Association Mapping of Seed Quality Traits Under Varying Conditions of Nitrogen Application in *Brassica juncea* L. Czern & Coss

Javed Akhatar, Mohini Prabha Singh, Anju Sharma, Harjeevan Kaur, Navneet Kaur, Sanjula Sharma, Baudh Bharti, V. K. Sardana and Surinder S. Banga*

DBT Centre of Excellence on Brassicas, Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India

Indian mustard (*Brassica juncea*) is a major source of vegetable oil in the Indian subcontinent. The seed cake left after the oil extraction is used as livestock feed. We examined the genetic architecture of oil, protein, and glucosinolates by conducting a genome-wide association study (GWAS), using an association panel comprising 92 diverse genotypes. We conducted trait phenotyping over 2 years at two levels of nitrogen (N) application. Genotyping by sequencing was used to identify 66,835 loci, covering 18 chromosomes. Genetic diversity and phenotypic variations were high for the studied traits. Trait performances were stable when averaged over years and N levels. However, individual performances differed. General and mixed linear models were used to estimate the association between the SNP markers and the seed quality traits. Population structure, principal components (PCs) analysis, and discriminant analysis of principal components (DAPCs) were included as covariates to overcome the bias due to the population stratification. We identified 16, 23, and 27 loci associated with oil, protein, and glucosinolates, respectively. We also established LD patterns and haplotype structures for the candidate genes. The average block sizes were larger on A-genome chromosomes as compared to the B-genome chromosomes. Genetic associations differed over N levels. However, meta-analysis of GWAS datasets not only improved the power to recognize associations but also helped to identify common SNPs for oil and protein contents. Annotation of the genomic region around the identified SNPs led to the prediction of 21 orthologs of the functional candidate genes related to the biosynthesis of oil, protein, and glucosinolates. Notable among these are: *LACS5* (A09), *FAD6* (B05), *ASN1* (A06), *GTR2* (A06), *CYP81G1* (B06), and *MYB44* (B06). The identified loci will be very useful for marker-aided breeding for seed quality modifications in *B. juncea*.

Keywords: Indian mustard, seed oil, seed protein, glucosinolates, genome-wide association study, SNP, quantitative trait loci
INTRODUCTION

Crop Brassicas are comprised of six economically important species belonging to the family Brassicaceae. These plants are cultivated as vegetables, fodder, edible oilseeds, or biofuel crops. Of these, Indian mustard (B. juncea L. Czern & Coss) is the most widely cultivated oilseed crop in India, with very high acreage (6.9 million hectares) and production (7.2 million metric tonnes) (USDA, 2018–2019). It is also cultivated in China, Southern Russia, and the Caspian steppes as a condiment, vegetable, and oilseed crop. B. juncea is an allotetraploid (AABB, 2n = 36) and it evolved through multiple hybridization events between B. rapa (AA, 2n = 20) and B. nigra (BB, 2n = 16) (Nagaharu, 1935). Breeding programs in this crop are focused on improving productivity and seed quality. Oil, protein, and glucosinolates (GSls) determine seed quality in Indian mustard. The oil content in brassicas range from 45 to 50% and is mainly made up of unsaturated fatty acids (Smooker et al., 2011), Palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2), linolenic (C18:3), eicosenoic (C20:1), and erucic (C22:1) are the key fatty acids in this crop (Abbiad and Leckband, 2011). Breeding programs emphasize on providing varieties with both high (for industrial applications) and low (suitable for human consumption) erucic acid content (Bearé-Rogers et al., 1971; Kaur et al., 2019). High oleic acid, low linoleic acid, or a combination of these have been developed in rapeseed-mustard crops (Appelqvist, 1971). Complex biochemical metabolic reactions with corresponding enzymes in fatty acid biosynthesis have been reviewed by Barker et al. (2007). FAE1 plays a key role in the synthesis of erucic acid (James and Dooner, 1990) and two orthologs – BnaA.FAE1.a and BnaC.FAE1.a – are present in B. napus, as confirmed by associative transcriptomics (Wu et al., 2008; Harper et al., 2012; Havlickova et al., 2018). A high amount of protein is stored in the seeds of B. juncea, and it is reported to be negatively correlated with the oil content (Grami and Stefansson, 1977). Seed cake left after the extraction of oil is used as livestock feed. Two classes of seed storage proteins are prevalent: legumin-type globulins (11S or 12S or cruciferin) and napin-type albumins (2S or napins) (Wanasundara, 2011). The primary structural proteins of Brassica oilseeds are oleosins or oil body proteins. Considerable metabolic proteins, such as lipid transfer proteins (LTP), protease inhibitors (Ceciliani et al., 1994), Ca²⁺ dependent-calmodulin binding proteins (Neumann et al., 1996), and dehydrins (Svensson et al., 2000), are well-documented in Brassica seeds. Extensive knowledge of the protein structure, structure-function relationships, and genetic control is important for optimal utilization of Brassica proteins and for developing new protein-based applications.

Glucosinolates (β-thioglucoside-N-hydroxysulfates), a endogenous allelochemical group and a subset of secondary metabolites, are only present in the family Brassicaceae (Zhang and Hamauzu, 2004). Glucosinolates are anticancer, antibacterial, anti-fungal, anti-oxidative, and allelopathic compounds (Latté et al., 2011). However, these are considered anti-nutritional components in food as they can affect the thyroid function (Walker and Booth, 2001). Therefore, it is advisable to reduce GSL content in the seeds, and maintain high GSL content in other tissues to prevent herbivore damage and pathogenic microbes. Over 200 GSLs are known in Brassica crops. Depending on the precursor amino acid involved, glucosinolates are categorized into aliphatic, aromatic, and indole types (Fahey et al., 2001; Clarke, 2010). Glucoiberin, progoitrin, epirogoitrin, glucoraphanin, sinigrin, glucoraphenin, glucosinapin, glucobarbarin, glucobrassicapin, glucorucin, glucobrassicin, and gluconasturtin are the most prevalent GSLs (Wittstock and Halkier, 2002). B. juncea contains significant amounts of aliphatic glucosinolates (Yadav and Rana, 2018). Most of the genes responsible for biosynthetic steps are now known in Arabidopsis thaliana (Sonderby et al., 2010). A group of R2R3 MYB transcription factors belonging to a single gene family within Arabidopsis is involved in the direct transcriptional regulation of GSLs' biosynthesis (Sonderby et al., 2010; Frerignam and Gogolashvili, 2014). Also, MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis by directly interacting with glucosinolate-related MYB24. Limited knowledge is available for the chain-elongated homophenylalanine-derived aromatic GSLs (Bhandari et al., 2015). Genes controlling aromatic GSL biosynthesis are still uncharacterized. An extensive study and regulation of GSL natural variations in Brassica requires a better understanding of its genetic system. All these traits are under the control of complex regulatory mechanisms, strongly affected by the environment. So, phenotyping in replicated and multi-environment trials are important (Becker and Léon, 1988; Wang et al., 2015). Quantitative trait loci (QTL) control complex traits in crop plants and these are better understood by genome wide association studies (GWAS). Associations between phenotypes and markers can be deduced from linkage disequilibrium (LD). GWAS requires the use of numerous genome-wide markers and it provides high mapping resolution by exploiting ancestral recombination events present in a plant species (Soto-Cerda and Cloutier, 2012). It is an alluring approach for the discovery of QTLs in plant genomes associated with phenological, morphological (Honsdorf et al., 2010; Akhatar and Banga, 2015), and seed quality traits (Li et al., 2014). Many studies have been conducted to understand the genetics underlying variations for oil, protein, and glucosinolate contents in B. napus and B. rapa. However, previous studies were primarily focused on phenotypic evaluations in greenhouse or growth chambers under uniform fertilizer applications. So, there is always a theoretical possibility of missing important loci if the crop growth conditions fail to include limiting environmental variables such as Nitrogen (N), a major limiting factor for Brassica production (Danesh-Shahraki et al., 2008). There is a positive correlation between soil N level and seed quality (Ahmadi and Bahrami, 2009). Both protein and oil content are impacted strongly by nitrogen availability as the plants produce particular stress proteins in response to sub-optimal conditions (Betley and Finch-Savage, 1996). For the present studies, we phenotyped 100 genotypes of B. juncea under real farm conditions. The crop was sown under optimal versus limited N fertilization over 2 years to generate N stress and G × N interactions. We then carried out GWAS to investigate the genetic architecture of seed quality-related traits and their stability across environments.
TABLE 1 | Descriptive statistics of the quality traits in B. juncea diversity panel during years 2015–2016 and 2016–2017 at two N-level.

