

RESEARCH ARTICLE

# Genome-Wide Analysis of Seed Acid Detergent Lignin (ADL) and Hull Content in Rapeseed (*Brassica napus* L.)

Jia Wang<sup>1</sup>✉, Hongju Jian<sup>1</sup>✉, Lijuan Wei<sup>1</sup>, Cunmin Qu<sup>1</sup>, Xinfu Xu<sup>1</sup>, Kun Lu<sup>1</sup>, Wei Qian<sup>1</sup>, Jiana Li<sup>1</sup>, Maoteng Li<sup>2</sup>, Liezhao Liu<sup>1\*</sup>

**1** College of Agronomy and Biotechnology, Southwest University, Beibei, Chongqing, China, **2** Institute of Resource Biology and Biotechnology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China

✉ These authors contributed equally to this work.

\* [liezhao2003@126.com](mailto:liezhao2003@126.com)



**OPEN ACCESS**

**Citation:** Wang J, Jian H, Wei L, Qu C, Xu X, Lu K, et al. (2015) Genome-Wide Analysis of Seed Acid Detergent Lignin (ADL) and Hull Content in Rapeseed (*Brassica napus* L.). PLoS ONE 10(12): e0145045. doi:10.1371/journal.pone.0145045

**Editor:** Harsh Raman, NSW Department of Primary Industries, AUSTRALIA

**Received:** August 6, 2015

**Accepted:** November 29, 2015

**Published:** December 16, 2015

**Copyright:** © 2015 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** Funding provided by National Natural Science Foundation of China, (31171584), pr. Liezhao Liu, Conceived and designed the experiments, analyzed the data, National Natural Science Foundation of China (CN), (31371655), pr. Liezhao Liu, 973 Program, (2015CB150201), Pr. Wei Qian, Contributed reagents/analysis tools and Earmarked Fund for Modern Agro-industry Technology Research System, (CARS-13), pr. Jiana Li, Contributed reagents/materials.

## Abstract

A stable yellow-seeded variety is the breeding goal for obtaining the ideal rapeseed (*Brassica napus* L.) plant, and the amount of acid detergent lignin (ADL) in the seeds and the hull content (HC) are often used as yellow-seeded rapeseed screening indices. In this study, a genome-wide association analysis of 520 accessions was performed using the Q + K model with a total of 31,839 single-nucleotide polymorphism (SNP) sites. As a result, three significant associations on the *B. napus* chromosomes A05, A09, and C05 were detected for seed ADL content. The peak SNPs were within 9.27, 14.22, and 20.86 kb of the key genes *BnaA.PAL4*, *BnaA.CAD2/BnaA.CAD3*, and *BnaC.CCR1*, respectively. Further analyses were performed on the major locus of A05, which was also detected in the seed HC examination. A comparison of our genome-wide association study (GWAS) results and previous linkage mappings revealed a common chromosomal region on A09, which indicates that GWAS can be used as a powerful complementary strategy for dissecting complex traits in *B. napus*. Genomic selection (GS) utilizing the significant SNP markers based on the GWAS results exhibited increased predictive ability, indicating that the predictive ability of a given model can be substantially improved by using GWAS and GS.

## Introduction

Increasing seed oil content has always been an important goal in the breeding of *B. napus*. Research has demonstrated that the yellow-seeded rapeseed is a desirable variety because of its quantities of oil and protein compared with the black-seeded rapeseed [1] as well as the association of the yellow seed color with a thinner seed coat [2], lower fiber content, and reduced amounts of anti-nutrients (lignin) [3]. Moreover, these characteristics make rapeseed meal a valuable source of high-quality feed. Therefore, the selection of a stable yellow-seed trait is one of the most essential breeding objectives for this crop. However, the development of yellow-

**Competing Interests:** The authors have declared that no competing interests exist.

seeded *B. napus* varieties with increased seed oil content and improved canola meal quality (i.e., lower fiber and higher protein contents) have been limited due to insufficient understanding of the genetic mechanisms underlying the formation of seed coat color in this species [4] and to the considerable influence of the environment on the yellow-seed trait [5]. Therefore, the reduced fiber content and thinner seed coat of the yellow-seeded varieties are considered important screening indices for breeding this trait.

Genome-wide association study (GWAS) is rapidly becoming the dominant paradigm for investigating the genetic basis of natural phenotypic variations. Although GWAS have been primarily used for research on human diseases, these studies have also been successful in mapping causal variants in many other organisms [6], including crop plants. The application of GWAS to detect quantitative trait loci (QTLs) controlling complex traits has become a popular approach for studying key characteristics of crop plants [7], such as rice [8], maize [9], wheat [10], barley [11], and potato [12]. Association mapping has also been effectively performed in *B. napus*. Thus far, due to the complexity of its genomic structure and the lack of high-quality molecular markers, the population structure and linkage disequilibrium (LD) of rapeseed (*B. napus* L.) are not well understood compared to those of other crop species [13]. The first association studies based on traits and markers in *B. napus* began in 2008. Based on orthologs of candidate genes for glucosinolate biosynthesis in *Arabidopsis*, 51 gene-linked simple sequence repeat (SSR) alleles were found to be associated with seed glucosinolate content in two sets of *B. napus* germplasms, with 94 and 46 genotypes separately [14]. Subsequently, numerous *B. napus* association analyses have been implemented [15–24]. The findings of these investigations indicate that it is feasible to develop molecular breeding markers related to agronomic traits found in association studies in order to analyze the genetic basis of complex agronomic traits in *B. napus*. With the development of the GWAS research method, the genome sequence of *B. napus* was completed [25], and the technology was improved to dramatically reduce the cost of the high-throughput single-nucleotide polymorphism (SNP) chip, thus making the GWAS method popular in rapeseed breeding research.

Light seed color and low fiber content are thought to coincide because the biochemical pathways leading to lignin and pigment synthesis have common precursors, such as *p*-cumarate [26]. Moreover, a correlation between seed color and fiber composition has been observed, particularly with the amount of acid detergent lignin (ADL) [27–29]. The lower seed hull content (HC) in yellow-seeded *B. napus* is partially attributed to reduced fiber quantity and the reduction of the palisade layer to half to two-thirds of its thickness compared with that of black-seeded varieties [30]. A major quantitative trait locus (QTL) influencing seed ADL content has been mapped to chromosome A9 of *B. napus* [27–29,31]. The *Bna.CCR1* homolog may interact with neighboring homologs of a *CAD* gene family member known to regulate seed-coat phenylpropanoid biosynthesis in *B. napus*. To date, only a few QTL reports, including some from our lab, have focused on seed HC [32–34], which is likely to be controlled by a number of genes and is therefore appropriate for QTL analyses [26]. The biosynthesis of lignin and its phenylpropanoid precursors in xylem and stem tissues has been studied extensively in numerous model and crop plants. In contrast, seed-coat phenylpropanoid biosynthesis is less understood, although seeds from many crops play a huge role in both livestock feeding and human nutrition [28]. In this study, we used a collection of 520 cultivars for association mapping and performed a GWAS for seed ADL content and HC. The aim of this work was to identify SNPs that are significantly associated with the quantity of seed ADL content and HC in mapping populations, to compare the findings with those from previous studies concerning the QTL mapping of seed ADL content [27–29,31], and to identify for candidate genes governing potentially important traits for improved *B. napus* breeding.

## Materials and Methods

### Field experiment and traits measurements

A total of 520 *B. napus* germplasm resources (98 yellow-seeded, 201 black-seeded, and 221 with intermediate colors) (S1 Table), collected from China spring and semi-winter accessions, were cultivated under natural growing conditions in the experimental farm of the Chongqing Engineering Research Center for Rapeseed, Southwest University in Beibei, Chongqing, China (106.40°E, 29.80°N) for two consecutive years. The accessions were arranged in a randomized complete block design with three replicates. In the growing periods from September 2012 to May 2013 (referred to as 2013) as well as September 2013 to May 2014 (referred to as 2014), each line was planted in two rows of 10 plants per row, with 30 cm between rows and a distance of 20 cm between plants within each row. Open pollinated seeds were collected from five randomly-chosen plants in each line at maturity for seed HC and NIRS analysis.

The determination of seed HC was performed as described by Dimov *et al.* [35]. Seed hulls were separated from the embryos using a dissecting needle and tweezers after the seeds were imbibed in water for 24 h. Both fractions were dried at  $105\pm 2^{\circ}\text{C}$  for 2 h before their dry weights were measured. The seed HC (in %) per sample of 50 seeds (approx. 200 mg) was determined. In addition, 13HC and 14HC represent the seed HC in 2013 and 2014, respectively.

