CSIRO PUBLISHING
Crop & Pasture Science
https://doi.org/10.1071/CP17161

Agricultural selection and presence–absence variation in spring-type canola germplasm

Annaliese S. Mason[A,B], *Pratibha Chauhan*[C], *Shashi Banga*[C], *Surinder S. Banga*[C], *Phil Salisbury*[D,E], *Martin J. Barbetti*[F], *and Jacqueline Batley*[B,G,H]

[A]Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany.
[B]School of Agriculture and Food Sciences and Centre for Integrative Legume Research, The University of Queensland, Brisbane, Qld 4072, Australia.
[C]Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana 141 001, India.
[D]Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, Vic. 3010, Australia.
[E]Department of Economic Development, Jobs, Transport and Resources, AgriBio, Centre for AgriBioscience, 5 Ring Road, La Trobe University, Bundoora, Vic. 3083, Australia.
[F]School of Agriculture and Environment and The UWA Institute of Agriculture, The University of Western Australia, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia.
[G]School of Biological Sciences and the UWA Institute of Agriculture, The University of Western Australia, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia.
[H]Corresponding author. Email: jacqueline.batley@uwa.edu.au

Abstract. *Brassica napus* (rapeseed, canola) is an important oilseed crop worldwide as well as a recent agricultural hybrid species, resulting from crosses between progenitor *B. rapa* (turnip) and *B. oleracea* (cabbage) species in the last few thousand years. No wild form of *B. napus* is known to exist, making *B. napus* an interesting model for studies of genetic and genomic evolution in a polyploid under agricultural selective pressure. We generated genotype (Illumina Infinium 60K Brassica array) and phenotype data for elite spring-type *B. napus* lines from Australia, China and India (only one line). Phenotypically, plant growth, silique development and flowering traits were more likely to differentiate Chinese germplasm, whereas resistance to blackleg disease, secondary branching and seed traits were more likely to differentiate Australian germplasm. Genetic differentiation between the Australian and Chinese populations was low ($F_{ST} = 0.035$). Genetic relationship was not a predictor of similarity in yield traits between lines. Presence–absence variants were detected across the population: variants shared by at least three lines were present in every chromosome in the *B. napus* genome, and large missing chromosome segments (>1 Mbp) putatively due to A–C genome translocations were observed on chromosomes A7, A10, C1, C2, C6, C8 and C9. Our results highlight that widespread presence–absence variation is usual in *B. napus*, and may suggest that phenotypic and genetic diversity are not closely linked within spring-type *B. napus* from Australia and China, although the low sample numbers in our study prevent strong conclusions. We propose that inbreeding and low levels of genetic diversity, coupled with exchanges between the A and C genomes, were major driving forces behind genome evolution in this recent agricultural crop species.

Additional keywords: *Brassica napus*, genetic diversity, phenotype, SNP genotyping.

Received 20 April 2017, accepted 4 July 2017, published online 14 August 2017

Introduction

Brassica napus (rapeseed, canola) is a major oilseed crop worldwide, and is hence of major agronomic interest. Allotetraploid *B. napus* ($2n = AACC$) is a purely agricultural species, with no known wild accessions, and is thought to have arisen via chance selection after natural hybridisations between *B. rapa* (turnip, Chinese cabbage, $2n = AA$) and *B. oleracea* (cabbage, cauliflower, broccoli, $2n = CC$) (Morinaga 1934; U 1935). Due to its recent origin (Chalhoub *et al*. 2014), it is considered an interesting model for studies of agricultural selective pressure in genome evolution, as well as for studies of allopolyploid evolution (Cifuentes *et al*. 2010; Mason and Snowdon 2016). Increasing amounts of evidence suggest that chromosomal exchanges between the 'A ' and 'C' genomes are common in *B. napus* (Osborn *et al*. 2003; Chalhoub *et al*. 2014), and that these exchanges may have direct impacts on phenotypic traits under agricultural selection (Zou *et al*. 2011; Liu *et al*. 2012; Chalhoub *et al*. 2014; Schiessl *et al*. 2014).

Development of genetic and genomic resources for this crop species has proceeded rapidly over the last decades, with

the release of a draft genome sequence in 2014 (Chalhoub *et al*. 2014), and recently additional resequencing (Schmutzer *et al*. 2015) and transcriptome data (He *et al*. 2015). Of great interest to the wider genetics and breeding community was the recent development and release of an Illumina 60K Infinium array for *B. napus* by an international consortium (Clarke *et al*. 2016). This combination of resources now permits high-throughput genotyping of large numbers of samples across more than 50 000 SNPs located on the sequenced genome. Large quantities of phenotypic data have also been produced for this crop species, and numerous quantitative trait loci (QTL) identified using linkage or association mapping approaches can now be integrated on to the reference genome sequence (Zhou *et al*. 2014; Körber *et al*. 2016; Liu *et al*. 2016; Xu *et al*. 2016).

We analysed allelic and phenotypic diversity and genomic presence/absence variation in a *B. napus* diversity set consisting of spring-type elite canola cultivars from Australia, China and India. Quantitative relationships between these variables were assessed, and association mapping carried out between yield- and disease-resistance related traits and SNP-genotyping data for these lines.

Materials and methods

Germplasm

Diverse *Brassica napus* lines from Australia (47), China (27) and India (2) (see Supplementary materials table S1, as available at the journal's website) were made available as part of a collaborative project agreement. These lines are described elsewhere (http://aciar.gov.au/files/node/13688/fr2011_11_pdf_15105.pdf, accessed 13 July 2017). Most assessed lines were derived from self-pollinated seed collected from field-grown plants at Punjab Agricultural University, Ludhiana, India (Supplementary materials table S1). These germplasm lines were self-pollinated for 3–4 generations, before being used for the studies reported in this manuscript. Remaining lines were sourced from the Australian Grains Genebank, glasshouse-grown self-pollinated seed at The University of Western Australia or field-grown open-pollinated seed at The University of Melbourne.

Phenotyping

Yield traits were assessed at Punjab Agricultural University (PAU), Ludhiana, Punjab, India for 64 lines during two consecutive years (2009–10 and 2010–11) under field conditions. The germplasm lines were grown with two replications in an 8×8 simple lattice design with row-to-row spacing of 45 cm and plant-to-plant spacing of 10 cm in 1.8 by 4.0-m plots. Yield traits assessed were: number of days to 50% flowering as estimated from seeding to initiation of flowering in ~50% of the plants in each plot (DF50); number of days from seeding to maturity as evident from physiological maturity of about >90% pods/plot (DM); plant height was measured as distance from ground level to the tip of main axis at cessation of flowering in cm (PH); main shoot length measured as distance from the last primary branch to the tip of main shoot at cessation of flowering in cm (MSL); number of productive silique bearing primary branches on the stem (PB); number of productive silique bearing secondary branches borne on the primary branches (SPB); number of siliquae on the main shoot (MSP); average silique length of five silique from the middle of the main shoot in cm (SL); average seed number in 10 random siliques (SSL); seed weight of 1000 seeds in grams (SS); seed yield in kg/ha (Yield); foliar chlorophyll content measured using SPAD-502 (PP); leaf area in cm^2 (LA); leaf area index (LAI) and percentage seed oil (Oil) and protein content (Protein) estimated using a Foss NIRS system model 6500 using standard protocols. The dust free, clean seeds of *B. napus* were packed properly in a ring cup of 2 mm thickness, having a diameter of 77 mm. The ring cup was then placed in a holder and scanned for oil and protein estimation. The oil and protein content was automatically calculated using the Canola calibration equation provided by Dr Rod Mailer (pers. comm.). Morphological traits (excluding DF50, DM and yield, which were assessed per plot) were assessed from five random competitive plants from the middle two rows per plot and averaged. Three representative plants per genotype were selected before initiation of bolting for estimation of leaf area. For this, length and width dimensions of three leaves per representative plant were used to calculate leaf area (length × width), and mean values so obtained per genotype were multiplied with a correction factor (0.68). LAI was estimated as the ratio of total upper leaf surface of vegetation to the surface area of the land on which the vegetation grows.

