Genetic characterization of a diversity panel, selected from IPK linseed (*Linum usitatissimum* L.) world Collection

**Ehsan Ataii 1, Aghafakhr Mirlohi 1\*, Mohammad R. Sabzalian 1, Negar Sharif Moghadam 1, Nafiseh Sadri 1, Parisa Jafari1, Aliakbar Mohammadi Mirik2**

1 Department of Agronomy and Plant Breeding, College of Agriculture, Isfahan University of Technology, Isfahan 84156 83111, Iran.

2Department of Plant Genetics and Production- Faculty of Agriculture - Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran

\* Corresponding author: Department of Agronomy and Plant Breeding, College of Agriculture, Isfahan University of Technology,

Isfahan 84156 83111, Iran

Telephone: +98313-3913450

Fax number: +98313-3912254

E-mail: mirlohi@iut.ac.ir

**Running title:** Genetic diversity in the IPK linseed world collection

**Abstract**

Despite all advances in breeding techniques, the fundamental role of studying and identifying new genetic material remains undeniable. A globally distributed diversity panel of linseed randomly selected from IPK gene bank (Gatersleben, Germany) plus a few other genotypes were field evaluated for yield per plant, plant height, number of branches and capsules per plant, capsule diameter, capsule seed number, 1000-seed weight, seed length and width, seed oil and protein over two years to investigate (i) existing genetic variability, (ii) possible correlation between traits and (iii) potential elite lines. Results revealed the existence of great variability in this diversity panel mainly attributable to the genetic factors (72.6%) rather than geographical origin (6.4%) or experimental years (20.9%). Expected genetic advance showed the potential selection gains were expected to be > 10% for eight traits if 5% of the accessions were selected. Among the traits, seed-size features had relatively higher broad sense heritability and showed the strongest correlation with seed yield. Five elite lines were identified, among which an Indian landrace had the highest seed yield potential.

***Keywords***: Genetic and phenotypic diversity; Agronomic traits; Seed quality traits; Genetic and phenotypic correlation; Diversity panel; Linseed; Flax

**1. Introduction**

Linseed (*Linum usitatisimum* L.) belongs to the Linaceae family, and its domestication in Near East dates back to more than 7000 years ago (Christopher A. Cullis, 2019). It is considered one of the founder crops due to its ancestral role in providing the raw material needed for human life (Chandrawati et al., 2017). Linseed has been anciently cultivated with two distinct purposes of seed oil and stem fiber (Christopher A. Cullis, 2019). Continuous manipulation to better achieve these two goals have led to the diversification of linseed into two main types: seed and fiber flax. The two are significantly different in terms of morphological traits and growth patterns. Generally, taller main stem with fewer branches are the main characteristics that distinguish the fiber type plants from the linseed types that produce more seeds (Diederichsen & Ulrich, 2009; Liu, Chen, Long, Shuai, & Long, 2011). However, based on a formal distinction of cultivated flax, four different convarieties exist for this plant: (I) genotypes with spontaneously opening capsules (convar. *crepitans*); (II) genotypes that are solely cultivated for stem fiber which have technical stem length over 70 cm with branches confined to the most upper stem part (convar. *elongatum*); (III) flax plants in this group grown only for seed use, considering their large seeds with 1000-seed weight over 9 g (convar. *mediterraneum*); and (IV) genotypes with intermediate characteristics which are used to produce both seed and fiber (convar. *usitatissimum*) (Diederichsen & Yong-Bi, 2006).

Linseed oil contains very high amounts of linolenic fatty acid (more than 50% of total fatty acids), which leads to high drying and oxidation rate; therefore, it is widely used in the paint, ink, putty, and varnish industries. Besides the historical utilization of flax fiber, in recent decades, the textile industry has also developed high-value products from flax fiber with applications in automotive, construction, biofuel, and pulp industries. As a result, industries currently dominate the world demand for flaxseed. Nevertheless, due to its exclusive nutrient contents including oil (almost 40%) protein (16 to 32%), dietary fiber (20 to 28%), and a considerable amount of lignan, recently, linseed is thriving in the human food and livestock industry too (Oomah, 2001; Herchi et al., 2012; You et al., 2017).

Currently, the most considerable amount of linseed cultivation in the world belongs to western Canada (seed), the cold regions of China (fiber and seed), north-central USA (seed), and Western Europe (fiber) (You et al., 2017). However, l, but still, finding varieties of linseed compatible with arid and semi-arid regions where almost half of the world agricultural lands and human population are located has become a breeding challenge (“2010–2020: UN Decade for Deserts and the Fight against Desertification,” n.d.). Understanding existing genetic variability in germplasm collections can significantly ease the reliable classification of genotypes and identification of the ones with potential utility in different breeding programs (Majidi & Zadhoush, 2014). Several studies have vastly investigated morphological and molecular genetic variability of important linseed traits in extensive world collections (Soto-Cerda, Diederichsen, Ragupathy, & Cloutier, 2013; You et al., 2017). However, depending on heritability, morpho phenological characteristics may differ among different populations of the same species under different ecological conditions (Talebi, Amini, Askary, Farahani, & Matsyura, 2020).

Studying various linseed collections under different ecological conditions can open the door to developing new varieties compatible with varied habitats. In this study, 120 globally distributed linseed accessions were phenotyped in an arid condition over two years. The objectives were to investigate: (1) the genetic diversity; (2) establishing a possible association between studied traits; and (3) to identify promising line(s) for arid conditions with yield stability over the years applicable to future studies and linseed breeding programs.

**2. Material and Methods**

***2.1. Plant material and experimental design***

The 120 linseed genotypes used in this study were diversity panel of 107 accessions randomly selected from IPK gene bank (Gatersleben, Germany) world collection, eight Iranian landraces, three Canadian breeding lines, a commercial cultivar (Flanders) and one landrace from India. All the accessions present in this diversity panel belonged to the intermediate flax group (*Linum usititassimum* L., convar. *usitatissimum*). The global geographical distribution of the genotypes included 14 regions of Australia (AUS), Russia (RUS), East Asia (EAS), South Asia (SAS), West Asia (WAS), East Africa (EAF), North Africa (NAF), Central Europe (CEU), Northern Europe (NEU), Southeast Europe (SEE), Southwest Europe (SWE), Western Europe (WEU), North America (NAM) and South America (SAM), which were containing 3, 4, 7, 6, 20, 5, 5, 22, 4, 4, 5, 8, 15, and 12 genotypes, respectively. For the purpose of this study, Russia was considered a separate geographical origin independent of Asia and Europe. Field evaluation of accessions was done in a randomized complete block design with two replications and over two years (February 29th to July 3rd in 2016 and March 5th to July 15th in 2017). Each year, accessions were planted in micro plots of 1.5 m length and 0.25 m width in two rows comprising about 300 individuals and were grown under a well-watered condition. All agronomic traits were evaluated on five single plants randomly selected from the center of micro plots (25 cm of each side was omitted to minimize the marginal effect) when genotypes reached physiological maturity stage. At the same phenological stage, total seeds of all plants in 1 m central part of micro plots were harvested and pooled to measure seed oil and crude protein contents in the laboratory.

The research farm at Isfahan University of Technology (51º 28' E Long., 32º 42' N Lat., 1624 m a.s.l.) is an arid region (Köppen Climate Classification System), with annual average temperature of 17 ºC and annual precipitation of 122 mm with no summer rain. The average monthly temperature in 2016 and 2017 varied from 5.57 ºC to 30.39 ºC and 5.54 ºC to 28.17 ºC, respectively. During the linseed growing season, the average air temperature and daily pan evaporation were 19.2 ºC and 7.67 mm/day in 2016 and 19.9 ºC and 8.68 mm/day in 2017. Also, the average relative humidity (RH) during the linseed growing season in 2016 and 2017 was 29.9 and 32.5%, respectively.

