Biochemical Genetics (2019) 57:98–116
https://doi.org/10.1007/s10528-018-9877-2

ORIGINAL ARTICLE

Genetic Diversity and Molecular Characterization of Iranian Durum Wheat Landraces (*Triticum turgidum durum* (Desf.) Husn.) Using DArT Markers

Farzad Fayaz[1] · **Mostafa Aghaee Sarbarzeh**[2] · **Reza Talebi**[1] · **Amin Azadi**[3]

Received: 6 January 2018 / Accepted: 14 July 2018 / Published online: 26 July 2018

Abstract

Durum wheat is grown mainly in rain-fed regions of Iran and the Mediterranean district under stressful conditions. Different environmental conditions and agricultural practices among ancient communities have led to the development of locally adapted genotypes known as landraces. Landraces are a valued source of genetic variety and show definite adaptation to local environmental conditions according to their home of origin. This study aimed to explore linkage disequilibrium (LD) analysis and the population structure and genetic diversity of Iranian durum wheat landraces. In this study, population structure and genome-wide LD were investigated in 129 durum landrace accessions using 1500 DArT markers. Both structure and discriminant analysis of principal components obviously subdivided the sample collection into seven distinct groups centered on key ancestors and regions of origin of the germplasm. Genetic diversity among the populations was primarily within population (68 vs. 32%). Mean LD values across the entire population sample decayed below r^2 of 0.11 after 1 cM. LD decay of genomes A and B of Iranian durum wheat landrace is approximately 2–3 cM ($r^2 = 0.11$) and approximately 0.5 cM ($r^2 = 0.12$), respectively. Altogether, low LD decay, a high number of subpopulations, and the high existence of genetic diversity among and within populations were characteristics of the Iranian durum landrace collection. Hence, the existing genetic diversity within the population can be associated with the very long evolutionary history of plants in Iran. The populations we studied are hence presented as a valuable resource that can be used in basic and applied research in durum wheat breeding.

Keywords Durum wheat · Linkage disequilibrium · Structure analysis · Genetic diversity · Landrace

---

✉ Farzad Fayaz
Far.Fayaz@gmail.com

Extended author information available on the last page of the article

Introduction

Durum wheat (*T. turgidum durum* $2n = 4x = 28$; BBAA genome) is the third main cereal crop worldwide and has a long history, with agriculture origins in the Near East 10,000 years ago (Badr et al. 2000). Any forthcoming improvement in durum wheat relies on the accessibility and identification of genetic resources specifically in a center of origin or diversity, such as the Fertile Crescent (the region where the first settled agricultural communities of the Middle East and Mediterranean basin are thought to have originated by the early 9th millennium BCE.). Iran, where the first cereals were domesticated by ancient civilizations, makes up a considerable share of this focus (Harlan 1986). Presently, durum wheat is grown mainly in rain-fed zones of Iran and the Mediterranean region under stressful conditions (Moragues et al. 2006). Different environmental conditions and agricultural practices among ancient communities have contributed to the development of locally adapted genotypes known as landraces. Many different terms have served as synonyms in the literature, including “primitive cultivar,” “traditional variety,” or even “conservation variety” (Lopes et al. 2015). Several descriptions have been developed for landraces, but an acceptable definition founded on information about their qualities, use, eco-geographic adaption, cultivation, and management procedures has yet to be agreed upon among all researchers (de Carvalho et al. 2013).

Previous studies have shown that landraces have different characteristics, including variations in morphological- and yield-related traits, quality traits, agro-ecological adaption, abiotic stresses, and resistance to pests that have evolved as a result of natural selection (Alipour et al. 2017). These characteristics are a collection of useful alleles (gene pool) helping landraces of various genotypes cope with environmental stresses (de Carvalho et al. 2013). In spite of recent progress in genetic engineering, landraces still have a role as a source of useful alleles for traditional and advanced breeding programs (Comadran et al. 2009). However, the extended gene pool of durum wheat remains poorly characterized, particularly in Iran.

In accordance with the definition of linkage disequilibrium (LD) is the “non-random association of alleles at different loci.” In other words, LD is study of correlations between polymorphisms (e.g., DArT markers) that are a symbol of population history and any genetic variation phenomena that have occurred previously (e.g., mutations and recombinations) (Flint-Garcia et al. 2003). The rate of LD in each population depends on (i) the generative biology of the organism (i.e., outbreeding vs. inbreeding) and (ii) the population history. Inbreeding plant types have high LD levels due to high levels of homozygosity, with non-random associations of alleles spanning large distances (Kovi et al. 2015). In maize (102 inbred lines), rapid intergenic LD decay ($r^2 < 0.1$ within 1.5 kb) has been reported by several researchers. However, LD spreads to larger than 0.1 kB for *adh1* and *y1* loci in elite maize populations as reported previously (Yan et al. 2009). Two synthetic populations that were derived from recurrent selection had different LD responses: one population showed an increase in LD over 12 generations, whereas the other showed a decrease. Therefore, the different degrees of LD decay depend on the levels of population bottlenecks, i.e., the progression from diverse landraces to diverse inbreds to elite inbreds (Laidò et al. 2014). In contrast, *Arabidopsis* (*Arabidopsis thaliana*) is a highly selfing species, and its LD extends much further. Diverse population accessions of *Arabidopsis* LD decay rates differ between 250 and 2500 kb (equivalent to 1–10 cM) (Flint-Garcia et al. 2003).

Durum wheat was tested for linkage disequilibrium (LD) using different markers distributed across the genome, and researchers have found various rates of LD decay. The LD between adjacent locus pairs extended ($r^2 > 0.2$) approximately 2–3 cM, on average, but some regions of the bread and durum wheat genomes showed high levels of LD ($r^2 = 0.7$ and 1.0, respectively) extending 41.2 and 25.5 cM, respectively (Somers et al. 2007). Hao et al. reported that the mean LD decay distance for the Chinese bread wheat landraces at the whole genome level was <5 cM, while a higher LD decay distance of 5–10 cM was seen in modern varieties (Hao et al. 2011). Another LD analysis study revealed that there was a shorter LD decay distance in Chinese winter wheat compared with other wheat germplasm collections. The LD decay distance range was estimated to be between 17.4 and 2.2 cM ($r^2 > 0.1$, $P < 0.001$) (Chen et al. 2012). Comparing three different materials (old landraces, semi-modern varieties, and modern varieties) on the B genome showed that LD decay increased from old cultivars to modern varieties (Horvath et al. 2009). Genetic diversity and LD were examined in 376 Asian and European accessions of bread wheat (*Triticum aestivum* L.) on chromosome 3B. LD analysis showed that the European population had a meaningfully higher mean r^2 value (0.23) than the Asian population (0.18), indicating a stronger LD in the European material (Hao et al. 2010).

