

Genetic diversity in the Red wild einkorn: *T. urartu* Gandilyan (Poaceae: Triticeae)

Bernard R. Baum · L. Grant Bailey

Received: 9 November 2011 / Accepted: 17 February 2012 / Published online: 25 March 2012
© Her Majesty the Queen in Right of Canada 2012

Abstract Red wild einkorn, *Triticum urartu*, is increasingly being recognized as a source of genetic material for the improvement of wheat grain quality and for conferring resistance to various diseases such as Powdery Mildew and Leaf Rust resistance including the virulent race Ug99. Two hundred and two samples of *T. urartu* collected throughout much of its distribution were investigated by amplified restriction length polymorphism, AFLPTM to estimate the genetic diversity within and among them. To infer the genetic structure of the populations the data were subjected to analyses of molecular variance, AMOVA. The analyses of the samples enabled us to assess the location(s) of the richest area(s) in genetic diversity of the species. This area is found in north-western Syria and the adjacent South Turkey. It was also found that the similarity among populations did not reflect on their geographic closeness.

Keywords AFLP · AMOVA · DNA · Genetic diversity · Population genetic analysis · *Triticum urartu*

Electronic supplementary material The online version of this article (doi:10.1007/s10722-012-9817-7) contains supplementary material, which is available to authorized users.

B. R. Baum (✉) · L. G. Bailey
Agriculture and Agri-Food Canada, Neatby Bldg.,
960 Carling Ave., Ottawa, ON K1A0C6, Canada
e-mail: bernard.baum@agr.gc.ca

Introduction

Recent interest, such as Rouse and Jin (2011) in the A-genome diploid relatives of wheat as a source of resistance genes stemmed from the urgent need to find sources of genetic resistance to use against *Puccinia graminis* f. sp. *tritici* race TTKSK, known as Ug99, a virulent race to many resistance genes bred into wheat cultivars. The genome of *T. urartu* Gandilyan (the authority of the species name sometimes attributed to Thumanjan by Gandilyan, thus Thumanjan ex Gandilyan, both nomenclaturally correct) also known as the Red wild einkorn, is designated as **A^u** and is believed to be the contributor to both the macaroni wheat and the bread wheat (Dvorak et al. 1993). It is different from *T. monococcum* L. with a differently designated genome (**A^m**). Moghaddam et al. (2000) investigated the genetic diversity of *T. urartu* by isozymes markers. The present contribution assesses the genetic diversity of *T. urartu* using AFLP markers.

As eloquently mentioned by Moghaddam et al. (2000) the main reason for the low interest in *T. urartu*, was that for a long time research emphasis was put on *T. monococcum* L. which was believed to have been the donor of the **A** haplome to macaroni and bread wheats ($2n = 4x = 28$ **AABB** and $2n = 6x = 42$ **AABBDD** respectively). A brief history of the studies leading to the role of *T. urartu* as the donor of the **A** haplome was also provided by Moghaddam et al. (2000). The diploid wheat *T. urartu*, because of its recently understood role as one of the progenitors of

wheat, has been increasingly the subject of a number of investigations of genetic diversity using molecular markers. Isoenzyme analysis was carried out by Smith-Huerta et al. (1989) along with *T. boeoticum* Boiss. Using DNA-RAPD Vierling and Nguyen (1992) investigated seven genotypes of *T. monococcum* and six genotypes of *T. urartu*. Isoenzyme analysis on samples from 23 populations of *T. urartu* was carried out by Moghaddam et al. (2000). DNA-AFLP was used in a study by Sasanuma et al. (2002) concerned with *Ae. speltoides* Tausch, *T. boeoticum*, *T. urartu*, *T. dicoccoides* (Koern. ex Aschers. et Graebn.) Schweinf. and *T. araraticum* Jacubz., using 144 plants among 16 accessions. In a similar study Mizumoto et al. (2002) investigated the nuclear and chloroplast diversity in *T. urartu*, *T. monococcum*, and *T. boeoticum*, *Ae. speltoides*, *Ae. tauschii* Cosson (donor of the **D** haplome), and in three bread wheat cultivars. In a previous study Baum and Bailey (2004) we were concerned with the 5S rDNA variation of *T. urartu*. The objective of this study was to assess the genetic diversity of *T. urartu* based on material from populations representing as much as possible the area of its distribution.