Previously published disease resistance phenotyping data was collated for the traits of blackleg (*Leptosphaeria maculans*) resistance (Li *et al*. 2008) and Sclerotinia (*Sclerotinia sclerotiorum*) resistance (Li *et al*. 2006, 2009). Blackleg resistance as a function of percentage survival was assessed over three different locations: in Victoria (Lmac_Vic.), South Australia (Lmac_SA) and Western Australia (Lmac_WA) (Li *et al*. 2008). Sclerotinia resistance was assessed as stem lesion length (cm) in a single field location (Li *et al*. 2006, 2009). A total of 76 lines had phenotypic data available (either yield traits or disease resistance data or both; Supplementary materials table S1).

Genotyping

Leaf tissue samples were collected from field-grown plants at Punjab Agricultural University, and DNA extracted using the CTAB method (Doyle 1990) before shipping to The University of Queensland, Australia. Samples were run on 1% agarose gel at The University of Queensland to check DNA quality (degradation and presence of contaminants) before genotyping, and DNA concentrations assessed using a Qubit fluorometer (ThermoFisher Scientific, Scoresby, Vic., Australia) according to manufacturer's instructions. For samples with poor quality DNA, additional samples from the same lines were sourced from The University of Western Australia, The University of Melbourne and from the Australian Grains Genebank (Supplementary materials table S1). The Illumina Infinium 60K Brassica array was used to genotype 1–3 DNA samples from individual plants of each genotype for which phenotype data was available, according to standard protocols as recommended by Illumina for the Infinium array system (www.illumina.com, accessed 13 July 2017). A single sample from each line with successful amplification (>80% SNP call rate)

and less than 20% heterozygosity was selected for further analysis; a total of 54 samples met these criteria.

Presence–absence variation

Common regions of presence–absence variation were defined by the presence of at least three contiguous SNPs with no call in at least three individuals in this population. Regions of presence–absence variation present in less than three individuals were defined as five contiguous SNPs with no call in lines with <10% missing values, and as 10 contiguous SNPs with no call in lines with >10% missing values. To define the boundaries of presence/absence variants, a maximum of two contiguous presence calls flanked by no calls were permitted, with a maximum of 10% calls in an absence variant region. These parameters were used to reduce the chance of incorrectly calling or defining absence variants due to missing values or false positives.

Statistical and association analysis

The R program for statistical computing (R version 3.1.3) was used for all data analysis steps (R Development Core Team 2015). The software program GAPIT (Genomic Association and Prediction Integrated Tool; Lipka *et al*. (2012)) was used with pre-set parameters to detect if any significant associations existed between SNPs and phenotypes in this population. Data was inputted using the 0/1/2 format recommended for SNP data containing heterozygous calls (Lipka *et al*. 2012). Homozygous SNPs (minor allele frequency <2%) and SNPs with greater than 11% missing values (excluding samples with lower amplification; Supplementary materials table S1) were removed before analysis, leaving 14 168 SNPs. Remaining missing values were imputed using column means via the function 'na.gam.replace' in the R 'gam' library. Global population structure was assessed via generation of a kinship matrix following VanRaden (2008) as implemented in the GAPIT package (Supplementary materials figure S1). Bootstrapped dendrograms (1000 iterations) were generated using the 'pvclust' function in the R 'pvclust' library with 'manhattan' and 'average' methodology parameters. Principle components analysis was performed using the 'PCA' function in the R package 'FactoMineR' (Lê *et al*. 2008). Global F_{ST} (theta value) for genetic divergence between Australian and Chinese subpopulations was calculated using the function 'calc_wcFst_spop_pairs' (Chan 2008) and locus-specific variance components and fixation indices between these two subpopulations was calculated using the function 'calc_wcFstats' (Chan 2008) according to the methods of Weir and Cockerham (1984).

Results

Phenotypic variation

Principle components analysis revealed that the assessed phenotypic traits in the population fell into several related groups, as evident by shared axes (Fig. 1). The first principle component explained 28.1% of the variance between measured phenotypes, whereas the second principle component explained 10.5%. Plant growth, silique development and flowering traits (SL, DF50, MSP, MSL, PH and DM) formed one clear group, more likely to be associated with Chinese germplasm, whereas

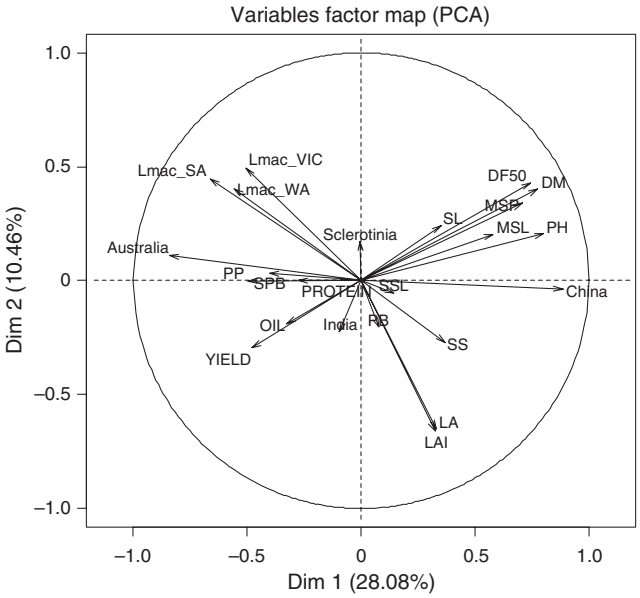

Fig. 1. Variables factor map for the first two principle components of a phenotypic principle components analysis of yield and yield-related traits in a *Brassica napus* diversity set of 76 lines from Australia (47 lines), India (2 lines) and China (27 lines). Yield traits assessed were: number of days from seeding to initiation of flowering in ~50% of the plants (DF50); days to maturity (DM); plant height in cm (PH); main shoot length in cm (MSL); number of primary branches (PB); number of secondary branches (SPB); number of siliquae on main shoot (MSP); silique length in cm (SL); average seed number in 10 random siliques (SSL); seed weight of 1000 seeds in grams (SS); seed yield in kg/ha (Yield); percentage seed oil content (Oil); foliar chlorophyll content measured using SPAD-502 (PP); leaf area in cm^2 (LA); leaf area index (LAI) and percentage seed protein content (Protein). Blackleg resistance as a function of percentage survival was assessed over three different locations: in Victoria (Lmac_Vic.), South Australia (Lmac_SA) and Western Australia (Lmac_WA). Sclerotinia resistance was assessed as stem lesion length (cm) in a single field location.

resistance to blackleg disease (Lmac_SA, Lmac_WA and Lmac_Vic), secondary branching and seed traits (SPB, PP, Protein, Oil and Yield) were more likely to be associated with Australian germplasm. Leaf area and LAI did not appear to show signs of differential selection between germplasm groups.

