Assessment of molecular diversity in doubled haploid lines of Camelina (Camelina sativa), as a new emerging oil crop

Mozafar Sadeghikian
  Razi University of Kermanshah: Razi University

Abdollah Najaphy (nadph@yahoo.com)
  Razi University  https://orcid.org/0000-0003-4990-7538

Danial Kahrizi
  Razi University of Kermanshah: Razi University

Hossein Rostami-Ahmadvandi
  AREEO: Agricultural Research Education and Extension Organization

Research Article

Keywords: Camelina, Genetic diversity, Doubled haploid lines, ISSR markers

DOI: https://doi.org/10.21203/rs.3.rs-435780/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Camelina [Camelina sativa (L.) Crantz], an oilseed crop, belongs to Brassicaceae family. Two unique features of camelina in comparison with the main oil crops are adaptation to different environments and also its unique oil composition. Development of doubled haploid plants is one of the essential methods for the crop improvement. This research was conducted to evaluate the genetic variation of 81 Camelina doubled haploid (DH) lines obtained from fifteen crosses by ISSR markers. The total number of amplified bands was 243, of which 239 bands (98.3%) showed polymorphism. The percentage of polymorphic bands (PPB) varied between 93.75 and 100. The size of the bands ranged from 50 to 1,700 base pairs. The informative ISSRs were identified by estimating marker features: polymorphism information content (PIC), effective multiplex ratio (EMR), marker index (MI) and resolving power (RP). Three markers had higher RP values (9.88, 8.5 and 7.46) and were the most informative markers to identify the DH lines. Cluster analysis based on Complete algorithm divided the lines into five groups, indicating relatively clear configuration from the geographic distribution patterns of the parents of the doubled haploid lines. Principal coordinate analysis (PCoA) classified the 81 camelina DH lines into six groups. The lines grouping by these two methods was similar to each other. The ISSR markers detected high polymorphism to reveal genetic variation of camelina DH lines. The findings of this research, along with biochemical traits, can improve classical and molecular breeding programs of camelina.

Introduction

The vegetable oils are mainly obtained from oilseeds such as soybean, sunflower, cottonseed, peanut and rapeseed, which are in high demand for water. However, these common oilseeds, despite their many benefits, have their limitations in terms of cultivation and climatic conditions [1].

Camelina (Camelina sativa) is from the Brassicaceae family and has been shown in many experiments to have much lower water requirements and greater resistance to spring cold than the other oil crops, especially canola [2]. Camelina is a crop native to Europe and South Asia and may grow as a weed in some farms. Its cultivation history dates back to 4000 years ago. The Former Soviet Union was the largest producer of this plant in the twentieth century by 300,000 hectares in 1950 [3]. Camelina oil possesses high omega-3 fatty acids so has many properties and uses in nutrition and health that may prevent many diseases and disorders. In industry, it is used as biofuel, resins, waxes, as well as for the production of cosmetics and pharmaceutical products. In oil refineries, industrial antioxidants, which are extremely dangerous to human health, are added to prevent oxidation as well as to increase the duration of the oil. However, due to the high content of alpha-tocopherol and high vitamin E in camelina oil, which are strong antioxidants, camelina does not need any additives to last [4].

89418 protein-coding genes have been identified in the camelina genome as a hexaploid plant (2n = 40, 750Mb) [5]. Unlike many crops, there was limited information about the genus Camelina and its biodiversity until 2005. The genome map of the camelina was drawn by [6]. Singh et al. [7] developed a genomic map using 768 SNP markers and the Illumina GoldenGate SNP array method. Vollmann et al. [8] performed DNA analysis using RAPD markers to investigate the variation of camelina accessions in Australia and reported
four distinct clusters. In this study, only 63% of the 30 markers studied were polymorphic. This indicates a low diversity of self-pollinating plants, such as camelina. Many camelina lines show little variation in biochemical traits. Creating variation and reviewing are essential for the camelina to consider positive characters and examine them in different environments.