Materials and methods

Materials and data acquisition methods

202 samples (Table S1 and Fig. 1) of *T. urartu* genomic DNA were received from Dr. Jan Dvorak (UC Davis, California, USA). These samples represent the distribution of *T. urartu* throughout most of its known range. To obtain DNA profiles, assays were carried out using AFLPTM Analysis System I (InvitrogenTM) according to the manufacturer's recommendation. First, 150 ng amounts of DNA per sample were digested to completion with EcoRI and MseI restriction enzymes. The resulting fragments were ligated to appropriate adaptors and pre-amplification of diluted ligations was performed prior to selective amplification. All EcoRI primers were labelled with $\gamma^{33}\text{PdATP}$ (Amersham) for this step. The following six selective amplification primer pairs were used (listed with their symbols used in this paper preceding them): A1: E-AACxM-CAA; E2: E-ACGxM-CAC; F6: E-ACTxM-CTC; F8: E-ACTxM-CTT; H2: E-AGGxM-CAC; H4: E-AGGxM-CAT. The completed reactions were

run on 7% polyacrylamide gels which were immediately dried and exposed to X-ray film at -80°C . Bands on the resulting autoradiographs were scored manually for presence/absence and a data matrix for each primer pair was assembled, and then the matrices were adjoined resulting in one data matrix.

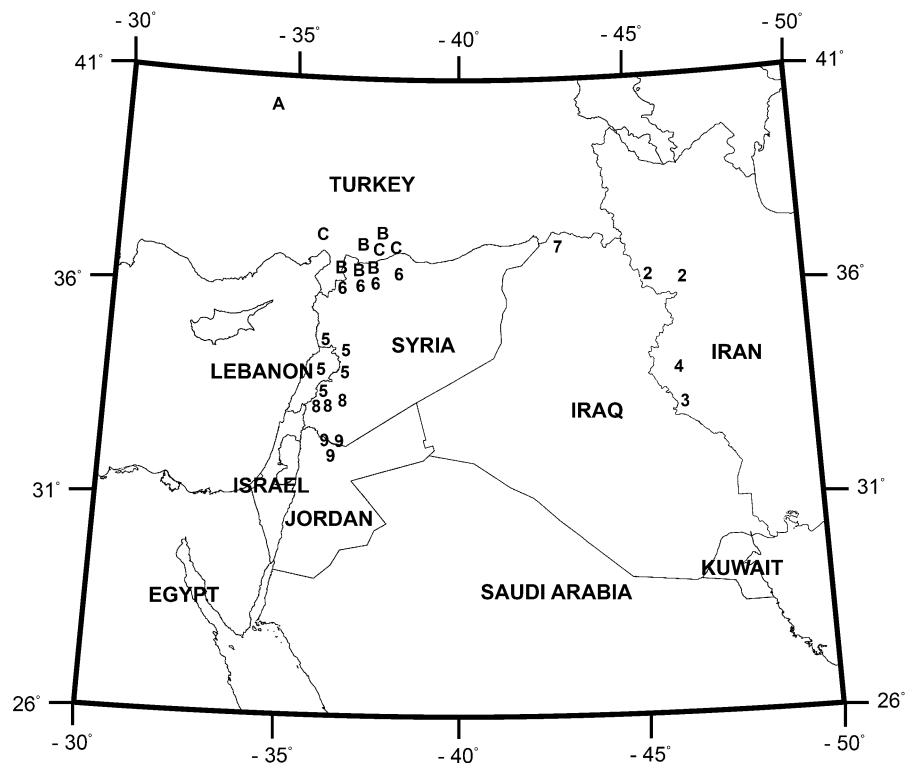
Data analysis

Genetic variation for each population and other standard diversity indices were first calculated. One statistic, gene diversity (H) is equal to "average heterozygosity" (Nei 1987: 177) of a population and was calculated as average heterozygosity over all loci. The expected gene diversity \hat{H} was also calculated. The other statistics calculated according to Nei (1973) were: H_s = average gene diversity within groups of populations; D_{st} = average gene diversity between groups of populations; H_t = H_s + D_{st} = gene diversity in the total sample, i.e. the species or the group consisting of all the populations; and G_{st} = estimate of gene flow between populations. Also for the entire *T. urartu* species the following estimates were computed: H ; H_c which is the gene diversity between the groups, H_s , and G_{cs} which is the same as G_{st} but between groups. For a summary of the genetic diversity analyses see the legend of Table 2.

To study the relationships between the populations we carried out a cluster analysis. Since the data consisted of a small number of individuals for most populations we computed the pairwise distances between them according to Nei (1978) and the resulting distance matrix was subjected to a UPGMA cluster analysis (Sokal and Michener 1958). Since at least one population was fairly large (Table 2) we also computed the pairwise distances according to Nei (1972) and subjected this to UPGMA clustering for comparison.

