

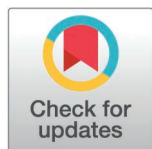
RESEARCH ARTICLE

Genome-wide association study uncovers key genomic regions governing agro-morphological and quality traits in Indian mustard [*Brassica juncea* (L.) Czern. and Coss.]

Manoj Kumar Patel¹, Navinder Saini^{1*}, Yashpal Taak¹, Sneha Adhikari¹, Rajat Chaudhary¹, Priya Pardeshi¹, Sudhakar Reddy Basu¹, Masochon Zimik¹, Sangita Yadav², K. K. Vinod¹, Sujata Vasudev¹, Devendra Kumar Yadava^{3*}

1 Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India, **2** Division of Seed Science and Technology, ICAR-Indian Agricultural Research Institute, New Delhi, India, **3** Indian Council of Agricultural Research, New Delhi, India

* dkygenet@gmail.com (DKY); navin12@gmail.com (NS)



OPEN ACCESS

Citation: Patel MK, Saini N, Taak Y, Adhikari S, Chaudhary R, Pardeshi P, et al. (2025) Genome-wide association study uncovers key genomic regions governing agro-morphological and quality traits in Indian mustard [*Brassica juncea* (L.) Czern. and Coss.]. PLoS One 20(4): e0322120. <https://doi.org/10.1371/journal.pone.0322120>

Editor: Aditya Pratap, ICAR-Indian Institute of Pulses Research, INDIA

Received: March 5, 2025

Accepted: March 15, 2025

Published: April 24, 2025

Copyright: © 2025 Patel 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 manuscript and its Supporting Information files.

Funding: The funding support of this work was provided by Incentivizing Research

Abstract

In Indian mustard, improving agro-morphological and quality traits through conventional methods are both cumbersome and resource-intensive. Marker-aided breeding presents a promising solution to these challenges. Hence, the present research aimed to identify genomic regions governing agro-morphological and quality traits using genome-wide association studies (GWAS). The GWAS panel comprised 142 diverse genotypes of Indian mustard were evaluated for 20 different agro-morphological and quality traits, revealing significant difference among genotypes. Subsequently, the GWAS panel genotyped using the *Brassica* 90K SNP array (Illumina). Structure and diversity analysis grouped the GWAS panel into 3 sub-populations or groups, and LD decay of 1.05 Mb was confirmed through genotypic analysis. GWAS using the BLINK model revealed a total of 49 marker-trait associations (MTAs), in which 28 and 21 MTAs were observed during *rabi* 2020–21 and *rabi* 2021–22 for various agro-morphological and quality traits, respectively. Amongst them, twelve MTAs demonstrated stable associations with the studied traits, including days to 50% flowering (DF), days to 100% flower termination (DFT), days to maturity (DM), plant height (PH), main shoot length (MSL), siliqua length (SL), seeds per siliqua (SPS), oil content (OC), and glucosinolates content (Glu) in both years. Moreover, *in silico* analysis of nearby regions of these stable SNPs revealed their association with 31 candidate genes known to be involved in various molecular, physiological, and biochemical pathways relevant to the studied traits. These genes can be further characterized and deciphered for more precise utilization in breeding programs in the future.

in Agriculture (IRA), ICAR, New Delhi and National Agricultural Higher Education Project - Centre for Advanced Agricultural Science and Technology (NAHEP-CAAST), ICAR, New Delhi. The funders played a significant role in defining the research mandates and providing essential facilities for the present study.

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

Introduction

Indian mustard [*Brassica juncea* (L.) Czern. & Coss.; AABB; $2n=36$], is an amphidiploid species of Brassicas belonging to the family Cruciferae. It is predominantly a self-pollinating crop; however, based on bee activities, up to 18% cross-pollination has been reported under natural field conditions [1]. Indian mustard is a predominant species among the rapeseed-mustard group of crops in India and accounts for more than 90% of its total acreage [2]. It is mainly used for edible oil, especially in India, China, and European countries. In addition, it is also utilized as condiments, salad, leafy vegetables, green manure, fodder, and biofuels [3,4]. Globally, India holds the third position in acreage as well as production after Canada and China. During 2022–23, the world's rapeseed-mustard seed production was about 87.06 million tones (mt) in 41.08 million hectares (mha) area with productivity of 2.08 t/ha [5]; whereas, in India, rapeseed-mustard production was 11.70 mt from an area of 9.2 mha with an average productivity of 1.27 t/ha [5]. Therefore, there is a huge scope (~0.81 t/ha) of improvement in seed productivity under Indian conditions.

In the context of Brassicas, the main challenge for present-day plant breeders is to develop high-yielding varieties to meet the demands of growing global populations. Therefore, breeding programmes of Indian mustard are mainly targeting improvement in seed productivity and oil quality. In Brassicas, seed yield is determined by various agro-morphological traits such as plant height, main shoot length, siliquae per plant, silique length, seeds per silique, etc. [6,7]; thus, these traits have significant role in improving the seed yield. On the other hand, oil or meal quality is determined by fatty acids, proteins, and glucosinolates constitution [4,8]. Amongst these, the fatty acid composition is a chief determinant of oil quality [9]. It has been reported that elevated amounts of saturated fats and erucic acid resulted in high cholesterol levels in humans [10]. Additionally, glucosinolates, another important determinant of oil quality, are considered as an anti-nutritional factor in animal feed, which exerts adverse effects on animals if present in a high proportion [11]. Therefore, in Brassicas, these traits are highly crucial for the improvement of oil quality. However, the desired modification of these traits using conventional means is very challenging and resource-demanding due to their complex inheritance and environmental influence. Therefore, the identification of genomic regions/quantitative trait loci (QTL) is of prime importance, as it will further aid in molecular breeding for efficient and accelerated crop improvement.

Advancements in molecular techniques have led to the evolution of markers from SSR to DArT and ultimately SNPs, which are now widely used in applications like QTL mapping, GWAS, and genomic selection [12,13]. Genotype-by-sequencing (GBS) is a popular SNP genotyping method, but high percentage of missing values, low sequencing coverage, unknown marker positions, and complex statistical analysis make its use limited [14,15]. These limitations are effectively addressed by array-based SNP genotyping, offering improved reliability and ease of analysis [16,17]. QTL mapping using biparental populations has limited precision due to restricted recombination, requiring additional efforts for fine mapping [3,18]. In

contrast, Association Mapping (AM) utilizes natural populations and historical recombination, providing higher resolution and capturing a broader genetic variation. GWAS, a key AM approach, has become vital for identifying the genetic basis of complex traits in crops [19–21]. The physical extent of linkage disequilibrium (LD) around a gene significantly influences the efficiency of association mapping. This extent is shaped by various factors, such as the rate of outcrossing, selection pressure on specific genomic regions, recombination rates, chromosomal location, and population size and structure [22]. Therefore, accounting for these factors is essential to ensure reliable results in association mapping. With the increasing adoption of whole-genome sequencing (WGS), the power and resolution of GWAS have been greatly enhanced [23]. Consequently, association mapping using SNP markers has emerged as one of the most effective strategies for identifying genomic regions and candidate genes.

In crucifers, several candidate genes associated with various agro-morphological and quality traits have been reported. For instance, *WRKY* transcription factors and *BTB/POZ* domain family genes (*BOP* genes) regulate flowering time in *Arabidopsis thaliana* [24,25]. Additionally, *Hd3b*, a rice ortholog of the *EFL3* gene in *A. thaliana*, has been found to cause delayed flowering under long-day field conditions in rice [26]. Genes such as *IAA17* and *B-box zinc finger protein 24* are involved in plant growth and development [27,28]. In *B. napus*, Pal et al. [29] identified the candidate gene *GATA15* associated with seeds per silique. For quality traits, *WRI1* and oleosin genes play a crucial role in oil accumulation [30,31]. In *Brassica juncea*, the *FAE* gene regulates erucic acid content [32,33] while, *MYB28*, *MYB29*, *CYP79F1*, *GSL-ELONG* and *GSL-ALK* genes are associated with glucosinolates biosynthesis [34]. Moreover, in Brassicas, GWAS played a paramount role in identifying the important gene(s) or genomic regions for different agro-morphological and quality parameters including flowering traits [35], plant height [35,36], primary branches per plant [36], silique length [37], seeds per silique and thousand seed weight [38], seed yield components [3,29,39], oil content [4,40,41], erucic acid [42] and glucosinolates [4,43,44]. However, in *B. juncea*, only a few reports on GWAS for agro-morphological [3,35,39] and quality traits [4,45] are available and to date, no study on GWAS using array-based SNP genotyping has been reported.

Building on this background, the present study utilized a germplasm panel of 142 genotypes for GWAS to identify MTAs and candidate genes for key agro-morphological and quality traits. The panel was evaluated across two consecutive seasons (*rabi*, 2020–2021 and *rabi*, 2021–2022) and genotyped using the *Brassica* 90K SNP array (Illumina). The study identified important stable MTAs, along with associated candidate genes, which hold great potential for advancing molecular breeding efforts in Indian mustard.

Materials and methods

Germplasm panel for GWAS

The GWAS panel consisting of 142 diverse genotypes of Indian mustard contains Indigenous lines, developed varieties, exotic lines, advanced breeding material as well as introgressed lines, which provided from Panjab Agricultural University, Ludhiana collected under the ICAR - National Agricultural Science Fund (NASF) project (S1 Table).

Design of experiments

Germplasm assembly of 142 Indian mustard genotypes was grown at research farm, ICAR-Indian Agricultural Research Institute, New Delhi during *rabi*, 2020–21 and *rabi*, 2021–22 in augmented block design with seven checks (DRMRIJ-31, PM-30, RLC-3, RH-749, NRCHB-101, PM-25 and PM-28) and four blocks. Check varieties included conventional as well as quality mustard varieties (low erucic acid/glucosinolates content, i.e., single zero and double zero type). In both seasons, each genotype had grown in two rows of 5m length with 45 cm × 15 cm spacing. The germplasm panel was evaluated for different agro-morphological and quality traits. To ensure better crop establishment, two irrigations were provided using the flood method. The first irrigation was applied at 35–40 days after sowing (DAS), and the second at 85–90 DAS. Fertilizers were applied at rates of 60 kg/ha N, 40 kg/ha P, and 40 kg/ha K. Phosphorus and potassium were incorporated

into the soil prior to sowing, while nitrogen was top-dressed at various growth stages. All other agronomic practices, including thinning and weeding, were carried out following the recommended package of practices.

Record of observations

Observations in the present investigation were recorded on 12 different agro-morphological traits, which include days to 50% flowering (DF), days to 100% flower termination (DFT), days to maturity (DM), plant height at maturity (PH; cm), main shoot length (MSL; cm), primary branches per plant (PBPP), secondary branches per plant (SBPP), siliquae per plant (SPP), silique length (SL; cm), seeds per silique (SPS), seed yield per plant (SYPP; g), biological yield per plant (BYPP; g) and 8 oil quality traits, i.e., oil content (OC; %), glucosinolates (Glu; $\mu\text{mol/g}$), and fatty acids content (%) namely, palmitic acid (PA), oleic acid (OA), linoleic acid (LliA), linolenic acid (LlnA), eicosenoic acid (EcA) and erucic acid (ErA). Observations for DF, DFT, and DM were recorded from the whole plot; while that of remaining traits were documented on five randomly selected plants from each genotype. Five randomly selected plants of each genotype were harvested at the time of maturity and sun-dried until no residual moisture remained, then using weighing balance BYPP was recorded in grams. After the record of biological yield, plants were threshed manually and their seed yield was recorded as SYPP using a weighing balance. Furthermore, phenotyping of seed-related parameters such as OC, Glu, and fatty acid contents was carried out on freshly threshed seeds from each selected plant of a genotype. Oil content has been estimated by the non-destructive method using Near InfraRed spectroscopy (Perten, DA 7250). Glucosinolates content was estimated by the simple spectrophotometric method suggested by Mawlong et al. [46]. In fatty acid profiling, the preparation of methyl esters was carried out as per the standard protocol suggested by Vasudev et al. [8]; while fatty acid peaks were captured on Gas Chromatograph (Perkin Elmer Claurus 500) and the amount of fatty acid is calculated by triangulation method.

Molecular analysis and SNP genotyping

Genomic DNA was extracted from tender leaves of 142 mustard genotypes using the CTAB (Cetyl Trimethyl Ammonium Bromide) method [47]. The DNA quality was analyzed using a 0.8% agarose gel with a known and standard DNA (uncut lambda DNA) and samples were quantified using Nanodrop (NanoDrop 2000, Thermo Scientific, USA). Approximately 5 μg DNA from each genotype was used for SNP genotyping using Brassica 90K SNP array, Illumina iScan, Infinium assays (AgriGenome Labs Pvt. Ltd.). A total of 77970 SNPs were obtained. Subsequently, genotypic data was subjected to quality check and filtering, wherein 15219 high-quality, polymorphic SNPs with a minor allele frequency of more than 5% and missing data less than 10% were retained for Linkage disequilibrium (LD) and GWAS analysis (S2 Table). However, for structure, diversity, and analysis of molecular variance (AMOVA), markers that are closely located and are in LD (<1.05 MB in the present study) were removed and 620 widely distributed SNPs, at a distance of ~1.05 MB were used.

