

# Analysis of genetic diversity, population structure and linkage disequilibrium in Iranian wheat landraces using SSR markers

Elham Zarei Abbasabad<sup>1</sup>, Seyed Abolghasem Mohammadi<sup>1,2\*</sup>,  
Mohammad Moghaddam<sup>1,2</sup> and Mohammad Reza Jalal Kamali<sup>3</sup>

<sup>1</sup>Department of Plant Breeding and Biotechnology, Faculty of Agriculture, University of Tabriz, Tabriz, Iran  
<sup>2</sup>Center of Excellence in Cereal Molecular Breeding, University of Tabriz, Tabriz, Iran and <sup>3</sup>CIMMYT (International Maize and Wheat Improvement Center), Karaj, Iran

Received 19 June 2015; Accepted 14 December 2015 – First published online 7 March 2016

## Abstract

Population structure and linkage disequilibrium (LD) were investigated in a set of 395 bread wheat landraces including 154 spring, 193 winter, two facultative wheat and 46 unknown growth type collected from different geographical regions of Iran. A total of 53 microsatellite markers distributed in three genomes of wheat were assayed for polymorphism. The 312 polymorphic alleles were served to estimate population structure and analyse the genome-wide LD. The number of alleles ranged from 2 to 18 with an average of 5.89 alleles/locus. Mean of polymorphic information content was 0.6 with a range of 0.15–0.86 and gene diversity varied from 0.16 to 0.88 with an average of 0.64. The population of landraces was highly structured and based on distance- and model-based cluster analyses the 395 landraces were assigned into eight subpopulations: SG1–SG8. Population structure estimates based on simple sequence repeat (SSR) marker data were quantified in a Q matrix and used in calculation of LD between pair of markers. A low overall LD level found in 12–13% (166) of the SSR marker pairs showed significant pairwise LD in  $r^2 \geq 0.01$  and  $P \leq 0.001$  and six pair showed  $r^2 \geq 0.05$  with  $P \leq 0.001$ . LD clearly decays within the genetic distance of 40–60 cM with  $r^2 \sim 0.05$ . The results of this study should provide valuable information for future association mapping using this wheat collection.

**Keywords:** gene diversity, linkage disequilibrium, microsatellite markers, population structure

## Introduction

Landraces are valuable sources to broaden the genetic base of crop plants. The development of new varieties from landrace populations is a viable strategy to improve yield and yield stability as well as resistance to biotic and abiotic stresses. Landraces harbour genes and as well gene complexes for quality traits and adapt to a wide range of low-input and organic farming systems (Jaradat, 1991;

Strelchenko *et al.*, 2008). Various types of data have been used for the analysis of genetic diversity and population structure in wheat germplasm. Among these, DNA markers especially simple sequence repeats (SSRs) have been extensively exploited for the analysis of genetic diversity and population structure, gene mapping, linkage disequilibrium (LD) analysis and marker assisted selection (Song *et al.*, 2005; Wang *et al.*, 2009; Hao *et al.*, 2011).

LD is a non-random association of alleles at two (more) loci, describing the condition with non-equal frequency of the haplotypes in a population at random combination of alleles at different loci. The extent of LD in natural and domesticated populations is mainly related to effective

\*Corresponding author. E-mail: mohammadi@tabrizu.ac.ir

recombination rate, mating system and population size (Flint-Garcia *et al.*, 2003; Gupta *et al.*, 2005).

In crop plants, the potential of exploiting LD to detect marker-trait associations was investigated for maize (Yu and Buckler, 2006; Belo *et al.*, 2008), wheat (Ravel *et al.*, 2006; Rhone *et al.*, 2007; Tommasini *et al.*, 2007), barley (Kraakman *et al.*, 2004, 2006; Rostoks *et al.*, 2006), sorghum (Hamblin *et al.*, 2004), ryegrass (Skot *et al.*, 2005; Xing *et al.*, 2007), soybean (Hyten *et al.*, 2009) and rice (Garris *et al.*, 2003). The published results suggest that association mapping is a valuable additional tool in the search for the detection of novel genes or quantitative trait locus for important agronomic characteristics.

