Assessment of genetic variation in *Linum* L. using SPAR markers

Zahra Noormohammadi*,
Mina Sakhaee¹,
Masoud Sheidai²,
Seyed Mehdi Talebi

¹ Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran
² Faculty of Biological Sciences, Shahid Beheshti University, Tehran, Iran
³ Department of Biology, Faculty of Sciences, Arak University, Arak

Flax (*Linum usitatissimum* subsp. *usitatissimum*) is an important crop plant cultivated both for its fiber and seed oil content. Uniform cultivation and selection practice may lead to genetic erosion and reduced genetic variability of Linum cultivars, which in turn makes the crop vulnerable to diseases and pathogens. Therefore we performed a genetic diversity analysis of few populations including both *Linum usitatissimum* subsp. *usitatissimum* L. and *L. Bienne* L. We used a combination of single primer amplification reaction markers, including RAPD, ISSR, and RAMPO molecular markers for genetic variability study. In general, all three molecular markers revealed a low degree of genetic variability in the materials studied; however, *L. bienne* had a higher level of genetic variability. Clustering and network analyses grouped the studied plants inter-mixed and did not separate the studied species or populations. STRUCTURE plot and K-Means clustering also identified these populations as genetically homogenous. These results indicate the need for introducing new linseed cultivars into the country for future hybridization and breeding programmes.

**Keywords:** genetic variability, K-Means clustering, Linum, STRUCTURE analysis, SPAR

**INTRODUCTION**

Flax (*Linum usitatissimum* subsp. *usitatissimum*) is a diploid, annual plant species. Since flax, and in particular fiber flax, has been such an important cultivated crop, it is a matter of great importance to conserve the genetic material of flax as widely as possible for future utilization in breeding. Presently, the largest producers of fiber flax are China, Russia, Belarus, and Ukraine, while oilseed production is important in India, Canada, China, United States, Germany, Argentina, United Kingdom, and Spain (Žiarovská et al., 2012).

Linum appears to have arisen in Eurasia, from which it spread to Africa, North America, South America, and Australasia (McDill et al., 2009).

The cultivation of wild flax dates back to around 9,000 B.C.E., as evidenced by the oldest archaeological flax seeds from excavation sites in Syria. *Linum bienne* Mill. (= *L. angustifolium* Huds.) is perhaps the oldest flax wild form cultivated. Its cultivation for fiber and seeds is believed to have led to the development of *L. usitatissimum* L., the modern cultivated flax, one of the early domesticated plants (McDill et al., 2009).

* Corresponding author. E-mail: marjannm@yahoo.com, z-nouri@srbiau.ac.ir
Divergent selection applied over thousands of years has resulted in fiber and linseed types of Linum which are the same species but differ considerably in morphology, anatomy, physiology, and agronomic performance (Soto-Cerda et al., 2013).

Linum species, such as the red-flowered *L. grandiflorum* Desf., the blue-flowered *L. Perenne* L., and the yellow-flowered *L. Flavum* L., are also of ornamental value. In recent years, ligansand α-linolenic acids of *Linum* have been explored for their usefulness in the treatment of cardiovascular diseases and cancer, particularly breast cancer (Sheidai et al., 2014). These medicinal applications have stimulated renewed interest in the studies related to the genetic and systematic relationships among Linum species.

Genetic diversity analysis is an important task in plant breeding as diversity in plant genetic resources (PGR) provides an opportunity for plant breeders to develop new and improved cultivars with desirable characteristics (Govindaraj et al., 2015). Genetic diversity helps organisms cope with current environmental variability (Freeland et al., 2011). Genetic diversity also reduces the incidence of unfavourable inherited traits due to inbreeding and produces a diverse array of genotypes that can resist diseases and pathogens (Freeland et al., 2011).

Population genetic studies can picture genetic variability of the species, the gene flow, inbreeding versus outbreeding, effective population size, and population bottleneck. These data are helpful in choosing effective management in conservative strategies, shed light on the current infra-specific taxonomic forms, and suggest a proper hybridization plan (Sheidai et al., 2014). Therefore, we performed a population genetic study of three available populations in either of *Linum usitatissimum* L. (cultivated flax) and *Linum austriacum* L. to provide data on the genetic structure and genetic variability in these species.

