A systems genomics and genetics approach to identify the genetic regulatory network for lignin content in *Brassica napus* seeds

Wentao Zhang[1], Erin E. Higgins[2], Stephen J. Robinson[2], Wayne E. Clarke[2], Kerry Boyle[1], Andrew G. Sharpe[3], Pierre R. Fobert[4] and Isobel A. P. Parkin[2]*

[1]Aquatic and Crop Resource Development, National Research Council of Canada, Saskatoon, SK, Canada, [2]Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, [3]Global Institute for Food Security (GIFS), University of Saskatchewan, Saskatoon, SK, Canada, [4]Aquatic and Crop Resource Development, National Research Council of Canada, Ottawa, ON, Canada

Seed quality traits of oilseed rape, *Brassica napus* (*B. napus*), exhibit quantitative inheritance determined by its genetic makeup and the environment via the mediation of a complex genetic architecture of hundreds to thousands of genes. Thus, instead of single gene analysis, network-based systems genomics and genetics approaches that combine genotype, phenotype, and molecular phenotypes offer a promising alternative to uncover this complex genetic architecture. In the current study, systems genetics approaches were used to explore the genetic regulation of lignin traits in *B. napus* seeds. Four QTL (qLignin_A09_1, qLignin_A09_2, qLignin_A09_3, and qLignin_C08) distributed on two chromosomes were identified for lignin content. The qLignin_A09_2 and qLignin_C08 loci were homologous QTL from the A and C subgenomes, respectively. Genome-wide gene regulatory network analysis identified eighty-three subnetworks (or modules); and three modules with 910 genes in total, were associated with lignin content, which was confirmed by network QTL analysis. eQTL (expression quantitative trait loci) analysis revealed four cis-eQTL genes including lignin and flavonoid pathway genes, *cinnamoyl-CoA-reductase* (*CCR1*), and *TRANSPARENT TESTA* genes *TT4*, *TT6*, *TT8*, as causal genes. The findings validated the power of systems genetics to identify causal regulatory networks and genes underlying complex traits. Moreover, this information may enable the research community to explore new breeding strategies, such as network selection or gene engineering, to rewire networks to develop climate resilience crops with better seed quality.

1 Introduction

Understanding the genetic and molecular architecture underlying complex quantitative traits or phenotypic variation is the central goal of current biology (Kliebenstein, 2009). Changing transcript abundance arising from sequence polymorphism is one possible underlying molecular mechanism for complex traits (Rockman and Kruglyak, 2006). A wide range of complex traits within multiple species have been shown to be controlled by their underlying gene expression (Kliebenstein, 2009). A large-scale analysis of the relationship between complex diseases and known non-synonymous single nucleotide polymorphisms (SNPs) found that an increased number of SNPs did not find a significant number of associations, further indicating the potential role of variation in gene expression in controlling phenotypic traits (Burton et al., 2007). Together, these findings support the hypothesis that gene expression acts as the intermediate molecular phenotype, connecting genotypic variation to phenotypic variation (Rockman and Kruglyak, 2006).

The traditional quantitative trait loci (QTLs) mapping approach has been widely used to dissect the molecular mechanism underlying complex traits. However, most complex traits are controlled by multiple genes (Kliebenstein, 2009), and multiple QTLs with marginal effects (Mackay et al., 2009), which can be difficult to dissect with traditional QTL analysis. With the advent of advanced high-throughput technologies, including expression arrays and genome sequencing technologies, a new approach emerged to meet the challenge of complex traits, gene expression quantitative trait locus (eQTL) analyses. This approach is characterized by its combination of population genomic and quantitative genetics (Jansen and Nap, 2001). Briefly, QTL analysis is applied to the transcript abundances measured from a well-genotyped segregating population, allowing genetic loci (eQTL) controlling variation of each transcript to be identified on a genome-wide level (Keurentjes et al., 2007). A locus of sequence polymorphism affecting gene expression that is tightly linked to the physical position of the gene itself is defined as a cis-eQTL; all other eQTLs act in trans (trans-eQTLs) (Brem et al., 2002). The heritability of transcript variation and the genetic effect on this variation can also be inferred from eQTL analysis. Studies from Arabidopsis (West et al., 2007; Keurentjes et al., 2007), yeast (Brem and Kruglyak, 2005), maize (Schadt et al., 2003), and barley (Potokina et al., 2007), revealed an uneven distribution of trans-eQTLs with some genomic regions identified as eQTL hotspots (Schadt et al., 2003), enriched with numerous trans-eQTLs. Identification and dissection of eQTL hotspots may help to identify underlying master regulators or pleiotropic genes, and concomitantly the affected downstream genes or biological pathways. Thus, global eQTL analysis has the potential to yield great insight into the complex genetic architecture of transcript variation, and when linked to phenotypic variation (pQTLs), molecular mechanisms underlying complex traits can be directly inferred.

Several *in silico* large-scale data analyses have demonstrated the power of gene co-expression networks in dissecting a range of biological problems such as secondary wall biosynthesis (Persson et al., 2005), glucosinolate accumulation (Hirai et al., 2007), and seed germination and dormancy (Bassel et al., 2011) at the systems level. Gene regulatory networks were also hypothesized to constitute important information that can be inferred from eQTL data (Jansen and Nap, 2001); yet, this component has often been overlooked. Studies on seed development of Arabidopsis demonstrated that seed filling was controlled by a dynamic and distinct developmental transcriptional program, and genes involved in central carbon transport and multiple other metabolic pathways were implicated (Ruuska et al., 2002). However, most eQTL analyses have been limited to single gene-level analysis.

Fibre in seed meal negatively affects digestion and processibility for animal feed, and reducing the seed fibre content in oilseed rape could result in increased seed oil and protein content (Wittkop et al., 2009; Snowdon et al., 2010; Liu et al., 2012, 2013; Yu et al., 2016; Behnke et al., 2018; Miao et al., 2019; Chao et al., 2022; Yusuf and Möllers, 2024). Therefore, the low fibre trait was identified as a high-value target for canola (*Brassica napus*) breeding programs. Crude fibre is mainly comprised of lignin, including neutral detergent fibre (NDF), acid detergent fibre (ADF), acid detergent lignin (ADL), and polysaccharides including cellulose and hemicellulose (Van Soest et al., 1991; Wittkop et al., 2012). Genetics studies revealed a few major QTLs for lignin traits concentrated on chromosomes A09, C09, and C05 (Badani et al., 2006; Liu et al., 2012, 2013; Stein et al., 2013; Yu et al., 2016; Behnke et al., 2018; Miao et al., 2019; Chao et al., 2022). Lignin biosynthetic pathway(s) are complex. Forward and reverse genetics and biochemical analyses have identified more than 600 involved genes, with a subset of key enzymes that include *phenylalanine ammonia-lyase1* (*PAL1*) and *PAL2*; *cinnamate 4-hydroxylase* (*C4H*); *4-coumarate: CoA ligase1* (*4CL1*) and *4CL2*; *caffeoyl-CoA O-methyltransferase1* (*CCoAOMT1*); *cinnamoyl-CoA reductase1* (*CCR1*); *ferulate 5-hydroxylase* (*F5H1*); *caffeic acid O-methyltransferase* (*COMT*); and *cinnamyl alcohol dehydrogenase6* (*CAD6*) (Vanholme et al., 2010). Previous studies also reported that QTLs for seed colour colocalize with lignin content in *B. napus* seeds (Badani et al., 2006; Liu et al., 2012, 2013; Yu et al., 2016; Chao et al., 2022). In Arabidopsis, genes in the flavonoid pathway, including those from the *TRANSPARENT TESTA* (TT) family (eg. *TT1-TT19*) and transcription factor complex MYB-bHLH-WD40, are key regulators controlling seed colour (Lepiniec et al., 2006; Xu et al., 2014, 2015). The biosynthetic pathways of lignin and pigments share the same precursors (Vanholme et al., 2012), indicating that the two complex biosynthetic pathways may act together to control both lignin and seed colour, and possibly other quality traits in *B. napus* seeds.

As Arabidopsis is a very close relative of *B. napus* within the Brassicaceae family, it is reasonable to infer that seed quality traits of *B. napus*, e.g., lignin, are also controlled by interactions of multiple metabolic and biological pathways. Therefore, a systems genetic approach that combines quantitative genetics, regulatory network analysis, and eQTL analysis would be more appropriate to investigate seed quality traits within *B. napus* (Kliebenstein, 2009 and Marshall-Colón and Kliebenstein, 2019; Kliebenstein, 2020). In the present study, we aimed to fully exploit the power of this

approach to decipher the regulatory network underlying lignin traits in *B. napus* seeds.