Multiallelic markers such as SSRs are also often used for association studies. These SSR markers have already been used for a study of population structure and LD-based association studies in wheat (Kruger *et al.*, 2004; Chen *et al.*, 2012). Estimates of LD decay in wheat at the whole genome level and with large genetic representations of wheat genotypes are of great value. Chao *et al.* (2007) used 242 genomic SSRs to estimate LD in a collection of 43 elite US wheat cultivars and reported that genome-wide LD estimates were generally less than 1 centiMorgans (cM) for genetically linked locus pairs, and that most of the LD was between loci less than 10 cM apart. A collection of 189 bread wheat landraces genotyped at 370 loci were used to examine LD across the genome, and it was found that LD mapping of wheat could be performed with SSRs to a resolution of less than 5 cM (Somers *et al.*, 2007).

The objectives of the present study were to estimate: (a) the levels and the structure of the SSR genetic diversity present in a collection of 395 Iranian wheat landraces; (b) the levels and patterns of LD between pairs of SSR loci; (c) the population structure of multilocus LD in the populations.

## Materials and methods

### Plant materials

A total of 395 wheat landraces from different geographical regions of Iran including 154 spring, 193 winter, two facultative wheat and 46 unknown growth type were kindly provided by International Center for the Improvement of Maize and Wheat (CIMMYT).

### DNA preparation and SSR analysis

DNA from fresh leaves of 10 plants of each genotype was extracted using cetyl trimethyl ammonium bromide method (Saghai-Marouf *et al.*, 1984). Polymerase chain reactions (PCR) for SSR analyses were performed in a total volume of 10  $\mu$ l containing 2  $\mu$ l of template DNA, 0.1  $\mu$ l dNTP (10 mM), 0.5  $\mu$ l of each primer, 0.3  $\mu$ l MgCl<sub>2</sub> (50 mM),

5.2  $\mu$ l 1 $\times$ PCRBuffer and 0.11  $\mu$ l *Taq* DNA polymerase. Thermocycling consisted of an initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 45 s, 48–63°C annealing temperature (depending upon the primer sets) for 45 s and 72°C extension for 1 min, with a final extension of 7 min at 72°C. PCR products were analysed by 4% polyacrylamide gel electrophoresis and visualized by Ethidium bromide staining. The chromosome locations of SSR primers were mainly taken from the National Bio Resource Project, Japan (<http://www.shigen.nig.ac.jp/wheat/komugi/maps/marker/map.jsp>), and the wheat genome map produced by Somers *et al.* (2004).

### Allele diversity and population structure

Gene diversity and heterozygosity were calculated at 53 SSR marker loci in 395 genotypes using Power-Marker software (Liu and Muse, 2005). Gene diversity ( $H$ ) and polymorphic information content (PIC) for each locus were calculated from allele frequencies using  $PIC = 1 - \sum p_i^2 - \sum 2p_i^2 p_j^2$  and  $H = 1 - \sum p_i^2$  formulas, where  $p_i$  and  $p_j$  are the frequency of the  $i$ th and  $j$ th alleles, respectively.

### Population structure analysis

Population structure was analysed with the STRUCTURE software v.2.3.4 (Pritchard *et al.*, 2000) using a Bayesian clustering approach to identify subpopulations, each possessing a characteristic set of allele polymorphisms, based on genotyping data from 53 SSRs. Independent simulations having 100,000 Markov Chain Monte Carlo (MCMC) replications and 100,000 burn-ins were performed with the number of subpopulations ( $k$ ) ranging from 1 to 12. To examine genetic relationships, unweighted pair-group method with arithmetic average (UPGMA) cluster based on Maximum Composite Likelihood distance was performed with MEGA software. In order to assess the further existence of a genetic structure within the identified clusters, estimation of posterior distribution of pairwise Wright's  $F_{ST}$  (Wright, 1951), a standardized measure of the genetic variance among populations, among 100,000 iterations were also plotted. Pairwise  $F_{ST}$  was calculated as  $F_{ST} = \text{var}(p_k) / \bar{p}(1 - \bar{p})$ , where  $p_k$  is the frequency of an allele in population  $k$ , and  $\bar{p}$  is the overall frequency of that allele across all subpopulations (Excoffier, 2001). The net nucleotide distance was calculated between all pairs of subpopulations to determine relationships between populations. This distance between populations A and B,  $D_{AB}$ , is calculated as  $D_{AB} = 1/L \sum_{l=1}^L \{1 - \sum_{j=1}^J \hat{p}_{A,j}^{(l)} \hat{p}_{B,j}^{(l)}\} - (H_A + H_B/2)$  is the (posterior mean) estimated allele frequency of allele  $j$  at locus  $l$  in population  $x$ ,  $L$  is the number of loci,  $J_l$  the number of alleles at locus  $l$  and where

$H_X = 1/L \sum_{i=1}^L \{1 - \sum_{j=1}^J \hat{p}_{X,j}^{(i)2}\}$ . In words, the net nucleotide distance is the average probability that a pair of alleles, one each from populations A and B are different, less the average within-population heterozygosities. The distance has the appropriate property that similar populations have distances near 0, and in particular,  $D_{AA} = 0$ .

