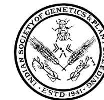
Indian J. Genet. Plant Breed., (2023); 83(2): 217-223
https://doi.org/10.31742/ISGPB.83.2.7
ISSN: 0975-6906



www.isgpb.org

RESEARCH ARTICLE

Whole genome resequencing of advanced introgression lines of *Brassica juncea* L. Czern. for characterizing alien introgression from *Diplotaxis erucoides* L. DC

Sharani Choudhury, Purva Asrani[$], Anamika Kashyap, Mahesh Rao, Lakshman Prasad[1], Usha Pant[2], Ashish K. Gupta and Ramcharan Bhattacharya[*]

Abstract

The productivity of Indian mustard *Brassica juncea* L., a major oilseed crop in India, is heavily inflicted due to the fungal disease *Alternaria* leaf spot caused by *Alternaria brassicae* L. Genetic resistance against this disease is not available in the mustard cultivars. Screening for resistance among the cruciferous wild relatives identified *Diplotaxis erucoides* L. as a resistant species that can be used for sourcing resistance genes against *Alternaria* disease. The introgression lines developed by wide hybridization between *D. erucoides* and *B. juncea* have been advanced to BC_2F_7 generation through a recurrent selection of resistant plants at three locations. Development of molecular markers linked to alien introgression and the trait is highly desirable for tracking the alien introgression and early selection of the desired plants. A study was conducted to develop STS markers associated with the introgressed alien DNA based on resequencing of a resistant and a susceptible introgression line. A step-wise bioinformatic analysis and identification of the species-specific DNA contigs followed by annotation led us to predict several putative STS markers. While the sequence data generated from more resistant plants and further analysis are in the process, initial wet lab validation endorsed the appropriateness of the strategy. The identified markers that showed species specificity need to be studied over the generations to establish their linkage with the trait of *Alternaria* resistance.

Keywords: *Alternaria* leaf spot, CWR, Indian mustard, introgression line, MAS

Introduction

Next to cereals, the second most important determinant of the agricultural economy in India is the oilseed crops. The rapeseed-mustard group of oilseed crop *Brassica* spp. represents a large share of the oilseed economy not only in India but also across the world. It covers 2.4 m ha area in the world (FAOSTAT 2020; https://www.fao.org/statistics/en/) and occupies 3rd position among the oilseed crops as per world statistics (Cartea et al. 2019). India occupies 3rd position in rapeseed-mustard production next to Canada and China. While India has attained self-sufficiency in the majority of agricultural produce, yet it depends on imports for more than 50% of its domestic demand for edible oil (Jat et al. 2019). Thus, productivity enhancement in rapeseed mustard has been the priority area in agricultural research. Indian mustard [*Brassica juncea* (L.) Czern.] occupies the largest acreage among the *Brassica* spp. The productivity of Indian mustard is crippled due to various biotic stresses, including significant damage by *Alternaria* leaf blight disease caused by *Alternaria brassicae*. The extent of yield loss due to *Alternaria* blight varies between 10-70% (Kumar and Kolte 2012). Chemical control is the only prevalent practice for mitigating damage despite huge environmental

ICAR-National Institute for Plant Biotechnology, New Delhi 110 012, India

[1]ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India

[2]G. B. Pant University of Agriculture and Technology, Pantnagar, Udham Singh Nagar 263 145, Uttrakhand, India

[$]Ruhr University of Bochum, Bochum, Germany

***Corresponding Author:** Ramcharan Bhattacharya, ICAR-National Institute for Plant Biotechnology, IARI Campus, New Delhi 110 012, India, E-Mail: rcbhattacharya1@gmail.com

How to cite this article: Choudhury S., Asrani P., Kashyap A., Rao M., Prasad K., Pant U., Gupta A.K. and Bhattacharya R. 2023. Whole genome resequencing of advanced introgression lines of Brassica juncea L. Czern. for characterizing alien introgression from Diplotaxis erucoides L. DC. Indian J. Genet. Plant Breed., **83**(2): 217-223.

Source of support: Nil

Conflict of interest: None.

Received: Sept. 2023 **Revised:** March 2023 **Accepted:** April 2023

© The Author(s). 2023 Open Access This article is Published by the Indian Society of Genetics & Plant Breeding, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org

consequences caused by fungicides. Developing genetic resistance in the cultivated varieties has been impeded due to lack of resistance source within the crossable germplasms and the complexity of the trait.