***2.2. Trait measurements***

A total of 11 traits included nine agronomic and two seed quality traits, were evaluated in the farm and laboratory in 2016 and 2017. Plant height (PLH) was measured at the maturity stage from the ground to the most upper part of the five sampled plant and the average of these measurements was recorded for each accession. Number of branches per plant (NBP) was recorded as the average number of lateral shoots counted on the main stem of five sampled plants. Similarly, number of capsules per plant (NCP) was determined by counting the total number of capsules on the five sampled plant and the average value was used for statistical analysis. Capsule diameter (CDI) was determined as the average diameter of 50 random capsules (10 random capsules from each of five sampled plants) using a caliper. Capsule seeds number (CSN) was recorded as the average seed number in 50 random capsules (10 random capsules from each of five sampled plants). Thousand seed weight (TSW) was measured by scaling 1000 harvested seeds of five sampled plants using a highly sensitive balance. Yield per plant (YPP) was considered the average weight of total harvested seeds on five randomly selected plants. Seed length (SLE) and seed width (SWI) were recorded as the length and width of 50 seeds randomly selected from total harvested seeds on five sampled plants. The seed oil content (OIL) was quantified utilizing petroleum ether, according to AOAC official method 945.16 (AOAC, 1980). The seed crude protein content (PRO) was measured on the remained meal from the oil extraction process performing Kjeldahl method based on AOAC official method 992.23 (AOAC, 1980).

***2.3. Data analysis and estimation of genetic parameters***

The following random linear model was used to analyze all data collected from field evaluation and laboratory measurements:

$$y\_{ij}=μ+G\_{i}+Y\_{j}+\left(GY\right)\_{ij}+ε\_{ij} , \left(i=1,2,…,g, j=1,2,…,y\right), $$

where, $y\_{ij}\~N\left(μ,σ\_{P}^{2}\right), G\_{i}\~N\left(0, σ\_{G}^{2}\right), Y\_{j}\~N\left(0,σ\_{Y}^{2}\right), (GY)\_{ij}\~N\left(0,σ\_{GY}^{2}\right)$ and $ε\_{ij}\~N\left(0,σ\_{e}^{2}\right)$. Based on this model $σ\_{P}^{2},σ\_{G}^{2},σ\_{Y}^{2},σ\_{GY}^{2}$ and $σ\_{e}^{2}$ represent variances of phenotype, genotype (G), Year (Y), G×Y and error, respectively.

 The multi-environment trial analysis in R (META-R) software was used to do a combined analysis of variance (ANOVA) for the collected data in two experimental years. By considering the power of META-R software for accurate computing of many statistical and genetic parameters in individual and combined environments, the phenotypic coefficient of variation ($\hat{PCV}$), genetic coefficient of variation ($\hat{GCV}$), and broad-sense heritability ($\hat{H}^{2}$) of each trait along with phenotypic and genetic correlations among traits as well as their statistical significance were all calculated using this software (Alvarado et al., 2020). To calculate the expected genetic advance for selection of studied traits (∆G, %) the equation $∆G=k\hat{σ}\_{G}{\sqrt{\hat{H}^{2}}}/{\overbar{x}}=k\hat{GCV}\sqrt{\hat{H}^{2}}$, was used. Here, *k* is considered the selection intensity and equal to 2.06, where 5% of the population with normal distribution were selected and, where $\overbar{x}$ representing the population mean of the studied traits (You et al., 2017). The JMP software was used for multivariate analysis, including clustering and factor analysis (FA), and for depicting bi-plots. FA was performed based on a principal component analysis on the studied traits, with an orthogonal rotational procedure (Varimax). The correlation matrix adequacy for FA was evaluated with Bartlett's test and Kaiser-Meyer-Olkin measure. The PROC VARCOMP was used to estimate the contribution of the experimental year and geographical origin of accessions to the variance of the studied traits (SAS, Cary, USA). Accordingly, a random model (y= geographical\_origin year) using the restricted maximum likelihood method (METHOD=REML) was applied to estimate variances attributable to the experimental year, geographical origin, and residual. The absolute values of each variance component were then converted to its proportion of the total variance. Cluster analysis was performed using the Ward method and based on Euclidean distances. The R (v2.5, http://cran.r-project.org/) package "*ggplot2*" was used to draw boxplots.

**3. Results**

***3.1. Genetic and Phenotypic Variation***

Combined ANOVA results revealed highly significant differences among accessions for all studied traits. Significant genotype × year interaction was observed for all the traits except for CDI, CSN, and SWI; though, the effect of year was only significant for NBP, NCP, and YPP (Table 1). As expected, the value of $\hat{GCV}$ was smaller than that of $\hat{PCV}$ for all the traits but for most traits the two values were close. Five traits that accounted for $\hat{PCV}$ and $\hat{GCV}$ values greater than 10% indicate their high genetic and phenotypic diversity. The average $\hat{PCV}$ for agronomic traits was 17.26%, while this amount was 6.62% for seed quality traits. Among agronomic traits, YPP, NBP, and NCP had the largest $\hat{PCV}$ values of 37.81, 30.34, and 27.00%, respectively. Neither of the two seed quality traits showed a considerable variation. According to expected genetic advance (∆G), the potential selection progress for eight traits estimated to be more than 10% if only five percent of individuals from this population were selected. This progress, in particular, will be high for YPP (35.12%), PLH (32.48%), NBP (31.89%), and TSW (25.26%) (Table 2).

 To determine the source of variability in this diversity panel of linseed, total phenotypic variations of all evaluated accessions were partitioned into components based on the experimental year, geographical region, and other factors for all the traits under study. The results showed that (Table 3), on average, the variance attributable to geographical origins (6.4%) and experimental years (20.9%) were considerably smaller than the variance caused mostly by genetic factors (72.6%). SWI and PLH were the only two traits that contributed more than 10% of variation within geographical origin. However, within each experimental year variation was mostly explained by TSW (~31%), YPP (~54%), NCP (~54.8%), and NBP (55.9%) (Figure 1).

***3.2. Broad-Sense Heritability***

The broad-sense heritability of a trait represents to what extent the environment and experimental error affect genotypes. This genetic measurement enables breeders to estimate the precision or repeatability of phenotypic selection in the breeding process (You, Booker, Duguid, Jia, & Cloutier, 2016). The average estimation of $\hat{H}^{2}$ for agronomic traits was 0.70, with the seed size-related traits showing relatively higher values. Grain quality traits with an average of 0.52 had a moderate $\hat{H}^{2}$. These calculated values of broad-sense heritability could be used to compare to which extent the environment affects different traits studied in this experiment. However, it is important to notice that these values were most likely overestimated because the present study was carried out across two years at one location.

***3.3. Phenotypic and genetic association among traits***

The phenotypic and genetic correlation (rp and rg, respectively) networks constructed by the pairwise associations between agronomic and seed quality traits are shown in Figures 2A and B, respectively. While it is almost unattainable to readily make a clear image of all positive and negative phenotypic and genetic correlations among all traits from the correlation matrix, this kind of visualization gives a comprehensive view of all associations between different pairs of traits in one glimpse. In this respect, each circle shows a trait, and the green and red lines represent the positive and negative correlations between them, respectively. Also the thickness of the lines represent the strength of the correlation between the attributes. As shown in Figure 2 and table 3, the sign for rg and rp are the same in almost all cases. However, rg was higher than rp in most cases regarding magnitude.