Australian and Chinese germplasm formed two overlapping, but distinct, clusters on the basis of phenotypic trait variation (Fig. 2). The two samples from India were intermediate, although slightly closer to the Australian groups based on the PCA results.

Moderate support for distinct phenotype clusters was obtained via hierarchical clustering analysis (Fig. 3). Although few higher branches had >95% support, several groups of more closely related lines could be observed, with much greater support for low-level relationships between individual lines. Separation by country of origin was not quantifiable, although some bias was suggested within clusters (e.g. one higher-level cluster contained 9/10 lines from Australia).

Genotypic variation

Genotypic differences did not appear to be related to either country of origin or to phenotype. Few lines within the diversity

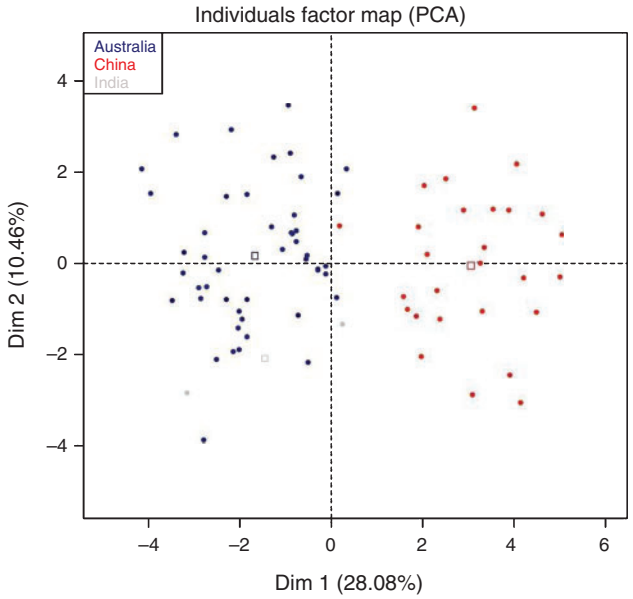

Fig. 2. Individuals factor map for the first two principle components of a phenotypic principle components analysis of yield and yield-related traits in a *Brassica napus* diversity set of 76 lines from Australia (47 lines), India (2 lines) and China (27 lines). Squares represent the midpoint of each country cluster. Yield traits assessed were: number of days from seeding to initiation of flowering in ~50% of the plants (DF50); days to maturity (DM); plant height in cm (PH); main shoot length in cm (MSL); number of primary branches (PB); number of secondary branches (SPB); number of siliquae on main shoot (MSP); silique length in cm (SL); average seed number in 10 random siliques (SSL); seed weight of 1000 seeds in grams (SS); seed yield in kg/ha (Yield); percentage seed oil content (Oil); foliar chlorophyll content measured using SPAD-502 (PP); leaf area in cm^2 (LA); leaf area index (LAI) and percentage seed protein content (Protein). Blackleg resistance as a function of percentage survival was assessed over three different locations: in Victoria (Lmac_Vic.), South Australia (Lmac_SA) and Western Australia (Lmac_WA). Sclerotinia resistance was assessed as stem lesion length (cm) in a single field location.

set were closely related, with no obvious subpopulation structure in the kinship matrix (Supplementary materials figure S1). PCA revealed no visible groupings within the population (Fig. 4), with only 9.4% and 7.1% of the variance explained by the first two principle components. Likewise, hierarchical clustering showed very low support for any differentiation between germplasm groups (Fig. 5). In fact, for several pairs of Australian and Chinese lines 100% support for the line from the other country as the closest genetic relative was obtained (Fig. 5). Global F_{ST} between the Australian and Chinese subpopulations was 0.0354; F_{ST} values between the Australian and Chinese subpopulations for individual SNP loci are presented in Supplementary materials table S2.

Several large regions of presence–absence variation were detected in the diversity set (Supplementary materials table S3). The most common was on chromosome C2, putatively representing a 5–10 Mbp duplication of a homeologous region on chromosome A2 with a corresponding deletion on chromosome C2. At least one small region of presence/absence variation was identified on every chromosome in the A and C genomes, with most events shared by more than one line (Figs 6, 7). Several much larger presence–absence variants (>1 Mbp) were also observed on chromosomes A7, A10, C1, C2, C6, C8 and C9 (Fig. 6). In the A genome one line had a deletion on A10 between 15 and 17 Mbp (end of chromosome) and another line had a deletion between 17 and 24 Mbp (end of chromosome). Five lines were missing a block between 32 and 35 Mbp on chromosome C1, and another line from 0 to 3 Mbp. Two lines had extremely large deletions comprising the majority of chromosome C2: one from ~8 Mbp to 45 Mbp (the end of the chromosome) and the other from ~20 Mbp to 45 Mbp. Three lines were missing from 25 to 35 Mbp on chromosome C6, one line from 6 to 9 Mbp on chromosome C8, one line from 0 to 8 Mbp on chromosome C9 and another line from 43 to 48 Mbp on the same chromosome.

Association analysis

No significant results were obtained for genotype–phenotype associations after correction for multiple testing. Presence–absence variants were scored irrespective of parent alleles to make a matrix of presence/absence of chromosome segments for all regions showing presence–absence variation in the population with >3 individuals with an absence variant. This genotype matrix was also tested for phenotype associations, but no significant results were found after correction for multiple testing.

Discussion

Genetic and phenotypic divergence between lines was not correlated in our germplasm set. Although this finding is worthy of further investigation given the small population size and high level of heterozygosity in these lines, it is not necessarily the case that genotypic and phenotypic diversity are related within a species (Kozak *et al*. 2011). *Brassica napus* has limited genetic diversity as a result of its formation as an allopolyploid from only a few hybridisation events between progenitor genotypes (Palmer *et al*. 1983; Allender and King 2010). Crop cultivars underwent further inbreeding as a result of intensive selection for reduced glucosinolates and erucic acid content in the 1970s, to produce modern ‘canola’ quality oil from rapeseed, as well as selection for yield traits and tolerances to biotic and abiotic stresses (Cowling 2007). Despite this low genetic diversity *B. napus* does have some population structure, with subpopulations of winter oilseed rape, summer/spring oilseed rape and swede (Bus *et al*. 2011), and some differentiation by country of origin (Chen *et al*. 2008; Xiao *et al*. 2012). However, recent evidence suggests that this population structure is not well supported genome-wide, with no strong division observable between either growth type or geographic origin when high density, evenly distributed markers are used to assess population genetic diversity (Wang *et al*. 2014). Instead, several localised selective sweeps have been hypothesised to differentiate the subpopulations based on differential selection for agriculturally significant traits (Wang *et al*. 2014). Our findings support this result, with genome-wide SNPs failing to clearly differentiate between Australian and Chinese spring-type lines, but with evidence for strong agricultural selective bias differentiating these two groups on the basis of phenotype, and some evidence of genetic