| Trait | Values | Y1 | Y2 | Y1 | Y2 |
|-------|--------|----|----|----|----|
|       |        | N0 | N100 | N0 | N100 |
| Oil (%) | Mean ± SE | 38.3 ± 0.2 | 37.4 ± 0.4 | 37.7 ± 0.3 | 37.6 ± 0.3 |
|        | Range   | (30.6–43.7) | (26.1–44.5) | (29.7–44.0) | (29.5 to 48.1) |
| CV     | 7.0     | 9.3  | 6.8  | 7.4  |
| H² (%) | 44.8    | 91.27 | 47.51 | 63.09 |
| PTN (%) | Mean ± SE | 29.3 ± 0.2 | 29.7 ± 0.2 | 29.2 ± 0.2 | 29.2 ± 0.2 |
|        | Range   | (25.0–34.8) | (25.2–34.4) | (25.5–33.8) | (23.0–33.1) |
| CV     | 6.9     | 7.4  | 5.6  | 5.6  |
| H² (%) | 53.36   | 85.56 | 34.74 | 54.17 |
| GSL (µmolg⁻¹) | Mean ± SE | 93.1 ± 1.5 | 82.1 ± 1.6 | 92.4 ± 1.3 | 78 ± 1.3 |
|        | Range   | (45.5–130.1) | (38.0–125.3) | (42.8–124.5) | (27.7–107.9) |
| CV     | 16.0    | 19.1 | 14.3 | 16.5 |
| H² (%) | 60.24   | 95.39 | 16.15 | 18.34 |

N0, with no added nitrogen; N100, with added nitrogen @100 kg/ha; Y1, year 1 (2015–2016); Y2, year 2 (2016–2017); SE, standard error; CV, coefficient of variation; PTN, proteins; GSL, glucosinolates; H², broad sense heritability.

MATERIALS AND METHODS

Plant Material, Field Trial, and Phenotyping

B. juncea diversity fixed foundation set (BjDFFS), comprising 100 accessions (S7/S8 inbred lines), constituted the experimental materials for the present investigations. These accessions were evaluated in the field trials conducted during the winters of 2015–2016 (Y1) and 2016–2017 (Y2) as per the alpha lattice design with two doses of nitrogen (N) application (N0: no added N and N100: added N @100 kg/ha). The experiment was conducted in the oilseeds research area, Punjab Agricultural University, Ludhiana, India (30.9010° N, 75.8573° E). Nitrogen was added in the form of urea (46% N). For N100 treatment, urea was applied in two split doses—half at the time of sowing and the remaining half at the stem elongation stage.

Oil, protein, and glucosinolates content in intact seeds were estimated by Near Infrared Reflectance Spectroscopy (NIRS), Model 6500 spectrophotometer, Foss-NIR Systems, Inc., Silver Spring, MD, United States. Existing NIRS-based calibration models (Sen et al., 2018) were used to estimate seed quality parameters. All the trait values were adjusted for seed moisture content. We used an average of five readings per replication for each test trait.

Statistical and Correlation Network Analysis

Statistical software SAS v9.3 was used for analyses of variance (ANOVA) to assess the significance of variance due to genotype, replication, years, N-levels, and all possible interactions between these (year \times genotype, N-level \times genotype, year \times N-level and replication \times N-level). The following model was used for this purpose.

\[
P_{ijkl} = \mu + G_i + R_j + Y_k + N_l + (Y \times G)_{kj} + (N \times G)_{lj}
+ (Y \times N)_{kl} + (R \times N)_{jl} + e_{ijkl}
\]

where \(P_{ijkl}\) is the phenotypic value of accession, \(i\) noted by the \(l\)th N-level in year \(k\), \(\mu\) the overall mean, \(G_i\) the effect of the accession \(i\), \(R_j\) the effect of replication \(j\), \(Y_k\) the effect of year \(k\), \(N_l\) the effect of the N-level \(l\), \((Y \times G)_{kj}\) the interaction term between year \(k\) and genotype \(j\), \((N \times G)_{lj}\) the interaction between N-level \(l\) and genotype \(j\), \((Y \times N)_{kl}\) the interaction between year \(k\) and N-level \(l\), \((R \times N)_{jl}\) the interaction term between replication \(j\) and N-level \(l\), and \(e_{ijkl}\) the random independent and identically distributed residual term. This approach was also used for characterizing the influence of the N-levels on the trait observations; multiyear and within-year models through the use of linear mixed models. Best unbiased linear predictors (BLUPs) were assessed to obtain datasets across the years and N-levels. These estimates and predictors extracted from both year and both N-level models were then used as input data for GWAS. Pairwise Pearson’s correlation coefficients (\(r\)) between the traits were estimated by using the R package “Hmisc”\(^1\). We visualized \(r\) network value with the R package “q-graph.”