Crop wild relatives (CWRs) are rich stock of resistant genes against many of diseases and pests (Bohra et al. 2021). The Brassicaceae family includes many wild species which possess useful agronomic traits such as pathogen resistance, insect resistance, CMS systems etc. (Saal et al. 2004; Warwick et al. 2009). Over the last decade several cases of successful introgression of the resistance genes from CWRs to cultivated germplasms have been documented (Hajjar and Hodgkin 2007). However, introgression of genes from CWRs to cultivated germplasms through wide hybridization has been a daunting task due to several factors, including lack of cross-compatibility, wide difference in ploidy level, loss of seed fertility and aneuploidy in the early back cross generations etc. Therefore, screening for the introgressed trait is not feasible in the early generations following wide hybridization. Secondly, for a polygenic trait the number of genes involved is large. Thus, screening in the segregating generations based on phenotyping does not necessarily identify plants with homogeneous constellations of introgressed genes. As a result, in spite of cumbersome phenotyping, genetic heterogeneity of the selected plants delay attaining stability of the introgressed trait. For overcoming such bottlenecks, developing molecular markers linked to the introgressed trait will be useful. Such molecular markers will not only hasten the recurrent selection process but will also lead the way towards identifying the introgressed genes and genomic segments responsible for the trait. Recently, there are many examples where a reverse approach based on whole genome sequencing has been used for identifying markers linked to the alien introgression and further testing them for their linkage with the trait (Vasupalli et al. 2017; Bimpong et al. 2011).

The cultivated Indian mustard, *B. juncea* suffers a substantial loss in crop productivity due to fungal pathogen *A. brassicae* damage. The Brassicaceae family hosts a large number of CWRs (Šamec and Salopek-Sondi 2019). A large throughput screening among the wild cruciferous species by Sharma et al. (2002) led to the identification of eight species, including *Diplotaxis erucoides* showing significant level of resistance against Alternaria disease. Subsequently, Bhat et al. (2006) developed an interspecific hybrid between *B. juncea* (4x, allopolyploid, AABB) and a diploid *D. erucoides* using *Brassica rapa* (2n) as the bridge species. After two successive backcrosses using *B. juncea* as the recurrent parent, the interspecific hybrids were advanced to BC_2F_7 through recurrent screening against *Alternaria* disease and isolation of the resistant plants. The objective of the present study was a whole genome resequencing-based approach for developing molecular markers linked to the alien introgression and the resistance trait.

Materials and methods

Plant material

BC_2F_7 plants of the introgression lines derived from the wide hybridization *B. juncea* x *D. erucoides* along with both the parents were grown in the experimental field of G.B. Pant University at Pantnagar, IARI Regional Station at Pusa, Bihar and ICAR-NIPB at New Delhi during the month of October-March, 2019-20. Earlier, wide hybridization was carried out by first crossing *D. erucoides* x *B. rapa* followed by crossing of the F_1s with *B. juncea.* Initially, F_1s were rescued by embryo rescue to overcome postzygotic incompatibility. Plant plants at Pantnagar and IARI regional station at Pusa were grown in natural field condition. At New Delhi the plants were grown in 10-inch earthen pots kept in a polyhouse with 20±2°C temperature, 85–90% RH, and natural photoperiod. At the plants' 3-4 leaf stage, thinning was carried, keeping only three healthy seedlings in each pot.

Screening for Alternaria resistance and generation advancement

Pantnagar (29.0229° N, 79.4879° E) being a hotspot for *Alternaria blight* disease in mustard, screening for disease resistance was carried out based on the natural infestation. The screening was carried out at New Delhi polyhouse facility under artificial epiphytotic conditions. The inoculum of *A. brassicae* was prepared using a virulent strain Ab1 (NCBI accession No. MN704652). The single spore culture method was performed to prepare the pure inoculum as Zhang et al. (2013) described. For this, Ab1 was cultured on potato dextrose agar (PDA) media and kept for incubation at 25 ± 2°C under 12 hours of alternating fluorescent light and dark period for 7 days. The spores were then collected by scraping the plates followed by suspending them in sterile distilled water. Any debris was removed from the spore culture by filtering through a muslin cloth. The stock solution was diluted with sterile double distilled water to set the spore concentration at 1×10^5 mL^{-1} by monitoring through a hemocytometer (Doullah et al. 2006) and used for the inoculation of the plants. For infecting the plants, the scratch method (Zhang et al. 2018) was used in which the leaf endodermis was gently scraped at the bottom side to enable the entry of *Alternaria* spores. Third, 4th and 5th leaves from the bottom were inoculated with one spot each side of the midrib. 20 μL spore solution was applied carefully upon each of the wounds. The disease severity was surveyed at the peak stage of the disease development at a scale ranging from 0 (No symptoms; Immune) to >50 (Higher symptoms; highly susceptible) as per the AICRP-RM scoring norm (Table 1). Disease severity in percent is represented in the form of the box plot.

