

## PLANT SCIENCES

A rare dominant allele *DYSOC1* determines seed coat color and improves seed oil content in *Brassica napus*Huaixin Li<sup>1†</sup>, Mingli Wu<sup>1†</sup>, Hongbo Chao<sup>2</sup>, Yongtai Yin<sup>1,3</sup>, Yutian Xia<sup>1</sup>, Xin Cheng<sup>1</sup>, Kang Chen<sup>1</sup>, Shuxiang Yan<sup>1</sup>, Xiaodong Wang<sup>4</sup>, Yiyi Xiong<sup>1</sup>, Jianjie He<sup>1</sup>, Shipeng Fan<sup>1</sup>, Yiran Ding<sup>1</sup>, Libin Zhang<sup>1,5</sup>, Haibo Jia<sup>1</sup>, Chunyu Zhang<sup>6\*</sup>, Maoteng Li<sup>1\*</sup>

Yellow seed coat color (SCC) is a valuable trait in *Brassica napus*, which is significantly correlated to high seed oil content (SOC) and low seed lignocellulose content (SLC). However, no dominant yellow SCC genes were identified in *B. napus*. In this study, a dominant yellow SCC *B. napus* N53-2 was verified, and then 58,981 eQTLs and 25 trans-eQTL hotspots were identified in a double haploid population derived from N53-2 and black SCC material Ken-C8. A rare dominant allele *DYSOC1* (dominant gene of yellow seed coat color and improved seed oil content 1) was subsequently cloned in a trans-eQTL hotspot that colocalized with SCC, SOC, and SLC QTL hotspot on ChrA09 through QTL fine mapping and multi-omics analysis. Transgenic experiments revealed that the expression of *DYSOC1* produced yellow SCC seeds with significantly increased SOC and decreased SLC. Our result provides a rare dominant yellow SCC allele in *B. napus*, which has excellent potential for yellow SCC and high SOC rapeseed breeding.

## INTRODUCTION

Rapeseed (*Brassica napus*, AACCC,  $2n = 38$ ) is one of the most important oilseed crops worldwide, which provides nutritious vegetable oils and proteins for humans and livestock (1, 2). *B. napus* is an allotetraploid that originates from the natural hybridization of *Brassica rapa* (AA,  $2n = 20$ ) and *Brassica oleracea* (CC,  $2n = 18$ ) (1). In previous studies, it has been concluded that the yellow seed coat color (SCC) seeds always exhibit more excellent properties than black SCC seeds, including a higher seed oil content (SOC), lower seed lignocellulose content (SLC), lower hull content (HC), and lower proanthocyanidin (PA) content (3–6). Therefore, yellow seeds produce more oil with less pigment and seed meal with less anti-nutrients, which are appreciated by both companies and customers (7–9). The breeding of yellow SCC *B. napus* varieties has been energetically conducted (10–12), for example, as a popular cultivar, the yellow SCC variety Yu-yellow No. 1 is cultivated on more than 10 million acres in central China (7, 13). However, *B. napus* varieties with yellow SCC remain relatively rare, and their breeding is still difficult due to the lack of dominant gene resources.

In *Arabidopsis*, yellow SCC results from reduced PA content in the seed coat. In brief, phenylalanine is transformed into *p*-coumaroyl-coenzyme A through the phenylpropanoid biosynthesis pathway, which is further processed for PA synthesis through the flavonoid biosynthesis pathway (14). In addition, lignin monomer synthesis nearly exclusively relies on the phenylpropanoid biosynthesis pathway (15). Many MYB transcription factors (TFs), along

with bHLH-MYB-WD40 (MBW) complexes, zinc finger proteins, and WRKY TFs, have been identified as the primary regulators during this process (16–18).

The genetic mechanism of yellow SCC is still unclear in *B. napus*, although it has been studied for decades. Most studies have indicated that the SCC in *B. napus* is controlled by at least three pairs of genes, and seeds appear yellow when all genes are homozygous recessive (19–21). In addition, an incompletely dominant locus that controls yellow SCC has been found in the synthetic line No. 2127-17 (22, 23). It is now widely accepted that yellow SCC in *B. napus* is a complex quantitative trait controlled by multiple loci (13, 24), and the quantitative trait locus (QTL) mapping for SCC, SOC, and SLC has been widely conducted (13, 25–27). For instance, Yan *et al.* (3) discovered 11, 4, and 12 QTLs for SOC, SCC, and HC, respectively, and one QTL for SOC was tightly linked with the QTLs for HC and SCC on the N8 linkage group (LG). Liu *et al.* (24) found a major SCC QTL on ChrA09 in an F<sub>2</sub> population, which is colocalized with major QTLs for SOC and SLC (5). The functions of key regulator genes (such as *TTG1*, *TT1*, *TT2*, and *TT8*), structural genes (such as *TT7*, *TT18*, and *TT10*), and transporter genes (*TT12*) of PA biosynthesis have been verified in *B. napus*, the silencing or mutation of which in black SCC materials usually leads to yellow SCC, higher SOC, and lower SLC (4, 28–30). However, only a few yellow SCC genes have been identified thus far (7, 31), all of which are recessive, and this recessive nature is inconvenient for yellow SCC rapeseed breeding.

Expression QTL (eQTL) mapping has been extensively used to elucidate the genetic architecture underlying gene expression variation, providing a robust framework for uncovering gene regulatory networks associated with complex quantitative traits (32). Integration analysis for QTL and eQTL has become increasingly prominent in the identification of causal genes linked to essential traits, as demonstrated in rice (33), soybean (34), wheat (35), and rapeseed (36). QTL fine mapping and map-based cloning are the most widely used strategies for identifying candidate genes within QTLs (5, 37–39). For example, the functional gene *GmFATA1B* was cloned through fine mapping from a SOC locus in soybean (40). In barley, the amino acid transporter gene *HvBlp*, which controls grain color, was isolated via QTL fine mapping and comparative genomic analysis (41).

<sup>1</sup>College of Life Science and Technology, Key Laboratory of Molecular Biophysics of the Ministry of Education, Huazhong University of Science and Technology, Wuhan 430074, China. <sup>2</sup>School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, China. <sup>3</sup>Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore 117604, Singapore. <sup>4</sup>Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Key Laboratory of Cotton and Rapeseed, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China. <sup>5</sup>Hubei Bioinformatics and Molecular Imaging Key Laboratory, Wuhan 430074, China. <sup>6</sup>National Key Lab of Crop Genetic Improvement, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China.

\*Corresponding author. Email: limaoteng426@hust.edu.cn (M.L.); zhchym@mail.hzau.edu.cn (C.Z.)

†These authors contributed equally to this work.

In our previous studies, QTL mapping for SCC, SOC, and SLC was conducted in the KN double haploid (DH) population derived from parental line N53-2 (which contained yellow SCC loci) and Ken-C8 (which varied considerably in SCC, SOC, and SLC compared with N53-2), and some major QTLs were obtained (5, 42, 43). In this study, eQTL analysis was conducted to further characterize the genetic mechanism of the SCC, SOC, and SLC variation in the KN DH population. Diversiform variation of eQTL distribution was observed, and a trans-eQTL hotspot overlapping with the major dominant pleiotropy QTL hotspot of SCC, SOC, and SLC was detected on ChrA09. A dominant allele, *DYSOC1* (dominant gene of yellow seed coat color and improved seed oil content 1), was subsequently cloned from this trans-eQTL hotspot through QTL fine mapping and map-based cloning, which was verified as a splendid rare locus for yellow SCC, high SOC, and low SLC *B. napus* breeding.

## RESULTS

### Genetic linkage map construction and eQTL identification

The yellow SCC *B. napus* N53-2 was crossed with three black SCC materials [Ken-C8, J2016, and Zhongshuang 11 (ZS11)]. All F<sub>1</sub> lines produced yellow SCC seeds (fig. S1), indicating a dominant yellow SCC phenotype in N53-2. A DH population was constructed using Ken-C8 and N53-2 as the parents for QTL mapping, and QTLs for SCC and other related traits were identified (5, 42, 43). In the present study, RNA sequencing (RNA-seq) was conducted on the selfed seeds of DH lines and the parental lines Ken-C8 and N53-2 at 28 days after flowering (DAF), which provided both expression levels of transcribed genes and a comprehensive set of single-nucleotide polymorphisms (SNPs) at the population level. In total, 1299.68-Gb clean data were obtained after eliminating the adaptor sequences and low-quality reads (table S1). After quality trimming, the reads were aligned to the *B. napus* Darmor-*bzh* reference genome, and 756,689 SNPs were detected consequently. The SNP data were further processed for genotyping to identify markers for genetic linkage map construction.

After high-quality SNP selection, bin division, and genetic linkage map construction, 3925 SNP markers were successfully assigned to 19 LGs, and a genetic linkage map with a total genetic length of 2933.03 cM was constructed (fig. S2). On average, the distance between the adjacent markers was 0.75 cM, and the map integrity was 99.57%, which ensured the accuracy of genotyping (table S2). Haplotype mapping and collinearity analysis further confirmed the high quality of the newly constructed linkage map (figs. S3 and S4). The fragments per kilobase of exon model per million mapped fragments (FPKM) of each gene was determined as a trait and processed to QTL mapping. In total, 58,981 eQTLs associated with the expression of 31,785 genes (called eGenes) were identified, with an average explained variance ( $r^2$ ) of 9.95% (Fig. 1 and table S3), including 5811 eQTLs that exhibited an  $r^2$  exceeding 20%. The median genetic distance and physical distance of all eQTLs were 3.13 cM and 697.85 kb, respectively (fig. S5, A and B).

According to the relative distance with corresponding eGenes, 5620 cis-eQTLs and 53,361 trans-eQTLs for 5392 and 29,026 eGenes were identified, respectively (Fig. 1 and fig. S5C). Among the eGenes, 2759, 26,393, and 2633 were regulated by only cis-eQTLs, only trans-eQTLs, and both of them, respectively (fig. S5D). Approximately half of the eGenes were affected by more than one eQTL, indicating a complex regulatory mechanism underlying expression variation in *B. napus* (fig. S6A).

### Characteristics of eQTLs distribution in the genome scale

The distribution of eQTLs varied considerably across the two sub-genomes and 19 chromosomes. Specifically, 68.77 and 53.98% of the cis- and trans-eQTLs, respectively, were located in the A sub-genome, in which ChrA06 and ChrA09 consisted of the most cis-eQTLs (608) and trans-eQTLs (4935), respectively (fig. S6B). Strong diagonal enrichment was observed on the basis of the relative positions of eGenes and their corresponding eQTLs throughout the genome. In addition, both the LOD score and  $r^2$  of cis-eQTLs were higher than those of trans-eQTLs (Fig. 2, A and B), indicating that cis-eQTLs tend to be more significant and explain more expression variation than the trans-eQTLs.

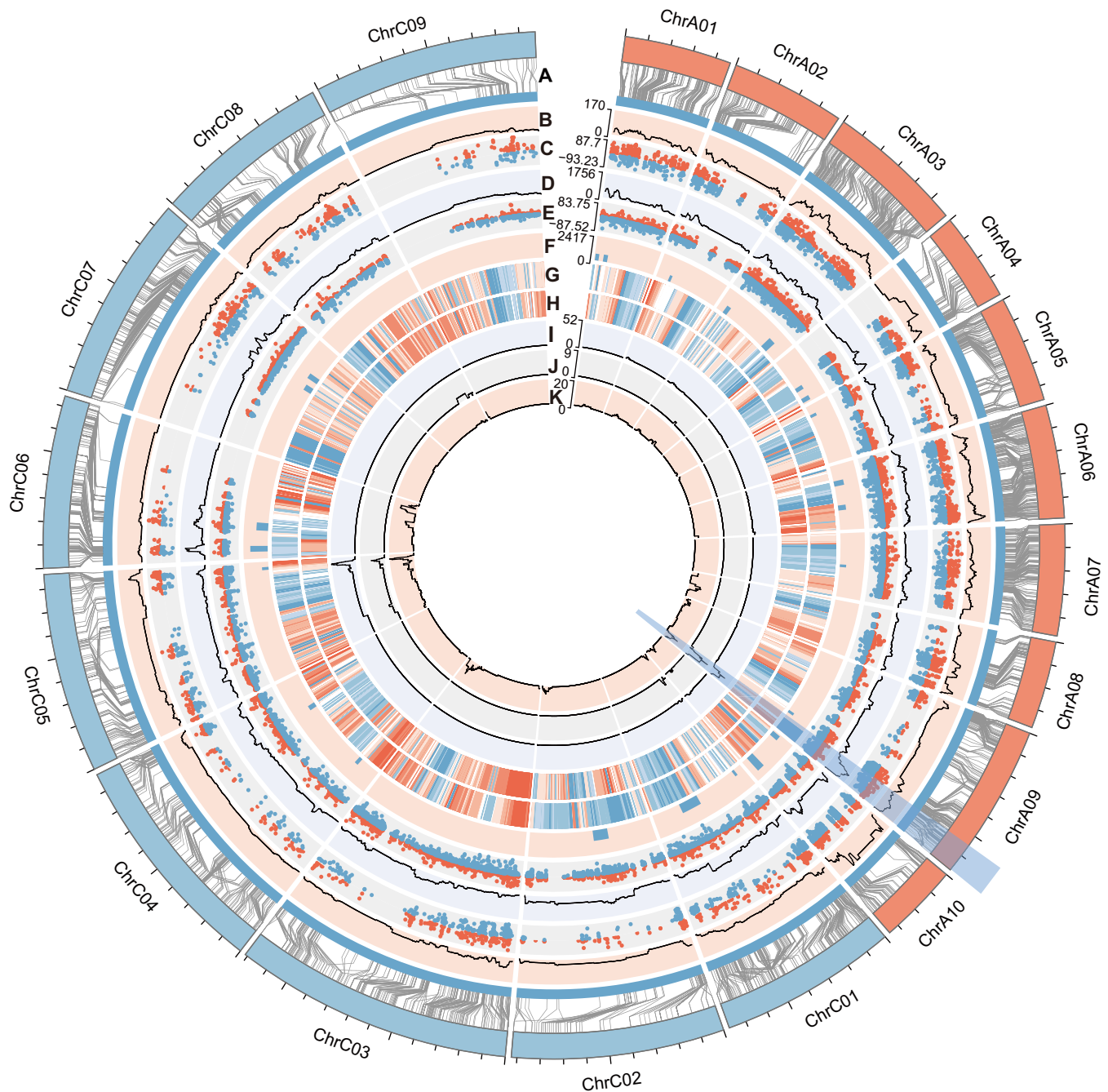
In our previous studies, genome re-sequencing and Hi-C were conducted on the two parents, Ken-C8 and N53-2 (42, 44). In the present study, 96.81% of the eGenes regulated by cis-eQTLs had sequence variations in the promoter (2 kb before ATG) and/or the coding regions between the two parents. A significant correlation between cis-eQTL and sequence variation was observed (Fig. 2C), indicating that cis-eQTLs were mainly caused by sequence variations within the corresponding eGenes and their promoter regions. The distributions of eGenes in the A and B compartments were also investigated in Ken-C8 and N53-2, revealing significant correlations between eQTLs and both the common and divergent compartments of the parents. Specifically, the eGene frequency of common A compartments (41.51%) was higher than that of common B compartments (34.45%) of the parents. In addition, using N53-2 as the benchmark, the eGene frequency of divergent A compartments (33.58%) was lower than that of divergent B compartments (35.70%) (Fig. 2D).