SNP Genotyping

Total genomic DNA was extracted from young leaves of 100 genotypes using a standard genomic DNA extraction procedure (Doyle and Doyle, 1990), with minor modifications. DNA samples were quantified by visual comparison to 
λ-DNA standards on ethidium bromide-stained agarose gels. The purity and concentration of the samples were calculated using a standards on ethidium bromide-stained agarose gels. The purity and concentration of the samples were calculated using a spectrophotometer at 260 and 280 nms. High quality DNA samples were genotyped by sequencing (GBS) on the Illumina\(^2\) HiSeq platform, which was outsourced to Novogene (HK) Company Limited, Hong Kong. GBS data was carried out for only 92 genotypes, as DNA from the remaining eight samples failed the quality test. SNP data calling was carried out using NGSEP (Next Generation Sequencing Experience Platform) GBS pipeline\(^3\) (Duitama et al., 2014). B. juncea reference genome v1.1\(^4\) was used for aligning 25× whole genome sequence of

---

\(^1\)http://biostat.mc.vanderbilt.edu/wiki/Main/Hmisc
\(^2\)https://github.com/NGSEP/NGSEPcore
\(^3\)http://brassicadb.org/brad
reference based on the oilseed type mustard cultivar as the available *B. juncea* reference genome assembly was based on a Chinese vegetable mustard genotype, Tumida. To construct a mock-up pseudomolecule reference, total SNPs were replaced in the reference genome using a perl script, PseudoMaker, as implemented in SEG-Map (Zhao et al., 2010). All 92 inbred lines were then aligned using the pseudomolecule genome reference and, using NGSEP-GBS pipeline, SNPs were identified. SNPs with a minor allele frequency (MAF) of >0.05, minimum allele proportion of 0.7, and minimum quality score of 30 were selected for GWAS. After this filtration, a total of 66,835 SNPs were left for association studies. These are uniformly distributed over 18 chromosomes. Imputation of SNPs was performed using fGENE v1.0.7 (Roshyara and Scholz, 2014) and BEAGLE v3.3.2 (Browning and Browning, 2007). We used R package “GAPIT” (genome association and prediction integrated tool) to transform SNP data into a numeric format (Lipka et al., 2012).

**Covariates Analysis**

Different covariates, i.e., population structure, principal components (PCs) analysis, and discriminant analysis of principal components (DAPCs), are often used to overcome the bias due to the population stratification on outcomes from GWAS. These allow for adjustments for population stratification. Population structure analysis was performed by using STRUCTURE v2.3.4 (Evanno et al., 2005), with the subgroups (K) ranging from 1 to 10. The Markov Chain Monte Carlo (MCMC) repetitions were set to 10,000. An optimum number of subgroups (Q) was selected based on the log probability of the data \([\ln P(D)]\) and ad hoc statistic \(\Delta K\) method. R package, GAPIT was used for the selection of the number of significant PCs with the largest eigenvalues based on all pairs of SNPs. Discriminant Analysis of Principal Components (DAPC) analysis was also used to reduce false positives by reducing the effects of population stratification which were implemented in the R package “adegenet.”

**Genome-Wide Association Analysis Based on SNP Genotyping**

Data for protein, oil, and glucosinolates content were normalized by arcsine transformation. Genome-wide association analysis was performed on normalized values of 92 genotypes and 66,835 SNPs. Phenotypic data were pooled across 2 years for N0, N100, and NP (average values of N0 and N100) levels. Kinship matrix data was generated using the MVP-package of the software R. Q, PCs, and DAPCs analysis were used as covariates in different GWAS analysis algorithms to reduce false positives by reducing the effects of population stratification. Marker trait associations (MTAs) were estimated by MVP (A Memory-efficient Visualization-enhanced and Parallel-accelerated Tool)\(^5\) with default settings to identify marker trait associations. Three methods (GLM, MLM, and FarmCPU) were used with Q, PCs, and DAPCs as covariates. Quantile-Quantile (Q-Q) plots were performed with \(-\log_{10}(P)\) of each observed SNP and the expected \(P\) value. These allowed us to select best fit algorithm and only

\(^{\text{ahttp://bowtie-bio.sourceforge.net/bowtie2}}\)

\(^{\text{bhttps://github.com/XiaoleiLiuBio/MVP}}\)
corresponding SNPs were retained for further analysis. The significance of the association between SNPs and traits was assessed based on an arbitrary threshold of \(-\log_{10}(P) \geq 3\).

**Functional Annotation and Linkage Disequilibrium (LD)**

25 kb region, upstream or downstream of peak SNPs were used for gene finding. To facilitate that, we used the software Blast2GO Pro (Götz et al., 2008) to BLAST against A. thaliana database. Annotation allowed us to identify candidates in the vicinity of peak SNPs. SNPs, common across N treatments, were also depicted. LD was estimated between annotated SNPs, calculating the square value of correlation coefficient \((r^2)\) between all pairs of markers by software TASSEL v5.2 (Bradbury et al., 2007) with LD type “sliding window” and LD window size “50.”

**Haplotype and Linkage Disequilibrium Analysis**

Haplotypes were generated from the annotated SNPs. The haplotype and linkage disequilibrium (LD) analysis was carried out by using the software Haploview v4.2 (Barrett et al., 2005). It uses an expectation maximization (EM) algorithm to calculate measures of LD and create a graphical representation of block definitions (Gabriel et al., 2002) to partition the region into segments of strong LD.

**Meta-Analysis**

The meta-analysis was conducted to test for differences in trait associations between two N-levels and 2 years. This test is based on comparing the differences between the two regression coefficients. We used the meta-analysis software METASOFT26. The diversity panel was first stratified by using software PLINK × Analysis of variance showed significant differences for the Traits of Variation for Seed Quality.

**RESULTS**

**Analysis of Variation for Seed Quality Traits**

Analysis of variance showed significant differences for the genotypes, years, and N-level \( \times \) genotype interactions for OIL, PTN, and GSLs (Supplementary Table 1). ANOVA also revealed highly significant differences in PTN and GSLs over N levels. Variation due to Y \( \times \) N interactions was significant for oil content alone. The summary data for oil, protein, and glucosinolates contents are presented in Table 1 and Figure 1. Average values for OIL remained stable over N levels and years. However, individual performances varied for both OIL and PTN contents over N and years. DJ-1-2 DT5, PLM-4, and DJ-27 DTA18 showed high OIL (\( \geq 43\% \)) at N0Y1. In comparison, DJ-1-2 DT5 and JC-1359-23-558 accumulated more OIL during N0Y2. T-26-15C-R1631 and JC-1359-23-558 performed better for N100Y1. The same was true for JLM-96 and JC-1359-23-558 at N100Y2. Heritability (\(H^2\)) values were generally high, but these were lower during Y1 as compared to Y2 at both N levels. PTN ranged from 23 to 35% over environments. TM117 (35%) and MCP-12-227 (34%) revealed maximum seed PTN for N0Y1, while MCP-12-227, TM117, DT 70, and PBR-357 had higher PTN (\( \geq 34\% \)) for N0Y2. DJ-1-2 DT2 and BAUSAM-2 had high PTN (\( \geq 34\% \)) at N100Y1. In comparison, DJ-1-2 DT2 and TM117 had maximum PTN (\( \geq 33\% \)) at N100Y2. \(H^2\) for PTN ranged from 35 to 86% over years and N-levels. \(H^2\) was lower during Y1 as contrasted to Y2 at both N-levels. Average GSL values varied and these were lower by \(\sim 11\) and 14 \(\mu\)mol/g during Y2 compared to Y1. The lowest GSL values were noted for the genotypes, JM-06006 (38 \(\mu\)mol/g) and JM-06026 (27.7 \(\mu\)mol/g) for N0Y2 and N100Y2 respectively. \(H^2\) values were lower at N100 compared to N0. Similar to OIL and PTN, \(H^2\) values for GSLs were lower for Y1 as compared to Y2. Oil and PTN were inversely correlated. GSL and PTN were positively associated at N0 (Supplementary Table 2 and Supplementary Figure 1).

**GWAS Analysis**

We used Q-Q plots to identify a best fit model (Supplementary Figure 2).

