Deconstructing molecular phylogenetic relationship among cultivated and wild Brassica species

Jebi Sudan · Ravinder Singh · Reetika Mahajan · Romesh K. Salgotra

Received: 24 January 2021 / Accepted: 29 April 2021 / Published online: 28 May 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract Brassica represents an agriculturally important and diverse group of oilseed crops with a long evolutionary history. Molecular markers played an important role in understanding origin and evolution of Brassica species. In the present research, both Single Nucleotide Polymorphisms (SNPs) and Simple Sequence Repeats (SSRs) developed from Brassica juncea L. Czern. were used to find out the phylogenetic relationship between different cultivated and wild Brassica species. A total of 88 SSR and 58 SNP markers were found to be functional across 38 genotypes belonging to ten different taxon groups. The polymorphic markers were able to group the genotypes into three different clusters and showed relatedness among different genomes based on genetic distances. The transferability of these markers serve the purpose of their quick use in cultivar identification, diversity and phylogenetic analysis in those orphan crops species where no or less genomic information is available.

Keywords Brassicaceae · ddRADseq · Molecular markers · Diversity pattern · Population structure · Evolutionary relationship

Introduction

Brassicaceae is a highly diverse family that comprises of agriculturally important species including model plant (Arabidopsis thaliana L. Heynh.), highly cultivated oilseed crops (Brassica napus L., Brassica rapa L., Brassica carinata A.Braun and Brassica juncea L. Czern.), widely utilized vegetable species (Brassica oleracea L.), eatable roots (Raphanus sativus L.) and various herbs (Brassica nigra L. K.Koch and Sinapis alba L.). This broad array of diversity is mainly because of genetic manipulations caused due to natural evolution of polyploids and amphidiploid hybrids, and traditional breeding including selection followed over the time (Edward et al. 2011). The genus Brassica has a long evolutionary history consisting of amphidiploid hybrids that are formed...
by the interspecific crosses between diploid progenitors.

The Brassicas had long suffered a narrow diversity bottleneck, and cross species incompatibility have led to limited use of related species for the improvement of Brassicas. Most of the interspecies and intergeneric hybrids between Brassica and its related species were carried out to transfer important traits from related wild species to some cultivated Brassica species. Among the Brassica spp., B. juncea, also known as Indian mustard, is a dominant species grown on Indian subcontinent for seed oil. It is an amphidiploid (2n = 4x = 36; AABB), thought to have developed through hybridization of its progenitors followed by the natural doubling of chromosomes (Nagaharu 1935). Among the three genomes (A, B and C), the AA genome has most variation in chromosome size and morphology and therefore most of the variation in B. juncea can be attributed due to AA genome (Kulak et al. 2002). The diversity present in Brassica spp. governs its usability for breeding purpose and further improvement of the cultivated crop. A plethora of molecular markers had been reported for the analysis of molecular diversity and population genetics studies. Among different marker systems used, SSRs (Simple Sequence Repeats) and SNPs (Single Nucleotide Polymorphisms) are more promising in relating the patterns of closeness in different Brassica species (Sudan et al. 2016; 2019; Li et al. 2019). As such, huge emphasis has been laid on the development of genic SSR and SNP markers since they find primary application in high-resolution genetic map construction, association mapping, genetic diagnostics, genetic diversity analysis, phylogenetic analysis, and characterization of genetic resources (Singh et al. 2018). A number of studies had been reported in which SSR markers have been developed in some Brassica spp. and their cross transferability had been explored in other Brassica spp. (Singh et al. 2016; Thakur et al. 2018; Li et al. 2019). However, no attempts had been made to explore the efficiency of cross species transferability of SNP as well as SSR markers from B. juncea to wild Brassica spp. The transferability of these markers serve the purpose of their quick use in cultivar identification, diversity and phylogenetic analysis in those orphan crops species where no or less genomic information is available.

In the present study, we used both SNP and SSR markers developed from Brassica juncea and evaluate them for population, genetic diversity and evolutionary relationship studies in wild Brassica spp.

Materials and methods

DNA extraction and dd-RAD library preparation

Young leaves from Brassica genotypes and wild Brassica relatives were used for the isolation of high quality genomic DNA using SGS buffer method (Sudan et al. 2017). A total of six B. juncea genotypes (Pusa tarak, Urvashi, RSIR 1, Zem 1, Donskaja 4 and EC 287,711) were used for dd-RAD library preparation using a modified protocol (Yang et al. 2016). The pooled dd-RAD library was sequenced on high throughput Illumina HiSeq 2000 generating 100 bp paired-end reads. The ddRAD-seq reads were subjected to a pipeline using CLC Genomics workbench to obtain a set of high quality SNPs from genic regions. A subset of SNP markers from B. juncea was used to check the efficiency of cross-species transferability of markers in 32 Brassica relatives (Table 1). Out of 32 genotypes, 14 belong to S. alba, five from B. rapa, four from R. sativus, two each of Eruca vesicaria, Eruca sativa and B. oleracea and one each from B. napus, B. tournefortii and Brassicaraphanus.

