Indian J. Genet. Plant Breed., (2022) 82(2): 177-185

ISSN: 0975-6906

doi: 10.31742/IJGPB.82.2.6

RESEARCH ARTICLE

www.isgpg.org

# Relationship and genetic diversity analysis of *Brassica juncea* and U tringle species using intron polymorphic markers

**Rajat Chaudhary[#], Yashpal[#], Priya Pardeshi, Joghee Nanjundan[2], Sangita Yadav[1], Sujata Vasudev, D. K. Yadava[3], Navinder Saini* and Rajat Chaudhary**

## Abstract

Intron Polymorphism (IP) markers were used to unravel the genetic variation and relationship among 26 genotypes representing six cultivated *Brassica* species described in the classical U triangle. One hundred and twenty-five *Arabidopsis thaliana*-derived IP markers were assayed and 90 to 100% cross-transferability was observed in the six *Brassica* species suggesting that IP markers were highly conserved during the evolution of different *Brassica* species. The number of alleles observed in species at each locus ranged from one to ten with an average of 2.89 alleles per primer pair and there was no consensus between the number of alleles amplified in diploid and tetraploid species. The size range of amplified alleles was 120-1250bp, which reflects enormous deletions/insertions in different alleles. In *B. juncea*, 100% cross-transferability had been obtained and 121 IP markers resulted in polymorphic amplicons with PIC value of 0.04 to 0.48. The dendrogram divided all the 26 genotypes into two groups composed of *B. napus/B. rapa/B. oleracea and B. carinata/B. nigra/B. juncea*. A-genome present in *B. juncea* and *B. napus/B. rapa* seems distinct from each other and hence provides a great opportunity for generating diversity through resynthesizing amphidiploids from different available sources of Agenome. The A and B genomes are more similar in comparison to C genome in tetra-diploid species.The evolutionary relationship established between various *Brassica* species would support in formulating suitable breeding approaches for widening the genetic base of *Brassica* amphidiploids by exploiting the genetic diversity found in diploid progenitor gene pools.

**Keywords:** *Arabidopsis, Brassica*, cross transferability, genetic diversity, intron polymorphism, markers

## Introduction

*Brassica* species, commonly called rapeseed-mustard, are the third most important oilseed crops of the world after soybean and palm. Canada, China, India, Japan, and Germany are the major rapeseed-mustard growing countries. Brassica is the second most important oilseed crop in India, next to soybean. Brassicas are widely studied model crops in plant taxonomy, evolutionary biology, biotechnology, modern genomics etc. Brassicas have undergone an intriguing biological journey through the evolutionary history of crop plants spread over millions of years. The present-day cultivated oilseed Brassicas consists of three diploid species *viz*. *B. rapa* (AA), *B. nigra* (BB), *B. oleracea* (CC), and three amphidiploid species *viz*.,*B. juncea* (AABB), *B. carinata* (BBCC), and *B. napus* (AACC). The classical U's triangle explains the relationship between the six major cultivated *Brassica* species. Nagaharu and Nagaharu (1935) deduced that three basic diploid *Brassica* species were probably the parents of subsequent amphidiploid ones. It is also interesting to note that the diploid species of *Brassica* are themselves mesopolyploid (Wang et al. 2011; Liu et al. 2014) and studying the history of origin, and evolutionary relationship among different *Brassica* species is a basic

Division of Genetics, [1]Division of Seed Science and Technology, ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India; [2]ICAR-Indian Agricultural Research Institute, Regional Station, Wellington 643 231, Tamil Nadu, India; [4]Indian Council of Agricultural Research, New Delhi 110 001, India

***Corresponding Author:** Navinder Saini, Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India, E-Mail: navin12@gmail.com

[#]First and second authors have equal contribution.

**How to cite this article:** Chaudhary R., Yashpal, Pardeshi P., Nanjundan J., Yadav S., Vasudev S., Yadava D.K. and Saini N. 2022. Relationship and genetic diversity analysis of *Brassica juncea* and U tringle species using intron polymorphic markers. Indian J. Genet. Plant Breed., **82**(2): 177-185 .

**Source of support:** ICAR-IRA-IARI-21-53/Incentivizing Project Mustard

**Conflict of interest:** None.

**Received:** Jan. 2022 **Revised:** April 2022 **Accepted:** May 2022

requirement that, in turn, have wider implication in modern genomics and other branches of research.

Polyploidy is one of the major factors contributing to the genomic structure and evolution in *Brassica* spp. Among the cultivated Brassicas, it has been revealed that diversity is higher in diploid species especially in *B. rapa* and *B. oleracea,* as compared to the polyploid species (Thakur et al. 2018). These diploid species are easily crossable with their amphidiploid species (Nikzad et al. 2020). This is the reason for the great genetic diversity present within *Brassica* species and most importantly, this offers greater advantages for the introgression of useful traits between these species. It has been found that the genetic material and its arrangement are highly conserved among closely related species. The genomic studies showed that *Brassica* species and *A. thaliana* have originated from a common ancestor and then diverged, 12.5-20.4 million years ago (Koo et al. 2011). Comparative genetic and physical mapping between *A. thaliana* and *Brassica* species revealed the conserved sequences and the co-linearity of genes. However, the variation in the gene content might have resulted from their diversion through insertion, deletion, and chromosomal rearrangements (Cheung et al. 2009; Lysak et al. 2005).

Among *Brassica* species, a complete genome has been sequenced for *B. oleracea* and *B. rapa*. However, very little genomic information is available for other members of the *Brassicaceae* family, particularly of *B. juncea* (AABB), *B. carinata* (BBCC) and *B. nigra* (BB). The plant genomes have genes with large introns. The alignment of spliced transcripts to the genomes has revealed a large diversity in the intron size. Despite being of diverse lengths, introns have been a major resource for molecular-marker development in several crop species (Poczai et al. 2011; Zhao et al. 2009) and have been recently leveraged to develop marker resources *viz*., Intron-spanning markers (ISMs) for legumes. There are several advantages of using Intron Polymorphic (IP) markers as they are co-dominant and multi-allelic in (Badoni et al. 2016; Panjabi et al. 2008), offer less expensive PCR-based assay, high resolvability, scorability, and reproducibility which make them an excellent marker system for determining phylogenetic relationships among closely related taxa. Sequence homology is found among the IP loci flanking regions of related species (Koo et al. 2011). A large number of IP markers have been developed from well-studied *Arabidopsis* and many of these markers have been shown to be applicable within and between the *Brassica* species (Panjabi et al. 2008, Sharma et al. 2016). This study used IP markers to unravel genetic variations in three diploids (*B. rapa, B. nigra* and *B. oleracea*) and three amphidiploid *Brassica* species. We evaluated the variation in the patterns of *Arabidopsis*-derived IP markers (Panjabi et al. 2008), amplification in terms of their cross-transferability, and allelic variation across *Brassica* species. This work will demonstrate the feasibility of IP markers in resolving the phylogenetic relationships of *Brassica* species and estimate the genetic diversity present in *B. juncea*, a very important edible oil yielding species in India.

## Materials and methods

### *Plant materials*

Two genotypes of each of the five *Brassica* species *viz. B. carinata* (BBCC), *B. napus* (AACC), *B. nigra* (BB), *B. rapa* (AA), *B. oleracea* (CC), and sixteen genotypes of *B. juncea* (AABB), cultivated mainly for oil were used in the present study for relationship and genetic diversity studies (Table 1). The *B. juncea* genotypes used in the present study were diverse for morphological, oil, and meal quality traits. Leaf samples from all the twenty-six genotypes were harvested and stored at 80°C in a deep freezer.

#### *Genomic DNA isolation, purification, and quantification*

Total genomic-DNA from young leaves was extracted and purified using standard CTAB method (Doyle and Doyle 1990). The quality and quantity of the extracted DNA were evaluated by determining the A260/A280 absorbance ratio by spectrophotometer (UV-Visible ElicoSpectrophotometer). DNA concentration and purity were also estimated using0.8% agarose gel electrophoresis and by comparing the known concentration of λ-DNA with the unknown samples. A portion of DNA was diluted in molecular grade water to a concentration of 10ng/µl and stored at -20°C.

#### *Intron Polymorphism (IP) markers and PCR analysis*

The sequence of IP primers used in the present study are listed in Supplementary Table S1(Panjabi et al. 2008). PCR was carried out in a 10 µL reaction cocktail with 25 ng of genomic DNA, 1X PCR buffer, 0.1 mM of each dNTP, 1U Taqpolymerase (Vivantis), and 10 pmol each of forward and reverse primer. Conditions for PCR amplification were as follows: 94°C for 4 minutes, then 40 cycles each at 94°C for 30 sec, 55°C for 30 sec and 72°C for 1 minute, followed by a final extension at 72°C for 5 minutes.

#### *PCR fragment separation, visualization and, data analysis*

Amplified DNA bands of all the 26 samples per primer were separated in a 4% high-resolution agarose gel (Amresco SFR™) containing 0.01% ethidium bromide, prepared in 1X TAE (40 mM Tris, 20 mM acetic acid, and 1 mM EDTA) using 100bp DNA ladder (G-biosciences) as a standard reference (Fig. 1). The amplified fragments of equal length had been considered as amplified from homologous loci. The numbers of bands were also considered as the number of paralogs of these genes. The total number of alleles identified in all the *Brassica* species under study was determined for each IP marker. For each IP marker, the total number of bands obtained in an individual genotype was considered as total loci and individual alleles scored of each locus/band.

**Table 1.** A list of genotypes used in the study

| Species | Genotypes | Pedigree/Collection | Genotypic characteristics |
|---|---|---|---|
| *Brassica juncea* | PDZM-31 (PDZ-1) | LES 1-27 × NUDHYJ-3 | Double zero variety |
| | RLC-3 | JM06003 × JM06020 | Double zero variety |
| | HEERA | ZYR- 4 × BJ 1058 | East European double zero germplasm line |
| | PM-30 | Bio-902×ZEM-1 | Brown seeded, single zero variety |
| | PUSA KARISHMA | PusaBarani×Zem- 1 | Yellow seeded, single zero variety |
| | BIO-YSR | Clipper/BH75/BK0019 | White rust resistance |
| | RE-8 | East European germplasm | East European germplasm line |
| | PUSA BOLD | Varuna × BIC -1780 | Full season variety (>135 days) |
| | RH749 | RH 781 × RH 9617 | Full season variety (>135 days) |
| | PM-25 | SEJ-8 × Pusa Jagannath | Short duration variety (<110 days) |
| | PM-28 | SEJ-8 × Pusa Jagannath | Short duration variety (<110 days) |
| | VARUNA | Selection from Varanasi local | Full season variety (>135 days) |
| | PUSA JAGANATH | Multi-cross between Varuna/inter-cross derivatives/Synthetic *Brassica juncea* | Full season variety (>135 days) |
| | NRCHB-101 | BL-4 × Pusa Bold | Short duration variety (<110 days) |
| | EC-766602 | Landrace | Early and dwarf genotype |
| | RC-275( LAYAPATA) | Indigenous collection | High glucosinolates, used for salad purpose |
| *Brassica napus* | GSL-1 | (NECN 13× Tribute) × NECN 13 | Canola type |
| | GSL-5 | (NECN 13× Tribute) × NECN 13 | Canola type |
| *Brassica carinata* | NPC-9 (Pusa Aditya) | Derived from the cross *Carinata* early mutant × HC 2 | Drought tolerant variety, suitable for rainfed conditions |
| | IGC -01 (PusaSwarnim) | HC 4 × Early mutant | Drought tolerant, white rust-resistant variety |
| *Brassica rapa* | RAPA-1 | TL-15 | An early maturing toria variety |
| | RAPA-2 | IC-332734 | |
| *Brassica nigra* | NIGRA -1 | IC-281862 | Tall plant type |
| | NIGRA -2 | IC-393266 | Tall plant type |
| *Brassica oleracea* | *Brassica oleracea* var. *botrytis* | PUSA MEGHNA | Early maturing cauliflower variety susceptible to black rot and downy mildew |
| | *Brassica oleracea* var. *capitata* | GOLDEN ACRE: A single plant selection from the German cultivar 'Ditmarscher' | Early variety of cabbage |

