestive LOD peaks are
140 also noticeable on A2, A3 and C9.

141

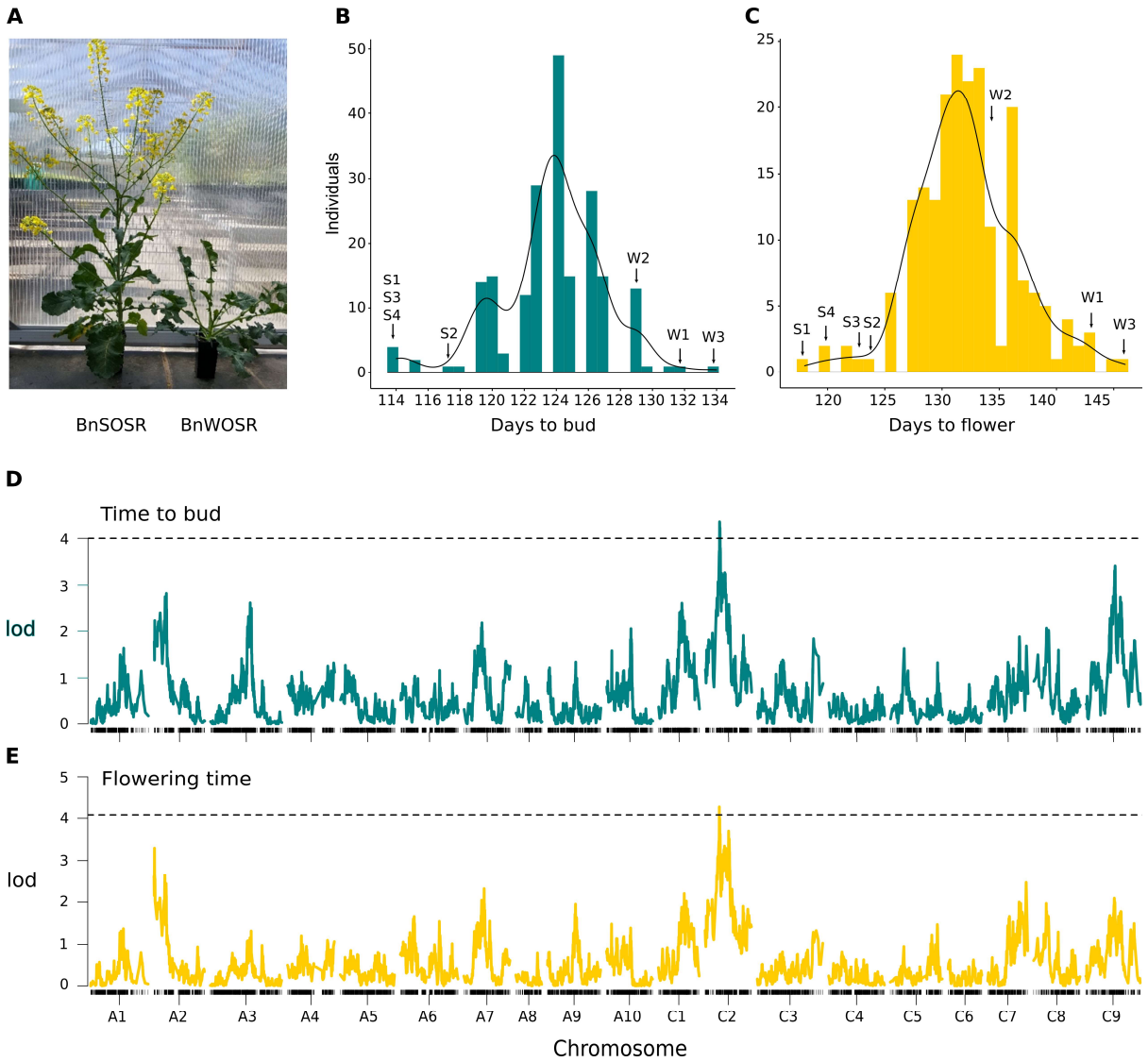

Figure 3. A) A photograph comparing development in parental individuals 126 six days after sowing. B) Histogram of mapping population phenotypes for budding time. C) Histogram of mapping population phenotypes for flowering time. Parental individuals S1-S4 of BnSOSR and W1-W3 of BnWOSR are indicated. Phenotype data is provided in Table S7. D) Genome-wide LOD plots for single QTL mapping of flowering time and budding time. Significant ($p<0.05$) thresholds based on permutation testing shown with broken lines.

150 # Discussion

151 ## Optimising genotyping-by-sequencing strategies

152 Genotyping with ddRAD was effective at generating a set of 14,630 high-quality SNPs for
153 linkage and QTL mapping. The findings of this study can help calibrate the number of
154 sequencing reads and genomic loci required for a range of research goals in *B. napus* and
155 related species. For GBS, researchers often aim to optimize the genome coverage by controlling
156 the ratio of reads sequenced to the number of loci generated. Here, the expected maximum
157 genome coverage breadth based on in silico digestion with the enzymes HpyCH4IV and HinfI
158 was 24.6%. However, the observed mean genome coverage breadth was lower at 18.03%. An
159 even greater inconsistency was reported when using HpyCH4IV and HinfI in Arabidopsis and
160 *Glycine max*, where the expected genome coverage breadth was 29.4% and 23.1% but the
161 observed experimental values were 4.45% and 3.33% respectively (*7*). The inconsistency has
162 been explained as a product of fragment size selection bias, redundant in silico loci, and
163 insufficient sequencing reads (*7*). In the *B. napus* population used here, the most important
164 factor determining coverage was the amount of sequencing reads available for each sample.
165 Indeed, 24 samples with high sequencing effort showed genome coverage breadth greater than
166 the expected 24.6% up to a maximum value of 36.68%. This is particularly surprising, as
167 coverage was calculated based only on reads aligned with high quality, which is expected to
168 substantially reduce coverage breadth. These findings suggest that, at least in *B. napus*,
169 igCoverage can underestimate the maximum achievable genome coverage breadth.