The quantity of ADL in the seeds was estimated using near-infrared reflectance spectroscopy (NIRS) with an NIR System 6500 and the WinISI II software (FOSS GmbH, Rellingen, Germany). Before scanning, the system was warmed up for at least 50 min. Then, the naturally dried samples were transferred into the sample cup. The sample was pressed gently into the sample table of the spectrometer and then scanned. The spectra between 1,100 and 2,498 nm were recorded, registering  $\log(1/R)$  absorbance values at 2-nm intervals for each sample. The phenotypic values of the ADL content (% seed dry mass) were extrapolated from NIR spectra using NIR calibrations developed in our lab specifically for the measurement of these traits (fiber content) in *B. napus*. The standard error of the calibration (SEC), the R-squared value (RSQ), the standard error of cross validation (SECV), the standard error of prediction corrected for bias SEP(c), and the value of 1 minus the variance ratio (1-VR) in this NIR calibration were 0.195, 0.979, 0.227, 0.172, and 0.971, respectively. The NIR-derived estimates for seed ADL content were averaged over three technical replicates.

### Genome-wide association analysis

In these lines, 5,000 SNPs (minor allele frequency [MAF]  $\geq 0.2$ ) were utilized to evaluate the population structure, and the 520 rapeseed lines were classified into two groups according to the results of the structure analysis. The relative kinship analysis revealed that the population of *B. napus* had a null or weak relationship, with greater than 80.9% pairwise relationship estimates between lines; values less than 0.05 and 55% were considered equal to 0. The level of LD in the rapeseed panel was low, with a distance of LD decays within 0.5 to 1 cM at the genomic level.

A trait—SNP association analysis was performed using the Q + K model with a total of 31,839 SNP sites (miss data < 20%, MAF  $\geq 0.05$ ); the Q + K model was implemented via a mixed linear model (MLM) [36,37] by a variance component estimation in TASSEL 5.1 [38]. To avoid the excessive correction from Bonferroni and relatively loose correction from 5% FDR [39], we performed a correction for multiple hypothesis testing by controlling the false discovery rates (FDRs) of 1% in this study. All *p*-values of the association analysis results were ranked from small to large. False discovery rates (FDRs) were calculated as  $[(m \times P)/n] \times 100\%$  [40], where *m* is the total number of SNPs, *P* is the *p*-value threshold for detecting significant

association, and  $n$  is the total number of significant associations per trait. If  $p < 0.01$  after the rectification, then the association between a SNP locus and a target trait was considered significant. We used the R package qqman [41] for Manhattan plots and quantile-quantile plots.

Haplotype blocks were constructed via the four-gamete rule with Haploview 4.2 [42]. The parameters were set as follows: the Hardy-Weinberg  $p$ -value cutoff was 0.001, the minimum genotype was 75%, the MAF was 0.05, and the maximum number of Mendel errors was 1.

### LD analysis for QTL determination interval

As described by Lund *et al.* [43], the analysis method mainly estimated the significant SNPs. In our populations, the mean linkage disequilibrium ( $r^2$ ) between adjacent SNPs was 0.23, and the median was 0.09.  $LD > 0.1$  was used as the determination standard. Each QTL determination interval was recorded as an R-QTL. The LD value between the first left SNP and the significant SNP was computed, with the significant SNP utilized as a reference standard. If  $LD < 0.1$ , no further calculations were performed, and this SNP was accepted as the left border of the R-QTL. If  $LD > 0.1$ , then the LD value between the next SNP and the significant SNP was calculated. The analysis continued until the LD of a SNP and the significant SNP reached  $< 0.1$ , at which point the SNP was accepted as the left border of the R-QTL. The same operation was performed for the right borders of the R-QTLs. When two significant SNPs were adjacent, the SNP on the left side was associated with the left R-QTL, whereas the SNP on the right side was associated with the right R-QTL based on the approach applied to determine the R-QTL interval boundaries of the left and right sides, respectively.

### Genome-wide prediction of seed ADL content and HC

We utilized the R package rrBLUP [44] for genome-wide prediction using ridge regression. To assess the effects of the varying sizes of the training populations, we randomly employed 40 to 80% of the experimental materials. To evaluate the potential of improving the performance of genome-wide prediction by selecting the most informative markers, we compared the predictive abilities of all SNP markers. An equal number of significant SNP markers were chosen randomly, and the significant SNP markers were selected based on their GWAS significance within the training population [45].

### Statistical analysis

Analyses of the means and coefficients of variation and the correlation of the evaluated traits were performed with SPSS 17.0. The effects of genotype, environment, and genotype by environment interaction ( $G \times E$ ) on phenotypic variation were assessed using PROC GLM of SAS 9.2. Upon removal of the outliers, a previously described procedure [46] was implemented to identify the optimal transformation of each trait, thus ensuring that the model assumptions of normally distributed error terms and constant variance were not violated [47]. The normality test of phenotypic traits using the UNIVARIATE procedure of SAS 9.2 was completed, and the Box-Cox transformation of the phenotypic data that did not adhere to a normal distribution was performed with Minitab 16.0.

## Results

### Phenotypic data analysis

Seed ADL content and HC of the *B. napus* association panel comprising 520 accessions was measured in three replicates in two consecutive years. As presented in Table 1, extensive phenotypic variations were observed in the descriptive statistics. The seed ADL content, which

**Table 1. Phenotypic variations in seed ADL content (ADL) and seed hull content (HC) in the *B. napus* panel.**

Traits	Range	Mean ±SD	CV(%) <sup>a</sup>	G	E	G×E	<i>h</i> <sup>2</sup> <sup>b</sup>
13ADL	0.85–7.66	4.42±2.16	48.87	**	**	**	88.62%
14ADL	0.44–8.07	4.37±2.32	53.10				
13HC	9.40–25.50	14.67±4.38	29.86	**	**	**	75.23%
14HC	9.29–20.83	14.81±4.05	27.35				

\*\* The values are significant at *P* < 0.01 for the effect of genotype (G), environment (E) and genotype by environment interaction (G×E) on phenotypic variance estimated by two-way ANOVA.

<sup>a</sup> CV is an abbreviation of coefficient of variation, which was estimated as the ratio of the standard deviation to the mean of all accessions.

<sup>b</sup> *h*<sup>2</sup> is broad-sense heritability;  $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{ge}^2 / n + \sigma_e^2 / nr) \times 100\%$ , where  $\sigma_g^2$  is the genetic variance,  $\sigma_{ge}^2$  is the variance due to the G × E interaction,  $\sigma_e^2$  represents the residual error, *n* is the number of environments (years), and *r* is number of replicates [48].

