Genetic Diversity And Population Structure of Indian Mustard (*Brassica Juncea*) Based On Morphological and Molecular Markers

Prashant Yadav (prashantnduat@gmail.com)
DRMR: Directorate of Rapeseed-Mustard Research https://orcid.org/0000-0002-5277-4572
Sushma Yadav
DRMR: Directorate of Rapeseed-Mustard Research
Anurag Mishra
IARI: Indian Agricultural Research Institute
Rajat Chaudhary
IARI: Indian Agricultural Research Institute
Arun Kumar
DRMR: Directorate of Rapeseed-Mustard Research
Hari Singh Meena
DRMR: Directorate of Rapeseed-Mustard Research
Pramod Kumar Rai
DRMR: Directorate of Rapeseed-Mustard Research

Research Article

Keywords: Indian mustard, Brassica juncea, edible oil, genetic diversity, molecular markers, population structure

Posted Date: November 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-904639/v1

Abstract

Rapeseed-mustard is one of the most important oilseed crops and providing a major source of edible oil in the world besides having other economic importance like leafy vegetables, ornamentals, and hedge crops. However, the genetic diversity present in the Brassica gene pool has not been investigated in detail. To address this problem, a study was conducted on 76 genotypes of B. juncea including cultivars, exotic lines, registered genetic stocks, advanced breeding lines, and germplasm lines. The genetic diversity was analyzed with the help of 50 polymorphic SSR and EST-SSR markers. For these genotype-marker combinations, a total of 126 alleles were amplified. Using molecular and phenotypic data, the dendrogram was constructed based on Jaccard's similarity coefficient and Manhattan dissimilarity coefficient and linkage algorithm UPGMA. All the genotypes were grouped into 5 clusters based on their dissimilarity matrix. Population structure analysis grouped the genotypes in 8 clusters and various degrees of admixture was also observed. The grouping of genotypes appears effective as per their pedigree. The marker data was found more accurate to characterize the diversity and study the population structure than the quantitative trait data. The results of the present investigation will provide useful information for the identification of important alleles for future studies and pave the way to enhance genetic gains in Indian mustard.

1. Introduction

Rapeseed-mustard crops are grown around the world under a spectrum of agro-climatic zones ranging from rainfed to irrigated conditions. In India, these crops are grown on normal to saline soil, in north-eastern to western hills, and in the different cropping systems. Among the various oilseed crops, Indian mustard (*B. juncea* L. Czern & Coss) is one of the most important edible oilseed crops contributing 9.3 MT production with a productivity of 1.5 t/ha during 2018-19 (Bodh et al., 2020). But due to yield ceiling and biotic and abiotic losses, there is a large gap between the supply and demand of edible oil in the country. The mustard germplasm is a rich source of valuable genes that can be used by plant breeders to develop improved varieties having resistance/tolerance to various biotic and abiotic stresses along with higher yield. The success of any breeding program depends upon genetic diversity in the germplasm, its intensive exploitation and utilization in crop improvement (Singh et al. 2016a; Mishra et al. 2019). Parents having genetic differences, mainly in terms of the number of genes and/or different allelic forms, resulted in good heterosis. Therefore, the assessment of genetic diversity among genotypes showing a significant impact on the success of the breeding program is important. The knowledge of population structure, similarity and divergence in *B. juncea* genotypes will be useful for breeders to predict the best parents for the *Brassica* improvement program.

To estimate the available diversity in the germplasm, several methods are available *viz*., evaluation of phenotypic variation, biochemical and DNA-based polymorphic analysis. However, phenotypic characterization is unreliable since it is affected by environment, labor demanding and numerically and phenologically limited (Duminil and Di Michele 2009). On the other hand, DNA based molecular markers are reproducible, ubiquitous, stable, and reliable (Song et al. 1988; Snowdon and Friedt 2004). There are

several classes of molecular markers available, among them, microsatellite markers (simple sequence repeats or SSR) are considered as most suitable due to their codominant and multiallelic nature, higher reproducibility, abundance and genome-wide coverage (Kumar et al. 2015). A number of SSRs and EST-SSR markers have been developed for Indian mustard varying in their degree of polymorphism (Shi et al. 2014; Singh et al. 2016a). Therefore, SSR markers covering the entire genome will help in the unbiased study of the genetic diversity of Indian mustard which in turn gives a molecular description of mustard cultivars.

The assessment of Indian mustard germplasm diversity has been carried out by several researchers earlier (Bansal et al. 2009; Ghosh et al. 2009; Singh et al. 2013b). However, most of these studies are based on a small number of genotypes with less effective RAPD markers and nondescript SSR markers. It has been hypothesized that the use of random markers for diversity analysis may not consider the functional variability present in the coding region of the genome (Zhang et al. 2010). Identification of variation occurring at a heterotic locus is crucial for plant breeding. Therefore, it is pertinent to study the genetic diversity by using random as well as EST-SSR markers which would perform their suitability to access diversity at the genetic level.

In the present study we selected 76 different genotypes including landraces, elite cultivars, mega-varieties, breeding lines and registered genetic stocks having common knowledge about their pedigree and variability with the objective to identify different heterotic groups. The knowledge gained would be useful to exploit heterosis and recombination advantages in the future for developing various kinds of mapping populations *viz*., recombinant inbred lines (RILs), near isogenic lines (NILs) and nested association mapping (NAM) population and other *Brassica* crop improvement programs.

2. Materials And Methods

2.1 Plant material

The experiment was conducted at ICAR-Directorate of Rapeseed Mustard Research, Bharatpur, Rajasthan, India, in randomized block design with three replications and two years of phenotypic data were recorded. A total of 76 mustard accessions taken in this study were obtained from the germplasm section of ICAR-Directorate of Rapeseed-Mustard Research, Bharatpur, Rajasthan, which represents landraces, popular cultivars, breeding lines and registered genetic stocks from different mustard breeding centers across the country (Table 1). The genotypes were sown in two rows of five meters each. The experiment was carried out with standard agronomic practices.

Table 1
The information of 76 Indian mustard genotypes used in this study.

S.N.	Genotype	Pedigree	From
1	PUSA KARIZSHMA (LES-39)	Pusa Basanti x ZEM 1	IARI, New Delhi
2	BPRQ–215	NUDHYJ-3 x Varuna	ICAR-DRMR, Bharatpur
3	HEERA (IC 296501)	ZYR-4 x BJ-1058	Nagpur University, Vadodara, Gujrat
4	PDZM-31 (PDZ–1)	LES-1-27 x NUDHYJ-3	IARI, New Delhi
5	PDZ–3	Pusa Karishma x EC-597325	IARI, New Delhi
6	PDZ–4	Pusa Mustard-21x EC-597325	IARI, New Delhi
7	RLC–2 (ELM-123)	-	PAU, Ludhiana
8	RLC–3	JM-06003 xJM-06020	PAU, Ludhiana
9	RLC–4	-	PAU, Ludhiana
10	PM–21 (LES-1-27)	Pusa Bold x ZEM 2	IARI, New Delhi
11	PM–30 (LES-43)	Bio 902 x ZEM 1	IARI, New Delhi
12	URVASI (RK-9501)	Varuna x Kranti	CSAUA & T, Kanpur, U P
13	RH–406	RH-6908 x RH-8812	CCSHAU, Hisar, Haryana
14	BPR–549–9 (IC 0595525)	Chapka x PCR-9301	ICAR-DRMR, Bharatpur
15	PM–28 (NPJ-124)	Sej-8 x Pusa Jagannathg	IARI, New Delhi
16	RH–819	Prakash x Bulk pollen	CCSHAU, Hisar, Haryana
17	PHR–2	-	Bharatpur, Rajasthan
18	PAB–9511 (IC 546946)	(RC-781 x Krishna)x(PHR-1 x Poorbi raya)	GBPUA&T, Pantnagar
19	RH–1230	RH-270 x B. alba	CCSHAU, Hisar, Haryana
20	RH–1235	RH-406 x B. alba	CCSHAU, Hisar, Haryana
21	Bio–YSR	Somaclone of B juncea	NRCPB, New Delhi
22	DRMRIJ–1206	DRMRIJ-31 x EC-597313	DRMR, Bharatpur, Raj
23	DRMRIJ–1237	Derived from DMH-1	DRMR, Bharatpur, Raj
24	DRMRIJ–1240	ZEM-2 x JGM-1-11	DRMR, Bharatpur, Raj
25	RH–749	RH-781 x RH-9617	CSSRI, Karnal, Haryana