At first glance, PLH, which was the most relevant trait indicating the fiber production ability of the intermediate genotypes in this study, shows a significant poor and negative phenotypic correlation with most other traits (Figure 2A). However, the genetic correlation between PLH and every other trait except PRO is significant, negative, and moderate (Figure 2B and Table 3). Seed yield components traits in linseed could be abstractly divided into seed size-related traits (TSW, CDI, SLE, and SWI) and seed number-related traits (NBP, NCP, and CSN). Our results showed that among these seven traits, YPP had the strongest positive phenotypic association with NCP (0.65\*\*) and TSW (0.58\*\*). However, genetically, YPP was more strongly correlated to TSW (0.69\*\*) than NCP (0.64\*\*). Also the genetic correlation among YPP and seed size-related traits was generally stronger than the seed number-related traits (Table 3). Interestingly, seed oil content was associated with all seed size-related traits and in contrast, completely independent of seed number-related traits both phenotypically and genetically (Table 3). Also, phenotypic and genetic correlation coefficients between OIL and YPP was significant, positive, and relatively high (rp=34\*\*, rg=0.53\*\*). The most underlying trait associated with PRO was OIL (rp=0.27\*\*, rg=0.19\*); though PLH also had a weak positive but highly significant genetic correlation coefficient with PRO (rg=0.25\*\*) (Figure 2 and Table 3).

***3.4. Diversity pattern within and between geographical origins***

Hierarchical cluster analysis classified the 120 linseed genotypes from 14 above-mentioned geographical origins into 11 different groups based on all studied traits in the first and second experimental years (Figure 3). Furthermore, as the dendrogram revealed, none of the clusters contained genotypes from an identical geographical area, indicating that the obtained grouping was fully independent of geographical origin. To establish further evidence, the Euclidean distances within and between all 14 geographical regions were computed. Results showed that, except for SAS and WEU, in all other regions, the intra-regional Euclidean distance was higher than inter-regional one (Table 4). Almost a similar intra-regional variation was observed for most of the areas. In most cases, the average distance among genotypes in a specific region was comparable to or even slightly higher than the average distance among that region and others with the exception of SAS and WEU. Accordingly, the average distance within (diagonal line in Table 4) and between the regions were similar and equal to 4.56±0.08 and 4.48±1.41, respectively. In this respect, SAS and WEU showed the most distinction to other regions, and also, the distance between them was significantly higher than other pairs of regions (Table 4).

 Factor analysis (FA) was performed to give a comprehensive but straightforward view of the relationships among different geographical origins and studied traits. The Kaiser-Meyer-Olkin measure (KMO) was 0.54, and the p-value of Bartlett’s test of sphericity was ˂0.001, indicating data sufficiency for FA and existence of relationships among variables. The first three factors were considered the most influential ones with respect to their eigenvalues ˃ 1 and cumulatively explaining ≈79% of the total variance (Table 5). Factor one contained PLH (negative sign), CDI, TSW, YPP, SLE, SWI, and OIL. Factor two and three particularly included CSN and NBP, respectively. YPP was almost equally distributed between factors one and two, and NCP was the variable in the situation of cross loading between factors two and three. Hence both YPP and NCP were considered as complex variables (Table 5). As illustrated in Figure 4, although most geographic origins centralized around the intersection of all three factors, few regions were more dispersed across all three factors; this was particularly evident for SAS and WEU. . In this respect, table 6 helps to understand better the bi-plot shown in Figure 4; the regions with higher positive values of factor one were owned bigger and heavier seeds and higher seed oil content along with the shorter main stem and *vice versa*. Also, the regions that acquired higher positive values of factors two and three were correlated to higher seed number-related traits.

***3.5. Distinction of elite genotypes***

Considering that the KMO was above 0.5 in all cases, the two years data individually and in combination were sufficient for FA. Also, the p-values of Bartlett’s test of sphericity for all FAs were lower than 0.001, indicating the existence of patterned relationships among the variables. In 2016, 2017, and the combination of both years’ data, the first two factors with an eigenvalue ˃1 cumulatively explained 54.11, 49.44, and 52.92% of total variation, respectively. Consequently these two factors were retained as the essential factors (Table 6). The first factor was named as “seed size,” considering that it was dominated by traits determining this feature including CDI, TSW, SLE, and SWI. The second factor was composed of NBP, NCP, and YPP, all representing the number of seeds in a single plant and its productivity, thus was named as “seed number” (Table 6). Accordingly, the genotypes with higher values of factor one are the ones with bigger and heavier seeds and *vice versa*; while, regarding factor two the higher positive values indicate the higher number of seeds.

 Genotypes distribution with respect to factors 1 and 2 are illustrated in Figures 5 A, B, and C for 2016, 2017, and combination of two years’ data, respectively. Based on 2016 data, two groups of genotypes were aggregated distinctively that are delimited with red circles. The first group, had a strong and positive association with both factors and mostly consisted of genotypes from WAS region. However, the genotypes of the second group, which had the highest negative values related to both factors, were mainly belonged to the WEU region. (Figure 5A). For 2017 data and the combination of two years data, the grouping of genotypes was mostly similar to 2016 results, except that eight genotypes were separated from group I and centralized with other intermediate genotypes having moderate levels of both factors (Figure 5B and C). In bi-plots drawn from 2016, 2017, and the combination of two years data, five genotypes designated with red arrows, were considered as elite lines for seed production purposes based on their highest amounts of factor one, two, or both(Figure 5A, B, and C).

 Among the five elite lines mentioned earlier, accession #108, a high-yielding oilseed flax from Canada (Flanders) (Rowland, Bhatty, & Kenaschuk, 1990), was considered a criterion genotypes regarding seed production. This genotype had the highest score regarding “seed number,” in 2016. However, this score dropped in 2017, but its moderate score of “seed size” was stable in both years. Reciprocally, for accessions #129 and #133, two other elite lines belonging to the IPK world collection, “seed number” score increased from average to the highest level from 2016 to 2017; while, their “seed size” score was at an average level across two experimental years (Figure 5A and B). Although the “seed number” scores of the three above mentioned elite lines showed some fluctuation over two experimental years but based on two years combined data, genotypes #129, #108, and #133 had the highest “seed number” scores, respectively, and an average “seed size” score (Figure 5C). The two remaining elite lines were genotypes #105 (an Indian landrace) and #15 (a line selected from IPK world collection), which had the highest “seed size” scores in 2016, respectively. Furthermore, in this year, genotype #105 had a strong positive association with “seed number” but genotype #15 was moderately in a positive relationship with this latent variable (Figure 5A). In 2017, genotype #105 was distinguished from the others regarding “seed size” and had an acceptable score of “seed number”, while, genotype #15 got closer to intermediate genotypes in this respect (Figure 5B). Among the 120 accessions studied, and based on two years combined datasets, genotype #105 was the most distinct one in respect to “seed size” and “seed number” latent factors (Figure 5C).