## LD

The genome-wide LD between all pairs of SSR alleles based on squared allele-frequency correlations ( $r^2$ ) was analysed based with 1000 permutations using software package TASSEL ver. 1.9.6 (Bradbury *et al.*, 2007). Loci were considered to be in significant LD if  $P \leq 0.001$ . LD decay scatter plots of syntenic  $r^2$  versus genetic distance (cM) between markers were generated. LD decay was calculated according to the method described in Brescghello and Sorrells (2006).

## Results

### SSR polymorphism

A total of 53 SSR loci were detected by 52 markers, including 4 Xwmc, 11 Xbarc and 37 Xgwm, which were located on 17 wheat chromosomes (Table S1). In total, 312 alleles were amplified at the 53 polymorphic SSR loci. Among the analysed SSR primer pairs, only primer Xbarc70 amplified two loci at the studied genotypes. The number of allelic variants detected per locus ranged from 2 (Xbarc186) up to 18 (Xwmc175), with a mean of 5.89. Major allele frequency ranged from 0.17 to 0.91, with an average of 0.48.

Polymorphic information content values ranged from 0.15 (Xgwm319) to 0.86 (Xgwm413) with an average 0.6. Gene diversity values of the 53 SSR loci ranged from 0.16 to 0.88 for Xgwm319 and Xgwm413 loci, respectively, with an average of 0.64. Considering the high degree of polymorphism, a set of SSR markers consisting of nine loci with gene diversity values of 0.80 or higher were selected. These loci were Xwmc327, Xgwm6, Xgwm35, Xgwm135, Xgwm400, Xgwm413, Xgwm480, Xgwm544 and Xgwm570, whose combinations provided distinct genotypes for all the studied landraces.

### Population structure

Population structure of the 395 landraces was estimated using STRUCTURE software based on 53 SSR markers. The number of subpopulations ( $K$ ) was identified based on  $\Delta K$  values. The highest delta  $K$  value is to be taken as optimum  $K$  (Figure S1). The number of clusters,  $K$ , was inferred to be eight (SG1 until SG2). The groups included 7.4, 12.7, 9.5, 16.8, 11.4, 14.6, 13 and 14.5% of the landraces,