Various types of genetic markers for studying the genotypic diversity of plant germplasm, including wheat landraces, have been developed and used. Diversity array technology (DArT) markers have been confirmed to be beneficial for the investigation of wheat and other crops over the previous decade (Nadeem et al. 2017; El-Esawi et al. 2018). Since DArT markers are biallelic and dominant, it is cost effective and time saving to use them, and they are employed in genome-wide genotyping (Jaccoud et al. 2001). The beneficial characteristics of such markers are responsible for successful use in genotyping in a variety of studies, in population structure studies, and for the genetic mapping of wheat (Akbari et al. 2006; Stodart et al. 2008; Zhang et al. 2011; Marone et al. 2012). The population stratification is clarified by population structure analysis and they help in the association between phenotypes and genotypes by using a better mixed model analysis. (Pritchard and Rosenberg 1999). Using DArT markers to analyze population structure encompasses the entire genome and offers a comprehensive depiction of the population. This study aimed to explore the LD analysis, population structure, and genetic diversity of Iranian durum wheat landraces.

Materials and Methods

Plant Material

In this study, Iran's National Gene Bank provided a collection of 150 diverse accessions of durum wheat (Fayaz et al. 2013). The obtained collection consisted of landraces and cultivated varieties. The panel was composed of 22 advanced cultivars, 120 landraces, and 8 common wheats in which most of the accessions were landraces, although some were cultivars. Table 1 provides the names, types, and origins of the collection.

Genotypic Data

Using the protocol recommended DArT company, adult plant leaves of ten individuals were used to extract the DNA of 150 durum wheat accessions (https://www.diversityarrays.com/files/DArT_DNA_isolation.pdf). Genotyping with DArT markers was executed by Triticarte Pty. Ltd. (Canberra, Australia; http://www.triticarte.com.au). A total of 1500 markers out of 2609 DArT markers tested yielded good quality calls and were found suitable for our analysis (DArT markers, alleles that had at a low frequency ($P < 0.05$) or over than 10% missing data were excluded from the analysis). Eight accessions of common wheat in germplasm collections were determined to evaluate structural analysis and the consistency of scoring for each DArT marker. We found 21 genotypes out of 150 accessions in the panel of genotypes performed badly in the assay, so they were disqualified from subsequent analysis. The information provided by Triticarte Pty. Ltd. determined the chromosomal locations of the DArT markers (http://www.triticarte.com.au/). DArT marker map positions on chromosomes based on Chinese spring wheat varieties (Peleg et al. 2008).

Statistical Analyses

Gene diversity, demarcated as the probability that two randomly selected alleles from the population are dissimilar (Weir 1996), was calculated using GenAlex (Moragues et al. 2006). The population groups were set based on DAPC results. To define the fitting population structure in the collection and estimate the number of subgroups in the panel, two approaches were implemented, first one based on Bayesian method by applying structure software, and second based on non-Bayesian method that is Discriminant Analysis of principal component (DAPC) was employed (Jombart et al. 2010). DAPC analysis was carried out by using R software version 3.2.3 (R Core Team 2014). Population structure was explored using a Bayesian clustering approach to infer the number of clusters (subpopulations) with the software structure version 2.3.4 (Pritchard et al. 2000). The algorithm tries to recognize genetically dissimilar subpopulations based on the patterns of DArT allele frequencies. 1500 DArT markers distributed on whole of chromosomes were designated for structure analysis (Falush et al. 2003). The number of subpopulations (k) was set from two to fifteen,

Table 1 A list of the genotypes used in this study

Origin	Number	Genotype ID or name
Azerbaijan, East	8	7, 444, 734, 724, 812, 739, 796, 719
Iran gene bank	5	717, 727, 744, 797, 92
Iran[a]	39	735, 107, 816, 130, 131, 743, 219, 752, 223, 805, 817, 803, 748, 740, 246, 771, 731, 774, 765, 775, 772, 806, 312, 746, 314, 811, 718, 787, 809, 361, 728, De (Dena)^b, K (Karghi)^b, A^b (Arya), 781, 783,730, 818 (Zardak),761
Commercial cultivars	17	762^b, 750^b, 788^b, 751^b, 782^b, 815^b, 777^b, 733^b,720^b, 741^b,793^b, 795^b, 726^b, Y (Yavaros)^b, 759^b, 778^b, 767^b
Kermanshah	17	204, 784, 792, 275, 776, 701, 747, 780, 722, 723, 737, 769, Sy (Syri and four), D8 (D-86-14), Se (Sefid), 705, 703
Kurdistan (Local check)	1	P (Parisht)
Khorasan	28	756, 790, 732, 706, 779, 821, 54, 736, 58, 710, 754, 316, 711, 319, 810, 791, 714, 721, 773, 708, 786, 813, 807, 789, 763, 822,713, 804
Khuzestan	7	757, 749, 716, 819, 704, 712, 707
Lorestan	28	702, 768, 74, 794, 715, 814, 766, 770, 802, 191, 738, 808, 745, 729, 785, 725, 758, 742, 700, 764, 798, 799, 753, 755, 350, 447, 709, 456
Total	150	

Genotypes with no symbol are Iranian landraces. [a]The province of origin for the genotypes in this group are unknown. [b]Commercial cultivars while almost no preceding information was used to delineate the clusters. Runs with $k = 2$ in 15 were repeated at least five times. For each run, the burn-in period and the simulation run length were both set at 10,000 replications of Markov Chain Monte Carlo (MCMC) with an admixture model and correlated allele frequency (Falush et al. 2003). The membership probability is the basis for applying a method used to estimate the different clusters and to assign individuals to subpopulations. We used the run that assigned all of the lines to a single cluster at a probability > 0.5 (set 0.5 as the threshold to define the membership of a genotype to a specific cluster or subpopulation). The probability of best fit into each number of assumed clusters (K) was estimated by an ad hoc statistic ΔK based on the rate of change in the log probability of data between consecutive K values (Evanno et al. 2005). For better detection number of subpopulations, the output analysis structure software was exported online web site Structure Harvester (Earl and vonHoldt 2012). Diversity structure pattern in the current population panel (with 1500 DArT markers) was analyzed using Discriminant Analysis of Principal Component (DAPC) method by R package "adegenet" as previously described by Garavito et al. 2016. Finally, to explain genetic relationship between the accessions, unrooted NJ trees were built with DARwin software version 6.0.16. (Perrier and Jacquemoud-Collet 2006). Tree bootstrap analysis was also computed with DARwin software using 100 iterations.

Genome-wide LD analysis was performed among the panel and subgroups by pair-wise comparison among the DArT markers using Tassel software version 3.0 (Bradbury et al. 2007). Using squared allele frequency and the correlations (r^2) between the pairs of loci (Flint-Garcia et al. 2003), LD was estimated. The loci were considered to be insignificant LD when $P < 0.001$; the rest of the r^2 values were not considered as edifying. Examining the normal LD decay in the entire genome among the panel, significant intra-chromosomal r^2 values were plotted against the genetic distance (cM) among markers.