2 Materials and methods

2.1 Plant materials

The *B. napus* lines used in this study include the two parental lines DH12075, a Canadian spring-type doubled haploid canola line (generated by Dr. Gerhard Rakow and Dr. Ginette Séguin-Swartz, Agriculture and Agri-Food Canada), and PSA12, a resynthesised *B. napus* line (generated by Dr. Monica Beschorner and Dr. Derek Lydiate, Agriculture and Agri-Food Canada), and 96 doubled haploid segregating (SG) lines generated from the cross of these two parental lines. PSA12 was developed from an interspecific cross between *B. oleracea ssp. alboglabra* line A12DH and *B. rapa* line Parkland Sunshine, and DH12075 was derived from a cross between varieties Cresor and Westar (Mayerhofer et al., 2005).

The developing seeds at 21 days post-anthesis were collected from *B. napus* lines grown under field conditions at the AAFC Saskatoon Research Farm in 2009. This time point was selected because approximately 20 days after flowering (DAF) had previously been identified as a critical stage for cell proliferation and oil deposition (O'Hara et al., 2002; Dong et al., 2004). Total RNA was extracted from the developing seeds using a method modified from that of Oñate-Sánchez and Vicente-Carbajosa (2008) and described in detail by Parkin et al. (2010). Seeds from field grown plants from two trials (2009 and 2010) were harvested for seed quality analysis.

2.2 Seed quality analysis

Seed quality traits including oil content (OilHD), protein, lignin content of NDF, ADF, and ADL, iodine value (IV), hydrogenated density (HD) total glucosinolate, and chlorophyll were estimated by near-infrared reflectance (NIR Systems Model 6500, FOSS, Eden Prairie, MN) following methods described by Yu et al. (2016). Results are presented as percentage of whole seed dry matter with zero moisture. The system was calibrated for oil content using the method of Raney and Serblowski (2007), using samples hexane extracted following the method described by Raney et al. (1987). Fatty acid composition of the extracted oil samples was also analysed using the Agilent 6890 GC-FID (Agilent Technologies, Santa Clara, CA) (OilEx).

2.3 Array processing and data analysis

cRNAs were amplified from total RNA samples (2 μg) and labelled with either cy3 or cy5 using the Quick Amp labelling kit, Two Colour (Agilent, Catalogue: 5190-0444) according to the manufacturer's instructions, followed by purification with the Qiagen RNeasy Mini Kit (Qiagen, Catalogue 74104) and quantification with a NanoDrop ND-1000. Purified cRNAs (2 μg)

were fragmented and hybridized to the Agilent 4x44K *Brassica* arrays (Parkin et al., 2010) for 17 h at 65°C with a rotation of 10 rpm according to the manufacturer's protocol. Arrays were washed with the Agilent Gene Expression Washing buffers according to the manufacturer's protocol and scanned with GenePix 4000B as described by Parkin et al. (2010). Four and two biological replications were performed for the parental lines and the SG lines, respectively, with dye-swaps applied to the biological replicates to minimize dye effects from the two colour systems (Lee et al., 2004).

The raw array datasets were extracted with Gene Pix Pro 6.0, and were firstly normalized by the lowess (Yang et al., 2002) method within arrays, followed by the quantile (Bolstad et al., 2003) method between arrays, using the R package Limma (Smyth, 2005). The normalized datasets of the SG lines were averaged across replications. Probes with intensities over 20 and with a frequency less than 30 percent from our datasets were treated as non-significant signal probes and removed, resulting in a final total of 29,752 quality probes. Finally, the normalized gene levels lower than 20 were adjusted to 20 as background intensity. The log2 transformed normalized data were treated as gene expression traits (e-traits) and were used to perform eQTL analysis.

2.4 Differential gene expression analysis of parental lines DH12075 and PSA12

Normalized and adjusted data of parental lines with four biological replications were imported to GeneSpring GX 10.0 (Agilent Technologies) and analysed differential gene expression with Student's t test. The p-value was adjusted with Benjamini-Hochberg's method and controlled at FDR = 0.05. Genes with adjusted p-value < 0.05 and fold changes equal or larger than 2.0 were claimed as significantly differentially expressed genes.

2.5 Genotyping and linkage map construction

Genomic DNA was extracted with a modified CTAB method and genotyped with simple sequence repeat (SSR) markers developed at the Saskatoon Research Centre, Agriculture and Agri-Food Canada (http://aafc-aac.usask.ca/BrassicaMAST/). The SSRs were analysed using fluorescently labelled tail PCR amplifications in 384-well format on an Applied Biosystems 3730xl. Genotype data was imaged using genographer (available at http://www.genographer.com/), and polymorphic loci were scored visually. The SNP genotyping data of the SG population were generated by the Brassica 60 K Infinium array from Clarke et al. (2016).

2.6 Linkage map construction and QTL mapping

Genetic maps were developed with Mapdisto (Lorieux, 2012) using 694 SSR markers and 1266 SNP (binned map) markers with a

cut-off recombination value of 0.3, and a threshold logarithm of odds (LOD) of 6.5. Physical positions of mapped markers were retrieved by BLASTN-searching SNP flanking sequences against the Darmor-bzh reference genome (Chalhoub et al., 2014).

QTL analysis was performed by composite interval mapping (CIM) as implemented in WinQTL Cartographer (Wang et al., 2007) with a walking speed of 1 cM and a window size of 10 cM. The significance of a QTL was claimed at the threshold of 5% significance level by the 3000-permutation test in WinQTL Cartographer (Churchill and Doerge, 1994; Basten et al., 2004; Wang et al., 2007). QTL within an interval of 10 cM were merged into one QTL, represented by the QTL peaks. QTL were visualized by the R package LinkageMapView (Ouellette et al., 2018).

2.7 Gene co-regulatory network construction

Normalized and filtered transcripts were used to construct a weighted gene co-regulatory network (WGCNA) with the step-by-step method implemented in the WGCNA R package (Zhang and Horvath, 2005; Langfelder and Horvath, 2008). The soft-thresholding power function analysis was applied correlation of adjacency of genes with the power value from 1 to 20 (Supplementary Figure S1). The optimized power value was 12, at which an approximal scale-free network was reached (Supplementary Figure S1). With this value, the adjacency matrix was transformed into a topological overlap matrix (TOM). Gene hierarchical clustering tree was generated by the dynamic tree-cutting method, hcluster. Modules with eigenvalue similarity >= 0.75 were merged, and modules with less than 30 genes were removed. Modules and their relationship to external traits were also obtained through the method within this package (Langfelder and Horvath, 2008). The *B. napus* seed network was generated with a hard cut-off value of |r|= 0.75 which made the network follow a roughly power-law distribution (Ravasz et al., 2002; Barabási and Oltvai, 2004). Identified gene interactions were imported into Cytoscape (Shannon et al., 2003) version 2.7.0 for visualization and additional analysis.

2.8 Module-trait association and module QTL (network QTL) analysis

The gene expression profile in each module was characterized by the eigenvalues of each module, and is referred to here as the module eigengene value (ME). The module-trait association was performed by Pearson correlation analysis between MEs and seed quality traits implemented within the WGCNA package. Eigengene value of the network can extract the key information (features) of the larger number of correlated genes within the network, and retain its most important variance of the network dataset, and reduce its dimensionality (Zhang and Horvath, 2005; Langfelder and Horvath, 2008). Thus, the ME that captures the summarized variance of the regulatory network can be treated as a classical phenotypic trait representing the correlated gene expression

patterns of the whole network. Subsequently, QTL mapping of MEs (network QTL) was performed as described above (section 2.6).

2.9 Gene annotation, functional and enrichment analysis

ESTs of array probes were used in BLASTP (E-value <1e−5, coverage >70%, identity >70%) against the *Arabidopsis* Information Resource protein database (TAIR 10). The BLASTP output and the *Arabidopsis* functional annotations were parsed with in-house scripts to derive functional annotation and relevant GO terms.

Gene function terms were characterized with Mapman (Usadel et al., 2005) ontology terms and mapped into different biological pathways with tools embedded in the Mapman software (Usadel et al., 2005). Gene function enrichment analysis within detected modules was performed with the Fisher's exact test and significance was claimed at p < 0.05.

