Whole genome diversity, population structure and linkage disequilibrium analysis of chickpea (*Cicer arietinum* L.) genotypes using genome-wide DArTseq-based SNP markers

Somayeh Farahani¹, Mojdeh Maleki¹, Rahim Mehrabi², Homayoun Kanouni³, Reza Talebi⁴*

1- Department of Plant Protection, Varamin-Pishva Branch, Islamic Azad University, Varamin, Iran
2- Department of Biotechnology, College of Agriculture, Isfahan University of Technology, P.O. Box 8415683111, Isfahan, Iran
3- Kordestan Agricultural and Natural Resources and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Sanandaj, Iran.
4- Department of Agronomy & Plant Breeding, College of Agriculture, Sanandaj Branch, Islamic Azad University, Sanandaj, Iran.

Corresponding Author:
Reza Talebi
Assistant professor
Islamic Azad University, Sanandaj Branch
Sanandaj, Iran.
P.O.Box:618
Email: srtalebi@yahoo.com, rezatalebi56@gmail.com
Abstract

Characterization of genetic diversity, population structure and linkage disequilibrium is prerequisite for proper management of breeding programs and conservation of genetic resources. In this study, 186 chickpea genotypes including advanced “Kabuli” breeding lines and Iranian landrace “Desi” chickpea genotypes were genotyped using DArTseq-Based SNP markers. Out of 3339 SNPs, 1152 markers with known chromosomal position were selected for genome diversity analysis. The number of mapped SNP markers varied from 52 (LG8) to 378 (LG4), with an average of 144 SNPs per linkage group. The chromosome size that covered by SNPs varied from 16236.36 kbp (LG8) to 67923.99 kbp (LG5), while LG4 showed higher number of SNPs, with an average of 6.56 SNPs per Mbp. Polymorphism information content (PIC) value of SNP markers ranged from 0.05 to 0.50, with an average of 0.32, while the markers on LG4, LG6 and LG8 showed higher mean PIC value than average. Un-weighted Neighbor Joining cluster analysis and Bayesian-based model population structure grouped chickpea genotypes into four distinct clusters. Principal component analysis (PCoA) and Discriminant Analysis of Principal Component (DAPC) results were consistent with that of the cluster and population structure analysis. Linkage disequilibrium (LD) was extensive and LD decay in chickpea germplasm was relatively low. A few markers showed $r^2\geq0.8$, while 2961 pairs of markers showed complete LD ($r^2=1$) and a huge LD block was observed on LG4. High genetic diversity and low kinship value between pairs of genotypes suggesting the presence of a high genetic diversity among studied chickpea genotypes. This study also demonstrated the efficiency of DArTseq-based SNP genotyping for large scale genome analysis in chickpea. The genotypic markers provided in this study are useful for various association mapping studies when combined with phenotypic data of different traits such as seed yield, abiotic and biotic stresses and therefore can be efficiently used in breeding programs to improve chickpea.

Key words: Chickpea, genetic diversity, Linkage Disequilibrium, DArTseq-SNP markers
1. Introduction