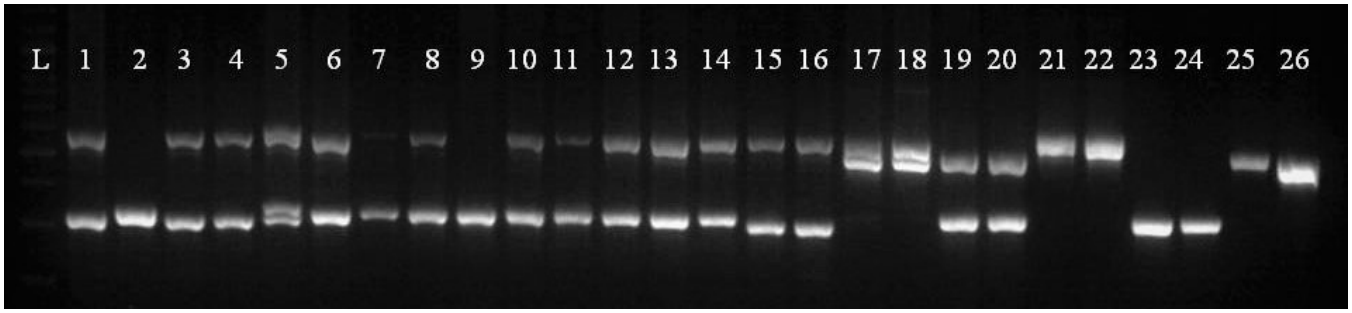


**Fig. 1. At5g24314,L.Ladder, 1.PDZM31, 2.RLC-3, 3.Heera, 4.PM-30, 5.Pusa Karishma, 6.Bio-YSR, 7.RE-8, 8.Pusa Bold, 9.RH749, 10.PM-25, 11.PM-28, 12.Varuna, 13.Pusa Jagannath, 14.NRCHB-101, 15.EC-766602, 16.RC-275, 17.GSL-1, 18.GSL-5, 19.NPC-09, 20.IGC-01, 21.Rapa-1, 22.Rapa-2, 23.Nigra-1, 24.Nigra-2, 25.Pusa Meghna and 26.Golden Acre**

The presence or absence of bands for each allele/locus were assigned as 1/0. Further, the data scored, based on the presence or absence of bands, creating a binary data matrix of 0 and 1 for each marker system, were evaluated using the MEGA 5.2 software (http:www.megasoftware.net/). The data matrices were used to estimate genetic

resemblance based on Jaccard's similarity coefficient and a dendrogram/PCA showing relationships along with 26 genotypes was created using UPGMA method (Rohlf 2000). The polymorphic information content (PIC) value of each IP marker was estimated (Botstein et al. 1980) using the formula; PIC = 1- $\Sigma(Pi)^2$, where Pi is the frequency of the $i^{th}$ allele of each IP marker.

The genotypic data of *B. juncea* were also analyzed using the model-based STRUCTURE v.2.3.4 software (Pritchard et al. 2000) to calculate the most probable number of clusters (K value). The K value was estimated by running an admixture and allied frequency model with K=1 to 10 (10 replication per K value); the burn-in time of each run and MCMC (the Monte Carlo Markov Chain) lengths were both set to 100,000. The online software STRUCTURE HARVESTER was used to determine the optimal number of K values (Earl and Vonholdt 2012). This program follows the ΔK method of Evanno et al. (2005).

## Results

### *Cross transferability of IP markers among Brassica species*

A subset of 125 IP markers selected from an earlier reported set of 1180 IP markers (Panjabi et al. 2008), developed from the intronic sequences of *Arabidopsis* genes, was used to study the cross-transferability and relationship among U triangle's *Brassica* species. The cross transferability of IP markers was found to be 100 percent (maximum) in *B. juncea* and minimum in *B. rapa* (92.8%) across the cultivated *Brassica* species used in this study (Fig. 2). The average percentage of cross-amplification of IP primers across the six species was found to be 97.38% in the present investigation. The number of loci amplified by each IP marker ranged from 1

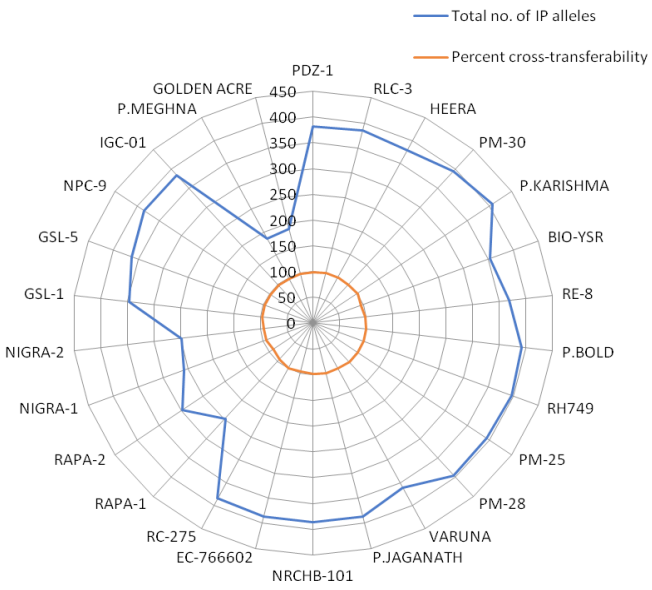


**Fig. 2. Number of cross-transferable *Arabidopsis*-derived IP markers in different genotypes of Brassicaceae.**

to 10 with an average of 2.78 loci per primer, being highest in *B. juncea* cv Pusa Karishma (3.26 per marker) and lowest in *B. oleracea* var. *botrytis* cv. Pusa Meghna (1.49 per marker). The size of amplified loci ranged from 120 to1250bp, indicating enormous deletions/insertions in different loci (Supplementary Table S2). No consensus was found between the number of loci amplified in diploid and in the tetraploid *Brassica* species. Interestingly, four IP markers (At1g65840, At2g18410, At4g26240, and At5g15930) were found to be monomorphic, both in diploid as well as in tetraploid species. However, only one allele was amplified in diploid species, whereas two loci were amplified in the tetraploid species. The IP marker 'At4g09760' of Protein kinase superfamily protein gene did not amplify any locus in the diploid progenitors. However, two loci were amplified in the tetraploid species *viz*. *B. juncea* and *B. napus, the* exception being *B. carinata*. Another set of six IP markers (At1g72420, At2g30130, At3g51100, At4g09760, At5g15930, and At5g52920) amplified in *B. juncea* (AABB) but did not amplify in *B. rapa* (AA), indicating changes at a genomic level during evolution.Interestingly, IP marker 'At2g38880' designed for 'Nuclear factor Y' genes, amplified the maximum number of loci in all the tetraploid species (e.g.,15 alleles in *B. juncea*) while it amplified eight loci in the diploid *B. rapa* and *B. nigra* and only two loci in *B. oleracea*.

### *In-silico analysis of IP markers*

A subset of about 20 IP markers was selected for *in-silico* analyses using NCBI Gene Bank (https://www.ncbi.nlm.nih.gov/) and Brassicaceae Database BRAD (http://brassicadb.cn) to assess the number of alleles present in the genome (Supplementary Table S3). These twenty IP primer sequences were blasted against the genomes of the diploid species and scored for the bands less than 5kb, taking into account both forward and reverse primer sequences. It was interesting to observe that there was more or less consensus between the numbers and the size of the amplified alleles and *in-silico* generated alleles in *B. oleracea* and *B. rapa* however, no consensus was observed for *B. nigra*. More number of alleles were amplified in PCR for *B. nigra*, as compared to *in-silico* identified alleles. Another interesting observation during the *in-silico* analysis was that many alleles having the same size,were found to be present two or three times, either on the same chromosome or on different chromosomes. At1g10840 amplified three alleles in PCR for all the genotypes of all the species. *In-silico* analysis also found the three alleles in diploid species, but two alleles of size 790bp and 661bp were found to be present in duplication at two positions on chromosome A9 and A8 respectively in *B. rapa*. Similarly, the same IP marker revealed two alleles of 695bp and 793bp at B3 and B2 in *B. nigra* and only 743bp at C8 in *B. oleracea*. Most of the IP markers showed a similar type of duplication in the genome in all the species.

### *Relationship among species*

IP markers were used to provide baseline evidence to clarify the possible origins of various diploid and amphidiploid *Brassica* species and to decipher the possible relationship. Molecular-genetic relationships among the *Brassica* species were analyzed through principal coordinate analysis (PCoA) based on a pair-wise distance matrix across all genotypes. All the genotypes were clustered in their respective species group. The distance between different clusters was variable (Fig. 3). The *B. oleracea* (CC) showed the maximum distance from *B. juncea* (AABB) genotypes. Among the diploid species, *B. rapa* (AA) and *B. olercea* (CC) were more similar (0.196) to each other, while *B. nigra* (0.163) was more dissimilar. The *B. nigra* (BB) was found almost at an equal distance from both the diploid species *B. rapa* (AA) and *B. olercea* (CC), meaning thereby that the distance between AA and CC-genomes is shorter, and though the distance between AA-BB and CC-BB is almost equal, it was more than the distance between AA-CC genomes. Among the tetraploids, *B. juncea* (AABB) was found closer to *B. carinata* (BBCC) than *B. napus* (AACC). The similarity between the progenitor and tatraploid was varied. *B. juncea* was found to be equidistant from *B. rapa* (0.347) and *B. nigra* (0.366), while *B. carinata* was found more closer to *B. nigra* (0.375) than *B. oleracea* (0.212). Similarly, *B. napus* was found more closer to *B. rapa* (0.380) than *B. oleracea* (0.225). The A and B-genome are more similar in comparison to C-genome in tetra-diploid species. The interspecies distance was found much higher between diploid species (0.162) as compared to tetraploid species (~0.447).

The intra-species distance was highest in *B. oleracea* followed by *B. rapa* (SupplementaryTable S4) indicating available variability within these species (Fig. 4). The *B. rapa* and *B. oleracea* genotypes used in the present study are also highly diverged morphologically. The least intra-species distance was observed in *B. carinata* genotypes.

### *Diversity in B. juncea*

The hundred percent cross-transferability had been obtained for *B. juncea,* where 125 IP markers showed successful amplification (Supplementary Table S5). The 125 IP markers amplified a total of 581 alleles, ranging from 2 (At1g19240) to 15 (At2g38880) alleles per marker (mean, 4.65; Supplementary Table S2). The average percentage of cross-amplification of IP primers was 98.31% with an average 3.07 loci per primer. The genetic similarity was ranged from 0.524 to 0.903. The genetic similarity coefficient was higher between PM-25 and PM-28 (0.903), indicating a close genetic relationship and a small genetic difference between them, while the genetic similarity coefficient between Varuna and Heera (0.524) was lower (Supplementary Table S4). PIC value ranged from 0.04 (At5g16210) to 0.48 (At4g01897) with an average of 0.22 and heterozygosity ranged from 0.06 (At1g18340) to 0.91 (At4g09760) with an average of 0.44. The diversity ranged from 0.08 to 0.32 with an average of 0.22.