S.N.	Genotype	Pedigree	From
26	RH–1301	RH-0202 x RH-0303	CSSRI, Karnal, Haryana
27	RH–555	RH-9615 x RH-9806	CSSRI, Karnal, Haryana
28	PUSA BOLD (PR–45)	Varuna x BIC-1780	IARI, New Delhi
29	PUSA TARAK (EJ–13)	Agra Local x Poorbi Raya	IARI, New Delhi
30	PUSA VIJAY (NPJ–93)	Synthetic Brassica juncea x VSL 5	IARI, New Delhi
31	PUSA AGARANI (SEJ–2)	Derived from a cross of B. juncea to a amphidilpoid	IARI, New Delhi
32	PUSA MAHAK (JD–6)	Pusa Bold x Glossy	IARI, New Delhi
33	PM–27 (EJ–17)	[(Divya x Pusa Bold) x PR-666 EPS x (PR- 704 EPS-2 x B-85)]	IARI, New Delhi
34	CS–54	B 380 x NDR 8603	CSSRI, Karnal, Haryana
35	PUSA JAIKISAN (BIO–902)	Somaclone of Varuna	IARI, New Delhi
36	PAB–9534 (IC 546947)	(RC-781 x Krishna) x (PHR-1 x Poorbiraya)	GBPUA&T, Pantnagar
37	WR–2019 (IC0598622)		Bharatpur, Rajasthan
38	DRMR–2035 (IC0598623)		Bharatpur, Rajasthan
39	RGN–48	RSM-204 x B-75	ARS, SKRAU, Sriganganagar
40	RRN–727	RW-01-02 x Patan67	ARS, Navgaon
41	DRMR–WFYSM–15	Mutant of Kranti	ICAR-DRMR, Bharatpur
42	PM–25 (NPJ–112)	Sej-8 x Pusa Jagannathg	IARI, New Delhi
43	B–15	-	Bharatpur, Rajasthan
44	B/K/S–72	-	Bharatpur, Rajasthan
45	Bio–Q–108	Bio-902 x Parkland	IARI, New Delhi
46	BT–15	-	-
47	EC–447055	-	Germany

S.N.	Genotype	Pedigree	From
48	EC–482983	-	Canada
49	EC–511589	-	Canada
50	EC 511690	-	Canada
51	EC–511718	-	Canada
52	EC–557025	-	Australia
53	Glossy–CCH	-	Bharatpur, Rajasthan
54	HP–11 (IC347667)	-	Palampur, HP
55	IC–571629	-	Bidar, Karnataka
56	IC–571646	-	Dharwad, Karnataka
57	JAIPUR–12 (IC511592)	-	Jaipur, Rajasthan
58	JCR–914 (IC427136)	-	Bilaspur, HP
59	NDRS–2017-1	(NDR8501 X DIRA313	NDUAT, Faizabad, UP
60	NE–68 (IC522378)	-	Dimapur, Nagaland
61	P–14 (IC346700)	-	Sangroor, Punjab
62	P–58 (IC346744)	-	Faridlot, Punjab
63	PBR–2004–06	(Varuna x MCN 23) x BSIPS 23	GBPUA&T, Pantnagar
64	PHR–2	-	Bharatpur, Rajasthan
65	RGN–157 (IC511523)	PCR 7 x DYS 7-1-2	ARS, SKRAU, Sriganganagar
66	RH–114	(DHR 9401 x Pusa Bold) xPCR 7	CCSHAU, Hisar
67	RH–0704	RH 8812 x (Pusa Basanti x RH 8814)	CCSHAU, Hisar
68	SN–55 (IC 426343)	-	Kaimnagar, Telangana
69	SN–56 (IC 426344)	-	Kaimnagar, Telangana
70	TM–1 (IC 296829)	-	Bose Institute, Kolkata
71	UP I–70 (IC346178)	-	G.B. Nagar, UP
72	UP I–87 (IC346195)	-	Faridabad, Haryana
73	UP II–119 (IC345973)	-	Ghaziabad, UP

S.N.	Genotype	Pedigree	From
74	UP II–123 (IC345977)	-	Meerut, UP
75	UP II–22 (IC345876)	-	Muzzafarnagar, UP
76	UP II–28 (IC345882)	-	Saharanpur, UP

2.2 Evaluation of morphological traits

Phenotypic data were taken from the five selected plants of the middle row from each replication plot such as days to flower initiation (DF), days to maturity (DM), number of primary branches (PB), number of secondary branches (SB), plant height (PH), main shoot length (MSL), number of silique on the main shoot (SMS), silique density (SD), silique length (SL), seeds per silique (SS), oil content (OC),1000-seeds weight (TSW), and yield per plant (YP). Oil content was estimated following the method described by Singh and coworkers (Singh et al. 2013a).

2.3 Selection of SSR markers

A set of 120 SSRs along with 20 EST-SSR markers developed in our laboratory were selected for the analysis (Singh et al. 2016a). These markers could effectively illustrate the diversity among the selected germplasm (Table S1).

2.4 SSR assay

Total genomic DNA was isolated by modified liquid N_2 free method (Siddiqui et al. 2011) and assayed with a total of 140 SSR + EST SSR markers. The PCR mixture contained 25–30 ng of template DNA, 5 pm of each primer, 0.05 mol dNTPs, 10x PCR buffer, and 0.5 U *Taq* polymerase in a reaction volume of 10 µl. the PCR cycle was set up as initial denaturation at 94°Cfor 5 min followed by 40 cycles (1 min denaturation at 94°C, 1 min annealing at 55–60°C and 1 min of extension at 72°C) and final extension of 72°C for 7 min. The PCR products were separated on a 3.5 % agarose gel and run for 3 h in 1x TAE buffer. PCR amplicons were visualized under UV transilluminator using EtBr.

2.5 Data analysis

All the statistical analysis was conducted using R software (Team 2021). The data was subjected to principal component analysis (PCA) based clustering since it had multi-collinear variables. The factors corresponding to significant PCs were used for cluster analysis using pair group distance with Euclidean similarity measures. The dendrogram was constructed using the software DARwin V6.0.021 (Perrier and Jacquemoud-Collet 2006). PCR amplicons were scored for their amplicon size for each primer genotype combination. The structure analysis was performed by using STRUCTURE V2.3.4 software (Pritchard et al. 2000; Hubisz et al. 2009).

3. Results

3.1 Morphological characterization

3.1.1 Phenotypic variance

Evaluation of phenotypic traits of Indian mustard genotypes showed a wide range of variation for traits like IF, DM, PH, PB, SB, MSL, SMS, SL, SS, TSW and YP evident by the coefficient of variation (Table 2). DF ranged from 30–74 days and observed earliest in dwarf genotype (30 days) whereas, latest in IC–571629 genotype (74 days). DM ranged from 103–168 days and the dwarf was the earliest maturing genotype (103 days) and Heera took longest (168 days) to mature. PH ranged from 103 cm (dwarf) – 273 cm (Heera); PB ranged from 3.4 (BioQ–108) – 13.3 (EC–557025) and SB ranged from 7.6 (RH–114) – 32 (RH–1230). MSL ranged from 32.5–91.6 cm and found shortest in PHR–2 and longest in RH–1230. SMS and SD ranged from 22.8–67.3 and 0.42–1.1, respectively. SL ranged from 2.3–7.2 cm and the shortest siliqua was observed in UPII–22 whereas, the longest siliques observed in RH–1230. The lowest number of SS observed in EC–511589 (7.6) whereas, the highest SS found in SN–55 (20.2). TSW ranges from 1.5g (IC–571629) to 7.8g (NE–68); OC ranged from 31.7% (NDRS–2017–1) to 42.2% (UPII–119) and YP ranged from 7.8g (IC–571629) to 23.5g (RH–749).