Chickpea (*Cicer arietinum* L.) is an important legume food crop that is currently cultivated in wide ranges of climatic regions across the world in more than 45 countries [1]. It is a second largest cultivated legume globally due to its high protein content and plays important role in human feed and nutritional security in most low income agricultural-based communities such as Asia and Africa [2]. Chickpea is a diploid (2n=2x=16) with approximate genome size of 931 Mbp [3] and comprised of two types; *desi* and *kabuli* cultivars that are distinctly different in agro-morphological characteristics such as seed shape, flower color, growth habit and genome composition that recently sequenced [2, 4, 5]. Both types of chickpea genotypes were grown worldwide, but *desi* type mainly cultivated in Ethiopia and Indian subcontinent [3]. The average world yield of chickpea is much lower than its potential yield under favorite conditions due to narrow genetic base of cultivated chickpea worldwide resulting in vulnerability of many cultivars to various biotic and abiotic stresses [6, 7]. Therefore, characterization of diverse germplasm is a fundamental prerequisite for plant breeders to select proper parental lines and utilize them into breeding programs. Classical breeding techniques based on morphological traits able to characterize genotypes based on their phenotypic characters, but these markers are limited in number, influenced by environment and often have epistatic interaction with other traits. Molecular markers reflect the genetic diversity at the DNA level and able to visualize the accurate genetic diversity between genotypes [8]. In chickpea, different DNA markers such as random amplified polymorphism DNA (RAPD) [9, 10, 11], inter-simple sequence repeat (ISSR) [12, 13], amplified fragment length polymorphism (AFLP) [14, 15] and simple sequence repeats (SSR) [16, 17] has been used for genetic diversity analysis in different germplasm.
During the last decade, single nucleotide polymorphism (SNP) markers have been developed and increasingly utilized as highly preferred molecular markers in various crop species because of their wide genome coverage, codominant inheritance, chromosome-specific location, low cost and fast tracking in compare to other PCR-based molecular markers [18, 19]. SNP markers mainly developed based on next generation sequencing technology. Fast development of SNP markers through genotyping-by-sequencing (GBS) has paved the road to facilitate genomics-assisted breeding through quantitative trait loci (QTL) and genome-wide association analysis in diverse crops [20, 21]. Recently, diversity array technology (DArT) developed a GBS method called “DArTseq” for genotyping with high density SNP in different crop species such as wheat [22], common bean [23], sesame [8], tomato [24], snake melon [25] as well as in chickpea [26]. In this study, we used DArTseq-based SNP markers for genetic diversity, population structure and linkage disequilibrium analysis in 186 chickpea genotypes comprised of advanced Kabuli breeding lines and landrace Desi accessions.

2. Materials and Methods

2.1. Plant materials and DNA extraction

The 186 chickpea genotypes (Supplementary Table S1), including 20 Iranian landrace Desi accessions and 166 Kabuli advanced breeding lines supplied by International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and International Center for Agriculture Research in the Dry Areas (ICARDA) were employed for genetic diversity, linkage disequilibrium and population structure analyses using DArTseq-based SNP markers. Fresh leaves of each genotype (pooled sample of ten plants per genotypes) were used for DNA extraction using the
cetyltrimethyl ammonium bromide (CTAB) method [27], with minor modification. DNA concentrations were measured using spectrophotometer and adjusted to 50 ng/µl.

**Genotyping by DArTseq platform**

All chickpea genotypes were genotyped using sequencing-based DArT genotyping platform. This method is based on methyl filtration and next-generation sequencing platforms as described before [28, 29]. Initially, we received 6678 SNPs from DArT Pty Ltd., which were polymorphic across chickpea genotypes. First, the markers with unknown chromosome position were removed from the analysis. Data set filtered for minor allele frequency (MAF) lower than 0.1 and also for missing data that higher than 20%. Chickpea is self-pollinated crop, therefore, SNP markers that showed heterozygosity more than 5%, were also removed from analysis. Overall, 1152 SNPs remained for further analysis of genetic diversity, population structure and linkage disequilibrium in studied chickpea genotypes (summarized in Table S2). Quality of SNP markers were determined by the parameters “reproducibility” and “call rate” as described previously [30].

**2.2. Data analysis**

Polymorphism information content (PIC) values for SilicoDArT and SNP markers were calculated using PowerMarker v.3.25 [31]. Genetic distance and kinship matrix between pair of 186 chickpea genotypes based on 1152 SNP markers were computed using the identity-by-state (IBS) method implemented in TASSEL v.5.2.37 [32]. Cluster analysis of chickpea genotypes based on Un-weighted Neighbor Joining (UNJ) [33] were imputed in DARwin ver 5.0 software [34]. Principal Component Analysis (PCA) for genotypes was imputed in TASSEL v.5.2.37 and the first two components were used for scatter plot distribution in XLSATA 2012 (Addinsof,
New York, USA; www.xlstat.com). Linkage disequilibrium (LD) for SNP markers was implemented in TASSEL v.5.2.37 and graphical LD decay imputed by GAPIT R package [35].