**4. Discussion**

***4.1. Genetic variability of the diversity panel and its application in linseed breeding***

The present study assessed nine agronomic and two seed quality traits essential to both breeders and farmers. Results revealed a sizeable genetic variability indicating a broad spectrum and flexibility for phenotypic performance-based selection particularly for YPP, NCP, NBP, and TSW, according to $\hat{PCV}$, $\hat{GCV}$, and ∆G. A comparative look at $\hat{H}^{2}$ and ∆G provided a general view of the type of gene action. Generally, a moderate to high $\hat{H}^{2}$ combined with a high ∆G indicates predominantly additive gene action in controlling a particular trait; while, a high $\hat{H}^{2}$ coupled with a low to moderate ∆G, suggests the significant role of non-additive gene action in governing that trait (Chozin, Sumardi, Sudjamiko, & Barchia, 2017). In this respect, YPP, TSW, NBP, and PLH showed moderate to high $\hat{H}^{2}$ and high amounts of ∆G. In contrast, high $\hat{H}^{2}$ for seed size-related traits including SLE, SWI, and CDI, and moderate $\hat{H}^{2}$ for seed quality traits were both accompanied by low ∆G. Considering that these results were mostly in close agreement with the extensive phenotypic evaluation of a broader collection (391 accessions) in eight different year-location environments in Canada (You et al., 2017), the additive gene action must play a significant role in the inheritance of YPP, TSW, NBP, and PLH. Non-additive gene actions seem to be involved in governing the seed size-related and seed quality traits. Using a diallel cross analysis, (Mohammadi, Saeidi, & Arzani, 2010) reported that additive gene action largely controls plant height, number of branches per plant, and thousand seed weight in linseed. However, seed yield per plant was dominantly controlled by both additive and non-additive gene actions and number of capsules per plant was under the dominant gene effects. Our results corroborate the previous findings suggesting that simple selection during breeding cycles is effective and may result in fast improvement of YPP, TSW, NBP, and PLH. However, fast progress may not be feasible by simple and direct selection for seed size-related and seed quality traits.

 ***4.2. Geographical patterns of variability of the diversity panel***

The present diversity panel of linseed comprised 120 accessions from 47 countries and 14 distinct geographical regions, and most of them were randomly selected from the IPK world collection (Figure 6). With the exception of four regions, results of cluster analysis, FA and Euclidean distances revealed a poor geographical patterns in this diversity panel. Genotypes from South and West Asia had bigger and heavier seeds along with high seed oil content and shorter main stem. In contrast, the majority accessions from Western Europe and Russia had low seed yield, tall main stem, and few branches, the essential descriptors of intermediate flax genotypes appropriate for fiber production.

 It is suggested that the center of origin for linseed is Middle East, India, Ethiopia, and the Mediterranean basin (Vavilov, 2009). Linseed has been domesticated first in the Middle East and was then chronologically spread to Europe, Nile valley, and other regions around the world (Hillman, 1975; van Zeist & Bakker-Heeres, 1975). It is also unanimously believed that the domestication of the oilseed type had occurred before the fiber type (Allaby, Peterson, Merriwether, & Fu, 2005; Fu & Allaby, 2010; Herbig & Maier, 2011). After domestication, only the oilseed type has been grown in South-West Asian countries, including Turkestan, Afghanistan, and India, whereas European countries, particularly Western European countries, mostly have cultivated fiber type (Badole, Zanwar, & Bodhankar, 2013; Baley, Le Duigou, Morvan, & Bourmaud, 2018). This pattern of domestication and cultivation mostly supports our results and suggests that continuous artificial selection based on regional demands has made a region-type attachment. In other words, these results suggested that the geographic origin had not naturally forced linseed plant to choose a growth type, but the pressure that came from the artificial selection of end-use traits with specific regional preference has made a few region-trait relationships.

Except for a minimum regional relationships between tall and high yielding genotypes with WEU and SAS regions respectively, the overall results suggested the existence of a weak geographical patterns in this diversity panel. Two studies using a Canadian linseed whole and core collection estimated that geographical origin explained just about 8.2 and 11% of the total variation based on molecular and morpho-phenological variations, respectively (Fu, 2005; You et al., 2017). Another study on the same Canadian core collection suggested the absence of population particular structure (Soto-Cerda et al., 2013), indicating its adequacy for association mapping. Our phenotypic based evaluation also demonstrated that just 6.5% of the total variation in the diversity panel of IPK world collection is attributable to the geographical origins, suggesting a high possibility for this population to be unstructured and may be appropriate for association mapping.

***4.3. The nature of traits association and its application in linseed breeding***

Genetic correlation between two traits may originates from a specific gene's pleiotropic effect, linkage disequilibrium between two distinct loci or both mechanisms (Falconer, 1981). In the present study a significant negative correlation was observed between PLH and every other trait. This is, however, not in agreement with observations reported by several other studies suggesting a positive correlation between plant height and seed yield and between plant height and most of the seed yield components in linseed (Adugna & Labuschagne, 2003; Copur, Gur, Karakus, & Demirel, 2006; Rahimi, Zarei, & Arminian, 2011; Zhang et al., 2014; Ibrar et al., 2016). This contradiction may be explained by the fact that germplasms used in almost all aforementioned studies were either limited in number or from narrow geographical distributions. In our study, the germplasm used was assembled of a collection originated from 14 different geographic regions with possible imposed selections based on end use purposes. Intended or unintended mixing of accessions from diverse subgroups with different allele frequencies during collection construction could create a false linkage disequilibrium leading to a population structure (Soto-Cerda & Cloutier, 2012). Hence, as a result of this germplasm assembly, a population structure may be formed in our diversity panel due to alleles' unequal distribution among subpopulations driven by diverse ancestries. Our results is also supported by the findings of Soto‐Cerda et al. (2014) studying a Canadian linseed world collection of 390 accessions demonstrating an inverse correlation between plant height and seed yield, number of capsules per area, and thousand seed weight.

 As observed in many crops, strong artificial selection for desired traits has dramatically affected LD during domestication bottlenecks resulting in miscellanies of large LD blocks, especially in chromosomal regions containing genes for agronomic characteristics. This has led to the genome architecture modification and genetic diversity reduction which finally created population structure (Soto-Cerda & Cloutier, 2012). Linseed has been anciently under strong selection for two distinct end-use traits of seed and fiber depending on specific regional preference. Fiber types have been selected for tall main stem and a lower number of branches, resulting in lower capsules and seed yield. Whereas, the seed type is exactly the opposite creating an unequal allele frequency between the two types. Also, it is well established that a self-mating system reduces recombination opportunities because individuals are most likely to be homozygous. Consequently, self-pollinated crops are capable of keeping LD for a much longer time comparing to cross-pollinated crops. As a result, genetic polymorphisms tend to remain correlated across many years (Soto-Cerda & Cloutier, 2012). Accordingly, the historical artificial selection in linseed for two different end use traits more likely led to a fixed genetic polymorphism of taller plants correlated with lower seed yield. The final result of this process might be a reduction in genetic diversity which in particular is evident for intermediate genotypes appropriate for dual-purpose uses. Besides, a regional preference for each end-use trait in linseed has led to some specific region-trait relationships that remained over a long time due to its self-mating system. Therefore, if the present diversity panel is going to be used for association mapping studies, its weak geographical pattern and the structural network of relationships that exist for PLH should be considered.

 A positive association was observed between OIL and seed size-related traits. This was expected considering that 75% of lipids are stored in cotyledons (Rubilar, Gutiérrez, Verdugo, Shene, & Sineiro, 2010), which occupies the majority of the mature seed space in linseed (Venglat et al., 2011). There was also positive correlations between all pairs of seed size-related traits and each of them with YPP. These positive correlations also make sense since seeds with higher diameters are bigger in size and heavier, consequently increasing the containing capsule’s diameter and the weight of thousand seeds. These all together may finally result in higher seed yield. Seed number related traits, except for CSN, also showed positive correlations with YPP and were almost independent of seed-size related traits. The decision to produce more small seeds versus less big seeds in plant species is a fundamental trade-off in resource allocation, termed as seed size/number trade-off (SSNT) (Leishman, 2001).

