ORIGINAL ARTICLE

Homoeologous non-reciprocal translocation explains a major QTL for seed lignin content in oilseed rape (*Brassica napus* L.)

Hanna Marie Schilbert[1,2] · Karin Holzenkamp[3] · Prisca Viehöver[1] · Daniela Holtgräwe[1] · Christian Möllers[3]

Abstract

Key message **A homoeologous non-reciprocal translocation was identified in the major QTL for seed lignin content in the low lignin line SGDH14. The lignin biosynthetic gene PAL4 was deleted.**

Abstract Oilseed rape is a major oil crop and a valuable protein source for animal and human nutrition. Lignin is a non-digestible, major component of the seed coat with negative effect on sensory quality, bioavailability and usage of oilseed rape's protein. Hence, seed lignin reduction is of economic and nutritional importance. In this study, the major QTL for reduced lignin content found on chromosome C05 in the DH population SGDH14 x Express 617 was further examined. SGDH14 had lower seed lignin content than Express 617. Harvested seeds from a F2 population of the same cross were additionally field tested and used for seed quality analysis. The F2 population showed a bimodal distribution for seed lignin content. F2 plants with low lignin content had thinner seed coats compared to high lignin lines. Both groups showed a dark seed colour with a slightly lighter colour in the low lignin group indicating that a low lignin content is not necessarily associated with yellow seed colour. Mapping of genomic long-reads from SGDH14 against the Express 617 genome assembly revealed a homoeologous non-reciprocal translocation (HNRT) in the confidence interval of the major QTL for lignin content. A homologous A05 region is duplicated and replaced the C05 region in SGDH14. As consequence several genes located in the C05 region were lost in SGDH14. Thus, a HNRT was identified in the major QTL region for reduced lignin content in the low lignin line SGDH14. The most promising candidate gene related to lignin biosynthesis on C05, *PAL4*, was deleted.

Introduction

Oilseed rape (*Brassica napus* L.) is one of the most important oil crops worldwide. The allopolyploid species *B. napus* (AACC, $2n = 38$) was formed by a spontaneous hybridisation and genome doubling between the two diploid species *Brassica rapa* (AA, $2n = 20$) and *Brassica oleracea* (CC,

Hanna Marie Schilbert and Karin Holzenkamp have contributed equally to this work.

✉ Hanna Marie Schilbert
  hschilbe@cebitec.uni-bielefeld.de

[1] Genetics and Genomics of Plants, CeBiTec and Faculty of Biology, Bielefeld University, Bielefeld, Germany

[2] Graduate School DILS, Bielefeld Institute for Bioinformatics Infrastructure (BIBI), Faculty of Technology, Bielefeld University, Bielefeld, Germany

[3] Department of Crop Sciences, Division of Crop Plant Genetics, Georg-August-University, Göttingen, Germany

$2n = 18$) more than 7500 years ago (Chalhoub et al. 2014; Lu et al. 2019; Nagaharu 1935). Recombination is one of the major sources of genetic variation and for novel allelic combinations. In many allopolyploid species homoeologous exchanges (HEs) can occur. Thus, mispairing between ancestral chromosomes can lead to deletions, duplications and translocations (Sharpe et al. 1995; Mason and Wendel 2020). Meiotic chromosome pairing that occurs between homoeologous chromosomes with a high sequence similarity leads to enhanced HEs and gene conversion events both initiated by the repair of double-strand breaks (DSB) throughout the *B. napus* genome. These HEs range in size from large segments to single SNPs. HEs are highly enriched in transcriptional regions generating novel gene combinations and can potentially cause phenotypic alterations (Chalhoub et al. 2014; Gaeta and Pires 2010; Stein et al. 2017; Zhang et al. 2020). HEs can result in reciprocal exchanges between homoeologous chromosomes during meiosis via crossover, where parts of the homoeologous chromosomes are exchanged between the two subgenomes. Alternatively, HEs can result as a non-reciprocal translocation of

chromosomal regions also known as homoeologous non-reciprocal translocations (HNRTs). These HNRTs are duplication/deletion events where an additional copy of a DNA sequence replaces its homoeologous copy leading to gain and loss of genetic material (reviewed in Gaeta and Pires (2010) and Mason and Wendel (2020)).

Synthetic polyploids and recent allopolyploids contain frequent HEs. For established *B. napus* (Chalhoub et al. 2014) and resynthesised *B. napus* (Szadkowski et al. 2010; Gaeta and Pires 2010), HEs are documented and indicate a diversification in the polyploid genome. In the *B. napus* genome, genomic segments and genes can be found in multiple copies. Two copies of the A and C subgenomes which share the most recent ancestry are primary homologues, while the remaining ones are secondary homologues and paralogous to one or the other copy (Nicolas et al. 2007; Mason and Wendel 2020). It was demonstrated that regions of primary homology show a high recombination rate of chromosomal rearrangements. This represents an up to 100-fold increase compared to the frequency of homologous recombination measured in euploid lines (Nicolas et al. 2007). Further, for *B. napus* a subgenome bias towards replacing larger C-subgenome segments by smaller, homoeologous A-subgenome segments was observed (Samans et al. 2017; Sochorová et al. 2017; Higgins et al. 2018). The reciprocal allele gain and loss observed between the A and C subgenomes ranged from small segments to entire chromosomes, with low quantity for aneuploidy across all gametes in *B. napus* (Higgins et al. 2018). HNRTs and HEs have been conclusively linked to phenotypic changes in *B. napus* like seed yield, seed quality, seed colour, seed fibre content, flowering time, disease resistance and fertility (Hurgobin et al. 2018; Schiessl et al. 2019; Stein et al. 2017). Having both adaptive and agronomic importance, HNRTs and HEs could play an important role in the breeding of improved allopolyploid crops.

Due to a growing demand of plant-based protein for food and feed (Ismail et al. 2020), the breeding aim in *B. napus* is increasing the seed oil and protein content (Si et al. 2003). However, in comparison to soybean meal, which is the most important vegetable protein source, oilseed rape meal has a lower protein content as well as higher anti-nutritive fibre compounds (Nesi et al. 2008). Therefore, genetic reduction of anti-nutritive and non-digestible compounds, such as lignin, is required to develop genotypes with improved nutritional quality for the production of optimised vegetable protein products (Wittkop et al. 2009). Lignin is mainly found in the seed coat. Hence, a low lignin content is associated with less seed coat phenolic compounds and a lighter seed colour (Carré et al. 2016; Wittkop et al. 2012). Although cultivars with some lower lignin content are available, breeding

efforts to develop competitive agronomic yellow seeded winter oilseed rape cultivars with low lignin contents were unsuccessful to date (Rahman and McVetty 2011; Simbaya et al. 1995).

In diverse populations a large number of quantitative trait loci (QTL) for reduced lignin content on A02, A05, A07, A09, C02, C03, C05 and C07 (Behnke et al. 2018; Wang et al. 2015; Körber et al. 2016) and yellow seed colour on A09 and C08 (Badani et al. 2006; Wang et al. 2017) or both traits (Stein et al. 2013; Liu et al. 2013) were detected on different linkage groups. A major QTL for low lignin content was found on C05 explaining 81% of the phenotypic variation in the doubled haploid (DH) population derived from the cross of genotype SGDH14 with inbred line 617 of the cultivar Express (Express 617) (Behnke et al. 2018). This QTL co-located with a QTL for oil and protein content with opposite additive effects. The DH population showed a large bimodal distribution for the acid detergent lignin (ADL, synonym: lignin) content ranging from 6.5 to 14.5%. SGDH14 was with 8.3% the low lignin parental genotype compared to Express 617 with 12%. The reduction in lignin content was accompanied by a significant reduction of the seed coat proportion (Behnke et al. 2018).

Lignin is a large ubiquitous aromatic polymer and a major cell wall component. There are three main types of lignin, *p*-hydroxyphenyl lignin (H-type lignin), syringyl lignin (S-type lignin) and guaiacyl lignin (G-type lignin). All of which are formed by polymerisation of different phenylpropanoids, predominantly the monolignols *p*-coumaryl, coniferyl and sinapyl alcohols varying in their degree of methoxylation (Fraser and Chapple 2011; Ding et al. 2021; Vanholme et al. 2010). The first step in the general phenylpropanoid metabolism is catalysed by phenylalanine ammonia-lyase (PAL) which plays an important role in the biosynthesis of flavonoid- and lignin-related metabolites. The *PAL* gene family in *Arabidopsis thaliana* consists of four paralogues named *AthPAL1 – AthPAL4* (Raes et al. 2003). Expression analysis of the *PAL* genes showed that *AthPAL3* is expressed at basal levels in stems, *AthPAL1*, *AthPAL2* and *AthPAL4* are expressed with high levels in stems of plants in later growth stages, while *AthPAL2* and *AthPAL4* are expressed in seeds (Fraser and Chapple 2011; Raes et al. 2003). In oilseed rape, the corresponding *B. napus* homologue *PAL4* was found as the key candidate gene associated with lignin biosynthesis, possessing a positive influence on seed lignin content (Behnke et al. 2018; Wang et al. 2015).

The objective of this work was to further investigate the major QTL found on C05 for low lignin content in the parental genotypes and progeny of the low fibre line SGDH14 and the high fibre line Express 617. In-depth sequence analysis

was conducted by mapping ONT long-reads derived from SGDH14 against the Express 617 genome assembly (Lee et al. 2020). RNA-Seq data of developing seeds were used to analyse transcriptional changes between the two genotypes. The sequencing results revealed a homologous non-reciprocal translocation (HNRT) of an A05 sequence fragment replacing a homoeologous fragment on C05, which is associated with a reduction of lignin content in SGDH14.

Materials and methods

Plant material

The plant material comprised the *B. napus* genotypes SGDH14 and Express 617 which were previously described by Behnke et al. (2018). The oil-rich genotype SGDH14 is derived from the cross between the cultivars Sollux and Gaoyou analysed by Zhao et al. (2005). Express 617, widely used in mapping populations, is the inbred line 617 of the German winter oilseed rape cultivar Express with low erucic acid, low glucosinolate and high oil content (Behnke et al. 2018; Lee et al. 2020). Phenotypic characteristics of SGDH14 in contrast to Express 617 are its high erucic acid and glucosinolate contents, as well as low lignin content in the seeds (Behnke et al. 2018). SGDH14 and Express 617 were used for nGBS, RNA-Seq analysis and SGDH14 was used for ONT sequencing. Moreover, previous published 60 K SNP data of Express 617 and SGDH14 and their corresponding DH population (SGDH, $n = 139$) was analysed (Behnke et al. 2018). Additional *B. napus* cultivars included in the Illumina 60 K SNP and nGBS analysis are Adriana, Sollux, SGEDH13, and Gaoyou. Moreover, a F2 population of the cross SGDH14 and Express 617 consisting of 195 lines, further referred to as SGEF2 population, was used for PCR analysis of the HNRT event and for the determination of the seed coat proportion.

Field experiments

The SGEF2 population was tested in a field experiment in one environment, Göttingen-Reinshof (51°29′51.6″N 9°55′55.2″E), in north-western Germany during the growing season 2019/2020. F1:2 seeds were sown in single row observation plots. Fertiliser, fungicides and insecticides were applied following standard schemes. Individual F2 plants ($n = 195$) were marked in the field and small amount of leaf material was collected at flowering time, freeze dried and used for DNA extraction (see polymerase chain reaction (PCR) method section). At maturity, F2:3 seeds were harvested from each open pollinated plant. Seeds were cleaned, dried to a moisture content of 5 to 8% and stored at a dry place at room temperature for further analysis.