Table 1. Scaling of disease severity based on its development as per AICRP RM

S.No.	Disease %	Disease reaction	Score
1	0	Immune	0
3	<5	Highly resistant	1–10
3	6–10	Resistant	11–25
4	11–25	Moderately resistant	26–50
5	26–50	Susceptible	51–75
6	>50	Highly susceptible	76–100

Whole Genome Resequencing and Data Analysis

NGS-based resequencing samples were collected from the plants grown at the polyhouse facility of NIPB, New Delhi. Based on the outcome of the disease scoring, two plants with contrasting phenotype, a moderately resistant plant designated as R and a susceptible plant as S were selected for the whole genome resequencing. Total DNA was isolated from the R and S lines following the CTAB method described by Doyle (1991). The plan of work for the whole genome resequencing and data analysis has been represented as a flowchart in Fig. 1. The sequence reads were developed through the Illumina HiSEQ X10 platform using paired-end chemistry. Quality check was performed using FastQC (version 0.10.1, www.bioinformatics.babraham.ac.uk/projects/fastqc/). All the developed reads were subjected to trimming to remove the low-quality reads and low-quality bases at the end by default parameters using trim galore (version 0.4.5). The quality reads were subjected to *de novo* assembly in CLCbio with default parameters and K-mer 30. Redundant reads were removed using cd-hit software (version 4.7, http://weizhongli-lab.org/cd-hit/). The identical contigs in S and R lines with 100% similarity were identified using local BLAST. These contigs were removed, and the remaining unique contigs were used to further analysis.

Gene annotation

Gene annotation for contigs was predicted using AUGUSTUS against *Arabidopsis thaliana* as a reference. Predicted genes were aligned on a complete NR database using BLASTx. Aligned genes were annotated using the UniProt database and KEGG database.

Validation of the primers for STS-marker development

Primers were designed for the selected contigs using IDT primer quest tool (https://www.idtdna.com/pages/tools/primerquest) and synthesized by GCC Biotech (GCC Biotech PVT. LTD., India). The PCR reactions were carried out using the genomic DNA under the following regime of thermal cycling: A. Initial denaturation at 98°C for 2 minutes; B. 40 cycles of denaturation at 96°C for 30 seconds, annealing at 55 and 50°C for 30 seconds and extension at 72°C for 1-minute; C. final extension at 72°C for 7 minutes. The components used for reaction mixture were: DNA 1-µL, forward and reverse primer 0.4 µL each, 2x EmeraldAmp® PCR master mix 5 µL, nuclease-free water 3.2 µL. PCR master mix was procured from Takara (Takara Bio Inc., Japan). PCR products were run in 1.2% agarose gel prepared in 1X TAE buffer. A housekeeping gene GAPDH (Glyceraldehyde 3-phosphate dehydrogenase) was used as the positive control for the PCR reaction. Generuler 1kb plus DNA ladder provided by Thermo Fisher Scientific (Thermo Fisher Scientific Corporation, US) has been used as DNA size markers.

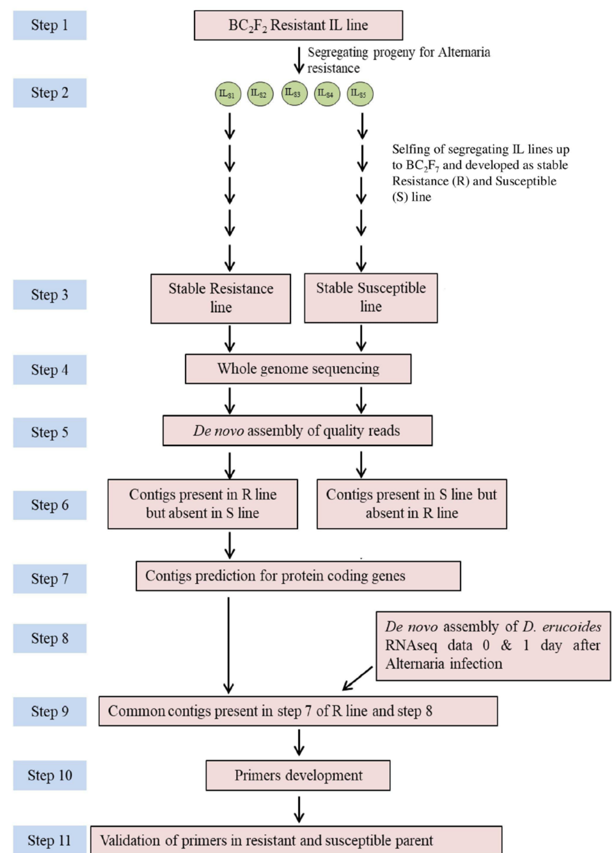


Fig. 1. Flowchart depicting the workflow of marker identification for *Alternaria* resistance

Results and discussion

Selection of the introgression lines for whole genome sequencing

The BC_2F_7 introgression lines were subjected to screening against *Alternaria brassicae* at three different locations GBPUAT, Pantnagar; IARI-RS, Bihar; and ICAR-NIPB Delhi, during the mustard season in 2019–20. Box plot depicting the percent disease severity at all the three locationsin 2019–20 has been shown in Fig. 2. In box plots, the percent disease severity of the IL lines varied across the three locations. It ranged from 5–75, 10–60 and 0–30% at Pantnagar, Delhi, and Bihar, respectively. The occurrence of the disease was