The seeds were procured from Department of Genetics and Plant Breeding, SKUAST- Jammu, India and Plant Gene Resources, Agriculture and Agri-Food, Canada.

SNP development and genotyping

A subset of 61 genic SNPs was selected by selecting 3–4 markers from each chromosome. The flanking sequences of each SNP locus were used for the designing of forward, reverse and iPLEX universal extension primer using AgenaCX assay design suite V2.0 software. The forward, reverse and iPLEX universal extension primers were diluted to 100 µM, 100 µM and 500 µM respectively, and were used for genotyping cultivated and wild Brassica species following Sequenom’s MassARRAY manufacturer protocol. The genotyping calls were evaluated using MassARRAY TYPER 4.0 software.
| S. No | Genotypes     | Taxon       | Genome composition | Genome size   | Country          |
|-------|---------------|-------------|--------------------|---------------|------------------|
| 1     | Kurikara      | B. rapa     | AA, 2n = 20        | 314.86 Mb     | Japan            |
| 2     | Siao-Baje-Tacaj | B. rapa    |                    |               | China            |
| 3     | PAK 85517     | B. rapa     |                    |               | Pakistan         |
| 4     | PAK 85552     | B. rapa     |                    |               | Pakistan         |
| 5     | PAK 85651     | B. rapa     |                    |               | Pakistan         |
| 6     | PAK 85484     | B. napus    | AACC, 2n = 38      | 912.19 Mb     | Pakistan         |
| 7     | Local selection-I | B. oleracea | CC, 2n = 18        | 514.43 Mb     | India            |
| 8     | Local selection-II | B. oleracea |                    |               | India            |
| 9     | Pusa Tarak    | B. juncea   | AABB, 2n = 36      | 954.86 Mb     | India            |
| 10    | Urvashi       | B. juncea   |                    |               | India            |
| 11    | RSPR-01       | B. juncea   |                    |               | India            |
| 12    | Zem-I         | B. juncea   |                    |               | Australia        |
| 13    | Donskaja-IV   | B. juncea   |                    |               | Russia           |
| 14    | EC 287711     | B. juncea   |                    |               | Sweden           |
| 15    | Dialba        | S. alba     | SS, 2n = 24        | NS            | NA               |
| 16    | Borowska      | S. alba     |                    |               | NA               |
| 17    | Albatross     | S. alba     |                    |               | Germany          |
| 18    | SRS 1297      | S. alba     |                    |               | Russia           |
| 19    | Asta          | S. alba     |                    |               | NA               |
| 20    | SRS 2754      | S. alba     |                    |               | Russia           |
| 21    | SRS 2755      | S. alba     |                    |               | Russia           |
| 22    | SRS 2757      | S. alba     |                    |               | Bulgaria         |
| 23    | SRS 2764      | S. alba     |                    |               | Korea            |
| 24    | SRS 2765      | S. alba     |                    |               | Korea            |
| 25    | SRS 2768      | S. alba     |                    |               | Russia           |
| 26    | Thorney       | S. alba     |                    |               | England          |
| 27    | Prekovska     | S. alba     |                    |               | Germany          |
| 28    | Kirby         | S. alba     |                    |               | Canada           |
| 29    | Novinka 515   | Raphanus sativus | RR, 2n = 18     | 392.71 Mb     | Russia           |
| 30    | Tetra Poznanska | Raphanus sativus |                    |               | Germany          |
| 31    | SRS 439       | Raphanus sativus |                    |               | Poland           |
| 32    | Nemex         | Raphanus sativus |                    |               | Poland           |
| 33    | SloboltS977/Ra32 | Brassicoraphanus | AARR, 2n = 38   | NS            | United Kingdom   |
| 34    | SRS 2855      | Eruca vesicaria | EE, 2n = 22     | NS            | Turkey           |
| 35    | SRS 2939      | Eruca vesicaria |                    |               | Pakistan         |
| 36    | SRS 2105      | B. tournefortii | TT, 2n = 20     | NS            | India            |
| 37    | Jhambha       | Eruca sativa | EE, 2n = 22       | 851 Mb        | Pakistan         |
| 38    | Ason          | Eruca sativa |                    |               | Pakistan         |

*NS- not sequenced
SSR development and genotyping

A subset of 100 SSR markers developed from *B. juncea* A-genome (unpublished data) was also used to study genetic variation in wild Brassica relatives. For SSR analysis, a two-step thermal amplification profile was followed during PCR amplification. First 10 cycles were carried out at $T_m-1$ and the next 25 cycles were carried out at $T_m-3$. The amplified products were resolved on 3.5% agarose gel and different alleles for each SSR locus were scored using 100 bp DNA ladder as a size reference.