The correlation between three-dimensional (3D) genome interactions and eQTLs was also evaluated. It was revealed that 6.11 and 13.32% eQTLs were aligned with the common and unique genome interactions of the parents, respectively, while only 2.12 and 6.29% of the randomly created eQTLs (RC-eQTLs) were aligned with the common and unique genome interactions of the parents, respectively, which were significantly lower than those of eQTLs (fig. S5, E and F). Therefore, a significant correlation between eQTLs and 3D genome interactions was detected in the present study.

### A trans-eQTL hotspot that colocalized with the QTL hotspot of SOC, SCC, and SLC was identified in ChrA09

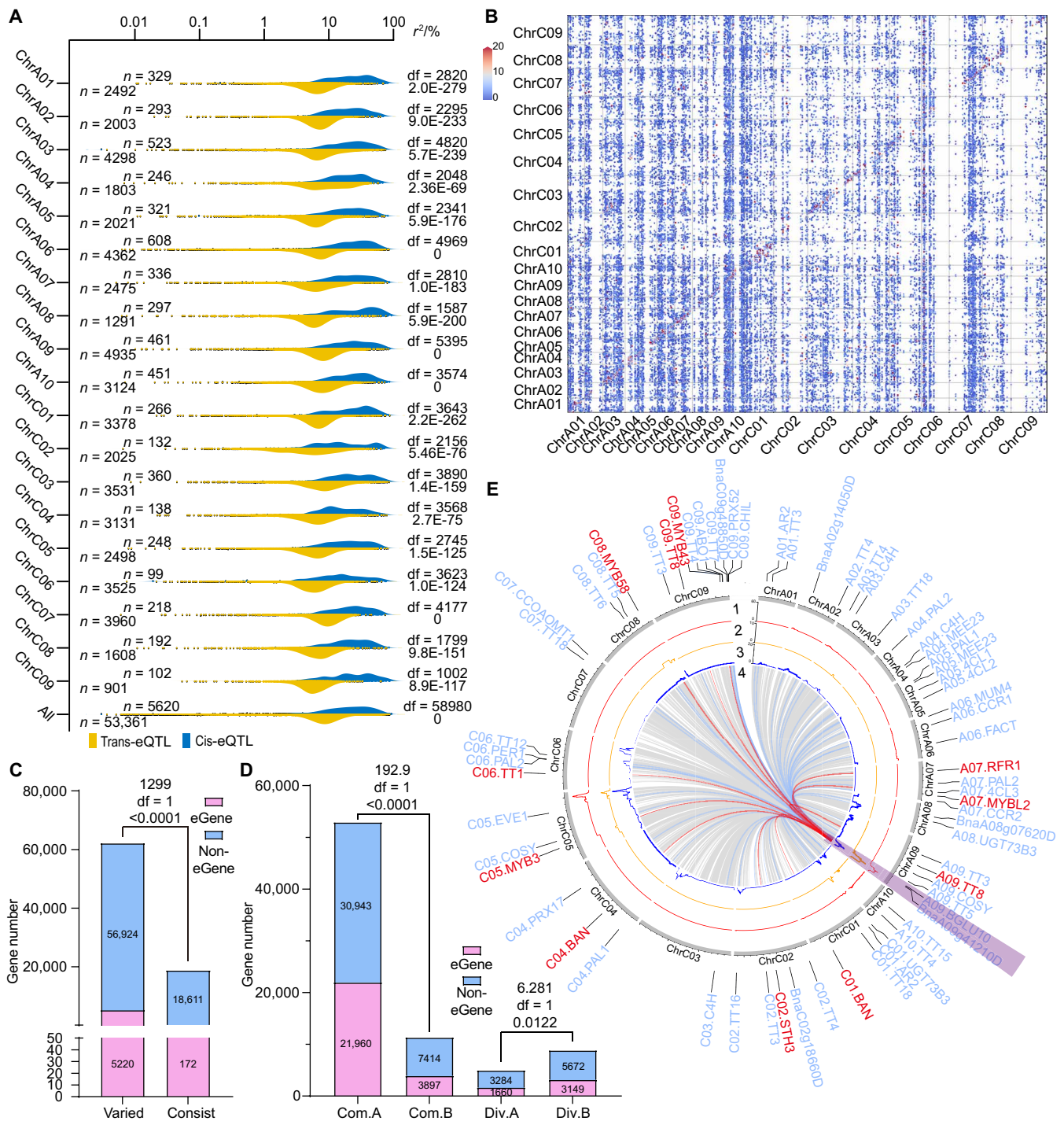
A total of 25 trans-eQTL hotspots were identified, each comprising 620 to 2420 trans-eQTLs (Fig. 1 and table S4). Among them, Hotspot 13, located at the end of ChrA09, had the most congregating trans-eQTLs (tables S4 and S5). Further analysis showed that the eGenes regulated by trans-eQTL hotspots were significantly enriched in different Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology terms (tables S6 to S30).

Hotspot 13 was colocalized with the QTL hotspots of SOC, SCC, and SLC [including neutral detergent fiber content (NDF), acid detergent fiber content (ADF), and acid detergent lignin content (ADL)] (Fig. 2E). In addition, the downstream eGenes of Hotspot 13 were significantly enriched in flavonoid and phenylpropanoid biosynthesis pathways (table S18 and fig. S7A). The trans-eQTLs of a set of enzyme genes of these pathways, such as *C4Hs* (*BnaA03g14010D*, *BnaA04g17570D*, and *BnaC03g16950D*), *PALs* (*BnaA05g07370D*, *BnaC04g08190D*, *BnaA04g04830D*, *BnaA07g16060D*, and *BnaC06g14510D*), and *4CLs* (*BnaA07g25210D*, *BnaA05g19760D*, and *BnaA05g15310D*), as well as regulator genes, such as *TT1* (*BnaC06g08390D*) and *TT8s* (*BnaA09g22810D* and *BnaC09g24870D*), were identified in Hotspot



**Fig. 1. The genetic LG construction and eQTL mapping.** (A) The collinearity analysis between genetic distance (inner circle) and physical distance (outer circle; the scale represents 5 Mb) of the genetic linkage map. (B) The density of *cis*-eQTLs. (C) The  $r^2$  of each *cis*-eQTL (red and blue points represent the *cis*-eQTLs with positive and negative additive effects, respectively). (D) The density of *trans*-eQTLs. (E) The  $r^2$  of each *trans*-eQTL (red and blue points represent the *trans*-eQTLs with positive and negative additive effects, respectively). (F) The distribution of the 25 *trans*-eQTL hotspots (Hotspot 13 was highlighted). (G and H) The distribution of A and B compartments in Ken-C8 (G) and N53-2 (H) genome (the heatmap was drawn by the compartment value, and the red and blue colors represent A and B compartments, respectively) (44). (I to K) The density of QTLs for SLC (I), SCC (J), and SOC (K) (5, 42, 43). Among all circles, the QTL and eQTL densities were calculated with a sliding window of 1 Mb and a step size of 100 kb.





**Fig. 2. The distribution of eQTLs.** (A) The  $r^2$  distribution of cis-eQTLs and trans-eQTLs across the 19 chromosomes and the entire genome. Each point represents one eQTL, and the degree of freedoms (df) (above) and P values (below) are annotated on the right [two-tailed one-way analysis of variance (ANOVA)]. In addition, the number of cis-eQTL (above) and trans-eQTL (below) are annotated on the left. (B) The positions of eQTLs (x axis) and their corresponding eGenes (y axis) on 19 chromosomes displayed by dot plot. The dot color represents the LOD score of each eQTL. (C) The distributions of cis-eQTL eGenes (and non-eGenes) among the genes with sequence variations ("varied") and without sequence variation ("consist") between Ken-C8 and N53-2. The chi-square, df, and P value are annotated above the histogram (chi-square test). (D) The distributions of eGenes (and non-eGenes) among the common ("Com.") and divergent ("Div.") A/B compartments of Ken-C8 and N53-2. The chi-square, df, and P value are annotated above the histogram (chi-square test). (E) The trans-eQTLs within Hotspot 13 (highlighted in purple) and the positions of their corresponding eGenes. The meaning of each circle is as follows: (1 to 3): the QTL density of SLC (1), SCC (2), and SOC (3) (5, 42, 43); (4) the trans-eQTLs located in Hotspot 13. The blue lines and labels outside represent the structural eGenes of phenylpropanoid and flavonoid biosynthesis pathways, and the red lines and labels represent the regulator genes of those pathways.



13 (Fig. 2E). Thus, there should be a pleiotropic gene that represses both the flavonoid and phenylpropanoid biosynthesis pathways and acts as an upstream regulator of SOC, SLC, and SCC in the meantime.

In addition, Hotspot 6 consisted of 779 trans-eQTLs for 772 eGenes that were significantly enriched in flavonoid biosynthesis (table S11 and fig. S7B). The eQTLs of several enzyme genes for flavonoid biosynthesis, such as *TT4* (*BnaC03g06120D*), *CHIs* (*BnaC07g45760D* and *BnaA10g25120D*), and *CAHs* (*BnaC04g41130D* and *BnaC03g16950D*), were identified within this hotspot (fig. S8A). Two *WD40* genes (*BnaA04g15460D* and *BnaA04g17210D*) were also identified in Hotspot 6 as candidate genes according to previous studies (16), and a cis-eQTL of *BnaA04g17210D* was detected (fig. S8C). Hotspot 25 consisted of 1119 trans-eQTLs for 1069 eGenes that were significantly enriched in fatty acid degradation, fatty acid biosynthesis, fatty acid beta-oxidation, and fatty acid catabolic process (table S30 and fig. S7C). Several enzyme genes involved in fatty acid metabolism, including *ECI3* (*BnaC08g09570D*), *ACX3* (*BnaC05g04350D*), *PED1* (*BnaC04g43560D*), and *LACSs* (*BnaC02g28920D* and *BnaA03g29320D*), were identified (fig. S8B). In addition, two *MYB* genes [including *BnaC07g25380D* (*MYB30*) and *BnaC07g27330D* (*MYB78*)] were identified in Hotspot 25 as candidate genes according to previous studies (45).

### A pleiotropy allele within Hotspot 13 was identified by the integration of eQTL and QTL fine mapping

Because Hotspot 13 was identified as the only trans-eQTL hotspot that colocalized with the QTL hotspot of SCC, SOC, and SLC in the KN DH population, the candidate gene within this region needed to be identified. A BC<sub>5</sub>F<sub>2</sub> near-isogenic line (NIL) population for Hotspot 13 containing 2276 individuals was constructed and QTL fine mapping for SOC, SCC, and SLC was conducted at one time in Wuhan (WH; semi-winter rape-producing area). The QTLs of all those traits were mapped to one highly coincident region that covered a 70-kb interval in the BC<sub>5</sub>F<sub>2</sub> population, and we named this region *uqA9-9*. It explained 52.02, 37.41, 61.38, 82.88, and 67.18% of the PV of SOC, SCC, NDF, ADF, and ADL, respectively (Fig. 3, A to C and E, and table S31). The BC<sub>5</sub>F<sub>2</sub> population was further categorized into three groups: the homozygotes that inherited the haplotype of *uqA9-9* from N53-2 (NIL<sup>N53-2</sup>) or Ken-C8 (NIL<sup>Ken-C8</sup>) and the heterozygotes (NIL<sup>H</sup>). The SOC of NIL<sup>N53-2</sup> was significantly higher than that of NIL<sup>Ken-C8</sup>, while the NDF, ADF, and ADL of NIL<sup>N53-2</sup> were significantly lower than those of NIL<sup>Ken-C8</sup> (Fig. 3D). The phenotype of NIL<sup>H</sup> was similar to that of NIL<sup>N53-2</sup>, while the phenotype of NIL<sup>Ken-C8</sup> was similar to that of Ken-C8 (Fig. 3D), indicating that the constructed BC<sub>5</sub>F<sub>2</sub> population had eliminated almost all background interference, and *uqA9-9* contained a dominant allele that could synchronously improve SCC, SOC, and SLC. The same phenomenon was also observed in the winter and spring rape-producing areas, indicating that *uqA9-9* is an environmentally stable locus (fig. S9).