2.10 eQTL analysis

The log transformed gene expression values were treated as a trait and analysed with the composite interval mapping (CIM) (Zeng, 1994) module "zmapqtl", which was implemented in QTL Cartographer version 1.17 (Basten et al., 2004) under the Linux system, with a walking speed of 1 cM and a window size of 10 cM. Due to the limited capacity of the computation, a global permutation threshold (GPT) described by West et al., 2007, was performed to obtain a threshold criterion for declaring a statistically significant eQTLs. The GPT value (p = 0.05) equals a LOD score of 4.75 for this eQTL analysis. eQTLs within 10 cM intervals were merged into one QTL represented by the QTL peak.

The sequence of the array probes was also used with BLASTN (E-value < 1e-10) against the Darmor-bzh reference genome (Chalhoub et al., 2014) and physical positions of the best-hit were retrieved. The eQTL was defined as cis- if the gene and the marker linked to it co-localized within a physical interval of 4 Mb; otherwise, they were defined trans-eQTLs. According to this definition, cis- and trans-eQTL of *B. napus* were identified.

3 Results

3.1 Characterization of QTL for lignin content

The lignin contents on the seeds of this DH population exhibited a normal distribution, assessed using the Shapiro-Wilk test (p < 0.05) (Supplementary Figure S2). The parental line DH12075 consistently showed higher lignin contents than the resynthesized PSA12 in different environments (Supplementary Figure S2). We consistently identified major QTLs for lignin content NDF, ADF, and ADL from two-years of field trials, on chromosomes C08 and A09 (Figure 1), while a minor QTL from

C01 was also identified in 2010. The NDF, ADF, and ADL shared the same QTLs on A09 and C08 (Figure 1), and since it has been suggested that investigation of ADF or NDF alone was sufficient to dissect the quantitative traits of fibre and lignin content (Cardinal et al., 2003), we therefore decided to use ADF as a representative trait to explore the genetic architecture underlying seed lignin traits. We defined three QTLs on A09: qLignin_A09_1 (comprised of qNDF_A09_1, qADF_A09_1 and qADL_A09_1); qLignin_A09_2 (from qNDF_A09_2, qADF_A09_2 and qADL_A09_2), and qLignin_A09_3 (from qNDF_A09_3, qADF_A09_3 and qADL_A09_3). The QTL on C08 was defined as qLignin_C08 (consisting of qNDF_C08, qADF_C08 and qADL_C08). qLignin_A09_1 spanned an interval of 130-145 cM and was located at 19-25 Mb on A09. It conferred effects with R^2 of 7% in

2009 and 12% in 2010. The qLignin_A09_2 conferred the largest effects in the two-year test, and localized between 145-160 cM with a physical interval of 25-30 Mb. The qLignin_A09_3 localized between 160-170 cM with an interval of 30-33 Mb on A09. The qLignin_C08 was found in the interval of 82-92 cM with the physical interval of 30-40 Mb on C08. The QTL qLignin_A09_1, qLignin_A09_3, and qLignin_C08 conferred a similar magnitude of effect on lignin content and with ~3-5% less than the qLignin_A09_2. Synteny analysis identified qLignin_A09_2 and qLignin_C08 are homologous QTL from the A and C subgenomes, respectively (Supplementary Figure S3). The alleles for all QTL on A09 that confer the higher level of lignin content are from parental line DH12075, while the alleles for the higher lignin QTL on C08 are from the synthetic parental line, PSA12.
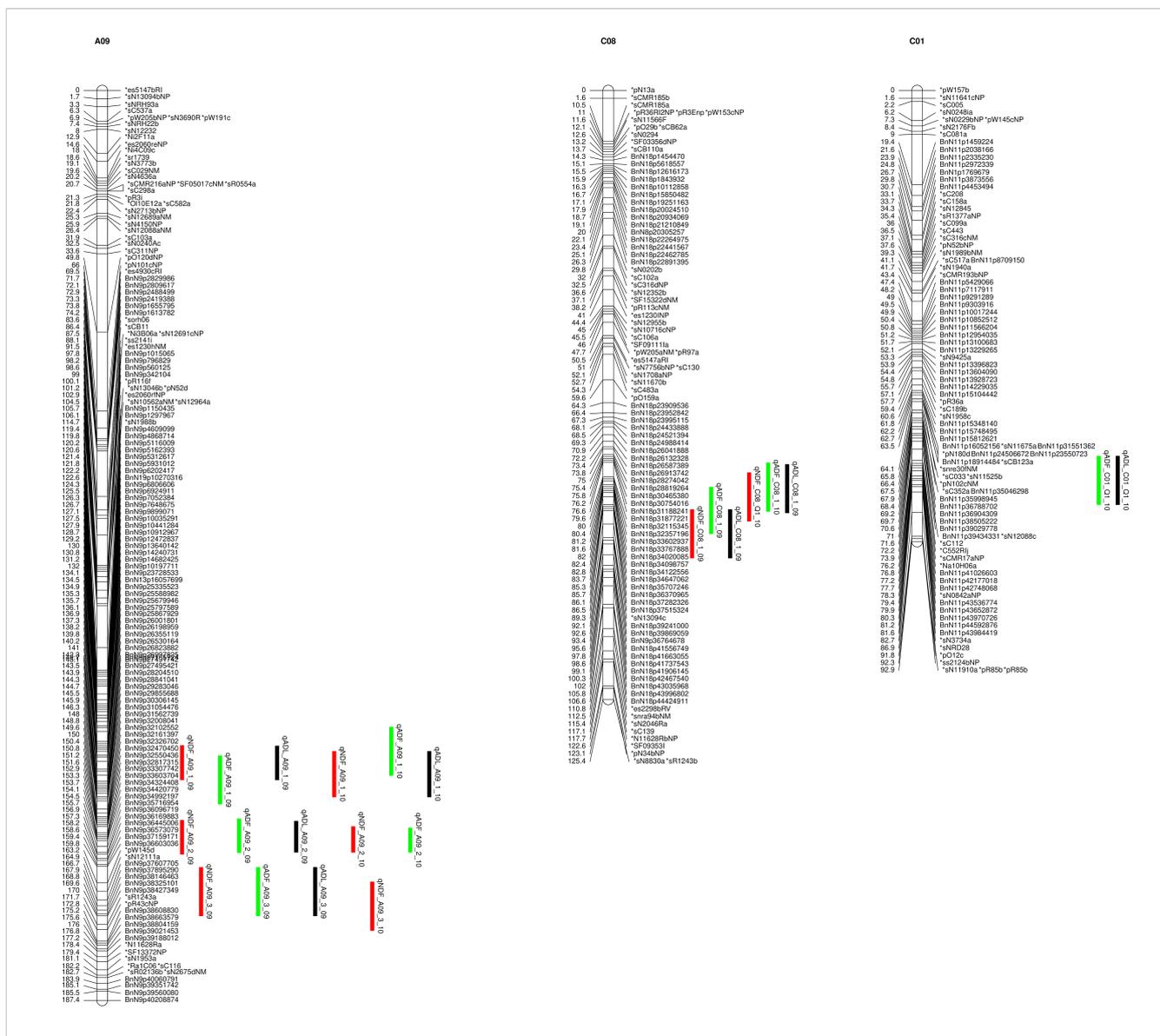
FIGURE 1
QTL for lignin content on identified on chromosomes A09 and C08 of the SG genetic map. NDF (neutral detergent fibre), indicated with red bars; ADF (acid detergent fibre), indicated by green bars, ADL (acid detergent lignin), indicated with black bars. 09 and 10 indicate the years 2009 and 2010.

3.2 Identification of genetic regulatory networks underlying lignin content

A total of 29,725 transcripts were used to build a co-regulatory network with WGCNA. The final co-expression network consisted of 12,670 nodes connected by 468,974 edges, containing ~60% of the information from the input datasets. The entire network was partitioned into 83 distinct gene modules (or subnetworks). A global view of the co-regulatory network with modules coloured distinctly on the network is shown in Figure 2. The number of genes within these subnetworks ranged from 30 to 7,303 genes.

Module-trait association analysis was performed to identify regulatory networks associated with seed quality traits in *B. napus* seeds; including oil content, protein, NDF, ADF, ADL, IV, glucosinolate, hydrogenated density, glucosinolate, and chlorophyll. Module 33 and module 35 were identified as significantly positively associated with NDF, ADF, and ADL, while module 67 was significantly negatively correlated with these three forms of lignin (Figure 3). As shown in Figure 3, the correlation coefficients between ADF and module 33, module 35, and module 67 were 0.6 (p-value: 3E-12), 0.6 (p-value: 9E-12), and -0.8 (p-value: 2e-22), respectively.