Seed quality analysis—determination of crude fibre contents

Near Infrared Reflectance Spectroscopy (NIRS) was applied to predict seed quality and fibre traits. NIRS measurements were performed using 3 g F2:3 seed samples of the SGEF2 population in small ring cups and the FOSS monochromator model 6500 (NIRSystem Inc., Silverspring, MD, USA). Between 400 and 2498 nm absorbance values log at 2 nm intervals were recorded to create a NIR spectrum for each sample. Using the WinISI software (Version 4.12.0.15440, FOSS Analytical A/S, Denmark) and applying the calibration raps2020.eqa provided by VDLUFA Qualitätssicherung NIRS GmbH (Kassel, Germany) oil and protein content were determined in percent at 91% seed dry matter. Protein content in the defatted meal (PidM) was calculated by using NIRS predicted seed oil and seed protein content (both at 91% dry matter) as: % protein in the defatted meal (PidM) = [% protein / (100 − % oil)] × 100. The sum of seed oil and protein content (oil + protein) was obtained by forming the sum of oil and protein content at 91% seed dry matter. Glucosinolate is determined as $\mu Mol \cdot g^{-1}$ at 91% seed dry matter. Erucic acid is determined in % of the oil. The fibre components of the Neutral Detergent Fibre (NDF), Acid Detergent Fibre (ADF) and lignin in the defatted meal (Van Soest et al. 1991) were estimated using the calibration equation valid600div7.eqa (Dimov et al. 2012). The calibration was extended by adding 252 selected samples with a wide variation in seed fibre components. The hemicellulose and cellulose contents were calculated by subtracting ADF from NDF and lignin from ADF contents, respectively. In validation, the standard error of prediction corrected for the bias (SEP(C)) of the extended calibration was 1.78%, 1.62%, 0.93%, 1.14% and 1.48% for NDF, ADF, lignin, hemicellulose and cellulose content, respectively, in the defatted seed meal.

Determination of other seed traits

The degree of seed colour pigmentation was visually scored according to a scale from (1) yellow, (3) yellow brown, (5) brown, (7) dark brown and (9) for black and using intermediate scores (Widiarsih et al. 2021). Seed coat content was determined following the method described by Dimov et al. (2012). The seed coat proportion in % was determined from 100 seeds (approx. 500 mg) per genotype. The thousand-kernel weight (TKW) was obtained from weight conversion of 500 seeds.

PCR and KASP assay

DNA was isolated following the standard protocol of innu-PREP PlantDNA Kit (Analytik Jena GmbH, Germany) using

the SLS lysis solution. All PCRs were performed in 25 µL reactions in Bio-Rad 96-well thermocyclers. Each reaction contained 16.8 µL ddwater (HPLC grad, Chemsolute, Th. Geyer GmbH & Co. KG Germany), 5 µL 5X Biozym Reaction Buffer, 0.2 µL Biozym HS Taq DNA Polymerase (5 u/ µL) (Biozym Scientific GmbH, Germany), 1 µL 10 mM per oligonucleotide and 1 µL 10 ng DNA was added. The used PCR strategies and corresponding oligonucleotides and applied PCR parameters are shown in Figure S1, Table S1 and S2 of Supplementary Information 1. Gel electrophoresis was performed on an 1.2% agarose gel and 100 V for 30 min. Based on provided sequences anchored KASP primers were ordered from LGC, Pindar Road, Hoddesdon, Herts, EN11 0WZ, UK. KASP assay was performed according to the manufacturer's instructions.

DNA was isolated following the standard protocol of innuPREP PlantDNA Kit (Analytik Jena GmbH, Germany) using the SLS lysis solution. All PCRs were performed in 25 µL reactions in Bio-Rad 96-well thermocyclers. Each reaction contained 16.8 µL ddwater (HPLC grad, Chemsolute, Th. Geyer GmbH & Co. KG Germany), 5 µL 5X Biozym Reaction Buffer, 0.2 µL Biozym HS Taq DNA Polymerase (5 u/ µL) (Biozym Scientific GmbH, Germany), 1 µL 10 mM per oligonucleotide and 1 µL 10 ng DNA was added. The used PCR strategies and corresponding oligonucleotides and applied PCR parameters are shown in Figure S1, Table S1 and S2 of Supplementary Information 1. Gel electrophoresis was performed on an 1.2% agarose gel and 100 V for 30 min. Based on provided sequences anchored KASP primers were ordered from LGC, Pindar Road, Hoddesdon, Herts, EN11 0WZ, UK. KASP assay was performed according to the manufacturer's instructions.

Identification of the C05 homologous region in the express 617 genome assembly

For the identification of the homologous C05 region, 30 kbp up- and downstream of the described borders of the C05 major QTL (Behnke et al. 2018) were extracted based on the Darmor-*bzh* reference genome sequence v4.1 (Chalhoub et al. 2014). Next, these sequences were used for a BLASTn against the Express 617 assembly (Lee et al. 2020) to infer the localisation of the respective homologous C05 region.

ONT genomic reads analyses of SGDH14

To analyse the major QTL for seed lignin content on chromosome C05 (Behnke et al. 2018) in SGDH14 on genomic level, Oxford Nanopore Technology (ONT) was used for the generation of long-reads derived from a single SGDH14 plant. Genomic DNA from one young leaf

was extracted as described before (Siadjeu et al. 2020). DNA quality was assessed via NanoDrop and PicoGreen (Fluostar) measurement, as well as running a 1% agarose gel. The DNA was prepared with the short-read eliminator (SRE) XL kit, following library construction with the LSK109 kit and then loaded onto a MinION/GridION R9.4 flow cell. After a nuclease flush another library, prepared with the LSK109 kit and derived from a sample prepared with the SRE-XL kit, was loaded onto the same flow cell. A total output of 3.4 Gbp passed reads were generated by this run. For the third library generation and sequencing on another MinION/GridION R9.4 flow cell, the DNA was used without prior short fragment depletion and a library was prepared with the LSK109 kit. A total output of 7.67 Gbp passed reads were generated by this run. All reads were filtered for a min qscore > 7. After basecalling with Guppy v4.2.3 (Wick et al. 2019), reads were mapped to the *B. napus* Express 617 assembly (Lee et al. 2020) using minimap2 (Li 2018) yielding ~ 10 × genome-wide coverage. Minimap2 was run in ONT mode (-ax map-ont) and no secondary alignments were allowed (–secondary = no). BAM files were sorted and indexed via samtools (Li et al. 2009). Manually curation of regions of interest were performed by using IGV (Robinson et al. 2011) and BLAST (Altschul et al. 1990). Synteny plots were generated with jcvi/MCscan (Tang et al. 2008).

Determination and molecular validation of HNRT border sequences

For the identification of the border sequences of the HNRT event, a k-mer approach was used. First, https://github.com/hschilbert/Major_low_ADL_QTLborder-spanning reads of SGDH14 were divided into 15 bp k-mers. Afterwards, each k-mer was assigned to C05, A05, or considered unspecific. If a k-mer was present in both subgenomes (unspecific) it was excluded from further analyses, yielding *bona fide* subgenome-specific k-mers. These k-mers were then mapped back to the border-spanning read sequences to narrow down the exact position of the left and right substitution border. Finally, manual curation was used to extract subgenome-specific SNPs around the borders of the HNRT. These SNPs were further used for oligonucleotide design to validate the HNRT borders via Sanger sequencing (Table S1, Supplementary Information 1).

Coverage analysis

For the genomic coverage analysis BAM derived coverage files were generated as described before (Pucker and Brockington 2018). Next, the genomic coverage of specific regions

was compared to the chromosome-wide coverage of the corresponding chromosome to identify duplicated or deleted regions (Figure S2 and S3, Supplementary Information 2).

Functional annotation and in silico gene prediction

Genes were functionally annotated by transferring the Araport11 functional annotation to the *B. napus* Express 617 gene models as described before (Schilbert et al. 2021). Not annotated genes in the Express 617 annotation were identified via an in silico gene prediction using AUGUSTUS v3.3.3 Web server (Stanke and Morgenstern 2005). The predicted gene structures can be accessed via Table S3 of Supplementary Information 3. The functional annotation was assigned via a BLASTp search against the Arabidopsis database. Finally, the genes were incorporated into the Express 617 structural annotation file (.gff3) for further analysis.

nGBS analysis

To identify SNPs and genotyping *Brassica napus* cultivars, normalised Genotyping-by-Sequencing (nGBS) was used (LGC, Berlin, Germany). Genomic DNA was extracted from leaves of the *B. napus* cultivars Express 617, SGDH14, Sollux, Adriana, SGEDH13, and Gaoyou. For each genotype 150 bp paired-end reads were generated via Illumina on the NextSeq500/550. After assessing read quality by FastQC (Andrews 2010), reads were mapped to the *B. napus* Express 617 assembly (Lee et al. 2020) using BWA-MEM v.0.7.13 (Li 2013). Default parameters were applied and the –M flag was set to avoid spurious mappings. Mapping statistics were calculated via the flagstat function of samtools (Li et al. 2009) (Table S4, Supplementary Information 3).

Gene expression analysis: RNA extraction, library construction, and sequencing

For the gene expression analysis total RNA was extracted with the innuPREP Plant RNA Kit according to the manufacturer's instructions with the PL Buffer (Analytik Jena, Germany). Seed samples were harvested at 35 days after flowering (DAF). Three biological replicates per genotype (SGDH14 and Express 617) were used. The RNA quality was validated using NanoDrop and Agilent DNF-472 HS RNA (15 nt) Kit (Agilent Technologies, CA, USA) on the 5200 Agilent system to confirm the purity, concentration, and integrity, respectively. Based on 2 µg of total RNA, sequencing libraries were constructed with removal of rRNA through PolyA selection for mRNA species (eukaryotic).

Paired-end sequencing of 150 bp was performed on an Illumina NovaSeq 6000 following the 2×150 bp configuration at the Sequencing Facility of GENEWIZ Germany GmbH, Leipzig, Germany.

RNA-Seq read mapping and differential gene expression analysis

Read quality was analysed via FastQC v0.11.9 (Andrews 2010). After trimming the reads with Trimmomatic v0.39 (Bolger et al. 2014), the reads revealed an overall good quality with a phred score of minimum 36. Reads were then mapped to the Express 617 genome assembly (Lee et al. 2020) by STAR v2.7.1a (Dobin et al. 2013) using basic mode allowing maximal 5% mismatches per read length and a minimum of 90% matches per read length (see Table S5 of Supplementary Information 3 for mapping statistics). The read mappings were used to manually correct the functional annotation of Express 617 including e.g. the previously not annotated A05 *PAL4* homologue and subsequently used for downstream analyses. We used featureCounts v2.0.1 (Liao et al. 2014) to generate count tables applying the -O, -B, -p, and -fraction parameters based on the improved Express 617 annotation (see above). Raw count tables were subjected to differentially expressed gene analysis via DESeq2 (Love et al. 2014) to identify differentially expressed genes (DEGs) between SGDH14 and Express 617 seeds (35 DAF). DEGs with a $\log_2$ fold change > 1 and an adjusted p-value < 0.05 were selected for downstream analysis.

Organ-specific gene expression analysis using newly generated and publicly available *B. napus* RNA-Seq data

For the analysis of organ-specific gene expression across various organs newly generated (see above) and publicly available *B. napus* RNA-Seq data sets were used as described before (Schilbert et al. 2021). Transcript abundance and corresponding count tables were analysed and generated via Kallisto v0.44.0 (Bray et al. 2016). Heatmaps were constructed as described previously (Schilbert et al. 2021).

Identification of *PAL* gene family in *B. napus* Express 617

The *PAL* homologues in *B. napus* Express 617 were identified via KIPEs as described previously (Pucker et al. 2020). To reduce the number of fragmented peptides

derived from possible mis-annotations, candidates with at least 40% sequence similarity to the query sequences (also known as bait sequences) were used for further analysis. The functional described PAL sequence collection described in KIPEs was used as bait peptide sequences. The proteome file of *B. napus* Express 617 was used as subject. The alignment was constructed with MAFFT v7 (Katoh and Standley 2013) and trimmed to minimal alignment column occupancy of 10%. Next, FastTree v2.1.10 (Price et al. 2009) was used to build a phylogenetic tree using 10,000 rounds of bootstrapping. Visualisation of the phylogenetic tree was done with FigTree v1.4.3 (Rambaut 2006) (Figure S4, Supplementary Information 4). The *B. napus PAL* sequences were classified based on the corresponding *A. thaliana* homologues and their amino acid and coding sequences can be found in Table S6 of Supplementary Information 4.