Factor analysis (FA) is customarily used to regroup variables that could not be meaningfully compared because of their different scales, into a limited number of clusters, usually called factors, based on common variance. Its primary purpose is to summarize data and subsequently easily interpret and understand the hidden relationships and patterns among variables and measurement units (Yong & Pearce, 2013). Surprisingly, the first two factors in this study, which explained a significant fraction of total variance, represented seed-size and seed-number related characteristics. The bi-plot graphic display of all 120 linseed genotypes under study based on first and second factors clearly depicted the existence of SSNT as most genotypes were aggregated in the center of the bi-plot (Figure 5A, B, and C). This may suggests a compromising strategy between the two factors for yield improvement in linseed. However, as the “seed size” factor was better able to disperse the genotypes on the bi-plot compared to “seed number” factor, indicating higher genetic variability for the seed-size related traits suggesting a tendency toward bigger seed size may be more effective during selection. From the five genotypes specified by red arrows on bi-plot chart, four were superior in respect to seed size, seed number or both, either in one or both of the experimental years (combined data of both years). However, as it is evident from the first to the second year of cultivation, almost all of them showed fluctuation in respect to both factors. Genotype #105 was exceptionally the most stable one in the two experimental years even when it was compared to the high-yielding commercial cultivar Flanders which was considered a criterion for seed yield. This may suggests the genetic potential of genotype #105 for yield increase in linseed breeding programs. Genotypes #133, #108, and #129 are the other candidates for this purpose.

 In conclusion, results suggested a broad range of variability for almost all the traits under study. Genetic diversity within sub-populations accounted for the majority of this variability and geographical origin explained a small fraction of the total variation. This diversity panel may be an appropriate source of genetic material for association mapping, considering that the geographical origin explained a small fraction of the total diversity in the diversity panel, and weak patterns were observed among geographical regions. However, it seems that the presence of two linseed morphotypes in this collection may have created a population structure based on false linkage disequilibrium between plant height and seed yield. This should be considered if this diversity panel is going to be used for association mapping. Phenotyping this globally distributed diversity panel resulted in identifying distinct genotypic groups with specific features and elite lines useful for linseed breeding programs and genetic studies.

**Acknowledgements**

This research was funded by Isfahan University of Technology.

**Conflicts of Interest**

The authors declare no conflicts of interest.

**Author Contributions**

A. Mirlohi and E. Ataii designed the research and coordinated the experiment. Ataii, N. Sharif Moghadam, N. Sadri, P. Jafari did the trial organization and phenotyping. Ataii performed all statistical analyses and wrote the manuscript. M. R. Sabzalian helped with statistical advice, edited and improved the manuscript together with A. Mohammadi Mirik. A. Mirlohi revised the manuscript critically. All authors read and approved the final version of the manuscript.

**References**

2010–2020: UN Decade for Deserts and the Fight against Desertification. (n.d.). Retrieved October 25, 2020, from https://www.un.org/en/events/desertification\_decade/value.shtml

Adugna, W., & Labuschagne, M. T. (2003). Association of linseed characters and its variability in different environments. *Journal of Agricultural Science*, *140*(3), 285–296. https://doi.org/10.1017/S0021859603003125

Allaby, R. G., Peterson, G. W., Merriwether, D. A., & Fu, Y. B. (2005). Evidence of the domestication history of flax (*Linum usitatissimum* L.) from genetic diversity of the *sad2* locus. *Theoretical and Applied Genetics*, *112*, 58–65. https://doi.org/10.1007/s00122-005-0103-3

Alvarado, G., Rodríguez, F. M., Pacheco, A., Burgueño, J., Crossa, J., Vargas, M., … Lopez-Cruz, M. A. (2020). META-R: A software to analyze data from multi-environment plant breeding trials. *Crop Journal*, *8*(5), 745–756. https://doi.org/10.1016/j.cj.2020.03.010

AOAC. (1980). *AOAC: Official Methods of Analysis*. Association of Official Agricultural Chemists. Washington, D.C.

Badole, S. L., Zanwar, A. A., & Bodhankar, S. L. (2013). Antihyperglycemic Potential of Secoisolaricinol Diglucoside. In R. R. Watson & V. R. Preedy (Eds.), *Bioactive Food as Dietary Interventions for Diabetes* (pp. 53–57). London: Elsevier Inc. https://doi.org/10.1016/B978-0-12-397153-1.00005-6

Baley, C., Le Duigou, A., Morvan, C., & Bourmaud, A. (2018). Tensile properties of flax fibers. In A. R. Bunsell (Ed.), *Handbook of Properties of Textile and Technical Fibres* (pp. 275–300). Elsevier. https://doi.org/10.1016/B978-0-08-101272-7.00008-0

Chandrawati, D., Singh, N., Kumar, R., Kumar, S., Singh, P. K., Yadav, V. K., … Yadav, H. K. (2017). Genetic diversity, population structure and association analysis in linseed (*Linum usitatissimum* L.). *Physiology and Molecular Biology of Plants*, *23*(1), 207–219. https://doi.org/10.1007/s12298-016-0408-5

Chozin, M., Sumardi, S., Sudjamiko, S., & Barchia, M. F. (2017). Genetic variability and traits association analyses on F2 generations for determination of selection criteria in Indonesian inland swamp rice breeding. *Australian Journal of Crop Science*, *11*(5), 535–541. https://doi.org/10.21475/ajcs.17.11.05.p317

Christopher A. Cullis. (2019). Origin and Induction of the Flax Genotrophs. In Christopher A. Cullis (Ed.), *Genetics and Genomics of Linum* (pp. 227–234). Cham: Springer. https://doi.org/10.1007/978-3-030-23964-0

Copur, O., Gur, M. A., Karakus, M., & Demirel, U. (2006). Determination of Correlation and Path Analysis among Yield Components and Seed Yield in Oil Flax Varieties (*Linum usitatissimum* L.). *Journal of Biological Sciences*, *6*(4), 738–743.

Diederichsen, A., & Ulrich, A. (2009). Variability in stem fibre content and its association with other characteristics in 1177 flax (*Linum usitatissimum* L.) genebank accessions. *Industrial Crops and Products*, *30*(1), 33–39. https://doi.org/10.1016/j.indcrop.2009.01.002

Falconer, D. S. (1981). *Introduction to Quantitative Genetics* (2nd ed.). London: Longman Press.

Fu, Y. B. (2005). Geographic patterns of RAPD variation in cultivated flax. *Crop Science*, *45*(3), 1084–1091. https://doi.org/10.2135/cropsci2004.0345

Fu, Y. B., & Allaby, R. G. (2010). Phylogenetic network of *Linum* species as revealed by non-coding chloroplast DNA sequences. *Genetic Resources and Crop Evolution*, *57*(5), 667–677. https://doi.org/10.1007/s10722-009-9502-7

Herbig, C., & Maier, U. (2011). Flax for oil or fibre? Morphometric analysis of flax seeds and new aspects of flax cultivation in Late Neolithic wetland settlements in southwest Germany. *Vegetation History and Archaeobotany*, *20*(6), 527–533. https://doi.org/10.1007/s00334-011-0289-z

Herchi, W., Arráez-Román, D., Boukhchina, S., Kallel, H., Segura-Carretero, A., & Fernández-Gutierrez, A. (2012). A review of the methods used in the determination of flaxseed components. *African Journal of Biotechnology*, *11*(4), 724–731. https://doi.org/10.5897/ajb11.984

Hillman, G. C. (1975). The plant remains from Tell Abu Hureyra. In A. M. T. Moore, G. C. Hillman, & A. J. Legge (Eds.), *The excavation of Tell abu Hureyra in Syria: a Preliminary Report* (pp. 70–73). London: Procedings of the Prehistoric Society. https://doi.org/10.1017/S0079497X00010902

Ibrar, D., Ahmad, R., Mirza, M. Y., Mahmood, T., Ahmad Khan, M., & Shahid Iqbal, M. (2016). Correlation and path analysis for yield and yield components in linseed (*linum usitatissimum* L.). *Journal of Agricultural Research*, *54*(2), 153–159.