Table 2
Variability (Mean, range and CV) observed of 76 genotypes of Brassica for phenotypic traits.

Trait	Mean	Range	CV(%)
DF	46.6	30 (Dwarf) — 74 (IC-571629)	17.32
DM	144.8	103 (Dwarf) — 168 (Heera)	9.06
PH	202.8	103 (Dwarf) — 273 (Heera)	16.77
PB	6.9	3.4 (Bio-Q-108) — 13.3 (EC-557025)	25.79
SB	17.4	7.6 (RH-114) — 32 (RH-1230)	33.49
MSL	65.7	32.5 (PHR-2) — 91.6 (RH-1230)	16.76
SMS	47.5	22.8 (BT-15) — 67.3 (RH-1235)	19.40
SD	0.7	0.42 (HP-11) — 1.1 (PHR-2)	19.65
SL	4.4	2.3 (UP-II-22) — 7.2 (RH-1230)	25.66
SS	14.2	7.6 (EC-511589) — 20.2 (SN-55)	17.68
TSW	4	1.5 (IC-571629) — 7.8 (NE-68)	29.46
OC	40	31.7 (NDRS-2017-1) — 42.2 (UP II-119)	3.88
YP	15.8	7.8 (IC-571629) — 23.5 (RH-749)	17.50

3.1.2 Correlation of traits

Correlation between different yield component traits was analyzed by calculating Pearson's correlation coefficient. Some traits viz., DF–DM, PH–SMS, PH–SD, PB–SB, SMS–SD, TSW–YP, and SL–YP were found positively correlated whereas, traits like MSL–SD, SD–TSW, and SD–SL were negatively correlated. The highest positive correlation was observed between TSW and YP (0.68). DF was found highly correlated with DM (0.51) whereas; PH was positively correlated with SMS (0.58), SD (0.56), and PB (0.46). PB was found highly correlated with SB (0.59) and SD (0.45) whereas, SB showed a high correlation with SMS (0.51). MSL showed a positive correlation with SL (0.47) and SMS (0.46) whereas, negatively correlated with SD (−0.42). SMS was highly correlated to PH (0.58) and SB (0.51). TSW was observed negatively correlated with SD (−0.51) whereas positively correlated with YP (0.68). SD was highly correlated with the component traits viz., SMS (0.58), PH (0.56), PB (0.45) and negatively correlated with MSL (−0.42). Yield component traits like OC and SS showed no significant correlation with any other component traits.

3.1.3 Principal component analysis (PCA)

Pearson's correlation coefficients between the phenotypic traits are presented in Fig. 1. It showed some highly correlated characters like PH, PB, SMS, and TSW which may bias diversity analysis as an effect of multiple–collinearity. Therefore, principal component analysis (PCA) was performed to estimate the variability independently. The important traits along with their eigenvalues obtained from PCA of the correlation matrix are given in Table 3 and contributions of different PCs are shown in Figure S1. A significant part of total variation (69.5%) was accounted by only four major PCs. The PC1 and PC2 together explained 44.6% of the total variability observed. The variables contributed to the first two PCs are given in Fig. 2. It showed that traits like SD, PB, and TSW contribute to PC1 significantly, while SL, MSL, SB, and SMS Showed significant contributions to PC2.

Table 3
Eigenvalues of correlation matrix and related statistics for the agronomic traits.

Eigen value	Cumulative Eigen value	Variability (%)	Cumulative %
3.193	3.193	24.565	24.565
2.615	5.808	20.114	44.680
1.997	7.085	15.361	60.041
1.230	9.035	9.465	69.506
0.978	10.013	7.522	77.028
0.764	10.777	5.874	82.902
0.638	11.415	4.906	87.808
0.465	11.880	3.578	91.386
0.381	12.261	2.929	94.316
0.344	12.605	2.648	96.964
0.223	12.820	1.714	98.678
0.156	12.984	1.203	99.881
0.015	13.000	0.119	100.000

3.1.4 Clustering of genotypes based on phenotype

Agglomerative clustering (bottom-up) of 76 genotypes on the basis of phenotypic data were performed by calculating the dissimilarity measure by Euclidean distance matrix using two linkage methods (i.e. complete and Ward). Both, Ward and complete linkage method analysis resulted in 5 groups named A-E and 1–5, respectively (Fig. 3). In the Ward's clustering method, group A is the largest group with 38 genotypes followed by group B (13 genotypes), group C (12), and group E (7), whereas D is the smallest group with 6 genotypes. In complete linkage clustering, group 3 is the largest with 25 genotypes followed by group 1 (20), group 4 (15), group 2 (9), and group 5 (7 genotypes). These major clusters corresponded to the clusters identified by the two-dimensional scaling of the first two PCs.

The first group is the largest group consisted of many significant variants like Pusa Bold, RH–749, UP–170 etc. The second group consisted DRMRIJ–1206, BioQ–108, NE68, SN55, etc while the third group contained B–15, HP11, RH–114, and IC–520747. Further, cluster analysis of all the four PCs based on the unweighted pair group averaging of Euclidean distance of the factor revealed three major clusters. These major clusters corresponded to the clusters identified by the two-dimensional scaling of the first two PCs. The contribution of genotypes to major PCs has shown in Table S2. Genotype Heera (6.7%), EC–557025 (6.7%), RH–114 (5.6%), and BT–15 (5.5%) contributed maximum to the first component whereas RH–

1230 (10%), IC–571629 (7.8%), Pusa Bold (4.4%), and RH–406 (4.4%) contributed maximum to the second component.

Most of the genotypes of group B of Ward linkage were grouped together into cluster 4 of complete linkage. Genotypes of group A (Ward linkage) were divided into group 1 and group 3 of clusters obtained by complete linkage. Group E of Ward cluster was highly correlated with group 5 of the complete cluster with 7 common genotypes viz., IC–520747, BT–15, JCR–914, Dwarf, HP–11, B–15, and RH–114. Group B, C and D correlated to groups 4, 1 and 2, respectively.

3.1.5 K-mean clustering of genotypes

The K-mean clustering method is used to optimally group the individual into K-groups. K-mean clustering of the mustard genotypes based on phenotypic data at k = 2, 3, 4 and 8 was given in Fig. 4. The optimal K value estimated was 4 (Figure S2). At k = 2, the genotypes were grouped into 2 distinct clusters but with higher k values, the clusters started to merge similar to that of population structure analysis.

3.2 Molecular Marker Characterization

The selected panel of SSR and EST–SSR markers showed allelic variation ranging from 1 to 5. Out of total 140 markers selected, 50 showed polymorphism. The PIC values of these markers ranged from 0.1 (cnum587aF) to 0.79 (gi258660353). The average PIC of these polymorphic markers was 0.47 and maximum numbers of alleles were found in cnu_m583aF.

3.2.1 Clustering of genotypes based on molecular markers

Clustering of the genotypes based on the 50 polymorphic markers generated two major groups and nine sub-groups (Fig. 5). Cluster A has a maximum of 23 genotypes followed by clusters I, B and D having 12, 11 and 10 genotypes, respectively. Cluster C and G have 10 and 6 genotypes and clusters E, F and H have two genotypes each. Most of the exotic lines viz., EC–511589, EC–447055, EC–482983, and EC–557025 were classified into cluster A, except EC–511690 and EC–511718 which were classified into clusters C and D, respectively. Cluster B mainly consists of varieties like Pusa Vijay, Pusa Bold, Pusa Agrani, Pusa Tarak, and Pusa Jaikisan and all developed at ICAR-Indian Agricultural Research Institute, New Delhi. Cluster D consists of genotypes having late crop maturity period in common viz., UP II–22 (167 d), UP I–87 (166 d), UP II–28 (163 d), NE–68 (162 d) and UP II–119 (152 d). Two genotypes, PM–28 and RH–704 were grouped into cluster E and having phenotypic similarity in traits like DF (43, 45), PB (6.6, 7.4), SB (25.6, 29.2), MSL (74.4, 70.2), SMS (48, 49.8), SD (0.6, 0.7), SL (6.5, 6.3) and SS (17.8, 16.4), respectively. Cluster H has two genotypes Pusa Mahak and BPRQ–215, both having early flowering in common but BPRQ–215 is a double zero quality line whereas Pusa Mahak has a considerably higher level of erucic acid and glucosinolate contents.