Development of a *PAL* specific marker

For the development of a marker, which can be applied in plant breeding, the subgenome-specific SNPs of the A05 and C05 *PAL4* homoeologues were harnessed. By using oligonucleotides (Table S1, Supplementary Information 1), which flank the subgenome-specific SNPs in the second exon, a subgenome-specific marker for the presence of the A05 and/or C05 *PAL4* homoeologue was developed.

Results

Phenotypic analysis – the F2 population derived from a cross of the low lignin line SGDH14 and high lignin line Express 617 reveals variation in fibre traits

SGDH14 had lower seed lignin content and higher contents of oil and the sum of oil and protein compared to Express 617 (Table 1). In the SGDH14 x Express 617 F2 population, namely SGEF2, the lignin content of seeds ranged from 3.6 to 13.8%, hemicellulose from 1.6 to 7.9% and cellulose from 12.5 to 17.7% in the defatted meal, respectively (Table 1). Particularly, the lignin content showed a bimodal, skewed distribution. The low and high lignin genotypes showed a 1:2 segregation (Fig. 1A). Moreover, F2 individuals with lower lignin contents than the low parental line SGDH14 were identified (Fig. 1A). For heterozygous and high lignin group this is more difficult to assess. Lignin exhibits a strong positive correlation to seed colour and seed coat in addition to negative correlation to hemicellulose (Fig. 1B). The oil and protein content ranged from 38.7 to 47.1% and from 17.1 to 22.3%, respectively. The protein content in the defatted meal ranged from 31.4 to 37.4% and the sum of oil and protein content varied between 60.6 and 66.1%. The thousand-kernel weight (TKW) ranges from 3.8 to 6.8 g and is significantly higher in the low lignin group (Table 1). All those traits showed a weak negative

Table 1 Phenotypic trait measurements in the SGEF2 population and the respective low and high lignin groups. Minimum, maximum, mean values and standard deviation (SD) for contents of seed oil (%), protein (%), protein in defatted meal (PidM in %), the sum of oil and protein (%), erucic acid (%), glucosinolate (μMol g^{-1} at 91% seed dry matter), the fibre fractions (% in defatted meal), thousand-kernel weight (TKW) (g), seed colour and seed coat content (%) of the SGEF2 population ($n = 195$). The lignin groups are divided at a lignin content of $< 7.5\%$ into low lignin group ($n = 67$) and high lignin group ($n = 128$). The mean values are shown. Significance levels were analysed with single factor ANOVA

Traits	SGEF2 population				Lignin groups			Parental genotypes	
	Min	Max	Mean	SD	Low		High	SGDH14	Express 617
Oil	38.7	47.1	43.3	1.2	43.8	**	43.0	46.2	42.7
Protein	17.1	22.3	19.6	0.8	19.6	ns	19.6	18.2	18.5
PidM	31.4	37.4	34.5	1.0	34.9	**	34.3	33.9	32.3
Oil + Protein	60.6	66.1	62.9	0.9	63.4	**	62.6	64.4	61.2
Erucic acid	0.0	27.4	13.8	6.0	13.4	ns	14.0	15.4	0.0
Glucosinolate	13.9	61.5	41.8	9.1	42.4	ns	41.4	44.1	13.7
Lignin	3.6	13.8	8.8	2.2	6.0	**	10.3	7.2	10.7
Hemicellulose	1.6	7.9	6.1	0.8	6.7	**	5.7	6.6	5.7
Cellulose	12.5	17.7	15.5	0.8	15.4	ns	15.5	15.4	16.0
TKW	3.8	6.8	5.2	0.5	5.3	**	5.1	4.5	4.7
Colour	6.0	9.0	8.3	0.9	7.5	**	8.8	7.8	9.0
Seed coat	11.2	25.3	15.2	2.2	14.1	**	16.1	12.1	16.3[§]

ns, *, ** shows not significant and significance with $p < 0.05$ und $p < 0.01$

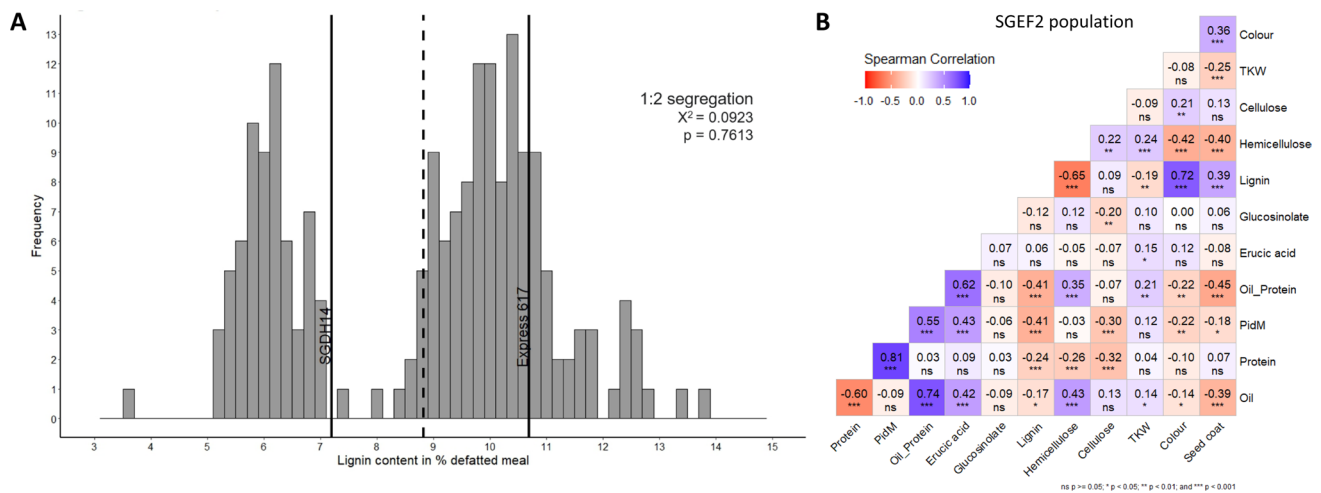[§]Data taken from Behnke et al. (2018)

Fig. 1 Phenotypic characterisation of the SGEF2 population. **A** Frequency distribution of lignin content in defatted meal of the SGEF2 population (F2:3 seeds). The lignin content of the parental genotypes SGDH14 and Express 617 is indicated by black vertical lines and the grey vertical dashed line marks the mean lignin content of the SGEF2 population ($n = 195$). The Chi-square result for a 1:2 segregation is displayed. The p-value is not significant so that H0, the 1:2 segregation, is accepted. **B** Correlation Matrix of the SGEF2 population (F2:3 seeds). Positive correlation is indicated in blue and negative correlation is indicated in red. The correlation coefficients are given for each trait. PidM is protein content in the defatted meal at 91% dry matter (color figure online)

correlation to lignin (Fig. 1B). Moreover, a positive correlation between lignin and darker seed colour as well as seed coat content was found (Fig. 1B). The pigmentation of SGEF2 seed ranged from dark brown (7.0) to black (9.0) (Table 1). Transgressive segregation was observed for most of the traits in the F2 population.

Comparing the low and high lignin groups of the SGEF2 population, the low lignin group showed significantly higher content of oil in the defatted meal and hemicellulose, as well as for the sum of oil and protein compared to the high lignin group. A strong positive correlation between seed lignin and oil was found in the low lignin group whereas protein showed a negative correlation (Fig. 2B). The high lignin group showed no correlation between oil and lignin (Fig. 2C), which in the overall population leads to a weak negative correlation between the two traits. A strong negative correlation between lignin and hemicellulose (Fig. 2C) was found in the high lignin group. In contrast a significant positive correlation was observed in the low lignin group (Fig. 2B). The whole population exhibits a strong negative correlation between lignin and hemicellulose (Fig. 1B). A weak positive correlation to cellulose could be identified in the low lignin group (Fig. 2B). A negative correlation between lignin and protein is found in the whole population as well as in both lignin groups. The high lignin group exhibited a significantly higher seed coat content (16.1%) than the low lignin group (14.1%; Table 1 and Fig. 2A). To summarise the SGEF2 population was extensively phenotypically characterised and showed a large variation in seed lignin content, which was also identified in the parents with

SGDH14 possessing a low lignin content and the Express 617 showing a high lignin content.

Detailed characterisation of the physical position of the major low lignin QTL

To analyse the genetic basis underlying the observed phenotypic variation of the parents, we used the genome assembly of the high lignin parent Express 617 (Lee et al. 2020) and generated genomic long-reads of the low lignin parent SGDH14. As a major QTL for low lignin content on C05 was identified before, based on the genome assembly of Darmor-*bzh* (Behnke et al. 2018), we first aimed to identify the syntenic genomic region in the Express 617 genome sequence. The physical position of the major low lignin QTL in the genome assembly of Express 617 ranged from 41,286,363 bp to 42,386,439 bp (~ 1,100 kbp) on C05, while in Darmor-*bzh* the QTL ranged from 39,713,411 bp to 40,616,480 bp (~ 903 kbp) (Fig. 3). Thus, the QTL region is ~ 197 kbp bigger in the Express 617 genome assembly. This finding can be explained by the higher continuity of the Express 617 long-read assembly as more repeats located in the QTL region are resolved whilst they are collapsed in the short-read Darmor-*bzh* assembly. To identify genomic differences located near or inside the major QTL between the parental genotypes, the genomic long-reads of SGDH14 were mapped against the Express 617 genome assembly. A ~ 208 kbp region ranging from 41,563 to 41,771 kbp on C05 located near the center of the major QTL revealed no read coverage in SGDH14 (Fig. 3; Figure S2 of Supplementary Information 2). Corresponding flanking reads of
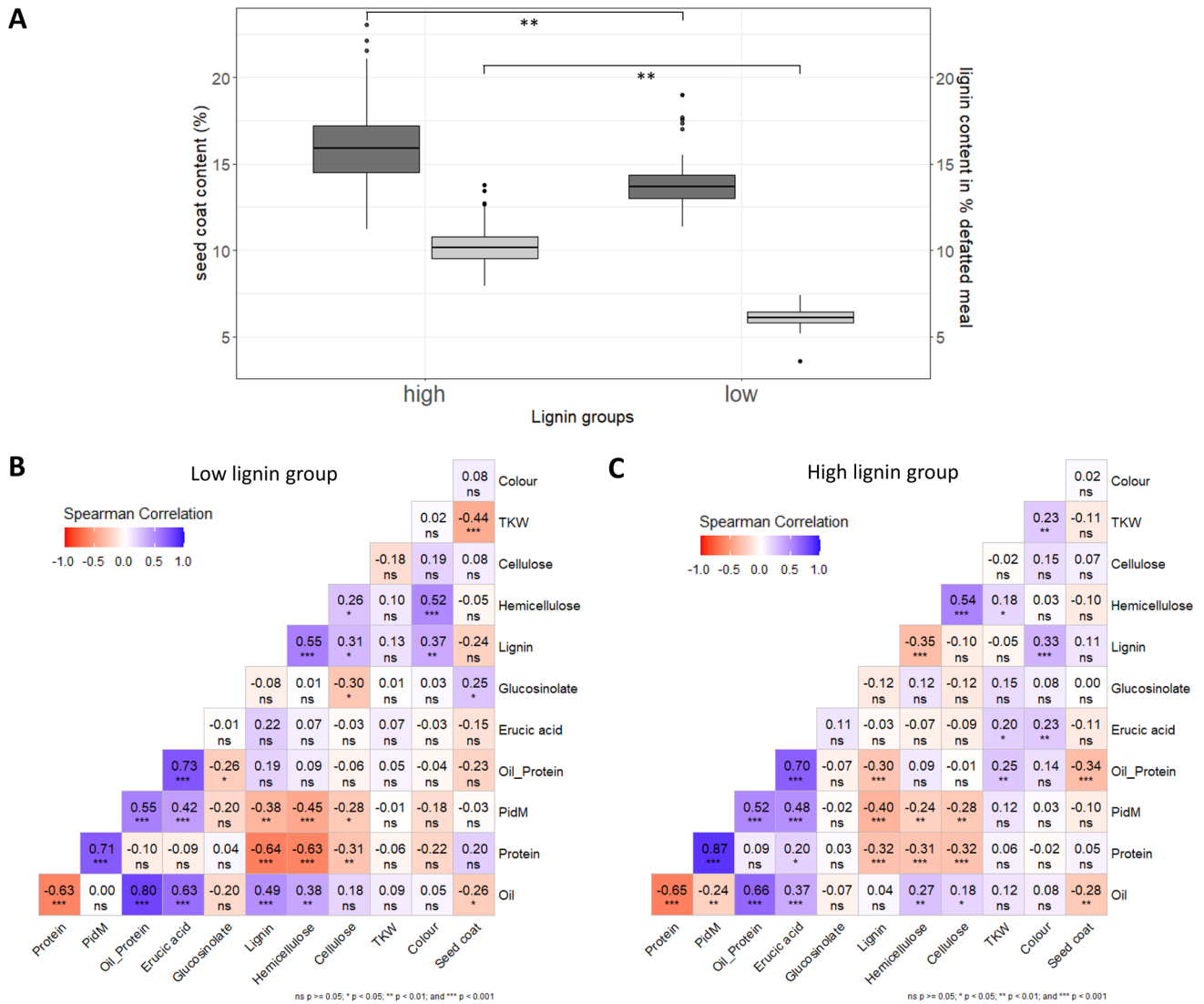
Fig. 2 Phenotypic characterisation of the high and low lignin groups of the SGEF2 population. **A** Seed coat and lignin content (%) in high and low lignin groups of the SGEF2 population. The high lignin group ($n = 128$) is shown on the left and the low lignin group ($n = 67$) is displayed on the right side of the figure. Dark grey coloured box-plots represent the seed coat content and the light grey boxplots show the lignin content. The significance level was calculated with a single-factor ANOVA. ** Shows significance $p < 0.01$. **B** Correlation Matrix of the low lignin group of the SGEF2 population (F2:3 seeds). Positive correlation is indicated in red and negative correlation is indicated in blue. The correlation coefficients are given for each trait. PidM is protein content in the defatted meal at 91% dry matter. **C** Correlation Matrix of the high lignin group of the SGEF2 population (F2:3 seeds). Positive correlation is indicated in blue and negative correlation is indicated in red. The correlation coefficients are given for each trait. PidM is protein content in the defatted meal at 91% dry matter (color figure online)