**Oil**

We recognized fifteen MTA's involving chromosomes A04, A06, A09, B05, B06, and B08 (Table 2 and Figure 2). The phenotypic variation explained by these loci ranged from 8.82 to 20.37%. Functional annotation predicted seven genes between 0.38 and 13.96 kb on either side of the peak SNPs. We envisaged CER26 at a distance of 2.08 kb from the SNP A04_6525658. Annotation also called GDSL lipase gene at a distance of 13.96 kb from a group of three SNPs (A06_26456373, A06_26456397, and A06_26456466). This gene was also envisaged close (0.38 kb) to A09_14737522. Another gene, LACS5 (long-chain acyl-coA synthetase 5), encoding a long-chain-fatty-acid-CoA ligase, was identified near SNPs A09_24082148, A09_24082198, and A09_24083336. We annotated FAD6 at a distance of 1.64 kb from the nearest of the associated SNPs: B05_5272794, B05_5272795, and B05_5281484. Atlg06090, a gene encoding delta-9 desaturase-like 1 protein (SNP in the intron), and ADS1 (2.31 kb) were called on chromosome B06 (B06_19315729). GWAS also allowed for recognition of four SNPs (B08_34645905, B08_34645977, B08_34645991, and B08_34646006) for association with Diri, a gene encoding bifunctional inhibitor/lipid-transfer protein/seeds storage 25 albumin superfamily protein, annotated at a distance of 3.83 kb from peak SNP.
### Table 2

| Trait | Candidate gene (distance from peak SNP in kb) | SNP IDs | NCBI ID | Chr. | Position | GWAS | Meta-analysis |
|-------|-----------------------------------------------|---------|---------|------|----------|------|---------------|
|       |                                               |         |         |      |          | –log₁₀(P) | PVE% | A.E. | MAF | Meta P value | N-level |
| Oil   | CER26 (2.08)                                  | A04_6252658 | 15236357 | A04 | 6252658 | 3.09 | 14.82 | 1.04 | 0.18 | 0.00 × 10⁻²⁵ | N100    |
|       | GDSL esterase/lipase gene (13.96)             | A06_26456373, 97, 466 | 152286051 | A06 | 26456373-6466 | 3.17 | 16.29 | 0.47 | 0.17 | 4.33 × 10⁻⁶ | N0, NP  |
|       | GDSL esterase/lipase gene (0.38)              | A09_14737522 | 15285051 | A09 | 14737522 | 3.19 | 17.1 | -1.12 | 0.3  | 2.44 × 10⁻⁶ | N0, NP  |
|       | LACS5 (within gene, in intronic region)       | A09_24082148, 2198, 3036 | 103228411 | A09 | 24082148-3036 | 3.39 | 15.88 | 0.45 | 0.3  | 1.44 × 10⁻⁶ | N0, NP  |
|       | FAD6 (1.84)                                   | B05_5272794, 95 | 15235766 | B05 | 5272794-2795 | 4.1  | 20.37 | 0.82 | 0.09 | 3.63 × 10⁻⁶ | N0, NP  |
|       | AtP4H3 (within gene, in exonic region)        | A06_9257535, 68 | 18394842 | A06 | 9257535-7568 | 3.74 | 16.11 | 0  | 0.12 | 9.18 × 10⁻⁵ | N100, NP |
|       | ASN1 (1.08)                                   | A06_15534622, 4718, 4750 | 15232776 | A06 | 15534622-4750 | 3.02 | 14.09 | -0.23 | 0.12 | 0.00 × 10⁻⁹ | N100, NP |
|       | GTR2 (6.84)                                   | A06_17974981, 5007, 5009, 5010 | 15219390 | A06 | 17974981-5010 | 3.3  | 15.04 | 0.35 | 0.12 | 0.00 × 10⁻⁷ | N100, NP |
|       | At2g41640 (9.24)                               | A06_27490668, 71, 93 | 18649935 | A06 | 27490668-0693 | 4.03 | 23.21 | 0.83 | 0.25 | 6.28 × 10⁻⁶ | N0, NP  |
|       | PSP1 (1.91)                                   | A09_30208138 | 30692244 | A09 | 30208138 | 3.08 | 15.94 | -1.05 | 0.11 | 9.18 × 10⁻⁵ | N100, NP |
|       | PSP1 (3.71)                                   | A09_30208138 | 30692244 | A09 | 30208138 | 3.08 | 15.94 | -1.05 | 0.11 | 9.18 × 10⁻⁵ | N100, NP |
|       | GDSL HB16 (2.29) SK1 (3.35)                   | A04_11732900, 2912, 2989, 3020, 3023, 3024, 3029 | 189303601 | A04 | 11732900-3029 | 3.99 | 15.48 | 2.37 | 0.19 | 7.99 × 10⁻⁷ | N0*    |
|       | AT2G35450 (1.20)                               | A05_5070339 | 79594244 | A05 | 5070339 | 3.08 | 13.17 | 7.37 | 0.05 | 1.34 × 10⁻⁵ | N0*    |
|       | CM1 (1.70)                                    | A06_24763290 | 9293913 | A06 | 24763290 | 3.05 | 13.75 | 2.73 | 0.5  | 6.59 × 10⁻⁷ | N0*    |
|       | CYP81G1 (12.97)                                | A06_9290634, 55, 70 | 24025648 | A06 | 9290634-0670 | 3.55 | 17.46 | -1.51 | 0.49 | 6.61 × 10⁻⁵ | N0*    |

*These had highest m-values and were non-significant for meta-analysis over N-levels. PTN, proteins; GSL, glucosinolates; Chr., chromosome; PVE, phenotypic variation explained; A.E., allelic effect; MAF, minor allele frequency.

**Protein**

We found 23 MTA’s on chromosomes A04, A06, A09, and B03 (Table 2 and Figure 2). The defined phenotypic variation ranged from 14.09 to 23.21%. Functional annotation predicted seven genes within 1.08–9.24 kb from respective peak SNPs. PASTICCINO-1 (pas1) was anticipated at 1.46 kb from SNPs A04_12420732 and A04_12425239. This gene encodes peptidylprolyl isomerase. We also envisage AtP4H3...
FIGURE 2 | Manhattan plots for Association Analysis of (A) oil, (B) protein, and (C) glucosinolates.

with SNPs (A06_9257535 and A06_9257568) within the exon. This gene codes for procollagen proline 4-dioxygenase. The associated SNPs were recognized in the exonic region of the gene. Three significant SNPs – A06_15534622, A06_15534718, and A06_15534750 – were present at a distance of 1.08 kb from ASN1, a gene encoding glutamine dependent-asparagine synthetase. GTR2 was envisioned at a distance of 6.84 kb from the closest of the associated
FIGURE 3 | Chromosome-wise linkage disequilibrium (LD) plot generated using Haploview. The plot also depicts the haplotypes block containing the annotated SNPs for OIL, PTN, and GSL. Dark gray colors in LD plots indicate strong LD between markers as estimated by $r^2$ values.

SNPs: A06_17974981, A06_17975007, A06_17975009, and A06_17975010. At2g41640, encoding glycosyltransferase family 61 protein, was predicted at a distance of 9.24 kb from SNPs A06_27490668, A06_27490671, and A06_27490693. We also detected PSP1 (PHOSPHOSERINE PHOSPHATASE 1) on the chromosome A09 at a distance of 3.71 kb from SNP (A09_30208138). This gene was also envisioned next to the peak SNP on the chromosome B03.