Diversity and Population structure analysis

The molecular data from SNP and SSR genotyping was arranged in a matrix format to calculate parameters of genetic diversity analysis using Powermarker v3.51 (Liu and Muse 2005). The molecular data was also used for phylogenetic tree construction using Darwin5 software (Perrier and Jacquemoud-Collet 2006). Dissimilarity coefficient was calculated and used for tree construction using unweighted neighbour-joining method. Bootstrap value for tree construction was determined by re-sampling loci at 1000 times. The K-value in population structure analysis was carried using default parameters mentioned in Evanno et al. 2005. The membership coefficient from STRUCTURE were then integrated into CLUMPP software to generate a Q matrix (Jakobsson and Rosenberg 2007), which was used to draw a bar plot using DISTRUCT software (Rosenberg 2004).

Results and Discussion

Across genome transferability of SNP and SSR markers

Both SSR and SNP markers were used on related species of *Brassica juncea* from Brassicaceae family to understand the genetic relationship between different taxon genomes based on the ability of markers to detect corresponding loci in related species. A total of 88 SSR and 58 SNP markers were found to be functional across 38 genotypes belonging to ten different taxon groups (Supplementary Table 1). The SNPs used in the current study have revealed high cross-genome transferability of more than 59% across the species. Interestingly, the SNPs developed from amphidiploid genome AABB were found to be highly transferable to distantly related species genomes such as EE, CC, RR and SS. The highest transferability (93.17%) was observed in case of close relatives - *B. napus* (AACC) and *B. oleracea* (CC); and lowest (63.8%) was observed for *R. sativus*. For *B. tournefortii* the cross transferability was observed to be 70.68%; although, Prakash and Narain (1971) designated *B. tournefortii* as D genome due to low cross compatibility, hybrid sterility and very little gene flow. When the SNP markers from individual genome of *B. juncea* (A and B) were taken into account, the B-genome SNP loci show more transferability to wild Brassica species as compared to A-genome derived markers where none of the two genomes are available (Fig. 1). A-genome derived markers from *B. juncea* shows high rate of transferability in all those species where A genome is present suggesting a high conservancy of this genome during the evolution of this crop.

The SSR markers from AA-genome were found to be highly transferable to the related genomes of different taxon included in the current study. The extent of transferability ranged from 78% to 100% (Supplementary Table 1). The high rate of SSRs transferability reveal highly conserved genomes in the most widely used taxon of the Brassicaceae family. The percent cross-transferability obtained in present investigation was in accordance with some recent studies (Thakur et al. 2018; Singh et al. 2018). However, when the transferability potential of both the markers were compared, the SSRs seems to be far better in studying the population genetics in orphan crops species by exploiting multi-allelism among the genomic resources of cultivated species. Moreover, as the SNP loci were selected from the genic region, so some of these regions might be highly conserved for a particular species.

Genome-wise allelic patterns

The number of alleles detected by SSR markers ranged from 89 alleles (DD genome; *B. tournefortii*) to 152 alleles (SS genome; *S. alba*). The SSR markers used in the present study have been developed using sequence information from diploid progenitor (*B. rapa*) contributing AA genome to *B. juncea*. Out of 88 SSR markers, 31 markers detected private alleles among six
taxon groups with the most private alleles among genotypes from SS (Sinapis alba) and EE (Eruca species) taxons. Taxons with AA, CC, AACC and DD genomes did not carry any private allele. Detection of highest number of 16 private alleles among SS taxon indicate either a selection pressure experienced by it or it could be due to it being a descendent of ‘nigra’ lineage.

The number of alleles detected by 41 SNPs among 10 taxon groups ranged from 35 (DD genome; B. tournefortii) to 75 (SS genome; Sinapis alba). Like SSRs, the SNP markers too detected highest number of alleles in DD genome of B. tournefortii; but contrary to high number of private alleles in SSRs, there were only three SNPs that were reported to have detected private alleles among SS (Sinapis alba) and RR (Raphanus sativus).