Comparative genomic analysis showed that 17 annotated genes were located in *uqA9-9*, of which 13 were expressed in the NIL seeds. Of the 17 genes, 9 genes had eQTLs, and 3 genes (*BnaA09g44670D*, *BnaA09g44710D*, and *BnaA09g44720D*) had cis-eQTLs (Fig. 3, F and G, and table S32). Meanwhile, the weighted gene co-expression network analysis (WGCNA) was conducted on the basis of the gene expression profiles of all DH lines and the corresponding trait data. A total of 391 and 183 genes that were highly significantly ( $P < 0.01$ ) correlated to SOC and SCC over all inspected microenvironments were detected, with 137 overlapping genes (tables S33 to S35). Among

these overlapping genes, *BnaA09g44670D* was the only gene located in *uqA9-9*, which had no homologous genes in *Arabidopsis thaliana* (table S32). Sanger sequencing confirmed that an 8-base pair (bp) Indel existed in the second exon of *BnaA09g44670D* in N53-2, which caused a frameshift mutation (Fig. 3, H to J). Conserved domain prediction indicated that the frameshifted *BnaA09g44670D* in N53-2 contained a Skp1 domain and was identified as a Skp1-like protein, which is described as the SKP1 components of the Skp1-Cul1-F-box protein (SCF) ubiquitin ligase complex (46–48). However, no significant domain was observed in the *BnaA09g44670D* protein of Ken-C8 (fig. S10, A and C). Together, we speculated that the N53-2 allele of *BnaA09g44670D* was the candidate gene within Hotspot 13 and named it *DYSOC1* according to its function.

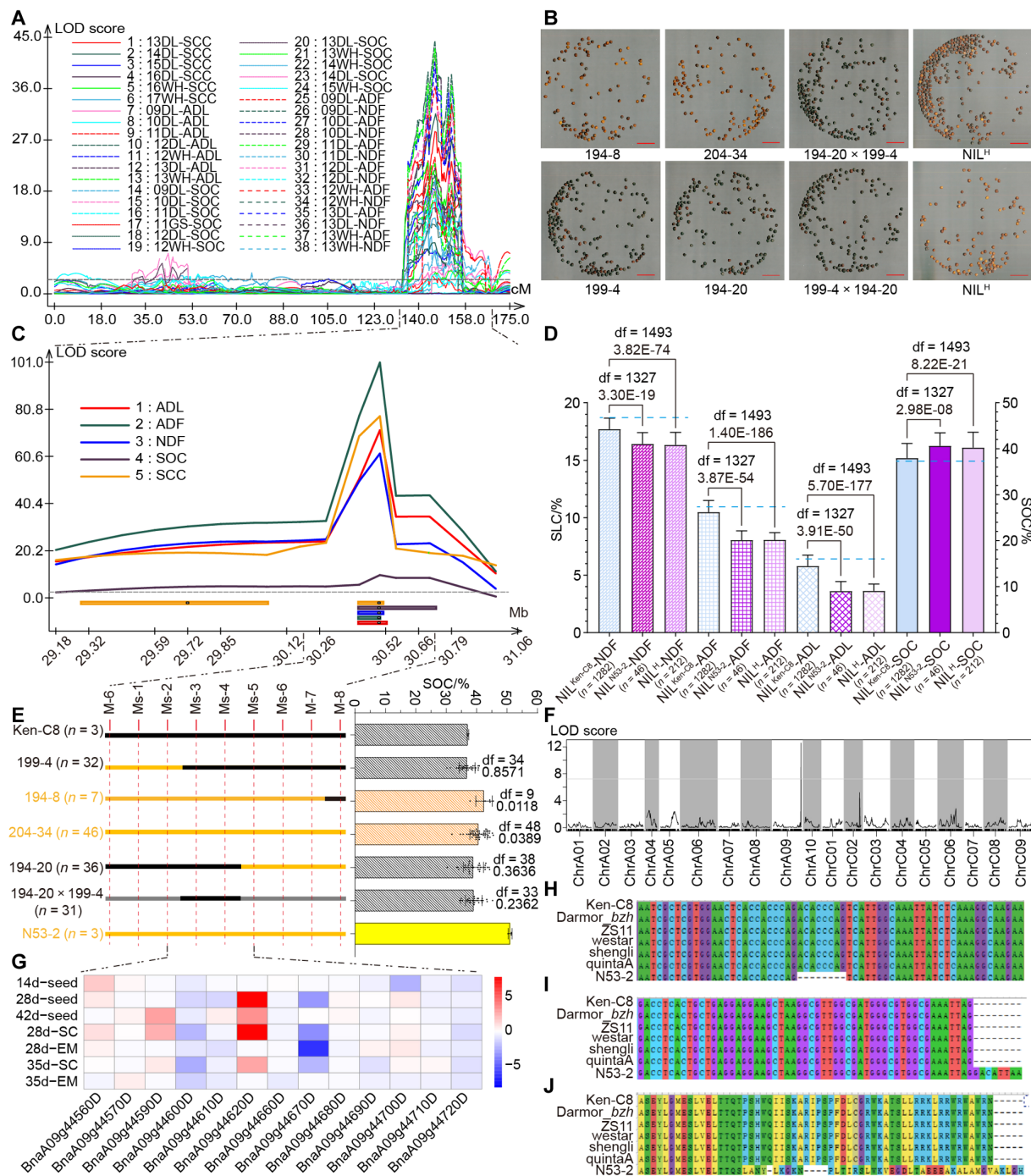
### Transgenic experiments confirmed that *DYSOC1* was a universal locus for SOC, SCC, and SLC improvement in *B. napus*

The full-length genomic sequence of *DYSOC1* (including its promoter) was amplified from N53-2 to construct the complementary vector, which was subsequently transformed into Ken-C8 through *Agrobacterium*-mediated hypocotyl transformation. Compared with wild-type Ken-C8 (Ken-C8<sup>WT</sup>), transgenic lines with *DYSOC1* (Ken-C8<sup>DYSOC1</sup>) represented yellow SCC (Fig. 4, A and B). Further phenotypic investigation revealed that the expression of *DYSOC1* in Ken-C8 led to an 8.80% increase in oil content in the embryo (EMOC), a 28.56% increase in oil content in the seed coat (SCOC), and a 25.37% reduction in HC, lastly resulting in a 9.91% increase in SOC (Fig. 4, M and P). In addition, compared with Ken-C8<sup>WT</sup>, the PA content, seed coat thickness (SCT), NDF, ADF, and ADL in Ken-C8<sup>DYSOC1</sup> were reduced by 69.30, 36.05, 17.14, 31.05, and 66.12%, respectively (Fig. 4, G, H, M, and P).

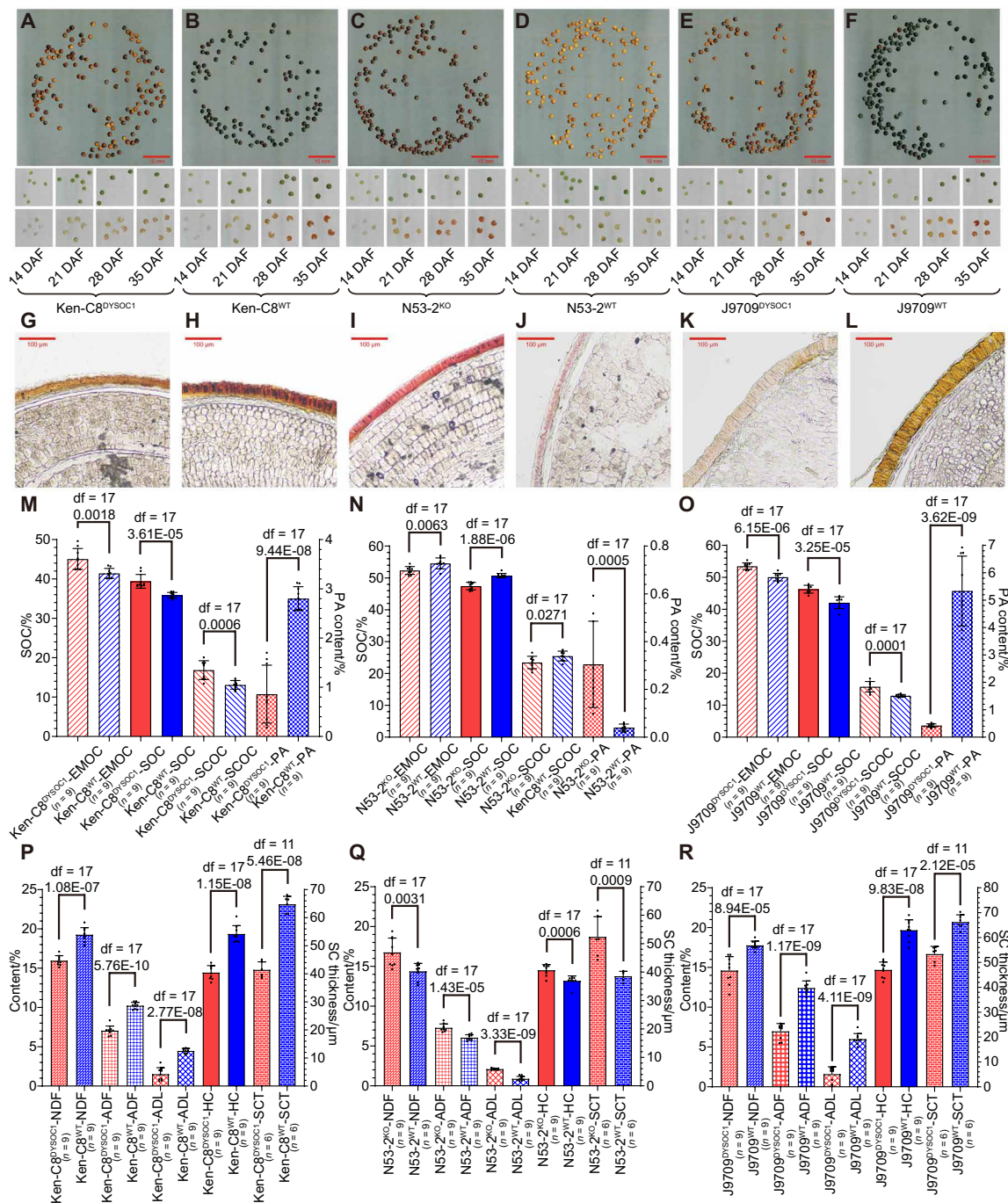
A CRISPR-Cas9 gene knockout vector of *DYSOC1* was constructed (fig. S10, A and B) and transformed into N53-2, and the homozygous knockout lines (N53-2<sup>KO</sup>) exhibited darker SCC compared with wild-type N53-2 (N53-2<sup>WT</sup>) (Fig. 4, C and D). Further phenotypic investigation revealed that N53-2<sup>KO</sup> lines had a 4.07% reduction in EMOC, a 7.90% reduction in SCOC, and a 10.10% increase in HC, which resulted in a 6.49% reduction in SOC (Fig. 4, N and Q). In addition, compared with N53-2<sup>WT</sup>, the PA content, SCT, NDF, ADF, and ADL of N53-2<sup>KO</sup> were increased by 659.14, 35.84, 16.50, 20.44, and 137.16%, respectively (Fig. 4, I, J, N, and Q).

To evaluate the universality of *DYSOC1*, we transformed the *DYSOC1* complementary vector into J9709, which manifested as black SCC and high SLC. The expression of *DYSOC1* in J9709 (J9709<sup>DYSOC1</sup>) resulted in yellow SCC, a 6.83% increase in EMOC, a 21.79% increase in SCOC, a 25.37% reduction in HC, and a 10.05% increase in SOC (Fig. 4, E, F, O, and R). Compared with wild-type J9709 (J9709<sup>WT</sup>), the PA content, SCT, NDF, ADF, and ADL of J9709<sup>DYSOC1</sup> were decreased by 92.05, 19.46, 17.83, 43.85, and 73.46%, respectively (Fig. 4, K, L, O, and R). The phenotypic changes mentioned above were consistent with those observed in the Ken-C8 transformation.

Therefore, we confirmed that *DYSOC1* was a universal locus for SOC, SCC, and ADL improvement. In addition, the *DYSOC1* promoter was subsequently amplified for the  $\beta$ -glucuronidase (GUS) stain experiment, demonstrating that *DYSOC1* was highly expressed in the seed coat, silique peel, and pseudo septum. It was slightly expressed in the flower, but no signals were found in nutritive organ, flower bud, or embryo (Fig. 5, H to M).