Glucosinolates
We identified 22 MTAs involving chromosomes A04, A05, A06, B03, B04, and B06 (Table 2 and Figure 2). The phenotypic variation explained varied from 13.17 to 17.46%. Functional annotation predicted 10 genes in the vicinity (1.20–15.71 kb) of the significant SNPs. Of these, eight SNPs (A04_11732900, 2912, 2989, 3020, 3023, 3024, 3028, and 3029), were located close to the predicted genes HB16 and SK1 at the respective distances of 2.29 and 3.35 kb from the peak SNPs. Both these encode shikimate kinase, which catalyzes a step of shikimate pathway for the synthesis of aromatic amino acids (tryptophan, tyrosine, and phenylalanine). We also established a gene AT2G35450 on chromosome A05 at a distance of 1.2 kb from SNP A05_5070339.

Chorismate mutase1 (CM1) was predicted close to two SNPs on chromosome A06 at a distance of 1.7 kb and three SNPs for chromosome B04 at a distance of 6.51 kb from the respective peak SNPs. Two significant SNPs, B03_19283625 and B03_19283626, were associated with JMT, encoding jasmonate O-methyltransferase. Three SNPs (B06_8990634, 55, and 70) were detected within the exon of the predicted gene, LINC4. This gene encodes branched chain-amino acid-transaminase. CYP81G1 and MYB44 were envisioned at distances of 12.97 and 15.71 kb from the respective significant SNPs present on chromosome B06.

LD Plot of Annotated SNPs
Haplotype analysis and pair-wise LD estimation was performed in the panel using 60 SNPs annotated for OIL, PTN, and GSL (Figure 3). Thirteen out of sixteen SNPs associated with OIL generated four haplotype blocks on chromosomes A06, A09, B05, and B08. We found six haplotype blocks on chromosomes A04, A06, and B03 for PTN with strong $r^2$ values. The remaining 16 SNPs produced three haplotype blocks on the chromosomes A04, B04, and B06 for GSLs.
META-ANALYSIS

For meta-analysis, we looked into 21 candidate genes that were significantly associated with OIL, PTN, and GSL. Meta-analysis allowed reconfirmation of most GWAS-SNPs for OIL and PTN, but with higher mapping resolution and an improved number of significant variants (over N levels) (Table 2 and Figure 4). At1g06090 and ADS1 were predicted repeatedly at both N levels (meta p = 1.67 × 10^{-7}) with B06_19315729. DIR1, predicted on chromosome B08, was significant at both N0 and NP in meta-analysis while this was significant at NP in GWAS. For PTN (Table 2), PASTICCINO1 was called twice on chromosome A04 at meta p value of 1.13 × 10^{-6}. AtP4H3, ASN1, and GTR2 (all on A06) were significant at N0 and NP following meta-analysis. These were significant at N100 alone for GWAS. SNP linked to At2g41640, a gene encoding glycosyltransferase family 61 protein, was significant at N0 and NP in meta-analysis (meta p = 8.72 × 10^{-6}). It was significant only at N0 level following GWAS. PSPI, with A09_30208138 as associated SNP, was significant at N0 and NP during meta-analysis (meta p = 6.28 × 10^{-6}). For glucosinolates, HB16 and SK1 were present close to eight SNPs associated with chromosome A04. These were significant at NP during GWAS. We recorded the highest

![Meta-analysis output file for oil (A), protein (B), and glucosinolates content (C) (one gene for each trait).](image-url)
A similar trend was observed for a similar trend was observed for

\[ \text{gene} \]

Table 2

Oil

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | B. juncea   | B. napus      |
|       |       | A6, A10     | A6, A10       |
|       |       | A6, A8, A10 | A6, A8, A10   |

GDSL

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A2, A2, A9   |               |
|       |       | A2, A2, A9   |               |

AT4G12510

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A3           | A3            |

AT4G13840

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A4           |               |

AT4G30950

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A1, A8       |               |

AT1G31070

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A8, A9       |               |

AT2G03360

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A6, A9       |               |

AT3G47340

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A6           |               |

AT3G54010

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A4           |               |

AT5G66280

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A1, A2, A6, A9|               |

GSL

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A6, A7       |               |

AT2G21940

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A6           |               |

AT2G35450

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A6, A9       |               |

AT4G40060

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A1, A8       |               |

AT5G65770

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A2, A2       |               |

AT5G67310

| Trait | AT ID | Gene/protein | Chromosome(s) |
|-------|-------|--------------|---------------|
|       |       | A7, A7       |               |

but non-significant m value (meta-analysis) for NP (Table 2).

A similar trend was observed for CM1 envisaged on chromosome A06 twice. We also recognized CM1 on chromosome B04 with an m-value of 10−73. It was significant at N0 for GWAS but non-significant following meta-analysis (highest m values at N0). Almost similar trends were recorded for LINCA, CYP81G1, and MYB44 on the chromosome B06.

In silico analysis of the genome sequences of B. napus and B. juncea allowed identification of multiple copies of the genes that predicted seed quality traits on the A genome common to both B. juncea and B. napus (Table 3). A fair degree of parallelism existed, but these were differences in their location on specific A genome chromosomes.

**DISCUSSION**

The development of high-yielding strains with increased oil and protein contents coupled with low meal glucosinolates in seeds is a major crop improvement goal in B. juncea. Simultaneous improvement for these interrelated traits is, however, challenging due to complex genetics and large environmental influences (Banga et al., 2015). Many investigations have been undertaken to understand the genetics of seed oil content in B. juncea (Mahmood et al., 2006; Ramchiary et al., 2007; Yadava et al., 2012; Rout et al., 2018), B. napus (Qiu et al., 2006; Cao et al., 2010; Zhao et al., 2012; Wang et al., 2013; Jiang et al., 2014; Körber et al., 2016; Fu et al., 2017), and B. carinata (Zhang et al., 2017). GWAS was also combined with transcriptome analyses to predict seven functional candidate genes for the seed oil content in B. napus (Xiao et al., 2019). However, there are only a few studies involving association mapping for seed quality traits in B. juncea. We investigated the genetics of seed quality traits through GWAS based on an association panel. Trait phenotyping was carried out at two doses of N application. ANOVA revealed highly significant differences in PTN and GSLs over N levels, but OIL was relatively stable. Relatively stable expression of oil content is largely in confirmation with an earlier report in mustard (Chauhan et al., 2012). GWAS and meta-GWAS allowed us to examine environment (N) specific SNPs and candidate genes related to OIL, PTN, and GSL. We identified several genes assigned to oil content and fatty acid synthesis. These were placed on both A-(A04, A06, A09) and B-(B05, B06, B08) genome chromosomes. Notable among these were CER26, GDSL lipase gene, and LACS5. CER26 has a role in fatty acid elongation (Pascal et al., 2013), while GDSL esterase/lipase (GLIP) gene(s) encode hydrolytic enzymes involved in development and morphogenesis as reported in rice, Arabidopsis, and maize (Akoh et al., 2004). GDSL lipase gene was first isolated from B. napus (Ling et al., 2006) and QTLs for esterase and lipases family protein have been reported in B. napus (Gu et al., 2017). LACS5 is a floral tissue-specific gene that contributes to oil biosynthesis (Shockey et al., 2002). FAD6 (called on chromosome B05) encodes delta-12 desaturase that aids in fatty acid desaturation (Mikkilineni and Rocheford, 2003). We envisaged delta9 desaturase (A9 DS) and ADS1 on chromosome B06 at all N levels. A9 DS, an intrinsic membrane protein, regulates the catalytic desaturation of saturated fatty acids at the C9 and C10 positions to form unsaturated fatty acids (Stukey et al., 1990). ADS1 has been used widely for genetically modifying saturated fatty acids in oilseed crops (Yao et al., 2003). Plant lipid-transfer proteins (LTPs) are able to reversibly bind and transport lipids in vitro (Kader et al., 1984). We predict gene encoding LTPs on chromosome B08 in contrast to their reported presence on chromosomes A04, A05, and A06 in B. napus (Wang et al., 2018.) A fair degree of parallelism existed for the genes associated with oil and fatty acid identified for A-genomes of B. juncea and B. napus. Candidate genes predicted in our studies were predominantly located on chromosomes A04, A06, and A09. Chromosome A09 was also important for oil content in B. napus (Xu et al., 2017; Xiao et al., 2019).