To validate that these modules are causal networks of lignin traits, a network QTL mapping approach was applied, utilizing linkage mapping to identify the genetic locus that controlled these modules. As shown in Figure 4, the QTL for module 33, qModule33_A09_1 overlapped with lignin pQTL qLignin_A09_1 on A09. The two QTLs for module 35, qModule35_A09_1 and qModule35_A09_2, colocalized with qLignin_A09_2 and qLignin_A09_3, respectively (Figure 4). The qModule67_A09_1 colocalized with qLignin_A09_2 and qModule_67_C08_1 collocated with qLignin_C08 (Figure 4). To further explore these
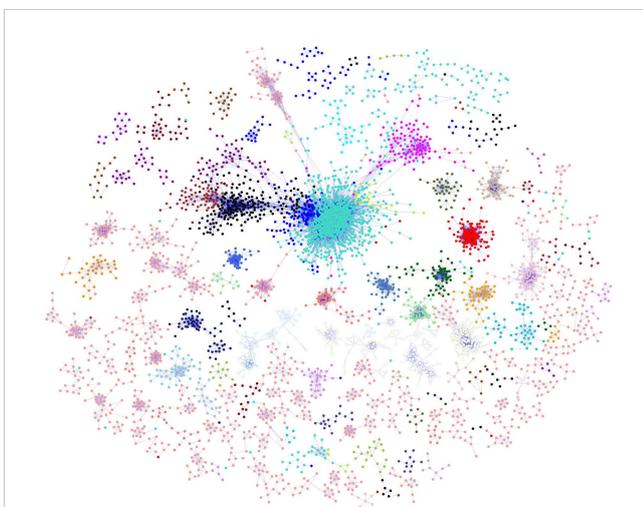
FIGURE 2
B. napus co-regulatory network. Nodes are probe sets from the customized *B. napus* array. The node colours depict the different modules identified from the network. Edges indicate significant co-expression between probe sets above a hard threshold. In order to improve the clarity of the display, only correlation coefficient (|r|>= 0.85) were shown on the network.

regulatory networks, genes were extracted from these three modules and a sub-network built, which we termed the lignin causative network (Figure 5).

3.3 Gene function and enrichment analysis of the genetic regulatory network for lignin content

From the metabolic overview map, the major metabolite pathways that corresponded with the change of lignin content were characterized by genes involved in secondary metabolism, lipid biosynthesis, and cell wall modification (Supplementary Figure S4). Gene function enrichment analysis revealed that the lignin causal network was enriched with genes related to secondary metabolites (p< 1.8-E04) (Supplementary Table S1). These secondary metabolism-related genes were mapped mainly on the flavonoid, phenylpropanoids, lignin, shikimate, and mevalonate (MVA) pathways (Supplementary Figure S5). Major genes involved in lignin and flavonoid pathways, including *BANYULS (BAN)*, *CCR1*, *C4H*, *PAL2*, and the *TT*-related genes, *TT3*, *TT4*, *TT6*, *TT7*, *TT8*, *TT18*, *TT19*, were found positively correlated with lignin as shown in Supplementary Table S2, Supplementary Figure S6.

Stress-related genes were also enriched in the regulatory network (p-value: 0.01) (Supplementary Table S1); therefore, we explored the patterns of transcripts involved in abiotic and biotic stress. An interesting change in gene expression patterns associated with biotic stress in response to the change of seed lignin was observed. Almost all secondary metabolite-related genes involved in biotic stress were positively correlated with the change of lignin, while many of the major genes involved in biotic stress, including WRKY transcription factors (*WRKY25* and *WRKY33*), ethylene-related AP2 transcription factors, ethylene-related genes, and MAP kinases, were negatively correlated with lignin accumulation (Supplementary Table S2, Supplementary Figure S7).

If the genes are causal for the trait, it can be hypothesized the eQTL of the gene will colocalize with the trait QTL (pQTL). Prioritized causal genes will be shown as cis-eQTLs, in which the gene controls the phenotypic variation by mediating its expression via cis-regulatory sequence variation. eQTL analysis of the 910 genes within the lignin regulatory network found 163 cis-eQTL and 491 trans-eQTL underlying the three lignin QTL of the A09 chromosome (Supplementary Table S3). There were 14 cis- and 136 trans- eQTL identified for the C08 lignin QTL (Supplementary Table S4). The lignin gene *CCR1* and seed colour genes *TT4*, *TT6*, and *TT8* were identified as cis-eQTLs that colocalized with the lignin QTL on A09. The key lignin genes, *PAL2* and *C4H*, and seed colour genes *TT3*, *TT7*, *TT18*, *TT19*, and *BAN* were found to be trans-eQTLs (Supplementary Tables S3, S4), and none of the cis-eQTLs of lignin and flavonoid genes from C08 QTL were found above the significance threshold value. Gene expression analysis between the two parental lines found that of the genes known to be important to the lignin pathway, only *CCR1* was significantly differentially expressed between the parental lines (Supplementary Table S5). DH lines with the eQTL of *CCR1* (based on the DH12075

FIGURE 3
Module-trait relationship between 83 modules and seed quality traits. Traits and rows represent the modules. Correlation co-efficient and p-value (in parenthesis) were shown in each cell. Correlation between modules and seed quality traits was presented by colours ranging from red (high positive correlation) to green (high negative correlation). Evaluated seed quality traits include oil content (OilHD, measured by NIR; OilEx, determined by GC-FID), protein, lignin content of NDF (neutral detergent fibre), ADF (acid detergent fibre), ADL (acid detergent lignin), iodine value (IV), hydrogenated density (HD), total glucosinolate (TotGSL), and chlorophyll.

marker type at the eQTL peak BnN9p34992197) have significantly higher group means of gene expression than lines without the eQTL (Supplementary Figure S8). The seed colour genes *TT8*, *TT6* and *TT19* were statistically differential expressed, but with a fold change of gene expression level between 1.5 to 2.0 (Supplementary Table S5). Lignin content was significantly associated with the expression of these genes with cis-QTL, *CCR1*, *TT6*, and *TT8* (Supplementary Figure S9).

4 Discussion

Complex traits or quantitative traits are typically controlled by complex genetic architecture consisting of hundreds to thousands of genes (Kliebenstein, 2009). These genes form a genetic regulatory network that controls the trait, which challenges the single gene or traditional QTL analysis. Therefore, we utilized a systems genetic approach that integrates the quantitative genetic, gene co-regulatory network, network-QTL, and eQTL analysis to uncover the causal regulatory network underlying lignin content in *B. napus* seeds.