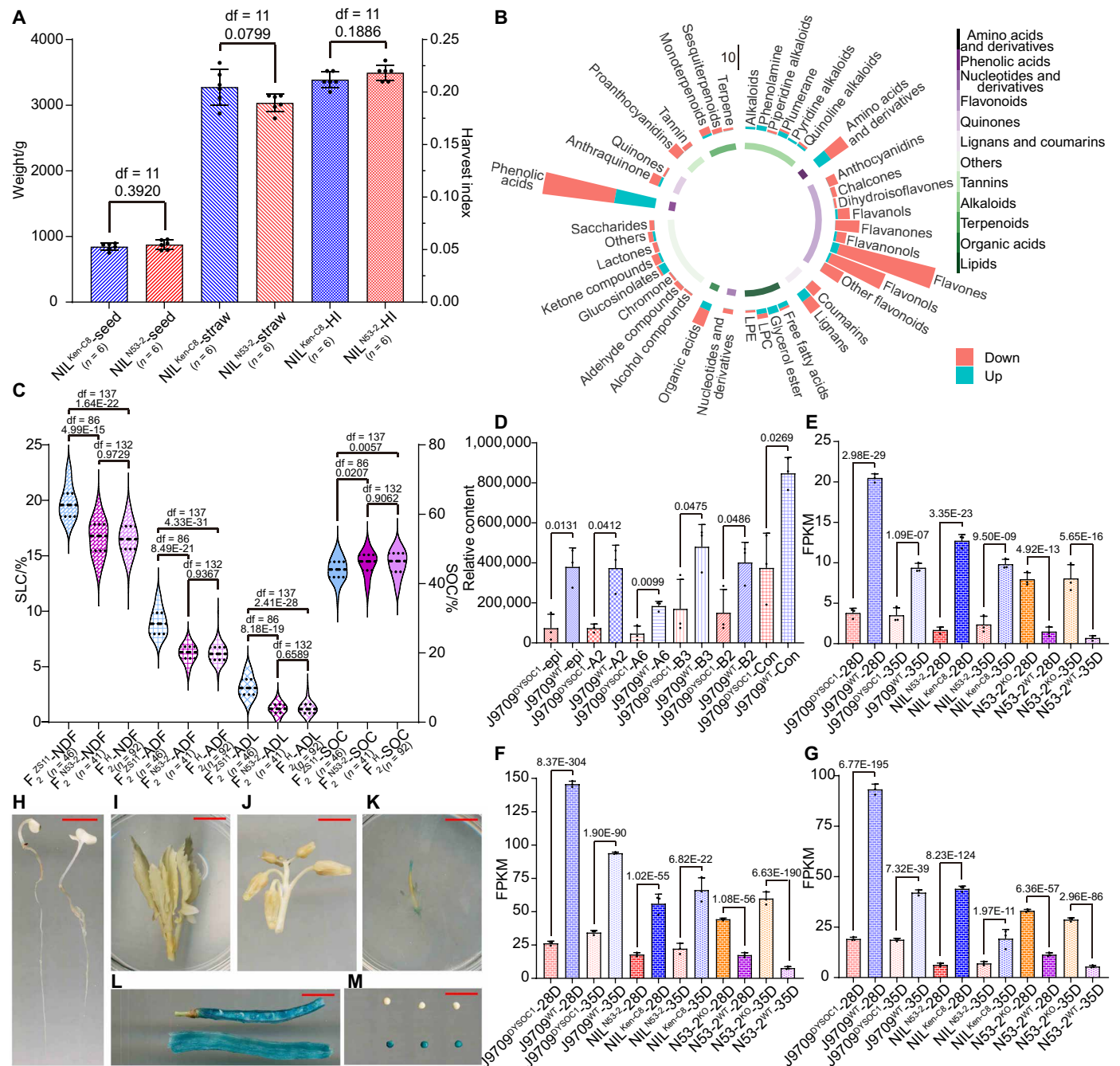


**Fig. 3. The QTL fine mapping and candidate gene identification of Hotspot 13.** (A) The QTL mapping of SOC, SCC, NDF, ADF, and ADL on ChrA9 in the KN DH population (5, 42, 43). (B) The matured seeds of BC<sub>5</sub>F<sub>2</sub> NIL lines. The red bars represent 10 mm. LOD, logarithm of the odds. (C) The QTL fine mapping of Hotspot 13 in the BC<sub>5</sub>F<sub>2</sub> NIL population. (D) The SOC (right y axis) and SLC (left y axis) of different NIL groups. The blue horizontal line denotes the corresponding trait value of Ken-C8. The histograms represent the average values, and the error bars represent the SDs. The df and P values are annotated above histograms (two-tailed one-way ANOVA). The number of individuals in each group is annotated behind the x axis. (E) The genetic background and phenotype of five typical NIL subgroups and their parents. Chromosome fragments inherited from Ken-C8 and N53-2 are marked as black and orange, respectively, and gray indicates heterozygous regions. Names and individual numbers are labeled on the left, and the label colors represent their SCC. The bar graph presents the SOC of those subgroups and their parents, and each point represents one individual. The histograms represent the average values, and the error bars represent SDs. The df and P values are annotated above histograms (compared with Ken-C8, two-tailed one-way ANOVA). (F) The eQTL mapping of *BnaA09g44670D*. (G) Expression profile of the 13 expressed genes within *uqA9-9* in the seed, seed coat (SC), and embryo (EM) of NIL lines. The heatmap was drawn by the log<sub>2</sub>(fold change) of the FPKM of NIL<sup>Ken-C8</sup> versus NIL<sup>N53-2</sup>. (H and I) The CDS sequence variation of *BnaA09g44670D* between N53-2 and other *B. napus* materials in the second exon (H) and the terminal region (I). (J) The amino acid sequence variation of *BnaA09g44670D* between N53-2 and other *B. napus* materials.



**Fig. 4. The phenotype of *DYSC1* transgenic lines.** (A to F) The matured seeds (the first row), intact immature seeds (the second row), and immature seed coats stained by vanillin (the third row) of Ken-C8<sup>DYSC1</sup> (A), Ken-C8<sup>WT</sup> (B), N53-2<sup>KO</sup> (C), N53-2<sup>WT</sup> (D), J9709<sup>DYSC1</sup> (E), and J9709<sup>WT</sup> (F). The red bars represent 10 mm. (G to L) The matured seed slices of Ken-C8<sup>DYSC1</sup> (G), Ken-C8<sup>WT</sup> (H), N53-2<sup>KO</sup> (I), N53-2<sup>WT</sup> (J), J9709<sup>DYSC1</sup> (K), and J9709<sup>WT</sup> (L) that stained by phloroglucinol. The red bars represent 100 μm. (M to O) The EMOC, SCOC, SOC, and PA content of Ken-C8<sup>DYSC1</sup> and Ken-C8<sup>WT</sup> (M), N53-2<sup>KO</sup> and N53-2<sup>WT</sup> (N), J9709<sup>DYSC1</sup> and J9709<sup>WT</sup> (O). Among them, EMOC, SOC, and SCOC are plotted on the left y axis, and PA content is plotted on the right y axis. Each point represents one independent transgenic line, and the histograms represent the average values. The error bars represent SDs, and the df and P values are annotated above histograms (two-tailed one-way ANOVA). In addition, the number of independent transgenic lines is annotated behind the x axis. Three repetitions were conducted for each line, and the average value was adopted. (P to R) The NDF, ADF, ADL, HC, and SCT of Ken-C8<sup>DYSC1</sup> and Ken-C8<sup>WT</sup> (P), N53-2<sup>KO</sup> and N53-2<sup>WT</sup> (Q), J9709<sup>DYSC1</sup> and J9709<sup>WT</sup> (R). Among them, NDF, ADF, ADL, and HC are plotted on the left y axis, and SCT is plotted on the right y axis. Each point represents one independent transgenic line, the histograms represent the average values, and the error bars represent SDs, and the df and P values are annotated above histograms (two-tailed one-way ANOVA). In addition, the number of independent transgenic lines is annotated behind the x axis. Three repetitions were conducted for each line, and the average value was adopted. Specially, the SCT of each material was measured in six randomly selected areas in (G) to (L), respectively.





**Fig. 5. *DYSOC1* restricted PA and lignin monomer synthesis in the seed coat.** (A) The seed weight (left y axis), straw weight (left y axis), and harvest index (right y axis) of NIL<sup>Ken-CB</sup> and NIL<sup>N53-2</sup> plots. Each point represents one plot, and the histograms represent the average values. The error bars represent SDs, and the df and P values are annotated above histograms (two-tailed one-way ANOVA). The numbers of plots are annotated behind the x axis. (B) DAMs detected between the seed coat of J9709<sup>DYSOC1</sup> and J9709<sup>WT</sup>. The pillar height represents the DAM numbers, with DAM classification annotated in the inner circle. (C) The SLC (left y axis) and SOC (right y axis) of F<sub>2</sub><sup>ZS11</sup>, F<sub>2</sub><sup>N53-2</sup>, and F<sub>2</sub><sup>H</sup> in the 22WH microenvironment. The top, middle, and bottom horizontal dashed lines of violin plots represent the first, second, and third quartiles of each group, respectively, and the df and P values are annotated above violin plots (two-tailed one-way ANOVA). The number of individuals of F<sub>2</sub><sup>ZS11</sup>, F<sub>2</sub><sup>N53-2</sup>, and F<sub>2</sub><sup>H</sup> are annotated behind the x axis. (D) The relative content of 4β-8-epicatechin (epi), procyanidin A2 (A2), procyanidin A6 (A6), procyanidin B2 (B2), procyanidin B3 (B3), and conifer-aldehyde (Con) in the seed coat of J9709<sup>DYSOC1</sup> and J9709<sup>WT</sup>. Each point represents one replication, the histograms represent the average values, and the error bars represent SDs. The P values are annotated above histograms (Wald test). (E to G) Expression level of C06.TT1 (E), A09.TT8 (F), and C09.TT8 (G). Each point represents one replication, the histograms represent the average values, and the error bars represent SDs. The P values are annotated above histograms (Wald test). (H to M) The expression profile of *DYSOC1* evaluated by the *GUS* stain in the seedling (H), leaf and stem (I), bud (J), flower (K), pseudoseptum (L, above), silique peel (L, below), embryo (M, above), and seed coat (M, below). The red bars represent 10 mm.

### DYSOC1 was promising for high SOC *B. napus* breeding

To evaluate the breeding value of *DYSOC1* in *B. napus*, F<sub>2</sub> populations were constructed using N53-2 and ZS11, a widely cultivated commercial variety in China that manifested black SCC and medium SLC and shared the same *BnaA09g44670D* DNA sequence as Ken-C8 (Fig. 3, H to J, and fig. S11). The F<sub>2</sub> populations were cultivated in 2022 and 2023 in Wuhan (abbreviated as 22WH and 23WH, respectively) and genotyped by the *DYSOC1*-specific molecular marker Ms-4 (fig. S12A) and then categorized into three groups: homozygotes containing *DYSOC1* (F<sub>2</sub><sup>N53-2</sup>), homozygotes lacking *DYSOC1* (F<sub>2</sub><sup>ZS11</sup>), and heterozygotes (F<sub>2</sub><sup>H</sup>). The SCC of F<sub>2</sub><sup>N53-2</sup> was brighter than that of F<sub>2</sub><sup>ZS11</sup>, as expected, and the SOC of F<sub>2</sub><sup>N53-2</sup> was significantly higher than that of F<sub>2</sub><sup>ZS11</sup> with an average increase of 4.04%. The NDF, ADF, and ADL of F<sub>2</sub><sup>N53-2</sup> were significantly lower than those of F<sub>2</sub><sup>ZS11</sup> by 16.04, 30.98, and 61.96%, respectively, in 22WH (Fig. 5C and fig. S12, B to J). The phenotype of F<sub>2</sub><sup>H</sup> was the same as that of F<sub>2</sub><sup>N53-2</sup>. A similar trend was also observed in 23WH (fig. S13A).

The major FA proportions of the F<sub>2</sub> population and transgenic lines were also analyzed. Although the content of some FAs significantly changed in the transgenic lines, most of these variances vanished when converted into the FA proportions (figs. S14 to S16). In addition, although the FA composition of ZS11 and N53-2 was diverse, no significant differences were observed in the major FA proportions among different genotypes of the F<sub>2</sub> population (figs. S11 and S13), as no side effects of *DYSOC1* were found in the major FA composition.

In the transgenic lines, other seed characteristics [seed protein content (SPC) and thousand seed weight] and plant architecture traits (plant height, branch height, main flower length, silique length, seed number per silique, main flower silique number, and branch number) were also investigated, showing no significant side effects of *DYSOC1* (figs. S17 and S18). The plot yields of NIL<sup>Ken-C8</sup> and NIL<sup>N53-2</sup> were also measured, revealing no significant differences in seed yield, straw yield, or harvest index (Fig. 5A and fig. S19).

### DYSOC1 down-regulated the phenylpropanoid and flavonoid biosynthesis pathways and restricted PA and lignin monomer synthesis

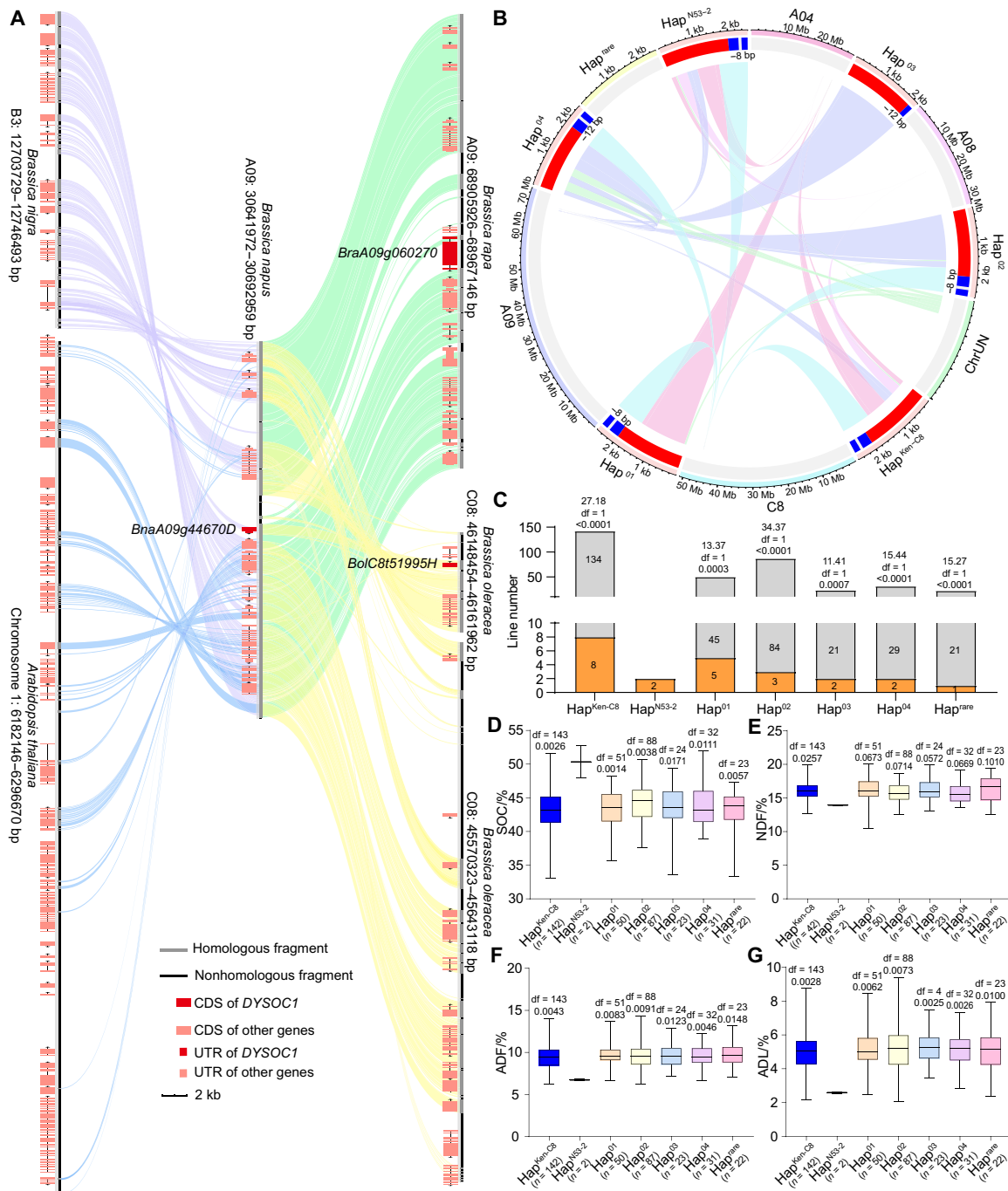
To characterize the mechanism of *DYSOC1* controlling SOC, SCC, and SLC, a widely targeted metabolomic analysis was conducted in the mature seed coat of J9709<sup>DYSOC1</sup> and J9709<sup>WT</sup>. Principal components analysis (PCA) and cluster tree analysis distinguished all replications of J9709<sup>DYSOC1</sup> and J9709<sup>WT</sup> (fig. S20, A and D). A total of 255 differential accumulated metabolites (DAMs) were screened from the 1989 detected metabolites (Fig. 5B and fig. S20, B and C). Among them, flavonoids comprised 41.57% of the DAMs (106), with 94.34% (100) being down-regulated in J9709<sup>DYSOC1</sup>. Following flavonoids, phenolic acids account for 18.43% of the DAMs (47), with 63.83% (30) being down-regulated in J9709<sup>DYSOC1</sup> (Fig. 5B). These results indicated that the phenylpropanoid and flavonoid biosynthesis pathways were largely down-regulated in the seed coat of J9709<sup>DYSOC1</sup>. In addition, the contents of eight of the 10 detected metabolites directly involved in the lignin monomer biosynthesis were largely decreased in J9709<sup>DYSOC1</sup>, including one DAM (coniferaldehyde). The contents of eight detected PAs were all reduced in J9709<sup>DYSOC1</sup>, including five DAMs (4β-8-epicatechin, procyanidin A2, procyanidin A6, procyanidin B3, and procyanidin B2) (Fig. 5D and fig. S20, E and F), which is consistent with the sharp decline in SLC and PA content in the seed coat of J9709<sup>DYSOC1</sup>.