SGDH14 were identified to be chimeric, having a C05 and an A05 sequence part. By separating the chimeric reads in subgenomic k-mers (Figure S3, Supplementary Information 2), the genomic border sequences were narrowed down and validated via Sanger sequencing (Fig. 3). Ultimately, these results showed that the homoeologous A05 sequence ranging from 27,121 kbp to 27,289 kbp was inserted between 41,563 kbp and 41,771 kbp on C05 in SGDH14 (Fig. 3). Further, this homoeologous A05 locus was identified to be duplicated, as the average coverage of the A05 homoeologous locus was ~twofold higher, compared to the average coverage of the whole A05 chromosome (Figure S2, Supplementary Information 2). Therefore, the name A05'' is used for sequence features located within this duplicated

homoeologous A05'' region (Fig. 3). A homoeologous non-reciprocal translocation (HNRT) is identified in SGDH14.

Importantly, in the Express 617 genome assembly on chromosome A05 at position 27,213,244 to 27,213,343 a stretch of ambiguous bases was identified. This is indicative of an assembly break that could not be resolved in the Express 617 genome assembly. In accordance, the ONT long-reads of SGDH14 revealed a highly repetitive region located at this position inside the HNRT region, which could not be spanned. Although border spanning chimeric reads reaching into this repeat could be identified, their length was not sufficient to span this large repeat (Figure S3, Supplementary Information 2). As the region sizes were estimated based on the Express 617 genome assembly, irrespective of this repeat, they are likely underestimating the true sizes. The nGBS data support all above described findings. In summary, in SGDH14 a ~ 208 kbp region on C05 of the major QTL is deleted and is substituted by a ~ 168 kbp duplication of the homoeologous A05'' sequence.

HNRT event in the SGDH14 x express 617 DH population explains seed lignin content

To analyse the association between the HNRT event and the lignin phenotype we screened a second independent population for the presence or absence of the HNRT event. We used the SGDH14 x Express 617 DH population of Behnke et al. (2018). SGDH14 shows the low lignin phenotype and harbours a HNRT event with a duplication of an A05 chromosome part and replacing the homoeologous C05 region. Analysis of 60 K Illumina SNP data revealed that several markers in the QTL region were scored as "failed" not only in SGDH14, but also in its ancestral genotypes Sollux and Gaoyou, as well as in SGEDH13 and Zheyou 50, but not in Express 617 and Adriana (Table S8, Table S9, Supplementary Information 5). Carefully studying the frequency distribution of "failed" scored markers in the C05 region in the original SGDH14 x Express 617 DH genotyping table, allowed a clear separation of DH lines with and without "failed"-markers, i.e. with and without HNRT. Comparison of the DH lines with and without the HNRT event showed a clear, significant difference in several quality traits (Table 2). The lignin content was significantly reduced in lines with the HNRT event, whereas hemicellulose and cellulose showed increased contents. Oil and protein content in defatted meal and the sum of oil and protein content were increased in the low lignin lines. The seed protein content was not different between the groups. These results were concordant with the results of the high and low lignin groups of the SGEF2 population (Table 1).

Candidate genes associated with phenylpropanoid and lignin biosynthesis

In order to identify candidate genes that might contribute to the observed lignin phenotype, e.g. involved in general phenylpropanoid and lignin biosynthesis, we screened the genomic regions affected by the HNRT event. Thus, the deleted region of the C05 chromosome and the substituted and thus duplicated A05'' region were inspected for candidate genes. Based on the Express 617 genome assembly, 207 genes are located inside the major low lignin QTL region on C05 (41,286 to 42,386 kbp) of which 32 genes are deleted by the HNRT event (41,563 to 41,771 kbp) (Table S10, Supplementary Information 6). Additionally, *C05p041220.1* (homologue of *AthSEC8*) and *C05p041550.1* (homologue of *AthPIP5K9*) are located at the genomic borders of the HNRT event, leading to the loss or disruption of 26 of its 27 exons in the case of *C05p041220.1* and part of the first exon in the case of *C05p041550.1*. The homoeologous A05 region (27,121 to 27,289 kbp) duplicated in SGDH14 contains 31 genes based on the Express 617 genome assembly and shows a large proportion of overlapping genes to its C05 homoeologous region (Table S11, Supplementary Information 6).

We analysed the organ-specific expression of candidate genes, which *A. thaliana* homologues are associated with lignin biosynthesis (Table S12, Supplementary Information 7). Most candidate genes were very low or not at all expressed in Express 617 and SGDH14 seeds (35 DAF) except for the C05 *B. napus* homologue *AthPAL4 C05p041260.1*, which showed the highest expression in Express 617 but not in the HNRT deleted region of SGDH14. Moreover, the *B. napus* homologues of *AthCESA3* (*C05p040780.1*), *AthGALAK* (*C05p040910.1*), *AthMUCI21* (*C05p041270.1*), and *AthDEX1* (*C05p042190.1*) were the highest expressed candidate genes during seed coat development. In line with the HNRT event, *AthPAL4* (*C05p041260.1*) and three other genes located in the HNRT region were identified as lower expressed in SGDH14 compared to Express 617 seeds (Table S13, Supplementary Information 7), namely the *B. napus* homologues of *AthSEC8* (*C05p041220.1*), *AthSDP6* (*C05p041230.1*) and *AthMUCI21* (*C05p041270.1*).

SGDH14 harbours two A05 BnaPAL4 homologues

Due to the allotetraploid origin of *B. napus* homoeologous gene copies are frequent and could compensate for the loss of one homoeologue gene copy. Synteny analysis and manual inspection of RNA-Seq read coverage revealed
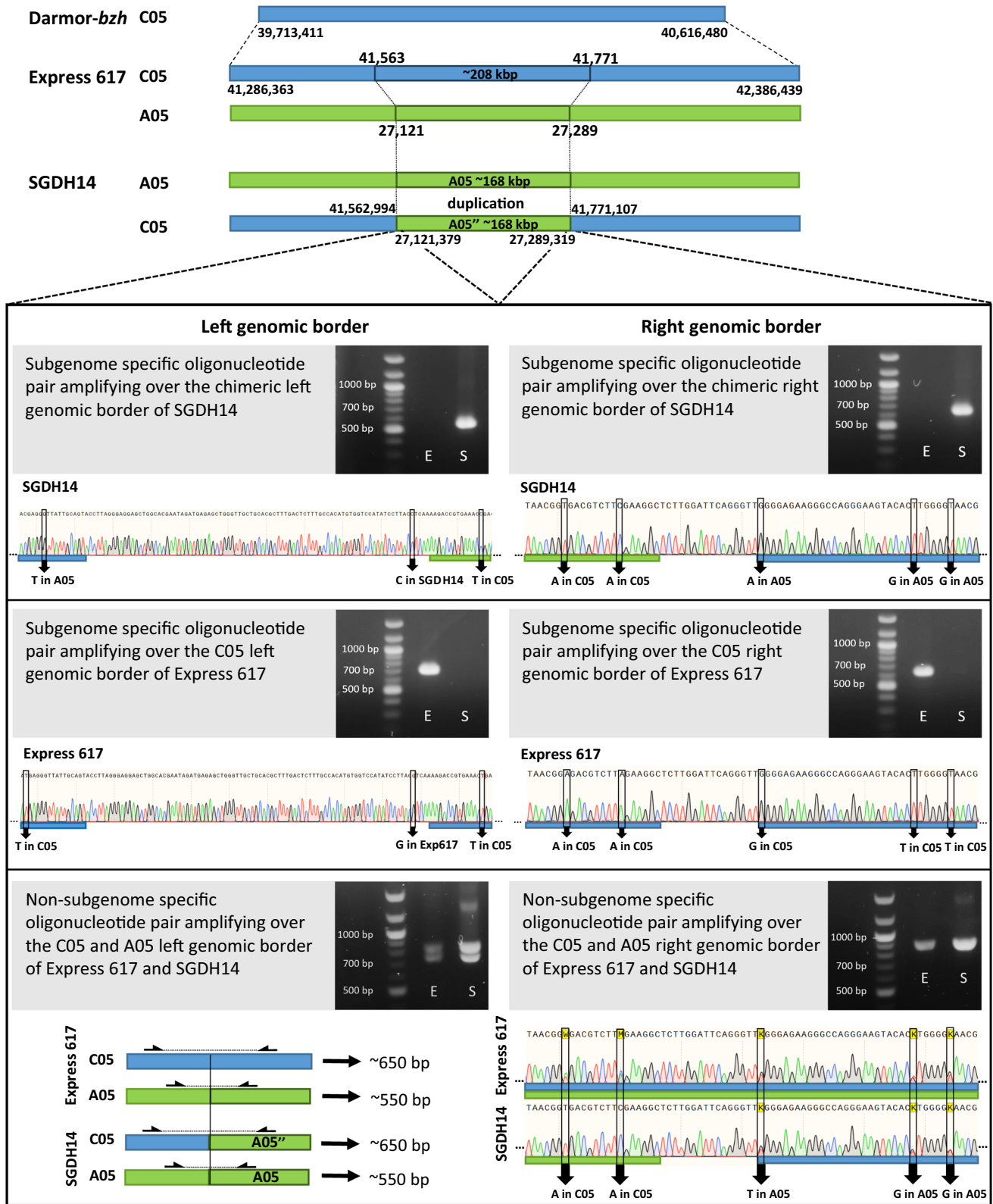
an A05 *PAL* homoeologue in Express 617, which was not included in the original Express 617 annotation and named *A05p031299.1* (Figure S5, Supplementary Information 4).