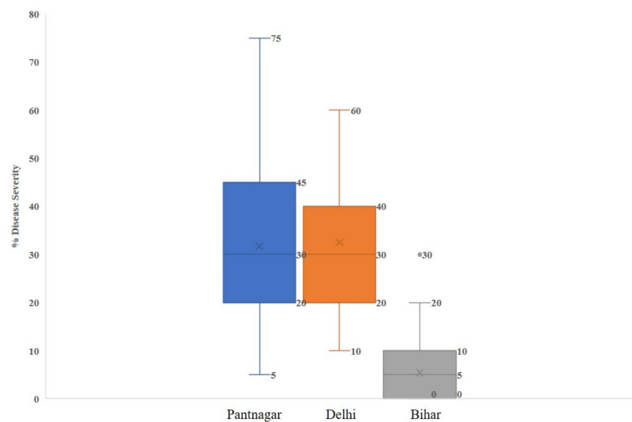


Fig. 2. Box plot distributions of per cent disease severity (2019-2020) of IL classified in 3 clusters according to the locations of data collection. In the box plots, median values are indicated as crosses within the boxes; the interquartile ranges (at 25 and 75%) are mentioned in the boxes; the whiskers (above or below the boxes) are mentioned with values beyond the interquartile ranges (<25% or >75%); minimum and maximum values are mentioned at the two ends

much more severe at Pantnagar compared to the other two locations. At Bihar, the disease severity was low due to low incidents of natural infestation. Therefore, the experimental plants for the whole genome resequencing were chosen from the Pantnagar field. Based on the result of phenotyping by disease scoring, a moderately resistant introgression line (IL41) with average score 20 and a susceptible introgression line (IL42) with average score 40 were selected for the genome sequencing. The two contrasting lines were siblings from a parent which was subjected to selfing for more than six generations. Thus, it is assumed that these progeny plants will be with mostly homozygous backgrounds except for the segregating disease-resistant genes and thus, comparable to near isogenic lines (NIL). NILs has been proven as a very useful genetic tool to associate genetic marker with the trait of interest. In case of quantitative resistance, the complex nature of genetic architecture can be effectively dissected using NILs (Kolkman et al. 2020). Recruiting NILs and ILs to explore the genetic regulator of other traits have also been reported in many examples (Lee et al. 2021; Xie et al. 2016).

Whole-genome resequencing of IL lines

The advent of NGS-based approaches has made the gene discovery effort much easier and more feasible. A low coverage whole genome sequencing of a resistant (R) and a susceptible (S) line was carried out on the HiSEQ X10 platform. A total of 4.9×10^8 and 4.5×10^8 reads were generated from the R and S lines, respectively, with an average read length of 150.6 bp (Supplementary Table S1). The quality trimming of these reads was done on CLCbio by using default parameters. After end trimming using CLCbio with default parameters a total of 4.9×10^8, and 4.5×10^8, quality reads were developed with an average length of 114 and 134.1 bp for R and S lines, respectively. The quality reads of both R and S lines were subjected to *de novo* assembly in CLCbio with default parameters and K-mer 30. A total of 670,851 and 343,363 contigs were developed for the R and S lines, respectively. The largest contig was 52,364 bp and 74,434 bp for the R and S line, respectively, and the smallest contig was 200 bp in both samples. The average and median contig lengths are 928.54 bp and 1736.14 bp and 361 bp and 504 bp for R and S lines, respectively. A total of 622,914,222 bp and 596,124,719 bp with 36.98 and 35.92% GC were assembled in R and S line, respectively (Supplementary Table S2).

Whole genome sequencing of plants with contrasting phenotypes has been deployed robustly for gene and marker identification for various traits (Wang et al. 2016). Mapping and identification of gene responsible for female-male sterility in *Brassica napus* (Teng et al. 2017), fresh seed dormancy in groundnut (Kumar et al. 2020), powdery mildew in cucumber (Xu et al. 2016), level of lycopene in watermelon (Lee et al. 2021) have reportedly sought WGS as the scientific approach. Thus, combining IL lines as tool and WGS as technique can give a great advantage towards defining the causal gene(s)/QTL(s) underlying a trait.

Identification of contigs specific to R and S lines

NCBI local blast was used to identify the common contigs in both R and S lines. The commonly found contigs were removed, and the rest of the contigs were used for further analysis. A total of 53,282 and 26,598 contigs were identified specific to R and S lines, respectively (Supplementary Table S3). The largest contig was 52,364 bp and 4,708 bp for the R and S line, respectively, and the smallest contig was 200 bp in case of both the samples. The average and median contig lengths are 340.06 bp and 284.89 bp and 251 bp and 265 bp for R and S lines, respectively. A total of 18,119,301 bp and 7,577,404 bp with 48.34 and 39.28 % GC were assembled in R and S line, respectively.