Most of the quality lines viz., PDZ–1, PDZ–2, PDZ–3, RLC–3, RLC–4, and Heera were categorized into cluster G except Pusa Karishma grouped in cluster A, BPRQ–215 grouped into cluster H and PM–21 and PM–30 grouped into cluster I. Custer I consists of genotypes primarily having biotic stress tolerance viz.,

Bio–YSR, DRMRIJ–1237 and DRMRIJ–1206 having tolerance against white rust and PAB–9511, PHR–2, RH–1230, and RH–1235 having various degree of tolerance against Alternaria blight disease. These markers showed a considerably higher level of genetic diversity among the mustard genotypes.

3.2.2 Population structure of genotypes

Population structure analysis of 76 genotypes classified the population into two groups and subdivided into 8 subgroups (Fig. 6). Structure analysis was performed at K2 to K10. Population structure at K8 correlated with the genotypic dendrogram and selected for further analysis. Genotypes having less than 70 individual ancestry proportions (q values) are considered admixture (Table S3). Out of 76, a total of 45 genotypes had admixture and genotypes like EC–511690, WR–2035, PM–27, PM–28, and BPR–5499 had a high level of admixture in it whereas Heera, PDZ–1, DRMRIJ–1237, DRMRIJ–1240, RH–749, WFYSM–15, NE–68, TM–1, and UP I–70 had the least amount of admixture with a Q value ≥ 0.96. Genotype Pusa Mahak had the highest Q value of 0.988 and grouped into K6.

The first group of STRUCTURE analysis (red) consisted mostly of the biotic stress-tolerant genotypes such as Bio–YSR, DRMRIJ–1237, and PAB–9511. The second (green) and seventh (Orange-yellow) group were the largest groups consisted of 16 genotypes each. Most of the exotic genotypes viz., EC–511718, EC–447055, EC–511589, EC–482983, and EC–511690 were clustered in the second (green) group. The third group (blue) had the remaining exotic line EC–557025 from Australia. Fourth group (yellow) had most of the germplasm lines from the core set. Most of the quality lines of Indian mustard viz., PDZ–1, PDZ–3, PDZ–4, RLC–2, RLC–4, BPR–Q215, and Heera were grouped in the fifth (pink) group whereas, other quality lines viz., RLC–2, PM–21, PM–30, and Pusa Karishma was grouped in the first (red) group. The sixth group (turquoise blue) contained two genotypes Pusa Mahak and PM-27 whereas the eighth group (brown) had single genotype RH–0704.

4. Discussion

The phenotypic and genetic diversity of 76 Indian mustard genotypes including germplasm lines, varieties, exotic lines, genetic stocks, and land races were studied by analyzing 13 yield component traits and 50 polymorphic microsatellite markers. The genotypes of Indian mustard exhibited significant variation for 13 yield and yield component traits (Table 2). Addition of exotic lines from Germany, Canada, and Australia along with indigenous genotypes developed at different *Brassica* breeding centers across the country as well as a core set of mustard germplasm further contributed to the diversity of the genotypic pool. Earlier studies have also shown similar variations in component traits on different sets of Indian mustard genotypes (Vaishnava et al. 2006; Yadava et al. 2009; Lodhi et al. 2016; Singh et al. 2016b).

Principal component analysis was used to trim down the variables to the significantly correlated variables. PCA showed that traits like SD, SB, MSL, YP, and TSW had contributed more in the first two PCs which contributed 44.6% variation in data. A study conducted on 167 Indian mustard germplasm lines showed 36.22% contribution of the first two PCs (Saleem et al. 2017). A strong correlation was observed

between TSW and YP. However, Ghosh et al. (Ghosh et al. 2019) found a strong correlation between yield/plant and number of siliqua/plant in a population of 47 Indian mustard having various degree of tolerance against *Alternaria brassicicola*. PCA grouped all the exotic germplasm into a separate cluster. Chen et al. (Chen et al. 2013) also found that the different genotypes of *B. juncea* with similar geographical distribution grouped into same cluster.

Ward and complete linkage based agglomerative cluster analysis on the basis of phenotypic data (DF, DM, SS, MSL, TSW *etc*) grouped the genotypes into 5 clusters. These traits were also used by other researchers for estimation of genetic diversity in Indian mustard (Gupta et al. 1991; Yadava et al. 2011, 2012; Singh et al. 2014). Singh et al. (Singh et al. 2013c) observed four clusters of genotypes by using the Manhattan dissimilarity coefficient in 44 Indian mustard germplasm lines. Saleem et al. (Saleem et al. 2017) did clustering by calculating Euclidean dissimilarity coefficient and observed 5 clusters in 167 Indian mustard accessions. The clustering of the genotypes based on both the methods grouped genotypes in a quite similar fashion except for group A, which mainly divided into group 1 and 3.

K-mean clustering method is generally used to study the population structure by better clustering of the individual into K-groups. The population is divided into 2 large groups at K = 2 and further grouping showed admixture in the population as clusters started to overlap. This pattern of K-mean clustering corresponds to the grouping of the population by STRUCTURE analysis. However, Stift et al. (Stift et al. 2019) found that STRUCTURE analysis is better to study the population structure that K-mean clustering method.

Previously, genetic diversity in Indian mustard at the molecular level has been carried out using isozyme markers (Kumar and Gupta 1985) and molecular markers (Jain et al. 1994; Huangfu et al. 2009; Singh et al. 2013b). PCR-based microsatellite markers (SSR, EST-SSR and ISSR) have high potential and widely used in genetic diversity studies in many crops, but have not been used extensively in *B. juncea* (Singh et al. 2016a). Grouping of genotypes based on genotypic data resulted in two major groups and 9 sub-groups. Yang et al. (Yang et al. 2018) also observed two major groups in 109 mustard (*B. juncea*) genotypes based on 170K SNP analyses. The germplasm lines procured from various international centers showed much diversity whereas genotypes from a single center showed lesser diversity. Likewise, in a study conducted by Chen et al. (Chen et al. 2013) where they found that in *B. juncea*, the genotypes with different geographical locations had greater diversity. The marker data were able to discriminate the landraces from other genotypes. In cluster B, most of the genotypes including Pusa Bold, Pusa Vijay, Pusa Agrani, Pusa Tarak, and Pusa Jaikisan are developed at ICAR-Indian Agricultural Research Institute, New Delhi. Similar grouping of genotypes was also found by Singh et al. (Singh et al. 2013b) in *B. juncea* by RAPD marker analysis. The quality mustard lines were grouped into cluster G and originated from the same center *viz.*, PDZ1, PDZ3 and PDZ4 from ICAR-IARI, New Delhi whereas RLC3 and RLC4 from Punjab Agricultural University, Ludhiana. These lines are either single low (low erucic acid) or double low (low erucic acid and low glucosinolate).

Structure analysis of the Indian mustard genotypes was conducted to understand the population structure in Indian mustard genotypes. The population of Indian mustard was divided into 8 groups on the basis of 50 polymorphic SSR markers. Unlike these results, Chen et al. (Chen et al. 2017) studied the population structure of 582 accessions of rapeseed (*B. napus*) with the help of 30 SSR markers and found 2 major groups. However, Gosh et al. (Ghosh et al. 2019) observed 6 clusters in the population of 47 Indian mustard genotypes based on 19 polymorphic SSR markers. Similar to our results, Chen et al. (Chen et al. 2020) conducted a genetic study in 537 accessions of rapeseed by using 30 SSR markers and observed 8 genetic clusters within 537 rapeseed accessions. Luo et al. (Luo et al. 2019) studied the genetic architecture of *Camelina sativa*, a distant relative of *B. juncea*, by genotyping by sequencing (GBS) and found 6,192 SNPs. The population STRUCTURE analysis based on these SNPs divided the 213 *C. sativa* accessions into 2 groups.