170 The high genome coverage breadth achieved using HpyCH4IV and HinfI indicates that these
171 enzymes are well suited for high-density sampling of genome-wide diversity in *B. napus*.
172 However, when sequencing effort is uneven between samples, high genome coverage breadth
173 can increase genotype missingness through allele dropout. If a locus is not sequenced in enough
174 individual samples (here the cut off was 50%) at sufficient depth, it is removed during SNP
175 calling or SNP filtering and becomes a missing genotype call. High levels of missingness are a
176 common characteristic of reduced representation sequencing (*3*) and can limit the usefulness
177 of genotype data in studies where genotype imputation is not possible (*33*). Nevertheless, as
178 shown in this study, by combining imputation with a high-density sampling of the genome, the
179 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* (*12, 34*). The linkage map constructed here was on average 1.9× larger in cM compared to six published *B. napus* linkage maps generated using different approaches with relatively similar marker densities (*12, 14, 16, 34-36*). 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_2 population, five of the compared studies used recombinant inbred lines or doubled haploid populations, which will likely suffer less 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_2 population also led to a smaller genetic map of roughly half the size (*16*). 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 (Parkin et al., 2005) on *B. napus* chromosomes A2, A3, A9, A10, C2 and C3. Here, we identified a single significant locus for flowering time and budding time on C2. These two phenotypes were significantly correlated, explaining the 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 (>1Mb), 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 (*37*). 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 (*38, 39*).

211 The QTL overlaps with a locus identified in an earlier study, which found that the 60K array
212 SNP Bn-scaff_18507_1-p889927 was associated with a QTL on C2 at position 33,936,984 on
213 v81 (position on Darmor v41: 26,548,393) explaining 6.36% of flowering time variation (*26*).
214 However, the QTL LOD peaks identified here are distant from this location. Among the known
215 flowering time genes on C2, *FT* (*40*), *FLC* (*41*) and *FY* (*42*) may have particularly substantial
216 effects. *FT* has been implicated in *B. napus* flowering time divergence (*28, 43*) and *FLC* has
217 been found to explain ~23% of flowering time variation in *B. napus* (*44*). *FY* is a suppressor of
218 the transcription factor *FLC* (*42, 45*), and lies in close physical proximity to the LOD peaks
219 identified in this study. However, the *FT* homolog on C2 is expressed at low levels in all
220 ecotypes tested (*46*), which suggests it may not be the candidate gene. Further, none of the
221 flowering time gene homologs occur within the budding time QTL region. Based on the QTL
222 results of this study, we therefore cannot conclusively determine whether the C2 homologs
223 and/or regulatory elements of *FY*, *FLC*, *FT* or another gene are underlying the detected QTL.
224 The F_2 experimental design is only the start of the discovery process because lack of
225 recombination between closely linked regions can hinder high-resolution mapping. Our results
226 show one significant QTL and additional suggestive regions of interest. It is likely that several
227 of these loci would need to be transferred into the desired genetic background to exploit
228 heterosis between spring and winter varieties.

229 Genotype errors and correction

230 We detected high pairwise genotype discordance within the duplicate parental samples.
231 Because genotype errors in either of the compared duplicate samples can lead to discordance,
232 the genotype error rate can be roughly estimated as half of the discordance (~6%). In line with
233 GBS results in a rice F_2 population (*47*), most errors can be attributed to undercalling of
234 heterozygous alleles (97.51%). The parental lines are homozygous, though residual
235 heterozygosity and mismapping can lead to heterozygous allele calls. The true error rate in the
236 progeny is therefore likely higher than in the parents, because F_2 populations contain an
237 expected 50% heterozygous alleles.

238 Despite the range of optimal depths reported in the literature, it is clear that calling
239 heterozygous SNPs accurately requires depths substantially higher than those required for
240 calling homozygous SNPs (*48, 49*). The moderate sequencing depths used in this study (9.41×)
241 may thus lead to inflation of sequencing noise and insufficient allele sampling, which can result

242 in undercalling of heterozygous alleles (*50*). The percentage of errors attributed to undercalling
243 of heterozygotes may even be an underestimate, as errors between apparently homozygous
244 alleles may be caused by conflicting erroneous genotype calls of a heterozygous allele.

245 The genotype error rates found here are higher than error rates reported in the literature, even
246 for heterozygous populations. For example, Malmberg et al. (2018a) analysed a heterozygous
247 *B. napus* panel with different skim sequencing coverages and filtered genotype calls using a
248 minimum read depth of 5. The authors found error rates of 2.1% error at 2 × sequencing
249 coverage and 4.2% error at 1 × sequencing coverage (*50*). Similarly, an error rate of 3% was
250 found using GBS in a bovine population with a minimum read depth of 5 (*51*). In a ddRAD
251 genotyping study using a mapping population of cichlid fishes, an investigation of genotype
252 errors found error rates of 4.41% at 8 × coverage (*52*). This suggests that GBS can lead to
253 higher genotype error rates than expected in samples such as F_2 mapping populations.

254 A substantial effect of genotype errors on linkage mapping was found. Cumulative inflation of
255 linkage map length is often caused by genotype errors that introduce spurious double
256 recombination events into the map (*53*). In addition, high levels of missing data and markers
257 with segregation distortion may also affect the mapping distance as these alter the calculated
258 recombination rate (*54*). It has been estimated that every 1% error rate in a marker adds
259 approximately two cM to the linkage map (*55*). Linkage map inflation has been previously
260 reported for GBS data in wheat (*53, 54*) and rice (*47*). In one of the studies on wheat, errors
261 inflated the linkage by up to 11 times (*53*). In a linkage and QTL mapping study of *B. rapa*
262 based on SNPs derived from GBS, high error rates (19.6%) were found and the resulting A
263 genome linkage map was inflated, spanning 4,802.52 cM (*56*).

264 Linkage and QTL mapping in major crops are commonly carried out using highly accurate
265 commercial genotyping arrays such as the Illumina Infinium Brassica 60K array (*57*).
266 Genotyping arrays may introduce sampling bias because they only genotype previously known
267 SNPs. A particular advantage of GBS over genotyping arrays is that regions missing from the
268 reference genome can be genotyped and used for linkage mapping. However, genotyping arrays
269 have the important advantages of more accurate calling of heterozygous genotypes and low
270 missing data. Although genotype arrays also produce errors that frequently involve
271 heterozygous sites, the error rates are likely lower at 1-2% (*58, 59*). Our results suggest that in
272 heterozygous populations, genotyping arrays will generate markers with substantially higher
273 accuracy than GBS. Here, to increase genotype accuracy for GBS data, genotype correction

274 was applied. We find that genotype correction substantially decreased genetic map inflation,
275 underlining the value of a correction step in heterozygous populations analysed using GBS at
276 low to moderate sequencing coverage.