As mentioned, due to the polyploidy of the camelina genome, the classical breeding methods are complicated. On the other hand, cross-species hybridization in this plant has not been successful. It seems that the production of double haploid plants isolated from the cultivation of anthers or microspores is a suitable method of creating diversity and improvement of this plant [9]. This research was conducted to assess the genetic diversity of 81 Camelina doubled haploid lines by ISSR markers.

**Materials And Methods**

**Plant materials**

In this study, 81 Camelina doubled haploid lines were studied in terms of genetic diversity. Table 1 shows the list of these lines, their parents and origins [10]. Camelina doubled haploid seeds were planted in October 2018 in the greenhouse of Faculty of Agricultural Sciences and Engineering, Razi University, Kermanshah, Iran.

**DNA purification**

DNA extraction was performed in the laboratory of Zagros Bioidea Company located in Science and Technology Park, Kermanshah, Iran. Leaf samples of young leaves of potted seedlings grown in the greenhouse were prepared and immediately transferred to a freezer at -80°C in liquid nitrogen (-196°C). DNA extraction from leaf samples was performed using a kit made by Zagros Bioidea Company.
| Cross N. | Cultivar          | Origin     | Cultivar         | Origin      | DH lines Numbers |
|---------|-------------------|------------|------------------|-------------|------------------|
| 1       | Voronezskij 349   | Russia     | Kirgizskij 1     | Kyrgyzstan  | 1, 2, 5, 37, 51, 62, 77, 86, 87, 93 |
| 2       | Omskij Mestnyj    | Russia     | Irkutskij Mestnyj | Irkutsk Region | 3, 6, 9, 32, 43, 60, 79 |
| 3       | Przybrodzka       | Poland     | Hoga             | Denmark     | 7, 14, 104      |
| 4       | Saratouskij       | Russia     | Bronowska        | Poland      | 24, 40, 96, 115 |
| 5       | Chulymskij        | Russia     | Omskij Mestnyj   | Russia      | 25, 41, 42, 68  |
| 6       | Krupnosemjannyj   | Russia     | Brzybrodzka II   | Poland      | 26, 78          |
| 7       | Came              | Germany    | Volynskaja       | Former Soviet Union | 69, 76, 98 |
| 8       | Boha              | Denmark    | Volynskaja       | Poland      | 23, 97          |
| 9       | Came              | Germany    | Omskij           | Former Soviet Union | 4, 35, 82, 88, 94, 95, 100 |
| 10      | Svalöf            | Sweden     | Ukrajinskij      | Former      | 27,             |
ISSR amplification

Ten ISSR markers were used to detect genetic variation at the molecular level among all doubled haploid lines (Table 2). PCR amplification was done according to Najaphy et al. [11] and Williams et al. [12] in a BIOER XP CYCLER thermocycler. Agarose gels staining and visualization were conducted using ethidium bromide and a UV transilluminator, respectively.

Statistical analysis

Amplified products were scored as (0) for the absence or (1) for the presence about each of the camelina DH lines. Jaccard's similarity coefficients were calculated based on the bivariate. Cluster analysis using
Complete method and Principal coordinate analysis (PCoA) were performed by NTSYS-pc software version 2.02.

The number of amplified and polymorphic bands, and the percentage of polymorphic bands (PPB) were measured for all of the markers. To determine the informative ISSRs to discriminate among camelina DH lines, polymorphism information content (PIC), effective multiplex ratio (EMR), marker index (MI) and resolving power (RP) were estimated. PIC was measured according to Mousapour Gorji et al. [13], as $\text{PIC} = 1 - p^2 - q^2$, where $p$ is the frequency of the present band and $q$ is the frequency of the absent band. DI (Diversity Index) was calculated as the average PIC value. EMR was recorded using the product of the number of polymorphic bands and the fraction of polymorphic bands [14]. MI index was measured according to Powell et al. [15] as the product of EMR and PIC. The formula $\text{RP} = \sum I_b$ was used to calculate RP, where $I_b$ is band informativeness and $I_b = 1 - [2 \times (0.5 - p)]$, where $p$ is the proportion of genotypes which contains the band [16].