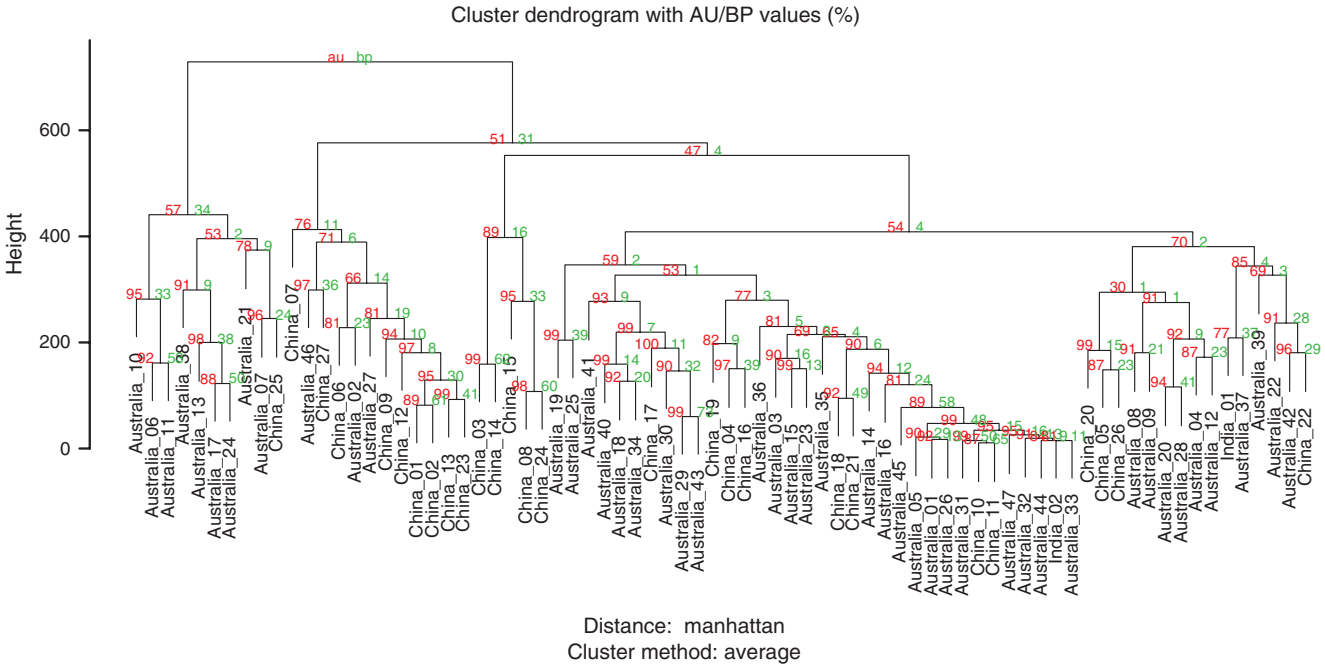

Fig. 3. Hierarchical clustering analysis for individual relationships based on yield and yield-related traits data in a *Brassica napus* diversity set of 76 lines from Australia (47 lines), India (2 lines) and China (27 lines). Yield traits assessed were: number of days from seeding to initiation of flowering in ~50% of the plants (DF50); days to maturity (DM); plant height in cm (PH); main shoot length in cm (MSL); number of primary branches (PB); number of secondary branches (SPB); number of siliquae on main shoot (MSP); silique length in cm (SL); average seed number in 10 random siliques (SSL); seed weight of 1000 seeds in grams (SS); seed yield in kg/ha (Yield); percentage seed oil content (Oil); foliar chlorophyll content measured using SPAD-502 (PP); leaf area in cm^2 (LA); leaf area index (LAI) and percentage seed protein content (Protein). Blackleg resistance as a function of percentage survival was assessed over three different locations: in Victoria (Lmac_Vic.), South Australia (Lmac_SA) and Western Australia (Lmac_WA). Sclerotinia resistance was assessed as stem lesion length (cm) in a single field location.

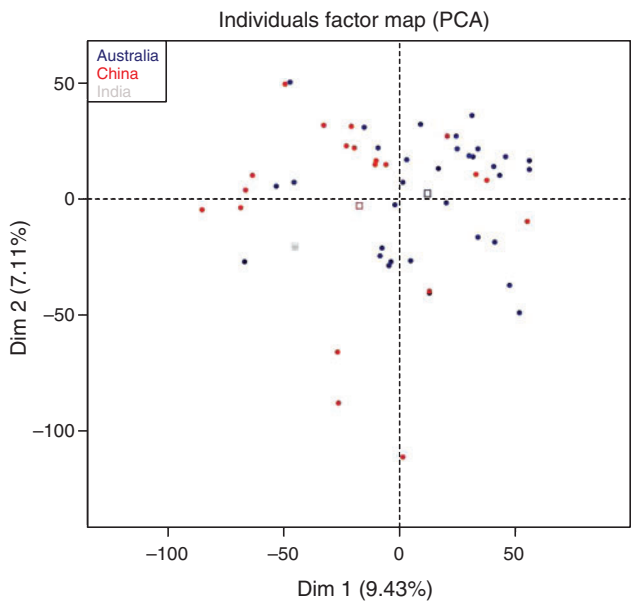

Fig. 4. Individuals factor map for the first two principle components of a principle components analysis of genome-wide single nucleotide polymorphism marker data in a *Brassica napus* diversity set of lines from Australia (33 lines), India (1 line) and China (20 lines). Squares represent the midpoint of each country cluster.

differentiation of the two subpopulations at particular genomic locations (Supplementary materials table S2). However, further investigation in larger germplasm sets should be undertaken to confirm this: intercrossing between the Australian and Chinese lines that make up the diversity set could also have muddied the waters. Possibly, the use of a SNP array for these analyses may also be influencing this result, as genetic diversity detection using SNP arrays is biased by the choice of SNPs used to construct the arrays (Mason *et al*. 2017), even though care was taken to minimise these effects and select an internationally representative marker set for the Illumina Infinium 60K *Brassica* array (Clarke *et al*. 2016).

As well as an overall correlation between yield traits and country of origin, some trends were also visible in the data. Lines originating from Australia were more likely to be similar in terms of resistance to blackleg disease, chlorophyll content, number of secondary branches and seed protein content, whereas lines originating from China were more likely to be similar in plant height and main shoot length, silique length and number of siliques, days to maturity and days to flowering. The similarity in blackleg resistance trends in Australian lines is unsurprising, given the importance of resistance to this disease under Australian growing conditions (Li *et al*. 2003; Kaur *et al*. 2009) and the targeted breeding that has occurred to develop blackleg-resistant lines in Australia (Salisbury *et al*. 1995). As well, blackleg is not a serious disease risk to crop production in China (West *et al*. 2001), so the differentiation of Australian and Chinese lines in