277 Conclusion

278 We report a QTL on chromosome C2 for flowering time and budding time in a *B. napus* winter
279 type *x* spring type cross. This QTL and the additional suggestive loci can be fine-mapped and
280 backcrossed into the parental varieties to facilitate flowering time control in hybrid spring type
281 *x* winter type varieties. An optimised combination of enzymes was also identified using in
282 silico analysis, and the resulting number of empirical ddRAD loci and SNPs demonstrate the
283 effectiveness of the enzyme combination HinfI and HpyCH4IV. In addition, we show that
284 ddRAD generates high levels of genotype errors that can impact linkage map construction. By
285 filtering SNPs and applying genotype imputation and correction an accurate map could be
286 constructed, allowing effective QTL analysis of flowering time. Further investigation of the
287 loci controlling flowering time and maturation will allow *B. napus* breeders to better exploit
288 variation in winter and spring types.

289

290 # Materials and methods

291 ## Plant material and phenotyping

292 The mapping population resulted from a cross between an early-flowering spring line
293 (BnSOSR) and a late-flowering winter line (BnWOSR) carried out by BASF. The F_2 population
294 consisting of 200 individuals as well as 4 BnSOSR and 3 BnWOSR parental replicates was
295 sown directly into 10x10x15 cm pots in a phytotron at the University of Western Australia,
296 Perth. Temperature was maintained between 18-22 °C. To ensure flowering occurred,
297 vernalization was initiated 63 days after sowing and plants were moved to a controlled
298 environment chamber with a constant temperature of 4 °C and an 8-hour light period. Plants
299 were watered twice a week by hand. After approximately 6 weeks, and a total of 108 days after
300 sowing, the plants were returned to the phytotron. This 6 week vernalization period is
301 sufficiently long to ensure that variation in flowering time in the F_2 population is not driven by
302 variation in vernalization response (*60, 61*). Each pot was provided with a dripper and
303 connected to the irrigation system. Plants were watered twice a day for 1 minute. A total of 125
304 ml of fertilizer with micro minerals was provided by hand every two weeks. The time of first
305 floral buds appearing and the date of first flower opening were recorded. All plants were grown
306 until seed set. To investigate whether phenotype data was normally distributed, Shapiro-
307 Wilke's tests were carried and histograms plotted using ggpubr 0.2.1 (*62*).

308 ## Restriction enzyme selection and digestion

309 DNA fragmentation was carried out by simultaneous digestion using two restriction enzymes.
310 Suitable restriction enzyme pairs, that created sticky or overhanging ends, were selected based
311 on their reaction buffer and incubation temperature compatibility to allow simultaneous
312 digestion. The software IgCoverage 1.0 (*7*) was used to carry out in silico digestion of the *B.*
313 *napus* genome (Darmor-bzh v8.1; *63*) using these enzyme pairs. The number of fragments
314 within the 100-600 bp size range with different ends (LengthDeFrag100-600), as well as the
315 expected percentage of genome coverage breadth generated by the selected 26 restriction
316 enzyme pairs, were then compared. Out of the 26 suitable enzyme pairs selected, 18 pairs
317 showed coverage breadth > 20% (Table S1). The enzyme pair HinfI and HpyCH4IV (New
318 England Biolabs, Ipswich, USA) was selected based on the number of fragments, genome
319 coverage, availability and cost per sample. This pair was predicted to generate 840,663

320 fragments with different ends within the 100-600 bp range, which covered 24.6% of the
321 genome. The suitability of the selected restriction enzyme pair was confirmed by digesting 400
322 ng of genomic DNA using 5 Units of each restriction enzyme and NEB CutSmart™ buffer
323 (10×) (New England Biolabs (NEB), Ipswich, USA). The reaction was incubated for four hours
324 at 37°C and the results were visualised using the LabChip GX Touch 24 (PerkinElmer,
325 Waltham, USA).

326 Adapter design

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

336 ddRAD library preparation

337 Genomic DNA was extracted from leaf material using the DNeasy Plant Mini Kit (QIAGEN,
338 Hilden, Germany) according to the manufacturer's protocol. DNA concentrations were
339 quantified using the broad range Qubit 3.0 Fluorometric assay (Invitrogen, Carlsbad, USA),
340 while DNA quality was assessed with the LabChip GX Touch 24 (PerkinElmer, Waltham,
341 USA). Modified versions of the Peterson et al. (2012) and Clark et al. (2014) protocols were
342 used to construct the ddRAD libraries. The extracted gDNA was normalised at 50 ng/µL and
343 200 ng of DNA of each sample was digested in a 20 µL reaction volume containing restriction
344 enzyme/s and recommended buffer. Digestion for the preparation of the ddRAD libraries was
345 carried out using HpyCH4IV (5 U) and Hinfl (5 U) in NEB CutSmart™ buffer. The reaction
346 was incubated at 37°C for 4 hours.

347 The digested DNA was ligated respectively to the unique barcoded and common adapters using
348 T4 DNA ligase (Thermo Invitrogen, Carlsbad, USA). An 18 µL master mix containing ligation
349 buffer, 200 U of T4 ligase and the common adapter was prepared and added directly to the 20

350 μL digest reaction, after which the individual barcoded adapters were added. The reaction was
351 incubated at 22°C for two hours, followed by 65°C for 20 minutes, then cooled to 4°C at a
352 ramp rate of 2°C per 90 seconds. To accommodate variation in DNA concentration and quality
353 the samples were not pooled after ligation but individually purified and double size selected to
354 enable enrichment of fragments between 250 bp and 800 bp. The total volume of the samples
355 was adjusted to 100 μL by adding 60 μL of nuclease free water. Double size selection was
356 carried out by adding 50 μL of a 1:4 (0.5X) mixture of AMPure XP Beads (Beckman Coulter,
357 Brea, USA) to PEG buffer (20% PEG w/v, 2.5M NaCl) to remove fragments >800 bp. The
358 supernatant was transferred to 20 μL of a 1:1 (0.7X) Ampure XP bead to PEG buffer mixture
359 to collect fragments >250 bp. The beads were washed using 80% ethanol and the fragments
360 eluted in 30 μL nuclease free water.

361 A 10 μL aliquot of the size selected DNA was used for PCR amplification. A 40 μL master mix
362 of Phusion Hot-Start High-Fidelity Master Mix Polymerase (Thermo Fisher Scientific,
363 Walthan, USA) and the Forward (0.5 μM) and Reverse primers (0.5 μM) was prepared.
364 Samples were amplified at 98°C for 2 minutes, followed by 15 cycles of 98°C for 15 seconds,
365 62°C for 30 seconds, 72°C for 30 seconds, and a final extension for 5 minutes at 72°C.
366 Amplified libraries were cleaned using 1.5X Ampure XP Beads to sample volume to remove
367 primer dimers. The resulting library DNA concentrations were determined using the High
368 Sensitivity (HS) Qubit 3.0 Fluoro metric assay. Library quality and fragment size distribution
369 were visualised using the LabChip GX Touch 24. Equimolar amounts of the prepared libraries
370 were pooled and loaded on a 1.5% agarose gel to enrich and select fragments between 300-700
371 bp. The DNA was recovered using the QIAquick Gel Extraction Kit (QIAGEN, Hilden,
372 Germany). The final library concentration, quality and size distribution were assessed again
373 and adjusted to 20 nM DNA using 10 nM Tris Buffer (pH 8.5, 0.1% Tween 20, 10 nM). The
374 final libraries were sent to the KCCG Core facility at the Garvan Institute for Medical Research
375 for paired end sequencing on the HiSeq X Ten platform.