Leishman, M. R. (2001). Does the seed size/number trade-off model determine plant community structure? An assessment of the model mechanisms and their generality. *Oikos*, *93*(2), 294–302. https://doi.org/10.1034/j.1600-0706.2001.930212.x

Liu, F.-H., Chen, X., Long, B., Shuai, R.-Y., & Long, C.-L. (2011). Historical and botanical evidence of distribution, cultivation and utilization of *Linum usitatissimum* L. (flax) in China. *Vegetation History and Archaeobotany*, *20*(6), 561–566. https://doi.org/10.1007/s00334-011-0311-5

Majidi, M. M., & Zadhoush, S. (2014). Molecular and morphological variation in a world-wide collection of safflower. *Crop Science*, *54*(5), 2109–2119. https://doi.org/10.2135/cropsci2013.12.0850

Mohammadi, A. A., Saeidi, G., & Arzani, A. (2010). Genetic analysis of some agronomic traits in flax (*Linum usitatissimum* L.). *Australian Journal of Crop Science*, *4*(5), 343–352.

Oomah, B. D. (2001). Flaxseed as a functional food source. *Journal of the Science of Food and Agriculture*, *81*(9), 889–894. https://doi.org/10.1002/jsfa.898

Rahimi, M. M., Zarei, M. A., & Arminian, A. (2011). Selection criteria of flax (*Linum usitatissimum* L.) for seed yield, yield components and biochemical compositions under various planting dates and nitrogen. *African Journal of Agricultural Research*, *6*(13), 3167–3175.

Rowland, G. G., Bhatty, R. S., & Kenaschuk, E. O. (1990). Flanders flax. *Canadian Journal of Plant Science, 70*(2), 543–544.

Rubilar, M., Gutiérrez, C., Verdugo, M., Shene, C., & Sineiro, J. (2010). Flaxseed as a source of functional ingredients. *Journal of Soil Science and Plant Nutrition*, *10*(3), 373–377. https://doi.org/10.4067/S0718-95162010000100010

Soto-Cerda, B. J., & Cloutier, S. (2012). Association mapping in plant genomes. In M. Caliskan (Ed.), *Genetic diversity in plants* (pp. 29–54). Rijeka: InTech. https://doi.org/10.5772/2640

Soto-Cerda, B. J., Diederichsen, A., Ragupathy, R., & Cloutier, S. (2013). Genetic characterization of a core collection of flax (*Linum usitatissimum* L.) suitable for association mapping studies and evidence of divergent selection between fiber and linseed types. *BMC Plant Biology*, *13*, 78. https://doi.org/10.1186/1471-2229-13-78

Soto‐Cerda, B. J., Duguid, S., Booker, H., Rowland, G., Diederichsen, A., & Cloutier, S. (2014). Genomic regions underlying agronomic traits in linseed (*Linum usitatissimum* L.) as revealed by association mapping. *Journal of Integrative Plant Biology*, *56*(1), 75–87. https://doi.org/10.1111/jipb.12118

Talebi, S. M., Amini, F., Askary, M., Farahani, S., & Matsyura, A. (2020). Seed morphology and fatty acids composition among Flax populations. *Brazilian Journal of Botany*, *43*(2), 355–365. https://doi.org/10.1007/s40415-020-00601-y

van Zeist, W., & Bakker-Heeres, J. A. H. (1975). Evidence for linseed cultivation before 6000 bc. *Journal of Archaeological Science*, *2*(3), 215–219. https://doi.org/10.1016/0305-4403(75)90059-X

Vavilov, N. (2009). Geographical regularities in the distribution of the genes of cultivated plants. *Comparative Cytogenetics*, *3*(1), 71–78. https://doi.org/10.3897/compcytogen.v3i1.10

Venglat, P., Xiang, D., Qiu, S., Stone, S. L., Tibiche, C., Cram, D., … Datla, R. (2011). Gene expression analysis of flax seed development. *BMC Plant Biology*, *11*(1), 74. https://doi.org/10.1186/1471-2229-11-74

Yong, A. G., & Pearce, S. (2013). A Beginner’s Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94. https://doi.org/10.20982/tqmp.09.2.p079

You, F. M., Booker, H. M., Duguid, S. D., Jia, G., & Cloutier, S. (2016). Accuracy of genomic selection in biparental populations of flax ( *Linum usitatissimum* L.). The Crop Journal, 4(4), 290–303. https://doi.org/10.1016/j.cj.2016.03.001

You, F. M., Jia, G., Xiao, J., Duguid, S. D., Rashid, K. Y., Booker, H. M., & Cloutier, S. (2017). Genetic variability of 27 traits in a core collection of flax (*Linum usitatissimum* L.). *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01636

Zhang, T., Lamb, E. G., Soto-Cerda, B., Duguid, S., Cloutier, S., Rowland, G., … Booker, H. M. (2014). Structural equation modeling of the Canadian flax (*Linum usitatissimum* L.) core collection for multiple phenotypic traits. *Canadian Journal of Plant Science*, *94*(8), 1325–1332. https://doi.org/10.4141/CJPS-2014-158

**Figure legends**

**Figure 1.** The bi-plot of the phenotypic variation of 11 traits explained by the experimental year and geographical origin, which were standardized to percentages (%) based on each trait's total variance. PLH: Plant height; NBP: Number of branches per plant; NCP: Number of capsules per plant; CDI: Capsule diameter; CSN: Capsule seeds number; TSW: Thousand seed weight; YPP: Yield per plant; SLE: Seed length; SWI: Seed width; OIL: Oil content; PRO: Protein content

**Figure 2.** Phenotypic (A) and genetic (B) correlation networks of 11 studied traits in the diversity panel of 120 linseed accessions over two years. PLH: Plant height; NBP: Number of branches per plant; NCP: Number of capsules per plant; CDI: Capsule diameter; CSN: Capsule seeds number; TSW: Thousand seed weight; YPP: Yield per plant; SLE: Seed length; SWI: Seed width; OIL: Oil content; PRO: Protein content. The red and green color of lines represent a negative and positive correlation, respectively. The thickness of lines indicates how strong the relevant correlation is.

**Figure 3.** Dendrogram derived from cluster analysis using a diversity panel of 120 linseed accessions, randomly selected from IPK flax world collection. The cluster analysis was performed based on the Ward method. Genotypes are colored according to their geographical origin.AUS: Australian; CEU: Central Europe; EAF: East Africa; EAS: East Asia; NAF: North Africa; NAM: North America; NEU: Northern Europe; RUS: Russia; SAM: South America; SAS: South Asia; SEE: Southeast Europe; SWE: Southwestern Europe; WAS: West Asia; WEU: Western Europe.

**Figure 4.** Dispersion bi-plot of 14 geographical origins according to the first three factors of the evaluated agronomic traits over two years. AUS: Australian; CEU: Central Europe; EAF: East Africa; EAS: East Asia; NAF: North Africa; NAM: North America; NEU: Northern Europe; RUS: Russia; SAM: South America; SAS: South Asia; SEE: Southeast Europe; SWE: Southwestern Europe; WAS: West Asia; WEU: Western Europe. PLH: Plant height; NBP: Number of branches per plant; NCP: Number of capsules per plant; CDI: Capsule diameter; CSN: Capsule seeds number; TSW: Thousand seed weight; YPP: Yield per plant; SLE: Seed length; SWI: Seed width; OIL: Oil content; PRO: Protein content.