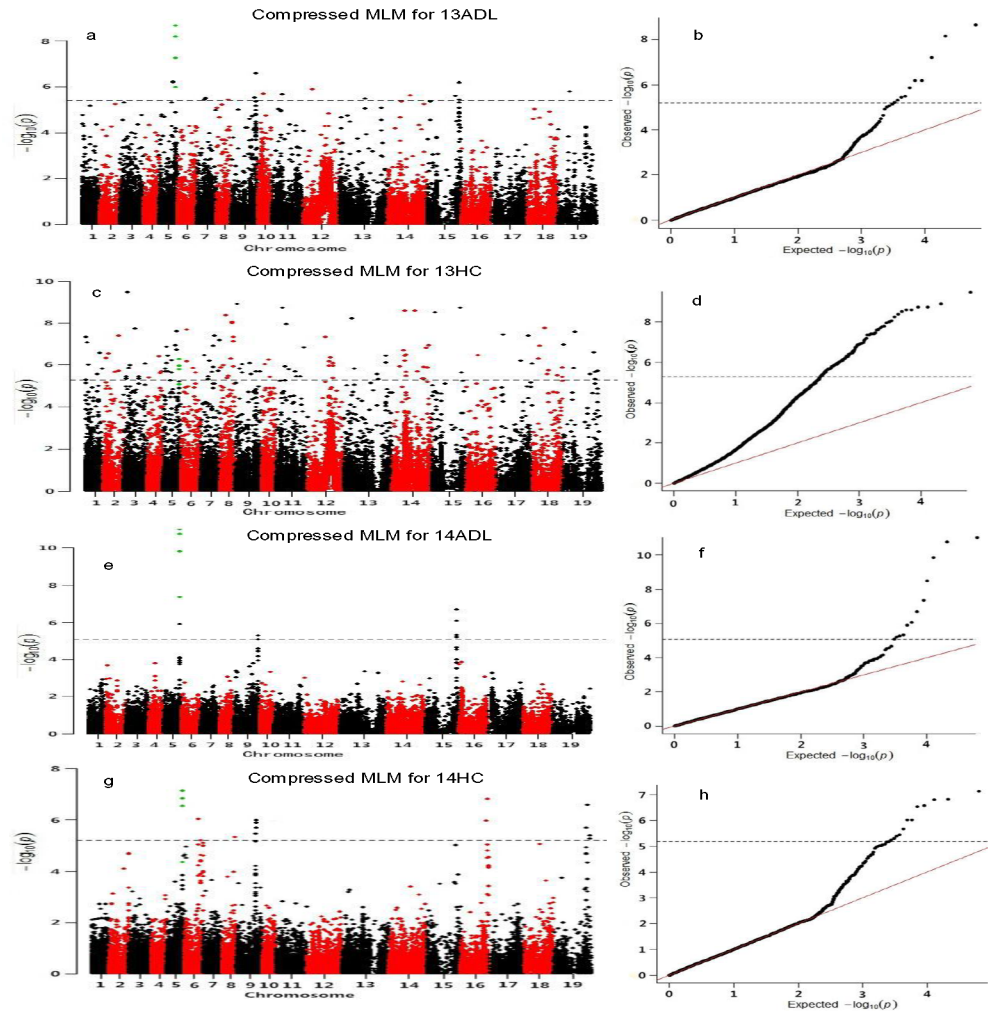
doi:10.1371/journal.pone.0145045.t001

ranged from 0.44 to 8.07 with an average of 4.37, had a maximum coefficient of variation of 53.10% in 2014. Conversely, the seed HC in 2014, which varied from 9.29 to 20.83 with an average of 14.81, exhibited the lowest coefficient of variation (27.35%). Seed ADL content and HC reached an extremely significant positive correlation in 2013 and 2014; their coefficients were 0.778 and 0.632, respectively (S1 Table). Seed ADL content and HC in 2013 and 2014 manifested a significant positive correlation with correlation coefficients of 0.567 and 0.467, respectively (S2 Table). A two-way analysis of variance (ANOVA) was performed on seed ADL content and HC using the SAS 9.2 software for the GWAS population. The genotype (G), environment (E), and genotype by environment interaction (G × E) exhibited significant effects on all these traits (*p* < 0.01) (Table 1 and S3 Table). The results indicate that seed ADL content was more stable than seed HC in both 2013 and 2014 because their broad-sense heritability coefficients were *h*<sup>2</sup> = 88.62% and 75.23%, respectively.

Upon removal of the outliers, a four-trait normality test was conducted using the SAS 9.2 UNIVARIATE procedure, and the normal distribution was measured across skewness and kurtosis. The analysis results indicate that 13ADL (skewness: -0.279; kurtosis: -0.024) and 14ADL (skewness: -0.021; kurtosis: -0.247) conformed to the normal distribution, but 13HC and 14HC did not. Thus, we first performed a Box-Cox data transformation of the two phenotypes (2013HC and 14HC) using the Minitab 16.0 software for GWAS to adhere to or approximate a normal distribution in the next step. The software applied the data transformation depending on the lambda value, and the obtained best Box-Cox transformation lambda values of 13HC and 14HC were 0.16 and 0.77, respectively. Through the Box-Cox transformation, 13HC and 14HC reached normal or obviously superior levels than those before the raw data transformation (S1 Fig).

### Genome-wide association studies for seed ADL content and HC

Using a unified mixed linear model that controlled the population structure and relative kinship, the removal of low-quality SNPs and those with a MAF ≥ 0.05 yielded 31,839 SNP data sets for association analysis. Significant SNP—trait associations at a 1% FDR were found for four traits: 13ADL, 13HC, 14ADL, and 14HC. The Manhattan plots of these traits are presented in Fig 1. The compressed MLM approach, which considered the genome-wide patterns of genetic relatedness, greatly reduced false positives, as illustrated in the quantile-quantile plots (Fig 1b, 1d, 1f and 1h). As detected by the Q + K models, in the ‘pseudomolecules’ of *B. napus* (Fig 1a, Table 2 and S4 Table), the seed ADL content (13ADL and 14ADL) was associated with three common significant regions located at 20.2 Mb, 29.8 Mb, and 40.2 Mb in the A05, A09, and C05 chromosomes, respectively. The peak SNP locus on A05 (rs11682;



**Fig 1. Genome-wide association studies of seed ADL content (ADL) and seed hull content (HC).** Manhattan plots of the compressed MLMs for ADL and HC. Negative  $\log_{10}$ -transformed  $p$ -values from a genome-wide scan are plotted against position on each of the 19 chromosomes. The black horizontal dashed line indicates the genome-wide significance threshold, and the green marker is the simultaneously detected locus. (a): 13ADL, (c): 13HC, (e): 14ADL, (g): 14HC. Quantile-quantile plot of the compressed MLMs for ADL and HC, (b): 13ADL, (d): 13HC, (f): 14ADL. (h): 14HC.

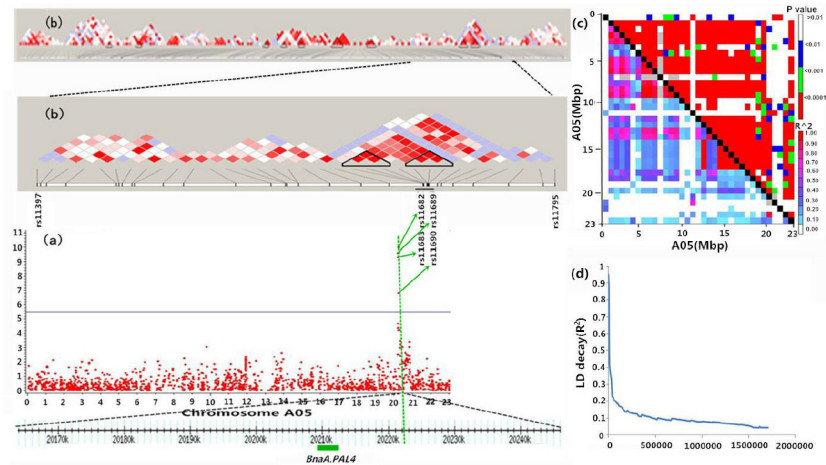
doi:10.1371/journal.pone.0145045.g001

20,222,542 bp;  $p$ -values was  $2.1 \times 10^{-9}$  and  $1.75 \times 10^{-11}$  in 2013 and 2014, respectively) was 9.27 kb from the key lignin biosynthesis gene *BraA.PAL4* (Fig 2 and Table 3). This peak SNP locus (rs11682) explained 9.4% and 13.9% of the total phenotypic variance for 13ADL and 14ADL, respectively, based on  $R^2$  values. Regarding seed HC, a total of 150 and 17 association signals were identified during the two experimental years with a 1% FDR from the compressed MLM, respectively (Table 2, S4 and S5 Tables). Three common significant regions for 2013 and 2014 were located at 20.2 Mb, 16.5 Mb, and 33.0 Mb in chromosomes A05, A08, and A09, respectively. The 150 association signals in 2013 were distributed across all chromosomes, and strong association signals with  $p < 5.0 \times 10^{-9}$  were noted in the physical regions of 4.2 Mb of A03, 2.1 Mb of A09, 7.5 Mb of C01, 35.4 Mb of C05, 14.3 Mb of C04, and 7.3 Mb of A09, accounting for 10.5%, 9.9%, 9.75%, 8.9%, 9.5%, and 9.3% of the phenotypic variation, respectively (S5 Table).

**Table 2. Genome-wide significant association signals of seed ADL content and HC.** Only 13ADL, 14ADL and 14HC are listed in this table. The significant association signals of 13HC are presented in [S5 Table](#).