Recently, interesting studies have been performed in our country on population genetic structure of Linum species, both cultivated and wild (see, for example, Sheidai et al., 2014; Afshar et al., 2015; Habibollahi et al., 2016 a, b). These authors used SSR (simple sequence repeats), EST-SSR (expressed sequence tags), ISSR (inter-simple sequence repeats) molecular markers and reported a low degree of genetic variability in the studied species. The single primer amplification region (SPAR) markers such as RAMPO RAPD and ISSR have been used in genetic variation studies. RAMPO (random amplified microsatellite polymorphism) molecular markers are efficient discrimination markers that have been recently used in crop genetic variability assessment studies (see for example, Chatti et al., 2008; Tamarzizt et al., 2015). In RAMPO, two types of primers are in use. Initially, DNA amplification was conducted using universal decamer oligonucleotides (RAPDs), followed by using primers that are complementary to simple sequence repeats used to perform RAMPO assays. Finally, this method consists of the combination of these two PCR steps (Tamarzizt et al., 2015).

**MATERIALS AND METHODS**

**Plant materials**

In the present study, 40 plants were collected from four geographical populations of Iran. Three populations of *Linum usitatissimum* L. (1 = Uromieh (West-Azarbayejan), 2 = Shiraz (Fars), and 3 = Saveh (Markazi), respectively), and one population of *Linum bienne* L. were studied.

**DNA extraction and molecular markers assay**

Fresh leaves were used randomly from 5–10 plants in each of the studied populations. CTAB-activated charcoal protocol was used to extract genomic DNA based on Krizman et al. (2006) protocol. Nine RAPD and 7 ISSR as well as 10 RAMPO loci were used (Table 1). PCR reactions were carried out in a 25 μl volume containing 10 mM Tris-HCl buffer at pH 8; 50 mM KCl; 1.5 mM MgCl2; 0.2 mM of each dNTP (Bioron, Germany); 0.2 μM of a single primer (used 2 primers used for RAMPO-PCR); 20 ng genomic DNA and 1 U of *Taq* DNA polymerase (Bioron, Germany). The amplification reactions were performed in Techne thermocycler
Genetic variation of *Linum* L. ... (Germany) with the following program: 5 min initial denaturation step 94 °C, followed by 40 cycles of 30 s at 94 °C; 1 min at 37, 50 and 38 °C for RAPD, ISSR and RAMPO respectively and 2 min at 72 °C. The reaction was completed by final extension step of 5 min at 72 °C. The amplification products were observed by running on 1% agarose gel, followed by GelRed™ Nucleic Acid Gel Staining. The fragment size was estimated by using a 100 bp molecular size ladder (Fermentas, Germany).

**Data analyses**

RAPD, ISSR and RAMPO bands obtained were coded as binary characters (presence = 1, absence = 0) and used for genetic diversity analysis. Parameter like Nei's gene diversity (H), Shannon information index (I), number of effective alleles, and percentage of polymorphism were determined (Weising, 2005; Freeland et al., 2011).

Nei's genetic distance among populations was used for Neighbor Joining (NJ) clustering and networking (Freeland et al., 2011; Huson, Bryant, 2006). These analyses were done by PAST v2.17 (Hamre et al., 2012), DARwin v5 (2012) and SplitsTree4 v4.13.1 (2013) software.

AMOVA (Analysis of molecular variance) test (with 1000 permutations) as implemented in GenAlex 6.4 (Peakall, Smouse, 2006), and Nei's Gst analysis as implemented in GenoDive v2 (2013) (Meirmans, Van Tienderen, 2004) were used to show the genetic difference of the populations. Moreover, genetic differentiation of populations was studied by G^ST^ = standardized measure of genetic differentiation (Hedrick, 2005), and D^est^ = Jost measure of differentiation (Jost, 2008).

The genetic structure of the populations was studied by Bayesian-based model STRUCTURE analysis (Pritchard et al., 2000), and maximum likelihood-based method of K-Means clustering of GenoDive v2 (2013). For STRUCTURE analysis, data were scored as dominant markers (Falush et al., 2007). In K-Means clustering, two summary statistics, pseudo-F and Bayesian Information Criterion (BIC), provide the best fit for k (Meirmans, 2012).

**RESULTS**

**Genetic diversity of the populations**

Genetic diversity parameters determined in the studied populations for all three molecular markers are presented in Table 2. In RAPD markers, population No. 4 (*L. biennae*) showed the highest degree of genetic polymorphism (30.94%), while population No. 1 (*Uromieh*) of *Linum usitatissimum* L. showed the lowest degree of genetic diversity (26.93%).