Genes identified in the contigs

A total of 268,156 and 196,536 genes were predicted in the assembled contigs of R and S line, respectively (Supplementary Table S4). In gene ontology study, out of the 268,156 predicted genes in the R line, 29,142 genes belonged to cellular components, 23,992 genes belonged to biological process, and 19,616 genes fell in the molecular function category (Fig. 3A). Similarly, out of the 196,536 predicted genes of S line, 25,276 were related to cellular component, 20288 genes belonged to the biological process, and 17,233 genes fell in the molecular function category (Fig. 3B). In the contigs specific to R lines a total of 1,424 genes were predicted whereas, 1,051 genes were predicted on the contigs specific to the S line (Supplementary Table S5). Functional categorization of the predicted genes in GO terms has been shown in Fig. 4.

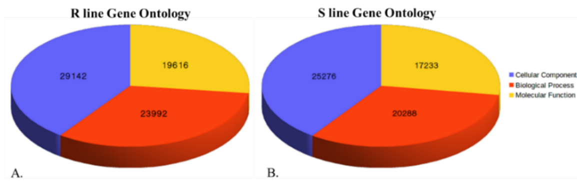


Fig. 3. Pie chart showing GO terms of total genes predicted in the contigs of the R-introgression and S-introgression lines. Share of genes belonging to cellular components, biological process and molecular function in R-IL (A) and S-IL (B)

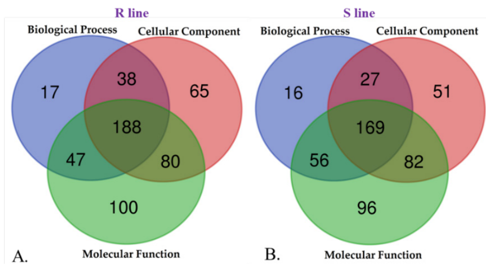


Fig. 4. Venn diagram representing the common and unique genes belonging to biological process, molecular function, and cellular components. GO classification of the annotated genes present only in resistant introgression R line (A) and susceptible introgression S line (B)

Identification of species specific contigs and primers

The 1424 predicted genes on the R line specific contigs were checked for their transcriptional activation in response to treatment by Alternaria infection in RNA-Seq data. The RNA-Seq data of a resistant and susceptible plant under healthy and Alternaria infection were already available in the laboratory (unpublished). In the transcriptome data, the differentially expressed genes (DEG) which mapped to *D. erucoides* origin in *de novo* assembly were compared with the predicted genes of R line specific contigs. A total of 60 genes were found common between these two data sets. This implies that these 60 genes are not only introgressed from *D. erucoides* (resistant parent) but are transcriptionally affected upon pathogen challenge. The involvement of many genes in the contigs of the resistant IL line indicated the use of complicated yet sophisticated and properly orchestrated machinery by the plant to counter the pathogen attack as mentioned by Andersen et al. (2018). By using IDT-Primer quest, primers were designed for selected genes among these contigs (Supplementary Table S6).

Validation of the predicted primers for polymorphic amplification

Primers were designed initially for a set of genes selected for PCR-based validation of their differential amplification in the two parents, *B. juncea* and *D. erucoides*. All the primers were subjected to PCR amplification at an annealing temperature 50° C using the DNA isolated from the parents.

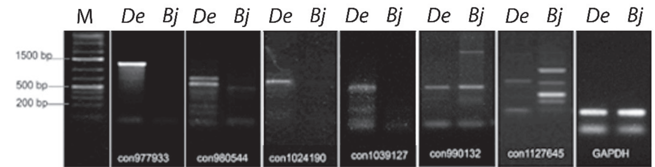


Fig. 5. Species-specific amplification of the selected primers in PCR based validation. PCR reaction was carried out at 50°C annealing temperature in *D. erucoides* (De) and *B. juncea* (Bj). PCR products were run in 1.2% agarose gel

The amplification pattern was analyzed in 1.2% agarose gel. Many of the selected primers showed amplification only in case of donor species *D. erucoides* and did not generate any amplification in case of *B. juncea* (Table 2 and Fig. 5). For some of the primers though amplification was found in both the parents, the pattern or sizes of the amplicons were polymorphic between the parents. The initial validation empirically endorsed the successful identification of a few putative STS markers associated with the donor DNA based on the low coverage WGS and the proposed pipeline of the bioinformatics analysis. However, more extensive prediction of flanking primers in and around the introgressed contigs and their large-scale validation for the development of STS-markers will be essential for further identification of the trait-linked markers.