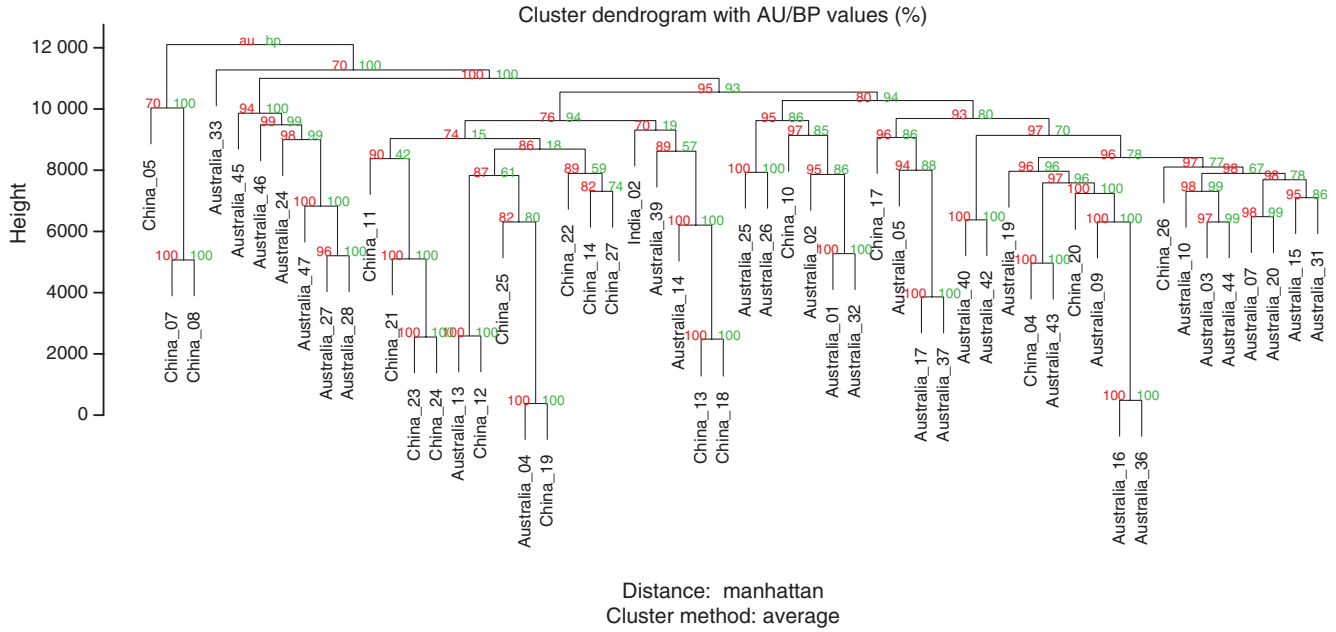

Fig. 5. Hierarchical clustering analysis for individual relationships based on genome-wide single nucleotide polymorphism marker data in a *Brassica napus* diversity set of lines from Australia (33 lines), India (1 line) and China (20 lines).

blackleg disease resistance phenotypes is almost certainly due to different agricultural selection pressures in these two germplasm groups. Other traits showing differentiation between Australian and Chinese groups, such as flowering time and days to maturity, are expected to be mainly due to localised adaptation to climatic factors such as temperature and daylength under the two different environments. Traits such as number of secondary branches and protein content in seeds may also reflect agricultural practices such as ease of mechanised harvesting (done in Australia but not in China) and economic desirability of high-protein seed meal for animal feed. That Sclerotinia disease-resistant phenotypes did not differentiate between China and Australia is not surprising, as the top 6, 16, 20, or 24 most resistant *B. napus* genotypes in the study of Li *et al.* (2008) included an equal number of genotypes each from China and from Australia (3, 8, 10 and 12, respectively). Li *et al.* (2008) found that cultivars from both countries offered similar expression levels and frequencies of resistance to Sclerotinia.

The release of the *B. napus* genome sequence revealed 17 translocated regions in sequenced cultivar 'Darmor' between the A and C genomes (relative to the diploid progenitor genomes), as well as several translocations in another six natural *B. napus* cultivars and a resynthesised (*B. rapa* × *B. oleracea*) line (Chalhoub *et al.* 2014). In *B. napus*, the observation of deletions using SNP markers almost always indicates the presence of a non-reciprocal homeologous translocation event; i.e. a duplication-deletion event whereby part of an A-genome chromosome has been replaced by part of a C-genome chromosome, resulting in two C genome copies (the duplication) and no A genome copies (the deletion), or vice versa (see Mason *et al.* 2017) for more details and explanatory figures). Using the SNP data to detect 'deletions' indicative of these translocation regions, we identified many of these same events (e.g. on C2 and C8), while additionally finding many more distributed across all A and C genome chromosomes. Although it is possible that some of these deletions are due instead to high levels of genetic divergence within small regions in these cultivars, such that the SNP probes are no longer able to bind, there is little evidence to support radically different rates of genomic evolution between different non-genic regions within the *B. napus* genome. Additionally, the criteria used for defining PAV in our study (contiguous SNPs, presence of deletions in more than one individual) make this an unlikely explanation relative to the presence of translocations. The same trend observed in Chalhoub *et al.* (2014) of increased deletions on the C genome was also observed in our study, as well as the distribution of translocation events towards the telomeres. Although most of the translocations were small (<1 Mbp), similar to the majority of those detected by Chalhoub *et al.* (2014), several much larger translocations were also identified. The most striking of these involved almost the entire C2 chromosome, from 8 – 45 Mbp; a length previously only identified in unstable synthetic (*B. rapa* × *B. oleracea*) lines (Szadkowski *et al.* 2010; Chalhoub *et al.* 2014) rather than in 'natural' *B. napus*. Two large A-genome translocations were also identified, one of which corresponded to the A07–C6 reciprocal translocation first identified in spring-type *B. napus* (Osborn *et al.* 2003), and where the individual in our study with the A07 deletion may have been derived from a cross between a line carrying the reciprocal translocation and a line without, in which case a non-reciprocal translocation (the observed 'deletion-duplication' event) could be generated. Another deletion event was observed on the end of A10, interestingly also where a duplication event was previously observed in a *B. napus* line used as a parent in an interspecific hybrid cross (Mason *et al.* 2015). Many translocation events we observed covered the