Three major QTLs were identified on A09 spanning a physical interval of approximately 13 Mb. All three QTLs conferred stable and relatively large effects on lignin content. QTLs for NDF, ADF, and ADL colocalized at the same genetic locus, indicating these three components of lignin shared the same genetic architecture. Numerous QTLs for seed quality traits including seed colour, oil, protein, and fibre were previously identified on the A09 chromosome (Wittkop et al., 2009; Liu et al., 2012, 2013; Yu et al., 2016; Behnke et al., 2018; Miao et al., 2019; Chao et al., 2022; Yusuf and Möllers, 2024), indicating that A09 is a hotspot that harbours several genes for these seed quality traits. However, this study and all previous research have used a relatively small population size, from ~100 to 300 lines, making it challenging to precisely identify QTLs within this hotspot. Future, larger and advanced-structured genetic populations could be used to dissect this hotspot. In addition, a stable and large effect homologous QTL to the qLignin_A09_2 was identified on C08, indicating both the A and C genome contributed to the variation of lignin content in our population. Currently, limited studies, including Badani et al. (2006) and Chao et al. (2022), have reported the QTL on C08 for seed colour and lignin content. The relative simplicity of the C08 QTL, compared to A09, makes it a better target to uncover genes underlying lignin QTL in the future. Furthermore, since the C08 QTL is homologous to the A09 lignin QTL, uncovering its genetic architecture may also shed light on the A09 hotspot.
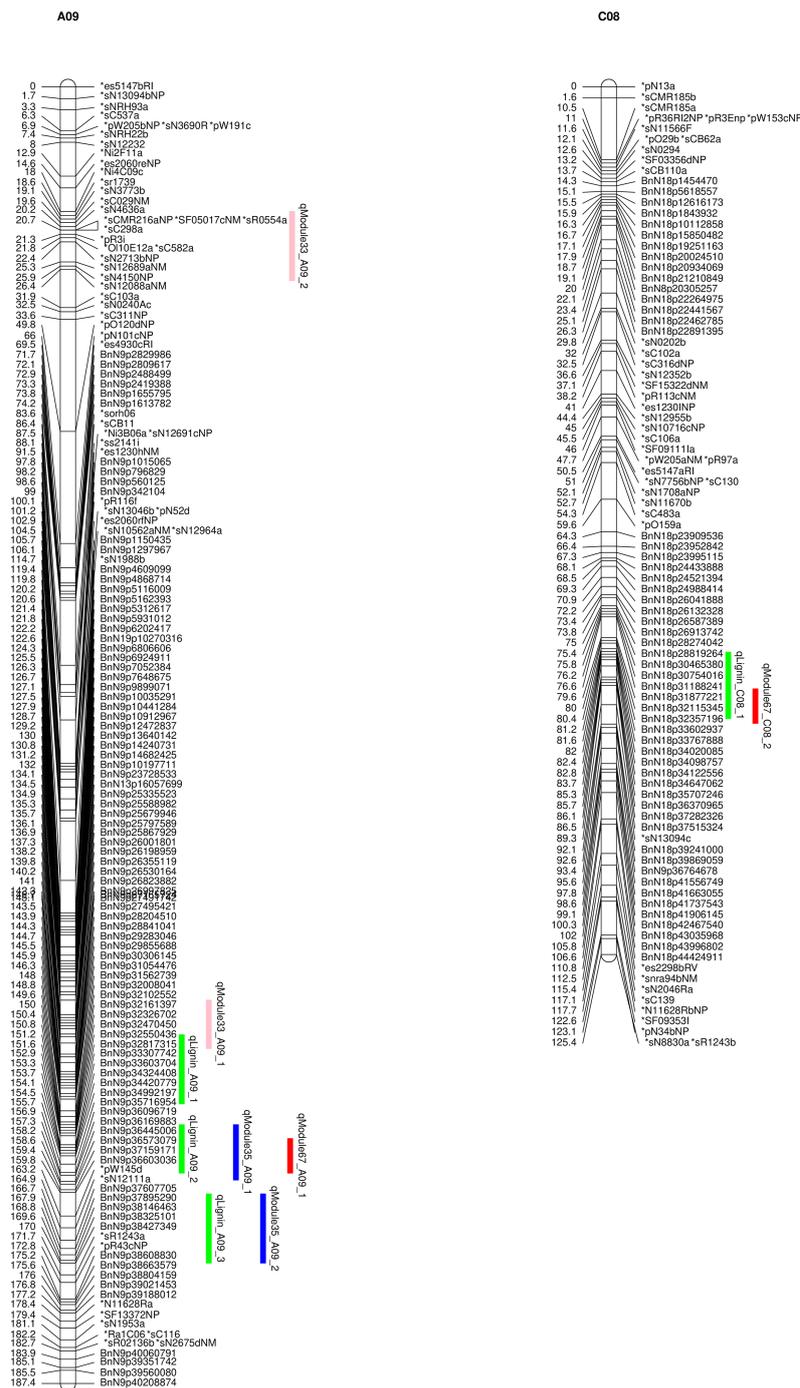
FIGURE 4

Linked lignin phenotype QTLs (pQTLs) to module QTLs (mQTLs). qLignin are lignin QTL, and qModule are module QTL.

The genome-wide transcriptome datasets for developed seeds of the SG population, allowed a seed development biological network to be successfully built (Figure 2). Biological network construction assumes that it will follow the topological scale-free property, in which the distribution of the nodes and edges follows a power-law distribution (Barabási and Oltvai, 2004). Although definitions of the properties of biological networks is still under discussion, the seed developmental regulatory network did exhibit the expected scale-free topology, described as a small number of nodes with a large number of connections, while most nodes have few connections (Figure 2). The distribution of the nodes and edges in our network exhibited a power-law distribution, supporting its topological free property, and was in agreement with the network property discovered in Arabidopsis (Atias et al., 2009; Mao et al., 2009),

FIGURE 5
Co-regulatory network associated with lignin. Nodes are transcripts identified from modules 33, 35 and 67. Edges indicate significant co-expression between probe sets above a hard threshold. Node colours depict different Mapman ontology terms (legend right).
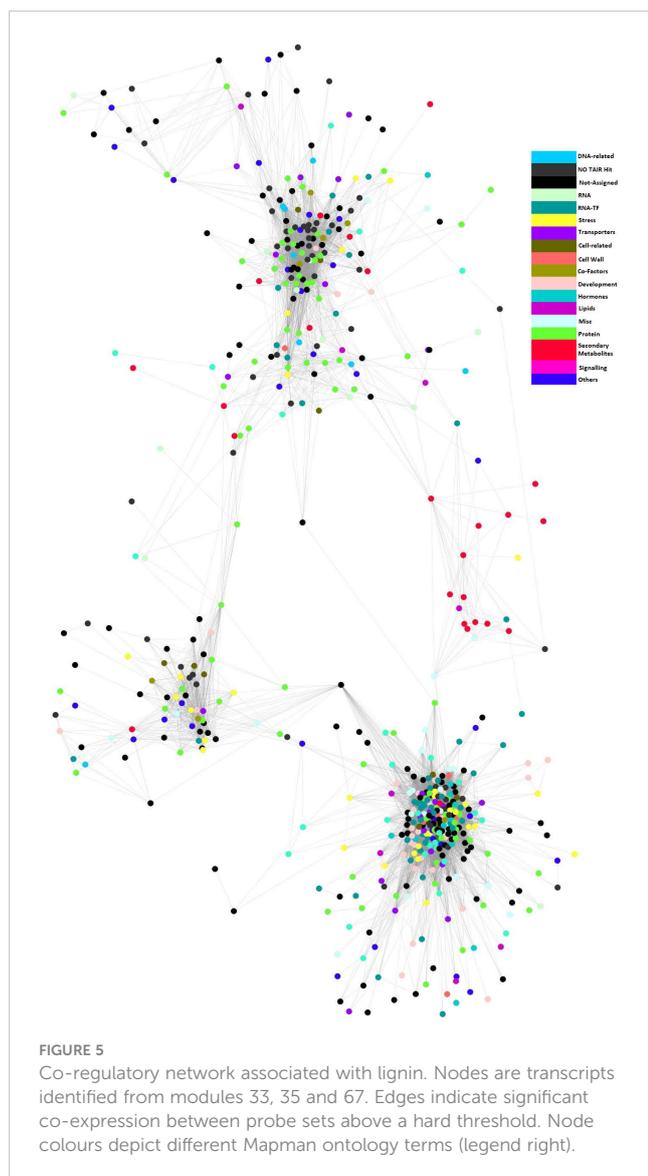
rice (Ficklin et al., 2010; Ficklin and Feltus, 2011) and maize (Ficklin and Feltus, 2011).

The structure of genome-wide co-expression networks are hierarchical with modular units (Aoki et al., 2007). Given that genes with similar biological functions are typically co-expressed together and formed functional modules, disassembling biological networks into functional modules or clusters is more informative and practical for biologists to interpret the network (Mutwil et al., 2010). Here, the seed network was divided into 83 computational biological modules. If phenotype variation is attributed to the variation of genes within the module, a correlation should be observed between modules and phenotypes. This was the case for two modules (modules 33 and 35) that were significantly positively correlated with lignin, and one (module 67) that was negatively coordinated with lignin. With network QTL analysis, it was found that lignin QTLs (pQTLs) and module QTLs shared the same

genetic loci on A09, supporting the suggestion that lignin variation was determined by the interactions of these three functional modules. These findings also support the hypothesis that entire plant networks can be inherited as a phenotypic trait (Marshall-Colón et al., 2019). Large-scale evaluation of 1,671 genes on agronomic performance in maize using a single gene engineering approach found that only one gene encoding a transcription factor, zmm28, showed promise, indicating the marginal effect of a single gene for the improvement of complex traits (Simmons et al., 2021). The defined lignin network could enable breeders to target a specific regulatory network and potentially use an engineering approach to rewire and fine-tune the network to improve crop performance.