Genes and pathways in the introgressed contigs

Gene annotations on the 60 contigs, which are specific to the resistant IL, predicted a large number of genes on the contigs. These predicted genes were further examined for their differential gene-expression in transcriptome data. The identified differentially expressed genes were found to encode diverse functions and pathways. The annotation data showed that some genes have seemed to be directly related to plant resistance while others are associated with the supportive role. Some of the prominent genes and regulatory sequences identified in the study included two of the receptor kinases (contig_942967, contig_978154), two Ca^{2+} signaling genes (contig_765179, contig_887631) and two stress-responsive promoters, namely, a NAC promoter (contig_1072352) and Rab16C promoter (contig_940851). Both promoters are known to be involved in plant stress signaling. In addition, several genes involved in oxidation-reduction process, secondary metabolite production, were also identified which were intricately associated to stress management mechanisms. Three of the genes were identified encoding chaperon protein and endopeptidases that provided the signature of cell homeostasis, taking place in the resistant plant for better survival. Not only stress-responsive genes, but 3 genes related to translation and part of the large ribosomal subunit have also been identified in the data. Among the genes already validated for their resistant parent-specific introgression and involvement in pathogen response, contig_980544 encoded for a lysosomal membrane-bound protein which plays

Table 2. PCR-based validation of the predicted primers for species-specific amplification at 50°C annealing temperature

Contig ID	Amplification in B. juncea	Amplification in D. erucoides	Polymorphism in amplification	Remarks
Contig 977933	√	-	Yes	Presence-absence polymorphism
Contig 980544	√	√	Yes	Length polymorphism
Contig 1024190	√	-	Yes	Presence-absence polymorphism
Contig 1039127	√	-	Yes	Presence-absence polymorphism
Contig 990132	√	√	Yes	Length polymorphism
Contig 1127645	√	√	Yes	RAPD-like polymorphism

role in protein localization and autophagy. The crucial role of autophagy for plant survival under biotic stress is well known (Sharma et al. 2012). Autophagy helps the host cell's survival by destroying damaged self-organelles (Signorelli et al. 2020). Lenz et al. (2011) has shown direct relation between disabled autophagy and compromised resistance against necrotrophic pathogen in Arabidopsis. Contig_1024190 has been found to encode an amino acid transporter. Amino acid transporter in plants contributes to reorganizing the amino acid pool in response to pathogen attack in favor of the survival of the host (Tünnermann et al. 2021). It has been reported that the *Rhg1* locus of soybean which imparts resistance against *Heterodera glycine* or cyst nematode of soybean, houses 3 genes, including one for an amino acid transporter Glyma18g02580 (Cook et al. 2012). This aspect has also been supported by Besnard et al. (2021) who demonstrated higher activation of an amino acid transporter contributing to an augmented level of SA-mediated resistance response in Arabidopsis against pathogen attack.

This piece of work demonstrates an attempt to develop molecular markers based on the introgressed contigs from the donor parent. Development of such markers validated for their linkage with the alien introgression and the trait will pave the way for efficient monitoring of trait introgression from a CWR. In conclusion, this study reveals the potential Alternaria resistance-related marker specific to resistant host *Diplotaxis erucoides*. These markers can be used in tracking the alien introgression after more extensive validation across the IL lines. Further, this strategy can be used to simultaneously map and transfer other traits in any other crop species of interest.

Author's contribution

Conceptualization of research (RB, MR, PS, AK); Designing of the experiments (RB, MR); Contribution of experimental materials (MR, RB, UP); Execution of field/lab experiments and data collection (SC, AK, PA); Analysis of data and interpretation (LP, AG); Preparation of the manuscript (SC, RB).

Acknowledgments

The authors extend sincere thanks to Dr. Naresh Vasupalli for his continuous support in execution of lab and field work as well as in manuscript preparation. We are also grateful to P.G. School, ICAR-IARI, DBT for the doctoral fellowship to SC, DBT project BT/IN/UK/PORI/ 03/AKP/18-19 for the financial support. Partial support from SERB CRG EMR/2017/003463 is also acknowledged. The authors are also thankful to NAHEP-CAAST project at ICAR-IARI for an overseas training fellowship to SC.

References

Andersen E. J., Ali S., Byamukama E., Yen Y. and Nepal M. P. 2018. Disease resistance mechanisms in plants. Genes, **9**: 339.

Besnard J., Sonawala U., Maharjan B., Collakova E., Finlayson S. A., Pilot G., Mcdowell J. and Okumoto S. 2021. Increased expression of UMAMIT amino acid transporters results in activation of salicylic acid dependent stress response. Front. Plant Sci., **11**: 606386.

Bhat S. R., Vijayan P., Ashutosh, Dwivedi K. K. and Prakash S. 2006. *Diplotaxis erucoides*-induced cytoplasmic male sterility in *Brassica juncea* is rescued by the *Moricandia arvensis* restorer: genetic and molecular analyses. Plant Breed., **125**: 150-155. doi: 10.1111/j.1439-0523.2006. 01184.x

Bimpong I. K., Serraj R., Chin J. H., Ramos J., Mendoza E. M., Hernandez J. E., Mendioro M. S. and Brar D. S. 2011. Identification of QTLs for drought-related traits in alien introgression lines derived from crosses of rice (*Oryza sativa* cv. IR64) × *O. glaberrima* under lowland moisture stress. J. Plant Biol., **54**: 237-250.

