

Analysis of single nucleotide polymorphism and simple sequence repeat markers for the estimates of genetic diversity by using two oilseed Brassica napus populations carrying genome contents of Brassica oleracea

Authors: Jiang, Junye, Kebede, Berisso, and Rahman, Habibur

Source: Canadian Journal of Plant Science, 102(3) : 679-689

Published By: Canadian Science Publishing

URL: <https://doi.org/10.1139/CJPS-2021-0204>

The BioOne Digital Library (<https://bioone.org/>) provides worldwide distribution for more than 580 journals and eBooks from BioOne's community of over 150 nonprofit societies, research institutions, and university presses in the biological, ecological, and environmental sciences. The BioOne Digital Library encompasses the flagship aggregation BioOne Complete (<https://bioone.org/subscribe>), the BioOne Complete Archive (<https://bioone.org/archive>), and the BioOne eBooks program offerings ESA eBook Collection (<https://bioone.org/esa-ebooks>) and CSIRO Publishing BioSelect Collection (<https://bioone.org/csiro-ebooks>).

Your use of this PDF, the BioOne Digital Library, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Digital Library content is strictly limited to personal, educational, and non-commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne is an innovative nonprofit that sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Analysis of single nucleotide polymorphism and simple sequence repeat markers for the estimates of genetic diversity by using two oilseed *Brassica napus* populations carrying genome contents of *Brassica oleracea*

Junye Jiang, Berisso Kebede, and Habibur Rahman

Abstract: Reliable estimates of genetic diversity among the accessions in a breeding population is important knowledge for use in breeding. Among the different types of molecular markers, single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) are largely used by breeders; however, our knowledge of the reliability of the estimates of genetic diversity based on these two types of markers in multiple populations is limited. In this study, a doubled haploid (DH) and an inbred population developed from *Brassica napus* × *Brassica oleracea* interspecific crosses were used for comparative analysis of these two types of markers. The estimates based on SNP and SSR markers showed a stronger correlation in the inbred population which was expected to carry greater genetic diversity as compared to the DH population. This inference was also evident from the analysis of different diversity groups (least, intermediate, and most similar) of these two populations for significant difference between the groups for six agronomic and seed quality traits, where this analysis failed to differentiate the diversity groups of the DH population for any of the traits. However, both marker types could differentiate the diversity groups of the inbred population for several traits. Furthermore, the six sub-populations of the inbred population could also be differentiated by both marker types. Thus, the results demonstrate the greater utility of the SSR and SNP markers in a genetically diverse population. This knowledge can be used while grouping a breeding population for diversity groups; however, caution needs to be taken while using the markers in a genetically narrow population.

Key words: genetic diversity, doubled haploid, inbred population, *B. napus* × *B. oleracea*, single nucleotide polymorphism, simple sequence repeat, Marker efficiency.

Résumé : Pour mieux exploiter la diversité génétique (DG) des obtentions dans une population utilisée pour l'hybridation, il faut pouvoir l'estimer. Parmi les différentes sortes de marqueurs moléculaires, les obtenteurs recourent souvent au polymorphisme mononucélotidique (SNP) et aux microsatellites (SSR). Cependant, on ignore dans quelle mesure les estimations de la DG reposant sur ces deux types de marqueurs sont fiables dans les populations multiples. Les auteurs ont utilisé une population d'haploïdes doubles (HD) et une population autogame de croisements interspécifiques entre *Brassica napus* et *Brassica oleracea* pour comparer les deux sortes de marqueurs. Les estimations fondées sur les marqueurs SNP et SSR sont fortement corrélées dans la population autogame, qui devrait présenter une plus grande DG que la population HD. La même inférence est évidente quand on analyse les groupes de diversité (spécimens les moins, modérément et les plus semblables) des deux populations, en quête de variations significatives entre les groupes pour six paramètres agronomiques ou relatifs à la qualité de la graine. L'analyse n'a cependant pas permis d'effectuer la distinction entre les groupes de diversité dans la population HD, peu importe le caractère. Les deux types de marqueurs permettent néanmoins de différencier ces groupes pour plusieurs caractères dans la population autogame. Selon ces résultats, les marqueurs SSR et SNP sont plus utiles avec une population génétiquement diversifiée. Sachant cela, on pourrait séparer une population destinée à l'hybridation par groupes de diversité. Quand la population se compose d'éléments génétiquement très voisins, on devra cependant faire preuve de prudence lors de l'usage de ces marqueurs. [Traduit par la Rédaction]

Mots-clés : diversité génétique, double haploïdie, population autogame, *B. napus* × *B. oleracea*, polymorphisme mononucélotidique, microsatellites, efficacité des marqueurs.

Received 20 September 2021. Accepted 23 January 2022.

J. Jiang, B. Kebede, and H. Rahman. Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G 2P5, Canada.

Corresponding author: Habibur Rahman (e-mail: habibur.rahman@ualberta.ca).

© 2022 The Author(s). Permission for reuse (free in most cases) can be obtained from [copyright.com](https://www.copyright.com).

Introduction

Canola, *Brassica napus* L. (AACC, $2n = 38$), is an important source of edible oil. Its demand in the world market is increasing; therefore, increasing the production of this crop is needed. For example, the Canola Council of Canada set a goal to increase its yield from 40.1 bu/acre (current) to 52 bu/acre by 2025 (Canola Council of Canada 2021). This crop species exists in three growth habitat types: the spring type, which is largely grown in Canada and Australia, the winter type grown in Europe, and the semi-winter type grown in China. These three types are genetically distinct; however, much less genetic diversity exists within each type. Of the two genomes of canola, genetic diversity in its C genome is relatively narrow as compared with its A genome (Bus et al. 2011; Gyawali et al. 2013). Therefore, broadening of genetic base of these three growth habit types, especially their C genome, is needed for continued improvement of this crop. The narrow genetic base of *B. napus* canola has been considered one of the impediments for continued improvement of this crop for seed yield and other traits (Cowling 2007; Zou et al. 2010).

Introduction of alleles from winter type (Butruille et al. 1999; Quijada et al. 2004; Kebede et al. 2010; Rahman 2017) and rutabagas (Shiranifar et al. 2020, 2021) into spring type *B. napus*, as well as from *Brassica rapa* L. (Li et al. 2013; Attri and Rahman 2017), *Brassica oleracea* L. (Li et al. 2014; Rahman et al. 2015, 2017; Iftikhar et al. 2018) and *Brassica juncea* (L.) Czern. or *Brassica carinata* A. Braun (Chatterjee et al. 2016) into *B. napus* through interspecific cross, have been accomplished by several researchers. Among these approaches, the *B. napus* × *B. oleracea* interspecific cross has especially been targeted to broaden the genetic base of the C genome of *B. napus*. Only a few research groups worked with this interspecific cross; therefore, very limited information is available on the breeding behavior of the progeny derived this cross. Li et al. (2014) used a kale and Bennett et al. (2012) used a Chinese kale accession in this interspecific cross despite the wide diversity exists in *B. oleracea* (Izzah et al. 2013).

The lack of effort on using *B. oleracea* in interspecific cross with *B. napus* is due to the difficulty of producing hybrids of these two species (Bennett et al. 2008; Iftikhar et al. 2018). Recently, Iftikhar et al. (2018) developed a few hundred *B. napus* lines from *B. napus* × *B. oleracea* interspecific crosses using different varieties of *B. oleracea*, and Bennett et al. (2012) also developed a population from this interspecific cross. These populations have been genotyped as well as phenotyped for different physiological, agronomic, and seed quality traits (Kebede and Rahman 2019; Nikzad et al. 2019). This is an important genetic resource for use in research to extend our knowledge on the utility of the *B. oleracea* alleles for the improvement of the C genome of *B. napus* canola.