To describe and compare the genetic variation for multiple AFLP loci (bands), and especially the partition of variance between groups and subgroups of populations (see below how the population groupings were explored), we carried out an Analysis of MOlecular VAriance (AMOVA) (Excoffier et al. 1992) justified by Huff et al. (1993) and in Peakall et al. (1995) for diploid dominant data such as RAPD as is the case for AFLP also, i.e. for binary type data. AMOVA consists of a hierarchical analysis of variance that partitions the total variance into covariance

Fig. 1 Collection sites of *T. urartu* samples from which populations were defined. 1 Former Yugoslavia (not shown); 2 Iran–Azerbaijan West; 3 Iran–Ilam; 4 Iran–Kermanshahan; 5 Lebanon–Baalbek; 6 Syria–Aleppo; 7 Syria–Al Hassaka; 8 Syria–Damascus; 9 Syria–Sweida; 10 Turkey (no specified locality, not shown); A Turkey–Çorum; B Turkey–Gaziantep; C Turkey–Sanli to Urfa



components (Excoffier 2000; Hartl and Clark 1997) due to intra-individual differences, inter-individual differences, inter-population differences and inter-regional differences. In other words the source of variation for several groups of populations is computed for among groups, among populations within groups, within populations and within and among regions. AMOVA as implemented in Excoffier and Lischer (2010) computes Euclidian distances. We also conducted an additional AMOVA based on Gower distances for comparison. A greater number of analyses were carried using GenAEx which is an “Add-In” module written by Peakall et al. (1995) for Microsoft® Excel™.

Specifically we carried out the following AMOVA analyses (Table): (1) 10 populations without regional structure, i.e. 1 region; (2) 3 populations without regional structure with the same statistics; (3) 9 populations with 3 regions; (4) 9 populations without regional structure, and used the following statistics PhiPT, PhiPTPV, PhiPTFD, PhiPTP, PhiPTL, PhiPTT (acronyms as in Peakall et al. 1995). PhiPT represents the correlation between individuals within a

population, relative to the total (of sampled individuals in our AFLP data). PhiPT (Φ_{pt}) is an analogue of F_{st} , and is also an estimate of population genetic differentiation provided by the GenAEx when binary data or haploid data are analyzed (Peakall et al. (1995). The AFLP data are binary and therefore suitable for analysis by Arlequin and GenAEx with its statistical tests, which according to Maguire et al. (2002) is best for binary data. Statistical testing by random permutation is facilitated in GenAEx to obtain an estimate of the value one would expect if the null hypothesis was true. We chose to use 99 permutations for the appropriate tests. For more details and for the calculations and tests see Peakall et al. (1995) who also provide differences from some calculations in Arlequin. For each measure and its formula please refer to Appendix 1, Table 2 in Peakall et al. (1995).

To explore how to group populations into regions (in population genetics parlance) we took two alternative approaches each yielding a number of possible groupings. The first was based on geography and a combination of geography together with ecology reasoning such as altitude and/or vegetation, whereas the second was

based on clustering of the haplotypes using the Gower (1971) general resemblance coefficient and Modeclus clustering where each cluster solution was followed by a discriminant analysis to justify and to validate the resulting cluster differences. The clustering approach is obviously based on haplotypes pairwise resemblance as opposed to geographical location and may therefore not reflect actual population content. Nevertheless, the clustering approach is justified by the specific method of clustering, i.e. Modeclus clustering (Sarle and An-Hsiang 1993), which has properties different than commonly used methods such as UPGMA. The UPGMA method attempts to produce compact hyperspherical clusters, attempts to equalize the variance among clusters (Sarle 1982). Modeclus instead results in clusters with unspecified shapes (sausage like) that are irregular in shape in hyperspace which might perhaps result in approximating the “real” populations. The different cluster solutions were also subjected to AMOVA using GenAIEx to compare with the AMOVA results based on populations and for reification.

Estimates of the various statistics were calculated with PopGene version 1.31 (Yeh and Boyle 1997) and population structure estimates and additional statistics were computed with Arlequin version 3.5.1.2 (Excoffier and Lischer 2010) and with GenAIEx (Peakall and Smouse 2006); the latter differs in philosophy and in some principles from the former. Cluster analysis of populations was carried out with NTSYS-pc version 2.1 (Rohlf 2000) whereas clustering by Modeclus and the distribution map were carried out with SAS version 9.1 (SAS Institute 2004).