In the present study, a total of 45 genotypes were having various degrees of admixture. Genotypes like EC–511690, WR–2035, PM–27, PM–28 and BPR–5499 had a high level of admixture in it whereas Heera, PDZ–1, DRMRIJ–1237, DRMRIJ–1240, RH–749, WFYSM–15, NE–68, TM–1, and UP I–70 had the least amount of admixture with a Q value ≥ 0.96. Chen et al. (Chen et al. 2017) also found varying levels of admixture in 79 genotypes out of 582 accessions of *B. napus*. In another study, out of 537 accessions, Chen et al. (Chen et al. 2020) found the highest level of admixture (94.1%) in Shibushi population of rapeseed. Similarly, Gosh et al. (Ghosh et al. 2019) recorded varying degrees of admixture in 16 genotypes out of 47 accessions of Indian mustard genotypes. Luo et al. (Luo et al. 2019) observed a low level of genetic diversity and a high level of admixture in *C. sativa* accessions due to less representation of wild genotypes in the available germplasm collection. The presence of this level of admixture observed in the current study may be due to the presence of heterozygosity of the improved genotypes.

5. Conclusions

The results of the present investigation showed that the genotypes originated from a particular center clustered together, showing a narrow genetic base in Indian mustard. Also, population structure of Indian mustard showed high level of admixture in some genotypes which revealed the already present genetic lineages mix. Therefore, there is a need for broadening the genetic base and diversifying mustard breeding and prebreeding program for the development of improved varieties and hybrids showing better heterosis.

Declarations

Acknowledgments: The authors gratefully acknowledge the Indian Council of Agricultural Research, New Delhi to provide funds and Germplasm section of ICAR-Directorate of Rapeseed-Mustard Research, Bharatpur for providing the germplasm.

Funding: The funding was provided by Indian Council of Agricultural research, New Delhi.

Conflicts of Interest: The authors declare no conflict of interest.

Author Contributions:

PY designed the experiments and PKR supervised the study. PY, SY, AK and HSM helped in conducting the experiments. AM and RC helped in data analysis and SY helped in writing the manuscript.

References

1. Bansal P, Kaur P, Banga SK, Banga SS (2009) Augmenting genetic diversity in Brassica juncea through its resynthesis using purposely selected diploid progenitors. Int J Plant Breed 3:41–45
2. Bodh PC, Yadav RK, Sharma A, Yadav M, Arora N, Dawar S, Prasad G (2020) Pocket Book of Agricultural Statistics. In: Dir. Econ. Stat. Dep. Agric. Coop. Farmers Welfare. Minist. Agric. Farmers Welfare, Gov. India. https://eands.dacnet.nic.in/PDF/Pocket Book 2019.pdf. Accessed 14 Jul 2021
3. Chen R, Hara T, Ohsawa R, Yoshioka Y (2017) Analysis of genetic diversity of rapeseed genetic resources in Japan and core collection construction. Breed Sci 67:239–247. https://doi.org/10.1270/jsbbs.16192
4. Chen R, Shimono A, Aono M et al (2020) Genetic diversity and population structure of feral rapeseed (Brassica napus L.) in Japan. PLoS One 15:1–14. https://doi.org/10.1371/journal.pone.0227990
5. Chen S, Wan Z, Nelson MN et al (2013) Evidence from genome-wide simple sequence repeat markers for a polyphyletic origin and secondary centers of genetic diversity of Brassica juncea in China and India. J Hered 104:416–427. https://doi.org/10.1093/jhered/est015
6. Duminil J, Di Michele M (2009) Plant species delimitation: A comparison of morphological and molecular markers. Plant Biosyst 143:528–542. https://doi.org/10.1080/11263500902722964
7. Ghosh KK, Haque ME, Parvin S et al (2009) Genetic diversity analysis in Brassica varieties through RAPD markers. Bangladesh J Agric Res 34:493–503
8. Ghosh S, Mazumder M, Mondal B et al (2019) Morphological and SSR marker-based genetic diversity analysis of Indian mustard (Brassica juncea L.) differing in Alternaria brassicicola tolerance. Euphytica 215:1–19. https://doi.org/10.1007/s10681-019-2523-1
9. Gupta VP, Sekhon MS, Satija DR (1991) Studies on genetic diversity, heterosis and combining ability in Indian mustard [Brassica juncea L.(Czern & Coss)]. Indian J Genet 51:448–453
10. Huangfu C, Song X, Qiang S (2009) ISSR variation within and among wild Brassica juncea populations: implication for herbicide resistance evolution. Genet Resour Crop Evol 56:913–924
11. Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332
12. Jain A, Bhatia S, Banga SS et al (1994) Potential use of random amplified polymorphic DNA (RAPD) technique to study the genetic diversity in Indian mustard (Brassica juncea) and its relationship to heterosis. Theor Appl Genet 88:116–122

13. Kumar M, Choi JY, Kumari N et al (2015) Molecular breeding in Brassica for salt tolerance: Importance of microsatellite (SSR) markers for molecular breeding in Brassica. Front Plant Sci 6:1–13. https://doi.org/10.3389/fpls.2015.00688
14. Kumar R, Gupta VP (1985) Isozyme studies in Indian mustard (Brassica juncea L.). Theor Appl Genet 70:107–110
15. Lodhi B, Thakral NK, Singh D et al (2016) Genetic diversity analysis in Indian mustard (Brassica juncea). J Oilseed Brassica 1:57–60
16. Luo Z, Brock J, Dyer JM et al (2019) Genetic diversity and population structure of a Camelina sativa spring panel. Front Plant Sci 10:1–12. https://doi.org/10.3389/fpls.2019.00184
17. Mishra A, Kumar P, Shamim M et al (2019) Genetic diversity and population structure analysis of Asian and African aromatic rice (Oryza sativa L.) genotypes. J Genet 98:. https://doi.org/10.1007/s12041-019-1131-0
18. Perrier X, Jacquemoud-Collet JP (2006) DARwin software: Dissimilarity analysis and representation for windows. Website http//darwin cirad fr/darwin [accessed 1 March 2013]
19. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
20. Saleem N, Jan SA, Atif MJ et al (2017) Multivariate based variability within diverse Indian mustard (Brassica juncea L.) genotypes. Open J Genet 7:69
21. Shi J, Huang S, Zhan J et al (2014) Erratum: Genome-wide microsatellite characterization and marker development in the sequenced brassica crop species (DNA Research (2014) 21 (53–68) DOI:10.1093/dnares/dst040). DNA Res 21:229. https://doi.org/10.1093/dnares/dsu011
22. Siddiqui SA, Rai PK, Kaushal P et al (2011) Efficient modified DNA extraction method for marker identification. Natl Acad Sci Lett 34:239–245
23. Singh BK, Mishra DC, Yadav S et al (2016a) Identification, characterization, validation and cross-species amplification of genic-SSRs in Indian Mustard (Brassica juncea). J Plant Biochem Biotechnol 25:410–420. https://doi.org/10.1007/s13562-016-0353-y
24. Singh J, Sharma PC, Sharma SK, Kumar A (2013a) Standardization of the fourier transform near-infrared reflectance spectroscopy for estimation of some oil quality parameters in mustard (Brassica spp.)
25. Singh KH, Shakya R, Mahawar R (2014) Genetic diversity and patterns of variation among indian mustard (Brassica juncea (L.) Czernj. & Cosson) varieties. SABRAO J Breed Genet 46:329–339
26. Singh KH, Shakya R, Thakur AK et al (2013b) Genetic diversity in indian mustard [Brassica juncea (L.) Czernj & Cosson] as revealed by agronomic traits and RAPD markers. Natl Acad Sci Lett 36:419–427. https://doi.org/10.1007/s40009-013-0149-8
27. Singh N, Vasudev S, Kumar Yadava D et al (2013c) Assessment of genetic diversity in Brassica juncea Brassicaceae genotypes using phenotypic differences and SSR markers. Rev Biol Trop 61:1919–1934