The STRUCTURE software analyzed the population structure and genetic relationship among *B. juncea* genotypes. The K-value was used to estimate the number of clusters of the genotypes based on the genotypic data. The K-value was plotted against delta K, which showed a sharp peak at K = 2 (Supplementary Table S6). The estimated linkage probability revealed the optimum at two sub-populations (pop1 and pop2) (Fig. 5). The grouping of *B. juncea* genotypes into two major groups by STRUCTURE was consistent with the PCA results. Population 1 consisted of all the conventional genotypes while population 2 had all the quality genotypes except RE-8. Among the quality (double zero/canola varieties) genotypes, Heera and EC-766602 were more diverse, while RLC-3 and Heera were found in the same

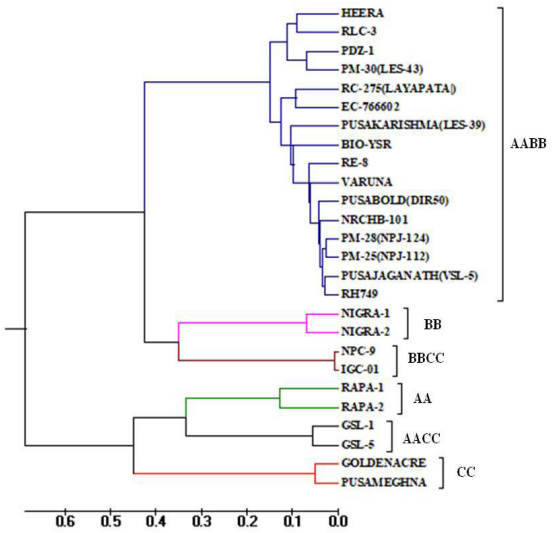


**Fig. 3.** Dendrogram of 26 genotypes depicting the genetic relationship among different species of Brassicaceae based on allelic data of IP markers

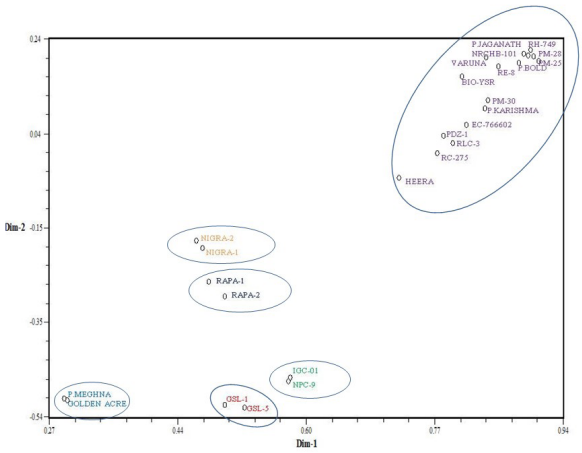


**Fig. 4.** 2D plots of 26 genotypes depicting the genetic relationship among different species of Brassicaceae based on allelic data of IP markers

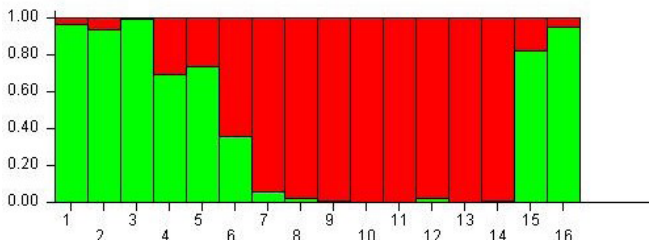


**Fig. 5. Structural analysis of 16 genotypes of B. juncea using IP marker at K=2. Each genotype is indicated by vertical bars. The color subsection (within vertical bars) shows the membership coefficient of the genotypes. The numbers 1–16 represent the same order of the genotype given in Table 1**

clade. Among the conventional genotypes, Pusa Jagannath and RH-749 (full season varieties with duration >135 days) were in the same clade. Similarly, PM-25 and PM-28, the popular short duration (<110 days) *B. juncea* varieties settled in the same clade (Fig.3). Out of 16 genotypes of *B. juncea*, Bio-YSR and PM-30 genotypes had the maximum amount of admixture as shown in structure analysis.

## Discussion

With the advancement of whole-genome sequencing techniques, the Intron-spanning markers of annotated genes have been developed successfully and effectively deployed in genotyping crop plants for assessing genetic relationships and association mapping (Badoni et al. 2016; Poczai et al. 2011; Zhao et al. 2009). Panjabi et al. (2008) developed 1180 IP markers from spanning intronic sequences of *Arabidopsis* genes that showed strong nucleotide conservation between *Arabidopsis* and the corresponding EST or Genome Survey Sequences (GSS) of any *Brassica* species. We have selected polymorphic 125 *Arabidopsis*-derived IP markers from the study of Panjabi et al. (2008); mostly located on the A and B-genomes (Supplementary Table S1). The cross transferability of IP markers was found to be 100 percent across the cultivated *Brassica* species used in this study (Fig.2). The present findings indicate that these IP sites were previously present or conserved in all the *Brassica* genotypes, further elucidating genome similarity and close relationship among these species. The genes where the IP markers were located involve a broad spectrum of molecular functions including transmembrane protein, homeodomain-like superfamily protein, ribosomal protein, protein precursors, isozyme, proteases, kinase, and so on (Rout et al. 2018). Therefore, the IP markers could well reflect the functional and structural genetic diversity between different *Brassica* species.

In recent years, the genomes of *Brassica* species have been sequenced and assembled (Zhang et al. 2018; Sun et al. 2019; Paritosh et al. 2021). During the evolution of *Brassica* species, the genomes underwent whole-genome triplication followed by a substantial genome reshuffling (Lysak et al. 2005). This genome triplication made genomes assembly more complicated. Hence, genome assembly is more accurate for the diploid species than the other tetraploid species. It was observed that there is more or less consensus between the number and size of the amplified alleles in PCR and *in-silico* generated alleles in the case of *B. oleracea* and *B. rapa*. However, no consensus was found for *B. nigra*. This could be due to the limited information for the genome assembly of *B. nigra* unlike *B. oleracea* and *B. rapa*. It was also observed during the *in-silico* analysis that many alleles with the same size were present two or three times, either on the same chromosome or on a different chromosome. The *Brassica* sequence contigs contain numerous examples of tandem arrays (Town et al. 2006).

The possible relationship among the various diploid and amphidiploid *Brassica* species was studied by using IP markers. Twenty-six genotypes taken in this study, which belong to six different *Brassica* species, were clustered into six main groups (clades). The diploid species were found to be almost at equal distances from each other. The genotypes of tetraploid species *B. juncea* (AABB) appear closer to *B. carinata* (BBCC) than *B. napus* (AACC) (Fig. 4). Therefore, this result suggested that B-genome might be less diverged in the studied genotypes when compared to A and C-genomes. It demonstrates that the A and C-genomes of oilseed *Brassica* species have undergone more genomic changes than B-genome after amphidiploidization and extensive cultivation. Grouping of *B. juncea* (AABB) and *B. napus* (AACC) under two separate groups in our study indicates inherent diversity between the A-genome of both species. A study reported that the A-genome of *B. juncea* and *B. napus* each had an independent origins (Takune et al. 2007) and this information may shed light on the unusual features of divergence in *Brassica*. Thus, introgression of individual A-genome types may be carried out to synthesize *Brassica* amphidiploids to achieve more diversity for breeding objectives.

The intra-species distance was highest in diploid species than tetraploid (Supplementary Table S5), indicating available variability within these species. The least intra-species distance was observed in *B. carinata* genotypes. Due to their global cultivation, *B. rapa* and *B. oleracea* have accumulated much more diversity. The *B. rapa* and *B. oleracea* accessions used in the present study are morphologically highly diverged. Rapa-1(Toria) genotype is bunching type and dwarf, on the other hand, Rapa-2 genotype is tall with bold seeds. Similarly, *B. oleracea* genotypes Pusa Meghna is cauliflower type and Golden Acre is cabbage type. That could be the reason for high genetic diversity despite the small sample size observed in *B. rapa* and *B. oleracea* when compared to other *Brassica* species. *B. carinata* cultivation is mostly restricted to a limited area of Africa and South Asia hence limited genetic variation exists compared to other U triangle's species (Khedikar et al. 2020; Seepaul et al. 2021).

Inter-species distances among these species elucidate their contribution to the evolutionary process.

The inter-species distance was found to be much higher between diploid species (>0.8) as compared to tetraploid species (~0.58). The *B. rapa* (AA) and *B. nigra* (BB) have very little genome similarity (0.17) because these diploid species originated and were cultivated in different regions. The cultivated *B. rapa* have originated in Europe and migrated to East and Central Asia (Arias and Pires 2012) and *B. nigra* originated in the Middle East (Amer et al. 2019). Genomic studies also revealed that *B. rapa* and *B. nigra* evolved from different lineages (Warwick and Black 1991; Pradhan et al. 1992). Among amphidiploid species, *B. juncea* and *B. carinata* were closer than the *B. napus*. It has been conclusively established that *B. nigra* contributed the cytoplasm to *B. carinata* and *B. juncea* although *B. juncea* has originated several times in independent hybridization events involving *B. rapa* as a cytoplasmic donor also, while *B. napus* has the *B. oleracea* cytoplasm only (Kaur et al.2014). The comparative genomic studies have also revealed the closeness of B-genome of *B. juncea* with *B. carinata* than the C-genome of *B. napus* and *B. carinata* (Song et al. 2021). It was quite interesting to note that when comparing the amphidiploids with their progenitors, *B. juncea* and *B. napus* are closer to diploid progenitors *B. rapa* than *B. nigra* and *B. oleracea,* respectively. It indicates the possibility of exchanging genetic material is more frequent from *B. rapa* to amphidiploids than other progenitor (Kaur et al. 2014). It is similar to studying genomic variability using SSR markers (Wang et al. 2011). The study also reported that *B. napus* had almost equal genetic distance with its ancestors *B. oleracea* (CC,0.551) and *B. rapa* (AA, 0.568). This could be due to the extensive breeding programs involving both (AA and CC) genomes frequently to improve the *B. napus*. In the present study, the *B. napus* genotypes grouped with *B. rapa* into *rapa* clade, whereas *B. nigra* and *B. carinata* were placed into *nigra* clade. Liu and Wang (2006) reported that in *B. napus* A-genome was more conserved while C-genome has been altered; and similarly, in *B. carinata*, B-genome was intact and C-genome was drastically modified. The highest interspecies distance between *B. nigra* and *B. napus* in our nuclear genome-based study is in conformity with the findings of plastid genome-based grouping of these two species in different clades by Arias and Pires (2012).

All the 125 IP markers showed successful amplification in six *Brassica* species and a 100% cross-transferability was obtained for *B. juncea* (Supplementary Table S5) and 121(97%) were found polymorphic. Polymorphic information content (PIC) is considered as one of the important features that could be used to assess the differentiation ability of the molecular markers (Botstein et al. 1980). The PIC value ranged from 0.04 (At2g43790) to 0.48 (At4g31720) with an average of 0.22 and heterozygosity ranged from 0.06 (At1g19240) to 0.91 (At5g14670) with an average of 0.44. The diversity ranged from 0.08 to 0.32 with an average of 0.22.

In STRUCTURE analysis, Delta K reached a maximum value at K=2, suggesting that the *B. juncea* genotypes would be divided into two subgroups. The sub-population-1 is composed of all the Canola quality genotypes and exotic germplasm, while the sub-population 2 consists of the conventional indigenous genotypes. The analysis performed using STRUCTURE, UPGMA and PCA yielded similar results, clustering *B. juncea* genotypes into 2 sub-populations. The quality genotypes (erucic acid <2% and glucosinolates < 30ppm) are derived from the East European germplasm and were grouped together. Out of 16 genotypes, Bio-YSR and PM-30 genotypes had the maximum amount of admixture. These two genotypes were developed as intermediate (20-30%) and low erucic acid (<2%) genotypes by crossing the Indian conventional genotypes with Canola quality East European genotype. The possible explanation for this may be the cross-hybridization or gene flow through conscious breeding efforts made by humans for crop improvement programs (Schilling et al. 2018).

Among the *B. juncea* genotypes used in this study, Heera and EC-766602 were more diverse due to their distinct geographical origin as EC-766602 is an East European genotype while Heera is an indigenous genotype. Genetic divergence among the genotypes may arise due to geographical separation or genetic barriers to crossability (Tiwari et al. 2022). The conventional varieties (erucic acid >2% and glucosinolates >30ppm) of Indian mustard developed by the different Indian universities and crop research institutes had narrow genetic diversity (<20%) and thereby all grouped together. A similar kind of narrow diversity among Indian mustard cultivars has been reported in the past (Singh et al. 2014; Sharma et al. 2020) which may be due to the use of common parentage in their pedigree (Chauhan et al. 2011).

The genotypes of different subgroups may carry diverse genes for different agronomical traits. Strategic use of diverse genotypes in the breeding program would allow a systematic expansion of these gene complexes to improve existing *Brassica* germplasm in terms of improved yield, more resistance/tolerance to major biotic and abiotic stresses, which are presently limiting the productivity of Indian rapeseed mustard cultivars.

## Authors' contribution

Conceptualization of Research (NS and DKY); Designing of experiments (RC,Y); Contribution of the experimental materials (DKY, SV, NS, Y); Execution of the field/lab experiments and data collection (RC, Y, PP, JN, SY); Analysis of the data and interpretation (NS, Y, and RC); Preparation of the manuscript (RC, Y, NS, SV).

## Supplementary materials

Supplementary Tables S1 to S6 are provided.

## Acknowledgement

Authors are grateful to get the financial grant received from the ICAR-IRA-IARI-21-53/Incentivizing project-Mustard for conducting the research.