Bohra A., Kilian B., Sivasankar S., Caccamo M., Mba C., McCouch S. R. and Varshney R. K. 2021. Reap the crop wild relatives for breeding future crops. Trends Biotechnol., **40**: 412-431.doi: 10.1016/j.tibtech.2021.08.009.

Cartea E., De Haro-Bailón A., Padilla G., Obregón-Cano S., Del Rio-Celestino M. and Ordás A. 2019. Seed oil quality of *Brassica napus* and *Brassica rapa* germplasm from north-western Spain. Foods (Basel, Switzerland), **8**: 292. doi: 10.3390/foods8080292

Cook D. E., Lee T. G., Guo X., Melito S., Wang K., Bayless A. M., Wang J., Hughes T. J., Willis D. K., Clemente T. E., Diers B. W., Jiang J., Hudson M. E. and Bent A. F. 2012. Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science, **338**: 1206-1209.

Doullah M. A. U., Meah M. B. and Okazaki K. 2006. Development of an effective screening method for partial resistance to *Alternaria brassicicola* (dark leaf spot) in *Brassica rapa*.Eur. J. Plant Pathol., **116**: 33–43. doi: 10.1007/s10658-006-9035-2

Doyle J. 1991. DNA protocols for plants. In Molecular techniques in taxonomy, Springer, Berlin, Heidelberg: 283-293.

Hajjar R. and Hodgkin T. 2007. The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica, **156**: 1-13.doi: 10.1007/s10681-007-9363-0

Jat R. S., Singh V. V., Sharma P. and Rai P. K. 2019. Oilseed Brassica in India: Demand, supply, policy perspective and future potential. OCL, **26**: 8.

Kolkman J. M., Strable J., Harline K., Kroon D. E., Wiesner-Hanks T., Bradbury P. J. and Nelson R. J. 2020. Maize introgression library provides evidence for the involvement of liguleless1 in resistance to northern leaf blight. G3: Genes Genom. Genet., **10**: 3611-3622.

Kumar B. and Kolte S. J. 2012. Progression of Alternaria blight of mustard in relation to components of resistance. Indian Phytopathol., **54**: 329-331

Kumar R., Janila P., Vishwakarma M. K., Khan A. W., Manohar S. S., Gangurde S. S., Variath M. T., Shasidhar Y., Pandey M. K. and Varshney R. K. 2020. Whole-genome resequencing-based QTL-seq identified candidate genes and molecular markers for fresh seed dormancy in groundnut. Plant Biotechnol. J., **18**: 992-1003.

Lee S., Park G., Choi Y., Park S., Kim H., Lee O., Kim T. and Park Y. 2021. Whole-genome resequencing of near-isogenic lines reveals a genomic region associated with high trans-lycopene contents in watermelon. Plants, **11**: 8.

Lenz H. D., Vierstra R. D., Nürnberger T. and Gust A. A. 2011. ATG7 contributes to plant basal immunity towards fungal infection. Plant signal. behave., **6**: 1040-1042.

Saal B., Brun H., Glais I. and Struss D. 2004. Identification of a *Brassica juncea*-derived recessive gene conferring resistance to *Leptosphaeria maculans* in oilseed rape. Plant Breed., **123**: 505-511.

Šamec D., andSalopek-Sondi B. 2019. Cruciferous (brassicaceae) vegetables. In Nonvitamin and nonmineral nutritional supplements: 195-202. Academic Press

Sharma G., Dinesh Kumar V., Haque A., BhatS. R., Prakash S. and Chopra V. L. 2002. Brassica coenospecies: a rich reservoir for genetic resistance to leaf spot caused by *Alternaria brassicae*. Euphytica, **125**: 411-417. doi: 10.1023/A:1016050631673

Sharma P., Jha A. B., Dubey R. S. andPessarakli M. 2012. Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. J. Bot., **2012**: 1-26.

Signorelli S., Tarkowski Ł. P., Van den Ende W. and Bassham D. C. 2019.Linking autophagy to abiotic and biotic stress responses. Trends Plant Sci., **24**: 413-430.

Teng C., Du D., Xiao L., Yu Q., Shang G. and Zhao Z. 2017. Mapping and identifying a candidate gene (Bnmfs) for female-male sterility through whole-genome resequencing and RNA-Seq in rapeseed (*Brassica napus* L.). Front. Plant Sci., **8**: 2086.

Tünnermann L., Colou J., Näsholm T. and Gratz R. 2022. To have or not to have: expression of amino acid transporters during pathogen infection. Plant mol. biol.,**109**: 413-425.

Vasupalli N., Rao M., Chamola R., Pant U., Bhattacharya R. and Bhat S. R. 2017. Development and validation of donor-specific STS markers for tracking alien introgressions into *Brassica juncea* (L.) Czern. Mol. Breed., **37**: 1-10

Wang Y., Xiao L., Guo S., An F. and Du D. 2016. Fine mapping and whole-genome resequencing identify the seed coat color gene in *Brassica rapa*. PLoS One, **11**: e0166464.