**Figure 5.** The bi-plot depicting the scatter of genotypes alongside the measured traits in 2016 (a), 2017 (b) and the combination of two years (c) based on factor analysis.

**Figure 6.** Geographic distribution of diversity panel of 120 linseed accessions randomly selected from IPK flax world collection.

**Table 1**. ANOVA table for studied traits in the diversity panel of 120 linseed genotypes over two years.

|  |  |  |  |
| --- | --- | --- | --- |
| Traits | Genotype | Year | Genotype × Year |
| PLH (cm) | 76.670\*\* | 2.443ns | 8.064\*\* |
| NBP | 0.209\*\* | 1.002\* | 0.227\*\* |
| NCP | 19.898\* | 152.320\* | 55.087\*\* |
| CDI (mm) | 0.065\*\* | 0.0001ns | 0.005ns |
| CSN | 0.316\*\* | 0.0001ns | 0.0001ns |
| TSW (g) | 0.276\*\* | 0.260ns | 0.113\*\* |
| YPP (g) | 0.022\*\* | 0.099\* | 0.042\*\* |
| SLE (mm) | 0.080\*\* | 0.001ns | 0.006\*\* |
| SWI (mm) | 0.018\*\* | 0.002ns | 0.001ns |
| OIL (%) | 2.918\*\* | 1.143ns | 2.614\*\* |
| PRO (%) | 2.268\*\* | 1.035ns | 4.831\*\* |

\* and \*\* represent significant at p˂0.05 and p˂0.01 respectively; ns represents not significant.

PLH: Plant height; NBP: Number of branches per plant; NCP: Number of capsules per plant; CDI: Capsule diameter; CSN: Capsule seeds number; TSW: Thousand seed weight; YPP: Yield per plant; SLE: Seed length; SWI: Seed width; OIL: Oil content; PRO: Protein content

**Table 2.** Estimates of genetic parameters and phenotypic performance of 11 studied traits in the diversity panel of linseed evaluated over two years.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Trait |  | Mean±std† | Range | $$\hat{PCV} (\%)$$ | $$\hat{GCV (\%)}$$ | $$\hat{H}^{2}$$ | ∆G (%) |
| PLH (cm) |  | 52.1±8.2 | 40.21 | 17.95 | 16.81 | 0.88 | 32.48 |
| NBP |  | 2.1±0.3 | 1.90 | 30.34 | 21.68 | 0.51 | 31.89 |
| NCP |  | 28.7±2.6 | 13.07 | 27.00 | 15.54 | 0.33 | 18.39 |
| CDI (mm) |  | 6.4±0.2 | 1.37 | 4.44 | 4.00 | 0.81 | 7.42 |
| CSN |  | 8.4±0.5 | 2.36 | 7.71 | 6.69 | 0.75 | 11.94 |
| TSW (g) |  | 3.7±0.5 | 3.19 | 16.59 | 14.26 | 0.74 | 25.27 |
| YPP (g) |  | 0.6±0.1 | 0.47 | 37.81 | 25.42 | 0.45 | 35.13 |
| SLE (mm) |  | 4.1±0.3 | 1.44 | 7.12 | 6.83 | 0.92 | 13.50 |
| SWI (mm) |  | 2.2±0.1 | 0.61 | 6.41 | 6.11 | 0.91 | 12.01 |
| OIL (%) |  | 30.0±1.3 | 6.96 | 7.22 | 5.69 | 0.62 | 9.23 |
| PRO (%) |  | 38.3±1.0 | 4.89 | 6.02 | 3.93 | 0.43 | 5.31 |

† Values represent means ±std (Standard deviation)

PLH: Plant height; NBP: Number of branches per plant; NCP: Number of capsules per plant; CDI: Capsule diameter; CSN: Capsule seeds number; TSW: Thousand seed weight; YPP: Yield per plant; SLE: Seed length; SWI: Seed width; OIL: Oil content; PRO: Protein content

**Table 3.** Phenotypic (upper diagonal) and genetic (lower diagonal) correlation coefficients among agronomic and seed quality traits on 120 genotypes of linseed over 2 years

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Traits | PLH | NBP | NCP | CDI | CSN | TSW | YPP | SLE | SWI | OIL | PRO |
| PLH |  | -0.26\*\* | -0.15ns | -0.17ns | -0.04ns | -0.44\*\* | -0.31\*\* | -0.39\*\* | -0.40\*\* | -0.29\*\* | 0.14ns |
| NBP | -0.44\*\* |  | 0.45\*\* | 0.17ns | -0.17ns | 0.17ns | 0.28\*\* | 0.25\* | 0.32\*\* | 0.04ns | -0.15ns |
| NCP | -0.40\*\* | 0.37\*\* |  | 0.10ns | -0.10ns | 0.14ns | 0.65\*\* | 0.11ns | 0.06ns | 0.03ns | -0.16ns |
| CDI | -0.20\* | 0.31\*\* | 0.16ns |  | -0.04ns | 0.63\*\* | 0.27\*\* | 0.68\*\* | 0.74\*\* | 0.15ns | 0.00ns |
| CSN | -0.05ns | -0.28\*\* | -0.08ns | -0.10ns |  | 0.06ns | 0.14ns | -0.17ns | -0.11ns | 0.04ns | 0.17ns |
| TSW | -0.54\*\* | 0.35\*\* | 0.21\* | 0.73\*\* | 0.11ns |  | 0.58\*\* | 0.78\*\* | 0.77\*\* | 0.48\*\* | 0.14ns |
| YPP | -0.57\*\* | 0.32\*\* | 0.64\*\* | 0.42\*\* | 0.29\*\* | 0.69\*\* |  | 0.36\*\* | 0.34\*\* | 0.35\*\* | 0.04ns |
| SLE | -0.42\*\* | 0.38\*\* | 0.15ns | 0.73\*\* | -0.20\* | 0.88\*\* | 0.50\*\* |  | 0.84\*\* | 0.36\*\* | 0.03ns |
| SWI | -0.44\*\* | 0.44\*\* | 0.05ns | 0.82\*\* | -0.12ns | 0.90\*\* | 0.56\*\* | 0.85\*\* |  | 0.36\*\* | -0.05ns |
| OIL | -0.38\*\* | 0.10ns | 0.00ns | 0.19\* | 0.13ns | 0.57\*\* | 0.53\*\* | 0.45\*\* | 0.48\*\* |  | 0.27\*\* |
| PRO | 0.25\* | -0.21\* | -0.20\* | 0.04ns | 0.40\*\* | 0.10ns | -0.05ns | 0.03ns | -0.08ns | 0.19\* |  |

\* and \*\* represent significant at p˂0.05 and p˂0.01 respectively; ns represents not significant.