## References

Amer W., Shoulkamy M., Faried A. and El-Baset A. 2019. Auto-taxonomy of *Brassica nigra* (L.) Koch (Brassicaceae) in Egypt. Egypt. J. Bot., **59**(2): 439-450.

Arias T. and Pires J. C. 2012. A fully resolved chloroplast phylogeny of the *Brassica* crops and wild relatives (Brassicaceae: Brassiceae): Novel clades and potential taxonomic implications. Taxon., **61**: 980-988.

Badoni S., Das S., Sayal Y. K., Gopalakrishnan S., Singh A. K., Rao A. R., Agarwal P., Parida S. K. and Tyagi A. K. 2016. Genome-wide generation and use of informative intron-spanning and intron-length polymorphism markers for high-throughput genetic analysis in rice. Sci. Rep., **6**: 23765.

Botstein D., White R. L., Skolnick M. and Davis R. W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet., **32**: 314-331.

Chauhan J. S., Singh K. H., Singh V. V. and Satyanshu K. 2011. Hundred years of rapeseed-mustard breeding in India: accomplishments and future strategies. Indian J. agric. Sci., **81**: 1093-1109.

Cheung F., Trick M., Drou N., Lim Y. P., Park J. Y., Kwon S. J. and Bancroft I. 2009. Comparative analysis between homoeologous genome segments of *Brassica napus* and its progenitor species reveals extensive sequence-level divergence. Plant Cell, **21**: 1912-1928.

Doyle J. J. and Doyle J. L. 1990. Isolation of plant DNA from fresh tissue. Focus, **2**: 13-15.

Earl D. A. and Vonholdt B. M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour., **4**: 359-361.

Evanno G., Regnaut S. and Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol., **14**: 2611-2620.

Kaur P., Banga S., Kumar N., Gupta S., Akhatar J. and Banga S. S. 2014. Polyphyletic origin of *Brassica juncea* with *B. rapa* and *B. nigra* (Brassicaceae) participating as cytoplasm donor parents in independent hybridization events. Am. J. Bot., **101**: 1157-1166.

Khedikar Y., Clarke W. E., Chen L., Higgins E. E., Kagale S., Koh C. S., Bennett R. and Parkin I. A. 2020. Narrow genetic base shapes population structure and linkage disequilibrium in an industrial oilseed crop, *Brassica carinata* A. Braun. Sci. Rep., **10**: 12629.

Koo D. H., Hong C. P., Batley J., Chung Y. S., Edwards D., Bang, J. W., Hur Y. and Lim Y. P. 2011. Rapid divergence of repetitive DNAs in *Brassica* relatives. Genomics, **97**(3): 173-185.

Liu A. H. and Wang J.B. 2006. Genomic evolution of *Brassica* allopolyploids revealed by ISSR marker. Genetic Res. Crop Evolut.,**53**:603-611.

Liu S., Liu Y., Yang X., Tong C., Edwards D., Parkin I. A., Zhao M., Ma J., Yu J., Huang S. and Wang X. 2014. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat. Commun., **5**(1): 1-11.

Lysak M. A., Koch M. A., Pecinka A. and Schubert I. 2005. Chromosome triplication found across the tribe Brassiceae. Genome Res., **15**: 516-525.

Nagaharu U. and Nagaharu N. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jpn. J. Bot., **7**(7): 389-452.

Nikzad A., Kebede B., Pinzon J., Bhavikkumar J., Wang X., Yang R. C. and Rahman H. 2020. Potential of the C genome of the different variants of *Brassica oleracea* for heterosis in spring *B. napus* Canola. Front. Plant Sci., **10**:1691.

Panjabi P., Jagannath A., Bisht N. C., Padmaja K. L., Sharma S., Gupt V. and Pental D. 2008. Comparative mapping of *Brassica juncea* and *Arabidopsis thaliana* using Intron Polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C *Brassica* genomes. BMC Genomics, **9**: 1-19.

Paritosh K., Yadava S. K., Singh P., Bhayana L., Mukhopadhyay A., Gupta V., Bisht N. C., Zhang J., Kudrna D. A., Copetti D. and Wing R. A. 2021. A chromosome-scale assembly of allotetraploid *Brassica juncea* (AABB) elucidates comparative architecture of the A and B genomes. Plant Biotechnol. J., **19**(3): 602.

Poczai P., Varga I., Bell N. E. and Hyvonen J. 2011. Genetic diversity assessment of bittersweet (*Solanum dulcamara*, Solanaceae) germplasm using conserved DNA derived polymorphism and intron-targeting markers. Ann. Appl. Biol., **159**: 141-153.

Pradhan A. K., Prakash S., Mukhopadhyay A. and Pental D. 1992. Phylogeny of *Brassica* and allied genera based on variation in chloroplast and mitochondrial DNA patterns: Molecular and taxonomical classifications are incongruous. Theor. Appl. Genet., **85**: 331-340.

Pritchard J. K., Stephens M. and Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics, **155**: 945-959.

Rohlf F. J. 2000. NTSYS-pc: Numerical taxonomy and multivariate analysis system.Version 2.1. Setauket: Exeter Software.

Rout K., Yadav B. G., Yadava S. K., Mukhopadhyay A., Gupta V., Pental D. and Pradhan A. K. 2018. QTL landscape for oil content in *Brassica juncea*: Analysis in multiple bi-parental populations in high and "0" erucic background. Front. Plant Sci., **9**: 1448.

Schilling M. P., Gompert Z., Li F. W., Windham M. D. and Wolf P. G. 2018. Admixture, evolution, and variation in reproductive isolation in the *Boecherapuberula* clade. BMC Evol. Biol., **18**: 61.

Seepaul R., Kumar S., Iboyi J. E., Bashyal M., Stansly T. L., Bennett R., Boote K. J., Mulvaney M. J., Small I. M., George S. and Wright D. L. 2021. *Brassica carinata*: Biology and agronomy as a biofuel crop. GCB Bioenergy, **13**(4): 582-599.

Sharma B. B., Kalia P., Yadava D. K., Singh D. and Sharma T. R. 2016. Genetics and molecular mapping of black rot resistance locus Xca1bc on chromosome B-7 in Ethiopian mustard (*Brassica carinata* A. Braun). PloSOne, **11**(3): e0152290.

Sharma D., Nanjundan J., Singh L., Singh S. P., Parmar N., Sujith Kumar M. S., Singh K. H., Mishra A. K., Ranbir Singh R., Verma K. S. and Thakur A. K. 2020. Genetic diversity in leafy mustard (*Brassica juncea* var. *rugosa*) as revealed by agro-morphological traits and SSR markers. Physiol. Mol. Biol.

Plants, **26**(10): 2005-2018.

Singh K. H., Shakya R. and Mahawar R. K. 2014. Genetic diversity and patterns of variation among Indian mustard (*Brassica juncea* (L.) Czernj. &Cosson) varieties. Sabrao J. Breed. Genet., **46**(2): 329-339.

Song X., Wei Y., Xiao D., Gong K., Sun P., Ren Y., Yuan J., Wu T., Yang Q., Li X., Nie F., Li N., Feng S., Pei Q., Yu T., Zhang C., Liu T., Wang X. and Yang J. 2021. *Brassica carinata* genome characterization clarifies U's triangle model of evolution and polyploidy in *Brassica*. Plant Physiol., **186**(1): 388-406.

Sun D., Wang C., Zhang X., Zhang W., Jiang H., Yao X., Liu L., Wen Z., Niu G. and Shan X. 2019. Draft genome sequence of cauliflower (*Brassica oleracea* L. var. *botrytis*) provides new insights into the C genome in *Brassica* species. Hortic. Res., **6**: 82.

Takune S., Kawahara T. and Ohnishi O. 2007. Phylogenetic relationships among cultivated types of *Brassica rapa* L. em. Metzg. as revealed by AFLP analysis. Genet. Resour. Crop. Evol., **54**: 279-285.

Thakur A. K., Singh K. H., Singh L., Nanjundan J., Khan Y. J. and Singh D. 2018. SSR marker variations in *Brassica* species provide insight into the origin and evolution of *Brassica* amphidiploids. Hereditas, **155**: 1-11.

Tiwari S., Singh Y., Upadhyay P. K. and Koutu G. K. 2022. Principal component analysis and genetic divergence studies for yield and quality-related attributes of rice restorer lines. Indian J. Genet. Plant Breed., **82**: 94-98.

Town C. D., Cheung F., Maiti R., Crabtree J., Haas B. J., Wortman J. R., Hine E. E., Althoff R., Arbogast T. S., Tallon L. J., Vigouroux M., Trick M. and Bancroft I. 2006. Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell, **18**(6): 1348-59.

Wang X. et al. 2011. *Brassica rapa* Genome Sequencing Project Consortium. The genome of the mesopolyploid crop species *Brassica rapa*. Nat. Genet., **43**: 1035-1039.

Warwick S. I. and Black L. D. 1991. Molecular systematics of *Brassica* and allied genera (subtribe Brassicinae, Brassiceae)-chloroplast genome and cytodeme congruence. Theor. Appl. Genet., **82**: 81-92.

Zhang L., Cai X., Wu J., Liu M., Grob S., Cheng F., Liang J., Cai C., Liu Z., Liu B. and Wang F. 2018. Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. Hortic. Res., **5**: 50.

Zhao X., Yang L., Zheng Y., Xu Z. and Wu W. 2009. Subspecies-specific Intron polymorphism markers reveal clear genetic differentiation in common wild rice (*Oryza rufipogon* L.) in relation to the domestication of cultivated rice (*O. sativa* L.). J. Genet. Genom., **36**: 435-442.

**Supplementary Table S1.** Primer Sequence (5′-3′) used in this study for molecular characterization of *Brassica* genotypes