Warwick S. I., Francis A. and Gugel R. K. 2009. Guide to wild germplasm of *Brassica* and allied crops (tribe Brassiceae, Brassicaceae). Canada: Agriculture and Agri-Food Canada, 1.

Xie L., Li F., Zhang S., Zhang H., Qian W., Li P., Zhang S. and Sun R. 2016. Mining for candidate genes in an introgression line by using RNA Sequencing: the anthocyanin overaccumulation phenotype in Brassica. Front. Plant Sci., **7**: 1245.

Xu Q., Shi Y., Yu T., Xu X., Yan Y., Qi X. and Chen X. 2016. Whole-genome resequencing of a cucumber chromosome segment substitution line and its recurrent parent to identify candidate genes governing powdery mildew resistance. PLoS One, **11**: e0164469.

Zhang K., Yuan-Ying S. and Cai L. 2013. An optimized protocol of single spore isolation for fungi. Cryptogam. Mycol., **34**: 349-356. doi: 10.7872/crym.v34.iss4.2013.349

Zhang M., Sun X., Cui L., Yin Y., Zhao X., Pan S. and WangW. 2018. The plant infection test: spray and wound-mediated inoculation with the plant pathogen *Magnaporthe grisea*. J. Vis. Exp., **138**: e57675.doi: 10.3791/57675.

Supplementary Table S1. Statistics of the NGS reads and trimmed quality IL line

S. No.	Sequenced IL lines	Total No. of reads	Avg. length	Number of reads after trim	Avg. length after trim
1	R line	490,302,012	150.6	490,297,347	114
2	S line	454,777,982	150.6	454,773,567	134.1

Supplementary Table S2. Details of de novo assembly of the reads derived from sequencing of R and S line

S.No	Feature	R line assembly	S line assembly
1	Total contigs	670,851	343,363
2	Total assembled bases	622,914,222	596,124,719
3	Largest contig size (bp)	52,364	74,434
4	smallest contig size (bp)	200	200
5	N10	9,334	19,784
6	N50	2,197	5,500
7	Percent GC	36.98	35.92
8	Median contig length (bp)	361	504
9	Average contig (bp)	928.54	1736.14

Supplementary Table S3. Details of contigs specific to R and S line

S.No	Feature	Contigs present in R line but absent in S line	Contigs present in S line but absent in R line
1	Total contigs	53,282	26598
2	Total assembled bases	18,119,301	7,577,404
3	Largest contig size (bp)	52,364	4,708
4	smallest contig size (bp)	200	200
5	N10	1,161	413
6	N50	311	281
7	Percent GC	48.34	39.28
8	Median contig length (bp)	251	265
9	Average contig (bp)	340.06	284.89

Supplementary Table S4. Statistics of gene annotation in the contigs of R and S line

S.No	Feature	R line annotation	S line annotation
1	Total genes	268,156	196,536
2	Total assembled bases	280,481,619	293,473,723
3	Largest contig size (bp)	17,839	22,398
4	smallest contig size (bp)	41	61
5	N10	4,263	5,409
6	N50	1,659	2,290
7	Percent GC	42.2	40.96
8	Median contig length (bp)	671	1,041
9	Average contig (bp)	1,045.96	1,493.23

Supplementary Table S5. Details of gene annotation in the R and S line specific contigs

S.No	Feature	Genes present in R line but absent in S line	Genes present in S line but absent in R line
1	Total genes	1,424	1,051
2	Total assembled bases	557,046	315,151
3	Largest contig size (bp)	2,721	4,651
4	smallest contig size (bp)	131	88
5	N10	1,529	543
6	N50	438	293
7	Percent GC	52.12	54.89
8	Median contig length (bp)	263	268
9	Average contig (bp)	391.18	299.86

Supplementary Table S6. Primer sequences for 6 contigs showing initial polymorphism in parents

Primer name	Primer sequence
contig_977933 F	GCCAAGTCGATCCTGTTCTT
contig_977933 R	AGCCGAACTTGACGTTGT
contig_980544 F	CGCGTTCAGGACGAAGAA
contig_980544 R	CTGGTGTGGGCGTTCAT
contig_990132 F	GACCGCGTCAACAACAAAG
contig_990132 R	CGTTGCCCTTCCGATCTC
contig_1024190 F	CTCCAGCAACCCTTACATGAT
contig_1024190 R	TGCATCTCGAGCAGCAAA
contig_1039127 F	CTGAGGTGTTCCCGGATGAAG
contig_1039127 R	ATCGCCGGCAAGAGGTG
contig_1127645 F	CTCTAACCTAAACCCACACCTC
contig_1127645 R	GATGAGCACCTTGGTGTACT