Results

An example of an AFLP autoradiograph is shown (Fig. 2). Of the 202 samples 198 were scored, resulting in 223 polymorphic loci out of 381. Almost all the haplotypes were unique in the data matrix (of the primers combined). The exceptions were in the population Syria–Aleppo where one haplotype was present in two individuals and another in three individuals. When only one individual was available in a population it was deleted, or incorporated with other populations depending on the population structure being analyzed (Table 1, Table S1).

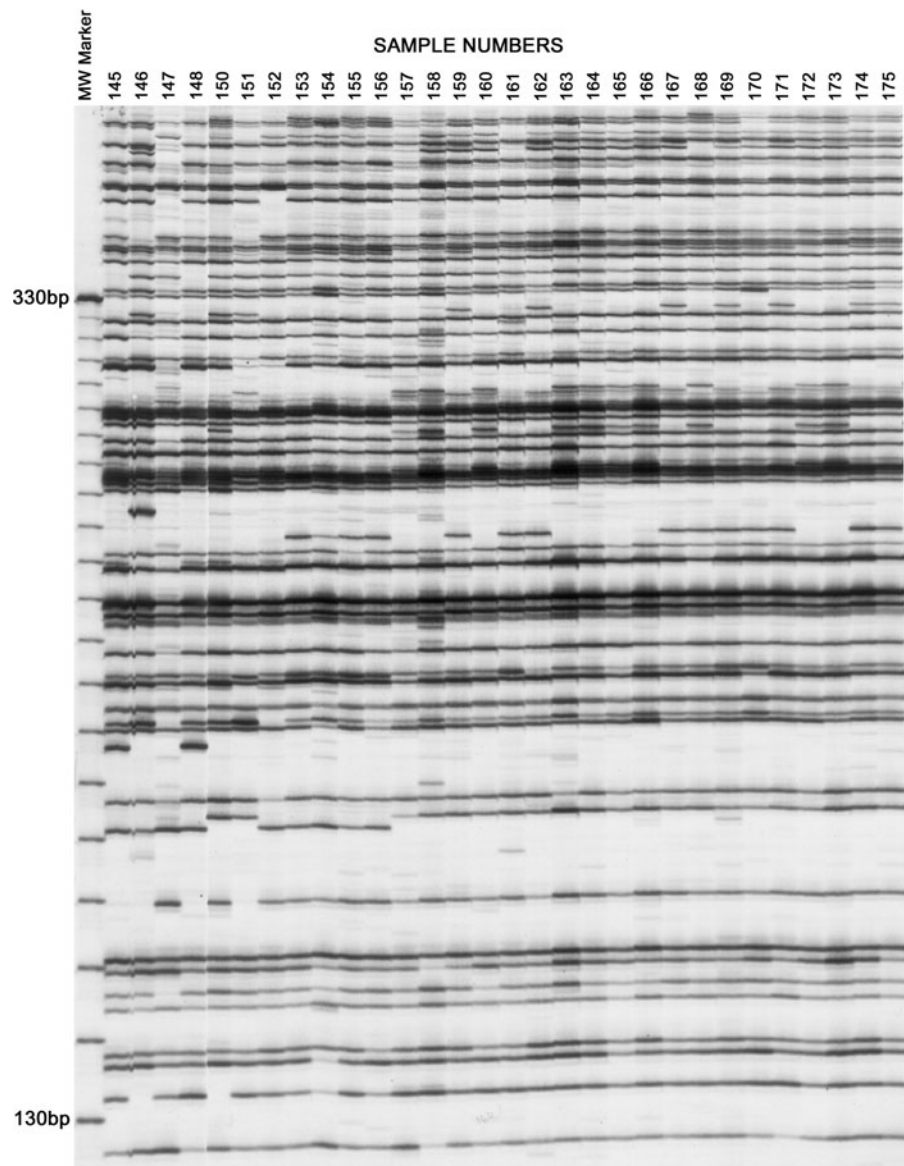
Getting populations into groupings as regions

Six possibilities of groupings of populations were summarized (Table S1). The first three were inferred based on geographical distribution (Fig. 1), i.e. based on latitude/longitude (Table S1). Population groupings 4–6 resulted from cluster analysis followed by discriminant analysis. Initial Modeclus cluster analysis pointed to 4, 3, and 2 clusters, which was inferred from the three regions of stability in the graph (Fig. 3; K values 17–20, 21–24 and 25–29 respectively). The three cluster solutions were subsequently analyzed using the specific k smoothing parameter appropriate for each grouping indicated in the legend of Table S1. These three groupings were initially judged to be unacceptable possibilities for the following reasons: some geographically defined populations, especially Syria–Aleppo exhibited an incoherent mixture of haplotypes (Table S1 grouping 4–6) and the classification results of the discriminant analyses (not shown) especially the cross-validation results which exhibited for at least one population values as low as 20% correct classification. Of the three possible population groupings based on geography or on a combination of geography and ecology the third grouping (marked * in Table S1) made the most intuitive sense. For instance populations 2, 3, 4 and 7 (Fig. 1) thrive in the same habitat, i.e. at the edges of the clearings of the *Quercus* forests, north and eastern fringes of the Fertile Crescent.