28. Singh R, Semwal DP, Bhatt KC (2016b) Genetic variability and diversity pattern for agromorphological traits in Indian mustard germplasm. Indian J Plant Genet Resour 29:52–58
29. Snowdon RJ, Friedt W (2004) Molecular markers in Brassica oilseed breeding: Current status and future possibilities. Plant Breed 123:1–8. https://doi.org/10.1111/j.1439-0523.2003.00968.x
30. Song KM, Osborn TC, Williams PH (1988) Brassica taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs) – 1. Genome evolution of diploid and amphidiploid species. Theor Appl Genet 75:784–794. https://doi.org/10.1007/BF00265606
31. Stift M, Kolář F, Meirmans PG (2019) Structure is more robust than other clustering methods in simulated mixed-ploidy populations. Heredity 123:429–441. https://doi.org/10.1038/s41437-019-0247-6
32. Team RC (2021) R: A language and environment for statistical computing
33. Vaishnava A, Sachan JN, Tewari SK (2006) Genetic divergence for important quantitative traits in Indian mustard {Brassica juncea (l.) czern and coss}. Agric Sci Dig 26:269–272
34. Yadava DK, Giri SC, Vignesh M et al (2011) Genetic variability and trait association studies in Indian mustard (Brassica juncea). Indian J Agric Sci 81:712
35. Yadava DK, Sapra RL, Sujata V et al (2009) Selection of high diversity with a minimal set of accessions from Indian mustard (Brassica juncea) germplasm collection. Indian J Agric Sci 79:552–554
36. Yadava DK, Singh N, Vasudev S et al (2012) Combining ability and heterobeltiosis for yield and yield-contributing traits in Indian mustard (Brassica juncea). Indian J Agri Sci 82:563–567
37. Yang J, Zhang C, Zhao N et al (2018) Chinese Root-type Mustard Provides Phylogenomic Insights into the Evolution of the Multi-use Diversified Allopolyploid Brassica juncea. Mol Plant 11:512–514. https://doi.org/10.1016/j.molp.2017.11.007
38. Zhang T, Ni XL, Jiang KF et al (2010) Relationship Between Heterosis and Parental Genetic Distance Based on Molecular Markers for Functional Genes Related to Yield Traits in Rice. Rice Sci 17:288–295. https://doi.org/10.1016/S1672-6308(09)60029-9

Figures

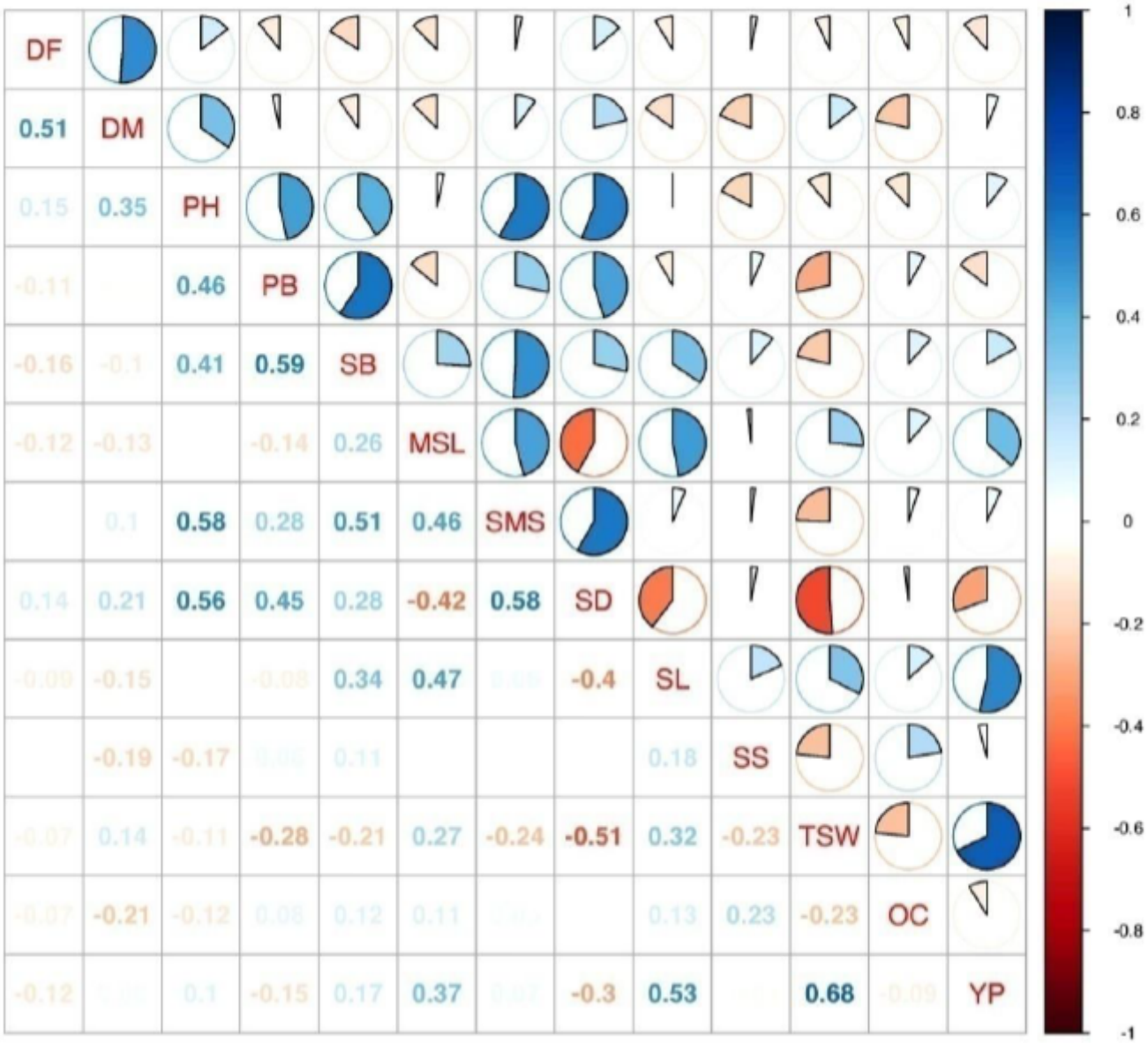

Figure 1

Phenotypic correlation matrix of yield and yield-contributing traits of Indian mustard genotype. Color scale indicates direction and magnitude of phenotypic correlation. Blue color indicates (From lighter to darker) positive correlation between pair of traits and red color indicates (From lighter to darker) negative correlation between pair of traits. White color indicates no correlation.