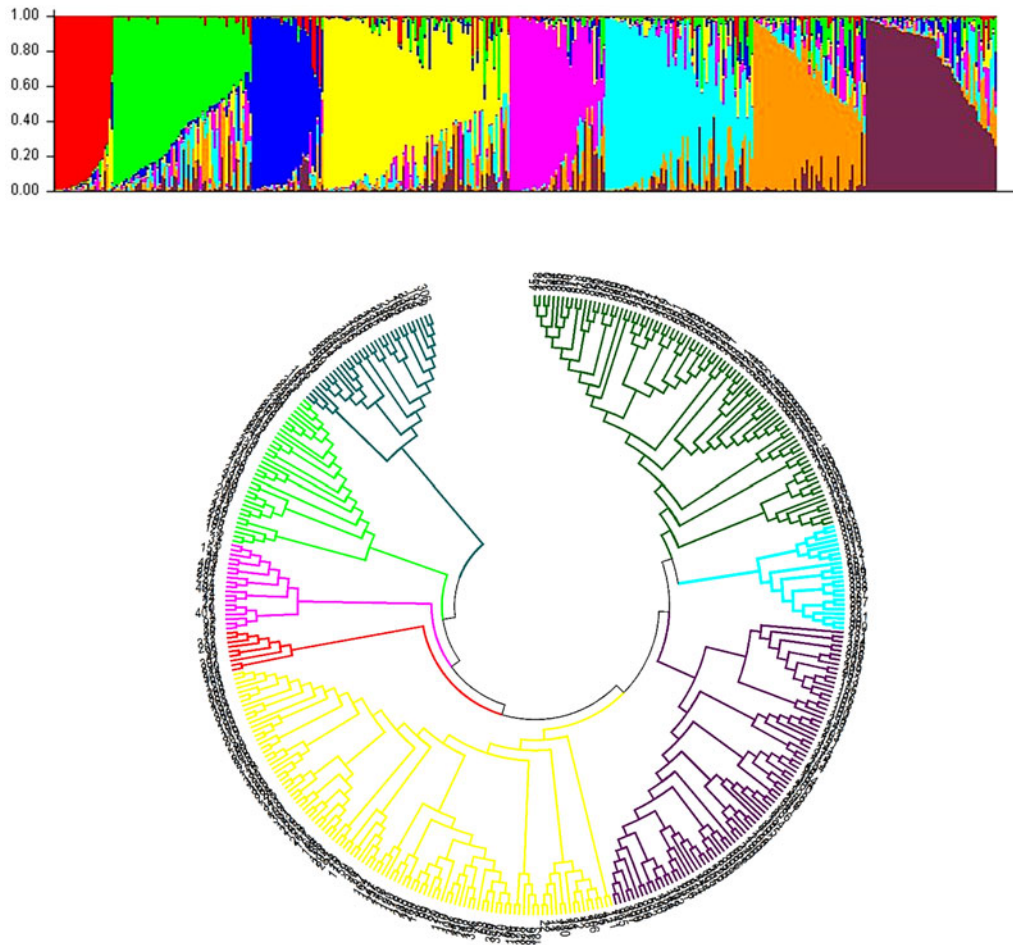
respectively. Landraces were spring, winter and facultative growth type and all the subgroups included spring and winter types. Two facultative landraces were in SG4 and SG5. SG4 was the largest group consisting of 41 winter and 30 spring landraces and SG1 with three winter, 18 spring and five unknown growth type landraces was the smallest group. SG1 and SG8 had the highest number of spring landraces. The population structure ascertained by model-based cluster analysis was in congruent with geographic ecotypes of the landraces. SG1 consisted of landraces of northern latitude 3616 and 3613 from Qazvin, Mashhad and Sabzevar. The landraces in second group were originated from Shiraz, Kerman, Neishabour, Mashhad and Urmia. The origin of most landraces in SG3 was northeastern Iranian cities including Birjand, Torbatejam, Damghan and Mashhad. In SG4, landraces from Qazvin and Sanandaj and in SG5 and SG7, landraces from Shahreza and Esfahan with 3201N and 3733N latitude were grouped. The most of the landraces in SG6 were originated from hot area of Iran including Tabas and Zahedan and SG8 consisted of more landraces from cold regions such as Hamedan and Zanjan. UPGMA cluster analysis also clearly divided the 395 landraces into eight groups. The dendrogram and bar plot were shown in Fig. 1. To quantify distances between the subpopulations and to further evaluate population structure, net nucleotide distances between pairs of sub-populations were computed (Table 1). The distances between pairs of sub-populations ranged from 0.0591 (between SG2 and SG6) to 0.3130 (between SG1 and SG5).

To assess further levels of genetic structure within the eight identified clusters, the estimate of posterior distribution of pairwise Wright's  $F_{ST}$  (Wright, 1951), a measure of the genetic variance among populations was also calculated using 100,000 permutations (Fig. 2).  $F_{ST}$  values between all groups were significant ( $P \leq 0.001$ ) and ranged from 0.14 to 0.28, supporting the existence of genetic structure.

### Pairwise LD and LD decay in Iranian wheat landraces

Across all 53 loci, 1378 locus pairs were detected in the Iranian wheat landraces collection. LD plot is shown in Figure S2. At significant threshold values ( $r^2 \geq 0.01$  and  $P \leq 0.001$ ), 12–13% of the SSR marker pairs showed significant pairwise LD in a total 166 pairwise and six pairs were detected with  $r^2 \geq 0.05$  and  $P \leq 0.001$ . At the highly significant threshold of  $r^2 \geq 0.1$ , only one SSR linked marker pair (Xbarc186 and Xbarc117) with  $r^2 = 0.226$  remained in LD (Table 2).

To visualize or depict the extent of LD, a plot of LD decay, which shows how LD declines with genetic



**Fig. 1.** Bar plot of the genetic composition of individuals based on 53 SSR markers generated by STRUCTURE 2.3.4 using the admixture model. Groups for each panel are represented by colours as indicated at the bottom (A). The dendrogram was resulted from genetic distance in MEGA software (B).