Data analysis

Phenotypic evaluation of GWAS panel

The season-wise estimation of the analysis of variance (ANOVA) and descriptive statistics of the phenotypic data were performed using the “augmented RCBD” R package (Version 4.2.2) [48]. Individual and combined best linear unbiased predictors (BLUPs) across the seasons (*rabi*, 2020–2021 and *rabi*, 2021–2022) were analyzed using software PB tools (version 1.4. 2014, *Biometrics and Breeding Informatics, PBGB Division, IIRRI*) considering genotypes as random effects. Subsequently, Individual BLUP values were used as phenotypic input for GWAS analysis. Whereas, combined BLUPs were used for Principal Component Analysis (PCA) of phenotypic data using R packages FactoMineR and Factoextra [49,50].

Structure, diversity, and AMOVA of GWAS panel

Identified 620 SNPs were subjected to structure analysis using structure software (v2.3.4) [51]. The allele admixture model was used for structure analysis, wherein the burn-in period and MCMC repeats were kept at 1,00,000. The number

of iterations was kept three and the number of sub-populations (k) ranged from 1 to 10. An appropriate number of delta k (sub-populations) was determined using Evanno method [52] with the help of the online tool Structure Harvester (<http://taylor0.biology.ucla.edu/structureHarvester>). To estimate the variance within and among sub-population, analysis of molecular variance (AMOVA) was performed using R based package “Poppr” [53]. Thereafter, neighbor-joining tree was constructed using Tassel 5 (v5.2.75) [54] for studying molecular diversity using the same set of SNPs.

LD decay analysis

Linkage disequilibrium (LD) values of filtered SNPs were estimated using the r^2 method via software Tassel 5 (v5.2.75) [54]. Further, the LD decay plot was obtained by plotting r^2 values against the physical distance between markers (in bp) using the R package as per [55] and a standard cutoff value of $r^2 = 0.1$ was followed to estimate LD decay [56].

GWAS analysis

GWAS analysis was done using R based package “Genome Association and Prediction Integrated Tool” (GAPIT) [57] using Bayesian-information and linkage-disequilibrium iteratively nested key-way model (BLINK) [58]. PCA using genotypic data is a good indicator of population structure in association mapping. Therefore, the first three principal components were generated and used as covariate in GWAS analysis using the GAPIT tool. The fitness of the GWAS model was determined by Q-Q plot plotted between observed and expected $-\log_{10}(p)$ values. The threshold for significant marker-trait associations (MTAs) was determined using Bonferroni corrected p -value (effective threshold of $-\log_{10}(p) = 5.48$). However, for quantitative traits, Bonferroni threshold can be too conservative due to the involvement of minor genes [59,60]. Therefore, a less stringent criterion (p -value < 0.001) was adopted to identify stable SNPs.

Gene annotations and *in silico* gene expression analysis

Stable-associated SNPs were blasted on whole genome shotgun contigs of *B. juncea* var. Varuna using the NCBI blast tool. The genomic region of 1.05 Mb flanking (upstream and downstream) the SNP position has been used and subjected to gene prediction using the online tool Softberry FGENESH [61]. The length of the region flanking the SNPs was taken as the value of linkage-disequilibrium decay (LD decay) in the current experiment, i.e., 1.05 Mb. Further, these predicted genes of *B. juncea* were annotated using *B. napus* var. De Ae genome (NCBI GENBANK) as a reference. The orthologs of annotated candidate genes were identified in the *Arabidopsis* genome using online tool Ortho DB [62] and these genes were subjected to *in silico* gene expression analysis using Klepikova *Arabidopsis* Atlas eFB Browser [63]. Subsequently, the putative function of candidate genes was predicted through literature search and significant *in silico* expression in the concerned tissues.

Results

Meteorological observations during *rabi* 2020–21 and 2021–22

Observations on weather parameters have provided the extent of meteorological variations observed during *rabi* 2020–21 and 2021–22 (Fig 1). A higher range of minimum and maximum temperatures was observed during *rabi* 2021–22, with a value of 1.5 to 23.8 °C and 11.6 to 38 °C, respectively; compared to that of *rabi* 2020–21. Similarly, higher rainfall was observed during *rabi* 2021–22 (mean = 1.66 mm and range = 0.00 to 69.2 mm) compared to *rabi* 2020–21 (mean = 0.40 mm and range = 0.00 to 16.9 mm). In *rabi* 2020–21, significant amount of rainfall was observed during 78–82 days after sowing (DAS; silique development stage), while in *rabi* 2021–22, it was more prominent during early germination (7 DAS), early vegetative (14 DAS), silique development (86–90 DAS; 103 DAS) and maturity phase (120 DAS). Conversely, mean sunshine duration was observed in *rabi* 2020–21 (5.34 hr), compared to *rabi* 2021–22 (5.23 hr) (S3 Table).

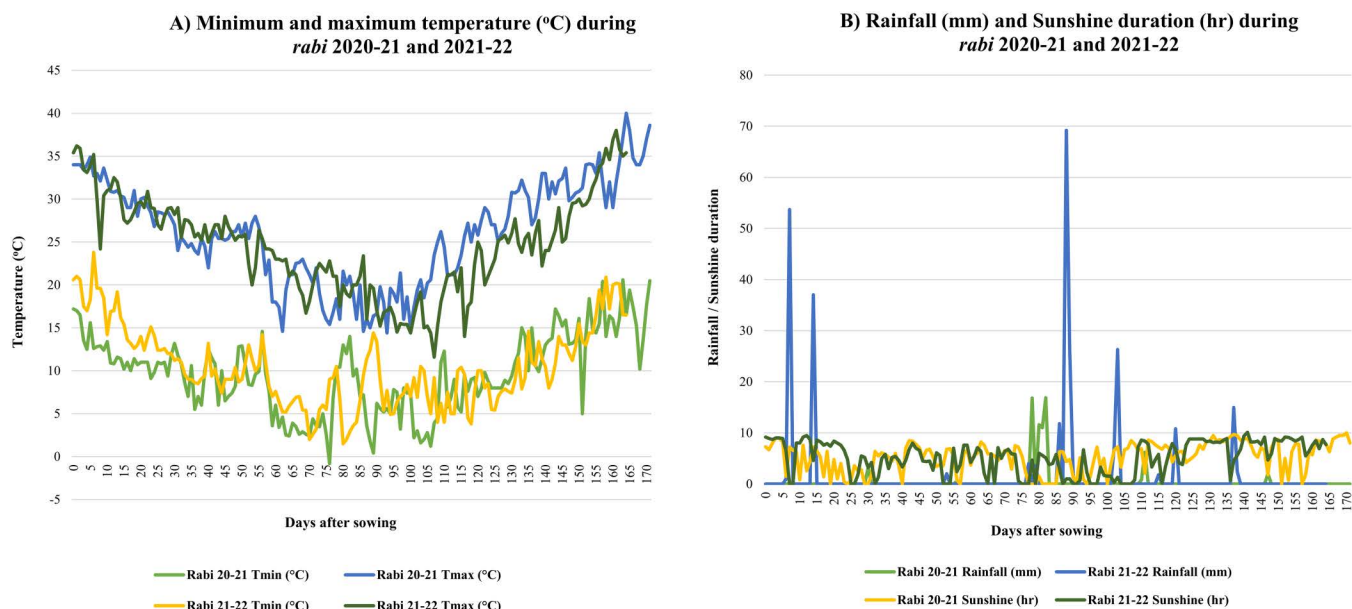


Fig 1. Meteorological variation observed during rabi 2020-21 and 2021-22; A) minimum and maximum temperature range, B) Rainfall pattern and Sunshine duration.

<https://doi.org/10.1371/journal.pone.0322120.g001>

Phenotypic evaluation of GWAS panel for agro-morphological and oil quality attributes

Season-wise (*Rabi*, 2020–21 and *Rabi*, 2021–22) analysis of variance (ANOVA) revealed significant differences among the genotypes for all of the studied traits (Table 1). It shows the existence of substantial variation among genotypes. In *rabi* 2020–21, higher mean values were observed for DF (53.82 days), DFT (98.00 days), MSL (79.12 cm), PBPP (4.62), SBPP (13.18), SL (3.67 cm), SYPP (18.55 g), BYPP (79.86 g), OC (38.40%), OA (23.70%), EcA (4.95%), and ErA (23.32%), while for *rabi* 2021–22, it was higher for DM (143.17 days), PH (234.04 cm), SPP (427.51), SPS (14.21), Glu (81.20 $\mu\text{mol/g}$), PA (9.40%), LliA (30.25%), and LlnA (15.27%). The summary statistics of the germplasm panel have shown a wide range for the evaluated traits suggesting greater diversity in the GWAS panel (Table 1). In *rabi* 2020–21, a wider range was observed for DF (38.68–112.39 days), DM (126.82–171.11 days), PH (141.25–295.59 cm), MSL (28.93–118.13 cm), PBPP (1.99–12.27), SBPP (7.18–34.95), BYPP (28.43–166.06 g), OC (28.89–46.13%), EcA (0.4–10.37%) and ErA (0.72–40.04%), while remaining traits such as DFT (65.61–143.04 days), SPP (201.45–693.99), SL (1.85–5.39 cm), SPS (8.08–19.93), SYPP (3.22–35.10 g), Glu (8.43–132.90 $\mu\text{mol/g}$), PA (5.53–14.99%), OA (7.98–47.09%), LliA (21.55–46.46%), and LlnA (9.76–23.40%) have shown a higher range in *rabi* 2021–22.

Furthermore, PCA of agro-morphological and oil quality attributes gave an insight into a better understanding of phenotypic variation. For agro-morphological traits, first two principal components, PC1 (51.6%) and PC2 (18.3%) cumulatively explain 69.9% of the total variation, (Fig 2A, Fig 2B). It has been observed that DF, DFT, DM, PH, MSL, PBPP, and SBPP have more contribution to PC1, while SYPP has more contribution to PC2. Interestingly, SPP, SL, SPS, and BYPP have considerable contributions to both PCs. In PC1, most of the traits (DF, DFT, DM, PH, PBPP, SBPP, SPP, SYPP, and BYPP) contributed positively, while MSL, SL, and SPS contributed negatively. Similarly, in PC2, PH, MSL, PBPP, SBPP, SPP, SL, SPS, SYPP, and BYPP made positive contribution, whereas flowering traits (DF, DFT, and DM) contributed negatively (Fig 2A). However, for oil quality attributes, first two PCs cumulatively explained 63.4% total variation, with individual contribution of 44.6 and 18.8% by PC1 and PC2, respectively (Fig 2C, Fig 2D). Individual fatty acids, e.g., OA, EcA, and ErA had more contribution towards PC1, while OC had more contribution to PC2. Moreover, the remaining traits

Table 1. Descriptive statistics of GWAS panel for different agro-morphological and quality traits evaluated during *Rabi*, 2020-21 and *Rabi*, 2021-22.

Traits	Genotypic MSS		CV		Mean		Range	
	<i>Rabi</i> 20-21	<i>Rabi</i> 21-22	<i>Rabi</i> 20-21	<i>Rabi</i> 21-22	<i>Rabi</i> 20-21	<i>Rabi</i> 21-22	<i>Rabi</i> 20-21	<i>Rabi</i> 21-22
DF	140.23 **	113.60 **	3.41	4.61	53.82	51.99	38.68–112.39	35.07–99.79
DFT	142.50 **	198.28 **	4.05	6.59	98.00	91.08	81.39–138.54	65.61–143.04
DM	48.79 **	52.80 **	2.29	2.21	141.04	143.17	126.82–171.11	121.39–164.39
PH	811.02 **	656.31 **	4.70	5.20	214.97	234.04	141.25–295.59	167.04–293.94
MSL	267.06 **	256.80 **	7.99	8.99	79.12	76.86	28.93–118.13	30.24–113.81
PBPP	2.38 **	1.09 **	11.08	11.02	4.62	4.61	1.99–12.27	2.98–9.52
SBPP	14.17 **	9.22 **	11.11	10.37	13.18	13.06	7.18–34.95	5.48–21.54
SPP	11739.22 **	10741.54 **	15.33	13.90	398.78	427.51	172.21–644.99	201.45–693.99
SL	0.48 **	0.58 **	6.54	8.64	3.67	3.58	2.37–5.16	1.85–5.39
SPS	2.85 *	3.62*	7.50	9.33	14.16	14.21	9.57–18.56	8.08–19.93
SYPP	40.21 **	47.20**	15.63	16.13	18.55	16.92	4.89–36.14	3.22–35.1
BYPP	582.27 **	660.87 **	14.27	14.82	79.86	79.65	28.43–166.06	39.7–151.29
OC	10.29 **	8.01 **	3.59	3.30	38.40	37.30	28.89–46.13	29.65–43.82
Glu	536.64 **	658.94 **	10.24	10.02	76.37	81.20	7.58–126.64	8.43–132.9
PA	2.61 **	4.04 **	7.36	9.29	7.72	9.40	3.81–11.95	5.53–14.99
OA	86.07 **	65.88 **	6.04	7.38	23.70	23.19	12.96–49.15	7.98–47.09
LliA	21.11 **	19.79 **	8.71	6.45	26.95	30.25	19.09–41.25	21.55–46.46
LlnA	5.24 **	7.70 **	7.53	6.17	13.28	15.27	8.73–21.56	9.76–23.4
EcA	4.50 **	4.65 **	10.05	10.00	4.95	3.61	0.4–10.37	0.2–9.31
ErA	113.53 **	88.79 **	7.65	8.51	23.32	18.59	0.72–40.04	0.3–35.21

DF = Days to 50% flowering, DFT = Days to 100% flower termination, DM = Days to maturity, PH = Plant height, MSL = Main shoot length, PBPP = Primary branches per plant, SBPP = Secondary branches per plant, SPP = Siliquae per plant, SL = Siliqua length, SPS = Seeds per siliqua, SYPP = Seed yield per plant, BYPP = Biological yield per plant, OC = Oil content, Glu = Glucosinolates, PA = Palmitic acid, OA = Oleic acid, LliA = Linoleic acid, LlnA = Linolenic acid, EcA = Eicosenoic acid, ErA = Erucic acid, Genotypic MSS = Genotypic mean sum of squares, CV = Coefficient of variation

*, ** indicate significant difference at 5% and 1% level of significance, respectively.