3. Results

3.1. SNP markers quality and diversity

The 186 chickpea genotypes analyzed by DArTseq-SNPs. The initial data set consisted of 3339 SNPs and after filtering data for some quality parameters including minor allele frequency lower than 0.1, missing data ≥10% and also SNPs with unknown chromosome position, a total of 1152 SNPs were selected for genome diversity analysis (Supplementary Table S2). The number of mapped SNP markers varied from 52 (LG8) to 378 (LG4), with an average of 144 SNPs per linkage group (Table 1). As shown in Table 1, the chromosome size that covered by SNPs varied from 16236.36 kbp (LG8) to 67923.99 kbp (LG5), while LG4 showed higher number of SNPs, with an average of 6.56 SNPs per Mbp (Table 1). Quality parameters such as call rate and average reproducibility were 0.97 and 0.98, respectively. PIC value of SNP markers ranged from 0.05 to 0.50, with an average of 0.32, while the markers on LG4, LG6 and LG8 showed higher mean PIC value than average (Table 1).

3.2. Genetic distance and relatedness between chickpea genotypes

Kinship coefficient between pairs of chickpea genotypes varied from -0.83 to 3 (on a scale of -3 to 3). Overall, 66% of the pairs of 186 chickpea genotypes had kinship values of ≤0.1 (Supplementary Table S3). Chickpea genotypes grouped into four distinct groups according to kinship matrix obtained by DArTseq-SNP markers (Fig 1). In order to identify the most similar pairs of genotypes, genetic distance matrix computed between pairs of genotypes, which varied from 0.2 to 0.56, with an average of 0.31 (Supplementary table S4). The large
proportion (88%) of pairs of genotypes showed genetic distance $\geq 0.25$ (Supplementary table S4). The Un-weighted Neighbor Joining cluster analysis based on DArTseq-SNP markers differentiated the 186 chickpea genotypes into four clusters (Fig 2). Clusters showed relatively same number of genotypes, although cluster III and IV were larger groups of chickpea genotypes compared to … . There were no relationships between cluster grouping and pedigree of chickpea genotypes, although most of Desi chickpea landraces grouped in cluster II. Most of the genotypes (which ones, can be more specific?) used in this study have been used as parental lines or have similar genetic background, so mixture of pedigree observed in all clusters.

3.3. Population structure and discriminate analysis of principal coordinate (DAPC)

Population structure analysis of the genotypes based on Bayesian model implemented in STRUCTURE software grouped genotypes at K=4 (Fig 3). Four sub-populations based on SNP markers showed relatively low genetic divergence among sub-populations (from 0.19 for POP4 to 0.28 for POP1), while high divergence between sub-populations was observed (Table 2). Genetic diversity among the populations based on net nucleotide distance revealed a higher distance between POP1 and POP2 compared to the genetic distance between POP1 and POP2 with POP3 and POP4 (Table 1). Mean fixation index of sub-populations ranged from 0.56 (POP1) to 0.65 (POP3) (Table 2). Principal component analysis (PCA) based on DArTseq-SNP markers revealed four distinct groups of chickpea genotypes and two principal components accounting for 75.18% of total variation (Fig 4). Discriminant Analysis of Principal Component (DAPC) also was employed to fine the fitting population structure based on DArTseq-SNP markers. The lowest BIC value was obtained at K=4, therefore, three discrimination function were detected which explained 30.17, 25.86 and 19.06% of variation
between sub-groups (Fig 5). Results from the DAPC analysis were consistent with the results from population structure analysis.

3.4. Linkage disequilibrium analysis

Distribution of LD within chromosomes based on 1152 DArTseq-SNP markers showed extensive LD decay, as in the entire population from 56325 marker pairs, 21660 (38.4%) intra-chromosomal pairs showed significant level ($P<0.001$) of LD. Mean $r^2$ value was 0.22, while the critical $r^2$ value was 0.33. The overall LD decay in chickpea germplasm was relatively low and few markers showed $r^2 \geq 0.8$. Nevertheless 2961 pairs of markers showed complete LD ($r^2=1$), although a huge LD block was observed on LG4 (Fig 6).