376 Adapter trimming and quality trimming

377 The Illumina bcl2fastq 2.20.0.422 pipeline (*64*) was used to convert base call files to FASTQ
378 format. Paired-end sequencing reads were demultiplexed using sabre 1.0 (*65*) with a single
379 mismatch allowed. Raw FASTQ files were trimmed of adapter sequences and low quality bases
380 with Trimmomatic 0.36 (*66*). For adapter trimming, a maximum mismatch score of 2 was used

381 for the adapter sequence, together with a palindrome clip score threshold of 30 and a simple
382 clip score threshold of 10. Low quality bases with a Phred+33 score below 3 were trimmed
383 from the start and end of the read. Sliding window trimming was carried out using a 4-base
384 wide window, trimming the bases when the average quality per base was below 15. The
385 Illumina TruSeq3-PE adapter list provided with Trimmomatic was used for adapter trimming.
386 All reads with fewer than 36 bases after trimming were discarded. Reads that were unpaired
387 after trimming were also discarded. Following read pre-processing, the untrimmed and
388 trimmed reads were analysed using the diagnostic tool fastQC (*67*). The results of fastQC for
389 each sample were then aggregated and summarised using multiQC (*68*). The multiQC report
390 was used to verify that adapters had been removed and read quality was high.

391 Aligning sequencing reads

392 Trimmed reads were mapped using BWA 0.7.17 with the BWA-MEM algorithm (*69*) to the *B.*
393 *napus* Darmor-bzh v8.1 assembly (*63*) using default parameters. After alignment, SAM files
394 were converted to BAM format using SAMtools 1.8 (*70*). Unmapped reads, supplementary
395 alignments and reads with a mapping quality below 20 were discarded. This filter removes
396 multi-mapping reads, which commonly occur in *B. napus* due to the homeologous regions of
397 its polyploid genome. The mapping results were analysed using SAMtools stats and mosdepth
398 0.2.3 (*71*). The number of ddRAD loci were calculated from mosdepth per-base output using
399 BEDTools 2.26.0 (*72*) to merge neighboring loci within 100bp.

400 Calling single nucleotide variants

401 Variants were called using GATK 3.6 (*73*). First, BAM alignments were indexed using
402 SAMtools, then HaplotypeCaller was used to call SNPs for each individual sample.
403 Genotyping was carried out using GATK GenotypeGVCF using default setting with auto index
404 creation and locking when reading rods disabled. Results per chromosome were merged using
405 GATK CatVariants. Variants were filtered using VCFtools 0.1.15 (*74*). Indels and multiallelic
406 SNPs were excluded *(--remove-indels --max-alleles 2 --min-alleles 2*). Before filtering SNPs
407 individuals with >0.9 missing genotypes were removed. To reduce the rate of heterozygous
408 alleles incorrectly called as homozygous alleles due to insufficient read depth, genotypes with
409 a depth <5 *(--minDP 5*) were set to missing. SNPs were discarded if they displayed a minor
410 allele frequency <0.05 (*--maf 0.05*) or when genotypes were not present in >80% of all

411 individuals (*--max-missing 0.8*). Genotype discordance was calculated with snpEff 4.3t (*75*)
412 using the duplicate samples for the parents (spring type BnSOSR with n=4, winter type
413 BnWOSR with n=3) with pairwise comparisons of genotypes for SNPs passing the above
414 filters. Heterozygosity per individual was calculated using VCFtools.

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

423 Genotypes were imputed and corrected using Genotype-Corrector 1.0 (*79*). This software uses
424 the order of SNPs on the genome reference and a sliding-window approach to impute and
425 correct genotypes based on neighboring genotypes in F_2 populations. Before correction, up to
426 eight consecutive homozygous SNPs within 150 bp genomic intervals in heterozygous regions
427 were binned into a single SNP with Genotype-Corrector qc_hetero. This helps prevent
428 miscorrection of heterozygous genotypes to homozygous genotypes when using the sliding
429 window approach. With the 20% missing SNPs used here, the expected accuracy of Genotype-
430 Corrector is >95%, based on empirical testing in crop mapping populations (*79*).

431 Linkage mapping

432 Linkage mapping was carried out using the MSTMap algorithm (*80*) implemented in the R
433 package ASMap (*81*). The sum of recombination events objective function was used to find the
434 optimal sequence of loci, and the p-value threshold for clustering markers into linkage groups
435 was set to $1e^{-23}$ based on evaluation of a range of values from $1e^{-14}$ to $1e^{-29}$. The kosambi
436 distance function was used to estimate genetic distances between SNPs, and rare recombination
437 events were treated as errors (*detectBadData=True*). Linkage groups were assigned
438 chromosome names based on marker positions on the reference genome. Small linkage groups
439 that did not represent an entire chromosome, were merged with chromosomal linkage groups
440 and genetic distances recalculated if an unambiguous assignment was possible using physical
441 marker positions. Linkage groups with <7 markers were discarded. The estimated pairwise

recombination fractions between markers were calculated using the rqtl function plotRF (*37*). 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 1,000 iterations (*82*). 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 (*83*) was based on the rqtl manual (*37*).

Flowering time genes

A total of 306 Arabidopsis flowering time (FT) genes from the FLOweRing Interactive Database (*84*) was downloaded from The Arabidopsis Information Resource (*85*). These genes include homologs of the known *B. napus* flowering time genes. BLAST+ 2.2.29 (*86, 87*) analysis of the FT genes against the reference genome was carried out to find gene homologs using a cut-off of $1e^{-6}$ (following *88*). 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.

Acknowledgements

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. This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

Author 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.

Data availability

All sequences have been deposited in SRA (PRJNA640838) with individual accessions listed in Table S8.

Code availability

Scripts used to for this study are available at https://github.com/ascheben/bn_gbs/.