<https://doi.org/10.1371/journal.pone.0322120.t001>

including Glu, PA, LliA, and LlnA exhibited remarkable contribution to both PCs. The contribution of most of the variables was positive to PC1 except for OA, LliA, and LlnA, whereas, only PA, LliA, and EcA contributed negatively to PC2 (Fig 2C). Directions of eigenvectors represented the correlation among variables. In the current study, SYPP was positively correlated with MSL, SBPP, SPP, SL, and BYPP. Flowering traits (DF, DFT, and DM) were positively correlated with each other. Conversely, quality parameters such as ErA was positively correlated with PA, EcA, and Glu, whereas negative correlation was observed with OA, LliA, and LlnA (Fig 2).

Population structure, diversity, analysis of molecular variance and LD decay of GWAS panel

The allele admixture model of structure analysis exhibited maximum Δk value at $k = 3$, indicating GWAS panel consists of three sub-populations (SP; Fig 3A, Fig 3B). SP1, SP2, and SP3 comprised 121, 9, and 12 genotypes, with respective allelic contributions of 73.80%, 8.60%, and 17.60% to the whole germplasm panel (Table 2). SP1 comprised most of the conventional genotypes, which included both exotic and indigenous types. Whereas, SP2 mainly composed of Indian germplasm (IC-597869, RC-371-1, and RC-132) and quality mustard (LES-54, ELM-123, RH-801, Pusa Karishma, LET-18, and TERI). Similarly, SP3 consists of Australian germplasm (AJ-11, JM-06010–1), quality mustard (PM-29, EC-597318, PDZ-1/PDZM-31, PDZ-11/PDZM-33, JC-33, and RLC-3) and European genotypes (Donskaza and Heera), etc. Moreover,

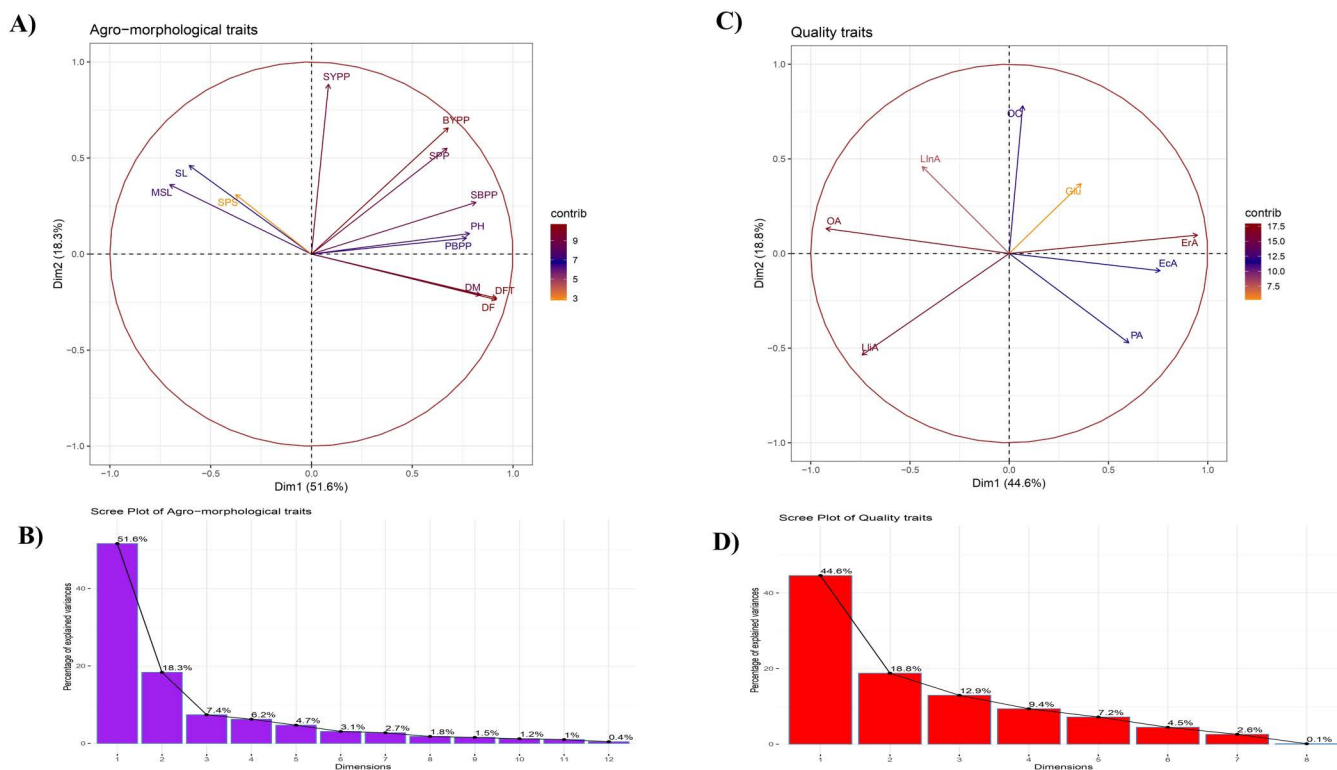


Fig 2. Principal component analysis (PCA); (A) PCA biplot of agro-morphological attributes, (B) Scree plot of agro-morphological traits showing total variation explained by individual PCs, (C) PCA biplot of quality attributes, (D) Scree plot of quality parameters showing total variation explained by individual PCs.

<https://doi.org/10.1371/journal.pone.0322120.g002>

the maximum average distance was observed in SP2 (0.59), whereas, it was least in SP1 (0.17). In contrast, the Fixation Index (F_{st}) was maximum in SP1 (0.76) and was lowest in the case of SP2 (0.15; Table 2). In the GWAS panel, 24 conventional genotypes of SP1, one Indian genotype (RC-371-1) of SP2, and two European genotypes (Heera and Donskaza) of SP3 did not have the allele admixtures while the rest of the genotypes contain allele admixtures (*alleles* from more than one sub-population). The population structure analysis provided information about the molecular relationship among genotypes. To partition total genotypic variation into within and among sub-populations molecular variance (AMOVA) analysis was performed, wherein 36.95% variation was observed among populations, and the remaining variation (63.05%) was attributed to within sub-populations (Table 2). Furthermore, to get a better insight into molecular diversity, a tree-based diversity analysis of the GWAS panel *via* neighbor-joining method was performed. It was observed that genotypes belonging to the same sub-population were more closely related; whereas, genotypes belonging to different sub-populations were more diverse. In diversity analysis, three major clusters were detected as observed in structure analysis (Fig 4). In the current study, the LD decay analysis revealed that up to 1.05Mb (1053225bp), the LD value persists, after which it begins to decay (Fig 5). Markers located within a physical distance of less than 1.05Mb tended to be in LD; therefore, this region was utilized for gene annotation of stable Marker-Trait Associations (MTAs).

Marker-trait associations (MTAs) of agro-morphological and quality attributes

GWAS analysis was conducted using the BLINK model in the R-based package GAPIT on various agro-morphological and oil quality attributes. A total of 49 SNPs were identified, with 28 and 21 SNPs detected in the *rabi* seasons of

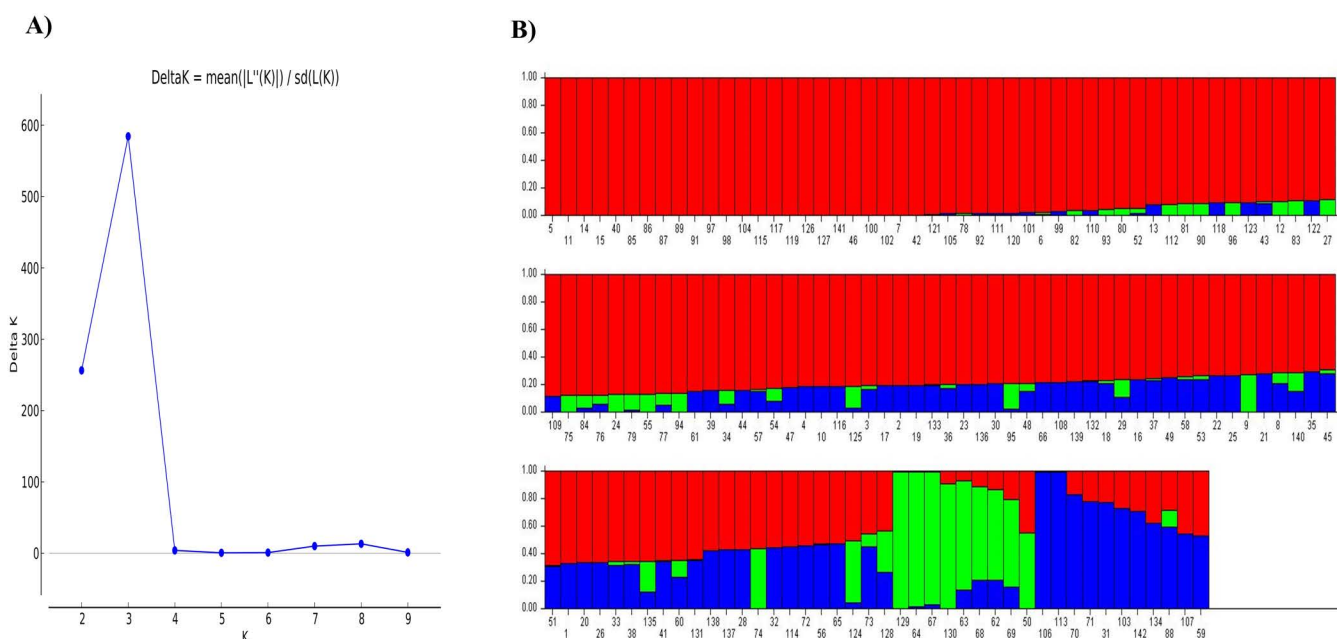


Fig 3. Population structure analysis; (A) Delta-K value to optimize number of sub-populations, (B) Grouping of GWAS panel based on sub-populations.

<https://doi.org/10.1371/journal.pone.0322120.g003>

Table 2. Analysis of molecular variance and population structure analysis.