Molecular genetic diversity

Out of total functional SNP markers, only 41 with an average missing data less than 40 % were considered for diversity analysis due to software requirement. As a result of biallelic nature of SNPs, a total of 82 alleles were amplified. The minor allele frequency ranged from 0.013 to 0.48 with an average of 0.235. The gene diversity and heterozygosity value was also able to identify the variability among the genotypes. The gene diversity ranged from 0.026 to 0.496 and heterozygosity level of markers ranged from 0.029 to 0.810 (Supplementary Table 2). The PIC (Polymorphism Information Content) value ranges from 0.025 to 0.375 with an average of 0.244 (Fig. 1).

The 88 polymorphic SSR markers were able to amplify 252 alleles among 38 genotypes and in many cases, the amplified PCR product is different from the expected size as obtained in case of species having A progenitor genome. The AA-genome SSR markers were used for the amplification of corresponding loci from AA-genome containing amphidiploid species B. juncea (AABB) and B. napus (AACC). The SSR analysis revealed that the alleles for nearly 74 % SSRs were amplified in different size (80-400 bp) range among the two amphidiploid genotypes. The results indicated that present day A-genome in these two amphidiploids is diverse from each other both at non-
coding (Thakur et al. 2018) and coding sites. This might be due to the fact that the AA-genome in these amphidiploids had evolved under different selection pressure after originating from the parental progenitor species (B. rapa). The number of alleles detected at each locus ranged from one to five with an average of 2.97 alleles per locus, with a size range of 80–400 bp reflecting a wide variation among repeat regions of different alleles. The gene diversity ranged from 0.027 to 0.694 and heterozygosity level of markers ranged from 0.026 to 0.833 (Supplementary Table 3). The PIC (Polymorphism Information Content) value ranged from 0.027 to 0.782 with an average of 0.478 (Fig. 2). The average SSRs PIC value is more than the SNPs which relies that SSRs are more informative as compared to SNP markers. Moreover, the high PIC values are also contributed due to SSRs being multi-allelic while SNPs are almost always bi-allelic in nature. The diversity within a collection of germplasm depends upon the degree of relatedness, origin of individual genotypes and the types of markers used to estimate allelic information at different loci. The SSR markers obtained from non-coding DNA tend to uncover higher genetic diversity than SNP markers that are obtained from highly conserved coding regions of a genome. However, SNPs from conserved regions are more likely to be involved in phenotype causal relationship.

Phylogenetic relationship and Population Structure

In order to see the efficacy of these two markers in determining the genetic distance between various species of Brassicaceae, a dendrogram based on unweighted neighbor joining method was constructed. Both the markers were able to grouped 38 genotypes into three major clusters {SS, (A, B, C, E, T) and (R, AR)} depending upon the difference in their genome composition. The cluster I, II and III represent SS genome group (S. alba); A-, B-, C-, E-, T- genome group and R- genome group genotypes respectively. The clustering indicated the ability of these molecular markers to form grouping of related genotypes from a genome with high level of accuracy. In case of SNP

---

![Fig. 2 SSR markers based analysis: (a) cross-transferability pattern among wild Brassica species, (b) various genetic diversity attributes, (c) UPGMA dendrogram showing genetic relationship among 38 genotypes, different colors represent different taxon and (d) structure analysis indicated genotypes grouping into three sub-populations based on membership coefficients indicated on vertical coordinate](image-url)
markers, the cluster I consists of genotypes from *S. alba* (SS) and cluster III contains genotypes from *R. sativus* (RR) and Brassicoraphanus (AARR) (Fig. 1c). As such the genetic distance of these three species is far away from the core species. *S. alba* seem closest to *B. napus* (0.472) and quite distinct from *B. juncea* (0.563) and *B. rapa* (0.548) (Supplementary Table 4). Brassicoraphanus was formed from the combination of two genome (AA and RR), but it shows more genetic closeness to *R. sativus* (RR) (0.243) when compared to AA genome species i.e. *B. rapa* (0.423) and *B. juncea* (0.510). The cluster II consists of genotypes from *E. sativa/vesicaria*, *B. juncea*, *B. tournefortii*, *B. oleracea*, *B. napus* and *B. rapa*. *E. sativa/vesicaria* was found to be closer to *B. juncea* (0.41) than to *B. rapa* (0.424), *B. napus* (0.432) and *B. oleracea* (0.514). Interestingly, *B. tournefortii* (TT genome) that showed low cross transferability of markers, was tend to found closer to *B. juncea* (0.432) than to *B. oleracea* (0.543) and other Brassica spp.