**Results**

**ISSR Polymorphism and marker parameters**

In this experiment, 10 ISSR markers were used individually. Marker parameters are summarized in Table 2. The total number of amplified bands was 243, of which 239 bands (98.3%) showed polymorphism. The character ranged between 16 (P4) and 33 (MM15). MM15 and P9 primers possessed the highest number of total amplified products as 32 and 31, respectively. The percentage of polymorphic bands (PPB) differed from 93.75 for P4 to 100 percent for primers P1, P6, P7, P8, P9, and MM14. The size of the bands varied between 50 and 1700 base pairs. The highest PIC was observed for MM16 primer (0.49) and P10 (0.48). PIC ranged between 0.25 and 0.49 with an average of 0.35.
Table 2: ISSR Marker parameters for genetic diversity evaluation of camelina DH lines

| RP | EMR | MI  | PIC | Total amplified bands | No. of polymorphic bands | PPB  | Sequence 3' → 5' | Primer |
|----|-----|-----|-----|-----------------------|--------------------------|------|------------------|--------|
| 4.8 | 20  | 6.8 | 0.34 | 20                    | 20                       | 100% | GAG (CAA)₅       | P1     |
| 3.36 | 14.06 | 4.64 | 0.33 | 16                    | 15                       | 93.75% | CTG (GT)₈ | P4     |
| 4.15 | 29  | 7.25 | 0.25 | 29                    | 29                       | 100% | (AG)₈GCC         | P6     |
| 4.87 | 28  | 8.12 | 0.29 | 28                    | 28                       | 100% | (AG)₈GTG         | P7     |
| 3.68 | 25  | 6.25 | 0.25 | 25                    | 25                       | 100% | (GA)₈ACC         | P8     |
| 5.62 | 31  | 9.30 | 0.30 | 31                    | 31                       | 100% | (GA)₈ATC         | P9     |
| 7.46 | 17.05 | 8.18 | 0.48 | 19                    | 18                       | 94.73% | (ATG)₅       | P10    |
| 4.72 | 23  | 7.59 | 0.33 | 23                    | 23                       | 100% | (CT)₈RA          | MM14   |
| 9.88 | 31.03 | 13.03 | 0.42 | 33                    | 32                       | 96.96% | (AC)₈YT     | MM15   |
| 8.5  | 17.05 | 8.35 | 0.49 | 19                    | 18                       | 94.73% | (AC)₈YA       | MM16   |
| -    | -   | -   | -   | 243                   | 239                      | -    |                  |        |
| 5.63 | 23.52 | 7.95 | 0.35 | 24.3                  | 23.9                     | 98.3% | Mean            |        |

PPB: Percentage of Polymorphic Bands; PIC: Polymorphism Information Content; MI: Marker Index; EMR: Effective Multiplex Ratio; RP: Resolving Power.

MI ranged from 4.64 to 13.03. The highest MI parameter were determined for ISSR primers P9 and MM16 which amplified more bands (32 and 31, respectively). EMR index depends on the percentage of
polymorphic bands and number of polymorphic bands. This feature marker varied from 14.06 to 31.03 with a mean of 23.52 (Table 2).

Three ISSR markers had higher resolving power (RP) values (9.88, 8.5 and 7.46 for MM15, MM16 and P10, respectively) and could be useful informative markers to identify DH lines (Table 2). This index varied from 3.36 to 9.88 with an average of 5.63 (Table 2).

**Genetic relationships and cluster analysis**

Jaccard's similarity coefficients using ISSR data for camelina doubled haploid lines were calculated. Genetic similarity ranged between 0.0.05 and 0.85 (Due to the large matrix, data not shown). The minimum similarity (0.05) was recorded between lines 27 and 45. These two DH lines obtained from separate crosses with different parents (Table 1). Doubled haploid lines 103 and 104 had the most similarity coefficient (0.85).