Relationship among populations

The 13 populations fell into identical major clusters whether using the Nei (1972) or Nei (1978) genetic distances (Fig. S1), but some differences were found within the major ones. Population 3 from Ilam was remote from the rest genetically as well as geographically. Population 2 from West Azerbaijan was closer to the Turkish populations from the Gaziantep area (populations 12 and 13 depicted as B and C in Fig. 1). Population 6 from the Aleppo areas fell closest to population 8 from the Damascus areas based on Nei's (1978) distances whereas it fell as closest to the Gaziantep population when clustering was based on the Nei (1972) distance coefficient. More details are shown in Fig. S1.

Fig. 2 Section of an autoradiograph of the AFLP gel resulting from using primer pairs E-AGG and M-CAT in the *T. urartu* study. *MW* molecular weight marker (in base pairs); samples 145–152 (149 excluded) from Turkey–Sanli Urfa and 153–175 from Syria–Aleppo, see Table 1



Population genetics estimates

The statistics for the accepted grouping (Table S1) are presented in Table 2. The mean observed mean genetic diversity estimates (H) are always far below the expected values (\hat{H}). The gene diversity in group (H_t) exhibits comparable values to those averaged from the H values of the populations. Population differentiation, i.e. G_{st} , is highest in group 2, intermediate in group 3 and very low in both groups 4 and 5. Most of the

variation, judged from H_s , is present in group 4 consisting of populations 6 (Syria–Aleppo), 12 (Turkey–Gaziantep) and 13 (Turkey Sanli Urfa). Clearly the highest amount of diversity is present in these three populations based on H (Table 2). This is perhaps the geographical center of the Fertile Crescent (Fig. 1). Not surprisingly the highest G_{st} was found in regions 4 (56%) in the northern part of the Fertile Crescent and 2 (49%) which consists of scattered and isolated populations at the eastern fringe of the Fertile Crescent.

Table 1 Design of AMOVA analyses and of genetic diversity analyses conducted in *T. urartu* populations and on groups into regions

Population number	Original sites	Number of samples	10 populations 1 region	Number of samples	3 populations 1 region	Number of samples	9 populations 1 region	Number of samples	9 populations, number of samples as to the left	3 regions numbered
1	Former_Yugoslavia	1		1	N_Syria-Turkey					
2	Iran-Azerbaijan West	2	Azerbaijan West	2	Iran-Iraq	2	Zagros	2	Zagros	1
3	Iran-Ilam	2	Ilam	2	Iran-Iraq	2	Zagros	2	Zagros	1
4	Iran-Kermanshahan	1		1	Iran-Iraq	1	Zagros	1	Zagros	1
5	Lebanon-Baalbek	5	Baalbek	5	Lebanon-Syria	5	Baalbek	5	Baalbek	2
6	Syria-Aleppo	102	Aleppo	102	N_Syria-Turkey	102	Aleppo	102	Aleppo	3
7	Syria-Al-Hassaka	1		1	Iran-Iraq	1	Zagros	1	Zagros	1
8	Syria-Damascus	3	Damascus	3	Lebanon-Syria	3	Damascus	3	Damascus	2
9	Syria-Sweida	11	Sweida	11	Lebanon-Syria	11	Sweida	11	Sweida	2
10	Turkey	4	Turkey	4	N_Syria-Turkey	4	Turkey	4	Turkey	3
11	Turkey-Çorum	4	Çorum	4	N_Syria-Turkey	4	Çorum	4	Çorum	3
12	Turkey-Gaziantep	34	Gaziantep	34	N_Syria-Turkey	34	Gaziantep	34	Gaziantep	3
13	Turkey-Sanlı Urfa	25	Sanli Urfa	25	N_Syria-Turkey	25	Sanli-Urfa	25	Sanli-Urfa	3
Total		195		195		194		194		