PLH: Plant height; NBP: Number of branches per plant; NCP: Number of capsules per plant; CDI: Capsule diameter; CSN: Capsule seeds number; TSW: Thousand seed weight; YPP: Yield per plant; SLE: Seed length; SWI: Seed width; OIL: Oil content; PRO: Protein content

**Table 4.** Phenotypic variance partitioning of 11 traits in terms of experimental year, geographic origin and other factors for the diversity panel of 120 linseed accessions.

|  |  |  |  |
| --- | --- | --- | --- |
| Trait | Year (%) | Geographical origin (%) | Other (%) |
| PLH | 2.45 | 19.80 | 77.75 |
| NBP | 55.90 | 2.24 | 41.86 |
| NCP | 54.83 | 0.69 | 44.48 |
| CDI | 0.00 | 6.85 | 93.15 |
| CSN | 0.45 | 0.00 | 99.55 |
| TSW | 30.97 | 8.11 | 60.93 |
| YPP | 54.06 | 1.75 | 44.19 |
| SLE | 1.70 | 8.49 | 89.81 |
| SWI | 7.43 | 12.79 | 79.78 |
| OIL | 12.37 | 2.74 | 84.89 |
| PRO | 9.96 | 7.37 | 82.67 |
| Average | 20.92 | 6.44 | 72.64 |

PLH: Plant height; NBP: Number of branches per plant; NCP: Number of capsules per plant; CDI: Capsule diameter; CSN: Capsule seeds number; TSW: Thousand seed weight; YPP: Yield per plant; SLE: Seed length; SWI: Seed width; OIL: Oil content; PRO: Protein content

**Table 5.** The Euclidean distance among 14 geographical origin of the linseed diversity panel.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NAC |  | AUS | CEU | EAF | EAS | NAF | NAM | NEU | RUS | SAM | SAS | SEE | SWE | WAS | WEU | Mean |
| 3 | AUS | 4.68 |  |  |  |  |  |  |  |  |  |  |  |  |  | 4.10 |
| 22 | CEU | 3.38 | 4.50 |  |  |  |  |  |  |  |  |  |  |  |  | 3.75 |
| 5 | EAF | 2.53 | 2.73 | 4.51 |  |  |  |  |  |  |  |  |  |  |  | 3.79 |
| 7 | EAS | **2.45** | 3.17 | 3.05 | 4.56 |  |  |  |  |  |  |  |  |  |  | 4.00 |
| 5 | NAF | 4.92 | 3.73 | 4.60 | 4.36 | 4.59 |  |  |  |  |  |  |  |  |  | 4.97 |
| 15 | NAM | 4.86 | 3.94 | 3.51 | 4.40 | 5.75 | 4.45 |  |  |  |  |  |  |  |  | 4.58 |
| 4 | NEU | 3.52 | 3.36 | 3.08 | 3.78 | 5.01 | 3.41 | 4.63 |  |  |  |  |  |  |  | 3.96 |
| 4 | RUS | 5.37 | 4.87 | 4.30 | 5.33 | 7.41 | 3.27 | 4.26 | 4.56 |  |  |  |  |  |  | 5.11 |
| 12 | SAM | 2.80 | 2.61 | 2.52 | 3.05 | 3.91 | 4.41 | 2.74 | 4.62 | 4.52 |  |  |  |  |  | 3.53 |
| 6 | SAS | 5.54 | 4.89 | 5.92 | 5.17 | 3.99 | 5.49 | 4.69 | 7.30 | 5.19 | 4.56 |  |  |  |  | 5.55 |
| 4 | SEE | 4.05 | 3.79 | 3.91 | 4.24 | 5.34 | 4.81 | 3.67 | 3.91 | 2.24 | 5.87 | 4.64 |  |  |  | 4.24 |
| 5 | SWE | 3.68 | 3.43 | 4.63 | 3.45 | 4.66 | 6.14 | 4.41 | 5.98 | 2.97 | 5.22 | 3.59 | 4.63 |  |  | 4.59 |
| 20 | WAS | 3.49 | 2.98 | 3.51 | 2.54 | 2.78 | 4.00 | 3.86 | 6.00 | 3.61 | 3.46 | 4.92 | 4.32 | 4.55 |  | 4.09 |
| 8 | WEU | 6.73 | 5.87 | 5.04 | 7.02 | 8.25 | 5.62 | 5.72 | 3.89 | 5.25 | **9.42** | 4.83 | 7.26 | 7.79 | 4.40 | 6.36 |

AUS: Australian; CEU: Central Europe; EAF: East Africa; EAS: East Asia; NAF: North Africa; NAM: North America; NEU: Northern Europe; RUS: Russia; SAM: South America; SAS: South Asia; SEE: Southeast Europe; SWE: Southwestern Europe; WAS: West Asia; WEU: Western Europe. The minimum and maximum distances are bolded

**Table 6.** Factor loadings after Varimax rotation for 11 studied traits in the combination of two years on 14 geographical origin.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Trait |  | Factor 1 |  | Factor 2 |  | Factor 3 |
| PLH (cm) |  | -0.85 |  | 0.07 | -0.06 |
| NBP |  | 0.14 |  | -0.06 |  | 0.90 |
| NCP |  | 0.04 |  | 0.74 |  | 0.55 |
| CDI (mm) |  | 0.88 |  | 0.25 |  | 0.03 |
| CSN |  | -0.08 |  | 0.85 |  | -0.35 |
| TSW (g) |  | 0.95 |  | 0.23 |  | 0.07 |
| YPP (g) |  | 0.55 |  | 0.75 |  | 0.15 |
| SLE (mm) |  | 0.93 |  | 0.05 |  | 0.12 |
| SWI (mm) |  | 0.95 |  | -0.04 |  | 0.16 |
| OIL (%) |  | 0.73 |  | -0.04 |  | 0.13 |
| PRO (%) |  | -0.21 |  | 0.07 |  | 0.14 |
| Variance proportion |  | 49.52 |  | 17.66 |  | 11.33 |
| Cumulative (%) |  | 49.52 |  | 67.18 |  | 78.51 |

PLH: Plant height; NBP: Number of branches per plant; NCP: Number of capsules per plant; CDI: Capsule diameter; CSN: Capsule seeds number; TSW: Thousand seed weight; YPP: Yield per plant; SLE: Seed length; SWI: Seed width; OIL: Oil content; PRO: Protein content

**Table 7.** Factor loadings after Varimax rotation for 11 studied traits in 2016, 2017 and combination of two years on 120 linseed genotypes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | 2016 |  | 2017 |  | Combined |
| Trait |  | Factor 1 | Factor 2 |  | Factor 1 | Factor 2 |  | Factor 1 | Factor 2 |
| PLH (cm) |  | -0.493 | -0.201 |  | -0.344 | -0.012 |  | -0.451 | -0.349 |
| NBP |  | 0.273 | 0.717 |  | 0.041 | 0.416 |  | 0.199 | 0.596 |
| NCP |  | 0.011 | 0.932 |  | -0.097 | 0.927 |  | -0.040 | 0.899 |
| CDI (mm) |  | 0.795 | 0.064 |  | 0.852 | 0.023 |  | 0.801 | 0.018 |
| CSN |  | -0.247 | -0.166 |  | 0.119 | 0.064 |  | -0.144 | 0.049 |
| TSW (g) |  | 0.783 | 0.187 |  | 0.848 | 0.218 |  | 0.865 | 0.231 |
| YPP (g) |  | 0.286 | 0.729 |  | 0.178 | 0.904 |  | 0.318 | 0.792 |
| SLE (mm) |  | 0.886 | 0.199 |  | 0.896 | -0.021 |  | 0.918 | 0.093 |
| SWI (mm) |  | 0.915 | 0.162 |  | 0.904 | -0.054 |  | 0.934 | 0.093 |
| OIL (%) |  | 0.398 | 0.065 |  | 0.297 | 0.240 |  | 0.472 | 0.153 |
| PRO (%) |  | -0.103 | -0.126 |  | -0.001 | -0.075 |  | 0.042 | -0.196 |
| Variance proportion |  | 38.30 | 15.82 |  | 31.96 | 17.49 |  | 37.52 | 15.40 |
| Cumulative (%) |  | 38.30 | 54.12 |  | 31.96 | 49.45 |  | 37.52 | 52.93 |

PLH: Plant height; NBP: Number of branches per plant; NCP: Number of capsules per plant; CDI: Capsule diameter; CSN: Capsule seeds number; TSW: Thousand seed weight; YPP: Yield per plant; SLE: Seed length; SWI: Seed width; OIL: Oil content; PRO: Protein content