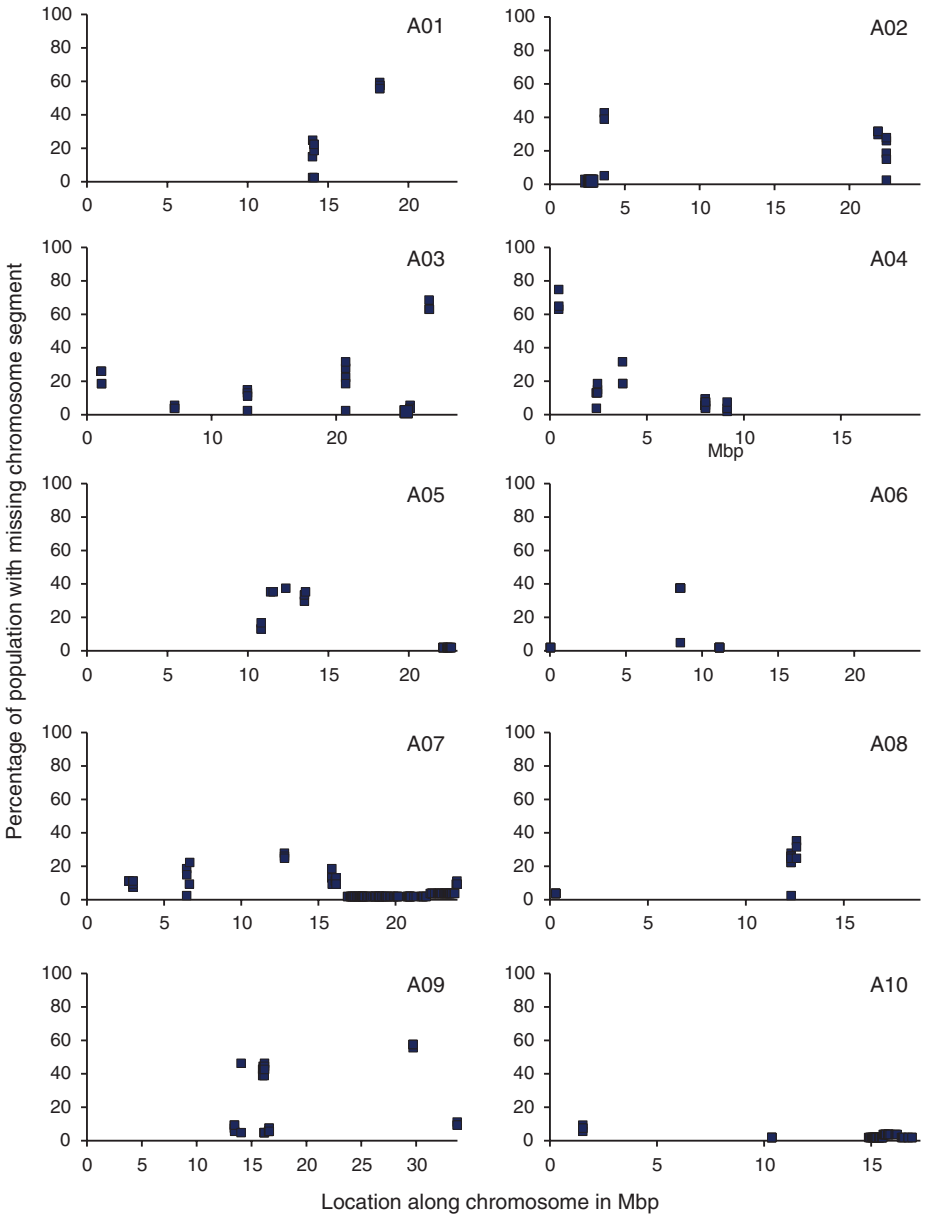

Fig. 6. Frequency and distribution of presence–absence variants in the *Brassica napus* A genome based on genome-wide single nucleotide polymorphism marker data in a diversity set of lines from Australia (33 lines), India (1 line) and China (20 lines).

same regions but had different lengths. These observations taken together may suggest that large translocation events occur infrequently in *B. napus*, perhaps initially at the sizes observed in synthetic *B. napus*, but tend to be reduced in size over time by intercrossing and selection pressure. These events are putatively tolerated without serious loss of viability as a result of dosage compensation effects between the A and C genomes, as has previously been observed in *B. napus* synthetics (Xiong *et al*. 2011). Rather than being fixed in the *B. napus* population, large A-C translocation events are probably quickly broken down as a result of recombination with unrearranged genotypes, with the remaining regions putatively being selected for due to positive influences on yield traits (Liu *et al*. 2012; Chalhoub *et al*. 2014).

The inability of our association mapping approach to pick up any statistically significant associations between traits and marker data after multiple testing correction is not predicted to be a fault of the association mapping method, which is known to work well in *B. napus* for mapping yield traits (Zou *et al*. 2010; Gajardo *et al*. 2015). Rather, the small population size and residual heterozygosity in lines in our study probably resulted in a lack of statistical power to detect these associations. The heterozygosity detected in these lines was clearly a result of outcrossing rather than a technical issue. Although the Illumina

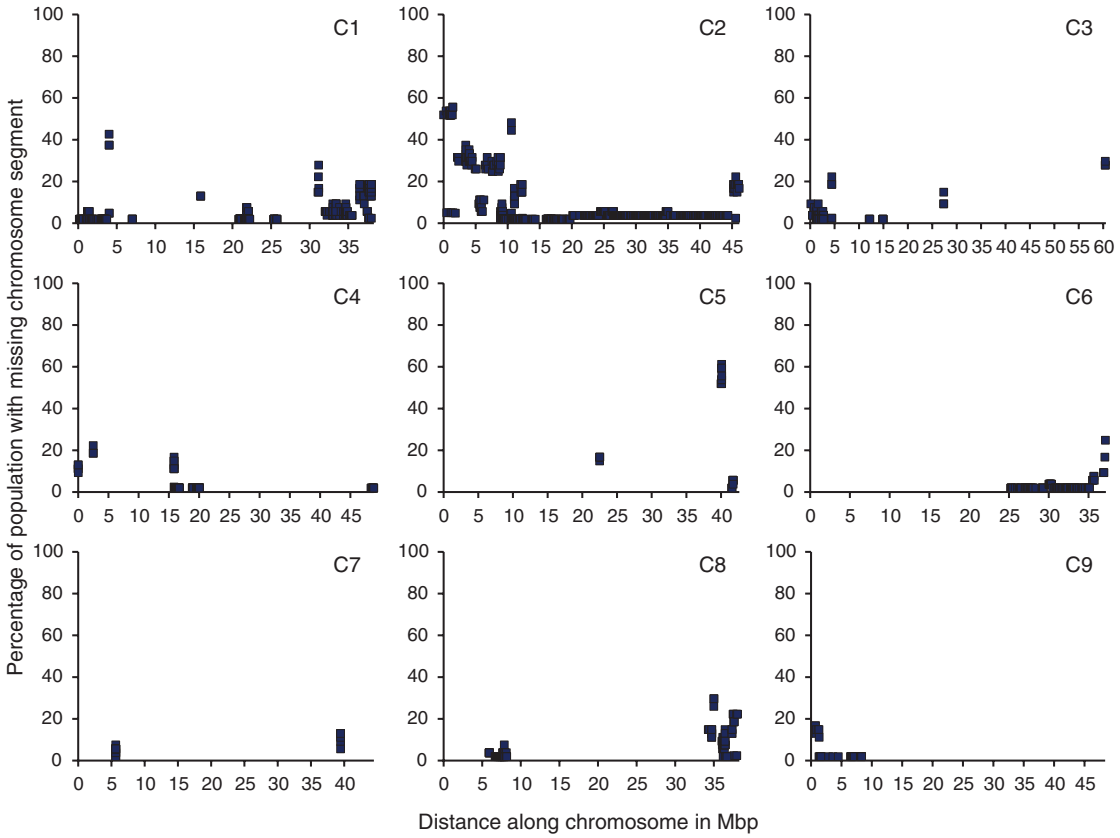

Fig. 7. Frequency and distribution of presence–absence variants in the *Brassica napus* C genome based on genome-wide single nucleotide polymorphism marker data in a diversity set of lines from Australia (33 lines), India (1 line) and China (20 lines).