Since modules 33, 35, and 67 were altogether associated with seed lignin, these three modules were assembled into one large lignin-related functional module, called "lignin causative network", to perform further analysis. This network was enriched with secondary metabolism-related genes. One group of these genes includes PAL2, C4H, and CCR1, encoding three key enzymes controlling critical steps of the lignin pathway (Vogt, 2010). The other group, which mainly consists of flavonoid pathway genes, is involved in the accumulation of flavonoids in the seed coat, such as TT3, TT7, TT6, TT8, TT18, TT19, and BAN. All these genes are significantly and positively associated with seed lignin variation. Enrichment of these secondary metabolite genes indicated the high reliability of this causative regulatory network underlying lignin content.

eQTL analysis was performed on genes within this network to identify the causal genes underlying the network for lignin content. The lignin gene, CCR1, was detected as a cis-eQTL, while both PAL and C4H were found as trans-eQTLs. Flavonoid genes including TT4, TT6, and TT8 were identified as cis-eQTLs, while the remaining transparent testa genes were trans-eQTLs. In addition, all these cis-eQTLs colocalize with the fibre QTL from chromosome A09, indicating they are prioritized as causal genes underlying this network (Kliebenstein, 2009). Liu et al. (2012) pinpointed CCR1 from chromosome A09 as the candidate gene for ADF in B. napus seeds. Zhang et al. (2022) identified a CCR-like gene, CCRL, from C07 as one candidate gene for seed content in B. napus. CCR1 has been found to encode the key enzyme for lignin accumulation in stems (Jones et al., 2001). The cis-controlled eQTL impacting the variation of CCR1, the co-localization with phenotype QTL, and the unique known gene from the monolignol pathway located on our lignin causative network as well as previous findings, indicated that CCR1 is one of the causative genes underlying the seed lignin trait. Of the three cis-regulated flavonoid genes, TT8 was found to be a key regulator for seed oil and colour in Arabidopsis and B. napus (Chen et al., 2014; Xu et al., 2014, 2015; Zhai et al., 2020; Zhang et al., 2022) and lignin in B. napus (Zhang et al., 2022). In addition, Zhai et al. (2020) found TT8, as a BHLH transcription factor, regulated both the flavonoid and lignin pathways. All these findings pinpoint TT8 as another key regulator in this network. Li et al. (2023) reported that TT4, which encodes an enzyme acting at an earlier step of the flavonoid pathway and located on chromosomes A03 and C02, is

key regulator for oil and protein content in *B. napus* seeds, while we identified *TT4* from A09 as playing a key role for lignin content. Therefore, it is hypothesized that multi-homologous and homoeologous *TT4* genes in different genotypes control seed quality traits. In addition, another earlier flavonoid biosynthetic gene *TT6* was identified as a potential new key regulator for lignin content in *B. napus*. Thus, the lignin traits associated with the A09 loci are controlled by a complex regulatory network with multiple genes.

Notably, we did not find any cis-eQTL from the key lignin and flavonoid genes associated with the homologous C08 lignin QTL. Network QTL analysis only found that the QTL of module 67 colocalized with the qLignin_C08 QTL. Genes for the development of *B. napus* seed are spatially and temporally regulated (Zhai et al., 2020). It is possible that the A and C genome coordinately regulate seed lignin content in a temporal manner, while the regulatory network built here, was only from one key time point, 21 days after flowering. We observed, under greenhouse conditions, that the A donor line of PSA12 flowered and matured earliest, followed by DH12075 and PSA12, while the C genome donor line of PSA12 flowered and matured very late (unpublished data), supporting this hypothesis, since positive alleles at the C08 locus were contributed by PSA12. A further spatial-temporal regulatory network built on seed development will be more powerful for dissecting the genetic architecture controlling seed quality traits and should be considered for future studies.

Another interesting finding of the lignin network was that it was enriched with biotic stress genes, including key transcript factors, *WRKY25*, *33* and *SIGMA FACTOR BINDING PROTEIN 1* (*SIB1*), which are involved in triggering resistance to the necrotrophic pathogen *Botrytis cinerea* in Arabidopsis (Andreasson et al., 2005; Zheng et al., 2006; Eulgem and Somssich, 2007; Pandey and Somssich, 2009; Lai et al., 2011). However, all these genes were significantly negatively associated with lignin accumulation, which is contradictory to the positive role of lignin on enhancing the resistance response (Vanholme et al., 2010). This paradoxical role of lignin in the plant defence mechanism compelled us to hypothesize that cell wall integrity plays a key role in this process. Reduced lignin affects the integrity of the wall which mimics disease infection, leading to the release of defence-related elicitors. When a normal plant is infected by disease, cell wall integrity is damaged and in response to this, the cell wall induces lignin biosynthesis to maintain its integrity. This theory is further supported by studies which have shown that cell wall damage directly triggers lignin biosynthesis (Denness et al., 2011).

It is worth noting a few limitations of the current study. First, the population size was small, and may not provide enough statistical power for identifying minor QTL. Second due to the limitation of array technology, many unknown genes within our causative network may be missing from our research. Third, this regulatory network was built only from one stage of seed development, and could not capture a complete temporal picture of development. New advanced structured genetic populations, and genome sequencing tools including long read and pangenome sequence analysis will empower these systems genetic approaches to uncover the genetic mechanism of complex traits.

5 Conclusion

An example of the power of systems genetics that combines quantitative genetic, regulatory network analysis, network QTL, and eQTL analysis to identify causal regulatory networks and causal genes for lignin traits was provided. Further insights were revealed into the complex genetic architecture that regulates lignin content, specifically that controlled by the locus on A09 of *B. napus*. The analyses provide more prioritized genes to target to improve seed quality trait in *B. napus*. Moreover, the research here may also provide a new avenue for crop communities to use regulatory networks to select or design crops with better performance and quality in a changing climate.

Data availability statement

The data presented in the study are deposited in the NCBI data repository, accession number GSE268725.

Author contributions

WZ: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. EH: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – review & editing. SR: Data curation, Methodology, Writing – review & editing. WC: Data curation, Methodology, Writing – review & editing. KB: Formal analysis, Visualization, Writing – review & editing. AS: Conceptualization, Funding acquisition, Writing – review & editing. PF: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – review & editing. IP: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1393621/full#supplementary-material

SUPPLEMENTARY TABLE 1
Gene Ontology (GO) Enrichment analysis of the fibre related regulatory network.

SUPPLEMENTARY TABLE 2
Gene function, module, and its correlation with acid detergent fibre (ADF). Significant correlations between gene expression levels and ADF were claimed at p-value < 0.05.

SUPPLEMENTARY TABLE 3
eQTLs overlapping with pQTLs identified from the fibre related network on A09. cis and trans, eQTL defined by B. napus cv. Darmor-bzh reference genome. cis*, cis-eQTL defined by the DH12075 reference genome.

SUPPLEMENTARY TABLE 4
eQTLs overlapping with pQTLs identified from the fibre related network on C08. cis and trans, eQTL defined by B. napus cv. Darmor-bzh reference genome. cis*, cis-eQTL defined by the DH12075 reference genome.

SUPPLEMENTARY TABLE 5
Differential expressed genes within the fibre regulatory network between parental lines DH12075 (DH) and PSA12 (PSA). Up indicated higher gene expression level in DH12075. Corrected p-value: p-value was adjusted with Benjamini-Hochberg's method and controlled at FDR (False Discovery Rate) = 0.05. Genes with more than 2- fold changes and corrected p-value < 0.05 were claimed as significantly differential expressed. Genes highlighted in red have a corrected p-value < 0.05, but a fold change between 2 - 1.5.

SUPPLEMENTARY FIGURE 1
Soft-thresholding powers test to identify the optimized value to fit the approximate scale free topology of the network. The left panel displayed the scale-free topology index (y-axis) with the tested soft-thresholding power (x-axis). The right panel displays the mean connectivity (degree, y-axis) with tested soft-thresholding power (x-axis).

SUPPLEMENTARY FIGURE 2
Distribution of lignin content in the years 2009 **(A-C)** and 2010 **(D-F)**. **(A, D)** acid detergent lignin (ADL); **(B, E)** neutral detergent fiber (NDF); **(C, F)** acid detergent fibre (ADF). Lignin contents of the parental lines PSA12 and DH12075 are highlighted with blue and red dashed lines, respectively. The normality of the distribution was assessed using the Shapiro-Wilk test, which yielded p values of **(A)** p = 0.001, **(B)** p = 0.032, **(C)** p = 0.003, **(D)** p = 0.001, **(E)** p = 0.0099, **(F)** p = 0.001, which all exceed the significance of p < 0.05.

SUPPLEMENTARY FIGURE 3
Synteny analysis of the lignin QTL qLignin_A09_2 and qLignin_C08. Synteny was analyzed and visualized with the Brassica napus Genome Browser (https://www.genoscope.cns.fr/brassicanapus/) with the Darmor-bzh reference genome (Chalhoub et al., 2014).

SUPPLEMENTARY FIGURE 4
Mapped transcripts from fibre network to the major metabolites map. Correlation between transcripts and seed ADF (acid detergent fibre) trait was presented by colors ranging from red (high positive correlation) to blue (high negative correlation).

SUPPLEMENTARY FIGURE 5
Dissection of transcripts on the major secondary metabolite pathways. Correlation between transcripts and seed ADF (acid detergent fibre) trait was presented by colors ranging from red (high positive correlation) to blue (high negative correlation).

SUPPLEMENTARY FIGURE 6
Mapped known genes to the lignin and flavonoid pathway. Correlation between transcripts and seed ADF (acid detergent fibre) were presented by colors ranging from red (high positive correlation) to blue (high negative correlation).

SUPPLEMENTARY FIGURE 7
Identification biotic stress signal pathway coordinated with the changes of lignin. Correlation between transcripts and seed (acid detergent fibre) trait was presented by colors ranging from red (high positive correlation) to blue (high negative correlation).

SUPPLEMENTARY FIGURE 8
Analysis of gene expression of double haploid (DH) lines with cis-eQTL of CCR1. The DH lines were separated into two groups based on the genotype of the eQTL peak marker *BnN9p34992197*: lines with the eQTL (DH12075 marker type) labeled in blue, and lines without the eQTL (PSA12 marker type) labeled with yellow. Student's t-test was used to test the significance of means between groups at the significance level of p < 0.05.

SUPPLEMENTARY FIGURE 9
Pearson correlation analysis of trait-gene expression. The log2 transformed gene expression levels of **(A)**. *cinnamoyl-CoA-reductase (CCR1)*; **(B)**. *TRANSPARENT TESTA 6 (TT6)*; and **(C)**. *TRANSPARENT TESTA 8 (TT8)* are shown in the Y-axis, with acid detergent fibre (ADF) in the X-axis. Parental line DH12075 is indicated in green, PSA12 is indicated in red, and Double haploid (DH) lines are indicated by black dots. R, Pearson Correlation Coefficient; and p, p-value, significance level was claimed at p < 0.05.

References

Andreasson, E., Jenkins, T., Brodersen, P., Thorgrimsen, S., Petersen, N. H., Zhu, S., et al. (2005). The MAP kinase substrate MKS1 is a regulator of plant defense responses. *EMBO J.* 24, 2579–2589.

Aoki, K., Ogata, Y., and Shibata, D. (2007). Approaches for extracting practical information from gene coexpression networks in plant biology. *Plant Cell Physiol.* 48, 381–390. doi: 10.1093/pcp/pcm013

Atias, O., Chor, B., and Chamovitz, D. A. (2009). Large-scale analysis of Arabidopsis transcription reveals a basal co-regulation network. *BMC Syst. Biol.* 3, 86. doi: 10.1186/1752-0509-3-86

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

Barabási, A. L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272

Bassel, G. W., Lan, H., Glaab, E., Gibbs, D. J., Gerjets, T., Krasnogor, N., et al. (2011). Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9709–9714. doi: 10.1073/pnas.1100958108

Basten, C. J., Weir, B. S., and Zeng, Z.-B. (2004). *QTL Cartographer Version 1.17* (Raleigh, NC: North Carolina State University).

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

Bolstad, B. M., Irizarry, R. A., Åstrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high-density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193. doi: 10.1093/bioinformatics/19.2.185

Brem, R. B., and Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1572–1577. doi: 10.1073/pnas.0408709102

Brem, R. B., Yvert, G., Clinton, R., and Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752–755. doi: 10.1126/science.1069516

Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., et al. (2007). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.* 39, 1329–1337. doi: 10.1038/ng.2007.17

Cardinal, A. J., Lee, M., and Moore, K. J. (2003). Genetic mapping and analysis of quantitative trait loci affecting fiber and lignin content in maize. *Theor. Appl. Genet.* 106, 866–874. doi: 10.1007/s00122-002-1136-5

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

Chao, H., Guo, L., Zhao, W., Li, H., and Li, M. (2022). A major yellow-seed QTL on chromosome A09 significantly increases the oil content and reduces the fiber content of seed in *Brassica napus*. *Theoret. Appl. Genet.* 135, 1293–1305. doi: 10.1007/s00122-022-04031-0

Chen, M., Xuan, L., Wang, Z., Zhou, L., Li, Z., Du, X., et al. (2014). TRANSPARENT TESTA8 inhibits seed fatty acid accumulation by targeting several seed development regulators in *Arabidopsis*. *Plant Physiol.* 165, 905–916. doi: 10.1104/pp.114.235507

Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971. doi: 10.1093/genetics/138.3.963

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

Denness, L., McKenna, J. F., Segonzac, C., Wormit, A., Madhou, P., Bennett, M., et al. (2011). Cell wall damage-induced lignin biosynthesis is regulated by a ROS- and jasmonic acid dependent process in *Arabidopsis thaliana*. *Plant Physiol.* 156, 1364–1374. doi: 10.1104/pp.111.175737

Dong, J., Keller, W. A., Yan, W., and Georges, F. (2004). Gene expression at early stages of *Brassica napus* seed development as revealed by transcript profiling of seed-abundant cDNAs. *Planta* 218, 483–491. doi: 10.1007/s00425-003-1124-2

Eulgem, T., and Somssich, I. E. (2007). Networks of WRKY transcription factors in defence signaling. *Curr. Opin. Plant Biol.* 10, 366–371. doi: 10.1016/j.pbi.2007.04.020

Ficklin, S. P., and Feltus, F. A. (2011). Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. *Plant Physiol.* 156, 1244–1256. doi: 10.1104/pp.111.173047

Ficklin, S. P., Luo, F., and Feltus, F. A. (2010). The association of multiple interacting genes with specific phenotypes in rice using gene coexpression networks. *Plant Physiol.* 154, 13–24. doi: 10.1104/pp.110.159459

Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., et al. (2007). Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Nat. Acad. Sci. U.S.A.* 104, 6478–6483. doi: 10.1073/pnas.0611629104

Jansen, R. C., and Nap, J. P. (2001). Genetical genomics: the added value from segregation. *Trends Genet.* 17, 388–391. doi: 10.1016/S0168-9525(01)02310-1

Jones, L., Ennos, A. R., and Turner, S. R. (2001). Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of Arabidopsis. *Plant J.* 26, 205–216. doi: 10.1046/j.1365-313x.2001.01021.x

Keurentjes, J. J., Fu, J., Terpstra, I. R., Garcia, J. M., Van Den Ackerveken, G., Snoek, L. B., et al. (2007). Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1708–1713. doi: 10.1073/pnas.0610429104

Kliebenstein, D. J. (2009). Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. *Annu. Rev. Plant Biol.* 60, 93–114. doi: 10.1146/annurev.arplant.043008.092114

Kliebenstein, D. J. (2020). Using networks to identify and interpret natural variation. *Curr. Opin. Plant Biol.* 54, 122–126. doi: 10.1016/j.pbi.2020.04.005

Lai, Z., Li, Y., Wang, F., Cheng, Y., Fan, B., Yu, J. Q., et al. (2011). Arabidopsis sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. *Plant Cell* 23, 3824–3841. doi: 10.1105/tpc.111.090571

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Lee, H. S., Wang, J., Tian, L., Jiang, H., Black, M. A., Madlung, A., et al. (2004). Sensitivity of 70-mer oligonucleotides and cDNAs for microarray analysis of gene expression in Arabidopsis and its related species. *Plant Biotechnol. J.* 2, 45–57. doi: 10.1046/j.1467-7652.2003.00048.x

Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N., et al. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* 57, 405–430. doi: 10.1146/annurev.arplant.57.032905.105252

Li, L., Tian, Z., Chen, J., Tan, Z., Zhang, Y., Zhao, H., et al. (2023). Characterization of novel loci controlling seed oil content in *Brassica napus* by marker metabolite-based multi-omics analysis. *Genome Biol.* 24, 141. doi: 10.1186/s13059-023-02984-z

Liu, L., Qu, C., Wittkop, B., Yi, B., Xiao, Y., He, Y., et al. (2013). A high-density SNP map for accurate mapping of seed fibre QTL in *Brassica napus* L. *PloS One* 8, e83052. doi: 10.1371/journal.pone.0083052

Liu, L., Stein, A., Wittkop, B., Sarvari, P., Li, J., Yan, X., et al. (2012). A knockout mutation in the lignin biosynthesis gene CCR1 explains a major QTL for acid detergent lignin content in *Brassica napus* seeds. *Theor. Appl. Genet.* 124, 1573–1586. doi: 10.1007/s00122-012-1811-0