| S.No. | Marker Name | Primer Sequence (5′-3′) | |
|---|---|---|---|
| | | Forward | Reverse |
| 1 | At1g02870 | TGCCAAGGAGTTCGAGC | AAAGGAGCACTCTTTCCATCT |
| 2 | At1g03910 | CTGAAGCCTATCGACGTCC | TCATCCCGAAGCTCTTCC |
| 3 | At1g10840 | GAGACCTTCATGAATTACCAGG | CTTCAGTCTCCAGTTCGGTC |
| 4 | At1g18340 | CTTCTCTCTTGTCAGGATCACTCTC | AACATACTGTTCCGGGCCATCT |
| 5 | At1g19240 | GCGAAATCGGCAGCTTCTTCTT | ACCCGTCTCTTAAGCGTATAACCCAC |
| 6 | At1g23440 | AGACTGCTGGGGAAGGTGCA | ACAACGGAAATGAGCTTCGTTCA |
| 7 | At1g30540 | GAGCCTATTCCGATCCTCG | GGTCTGAGGGATGATTAACACC |
| 8 | At1g31812 | ACGTCTAATCATCATGGGTTTGA | GCTGAACATCCCAGGACGAC |
| 9 | At1g34270 | GAGATGGAGCTTGGAAATGG | GGGAAACCTTCTTCCATGAC |
| 10 | At1g35680 | TGGCTGAACCGGAGACTAC | TTTGTGCCGACAAGCAAC |
| 11 | At1g48440 | CTGCTGCTTCCATCGTCG | GCTCCAGGTCCTTGTTGTACTTG |
| 12 | At1g50240 | CGGATTCCCAACCCAAGAG | GCGTCAGTAAAGTCTAGCAGTGC |
| 13 | At1g57680 | TGCTCAGATCTTCCCCGTT | ATCCACGCATGCTCTCTACCA |
| 14 | At1g65440 | GAGAGCCTGTTCATGSAGATCC | CTTGGAGMAGCAGATAGTCATCTTC |
| 15 | At1g65840 | GGGAACACTCCATCTTGTTTG | GAATTCTCTTGAATGCATCTCC |
| 16 | At1g67060 | CGGTTGCTTGCAAGCTTAC | TTCATTTGCATTCACCTCCA |
| 17 | At1g67170 | ACAGGAGTACCAGCAGTGCA | CTCCTTGATAAGGGTATTGAGCTG |
| 18 | At1g67250 | GGATCTCACCCACTCGAATC | CGTGGTGGAAGTCTACCGG |
| 19 | At1g68310 | GGTGGAACATTGTAGCATGG | ACGTTTGGATTCTCTAGTGCAG |
| 20 | At1g70350 | TTACAYAGYGTTACTGCTTCAGCYTT | CGGAAACWAYTTTCTTGGAATCT |
| 21 | At1g71950 | GGTCCACATCATCTACACCGAG | GCTCGGCACAACTTGAATCAC |
| 22 | At1g72020 | CAARGMTCATAGCTTCATCCG | CTTCTTAGCTCCAGTCGAGTG |
| 23 | At1g72420 | GGAACATTGAAGAATGGATACCAC | CGCATTCCACAGAAACCACT |
| 24 | At1g72710 | GGATCATTCGGAGAGATCTATC | ACTGAGGATGCTTAGTTTTCAC |
| 25 | At1g72890 | AGAGTGGCCGTCCGATTCC | CGACCAGCTTCGAGTCATC(T/A)TC |
| 26 | At1g75330 | CCTCCTCCGTCACGTCCC | GCYTCGGCCATGATTTGGC |
| 27 | At1g76200 | GCAAAGCAAGATGGTCCTGTAG | CGCCATGATCTTCAAAGGTTC |
| 28 | At1g76540 | GCTCGCGATCCYCACATC | GGGAGAGTGAAAGCTCTGGC |
| 29 | At1g77550 | GAATTATCGTCTTGCGGATG | GAACCCAGTTCATCCATCAC |
| 30 | At1g78010 | GGCAGTGTTGTTGACGAGGTAG | CATTCTCAGCTTGCGTAAGGTC |
| 31 | At1g78380 | CTGGGTTGATTTCATCGACAAG | CCTGTCTCTTGTTCCTCACCC |
| 32 | At1g78560 | GCTGGCTTTGTGTTGCAG | CCTGGACAGCAACCAACC |
| 33 | At1g78810 | GCATTCGCTTGCTATTCACAAG | GCATTCTGCTGCATTTGTAACAC |
| 34 | At1g79040 | TCATCGGTGACGTTGAAACC | ACAAGAAGAGCTCCTCCGGC |
| 35 | At1g79950 | CTATCTGGAAGCATGTGGTACAGTC | CGGAAAAATCGAGCAAGGTC |
| 36 | At2g01640 | GGGCAAACCGAGCAAAGG | GGAAAAGATCATTCGCCTTTAG |
| 37 | At2g03870 | GACAAGGGTGTTCAAGTTAAGCT | AGTTCCATCAGTTGGTGACAC |
| 38 | At2g06510 | GATCGGGTTAAGTCAGGACA | ATGGTCTCCATGTTCAGCAC |
| 39 | At2g16860 | GAATCCTTACCACGAGTGTGT | TGGGGTTCTTGGTTATCCTC |
| 40 | At2g17420 | GTACTACGAAGACGAGCGTGG | TCCAAAGCGGCCATGCA |
| 41 | At2g18410 | GCTCGCGCCAGCTCTCAC | GGCATGGATGAATGTCTTAGCA |
| 42 | At2g18900 | AAG(A/T)TTTGGG(A/C)GTCTGAACCA | AGAGAAAGCAGCAGCAGTCA |
| 43 | At2g19260 | ATATGCGGAAAGTGGCGTAG | TCGTGTTTTGGCGTCTGA |
| 44 | At2g19450 | GTGGCTCTGCATGTTCTACTGC | GACGACAAGGAACTGCGATG |

| S.No. | Marker Name | Primer Sequence (5′-3′) | |
|---|---|---|---|
| | | Forward | Reverse |
| 45 | At2g20490 | GACGAAGATGTATCTTCAGTGCT | CCGCTTCTTCAACAAAACTCTCT |
| 46 | At2g23930 | AGGTCAGCCTCCGGATCT | TCAACAACAAGATTCATGAACTG |
| 47 | At2g24765 | TCAAGTTTCAGGTCTGGGATTTAG | CCTGATGCTCGTTTGTCCAC |
| 48 | At2g30130 | ATGGGCGGTCCCGGAT | AGAATCTCGGCTTGAGCCACT |
| 49 | At2g30350 | AACCCTAGATCCCGTCGG | GGATGCTGCCAAGCCCA |
| 50 | At2g33040 | CgCgAAgggAggCgTCTC | TCACACTCTTCATgCggTTACgC |
| 51 | At2g34860 | CTCTTCGCAACCCTTCGC | GCTGGAGATAAGCACACACATG |
| 52 | At2g35790 | CGCCGTCTCTCTTCCTCC | GATGCCCACGTTCTCTGG |
| 53 | At2g36930 | GCGTAAACCAGAGACGGAGA | GAGTTGTGAGTGTGGCGC |
| 54 | At2g38130 | TGGAATCTGGCTGTGAAGAG | TGAACCCGAGTCTCCCATATAG |
| 55 | At2g38880 | CAggAATgCgTCTCTgAgTTCAT | CTTCTCCTTCTggCACTTATCACT |
| 56 | At2g40765 | GGCGCCATATGGGTCGT | TGTCTTCTTAATCCAATCAAATGG |
| 57 | At2g43790 | TgCTTCgTCACATggATCATgA | gATCTgATggAgATCAgTgTCCATT |
| 58 | At2g44970 | CCCACATTCCTTAAGCTCCAG | CGGCTTCATTAAGAGCTGGT |
| 59 | At2g45790 | gCAAgATATCTgAgCAgCTTggC | CCAAggTgCAgCTTCAggCT |
| 60 | At2g46390 | AGATGATCTACCGAAAGTGGAGT | CACTTGTTTGATGAGACATTCTTCT |
| 61 | At3g01060 | TCGGTCGACCCATGTCCGT | TTGTCTTGGCGCCACCCATA |
| 62 | At3g02420 | GGCTTCCTTCTGATTATCTCTCTG | GGAGTATTCAAGAAGGGAGCGT |
| 63 | At3g02860 | GTACAACGAATCAGACCAGCCT | GAAGAATCCCTCTGGAAGAGGT |
| 64 | At3g08690 | GGATCCTCCTTCTAACTGCAGC | GATGGTAAGAGCAGGGCTCCA |
| 65 | At3g09925 | ATATGGATTGGTTCGGCTGC | GGGAAGTAGACAGGCCAGTTGTA |
| 66 | At3g10572 | TGCTCTATGTATGAAGAGGCAGAG | CCACAACTTCCATGTACTCATCAA |
| 67 | At3g12260 | CACCATCATGGACATTTACAATCT | CAGCTCTTCCATTCCCTTGAA |
| 68 | At3g13120 | TCCAGAAACTCTCGATGAACCC | GAGGCACCCAGTATGATCTAAGC |
| 69 | At3g15190 | GCTTCTCTCAGAGCGTTTCTCAA | GCCCATCGAGTGCTTCCAA |
| 70 | At3g16760 | AGTCAAGCACATTCTCATCAGGA | CGTCACATTCTTCT(C/T)GTCGTG |
| 71 | At3g23980 | CAAAGGGATCTTGATGCTTCA | TCCATATCATCTTTTAGTTGGTTGAC |
| 72 | At3g24800 | TTCTCTGCTGCGTTTGCCT | GGAAAGTGAACGTAAGGGTCTCTAC |
| 73 | At3g51100 | CAgCTTCAAgCAACACTTCATC | AAgTTCTTCATCACCCCTgC |
| 74 | At3g54130 | CCGATCTCGACGGGAAGGAGCG | GGGTCTATCTGCGCAGGCTCTGCA |
| 75 | At3g55430 | GCTGGTGTCAGAGATGTTAAGG | CAGGTCATAGGCTATGATGTTAAG |
| 76 | At3g63420 | AAGCACATGATCCTTGCGGAGC | CTTCAAACCACCGGTCCCATCC |
| 77 | At4g01310 | TCTCCTTCGCTTCTGCAGTCTT | TTCTTGATCGCTTTTAGCCGTT |
| 78 | At4g01897 | AGTACGGAACAAGAATGGGAAGAGT | GACCATAGAAGTGAGGGAAGCTATT |
| 79 | At4g05530 | GCATCGACGCAAGGGAT | GGATAAGATTGGGTCAGTAGATGG |
| 80 | At4g09760 | CGATGTAGGAAGCTTCTTGTTCG | GCTGATATGTTGGGATCACGA |
| 81 | At4g11790 | CAATAGCCAGACTGGATCTTTTAGC | CTTCATGAACCACAGTAATACCCTT |
| 82 | At4g15520 | CTGCGTTTCTTCTCGGTAATGAG | GTAACATTCAAAGAGGCAGTGCC |
| 83 | At4g15802 | GTCCAAAATCTTCTCCAGCAGA | CTCATTGATTCTGCCCTCCCAT |
| 84 | At4g16180 | CATCACAGTGCTTTCCTTGACGTA | AACCTCTAGTGCAGTTCCAGC |
| 85 | At4g17050 | CGAATCCAACACTCTCTCCTTC | GGAGGTAAACCTGAACTTGACATT |
| 86 | At4g20150 | CAATGAGCGCCGTTACTTTG | TCCATAACTAATCCCGATCTTCAT |
| 87 | At4g20410 | CTTCGGATCCTGACAAAATGATG | GCATCTCTTGTCCTTTTGAAGC |
| 88 | At4g21720 | ATGATACCGCAGCAA TGGACG | GAGTCACTATCGGCATCACATCC |

| S.No. | Marker Name | Primer Sequence (5′-3′) | |
|---|---|---|---|
| | | Forward | Reverse |
| 89 | At4g23860 | GGCATCTGGTGTTTTCGAGGA | TTGCCCGGGAGAAGCTTACA |
| 90 | At4g25140 | CAAGTCTAGGCAGAT TGCTAAAGCT | CAGCTTCATCCTTGCACTGTC |
| 91 | At4g26240 | CAGCAGTTCATGTCACCATGGGA | AGCAAGCCATGATAGCGATCCA |
| 92 | At4g26840 | CACATCAATCTCAAGGTCAAGG | ACAGATTGCCTGTCACAGTAAGC |
| 93 | At4g27490 | TGATGTATGACCTTATCACAGCTGT | CAATCACGCATGATCTCTCCAA |
| 94 | At4g30790 | AGCCAGTCTGAG CAGAGAGTTGG | AGCAGTTCTCCAAATGCGCAC |
| 95 | At4g31130 | GCCWGCGTCTGGGACTCCAAT | CATCCCTGGAGGAAGAGTTYCCATC |
| 96 | At4g31720 | CAATCTGGTGAGGCA AAGCATGA | GCAACATCTGCAACAAACTTTTGTG |
| 97 | At5g14105 | CTGGCTCGAAGAGGAATCTTCC | GGGATGTCCTTCTTCGGAACC |
| 98 | At5g14670 | GAAATCGTCACCACCATCCC | CCAGATGTGGCACAAGTGC |
| 99 | At5g14920 | CTGCTTCAAATGAGGA(G/A)TCCAA | AGTCCCACATAAAGACACGCAA |
| 100 | At5g15280 | ATGGCGTCGTTGCTTATCAG | CTCTCTCCGAGCCAAACCTC |
| 101 | At5g15400 | TGGTTAGTCATGAGTGGTGGGTT | GGACGACGTTCCGATGCT |
| 102 | At5g15930 | CCCCAATGGAAGGATTGTATC | CTGTTTCATTTCTTCCAAACGC |
| 103 | At5g16090 | TCGCTTCGAGATCCAAGT | CTCCCATCACGGTCTCTATG |
| 104 | At5g16210 | TTCGTCTAGCACAGGAGGATATT | CGTTCATTGTTCTTCAAAGGC |
| 105 | At5g16260 | GCGCTACATGTTCAGTCCAGC | GCAGCGTGGTTCACTGATCC |
| 106 | At5g16400 | CCTCGACATGTACACTCAGTG | GGAACCACTCTAATTCCTAGCTC |
| 107 | At5g17300 | CgTTgACTACgAATgTTCAg | CATCAgTCCATCTCTCTCTTTC |
| 108 | At5g20010 | CTTCAAgCTCgTCATYgTTgg | gCACACTggCCATggATg |
| 109 | At5g22340 | GCTCTGGTTAAGGGAGATTCTCG | AAGCTAGTAAGCAATATCCGCTGC |
| 110 | At5g22640 | ATTGAGGAGTTTCTTCAGTGGGT | GACATCCATTTCCAACCACTT |
| 111 | At5g24314 | CAGAGAGATGATAATGGACGCC | CCACAAACGGAACTCTTTGC |
| 112 | At5g26360 | GCTTCTTGATGCTGGTGGA | GGCAGATGACTGTGGGATG |
| 113 | At5g27740 | CTTgCCgTCTCATCTTATgCTg | gCAACgAACTCCAACACTTTCAC |
| 114 | At5g35360 | ACCGTCAGAGTTTGTGAAACTG | GAAGGTGCTTCTTCCAGCAAC |
| 115 | At5g37580 | CATggCATATATCAggAgACTgAg | gCCTCCATTgAgTTCCATCC |
| 116 | At5g40650 | TggAACCCTgAgTCTTATCTTgg | TCTCgACTgTCgCTTATC CAC |
| 117 | At5g40670 | CTCAgCTgATTTTgAATTTCCg | CATTTgCAgCCACAggTA TCA T |
| 118 | At5g43150 | AGAGTGGTGGAACACGATGG | CTTCGTACTCGCAAGAACTCAC |
| 119 | At5g45610 | GCGGTTTCAATGACGCG | GATAATTTGAGGTGCGCCAGAT |
| 120 | At5g47040 | GGGAGTAGCAGCTCGTGCTCT | TCTTAAACTggCgTgACAgTgC |
| 121 | At5g49510 | GAAGCAAGGAAGGGTACTGG | CCAACCACAAACACACTGAGTC |
| 122 | At5g52920 | GAGACACTTGAATGTCCGAGG | CATGCTACGGCATAGGTTAATG |
| 123 | At5g58730 | CATTGTTAGGGAACCTCTTTGG | CTGCTTCTTCCTTCTCTGTACCTC |
| 124 | At5g61500 | GCGATAATCTCGTCTCCAAGTG | GCTTCATAATCCTCTGCCACAG |
| 125 | At5g63520 | GGAACATCGTCTGCAATCG | GTAGCCAATGGTCAACAGTATAGC |