To date, different types of molecular markers such as restriction fragment length polymorphism (Pradhan et al. 2003), random amplified polymorphic DNA (Teklewold and Becker 2006), sequence-related amplified polymorphism and different simple sequence repeat (SSR) markers (Shiranifar et al. 2020; Summanwar et al. 2021) have been used in breeding to understand the *Brassica* gene pools. In recent years, the availability of *Brassica* genome sequences and high-throughput sequencing technologies has enabled generation of a large number of single nucleotide polymorphism (SNP) markers for use in breeding (Scheben et al. 2019). Of the different types of molecular markers, SSR and SNP markers are currently widely used for different breeding applications including assessment of genetic diversity; however, our knowledge of the reproducibility of the diversity estimates by using these markers and the relevance of this diversity with different traits in *Brassica* is limited.

The objectives of this study were to assess the reliability of the estimates of genetic diversity based on SSR and SNP markers in two oilseed *B. napus* populations, derived from *B. napus* × *B. oleracea* interspecific crosses carrying varying extent of allelic diversity, and to further extend our knowledge of the effect of this diversity on different agronomic and seed quality traits.

Materials and Methods

Plant materials

Two oilseed *B. napus* populations carrying a different level of genetic diversity in their C genome were used in this study. This included a doubled haploid (DH) and an inbred population. The DH population included 88 lines derived from F_1 of Hi-Q × RIL-144, where Hi-Q is a spring *B. napus* canola (AACC, $2n = 38$) cultivar and the RIL-144 is a F_6 *B. napus* line developed from Hi-Q × *B. oleracea* var. *alboglabra* (Chinese kale; germplasm accession for research) (CC, $2n = 18$) interspecific cross (Rahman et al. 2017). Based on this pedigree, it was expected that the DH population would segregate only part of its C genome.

The inbred population included 174 lines, where 86 were F_7 lines derived from the crossing of a single elite spring *B. napus* canola line A04-73NA to six *B. oleracea* accessions belonging to four varieties of this species: viz. var. *alboglabra* line NRC-PBI (line maintained in NRC, Saskatoon, SK) (abbreviation: nrc); var. *botrytis* cv. BARI cauliflower (cau); var. *capitata* cvs. Badger Shipper (bad), Bindsachsener (bin), and Balbro (bal); and var. *italica* cv. Premium Crop (pre). A total of 88 lines were BC_1F_6 lines derived from crossing of the abovementioned F_1 s to the *B. napus* parent A04-73NA (Nikzad et al. 2019). Based on this pedigree, the inbred population was expected to carry greater diversity in their C genome as compared with the DH population.

The DH population was evaluated for the following agronomic and seed-quality traits in nine field trials:

seed yield (kg ha^{-1} , abbreviation: YIELD), days to flowering (DTF), seed oil content (%), SOC, seed protein content (%), SPC, and glucosinolate ($\mu\text{mol g}^{-1}$ seed, GLS). The DH lines were also evaluated in a separate experiment in a growth chamber for DTF at 10 h (DTF10H), 14 h (DTF14H), 16 h (DTF16H) and 18 h (DTF18H) photoperiods, and for leaf dry weight (g), total biomass (g), and shoot/root weight ratio (g). The details of the field trials for agronomic and seed quality traits of the DH population can be found in [Rahman et al. \(2017\)](#) and [Rahman and Kebede \(2021\)](#), for the growth chamber experiments for photosensitivity in [Rahman et al. \(2018\)](#), and for root and aboveground biomass in [Kebede and Rahman \(2019\)](#).

The inbred population was evaluated in 10 field trials during the period of 2015–2018 for the following traits: yield (kg ha^{-1}), DTF, days to maturity, duration of flowering, plant height (cm), grain filling period (GFP), SOC (%), SPC (%), and GLS ($\mu\text{mol g}^{-1}$ seed) content. The details of the field trial and seed quality analysis have been reported elsewhere ([Nikzad et al. 2019](#)).

Genetic similarity and difference

Genotypic data of the 88 DH lines were obtained using 113 SSR ([Rahman et al. 2017](#)) and 2420 SNP markers ([Kebede and Rahman 2019](#)), and data from the 174 inbred lines were obtained using 103 SSR ([Nikzad et al. 2019](#)) and 8091 SNP markers; these data were used to estimate genetic similarity between all possible pairs of lines of these two populations. The SNP data of the inbred population was obtained by using targeted genotyping by sequencing technique ([Nikzad 2020](#)). Similarity coefficients (SCs) between the pairs of lines of the DH and inbred population was calculated as 1 minus genetic distance which was calculated based on Nei's genetic distance method ([Nei 1972](#)). For the 88 DH lines, SCs for a total of 3828 pairs of lines ($88 \times 87/2 = 3828$) could be obtained, while for the 174 inbred lines, SCs for a total of 15 051 pairs of lines ($174 \times 173/2 = 15\,051$) could be obtained. Pearson's correlations between the SSR- and SNP-based SCs of the pairs of lines of the two populations was calculated using the 'cor.test' function, and the result was viewed using 'ggplot2' package ([Wickham 2011](#)) and 'ggplot' function of the software program R ([R Development Core Team 2015](#)). The genetic diversity information of the DH and inbred populations and genetic differentiation of the inbred lines of the six crosses by SNP and SSR markers were calculated by GenALEX version 6.5 ([Peakall and Smouse 2012](#)).

Least square mean differences (MDs) of the phenotypic data

Least square means of all phenotypic traits of the DH and inbred populations were calculated using the software program SAS ([Rodriguez 2011](#)). The least square MDs for each of the 3828 pairs of lines of the DH population and 15 501 pairs of lines of the inbred population was calculated using Excel.

Relationship between SC and MD for the two marker types

To investigate whether the SNP- and SSR-based genetic similarity and difference would show a consistent pattern with the performance of the lines for different agronomic and seed quality traits, we partitioned the 3828 and 15 501 possible pairs of the DH and inbred lines into three groups for each marker type based on the SC values and tested the groups for significant difference. The three groups included the least similar (1/3 lowest SC) group, intermediate (1/3 middle SC), and most similar group (1/3 highest SC). This grouping was done based on the hypothesis that the diversity estimate between the pairs of lines will reflect in their phenotypic performance. The difference in performance between the most similar lines will be the least, while the difference in performance between the least similar lines will be the greatest; thus, the three similarity groups will be significantly differing for their difference in performance. The average MD of each of the three groups was calculated and tested for significant difference, and the results presented as boxplots using 'ggplot' function of R software ([R Development Core Team 2015](#)). Tukey's tests for significant difference between the three groups were carried out by least significant difference using the 'Test' function from 'agricolae' package ([de Mendiburu 2013](#)) of R language software ([R Development Core Team 2015](#)). The significance level used in this statistical analysis was $p < 0.05$.

Analysis of genetic differentiation and multivariate analysis of the inbred population by SNP and SSR markers

To understand the extent of genetic diversity in the *B. oleracea* gene pool using SNP and SSR markers, the genetic variance among the six crosses, as well as within the inbred population, was estimated by analysis of molecular variance (AMOVA) using GenALEX version 6.5 ([Peakall and Smouse 2012](#)). In addition to this, the pairs of crosses were subjected to multivariate analysis. This analysis was carried out for 15 ($6 \times 5/2 = 15$) possible pairs of combinations of the six crosses for the SNP and SSR marker data separately. The biplots of principal component analysis (PCA) were evaluated for separation of every pairs of crosses; therefore, 15 biplots of PCA were generated for each of the SNP and SSR markers. The PCA was conducted using 'ggplot2' package, 'prcomp' and 'ggbiplot' functions of R language software ([R Development Core Team 2015](#)).

Results

Correlation of SCs based on SSR and SNP data and genetic diversity information of the two populations

Coefficient of correlation (r) for the SCs based on SSR and SNP marker data was 0.27 ($p < 0.001$; $R^2 = 0.073$) in the DH population of Hi-Q \times RIL-144 ([Fig. 1A](#)); however, a stronger correlation ($r = 0.59$; $p < 0.001$) with a greater R^2 value ($R^2 = 0.35$) was obtained for these two types of marker data in the inbred population derived from six

Fig. 1. Correlation of similarity coefficients (SCs) between pairs of lines estimated by single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) markers in (A) a doubled haploid (DH) (88 lines) and (B) an inbred (174 lines) population of *Brassica napus* L. derived from *B. napus* × *B. oleracea* interspecific crosses. SC = 1 – genetic distance between the pairs of lines; n = number of pair-wise combinations of the lines for which SC values were estimated. In both populations, *p* values smaller than 2.2e-16 indicates significant correlation between the SNP- or SSR-based SCs. [Colour online.]