To investigate the effect of *DYSOC1* at the transcriptional level, transcriptomic analysis was conducted in the 28- and 35-DAF seed coat of J9709<sup>DYSOC1</sup>, J9709<sup>WT</sup>, N53-2<sup>KO</sup>, and N53-2<sup>WT</sup> (fig. S21). Together with the RNA-seq data of the seed coat of NIL<sup>Ken-C8</sup> and NIL<sup>N53-2</sup>, almost all KEGG orthologies within phenylpropanoid and flavonoid biosynthesis pathways that participated in PA synthesis were down-regulated by *DYSOC1* (fig. S22). Furthermore, KEGG enrichment analysis of differentially expressed genes (DEGs) of NIL<sup>Ken-C8</sup> versus NIL<sup>N53-2</sup>, J9709<sup>WT</sup> versus J9709<sup>DYSOC1</sup>, and N53-2<sup>WT</sup> versus N53-2<sup>KO</sup> in the 28- and 35-DAF seed coat revealed that phenylpropanoid and flavonoid biosynthesis pathways were consistently significantly enriched (fig. S23). A total of 26 common down-regulated and 125 common up-regulated DEGs were identified through all six comparisons (fig. S24). Among them, *A9.TT8*, *C9.TT8*, and *C6.TT1* emerged as down-regulated pivotal facilitators of PA synthesis that have been reported before (29, 30, 49, 50) (Fig. 5, E to G), which was consistent with the eQTL hotspot analysis (Fig. 2E).

### Evolution and haplotype distribution of *DYSOC1* in *B. napus*

As previously mentioned, *DYSOC1* was identified as a Skp1-like protein. BLAST analysis of *DYSOC1* in Phytozome (<https://phytozome-next.jgi.doe.gov/>) revealed that the homologous genes were only detected in *B. napus* and its ancestors (*B. rapa* and *B. oleracea*) (51). In addition, the sequence identity of *DYSOC1* and other known Skp1-like proteins in *A. thaliana* was less than 40% (table S36 and fig. S25), indicating that *DYSOC1* is an orphan gene that evolved within *Brassica*. To further address the evolution of *DYSOC1*, the micro-synteny of *DYSOC1* and its 25-kb flanks were analyzed among *B. napus* (AACC), *B. rapa* (AA), *Brassica nigra* (BB), *B. oleracea* (CC), and *A. thaliana* genomes (Fig. 6A and fig. S26). *A. thaliana* does not have any *DYSOC1* homologous sequence, and *B. nigra* contained only a fragment of the *DYSOC1* coding sequence (CDS), which was not addressed as a gene. *B. rapa* contained the entire *DYSOC1* CDS, while the second exon was interrupted by a 3872-bp insertion that contained a predicted transposable element. *B. oleracea* contained the intact *DYSOC1* CDS, while its promoter was not consistent with that of *B. napus* (Fig. 6A and table S37).

To evaluate the distribution of *DYSOC1* in *B. napus*, a natural population containing 357 lines was collected, and *BnaA09g44670D* and its promoter were amplified in these lines for haplotype analysis. Five main haplotypes (frequency > 5%) were detected, among which one was identical to that of Ken-C8 (named Hap<sup>Ken-C8</sup>), and the other four were named Hap<sup>01</sup> to Hap<sup>04</sup>. Besides, six rare haplotypes (frequency < 5%) were also isolated, among which one was identical to that of N53-2 (named Hap<sup>N53-2</sup>, which is the *DYSOC1* allele), and the other five were all classified as Hap<sup>rare</sup> (table S38). Compared with the ancestors *B. rapa* and *B. oleracea*, *BnaA09g44670D* in *B. napus* experienced multiple SVs (chromosome structure variations) across at least four chromosomes, and plentiful haplotypes were consequently formed in the natural population (Fig. 6B). The phenotype of each *B. napus* line in the natural population was measured, which indicated that only Hap<sup>N53-2</sup> was closely linked with yellow SCC and low SLC. The SOC of Hap<sup>N53-2</sup> was significantly higher than that of other haplotypes (Fig. 6, C to G). However, no significant differences were observed in other traits between Hap<sup>N53-2</sup> and other haplotypes (fig. S27).



**Fig. 6. The micro-synteny and haplotype analyses of *DYSOC1*.** (A) The micro-synteny of *DYSOC1* and its 25-kb flanks with *B. rapa*, *B. nigra*, *B. oleracea*, and *A. thaliana* genome. The query sequence was homologous with the following regions: "A09: 68.91–68.97 Mb," "B3: 12.70–12.75 Mb," "C08: 45.57–45.64 Mb and 46.15–46.16 Mb," and "chromosome 1: 6.18–6.30 Mb" of *B. rapa*, *B. nigra*, *B. oleracea*, and *A. thaliana* genome, respectively. The links indicate the best-aligned position of each query sequence segment relative to the refs. The name of *DYSOC1* and its homology genes are annotated. UTR, untranslated region. The red sections represent promoters, the blue sections represent exons, and the gaps between the blue sections represent introns. The length of Indels in the exons between Hap<sup>Ken-C8</sup> and other haplotypes is also annotated. (B) The variation among six *BnaA09g44670D* haplotypes (Hap<sup>Ken-C8</sup>, Hap<sup>N53-2</sup>, and Hap<sup>01-04</sup>) and their formation through SVs. The links represent the best-aligned position of each haplotype sequence segment relative to the merged AACC genome. The red sections represent promoters, the blue sections represent exons, and the gaps between the blue sections represent introns. The length of Indels in the exons between Hap<sup>Ken-C8</sup> and other haplotypes is also annotated. (C) The yellow SCC frequency of the different *BnaA09g44670D* haplotypes. The frequencies are annotated in the histograms, and the chi-square, df, and *P* value are annotated above the histogram (compared with Hap<sup>N53-2</sup>, chi-square test). (D to G) The SOC (D), NDF (E), ADF (F), and ADL (G) variations of different *BnaA09g44670D* haplotypes in the *B. napus* natural population. The error bars represent the min and max values of each haplotype, and the top, middle, and bottom parallel lines represent the first, second, and third quartiles of each haplotype, respectively. The df and *P* values are annotated above histograms (compared with Hap<sup>N53-2</sup>, two-tailed one-way ANOVA). In addition, the number of lines of each haplotype is annotated behind the x axis. Two replications were cultivated for each line, and the average trait value was adopted.



## DISCUSSION

High SOC, yellow SCC, and low SLC are valuable traits for *B. napus* breeding. Many independent studies have revealed that yellow SCC, high SOC, and low SLC are closely correlated in *B. napus* (3, 5, 52). Previous studies indicated that yellow SCC is mainly controlled by recessive genes in *B. napus* (19–21), and all cloned yellow SCC genes are recessive, which limits their application in breeding. For instance, *BnaO9MYB47a* promotes flavonoid biosynthesis, and its expression results in black SCC (7). Now, only a few dominant loci for yellow SCC have been reported (22, 23), and no dominant genes have been successfully cloned.

Over the last three decades, many QTLs controlling SCC, SOC, SLC, and other related traits have been reported (3, 26, 53–56), with a few being characterized (7, 36, 57, 58). For example, *BnaC07.CCRL* was identified as a positive regulator of lignin biosynthesis by combining genome- and transcriptome-wide association studies. Knockdown of *Bna.CCRLs* led to a 4.2% increase in SOC, as well as a 17.6 and 13.1% reduction in SLC and SCT, respectively (58). Tan *et al.* (36) observed two positive regulators of SOC (*NAC13* and *SCL31*) in *B. napus*, which increased the number of lipid droplets and raised SOC by 10.56 and 8.33%, respectively, in *Arabidopsis*. However, the pleiotropic QTLs are still under characterization, and no genes that simultaneously control SCC, SOC, and SLC have been identified through QTL fine mapping in *B. napus*.

In recent years, the combination of eQTL and QTL mapping has been widely applied to characterize important agronomic traits in crops (33–36). In this study, eQTL analysis was conducted using a DH population derived from Ken-C8 and N53-2. In total, 58,981 eQTLs for 31,785 eGenes were identified, and approximately half of the eGenes were regulated by more than one eQTL. The cis-eQTLs explained more PV than the trans-eQTLs, consistent with previous studies in *B. napus* and other species (33–36, 59). In addition, a significant relationship between eQTLs, sequence variation, and the 3D genome was reported in the present study.

Trans-eQTLs usually affect the expression of eGenes in the form of TFs or certain metabolites. Therefore, many trans-eQTLs cluster together and form eQTL hotspots in certain regions (60). In this study, 25 trans-eQTL hotspots were identified genome-wide, which contained key regulators of certain pathways (such as flavonoid biosynthesis and fatty acid degradation pathways). Notably, Hotspot 13, located at the end of ChrA09, regulated the phenylpropanoid and flavonoid biosynthesis pathways and affected SCC, SOC, and SLC pleiotropically (5, 42, 43). A dominant allele *DYSOC1* was cloned from Hotspot 13 through the integration of eQTL analysis, QTL fine mapping, and map-based cloning. *DYSOC1* repressed the expression of *Bn.TT8s* (*BnaA09g22810D* and *BnaC09g24870D*) and *Bn.TT1* (*BnaC06g08390D*) and down-regulated phenylpropanoid and flavonoid biosynthesis pathway genes in the seed coat of *B. napus*, subsequently leading to yellow SCC, low SLC, and thin SCT. Consequently, *DYSOC1* universally improved SOC by up to 10.05%. It is an important advance in the cloning of the dominant yellow SCC gene as well as the pleiotropic gene for SCC, SOC, and SLC in *B. napus*.

Previous studies have demonstrated that PA and lignin synthesis are mainly regulated by TFs, such as MYB, MBW, and WRKY (16, 61). However, regulation through the ubiquitin ligase pathway is rarely reported. Identified as a Skp1-like protein, *DYSOC1* is

anticipated to form SCF ubiquitin ligase complexes with F-box proteins and CUL proteins (62). In these complexes, F-box proteins recognize target proteins and direct them to the CUL1/RBX1 core for ubiquitylation and subsequent degradation (46). As far as we know, a few F-box proteins, such as OsFBK1 (63), KFB01, KFB20, and KFB50 (64), have been identified to mediate the ubiquitylation and degradation of enzymes involved in lignin and PA synthesis. However, the regulatory roles of Skp1-like proteins in these pathways have not yet been investigated. This study presents evidence of Skp1-like protein (*DYSOC1*) regulating SCC. Therefore, *DYSOC1* was supposed to interact with a specific F-box protein, which, in turn, mediates the ubiquitination and degradation of key activating factors involved in the phenylpropanoid and flavonoid biosynthesis pathways.

Overexpressing positive regulators and knocking out negative regulators are possible strategies for improving SOC, SCC, and SLC in *B. napus* (65–70), but their application in breeding is still challenging. Therefore, searching for naturally dominant alleles is the most effective breeding strategy. As a universally dominant karyogene, *DYSOC1* can be easily induced into any black SCC parent of the crossbreeding through hybridization and then applied to rapeseed crossbreeding (fig. S28). The yellow SCC can serve as a clear phenotypic marker during this process to ensure accuracy. In the present study, the NIL lines and hybridization between N53-2 and ZS11 demonstrated that *DYSOC1* improved SCC, SOC, and SLC in both heterozygotes and homozygotes. More preciously, no unsatisfactory effects were observed on FA composition, plant architecture, or seed yield due to *DYSOC1*. The rarity of Hap<sup>N53-2</sup> in the natural *B. napus* population (2 of the 357 lines) further highlights the potential of *DYSOC1* allele in *B. napus* breeding. Although some researchers expressed concerns about the shortcomings of yellow SCC seeds (71, 72), these challenges are expected to be addressable through suitable breeding strategies and locus selection (fig. S28). On the basis of the existing results, the widespread application of *DYSOC1* could substantially increase rapeseed oil production.