Seed meal of B. juncea is protein-rich (Grami and Stefansson, 1977; Das et al., 2009) and it is used as livestock feed depending on its nutritional value (Wanasundara, 2011). We anticipated many genes associated with plant development, transport, and protein synthesis on chromosomes A04, A06, A09, and B03. Our
in silico analysis also predicted corresponding orthologs for A03, A04, A06, and A09 for A genome of B. napus. PASTICCINO1, important for coordinating cell division and differentiation during plant development (Harrar et al., 2003), was envisioned on chromosome A04 at all N levels. Glutamine-dependent asparagine synthetase (ASN1) was envisaged on chromosome A06. It is required in nitrogen storage and transport (Canales et al., 2012). Transgenic Arabidopsis lines over-expressing ASN1 produced high seed-soluble protein (Lam et al., 2003). We also identified that GTR2 is a glucosinolate-specific transporter. It regulates the loading of glucosinolates from the apoplast into the phloem in Arabidopsis (Nour-Eldin et al., 2012). Two copies of a functional nucleotidyltransferase gene, PSPI, were annotated on chromosomes A09 and B03. PSPI encodes UDP-N-acetylglucosamine diphosphorylase 1, a fundamental precursor for glycoprotein and glycolipid synthesis (Yang et al., 2010).

Glucosinolates are the most extensively studied defense-related secondary metabolites, produced exclusively in the family Brassicaceae (Halkier and Gershenzon, 2006; Yatusevich et al., 2010). Very large numbers of genes encoding various steps of glucosinolates biosynthesis have been predicted in B. rapa (Wang et al., 2011) and B. napus (Chalhoub et al., 2014). Expression of these genes is subject to influence by temperature, nitrogen, and sulfur (Barthet and Daun, 2011). We recorded eight genes on chromosomes A04, A05, A06, B03, B04, and B06. These include HB16 and SK1, which encodes shikimate kinase (SK). It is an enzyme of shikimate pathway that directs carbon from the central metabolism pool to a broad range of secondary metabolites (Fucile et al., 2008; Tzin and Galili, 2010; Tohge et al., 2013; Averesch and Krömer, 2018). CM1, predicted on chromosomes A06 and B04, is associated with the biosynthesis of phenylalanine and tyrosine. It also participates in triptophan biosynthesis (Westfall et al., 2014; Qian et al., 2019). JMT was annotated on chromosome B03. It is critical for jasmonate-dependent induction of indole glucosinolates in Arabidopsis (Brader et al., 2001; Ku et al., 2016). LINC4 was predicted on B06. It encodes branched-chain-α-amino-acid transaminase. This enzyme catalyzes the conversion of branched-chain amino acids and α-ketoglutarate into branched chain α-keto acids and glutamate (De Kraker et al., 2007). We identified CYP81G1 on chromosome B06. It is involved in the metabolic processes of indole glucosinolates (Halkier, 1999). MYB44 (AT4G37260) was envisioned on chromosome B06. MYB44 regulates the expression of most GSL biosynthesis genes in partnership with EIN2 (Li et al., 2013). In silico analysis of B. juncea genome assembly revealed four orthologous copies (A02, A07, A09, and B06) with a coding sequence comparable to MYB44. Three major QTLs are known for GSL content in B. napus (Howell et al., 2003; Li et al., 2014; Lu et al., 2014; Qu et al., 2015). These colocalized with three orthologs of the Arabidopsis MYB28 on chromosomes A09, C02, and C09 (Chalhoub et al., 2014; Lu et al., 2014; Wang et al., 2018).

To summarize, we identified 21 orthologs of the functional candidate genes related to the biosynthesis of OIL, PTN, and GSL. As was expected for a polyploid crop, in silico analysis of the reference genome sequence revealed multiple copies of predicted genes on different A- and B- genome chromosomes. We were also able to establish LD patterns and haplotype structures for the candidate genes. The average block sizes were larger on A-genome chromosomes as compared to the B- genome chromosomes. Genetic associations differed over N levels and meta-analysis of GWAS datasets not only improved the power to detect associations but also helped to identify common SNPs. N0 proved better to unravel subtle variations for OIL. In contrast, evaluation at N100 appeared suitable for investigating PTN.

**DATA AVAILABILITY STATEMENT**

The datasets generated for this study can be found in the NCBI under bioproject PRJNA639209.

**AUTHOR CONTRIBUTIONS**

SB developed genetic resources and designed and supervised the research, edited the manuscript. VS helped in the conduct of field trials. SS performed biochemical analysis. JA conducted statistical and bioinformatics analysis. BB helped with statistical analysis. MS and AS annotated the results. MS, AS, HK, JA, and NK interpreted the results and wrote the manuscript. All authors have read and approved the version of manuscript.

**FUNDING**

The studies were financially supported by the Department of Biotechnology, Government of India in the form of the Centre of Excellence and Innovation in Biotechnology “Germplasm enhancement for crop architecture and defensive traits in Brassica juncea L. Czern. and Coss.” SB also acknowledges salary support from the Indian Council of Agricultural Research under ICAR National Professor Project “Broadening the genetic base of Indian mustard (Brassica juncea) through alien intrогressions and germplasm enhancement.”

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00744/full#supplementary-material

**TABLE S1 | Analysis of Variance (ANOVA).**

**TABLE S2 | Correlation (r) of the quality traits across years (Y1 & Y2) at two N-levels: N0 and N100.**

**FIGURE S1 | A correlation network plot based on Pearson’s correlations (r) for seed quality traits across years and N-levels. Black connecting lines denote positive and red connecting lines denote negative correlations.**

**FIGURE S2 | Q-Q plots for selection of the best fitted GWAS algorithm.**
References

Abdali, A., and Leckband, G. (2011). Rapseed breeding for oil content, quality, and sustainability. Eur. J. Lipid Sci. Technol. 113, 1198–1206. doi: 10.1002/ejlt. 201100053

Ahmadi, M., and Bahrami, M. J. (2009). Yield and Yield Components of Rapeseed as Influenced by Water Stress at Different Growth Stages and Nitrogen Levels. Am. Eur. J. Agric. Environ. Sci. 5, 755–761.