Source	Df	SS	MSS	Estimated Variance	% contribution
Among populations	2	5752.39	2876.20	141.26	36.95
Within populations	139	33507.95	241.06	241.06	63.05
Total	141	39260.34	278.44	382.32	100
Sub-populations	Number of individuals	Average distance (He)	Fst	Overall proportion	
SP1	121	0.17	0.76	73.80%	
SP2	09	0.59	0.15	8.60%	
SP3	12	0.28	0.65	17.60%	

Df = Degrees of freedom, SS = Sum of squares, MSS = Mean sum of squares, % = Percent contribution of total variance

<https://doi.org/10.1371/journal.pone.0322120.t002>

2020–21 and 2021–22, respectively (Table 3, Fig 6 and 7). Among these, 16 and 9 SNPs were specific to the *rabi* 2020–21 and 2021–22, respectively, while the remaining 12 SNPs were common to both seasons and were utilized for further analysis. In the *rabi* 2020–21, two MTAs were identified for each of DF (Bn-A03-p19798208 and Bn-scaff_18360_1), DFT (Bc-B6-p26812452 and Bn-A05-p21930978), SL (Bn-A03-p21210189 and Bn-A08-p19365594), SPS (Bn-A03-p21615990 and Bn-A06-p2234696), and LlnA (Bj-B7-p38672208 and Bn-A02-p3300731). Conversely, single SNPs were associated with traits including DM (Bn-A09-p36455112), PH (Bn-A09-p36455112), MSL (Bn-A03-p28109103), OC (Bn-A02-p636428), and Glu (Bn-A08-p8658074). Additionally, traits such as OA exhibited seven MTAs (Bn-A03-p15957657, Bj-B4-p25774828, Bj-B4-p9000035, Bj-B4-p17202410, Bj-B3-p25944867, Bj-B1-p2664020, and Bj-B8-p41554126), while ErA displayed six (Bj-B7-p19658501, Bn-A02-p25655528, Bn-A08-p10822662, Bj-B4-p17190386, Bj-B6-p37724780, and Bn-A02-p5612147). In *rabi* 2020–21, two SNPs were identified for each of DFT (Bc-B6-p26812452 and Bn-A05-p21930978), SL (Bn-A03-p21210189 and Bn-A08-p19365594), SPS (Bn-A03-p21615990

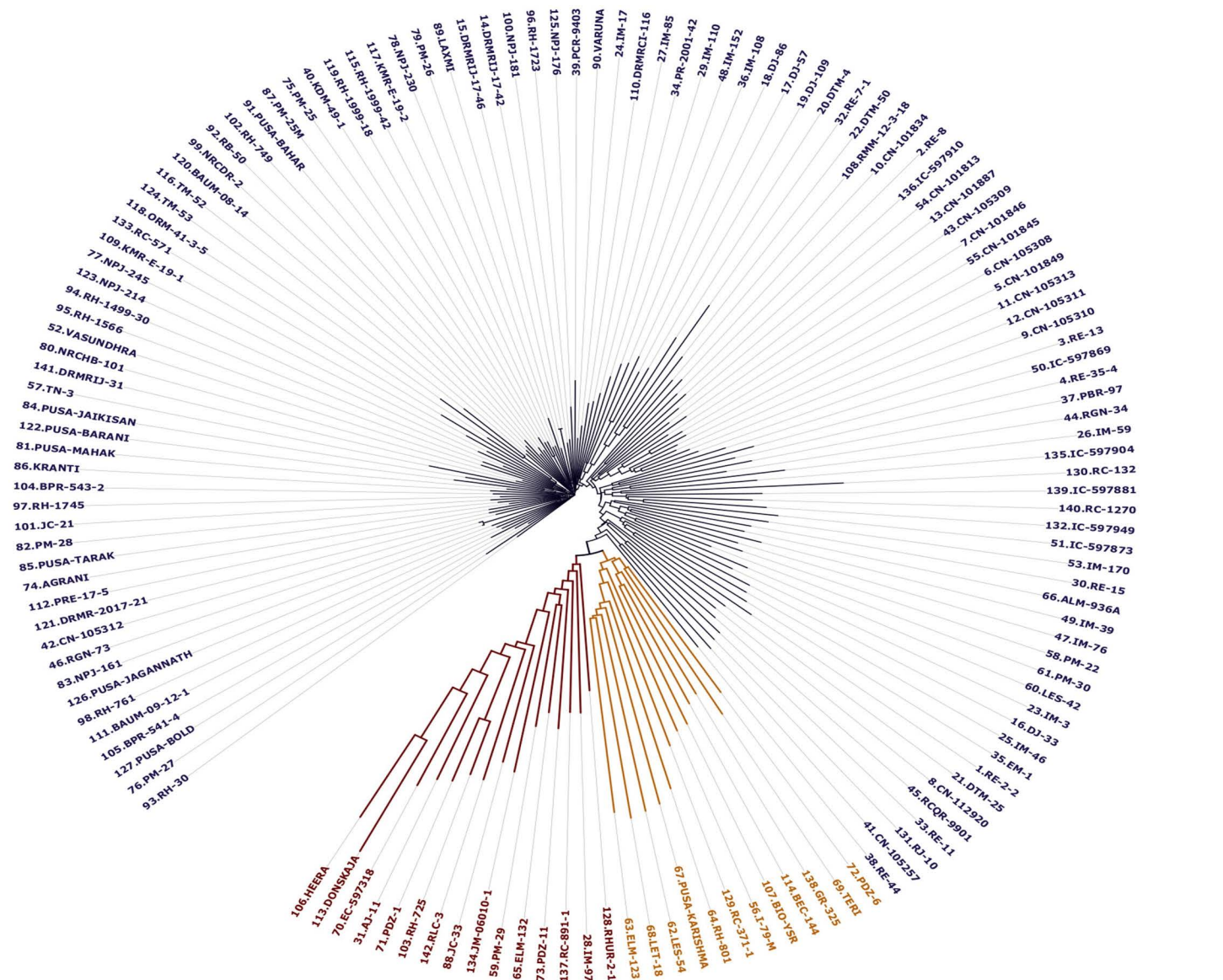


Fig 4. Neighbour joining diversity tree showing molecular diversity among GWAS panel. Genotypes highlighted with different colours denoting different clusters.

<https://doi.org/10.1371/journal.pone.0322120.g004>

and Bn-A06-p2234696), OA (Bn-A08-p4254041 and Bj-B8-p3030447), EcA (Bn-A08-p7563202 and Bj-B3-p25039645) and ErA (Bj-B8-p41554126 and Bn-A08-p4001683), while single SNPs were found for DF (Bn-A03-p19798208), DM (Bn-A09-p36455112), PH (Bc-B3-p317964), MSL (Bn-A03-p28109103), SPP (Bj-B5-p19074770), BYPP (Bn-A10-p15838932), OC (Bn-A02-p636428), Glu (Bn-A08-p8658074), and PA (Bj-B3-p9090824) (Table 3).

Gene annotations of nearby regions of stable SNPs and *in silico* expression analysis of candidate genes

Among 49 associated SNPs, a total of 12 SNPs were consistent in both seasons for studied traits. Ten SNPs have shown association with 7 agro-morphological traits *viz.*, DF, DFT, DM, PH, MSL, SL, and SPS, and two stable MTAs were recorded

for 2 quality attributes, i.e., OC and Glu. These SNPs were considered stable MTAs and used for further analysis. In the present study, gene annotations of stable SNPs considering the LD identified several candidate genes for studied traits (Table 4). For instance, in DF, SNP Bn-A03-p19798208 was associated with three significant candidate genes, including *BTB/POZ domain-containing protein*, *WRKY transcription factor 1*, and *Heading date 3b*, which exhibited notably higher *in silico* expression in flowering tissues, indicating their crucial role in flowering regulation. (Table 4). Similarly, DFT revealed two stable SNPs, Bn-A05-p21930978 and Bc-B6-p26812452, closely located with 5 (*Photoperiod-independent early flowering 1*, *agamous-like MADS-box protein AP1*, *FBD-associated F-box protein*, *REVEILLE 8* and *FAR1-related sequence 5*) and 4 (*Ethylene-responsive transcription factor 9*, *auxin-responsive protein IAA10*, *pectate lyase 1* and *ethylene response sensor 2*) candidate genes, respectively, showing maximum expression either in flowering or mature tissues as identified through *in silico* analysis. Moreover, stable SNP Bn-A09-p36455112 of DM was closely linked with 5 candidate genes (*Sialyltransferase-like protein 1*, *early nodulin-like protein 1*, *axial regulator YABBY 2*, *ethylene-responsive transcription factor ERF023*, and *ethylene-responsive transcription factor 1*). The *in-silico* analysis revealed that these were highly expressive in flowering and later stages of the plant, suggesting the putative role in plant maturity. PH-associated SNP Bc-B3-p317964 had two linked genes, *IAA17* and *B-box zinc finger protein 24*, specifically showing higher expression at seedling hypocotyl and senescent internode, respectively. In the case of MSL, SNP Bn-A03-p28109103 harbored 3 candidate genes viz., *cup-shaped cotyledon 3*, *SMAX1*, and *short hypocotyl in white light 1*, predominantly expressing in terminal and mature tissues of the plant identified through *in silico* expression analysis. Additionally, SL exhibited two stable associations, Bn-A03-p21210189 and Bn-A08-p19365594, linked to *GATA transcription factor 25* and *E3 ubiquitin-protein ligase BOI*, respectively, with significant *in silico* gene expression at pod of the senescent silique and carpels of the young flower, respectively. Conversely, SNPs Bn-A03-p21615990 and Bn-A06-p2234696 were consistently associated with SPS, closely linked with three candidate genes (*Nuclear fusion defective 4*, *GATA transcription factor 25*, *APRR1*) and one candidate gene (*GATA transcription factor 11*), respectively, within the LD block, showing significant *in silico* expression in seed-bearing tissues such as ovules and silique. Furthermore, two stable MTAs were observed for quality attributes, one each for OC and Glu. OC was associated with SNP Bn-A02-p636428 carrying candidate genes *WR11* and *Oleosin*, exhibiting higher *in silico*

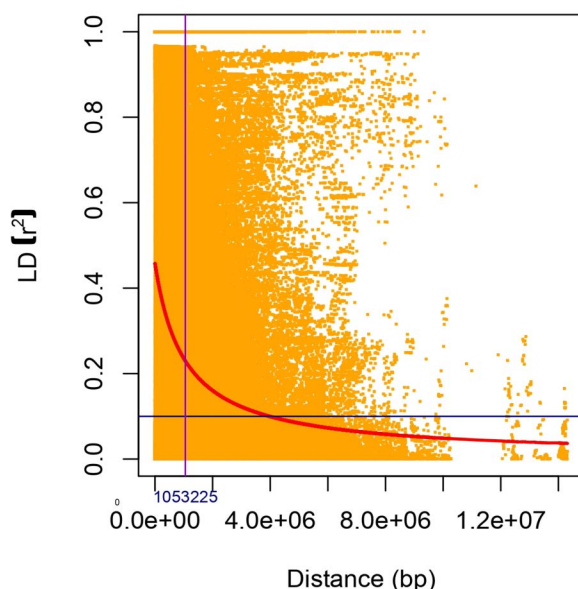


Fig 5. LD decay plot of GWAS panel using SNP markers.

<https://doi.org/10.1371/journal.pone.0322120.g005>

expression in seeds. Similarly, the candidate gene, *Transcription factor MYB123-1.1* linked to Glu MTA Bn-A08-p8658074 has also shown maximum expression in seeds (Table 4).

Discussion

GWAS panel in the present investigation comprises of advanced breeding lines, introgression lines (ILs), released varieties, and indigenous as well as exotic collections. Analysis of variance (ANOVA) is the initial step of any experimental design to determine the significance of treatments under study. In the present study, the significant genotypic variation for all the studied traits suggested that the genotype collection carries vast genetic variability, which could be efficiently utilized in the GWAS. Germplasm accessions having wide variability are considered as a potential source for genetic improvement. Hence, enhancing quantitative traits like agro-morphological and quality attributes require the identification of genetic variations present in the germplasm, followed by their utilization in breeding programs [39]. In the present study, the *rabi* 2020–21 demonstrated superior mean performance and a wide range for most of the studied traits as compared to *rabi* 2021–22, indicating the positive influence of its meteorological parameters on the trait performance. Conversely, the *rabi* 2021–22 experienced a higher temperature range, irregular and frequent rainfall during later growth stages, and reduced sunshine duration, rendering it unfavourable for most of the studied traits. PCA is regarded as an effective method for extracting key information from phenotypically complex traits with high correlations, while preserving the original data integrity [64]. In the present investigation, major variation was explained by the first two PCs (PC1 and PC2) for agro-morphological and quality parameters, suggesting the efficiency of PCA in data visualization by dimension reduction.

The advancement of molecular markers has facilitated the application of various mapping approaches for diverse traits across plant species. For instance, these approaches have been used for mapping micronutrients, grain quality, and agronomic traits [65], abiotic stress tolerance [66,67], and grain yield components [68] in bread wheat. Similarly, ionome-related QTLs have been mapped in *A. thaliana* [69], and candidate genes for phenolic acids and flavonoids [70], as well as agronomic and yield-related traits under drought stress [71] have been identified in rapeseed. Moreover, SNP genotyping has revolutionized molecular breeding, as it is widely used for its efficiency and precision [15]. The array-based genotyping preferred over GBS due to their reliability, fixed loci, and simpler analysis [16,17]. Therefore, in the current study, genome-wide association study (GWAS) has been undertaken using 90K *Brassica* array-based genotyping. These SNPs were widely distributed throughout the genome of *B. juncea*. Structure analysis using STRUCTURE software for 142 genotypes of *B. juncea* showed that the GWAS panel consisted of three sub-populations. Diversity analysis using the neighbor-joining method also suggested three major clusters, confirming the three sub-populations in the GWAS panel. It is important to note that genotypes that belong to the same sub-populations were more closely related in diversity analysis as compared to those that occur in different sub-populations. Genotypic diversity and structure analysis have provided the evolutionary significance of genotypes. For instance, genotypes belonging to group-I were mostly of Indian origin and conventional type. Group-II genotypes were intermediate type, mostly quality mustard, that shares genes from both Indian and exotic germplasms (Australian and European). However, the third group contains European germplasms and Indian-quality genotypes derived from European mustard, containing the majority of genes from European parents. Moreover, the allele admixture as observed in the present study suggests the gene flow among the sub-populations. Therefore, this diverse panel can provide more valuable implications as compared to bi-parental populations, by utilizing maximum allelic diversity and historic recombination [72].

The total genotypic variance observed in the GWAS panel was partitioned using AMOVA, which has more within-sub-population variation compared to among sub-populations, which indicated the existence of sufficient diversity within a sub-population. *Fst* values give the idea of the degree of differentiation among populations [73]. The *Fst* values of SP1(0.76) and SP3 (0.65) have shown very strong differentiation, while SP2 (0.15) has shown moderate differentiation from the original population.