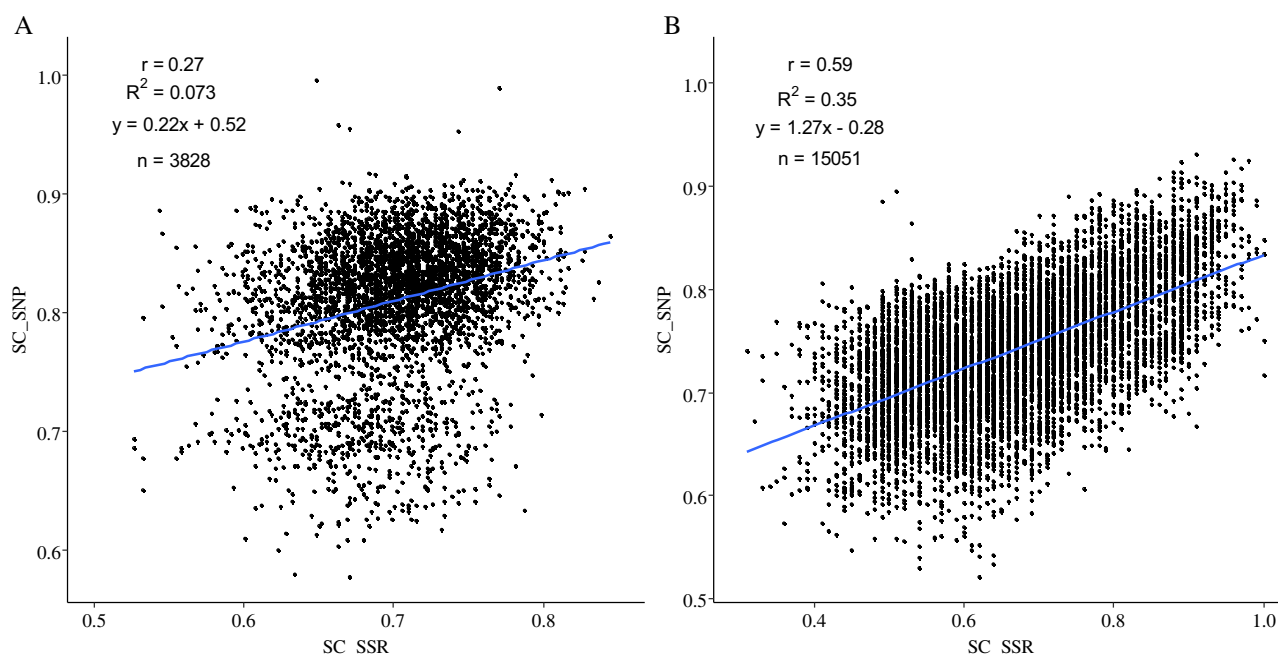


Table 1. Genetic diversity information (mean ± standard error) of DH and inbred populations of *Brassica napus* L. by SNP and SSR.

| Population | Marker | Na | Ne | I |
|------------|--------|-------------|-------------|-------------|
| DH | SNP | 1.98 ± 0.00 | 1.81 ± 0.01 | 0.62 ± 0.00 |
| | SSR | 1.98 ± 0.01 | 1.63 ± 0.03 | 0.53 ± 0.02 |
| Inbred | SNP | 1.61 ± 0.00 | 1.12 ± 0.00 | 0.15 ± 0.00 |
| | SSR | 1.85 ± 0.11 | 1.34 ± 0.05 | 0.35 ± 0.03 |

Note: I, Shannon's Information Index; Na, no. of different alleles; Ne, no. of effective alleles; DH, doubled haploid; SNP, single nucleotide polymorphism; SSR, simple sequence repeat.

B. napus × *B. oleracea* interspecific crosses (Fig. 1B). According to the Shannon's Information Index (Shannon 2001), the number of different alleles and expected alleles of DH population was 0.62, 1.98, and 1.81 for SNP and 0.53, 1.98, and 1.63 for SSR, respectively (Table 1). For the inbred population, values of these three parameters were 0.15, 1.61, and 1.12 for SNP and 0.35, 1.85, and 1.34 for SSR, respectively (Table 1).

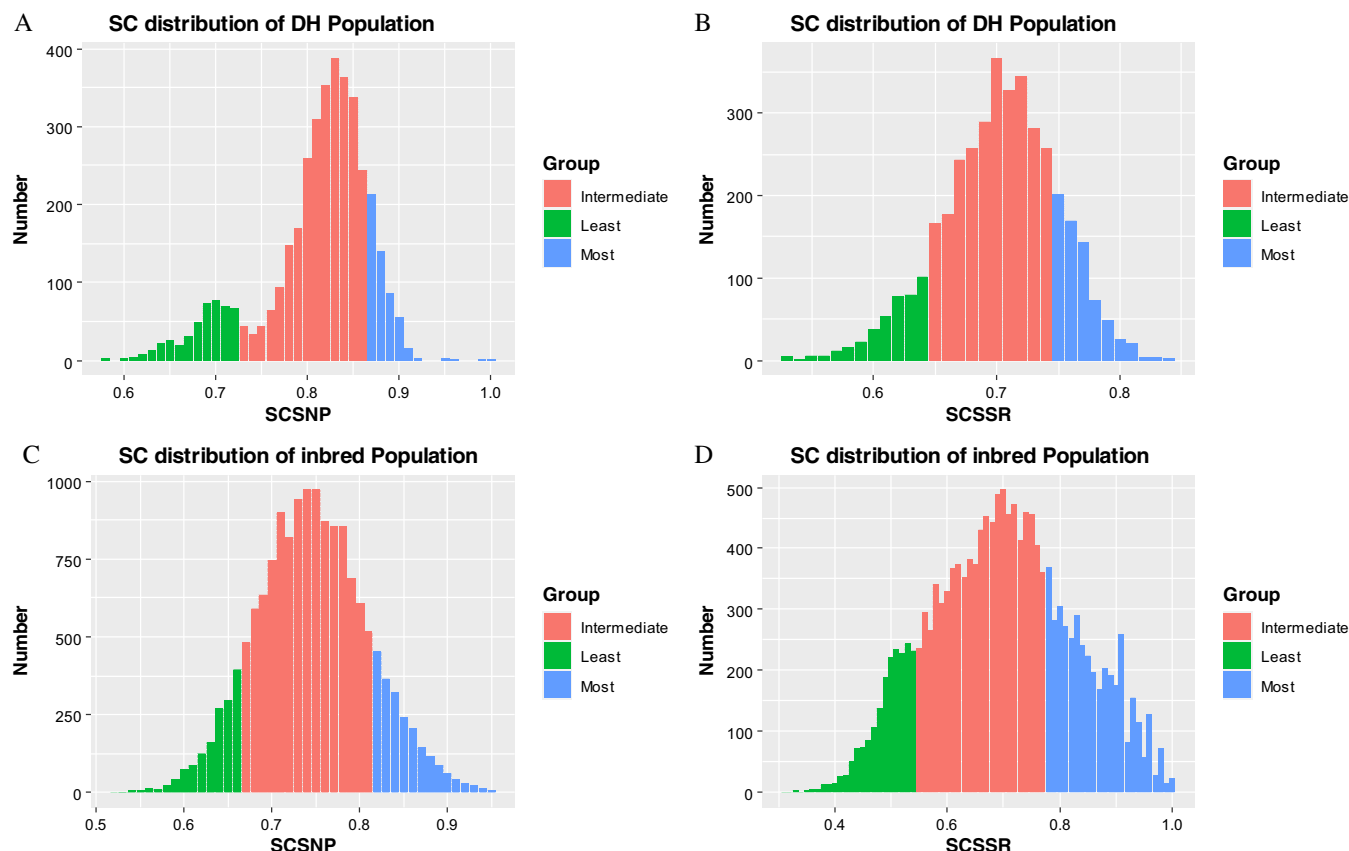
Comparison of MD of three genetic groups