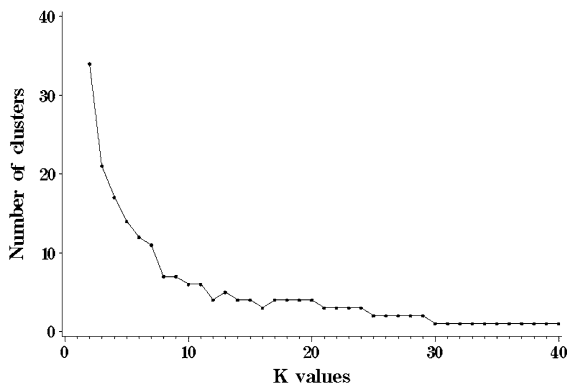


Fig. 3 Modeclus preliminary cluster analysis. Plot of the number of clusters against k nearest-neighbor values; see text

AMOVA

The main AMOVA results following the various analytical designs (Table 1) were depicted in the following pie charts (Fig. 4a–g). When 10 populations were subjected to AMOVA the variance components within populations was 71% and among populations 29% (Fig. 4a). When the 13 populations were pulled together as three populations as indicated the results only slightly differ (Fig. 4b). The results are similar for nine populations given that four original populations from the Zagros Mountains area were pulled together (Fig. 4c). When the same nine populations were divided into three regions the variance components among regions was 16%, among populations was 19% and within populations 65% (Fig. 4d). Regarding the Modeclus cluster solutions, with four clusters the percentages of molecular variance are 70% within populations and 30% among populations (Fig. 4e). The AMOVA results of the three Modeclus cluster solutions, to be compared with the previously ones in the Discussion section, are as follows. In the four clusters solution the percentages of molecular variance were 70% within populations and 30% among populations; and respectively in the three clusters solution 77 and 23% (Fig. 4f) and in the two clusters solution 83 and 17% (Fig. 4g).

AMOVA results using Arlequin with the option of using squared Gower distances instead of squared Euclidian distances yielded similar results in general, for instance with the population structure of nine populations and three regions (67.39% within populations, 21.47% among populations within groups and 11.14% among groups).

Discussion

The center of diversity of *T. urartu* is found evidently in the northern tip of the Fertile Crescent (Fig. 1) in an area and around the Syrian Turkish border, based on the population genetic estimates and the AMOVA analyses. However, this does not tell us anything about the kind of variation. A gene of interest for a specific purpose may indeed be found outside the center of diversity. Based on the clustering of haplotypes by Modeclus the population Syria–Aleppo is the most diverse one as it contains haplotypes of most of the clusters in the three different grouping solutions (groupings 4–6 in Table S1). A careful examination shows that the population Turkey–Sanli Urfa is closest to the populations in Iran, Lebanon and the extreme South Syria (Syria–Sweida population), which contain the majority of haplotypes that belong to cluster 4 in grouping solution 4 for example. This indicates a more complex dissemination and origin of the different haplotypes and thus of gene content. Evidently this indicates that for gene conservation and use for improvement it is desirable to put emphasis on the entire area of distribution instead of concentrating on the center of diversity.

Smith-Huerta et al. (1989) reported a low genetic diversity in four populations of *T. urartu* using star gel electrophoresis. Three of those populations were close to population Syria–Al Hassaka in our study (Fig. 1 No 7) and one from Lebanon–Baalbek (Fig. 1 No 5). Our values are also low but slightly higher than those in Smith-Huerta et al. (1989) primarily due to the higher multiplex ratio of AFLP, although they (Smith-Huerta et al. 1989) reported that Yaghoobi-Sarray (1979) found higher values than theirs probably due to the different enzyme system used. As a highly self-pollinated species one is expected to find low diversity (Nevo 1978; Hamrick et al. 1979; Smith-Huerta 1986).