References

1. 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. Nature Reviews Genetics 17, 81-92. https://doi.org/10.1038/nrg.2015.28

2. 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

3. J. A. Poland, T. W. Rife (2012) Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome 5, 92-102. https://doi.org/10.3835/plantgenome2012.05.0005

4. S. Deschamps, V. Llaca, G. D. May (2012) Genotyping-by-sequencing in plants. Biology 1, 460-483. https://doi.org/10.3390/biology1030460

5. 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

6. 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

7. 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. G3: Genes|Genomes|Genetics 6, 845. https://doi.org/10.1534/g3.115.025775

8. J. G. Uitdewilligen, A. M. Wolters, B. D'Hoop B, 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

9. H. Sonah, L. O'Donoughue, 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

10. M. M. Malmberg, L. W. Pembleton, R. C. Baillie, M. C. Drayton, S. Sudheesh, S. Kaur, H. Shinozuka, 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

11. S. J. Yang, B. Zhang, G. Liu, B. H. Hong, J. S. Xu, X. Chen, B. Wang, Z. K. Wu, F. Hou, X. P. Yue *et al.* (2018) A comprehensive and precise set of intervarietal substitution lines to identify candidate genes and quantitative trait loci in oilseed rape (*Brassica napus* L.). Theor. Appl. Genet. 131, 2117-2129. https://doi.org/10.1007/s00122-018-3140-4

525 12. X. Chen, X. Li, B. Zhang, J. Xu, Z. Wu, B. Wang, H. Li, M. Younas, L. Huang, Y. Luo
526 *et al.* (2013) Detection and genotyping of restriction fragment associated
527 polymorphisms in polyploid crops with a pseudo-reference sequence: a case study in
528 allotetraploid *Brassica napus*. BMC Genomics 14, 346-346.
529 https://doi.org/10.1186/1471-2164-14-346

530 13. Q. H. Zhou, C. Zhou, W. Zheng, A. S. Mason, S. Y. Fan, C. J. Wu, D. H. Fu, Y. J.
531 Huang (2017) Genome-wide SNP markers based on SLAF-seq uncover breeding traces
532 in rapeseed (*Brassica napus* L.). Frontiers in Plant Science 8.
533 https://doi.org/10.3389/fpls.2017.00648

534 14. I. Bancroft, C. Morgan, F. Fraser, J. Higgins, R. Wells, L. Clissold, D. Baker, Y. Long,
535 J. L. Meng, X. W. Wang *et al.* (2011) Dissecting the genome of the polyploid crop
536 oilseed rape by transcriptome sequencing. Nat. Biotechnol. 29, 762-766.
537 https://doi.org/10.1038/nbt.1926

538 15. B. Li, J. Gao, J. Chen, Z. Wang, W. Shen, B. Yi, J. Wen, C. Ma, J. Shen, T. Fu *et al.*
539 (2019) Identification and fine mapping of a major locus controlling branching in
540 *Brassica napus*. Theor. Appl. Genet. https://doi.org/10.1007/s00122-019-03506-x

541 16. F. Sun, J. Liu, W. Hua, X. Sun, X. Wang, H. Wang (2016) Identification of stable QTLs
542 for seed oil content by combined linkage and association mapping in *Brassica napus*.
543 Plant Sci. 252, 388-399. https://doi.org/10.1016/j.plantsci.2016.09.001

544 17. A. Abe, S. Kosugi, K. Yoshida, S. Natsume, H. Takagi, H. Kanzaki, H. Matsumura, K.
545 Yoshida, C. Mitsuoka, M. Tamiru *et al.* (2012) Genome sequencing reveals
546 agronomically important loci in rice using MutMap. Nat. Biotechnol. 30, 174-178.
547 https://doi.org/10.1038/nbt.2095

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

551 19. M. Tadege, C. C. Sheldon, C. A. Helliwell, P. Stoutjesdijk, E. S. Dennis, W. J. Peacock
552 (2001) Control of flowering time by *FLC* orthologues in *Brassica napus*. Plant J. 28,
553 545-553. https://doi.org/10.1046/j.1365-313X.2001.01182.x

554 20. I. A. Parkin, S. M. Gulden, A. G. Sharpe, L. Lukens, M. Trick, T. C. Osborn, D. J.
555 Lydiate (2005) Segmental structure of the *Brassica napus* genome based on
556 comparative analysis with *Arabidopsis thaliana*. Genetics 171, 765-781.
557 https://doi.org/10.1534/genetics.105.042093

558 21. L. S. Robert, F. Robson, A. Sharpe, D. Lydiate, G. Coupland (1998) Conserved
559 structure and function of the Arabidopsis flowering time gene *CONSTANS* in *Brassica*
560 *napus*. Plant Mol. Biol. 37, 763-772. https://doi.org/10.1023/A:1006064514311

561 22. H. J. Jian, A. X. Zhang, J. Q. Ma, T. Y. Wang, B. Yang, L. S. Shuang, M. Liu, J. N. Li,
562 X. F. Xu, A. H. Paterson *et al.* (2019) Joint QTL mapping and transcriptome sequencing
563 analysis reveal candidate flowering time genes in *Brassica napus* L. BMC Genomics
564 20, 21. https://doi.org/10.1186/s12864-018-5356-8

565 23. B. J. Li, W. G. Zhao, D. R. Li, H. B. Chao, X. P. Zhao, N. Ta, Y. H. Li, Z. B. Guan, L.
566 X. Guo, L. N. Zhang *et al.* (2018) Genetic dissection of the mechanism of flowering
567 time based on an environmentally stable and specific QTL in *Brassica napus*. Plant Sci.
568 277, 296-310. https://doi.org/10.1016/j.plantsci.2018.10.005
569

24. 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

25. 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. Frontiers in Plant Science 9, 390. https://doi.org/10.3389/fpls.2018.00390

26. 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

27. 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

28. 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

29. T. C. Osborn, C. Kole, I. A. Parkin, A. G. Sharpe, M. Kuiper, D. J. Lydiate, M. Trick (1997) Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. Genetics 146, 1123-1129.