(centiMorgans, cM) was derived. Genome wide LD at  $r^2 = 0.2$  reduced to  $\sim 1\text{--}2$  cM and we observed a significant ( $r^2 = 0.2$ ) LD between one pair of SSR loci within 1.93 cM distance. However, LD clearly decays within the genetic distance of 40–60 cM with  $r^2 \sim 0.05$ . Significant LD between unlinked markers suggests the existence of LD generating factors other than linkage in the landraces (Fig. 3).

## Discussion

### *Genetic diversity and population structure in Iranian wheat landraces*

The level of polymorphism detected was relatively high, with an average of 5.88 alleles per locus and many alleles for some loci. A similar number of alleles was found for 60 hexaploid wheat cultivars from Eastern Europe, genotyped at 42 SSR loci (Stachel *et al.*, 2000). Analysis of genetic

diversity in 250 Chinese bread wheat accessions, including 93 modern varieties and 157 landraces revealed amplification of an average 13.1 alleles per locus with a mean genetic diversity of 0.65 and higher genetic diversity (0.64) in landraces compared with modern varieties (0.628) (Hao *et al.*, 2011). Other estimates of mean allele number per locus from previous studies using SSR markers on more diverse germplasm were: 18.1 alleles at 26 loci on 998 accessions of hexaploid wheat from Gatersleben germplasm collection (Huang *et al.*, 2002); 5.6 alleles at 70 loci on 58 durum cultivars of diverse geographical origins, including old cultivars (Maccaferri *et al.*, 2003) and 5.05 alleles at 269 loci on 90 Chinese winter wheat (Chen *et al.*, 2012). Our estimates of SSR allele diversity in wheat were comparable with estimates reported for other grasses, when the sampling breadth is considered. Differences in results reported by various studies could be because of differences in number of genotypes studied, their genetic background and number of markers used and techniques applied to



**Table 1.** Allele-frequencies divergence among populations (net nucleotide distance), computed using point estimates on structure

Sub-populations	SG1	SG2	SG3	SG4	SG5	SG6	SG7	SG8
SG1	–	0.2622	0.3008	0.2107	0.3130	0.2416	0.2409	0.2767
SG2		–	0.1031	0.0999	0.1292	0.0591	0.0679	0.0827
SG3			–	0.1489	0.1617	0.1243	0.1159	0.0937
SG4				–	0.1538	0.0813	0.0947	0.1250
SG5					–	0.1428	0.1405	0.1534
SG6						–	0.0621	0.0908
SG7							–	0.0789

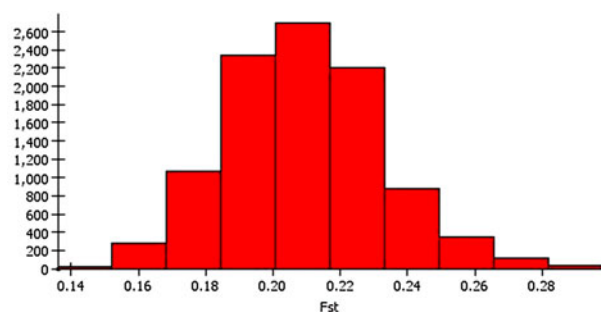
detect polymorphism. For example, high number of alleles per locus (18.1) in Huang *et al.* (2002) study could be explained by the larger number of genotypes analysed in their study. In addition, Hai *et al.* (2007) assumed that the higher genetic diversity observed in some germplasms may be because of the presence of relatively more rare alleles, resulting from less stringent selection applied in those germplasms. They suggested that breeding practice affected the total number of alleles and rare alleles.

Population structure is one of several important factors that strongly influence LD. In the study of flowering time locus in maize a suite of polymorphisms in the maize dwarf8 gene was significantly associated with variation in flowering time (Thornsberry *et al.*, 2001). The incidence of false positives created by population structure was reduced by up to 8% as a result of the Pritchard method. Using these statistical methods in an association test allowed researchers to improve their resolution from the level of a 20-cM region to that of an individual gene. The methodological advances that estimate the effects of population structure-induced linkage disequilibria should allow the use of association testing in a much wider context, enabling the use of this very powerful technique (Flint-Garcia *et al.*, 2003). Population structure analyses have indicated that wheat landraces can be categorized by geographical origin divided to eight subpopulations. In total, 395

landraces were assigned by STRUCTURE to SG1 until SG8 and this result was similar to UPGMA result. Chen *et al.* (2012) using 269 SSR markers in 90 elite Chinese winter wheat reported that 90 Chinese wheat accessions could be divided into three subgroups based on STRUCTURE, UPGMA cluster and principal coordinate analyses. They also noted that the population structure derived from STRUCTURE clustering was positively correlated to some extent with geographic eco-type.