The respective gene was identified as *PAL4* homologue by in silico gene prediction followed by a BLASTp search. The organ-specific expression analysis revealed that both

◀**Fig. 3** SGDH14 reveals a homoeologous non-reciprocal translocation in the major low lignin QTL region on C05. (Top) The location of the major low lignin QTL on C05 (blue) in the Darmor-*bzh*- and the Express 617 genome assembly is shown and the A05 (green) to C05 homoeologous non-reciprocal translocation region in SGDH14 is marked with a dark green outlined rectangle and dotted lines. (Bottom) Subgenome-specific and non subgenome-specific oligonucleotides were used for the validation of the substitution borders. The amplified products using Express 617 (E) and SGDH14 (S) genomic DNA, as well as sections of corresponding representative Sanger sequencing results are shown. Figure is not to scale (color figure online)

BnaPAL4 homoeologues are much higher expressed in developing seeds than the other identified paralogues of Express 617 (Table S7, Supplementary Information 4).

In SGDH14, two A05 *PAL4* copies (A05 and A05'' *PAL4*) were identified by inspecting SNPs of the genomic and transcriptomic SGDH14 read mappings against the Express 617 genome assembly. In total, 16 SNPs (12 in exon No. 3, 3 in 3'UTR, 1 upstream of 5'UTR) revealed a ~ 50:50 ratio of reference vs alternative alleles (Fig. 4, Figure S6 of Supplementary Information 4). Twelve of these SNPs are located in the third exon, of which six SNPs are identical to the C05 *PAL* allele, while the remaining six were specific for the new A05'' *PAL4* allele as they differ from the A05 and C05 *PAL4* allele (Figure S6, Supplementary Information 4). Some of these SNPs result in amino acid exchanges compared to the A05 PAL4 amino acid sequence, of which three are specific to the A05'' PAL4 (V586I, D596E, V638I), while the remaining two (E610G, V629G) are identical to the C05 PAL4 sequence. Analysis of the putative promoter sequence using 1 kbp upstream of the translational start side showed only one SNP differentiating the A05'' *PAL* from the A05 *PAL* allele, while several subgenome-specific SNPs were identified (Figure S6, Supplementary Information 4). Finally, a marker for the presence or absence of the C05 *PAL4* homoeologue was designed by using non subgenome-specific oligonucleotides. Sequencing the PCR amplicons derived from the second exon showed the absence of C05 specific SNPs in SGDH14, while Express 617 revealed heterozygous SNPs for each homoeologue (Fig. 4). In accordance with the genomic results stressing the loss of the C05 *PAL4* homologue in SGDH14, the C05 *PAL4* was not expressed in SGDH14 but in Express 617. The two A05 *PAL4* homologues of SGDH14 revealed a ~ 1.5 higher expression compared to the combined expression of the C05 and A05 *PAL4* homoeologues of Express 617 (Table S7, Supplementary Information 4).

Discussion

HNRT in *B. napus* and phenotypic changes

Widespread segmental duplications, deletions and homologous chromosome exchanges are already identified in a wide range of natural and synthetic genotypes of *B. napus*. Due to the presence of two additive subgenomes with high degree of homoeology, the meiotic chromosome pairing between homoeologous chromosomes causes the occurrence of homoeologous exchange (HE) and gene conversions. These homoeologous exchanges are the main drivers of diversification in polyploid genomes (Szadkowski et al. 2010; Chalhoub et al. 2014; Gaeta and Pires 2010). Changing the genome may subsequently cause modifications in the phenotypic appearance. The impact of homologous non-reciprocal translocation (HNRT) and HEs on the phenotype in *B. napus* were shown earlier. Seed yield, seed quality, seed colour, seed fibre content, flowering time, disease resistance and fertility were affected by homoeologous exchanges (Hurgobin et al. 2018; Schiessl et al. 2019; Stein et al. 2017). Looking at fibre content, Stein et al. (2013) concluded in their *B. napus* QTL mapping population that a HE causing a deletion or conversion is responsible for a lower seed fibre content and a yellow-coloured seed. The HE and subsequently important candidate genes were identified on chromosome A09. With higher marker density and sequence coverage, the borders and genetic position of the HNRT were confirmed (Stein et al. 2017). In the present study, the major QTL for lignin content on chromosome C05 (Behnke et al. 2018) was analysed in depth. With ONT- and RNA-Seq data, it was shown that HNRT of a A05 region replaces a homologous C05 region in SGDH14. This translocation was accompanied by a segmental C05 chromosome deletion and is responsible for the observed reduction in lignin content. The replacement of the C05 *BnaPAL4* by a A05 *BnaPAL4* homologue was identified as the likely cause for the observed changes in lignin content.

A chromosome segment replaces C chromosome segment

As shown in previous studies, larger C-subgenome segments are replaced by smaller, homoeologous A-subgenome segments, driving postpolyploidisation genome size reduction (Samans et al. 2017; Sochorová et al. 2017; Higgins et al. 2018). Further, most of those chromosomal rearrangements occurred between regions of primary homology (Nicolas et al. 2007). In the present study, the allotetrapolyploid *B. napus* shows similar chromosomal rearrangements with a subgenome bias. In SGDH14 the 208 kbp large C05 chromosome segment was replaced by a smaller at least 168 kbp

Table 2 Phenotypic trait measurements in the SGDH14 x Express 617 DH population and the respective lines with and without HNRT. The groups with and without HNRT were formed by dividing the DH lines according to their SNP marker data in the C05 QTL region. Genotypes with HNRT showed several markers as "failed" in this region, equal to missing marker pattern of SGDH14. Oil, protein and crude fibre contents of the genotypes with the HNRT ($n = 71$) and without HNRT ($n = 67$) in the SGDH14 x Express 617 DH population of Behnke et al. (2018) are listed. The significance level was calculated with a single factor ANOVA

Traits	Range with HNRT	Range without HNRT	Genotypes with HNRT		Genotypes without HNRT
Oil	43.0–50.8	42.5–48.6	47.0	**	45.6
Protein	16.6–19.5	16.6–19.5	17.9	ns	17.9
PidM	30.3–36.9	30.3–36.3	33.8	**	32.9
Oil + Protein	60.8–68.3	60.9–67.0	64.9	**	63.5
Erucic acid	0.0–51.0	0.0–49.5	22.9	ns	20.2
Glucosinolate	18.2–74.6	18.7–69.0	53.1	ns	49.4
Lignin	6.1–9.8	8.4–14.0	7.3	**	11.4
Hemicellulose	11.5–15.1	11.0–13.3	14.0	**	12.4
Cellulose	10.7–12.5	9.9–11.6	11.4	**	10.8
TKW	4.4–6.0	4.1–5.8	5.1	ns	5.0

Oil, Protein, Oil + Protein in % 9% moisture, PidM = Protein in % in defatted meal; Hemicellulose, Cellulose und Lignin (% in defatted meal); **Shows significance with $p < 0.01$
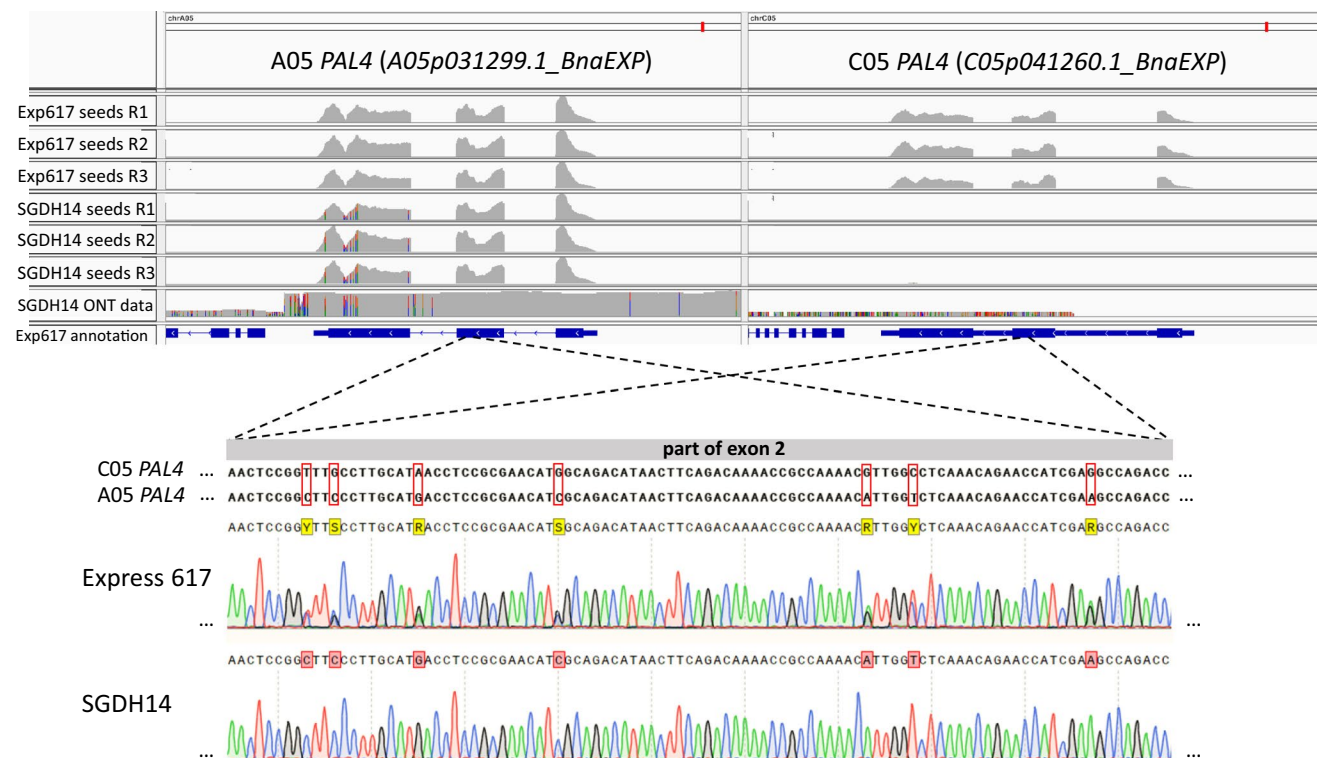
Fig. 4 SGDH14 harbours a second A05 *PAL4* homologue, namely A05'' *PAL4*, expressed in seeds, while the C05 *PAL4* is deleted. (Top) RNA-Seq data from seeds (35 DAF) of Express 617 and SGDH14 as well as ONT read mapping of SGDH14 against the Express 617 genome assembly. The three biological replicates of the RNA-Seq data are marked with R1, R2, and R3, respectively. Positions of SNPs differentiating from the Express 617 genome assembly are coloured as follows: adenine green, thymine red, cytosine blue, and guanine black. Twelve of sixteen identified SNPs are located in the third exon, of which six were specific for the new A05'' *PAL4* allele as they differ from the A05 and C05 *PAL4* allele. (Bottom) Sequencing results of a part of the *PAL4* second exon using oligonucleotides, which can bind to both subgenomes, with Express 617 and SGDH14 gDNA. The C05 *PAL4* second exon sequence was used as reference and subgenome-specific SNPs are marked with a red rectangle. A representative sequenced section is shown (color figure online)

segment of the homoeologous A05 chromosome. The exact size of this segment could not be determined since a large repeat in this sequence region was not spanned by the long-read sequencing data. In the Express 617 genome assembly this repeat is marked as a stretch of ambiguous bases (Lee et al. 2020). Although the Express 617 genome assembly is based on 54.5× coverage with Pacific Biosciences long-reads, linked reads, optical map-, and high-density genetic map data (Lee et al. 2020) this repeat cluster could not be resolved. With the long-read data of SGDH14 the repeat could also not be resolved due to the lack of spanning reads. To determine the exact size of the HNRT, longer reads which span the repeat cluster are necessary.