## MATERIALS AND METHODS

### Plant materials and field experiment

The KN DH population and its parental lines, Ken-C8 and N53-2, were first described by Wang *et al.* (73). Ken-C8, the male parent of the KN DH population, is a spring *B. napus* manifested low SOC (around 40%), high SLC, and black SCC. In contrast, N53-2, the female parent of the KN DH population, is a winter *B. napus* manifested high SOC (over 50%), low SLC, and yellow SCC. Besides, ZS11 is a widely cultivated commercial variety in China, and J9709 is a widely used spring *B. napus* line in transgenic research.

In this study, all plant materials were cultivated in the field under natural conditions in Wuhan, Hubei Province (semi-winter rapeseed-producing area, 113.68°E, 30.58°N) unless otherwise stated. The KN DH population was cultivated in Yangling (winter rapeseed-producing area, 108.08°E, 34.27°N) in 2019 for eQTL sample collection. The BC<sub>5</sub>F<sub>2</sub> NIL lines were cultivated in Wuhan, Yangling, and Xining (spring rapeseed-producing area, 101.49°E, 36.34°N). The field management adhered to the regular agricultural requirements described by Chao *et al.* (42).

### Sample collection and the RNA-seq of the KN DH population

The 28-DAF selfed seeds were collected from 156 different DH lines and the parental lines Ken-C8 and N53-2 for RNA extraction, library

construction, and RNA-seq analysis following the reported procedures (74). For quality control, all raw reads were filtered to remove low-quality reads, adapter sequences, and reads with  $Q < 30$  bases. The clean data of tall samples were mapped to the *B. napus* Darmorbzh reference genome using STAR software, and these mapped reads were used for FPKM calculation and subsequent analysis.

### SNP calling and genotyping

SNP loci were identified and selected by Genome Analysis Toolkit (GATK) software (75), and the low-quality SNPs were filtered out. The retained SNP markers were used for genotyping to facilitate subsequent genetic analysis.

The genotype encoding followed the two alleles encoding rules, which were commonly used in genetic research (74). All polymorphism markers were coded into eight segregating patterns (ab × cd, ef × eg, hk × hk, lm × ll, nn × np, aa × bb, ab × cc, and cc × ab), and the markers with the segregation pattern of aa × bb were selected to match with the DH population used in this study. After SNP selection and bin division, a total of 4005 effective SNP markers were selected for the linkage analysis.

### Linkage map construction, evaluation, and eQTL analysis

HighMap software was used to arrange the SNPs, and the genetic distances among SNPs were calculated using the Kosambi mapping function (76). In total, 3925 of the 4005 effective SNP markers were assigned to 19 LGs. The maximum likelihood method, SMOOTH algorithm, and Detail MSTmap were used to construct the genetic linkage map and correct genotyping errors.

The Composition Internal Mapping (CIM) method and the RQTL software were used for eQTLs mapping. The 1000 permutation tests were performed to determine the thresholds of eQTLs. According to the relative distance between eQTLs and the corresponding eGenes, the eQTLs were further classified as cis-eQTLs (<1 Mb) or trans-eQTLs (>1 Mb or located on different chromosomes). The eQTL density was calculated with a sliding window of 1-Mb window and a step size of 100 kb, and the 95% value served as the threshold for trans-eQTL hotspot identification.

### Weighted gene co-expression network analysis

The SOC and SCC of the KN DH population in different microenvironments have been published before (5, 42). In the present study, the co-expression network analysis was performed using the R package WGCNA v1.7.3. Specifically, the “adjacency” function in WGCNA was used to calculate the adjacencies with a soft-thresholding power of seven. To identify modules of co-expressed genes, the function “TOMsimilarity” was used to calculate the topological overlap matrix (TOM) between all genes. Then, the TOM values were used as input for average linkage hierarchical clustering. The clustered gene trees were cut into different modules using the dynamic shearing method. The module eigengene (ME) was summarized for each module. Last, association analyses were conducted between merged ME and phenotypic traits (77).

### RC.eQTL generation

The distribution of eQTL length and the number of eQTLs for each eGenes were first analyzed. Then, a series of genomic regions were randomly generated across the 19 chromosomes and matched with 1000 randomly selected genes on the basis of the observed eQTL

length and number distribution. These randomly selected genes and corresponding randomly generated genomic regions were regarded as the RC.eQTLs.

### Trait investigation

For NIL and the  $F_2$  populations of ZS11 × N53-2, the matured seeds of each individual were collected for seed traits analysis. As for the *B. napus* natural population, transgenic lines, and their corresponding WT lines, each line was cultivated for two rows. For each line, the matured seeds from five healthy individuals were pooled for seed traits analysis. In addition, the *B. napus* natural population was cultivated for two replications.

In the present study, the SOC, SLC, SPC, and FA composition of the NIL lines,  $F_2$  populations of ZS11 × N53-2, and the *B. napus* natural population were measured by the NIRS method using NIRS Systems 5000 (Foss, Denmark) (78–80). The SPC and SLC of the transgenic lines and their corresponding WT lines were also analyzed by NIRS, while their SOC and FA composition were measured by gas chromatography (GC) as previously described (81). Total FAs were extracted using 2.5%  $H_2SO_4$ -methanol, with heptadecanoic acid as the internal reference. Major FA content was measured using the 7890A chromatographic column (Agilent, USA) in the gas chromatography system (Agilent, USA) following the instructions. Three repetitions were conducted for each measurement, and the average values were adopted. The SOC, SLC, SPC, HC, PA content, and FA content described in the present study represent the mass ratio of each component relative to the total seed weight.

### SCT and lignin content analysis of seed coat via hard-tissue slicing and phloroglucinol stain

The mature seeds were fixed in 70% Formalin-Aceto-Alcohol (FAA) solution for 2 days, followed by dehydration, methyl methacrylate (MMA) infiltration, and MMA embedding. The seeds were sectioned into 10- $\mu$ m slices in the middle and placed in a 60°C oven overnight. Slices were stained with phloroglucinol staining solution (Servicebio, China) for 30 s after rehydration and mordanting. The excess staining solution was removed, and the images were captured within 5 min.

### PA content analysis in the seed coat of *B. napus*

The PA of mature *B. napus* seed coats was extracted using 2.5% butanol-HCl (82). Meanwhile, for the standard curve, the PA standard (Macklin, Shanghai, China) was dissolved in methanol and extracted following the same protocol. The absorbance of the mixture was measured at 545 nm using a FlexStation 3 multimode reader (Molecular Devices, USA) (figs. S29 and S30). The PA concentration of the extracted seed coat solutions was calculated on the basis of the standard curve.

Vanillin turns red upon binding to PA. The 1% vanillin staining solution was prepared as reported before (72) and balanced with 12 mol/L HCl before use. The immature seed coats were soaked in the staining solution for 10 min at room temperature before images were captured.

### NIL construction and QTL fine mapping

The genetic background of each individual of the KN DH population was screened, and two lines (I-094 and I-179) that inherited the superior genotype of Hotspot 13 from N53-2 were screened out, which were continuously backcrossed with Ken-C8 five times to

construct the BC<sub>5</sub> population. Subsequently, the BC<sub>5</sub>F<sub>2</sub> population was constructed by selfing the BC<sub>5</sub> individuals. In addition, to increase the genetic diversity of the BC<sub>5</sub>F<sub>2</sub> population and to improve the accuracy of QTL fine mapping, two BC<sub>5</sub> individuals (194-20 and 199-4) with interchanged chromosomes within Hotspot 13 and exhibited black SCC were crossed, and their hybrids were also used for BC<sub>5</sub>F<sub>2</sub> population construction.

Meanwhile, the Indels within Hotspot 13 between the genomes of Ken-C8 and N53-2 were screened out to develop codominant molecular markers, as reported before (44). Six molecular markers were selected for QTL fine mapping (table S39). The fine mapping of Hotspot 13 in the BC<sub>5</sub>F<sub>2</sub> population was subsequently processed using WinQTLCart2.5 software and the CIM method, as described before (42).

### Vector construction and the transformation of *B. napus*

The unabridged sequence of *DYSOC1*, including its native promoter, was amplified from the genomic DNA of N53-2 and cloned into the pCAMBIA1303 plasmid to construct the complementary vector. To guarantee that *DYSOC1* was driven by its native promoter, the *CaMV35S* promoter in the pCAMBIA1303 plasmid was eliminated. The CRISPR-Cas9 target site was designed using the online tool (<http://crispr.hzau.edu.cn/CRISPR2/>) (83), and the gene knockout vector was constructed using pHSN401 plasmid as the backbone.

The complementary vector and CRISPR-Cas9 gene knockout vector were transformed into *B. napus* via *Agrobacterium*-mediated hypocotyl transformation (81). In consideration of the hygromycin B resistance gene in those vectors, the empty pCAMBIA1303 and pHSN401 plasmid were also transformed into *B. napus*, which were referred to as wild-type in this study for unambiguous demonstration.

### Identification of positively transformed plantlets and CRISPR-Cas9 on-target mutants in *B. napus*

The genomic DNA of each transgenic *B. napus* individual was extracted using the NuClean PlantGen DNA Kit (ComWin, Beijing, China) according to the manual. Vector-specific primers were designed to identify positive transgenic plantlets via polymerase chain reaction. As for *DYSOC1* gene knockout plantlets, the CRISPR-Cas9 target site of *DYSOC1* was amplified for Sanger sequencing, and the on-target mutations were screened. Nine lines with different *DYSOC1* on-target mutations were identified (fig. S10B), and nine independent positively transformed plantlets were generated for each complementary vector transgenic experiment. The primers mentioned above have been listed in table S39. Those transgenic lines were subcultured to the T3 generation for trait measurement.

### The expression pattern analysis of *DYSOC1*

The promoter of *DYSOC1* was amplified from the genomic DNA of N53-2 and cloned into the pCAMBIA1305 plasmid with the *GUS* gene on the heels, and the *CaMV35S* promoter was removed beforehand. The *DYSOC1* promoter reconstituted *GUS* vector was transformed into J9709 via *Agrobacterium*-mediated hypocotyl transformation. Positively transformed plant was screened out as described above and subcultured to T3 generation. Tissues from these plants were collected and stained with the GUS Stain Kit (Coolaber, Beijing, China) overnight under room temperature following the instruction. Images were captured after decolorizing with absolute ethanol.

### RNA-seq analysis of NIL and transgenic materials

For seed coat and embryo samples, the seed coat and embryo were separated and immediately soaked in the RNA storage solution (Invitrogen, USA). The RNA-seq samples were named after “line name-stage-tissue-replication”; in addition, “seed coat,” “embryo,” and “DAF” were abbreviated as “SC,” “EM,” and “d,” respectively. For example, “NIL<sup>N53-2</sup>-35d-SC-1” represents the first replication of the 35-DAF seed coat of NIL<sup>N53-2</sup> (tables S40 and S41).

Total RNA extraction, library construction, and sequencing were performed as described before. The clean reads were mapped to Darmor\_bzh *B. napus* reference genome, and DEGs were identified using the Deseq2 R package with  $|\log_2(\text{fold change})| > 1$ ,  $P$  value  $< 0.05$  and  $P$  adj. value  $< 0.05$  in the meantime. The comparisons between different samples were presented as “A versus B,” indicating the relative expression level of genes in sample “B” compared with that in sample “A.”

### Widely targeted metabolomic analysis processed by UPLC-MS/MS

Three independent lines of J9709<sup>WT</sup> and J9709<sup>DYSOC1</sup> were randomly selected, and the seed coats of their mature seeds were collected for metabolomics analysis. All samples were frozen with liquid nitrogen, vacuum freeze dried using a Scientz-100F Lyophilizer (Xinzhi, China), and ground into powder. The sample powder was extracted with 70% methanol solution and filtered with 0.22- $\mu\text{m}$  microfiltration membrane before ultra-performance liquid chromatography (UPLC) (ExionLC AD, Sciex, USA) and tandem mass spectrometry (MS/MS) (Applied Biosystems 4500 QTRAP, USA). The detailed parameters of UPLC and MS/MS followed the reported protocol (84). The DAMs were selected with  $|\log_2(\text{fold change})| > 1$  and  $P$  value  $< 0.05$ .

### Statistical analysis

The two-tailed one-way analysis of variance (ANOVA) assessed the significance of trait variation between different materials or genotypes. The chi-square test was used to evaluate the significance of the correlation between eQTL and gene sequence variation. The chi-square test was used to assess the significance of the correlation between yellow SCC and different *DYSOC1* haplotypes. The significant level was classified as follows:  $P \leq 0.05$  (significant, “\*”),  $P$  value  $\leq 0.01$  (highly significant, “\*\*”), and  $P > 0.05$  (nonsignificant, “n.s.”).

### The evolution analysis of *DYSOC1*

The conserved domain and tertiary structure of *DYSOC1* were predicted using the Conserved Domain Database ([www.ncbi.nlm.nih.gov/cdd](http://www.ncbi.nlm.nih.gov/cdd)) and AlphaFold2 (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>) websites, respectively. Homologous sequences of *DYSOC1* and the 25-kb flanks were searched in the *B. rapa* (Chiifu.v4), *B. nigra* (Ni100 LR v2.0), *B. oleracea* (HDEM.v0), and *A. thaliana* (TAIR10) genome using BLAST (<https://ftp.ncbi.nlm.nih.gov/blast/executables/blast/>) with default parameters. The *B. rapa*, *B. nigra*, and *B. oleracea* genomes were downloaded from BnIR (<https://yanglab.hzau.edu.cn/BnIR/>) (85), and the *A. thaliana* genome was downloaded from TAIR ([www.arabidopsis.org/](http://www.arabidopsis.org/)) (86). The syntenic region of the entire query in each ref genome was first determined, and only the best-aligned sequence of each query segment within the syntenic region was retained for the micro-syteny presentation. In addition, the unabridged BLAST



output was presented in fig. S26. The sequence of each *BnaA09g44670D* haplotype was blasted against the merged AACC genome (merged *B. rapa* and *B. oleracea* genome), and the best-aligned sequence of each query segment relative to the ref genome was retained to evaluate the origin of each *BnaA09g44670D* haplotype.