Results

Population Diversity

Of the 2609 DArT markers for 129 genotypes, about 29.8 and 13.2% were missing data and lacked polymorphism, respectively. After the deletion of data with more than 10% missing and polymorphism less than 5%, about 1500 markers were selected for analysis. The average observed and expected heterozygosity was estimated to be 0.22 and 0.24, respectively (Table 2). The highest number of alleles and effective alleles were in the fifth population, geographically undefined. In total, the number of effective alleles was higher in local subpopulations as compared to the fourth subpopulations (breeding cultivars), which was a predictable issue. The comparison of observed heterozygosity (Ho) and expected heterozygosity (HexP) in 7 subpopulations showed that probably none of the subgroups were in the Hardy–Weinberg equilibrium (due to the inequality of Ho and HexP) (Table 2). The highest heterozygosity was found in populations 5 and 2 (values 0.33 and 0.28) and the lowest in subgroups 4 and 7 (values 0.17 and 0.173). The Shannon

Table 2 Genetic population parameters variation among seven groups identified by DAPC analysis on a diversity panel of 129 Iranian durum wheat landraces and cultivars

Pop	Geographic source	Na	Ne	I	Ho	HexP	Fst
1	Azer (20)	1.53	1.34	0.33	0.21	0.22	0.86
2	Common wheat (7)	1.69	1.5	0.42	0.28	0.34	0.92
3	Khorsan (9)	1.21	1.3	0.27	0.18	0.2	0.94
4	Advanced cultivars (43)	1.32	1.28	0.26	0.17	0.17	0.76
5	Unknown (18)	1.93	1.57	0.5	0.34	0.35	0.8
6	Lor-Kerm (18)	1.41	1.33	0.3	0.19	0.2	0.82
7	Common wheat (4)	1.22	1.3	0.25	0.17	0.23	0.97
Total	129	1.47	1.37	0.33	0.22	0.25	0.08

diversity index also confirmed these results. The Fst values in each subpopulation were approximately close to 1, which can be due to the complete differentiation of subpopulations (zero value means a random and free crossing between the two subpopulations and 1 value means no genetic diversity exchange between the two subpopulations). Conversely, Fst values in each subpopulation can also indicate the similarity of genotypes in each subgroup. If Fst is small, it means that the allele frequency is similar in each subpopulation, and if Fst is large, it indicates different allele frequencies in each subpopulation (Holsinger and Weir 2009). Therefore, subpopulations were completely distinct from the genetic structure and had varying allelic frequencies and variance within each subpopulation (Table 2). The molecular variance analysis table (AMOVA) (Table 3) showed genetic variations between and within populations. The genetic variation between and within subpopulations were 68 and 32%, respectively, which clearly reflects the genetic diversity in the whole population. In other words, in the studied population, there is diversity between and within the population, confirming Fst values and the results obtained from the population structure analysis in this study.

Structure Analysis

The population structure of the panel of 129 durum wheat accessions was examined by means of 1500 DArT markers and a model-based approach in STRUCTURE. The analysis of population divided the population into seven groups. As shown in the ΔK diagram, the ΔK value was maximum at $k=7$, indicating that the population was divided into seven groups (Fig. 1). The first group (red, with 11 genotypes)

Table 3 Analysis of molecular variance (AMOVA) results showing the percent of diversity between and within subpopulations

Source	*df*	SS	MS	Est. var.	%
Among pops	6	7023.428	1170.571	61.204	32
Within pops	122	16,064.672	131.678	131.678	68
Total	128	23,088.101		192.882	100

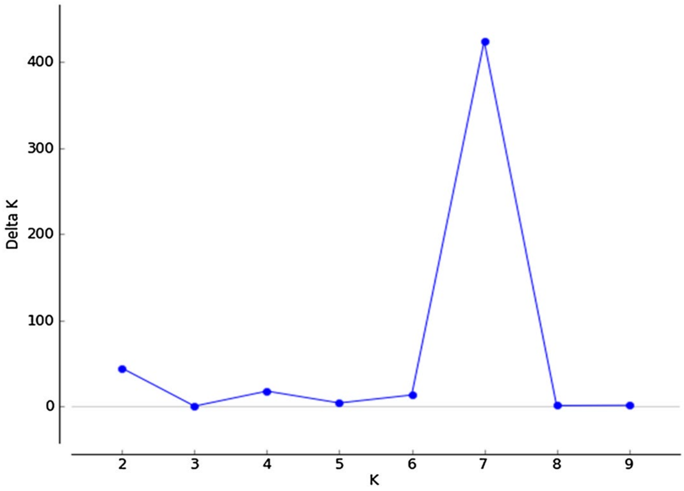

Fig. 1 Structure results using 1500 DArT markers. ΔK values calculated for $K = 1–9$ to determine the number of groups in the Iranian durum wheat landraces

contained common wheat genotypes (Fig. 2a). The second group (green, consisting of 30 genotypes) mainly contained native Iranian genotypes from the geographical area of Lorestan and Kermanshah. The third group (blue, consisting of 15 genotypes) was mainly identified as that of the Khorasan region. The fourth group (yellow, consisting of 13 genotypes) was genetically highly heterogeneous and had genotypes of different geographical origin. The fifth group (purple with 15 genotypes) included native genotypes obtained from East Azarbaijan. The sixth group (turquoise, with 14 genotypes) was heterogeneous like the second group and contained the geographical region of Lorestan and Kermanshah. Finally, the seventh group (bright brown, with 27 genotypes) carried mainly genotypes of advanced cultivars (Fig. 2a). Figure 2b shows the tree diagram of genotypes studied by the NJ method, consistent with the output of the structure software. To facilitate comparison, each branch of the tree graph is shown in the same color as in the structure analysis. At the end of the red tree, the common wheat genotypes are placed that are completely different from other groups. The second group, shown in green, indicates the landrace genotypes of Lorestan and Kermanshah. The third group, shown in blue, identified local genotypes of Khorasan. The fourth group, located in the middle between the common wheat group and other groups (durum landrace wheat) and marked yellow, is the heterogeneous group followed by structure analysis. The fifth group with a purple color contains local Iranian genotypes from the East Azarbaijan province. The sixth and seventh groups, in turquoise and light brown, revealed coordination with subgroups traced in structure.