Method for QTL and HE detection

The major QTL C05 was genotyped and mapped previously with the Illumina Infinium Brassica 60 K SNP array with a total number of 58,464 markers. The relevant region on C05 with the deleted fragment marked is shown in Table S8 along with the most likely and second most likely position of the SNP markers of Express 617 in Table S9 of Supplementary Information 5. Within the deleted C05 region only two SNP markers were polymorphic (Behnke et al. 2018). Other SNP markers within the C05 confidence interval were scored in SGDH14 as "failed" markers, indicating the HNRT. The DH population of Behnke et al. (2018) did not show a significant segregation distortion for the presence ($n=71$) and absence ($n=67$) of the HNRT. The DH lines with the HNRT had a significantly lower seed lignin content than the ones without HNRT (Table 2).

When did the HNRT event occur in SGDH14? To answer this question, the Illumina 60 K SNP marker data of the parental genotypes Sollux and Gaoyou of SGDH14 were analysed. Based on the SNP marker data it became obvious that the same SNP markers gave "failed" results for the HNRT region on C05 in both parents. Furthermore, SGEDH13, an offspring of SGDH14 x Express 617, showed the same lignin QTL on C05 with 61.3% phenotypic variance, a bimodal distribution for seed lignin content and the same "failed" results for the HNRT region on C05 (Yusuf et al. 2022). These are strong indications, that the offspring SGEDH13 also carries the described HNRT event and can be used for commercial breeding. The same was found for the Chinese cultivar Zheyou 50. Yusuf et al. (2022) discussed that Zheyou 50 may be an offspring from the SGDH14 x Express 617 cross, since it originates from the Zhejiang Academy of Agricultural Sciences as SGDH14, and carries the same major QTL on C05 and is of canola quality. Although the HNRT event in Sollux, Gaoyou and Zheyou 50 was not confirmed by sequencing, it appears likely that the HNRT event occurred much earlier and

indicates that a positive selection may have led to its enrichment. The nGBS data indicated for Sollux and Gaoyou the HNRT event, but the exact borders could not be determined. The HNRT in Sollux is potentially even bigger (~610 kbp) than in SGDH14. Yield testing of selected high and low lignin bulks of the crosses SGDH14 x Express 617 and of Adriana x SGEDH13 recently indicated an advantage of the low lignin over the high lignin bulks, since a higher yield could be detected (Holzenkamp et al. 2022).

As a consequence of genomic rearrangements, skewed marker segregation patterns often prevent standard mapping procedures accurately localising QTL and causal genes in these regions (Stein et al. 2017). It was proposed to use SNP marker data from loci spanning deletions or duplications, which would be discarded in standard mapping procedures (Stein et al. 2017; Gabur et al. 2018). With this data structural chromosome variation can be identified in QTL analysis and single nucleotide absence polymorphisms (SNaPs) can provide an added value. SNaPs should be used when one parent and progenies of the mapping population show missing markers. The search for presence–absence variation should be implemented in a routine screening to find diversity in breeding material (Stein et al. 2017; Gabur et al. 2018). Surprisingly, the HNRT event was accurately localised by the two SNP markers Bn-scaff_17441_1-p950045 and Bn-scaff_17441_3-p28950. SNP Bn-scaff_17441_1-p950045 was mapped next to Bn-scaff_17441_3-p28950, which are located within the confidence interval of the major lignin QTL on C05 identified by Behnke et al. (2018). Because Bn-scaff_17441_1-p950045 is located within the deleted C05 region based on its physical position in the Express 617 assembly, it should have given a "failed" result for about half of the DH population as the other markers in the HNRT region. However, this was not the case. This marker segregated 1 to 1 in the DH population (Table S8 in Supplementary Information 5). Obviously, Bn-scaff_17441_1-p950045 detected in homozygous DH lines with the C05 deletion rather the homoeologous SNP on the SGDH14 A05 chromosome, resulting in a 1 to 1 segregation which fits to the presence/absence of the HNRT in the DH population. The same may be valid for the second SNP marker Bn-scaff_17441_1-p958418. The physical position of the second QTL flanking marker Bn-scaff_17441_3-p28950 is located downstream of the C05 deleted region.

Lignin and seed coat content

In the present study, the F2:3 seeds of the SGEF2 population showed a similar bimodal distribution of lignin content as in the DH population SGDH14 x Express 617 of Behnke et al. (2018). Although the SGEF2 population was only tested in one field environment, the results were comparable to those of Behnke et al. (2018). The chi-squared

test revealed a 1:2 segregation, which deviates from the expected 1:3 segregation in F2 populations for a recessive mutation. This indicates a segregation distortion towards homozygous genotypes with HNRT. This segregation distortion towards presence of HNRT is tentatively also observed in the DH population with 71:67 segregation for presence absence of HNRT. This may be explained by chance and by lack of effects during the in vitro culture effects of isolated microspore culture. It was not possible to clearly identify and separate homozygous and heterozygous F2-plants for the HNRT translocation, because the designed PCR Primers were not specific enough. The genetic causes for the segregation distortion in the SGEF2 population remains unclear. Later on, designed anchored KASP INDEL markers were able to detect the A05/C05 insertion using subgenome specific polymorphisms on the right border (Fig. 3, Figure S7 and S8 of Supplementary information 8). This marker could, however, not anymore be applied to the DNA samples of the F2 plant population.

In general, yellow seeded cultivars show a reduced lignin and seed coat content compared to black seeded cultivars (Wang et al. 2015; Ding et al. 2021; Rahman and McVetty 2011). Black seeded cultivars have higher lignin contents than yellow-seeded cultivars (Wittkop et al. 2009). The positive correlation between seed lignin and seed colour which was found for the SGEF2 population, was also reported by Wittkop et al. (2009) for a genetic diverse oilseed rape panel. However, the seed material in this study revealed a rather low variation in seed colour, despite showing a wide variation in lignin content. The pigmentation of F2:3 seeds ranged only from dark brown to black, with a tendency to a lighter colour for the low seed lignin genotypes. Hence, the variation in lignin content should be correlated with the seed coat proportion. It was already previously reported, that the seed coat in high lignin lines was thicker than that of low lignin lines (Behnke et al. 2018; Ding et al. 2021; Wang et al. 2015). The seed coat content of SGDH14 (12.1%) is significantly reduced by a quarter compared to Express 617 (16.3%) (Behnke et al. 2018). Similar results were determined in this study with the SGEF2 population. The high and the low lignin groups of F2:3 seeds showed significant differences in seed coat content (Fig. 1C).

The seed coat consists mainly of three layers from inside to outside: the mucous epidermal cells, the palisade and the endothelial layers (Moïse et al. 2005). Further, Ding et al. (2021) suggested that lignin is deposited in the palisade cells, thereby increasing the thickness of the palisade layer and the overall thickness of the cell wall. This can explain the seed coat thickness, but not the seed coat colour, since in our material the seed colour is similar despite varying seed lignin contents. The material of Yin et al. (2022) showed similar results with no correlation between seed coat thickness and seed colour. Generally, the seed coat should be thin

to reach high oil contents. Those genotypes with a thinner seed coat are mostly yellow seeded cultivars. In the present study, this was not the case since SGDH14 showed a dark seed colour and a high oil content. Hu et al. (2013) showed similar results with black seeded oilseed rape lines which were reported to have high oil contents. It was shown by Iwanowska et al. (1994), that the innermost layer of the inner integument, known as the endothelium, accumulates phenolic compounds starting from the early stages of embryo development. Lignins and proanthocyanidins (PAs) are phenolic compounds, which biosynthetic pathways share common enzymes as they are both derived from Phe through the general phenylpropanoid pathway. PAs are derived from the flavonoid biosynthesis and are the major pigments responsible for the seed colour in *A. thaliana* and *Brassica* species pathway (Xu et al. 2014). The biosynthetic pathways of lignin and PA diverge at the level of *p*-coumaroyl-CoA, which is used as substrate for the initial steps of the flavonoid and lignin biosynthesis. Seed pigments in mature dry seed of *Brassica* species are primarily oxidised PAs (Durkee 1971), which are responsible for the dark seed colour developed during seed maturation (Moïse et al. 2005). SGDH14 exhibits low lignin contents, but similar seed coat pigmentation as the high lignin line Express 617. The absence of C05 *BnaPAL4* expression and thus likely activity in SGDH14 could be an indication that the C05 *BnaPAL4* is specific for the lignin biosynthesis, while the A05 *BnaPAL4* could be involved in PAs synthesis.

In the present study, the high and low lignin groups (Table 1) showed clearly differentiating lignin contents. We identified F2 individuals in the SGEF2 population with lower lignin contents than in the low lignin parental line SGDH14. This finding suggests that epistatic effects caused by additional minor loci of Express 617 leads to comparatively lower lignin contents in F2 plants in comparison to SGDH14. Additional minor QTL may have been non-significant and hence remained undetected in the publication of Behnke et al. (2018). The overall lignin content in those genotypes was still higher than in the study of Ding et al. (2021) (with min. 0.94%; max. 4.7% lignin). The lignin content and seed coat colour of our low lignin group with the high lignin group of Ding et al. (2021), were similar. This leads to the assumption that yellow seed coat colour appears, when the lignin content drops below 2%. Moreover, a negative correlation between seed lignin and hemicellulose content was detected. The low lignin lines had a significantly higher hemicellulose content compared to high lignin lines (Table 1). In previous studies it was reported that the synthesis of cellulose and hemicellulose redirects photosynthetic assimilates from oil and protein into sugar biosynthesis, which may result in reduced contents of these two compounds (Wang et al. 2015; Liu et al. 2013). The hypothesis is, that assimilates

which are not used in the general phenylpropanoid pathway to produce lignin are redirected into hemicellulose production. Similar redirection to hemicellulose synthesis was reported in rice (Ambavaram et al. 2011) and in aspen trees (Hu et al. 1999). Obviously, there is a crosstalk between the lignin and hemicellulose biosynthetic pathway, which seems not to be the case for cellulose.

The thinner the seed coat of *B. napus* cultivars, the larger the proportion occupied by the embryo, which leads to increasing contents of oil and protein (Slominski et al. 1999). The seed lignin content is confirmed to be significantly negatively correlated with seed oil content (Behnke et al. 2018; Miao et al. 2019). Similar results could be obtained in this study, where we found that oil content did show a significant difference between the low and high lignin groups. The low lignin group had a higher oil content. This was expected when compared to the parental genotypes and to the results of Behnke et al. (2018). Further, protein in defatted meal (%) and the sum of oil and protein content (%) showed a significant difference between the lignin groups, where the low lignin group contains significantly more. Similar results were observed in Behnke et al. (2018). However, the protein content did not show a significant difference between the groups, as in Behnke et al. (2018). Thus, reducing seed lignin content is associated with a reduction in seed coat content and increased oil content, while simultaneously improving the nutritive value of oilseed rape meal through a reduced fibre content.

Candidate genes and ONT-technology

Genomic and transcriptomic data were used to unravel the genetic basis underlying the contrasting seed lignin content phenotypes between the parental genotypes SGDH14 and Express 617. In accordance with metabolic studies surveying the flavonoid – and general phenylpropanoid pathway in yellow and black seeded *B. napus*, 35 days after flowering (DAF) was chosen for the expression analysis, since those contents were significantly higher at 35 DAF in black seeds compared to yellow seeds (Wang et al. 2018). In this study, the low lignin and reduced fibre line SGDH14 lack the *PAL4, SEC8, MUCI21* and *SDP6* homologues located on C05 compared to the high lignin line Express 617 due to the HNRT event. *PAL4* is the most promising candidate gene for the lignin reduction, since its encoded protein catalyses the first rate limiting step in the general phenylpropanoid pathway leading to lignin synthesis (Fraser and Chapple 2011). Moreover, previous studies have proposed *PAL* as the key enzyme associated with lignin content in oilseed rape (Behnke et al. 2018; Ding et al. 2021; Wang et al. 2015). The expression of *PAL* was significantly higher in high lignin lines than in low lignin lines (Ding et al. 2021).