Trait	Chromosome	Position	Major allele	Minor allele	Minor allele frequency	p-value	Contribution (%)
13ADL	A05	3207113	G	A	0.45	1.61×10 <sup>-06</sup>	6.369
	A05	16983454	G	A	0.35	6.28×10 <sup>-07</sup>	6.8
	A05	20222542	T	C	0.43	2.21×10 <sup>-09</sup>	9.409
	A05	20222599	C	T	0.44	6.73×10 <sup>-09</sup>	8.892
	A05	20253712	A	G	0.42	5.97×10 <sup>-08</sup>	7.882
	A09	29882802	T	G	0.30	2.90×10 <sup>-06</sup>	6.973
	A09	30668239	A	G	0.41	6.27×10 <sup>-06</sup>	6.666
	A09	31262997	A	G	0.15	2.78×10 <sup>-06</sup>	6.12
	A10	6685826	G	T	0.34	2.02×10 <sup>-06</sup>	6.266
	C01	12655116	G	A	0.19	2.07×10 <sup>-06</sup>	6.254
	C02	12309497	G	A	0.20	1.32×10 <sup>-06</sup>	6.459
	C04	27480077	C	A	0.25	2.44×10 <sup>-06</sup>	5.528
	C05	40211108	C	T	0.31	6.51×10 <sup>-07</sup>	6.111
	C09	15755706	A	C	0.39	1.57×10 <sup>-06</sup>	6.381
	14ADL	A05	20222542	T	C	0.43	1.75×10 <sup>-11</sup>
A05		20222599	C	T	0.44	1.53×10 <sup>-10</sup>	12.693
A05		20252156	A	G	0.40	4.37×10 <sup>-08</sup>	9.535
A05		20253712	A	G	0.42	1.01×10 <sup>-11</sup>	14.232
A05		20293807	A	C	0.44	1.28×10 <sup>-06</sup>	6.428
A09		30752816	C	T	0.10	5.18×10 <sup>-06</sup>	5.701
A09		29882802	T	G	0.30	7.28×10 <sup>-07</sup>	7.386
C05		39839348	C	T	0.37	4.71×10 <sup>-06</sup>	5.747
C05		40125110	C	T	0.23	4.70×10 <sup>-06</sup>	6.973
C05		40132442	T	C	0.23	8.36×10 <sup>-07</sup>	7.914
C05		40211108	C	T	0.31	1.96×10 <sup>-07</sup>	7.886
C05		40336201	T	C	0.37	6.33×10 <sup>-06</sup>	6.811
14HC	A05	20222542	T	C	0.43	7.42×10 <sup>-08</sup>	7.768
	A05	20222599	C	T	0.44	1.49×10 <sup>-07</sup>	7.447
	A05	20252156	A	G	0.40	1.48×10 <sup>-06</sup>	8.635
	A05	20253712	A	G	0.42	2.86×10 <sup>-07</sup>	7.148
	A06	17448325	T	C	0.13	9.32×10 <sup>-07</sup>	6.924
	A06	20553941	A	C	0.45	6.52×10 <sup>-06</sup>	6.049
	A08	16530504	A	C	0.09	4.85×10 <sup>-06</sup>	5.854
	A09	24696779	T	C	0.32	3.56×10 <sup>-06</sup>	6.327
	A09	29882802	T	G	0.30	7.28×10 <sup>-06</sup>	6.386
	A09	33496928	T	C	0.29	1.26×10 <sup>-06</sup>	6.531
	A09	33029776	T	G	0.27	9.34×10 <sup>-07</sup>	6.606
	C06	33566777	T	G	0.21	8.64×10 <sup>-07</sup>	6.453
	C06	35525816	C	T	0.06	1.57×10 <sup>-07</sup>	7.421
	C09	41831166	T	C	0.07	2.11×10 <sup>-06</sup>	6.438
	C09	44555029	A	C	0.21	2.63×10 <sup>-07</sup>	7.138
	C09	46608012	G	A	0.08	5.68×10 <sup>-06</sup>	6.172
	C09	46931937	T	A	0.07	4.16×10 <sup>-06</sup>	6.275

doi:10.1371/journal.pone.0145045.t002



**Fig 2. Genome-wide association scan for seed ADL content and the genomic landscape of the major locus on A05.** (a) Association signals of ADL (2014) values on A05. The top of the panel shows an R-QTL region based on its significant SNPs, the positions of which are indicated by vertical green lines. Negative log<sub>10</sub>-transformed *p*-values from the compressed MLM are plotted on the vertical axis. The blue horizontal line indicates the 1% FDR-adjusted significance threshold ( $6.27 \times 10^{-6}$ ); the bottom of the panel indicates the related candidate genes marked with green boxes in the R-QTL region. One previously identified gene, *PAL4*, was significantly associated with seed ADL content. (b) The distribution of the linkage disequilibrium (LD) blocks of the major locus on A05. (c) LD analysis of A05. (d) The LD decay of A05.

doi:10.1371/journal.pone.0145045.g002

The loci on A05 were repeatedly detected in 2013 and 2014 both for seed ADL content and HC, indicating that a major locus responsible for seed ADL content may be located in this genomic region (Fig 1). LD analysis of A05 was performed using 1,615 SNPs to achieve a comprehensive view of the genomic landscape of this major locus. The LD of A05 over physical distances is depicted in Fig 2. The linkage disequilibrium blocks were unevenly distributed on A05, and the peak SNP associated with seed ADL content was located at one end of A05. The

**Table 3. Candidate genes within R-QTL of SNPs most highly associated with seed ADL content.**

SNP(s)	Chromosome	Physical interval	Candidate gene(s)
rs11682 rs11683	A05	20169332..20243410	BnA.SEC8 BnaA.PAL4 BnaA05g28490D BnaA.CESA3 BnaA.GPAT5
rs11690 rs11689	A05	20243410..20265627	BnaA05g28570D
rs11282	A05	16961778..17077909	BnaA05g22190D
rs12071	A05	3195182..3222564	BnaA05g06020D BnaA05g06000D
rs11697	A05	20279439..20340965	BnaA.CCR1
rs21364	A09	31213064..31297302	BnaA09g45870D
rs21157	A09	29790101..29917465	BnaA.CAD2 BnaA.CAD3
rs21223	A09	30726019..30755421	CHS1 BnaA09g44870D
rs23199	A10	6630309..6686572	BnaA10g08080D BnaA10g08070D
rs35374	C01	12583623..12804840	BnaC01g18220D BnaC01g18230D
rs29646	C02	12058887..12315563	BnaC02g16830D
rs42895	C04	27304035..27517912	BnaC04g26130D BnaC04g26090D
rs37338 rs37335 rs37333	C05	40099068..40267266	BnaC.CCR1 BnaC05g43300D
rs35157	C05	35235636..35317212	BnaC05g36040D BnaC05g36080D BnaC05g36150D
rs37369	C05	39828480..39936267	BnaC05g42520D BnaC05g42570D BnaC05g42630D
rs37408	C05	40322746..40371177	BnaC05g43530D BnaC.MYB83
rs39358	C09	15753918..15758520	BnaC09g18860D

doi:10.1371/journal.pone.0145045.t003



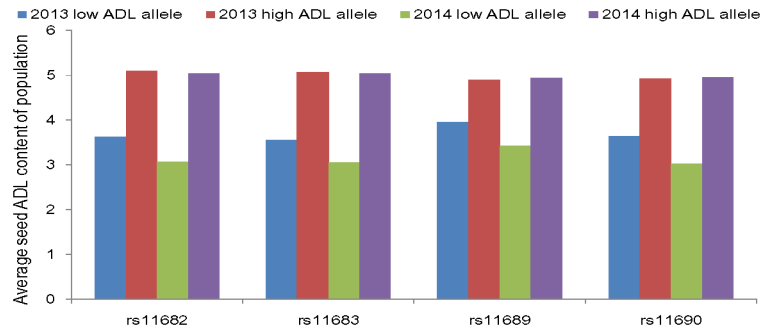
four peak SNPs were located in a similar position on A05 but were not in the same LD block (Fig 2b). Therefore, understanding the LD decay of the A05 genomic region would be facilitated by the determination of candidate genes and beneficial haplotypes. The decay of LD with a physical distance between SNPs occurred at 90 kb on A05 ( $r^2 = 0.2$ ) (Fig 2c). The first and third peak SNPs (rs11682 and rs11683) were involved in a 74-kb LD block that encompassed five SNP markers, but the second and fourth peak SNPs (rs11690 and rs11689) were incorporated in a 22-kb LD block that consisted of three SNP markers. This result suggests that the key lignin biosynthesis gene linked with SNPs in A05 might be the major genetic locus responsible for natural variation in rapeseed ADL content. Accessions with a cytosine (C) allele at the first peak SNP (rs11682) manifested, on average, a 40.5% and 64.1% reduced seed ADL content compared with accessions with a thymine (T) allele in 2013 and 2014, respectively. The minor allele (C) was represented in 43% of the 520 accessions. Accessions with a G allele at the second peak SNP (rs11690) had, on average, 45.7% and 62.9% reduced seed ADL content compared with accessions with an adenine (A) allele in 2013 and 2014, respectively. The minor allele (G) was represented in 42% of the 520 accessions. Accessions with a cytosine (C) allele at the third peak SNP (rs11683) had, on average, 42.9% and 64.6% reduced seed ADL content compared with accessions with a thymine (T) allele in 2013 and 2014, respectively. The minor allele (T) was represented in 44% of the 520 accessions. Accessions with an A allele at the fourth peak SNP (rs11689) had, on average, 24.1% and 44.1% reduced seed ADL content compared with accessions with a guanine (G) allele in 2013 and 2014, respectively. The minor allele (G) was represented in 40% of the 520 accessions (Table 2 and Fig 3).