DH population

The SCs of the pairs of lines estimated based on SNP markers varied from 0.58 to 1.00 (Fig. 2A). In this case, the 3828 possible pairs of lines were placed into least similar group with SCs of 0.58 to 0.72, intermediate group with SCs of 0.73 to 0.86, and most similar group with SCs of 0.87 to 1.00. In the case of SSR markers, SCs varied from

0.53 to 0.84 (Fig. 2B), where the pairs of lines with SCs of 0.53 to 0.64, 0.65 to 0.74, and 0.75 to 0.84 were included in the least, intermediate, and most similar groups, respectively. No consistent trend for significant difference between the three similarity groups based on both SNP or SSR marker could be found for any of the traits (Fig. 3A). However, in a few cases, significant (*p* < 0.05) difference between the least and most similar groups was found, such as for the traits DTF16H, total biomass, and shoot/root weight ratio by SNP, and for SOC and DTF18H by SSR. The significant difference between these two groups could also be seen for DTF based on both SNP and SSR markers; however, in the opposite direction, i.e., for the SNP-based similarity, the least similar group had greater mean value for the trait than the most similar group, while for the SSR-based similarity, the most similar group had the greater mean value (Fig. 3A).

Fig. 2. Frequency distribution of genetic similarity coefficients (SCs) of the possible pairs of lines of (A and B) the doubled haploid (DH) and (C and D) inbred populations of *Brassica napus* L. developed from *B. napus* × *B. oleracea* interspecific crosses. (A) A total of 3828 SC values ranging from 0.58 to 1.00 estimated by single nucleotide polymorphism (SNP) markers in the DH population; (B) 3828 SC values ranging from 0.53 to 0.84 estimated by simple sequence repeat (SSR) markers in the DH population; (C) 15 051 SC values ranging from 0.52 to 0.95 estimated by SNP markers in the inbred population; (D) 15 051 SC values ranging from 0.31 to 1.00 estimated by SSR markers in the inbred population. [Colour online.]



Inbred population

SCs revealed by SNP markers varied from 0.52 to 0.95 (Fig. 2C). The 15 051 possible pairs of lines were grouped into least similar group with SCs of 0.6 to 0.71, intermediate similar group with SCs of 0.72 to 0.82, and most similar group with SCs of 0.83 to 0.95. The SSR-based SCs showed wider variation in this population as compared with the DH population and varied from 0.31 to 1.00 (Fig. 2D). In this case, the least, intermediate, and most similar groups had SCs of 0.31 to 0.54, 0.55 to 0.77, and 0.78 to 1.00, respectively. In this population, six phenotypic traits, DTF, grain filling period, plant height, SOC, SPC, and GLS, showed a consistent pattern for both marker types where the greatest value was found for the least similar group and the least value for the most similar group; intermediate group was significantly different ($p < 0.05$) from these two groups (Fig. 3B). In case of days to maturity, the least similar group was found to be the lowest ($p < 0.05$) among the three groups based on both SNP and SSR markers (Fig. 3B).

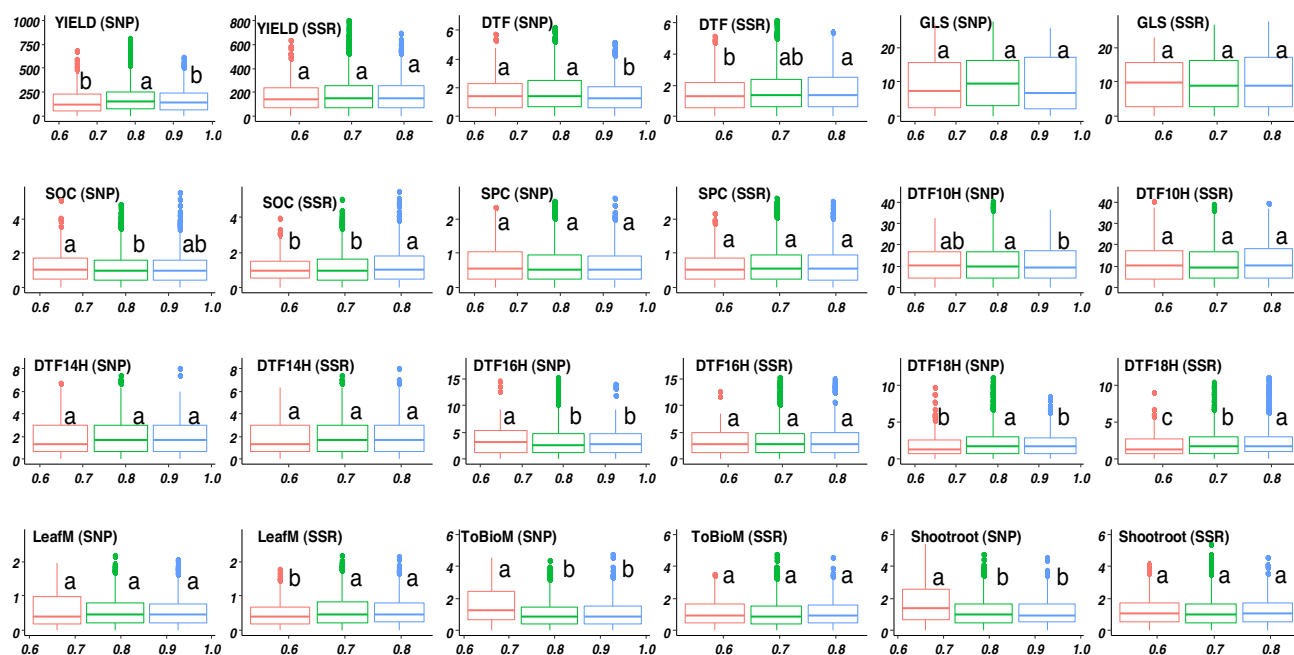
Analysis of genetic differentiation and PCA of the inbred lines by SNP and SSR markers

Results from AMOVA based on SNP and SSR data showed a consistent pattern and demonstrated the existence of significant variation among the populations of the six crosses as well as within the population of the individual crosses ($p < 0.001$). The great majority of the variance was accounted for by the population within the crosses (90% for SNP, 80% for SSR) while the populations between the six crosses accounted for only 10%–20% of the variance (10% for SNP, 20% for SSR) (Table 2).

PCA was performed using marker data of all possible pairs of the six crosses to understand genetic similarity among the six populations. Of the 15 ($6 \times 5/2 = 15$) possible pairs of cross combinations, the populations of 14 pairs could be separated by both SNP and SSR markers (Figs. 4A and 4B); however, the inbred lines of *B. napus* A04-73NA × *B. oleracea* var. *albuglabra* line NRC-PBI (nrc) and *B. napus* A04-73NA × *B. oleracea* var. *botrytis* cv. BARI

Fig. 3. Boxplot comparison of the difference between pairs of *Brassica napus* L. of two populations (doubled haploid (DH) and inbred) derived from *B. napus* × *B. oleracea* interspecific crosses. (A) DH lines belonging to least (0.58 to 0.72), intermediate (0.73 to 0.86), and most (0.87 to 1.00) similar groups revealed by single nucleotide polymorphism (SNP) markers, and the least (0.53 to 0.64), intermediate (0.65 to 0.74) and the most (0.75 to 0.84) similar groups revealed by simple sequence repeat (SSR) markers; (B) inbred lines belonging to least (0.52 to 0.66), intermediate (0.67 to 0.81), and most (0.82 to 0.95) similar groups revealed by SNP markers, and the least (0.31 to 0.54), intermediate (0.55 to 0.77), and the most (0.78 to 1.00) similar groups revealed by SSR markers for different agronomic and seed quality traits. X-axis = similarity coefficient; Y-axis = least square means (LSmean) difference between pairs of lines. Different lowercase letters (a, b, c) indicate significant difference between the groups ($p < 0.05$). DTF, days to flowering; DOF, duration of flowering; DTM, days to maturity; GFP, grain-filling period; PH, plant height; SOC, seed oil (%); SPC, seed protein (%); GLS, glucosinolate in $\mu\text{mol g}^{-1}$ seed; DTF10h, DTF14h, DTF16h, and DTF18h, days to flowering at 10, 14, 16, and 18 h photoperiod, respectively; LeafM, leaf dry weight (g); ToBioM, total biomass (g); Shootroot, ratio of shoot and root weight (g). [Colour online.]

A



B

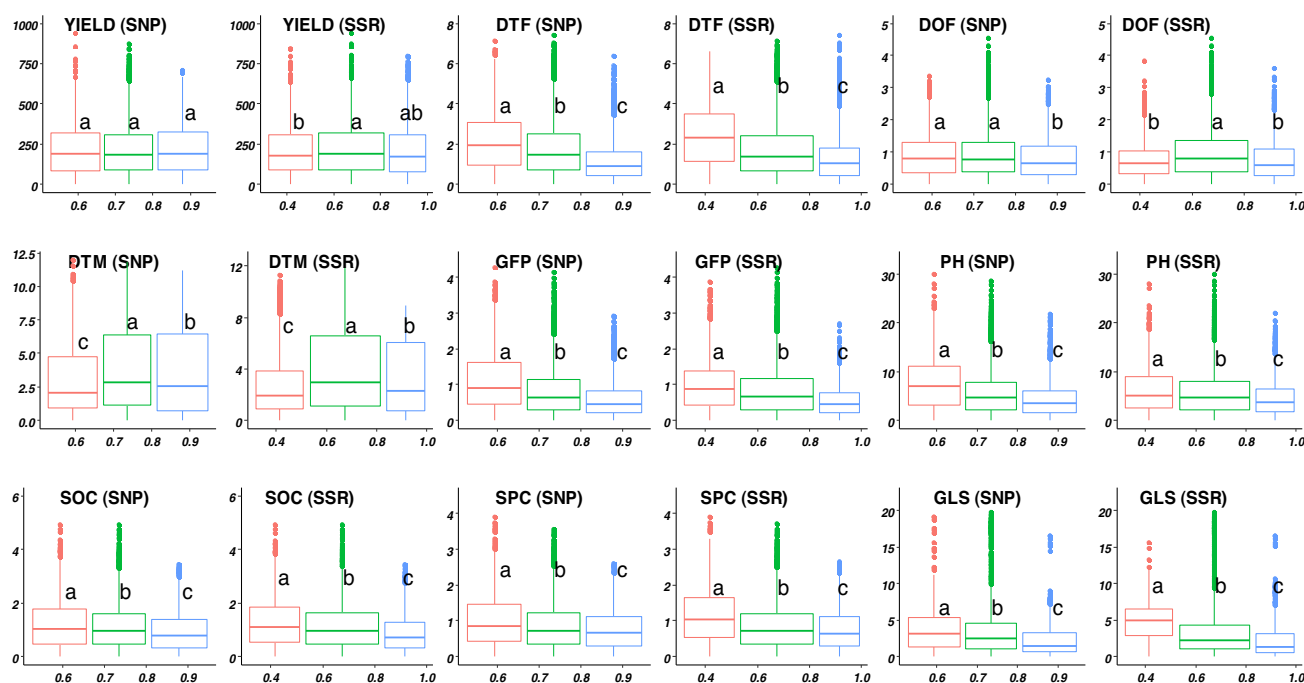


Table 2. Analysis of genetic differentiation among the inbred *Brassica napus* population derived from six *B. napus* × *B. oleracea* interspecific crosses.

| Based on SSR markers | | | | | | |
|----------------------|-----|------------|----------|-----------|------|----------|
| Source | df | SS | MS | Est. Var. | % | <i>p</i> |
| Among crosses | 5 | 33177.274 | 6635.455 | 87.604 | 10% | <0.001 |
| Within crosses | 168 | 264190.474 | 1572.562 | 786.281 | 90% | <0.001 |
| Within individual | 174 | 0 | 0 | 0 | 0% | <0.001 |
| Total | 347 | 297367.747 | | 873.885 | 100% | |
| Based on SNP markers | | | | | | |
| Source | df | SS | MS | Est. Var. | % | <i>p</i> |
| Among crosses | 5 | 1164.672 | 232.934 | 3.541 | 20% | <0.001 |
| Within crosses | 168 | 4752.512 | 28.289 | 14.144 | 80% | <0.001 |
| Within individual | 174 | 0 | 0 | 0 | 0% | <0.001 |
| Total | 347 | 5917.184 | | 17.685 | 100% | |

Note: %, percentage of total variance contributed by each component; df, degrees of freedom; Est. Var, estimates of variance components; MS, mean squares; *p* value, probability value; SS, sum of squares; *B. napus*, *Brassica napus* L.; *B. oleracea*, *Brassica oleracea* L.; SNP, single nucleotide polymorphism; SSR, simple sequence repeat.

cauliflower (cau) did not show obvious separation by either SNP or SSR markers (Figs. 4A and 4B). Thus, the results further demonstrated the utility of these two types of markers for genetic diversity analysis in this inbred population.

Discussion

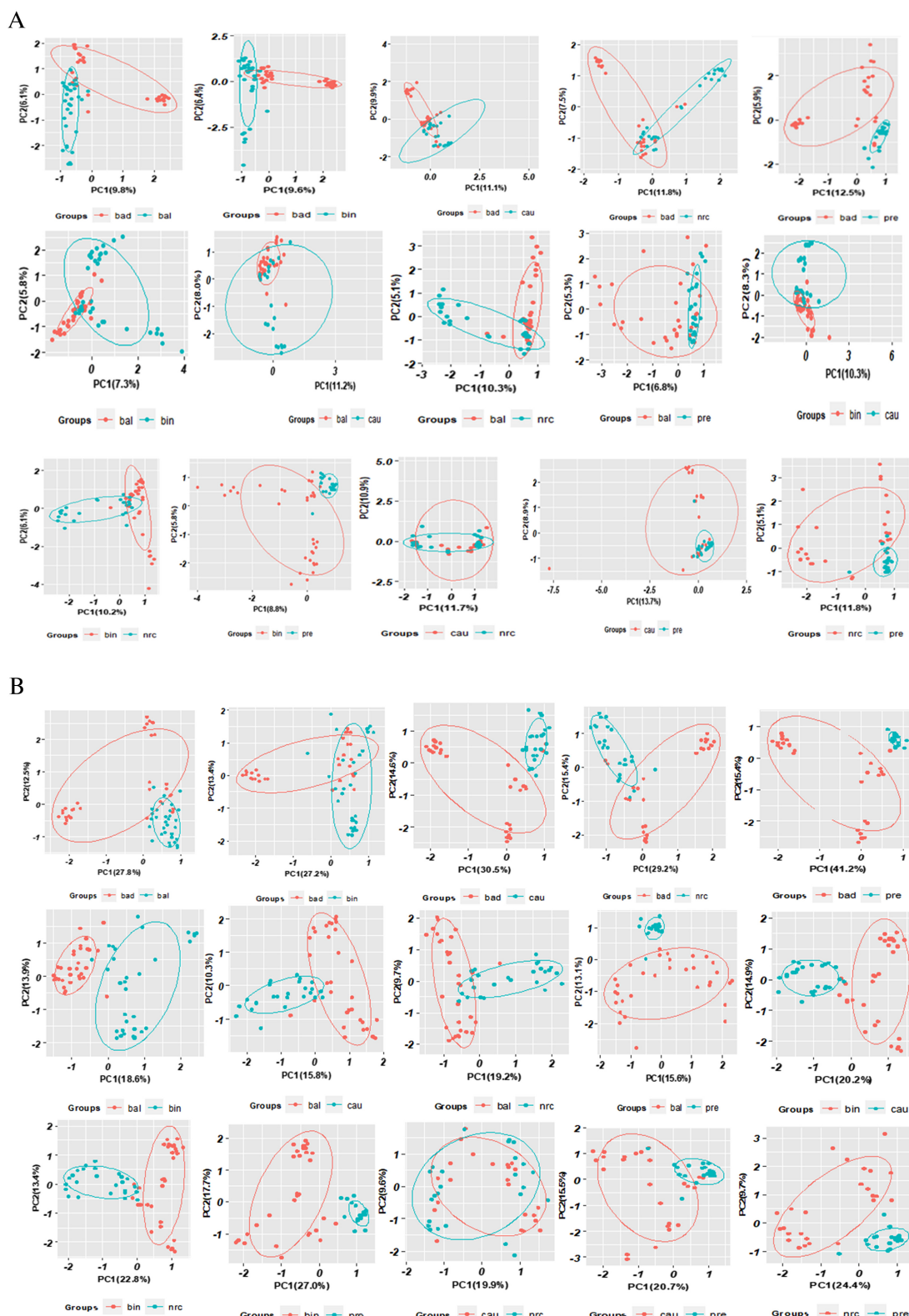
In this study, we used two populations, the DH population of Hi-Q × RIL-144 where the RIL-144 carried genome content of a single *B. oleracea* line and the inbred population of six *B. napus* × *B. oleracea* crosses, to investigate the utility of the SNP and SSR markers for the estimates of genetic diversity. Based on pedigree information, it was anticipated that the DH population carried relatively low genetic diversity as compared with the inbred population. The existence of lower genetic diversity in the DH population was also evident from a relatively narrow variation for the SC values based on SSR markers (Figs. 2B vs. 2D).