To investigate genetic relationships among camelina DH lines and to classify the lines, cluster analysis was performed using the Jaccard's similarity coefficients matrix. Various clustering methods were tested and finally Complete method was selected as the most appropriate due to the lowest rate of chaining and a considerable cophenetic correlation coefficient of $r = 0.89$. The dendrogram grouped the camelina DH lines into five clusters (Figure 1):

- **Group I**: DH Lines 1, 6, 9, 12, 13, 21, 23, 24, 40, 7.
- **Group II**: DH lines 2, 3, 103, 104, 109, 64, 65, 72, 74, 81, 92, 98, 95, 96, 97, 105, 85, 100, 86, 88, 93, 94, 26, 41, 55, 36, 44, 68, 69, 75, 82, 83, 43, 53, 51, 54, 45, 52, 50, 76, 77, 78, 79, 28, 33, 35, 32, 59, 60, 62, 61. **Group III**: DH lines 4, 25, 56, 115, 15, 87, 57, 16, 80.
- **Group IV**: DH lines 5, 10, 14, 67, 107.
- **Group V**: DH lines 27, 37, 42, 84, 101, 102.

**Principal coordinates analysis based on ISSR markers**

Principal coordinate analyses (PCoA) was carried out using the marker system (Figure 2). The Latent roots (Eigen values) calculated by ISSRs data showed that the first five principal coordinates accounted for 25.27, 3.61, 2.74, 2.34, and 2.10% of total variability, respectively (36.06% cumulatively). PCoA classified the 81 camelina DH lines into six distinct groups. Most of the lines that came together in the scatter plot were in the same cluster dendrogram (Figure 2).

**Discussion**

In the presented research, the genetic variation of Camelina doubled haploid (DH) lines was assessed by ISSR markers. The set of 81 DH lines, obtaining from fifteen different crosses, possessed the parents from The Former Soviet Union and some parts of Europe.
The efficiency of the used markers was assessed by calculating their PIC parameter. The high average polymorphism (98.3%) among camelina DH lines and mean PIC value of 0.35 (Diversity Index = 0.35) in our study indicates considerable genetic variation among the lines and high efficiency of the ISSR markers to distinguish the variation. An average polymorphism information content, i.e. Diversity Index (DI) = 0.29 indicating moderate genetic variation of *Camelina sativa* spring accessions using SNPs was reported by Luo et al. [17]. Najaphy et al. [11] detected 80.2% polymorphism among 30 wheat accessions and an average PIC value of 0.22 using ISSR markers. Rostami Ahmadvandi et al. [18] recorded a mean PIC value of 0.39 for ISSR markers in cumin.

EMR based on the product of the number of polymorphic bands and the fraction of polymorphic bands; and MI index as the product of EMR and PIC, were the two other parameters recorded for the marker system. The primers that produced more polymorphic bands had a higher EMR record. On the other hand, most of the primers with higher PIC showed higher MI value (Table 2). Different reports have been used MI and EMR parameters to assess the discriminatory characteristic of molecular markers in some plants such as wheat [11], apricot [14] and Thyme [19].

The ability of primers to identify the DH lines was also assessed by determining resolving power. This index ranged between 3.36 and 9.88 with a mean of 5.63 (Table 2). Three ISSR markers: M15, MM16 and P10 had higher resolving power (RP) values 9.88, 8.5 and 7.46, respectively) and could be useful informative markers to distinguish DH lines (Table 2). RP parameter presents a partly precise estimate of the number of accessions that can be distinguished using a marker. The index may provide no knowledge about the ability of a marker to show the phylogenetic relationships of the accessions [20]. Some reports while calculating the ISSR parameters in the study of different crops genetic diversity, introduced the resolution power and marker index as indicator parameters for informative primers [11, 19].