Abbreviations: Na = Mean number of alleles, Ne = Number of effective alleles, I = Shannon Information Index, He = Gene diversity.
Table 2. Genetic diversity parameters in Linum usitatissimum, and L. bienne

<table>
<thead>
<tr>
<th>Population</th>
<th>Na</th>
<th>Ne</th>
<th>I</th>
<th>He</th>
<th>P%</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAPD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pop1</td>
<td>0.539</td>
<td>1.039</td>
<td>0.064</td>
<td>0.033</td>
<td>26.93</td>
</tr>
<tr>
<td>Pop2</td>
<td>0.580</td>
<td>1.040</td>
<td>0.067</td>
<td>0.034</td>
<td>29.02</td>
</tr>
<tr>
<td>Pop3</td>
<td>0.577</td>
<td>1.037</td>
<td>0.064</td>
<td>0.032</td>
<td>28.87</td>
</tr>
<tr>
<td>Pop4</td>
<td>0.619</td>
<td>1.045</td>
<td>0.074</td>
<td>0.038</td>
<td>30.94</td>
</tr>
<tr>
<td>ISSR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pop1</td>
<td>0.590</td>
<td>1.046</td>
<td>0.074</td>
<td>0.039</td>
<td>29.50</td>
</tr>
<tr>
<td>Pop2</td>
<td>0.581</td>
<td>1.046</td>
<td>0.073</td>
<td>0.038</td>
<td>29.05</td>
</tr>
<tr>
<td>Pop3</td>
<td>0.529</td>
<td>1.036</td>
<td>0.061</td>
<td>0.031</td>
<td>26.48</td>
</tr>
<tr>
<td>Pop4</td>
<td>0.919</td>
<td>1.066</td>
<td>0.109</td>
<td>0.056</td>
<td>45.95</td>
</tr>
<tr>
<td>RAMPO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pop1</td>
<td>0.942</td>
<td>1.065</td>
<td>0.110</td>
<td>0.056</td>
<td>47.00</td>
</tr>
<tr>
<td>Pop2</td>
<td>0.734</td>
<td>1.052</td>
<td>0.086</td>
<td>0.044</td>
<td>36.68</td>
</tr>
<tr>
<td>Pop3</td>
<td>0.479</td>
<td>1.052</td>
<td>0.076</td>
<td>0.042</td>
<td>23.95</td>
</tr>
</tbody>
</table>

and P% = polymorphism. (Populations 1–4 are: Urmia, Shiraz, Saveh (Linum usitatissimum), and Linum bienne).

In ISSRs, population No. 4 (L. bienne) had a low degree of genetic polymorphism (45.95%), while population No. 3 (Saveh) of Linum usitatissimum L. showed the lowest degree of genetic diversity (26.93%).

In the case of RAMPOs, we could not get a result for L. bienne, therefore a comparison was made between Linum usitatissimum populations. Population No. 1 (Uromieh) showed the highest genetic diversity (47.00%, Table 2).

Population genetic differentiation

AMOVA and Gs analysis revealed a significant difference among the studied populations for all three molecular markers examined. For example in RAPDs, AMOVA produced PhiPT = 0.04, P = 0.01.

AMOVA also revealed that 96% of total genetic variability was due to diversity within a population and only 4% was due to genetic differentiation among populations. These results indicated the presence of almost a homogeneous genetic group.

Genetic affinity of the populations

Since the NJ tree and Net network analyses produced similar results for RAPDs and ISSRs, only ISSR Network is presented and discussed (Fig. 1). In both molecular markers we obtained a mixture of genotypes and no single population was

![Fig. 1. Median-Joining network of ISSR data in L. usitatissimum and Linum bienne populations](image)
separated from the others. The samples of both Linum species were placed inter-mixed. This means that the two species are closely related and are not very divergent in parts of their genome that are revealed by RAPD and ISSR molecular markers.

In case of RAMPOs, in which only the populations of *L. usitatissimum* were studied, the mixture of samples from all three populations occurred again (Fig. 2).

**Population genetic structure**

K-Means clustering produced $k = 2$ according to pseudo-F and $k = 1$ according to BIC. Therefore, we have one or two distinct genetic groups. Similar result was obtained by STRUCTURE analysis (Fig. 3).