Ahkbari, I., and Banga, S. S. (2015). Genome-wide association mapping for grain yield components and root traits in Brassica juncea (L.) Czern & Coss. Mol. Breed. 35, 48–54. doi: 10.1007/s10880-015-0230-8

Akoh, C. C., Lee, G. C., Liaw, Y. C., Huang, T. H., and Shaw, J. F. (2004). GDSL family of serine esterases/lipases. Prog. Lipid Res. 43, 534–552. doi: 10.1016/j.plipres.2004.09.002

Appelqvist, A. L. (1971). Lipids in Cruciferae. VIII. The fatty acid composition of oil-rape (Brassica napus L.) protein isolate and recovery of phenolic compounds by ultrafiltration. Ind. Eng. Chem. Res. 48, 4939–4947. doi: 10.1021/ie010047q

De Kraker, J. W., Luck, K., Teztor, S., Tokuhisa, J. G., and Ghershenzon, J. (2007). Two arabidopsis genes (IMPS1 and IMPS2) encode isopropylmalate synthase, the branchpoint step in the biosynthesis of leucine. Plant Physiol. 143, 970–986. doi: 10.1104/pp.106.085555

Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. Focus 12, 13–15. doi: 10.3389/f11.2007.01179

Duitama, J., Quintero, J. C., Cruz, D. F., Quintero, C., Humman, G., Fouliqué-Moreno, M. R., et al. (2014). An integrated framework for discovery and genotyping of genomic variants from high-throughput sequencing experiments. Nucleic Acids Res. 42, e44. doi: 10.1093/nar/gkt1381

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02533.x

Fahy, I. W., Zalcman, A. T., and Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochemistry 56, 5–51. doi: 10.1016/S0031-9422(00)00316-3

Frigerrimann, H., and Gigolashvili, T. (2014). MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in Arabidopsis thaliana. Mol. Plant 7, 814–828. doi: 10.1093/mp.sau004

Fu, Y., Zhang, D., Gleeson, M., Zhang, Y., Lin, B., Hua, S., et al. (2017). Analysis of QTL for seed oil content in Brassica napus by association mapping and QTL mapping. Euphytica 213:17. doi: 10.1007/s10681-016-1817-1819

Fucile, G., Falconer, S., and Christendat, D. (2008). Evolutionary diversification of plant shikimate kinase gene duplicates. PLoS Genet. 4:e1000292. doi: 10.1371/ journal.pgen.1000292

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, b., et al. (2002). The structure of haplotype blocks in the human genome. Science 296, 2225–2229. doi: 10.1126/science.1069424

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420–3435. doi: 10.1093/nar/gkn176

Grani, B., and Stefansson, B. R. (1977). Paternal and maternal effects on protein and oil content in summer rape. Can. J. Plant Sci. 57, 945–949. doi: 10.4141/cps77-135

Gu, J., Chao, W., Wang, H., Li, Y., Li, D., and Xiang, J. (2017). Identification of the Relationship between Oil Body Morphology and Oil Content by Microstructure Comparison Combining with QTL Analysis in Brassica napus. Front Plant Sci. 7:1899. doi: 10.3389/fpls.2016.01899

Halkier, B. (1999). “glucosinolates”, in Naturally Occurring Glucosides, ed. R. Ikan, (New York, NY: John Wiley & Sons Ltd.), 193–223.

Halkier, B. A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. Annu. Rev. Plant Biol. 57, 303–333. doi: 10.1146/annurev.arplant.57.032905.105228

Han, B., and Eskin, E. (2012). Interpreting meta-analyses of genome-wide association studies. PLoS Genet. 8:e1002555. doi: 10.1371/journal.pgen.1002555

Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., et al. (2012). Associative transcriptomics of traits in the polyploid crop species Brassica napus. N. J. Genet. Genomics 37, 798–802. doi: 10.1016/j.jbg.2001.08.002

Harran, Y., Blecic, Y., Bellini, C., and Faure, J. D. (2003). Hormonal control of cell proliferation requires PASTICCINO genes. Plant Physiol. 132, 1217–1227. doi: 10.1104/pp.102.019026
Mikkilineni, V., and Rocheford, T. R. (2003). Sequence variation and genomic frontiers in genetics | www.frontiersin.org 13

Honsdorf, N., Becker, H. C., and Ecke, W. (2010). Association mapping for monodora charantia and ricinus communis seeds and determination of sites phosphorylated by plant cac2-dependent protein kinase. Biochim. Biophys. Acta Protein Struct. Mol. Enzymol. 1298, 223–240. doi: 10.1016/j.sbi.2010.06.013

Nour-Eldin, H. H., Andersen, T. G., Burow, M., Madsen, S. R., Jorgensen, M. E., Olsen, C. E., et al. (2012). NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. Nature 488, 531–534. doi: 10.1038/nature11285

Pascal, S., Bernard, A., Sorel, M., Pervent, M., Vile, D., Haslam, R. P., et al. (2013). The Arabidopsis ccr26 mutant, like the ccr mutant, is specifically affected in the very long chain fatty acid elongation process. Plant J. 73, 733–746. doi: 10.1111/tpj.12060

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. doi: 10.1086/512005

Qian, Y., Lynch, J. H., Guo, L., Rhodes, D., Morgan, J. A., and Dudareva, N. (2019). Completion of the cytosolic post-chorismate phenylalanine biosynthetic pathway in plants. Nat. Commun. 10:15. doi: 10.1038/s41467-018-07969-7962

Qu, D., Morgan, C., Shi, J., Long, Y., Liu, J., Li, R., et al. (2006). A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. Theor. Appl. Genet. 114, 67–80. doi: 10.1007/s00122-006-0411-4

Qu, C. M., Li, S. M., Duan, X. J., Fan, J. H., Jia, L. D., Zhao, H. Y., et al. (2015). Identification of candidate genes for seed glucosinolate content using association mapping in Brassica napus L. Genes 6, 1215–1229. doi: 10.3390/genes6041215

Ramsay, N., Bišt, N. C., Gupta, V., Mukhopadhyay, A., Arumugam, N., Sodhi, Y. S., et al. (2007). QTL analysis reveals context-dependent loci for seed glucosinolate trait in the oilseed Brassica juncea:importance of recurrent selection backcross scheme for the identification of “true” QTL. Theor. Appl. Genet. 116, 77–85. doi: 10.1007/s00122-006-0648-4

Roshnyara, N. R., and Schutz, M. (2014). fEGENE: a versatile tool for processing and transforming SNP datasets. PLoS One 9:e97589. doi: 10.1371/journal.pone.0097589

Rout, K., Yadav, B. G., Yadava, S. K., Mukhopadhyay, A., Gupta, V., Pental, D., et al. (2018). QTL landscape for oil content in Brassica juncea: analysis in multiple bi-parental populations in high and “0” erucic background. Front. Plant Sci. 9:1448. doi: 10.3389/fpls.2018.01448