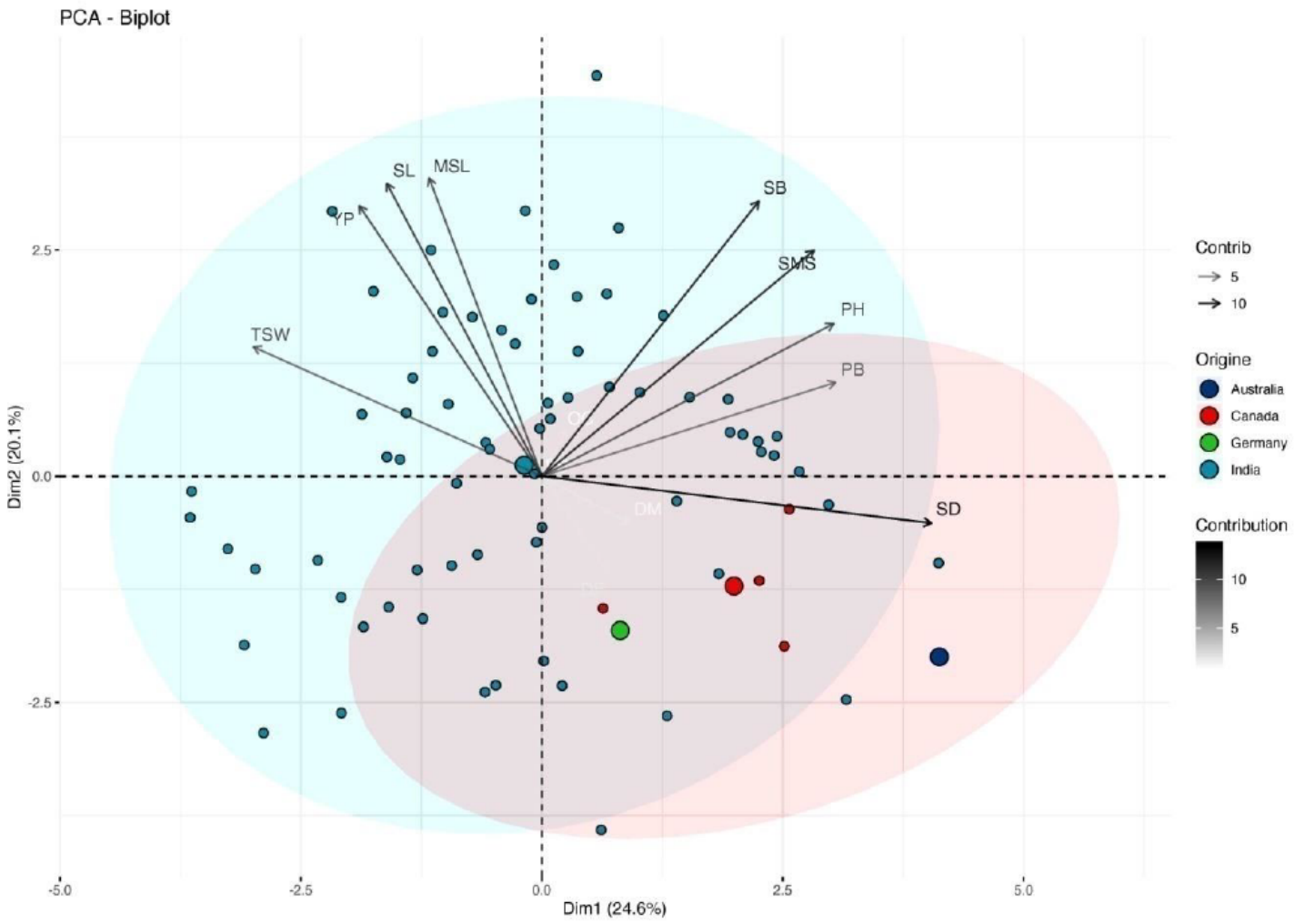

Figure 2

Principal component analysis (PCA) of yield component traits showing their contribution towards total variation in Brassica juncea genotypes.

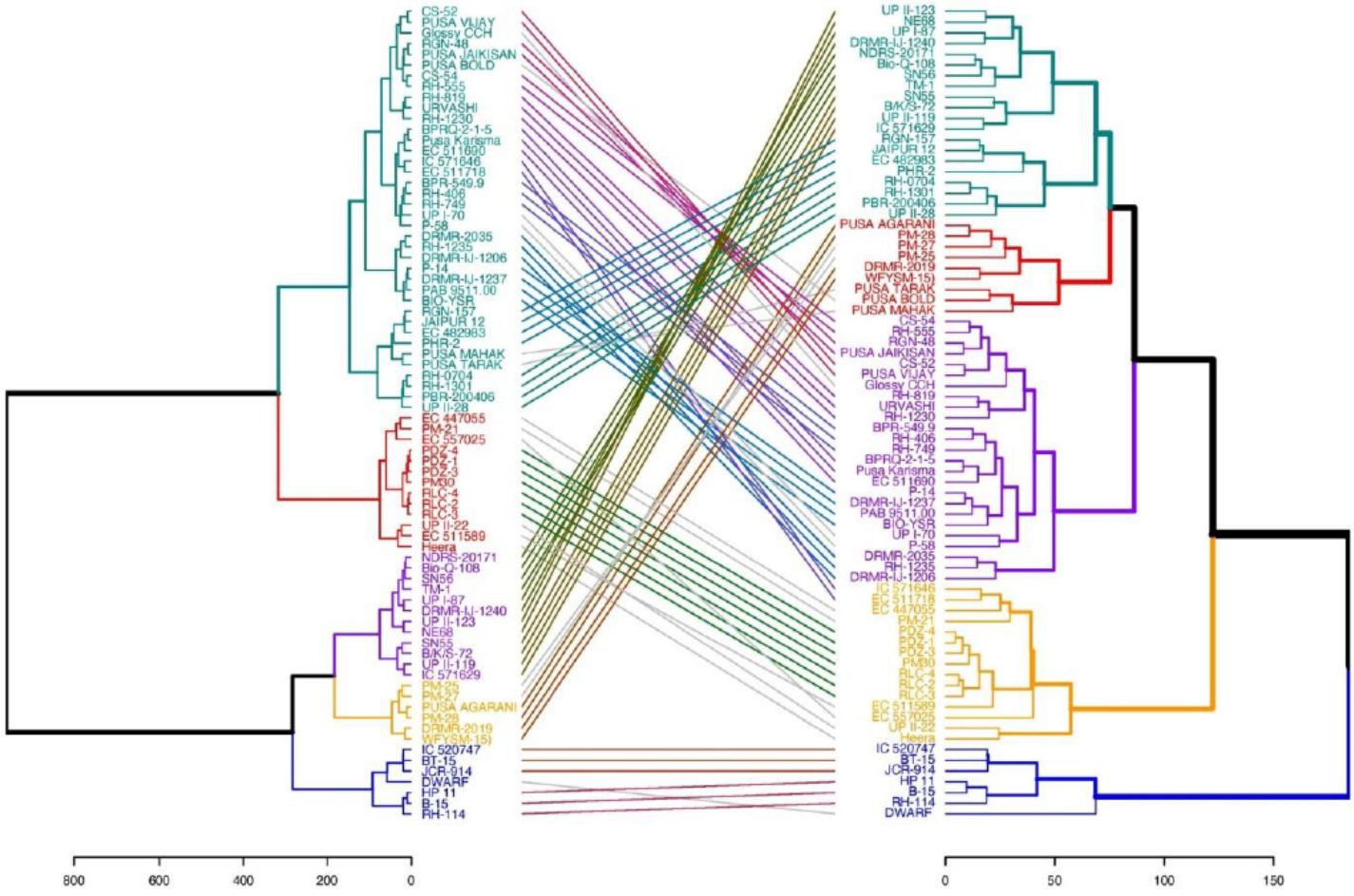

Figure 3

Comparative dendrogram of 76 Indian mustard genotypes based on Completed linkage and Ward's analysis of phenotypic data.