30. 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

31. 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

32. L. A. Ouellette, R. W. Reid, S. G. Blanchard, Jr., C. R. Brouwer (2017) LinkageMapView - Rendering High Resolution Linkage and QTL Maps. Bioinformatics 34, 306-307. https://doi.org/10.1093/bioinformatics/btx576

33. H. Alipour, G. H. Bai, G. R. Zhang, M. R. Bihamta, 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

34. 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. Frontiers in Plant Science 6, 1164. https://doi.org/10.3389/fpls.2015.01164

35. W. Ecke, A. Kampouridis, K. Ziese-Kubon, A. C. Hirsch (2015) Identification and genetic characterization by high-throughput SNP analysis of intervarietal 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

615 36. W. E. Clarke, E. E. Higgins, J. Plieske, R. Wieseke, C. Sidebottom, Y. Khedikar, J.
616 Batley, D. Edwards, J. L. Meng, R. Y. Li *et al.* (2016) A high-density SNP genotyping
617 array for *Brassica napus* and its ancestral diploid species based on optimised selection
618 of single-locus markers in the allotetraploid genome. Theor. Appl. Genet. 129, 1887-
619 1899. https://doi.org/10.1007/s00122-016-2746-7

620 37. K. W. Broman, H. Wu, S. Sen, G. A. Churchill (2003) R/qtl: QTL mapping in
621 experimental crosses. Bioinformatics 19, 889-890.
622 https://doi.org/10.1093/bioinformatics/btg112

623 38. J. Dupuis, D. Siegmund (1999) Statistical methods for mapping quantitative trait loci
624 from a dense set of markers. Genetics 151, 373-386.

625 39. B. Mangin, B. Goffinet, A. Rebai (1994) Constructing confidence intervals for QTL
626 location. Genetics 138, 1301-1308.

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

630 41. S. D. Michaels, R. M. Amasino (1999) *FLOWERING LOCUS C* encodes a novel
631 MADS domain protein that acts as a repressor of flowering. Plant Cell 11, 949-956.
632 https://doi.org/10.1105/tpc.11.5.949

633 42. G. G. Simpson, P. P. Dijkwel, V. Quesada, I. Henderson, C. Dean (2003) *FY* is an RNA
634 3 ' end-processing factor that interacts with *FCA* to control the *Arabidopsis* floral
635 transition. Cell 13, 777-787. https://doi.org/10.1016/S0092-8674(03)00425-2

636 43. D. Z. Wu, Z. Liang, T. Yan, Y. Xu, L. J. Xuan, J. Tang, G. Zhou, U. Lohwasser, S. J.
637 Hua, H. Y. Wang *et al.* (2019) Whole-genome resequencing of a worldwide collection
638 of rapeseed accessions reveals the genetic basis of ecotype divergence. Molecular Plant
639 12, 30-43. https://doi.org/10.1016/j.molp.2018.11.007

640 44. R. Raman, S. Diffey, J. Carling, R. B. Cowley, A. Kilian, D. J. Luckett, H. Raman
641 (2016) Quantitative genetic analysis of grain yield in an Australian *Brassica napus*
642 doubled-haploid population. Crop Pasture Sci 67, 298-307.
643 https://doi.org/10.1071/Cp15283

644 45. W. Feng, S. D. Michaels (2011) Dual roles for *FY* in the regulation of *FLC*. Plant
645 Signaling & Behavior 6, 703-705.

646 46. J. Wang, C. J. Hopkins, J. N. Hou, X. X. Zou, C. N. Wang, Y. Long, S. Kurup, G. J.
647 King, J. L. Meng (2012) Promoter variation and transcript divergence in Brassicaceae
648 lineages of *FLOWERING LOCUS T*. PLoS One 7, e47127.
649 https://doi.org/10.1371/journal.pone.0047127

650 47. T. Furuta, M. Ashikari, K. K. Jena, K. Doi, S. Reuscher (2017) Adapting genotyping-
651 by-sequencing for rice F2 populations. G3-Genes Genom Genet 7, 881-893.
652 https://doi.org/10.1534/g3.116.038190

653 48. D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G.
654 Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell *et al.* (2008) Accurate whole
655 human genome sequencing using reversible terminator chemistry. Nature 456, 53-59.
656 https://doi.org/10.1038/nature07517

657 49. T. Maruki, M. Lynch (2017) Genotype calling from population-genomic sequencing
658 data. G3-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. Frontiers in Plant Science 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. Saintenac, 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-Us 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 Hortic 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/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/10.1093/bioinformatics/btw354

69. H. Li (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, 1303.3997. https://doi.org/

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 (2020) Accessed: 25 June 2020. 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. 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

80. 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

81. 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

82. G. A. Churchill, R. W. Doerge (1994) Empirical threshold values for quantitative trait mapping. Genetics 138, 963-971.

83. RStudio Team (2015) RStudio: Integrated Development for R. Accessed: 11 November 2017. http://www.rstudio.com/

84. 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, D1167-1171. https://doi.org/10.1093/nar/gkv1054

85. 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

86. 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

87. 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

88. 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*. Frontiers in Plant Science 3, 297. https://doi.org/10.3389/fpls.2012.00297

Supplemental tables

Table S1. 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; DeFrag100-600: 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. https://github.com/ascheben/bn_gbs/blob/master/supplemental_tables/Table_S1.xlsx

Table S2. Linkage map constructed using ASMap with corrected and imputed ddRAD markers derived from an BnSOSR x BnWOSR F2 population. https://github.com/ascheben/bn_gbs/blob/master/supplemental_tables/Table_S2.xlsx

Table S3. 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 60K genotyping array (Brassica 60k). https://github.com/ascheben/bn_gbs/blob/master/supplemental_tables/Table_S3.xlsx

Table S4. Phenotypic variation in flowering time and time to bud in the F2 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). https://github.com/ascheben/bn_gbs/blob/master/supplemental_tables/Table_S4.xlsx

Table S5. 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. https://github.com/ascheben/bn_gbs/blob/master/supplemental_tables/Table_S5.xlsx

Table S6. Candidate flowering time genes within the QTL interval detected on chromosome C2. https://github.com/ascheben/bn_gbs/blob/master/supplemental_tables/Table_S6.xlsx

Table S7. Phenotype data for flowering time and budding time in the mapping population and parents. https://github.com/ascheben/bn_gbs/blob/master/supplemental_tables/Table_S7.xlsx

Table S8. Sample list for the 192 progeny and 7 parental individuals used in this study. https://github.com/ascheben/bn_gbs/blob/master/supplemental_tables/Table_S8.xlsx