The number of markers needed for association mapping depends on the rate of LD decay, which is determined by the genetic distance between markers [74]. It has been observed that the mode of pollination affects the LD decay rate, i.e.,

Table 3. List of marker-trait associations identified during *Rabi*, 2020–21 and *Rabi*, 2021–22.

Season	Trait	SNP	Allele	Chrom	Position	p-value	MAF	Effect	PVE%
<i>Rabi</i> , 2020–21	DF	Bn-A03-p19798208	A/C	A03	19798208	6.08×10^{-15}	0.15	7.90	21.24
		Bn-scaff_18360_1	A/C	NA	1296271	1.42×10^{-11}	0.09	7.46	27
	DFT	Bc-B6-p26812452	T/C	B06	26812452	7.12×10^{-10}	0.12	-7.83	20.27
		Bn-A05-p21930978	A/G	A05	21930978	6.43×10^{-08}	0.23	-4.89	15.76
	DM	Bn-A09-p36455112	T/C	A09	36455112	4.31×10^{-08}	0.10	3.10	48.89
	PH	Bc-B3-p317964	A/C	B03	317964	2.40×10^{-09}	0.23	11.89	29.02
	MSL	Bn-A03-p28109103	A/G	A03	28109103	0.00004	0.10	10.17	NA
	SL	Bn-A03-p21210189	A/C	A03	21210189	0.0002	0.11	-0.35	NA
		Bn-A08-p19365594	A/G	A08	19365594	0.0005	0.15	-0.29	NA
	SPS	Bn-A03-p21615990	A/C	A03	21615990	0.0005	0.15	-0.66	NA
		Bn-A06-p2234696	A/G	A06	2234696	0.0005	0.12	0.70	NA
	OC	Bn-A02-p636428	T/G	A02	636428	0.0006	0.50	-0.43	NA
	Glu	Bn-A08-p8658074	T/C	A08	8658074	0.0004	0.10	-12.22	NA
	OA	Bn-A03-p15957657	T/G	A03	15957657	1.57×10^{-11}	0.10	6.85	15.14
		Bj-B4-p25774828	A/G	B04	25774828	8.57×10^{-10}	0.24	3.55	4.85
		Bj-B4-p9000035	T/G	B04	9000035	3.49×10^{-09}	0.21	3.29	4.39
		Bj-B4-p17202410	A/G	B04	17202410	1.70×10^{-08}	0.12	4.44	7.7
		Bj-B3-p25944867	A/G	B03	25944867	2.85×10^{-08}	0.18	4.37	6.04
		Bj-B1-p2664020	T/C	B01	2664020	1.17×10^{-06}	0.42	-2.10	1.53
		Bj-B8-p41554126	A/G	B08	41554126	2.63×10^{-06}	0.14	-2.86	3.33
		Bj-B7-p38672208	A/G	B07	38672208	2.22×10^{-08}	0.27	-1.05	11.6
	ErA	Bn-A02-p3300731	A/C	A02	3300731	1.25×10^{-06}	0.08	1.36	27.23
		Bj-B7-p19658501	A/G	B07	19658501	1.06×10^{-08}	0.23	5.16	7.98
		Bn-A02-p25655528	A/C	A02	25655528	1.21×10^{-08}	0.14	-4.39	9
		Bn-A08-p10822662	A/C	A08	10822662	1.51×10^{-08}	0.26	-3.86	5.87
		Bj-B4-p17190386	T/C	B04	17190386	8.60×10^{-08}	0.12	-4.22	8.27
		Bj-B6-p37724780	T/G	B06	37724780	1.35×10^{-07}	0.09	4.06	6.67
		Bn-A02-p5612147	T/C	A02	5612147	2.68×10^{-06}	0.11	-3.75	5.4
		Bn-A03-p19798208	A/C	A03	19798208	0.001	0.15	5.19	NA
<i>Rabi</i> , 2021–22	DFT	Bc-B6-p26812452	T/C	B06	26812452	8.18×10^{-10}	0.12	-7.86	20.15
		Bn-A05-p21930978	A/G	A05	21930978	7.88×10^{-08}	0.23	-4.89	15.51
	DM	Bn-A09-p36455112	T/C	A09	36455112	0.0002	0.10	2.83	NA
	PH	Bc-B3-p317964	A/C	B03	317964	0.0001	0.23	10.27	NA
	MSL	Bn-A03-p28109103	A/G	A03	28109103	9.68×10^{-10}	0.10	11.97	50.22
	SPP	Bj-B5-p19074770	T/C	B05	19074770	1.21×10^{-08}	0.17	-39.12	35.37
	SL	Bn-A03-p21210189	A/C	A03	21210189	0.0005	0.11	-0.34	NA
		Bn-A08-p19365594	A/G	A08	19365594	0.0005	0.15	-0.30	NA
	SPS	Bn-A03-p21615990	A/C	A03	21615990	0.0008	0.15	-0.64	NA
		Bn-A06-p2234696	A/G	A06	2234696	0.0001	0.12	0.79	NA
	BYPP	Bn-A10-p15838932	T/G	A10	15838932	8.55×10^{-07}	0.44	-6.67	12.96
	OC	Bn-A02-p636428	T/G	A02	636428	0.0003	0.50	-0.44	NA
	Glu	Bn-A08-p8658074	T/C	A08	8658074	0.0009	0.10	-11.45	NA
	PA	Bj-B3-p9090824	A/G	B03	9090824	1.50×10^{-08}	0.24	-0.89	31.61
	OA	Bn-A08-p4254041	T/G	A08	4254041	6.78×10^{-08}	0.12	-2.83	36.83
		Bj-B8-p3030447	T/C	B08	3030447	8.76×10^{-07}	0.15	-3.72	19.95
	EcA	Bn-A08-p7563202	T/C	A08	7563202	3.69×10^{-07}	0.11	1.37	17.46
		Bj-B3-p25039645	A/C	B03	25039645	7.00×10^{-07}	0.15	0.98	14.08
	ErA	Bj-B8-p41554126	A/G	B08	41554126	3.34×10^{-07}	0.14	3.62	29.99
		Bn-A08-p4001683	T/G	A08	4001683	1.16×10^{-06}	0.16	-5.30	1.3

Chrom = Chromosome, MAF = Minor allele frequency, PVE% = Percent phenotypic variation explained

<https://doi.org/10.1371/journal.pone.0322120.t003>

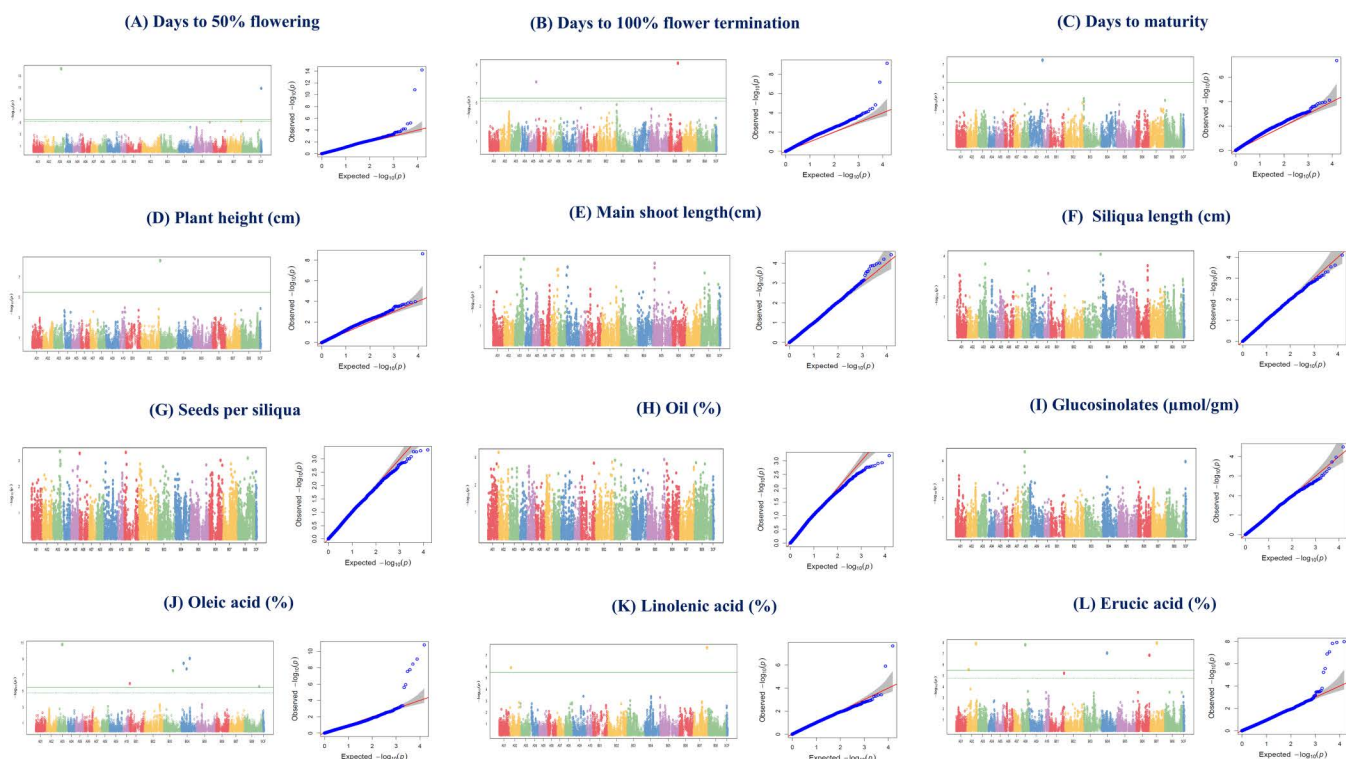


Fig 6. Manhattan and QQ plots showing MTAs for different agro-morphological traits during *rabi* 2020-21.

<https://doi.org/10.1371/journal.pone.0322120.g006>

self-pollinating crops have longer LD blocks and decay slower compared to cross-pollinating crops. For instance, wheat exhibited LD decay at a distance of 7.15 Mb [75], whereas maize decayed at a distance of about 2 kb [76]. Similarly, in this study, moderate LD decay (1.05 Mb) was observed in *B. juncea*, as the crop has shown considerable cross-pollination based on bee activities. This region was subsequently used for gene annotations. Similarly, Vos et al. [77] suggested that LD decay can be used to determine the size of a candidate gene region.

To identify genomic regions for quantitative traits, genome-wide association studies (GWAS) is considered a powerful approach [21]. The season-wise GWAS detected 28 MTAs detected during *rabi*, 2020–21 and 21 MTAs during *rabi*, 2021–22 for various agro-morphological and quality attributes. Some MTAs were specific to only one season, indicating highly influenced by the environment. Such SNPs would not be reproducible; thus, cannot be used in the MAB. It has been observed that the threshold for GWAS using Bonferroni-correction is too stringent which leads to false negatives [78]. Therefore, we adapted less stringent criteria ($p < 0.001$), similar to those utilized by Devate et al. [75] and Mroz et al. [60], for the identification of stable MTAs. Interestingly, as per our assumptions, when consistently occurring SNPs with $p < 0.001$ were subjected to gene annotations, some remarkable genes such as *WRI1* and *Oleosin* in case of OC MTA Bn-A02-p636428, *TF MYB123-1.1* in Glu MTA Bn-A08-p8658074 etc. were identified. Therefore, if any SNP occurs consistently across the environments, but falls below the Bonferroni threshold; then, using a less stringent approach will be rewarding.