### LD analysis for the QTL determination interval and candidate genes potentially related to seed ADL content

We identified the QTL determination interval using the significant association signals of seed ADL content with known key genes of lignin biosynthesis, and a total of 17 R-QTLs were identified (Table 3). The R-QTLs were 4.60 to 256.68 kb in length with an average length of ~95.73 kb in chromosomal regions near the identified loci with an LD decay distance of 0.5 Mb ( $r^2 = 0.1$ ). This average determination interval was much smaller than that delineated by traditional linkage mapping. The major locus on A05 was divided into two R-QTLs. The 74.08-kb QTL determination interval contained two peak SNPs (rs11682 and rs11683), whereas the other R-QTL encompassed two additional peak SNPs (rs11690 and rs11689) spanning a 22.22-kb determination interval, which was identical to the result for the LD blocks. Three additional R-QTLs were noted on A05. Within these R-QTL regions bracketed by significant SNPs, a total of 32 potential candidate genes related to the trait were noted (Table 3). The peak SNP contained two identical genotype SNPs (rs11682 and rs11683) on A05 located in gene *BnaA05g28490D*, the product of which is involved in microtubule-based movement and aligns most closely with *AT3G10310* from *Arabidopsis thaliana*. The key lignin biosynthesis gene *BnaA.PAL4* was 9.27 kb from the peak SNP rs11682. In addition, various candidate genes responsible for lignin biosynthesis or seed-coat development were identified in the R-QTL of the peak SNP rs11682. *BnaA.CCR1* was identified in another R-QTL associated with the significant SNP rs11697. Moreover, the peak SNPs rs37333, rs37335, and rs37338 on C05 were located within 168.20 kb of an R-QTL, and *BnC.CCR1* was 20.86 kb from rs37338 (Table 3 and S2a Fig).

### The major locus on A09 for seed ADL content detected via GWA mapping compared with that determined by QTL mapping

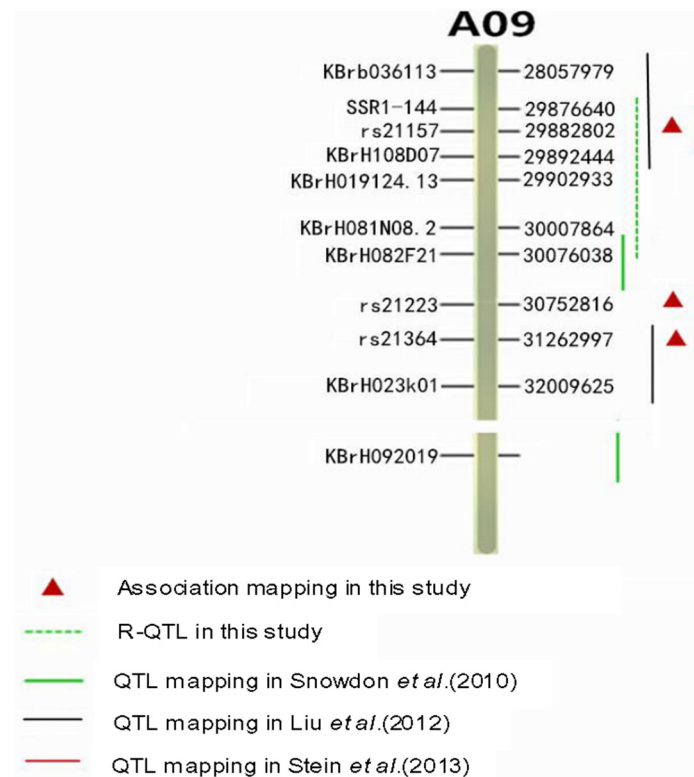
To compare the positions determined by GWA mapping and QTL mapping, we used common markers from previous studies to anchor the latest *B. napus* genome (v4.1) via local BLAST



**Fig 3. Average seed ADL content of accessions carrying the same allele of the significant SNPs on A05 in *B. napus*.**

doi:10.1371/journal.pone.0145045.g003

searching. The results are presented in Fig 4. In total, five QTLs from QTL mapping and three signals from GWA mapping associated with seed ADL content distributed on chromosome A09 were analyzed. The peak SNP rs21157 was 14.22 kb from *BnaA.CAD2/BnaA.CAD3* in the marker intervals of *ssr1-144* and *KBrH108D07*, where they overlapped with the QTL that was reported by Snowdon *et al.* [27] and Stein *et al.* [31] based on seed ADL content (S2b Fig). However, the marker *KBrH092019*, which encompasses the key lignin biosynthesis gene *BnaA.CCR1* and was communicated by Liu *et al.* [28,29] and Stein *et al.* [31] was anchored on a scaffold on A09.



**Fig 4. Overlapping or linkage relationships among seed ADL content quantitative trait loci (QTLs) in this and previous studies.**

doi:10.1371/journal.pone.0145045.g004

In addition, our R-QTLs overlapped with previous QTLs on A09. The R-QTL that contained *BnaA.CAD2/BnaA.CAD3* located in the determination interval of the peak SNP rs21157 overlapped with previous QTLs reported by Snowdon *et al.* [27] and Stein *et al.* [31] and was only 90.34 kb from the QTL described by Liu *et al.* [28,29]. These results indicate that our methods using GWA mapping generate clear association signal intervals that potentially enable the screening of candidate genes.

### Loci associated with seed HC

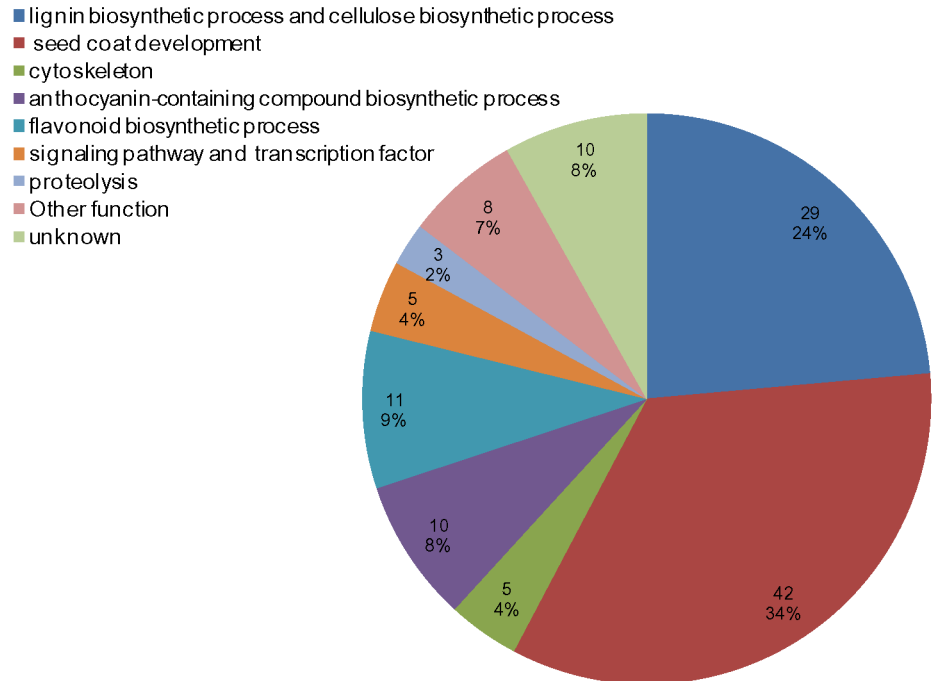
Although 167 loci were highly significantly associated with seed HC in 2013 and 2014, only 4 loci were detected in both years. Given that the genotype and the environment together affect seed HC, this result makes biological sense because the latter had a relatively small value of broad-sense heritability ( $h^2 = 75.23\%$ ). To evaluate whether GWA mapping performed high-efficiency detection, we filtered candidate genes for each identified locus. All potential candidate genes 250 kb upstream or downstream of the lead SNPs ( $r^2 = 0.1$ ) of the loci are listed in [S6 Table](#). Among these 123 filtered candidate genes, 68% of these genes associated with seed-coat development and fiber compound biosynthesis, and 17% of these genes were related to the phenylpropanoid pathway of flavonoid biosynthetic processes and anthocyanin accumulation ([Fig 5](#)), indicating that our approach can detect candidate genes related to seed HC with a relatively high detection efficiency.