**Supplementary Table S2.** Details of Intron polymorphism (IP) markers used in the study

| S.No. | Marker Name | Number of alleles (Na) | | | | | | PIC | He | Putative Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | B. juncea | B. napus | B. carinata | B. rapa | B. nigra | B. oleracea | | | |
| 1 | At1g02870 | 3 | 3 | 3 | 2 | 1 | 1 | 0.29 | 0.48 | Nucleolar-like protein |
| 2 | At1g03910 | 3 | 1 | 2 | 3 | 2 | 2 | 0.14 | 0.18 | Cactus-binding carboxy-terminal, cactin |
| 3 | At1g10840 | 4 | 3 | 4 | 3 | 3 | 1 | 0.03 | 0.25 | Translation initiation factor 3 subunit H1 |
| 4 | At1g18340 | 2 | 1 | 1 | 1 | 2 | 2 | 0.06 | 0.06 | Basal transcription factor complex subunit-like protein |
| 5 | At1g19240 | 3 | 1 | 1 | 1 | 3 | 2 | 0.39 | 0.82 | Transmembrane protein |
| 6 | At1g23440 | 4 | 2 | 3 | 3 | 3 | 4 | 0.21 | 0.26 | Peptidase C15, pyroglutamyl peptidase I-like protein |
| 7 | At1g30540 | 4 | 1 | 2 | 1 | 3 | 1 | 0.39 | 0.71 | Actin-like ATPase superfamily protein |
| 8 | At1g31812 | 4 | 3 | 3 | 3 | 1 | 2 | 0.25 | 0.56 | Acyl-CoA-binding protein 6 |
| 9 | At1g34270 | 4 | 3 | 3 | 2 | 2 | 2 | 0.15 | 0.33 | Exostosin family protein |
| 10 | At1g35680 | 4 | 3 | 2 | 2 | 1 | 1 | 0.28 | 0.72 | Ribosomal protein L21 |
| 11 | At1g48440 | 5 | 3 | 4 | 3 | 2 | 1 | 0.17 | 0.42 | B-cell receptor-associated 31-like protein |
| 12 | At1g50240 | 5 | 3 | 3 | 2 | 2 | 2 | 0.29 | 0.47 | Kinase family with ARM repeat domain-containing protein |
| 13 | At1g57680 | 5 | 4 | 6 | 3 | 4 | 3 | 0.05 | 0.40 | Plasminogen activator inhibitor |
| 14 | At1g65440 | 3 | 2 | 3 | 1 | 3 | 2 | 0.41 | 0.50 | Global transcription factor group B1 |
| 15 | At1g65840 | 2 | 1 | 1 | 1 | 0 | 1 | 0.12 | 0.56 | Polyamine oxidase 4 |
| 16 | At1g67060 | 3 | 1 | 2 | 1 | 3 | 1 | 0.07 | 0.33 | Peptidase M50B-like protein |
| 17 | At1g67170 | 3 | 2 | 3 | 4 | 4 | 2 | 0.24 | 0.31 | Sarcolemmal membrane-associated protein |
| 18 | At1g67250 | 6 | 3 | 6 | 2 | 4 | 2 | 0.19 | 0.46 | Proteasome maturation factor UMP1 |
| 19 | At1g68310 | 5 | 2 | 5 | 3 | 4 | 2 | 0.12 | 0.16 | MIP18 family protein |
| 20 | At1g70350 | 2 | 0 | 1 | 1 | 0 | 2 | 0.34 | 0.80 | Hypothetical protein |
| 21 | At1g71950 | 5 | 3 | 5 | 4 | 3 | 2 | 0.19 | 0.54 | Proteinase inhibitor |
| 22 | At1g72020 | 8 | 4 | 5 | 4 | 5 | 2 | 0.19 | 0.40 | TonB-dependent heme receptor A |
| 23 | At1g72420 | 3 | 3 | 3 | 0 | 2 | 1 | 0.20 | 0.22 | NADH:ubiquinone oxidoreductase intermediate-associated protein 30 |
| 24 | At1g72710 | 4 | 2 | 3 | 3 | 3 | 1 | 0.14 | 0.33 | CKL2 - Casein kinase 1-like protein 2 |
| 25 | At1g72890 | 4 | 3 | 3 | 2 | 3 | 1 | 0.06 | 0.28 | Disease resistance protein |
| 26 | At1g75330 | 3 | 1 | 2 | 2 | 2 | 3 | 0.16 | 0.35 | OTC - Ornithine carbamoyltransferase, chloroplastic; |
| 27 | At1g76200 | 8 | 6 | 3 | 4 | 3 | 1 | 0.15 | 0.54 | NADH dehydrogenase ubiquinone 1 beta subcomplex subunit |
| 28 | At1g76540 | 5 | 3 | 4 | 4 | 4 | 2 | 0.08 | 0.19 | CDKB2;1 - Encodes a cyclin-dependent protein kinase |
| 29 | At1g77550 | 6 | 2 | 2 | 2 | 1 | 1 | 0.18 | 0.57 | Tubulin-tyrosine ligases; Involved in protein modification process |
| 30 | At1g78010 | 4 | 2 | 2 | 2 | 3 | 1 | 0.20 | 0.35 | tRNA modification GTPase, |
| 31 | At1g78380 | 4 | 4 | 2 | 4 | 2 | 3 | 0.16 | 0.39 | GSTU19 - A glutathione transferase that is a member of Tau GST gene family. |

| S.No. | Marker Name | Number of alleles (Na) | | | | | | PIC | He | Putative Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | B. juncea | B. napus | B. carinata | B. rapa | B. nigra | B. oleracea | | | |
| 32 | At1g78560 | 5 | 2 | 2 | 2 | 2 | 2 | 0.25 | 0.51 | Sodium/metabolite co-transporter BASS1, chloroplastic |
| 33 | At1g78810 | 2 | 2 | 3 | 2 | 1 | 2 | 0.11 | 0.12 | Small nuclear ribonucleoprotein family protein |
| 34 | At1g79040 | 7 | 5 | 4 | 5 | 1 | 2 | 0.18 | 0.30 | PSBR - Encodes for the 10 kDaPsbR subunit of photosystem II (PSII). |
| 35 | At1g79950 | 5 | 3 | 6 | 1 | 4 | 1 | 0.17 | 0.31 | Regulator of telomere elongation helicase 1 homolog |
| 36 | At2g01640 | 2 | 2 | 3 | 3 | 1 | 1 | 0.17 | 0.87 | Elongator complex protein |
| 37 | At2g03870 | 5 | 3 | 1 | 4 | 1 | 3 | 0.26 | 0.74 | EMB2816 - Small nuclear ribonucleoprotein family protein |
| 38 | At2g06510 | 4 | 3 | 2 | 4 | 1 | 1 | 0.31 | 0.87 | RPA1A - Replication protein A 70 kDa DNA-binding subunit A |
| 39 | At2g16860 | 4 | 2 | 3 | 2 | 1 | 1 | 0.23 | 0.38 | membrane bound O-acyl transferase (MBOAT) family protein |
| 40 | At2g17420 | 4 | 3 | 2 | 3 | 2 | 2 | 0.09 | 0.31 | NTRA - NADPH-dependent thioredoxin reductase A |
| 41 | At2g18410 | 3 | 1 | 0 | 1 | 1 | 1 | 0.40 | 0.62 | Elongator complex protein 5 |
| 42 | At2g18900 | 3 | 1 | 3 | 2 | 3 | 1 | 0.16 | 0.35 | Transducin/WD40 repeat-like superfamily protein |
| 43 | At2g19260 | 4 | 5 | 2 | 3 | 1 | 2 | 0.18 | 0.49 | RING/FYVE/PHD zinc finger superfamily protein |
| 44 | At2g19450 | 2 | 1 | 3 | 1 | 2 | 1 | 0.15 | 0.17 | TAG1 - Membrane bound O-acyl transferase (MBOAT) family protein |
| 45 | At2g20490 | 6 | 5 | 5 | 5 | 3 | 1 | 0.23 | 0.61 | NOP10 - H/ACA ribonucleoprotein complex subunit 3-like protein |
| 46 | At2g23930 | 9 | 4 | 6 | 3 | 4 | 2 | 0.16 | 0.43 | SNRNP-G - Probable small nuclear ribonucleoprotein G |
| 47 | At2g24765 | 7 | 4 | 7 | 4 | 4 | 2 | 0.17 | 0.25 | ARF3 - ADP-ribosylation factor 3 |
| 48 | At2g30130 | 4 | 2 | 3 | 0 | 2 | 1 | 0.21 | 0.61 | ASL5 - Lateral organ boundaries (LOB) domain family protein |
| 49 | At2g30350 | 4 | 3 | 4 | 3 | 2 | 1 | 0.22 | 0.35 | Structure-specific endonuclease subunit SLX1 homolog |
| 50 | At2g33040 | 5 | 5 | 4 | 4 | 2 | 5 | 0.38 | 0.71 | ATP3 - ATP synthase subunit gamma |
| 51 | At2g34860 | 4 | 4 | 4 | 2 | 4 | 1 | 0.24 | 0.35 | DnaJ/Hsp40 cysteine-rich domain superfamily protein |
| 52 | At2g35790 | 6 | 5 | 3 | 5 | 4 | 1 | 0.16 | 0.18 | Alpha/beta-Hydrolases superfamily protein |
| 53 | At2g36930 | 6 | 3 | 4 | 3 | 3 | 1 | 0.19 | 0.44 | Zinc finger (C2H2 type) family protein |
| 54 | At2g38130 | 6 | 3 | 4 | 2 | 2 | 3 | 0.16 | 0.72 | TMAK3 - Acyl-CoA N-acyltransferases (NAT) superfamily protein |
| 55 | At2g38880 | 15 | 10 | 9 | 8 | 8 | 2 | 0.26 | 0.56 | Nuclear factor Y |
| 56 | At2g40765 | 7 | 3 | 4 | 3 | 2 | 4 | 0.36 | 0.55 | Ubiquitin-conjugating enzyme 11 |
| 57 | At2g43790 | 8 | 5 | 5 | 6 | 4 | 3 | 0.24 | 0.45 | MPK6 - Mitogen-activated protein kinase 6 |
| 58 | At2g44970 | 5 | 1 | 2 | 3 | 1 | 1 | 0.20 | 0.53 | Alpha/beta-Hydrolases superfamily protein |
| 59 | At2g45790 | 7 | 5 | 3 | 6 | 3 | 3 | 0.30 | 0.45 | Phosphomannomutase |
| 60 | At2g46390 | 5 | 4 | 3 | 4 | 1 | 2 | 0.09 | 0.42 | Succinate dehydrogenase subunit |
| 61 | At3g01060 | 4 | 3 | 2 | 5 | 2 | 1 | 0.31 | 0.54 | Lysine-tRNA ligase |
| 62 | At3g02420 | 5 | 4 | 4 | 4 | 3 | 3 | 0.27 | 0.46 | Dihydroflavonol 4-reductase/flavanone protein |