Genetic similarity matrix of camelina doubled haploid lines based on Jaccard’s similarity coefficients using ISSR data was calculated (Table 1). The most similarity coefficient (0.85) was found between lines 103 and 104. Interestingly, the two DH lines were also originated from different crosses by the parents from different countries. Line 103 had been obtained from the cross between Svalöf (originated from Sweden) and Ukrajinskij (from the Former Soviet Union). DH line 104 originated from a cross between Przybrodzka (from Poland) and Hoga (from Denmark).

Cluster analysis using Complete method depicted genetic relationships and classified the lines into five groups. The clustering method indicated relatively clear configuration from the geographic distribution patterns of the parents of the doubled haploid lines. Nevertheless, there is a mixture of DH lines with different parents especially in Group II (Fig. 1, Table 1). The phenomenon may related to genetic interchange between the areas where are placed near to or overlapping each other in the Former Soviet Union and some parts of Europe. The relationships between geographic distribution and genetic variation has been shown in some other studies [11, 17, 19, 21, 22]. Ghamkhar et al. [23] reported high genetic variation of fifty three camelina genotypes from Ukraine.
Principal coordinate analysis (PCoA) was performed on ISSR data to develop a predictive function using independent new variables, based on the primary correlated variables and to compare with clustering method. By the analysis, six distinct groups were revealed. The output of grouping DH lines obtained from PCoA was partly comparable and consistent with the cluster analysis (Fig. 2). It is noteworthy that most of the camelina doubled haploid lines that occupied positions together in the scatter plot of PCoA, were classified in the same grouping dendrogram.

Doubled haploids is one of the applied mapping population that can facilitate QTL mapping and molecular breeding programs. The identification of genetic diversity and relationships among *Camelina sativa* doubled haploid lines is useful in breeding efforts. In our investigation, inter simple sequence repeats (ISSR) markers were employed to explore the genetic variation of *C. sativa* doubled haploid lines. The ISSRs generated high polymorphism for assessing genetic diversity and can be used in mapping QTLs and distinctive fingerprinting investigations of camelina accessions. Our results showed that the *C. sativa* DH lines were genetically diverse. By the ISSRs analysis, clustering of the lines according to the geographic distribution patterns of the parents of the doubled haploid lines was depicted. PIC, MI and RP are suggested as maker features to identify informative ISSRs. The molecular diversity evaluated in this research, along with biochemical traits of camelina can improve classical and molecular breeding programs of the crop.

**Declarations**

Mozafar Sadeghikian: Investigation, Data analysis, Writing – review & editing.

Abdollah Najaphy: Supervision, Investigation, Writing – review & editing.

Danial Kahrizi: Supervision, Investigation, Writing – review & editing.

Hossein Rostami-Ahmadvandi: Co-supervision, Writing – review & editing.

All authors have read and agreed to the published version of the manuscript.

**Compliance with Ethical Standards**

**Conflict of interest**

All the authors declare that they have no conflicts of interest.

**References**

1. Shonnard DR, Williams L, Kalnes TN (2010) Camelina-derived jet fuel and diesel: Sustainable advanced biofuels. Environ Prog Sustainable Energy 29: 382-392

2. McVay KA (2008) Camelina production in Montana. Copyright © 2008 MSU Extension.
3. Gehringer A (2010) Development of camelina (Camelina sativa Crtz.) genotypes and winter rapeseed (Brassica napus L.) hybrids for marginal locations. Doctoral dissertation, Justus Liebig University, Giessen.

4. Kahrizi D, Rostami-Ahmadvandi H, Akbarabadi A (2015) Feasibility cultivation of camelina (Camelina sativa) as medicinal-oil plant in rain fed conditions in Kermanshah-Iran's first report. J Medicin Plants By-products 2: 215-218

5. Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, Clarke C, Higgins EE, Huebert T, Sharpe AG, Parkin IA (2014) The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. Nat Commun 5: 3706

6. Gehringer A, Friedt W, Luhs W, Snowdon, RJ (2006) Genetic mapping of agronomic traits in false flax (Camelina sativa subsp. sativa). Genome 49: 1555-1563