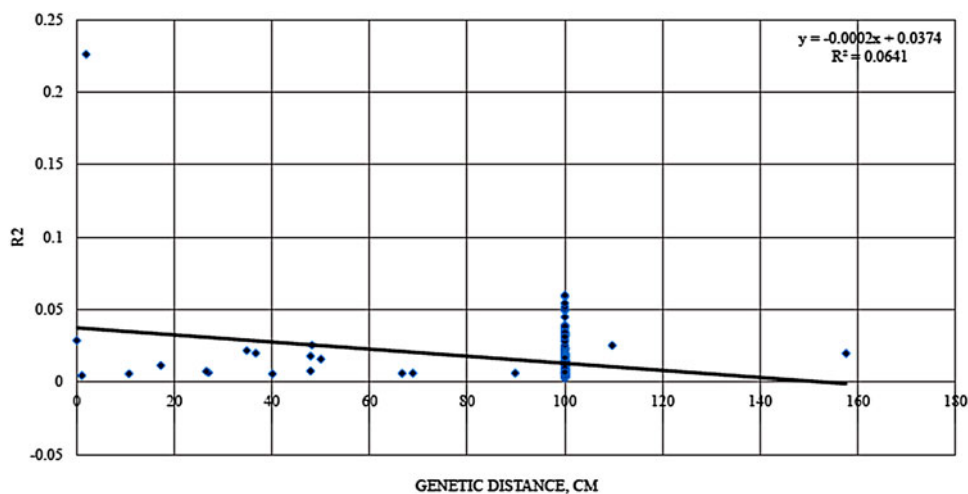
## LD

The pattern of LD decay determines the marker density required for and the level of resolution that may be obtained in an association study (Flint-Garcia *et al.*, 2003). A higher level of LD might be expected in hexaploid wheat compared with maize and sorghum because of the rapid rate of inbreeding in wheat with a high degree of self-pollination (Flint-Garcia *et al.*, 2003). Numerous studies in wheat have reported LD decay distances. Mean genome-wide LD decay was estimated to be 10 cM ( $r^2 \geq 0.1$ ) in 205 US elite wheat breeding lines (Zhang *et al.*, 2010). Hao *et al.* (2011) reported a wider average LD decay in Chinese modern varieties compared with landraces across the whole genome for locus pairs with  $r^2 \geq 0.05$  ( $P \leq 0.001$ ). In their study, mean LD decay distance for 157 landraces at the whole genome level was 5 cM compared with 5–10 cM for the 93 modern varieties investigated. For 189 bread wheat and 93 durum wheat accessions, the LD between adjacent locus pairs ( $r^2 \geq 0.2$ ) extended  $\sim 2$ –3 cM on average, and it was suggested that LD mapping of wheat could be performed with SSRs to a resolution of 5 cM (Somers *et al.*, 2007). From an analysis of 242 genomic SSRs among 43 elite US wheat cultivars, it was reported that genome-wide LD estimates generally were less than 1 cM for genetically linked locus pairs with  $r^2 \geq 0.2$  ( $P \leq 0.01$ ), and that most LD was less than 10 cM apart between loci (Chao *et al.*, 2007). Chen *et al.* (2012) in 90 elite Chinese winter wheat reported 17.4 cM ( $r^2 \geq 0.2$ )

**Fig. 2.** The distribution of the pairwise  $F_{ST}$  values between eight subpopulations over 100,000 iterations.

**Table 2.** The pairwise genome-wide linkage disequilibrium between pairs of simple sequence repeat markers

	Number of pairwise comparisons	$P \leq 0.01$	$P \leq 0.001$	$r^2 \geq 0.01$	$r^2 \geq 0.05$	$r^2 \geq 0.1$
All markers	1378	407	244	166	6	1
Linked markers	75	27	20	12	1	1
Unlinked markers	1303	380	224	154	5	0

**Fig. 3.** Scatter plot of significant  $r^2$  values and genetic distance (cM) ( $P \leq 0.001$ ) of locus pairs on whole genomes in landraces. The logarithmic trend line represents the average LD decay.

as the maximum LD decay distance and whole genome LD decay distance was  $\sim 2.2$  cM ( $r^2 \geq 0.2$ ,  $P \geq 0.001$ ). Nielsen *et al.* (2014) using DArT technology in 94 European hexaploid bread wheat reported mean  $r^2$  value of 0.080 for all marker-pairs and 0.271 for significant marker pairs and LD decay for the total population was 23 cM. These differences are due both to variations in material type and quantity as well as differences in  $r^2$  values (i.e.  $r^2 \geq 0.05$ , 0.1, or 0.2) used for the estimations. Our results showed genome wide LD at  $r^2 = 0.2$  reduced to  $\sim 1$ –2 cM and we observed a significant ( $r^2 = 0.2$ ) LD between one pair of SSR loci within 1.93 cM distance. However, LD decays within the genetic distance of 40–60 cM with  $r^2 \sim 0.05$  (Fig. 3).