| S.No. | Marker Name | Number of alleles (Na) | | | | | | PIC | He | Putative Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | B. juncea | B. napus | B. carinata | B. rapa | B. nigra | B. oleracea | | | |
| 63 | At3g02860 | 6 | 2 | 5 | 4 | 4 | 2 | 0.19 | 0.62 | Zinc ion binding protein |
| 64 | At3g08690 | 8 | 5 | 4 | 6 | 3 | 2 | 0.25 | 0.51 | UBC11 - Ubiquitin-conjugating enzyme E2 11 |
| 65 | At3g09925 | 5 | 4 | 2 | 4 | 3 | 2 | 0.27 | 0.80 | Pollen Ole e 1 allergen and extensin family protein |
| 66 | At3g10572 | 4 | 4 | 2 | 4 | 2 | 3 | 0.16 | 0.38 | 3-phosphoinositide-dependent protein kinase-1 |
| 67 | At3g12260 | 7 | 5 | 5 | 8 | 4 | 2 | 0.18 | 0.45 | LYR family of Fe/S cluster biogenesis protein |
| 68 | At3g13120 | 7 | 5 | 4 | 3 | 3 | 2 | 0.26 | 0.44 | Ribosomal protein S10p/S20e family protein |
| 69 | At3g15190 | 7 | 6 | 7 | 3 | 5 | 3 | 0.09 | 0.11 | PRPS20 - Chloroplast 30S ribosomal protein S20 |
| 70 | At3g16760 | 2 | 2 | 2 | 2 | 0 | 1 | 0.34 | 0.64 | Tetratricopeptide repeat (TPR)-like superfamily protein |
| 71 | At3g23980 | 5 | 5 | 5 | 3 | 2 | 2 | 0.28 | 0.53 | BLI - Protein BLISTER |
| 72 | At3g24800 | 4 | 2 | 1 | 1 | 1 | 2 | 0.12 | 0.48 | PRT1 - E3 ubiquitin-protein ligase PRT1 |
| 73 | At3g51100 | 2 | 2 | 0 | 0 | 3 | 1 | 0.17 | 0.18 | Histone-lysine N-methyltransferase |
| 74 | At3g54130 | 6 | 5 | 3 | 5 | 3 | 0 | 0.21 | 0.53 | Josephin family protein |
| 75 | At3g55430 | 5 | 4 | 4 | 3 | 2 | 2 | 0.33 | 0.45 | Involved in carbohydrate metabolic process |
| 76 | At3g63420 | 4 | 3 | 2 | 3 | 2 | 3 | 0.22 | 0.34 | GG1 - Guanine nucleotide-binding protein subunit gamma 1 |
| 77 | At4g01310 | 4 | 2 | 2 | 1 | 2 | 1 | 0.30 | 0.62 | 50S ribosomal protein L5, chloroplastic |
| 78 | At4g01897 | 4 | 4 | 2 | 4 | 1 | 2 | 0.48 | 0.69 | TBP-associated factor II 15 |
| 79 | At4g05530 | 4 | 5 | 2 | 3 | 0 | 2 | 0.27 | 0.42 | IBR1 - Short-chain dehydrogenase/reductase SDRA |
| 80 | At4g09760 | 2 | 2 | 0 | 0 | 0 | 0 | 0.39 | 0.91 | Protein kinase superfamily protein |
| 81 | At4g11790 | 4 | 2 | 4 | 2 | 3 | 1 | 0.24 | 0.39 | Gibberellin-regulated family protein |
| 82 | At4g15520 | 4 | 5 | 4 | 4 | 4 | 2 | 0.05 | 0.25 | tRNA/rRNA methyltransferase (SpoU) family protein |
| 83 | At4g15802 | 5 | 4 | 4 | 2 | 3 | 2 | 0.35 | 0.69 | HSBP - Encodes a protein with similarity to heat shock factor binding proteins |
| 84 | At4g16180 | 3 | 3 | 4 | 3 | 4 | 2 | 0.07 | 0.08 | Plant adhesion molecule 1 |
| 85 | At4g17050 | 6 | 4 | 3 | 2 | 2 | 2 | 0.34 | 0.45 | UGLYAH - (S) - ureidoglycineaminohydrolase |
| 86 | At4g20150 | 7 | 3 | 4 | 5 | 3 | 2 | 0.33 | 0.67 | HEAT repeat-containing protein |
| 87 | At4g20410 | 4 | 1 | 1 | 1 | 4 | 1 | 0.32 | 0.69 | GSNAP - Gamma-soluble NSF attachment protein |
| 88 | At4g21720 | 4 | 2 | 2 | 2 | 3 | 1 | 0.22 | 0.44 | Thioredoxin F2 |
| 89 | At4g23860 | 4 | 1 | 3 | 1 | 2 | 1 | 0.37 | 0.70 | Homeodomain-like superfamily protein |
| 90 | At4g25140 | 6 | 5 | 2 | 4 | 4 | 2 | 0.16 | 0.37 | Oleosin1, a protein found in oil bodies |
| 91 | At4g26240 | 3 | 1 | 1 | 1 | 1 | 1 | 0.24 | 0.48 | MORN repeat-containing protein |
| 92 | At4g26840 | 6 | 5 | 7 | 4 | 3 | 1 | 0.15 | 0.18 | SUMO1 - Small ubiquitin-related modifier 1 |
| 93 | At4g27490 | 5 | 4 | 3 | 1 | 2 | 1 | 0.23 | 0.73 | 3′-5′-exoribonuclease family protein |
| 94 | At4g30790 | 7 | 3 | 2 | 4 | 4 | 2 | 0.17 | 0.34 | Autophagy-related protein 11 |

| S.No. | Marker Name | Number of alleles (Na) | | | | | | PIC | He | Putative Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | B. juncea | B. napus | B. carinata | B. rapa | B. nigra | B. oleracea | | | |
| 95 | At4g31130 | 4 | 2 | 2 | 2 | 1 | 1 | 0.26 | 0.39 | Keratin-associated protein |
| 96 | At4g31720 | 4 | 3 | 6 | 4 | 4 | 2 | 0.09 | 0.09 | TAFII15 - Transcription initiation factor TFIID subunit 10 |
| 97 | At5g14105 | 8 | 3 | 3 | 2 | 6 | 1 | 0.36 | 0.52 | PQ-loop repeat family protein |
| 98 | At5g14670 | 5 | 3 | 3 | 4 | 5 | 2 | 0.05 | 0.40 | ARFA1B - ADP-ribosylation factor-like protein |
| 99 | At5g14920 | 5 | 2 | 4 | 5 | 4 | 2 | 0.22 | 0.32 | Gibberellin-regulated family protein |
| 100 | At5g15280 | 3 | 1 | 2 | 2 | 1 | 1 | 0.27 | 0.22 | Mitochondrial; Pentatricopeptide repeat (PPR) superfamily protein; |
| 101 | At5g15400 | 5 | 2 | 4 | 2 | 1 | 2 | 0.17 | 0.31 | Ubiquitin conjugation factor E4 |
| 102 | At5g15930 | 2 | 0 | 2 | 0 | 1 | 1 | 0.31 | 0.44 | PAM1 - Encodes a putative plant adhesion molecule |
| 103 | At5g16090 | 3 | 5 | 2 | 3 | 2 | 3 | 0.27 | 0.34 | RAD23 UV excision repair family protein |
| 104 | At5g16210 | 3 | 3 | 3 | 1 | 3 | 1 | 0.04 | 0.33 | MAP kinase 6 |
| 105 | At5g16260 | 4 | 3 | 1 | 2 | 1 | 2 | 0.38 | 0.11 | RNA binding protein ELF9 |
| 106 | At5g16400 | 2 | 3 | 1 | 2 | 1 | 3 | 0.30 | 0.34 | Thioredoxin F2, chloroplastic |
| 107 | At5g17300 | 5 | 3 | 4 | 1 | 2 | 2 | 0.24 | 0.50 | Homeodomain-like superfamily protein |
| 108 | At5g20010 | 5 | 5 | 3 | 4 | 2 | 3 | 0.40 | 0.59 | RAN-1 - GTP-binding nuclear protein Ran-1 |
| 109 | AT5g22340 | 2 | 2 | 1 | 2 | 1 | 2 | 0.39 | 0.47 | Ribosomal protein S10p/S20e family protein |
| 110 | At5g22640 | 4 | 2 | 2 | 1 | 1 | 2 | 0.30 | 0.46 | MORN (Membrane Occupation and Recognition Nexus) repeat-containing protein |
| 111 | At5g24314 | 5 | 2 | 2 | 1 | 1 | 1 | 0.28 | 0.75 | PTAC7 - Protein PLASTID TRANSCRIPTIONALLY ACTIVE 7 |
| 112 | At5g26360 | 5 | 5 | 3 | 3 | 1 | 3 | 0.20 | 0.55 | TCP-1/cpn60 chaperonin family protein; Molecular chaperone |
| 113 | At5g27740 | 6 | 5 | 4 | 2 | 4 | 2 | 0.26 | 0.55 | EMB2775 - ATPase family associated with various cellular activities (AAA) |
| 114 | At5g35360 | 5 | 4 | 3 | 4 | 2 | 1 | 0.25 | 0.31 | CAC2 - Acetyl Co-enzyme a carboxylase biotin carboxylase subunit |
| 115 | At5g37580 | 5 | 4 | 2 | 3 | 1 | 1 | 0.11 | 0.49 | Josephin family protein |
| 116 | At5g40650 | 5 | 6 | 7 | 5 | 3 | 2 | 0.14 | 0.55 | SDH2-2 - succinate dehydrogenase complex |
| 117 | At5g40670 | 7 | 3 | 2 | 2 | 2 | 2 | 0.32 | 0.57 | PQ-loop repeat family protein / transmembrane family protein |
| 118 | At5g43150 | 4 | 5 | 3 | 4 | 0 | 1 | 0.00 | 0.00 | Elongation factor |
| 119 | At5g45610 | 2 | 3 | 2 | 2 | 1 | 1 | 0.11 | 0.44 | SUV2 - protein dimerization |
| 120 | At5g47040 | 8 | 3 | 2 | 3 | 2 | 1 | 0.23 | 0.50 | lon protease 2 |
| 121 | At5g49510 | 4 | 5 | 4 | 3 | 2 | 2 | 0.20 | 0.37 | PFD3 - Probable prefoldin subunit 3 |
| 122 | At5g52920 | 4 | 2 | 5 | 0 | 2 | 2 | 0.24 | 0.34 | PKP-BETA1 - A dominant chloroplast pyruvate kinase beta subunit |
| 123 | At5g58730 | 3 | 3 | 2 | 1 | 1 | 2 | 0.31 | 0.49 | Mik - pfkB-like carbohydrate kinase family protein |
| 124 | At5g61500 | 4 | 5 | 3 | 2 | 2 | 3 | 0.15 | 0.17 | ATG3 - Autophagy-related protein 3 |
| 125 | At5g63520 | 4 | 4 | 3 | 3 | 3 | 2 | 0.09 | 0.23 | F-box/LRR-repeat protein |

Na: Number of alleles, PIC: Polymorphism information content, He: Expected Heterozygosity