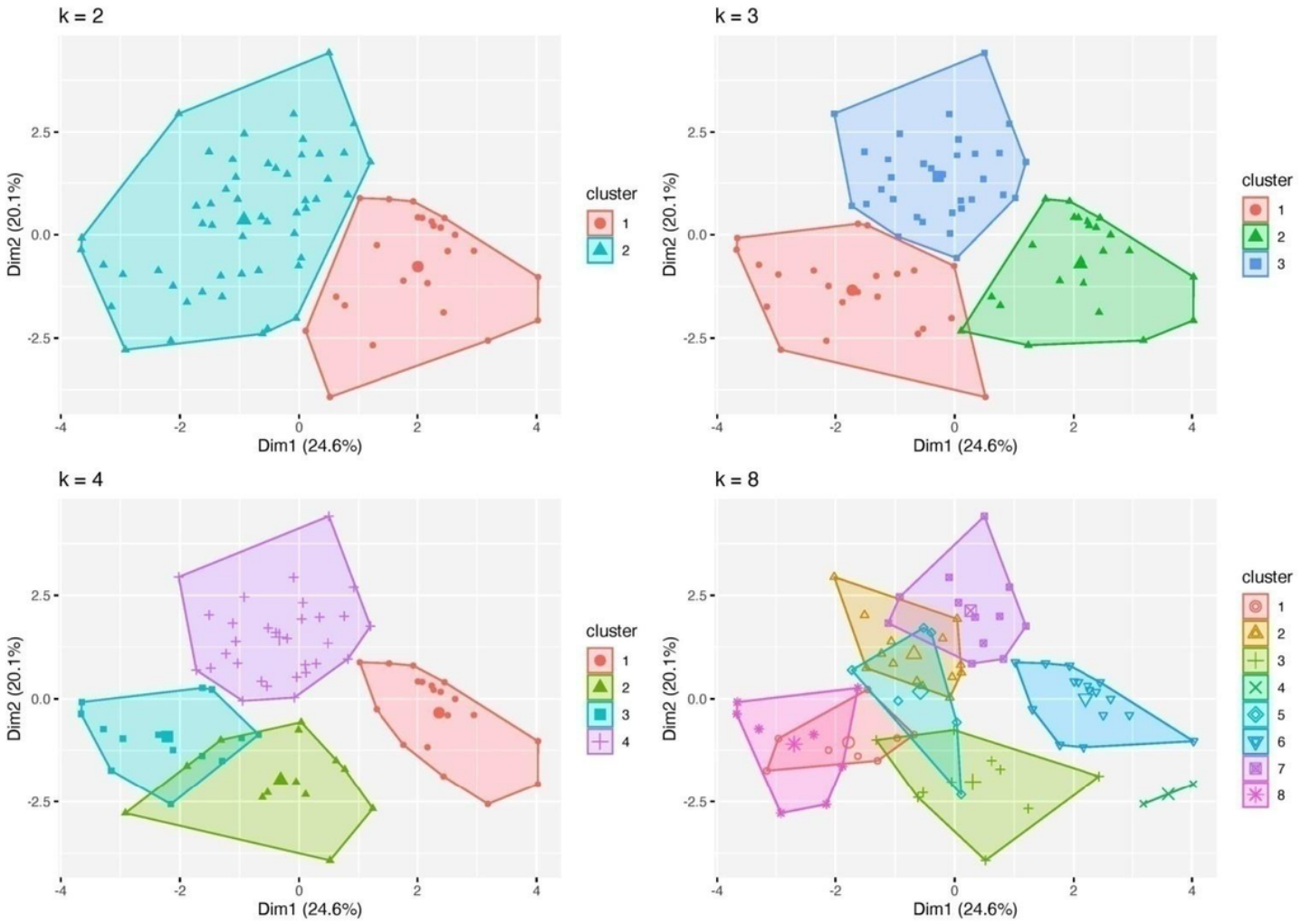

Figure 4

K-mean clustering analysis of 76 Indian mustard genotypes on the basis of their phenotypes at K=2, 3 4 and 8

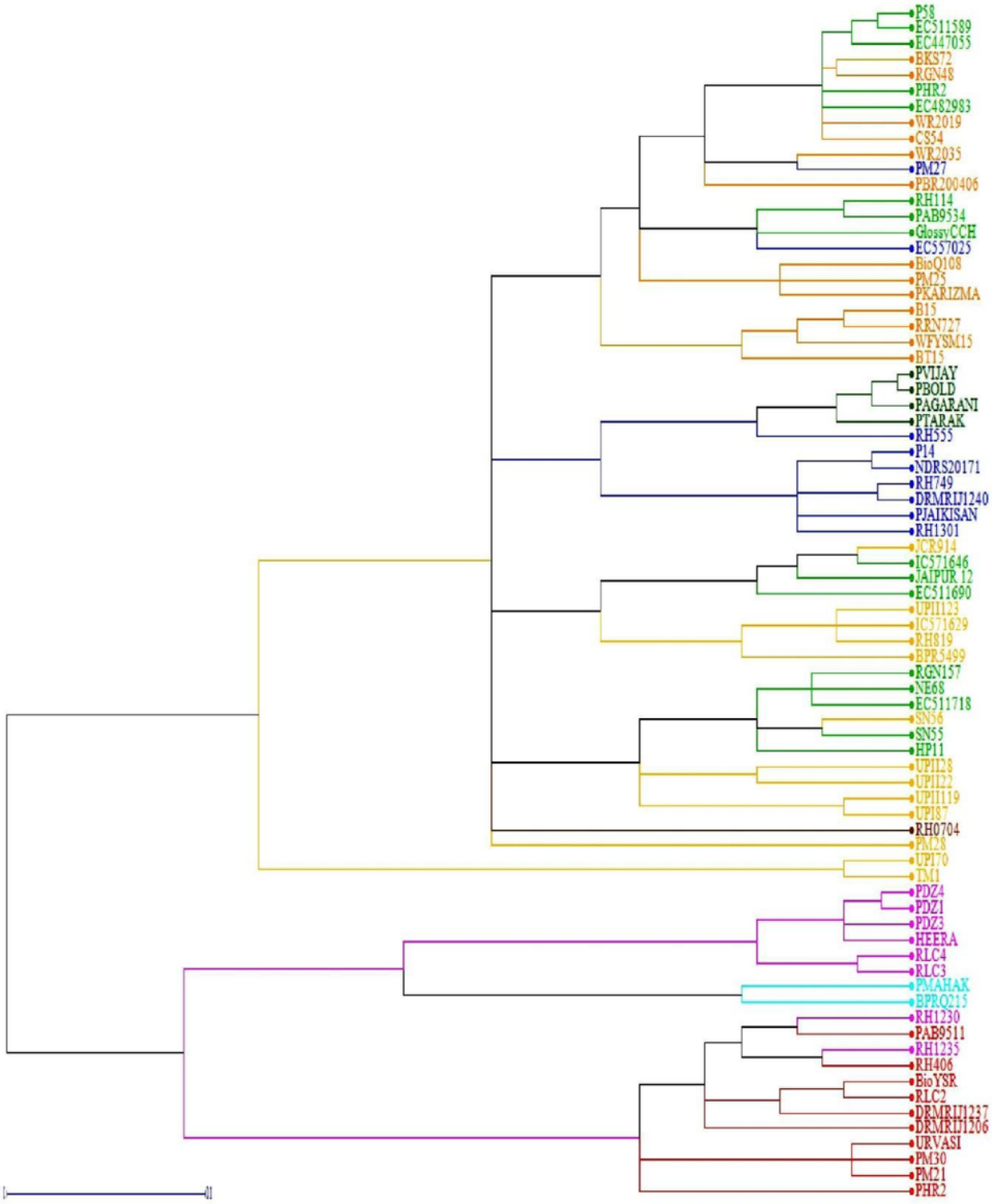

Figure 5

Dendrogram showing genetic relationship in between 76 Brassica genotypes based on genotypic data. The identified groups are labeled with A, B, C, D, E, F, G, H and I.

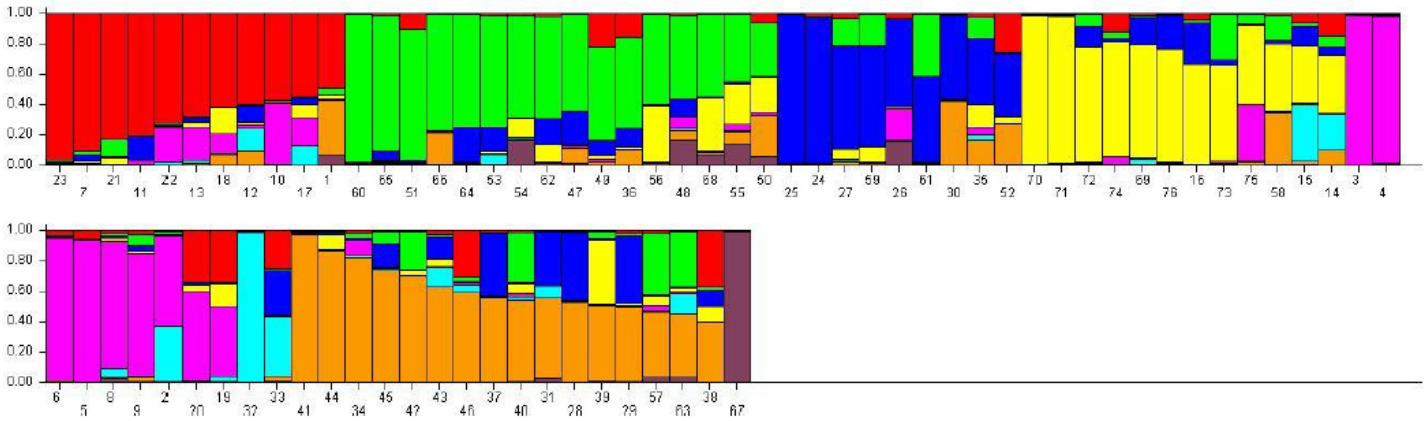

Figure 6

Population structure of 76 Brassica genotypes based on the analysis of genotypic data. Each genotype indicated by vertical bars. The color subsection (within vertical bars) shows the membership coefficient of the genotypes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx