**Analysis of SNP and SSR markers for the estimates of genetic diversity by using two oilseed Brassica napus populations carrying genome contents of Brassica oleracea**

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>Canadian Journal of Plant Science</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>CJPS-2021-0204.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>13-Jan-2022</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Jiang, Junye; Alberta Agriculture and Rural Development, Kebede, Berisso; University of Alberta, Department of Agricultural, Food and Nutritional Science Rahman, Habibur; University of Alberta, Department of Agricultural, Food and Nutritional Science</td>
</tr>
<tr>
<td>Is the invited manuscript for consideration in a Special Issue?:</td>
<td>Not applicable (regular submission)</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Genetic diversity, B. napus × B. oleracea, SNP, SSR, Marker efficiency</td>
</tr>
</tbody>
</table>
Analysis of SNP and SSR 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*

Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G 2P5, Canada

*Corresponding author email: habibur.rahman@ualberta.ca

**Abstract:** Reliable estimates of genetic diversity (GD) among the accessions in a breeding population is important for use of this knowledge in breeding. Among the different types of molecular markers, SNP and SSR are largely used by breeders; however, our knowledge of the reliability of the estimates of GD 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 GD 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*, SNP, SSR, Marker efficiency

**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 habit types: the spring type, which is largely grown in Canada and Australia, the winter type grown in the European Union, 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 to 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 as one of the impediments for continued improvement of this crop for seed yield and other traits (Cowling 2007, Zhou et al. 2010).

Introduction of alleles from winter type (Butruille et al. 1999, Quijada et al. 2004, Kebele et al. 2010) and rutabagas (Shiranifar et al. 2020, 2021) into spring type *B. napus*, as well as from *B. rapa* (Li et al. 2013, Attri et al. 2017), *B. oleracea* (Li et al. 2014, Rahman et al. 2015,
2017, Iftikhar et al. 2018) and *B. juncea* or *B. carinata* (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 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 apparently due to the difficulty of producing hybrids of these two species (Iftikhar et al. 2018, Bennett et al. 2008). 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 (Nikzad et al. 2019, Kebede 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 (RFLP) (Pradhan et al. 2003), random amplified polymorphic DNA (RAPD) (Teklewold et al. 2006), sequence-related amplified polymorphism (SRAP) (2) and simple sequence repeats (SSRs) markers (Shiranifar et al. 2020, Summanwar et al. 2006) have been used to understand the Brassica gene pools for use in breeding. In recent years, the availability of Brassica genome sequences and high-throughput sequencing technologies has enabled to
generate 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 F₁ of Hi-Q × RIL-144, where Hi-Q is a spring B. napus canola (AACC, 2n = 38) cultivar and the RIL-144 is a F₆ 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 will be segregating only for a part of its C genome. On the other hand, the inbred population included 174 lines, where 86 were F₇ 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), var. *italica* cv. Premium Crop (pre), and 88 were BC$_{3}$F$_{6}$ lines derived from crossing of the above-mentioned 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 to 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 (μmol g$^{-1}$ seed, GLS). The DH lines were also evaluated in a separate experiment in a growth chamber for days to flowering at 10 h (DTF10H), 14 h (DTF14H), 16 h (DTF16H) and 18 h (DTF18H) photoperiods, and for leaf dry weight (g, LeafM), total biomass (g, ToBioM) and shoot and root weight (g, Shootroot). 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 to 2018 for the following traits: Seed yield (kg ha$^{-1}$, YIELD), days to flowering (DTF), days to maturity (DTM), duration of flowering (DOF), plant height (cm, PH), grain filling period (GFP), seed oil (%, SOC), protein (%, SPC) and glucosinolate (μmol g$^{-1}$ seed, GLS) contents. The details of the field trial and seed quality analysis have been reported elsewhere (Nikzad et al. 2019).

Genetic similarity/difference

Genotypic data of the 88 DH lines obtained by using 113 SSR (Rahman et al. 2017, 2021) and 2,420 SNP markers (Kebede and Rahman 2019), and of the 174 inbred lines obtained by using
103 SSR (Nikzad et al. 2019) and 8,091 SNP markers 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 (tGBS) 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). In case of the 88 DH lines, SCs for a total of 3,828 pairs of lines (88 × 87 / 2 = 3,828) could be obtained, while for the 174 inbred lines, SCs for a total of 15,051 pairs lines (174 × 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 ‘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 6.5 (Rod et al. 2012).

**Least square mean differences of the phenotypic data**

Least square means (LSmeans) of all phenotypic traits of the DH and inbred populations were calculated using the software program SAS (Rodriguez 2011). The least square mean differences (MD) for each of the 3,828 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/difference would show a consistent pattern with the performance of the lines for different agronomic and seed quality traits, we partitioned the 3,828 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 LSD 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 6.5 (Rod et al. 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 SC’s 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 × 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 *B. napus* × *B. oleracea* interspecific crosses (Fig. 1B). The Shannon’s Information Index (Shannon 2001), number of different alleles and expected alleles of DH population was 0.62, 1.98, 1.81 for SNP and 0.53, 1.98, 1.63 for SSR, respectively (Table 1). For the inbred population, values of these three parameters were 0.15, 1.61, 1.12 for SNP and 0.35, 1.85, 1.34 for SSR, respectively (Table 1).

Comparison of mean difference (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 3,828 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 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, ToBioM and Shootroot 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).

**Inbred population:** SCs revealed by SNP markers varied from 0.52 to 0.95 (Fig. 2C). In this case, the 15,051 possible pairs of lines could be 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 to 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, GFP, PH, 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 DTM, 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 to 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 × 5 / 2 = 15) possible pairs of cross combinations, the populations of the 14 pairs could be separated by both SNP and SSR markers (Fig 4A and 4B). However, the inbred lines of *B. napus* A04-73NA × *B. oleracea* var. *alboglabra* line NRC-PBI (nrc) and *B. napus* A04-73NA × *B. oleracea* var. *botrytis* cv. BARI cauliflower (cau) did not show obvious separation by either SNP or SSR markers (Fig 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, viz. 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 carries relatively low genetic diversity as compared to 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 (Fig. 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 RAPD, inter-simple sequence repeat (ISSR) and sequence tagged microsatellite site (STMS) 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 4,680 SNP markers and found similar efficiency of these two markers. Van Inghelandt et al. (2010) evaluated 1,537 elite maize inbred lines with 359 SSR and 8,244 SNP markers and reported a strong correlation ($r = 0.87$) between the estimates of genetic distance by using these two types of markers. 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 (1,395) 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 to 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 2,420 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 apparently 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 population where significant difference between the three groups could be found for several traits. In this case, diversity in the population apparently reflected from 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 (Rahman and Kebede 2021; Nikzad 2020) on 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 seed oil content 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 (Rahman et al. 2017, Li et al. 2014, 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 by 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 to the other accessions (Fig 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, cautions need 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 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.

**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.

Authors’ 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 has been collected by a graduate student and has been 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.

Conflict of interest

The authors declare that there is no conflict of interest.

ORCID

Habibur Rahman https://orcid.org/0000-0002-6736-6520

References:


Rahman, M. H., 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. The Plant Genome, e20078.


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

<table>
<thead>
<tr>
<th>Population</th>
<th>Marker</th>
<th>Na</th>
<th>Ne</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH</td>
<td>SNP</td>
<td>1.98 ± 0.00</td>
<td>1.81 ± 0.01</td>
<td>0.62 ± 0.00</td>
</tr>
<tr>
<td></td>
<td>SSR</td>
<td>1.98 ± 0.01</td>
<td>1.63 ± 0.03</td>
<td>0.53 ± 0.02</td>
</tr>
<tr>
<td>Inbred</td>
<td>SNP</td>
<td>1.61 ± 0.00</td>
<td>1.12 ± 0.00</td>
<td>0.15 ± 0.00</td>
</tr>
<tr>
<td></td>
<td>SSR</td>
<td>1.85 ± 0.11</td>
<td>1.34 ± 0.05</td>
<td>0.35 ± 0.03</td>
</tr>
</tbody>
</table>

Na = No. of different alleles; Ne = No. of Effective alleles; I = Shannon's Information Index;
Table 2. Analysis of genetic differentiation among the inbred *Brassica napus* population derived from six *B. napus × B. oleracea* interspecific crosses.

A. Based on SSR markers

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>Est. Var.</th>
<th>%</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among crosses</td>
<td>5</td>
<td>33177.274</td>
<td>6635.455</td>
<td>87.604</td>
<td>10%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Within crosses</td>
<td>168</td>
<td>264190.474</td>
<td>1572.562</td>
<td>786.281</td>
<td>90%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Within individual</td>
<td>174</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Total</td>
<td>347</td>
<td>297367.747</td>
<td>873.885</td>
<td></td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

B. Based on SNP markers

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>Est. Var.</th>
<th>%</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among crosses</td>
<td>5</td>
<td>1164.672</td>
<td>232.934</td>
<td>3.541</td>
<td>20%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Within crosses</td>
<td>168</td>
<td>4752.512</td>
<td>28.289</td>
<td>14.144</td>
<td>80%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Within individual</td>
<td>174</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Total</td>
<td>347</td>
<td>5917.184</td>
<td>17.685</td>
<td></td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

Note: df = degrees of freedom; SS = sum of squares; MS = mean squares; Est. Var = estimates of variance components; % = percentage of total variance contributed by each component; $P$-value = probability value.
Fig. 1. Correlation of similarity coefficients (SCs) between pairs of lines estimated by SNP and SSR markers in a (A) DH (88 lines) and (B) an inbred (174 lines) population of *Brassica napus* derived from *B. napus × B. oleracea* interspecific crosses. Similarity coefficient = 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.
Fig. 2 Frequency distribution of genetic similarity coefficients (SCs) of the possible pairs of lines of the doubled haploid (DH) (A and B) and inbred (C and D) populations of *Brassica napus* developed from *B. napus × B. oleracea* interspecific crosses. A. 3,828 SC values ranging from 0.58 to 1.00 estimated by SNP markers in the DH population; B. 3,828 SC values ranging from 0.53 to 0.84 estimated by 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.
Fig. 3. Boxplot comparison of the difference between pairs of *Brassica napus* of two populations (doubled haploid and inbred) derived from *B. napus × B. oleracea* interspecific crosses. A. Doubled haploid (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 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 SSR markers,
and 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 = Lsmean difference between pairs of lines. Different alphabets (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 μmol g⁻¹ seed; DTF10h, DTF14h, DTF16h and DTF18h = days to flowering at 10-, 14-, 16- and 18-hours photoperiod, respectively; LeafM = leaf dry weight (g); ToBioM = total biomass (g); Shootroot = ratio of shoot and root weight (g).
Fig. 4. 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. 8,091 SNP markers, and B. 103 SSR markers. Crosses are indicated by abbreviation of the six *B. oleracea* accessions involved in the