Supplemental figures

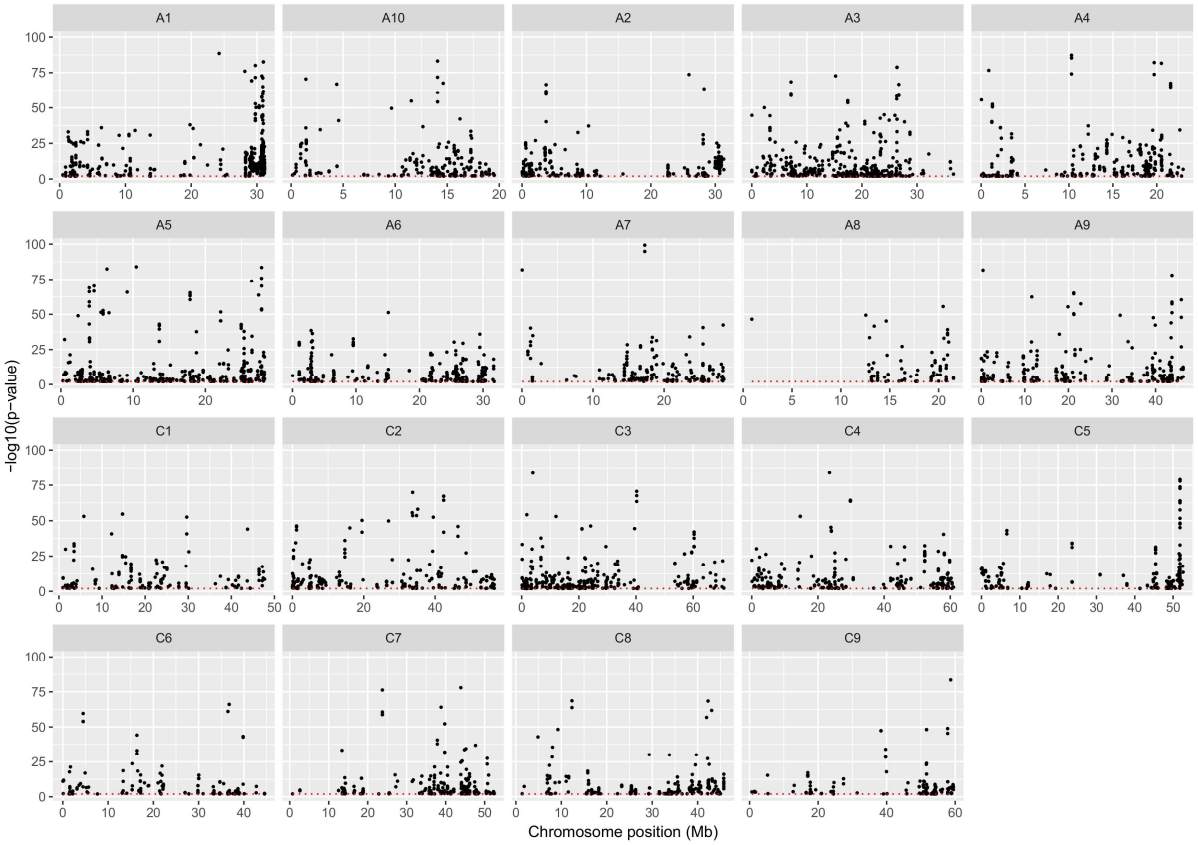

Figure S1. Segregation distorted loci across all chromosomes. The significance threshold (p>0.01) is shown as a red dotted line.

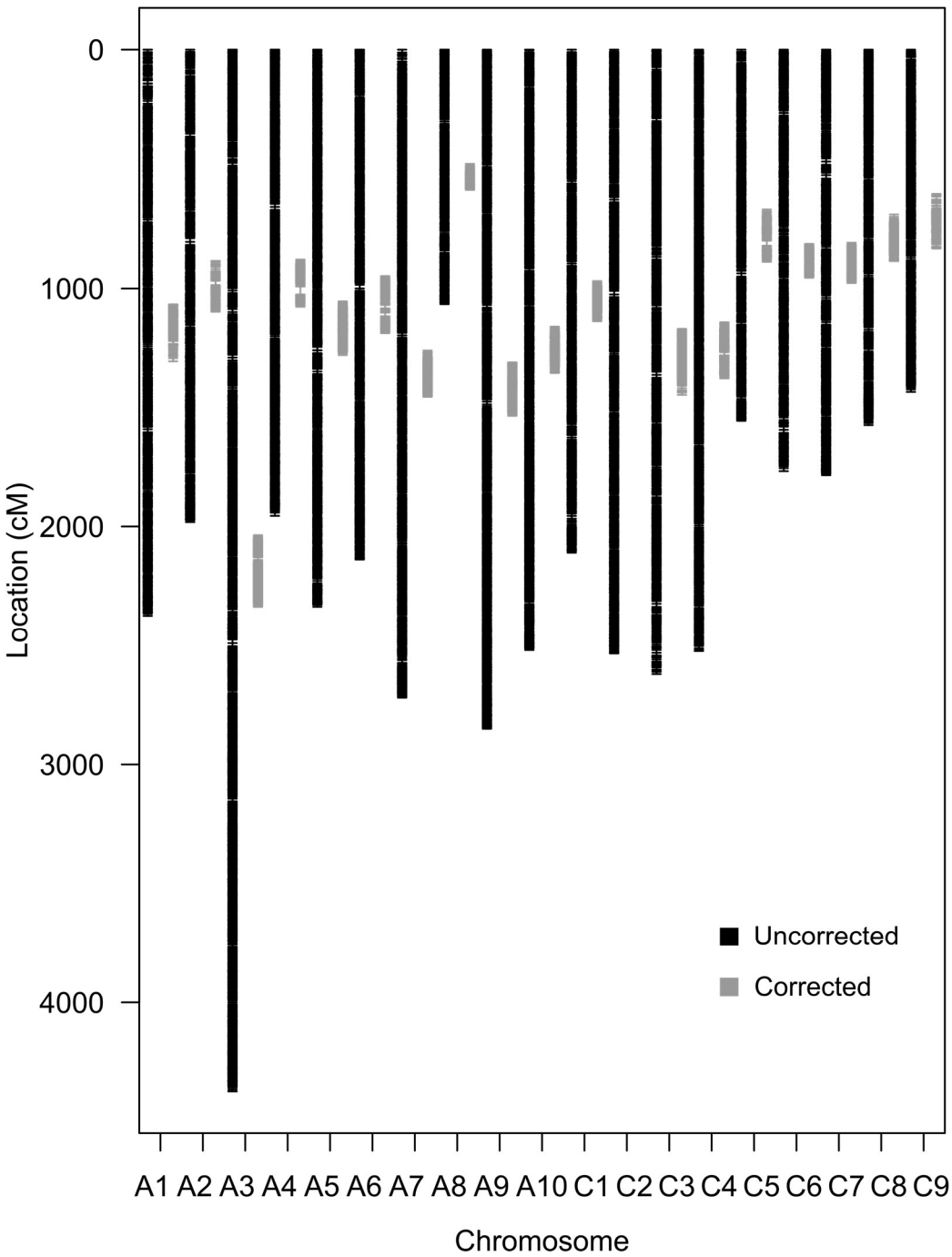

Figure S2. Comparison of linkage group sizes in an uncorrected genetic map and a corrected genetic map. Corrected linkage groups are aligned centrally to uncorrected groups.