The results of DAPC were more or less similar to the results of the structure method (Fig. 3a). The graph (Fig. 3b, middle) clearly shows that the lowest BIC value was obtained at $K = 7$, indicated by “x” on the graph. A total of 6 discrimination functions were detected (Fig. 3b). The first six main components explained

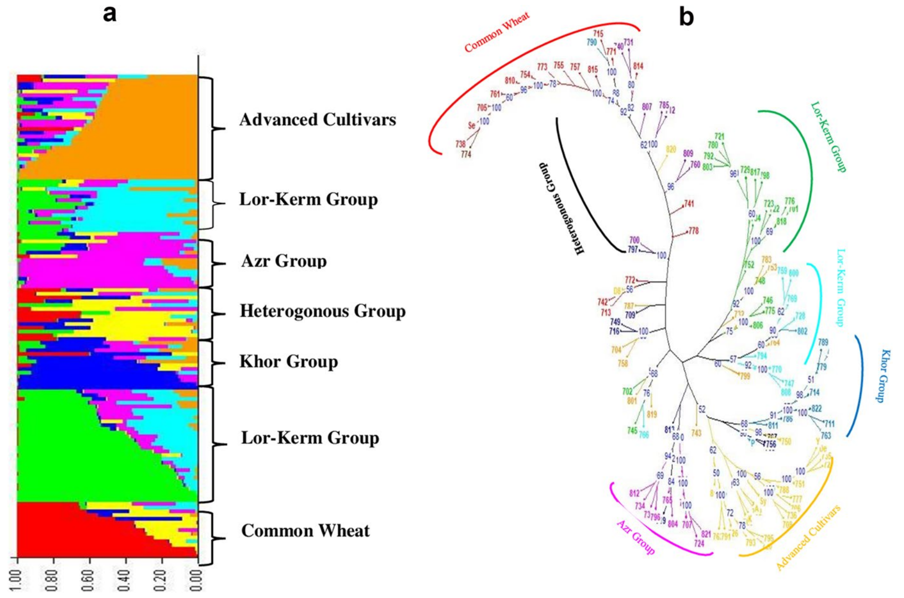

Fig. 2 Population sub-structuring in the collection population. **a** Bayesian clustering of the 129 durum wheat accessions into seven defined groups based on 1500 DArT markers. The number of accessions per group and their respective geographical origin are presented. **b** Neighbor-joining tree based on 1500 DArT markers evaluations. Unrooted tree using the Neighbor-joining algorithm based on Nei's genetic distances between 129 durum landraces wheat. The color patterns are equivalent to the structural analysis. *Azr* Azerbaijan, *Ir* Iran, *Kerm* Kermanshah, *Khor* Khorasan, *Lor* Lorestan (Color figure online)

10.77, 9.33, 7.38, 4.29, 3.1, and 2.51% of the total variance, and ultimately entered the final equation. The optimum number of main components was about 40 principal components, whose validity was verified using the xvalDAPC function (Fig. 3b).

The first group (dark blue, with 20 genotypes) was related to local Iranian genotypes from East Azarbaijan province. The second group (purple, with 7 genotypes) carried a number of common wheat genotypes. The third group (green, with 9 genotypes) consisted of isolated genotypes of the Khorasan area. The fourth group (orange, with 43 genotypes) consisted mainly of genotypes in the geographical area of Lorestan and Kermanshah. The fifth group (red, and 18 genotypes) was diverse genotypes of different geographical origins. The sixth group (blue, and 28 genotypes) was associated with advanced cultivars. The seventh group (bright brown color, with only 4 genotypes) consisted of the remaining genotypes of the second group, common wheat (Fig. 3a).

A comparison of the results of the two methods of structure analysis and DAPC showed high matching. In summary, Group 4 in DAPC was divided into two groups (Groups 2 and 6) in structured analysis. Groups 4 and 5 were matched in DAPC and structure analysis. On the other hand, the second group was divided into two groups in structure analysis (which carried common wheat groups) in DAPC (Groups 2 and 7).

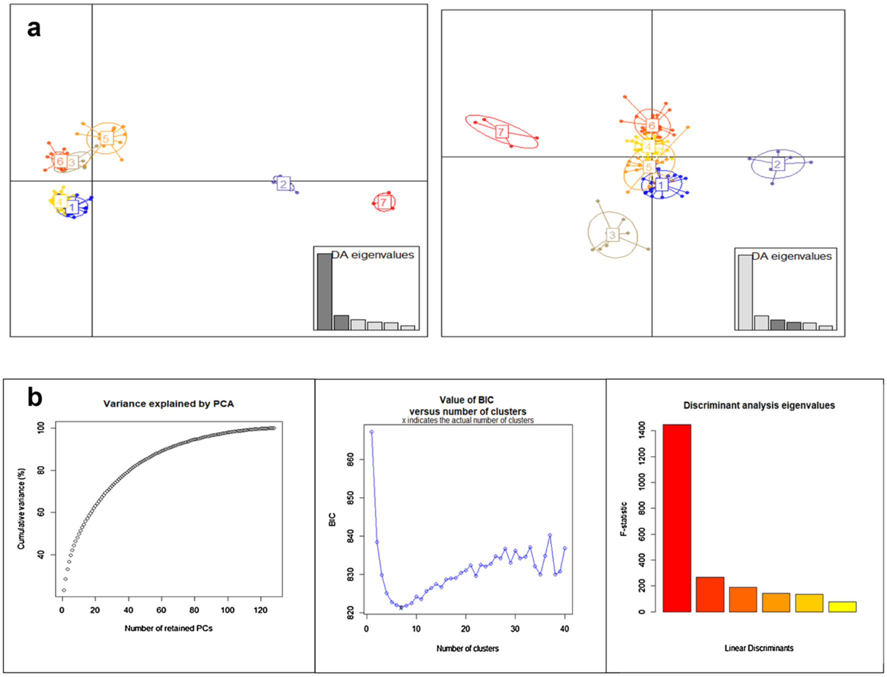

Fig. 3 Genetic structure of the durum wheat population evaluated with 1500 DArT markers. Scatter plot from the DAPC analysis accomplished with 129 *T. durum* accessions. **a** Discriminant axes 1 and 2 (left) and 3 and 4 (right) representing the seven groups (inertia ellipses) determined by the DAPC. Group 1 (dark blue circle, with 20 genotypes) includes individuals from East Azaribajan. Group 2 (purple circle, with 7 genotypes) enclosed number of common wheats. Group 3 (Green circles with 9 genotypes) represents the accessions from Khorasan province. Group 4 (Orange circles with 43 genotypes) was biggest group, mainly made up the accessions from Lorstan and Kermanshah. Group 5 (red circle, with 18 genotypes) is composed of divers and heterogenic Landrace accession with different geographical sources. Group 6 (blue circle, with 28 genotypes) is equivalent to Advanced cultivars and finally Group 7 (brown circle, with 4 genotypes) included the rest of common wheat in this cluster. The rest of common wheat that not clustered in Group 2 was in this Group categorized. **b** The percentage of cumulative variance for the retained PCA eigenvectors(left), the Bayesian information criterion (BIC) used to determine the optimal *k* number of clusters (blue dot) “x” sign show the number of subpopulations (middle), and the *F* statistic of the DAPC (Right) are depicted (Color figure online)