4. Discussion

Characterization of genetic diversity in crop species is prerequisite for efficient conservation and utilization of germplasm and developing breeding programs [10, 37]. Iran, Afghanistan, Turkey Indian subcontinent and Lebanon contain a large number of chickpea landraces and have been previously identified as the centers of origin and/or diversity of chickpea by Vavilov [38]. In most chickpea growing areas worldwide are under biotic and abiotic stresses resulting in the low seed yield production. This is mainly due to narrow genetic base and lack of desirable traits in cultivated genotypes [37, 39]. Therefore, incorporation of desirable traits with high rate of allelic frequencies and transgressive segregation through introduction of diverse genotypes from diverse sources into breeding programs is required to improve tolerance to various stresses and to maximize seed yield and quality [26, 40]. To date, different molecular markers have been utilized for genetic diversity analysis in chickpea [11, 12, 13, 14, 17]. DArTseq-SNP markers conducted by GBS technology is a rapid, low cost and efficient method for genotyping providing an broad genome coverage and therefore has
been increasingly used in different plants species [23, 24, 25] as well as in chickpea [26]. In this study, we employed DArTseq-SNP markers for population structure and LD analysis in a set of 186 diverse chickpea genotypes. Among 3339 generated SNPs, 1152 markers (34.5%) employed for further analysis. The average call rate and reproducibility of SNP markers was 0.97 and 0.98, respectively, which was higher than to values previously reported in watermelon [41], common bean [23] and was consistent with the value reported for wheat [42, 43]. Average PIC value of SNP markers was 0.32 and more than half of the markers showed PIC value higher than 0.25, which suggests the sufficient efficiency of these markers that has been reported previously for DArTseq-SNP markers in chickpea [19, 26]. Physical distribution of the mapped markers with known positions on chickpea linkage groups (Table 1) showed that LG4 had a higher marker density compared to other LGs which is consistent with previous studies [19, 26]. Average inter-chromosomal LD decay of SNP markers showed long distances for marker pairs in LD (Fig 6), which may be attributed to genetic admixture apart from the genetics or physical distances as has been previously reported in chickpea and other crops [26, 44]. Kinship values calculated between pairs of genotypes are a reliable factor for understanding the extent relatedness between chickpea genotypes. It is generally accepted that the kinship value close to 0.5 or higher refer to highly similar genotypes [24, 45]. Average genetic distance (GD) between pairs of genotypes was 0.31, while 88% of pairs of genotypes showed GD more than 0.25 indicating that high genetic variability were presented among chickpea genotypes. Cluster analysis, principal component and model-based population structure analyses based on DArTseq- SNP markers revealed three and four distinct groups of chickpea genotypes although weak correlation between cluster grouping and pedigree of chickpea genotypes was identified. This might be due to the
fact that these genotypes have been utilized in different breeding programs and thus they have same parental in their pedigree, as shown in Table S1 that FLIP98-28C and FLIP98-52C included in pedigree of large number of chickpea genotypes. Therefore, it is more likely that recent breeding activities and incorporating same genotypes in parental crossing as well as domestication and selection of similar chickpea genotypes by farmers during past centuries had a significant impact on global chickpea genetic structure resulting in genotypic admixture as shown in this study and previous reports [46, 47. Therefore, breeding activities with different strategies and incorporating same genotypes in parental crossing may led to significant impact on global chickpea genetic structure [48]. Interestingly many of the genotypes used in this research were studied for the first time and has not been previously used in breeding programs. Therefore, these genotypes are appropriate novel sources that may possess useful genes and can be used in breeding programs to broaden chickpea gene pool.

5. Conclusion

In this study, we employed DArTseq-SNP markers for genome diversity, population structure and LD analyses in a mini-core collection of advanced chickpea breeding lines and landraces. This information can be used in genome-wide association and marker-assisted selection in chickpea breeding programs. DArTseq-SNP markers that used in this study can be used for the identification of molecular markers linked with various morphological traits or resistance genes for biotic and abiotic stresses. The appropriate DArTseq-SNP markers for different gene/traits of interest can be used for cloning and designing kompetitive allele specific PCR (KASP) in chickpea.
Supplementary materials:

Table S1: List and pedigree of 186 chickpea genotypes used for DArTseq-SNP genotyping.
Table S2: List of DArTseq-SNP marker used in this study.
Table S3: Kinship matrix between pair of 186 chickpea genotypes based on DArTseq-SNP markers.
Table S4: Genetic distance matrix of 186 chickpea genotypes based on DArTseq-SNP markers.

Author Contributions: S.F carried out the experiment and prepared the draft of manuscript. R.T and R.M conceived and design the experiment, wrote manuscript and analysis the data. H.K prepared the seed material and design the experiment. M.M contributed in data analysis.

Funding: The author(s) received no financial support for the research, authorship, and/or publication of this article.

Acknowledgment: The study was conducted as a part of PhD thesis of first author and supported by Islamic Azad University, Varamin-Pishva Branch, Iran. We are grateful to Kordestan Agricultural and Natural Resources and Education Center for providing some wheat genotypes used in this study.