### Genomic selection for seed ADL content and HC

In our study, genomic selection was performed using phenotypic and SNP genotype data. As depicted in [Fig 6](#), we established the choice model with a random 40 to 80% fraction of the experimental materials and all of the 31,839 SNP markers. The predictive ability and standard deviation increased with the size of the reference population. A higher predictive ability and a lower standard deviation were optimal; therefore, 60% of the reference population was selected based on both predictive abilities and standard deviation. All markers were used for the prediction, and the predictive abilities were only 0.27 and 0.36 in seed ADL content and HC, respectively, indicating that not all markers were suitable for efficient prediction. Then, we selected significant association markers identified by GWAS whose predictive abilities were 0.8 for seed ADL content and 0.64 for seed HC, and an equal number of the significant association markers were randomly selected and recorded. The predictive efficiency of the randomly selected associate markers was low, and they manifested a reduced ability to predict seed ADL content and HC. The findings of this examination indicate that significant association markers have a higher predictive ability for seed ADL content and HC. In conclusion, the significant loci detected in GWAS improved the predictive ability for seed ADL content and HC.

### Discussion

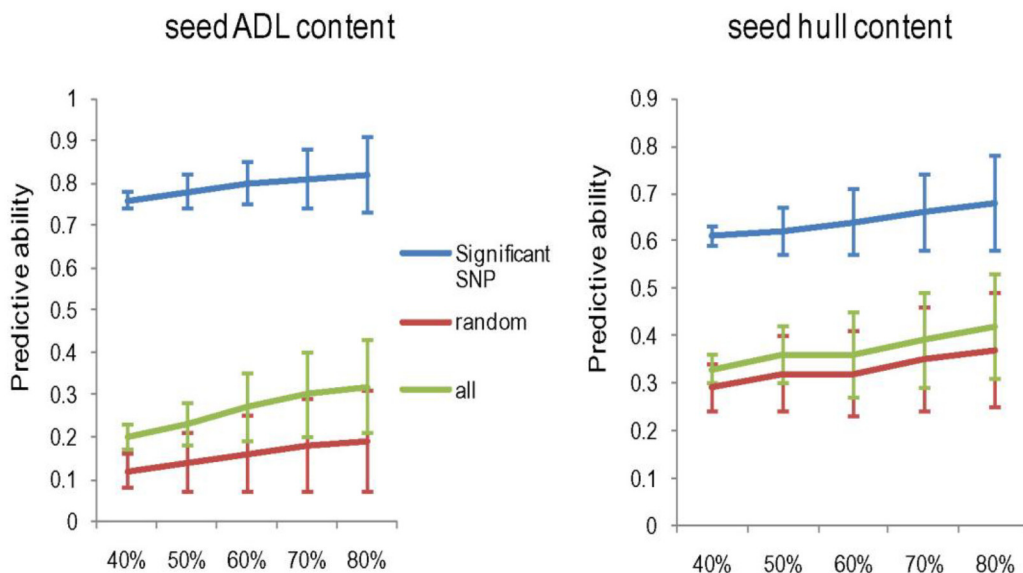
Rapeseed seed-coat thickness varies depending on variety and seed-coat color. An anatomical analysis of the seed coat indicates that enlarged cells of the palisade and sponge tissues cause the seed coat of black-seeded rapeseed to thicken, and the color depth of the seed coat is mainly determined by the pigment type and quantity in palisade tissue cells [30]. Reducing seed-coat thickness may decrease or even eliminate the accumulation of pigments contributing to increased oil content output and good production quality. Lignin in yellow-seeded rapeseed is one of the key factors that leads to a seed-coat ratio lower than that of the black-seeded rapeseed [49]. Therefore, reducing seed ADL content can indirectly thin seed coats, not only increasing oil content but also improving the nutritive value of rapeseed meal [29].



**Fig 5. Functional category annotations for 123 candidate genes and their respective percentages identified via GWAS as significantly associated with seed hull content (HC) in *B. napus*.**

doi:10.1371/journal.pone.0145045.g005

The reduced fiber content and thinner seed coat of rapeseed are considered important screening indices for breeding. Previous studies on seed ADL content and HC have been conducted mainly using QTL mapping. Thus, we utilized GWAS for the purposes of this study. The results obtained in the two consecutive experimental years indicated that significant association signals of seed ADL content were found not only on the recognized chromosome A09



**Fig 6. Predictive ability of genome selection with different reference populations (x-axis) and SNP markers (y-axis).**

doi:10.1371/journal.pone.0145045.g006

but also on chromosomes A05 and C05. We were particularly interested in the significant association signals of ADL on chromosome A05, which were repeatedly noted for seed HC, reflecting the significant positive correlation between seed ADL content and HC. These results provide a theoretical basis for research on the genes affecting seed ADL content and HC. However, many shortcomings should be noted in our study, such as the small population size and low genetic diversity of these landraces within China, both of which can limit the potential of GWAS. The peak signals of the GWAS loci are often close to but not within the range of known genes due to the low density of SNPs around the candidate genes. This work is a preliminary GWAS for seed ADL content and HC in Chinese rapeseed. On one hand, expanding the population size using worldwide accessions is necessary, particularly concerning the yellow-seeded lines. On the other hand, the candidate genes need to be functionally verified via sequence and expression analyses or even gene transformation to confirm whether they are genuine.

The QTLs from preliminary mapping results often spanned  $\sim 10$  to 30cM [50] and contained hundreds of genes. Association mapping based on LD has been successfully used for exploring trait-associated loci in animals and plants, but this method cannot provide the confidence interval of statistical significance ensured by traditional QTL mapping. Therefore, association mapping and traditional QTL mapping are complementary methods [18], and their integration will substantially promote the analysis of complex quantitative traits. At present, nested association mapping (NAM) [51] and multiparent advanced generation intercross (MAGIC) [52], based on linkage mapping and GWA mapping are popular methods. If LD exists, it can be detected among the significant SNP markers and QTLs. Thus, we applied the LD analysis method in this study to infer the approximate range of the QTL position through an assessment of the LD degree between the significant SNP markers and their surrounding SNP markers. Our results indicate that the average range of R-QTL was approximately 92.32 kb, which is considerably smaller than that of traditional linkage mapping, and the R-QTLs can therefore provide a valuable reference for the next step of marker-assisted selection. We further analyzed the significant association signals on chromosomes A05, A09, and C05. Our findings indicate that all three R-QTL regions of the three peak SNPs contained genes related to lignin synthesis, particularly the R-QTL containing rs11682 and rs11683. Four key genes, *BnaA.SEC8*, *BnaA.PAL4*, *BnaA.CESA3*, and *BnaA.GPAT5*, associated with the phenylpropanoid-lignin pathways and seed-coat development were identified in this R-QTL region. Moreover, a comparison of our GWAS and QTL mapping results revealed a common chromosomal region on A09, and a key lignin biosynthesis gene *BnaA.CAD2/BnaA.CAD3* was discovered in this region. We are particularly interested in the significant association signals on chromosome A05 because it might be a potential locus with enormous importance for seed ADL content. In addition, we previously established that the ADL-related QTLs from QTL mapping were inherited from the yellow-seeded parent. The different seed ADL content signals from GWAS mapping were from diverse seed ADL content accessions. The accuracy of trait scoring is exceedingly important, and the outcomes will directly affect the reliability of GWAS. The identification of major loci in this study will provide the genetic resources and markers needed for the selection of yellow- and black-seeded lines with reduced lignin content and a thinner seed coat.

In this study, we focused on the significant association signals of seed ADL content rather than seed HC given the poor reproducibility of seed HC in the two-year experimental period. All the results reasonably support the finding that the loci detected in our study are located close to candidate genes for controlling seed ADL content and that these tightly associated SNPs are of significant benefit for the design of molecular markers for breeding *B. napus* to enhance production and improve its genetics. We propose two explanations for the poor

reproducibility of seed HC between the two years in our study. First, the seed embryo was not completely stripped, causing experimental errors. Second, the experimental plants experienced serious crop lodging in 2013, and the seed HC could have been strongly influenced by the decreased seed quality. Furthermore, most significant association signals for seed HC in 2013 were from environmentally specific loci.

Genomic selection (GS) is an important aspect of plant breeding. In this study, we used the R package rrBLUP to conduct a genome-wide prediction for seed ADL content and HC. The results indicate that genomic selection using the significant SNP markers identified via GWAS had a higher predictive ability, indicating that the predictive ability of a given model can be improved by combining GWAS and GS. Molecular marker-assisted selection is the main aim of molecular breeding, using close linked markers for molecular marker-assisted selection can ensure the accuracy of the selection and improve the efficiency of the selection [53]. SSR markers can be developed according to the sequence around the significant associated SNP loci, and can be easily used in laboratory for the low seed ADL content and HC with high seed protein content in *B. napus* breeding. Furthermore, the low seed ADL content and HC of rapeseed can be considered the important screening indices for high seed protein content of the defatted meal rather than seed-coat color.

At present, GWAS investigation have entered a post-GWAS world [54]; the most contentious and difficult issues in our GWAS were the interaction analyses of  $G \times G$  and  $G \times E$  as well as the analysis of regulatory networks. Further work to analyze the population differentiation index, genetic diversity, and selective sweep of this GWAS population is ongoing, and an analysis for candidate genes can be conducted in the near future. Along with the emergence of new sequencing technologies and the continuing decline in sequencing costs, genome-wide association research will find a wider application in the investigations of *B. napus* and other crops.

## Conclusions

With the GWAS Q + K model, significant association signals for seed ADL content were found not only on the recognized chromosome, A09, but also on chromosomes A05 and C05 in two experimental years. A comparison of our GWAS results and those of previous linkage mappings revealed a common chromosomal region on A09 and a major locus on A05 associated with seed ADL content, as detected by GWA mapping. Various key lignin biosynthesis genes were identified in the R-QTLs of seed ADL content based on the LD analysis. In total, 123 candidate genes associated with seed HC were identified by GWA mapping. Greater than 65% of these genes are involved in the phenylpropanoid pathway and are transparent testa genes. The beneficial allele and candidate genes will be useful in rapeseed breeding for achieving the ideal yellow-seeded rapeseed variety through a molecular design approach.

## Supporting Information

**S1 Fig. The optimal Box-Cox transformation of 13HC and 14HC.**  
(TIF)

**S2 Fig. Genome-wide association scan for seed ADL content on A05.** The top of the panel shows an R-QTL region based on a significant SNP, whose position is indicated by a vertical green line. The negative log<sub>10</sub>-transformed *p*-values from the compressed MLM are plotted on the vertical axis. The blue horizontal lines indicate the 1% FRD-adjusted significance threshold ( $6.27 \times 10^{-6}$ ), and the bottom of the panel shows the related candidate genes with green boxes in the R-QTL region. Two previously identified genes, (a) *CCR1* and (b) *CAD2/CAD3*, were

significantly associated with seed ADL content.  
(TIF)

**S1 Table. *Brassica napus* accession information in this study.**  
(XLSX)

**S2 Table. Correlation analysis for seed ADL content and HC over two years.**  
(DOCX)

**S3 Table. ANOVA of seed ADL content and HC phenotypes.**  
(DOCX)

**S4 Table. The sequence data of significant association SNPs for seed ADL content and HC.**  
(XLSX)

**S5 Table. Genome-wide significant association signals for seed hull content (13HC).**  
(DOCX)

**S6 Table. Candidate genes near the genomic regions most highly associated with seed HC in *B. napus*.**  
(DOCX)

## Acknowledgments

Elizabeth F. and Nicole D., who from AJE, have edited this document.

## Author Contributions

Conceived and designed the experiments: LL. Performed the experiments: JW HJ LW. Analyzed the data: JW LW LL. Contributed reagents/materials/analysis tools: CQ XX KL WQ JL ML. Wrote the paper: JW.

## References

1. Ochodzki P, Piotrowska A, Sorensen H Physical properties and chemical composition of whole seeds, hulls and meals of yellow-seeded rape grown in Poland. In: 11th International Rapeseed Congress, Copenhagen, 6–10 July 2003. The Royal Veterinary and Agricultural University, pp 5–7
2. Wittkop B, Snowdon RJ, Friedt W. Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. *Euphytica*. 2009; 170: 131–140.
3. Simbaya J, Slominski BA, Rakow G, Campbell LD, Downey RK, Bell JM. Quality Characteristics of Yellow-Seeded Brassica Seed Meals—Protein, Carbohydrates, and Dietary Fiber Components. *Journal of Agricultural and Food Chemistry*. 1995; 43: 2062–2066.
4. Qu C, Fu F, Lu K, Zhang K, Wang R, Xu X, et al. Differential accumulation of phenolic compounds and expression of related genes in black- and yellow-seeded *Brassica napus*. *J Exp Bot*. 2013; 64: 2885–2898. doi: [10.1093/jxb/ert148](https://doi.org/10.1093/jxb/ert148) PMID: [23698630](https://pubmed.ncbi.nlm.nih.gov/23698630/)
5. Fu FY, Liu LZ, Chai YR, Chen L, Yang T, Jin MY, et al. Localization of QTLs for seed color using recombinant inbred lines of *Brassica napus* in different environments. *Genome*. 2007; 50: 840–854. PMID: [17893725](https://pubmed.ncbi.nlm.nih.gov/17893725/)
6. Seren U, Vilhjalmsdottir BJ, Horton MW, Meng DZ, Forai P, Huang YS, et al. GWAPP: A Web Application for Genome-Wide Association Mapping in Arabidopsis. *Plant Cell*. 2012; 24: 4793–4805. doi: [10.1105/tpc.112.108068](https://doi.org/10.1105/tpc.112.108068) PMID: [23277364](https://pubmed.ncbi.nlm.nih.gov/23277364/)
7. Pauli D, Muehlbauer GJ, Smith KP, Cooper B, Hole D, Obert DE, et al. Association Mapping of Agronomic QTLs in US Spring Barley Breeding Germplasm. *Plant Genome*. 2014; 7:
8. Huang XH, Wei XH, Sang T, Zhao QA, Feng Q, Zhao Y, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*. 2010; 42: 961–U976. doi: [10.1038/ng.695](https://doi.org/10.1038/ng.695) PMID: [20972439](https://pubmed.ncbi.nlm.nih.gov/20972439/)
9. Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES. Genetic diversity and selection in the maize starch pathway. *Proc Natl Acad Sci U S A*. 2002; 99: 12959–12962. PMID: [12244216](https://pubmed.ncbi.nlm.nih.gov/12244216/)

10. Ravel C, Praud S, Canaguier A, Dufour P, Giancola S, Balfourier F, et al. DNA sequence polymorphisms and their application to bread wheat quality. *Euphytica*. 2007; 158: 331–336.
11. Stracke S, Haseneyer G, Veyrieras JB, Geiger HH, Sauer S, Graner A, et al. Association mapping reveals gene action and interactions in the determination of flowering time in barley. *Theor Appl Genet*. 2009; 118: 259–273. doi: [10.1007/s00122-008-0896-y](https://doi.org/10.1007/s00122-008-0896-y) PMID: [18830577](https://pubmed.ncbi.nlm.nih.gov/18830577/)
12. Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Schuler K. Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Molecular Breeding*. 2004; 13: 93–102.
13. Xiao YJ, Cai DF, Yang W, Ye W, Younas M, Wu JS, et al. Genetic structure and linkage disequilibrium pattern of a rapeseed (*Brassica napus* L.) association mapping panel revealed by microsatellites. *Theoretical and Applied Genetics*. 2012; 125: 437–447. doi: [10.1007/s00122-012-1843-5](https://doi.org/10.1007/s00122-012-1843-5) PMID: [22437490](https://pubmed.ncbi.nlm.nih.gov/22437490/)
14. Hasan M, Friedt W, Pons-Kuehnemann J, Freitag NM, Link K, Snowdon RJ. Association of gene-linked SSR markers to seed glucosinolate content in oilseed rape (*Brassica napus* ssp. *napus*). *Theoretical and Applied Genetics*. 2008; 116: 1035–1049. doi: [10.1007/s00122-008-0733-3](https://doi.org/10.1007/s00122-008-0733-3) PMID: [18322671](https://pubmed.ncbi.nlm.nih.gov/18322671/)
15. Honsdorf N, Becker HC, Ecke W. Association mapping for phenological, morphological, and quality traits in canola quality winter rapeseed (*Brassica napus* L.). *Genome*. 2010; 53: 899–907. doi: [10.1139/G10-049](https://doi.org/10.1139/G10-049) PMID: [21076505](https://pubmed.ncbi.nlm.nih.gov/21076505/)
16. Zou J, Jiang CC, Cao ZY, Li RY, Long Y, Chen S, et al. Association mapping of seed oil content in *Brassica napus* and comparison with quantitative trait loci identified from linkage mapping. *Genome*. 2010; 53: 908–916. doi: [10.1139/G10-075](https://doi.org/10.1139/G10-075) PMID: [21076506](https://pubmed.ncbi.nlm.nih.gov/21076506/)
17. Jestin C, Lode M, Vallee P, Domin C, Falentin C, Horvais R, et al. Association mapping of quantitative resistance for *Leptosphaeria maculans* in oilseed rape (*Brassica napus* L.). *Molecular Breeding*. 2011; 27: 271–287.
18. Rezaeizad A, Wittkop B, Snowdon R, Hasan M, Mohammadi V, Zali A, et al. Identification of QTLs for phenolic compounds in oilseed rape (*Brassica napus* L.) by association mapping using SSR markers. *Euphytica*. 2011; 177: 335–342.
19. Wang NA, Qian W, Suppanz I, Wei LJ, Mao BZ, Long Y, et al. Flowering time variation in oilseed rape (*Brassica napus* L.) is associated with allelic variation in the FRIGIDA homologue BnaA.FRI.a. *Journal of Experimental Botany*. 2011; 62: 5641–5658. doi: [10.1093/jxb/err249](https://doi.org/10.1093/jxb/err249) PMID: [21862478](https://pubmed.ncbi.nlm.nih.gov/21862478/)
20. Fritsche S, Wang XX, Li JQ, Stich B, Kopisch-Obuch FJ, Endrigkeit J, et al. A candidate gene-based association study of tocopherol content and composition in rapeseed (*Brassica napus*). *Frontiers in Plant Science*. 2012; 3:
21. Wang XX, Zhang CY, Li LJ, Fritsche S, Endrigkeit J, Zhang WY, et al. Unraveling the Genetic Basis of Seed Tocopherol Content and Composition in Rapeseed (*Brassica napus* L.). *Plos One*. 2012; 7:
22. Fomeju BF, Falentin C, Lassalle G, Manzanares-Dauleux MJ, Delourme R. Homoeologous duplicated regions are involved in quantitative resistance of *Brassica napus* to stem canker. *Bmc Genomics*. 2014; 15:
23. Li F, Chen BY, Xu K, Wu JF, Song WL, Bancroft I, et al. Genome-Wide Association Study Dissects the Genetic Architecture of Seed Weight and Seed Quality in Rapeseed (*Brassica napus* L.). *DNA Research*. 2014; 21: 355–367. doi: [10.1093/dnares/dsu002](https://doi.org/10.1093/dnares/dsu002) PMID: [24510440](https://pubmed.ncbi.nlm.nih.gov/24510440/)
24. Raman H, Raman R, Kilian A, Detering F, Carling J, Coombes N, et al. Genome-Wide Delineation of Natural Variation for Pod Shatter Resistance in *Brassica napus*. *Plos One*. 2014; 9:
25. Chalhoub B, Denoeud F, Liu SY, Parkin IAP, Tang HB, Wang XY, et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*. 2014; 345: 950–953. doi: [10.1126/science.1253435](https://doi.org/10.1126/science.1253435) PMID: [25146293](https://pubmed.ncbi.nlm.nih.gov/25146293/)
26. Badani AG, Snowdon RJ, Wittkop B, Lipsa FD, Baetzel R, Horn R, et al. 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*. 2006; 49: 1499–1509. PMID: [17426765](https://pubmed.ncbi.nlm.nih.gov/17426765/)
27. Snowdon RJ, Wittkop B, Rezaidat A, Hasan M, Lipsa F, Stein A, et al. Regional association analysis delineates a sequenced chromosome region influencing antinutritive seed meal compounds in oilseed rape. *Genome*. 2010; 53: 917–928. doi: [10.1139/G10-052](https://doi.org/10.1139/G10-052) PMID: [21076507](https://pubmed.ncbi.nlm.nih.gov/21076507/)
28. Liu LZ, Stein A, Wittkop B, Sarvari P, Li JN, Yan XY, et al. A knockout mutation in the lignin biosynthesis gene CCR1 explains a major QTL for acid detergent lignin content in *Brassica napus* seeds. *Theoretical and Applied Genetics*. 2012; 124: 1573–1586. doi: [10.1007/s00122-012-1811-0](https://doi.org/10.1007/s00122-012-1811-0) PMID: [22350089](https://pubmed.ncbi.nlm.nih.gov/22350089/)
29. Liu L, Qu C, Wittkop B, Yi B, Xiao Y, He Y, et al. 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](https://doi.org/10.1371/journal.pone.0083052) PMID: [24386142](https://pubmed.ncbi.nlm.nih.gov/24386142/)



30. Stringam GR, McGregor DI, Pawlowski SH Chemical and morphological characteristics associated with seed coat color in rapeseed. In: Proceedings of the 4th international rapeseed conference, Gies-sen, 1974. vol 2. pp 99–108
31. Stein A, Wittkop B, Liu LZ, Obermeier C, Friedt W, Snowdon RJ. 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 Breeding*. 2013; 132: 382–389.
32. Jin MY, Li JN, Fu FY, Zhang ZS, Zhang XK, Liu LZ. QTL Analysis of Oil and Hull Content in *Brassica napus* L. *Scientia Agricultura Sinica*. 2007; 40: 677–684.
33. Yan XY, Li JN, Fu FY, Jin MY, Chen L, Liu LZ. Co-location of seed oil content, seed hull content and seed coat color QTL in three different environments in *Brassica napus* L. *Euphytica*. 2009; 170: 355–364.
34. Liu HF. QTL Analysis of seed coat lignin and Hull Content in *Brassica napus* L. *Agriculture and Technology*. 2014; 34: 22–23.
35. Dimov Z, Suprianto E, Hermann F, Mollers C. Genetic variation for seed hull and fibre content in a collection of European winter oilseed rape material (*Brassica napus* L.) and development of NIRS calibrations. *Plant Breeding*. 2012; 131: 361–368.
36. Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*. 2006; 38: 203–208. PMID: [16380716](#)
37. Zhang ZW, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*. 2010; 42: 355–U118. doi: [10.1038/ng.546](#) PMID: [20208535](#)
38. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007; 23: 2633–2635. PMID: [17586829](#)
39. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. *Science*. 2004; 305: 525–528. PMID: [15273396](#)
40. Olsen HG, Hayes BJ, Kent MP, Nome T, Svendsen M, Larsgard AG, et al. Genome-wide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. *Animal Genetics*. 2011; 42: 466–474. doi: [10.1111/j.1365-2052.2011.02179.x](#) PMID: [21906098](#)
41. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots (2014) <http://dx.doi.org/10.1101/005165>. Accessed 14-May
42. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21: 263–265. PMID: [15297300](#)
43. Lund MS, Sahana G, de Koning DJ, Su G, Carlborg O. Comparison of analyses of the QTLMAS XII common dataset. I: Genomic selection. *BMC Proc*. 2009; 3 Suppl 1: S1. PMID: [19278535](#)
44. Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome*. 2011; 4: 250–255.
45. Slavov GT, Nipper R, Robson P, Farrar K, Allison GG, Bosch M, et al. Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass *Miscanthus sinensis*. *New Phytologist*. 2014; 201: 1227–1239. doi: [10.1111/nph.12621](#) PMID: [24308815](#)
46. Box GEP, Cox DR. An Analysis of Transformations. *Journal of the Royal Statistical Society Series B-Statistical Methodology*. 1964; 26: 211–252.
47. Angelovici R, Lipka AE, Deason N, Gonzalez-Jorge S, Lin HN, Cepela J, et al. Genome-Wide Analysis of Branched-Chain Amino Acid Levels in Arabidopsis Seeds. *Plant Cell*. 2013; 25: 4827–4843. doi: [10.1105/tpc.113.119370](#) PMID: [24368787](#)
48. Cai DF, Xiao YJ, Yang W, Ye W, Wang B, Younas M, et al. Association mapping of six yield-related traits in rapeseed (*Brassica napus* L.). *Theoretical and Applied Genetics*. 2014; 127: 85–96. PMID: [24121524](#)
49. Liang YL, Liang Y, Li JN, Chen L. Difference between yellow and black seed coat in *B.napus*. *Chinese journal of oil crop sciences*. 2002; 24: 13–18.
50. Flint-Garcia SA, ThUILlet AC, Yu JM, Pressoir G, Romero SM, Mitchell SE, et al. Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant Journal*. 2005; 44: 1054–1064. PMID: [16359397](#)
51. McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li HH, Sun Q, et al. Genetic Properties of the Maize Nested Association Mapping Population. *Science*. 2009; 325: 737–740. doi: [10.1126/science.1174320](#) PMID: [19661427](#)

52. Cavanagh C, Morell M, Mackay I, Powell W. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology*. 2008; 11: 215–221. doi: [10.1016/j.pbi.2008.01.002](https://doi.org/10.1016/j.pbi.2008.01.002) PMID: [18295532](https://pubmed.ncbi.nlm.nih.gov/18295532/)
53. Andersen JR, Lubberstedt T. Functional markers in plants. *Trends Plant Sci*. 2003; 8: 554–560. PMID: [14607101](https://pubmed.ncbi.nlm.nih.gov/14607101/)
54. Nuzhdin SV, Friesen ML, McIntyre LM. Genotype-phenotype mapping in a post-GWAS world. *Trends Genet*. 2012; 28: 421–426. doi: [10.1016/j.tig.2012.06.003](https://doi.org/10.1016/j.tig.2012.06.003) PMID: [22818580](https://pubmed.ncbi.nlm.nih.gov/22818580/)