Linkage Disequilibrium

LD Analysis

As mentioned before, LD (linkage disequilibrium) is the non-random association of alleles at different loci, which can be influenced by genetic linkage, genetic drift, population structure, or pressure of population selection. Then, the rate of LD, determines the events that occurred in the genome over time. The results are based on 1500 DArT markers and 129 samples of genotypes, which were used to determine LD. It should be noted that the critical value r^2 LD, or basal LD, was calculated from the inter-chromosomal analysis. The intersect point of the LOESS curve with the baseline in each panel can determine the critical amount of r^2 as the LD decay of each population. The r^2 values of the whole genome, genome A, and genome B, were calculated as 0.11, 0.11, and 0.12, respectively, beyond which LD is assumed to be triggered by genetic linkage. In the studied collection, the distribution of non-linkage r^2 values was calculated to be approximately 0.11. The amount of LD decay of genome A of Iranian durum wheat landrace is approximately 2–3 cM ($r^2=0.11$), indicating a very low amount compared to the value of improved varieties. Studying the graph indicates that a single linkage block (approximately 6 markers) in the area between 90 and 100 cM is associated with this ($r^2=0.65–0.78$); although, an increase in the genetic distance of these markers seems significant (because a genetic distance of more than 50 cM is usually acting independently). These markers are as follows: wpt-3965, wpt-731010, wpt-730729, wpt-667780, wpt-733764, and wpt-664733 (Fig. 4). This could be a result from the principal structure in the studied subpopulations or natural selection over time. Studying the graph of the genome LD B shows that LD decay is approximately 0.5 cM ($r^2=0.12$), representing a very low LD when it is compared to genome A (Fig. 5). In this graph, other blocks of molecular markers can be seen in the area of 60–70 cM in which the values are $r^2=0.4–0.54$. The LD graph for the whole genome (genome A and B) in the total amount of LD decay displays approximately 1 cM ($r^2=0.11$) (Fig. 6). This value is similar to trends observed in genome B. Additionally, as seen in this graph, the same block is observed in genome A.

For a better analysis of genome-wide LD in A, B and the entire pairs of genome loci, they were grouped based on inter-marker chromosomal intervals. In this case, the locus pairs were classified into four groups based on their (cM) distance. In this case, the distances are as follows: 0–10 cM (firmly linked markers), 11–20 cM (relatively linked markers), 21–50 cM (markers that have joined in loosely), and 50> cM (independent markers). Significant loci ($P<0.001$) of the genome and mean r^2

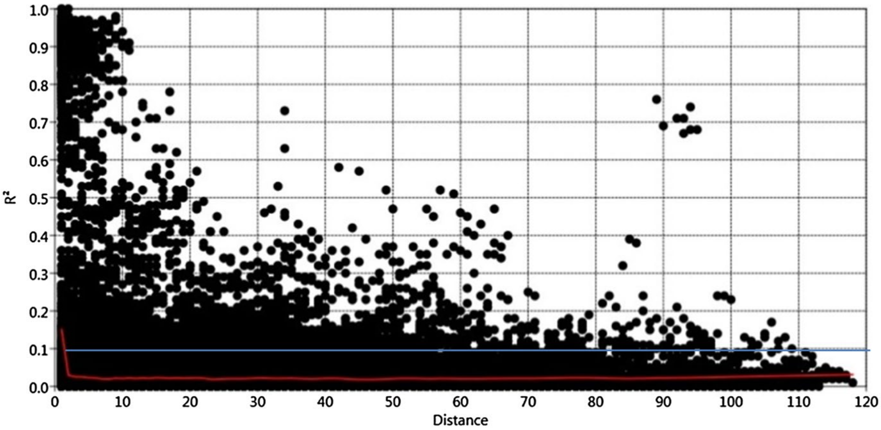

Fig. 4 Intra-chromosomal LD (r^2) decay of marker pairs overall for genome A chromosomes as a function of genetic distance (cM). The horizontal line (blue line) indicates the 95th percentile distribution of unlinked r^2 ($r^2=0.11$). The LOESS fitting curve (red line) illustrates the LD decay (Color figure online)

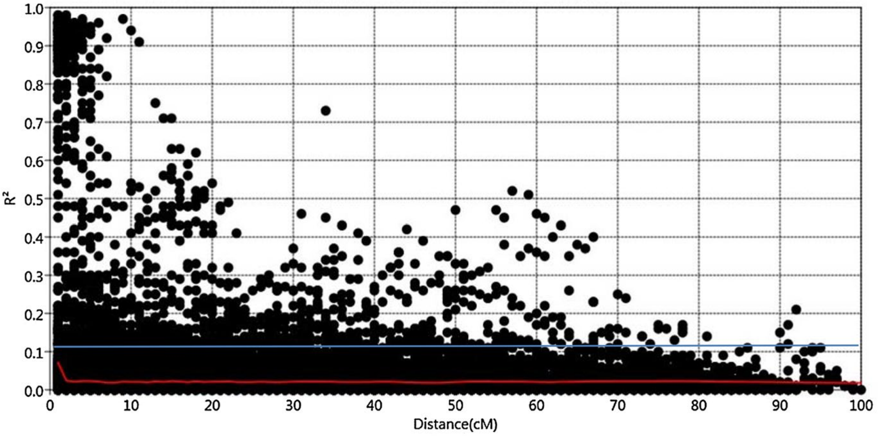

Fig. 5 Intra-chromosomal LD (r^2) decay of marker pairs overall for genome B chromosomes as a function of genetic distance (cM). The horizontal line (blue line) indicates the 95th percentile distribution of unlinked r^2 ($r^2=0.12$). The LOESS fitting curve (red line) illustrates the LD decay (Color figure online)

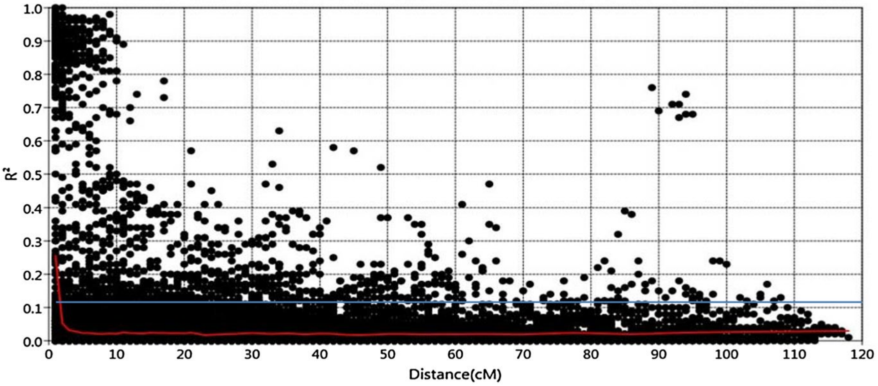

Fig. 6 Intra-chromosomal LD (r^2) decay of marker pairs over all chromosomes, genomes A and B, as a function of genetic distance (cM). The horizontal line (blue line) indicates the 95th percentile distribution of unlinked r^2 ($r^2=0.11$). The LOESS fitting curve (red line) illustrates the LD decay (Color figure online)

values for all groups of markers for the entire genome, genome A, and B are accessible in Table 4. In the whole genome, only 3.96% of the entire pair loci contained significant LD. The distance between the loci indicates the decline in the percentage of the significant loci pairs. Additionally, 16.77% of tightly linked markers demonstrated a significant r^2. Correspondingly, 9.6% of the moderately linked markers embraced approximately 8.2% of the loosely linked markers, and 7.3% of independent markers had significant LD. The proportion of r^2 values exceeding the basal LD level was 0.11. The tightly linked markers declined to moderately linked markers, while the value declined from 15.26 to 7.96%; this value for the independent