Lorieux, M. (2012). MapDisto: fast and efficient computation of genetic linkage maps. *Mol. Breed.* 30, 1231–1235. doi: 10.1007/s11032-012-9706-y

Mackay, T. F. C., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10, 565–577. doi: 10.1038/nrg2612

Mao, L., VanHemert, J. L., Dash, S., and Dickerson, J. A. (2009). Arabidopsis gene coexpression network and its functional modules. *BMC Bioinf.* 10, 346. doi: 10.1186/1471-2105-10-346

Marshall-Colón, A., and Kliebenstein, D. J. (2019). Plant networks as traits and hypotheses: moving beyond description. *Trends Plant Sci.* 24, 840–852. doi: 10.1016/j.tplants.2019.06.003

Mayerhofer, R., Wilde, K., Mayerhofer, M., Lydiate, D., Bansal, V. K., Good, A. G., et al. (2005). Complexities of chromosome landing in a highly duplicated genome: toward map-based cloning of a gene controlling blackleg resistance in *Brassica napus*. *Genetics* 171, 1977–1988. doi: 10.1534/genetics.105.049098

Miao, L., Chao, H., Chen, L., Wang, H., Zhao, W., Li, B., et al. (2019). Stable and novel QTL identification and new insights into the genetic networks affecting seed fiber traits in *Brassica napus*. *Theor. Appl. Genet.* 132, 1761–1775. doi: 10.1007/s00122-019-03313-4

Mutwil, M., Usadel, B., Schütte, M., Loraine, A., Ebenhöh, O., and Persson, S. (2010). Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. *Plant Physiol.* 152, 29–43. doi: 10.1104/pp.109.145318

O'Hara, P., Slabas, A. R., and Fawcett, T. (2002). Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. *Plant Physiol.* 129, 310–320. doi: 10.1104/pp.010956

Oñate-Sánchez, L., and Vicente-Carbajosa, J. (2008). DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. *BMC Res. Notes* 1, 93. doi: 10.1186/1756-0500-1-93

Ouellette, L. A., Reid, R. W., Blanchard, S. G., and Brouwer, C. R. (2018). LinkageMapView-rendering high-resolution linkage and QTL maps. *Bioinformatics* 34, 306–307. doi: 10.1093/bioinformatics/btx576

Pandey, S. P., and Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiol.* 150, 1648–1655. doi: 10.1104/pp.109.138990

Parkin, I. A. P., Clarke, W. E., Sidebottom, C., Zhang, W., Robinson, S. J., Links, M. G., et al. (2010). Towards unambiguous transcript mapping in the allotetraploid *Brassica napus*. *Genome* 53, 929–938. doi: 10.1139/G10-053

Persson, S., Wei, H., Milne, J., Page, G. P., and Somerville, C. R. (2005). Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc. Natl. Acad. Sci. U.S.A.* 102, 8633–8638. doi: 10.1073/pnas.0503392102

Potokina, E., Druka, A., Luo, Z., Wise, R., Waugh, R., and Kearsey, M. J. (2007). Gene expression quantitative trait locus analysis of 16 000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. *Plant J.* 53, 90–101. doi: 10.1111/j.1365-313X.2007.03315.x

Raney, J. P., Love, H. K., Rakow, G., and Downey, R. K. (1987). An apparatus for rapid preparation of oil and oil-free meal from Brassica seed. *Lipid Fett.* 89, 235–237. doi: 10.1002/lipi.19870890608

Raney, J. P., and Serblowski, G. (2007). "Inserts for FOSS NIRS 6500 spinning ring cups," in *Proceedings of the 12th International Rapeseed Congress, Wuhan, China*, 26–30 Mar. 2007, Vol. 5. 39–42 (Monmouth Junction, NJ: Science Press USA Inc.).

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Baraba´si, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555. doi: 10.1126/science.1073374

Rockman, M. V., and Kruglyak, L. (2006). Genetics of global gene expression. *Nat. Rev. Genet.* 7, 862–872. doi: 10.1038/nrg1964

Ruuska, S. A., Girke, T., Benning, C., and Ohlrogge, J. B. (2002). Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell* 14, 1191–1206. doi: 10.1105/tpc.000877

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., et al. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297–302. doi: 10.1038/nature01434

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Simmons, C. R., Lafitte, H. R., Reimann, K. S., Brugie`re, N., Roesler, K., Albertsen, M. C., et al. (2021). Successes and insights of an industry biotech program to enhance maize agronomic traits. *Plant Sci.* 307, 110899. doi: 10.1016/j.plantsci.2021.110899

Smyth, G. K. (2005). "Limma: Linear models for microarray data," in *Bioinformatics and computational biology solutions Using R and Bioconductor*. Eds. R. Gentlemen, V. Carey, S. Dudoit, R. Irizarry and W. Huber (Springer, New York), 397–420.

Snowdon, R. J., Wittkop, B., Rezaidad, A., Hasan, M., Lipsa, F. D., Stein, A., et al. (2010). Regional association analysis delineates a sequenced chromosome region influencing antinutritive seed meal compounds in oilseed rape. *Genome* 53, 917–928. doi: 10.1139/G10-052

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

Usadel, B., Nagel, A., Thimm, O., Redestig, H., Blaesing, O. E., Palacios-Rojas, N., et al. (2005). Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. *Plant Physiol.* 138, 1195–1204. doi: 10.1104/pp.105.060459

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

Vanholme, R., Storme, V., Vanholme, B., Sundin, L., Christensen, J. H., Goeminne, G., et al. (2012). A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis. Plant Cell* 24, 3506–3529. doi: 10.1105/tpc.112.102574

Van Soest, P. J., Robertson, J. B., and Lewis, B. A. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *J. Dairy. Sci.* 74, 3583–3597. doi: 10.3168/jds.S0022-0302(91)78551-2

Vogt, T. (2010). Phenylpropanoid biosynthesis. *Mol. Plant* 3, 2–20. doi: 10.1093/mp/ssp106

Wang, W., Basten, C. J., and Zeng, Z. B. (2007). *Windows QTL Cartographer 2.5* (Raleigh, NC: Department of Statistics, NCSU).

West, M. A., Kim, K., Kliebenstein, D. J., Van Leeuwen, H., Michelmore, R. W., Doerge, R. W., et al. (2007). Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. *Genetics* 175, 1441–1450. doi: 10.1534/genetics.106.064972

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

Wittkop, B., Snowdon, R. J., and Friedt, W. (2012). New NIRS calibrations for fibre fractions reveal broad genetic in *Brassica napus* seed quality. *J. Agric. Food Chem.* 60, 224–2256. doi: 10.1021/jf204936f

Xu, W., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* 20, 176–185. doi: 10.1016/j.tplants.2014.12.001

Xu, W., Grain, D., Bobet, S., Le Gourrierec, J., Thevenin, J., Kelemen, Z., et al. (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.* 202, 132–144. doi: 10.1111/nph.12620

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., et al. (2002). Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* 30, e15. doi: 10.1093/nar/30.4.e15

Yu, B., Boyle, K., Zhang, W., Robinson, S. J., Higgins, E., Ehman, L., et al. (2016). Multi-trait and multi-environment QTL analysis reveals the impact of seed colour on seed composition traits in *Brassica napus. Mol. Breed.* 36, 111. doi: 10.1007/s11032-016-0521-8

Yusuf, A. O., and Möllers, C. (2024). Inheritance of cellulose, hemicellulose and lignin content in relation to seed oil and protein content in oilseed rape. *Euphytica* 220, 5. doi: 10.1007/s10681-023-03264-4

Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1468. doi: 10.1093/genetics/136.4.1457

Zhai, Y., Yu, K., Cai, S., Hu, L., Amoo, O., Xu, L., et al. (2020). Targeted mutagenesis of bnTT8 homologs controls yellow seed coat development for effective oil production in *Brassica napus* L. *Plant Biotechnol. J.* 18, 1153–1168. doi: 10.1111/pbi.13281

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4, 17. doi: 10.2202/1544-6115.1128

Zhang, Y., Zhang, H., Zhao, H., Xia, Y., Zheng, X., Fan, R., et al. (2022). Multi-omics analysis dissects the genetic architecture of seed coat content in *Brassica napus. Genome Biol.* 23, 86. doi: 10.1186/s13059-022-02647-5

Zheng, Z., Qamar, S. A., Chen, Z., and Mengiste, T. (2006). Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *Plant J.* 48, 592–605. doi: 10.1111/j.1365-313X.2006.02901.x