The other candidate genes were related to fibre traits. SEC8 is involved in the localised deposition of seed coat pectin and formation of the central columella of seed epidermal cells (Kulich et al. 2010). *SEC8* and *PAL4* were also related to seed coat development in the study of Zhang et al. (2022). SDP6 is required for glycerol catabolism and is essential for post germinative growth and seedling establishment (Quettier et al. 2008). MUCI21 is necessary for the adherence of mucilage pectin to microfibrils made by Cellulose Synthase 5 (*CESA5*) to form an adherent mucilage layer on the seed coat (Ralet et al. 2016).

In complex allopolyploids like *B. napus*, pairing of homologous chromosomes occurring during meiosis is common and can lead to fragment exchanges between homoeologous chromosomes, resulting in widespread segmental deletions and duplications (Chalhoub et al. 2014). To clearly distinguish highly similar gene sequences and genomic rearrangements from another, long-read sequencing technology is valuable (Lee et al. 2020). In this study, the detection of the HNRT in SGDH14 was possible by long-reads derived from SGDH14 DNA and read mapping against the high quality long-read genome assembly of the winter *B. napus* cultivar Express 617 (Lee et al. 2020). Behnke et al. (2018) used Illumina 60 K SNP markers and the Darmor-*bzh* genome assembly which had a higher level of fragmentation since it was constructed prior to the spread of long-read sequencing technologies. This did not allow the immediate detection of the HNRT in SGDH14. With the new long-read technology the sequence specificity of homoeologous chromosomes is enhanced, interference of repetitive sequences is reduced and HE breakpoints can be detected (Lee et al. 2020).

BnaPAL4 and paralogues

This study revealed *PAL4* as the most promising candidate gene, which is involved in the general phenylpropanoid pathway and therefore in the lignin synthesis. Since, *B. napus* and *A. thaliana* have multiple *PAL* paralogues (Fraser and Chapple 2011), their seed specificity needs to be determined. In *A. thaliana,* expression analysis showed that *AthPAL1*, *AthPAL2* and *AthPAL4* are expressed in stems of plants in later growth stages and *AthPAL2* and *AthPAL4* are expressed in seeds (Fraser and Chapple 2011; Raes et al. 2003). In this study, in total 14 *PAL* homologues including four *PAL1*, six *PAL2*, two *PAL3*, and two *PAL4* homologues across eight chromosomes were identified in the Express 617 genome assembly (Lee et al. 2020). The organ-specific expression in developing *B. napus* seeds showed that both *PAL4* homoeologues are much higher expressed than the other paralogues. Ding et al. (2021) found similar results in comparing the seed coat expression in high to low lignin cultivars and stated that *BnaPAL4* was significantly higher expressed in

the high lignin lines than in the low lignin lines. Hong et al. (2017) found several DEGs involved in phenylpropanoid and flavonoid biosynthesis during seed development when comparing yellow and black seeded *B. napus* cultivars. The expression results show that *BnaPAL4* C05 was downregulated in low lignin lines as expected (Hong et al. 2017).

In *B. napus* cultivars, typically two *PAL4* homoeologues can be found on chromosomes C05 and A05. In this study, a HNRT in SGDH14 led to the deletion of *BnaPAL4* on C05 and its replacement by a copy of the A05 *BnaPAL4*. The two A05 *BnaPAL4* copies of SGDH14 revealed a ~ 1.5 higher expression compared to the combined C05 and A05 *BnaPAL4* expression of Express 617. Thus, although SGDH14 lacks the C05 *PAL4* homoeologue, it still harbours two A05 *PAL4* copies which are expressed in seeds. But apparently these cannot complement SGDH14's low lignin phenotype. Several hypotheses may explain the effect on lignin content of the two *BnaPAL4* A05 homologues in SGDH14. As stated previously, the C05 *BnaPAL4* and A05 *BnaPAL4* might be involved in different pathways. While the C05 *BnaPAL4* could be specific for the seed lignin biosynthesis, the A05 *BnaPAL4* could be primarily involved in other pathways, e.g. in PA biosynthesis. The analysis of seed coat layers spatio transcription patterns would enable to investigate expression of the two *BnaPAL4* A05 (and A05'') homologues in SGDH14. They are expressed in the inner integument, endothelium, of the seed coat where PAs are synthesised (Moïse et al. 2005; Auger et al. 2010) while not being expressed in the palisade cells where lignin is synthesised (Jiang and Deyholos 2010; Cosgrove and Jarvis 2012), i.e. they might not contribute to lignin biosynthesis. We observed several SNPs in the putative promoter sequence between the A05 and C05 *BnaPAL4* homoeologues, which might affect the binding of different transcription factors or regulators. This could in turn determine different pathway specificities of the A05 and C05 *BnaPAL4* homoeologues. We propose that *BnaPAL4* on chromosome C05 is stronger associated with the lignin biosynthesis in the seed coat than its homoeologue on A05. Yin et al. (2022) found a similar divergence in function in accumulation of lignins, flavonoids and glucosinolates of two types of cinnamoyl-CoA reductase genes in *B. napus*. Finally, besides *PAL4*, additional candidate genes like *SEC8*, *SDP6*, and *MUCI21* might play a more important role in the control of seed lignin content than initially thought. However, clearly further functional genomic research is necessary to decipher the role of the identified candidate genes and especially of both *BnaPAL4* homoeologues in lignin biosynthesis in oilseed rape.

Declarations

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ambavaram MMR, Krishnan A, Trijatmiko KR, Pereira A (2011) Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. Plant Physiol 155(2):916–931. https://doi.org/10.1104/pp.110.168641

Andrews S (2010) FastQC. A quality control tool for high throughput sequence data, Brbraham Bioinformatics

Auger B, Marnet N, Gautier V, Maia-Grondard A, Leprince F, Renard M, Guyot S, Nesi N, Routaboul J-M (2010) A detailed survey of seed coat flavonoids in developing seeds of *Brassica napus* L. J Agric Food Chem 58(10):6246–6256. https://doi.org/10.1021/jf903619v

Badani AG, Snowdon RJ, Wittkop B, Lipsa FD, Baetzel R, Horn R, de Haro A, Font R, Lühs W, Friedt W (2006) Colocalization of a partially dominant gene for yellow seed colour with a major QTL influencing acid detergent fibre (ADF) content in different crosses of oilseed rape (*Brassica napus*). Genome 49(12):1499–1509. https://doi.org/10.1139/g06-091

Behnke N, Suprianto E, Möllers C (2018) A major QTL on chromosome C05 significantly reduces acid detergent lignin (ADL) content and increases seed oil and protein content in oilseed rape (*Brassica napus L.*). Theor Appl Genet 131(11):2477–2492. https://doi.org/10.1007/s00122-018-3167-6

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinform (oxford, England) 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34(5):525–527. https://doi.org/10.1038/nbt.3519

Carré P, Citeau M, Robin G, Estorges M (2016) Hull content and chemical composition of whole seeds, hulls and germs in cultivars of rapeseed (*Brassica napus*). OCL 23(3):A302. https://doi.org/10.1051/ocl/2016013

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Wang X, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker P (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345(6199):950–953

Cosgrove DJ, Jarvis MC (2012) Comparative structure and biomechanics of plant primary and secondary cell walls. Front Plant Sci. https://doi.org/10.3389/fpls.2012.00204

Dimov Z, Suprianto E, Hermann F, Möllers C (2012) Genetic variation for seed hull and fibre content in a collection of European winter oilseed rape material (*Brassica napus L.*) and development of NIRS calibrations. Plant Breeding 131(3):361–368. https://doi.org/10.1111/j.1439-0523.2012.01951.x

Ding Y, Yu S, Wang J, Li M, Qu C, Li J, Liu L (2021) Comparative transcriptomic analysis of seed coats with high and low lignin contents reveals lignin and flavonoid biosynthesis in *Brassica napus*. BMC Plant Biol 21(1):246. https://doi.org/10.1186/s12870-021-03030-5

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics. https://doi.org/10.1093/bioinformatics/bts635

Durkee AB (1971) The nature of tannin in rapeseed (*Brassica campestris*). Phytochemistry. https://doi.org/10.1016/0031-9422(71)85028-8

Fraser CM, Chapple C (2011) The phenylpropanoid pathway in Arabidopsis. The Arabidopsis Book 9:0152. https://doi.org/10.1199/tab.0152

Gabur I, Chawla HS, Liu X, Kumar V, Faure S, von Tiedemann A, Jestin C, Dryzska E, Volkmann S, Breuer F, Delourme R, Snowdon R, Obermeier C (2018) Finding invisible quantitative trait loci with missing data. Plant Biotechnol J 16(12):2102–2112. https://doi.org/10.1111/pbi.12942

Gaeta RT, Pires CJ (2010) Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol 186(1):18–28. https://doi.org/10.1111/j.1469-8137.2009.03089.x

Higgins EE, Clarke WE, Howell EC, Armstrong SJ, Parkin IAP (2018) Detecting de novo homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array.

G3 Bethesda, Md 8(8): 2673–2683 doi: https://doi.org/10.1534/g3.118.200118

Holzenkamp K, Link W, Möllers C (2022) Effect of a low seed lignin content on seed yield and quality in winter oilseed rape (*Brassica napus*). Plant Breeding. https://doi.org/10.1111/pbr.13059

Hong M, Hu K, Tian T, Li X, Chen L, Zhang Y, Yi B, Wen J, Ma C, Shen J, Fu T, Tu J (2017) Transcriptomic analysis of seed coats in yellow-seeded *Brassica napus* reveals novel genes that influence proanthocyanidin biosynthesis. Front Plant Sci 8:1674. https://doi.org/10.3389/fpls.2017.01674

Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai CJ, Chiang VL (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nat Biotechnol 17(8):808–812. https://doi.org/10.1038/11758

Hu Z-Y, Hua W, Zhang L, Deng L-B, Wang X-F, Liu G-H, Hao W-J, Wang H-Z (2013) Seed structure characteristics to form ultrahigh oil content in rapeseed. PLoS ONE 8(4):e62099. https://doi.org/10.1371/journal.pone.0062099

Hurgobin B, Golicz AA, Bayer PE, Chan C-KK, Tirnaz S, Dolatabadian A, Schiessl SV, Samans B, Montenegro JD, Parkin IAP, Pires JC, Chalhoub B, King GJ, Snowdon R, Batley J, Edwards D (2018) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. Plant Biotechnol J 16(7):1265–1274. https://doi.org/10.1111/pbi.12867

Ismail BP, Senaratne-Lenagala L, Stube A, Brackenridge A (2020) Protein demand: review of plant and animal proteins used in alternative protein product development and production. Animal Front: Rev Magaz Animal Agriculture 10(4):53–63. https://doi.org/10.1093/af/vfaa040

Iwanowska A, Tykarska T, Kuras M, Zobel AM (1994) Localization of phenolic compounds in the covering tissues of the embryo of *Brassica napus* (L) during different stages of embryogenesis and seed maturation. Ann Bot 74(4):313–320. https://doi.org/10.1006/anbo.1994.1123

Jiang Y, Deyholos MK (2010) Transcriptome analysis of secondary-wall-enriched seed coat tissues of canola (*Brassica napus L.*). Plant Cell Rep 29(4):327–342. https://doi.org/10.1007/s00299-010-0824-x

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772–780. https://doi.org/10.1093/molbev/mst010

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

Kulich I, Cole R, Drdová E, Cvrcková F, Soukup A, Fowler J, Zárský V (2010) Arabidopsis exocyst subunits SEC8 and EXO70A1 and exocyst interactor ROH1 are involved in the localized deposition of seed coat pectin. New Phytol 188(2):615–625. https://doi.org/10.1111/j.1469-8137.2010.03372.x

Lee H, Chawla HS, Obermeier C, Dreyer F, Abbadi A, Snowdon R (2020) Chromosome-scale assembly of winter oilseed rape *Brassica napus*. Front Plant Sci 11:496. https://doi.org/10.3389/fpls.2020.00496

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352

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

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic

features. Bioinform (oxford, England) 30(7):923–930. https://doi.org/10.1093/bioinformatics/btt656

Liu L, Qu C, Wittkop B, Yi B, Xiao Y, He Y, Snowdon RJ, Li J (2013) A high-density SNP map for accurate mapping of seed fibre QTL in *Brassica napus L.* PLoS ONE 8(12):e83052. https://doi.org/10.1371/journal.pone.0083052

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15(12):550. https://doi.org/10.1186/s13059-014-0550-8

Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, Zhang C, Chen Z, Xiao Z, Jian H, Cheng F, Zhang K, Du H, Cheng X, Qu C, Qian W, Liu L, Wang R, Zou Q, Ying J, Xu X, Mei J, Liang Y, Chai Y-R, Tang Z, Wan H, Ni Y, He Y, Lin N, Fan Y, Sun W, Li N-N, Zhou G, Zheng H, Wang X, Paterson AH, Li J (2019) Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. Nat Commun 10(1):1154. https://doi.org/10.1038/s41467-019-09134-9

Mason AS, Wendel JF (2020) Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. Front Genet 11:1014. https://doi.org/10.3389/fgene.2020.01014

Miao L, Chao H, Chen L, Wang H, Zhao W, Li B, Zhang L, Li H, Wang B, Li M (2019) Stable and novel QTL identification and new insights into the genetic networks affecting seed fiber traits in *Brassica napus*. TAG. Theoretical and applied genetics. Theoretische Und Angewandte Genetik 132(6):1761–1775. https://doi.org/10.1007/s00122-019-03313-4

Moïse JA, Han S, Gudynaitę-Savitch L, Johnson DA, Miki BLA (2005) Seed coats: structure, development, composition, and biotechnology. In vitro cell. Dev Biol Plant 41(5):620–644. https://doi.org/10.1079/IVP2005686

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

Nesi N, Delourme R, Brégeon M, Falentin C, Renard M (2008) Genetic and molecular approaches to improve nutritional value of *Brassica napus L.* seed. CR Biol 331(10):763–771. https://doi.org/10.1016/j.crvi.2008.07.018

Nicolas SD, Le Mignon G, Eber F, Coriton O, Monod H, Clouet V, Huteau V, Lostanlen A, Delourme R, Chalhoub B, Ryder CD, Chèvre AM, Jenczewski E (2007) Homeologous recombination plays a major role in chromosome rearrangements that occur during meiosis of *Brassica napus* haploids. Genetics 175(2):487–503. https://doi.org/10.1534/genetics.106.062968

Price MN, Dehal PS, Arkin AP (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26(7):1641–1650. https://doi.org/10.1093/molbev/msp077

Pucker B, Brockington SF (2018) Genome-wide analyses supported by RNA-Seq reveal non-canonical splice sites in plant genomes. BMC Genom 19(1):980. https://doi.org/10.1186/s12864-018-5360-z

Pucker B, Reiher F, Schilbert HM (2020) Automatic Identification of Players in the Flavonoid Biosynthesis with Application on the Biomedicinal Plant *Croton tiglium*. Plants (Basel, Switzerland) https://doi.org/10.3390/plants9091103

Quettier A-L, Shaw E, Eastmond PJ (2008) SUGAR-DEPENDENT6 encodes a mitochondrial flavin adenine dinucleotide-dependent glycerol-3-p dehydrogenase, which is required for glycerol catabolism and post germinative seedling growth in Arabidopsis. Plant Physiol 148(1):519–528. https://doi.org/10.1104/pp.108.123703

Raes J, Rohde A, Christensen JH, van de Peer Y, Boerjan W (2003) Genome-wide characterization of the lignification toolbox in Arabidopsis. Plant Physiol 133(3):1051–1071. https://doi.org/10.1104/pp.103.026484

Rahman M, McVetty P (2011) A review of Brassica seed color. Can J Plant Sci 91(3):437–446. https://doi.org/10.4141/cjps10124

Ralet M-C, Crépeau M-J, Vigouroux J, Tran J, Berger A, Sallé C, Granier F, Botran L, North HM (2016) Xylans provide the structural driving force for mucilage adhesion to the arabidopsis seed coat. Plant Physiol 171(1):165–178. https://doi.org/10.1104/pp.16.00211

Rambaut A (2006) FigTree. https://github.com/rambaut/figtree/releases

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29(1):24–26. https://doi.org/10.1038/nbt.1754

Samans B, Chalhoub B, Snowdon RJ (2017) Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species *Brassica napus*. The Plant Genome. https://doi.org/10.3835/plantgenome2017.02.0013

Schiessl S-V, Katche E, Ihien E, Chawla HS, Mason AS (2019) The role of genomic structural variation in the genetic improvement of polyploid crops. The Crop Journal 7(2):127–140. https://doi.org/10.1016/j.cj.2018.07.006

Schilbert HM, Schöne M, Baier T, Busche M, Viehöver P, Weisshaar B, Holtgräwe D (2021) Characterization of the *Brassica napus* flavonol synthase gene family reveals bifunctional flavonol synthases. Front Plant Sci 12:733762. https://doi.org/10.3389/fpls.2021.733762

Sharpe AG, Parkin IAP, Keith DJ, Lydiate DJ (1995) Frequent non-reciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome. https://doi.org/10.1139/g95-148

Si P, Mailer RJ, Galwey N, Turner DW (2003) Influence of genotype and environment on oil and protein concentrations of canola (*Brassica napus L.*) grown across southern Australia. Aust J Agric Res 54(4):397. https://doi.org/10.1071/AR01203

Siadjeu C, Pucker B, Viehöver P, Albach DC, Weisshaar B (2020) High contiguity de novo genome sequence assembly of trifoliate yam (*Dioscorea dumetorum*) USING long read sequencing. Genes 11(3):274. https://doi.org/10.3390/genes11030274

Simbaya J, Slominski BA, Rakow G, Campbell LD, Downey RK, Bell JM (1995) Quality characteristics of yellow seeded brassica seed meals protein, carbohydrates and dietary fiber components. J Agric Food Chem 43:2062–2066. https://doi.org/10.1021/jf00056a020

Slominski B, Simbaya J, Campbell L, Rakow G, Guenter W (1999) Nutritive value for broilers of meals derived from newly developed varieties of yellow-seeded canola. Anim Feed Sci Technol 78(3–4):249–262. https://doi.org/10.1016/S0377-8401(99)00003-6

Sochorová J, Coriton O, Kuderová A, Lunerová J, Chèvre A-M, Kovařík A (2017) Gene conversion events and variable degree of homogenization of rDNA loci in cultivars of *Brassica napus*. Ann Bot 119(1):13–26. https://doi.org/10.1093/aob/mcw187

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. https://doi.org/10.1093/nar/gki458

Stein A, Wittkop B, Liu L, Obermeier C, Friedt W, Snowdon RJ (2013) Dissection of a major QTL for seed colour and fibre content in *Brassica napus* reveals colocalization with candidate genes for phenylpropanoid biosynthesis and flavonoid deposition. Plant Breed 132(4):382–389. https://doi.org/10.1111/pbr.12073

Stein A, Coriton O, Rousseau-Gueutin M, Samans B, Schiessl SV, Obermeier C, Parkin IAP, Chèvre A-M, Snowdon RJ (2017) Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. Plant Biotechnol J 15(11):1478–1489. https://doi.org/10.1111/pbi.12732

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 A-M (2010) The first meiosis of resynthesized *Brassica napus*, a genome blender. New Phytol 186(1):102–112. https://doi.org/10.1111/j.1469-8137.2010.03182.x

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. Science 320(5875):486–488. https://doi.org/10.1126/science.1153917

Van Soest P, Robertson J, Lewis B (1991) Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. Dairy Sci.

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol 153(3):895–905. https://doi.org/10.1104/pp.110.155119

Wang J, Jian H, Wei L, Qu C, Xu X, Lu K, Qian W, Li J, Li M, Liu L (2015) Genome-wide analysis of seed acid detergent lignin (ADL) and hull content in rapeseed (*Brassica napus L.*). PLoS ONE 10(12):e0145045. https://doi.org/10.1371/journal.pone.0145045

Wang J, Xian X, Xu X, Qu C, Lu K, Li J, Liu L (2017) Genome-wide association mapping of seed coat Color in *Brassica napus*. J Agric Food Chem 65(26):5229–5237. https://doi.org/10.1021/acs.jafc.7b01226

Wang Y, Meng G, Chen S, Chen Y, Jiang J, Wang Y-P (2018) Correlation analysis of phenolic contents and antioxidation in yellow- and black-seeded *Brassica napus*. Molecules Basel, Switzerland. https://doi.org/10.3390/molecules23071815

Wick RR, Judd LM, Holt KE (2019) Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol 20(1):129. https://doi.org/10.1186/s13059-019-1727-y

Widiarsih S, Nagel M, Börner A, Feussner K, Feussner I, Möllers C (2021) Inheritance of seed quality and seed germination in two doubled haploid populations of oilseed rape segregating for acid detergent lignin (ADL) content. Euphytica. https://doi.org/10.1007/s10681-021-02891-z

Wittkop B, Snowdon RJ, Friedt W (2009) Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. Euphytica 170(1–2):131–140. https://doi.org/10.1007/s10681-009-9940-5

Wittkop B, Snowdon RJ, Friedt W (2012) New NIRS calibrations for fiber fractions reveal broad genetic variation in *Brassica napus* seed quality. J Agric Food Chem 60(9):2248–2256. https://doi.org/10.1021/jf204936f

Xu W, Grain D, Bobet S, Le Gourrierec J, Thévenin J, Kelemen Z, Lepiniec L, Dubos C (2014) Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB–bHLH–WDR complexes and their targets in Arabidopsis seed. New Phytol. https://doi.org/10.1111/nph.12620

Yin N, Li B, Liu X, Liang Y, Lian J, Xue Y, Qu C, Lu K, Wei L, Wang R, Li J, Chai Y (2022) Two types of cinnamoyl-CoA reductase function divergently in accumulation of lignins, flavonoids and glucosinolates and enhance lodging resistance in *Brassica napus*. The Crop J 10(3):647–660. https://doi.org/10.1016/j.cj.2021.10.002

Yusuf AO, Richter JC, Möllers C (2022) Genetic variation and QTL analysis of saturated fatty acids in two doubled haploid populations of oilseed rape (*Brassica napus* L.). Euphytica 218(7):88. https://doi.org/10.1007/s10681-022-03043-7

Zhang Z, Gou X, Xun H, Bian Y, Ma X, Li J, Li N, Gong L, Feldman M, Liu B, Levy AA (2020) Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. Proc Natl Acad Sci USA 117(25):14561–14571. https://doi.org/10.1073/pnas.2003505117

Zhang Y, Zhang H, Zhao H, Xia Y, Zheng X, Fan R, Tan Z, Duan C, Fu Y, Li L, Ye J, Tang S, Hu H, Xie W, Yao X, Guo L (2022) Multi-omics analysis dissects the genetic architecture of seed coat content in *Brassica napus*. Genome Biol 23(1):86. https://doi.org/10.1186/s13059-022-02647-5

Zhao J, Becker HC, Zhang D, Zhang Y, Ecke W (2005) Oil content in a European × chinese rapeseed population QTL with additive and epistatic effects and their genotype-environment interactions. Crop Breeding, Gen Cytol 45:51–59. https://doi.org/10.2135/cropsci2005.0051a