The homology between DYSOC1 and the *A. thaliana* Skp1-like proteins was calculated using the BLAST. The phylogenetic tree was constructed using MEGA 10 software and the maximum likelihood method under default parameters (87).

## Supplementary Materials

### The PDF file includes:

Supplementary Text

Figs. S1 to S30

Legends for tables S1 to S41

### Other Supplementary Material for this manuscript includes the following:

Tables S1 to S41

## REFERENCES AND NOTES

- B. Chalhouf, F. Denoed, S. Liu, I. A. Parkin, H. Tang, X. Wang, J. Chiquet, H. Belcram, C. Tong, B. Samans, M. Correa, S. C. Da, J. Just, C. Falentin, C. S. Koh, I. Le Clainche, M. Bernard, P. Bento, B. Noel, K. Labadie, A. Alberti, M. Charles, D. Arnaud, H. Guo, C. Daviaud, S. Alamey, K. Jabbari, M. Zhao, P. P. Edger, H. Chelaifa, D. Tack, G. Lassalle, I. Mestiri, N. Schnel, M. C. Le Paslier, G. Fan, V. Renault, P. E. Bayer, A. A. Golicz, S. Manoli, T. H. Lee, V. H. Thi, S. Chalabi, Q. Hu, C. Fan, R. Tollenaere, Y. Lu, C. Battail, J. Shen, C. H. Sidebottom, X. Wang, A. Canaguier, A. Chauveau, A. Berard, G. Deniot, M. Guan, Z. Liu, F. Sun, Y. P. Lim, E. Lyons, C. D. Town, I. Bancroft, X. Wang, J. Meng, J. Ma, J. C. Pires, G. J. King, D. Brunel, R. Delourme, M. Renard, J. M. Aury, K. L. Adams, J. Batley, R. J. Snowdon, J. Tost, D. Edwards, Y. Zhou, W. Liu, A. G. Sharpe, A. H. Paterson, C. Guan, P. Wincker, Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950–953 (2014).
- K. Lu, L. Wei, X. Li, Y. Wang, J. Wu, M. Liu, C. Zhang, Z. Chen, Z. Xiao, H. Jian, F. Cheng, K. Zhang, H. Du, X. Cheng, C. Qu, W. Qian, L. Liu, R. Wang, Q. Zou, J. Ying, X. Xu, J. Mei, Y. Liang, Y. R. Chai, Z. Tang, H. Wan, Y. Ni, Y. He, N. Lin, Y. Fan, W. Sun, N. N. Li, G. Zhou, H. Zheng, X. Wang, A. H. Paterson, J. Li, Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nat. Commun.* **10**, 1154 (2019).
- X. Y. Yan, J. N. Li, F. Y. Fu, M. Y. Jin, L. Chen, L. Z. Liu, Co-location of seed oil content, seed hull content and seed coat color QTL in three different environments in *Brassica napus* L. *Euphytica* **170**, 355–364 (2009).
- H. Li, K. Yu, Z. Zhang, Y. Yu, J. Wan, H. He, C. Fan, Targeted mutagenesis of flavonoid biosynthesis pathway genes reveals functional divergence in seed coat colour, oil content and fatty acid composition in *Brassica napus* L. *Plant Biotechnol. J.* **22**, 445–459 (2024).
- H. Chao, L. Guo, W. Zhao, H. Li, M. Li, A major yellow-seed QTL on chromosome A09 significantly increases the oil content and reduces the fiber content of seed in *Brassica napus*. *Theor. Appl. Genet.* **135**, 1293–1305 (2022).
- J. Jiang, S. Zhu, Y. Yuan, Y. Wang, L. Zeng, J. Batley, Y. P. Wang, Transcriptomic comparison between developing seeds of yellow- and black-seeded *Brassica napus* reveals that genes influence seed quality. *BMC Plant Biol.* **19**, 203 (2019).
- C. Qu, M. Zhu, R. Hu, Y. Niu, S. Chen, H. Zhao, C. Li, Z. Wang, N. Yin, F. Sun, Z. Chen, S. Shen, G. Shang, Y. Zhou, X. Yan, L. Wei, L. Liu, B. Yi, J. Lian, J. Li, Z. Tang, Y. Liang, X. Xu, R. Wang, J. Yin, H. Wan, H. Du, W. Qian, Y. Chai, Q. Zhou, Y. He, S. Zhong, X. Qiu, H. Yu, H. M. Lam, K. Lu, F. Fu, J. Li, Comparative genomic analyses reveal the genetic basis of the yellow-seed trait in *Brassica napus*. *Nat. Commun.* **14**, 5194 (2023).
- G. Bhinder, S. Sharma, H. Kaur, J. Akhatar, M. Mittal, S. Sandhu, Genomic regions associated with seed meal quality traits in *Brassica napus* germplasm. *Front. Plant Sci.* **13**, 882766 (2022).
- M. Shirzadegan, G. Röbbelen, Influence of seed color and hull proportion on quality properties of seeds in *Brassica napus* L. *Fette Seifen Anstrichm.* **87**, 235–237 (1985).
- M. H. Rahman, M. Joersbo, M. H. Poulsen, Development of yellow-seeded *Brassica napus* of double low quality. *Plant Breed.* **120**, 473–478 (2001).
- J. Meng, S. Shi, L. Gan, Z. Li, X. Qu, The production of yellow-seeded *Brassica napus* (AACC) through crossing interspecific hybrids of *B. campestris* (AA) and *B. carinata* (BBCC) with *B. napus*. *Euphytica* **103**, 329–333 (1998).
- M. H. Rahman, Production of yellow-seeded *Brassica napus* through interspecific crosses. *Plant Breed.* **6**, 463–472 (2008).
- L. Z. Liu, J. L. Meng, N. Lin, L. Chen, Z. L. Tang, X. K. Zhang, J. N. Li, QTL mapping of seed coat color for yellow seeded *Brassica napus*. *Yi Chuan Xue Bao* **33**, 181–187 (2006).
- I. Appelhagen, K. Thiedig, N. Nordholt, N. Schmidt, G. Huep, M. Sagasser, B. Weisshaar, Update on transparent testa mutants from *Arabidopsis thaliana*: Characterisation of new alleles from an isogenic collection. *Planta* **240**, 955–970 (2014).
- T. Vogt, Phenylpropanoid biosynthesis. *Mol. Plant* **3**, 2–20 (2010).
- W. Xu, C. Dubos, L. Lepiniec, Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* **20**, 176–185 (2015).
- D. Ma, C. P. Constabel, MYB repressors as regulators of phenylpropanoid metabolism in plants. *Trends Plant Sci.* **24**, 275–289 (2019).
- C. S. Johnson, B. Kolevski, D. R. Smyth, TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor. *Plant Cell* **14**, 1359–1375 (2002).
- A. Van Deynze, K. P. Pauls, The inheritance of seed colour and vernalization requirement in *Brassica napus* using doubled haploid populations. *Euphytica* **74**, 77–83 (1993).
- M. Rahman, G. Li, D. Schroeder, P. B. E. McVetty, Inheritance of seed coat color genes in *Brassica napus* (L.) and tagging the genes using SRAP, SCAR and SNP molecular markers. *Mol. Breed.* **26**, 439–453 (2010).
- M. Shirzadegan, Inheritance of seed color in *Brassica napus* L. *Z. Pflanzenzuecht.* **96**, 140–146 (1986).
- S. Xiao, J. Xu, Y. Li, L. Zhang, S. Shi, S. Shi, J. Wu, K. Liu, Generation and mapping of SCAR and CAPS markers linked to the seed coat color gene in *Brassica napus* using a genome-walking technique. *Genome* **50**, 611–618 (2007).
- X. P. Liu, J. X. Tu, B. Y. Chen, T. D. Fu, Identification and inheritance of a partially dominant gene for yellow seed colour in *Brassica napus*. *Plant Breed.* **124**, 9–12 (2005).
- F. Liu, H. Chen, L. Yang, L. You, J. Ju, S. Yang, X. Wang, Z. Liu, QTL mapping and transcriptome analysis reveal candidate genes regulating seed color in *Brassica napus*. *Int. J. Mol. Sci.* **24**, 9262 (2023).
- A. G. Badani, R. J. Snowdon, B. Wittkop, F. D. Lipsa, R. Baetzler, R. Horn, A. De Haro, R. Font, W. Luhs, W. Friedt, 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 (2006).
- N. Behnke, E. Suprianto, C. Mollers, 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 (2018).
- N. Raboanahiry, H. Chao, L. Guo, J. Gan, J. Xiang, M. Yan, L. Zhang, L. Yu, M. Li, Synteny analysis of genes and distribution of loci controlling oil content and fatty acid profile based on QTL alignment map in *Brassica napus*. *BMC Genomics* **18**, 776 (2017).
- T. Xie, X. Chen, T. Guo, H. Rong, Z. Chen, Q. Sun, J. Batley, J. Jiang, Y. Wang, Targeted knockout of *BnTT2* homologues for yellow-seeded *Brassica napus* with reduced flavonoids and improved fatty acid composition. *J. Agric. Food Chem.* **68**, 5676–5690 (2020).
- Y. Zhai, K. Yu, S. Cai, L. Hu, O. Amoo, L. Xu, Y. Yang, B. Ma, Y. Jiao, C. Zhang, M. Khan, S. U. Khan, C. Fan, Y. Zhou, Targeted mutagenesis of *BnTT8* homologs controls yellow seed coat development for effective oil production in *Brassica napus* L. *Plant Biotechnol. J.* **18**, 1153–1168 (2020).
- J. Lian, X. Lu, N. Yin, L. Ma, J. Lu, X. Liu, J. Li, J. Lu, B. Lei, R. Wang, Y. Chai, Silencing of *BnTT1* family genes affects seed flavonoid biosynthesis and alters seed fatty acid composition in *Brassica napus*. *Plant Sci.* **254**, 32–47 (2017).
- F. Wang, J. He, J. Shi, T. Zheng, F. Xu, G. Wu, R. Liu, S. Liu, Embryonal control of yellow seed coat locus *ECY1* is related to alanine and phenylalanine metabolism in the seed embryo of *Brassica napus*. *G3* **6**, 1073–1081 (2016).
- Z. Li, P. Wang, C. You, J. Yu, X. Zhang, F. Yan, Z. Ye, C. Shen, B. Li, K. Guo, N. Liu, G. N. Thyssen, D. D. Fang, K. Lindsey, X. Zhang, M. Wang, L. Tu, Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton. *New Phytol.* **226**, 1738–1752 (2020).
- H. Wei, X. Wang, Z. Zhang, L. Yang, Q. Zhang, Y. Li, H. He, D. Chen, B. Zhang, C. Zheng, Y. Leng, X. Cao, Y. Cui, C. Shi, Y. Liu, Y. Lv, J. Ma, W. He, X. Liu, Q. Xu, Q. Yuan, X. Yu, T. Wang, H. Qian, X. Li, B. Zhang, H. Zhang, W. Chen, M. Guo, X. Dai, Y. Wang, X. Zheng, L. Guo, X. Xie, Q. Qian, L. Shang, Uncovering key salt-tolerant regulators through a combined eQTL and GWAS analysis using the super pan-genome in rice. *Natl. Sci. Rev.* **11**, e43 (2024).
- X. Yuan, X. Jiang, M. Zhang, L. Wang, W. Jiao, H. Chen, J. Mao, W. Ye, Q. Song, Integrative omics analysis elucidates the genetic basis underlying seed weight and oil content in soybean. *Plant Cell* **36**, 2160–2175 (2024).
- G. Ai, C. He, S. Bi, Z. Zhou, A. Liu, X. Hu, Y. Liu, L. Jin, J. Zhou, H. Zhang, D. Du, H. Chen, X. Gong, S. Saeed, H. Su, C. Lan, W. Chen, Q. Li, H. Mao, L. Li, H. Liu, D. Chen, K. Kaufmann, K. F. Alazab, W. Yan, Dissecting the molecular basis of spike traits by integrating gene regulatory networks and genetic variation in wheat. *Plant Commun.* **5**, 100879 (2024).
- Z. Tan, Y. Peng, Y. Xiong, F. Xiong, Y. Zhang, N. Guo, Z. Tu, Z. Zong, X. Wu, J. Ye, C. Xia, T. Zhu, Y. Liu, H. Lou, D. Liu, S. Lu, X. Yao, K. Liu, R. J. Snowdon, A. A. Golicz, W. Xie, L. Guo, H. Zhao,