In summary, this study illustrates that genetic diversity in Iranian wheat landraces, was mediocre and lower than that of many other wheat germplasm collections. However we can use this collection to identify resistance and other important genes source. Using landraces from other countries and wild genotypes will be useful to increase genetic diversity. The large number of SSR markers can improve results suggesting that, use large number of markers in future studies. The population structure revealed using STRUCTURE analysis of SSR markers distributed throughout the wheat genome was more accurate than using geographic ecotype information and was available for investigating

population genetic structure. LD analysis revealed that there are significant associations between unlinked and some of linked markers and LD decay distance is different from other wheat collections. For obtaining extra information, association mapping is necessary.

### Supplementary material

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S1479262115000684>

### Acknowledgements

This research was supported by grants from the Center of Excellence in Cereal Molecular Breeding, University of Tabriz, Tabriz, Iran. The seed of Iranian wheat landraces was generously provided by the International Maize and Wheat Improvement Center (CIMMYT).

### References

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y and Buckler ES (2007) TASSEL: software for association mapping

- of complex traits in diverse samples. *Bioinformatics* 23: 2633–2635.
- Belo A, Zheng P, Luck S, Shen B, Meyer DJ, Li B, Tingey S and Rafalski A (2008) Whole genome scan detects an allelic variant of *fad2* associated with increased oleic acid levels in maize. *Mol Genet Genomics* 279: 1–10.
- Breseghele F and Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172: 1165–1177.
- Chao S, Zhang WJ, Dubcovsky J and Sorrells M (2007) Evaluation of genetic diversity and genome-wide linkage disequilibrium among U.S. wheat (*Triticum aestivum* L.) germplasm representing different market classes. *Crop Science* 47: 1018–1030.
- Chen X, Min D, Yasir TA and Hu YG (2012) Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. *PLoS ONE* 7: e44510. doi: 10.1371/journal.pone.0044510.
- Excoffier L (2001) Analysis of population subdivision. In: Balding D, Bishop M and Cannings C (eds) *Handbook of Statistical Genetics*. New York: John Wiley & Sons, pp. 271–307.
- Flint-Garcia SA, Thornsberry JM and Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* 54: 357–374.
- Garris AJ, McCouch SR and Kresovich S (2003) Population structure and its effects on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice *Oryza sativa* L. *Genetics* 165: 759–769.
- Gupta PK, Rustgi S and Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology* 57: 461–485.
- Hai L, Wagner C and Friedt W (2007) Quantitative structure analysis of genetic diversity among spring bread wheats (*Triticum aestivum* L.) from different geographical regions. *Genetica* 130: 213–225.
- Hamblin MT, Mitchell SE, White GM, Gallego J, Kukatla R, Wing RA, Paterson AH and Kresovich S (2004) Comparative population genetics of the panicoid grasses: sequence polymorphism linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. *Genetics* 167: 471–483.
- Hao CY, Wang LF, Ge HM, Dong YC and Zhang XY (2011) Genetic diversity and linkage disequilibrium in Chinese bread wheat (*Triticum aestivum* L.) revealed by SSR markers. *PLoS ONE* 6: e17279. doi: 10.1371/journal.pone.0017279.
- Huang XQ, Borner A, Roder MS and Ganai MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. *Theoretical and Applied Genetics* 105: 699–707.
- Hyten DL, Choi IY, Song Q, Shoemaker RC, Nelson RL, Costa JM, Specht JE and Cregan PB (2007) Highly variable patterns of linkage disequilibrium in multiple Soybean populations. *Genetics* 175: 1937–1944.
- Jaradat AA (1991) Levels of phenotypic variation for developmental traits in landrace genotypes of durum wheat (*Triticum turgidum* ssp. *Turgidum* L. conv. *durum* (Desf) MK) from Jordan. *Euphytica* 51: 265–271.
- Kraakman ATW, Martinez F, Mussiraliyev B, Van Eeuwijk FA and Niks RE (2006) Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. *Molecular Breeding* 17: 41–58.