Infinium SNP array does produce higher levels of heterozygous calls with poor sample amplification (Mason *et al*. 2017), in this case heterozygosity was present in large blocks indicative of recombination events, rather than randomly distributed throughout the genome. Unexpectedly high frequencies of heterozygosity are commonly observed in *Brassica* inbred and doubled-haploid lines (Bayer *et al*. 2015), as inadvertent outcrossing via pollen contamination (Arús *et al*. 1982; Pascher *et al*. 2010) and volunteer seeds in the field (Friesen *et al*. 2003) are common vectors for seed lot contamination. This problem is not new, as no enforced self-pollination system is able to perfectly exclude foreign pollen while still maintaining air flow to the developing racemes in *Brassica*. Although *B. napus* is self-compatible, frequencies of outcrossing under bee-pollination may be 80% or more (Cresswell 1994). Acknowledging this problem in *B. napus*, particularly in field trials or genetic studies of elite crop cultivars, rather than excluding heterozygous samples from publications, may give rise to better solutions for self-pollination or a wider understanding of how pollen flow, heterozygosity and outcrossing dictate crop phenotypes in the field.

Author contributions

SSB, SB, JB and PS conceived of the study and experimental design/material, PC carried out field-based phenotyping and DNA extractions, ASM analysed and interpreted the data and wrote the manuscript, and SSB, SB, JB, MB and PS contributed to manuscript revisions.

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Acknowledgements

ASM is funded by DFG grant MA6473/1-1. Phenotyping was carried out with financial assistance received from the Indian Council of Agricultural research (ICAR), the Australian Centre for International Agricultural Research (ACIAR) and the Grains Research and Development Corporation (GRDC), Australia, through a collectively funded research project ‘Oilseed *Brassica* improvement in China, India and Australia’.

References

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

Arús P, Tanksley SD, Orton TJ, Jones RA (1982) Electrophoretic variation as a tool for determining seed purity and for breeding hybrid varieties of *Brassica oleracea*. *Euphytica* **31**, 417–428. doi:10.1007/BF00021659

Bayer PE, Ruperao P, Mason AS, Stiller J, Chan C-KK, Hayashi S, Long Y, Meng J, Sutton T, Visendi P, Varshney RK, Batley J, Edwards D (2015)

High-resolution skim genotyping by sequencing reveals the distribution of crossovers and gene conversions in *Cicer arietinum* and *Brassica napus*. *Theoretical and Applied Genetics* **128**, 1039–1047. doi:10.1007/s00122-015-2488-y

Bus A, Korber N, Snowdon RJ, Stich B (2011) Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. *Theoretical and Applied Genetics* **123**, 1413–1423. doi:10.1007/s00122-011-1676-7

Chalhoub B, Denoeud F, Liu SY, Parkin IAP, Tang HB, Wang XY, Chiquet J, Belcram H, Tong CB, Samans B, Correa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao MX, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier MC, Fan GY, Renault V, Bayer PE, Golicz AA, Manoli S, Lee TH, Thi VHD, Chalabi S, Hu Q, Fan CC, Tollenaere R, Lu YH, Battail C, Shen JX, Sidebottom CHD, Wang XF, Canaguier A, Chauveau A, Berard A, Deniot G, Guan M, Liu ZS, Sun FM, Lim YP, Lyons E, Town CD, Bancroft I, Wang XW, Meng JL, Ma JX, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury JM, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou YM, Hua W, Sharpe AG, Paterson AH, Guan CY, Wincker P (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950–953. doi:10.1126/science.1253435

Chan E (2008) Handy R functions for genetics research. Github repository. Available at: https://github.com/ekfchan/evachan.org-Rscripts

Chen S, Nelson MN, Ghamkhar K, Fu T, Cowling WA (2008) Divergent patterns of allelic diversity from similar origins: the case of oilseed rape (*Brassica napus* L.) in China and Australia. *Genome* **51**, 1–10. doi:10.1139/G07-095

Cifuentes M, Grandont L, Moore G, Chèvre AM, Jenczewski E (2010) Genetic regulation of meiosis in polyploid species: new insights into an old question. *New Phytologist* **186**, 29–36. doi:10.1111/j.1469-8137.2009.03084.x

Clarke WE, Higgins EE, Plieske J, Wieseke R, Sidebottom C, Khedikar Y, Batley J, Edwards D, Meng J, Li R, Lawley CT, Pauquet J, Laga B, Cheung W, Iniguez-Luy F, Dyrszka E, Rae S, Stich B, Snowdon RJ, Sharpe AG, Ganal MW, Parkin IAP (2016) A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theoretical and Applied Genetics* **129**, 1887–1899. doi:10.1007/s00122-016-2746-7

Cowling WA (2007) Genetic diversity in Australian canola and implications for crop breeding for changing future environments. *Field Crops Research* **104**, 103–111. doi:10.1016/j.fcr.2006.12.014

Cresswell JE (1994) A method for quantifying the gene flow that results from a single bumblebee visit using transgenic oilseed rape, *Brassica napus* L. cv. Westar. *Transgenic Research* **3**, 134–137. doi:10.1007/BF01974092

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

Friesen LF, Nelson AG, Van Acker RC (2003) Evidence of contamination of pedigreed canola (*Brassica napus*) seedlots in western Canada with genetically engineered herbicide resistance traits. *Agronomy Journal* **95**, 1342–1347. doi:10.2134/agronj2003.1342

Gajardo HA, Wittkop B, Soto-Cerda B, Higgins EE, Parkin IAP, Snowdon RJ, Federico ML, Iniguez-Luy FL (2015) Association mapping of seed quality traits in *Brassica napus* L. using GWAS and candidate QTL approaches. *Molecular Breeding* **35**, 143. doi:10.1007/s11032-015-0340-3

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

Kaur S, Cogan NOI, Ye G, Baillie RC, Hand ML, Ling AE, Mcgearey AK, Kaur J, Hopkins CJ, Todorovic M, Mountford H, Edwards D, Batley J, Burton W, Salisbury P, Gororo N, Marcroft S, Kearney G, Smith KF, Forster JW, Spangenberg GC (2009) Genetic map construction and QTL mapping of resistance to blackleg (*Leptosphaeria maculans*) disease in Australian canola (*Brassica napus* L.) cultivars. *Theoretical and Applied Genetics* **120**, 71–83. doi:10.1007/s00122-009-1160-9

Körber N, Bus A, Li JQ, Parkin IAP, Wittkop B, Snowdon RJ, Stich B (2016) Agronomic and seed quality traits dissected by genome-wide association mapping in *Brassica napus*. *Frontiers in Plant Science* **7**, 386. doi:10.3389/fpls.2016.00386

Kozak M, Bocianowski J, Liersch A, Tartanus M, Bartkowiak-Broda I, Piotto FA, Azevedo RA (2011) Genetic divergence is not the same as phenotypic divergence. *Molecular Breeding* **28**, 277–280. doi:10.1007/s11032-011-9583-9

Lê S, Josse J, Husson F (2008) FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software* **25**, 1–18. doi:10.18637/jss.v025.i01

Li H, Sivasithamparam K, Barbetti MJ (2003) Breakdown of a *Brassica rapa* subsp. *sylvestris* single dominant blackleg resistance gene in *B. napus* rapeseed by *Leptosphaeria maculans* field isolates in Australia. *Plant Disease* **87**, 752. doi:10.1094/PDIS.2003.87.6.752A