7. Singh R, Bollina V, Higgins EE, Clarke WE, Eynck C, Sidebottom C, Gugel R, Snowdon R, Parkin IA (2015) Single nucleotide polymorphism identification and genotyping in Camelina sativa. Mol Breed 35: 1-13

8. Vollmann J, Grausgruber H, Stift G, Dryzhyruk V, Lelley T (2005) Genetic diversity in camelina germplasm as revealed by seed quality characteristics and RAPD polymorphism. Plant Breed 124: 446–453

9. Vollmann J, Eynck C (2015) Camelina as a sustainable oilseed crop: Contributions of plant breeding and genetic engineering. Biotechnol J 10: 525-535

10. Soorni J, Kazemitabar SK, Kahrizi D, Dehestani A, Bagheri N (2017) Screening of camelina (Camelina sativa L.) doubled haploid lines for freezing tolerance in the seedling stage. Genetika 49: 173-181

11. Najaphy A, Ashrafi Parchin R, Farshadfar E (2011) Evaluation of genetic diversity in wheat cultivars and breeding lines using inter simple sequence repeat markers. Biotechnol Biotechnol Equip 4: 2634-2638

12. Williams JG, Kubelik KAR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18: 6531- 6553

13. Mousapour Gorji A, Poczai P, Polgar Z, Taller J (2011) Efficiency of arbitrarily amplified dominant markers (SCOT, ISSR and RAPD) for diagnostic fingerprinting in tetraploid potato. Am J Pot Res 88: 226–237

14. Kumar M, Mishra GP, Singh R, Kumar J, Naik PK, Singh, SB (2009) Correspondence of ISSR and RAPD markers for comparative analysis of genetic diversity among different apricot genotypes from cold arid deserts of trans-Himalayas. Physiol Mol Biol Plants 15: 225-236

15. Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingeys S, Rafalsky A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2: 225-238

16. Altıntas S, Toklu F, Kafkas S, Kilian B, Brandolini A, OZkan H (2008) Estimating genetic diversity in durum and bread wheat cultivars from Turkey using AFLP and SAMPL markers. Plant Breed 127: 9-14

17. Luo Z, Brock J, Dyer JM, Kutchan T, Schachtman D., Augustin M, Ge Y, Fahlgren N, Abdel-Haleem H (2019) Genetic diversity and population structure of a Camelina sativa spring panel. Front Plant Sci 10: 1-12
18. Rostami-Ahmadvandi H, Cheghamirza K, Kahrizi D, Bahraminejad, S (2013) Comparison of morpho-agronomic traits versus RAPD and ISSR markers in order to evaluate genetic diversity among *Cuminum cyminum* L. accessions. Aust J Crop Sci 7: 361-367

19. Yousefi V, Najaphy A, Zebarjadi A, Safari H (2015) Molecular characterization of thymus species using ISSR markers. J Anim Plant Sci 25: 1087-1094

20. Prevost A, Wilkinson MJ (1999) A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theor Appl Genet 98: 07-112

21. Fracaro F, Echeverrigaray S (2006) Genetic variability in Hesperozygis ringens Benth. (Lamiaceae), an endangered aromatic and medicinal plant of Southern Brazil. Biochem Genet 44: 479-490

22. Agostini G, Echeverrigaray S, Souza-Chies TT (2008) Genetic relationships among South American species of *Cunila* D. Royenex L. based on ISSR. Plant Sys Evol 274: 135-141

23. Ghamkhar K, Croser J, Aryamanesh N, Campbell M, Kon’kova N, Francis C (2010) Camelina (*Camelina sativa* L.) Crantz as an alternative oilseed: molecular and ecogeographic analyses. Genome 53: 558–567

**Figures**
Figure 1

Complete Dendrogram of 81 Camelina doubled haploid lines based on ISSR marker data.
Figure 2

Plot of camelina double haploid lines by principal coordinate analysis using the ISSR data.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Coverletter.docx