Table 4 LD overview of the whole genome, A, and B genomes in Iranian landraces accessions

Distance A genome (cM)	Total pairs	% significant	Significant pairs*	Mean r^2	Pairs in complete LD	% of pairs in LD > 0.11	Mean of $r^2 > 0.11$
0–10	3175	19.81	629	0.123	3	17.98	0.578
11–20	2511	10.91	274	0.043	0	8.76	0.2337
21–50	4232	5.4	230	0.035	0	5.2	0.233
> 50	2424	7.7	187	0.034	0	6	0.21
Total	12,342	11.95	1467	0.059	3	10.19	0.379
Distance B genome (cM)	Total pairs	% significant	Significant pairs	Mean r^2	Pairs in complete LD	% of pairs in LD > 0.12	Mean of $r^2 > 0.12$
0–10	3625	14.12	512	0.077	0	12	0.471
11–20	2925	8.4	246	0.041	0	6.52	0.295
21–50	5028	7.69	387	0.033	0	5.65	0.211
> 50	1876	6.87	129	0.029	0	4.7	0.233
Total	13,454	9.46	1274	0.047	0	7.41	0.342
Distance whole genome (cM)	Total pairs	% significant	Significant pairs	Mean r^2	Pairs in complete LD	% of pairs in LD > 0.11	Mean of $r^2 > 0.11$
0–10	6800	16.77	1141	0.01	3	15.26	0.519
11–20	5436	9.6	520	0.042	0	7.96	0.255
21–50	9260	8.2	764	0.034	0	6.7	0.201
> 50	4300	7.3	316	0.034	0	5.8	0.212
Total	25,796	10.62	2741	0.053	3	9.2	0.351

*Significant r^2 marker pairs P value < 0.001

markers was less than 5.8%. The mean for r^2 values decreased from 0.12 to 0.05, as the former shows the mean for the closely linked marker loci and the latter mean belongs to the unlinked marker pairs. The same trend was seen in genomes A and B. The data showed a noticeable reduction in the amount and percentage of significant r^2 in the 0–10 cM interval.

Other factors studied in the LD table show a similar decreasing trend at the mentioned class distances. The average value of r^2 was higher than the established baseline. The amount of r^2 at the 0–10 cM distance was reduced by half of the original value, and then, the declining process continued, although the decreasing trend became lower. Accordingly, the amount of LD decay for the whole genome is estimated at approximately 1 cM. Thus, according to the above r^2 values, the declining trend and its pattern will be approved. The variability magnitude was calculated by means of r^2 values comparing all pairs of loci in the whole genome, genome A, and genome B and showed similar values (0.351, 0.379, and 0.342, respectively).

Discussion

Local varieties of durum wheat have been grown under rain-fed conditions in different areas of the country, especially in the northwest to the south of Iran for many years. The area under rain-fed cultivation has been estimated to be approximately 250 thousand hectares in recent years. Before 1991, irrigated durum wheat cultivation was not widespread, but a decade later (2000), after the introduction of Yavaros-79 and the introduction of three varieties of durum wheat seed including Simineh, Aria, and Karkheh for cultivation in irrigated agriculture, irrigated durum wheat cultivation exceeded 200 thousand hectares, representing an extensive compatibility with Iran's climate, especially in dry weather regions of southern Iran.

Since the Iran pasta industries need semolina flour, which is obtained from the seeds of durum wheat, it is necessary to identify the lines and compatible varieties with high performance yield among local genotypes. The exact areas in which local genotypes are cultivated in Iran are unknown, but these genotypes are obviously planted in dry areas with rain-fed cultivation conditions. Comprising such a large area of cultivated lands in Iran, these genotypes have adequate opportunities to be exposed to a variety of environmental stresses (particularly drought) and non-environmental stresses that result in natural selection of the gene pool. This situation has provided more useful genes for a higher degree of adaptability. It is possible that most of these genotypes have been lost since the time of gathering for different reasons; subsequently, collection, identification, and recovery centers seem essential. Developing breeding programs and providing the necessary capacity to create new varieties can be considered among the objectives of these centers.

The DArT molecular marker has been used in many studies for evaluation of genetic diversity. Out of more than 1500 polymorphism markers, 129 genotypes of Iranian landrace durum wheat populations were selected for the study. A total of 38% of the population showed polymorphic marker comparisons to previous studies on barley (*Hordeum vulgare* L.) (10.4%; (Ruiz et al. 2012)) and wheat varieties ((Akbari et al. 2006); 14.6%) and is considered a substantial number. These values show that Iranian landrace genotypes have considerable diversity and that molecular markers can successfully detect this diversity. However, it should be noted that molecular diversity can be seen more among the populations rather than within the studied populations (68 vs. 32%) (Table 3).

The structure and R software show 7 subpopulations, and this value is lower in comparison with those found in studies conducted on local Afghan populations (15 subpopulations) and other studies (9 subpopulations) (Ruiz et al. 2012). However, this result is consistent with those found in studies conducted by Novoselović et al. (2016), Maccaferri et al. (2005) (5 and 6 populations) and Sohail et al. (2012) for *A. taushii* (3 populations). However, the number of subpopulations obviously depends on the composition of the panel. The disruption of Iranian durum wheat population's structure because of seed exchange between local farmers or assaults and chaos due to wars in the ancient history of Iran are the chief causes of disturbances of populations.