Conflicts of Interest: The authors declare no conflict of interest.
References


35. Van Raden, P.M.; Efficient methods to compute genomic predictions. *J. Dairy. Sci.* **2008**, *91*, 4 414–4423


Table 1. Polymorphism information content (PIC), call rate, average reproducibility and distribution of DArTseq-SNPs on chickpea chromosomes.

<table>
<thead>
<tr>
<th>Linkage group (Chromosome)</th>
<th>Number of SNPs</th>
<th>Chromosome size (kbp)</th>
<th>Mean of SNPs per Mbp</th>
<th>PIC range (Mean)</th>
<th>Call rate</th>
<th>Average reproducibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG1</td>
<td>192</td>
<td>44634.56</td>
<td>4.20</td>
<td>0.05-0.49 (0.23)</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>LG2</td>
<td>89</td>
<td>36915.99</td>
<td>2.41</td>
<td>0.05-0.49 (0.32)</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>LG3</td>
<td>105</td>
<td>61351.17</td>
<td>1.71</td>
<td>0.05-0.49 (0.30)</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>LG4</td>
<td>378</td>
<td>57562.47</td>
<td>6.56</td>
<td>0.05-0.50 (0.36)</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>LG5</td>
<td>74</td>
<td>67923.99</td>
<td>1.08</td>
<td>0.09-0.48 (0.31)</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>LG6</td>
<td>141</td>
<td>63087.8</td>
<td>2.34</td>
<td>0.05-0.50 (0.34)</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>LG7</td>
<td>121</td>
<td>54252.93</td>
<td>2.23</td>
<td>0.05-0.50 (0.31)</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>LG8</td>
<td>52</td>
<td>16236.36</td>
<td>3.20</td>
<td>0.06-0.49 (0.34)</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>Total</td>
<td>1152</td>
<td>67923.99</td>
<td>16.96</td>
<td>0.05-0.50 (0.32)</td>
<td>0.97</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table 2. Genetic divergence among (Net Nucleotide Distance) and within (expected heterozygosity) population, proportion of membership and mean value of Fst observed from the study of population structure of 186 chickpea genotypes using DArTseq-SNP markers.

<table>
<thead>
<tr>
<th>Population</th>
<th>POP2</th>
<th>POP3</th>
<th>POP4</th>
<th>Expected Heterozygosity</th>
<th>% Of membership</th>
<th>Mean fixation index (Fst)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POP1</td>
<td>0.44</td>
<td>0.25</td>
<td>0.26</td>
<td>0.28</td>
<td>0.20</td>
<td>0.56</td>
</tr>
<tr>
<td>POP2</td>
<td>0.33</td>
<td>0.30</td>
<td>0.30</td>
<td>0.22</td>
<td>0.19</td>
<td>0.62</td>
</tr>
<tr>
<td>POP3</td>
<td>0.16</td>
<td>0.17</td>
<td>0.36</td>
<td>0.17</td>
<td>0.36</td>
<td>0.65</td>
</tr>
<tr>
<td>POP4</td>
<td>0.19</td>
<td>0.26</td>
<td>0.26</td>
<td>0.19</td>
<td>0.26</td>
<td>0.63</td>
</tr>
</tbody>
</table>
Figure 1. Heatmap plot of kinship matrix displaying relationships of 186 chickpea genotypes based on DArTseq-SNP markers. The details of members of these groups are presented in Table S3.
Figure 2. The neighbor-joining cluster analysis using DArTseq-SNP markers for grouping 186 chickpea genotypes.
Figure 3. Determination of the optimal value of K and population structure of 186 chickpea genotypes using DArTseq-SNP markers
Figure 4. Principal coordinate analysis of 186 chickpea genotypes based on DArTseq-SNP markers.

Figure 5. The percentage of cumulative variance for the retained PCA eigen vectors and Scatter plot from the DAPC analysis for 186 chickpea genotypes used to determine the optimal $k$ number of clusters using DArTseq-SNP markers.
Figure 6. Linkage disequilibrium (LD) measured $r^2$ plotted vs. the physical map (Mb) between pairs of DArTseq-SNP markers in a panel of 186 chickpea genotypes, which show the huge LD decay on LG4.