Li CX, Li H, Sivasithamparam K, Fu TD, Li YC, Liu SY, Barbetti MJ (2006) Expression of field resistance under Western Australian conditions to *Sclerotinia sclerotiorum* in Chinese and Australian *Brassica napus* and *Brassica juncea* germplasm and its relation with stem diameter. *Australian Journal of Agricultural Research* **57**, 1131–1135. doi:10.1071/AR06066

Li CX, Wratten N, Salisbury PA, Burton WA, Potter TD, Walton G, Li H, Sivasithamparam K, Banga SS, Banga S, Singh D, Liu SY, Fu TD, Barbetti MJ (2008) Response of *Brassica napus* and *B. juncea* germplasm from Australia, China and India to Australian populations of *Leptosphaeria maculans*. *Australasian Plant Pathology* **37**, 162–170. doi:10.1071/AP08005

Li CX, Liu SY, Sivasithamparam K, Barbetti MJ (2009) New sources of resistance to *Sclerotinia* stem rot caused by *Sclerotinia sclerotiorum* in Chinese and Australian *Brassica napus* and *B. juncea* germplasm screened under Western Australian conditions. *Australasian Plant Pathology* **38**, 149–152. doi:10.1071/AP08087

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399. doi:10.1093/bioinformatics/bts444

Liu LZ, Stein A, Wittkop B, Sarvari P, Li JN, Yan XY, Dreyer F, Frauen M, Friedt W, Snowdon RJ (2012) A knockout mutation in the lignin biosynthesis gene CCR1 explains a major QTL for acid detergent lignin content in *Brassica napus* seeds. *Theoretical and Applied Genetics* **124**, 1573–1586. doi:10.1007/s00122-012-1811-0

Liu S, Fan CC, Li JN, Cai GQ, Yang QY, Wu J, Yi XQ, Zhang CY, Zhou YM (2016) A genome-wide association study reveals novel elite allelic variations in seed oil content of *Brassica napus*. *Theoretical and Applied Genetics* **129**, 1203–1215. doi:10.1007/s00122-016-2697-z

Mason AS, Snowdon RJ (2016) Oilseed rape: learning about ancient and recent polyploid evolution from a recent crop species. *Plant Biology*. doi:10.1111/plb.12462In press

Mason AS, Takahira J, Atri C, Samans B, Hayward A, Cowling WA, Batley J, Nelson MN (2015) Microspore culture reveals complex meiotic behaviour in a trigenomic *Brassica* hybrid. *BMC Plant Biology* **15**, 173. doi:10.1186/s12870-015-0555-9

Mason AS, Higgins EE, Snowdon RJ, Batley J, Stein A, Werner C, Parkin IAP (2017) A user guide to the Brassica 60K Illumina Infinium (TM) SNP genotyping array. *Theoretical and Applied Genetics* **130**, 621–633. doi:10.1007/s00122-016-2849-1

Morinaga T (1934) Interspecific hybridisation in *Brassica* VI. The cytology of F_1 hybrids of *B. juncea* and *B. nigra*. *Cytologia* **6**, 62–67. doi:10.1508/cytologia.6.62

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

Palmer JD, Shields CR, Cohen DB, Orton TJ (1983) Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. *Theoretical and Applied Genetics* **65**, 181–189. doi:10.1007/BF00308062

Pascher K, Macalka S, Rau D, Gollmann G, Reiner H, Glossl J, Grabherr G (2010) Molecular differentiation of commercial varieties and feral populations of oilseed rape (*Brassica napus* L.). *BMC Evolutionary Biology* **10**, 63. doi:10.1186/1471-2148-10-63

R Development Core Team (2015) 'R: A language and environment for statistical computing.' 2.10.1 edn. (R Foundation for Statistical Computing: Vienna, Austria)

Salisbury PA, Ballinger DJ, Wratten N, Plummer KM, Howlett BJ (1995) Blackleg disease on oilseed *Brassica* in Australia - a review. *Australian Journal of Experimental Agriculture* **35**, 665–672. doi:10.1071/EA9950665

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

Schmutzer T, Samans B, Dyrska E, Lespinasse D, Micic Z, Abel S, Duchscherer P, Breuer F, Abbadi A, Leckband G, Snowdon RJ, Scholz U (2015) Species-wide genome sequence and nucleotide polymorphism datasets from the model allopolyploid plant *Brassica napus*. *Scientific Data* **2**, 150072. doi:10.1038/sdata.2015.72

Szadkowski E, Eber F, Huteau V, Lodé M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ, Chalhoub B, Jenczewski E, Chèvre AM (2010) The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytologist* **186**, 102–112. doi:10.1111/j.1469-8137.2010.03182.x

U (1935) Genome-analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* **7**, 389–452.

VanRaden P (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414–4423. doi:10.3168/jds.2007-0980

Wang N, Li F, Chen BY, Xu K, Yan GX, Qiao JW, Li J, Gao GZ, Bancroft I, Meng JL, King GJ, Wu XM (2014) Genome-wide investigation of genetic changes during modern breeding of *Brassica napus*. *Theoretical and Applied Genetics* **127**, 1817–1829. doi:10.1007/s00122-014-2343-6

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

West JS, Kharbanda PD, Barbetti MJ, Fitt BDL (2001) Epidemiology and management of *Leptosphaeria maculans* (phoma stem canker) on oilseed rape in Australia, Canada and Europe. *Plant Pathology* **50**, 10–27. doi:10.1046/j.1365-3059.2001.00546.x

Xiao YJ, Cai DF, Yang W, Ye W, Younas M, Wu JS, Liu KD (2012) Genetic structure and linkage disequilibrium pattern of a rapeseed (*Brassica napus* L.) association mapping panel revealed by microsatellites. *Theoretical and Applied Genetics* **125**, 437–447. doi:10.1007/s00122-012-1843-5

Xiong Z, Gaeta RT, Pires JC (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proceedings of the National Academy of Sciences, USA* **108**, 7908–7913.

Xu LP, Hu KN, Zhang ZQ, Guan CY, Chen S, Hua W, Li JN, Wen J, Yi B, Shen JX, Ma CZ, Tu JX, Fu TD (2016) Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.). *DNA Research* **23**, 43–52.

Zhou Q-H, Fu D-H, Mason AS, Zeng Y-J, Zhao C-X, Huang Y-J (2014) *In silico* integration of quantitative trait loci for seed yield and yield-related traits in *Brassica napus*. *Molecular Breeding* **33**, 881–894. doi:10.1007/s11032-013-0002-2

Zou J, Jiang CC, Cao ZY, Li RY, Long Y, Chen S, Meng JL (2010) Association mapping of seed oil content in *Brassica napus* and comparison with quantitative trait loci identified from linkage mapping. *Genome* **53**, 908–916. doi:10.1139/G10-075

Zou J, Fu DH, Gong HH, Qian W, Xia W, Pires JC, Li RY, Long Y, Mason AS, Yang TJ, Lim YP, Park BS, Meng JL (2011) *De novo* genetic variation associated with retrotransposon activation, genomic rearrangements and trait variation in a recombinant inbred line population of *Brassica napus* derived from interspecific hybridization with *Brassica rapa*. *The Plant Journal* **68**, 212–224. doi:10.1111/j.1365-313X.2011.04679.x