For example, the STRUCTURE plot of ISSR data revealed very similar genetic content in both studied species (similarly coloured segments). The only difference was the proportion

![Median-Joining network of RAMPO data in *L. usitatissimum* populations](image1.png)

**Fig. 2.** Median-Joining network of RAMPO data in *L. usitatissimum* populations

![STRUCTURE plot of ISSR data in *L. usitatissimum* (populations 1–3) and *Linum bienne* (population 4) populations](image2.png)

**Fig. 3.** STRUCTURE plot of ISSR data in *L. usitatissimum* (populations 1–3) and *Linum bienne* (population 4) populations
of differently coloured segments in these populations. In this regard *Linum bienne* differed from *L. usitatissimum* populations. This high degree of genetic similarity between the two species and also among different populations in *L. usitatissimum* indicated either the presence of ancestral shared alleles or an ongoing gene flow among these populations.

Based on the likelihood method, the population assignment test performed on the studied species is presented in Table 3. Based on the ratio of likelihoods, 16 out of 40 plants were inferred to be from another population and were genetically more similar to them. For example, plants No. 5, 8, 9, and 10 of population 1 were inferred to be from populations 2 and 3. (Mostly they were genetically similar to population 2). Similarly, population 2 had some plants that were inferred to be from population 1. The same holds true for populations 3 and 1, as well as populations 4 and 1. These results are in agreement with STRUCTURE analysis in showing a high degree of genetic admixture among the studied species and populations.

Abbreviations: Lik_max = Maximum likelihood, Lik_home = Home likelihood, and Lik_ratio = Ratio of Home likelihood/ Maximum likelihood.

**DISCUSSION**

Determination of genetic variability both within and among populations is one of the major findings in population genetic analysis. Genetic diversity is of fundamental importance in the continuity of a species as it provides the necessary adaptation to the prevailing biotic and abiotic environmental conditions, and changes in the genetic composition may help to overcome the environmental changes (Sheidai et al., 2014). In general we found a low degree of genetic diversity in the studied *Linum* species and populations.

<table>
<thead>
<tr>
<th>Individual</th>
<th>Current</th>
<th>Inferred</th>
<th>Lik_max</th>
<th>Lik_home</th>
<th>Lik_ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Pop001</td>
<td>Pop003</td>
<td>-36.015</td>
<td>-42.928</td>
<td>13.826</td>
</tr>
<tr>
<td>8</td>
<td>Pop001</td>
<td>Pop002</td>
<td>-31.021</td>
<td>-49.208</td>
<td>36.374</td>
</tr>
<tr>
<td>9</td>
<td>Pop001</td>
<td>Pop002</td>
<td>-45.432</td>
<td>-48.109</td>
<td>5.353</td>
</tr>
<tr>
<td>10</td>
<td>Pop001</td>
<td>Pop002</td>
<td>-32.597</td>
<td>-42.928</td>
<td>20.663</td>
</tr>
<tr>
<td>14</td>
<td>Pop002</td>
<td>Pop001</td>
<td>-54.714</td>
<td>-64.842</td>
<td>20.257</td>
</tr>
<tr>
<td>18</td>
<td>Pop002</td>
<td>Pop001</td>
<td>-25.239</td>
<td>-28.067</td>
<td>5.657</td>
</tr>
<tr>
<td>20</td>
<td>Pop002</td>
<td>Pop001</td>
<td>-58.709</td>
<td>-60.439</td>
<td>3.462</td>
</tr>
<tr>
<td>23</td>
<td>Pop003</td>
<td>Pop001</td>
<td>-85.175</td>
<td>-86.646</td>
<td>2.942</td>
</tr>
<tr>
<td>24</td>
<td>Pop003</td>
<td>Pop004</td>
<td>-83.055</td>
<td>-83.602</td>
<td>1.094</td>
</tr>
<tr>
<td>25</td>
<td>Pop003</td>
<td>Pop001</td>
<td>-18.611</td>
<td>-24.67</td>
<td>12.118</td>
</tr>
<tr>
<td>28</td>
<td>Pop003</td>
<td>Pop001</td>
<td>-15.603</td>
<td>-19.489</td>
<td>7.773</td>
</tr>
<tr>
<td>29</td>
<td>Pop003</td>
<td>Pop002</td>
<td>-17.58</td>
<td>-19.489</td>
<td>3.819</td>
</tr>
<tr>
<td>31</td>
<td>Pop004</td>
<td>Pop001</td>
<td>-58.76</td>
<td>-59.146</td>
<td>0.771</td>
</tr>
<tr>
<td>37</td>
<td>Pop004</td>
<td>Pop001</td>
<td>-101.594</td>
<td>-102.787</td>
<td>2.387</td>
</tr>
<tr>
<td>38</td>
<td>Pop004</td>
<td>Pop001</td>
<td>-124.425</td>
<td>-126.303</td>
<td>3.754</td>
</tr>
<tr>
<td>40</td>
<td>Pop004</td>
<td>Pop003</td>
<td>-61.671</td>
<td>-74.541</td>
<td>25.74</td>
</tr>
</tbody>
</table>
STRUCTURE plot and genetic diversity analysis revealed lesser genetic variability among cultivated flax and a higher degree of genetic variability within its wild relatives that is in agreement with other studies using SSR, EST-SSR, ISSR, and retrotransposon-based markers (IRAP) (Fu et al., 2002; Smy‘kal et al., 2011; Žiarovská et al., 2012; Habibollahi et al., 2015).