- Kraakman ATW, Niks RE, Van der Berg PMMM, Stam P and Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168: 435–446.
- Kruger SA, Able JA, Chalmers KJ and Langridge P (2004) Linkage disequilibrium analysis of hexaploid wheat. In: *Plant & Animal Genomes XII Conference*, 10–14 January, Town & Country Convention Center, San Diego, CA, p. 321.
- Liu K and Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128–2129.
- Maccaferri M, Sanguineti MC, Donini P and Tuberosa R (2003) Microsatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. *Theoretical and Applied Genetics* 107: 783–797.
- Nielsen NH, Backes G, Stougaard J, Andersen SU and Jahoor A (2014) Genetic diversity and population structure analysis of European hexaploid bread wheat (*Triticum aestivum* L.) Varieties. *PLoS ONE* 9: e94000. doi: 10.1371/journal.pone.0094000.
- Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Ravel C, Praud S, Murigneux A, Linossier L, Dardevet M, Balfourier F, Dufour PH, Brunel D and Charmet G (2006) Identification of *Glu-B1-1as* a candidate gene for the quantity of high-molecular-weight glutenin in bread wheat (*Triticum aestivum* L.) by means of an association study. *Theoretical and Applied Genetics* 112: 738–743.
- Rhone B, Raquin AL and Goldringer I (2007) Strong linkage disequilibrium near the selected *Yr17* resistance gene in a wheat experimental population. *Theoretical and Applied Genetics* 114: 787–802.
- Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ and Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proceedings of the National Academy of Sciences of the United States of America* 103: 18656–18661.
- Saghai-Marouf MA, Soliman KM, Jorgensen RA and Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences of the United States of America* 81: 8014–8018.
- Skot L, Humphreys MO, Armstead I, Heywood S, Skot KP, Sanderson R, Thomas ID, Sanderson R, Chorlton KH and Hamilton NRS (2005) An association mapping approach to identify lowering time genes in natural populations of *Lolium perenne* (L.). *Molecular Breeding* 15: 233–245.
- Somers DJ, Banks T, Depauw R, Fox S, Clarke J, Pozniak C and McCartney C (2007) Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. *Genome* 50: 557–567.
- Somers DJ, Isaac P and Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* 109: 1105–1114.
- Song QJ, Shi JR, Singh S, Fickus EW, Cosra JM, Lewis J, Gill BS, Ward R and Cregan PB (2005) Development and mapping of microsatellite (SSR) markers in wheat. *Theoretical and Applied Genetics* 110: 550–560.
- Stachel M, Lelley T, Grausgruber H and Vollmann J (2000) Application of microsatellites in wheat (*Triticum aestivum* L.) for studying genetic differentiation caused by selection for adaptation and use. *Theoretical and Applied Genetics* 100: 242–248.
- Strelchenko P, Street K, Mitrofanova O, Hill H, Henry R and Mackay M (2008) Comparative assessment of wheat landraces from AWCC, ICARDA and VIR germplasm collections

- based on the analysis of SSR markers. In: *11th International Wheat Genetics Symposium*, 24–29 August 2008, Brisbane, Australia, pp. 309–311.
- Thomsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D and Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics* 28: 286–289.
- Tommasini L, Schnurbusch T, Fossati D, Mascher F and Keller B (2007) Association mapping of *Stagonospora nodorum* blotch resistance in modern European winter wheat varieties. *Theoretical and Applied Genetics* 115: 697–708.
- Wang ML, Barkley NA and Jenkins TM (2009) Microsatellite markers in plants and insects. Part I: applications of biotechnology. *Genes, Genomes and Genomics* 3: 54–67.
- Wright S (1951) The genetical structure of populations. *Annals of Eugenics* 15: 323–354.
- Xing Y, Frei U, Schejbel B, Asp T and Lubberstedt T (2007) Nucleotide diversity and linkage disequilibrium in 11 expressed resistance candidate genes in *Lolium perenne*. *BMC Plant Biology* 7: 43.
- Yu JM and Buckler ES (2006) Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology* 17: 1–6.
- Zhang DD, Bai GH, Zhu CS, Yu JM and Carver BF (2010) Genetic diversity, population structure and linkage disequilibrium in U.S. and Elite Winter Wheat. *Plant Genome* 3: 117–127.