A comparison of both methods of structure analysis and discriminant analysis of principal components (DAPC) with morphological clustering and yield-related traits in a previous study (Fayaz et al. 2013) also showed that clusters 6 and 7 in DAPC and structure analysis, respectively, were associated with Group 1, derived from morphological analysis (advanced cultivars). As mentioned in the previous study, this category includes advanced cultivars with characteristics such as high beta-carotene, high number of seeds in spike, high harvest index, and low plant height. Also, clusters 2 and 6 in structural decomposition analysis and Group 4 in DAPC can be attributed to the morphological cluster 2, characterized by high plant height and number of spikes per square meter. Interestingly, the morphological cluster 2 was mainly derived from the geographical area of Lorestan and Kermanshah. The third cluster in structure analysis and DAPC, which consists mainly of Khorasan landrace genotypes, was grouped as 4 and 5 in morphological cluster analysis, with morphological characteristics such as low beta-carotene seed content, high thousand seed weight, high number of tillers, and low function. Simply put, comparisons of genetic groupings (structure, DAPC, and NJ) and morphological clustering showed that landrace genotypes with geographical origin of Khorasan were almost present in all categories, as well as genetic, and morphological groups.

A comparison of landrace genotypes of the third group (local genotypes of Khorasan Fst = 0.94) and the sixth group (Lorestan and Kermanshah groups Fst = 0.82) indicated the diversity of this group, genotypes in eastern Iran were compared with those in western Iran. This could be due to the fact that most of the genotypes in Khorasan were related to allelic variance in Khorasan, compared to Lorestan and Kermanshah (Holsinger and Weir 2009). The landrace populations of East Azarbaijan also had high Fst = 0.86, but were not considered in comparison, due to the limited number of genotypes (Table 2).

Existence of LD decay values that are approximately 1–2 cM ($r^2 = 0.11$) in Iranian traditional genotypes is another distinctive feature of the studied population. At a distance of 3–2 cM r^2, the value is reduced by one-half or one-third of the original values, which can be ranked as a sharp decline. The values calculated in this study are consistent with scientific findings for local varieties by (Hao et al. 2011) (LD < 5 cM) and (Chen et al. 2012) with $r^2 > 0.1$ (2.2–17.4 cM).

One of the reasons for the small amount of LD decay in the studied population ($r^2 = 0.11$, 2–3 cM) is probably the variation between populations in comparison with the variation within populations. Since the amount of cross pollination in wheat is less than 5%, it is expected that this leads to increases in the homozygous value and genetic purity, which in turn upsurges the amount of LD decay, as seen for other self-pollinated plants such as Arabidopsis (Flint-Garcia et al. 2003). Comparing this population with Asian and European wheat populations studied by (Hao et al. 2011), the results are largely consistent with the level of r^2 values ($r^2 = 0.11$–0.12). Compared with the European populations and, in particular, improved varieties for LD decay, in this study, $r^2 = 0.18$ for the Asian population shows a high value, which is approximately 17.4 cM or more. This clearly indicates the difference between breeding programs for the selection of improved and local varieties. Local landrace genetic variations are subject to different factors in their evolutionary history. In self-pollinated plants, cross pollination and mutations (that enhance the fitness rate) increase the diversity of the population, while natural selection or selection by humans and genetic drift can contribute to variation within the populations (Ennos 1994). Existence diversity within populations is due to mutuality with the self-pollinated plants; however, the studies conducted by using protein markers in Pakistanian local genotypes (Tahir et al. 1996) and the Italian Emmer wheat (subsp. *dicoccum* [Schrank ex Schuu bl *Triticum turgidum* L]) RAPD molecular markers show a high level of consistency (Barcaccia et al. 2002). Hence, increasing the genetic diversity within the population can be associated with the very long evolutionary history of plants in Iran (Dreisigacker et al. 2005). On the other hand, local Iranian genotypes may have been moved from one geographical area to other locations, even in small amounts. Thus, they may have been subjected to a founder effect.

In the present study, 129 genotypes were identified as 7 subpopulations, and the group diversity showed advantages to the within population diversity (although the within population variation was also significant). This study showed a high diversity of Iranian durum wheat as compared to a previous study that was accomplished on Iranian landrace common wheat genotypes (Alipour et al. 2017). Therefore, the present population can be a significant source of breeding programs and adaptation to climate conditions in future. Such information may be useful for collecting and preserving plant material in the future, and it could play a major role in higher allelic diversity and in the protection of rare alleles in landrace Iran's populations. Increasing the number of local genotypes obtained in Iran and comparing them with sources in neighboring countries and the world's resources could considerably contribute to breeding programs and the selection of plant material in the restore program. A closer examination of plant material to take into account molecular markers based on next-generation sequencing can provide more accurate information and reveal further insight into the internal dynamics of population genetics and its ecotypes. Of course, further studies can help to develop the prospects of such information.

Conclusion

In this study, Iran's local durum wheat varieties were considered and investigated. Approximately 129 genotypes with a total of 1500 polymorphic genetic DArT markers were analyzed. High genetic diversity among and within subpopulations was identified (about 7 subpopulations). The identified subpopulations were affected to some extent by the geographical conditions of the area where they were collected; however, this situation is not constant but is relatively associated with exchange of seeds between geographical communities and processes that evolved over time. Low levels of LD decay in native genotypes may represent the mixture of each subpopulation. Thus, investigating within populations can result in identifying very rich allele genotypes, especially exclusive allele resistance to a variety of environmental and non-environmental stresses (particularly drought). Better evaluation of these populations could be the basis for modern breeding programs to identify new genes. For example, the local and native Kurdistan genotype named "Parisht" (with low yield but strong resistance to unfavorable environmental conditions) could be useful. Therefore, this type of germplasm resource can serve as an excellent basis to promote new breeding programs and to produce new varieties with a promising future.

Acknowledgements This research project was supported by grants from Islamic Azad University Sanandaj branch (Project Number 11-12-1-37345). We gratefully acknowledge the Iran National Gene Bank and Dr. Mostafa Aghaee for providing the germplasm for this study.

References

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113(8):1409–1420

Alipour H, Bihamta MR, Mohammadi Peyghambari VSA, Bai G, Zhang G (2017) Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. Front Plant Sci 8(2017):1293. https://doi.org/10.3389/fpls.2017.01293

Badr A, Sch R, El Rabey H, Effgen S, Ibrahim H, Pozzi C, Rohde W, Salamini F (2000) On the origin and domestication history of barley (*Hordeum vulgare*). Mol Biol Evol 17(4):499–510

Barcaccia G, Molinari L, Porfiri O, Veronesi F (2002) Molecular characterization of emmer (*Triticum dicoccon* Schrank) Italian landraces. Genet Resour Crop Evol 49(4):417–428

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Chen X, Min D, Yasir TA, Hu YG (2012) Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. PloS ONE 7(9):e44510

Comadran J, Thomas W, Van Eeuwijk F, Ceccarelli S, Grando S, Stanca A, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A (2009) Patterns of genetic diversity and linkage disequilibrium in a highly structured *Hordeum vulgare* association-mapping population for the Mediterranean basin. Theor Appl Genet 119(1):175–187

de Carvalho MAP, Bebeli PJ, Bettencourt E, Costa G, Dias S, Dos Santos TM, Slaski JJ (2013) Cereal landraces genetic resources in worldwide GeneBanks: a review. Agron Sustain Dev 33(1):177–203