The stable SNPs when subjected to *in silico* analysis, various candidate genes were predicted. In the current study, SNP Bn-A03-p19798208 identified candidate genes for DF such as *BTB/POZ domain* family genes (*BOP* genes) delay the flowering under short-day conditions in *A. thaliana*, if present in mutant form [24]. Likewise, *WRKY transcription factors* regulates the flowering [25]. However, the third candidate gene of DF, i.e., *Hd3b*, a rice ortholog of the *EFL3* gene of

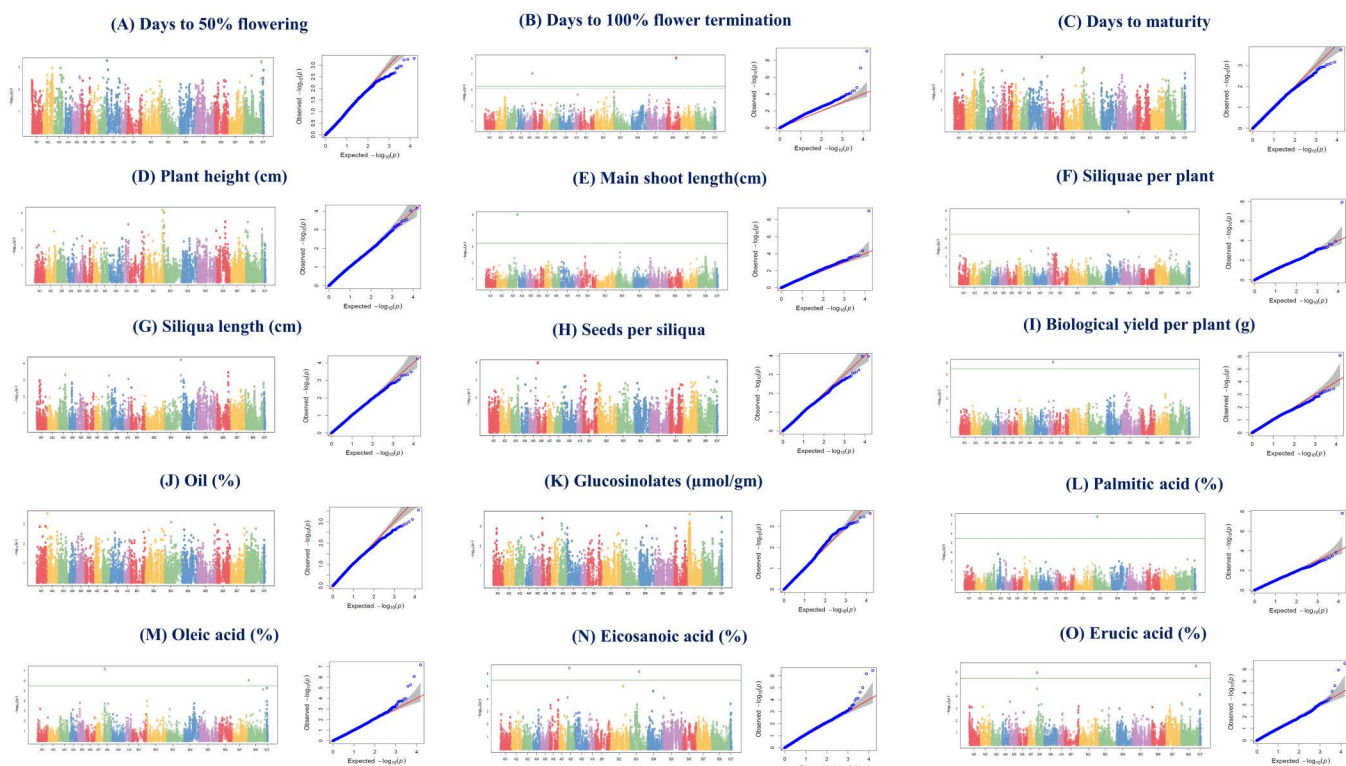


Fig 7. Manhattan and QQ plots showing MTAs for different agro-morphological traits during rabi 2021–22.

<https://doi.org/10.1371/journal.pone.0322120.g007>

Arabidopsis, causes delayed flowering under long-day field conditions in rice [26]. Similar to present findings, SNP Bn-A03-p1050893, was also located on A03 chromosome and associated with flowering time in *B. napus* [79].

Two stable MTAs (Bn-A05-p21930978 and Bc-B6-p26812452) were reported for DFT. Furthermore, Shah et al. [79] reported an SNP Bn-A05-p2497466, located on A05 chromosome and associated with flowering time in *B. napus*, similar to the current association (Bn-A05-p21930978) on the same chromosome. Candidate genes identified for DFT have either played an important role in the regulation of flowering time [80–82], cell senescence/abscission [83] or involved in floral organ formation [84,85]. In addition to these reports, *in silico* tissue expression analysis of these genes, provided supportive evidence that these genes have significant role in the regulation of flowering in Crucifers. Therefore, these genes might be involved in various metabolic pathways and ultimately determining the flowering duration. Similar to DF and DFT, DM MTA (Bn-A09-p36455112) identified candidate genes which involved either in flowering control [86–88] or in cell senescence or programmed cell death [89]. Moreover, *in silico* analysis revealed that these genes were highly expressive in later stages of plant development (in mature flowers and leaves), suggesting their active involvement in plant maturity. Interestingly, the involvement of some important flowering genes in DM suggests the genetic basis of positive correlation of flowering time (DF, DFT) with maturity as observed during the phenotypic screening of the germplasm panel. Candidate genes identified for PH (*IAA17* and *B-box zinc finger protein 24*) involved in the growth and development of plants [27,28]. Conversely, that of main shoot length viz., *cup-shaped cotyledon 3*, *SMA1*, and *short hypocotyl in white light 1* were reported to affect shoot apical meristem, photomorphogenesis and hypocotyl elongation in *Arabidopsis*, respectively [90–92]. Similar to these reports, the maximum *in silico* expression of these genes was observed in seedling hypocotyl and internodal regions, suggesting their crucial role in shoot elongation. Two stable MTAs, Bn-A03-p21210189 and Bn-A08-p19365594, were identified for SLs, associated with *GATA transcription factor 25* and *E3 ubiquitin-protein ligase*

Table 4. . Gene annotations of stable SNPs and *in silico* tissue expression analysis of identified candidate genes.

Traits	Associated SNPs	FA	AA	Candidate genes	Gene ID	<i>A. thaliana</i> Orthologs	Expressing Tissues	Expres- sion level	Con- trol	Fold- Change
DF	Bn-A03-p19798208	A	C	<i>BTB/POZ domain-containing protein</i>	111214206	AT2G04740	Flowers	7.28	5.23	1.39
				<i>WRKY transcription factor 1</i>	106426741	AT5G35750	Axis of inflorescence	8.79	5.17	1.7
				<i>Heading date 3b</i>	106439656	AT3G21320	Anthers of the mature flower	21.76	0.02	1279.91
DFT	Bn-A05-p21930978	G	A	<i>Photoperiod-independent early flowering 1</i>	106452362	AT3G12810	Carpels of the young flowers	10.24	4.39	2.33
				<i>Agamous-like MADS-box protein AP1</i>	106412766	AT1G69120	Sepals of the young flower	99.08	0.03	3416.62
				<i>FBD-associated F-box protein</i>	125608703	AT1G05080	Anthers of the young flower	5.9	0.02	245.82
				<i>REVEILLE-8</i>	111215777	AT3G09600	Opened anthers	16.43	0.45	36.19
				<i>FAR1-related sequence 5</i>	106452929	AT3G07500	Senescent silique	12.98	1.75	7.42
	Bc-B6-p26812452	T	C	<i>Ethylene-responsive transcription factor 9</i>	106398162	AT1G03800	Young seeds	1.14	0.2	5.63
				<i>Auxin-responsive protein IAA10</i>	106402229	AT1G04100	Pod of the senescent silique	5.99	1.49	4.01
				<i>Pectate lyase 1</i>	106370994	AT1G04680	Petals of the mature flower	107.66	16.98	6.34
				<i>Ethylene response sensor 2</i>	106371023	AT1G04310	Petals of the mature flower	18.78	4.96	3.79
DM	Bn-A09-p36455112	T	C	<i>Sialyltransferase-like protein 1</i>	106420237	AT1G08660	Flowers	20.98	11.05	1.9
				<i>Early nodulin-like protein 1</i>	106419837	AT1G08500	Petals of the mature flower	43.06	1.4	30.69
				<i>Axial regulator YABBY 2</i>	106419835	AT1G08465	Sepals of the mature flower	21.91	4.05	5.41
				<i>Ethylene-responsive transcription factor ERF023</i>	106381811	AT1G01250	Petals of the mature flower	40.05	2.5	16.04
				<i>Ethylene-responsive transcription factor 1</i>	106434150	AT4G17500	Leaf Petiole of the senescent leaf	273.82	17.04	16.07
PH	Bc-B3-p317964	A	C	<i>Auxin-responsive protein IAA17</i>	125591576	AT1G04250	Seedling Hypocotyl	31.58	3.61	8.74
				<i>B-box zinc finger protein 24</i>	106365696	AT1G06040	Senescent internode	67.59	28.07	2.41
MSL	Bn-A03-p28109103	G	A	<i>Cup-shaped cotyledon 3</i>	106386092	AT4G28530	Senescent internode	8.92	0.23	39.65
				<i>SMAX1</i>	106441046	AT4G30350	Sepals of the young flower	62.25	16.04	3.88
				<i>Short hypocotyl in white light 1</i>	106441155	AT4G33780	Stigmatic tissue	43.57	10.24	4.25
SL	Bn-A03-p21210189	A	C	<i>GATA transcription factor 25</i>	106439837	AT4G24470	Pod of the senescent silique	10.98	9.15	1.2
	Bn-A08-p19365594	A	G	<i>E3 ubiquitin-protein ligase BOI</i>	106362423	AT1G10650	Carpels of the young flower	12.01	7.11	1.69

(Continued)

Table 4. (Continued)

Traits	Associated SNPs	FA	AA	Candidate genes	Gene ID	<i>A. thaliana</i> Orthologs	Expressing Tissues	Expres-sion level	Con-trol	Fold-Change
SPS	Bn-A03-p21615990	A	C	<i>Nuclear fusion defective 4</i>	106439838	AT4G34950	Silique	30.45	6.88	4.43
				<i>GATA transcription factor 25</i>	106439837	AT4G24470	Ovules	13.0	9.15	1.42
				<i>Two-component response regulator APRR1</i>	125607080	AT5G61380	Pod of the senes-cent silique	63.88	36.04	1.77
	Bn-A06-p2234696	G	A	<i>GATA transcription factor 11</i>	106346369	AT1G08010	Ovules	20.5	9.49	2.16
OC	Bn-A02-p636428	T	G	<i>Ethylene-responsive tran-scription factor WRI1</i>	106352859	AT3G54320	Seeds	32.34	2.45	13.19
				<i>Oleosin</i>	106352688	At5g07260	Seeds	32.29	0.01	2152.93
Glu	Bn-A08-p8658074	C	T	<i>Transcription factor MYB123-1.1</i>	106418890	AT5G35550	Seeds	60.06	0.14	444.88

FA = Favourable allele, AA = Alternate allele

<https://doi.org/10.1371/journal.pone.0322120.t004>

BOI genes, respectively, indicating higher *in silico* expression in siliqua developing tissues. Conversely, SPS had two stable MTAs; Bn-A03-p21615990 associated with *nuclear fusion defective 4*, *GATA transcription factor 25* and *APRR1* genes, and Bn-A06-p2234696 linked with *GATA transcription factor 11*, which have shown maximum *in silico* expression in seed-bearing tissues (ovules and siliqua). In support of the present investigation, Pal et al. [29] reported the candidate gene GATA15, associated with SPS, located on chromosome A06 in *B. napus*. Similarly, Khan et al. [38] identified two MTAs (Bn-A03-p12353370 and Bn-A03-p13109267) on A03 and one MTA (Bn-A06-p24204030) on A06, associated with SPS.

The close proximity of SL SNP Bn-A03-p21210189 and SPS SNP Bn-A03-p21615990 share a common candidate gene (*GATA transcription factor 25*) suggesting pleiotropy as the genetic basis of positive trait correlation. Additionally, stable MTAs were found for OC and Glu, with SNP Bn-A02-p636428 linked to *WRI1* and *oleosin* genes regulating oil accumulation [30,31]. Conversely, the SNP associated with Glu (Bn-A08-p8658074) had the candidate gene *Transcription factor MYB123-1.1*, known for regulating glucosinolates biosynthesis [93]. These candidate genes exhibit maximum *in silico* expression in seeds, indicating their pivotal role in determining both seed oil quantity and quality.

Conclusion

The present study reveals substantial variability within the GWAS panel, highlighting extensive diversity across the evaluated traits. Notably, this investigation represents the first application of array-based genotyping in *Brassica juncea*, signifying a significant milestone in genetic research. The germplasm panel, comprising three distinct sub-populations as determined through structure and neighbor-joining diversity analyses, provides a valuable resource for future studies. Furthermore, this research identifies critical genomic regions influencing the inheritance of the analyzed traits, offering insights for advancing molecular breeding and genetic improvement efforts. In this study, twelve stable marker-trait associations (MTAs) were identified for key traits, including DF (Bn-A03-p19798208), DFT (Bn-A05-p21930978 and Bc-B6-p26812452), DM (Bn-A09-p36455112), PH (Bc-B3-p317964), MSL (Bn-A03-p28109103), SL (Bn-A03-p21210189 and Bn-A08-p19365594), SPS (Bn-A03-p21615990 and Bn-A06-p2234696), OIL (Bn-A02-p636428), and Glu (Bn-A08-p8658074). These MTAs harbor important candidate genes that exhibited significant expression in relevant tissues based on *in silico* analysis, suggesting their potential roles in the respective traits. This study provides a foundation for further research, with the identified candidate genes offering promising targets for molecular breeding programs

after validation through marker analysis. Additionally, these genes can be explored through gene cloning to gain deeper insights into the underlying genetic variation and trait regulation.

Supporting information

S1 Table. List of 142 *B. juncea* genotypes used for Genome Wide Association Studies.

(DOCX)

S2 Table. List of 15219 SNP markers used for Genome Wide Association Studies.

(XLSX)

S3 Table. Meteorological data recorded during growing period in *rabi* 2020–21 and 2021–22.

(XLSX)

Acknowledgements

Authors are grateful to ICAR - National Agricultural Science Fund (NASF) project for providing plant materials used in the present investigation.