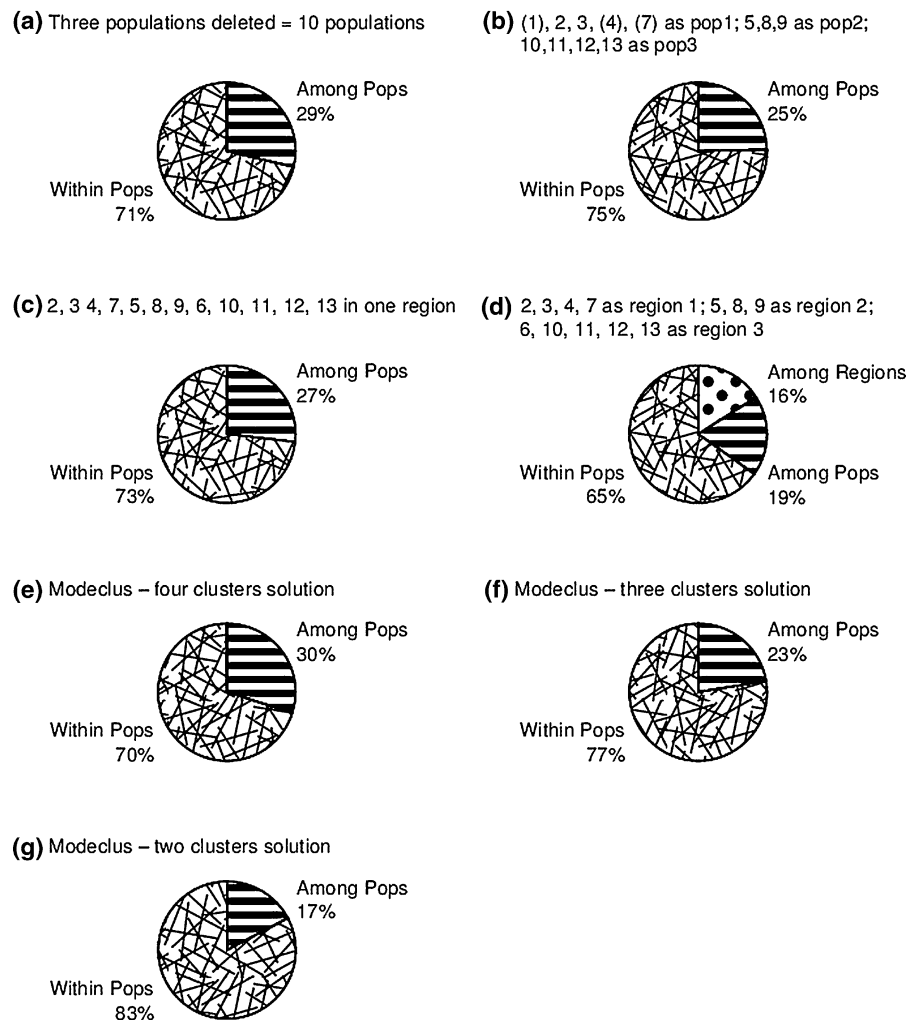
Moghaddam et al. (2000) devoted a whole genetic diversity study to *T. urartu* populations using isoenzyme markers at eight polymorphic loci. A number of their population sites were near or the same as ours but the Syria–Aleppo population was missing from Moghaddam et al. (2000) and six among the 23 populations were near the Syria–Al Hassaka locality in our study (Fig. 1 No 7) but on the Turkish side. Their two Iranian populations were located further East than ours and therefore from a dryer area. Overall their percent polymorphic loci ranged similar to ours. The diversities

Table 2 Population genetic estimates for *T. urartu* from samples in the area of distribution

Population	Sample size	Polymorphic loci (%)	H	\hat{H}	I	Hr	Hc	Hs	Gst	Gcs	Nm	Nm(Gcs)
1. Former Yugoslavia**	1	0		1.0 ± 0.0								
2. Iran–Azerbaijan West	2	30 (7%)	0.0557 ± 0.1417	1.0 ± 0.5	0.0814 ± 0.2058							
3. Iran–Ilam	2	101 (26%)	0.1875 ± 0.2057	1.0 ± 0.5	0.2739 ± 0.3017							
4. Iran–Kermanshahan*	1	0		1.0 ± 0.0								
5. Lebanon–Baalbek	5	74 (19%)	0.1191 ± 0.1854	1.0 ± 0.1265	0.1776 ± 0.2671							
6. Syria–Aleppo	105	199 (52%)	0.2053 ± 0.1680	0.9993 ± 0.0015	0.3252 ± 0.2364							
7. Syria–Al-Hassaka*	1	0		1.0 ± 0.0								
8. Syria–Damascus	3	46 (5%)	0.0804 ± 0.1638	1.0 ± 0.2722	0.1185 ± 0.2376							
9. Syria–Sweida	11	87 (10%)	0.1305 ± 0.1866	1.0 ± 0.0388	0.1965 ± 0.2694							
10. Turkey -	4	116 (30%)	0.1426 ± 0.2017		0.2084 ± 0.2886							
11. (A) Turkey–Çorum	4	116 (30%)	0.1694 ± 0.2099	1.0 ± 0.0625	0.2476 ± 0.2991							
12. (B) Turkey–Gaziantep	34	199 (52%)	0.2587 ± 0.1756	1.0 ± 0.0071	0.3967 ± 0.2393							
13. (C) Turkey–Sanli Urfu	25	119 (31%)	0.1773 ± 0.1992	1.0 ± 0.0113	0.2661 ± 0.2849							
14. Region 2, Pops 2, 3, 4, 7	6	167 (49%)	0.2631 ± 0.1826	1.0 ± 0.0962	0.3958 ± 0.2579	0.2376 ± 0.0309		0.0608 ± 0.0037	0.7439		0.1721	
15. Region 3, Pops 5, 8, 9	19	133 (34%)	0.1747 ± 0.1827	1.0 ± 0.0171	0.2702 ± 0.2638	0.1922 ± 0.0389		0.1100 ± 0.0166	0.4277		0.6689	
16. Region 4, Pops 6, 12, 13	164	215 (56%)	0.2337 ± 0.1623	0.9997 ± 0.0007	0.3685 ± 0.2220	0.2442 ± 0.0279		0.2138 ± 0.0218	0.1246		3.5143	
17. Region 5, Pops 10, 11	8	116 (30%)	0.1895 ± 0.2027	1.0 ± 0.0625	0.2817 ± 0.2911	0.1895 ± 0.0411		0.1560 ± 0.0309	0.1768		2.3282	
18. Species	198	223 (58%)	0.2437 ± 0.1649	0.9998 ± 0.0006	0.3826 ± 0.2206	0.2459 ± 0.0278	0.1174 ± 0.0063	0.2030 ± 0.0173	0.1746	0.4214	2.3637	0.6865