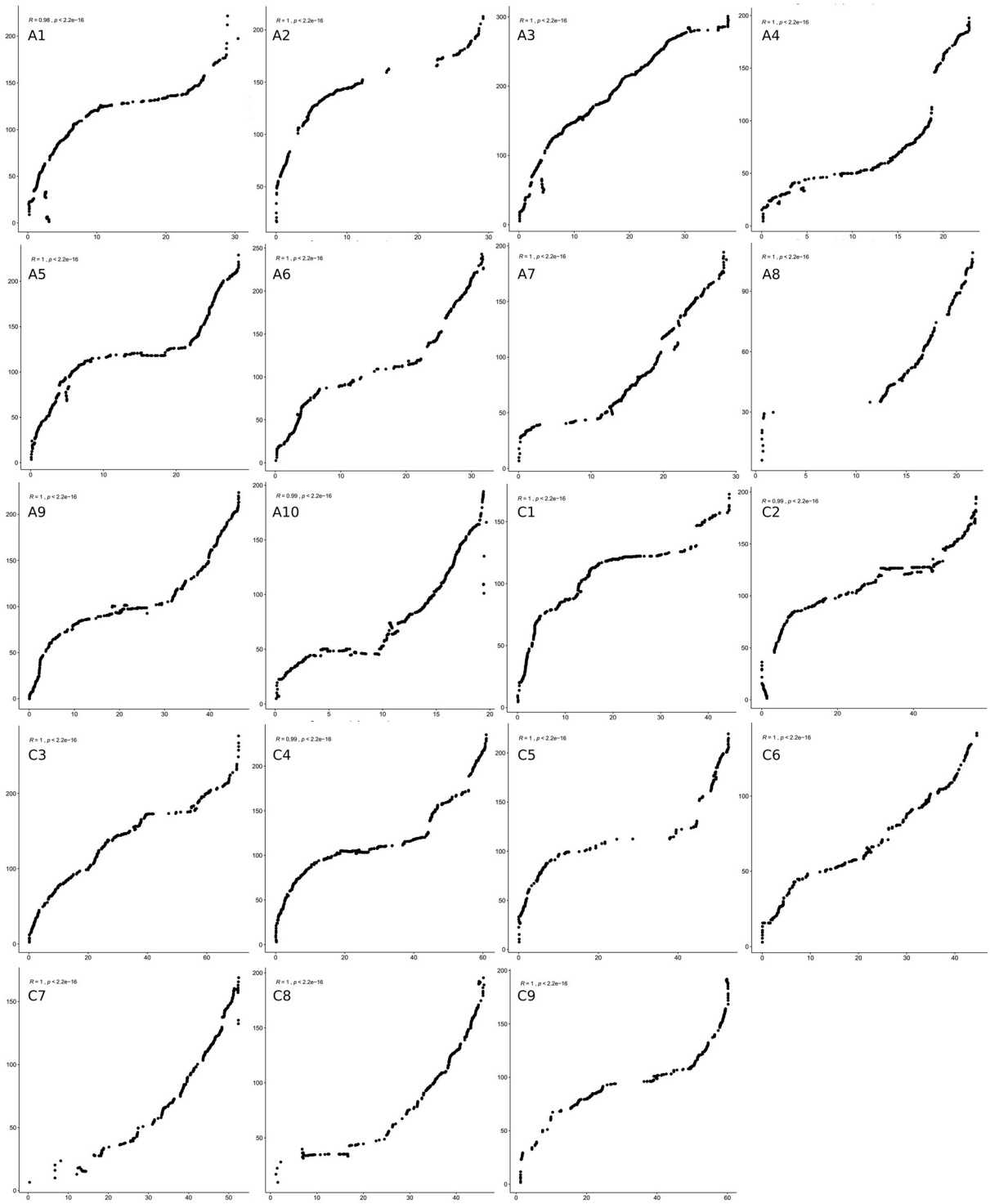

Figure S3. Physical (x-axis) and genetic (y-axis) marker positions on all chromosomes in Mb and cM respectively. Spearmans's rank correlation test result shown in the top left corner of each plot.

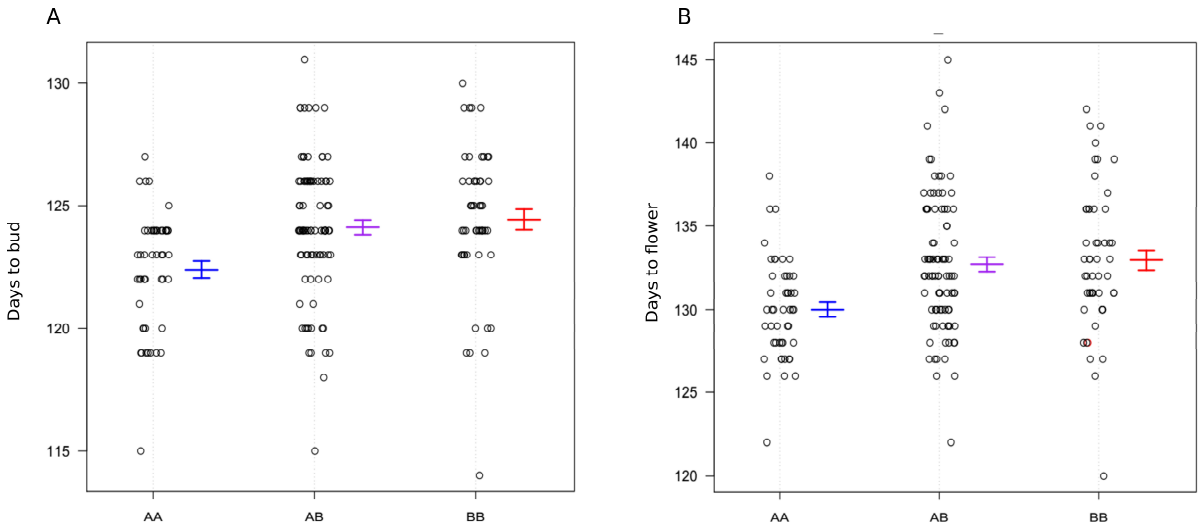

Figure S4. Effect plot for the budding time (A) and flowering time (B) QTL on C2 at positions 4,673,904 and 4,655,461 respectively. The 'AA' genotype is BnSOSR and the 'BB' genotype is BnWOSR.