In order to maintain and exploit the genetic resources efficiently, an understanding of the amount and distribution of the genetic variation within and among accessions in a collection is required (Žiarovská et al., 2012). Molecular markers are highly useful both to identification of potentially novel genotypes among the many flax accessions and to the assessment of the genetic diversity for both germplasm management and core collection assembly (Smy‘kal et al., 2011).

Different measures are used to quantify the distribution of genetic variation among individuals within populations, and among populations. These measures describe how genetic variation is partitioned within species. These include Wright’s inbreeding coefficient (Fst) and Nei’s coefficient of gene variation (Gst, Falk et al., 2001).

Values of Fst and Gst vary from 0 to 1. GST in particular has a number of useful properties: it can be used for one or many loci, mutation rates do not alter the statistic significantly, the exact number of populations need not be specified, and the statistic is relatively responsive to changes in allele frequencies in time. Although they have important conceptual differences, in practice Fst and Gst are used in similar fashion as indices of the genetic difference among populations (Crow, 1986).

In species with low Gst (approaching 0), the majority of variation is found within populations; individuals within populations are likely to be genetically different, but each population contains the same complement of alleles in similar frequencies where Gst is high (approaching 1). Individuals within a population are relatively similar but populations are significantly different (Freeland et al., 2011).

In the present investigation we used both measures of genetic differentiation. They revealed a low value for both estimated pair-wise Fst and Gst between populations. These low values indicate either a high degree of gene flow among the studied populations or the presence of many common shared ancestral alleles in them (Sheidai et al., 2014, 2015).

The distribution of genetic variation within and among species is strongly linked to life-history traits, particularly to the dispersal and reproductive mode (Hamrick et al. 1991). Plant species that disperse their genes widely and frequently will tend, other things being equal, to have a lower Gst value (i.e., populations will be more similar). Linum is a predominantly selfing species therefore we cannot expect extensive open pollination and wide dispersal in this species. Moreover, the plants studied are cultivated in different cities with no chance for gene exchange.

K-Means clustering also showed the presence of one or two genetic groups and the STRUCTURE plot revealed very close genetic similarities between the studied species and populations due to the common shared alleles. Moreover, population assignment test revealed a high degree of inter-population genomic admixture. We therefore consider the ancestral conserved common alleles responsible for the high degree of genetic similarity observed in these species and populations. These results are in agreement with the study of Soto-Cerda et al. (2013), who investigated 407 flax accessions by using 448 microsatellite loci. These accessions were grouped in six major clusters, which showed weak general population differentiation (Fst = 0.094) and weak pairwise population differentiation among sub-groups.

Therefore, the present study supported a close genetic affinity between L. usitatissimum and Linum bienne, and revealed the presence of similar genetic content in the populations of L. usitatissimum. This means that we need to broaden the genetic pool of this species in Iran if we want to obtain superior cultivars with suitable genetic content and agronomic traits. We should introduce some foreign cultivars into the country.
and perform genetic and phenotypic diversity analyses on them to select the most divergent cultivars for future hybridization and selection.

Received 10 October 2016
Accepted 19 January 2017

References