SSR markers also showed nearly the same clustering pattern (Fig. 1c) but however the genetic distance between the species was large when compared to SNP markers (Supplementary Table 5). As the SNP markers in the present study were derived from genic region and SSRs were mostly obtained from non-genic region, so the SNP loci show more conserve nature due to low mutation rate in the genic region as compared to non-genic region. Population structure estimated using STRUCTURE V2.3.4 software under the Hardy-Weinberg Equilibrium also clustered 38 genotypes into three (SSR markers) and seven (SNP markers) groups based on the maximum likelihood and delta K (ΔK) (Figs. 1d and 2d).

**Conclusions**

This is the first report of use of both SSR and SNP markers derived from *B. juncea* in wild Brassica for their phylogenetic relationship with the core species. Interestingly, the markers from both A and B genome are able to show amplification in diverse genomes such as E, R and T, thus indicating that various species in the family Brassicaceae are closely related and share some unique sequences that remain conserved during the course of evolution. Due to their multi-allelic nature, the SSR markers were found to be more informative to study relatedness among wild and cultivated species. The present study also reported molecular markers that would be useful in orphan Brassica crops for different genetic studies including fine mapping and association analysis.

**Acknowledgements** We would like to thank Dr. Manmohan Sharma of School of Biotechnology, SKUAST-Jammu for generously sharing aliquots of SSR markers.

**Authors’ contributions** RS design the experiment; JS and RS performed the experiments and written the manuscript; RM and RKS helped in the analysis of various parameters of genetic diversity.

**Funding** The present work was supported by the funds received by Dr. Ravinder Singh as research grant (BT/PR3946/AGR/2/839/2011 from 2013 to 2016) from Department of Biotechnology, Government of India, New Delhi.

**Data availability** Supplementary data to this article is available under appendix.

**Code Availability** Not applicable.

**Declarations**

**Conflict of interest** The authors declare that they have no competing interests.

**Ethics approval** Not applicable.

**Consent for publication** All the authors agreed to publish in this journal and in present form.

**References**

Edwards D, Batley J, Parkin I, Kole C (2011) Genetics, genomics and. breeding of oilseed Brassicas.CRC Press, US

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14(8):2611–2620

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801–1806

Kulak S, Hasterok R, Maluszynska J (2002) Karyotyping of Brassica amphidiploids using 5S and 25S rDNA as chromosomes markers. Hereditas. 136:144–150

Li P, Su T, Yu S, Wang H, Wang W, Yu Y, Zhang D, Zhao X, Wen C, Zhang F (2019) Identification and development of a core set of informative genic SNP markers for assaying genetic diversity in Chinese cabbage. Horticulture, Environment Biotechnol 60(3):411–425
Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21(9):2128–2129

Nagaharu U (1935) Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. Japan J Botany 7:389–452

Perrier X, Jacquemoud-Collet JP (2006) DARwin software: Dissimilarity analysis and representation for windows. http://darwin.cirad.fr/darwin

Prakash S, Narain A (1971) Genomic status of Brassica tournefortii Gouan. Theor Appl Genet 41(5):203–204

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4(1):137–138

Singh BK, Choudhary SB, Yadav S, Malhotra EV, Rani R, Ambawat S, Pandey A, Kumar R, Kumar S, Sharma HK, Singh DK (2018) Genetic structure identification and assessment of interrelationships between Brassica and allied genera using newly developed genic-SSRs of Indian Mustard (Brassica juncea L.). Ind Crops Prod 113:111–120

Sudan J, Khajuria P, Gupta SK, Singh R (2016) Analysis of molecular diversity in Indian and Exotic genotypes of Brassica juncea using SSR markers. Indian J Genet 76(3):361–364

Sudan J, Raina M, Singh R, Mustafiz A, Kumari S (2017) A modified protocol for high-quality DNA extraction from seeds rich in secondary compounds. J Crop Improve 31(5):637–647

Sudan J, Singh R, Sharma S, Salgotra RK, Sharma V, Singh G, Sharma I, Sharma S, Gupta SK, Zargar SM (2019) ddRAD sequencing-based identification of inter-genepool SNPs and association analysis in Brassica juncea. BMC Plant Biol 19(1):1–15

Thakur AK, Singh KH, Singh L, Nanjundan J, Khan YJ, Singh D (2018) SSR marker variations in Brassica species provide insight into the origin and evolution of Brassica amphidiploids. Hereditas 155(1):6

Yang GQ, Chen YM, Wang JP, Guo C, Zhao L, Wang XY, Guo Y, Li L, Li DZ, Guo ZH (2016) Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. Plant Methods 12(1):39

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.