Author contributions

Conceptualization: Navinder Saini, Sujata Vasudev, Devendra Kumar Yadava.

Data curation: Manoj Kumar Patel, Rajat Chaudhary, Priya Pardeshi.

Formal analysis: Manoj Kumar Patel, K. K. Vinod.

Investigation: Navinder Saini, Devendra Kumar Yadava.

Methodology: Navinder Saini, Yashpal Taak.

Resources: Devendra Kumar Yadava.

Software: Manoj Kumar Patel, Sudhakar Reddy Basu, K. K. Vinod.

Supervision: Sujata Vasudev, Devendra Kumar Yadava.

Visualization: Navinder Saini, Devendra Kumar Yadava.

Writing – original draft: Manoj Kumar Patel, Masochon Zimik.

Writing – review & editing: Navinder Saini, Sneha Adhikari, Sangita Yadav, Sujata Vasudev.

References

1. Rakow G, Woods DL. Outcrossing in rape and mustard under saskatchewan prairie conditions. *Can J Plant Sci.* 1987;67(1):147–51. <https://doi.org/10.4141/cjps87-017>
2. Chand S, Prakash Patidar O, Chaudhary R, Saroj R, Chandra K, Kamal Meena V, et al. Rapeseed-mustard breeding in india: scenario, achievements and research needs. *Brassica Breed Biotechnol.* 2021. <https://doi.org/10.5772/intechopen.96319>
3. Akhatar J, Banga SS. Genome-wide association mapping for grain yield components and root traits in brassica juncea (l.) czern & coss. *Mol Breed.* 2015;35(1). <https://doi.org/10.1007/s11032-015-0230-8>
4. Akhatar J, Singh MP, Sharma A, Kaur H, Kaur N, Sharma S, et al. Association mapping of seed quality traits under varying conditions of nitrogen application in brassica juncea l. czern & coss. *Front Genet.* 2020;11:744. <https://doi.org/10.3389/fgene.2020.00744> PMID: [33088279](https://pubmed.ncbi.nlm.nih.gov/33088279/)
5. Anonymous. United States Department of Agriculture (USDA) foreign agricultural service. <https://apps.fas.usda.gov/psdonline/app/index.html#/app/advQuery>. 2023.
6. Chen W, Zhang Y, Liu X, Chen B, Tu J, Tingdong F. Detection of QTL for six yield-related traits in oilseed rape (*Brassica napus*) using DH and immortalized F(2) populations. *Theor Appl Genet.* 2007;115(6):849–58. <https://doi.org/10.1007/s00122-007-0613-2> PMID: [17665168](https://pubmed.ncbi.nlm.nih.gov/17665168/)

7. Saroj R, Soumya SL, Singh S, Sankar SM, Chaudhary R, Yashpal, et al. Unraveling the relationship between seed yield and yield-related traits in a diversity panel of brassica juncea using multi-traits mixed model. *Front Plant Sci.* 2021;12:651936. <https://doi.org/10.3389/fpls.2021.651936> PMID: [34017349](#)
8. Vasudev S, Yadava D, Malik D, Tanwar R, Prabhu KV. A simplified method for preparation of fatty acid methyl esters of Brassica oil. *Indian J Genet Plant Breed.* 2008;68:456–8.
9. McVetty PBE, Scarth R. Breeding for improved oil quality in brassica oilseed species. *J Crop Prod.* 2002;5(1–2):345–69. https://doi.org/10.1300/j144v05n01_14
10. Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Rosner BA, et al. Dietary fat intake and the risk of coronary heart disease in women. *N Engl J Med.* 1997;337(21):1491–9. <https://doi.org/10.1056/NEJM199711203372102> PMID: [9366580](#)
11. Mithen RF, Dekker M, Verkerk R, Rabot S, Johnson IT. The nutritional significance, biosynthesis and bioavailability of glucosinolates in human foods. *J Sci Food Agric.* 2000;80(7):967–84. [https://doi.org/10.1002/\(sici\)1097-0010\(20000515\)80:7<967::aid-jsfa597>3.0.co;2-v](https://doi.org/10.1002/(sici)1097-0010(20000515)80:7<967::aid-jsfa597>3.0.co;2-v)
12. Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S. SNP markers and their impact on plant breeding. *Int J Plant Genomics.* 2012;2012:728398. <https://doi.org/10.1155/2012/728398> PMID: [23316221](#)
13. Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, et al. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genomics.* 2006;7:206. <https://doi.org/10.1186/1471-2164-7-206> PMID: [16904008](#)
14. Elbasyoni IS, Lorenz AJ, Guttieri M, Frels K, Baenziger PS, Poland J, et al. A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant Sci.* 2018;270:123–30. <https://doi.org/10.1016/j.plantsci.2018.02.019> PMID: [29576064](#)
15. Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, et al. Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front Genet.* 2016;7:221. <https://doi.org/10.3389/fgene.2016.00221> PMID: [28083016](#)
16. Yu G, Cui Y, Jiao Y, Zhou K, Wang X, Yang W, et al. Comparison of sequencing-based and array-based genotyping platforms for genomic prediction of maize hybrid performance. *The Crop Journal.* 2023;11(2):490–8. <https://doi.org/10.1016/j.cj.2022.09.004>
17. Chen H, Xie W, He H, Yu H, Chen W, Li J, et al. A high-density SNP genotyping array for rice biology and molecular breeding. *Mol Plant.* 2014;7(3):541–53. <https://doi.org/10.1093/mp/sst135> PMID: [24121292](#)
18. Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica.* 2005;142(1–2):169–96. <https://doi.org/10.1007/s10681-005-1681-5>
19. Nordborg M, Tavaré S. Linkage disequilibrium: what history has to tell us. *Trends Genet.* 2002;18(2):83–90. [https://doi.org/10.1016/s0168-9525\(02\)02557-x](https://doi.org/10.1016/s0168-9525(02)02557-x) PMID: [11818140](#)
20. Yu J, Buckler ES. Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol.* 2006;17(2):155–60. <https://doi.org/10.1016/j.copbio.2006.02.003> PMID: [16504497](#)
21. Zhu C, Gore M, Buckler ES, Yu J. Status and prospects of association mapping in plants. *The Plant Genome.* 2008;1(1). <https://doi.org/10.3835/plantgenome2008.02.0089>
22. Oraguzie NC, Wilcox PL. An overview of association mapping. *association mapping in plants.* New York, NsY: Springer New York; 2007;1–9. [doi:10.1007/978-0-387-36011-9_1](https://doi.org/10.1007/978-0-387-36011-9_1)
23. Gupta PK, Kulwal PL, Jaiswal V. Association mapping in plants in the post-GWAS genomics era. *Adv Genet.* 2019;104:75–154. <https://doi.org/10.1016/bs.adgen.2018.12.001> PMID: [31200809](#)
24. Norberg M, Holmlund M, Nilsson O. The blade on petiole genes act redundantly to control the growth and development of lateral organs. *Development.* 2005;132(9):2203–13. <https://doi.org/10.1242/dev.01815> PMID: [15800002](#)
25. Ma Z, Li W, Wang H, Yu D. WRKY transcription factors WRKY12 and WRKY13 interact with SPL10 to modulate age-mediated flowering. *J Integr Plant Biol.* 2020;62(11):1659–73. <https://doi.org/10.1111/jipb.12946> PMID: [32396254](#)
26. Monna L, Lin X, Kojima S, Sasaki T, Yano M. Genetic dissection of a genomic region for a quantitative trait locus, Hd3, into two loci, Hd3a and Hd3b, controlling heading date in rice. *Theor Appl Genet.* 2002;104(5):772–8. <https://doi.org/10.1007/s00122-001-0813-0> PMID: [12582636](#)
27. Huang J, Zhao X, Weng X, Wang L, Xie W. The rice B-box zinc finger gene family: genomic identification, characterization, expression profiling and diurnal analysis. *PLoS One.* 2012;7(10):e48242. <https://doi.org/10.1371/journal.pone.0048242> PMID: [23118960](#)
28. Rinaldi MA, Liu J, Enders TA, Bartel B, Strader LC. A gain-of-function mutation in IAA16 confers reduced responses to auxin and abscisic acid and impedes plant growth and fertility. *Plant Mol Biol.* 2012;79(4–5):359–73. <https://doi.org/10.1007/s11103-012-9917-y> PMID: [22580954](#)
29. Pal L, Sandhu SK, Bhatia D, Sethi S. Genome-wide association study for candidate genes controlling seed yield and its components in rapeseed (*Brassica napus* subsp. *napus*). *Physiol Mol Biol Plants.* 2021;27(9):1933–51. <https://doi.org/10.1007/s12298-021-01060-9> PMID: [34629771](#)
30. Wu X, Liu Z, Hu Z, Huang R. BnWRI1 coordinates fatty acid biosynthesis and photosynthesis pathways during oil accumulation in rapeseed. *J Integr Plant Biol.* 2014;56(6):582–93. <https://doi.org/10.1111/jipb.12158> PMID: [24393360](#)
31. Jia Y, Yao M, He X, Xiong X, Guan M, Liu Z, et al. Transcriptome and regional association analyses reveal the effects of oleosin genes on the accumulation of oil content in brassica napus. *Plants (Basel).* 2022;11(22):3140. <https://doi.org/10.3390/plants11223140> PMID: [36432869](#)
32. Saini N, Singh N, Kumar A, Vihan N, Yadav S, Vasudev S, et al. Development and validation of functional CAPS markers for the FAE genes in Brassica juncea and their use in marker-assisted selection. *Breed Sci.* 2016;66(5):831–7. <https://doi.org/10.1270/jsbbs.16132> PMID: [28163599](#)

33. Saini N, Koramutla MK, Singh N, Singh S, Singh R, et al. Promoter polymorphism in FAE1.1 and FAE1.2 genes associated with erucic acid content in Brassica juncea. *Mol Breeding*. 2019;39(5). <https://doi.org/10.1007/s11032-019-0971-x>
34. Bisht NC, Gupta V, Ramchiary N, Sodhi YS, Mukhopadhyay A, Arumugam N, et al. Fine mapping of loci involved with glucosinolate biosynthesis in oilseed mustard (*Brassica juncea*) using genomic information from allied species. *Theor Appl Genet*. 2009;118(3):413–21. <https://doi.org/10.1007/s00122-008-0907-z> PMID: 18979082
35. Akhatar J, Goyal A, Kaur N, Atri C, Mittal M, Singh MP, et al. Genome wide association analyses to understand genetic basis of flowering and plant height under three levels of nitrogen application in Brassica juncea (L.) Czern & Coss. *Sci Rep*. 2021;11: 4278. <https://doi.org/10.1038/s41598-021-83689-w>
36. Li F, Chen B, Xu K, Gao G, Yan G, Qiao J, et al. A genome-wide association study of plant height and primary branch number in rapeseed (*Brassica napus*). *Plant Sci*. 2016;242:169–77. <https://doi.org/10.1016/j.plantsci.2015.05.012> PMID: 26566834
37. Dong H, Tan C, Li Y, He Y, Wei S, Cui Y, et al. Genome-wide association study reveals both overlapping and independent genetic loci to control seed weight and silique length in brassica napus. *Front Plant Sci*. 2018;9:921. <https://doi.org/10.3389/fpls.2018.00921> PMID: 30073005
38. Khan SU, Yangmiao J, Liu S, Zhang K, Khan MHU, Zhai Y, et al. Genome-wide association studies in the genetic dissection of ovule number, seed number, and seed weight in Brassica napus L. *Ind Crops Prod*. 2019;142:111877. <https://doi.org/10.1016/j.indcrop.2019.111877>
39. Sandhu SK, Pal L, Kaur J, Bhatia D. Genome wide association studies for yield and its component traits under terminal heat stress in Indian mustard (*Brassica juncea* L.). *Euphytica*. 2019;215: 188. <https://doi.org/10.1007/s10681-019-2489-z>
40. Wu Z, Wang B, Chen X, Wu J, King GJ, Xiao Y, et al. Evaluation of linkage disequilibrium pattern and association study on seed oil content in brassica napus using ddrad sequencing. *PLoS One*. 2016;11(1):e0146383. <https://doi.org/10.1371/journal.pone.0146383> PMID: 26730738
41. Tang S, Zhao H, Lu S, Yu L, Zhang G, Zhang Y, et al. Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in Brassica napus. *Mol Plant*. 2021;14(3):470–87. <https://doi.org/10.1016/j.molp.2020.12.003> PMID: 33309900
42. Li F, Chen B, Xu K, Wu J, Song W, Bancroft I, et al. Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). *DNA Res*. 2014;21(4):355–67. <https://doi.org/10.1093/dnares/dsu002> PMID: 24510440
43. Gajardo HA, Wittkop B, Soto-Cerda B, Higgins EE, Parkin IAP, Snowdon RJ, et al. Association mapping of seed quality traits in brassica napus L. using gwas and candidate qtl approaches. *Mol Breeding*. 2015;35(6). <https://doi.org/10.1007/s11032-015-0340-3>
44. Liu S, Huang H, Yi X, Zhang Y, Yang Q, Zhang C, et al. Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of Brassica napus by genome-wide association study. *Plant Biotechnol J*. 2020;18(6):1472–84. <https://doi.org/10.1111/pbi.13314> PMID: 31820843
45. Tandayu E, Borpatragohain P, Mauleon R, Kretschmar T. Genome-wide association reveals trait loci for seed glucosinolate accumulation in indian mustard (*brassica juncea* l.). *Plants (Basel)*. 2022;11(3):364. <https://doi.org/10.3390/plants11030364> PMID: 35161346
46. Mawlong I, Sujith Kumar MS, Gurung B, Singh KH, Singh D. A simple spectrophotometric method for estimating total glucosinolates in mustard de-oiled cake. *International J Food Properties*. 2017;20(12):3274–81. <https://doi.org/10.1080/10942912.2017.1286353>
47. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 1987;19(1):11–5.
48. Aravind J, Mukesh SS, Wankhede DP, Kaur V. Augmentedrcbd: Analysis of augmented randomised complete block designs. R package version 01 59000 <https://aravindj.github.io/augmentedRCBD/><https://cran.rproject.org/-package=augmentedRCBD>. 2021.
49. Husson F, Josse J, Le S, Mazet J, Husson MF. Package ‘FactomineR’. *R Pack*. 2016;96:1–106.
50. Kassambara A, Mundt F. Extract and visualize the results of multivariate data analyses. 2020.
51. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155(2):945–59. <https://doi.org/10.1093/genetics/155.2.945> PMID: 10835412
52. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005;14(8):2611–20. <https://doi.org/10.1111/j.1365-294X.2005.02553.x> PMID: 15969739
53. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*. 2014;2:e281. <https://doi.org/10.7717/peerj.281> PMID: 24688859
54. 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(19):2633–5. <https://doi.org/10.1093/bioinformatics/btm308> PMID: 17586829
55. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci U S A*. 2001;98(20):11479–84. <https://doi.org/10.1073/pnas.201394398> PMID: 11562485
56. Jan HU, Guan M, Yao M, Liu W, Wei D, Abbadi A, et al. Genome-wide haplotype analysis improves trait predictions in Brassica napus hybrids. *Plant Sci*. 2019;283:157–64. <https://doi.org/10.1016/j.plantsci.2019.02.007> PMID: 31128685
57. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: Genome association and prediction integrated tool. *Bioinformatics*. 2012;28(18):2397–9. <https://doi.org/10.1093/bioinformatics/bts444> PMID: 22796960
58. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience*. 2019;8(2):giy154. <https://doi.org/10.1093/gigascience/giy154> PMID: 30535326