Comparative analysis of different types of markers for evaluation of genetic diversity (or relatedness) has been carried out in different crops. For example, Davierwala et al. (2000) assessed genetic diversity among 42 Indian elite rice cultivars using random amplified polymorphic DNA, inter-simple sequence repeat, and sequence-tagged microsatellite site markers and found a better estimate of genetic relationship among the cultivars while using all three markers as opposed to using a single marker. This was apparently due to a greater genome coverage by the data of the three markers. Chen et al. (2017) conducted genetic diversity analysis among 150 jujube accessions collected from all over China by using 24 SSR and 4680 SNP markers and found similar efficiency of these two markers. Van Inghelandt et al. (2010) evaluated 1537

elite maize inbred lines with 359 SSR and 8244 SNP markers and reported a strong correlation ($r = 0.87$) between the estimates of genetic distance using these two types of markers. A similar result has also been reported by Filippi et al. (2015) while evaluating 37 sunflower inbred lines from germplasm bank by 42 SSR and 384 SNP markers. In contrast, Würschum et al. (2013) did not find significant correlation between the estimates of genetic similarity based on SNP and SSR markers ($r = 0.008$; $p = 0.32$) while evaluating 172 elite European winter wheat, which might be due to the use of a limited number of SNP (1395) and SSR markers (91) for a crop with large genome (~17 Gb).

Among the two types of markers, SSR provide greater information of genetic diversity in maize (Van Inghelandt et al. 2010; Yang et al. 2011). According to Van Inghelandt et al. (2010), about 10 times greater number of SNP markers, as compared with the number of SSR markers, should be used for a reliable estimate of population structure and genetic diversity in this crop. Our comparative analysis of these two types of markers in two populations differing for the extent of genetic diversity showed a stronger correlation for the estimates in the population carrying greater diversity. We used 113 SSR and 2420 SNP markers to genotype the DH population. The occurrence of less strong correlation in this population ($r = 0.27$ in DH vs. $r = 0.59$ in inbred) for the estimate of genetic similarity based on these two types of markers is due to small genetic variation in this population, as could be inferred from pedigree history of this population. This is also evident from the lack of significant difference between the MD of the three similarity groups of the DH population for most of the traits based on both marker types. This is in contrast to the inbred

Fig. 4. Principal component analysis (PCA) of the inbred populations of derived from six *B. napus* × *B. oleracea* interspecific crosses for all possible pairs of crosses as revealed by (A) 8091 single nucleotide polymorphism (SNP) markers, and (B) 103 simple sequence repeat (SSR) markers. Crosses are indicated by abbreviation of the six *B. oleracea* accessions involved in the crosses: nrc = *B. oleracea* var. *alboglabra* line NRC-PBI; cau = var. *botrytis* cv. BARI cauliflower; bad = var. *capitata* cv. Badger Shipper; bin = var. *capitata* cv. Bindsachsener; bal = var. *capitata* cv. Balbro; pre = var. *italica* cv. Premium Crop. [Colour online.]



population where significant difference between the three groups could be found for several traits. In this case, diversity in the population apparently reflected both diversity for the marker alleles as well as for the alleles contributing to the traits.

As mentioned above, from pedigree information it is highly likely that the DH population would have a narrow genetic base. This was also evident from our previous study (Nikzad et al. 2020; Rahman and Kebede 2021) on quantitative trait locus (QTL) mapping of different agronomic and seed quality traits in these two populations. For example, only a single QTL on chromosome C5 could be detected for SOC by using the DH population (Rahman and Kebede 2021) while at least three QTL including the C5 QTL could be detected by using the inbred population (Nikzad 2020). Thus, the results from our study suggest that cautions need to be taken while estimating genetic diversity in a genetically narrow population by using a limited number of markers, especially when the markers are not from the genes reflecting the alleles.

Introgression of favorable alleles from exotic germplasm, including allied species, is important for broadening the genetic base of *B. napus* canola for continued improvement of this crop (Rahman 2013). This has been demonstrated by different researchers using different gene pools, such as by using winter canola (Butruille et al. 1999; Quijada et al. 2004; Kebede et al. 2010), *B. oleracea* (Li et al. 2014; Rahman et al. 2017; Rahman and Kebede 2021) and *B. rapa* (Qian et al. 2005) to improve the performance of spring canola. Therefore, an understanding of the different gene pools and their use in breeding is important for continued improvement of this crop. By using SSR markers, Nikzad et al. (2020) demonstrated that the six *B. oleracea* accessions that were included in the six crosses of the inbred population are genetically distinct. Our study, using both SSR and SNP markers and pair-wise comparison of the six populations derived from the six crosses involving the six accessions, further supported that these *B. oleracea* are genetically distinct; however, the Chinese kale and cauliflower accessions are found to be genetically close as compared with the other accessions (Figs. 4A and 4B). Thus, this *B. oleracea* gene pool can be an important resource for trait improvement in *B. napus*.

In conclusion, by using two populations carrying varying extent of genetic diversity, we demonstrated that both SSR and SNP markers can be used for estimating genetic diversity among the lines of a breeding population derived from crossing of genetically diverse parents; however, these two types of markers will have limited utility when used in a population carrying narrow diversity. This finding has also been supported from the analysis of morphological and seed quality data from replicated field trials of these two populations. Therefore, caution needs to be taken while using either of these two markers in a genetically narrow population;

however, the use of a greater number of markers or functional marker may compromise this limitation. To the best of our knowledge, this is the first study to compare the efficiency of these two types of markers by using both genotypic and phenotypic data and multiple populations. Knowledge from this study can be applied for characterization of the breeding materials.

Conflict of Interest

The authors declare that there is no conflict of interest.

Author Contributions

Junye Jiang carried out statistical analysis of the molecular marker and phenotypic data, and prepared the first draft of the manuscript; Berisso Kebede carried out some of the data analysis and collected some of the phenotypic data, and the other data were collected by a graduate student and published in a different form after different types of analyses; Habibur Rahman conceived the original research and developed the research plan, supervised the research, helped in writing and interpretation of the results, and provided extensive support on the manuscript for further improvement.