**Supplementary Table S3.** Number of alleles found for each primer

| S.No. | Primer | In Silico | | | In PCR | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | B.rapa | B.nigra | B.oleracea | B. rapa | B. nigra | B. oleracea | B. juncea | B. napus | B. carinata |
| 1 | At1g02870 | 1 | 0 | 1 | 2 | 1 | 1 | 3 | 3 | 3 |
| 2 | At1g03910 | 1 | 0 | 1 | 3 | 2 | 2 | 3 | 1 | 2 |
| 3 | At1g10840 | 2 | 3 | 2 | 3 | 3 | 1 | 4 | 3 | 4 |
| 4 | At1g18340 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 |
| 5 | At1g19240 | 1 | 2 | 2 | 1 | 3 | 2 | 3 | 1 | 1 |
| 6 | At1g23440 | 2 | 3 | 3 | 3 | 3 | 4 | 4 | 2 | 3 |
| 7 | At1g30540 | 1 | 1 | 1 | 1 | 3 | 1 | 4 | 1 | 2 |
| 8 | At1g31812 | 2 | 1 | 2 | 3 | 1 | 2 | 4 | 3 | 3 |
| 9 | At1g34270 | 2 | 1 | 1 | 2 | 2 | 2 | 4 | 3 | 3 |
| 10 | At1g35680 | 1 | 0 | 2 | 2 | 1 | 1 | 4 | 3 | 2 |
| 11 | At1g48440 | 1 | 0 | 1 | 3 | 2 | 1 | 5 | 3 | 4 |
| 12 | At1g50240 | 1 | 1 | 2 | 2 | 2 | 2 | 5 | 3 | 3 |
| 13 | At1g57680 | 1 | 2 | 2 | 3 | 4 | 3 | 5 | 4 | 6 |
| 14 | At1g65440 | 0 | 0 | 1 | 1 | 3 | 2 | 3 | 2 | 3 |
| 15 | At1g65840 | 1 | 0 | 3 | 1 | 0 | 1 | 2 | 1 | 1 |
| 16 | At1g67060 | 1 | 1 | 1 | 1 | 3 | 1 | 3 | 1 | 2 |
| 17 | At1g67170 | 1 | 3 | 1 | 4 | 4 | 2 | 3 | 2 | 3 |
| 18 | At1g67250 | 1 | 0 | 0 | 2 | 4 | 2 | 6 | 3 | 6 |
| 19 | At1g68310 | 2 | 2 | 2 | 3 | 4 | 2 | 5 | 2 | 5 |
| 20 | At1g70350 | 1 | 0 | 2 | 1 | 0 | 2 | 2 | 0 | 1 |

**SupplementaryTable S4.** Genetic distances between different species of family Brassicacaeae in present investigation

| | PDZ-1 | RLC-3 | | PM-30 | P. KA-RISH-MA | BIO-YSR | RE-8 | P. | | PM-25 | PM-28 | VAR-UNA | P. JA-GANATH | NRCHB-101 | EC- | RC-275 | RAPA-1 | RAPA-2 | NIG-RA-1 | NIG-RA-2 | GSL-1 | GSL-5 | NPC-9 | IGC-01 | P. MEGH-NA | GOLD-EN ACRE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PDZ-1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | |
| RLC-3 | 0.708 | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| HEERA | 0.706 | 0.736 | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| PM-30 | 0.788 | 0.701 | 0.662 | 1 | | | | | | | | | | | | | | | | | | | | | | |
| P.KARISHMA | 0.690 | 0.729 | 0.698 | 0.712 | 1 | | | | | | | | | | | | | | | | | | | | | |
| BIO-YSR | 0.637 | 0.655 | 0.605 | 0.702 | 0.716 | 1 | | | | | | | | | | | | | | | | | | | | |
| RE-8 | 0.640 | 0.625 | 0.565 | 0.703 | 0.709 | 0.731 | 1 | | | | | | | | | | | | | | | | | | | |
| P.BOLD | 0.640 | 0.650 | 0.539 | 0.731 | 0.704 | 0.697 | 0.799 | 1 | | | | | | | | | | | | | | | | | | |
| RH749 | 0.650 | 0.671 | 0.565 | 0.754 | 0.737 | 0.705 | 0.807 | 0.879 | 1 | | | | | | | | | | | | | | | | | |
| PM-25 | 0.663 | 0.680 | 0.572 | 0.753 | 0.732 | 0.711 | 0.819 | 0.887 | 0.890 | 1 | | | | | | | | | | | | | | | | |
| PM-28 | 0.636 | 0.667 | 0.557 | 0.730 | 0.740 | 0.720 | 0.810 | 0.847 | 0.877 | 0.903 | 1 | | | | | | | | | | | | | | | |
| VARUNA | 0.610 | 0.609 | 0.524 | 0.683 | 0.671 | 0.718 | 0.748 | 0.778 | 0.785 | 0.823 | 0.828 | 1 | | | | | | | | | | | | | | |
| P.JAGANATH | 0.638 | 0.658 | 0.561 | 0.729 | 0.743 | 0.703 | 0.793 | 0.852 | 0.896 | 0.877 | 0.896 | 0.798 | 1 | | | | | | | | | | | | | |
| NRCHB-101 | 0.645 | 0.658 | 0.545 | 0.745 | 0.724 | 0.719 | 0.793 | 0.835 | 0.860 | 0.855 | 0.882 | 0.806 | 0.865 | 1 | | | | | | | | | | | | |
| EC-766602 | 0.587 | 0.590 | 0.524 | 0.653 | 0.649 | 0.624 | 0.693 | 0.717 | 0.732 | 0.739 | 0.724 | 0.680 | 0.723 | 0.723 | 1 | | | | | | | | | | | |
| RC-275 | 0.570 | 0.616 | 0.567 | 0.617 | 0.631 | 0.616 | 0.623 | 0.648 | 0.658 | 0.660 | 0.665 | 0.593 | 0.663 | 0.674 | 0.734 | 1 | | | | | | | | | | |
| RAPA-1 | 0.347 | 0.348 | 0.329 | 0.361 | 0.351 | 0.345 | 0.339 | 0.360 | 0.362 | 0.368 | 0.360 | 0.343 | 0.364 | 0.361 | 0.394 | 0.408 | 1 | | | | | | | | | |
| RAPA-2 | 0.361 | 0.367 | 0.363 | 0.374 | 0.367 | 0.360 | 0.365 | 0.375 | 0.372 | 0.388 | 0.368 | 0.357 | 0.374 | 0.366 | 0.402 | 0.428 | 0.559 | 1 | | | | | | | | |
| NIGRA-1 | 0.345 | 0.329 | 0.335 | 0.328 | 0.363 | 0.337 | 0.351 | 0.351 | 0.362 | 0.359 | 0.360 | 0.335 | 0.353 | 0.367 | 0.376 | 0.349 | 0.172 | 0.173 | 1 | | | | | | | |
| NIGRA-2 | 0.338 | 0.327 | 0.331 | 0.329 | 0.342 | 0.333 | 0.341 | 0.359 | 0.358 | 0.352 | 0.348 | 0.330 | 0.355 | 0.358 | 0.358 | 0.348 | 0.162 | 0.165 | 0.687 | 1 | | | | | | |
| GSL-1 | 0.356 | 0.364 | 0.358 | 0.370 | 0.357 | 0.314 | 0.354 | 0.357 | 0.358 | 0.369 | 0.367 | 0.330 | 0.358 | 0.348 | 0.351 | 0.352 | 0.345 | 0.398 | 0.189 | 0.179 | 1 | | | | | |
| GSL-5 | 0.377 | 0.385 | 0.379 | 0.384 | 0.385 | 0.333 | 0.373 | 0.378 | 0.377 | 0.392 | 0.381 | 0.349 | 0.382 | 0.359 | 0.382 | 0.376 | 0.368 | 0.407 | 0.201 | 0.186 | 0.804 | 1 | | | | |
| NPC-9 | 0.403 | 0.427 | 0.421 | 0.412 | 0.425 | 0.380 | 0.405 | 0.424 | 0.410 | 0.442 | 0.432 | 0.393 | 0.423 | 0.421 | 0.442 | 0.441 | 0.259 | 0.268 | 0.384 | 0.380 | 0.383 | 0.426 | 1 | | | |
| IGC-01 | 0.398 | 0.429 | 0.421 | 0.409 | 0.428 | 0.388 | 0.410 | 0.429 | 0.415 | 0.447 | 0.443 | 0.393 | 0.429 | 0.423 | 0.439 | 0.438 | 0.262 | 0.271 | 0.378 | 0.371 | 0.383 | 0.417 | 0.969 | 1 | | |
| P.MEGHNA | 0.203 | 0.212 | 0.180 | 0.198 | 0.198 | 0.167 | 0.194 | 0.212 | 0.205 | 0.209 | 0.208 | 0.192 | 0.199 | 0.204 | 0.204 | 0.183 | 0.193 | 0.208 | 0.175 | 0.160 | 0.234 | 0.234 | 0.215 | 0.218 | 1 | |
| GOLDENA-CRE | 0.197 | 0.211 | 0.181 | 0.202 | 0.204 | 0.168 | 0.188 | 0.203 | 0.194 | 0.205 | 0.207 | 0.194 | 0.201 | 0.201 | 0.193 | 0.187 | 0.179 | 0.203 | 0.155 | 0.141 | 0.242 | 0.244 | 0.207 | 0.207 | 0.685 | 1 |

**Supplementary Table S5.** Per cent cross transferability and total number of alleles found in 26 Brassica genotypes for all the amplified primer sets

| Genotypes | Total no. alleles | Number of amplified primers | Per cent cross-transferability | AVG/marker |
|---|---|---|---|---|
| PDZM-1 | 382 | 123 | 98.40 | 3.06 |
| RLC-3 | 385 | 124 | 99.20 | 3.08 |
| HEERA | 379 | 124 | 99.20 | 3.03 |
| PM-30 | 394 | 123 | 98.40 | 3.15 |
| P.KARISHMA | 407 | 124 | 99.20 | 3.26 |
| BIO-YSR | 353 | 118 | 94.40 | 2.82 |
| RE-8 | 369 | 122 | 97.60 | 2.95 |
| P.BOLD | 392 | 123 | 98.40 | 3.14 |
| RH749 | 397 | 125 | 100.00 | 3.18 |
| PM-25 | 393 | 124 | 99.20 | 3.14 |
| PM-28 | 395 | 125 | 100.00 | 3.16 |
| VARUNA | 360 | 123 | 98.40 | 2.88 |
| P.JAGANATH | 386 | 125 | 100.00 | 3.09 |
| NRCHB-101 | 386 | 123 | 98.40 | 3.09 |
| EC-766602 | 386 | 120 | 96.00 | 3.09 |
| RC-275 | 384 | 123 | 98.40 | 3.07 |
| RAPA-1 | 247 | 119 | 95.20 | 1.98 |
| RAPA-2 | 297 | 113 | 90.40 | 2.38 |
| NIGRA-1 | 258 | 117 | 93.60 | 2.06 |
| NIGRA-2 | 248 | 117 | 93.60 | 1.98 |
| GSL-1 | 346 | 120 | 96.00 | 2.77 |
| GSL-5 | 363 | 122 | 97.60 | 2.90 |
| NPC-9 | 384 | 122 | 97.60 | 3.07 |
| IGC-01 | 384 | 122 | 97.60 | 3.07 |
| P.MEGHNA | 186 | 121 | 96.80 | 1.49 |
| GOLDEN ACRE | 188 | 123 | 98.40 | 1.50 |

**Supplementary Table S6.** The Evanno table output after running STRUCTURE HRVESTER

| K | Reps | Mean LnP(K) | StdevLnP(K) | Ln′(K) | \|Ln″(K)\| | Delta K |
|---|---|---|---|---|---|---|
| 1 | 3 | -1705.87 | 2.909181 | — | — | — |
| 2 | 3 | -1574.13 | 0.814453 | 131.7333 | 134.8333 | 165.5508 |
| 3 | 3 | -1577.23 | 12.22307 | -3.1 | 12.5 | 1.022657 |
| 4 | 3 | -1592.83 | 9.096336 | -15.6 | 13.26667 | 1.458463 |
| 5 | 3 | -1595.17 | 22.52872 | -2.33333 | 48.1 | 2.135052 |
| 6 | 3 | -1645.6 | 12.5048 | -50.4333 | 65.56667 | 5.24332 |
| 7 | 3 | -1630.47 | 28.90023 | 15.13333 | 25.36667 | 0.877732 |
| 8 | 3 | -1640.7 | 16.88579 | -10.2333 | 24.7 | 1.462768 |
| 9 | 3 | -1626.23 | 21.92951 | 14.46667 | 38.03333 | 1.734345 |
| 10 | 3 | -1649.8 | 80.12871 | -23.5667 | — | — |