- Comprehensive transcriptional variability analysis reveals gene networks regulating seed oil content of *Brassica napus*. *Genome Biol.* **23**, 233 (2022).
37. L. Liu, A. Stein, B. Wittkop, P. Sarvari, J. Li, X. Yan, F. Dreyer, M. Frauen, W. Friedt, R. J. Snowden, 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 (2012).
  38. A. H. Paterson, E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln, S. D. Tanksley, Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**, 721–726 (1988).
  39. S. D. Tanksley, S. Grandillo, T. M. Fulton, D. Zamir, Y. Eshed, V. Petiard, J. Lopez, T. Beck-Bunn, Advanced backcross QTL analysis in a cross between an elite processing. *Theor. Appl. Genet.* **92**, 213–224 (1996).
  40. Z. Cai, P. Xian, Y. Cheng, Y. Yang, Y. Zhang, Z. He, C. Xiong, Z. Guo, Z. Chen, H. Jiang, Q. Ma, H. Nian, L. Ge, Natural variation of *GmFATA1B* regulates seed oil content and composition in soybean. *J. Integr. Plant Biol.* **65**, 2368–2379 (2023).
  41. B. Li, Y. Jia, L. Xu, S. Zhang, Z. Long, R. Wang, Y. Guo, W. Zhang, C. Jiao, C. Li, Y. Xu, Transcriptional convergence after repeated duplication of an amino acid transporter gene leads to the independent emergence of the black husk/pericarp trait in barley and rice. *Plant Biotechnol. J.* **22**, 1282–1298 (2024).
  42. H. Chao, H. Wang, X. Wang, L. Guo, J. Gu, W. Zhao, B. Li, D. Chen, N. Raboanatahary, M. Li, Genetic dissection of seed oil and protein content and identification of networks associated with oil content in *Brassica napus*. *Sci. Rep.* **7**, 46295 (2017).
  43. L. Miao, H. Chao, L. Chen, H. Wang, W. Zhao, B. Li, L. Zhang, H. Li, B. Wang, M. Li, 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 (2019).
  44. L. Zhang, L. Liu, H. Li, J. He, H. Chao, S. Yan, Y. Yin, W. Zhao, M. Li, 3D genome structural variations play important roles in regulating seed oil content of *Brassica napus*. *Plant Commun.* **5**, 100666 (2023).
  45. S. Raffaele, F. Vailliau, A. Léger, J. Joubès, O. Miersch, C. Huard, E. Blée, S. Mongrand, F. Domergue, D. Roby, A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in *Arabidopsis*. *Plant Cell* **20**, 752–767 (2008).
  46. Z. Hua, R. D. Vierstra, The cullin-RING ubiquitin-protein ligases. *Annu. Rev. Plant Biol.* **62**, 299–334 (2011).
  47. H. R. Woo, K. M. Chung, J. H. Park, S. A. Oh, T. Ahn, S. H. Hong, S. K. Jang, H. G. Nam, ORE9, an F-box protein that regulates leaf senescence in *Arabidopsis*. *Plant Cell* **13**, 1779–1790 (2001).
  48. C. E. Stebbins, W. J. Kaelin Jr., N. P. Pavletich, Structure of the VHL-ElonginC-ElonginB complex: Implications for VHL tumor suppressor function. *Science* **284**, 455–461 (1999).
  49. N. Nesi, I. Debeaujon, C. Jond, G. Pelletier, M. Caboche, L. Lepiniec, The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques. *Plant Cell* **12**, 1863–1878 (2000).
  50. I. Appelhagen, G. H. Lu, G. Hup, E. Schmelzer, B. Weisshaar, M. Sagasser, TRANSPARENT TESTA1 interacts with R2R3-MYB factors and affects early and late steps of flavonoid biosynthesis in the endothelium of *Arabidopsis thaliana* seeds. *Plant J.* **67**, 406–419 (2011).
  51. D. M. Goodstein, S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, D. S. Rokhsar, Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).
  52. Y. Ding, S. Yu, J. Wang, M. Li, C. Qu, J. Li, L. Liu, Comparative transcriptomic analysis of seed coats with high and low lignin contents reveals lignin and flavonoid biosynthesis in *Brassica napus*. *BMC Plant Biol.* **21**, 246 (2021).
  53. W. Ecke, M. Uzunova, K. Weissleder, Mapping the genome of rapeseed (*Brassica napus* L.). II. Localization of genes controlling erucic acid synthesis and seed oil content. *Theor. Appl. Genet.* **91**, 972–977 (1995).
  54. L. Guo, H. Chao, Y. Yin, H. Li, H. Wang, W. Zhao, D. Hou, L. Zhang, C. Zhang, M. Li, New insight into the genetic basis of oil content based on noninvasive three-dimensional phenotyping and tissue-specific transcriptome in *Brassica napus*. *Biotechnol. Biofuels Bioprod.* **16**, 88 (2023).
  55. L. Li, Z. Tian, J. Chen, Z. Tan, Y. Zhang, H. Zhao, X. Wu, X. Yao, W. Wen, W. Chen, L. Guo, Characterization of novel loci controlling seed oil content in *Brassica napus* by marker metabolite-based multi-omics analysis. *Genome Biol.* **24**, 141 (2023).
  56. J. Wang, X. Xian, X. Xu, C. Qu, K. Lu, J. Li, L. Liu, Genome-wide association mapping of seed coat color in *Brassica napus*. *J. Agric. Food Chem.* **65**, 5229–5237 (2017).
  57. H. M. Schilbert, K. Holzenkamp, P. Viehvor, D. Holtgrawe, C. Mollers, Homoeologous non-reciprocal translocation explains a major QTL for seed lignin content in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* **136**, 172 (2023).
  58. Y. Zhang, H. Zhang, H. Zhao, Y. Xia, X. Zheng, R. Fan, Z. Tan, C. Duan, Y. Fu, L. Li, J. Ye, S. Tang, H. Hu, W. Xie, X. Yao, L. Guo, Multi-omics analysis dissects the genetic architecture of seed coat content in *Brassica napus*. *Genome Biol.* **23**, 86 (2022).
  59. X. Wang, Q. Chen, Y. Wu, Z. H. Lemmon, G. Xu, C. Huang, Y. Liang, D. Xu, D. Li, J. F. Doebley, F. Tian, Genome-wide analysis of transcriptional variability in a large maize-teosinte population. *Mol. Plant* **11**, 443–459 (2018).
  60. C. Liu, X. Zhu, J. Zhang, M. Shen, K. Chen, X. Fu, L. Ma, X. Liu, C. Zhou, D. X. Zhou, G. Wang, eQTLs play critical roles in regulating gene expression and identifying key regulators in rice. *Plant Biotechnol. J.* **20**, 2357–2371 (2022).
  61. I. Appelhagen, O. Jahns, L. Bartelniewoehner, M. Sagasser, B. Weisshaar, R. Stracke, Leucoanthocyanidin Dioxygenase in *Arabidopsis thaliana*: Characterization of mutant alleles and regulation by MYB-BHLH-TTG1 transcription factor complexes. *Gene* **484**, 61–68 (2011).
  62. E. P. Risseuw, T. E. Daskalchuk, T. W. Banks, E. Liu, J. Cotelesage, H. Hellmann, M. Estelle, D. E. Somers, W. L. Crosby, Protein interaction analysis of SCF ubiquitin E3 ligase subunits from *Arabidopsis*. *Plant J.* **34**, 753–767 (2003).
  63. P. Borah, J. P. Khurana, The OsFBK1 E3 ligase subunit affects anther and root secondary cell wall thickenings by mediating turnover of a Cinnamoyl-CoA reductase. *Plant Physiol.* **176**, 2148–2165 (2018).
  64. X. Zhang, M. Gou, C. J. Liu, *Arabidopsis* Kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. *Plant Cell* **25**, 4994–5010 (2013).
  65. K. Zhu, N. Li, X. Zheng, R. Sarwar, Y. Li, J. Cao, Z. Wang, X. Tan, Overexpression the *BnLACS9* could increase the chlorophyll and oil content in *Brassica napus*. *Biotechnol. Biofuels Bioprod.* **16**, 3 (2023).
  66. S. Fenyk, H. K. Woodfield, T. B. Romsdahl, E. J. Wallington, R. E. Bates, D. A. Fell, K. D. Chapman, T. Fawcett, J. L. Harwood, Overexpression of phospholipid: Diacylglycerol acyltransferase in *Brassica napus* results in changes in lipid metabolism and oil accumulation. *Biochem. J.* **479**, 805–823 (2022).
  67. Z. Xiao, F. Tang, L. Zhang, S. Li, S. Wang, Q. Huo, B. Yang, C. Zhang, D. Wang, Q. Li, L. Wei, T. Guo, C. Qu, K. Lu, Y. Zhang, L. Guo, J. Li, N. Li, The *Brassica napus* fatty acid exporter FAX1-1 contributes to biological yield, seed oil content, and oil quality. *Biotechnol. Biofuels* **14**, 190 (2021).
  68. S. Tang, N. Guo, Q. Tang, F. Peng, Y. Liu, H. Xia, S. Lu, L. Guo, Pyruvate transporter *BnaBASS2* impacts seed oil accumulation in *Brassica napus*. *Plant Biotechnol. J.* **20**, 2406–2417 (2022).
  69. J. Liu, W. Hua, G. Zhan, F. Wei, X. Wang, G. Liu, Increasing seed mass and oil content in transgenic *Arabidopsis* by the overexpression of *wri1*-like gene from *Brassica napus*. *Plant Physiol. Biochem.* **48**, 9–15 (2010).
  70. Y. Guo, Y. Huang, J. Gao, Y. Pu, N. Wang, W. Shen, J. Wen, B. Yi, C. Ma, J. Tu, T. Fu, J. Zou, J. Shen, *CIPK9* is involved in seed oil regulation in *Brassica napus* L. and *Arabidopsis thaliana* (L.) Heynh. *Biotechnol. Biofuels* **11**, 124 (2018).
  71. Z. H. Zhu, A. Sami, Z. P. Chen, M. Fatima, W. Y. Zheng, Q. Q. Xu, Y. H. Lei, X. Z. Jin, H. Zhang, Y. Li, Y. Yu, K. J. Zhou, Effects of microscopic testa color and morphology on the water uptake ability and drought tolerance of germination-stage rapeseed (*Brassica napus* L.). *Bioengineered* **12**, 9341–9355 (2021).
  72. I. Debeaujon, K. M. Leon-Kloosterziel, M. Koornneef, Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Physiol.* **122**, 403–414 (2000).
  73. X. Wang, H. Wang, Y. Long, D. Li, Y. Yin, J. Tian, L. Chen, L. Liu, W. Zhao, Y. Zhao, L. Yu, M. Li, Identification of QTLs associated with oil content in a high-oil *Brassica napus* cultivar and construction of a high-density consensus map for QTLs comparison in *B. napus*. *PLOS ONE* **8**, e80569 (2013).
  74. J. Li, Y.-c. Xu, Z.-h. Wang, Construction of a high-density genetic map by RNA sequencing and eQTL analysis for stem length and diameter in *Dendrobium* (*Dendrobium nobile* × *Dendrobium wardianum*). *Ind. Crops Prod.* **128**, 48–54 (2019).
  75. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytzky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
  76. D. Liu, C. Ma, W. Hong, L. Huang, M. Liu, H. Liu, H. Zeng, D. Deng, H. Xin, J. Song, C. Xu, X. Sun, X. Hou, X. Wang, H. Zheng, Construction and analysis of high-density linkage map using high-throughput sequencing data. *PLOS ONE* **9**, e98855 (2014).
  77. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
  78. B. Wittkop, R. J. Snowden, W. Friedt, New NIRS calibrations for fiber fractions reveal broad genetic variation in *Brassica napus* seed quality. *J. Agric. Food Chem.* **60**, 2248–2256 (2012).
  79. G. Li, S. Xiuli, J. Liang, W. Gaoquan, X. Jiwei, W. Zelan, F. Tingdong, Establishment of math models of NIRS analysis for oil and protein contents in seed of *Brassica napus*. *Sci. Agric. Sin.* **36**, 1609–1613 (2003).
  80. S. Tang, H. Zhao, S. Lu, L. Yu, G. Zhang, Y. Zhang, Q. Y. Yang, Y. Zhou, X. Wang, W. Ma, W. Xie, L. Guo, Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. *Mol. Plant* **14**, 470–487 (2021).

81. K. Zhang, L. Nie, Q. Cheng, Y. Yin, K. Chen, F. Qi, D. Zou, H. Liu, W. Zhao, B. Wang, M. Li, Effective editing for lysophosphatidic acid acyltransferase 2/5 in allotetraploid rapeseed (*Brassica napus* L.) using CRISPR-Cas9 system. *Biotechnol. Biofuels* **12**, 225 (2019).
82. N. Nesi, M. O. Lucas, B. Auger, C. Baron, A. Lecureuil, P. Guerche, J. Kronenberger, L. Lepiniec, I. Debeaujon, M. Renard, The promoter of the *Arabidopsis thaliana* *BAN* gene is active in proanthocyanidin-accumulating cells of the *Brassica napus* seed coat. *Plant Cell Rep.* **28**, 601–617 (2009).
83. Y. Lei, L. Lu, H. Y. Liu, S. Li, F. Xing, L. L. Chen, CRISPR-P: A web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol. Plant* **7**, 1494–1496 (2014).
84. M. Li, Y. Haixia, M. Kang, P. An, X. Wu, H. Dang, X. Xu, The arachidonic acid metabolism mechanism based on UPLC-MS/MS metabolomics in recurrent spontaneous abortion rats. *Front. Endocrinol.* **12**, 652807 (2021).
85. Z. Yang, S. Wang, L. Wei, Y. Huang, D. Liu, Y. Jia, C. Luo, Y. Lin, C. Liang, Y. Hu, C. Dai, L. Guo, Y. Zhou, Q. Y. Yang, BnIR: A multi-omics database with various tools for *Brassica napus* research and breeding. *Mol. Plant* **16**, 775–789 (2023).
86. L. Reiser, E. Bakker, S. Subramaniam, X. Chen, S. Sawant, K. Khosa, T. Prithvi, T. Z. Berardini, The *Arabidopsis* information resource in 2024. *Genetics* **227**, iyae027 (2024).
87. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

**Acknowledgments:** We thank LetPub ([www.letpub.com.cn](http://www.letpub.com.cn)) for its linguistic assistance during the preparation of this manuscript. **Funding:** This study was supported by the National Key Research and Development Program of China (2023YFD1201401) and the National Natural Science Foundation of China (32272067 and 32072098). **Author contributions:** Conceptualization: M.L. and Y.Y. Methodology: H.L., H.C., X.W., and Y.Xiong. Investigation: S.F., K.C., S.Y., and Y.D. Visualization: Y.Xia, X.C., and J.H. Supervision: M.L., C.Z., H.J., and L.Z. Writing—original draft: H.L. and M.W. Writing—review and editing: M.L. and H.L. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The raw data of the RNA-seq analysis of the KN DH lines have been deposited in the Sequence Read Archive (SRA) Database in the National Center for Biotechnology Information (NCBI) ([www.ncbi.nlm.nih.gov/sra/](http://www.ncbi.nlm.nih.gov/sra/)) with the accession number PRJNA1103768. The raw data of the RNA-seq analysis of NIL and transgenic materials have been deposited in the SRA Database in NCBI with the accession number PRJNA1049609. The Python code for RC.eQTL generation has been posted on GitHub (<https://github.com/alex010502/random-eQTL-generator>) and Zenodo (<https://doi.org/10.5281/zenodo.14044435>).

Submitted 27 August 2024  
Accepted 27 November 2024  
Published 3 January 2025  
10.1126/sciadv.ads7620