Acknowledgements

Habibur Rahman gratefully acknowledges the Natural Sciences and Engineering Research Council of Canada (NSERC) (grant numbers NSERC CRDPJ 419391-11 Rahman, 298778-2011 RGPIN), Alberta Innovates Bio Solutions (AI-Bio) (grant number 2011F006R), the Alberta Crop Industry Development Fund (ACIDF) (grant number 2011F006R), the Alberta Canola Producers Commission (ACPC) (grant number 2011F006R) and Nutrien AgSolutions (grant number NSERC CRDPJ 419391-11 Rahman) for financial support to this research. The authors thank An Vo, Salvador Lopez, and other staff from the Canola Program for assistance in various routine works.

References

- Attri, R., and Rahman, M.H. 2017. Introgression of allelic diversity from genetically distinct variants of *Brassica rapa* into *Brassica napus* canola and inheritance of the *B. rapa* alleles. *Crop Pasture Sci.* **69**: 94–106. doi:10.1071/CP17193.
- Bennett, R.A., Séguin-Swartz, G., and Rahman, H.M. 2012. Broadening genetic diversity in Canola using the C-Genome species *Brassica oleracea* L. *Crop Sci.* **52**: 2030–2039. doi:10.2135/cropsci2011.11.0580.
- Bennett, R.A., Thiagarajah, M.R., King, J.R., and Rahman, M.H. 2008. Interspecific cross of *Brassica oleracea* var. alboglabra and *B. napus*: effects of growth condition and silique age on the efficiency of hybrid production, and inheritance of erucic acid in the self-pollinated backcross generation. *Euphytica*, **164**: 593–601. doi:10.1007/s10681-008-9788-0.
- Bus, A., Körber, N., Snowdon, J.R., and Stich, B. 2011. Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. *Theor. Appl. Genet.* **123**: 1413–1423. doi:10.1007/s00122-011-1676-7. PMID:21847624.

- Butruille, D.V., Guries, R.P., and Osborn, T.C. 1999. Increasing yield of spring oilseed rape hybrids through introgression of winter germplasm. *Crop Sci.* **39**: 1491–1496. doi:[10.2135/cropsci1999.3951491x](https://doi.org/10.2135/cropsci1999.3951491x).
- Chatterjee, D., Banga, S., Gupta, M., Bharti, S., Salisbury, A.P., and Banga, S.S. 2016. Resynthesis of *Brassica napus* through hybridization between *B. juncea* and *B. carinata*. *Theor. Appl. Genet.* **129**: 977–990. doi:[10.1007/s00122-016-2677-3](https://doi.org/10.1007/s00122-016-2677-3). PMID:[26849238](https://pubmed.ncbi.nlm.nih.gov/26849238/).
- Chen, W., Hou, L., Zhang, Z., Pang, X., and Li, Y. 2017. Genetic diversity, population structure, and linkage disequilibrium of a core collection of *Ziziphus jujuba* assessed with genome-wide SNPs developed by genotyping-by-sequencing and SSR markers. *Front. Plant Sci.* **8**: 575. PMID:[28458680](https://pubmed.ncbi.nlm.nih.gov/28458680/).
- Canola Council of Canada. 2021. Our plan to grow the Canadian canola industry. (Internet, accessed Jan 11, 2022). <https://www.canolacouncil.org/about-us/strategic-plan/>.
- Cowling, A.W. 2007. Genetic diversity in Australian canola and implications for crop breeding for changing future environments. *Field Crops Res.* **104**: 103–111. doi:[10.1016/j.fcr.2006.12.014](https://doi.org/10.1016/j.fcr.2006.12.014).
- Daviewala, A.P., Chowdari, K.V., Kumar, S., Reddy, A.P.K., Ranjekar, P.K., and Gupta, V.S. 2000. Use of three different marker systems to estimate genetic diversity of Indian elite rice varieties. *Genetica*, **108**: 269–284. doi:[10.1023/A:1004160232679](https://doi.org/10.1023/A:1004160232679). PMID:[11294614](https://pubmed.ncbi.nlm.nih.gov/11294614/).
- de Mendiburu, F. 2013. Statistical procedures for agricultural research. Package ‘Agricolae,’ version 1.4-4. Comprehensive R Archive Network, Institute for Statistics and Mathematics, Vienna, Austria.
- Filippi, C.V., Aguirre, N., Rivas, J.G., Zubrzycki, J., Puebla, A., Cordes, D., et al. 2015. Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biol.* **15**: 52. doi:[10.1186/s12870-014-0360-x](https://doi.org/10.1186/s12870-014-0360-x). PMID:[25848813](https://pubmed.ncbi.nlm.nih.gov/25848813/).
- Gyawali, S., Hegedus, D.D., Parkin, I.A.P., Poon, J., Higgins, E., Horner, K., et al. 2013. Genetic diversity and population structure in a world collection of *Brassica napus* accessions with emphasis on South Korea, Japan, and Pakistan. *Crop Sci.* **53**: 1537–1545.
- Iftikhar, R., Wang, X., and Rahman, H. 2018. Broadening the genetic base of *Brassica napus* canola by interspecific crosses with different variants of *B. oleracea*. *Euphytica*, **214**: 133. doi:[10.1007/s10681-018-2213-4](https://doi.org/10.1007/s10681-018-2213-4).
- Izzah, N.K., Lee, J., Perumal, S., Park, J.Y., Ahn, K., Fu, D., et al. 2013. Microsatellite-based analysis of genetic diversity in 91 commercial *Brassica oleracea* L. cultivars belonging to six varietal groups. *Genet. Resour. Crop Evol.* **60**: 1967–1986. doi:[10.1007/s10722-013-9966-3](https://doi.org/10.1007/s10722-013-9966-3).
- Kebede, B., and Rahman, H. 2019. QTL mapping of root and aboveground biomass in the *Brassica* C genome using a *B. napus* population carrying genome content introgressed from *B. oleracea*. *Mol. Breeding*, **39**: 169. doi:[10.1007/s11032-019-1064-6](https://doi.org/10.1007/s11032-019-1064-6).
- Kebede, B., Thiagarajah, M.R., Zimmerli, C., and Rahman, M.H. 2010. Improvement of open-pollinated spring rapeseed (*Brassica napus* L.) through introgression of genetic diversity from winter rapeseed. *Crop Sci.* **50**: 1236–1243. doi:[10.2135/cropsci2009.06.0352](https://doi.org/10.2135/cropsci2009.06.0352).
- Li, Q., Mei, J., Zhang, Y., Li, J., Ge, X., Li, Z., and Qian, W. 2013. A large-scale introgression of genomic components of *Brassica rapa* into *B. napus* by the bridge of hexaploid derived from hybridization between *B. napus* and *B. oleracea*. *Theor. Appl. Genet.* **126**: 2073–2080. doi:[10.1007/s00122-013-2119-4](https://doi.org/10.1007/s00122-013-2119-4). PMID:[23699961](https://pubmed.ncbi.nlm.nih.gov/23699961/).
- Li, Q., Zhou, Q., Mei, J., Zhang, Y., Li, J., Li, Z., et al. 2014. Improvement of *Brassica napus* via interspecific hybridization between *B. napus* and *B. oleracea*. *Mol. Breeding*, **34**: 1955–1963. doi:[10.1007/s11032-014-0153-9](https://doi.org/10.1007/s11032-014-0153-9).
- Nei, M. 1972. Genetic distance between populations. *Am. Nat.* **106**: 283–292.
- Nikzad, A. 2020. Potential of the *Brassica oleracea* gene pool for the improvement of spring *B. napus* canola. A Ph.D. thesis. Department of Agricultural, Food and Nutritional Science, University of Alberta.
- Nikzad, A., Kebede, B., Bhavikkumar, J., and Rahman, H. 2020. Study of the genetic structure of a *Brassica napus* canola population derived from six interspecific crosses of *B. napus* × *B. oleracea*. *Can. J. Plant Sci.* **101**: 315–327. doi:[10.1139/cjps-2020-0059](https://doi.org/10.1139/cjps-2020-0059).
- Nikzad, A., Kebede, B., Pinzon, J., Bhavikkumar, J., Yang, R.C., and Rahman, H. 2019. Potential of the C genome of different variants of *Brassica oleracea* for the improvement of agronomic and seed quality traits of *B. napus* canola. *Crop Sci.* **59**: 2608–2620. doi:[10.2135/cropsci2019.05.0304](https://doi.org/10.2135/cropsci2019.05.0304).
- Peakall, R., and Smouse, P.E. 2012. GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, **28**: 2537–2539.
- Pradhan, K., Gupta, V., Mukhopadhyay, A., Arumugam, N., Sodhi, Y., and Pental, D. 2003. A high-density linkage map in *Brassica juncea* (Indian mustard) using AFLP and RFLP markers. *Theor. Appl. Genet.* **106**: 607–614. doi:[10.1007/s00122-002-1083-1](https://doi.org/10.1007/s00122-002-1083-1). PMID:[12595988](https://pubmed.ncbi.nlm.nih.gov/12595988/).
- Qian, W., Chen, X., Fu, D., Zou, J., and Meng, J. 2005. Intersubgenomic heterosis in seed yield potential observed in a new type of *Brassica napus* introgressed with partial *Brassica rapa* genome. *Theor. Appl. Genet.* **110**: 1187–1194. doi:[10.1007/s00122-005-1932-9](https://doi.org/10.1007/s00122-005-1932-9). PMID:[15806350](https://pubmed.ncbi.nlm.nih.gov/15806350/).
- Quijada, P., Udall, J.A., Polewicz, H., Vogelzang, R.D., and Osborn, T.C. 2004. Phenotypic effects of introgressing winter winter germplasm into hybrid spring canola (*Brassica napus* L.). *Crop Sci.* **44**: 1982–1989. doi:[10.2135/cropsci2004.1982](https://doi.org/10.2135/cropsci2004.1982).
- R Development Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria.
- Rahman, M.H. 2013. Review: breeding spring canola (*Brassica napus* L.) by the use of exotic germplasm. *Can. J. Plant Sci.* **93**: 363–373. doi:[10.4141/cjps2012-074](https://doi.org/10.4141/cjps2012-074).
- Rahman, M.H. 2017. UA AlfaGold clearfield herbicide-tolerant spring *Brassica napus* canola developed from winter × spring canola cross. *Can. J. Plant Sci.* **97**: 144–146.
- Rahman, M.H., and Kebede, B. 2021. Mapping of seed quality traits in the C genome of *Brassica napus* by using a population carrying genome content of *B. oleracea* and their effect on other traits. *Plant Genome*, e20078. doi:[10.1002/tpg2.20078](https://doi.org/10.1002/tpg2.20078).
- Rahman, M.H., Bennett, A.R., and Ginette Séguin-Swartz, G. 2015. Broadening genetic diversity in *Brassica napus* canola: development of canola-quality spring *B. napus* from *B. napus* × *B. oleracea* var. *alboglabra* interspecific crosses. *Can. J. Plant Sci.* **95**: 29–41. doi:[10.4141/cjps-2014-017](https://doi.org/10.4141/cjps-2014-017).
- Rahman, M.H., Bennett, A.R., and Kebede, B. 2017. Mapping of days to flower and seed yield in spring oilseed *Brassica napus* carrying genome content introgressed from *Brassica oleracea*. *Mol. Breeding*, **37**: 5. doi:[10.1007/s11032-016-0608-2](https://doi.org/10.1007/s11032-016-0608-2).
- Rahman, M.H., Bennett, A.R., and Kebede, B. 2018. Molecular mapping of QTL alleles of *Brassica oleracea* affecting days to flowering and photosensitivity in spring *Brassica napus*. *PLoS ONE*, **13**: e0189723. doi:[10.1371/journal.pone.0189723](https://doi.org/10.1371/journal.pone.0189723). PMID:[29320498](https://pubmed.ncbi.nlm.nih.gov/29320498/).
- Rodriguez, R.N. 2011. “SAS,” Wiley Interdiscip. Rev.: Comput. Stat. **3**: 1–11. doi:[10.1002/wics.131](https://doi.org/10.1002/wics.131).
- Scheben, A., Verpaalen, B., Lawley, C.T., Bayer, P.E., Batley, J., and Edwards, D. 2019. CropSNPdb: a database of SNP array