Sen, R., Sharma, S., Kaur, G., and Banga, S. S. (2018). Near-infrared reflectance spectroscopy calibrations for assessment of oil, phenols, glucosinolates and fatty acid content in the intact seeds of oilseed Brassica species. J. Sci. Food Agric. 98, 4050–4057. doi: 10.1002/jsfa.8919

Shockley, J. M., Fulda, M. S., and Browse, J. A. (2002). Arabidopsis contains nine long-chain acyl-coenzyme A synthetase genes that participate in fatty acid and glycerolipid metabolism. Plant Physiol. 129, 1710–1722. doi: 10.1104/pp.103.020123

Smooker, A. M., Wells, R., Morgan, C., Beaudoin, F., Cho, K., Fraser, F., et al. (2011). The identification and mapping of candidate genes and QTL involved in the fatty acid desaturation pathway in Brassica napus. Theor. Appl. Genet. 122, 1075–1090. doi: 10.1007/s00122-010-1512-1515

Sonderby, I. E., Burow, M., Rowe, H. C., Kliebenstein, D. J., Halkier, B. A., Senderby, I. E., et al. (2010). A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in Arabidopsis. Plant Physiol. 153, 348–363. doi: 10.1104/pp.109.149286

Soto-Cerda, J. B., and Cloutier, S. (2012). “Association mapping in plant genomes,” in Genetic Diversity in Plants, eds A. Muhammed, R. Aksel, R. Aksel, and R. C. Von Borstel, (London: InTech), doi: 10.5772/33005

Stukey, J. E., McDonough, V. M., and Martin, C. E. (1990). The OLE1 gene of Trichoplusia ni analgesia in two broccoli cultivars following jasmonate treatment of broccoli - An overview. Arabidopsis Biochem. 98:1153. doi: 10.1002/jsfa.8919

The impact of reducing fatty acid desaturation on the composition and thermal stability of rapeseed oil. Theor. Appl. Genet. 113, 1211–1220. doi: 10.1007/s00122-006-0370-5

Theor. Appl. Genet. 127, 957–968. doi: 10.1007/s00122-014-2271-2275

Theor. Appl. Genet. 114, 411–416. doi: 10.1111/j.1399-5156.2008.0097589

Theor. Appl. Genet. 114, 67–80. doi: 10.1007/s00122-006-0411-4

Theor. Appl. Genet. 116, 77–85. doi: 10.1007/s00122-006-0648-4

Theor. Appl. Genet. 116, 77–85. doi: 10.1007/s00122-006-0648-4

Theor. Appl. Genet. 116, 77–85. doi: 10.1007/s00122-006-0648-4

Theor. Appl. Genet. 116, 77–85. doi: 10.1007/s00122-006-0648-4

Theor. Appl. Genet. 116, 77–85. doi: 10.1007/s00122-006-0648-4
Wang, X., Long, Y., Yin, Y., Zhang, C., Gan, L., Liu, L., et al. (2015). New insights into the genetic networks affecting seed fatty acid concentrations in *Brassica napus*. BMC Plant Biol. 15:91. doi: 10.1186/s12870-015-0475-4

Wang, X., Long, Y., Yin, Y., Zhang, C., Gan, L., Liu, L., et al. (2015). New insights into the genetic networks affecting seed fatty acid concentrations in *Brassica napus*. BMC Plant Biol. 15:91. doi: 10.1186/s12870-015-0475-4

Wittstock, U., and Halkier, B. A. (2002). Glucosinolate research in the *Arabidopsis era*. Trends Plant Sci. 7, 263–270. doi: 10.1016/S1360-1385(02)02273-2

Wu, G., Wu, Y., Xiao, L., Li, X., and Lu, C. (2008). Zero erucic acid trait of rapeseed (*Brassica napus L.*) results from a deletion of four base pairs in the fatty acid elongase 1 gene. Theor. Appl. Genet. 116, 491–499. doi: 10.1007/s00122-007-0685-x

Xiao, Z., Zhang, C., Tang, F., Yang, B., Zhang, L., Liu, J., et al. (2019). Identification of candidate genes controlling oil content by combination of genome-wide association and transcriptome analysis in the oilseed crop *Brassica napus*. Biotechnol. Biofuels 12:206. doi: 10.1186/s13068-019-1557-x

Xu, P., Cao, S., Hu, K., Wang, X., Huang, W., Wang, G., et al. (2017). Trilocular phenotype in *Brassica juncea* L. resulted from interruption of CLAVATA1 gene homologue (BjMCI) transcription. Sci. Rep. 7:3498.

Yadav, M., and Rana, J. S. (2018). Quantitative analysis of Sinigrin in *Brassica juncea*. J. Pharmacogn. Phytochem. 7, 948–954.

Yadava, S. K., Arumugam, N., Mukhopadhyay, A., Sadhi, Y. S., Gupta, V., Pental, D., et al. (2012). QTL mapping of yield-associated traits in *Brassica juncea*: meta-analysis and epistatic interactions using two different crosses between east European and Indian gene pool lines. Theor. Appl. Genet. 125, 1533–1564. doi: 10.1007/s00122-012-1934-1933

Yang, T., Echols, M., Martin, A., and Bar-Peled, M. (2010). Identification and characterization of a strict and a promiscuous N-acetylglucosamine-1-P uridylyltransferase in *Arabidopsis*. Biochem. J. 430, 275–284. doi: 10.1042/BJ20100315

Yao, K., Bacchetti, R. G., Lockhart, K. M., Friesen, L. J., Potts, D. A., Covello, P. S., et al. (2003). Expression of the Arabidopsis ADS1 gene in *Brassica juncea* results in a decreased level of total saturated fatty acids. Plant Biotechnol. J. 1, 221–229. doi: 10.1046/j.1467-7652.2003.00021.x

Yatusevich, R., Mugford, S. G., Matthewman, C., Gigolashvili, T., Frerigmann, H., Delaney, S., et al. (2010). Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in *Arabidopsis thaliana*. Plant J. 62, 1–11. doi: 10.1111/j.1365-313X.2009.04118.x

Zhang, D., and Hamamura, Y. (2004). Phenolics, ascorbic acid, carotenoids and antioxidative activity of broccoli and their changes during conventional and microwave cooking. Food Chem. 88, 503–509. doi: 10.1016/j.foodchem.2004.01.065

Zhang, W., Hu, D., Raman, R., Guo, S., Wei, Z., Shen, X., et al. (2017). Investigation of the genetic diversity and quantitative trait loci accounting for important agronomic and seed quality traits in Brassica carinata. Front. Plant Sci. 8:307. doi: 10.3389/fpls.2017.00307

Zhao, Q., Huang, X., Lin, Z., and Han, B. (2010). SEG-Map: a novel software for genotype calling and genetic Map construction from next-generation sequencing. Rice 3, 98–102. doi: 10.1007/s12284-010-0951-x

Zhao, Y. T., Wang, M., Fu, S. X., Yang, W. C., Qi, C. K., and Wang, X. J. (2012). Small RNA profiling in two *Brassica napus* cultivars identifies MicroRNAs with oil production- and development-correlated expression and new small RNA classes. Plant Physiol. 158, 813–823. doi: 10.1104/pp.111.187666

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

**Copyright © 2020 Akhatar, Singh, Sharma, Kaur, Kaur, Sharma, Bharti, Sardana and Banga. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.**