59. Haikka H, Manninen O, Hautsalo J, Pietilä L, Jalli M, Veteläinen M. Genome-wide association study and genomic prediction for fusarium graminearum resistance traits in nordic oat (*Avena sativa* L.). *Agronomy*. 2020;10(2):174. <https://doi.org/10.3390/agronomy10020174>
60. Mroz T, Dieseth JA, Lillemo M. Grain yield and adaptation of spring wheat to Norwegian growing conditions is driven by allele frequency changes at key adaptive loci discovered by genome-wide association mapping. *Theor Appl Genet*. 2023;136(9):191. <https://doi.org/10.1007/s00122-023-04424-9> PMID: 37589760
61. Solovyyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol*. 2006;7 Suppl 1(Suppl 1):S10.1-12. <https://doi.org/10.1186/gb-2006-7-s1-s10> PMID: 16925832
62. Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, et al. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res*. 2023;51(D1):D445–51. <https://doi.org/10.1093/nar/gkac998> PMID: 36350662
63. Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J*. 2016;88(6):1058–70. <https://doi.org/10.1111/tpj.13312> PMID: 27549386
64. Ringnér M. What is principal component analysis?. *Nat Biotechnol*. 2008;26(3):303–4. <https://doi.org/10.1038/nbt0308-303> PMID: 18327243
65. Shariatpour N, Heidari B, Tahmasebi A, Richards C. Comparative genomic analysis of quantitative trait loci associated with micronutrient contents, grain quality, and agronomic traits in wheat (*Triticum aestivum* L.). *Front Plant Sci*. 2021;12:709817. <https://doi.org/10.3389/fpls.2021.709817> PMID: 34712248
66. Tahmasebi S, Heidari B, Pakniyat H, McIntyre CL. Mapping QTLs associated with agronomic and physiological traits under terminal drought and heat stress conditions in wheat (*Triticum aestivum* L.). *Genome*. 2017;60(1):26–45. <https://doi.org/10.1139/gen-2016-0017> PMID: 27996306
67. Salarpour M, Abdolshahi R, Pakniyat H, Heidari B, Aminizadeh S. Mapping quantitative trait loci for drought tolerance/susceptibility indices and estimation of breeding values of doubled haploid lines in wheat. *Crop & Pasture Sci*. 2021;72(7):500–13. <https://doi.org/10.1071/cp20526>
68. Heidari B, Sayed-Tabatabaei BE, Saeidi G, Kearsey M, Suenaga K. Mapping QTL for grain yield, yield components, and spike features in a doubled haploid population of bread wheat. *Genome*. 2011;54(6):517–27. <https://doi.org/10.1139/g11-017> PMID: 21635161
69. Shariatpour N, Heidari B, Ravi S, Stevanato P. Genomic analysis of ionome-related QTLs in *Arabidopsis thaliana*. *Sci Rep*. 2021;11(1):19194. <https://doi.org/10.1038/s41598-021-98592-7> PMID: 34584138
70. Salami M, Heidari B, Batley J, Wang J, Tan X-L, Richards C, et al. Integration of genome-wide association studies, metabolomics, and transcriptomics reveals phenolic acid- and flavonoid-associated genes and their regulatory elements under drought stress in rapeseed flowers. *Front Plant Sci*. 2024;14:1249142. <https://doi.org/10.3389/fpls.2023.1249142> PMID: 38273941
71. Salami M, Heidari B, Alizadeh B, Batley J, Wang J, Tan X-L, et al. Dissection of quantitative trait nucleotides and candidate genes associated with agronomic and yield-related traits under drought stress in rapeseed varieties: integration of genome-wide association study and transcriptomic analysis. *Front Plant Sci*. 2024;15:1342359. <https://doi.org/10.3389/fpls.2024.1342359> PMID: 38567131
72. Voss-Fels KP, Qian L, Parra-Londono S, Uptmoor R, Frisch M, Keeble-Gagnère G, et al. Linkage drag constrains the roots of modern wheat. *Plant Cell Environ*. 2017;40(5):717–25. <https://doi.org/10.1111/pce.12888> PMID: 28036107
73. Abdel-Haleem H, Luo Z, Szczepanek A. Genetic diversity and population structure of the USDA collection of *Brassica juncea* L. *Industrial Crops and Products*. 2022;187:115379. <https://doi.org/10.1016/j.indcrop.2022.115379>
74. Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD. The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics*. 2007;177(4):2223–32. <https://doi.org/10.1534/genetics.107.079616> PMID: 17947413
75. Devate NB, Krishna H, Parmeshwarappa SKV, Manjunath KK, Chauhan D, Singh S, et al. Genome-wide association mapping for component traits of drought and heat tolerance in wheat. *Front Plant Sci*. 2022;13:943033. <https://doi.org/10.3389/fpls.2022.943033> PMID: 36061792
76. Huang X, Han B. Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol*. 2014;65:531–51. <https://doi.org/10.1146/annurev-arplant-050213-035715> PMID: 24274033
77. Vos PG, Paulo MJ, Voorrips RE, Visser RGF, van Eck HJ, van Eeuwijk FA. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theor Appl Genet*. 2017;130(1):123–35. <https://doi.org/10.1007/s00122-016-2798-8> PMID: 27699464
78. Rice TK, Schork NJ, Rao DC. Methods for handling multiple testing. *Adv Genet*. 2008;60:293–308. [https://doi.org/10.1016/S0065-2660\(07\)00412-9](https://doi.org/10.1016/S0065-2660(07)00412-9) PMID: 18358325
79. Shah S, Weinholdt C, Jedrusik N, Molina C, Zou J, Große I, et al. Whole-transcriptome analysis reveals genetic factors underlying flowering time regulation in rapeseed (*Brassica napus* L.). *Plant Cell Environ*. 2018;41(8):1935–47. <https://doi.org/10.1111/pce.13353> PMID: 29813173
80. Ritter A, Iñigo S, Fernández-Calvo P, Heyndrickx KS, Dhondt S, Shi H, et al. The transcriptional repressor complex FRS7-FRS12 regulates flowering time and growth in *Arabidopsis*. *Nat Commun*. 2017;8(1). <https://doi.org/10.1038/ncomms15235>
81. Yan J, Li X, Zeng B, Zhong M, Yang J, Yang P, et al. FKF1 F-box protein promotes flowering in part by negatively regulating DELLA protein stability under long-day photoperiod in *Arabidopsis*. *J Integr Plant Biol*. 2020;62(11):1717–40. <https://doi.org/10.1111/jipb.12971> PMID: 32427421
82. Tian Y-Y, Li W, Wang M-J, Li J-Y, Davis SJ, Liu J-X. Reveille 7 inhibits the expression of the circadian clock gene early flowering 4 to fine-tune hypocotyl growth in response to warm temperatures. *J Integr Plant Biol*. 2022;64(7):1310–24. <https://doi.org/10.1111/jipb.13284> PMID: 35603836
83. He D, Liang R, Long T, Yang Y, Wu C. Rice *rbh1* encoding a pectate lyase is critical for apical panicle development. *Plants (Basel)*. 2021;10(2):271. <https://doi.org/10.3390/plants10020271> PMID: 33573206

84. Elliott RC, Betzner AS, Huttner E, Oakes MP, Tucker WQ, Gerentes D, et al. Aintegumenta, an apetala2-like gene of arabidopsis with pleiotropic roles in ovule development and floral organ growth. *Plant Cell*. 1996;8(2):155–68. <https://doi.org/10.1105/tpc.8.2.155> PMID: [8742707](https://pubmed.ncbi.nlm.nih.gov/8742707/)
85. Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, Ditta G, et al. MADS-box protein complexes control carpel and ovule development in arabidopsis. *Plant Cell*. 2003;15(11):2603–11. <https://doi.org/10.1105/tpc.015123> PMID: [14555696](https://pubmed.ncbi.nlm.nih.gov/14555696/)
86. Mashiguchi K, Asami T, Suzuki Y. Genome-wide identification, structure and expression studies, and mutant collection of 22 early nodulin-like protein genes in Arabidopsis. *Biosci Biotechnol Biochem*. 2009;73(11):2452–9. <https://doi.org/10.1271/bbb.90407> PMID: [19897921](https://pubmed.ncbi.nlm.nih.gov/19897921/)
87. Deng Y, Wang W, Li W-Q, Xia C, Liao H-Z, Zhang X-Q, et al. Male gametophyte defective 2, encoding a sialyltransferase-like protein, is required for normal pollen germination and pollen tube growth in Arabidopsis. *J Integr Plant Biol*. 2010;52(9):829–43. <https://doi.org/10.1111/j.1744-7909.2010.00963.x> PMID: [20738727](https://pubmed.ncbi.nlm.nih.gov/20738727/)
88. Ali S, Khan N, Xie L. Molecular and hormonal regulation of leaf morphogenesis in arabidopsis. *Int J Mol Sci*. 2020;21(14):5132. <https://doi.org/10.3390/ijms21145132> PMID: [32698541](https://pubmed.ncbi.nlm.nih.gov/32698541/)
89. Koyama T. The roles of ethylene and transcription factors in the regulation of onset of leaf senescence. *Front Plant Sci*. 2014;5:650. <https://doi.org/10.3389/fpls.2014.00650> PMID: [25505475](https://pubmed.ncbi.nlm.nih.gov/25505475/)
90. Aida M, Ishida T, Tasaka M. Shoot apical meristem and cotyledon formation during Arabidopsis embryogenesis: interaction among the cup-shaped cotyledon and shoot meristemless genes. *Development*. 1999;126(8):1563–70. <https://doi.org/10.1242/dev.126.8.1563> PMID: [10079219](https://pubmed.ncbi.nlm.nih.gov/10079219/)
91. Bhatia S, Gangappa SN, Kushwaha R, Kundu S, Chattopadhyay S. Short hypocotyl in white light1, a serine-arginine-aspartate-rich protein in arabidopsis, acts as a negative regulator of photomorphogenic growth. *Plant Physiol*. 2008;147(1):169–78. <https://doi.org/10.1104/pp.108.118174> PMID: [18375596](https://pubmed.ncbi.nlm.nih.gov/18375596/)
92. Osnato M. Not too short and not too long: Smax1 optimizes hypocotyl length at warmer temperature. *Plant Cell*. 2022;34(7):2580–1. <https://doi.org/10.1093/plcell/koac125> PMID: [35526157](https://pubmed.ncbi.nlm.nih.gov/35526157/)
93. Seo M-S, Kim J. Understanding of myb transcription factors involved in glucosinolate biosynthesis in brassicaceae. *Molecules*. 2017;22(9):1549. <https://doi.org/10.3390/molecules22091549>