- data for Brassica crops and hexaploid bread wheat. *Plant J.* **98**: 142–152. doi:[10.1111/tpj.14194](https://doi.org/10.1111/tpj.14194). PMID:[30548723](https://pubmed.ncbi.nlm.nih.gov/30548723/).
- Shannon, C.E. 2001. A mathematical theory of communication. *Bell Syst. Tec. J.* **27**: 379–423. doi:[10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x).
- Shiranifar, B., Hobson, N., Kebede, B., Yang, R.C., and Rahman, H. 2020. Potential of rutabaga (*Brassica napus* var. *napobrassica*) gene pool for use in the breeding of *B. napus* canola. *Crop Sci.* **60**: 157–171. doi:[10.1002/csc2.20074](https://doi.org/10.1002/csc2.20074).
- Shiranifar, B., Hobson, N., Kebede, B., Yang, R.C., and Rahman, H. 2021. Potential of rutabaga (*Brassica napus* var. *napobrassica*) gene pool for use in the breeding of hybrid spring *Brassica napus* canola. *Plant Breeding*, **140**: 305–319. doi:[10.1111/pbr.12895](https://doi.org/10.1111/pbr.12895).
- Summanwar, A., Basu, U., Kav, N., and Rahman, M.H. 2021. Identification of lncRNAs in response to infection by *Plasmodiophora brassicae* in *Brassica napus* and development of lncRNA-based SSR markers. *Genome*, **64**: 547. doi:[10.1139/gen-2020-0062](https://doi.org/10.1139/gen-2020-0062). PMID:[33170735](https://pubmed.ncbi.nlm.nih.gov/33170735/).
- Teklewold, A., and Becker, H. 2006. Geographic pattern of genetic diversity among 43 Ethiopian mustard (*Brassica carinata* A. Braun) accessions as revealed by RAPD analysis. *Genet. Resour. Crop Evol.* **53**: 1173–1185. doi:[10.1007/s10722-005-2011-4](https://doi.org/10.1007/s10722-005-2011-4).
- Van Inghelandt, D., Melchinger, A.E., Lebreton, C., and Stich, B. 2010. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor. Appl. Genet.* **120**: 1289–1299. doi:[10.1007/s00122-009-1256-2](https://doi.org/10.1007/s00122-009-1256-2).
- Wickham. 2011. ggplot2. Wiley Interdiscip. Rev. Comput. **3**: 180–185. doi:[10.1002/wics.147](https://doi.org/10.1002/wics.147).
- Würschum, T., Langer, S.M., Longin, C.F.H., Korzun, V., Akhunov, E., Ebmeyer, E., et al. 2013. Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *Theor. Appl. Genet.* **126**: 1477–1486. doi:[10.1007/s00122-013-2065-1](https://doi.org/10.1007/s00122-013-2065-1). PMID:[23429904](https://pubmed.ncbi.nlm.nih.gov/23429904/).
- Yang, X., Xu, Y., Shah, T., Li, H., Han, Z., Li, J., and Yan, J. 2011. Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. *Genetica*, **139**: 1045–1054. doi:[10.1007/s10709-011-9606-9](https://doi.org/10.1007/s10709-011-9606-9). PMID:[21904888](https://pubmed.ncbi.nlm.nih.gov/21904888/).
- Zou, J., Zhu, J., Huang, S., Tian, E., Xiao, Y., Fu, D., et al. 2010. Broadening the avenue of intersubgenomic heterosis in oil-seed *Brassica*. *Theor. Appl. Genet.* **120**: 283–290. doi:[10.1007/s00122-009-1201-4](https://doi.org/10.1007/s00122-009-1201-4). PMID:[19911158](https://pubmed.ncbi.nlm.nih.gov/19911158/).