Dreisigacker S, Zhang P, Warburton M, Skovmand B, Hoisington D, Melchinger A (2005) Genetic diversity among and within CIMMYT wheat landrace accessions investigated with SSRs and implications for plant genetic resources management. Crop Sci 45(2):653–661

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361. https://doi.org/10.1007/s12686-011-9548-7

El-Esawi MA, Witczak J, Abomohra AE-F, Ali HM, Elshikh MS, Ahmad M (2018) Analysis of the genetic diversity and population structure of Austrian and Belgian wheat germplasm within a regional context based on DArT Markers. Genes 9(1):47

Ennos R (1994) Estimating the relative rates of pollen and seed migration among plant populations. Heredity 72(3):250–259

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164(4):1567–1587

Fayaz F, Mardi M, Aghaee M, Darvish F, Talebi R (2013) Phenotypic diversity analysis of grain yield and yellow pigment content in germplasm collected from Iranian durum wheat (*Triticum turgidum* L.) landraces. Arch Agron Soil Sci 59:1339–1357

Flint-Garcia SA, Thornsberry JM, Buckler IV ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54(1):357–374

Garavito A, Montagnon C, Guyot R, Bertrand B (2016) Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. BMC Plant Biol 16:242. https://doi.org/10.1186/s12870-016-0933-y

Hao C, Perretant MR, Choulet F, Wang L, Paux E, Sourdille P, Zhang X, Feuillet C, Balfourier F (2010) Genetic diversity and linkage disequilibrium studies on a 3.1-Mb genomic region of chromosome 3B in European and Asian bread wheat (*Triticum aestivum* L.) populations. Theor Appl Genet 121(7):1209–1225

Hao C, Wang L, Ge H, Dong Y, Zhang X (2011) Genetic diversity and linkage disequilibrium in Chinese bread wheat (*Triticum aestivum* L.) revealed by SSR markers. PLoS ONE 6(2):e17279

Harlan JR (1986) Plant domestication: diffuse origins and diffusions. The origin and domestication of cultivated plants. Elsevier, Amsterdam, pp 21–34

Holsinger K, Weir B (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet 10:639–650

Horvath A, Didier A, Koenig J, Exbrayat F, Charmet G, Balfourier F (2009) Analysis of diversity and linkage disequilibrium along chromosome 3B of bread wheat (*Triticum aestivum* L.). Theor Appl Genet 119(8):1523–1537

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29(4):e25–e25

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11(1):94

Kovi MR, Fjellheim S, SandveS R, Larsen A, Rudi H, Asp T, Kent MP, Rognli OA (2015) Population structure, genetic variation, and linkage disequilibrium in perennial ryegrass populations divergently selected for freezing tolerance. Front Plant Sci 6:929

Laidò G, Marone D, Russo MA, Colecchia SA, Mastrangelo AM, De Vita P et al (2014) Linkage disequilibrium and genome-wide association mapping in tetraploid wheat (*Triticum turgidum* L.). PLoS ONE 9(4):e95211. https://doi.org/10.1371/journal.pone.0095211

Lopes MS, El-Basyoni I, Baenziger PS, Singh S, Royo C, Ozbek K, Aktas H, Ozer E, Ozdemir F, Manickavelu A, Ban T, Vikram P (2015) Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. J Exp Bot 66(12):3477–3486. https://doi.org/10.1093/jxb/erv122

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15(3):271–290

Marone D, Panio G, Ficco DB, Russo MA, De Vita P, Papa R, Rubiales D, Cattivelli L, Mastrangelo AM (2012) Characterization of wheat DArT markers: genetic and functional features. Mol Genet Genom 287(9):741–753

Moragues M, del Moral LFG, Moralejo M, Royo C (2006) Yield formation strategies of durum wheat landraces with distinct pattern of dispersal within the Mediterranean basin I: yield components. Field Crops Res 95(2):194–205

Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, Hatipoğlu R, Ahmad F, Alsaleh A, Labhane N (2017) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol Biotechnol Equip 32:261–285

Novoselović D, Bentley AR, Šimek R, Dvojković K, Sorrells ME, Gosman N, Horsnell R, Drezner G, Šatović Z (2016) Characterizing Croatian wheat germplasm diversity and structure in a European context by DArT markers. Front Plant Sci 7:184. https://doi.org/10.3389/fpls.2016.00184

Peleg Z, Saranga Y, Suprunova T et al (2008) Genetic diversity and population structure of tetraploid wheats (*Triticum turgidum* L.) estimated by SSR, DArT and pedigree data. Theor Appl Genet 117:103. https://doi.org/10.1007/s00122-008-0756-9

Perrier X, Jacquemoud-Collet J (2006) DARwin: dissimilarity analysis and representation for Windows 5.0.158. CIRAD, Montpellier

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65(1):220–228

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959

R Core Team (2014) R: a language and environment for statistical computing. R Core Team, Vienna

Ruiz M, Giraldo P, Royo C, Villegas D, Aranzana MJ, Carrillo JM (2012) Diversity and genetic structure of a collection of Spanish durum wheat landraces. Crop Sci 52(5):2262–2275

Sohail QTS, Kilian A, Eltayeb AE, Tanaka H, Tsujimoto H (2012) Development of diversity array technology (DArT) markers for assessment of population structure and diversity in *Aegilops tauschii*. Breed Sci 62(1):38–45

Somers DJ, Banks T, DePauw R, Fox S, Clarke J, Pozniak C, McCartney C (2007) Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome 50(6):557–567

Stodart B, Mackay M, Raman H (2008) Assessment of molecular diversity in landraces of bread wheat (*Triticum aestivum* L.) held in an ex situ collection with diversity arrays technology (DArT™). Crop Pasture Sci 58(12):1174–1182

Tahir M, Turchetta T, Anwar R, Lafiandra D (1996) Assessment of genetic variability in hexaploid wheat landraces of Pakistan based on polymorphism for HMW glutenin subunits. Genet Resour Crop Evol 43(3):211–220

Weir B (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer Assoc. Inc., Sunderland, MA

Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD et al (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS ONE 4:e8451

Zhang L, Liu D, Guo X, Yang W, Sun J, Wang D, Sourdille P, Zhang A (2011) Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers. BMC Genet 12(1):42

Affiliations

Farzad Fayaz[1] · Mostafa Aghaee Sarbarzeh[2] · Reza Talebi[1] · Amin Azadi[3]

[1] Department of Agronomy & Plant Breeding, College of Agriculture, Islamic Azad University, Sanandaj Branch, Sanandaj, Iran

[2] Seed and Plant Improvement Institute, Agriculture Research Education and Extension Organization (AREEO), Karaj, Iran

[3] Department of Plant Breeding, Yadegar-e-Imam Khomeini (RAH), Islamic Azad University, Shahre Rey Branch, Tehran, Iran