Population 1–13 samples collected in different sites, i.e. colonies; 14–17 groups of populations; 18 all samples belonging to *T. urartu*. * included in 14; ** included in 18; *H* mean gene diversity (Nei 1973) in a population, group or species (followed by standard error (SE)); \hat{H} expected gene diversity for diploid data (Nei 1987: 180); *I* Shannon index (Lewontin 1972) followed by SE; *Hr* Gene diversity in group; *Hs* average gene diversity within populations; *Gst* population differentiation; *Gcs* same as *Gst* but between groups; *Nm* Estimate of number of migrants between population within groups (computed from *Gst*) or between groups within species (computed from *Gcs*). See Fig. 1 for distribution (Populations 11, 12, 13 are labelled A, B, C respectively in the figure)

Fig. 4 AMOVA analyses of populations of *T. urartu*. See text



in their (Moghaddam et al. 2000) study were generally lower than ours probably attributed to the differences in the marker systems. As we did not group the population by countries, but by geographical-ecological considerations, we were unable to make meaningful comparisons with Moghaddam et al. results, although their inter-population component of diversity accounted for 40.7% of the total compared to 29 and 59% component within population compared to 65–75% in ours. The differences between the two studies were likely due to a combination of sampling and the marker systems used.

Sasanuma et al. (2002) analyzed 37 plants in total belonging to four accessions, as part of a study with a different aim than ours and that also included *T. boeoticum*, *T. dicoccoides*, *T. araraticum* and *Ae. speltoides*. Their (Sasanuma et al. 2002) reported percentage polymorphic bands, and mean gene

diversity (their Table 4) were roughly comparable to ours, but the within population variance component (their Table 5) was 7.1% remarkably different than ours—65–75% Fig. 2). This difference in the within population variance component may be due to the difference between the two studies because they dealt with differences between species, and found that the variance component between *T. urartu* and *T. boeoticum* was 69.5% whereas the among populations component—23.5% (in their Table 5) was similar to ours—25–29%.

Based on pairwise similarity analysis of the AFLP banding, Mizumoto et al. (2002) found a much lower diversity in *T. urartu* and in the other two diploid wheats compared to the polyploid species in their study. However, Mizumoto et al. (2002) found the highest diversity in the chloroplast genome of

T. urartu. But, their comparison of nuclear DNA with chloroplast DNA is not accurate because the former was based on AFLP of total DNA (including chloroplast DNA) whereas the latter was based on chloroplast specific SSLP primers.

In our opinion, to expect differences in population genetic structure among countries as Moghaddam et al. (2000) did would not be appropriate for the reason that species are adapted to specific eco-geographical conditions, not to political boundaries. Sasanuma et al. (2002) found that no such relationship exists but thought that those expectations failed to materialize due to the small number of accessions used in their study. As erstwhile mentioned, clustering of haplotypes is designed to reflect relationships between haplotypes by pairwise resemblance and not expected to assess relationships between geography and distribution. This became obvious especially in the Syria–Aleppo population in the clustering results (Table S1, grouping 4–6). Furthermore, simple population genetic interpretation of phenetic cluster diagrams is problematical (Hollingsworth and Ennos 2004).

In the comparison of the AMOVA results of the different designs (Table 1) with the AMOVA results of the three different Modeclus cluster solutions the following observations can be made. The percentages of the molecular variance of the Modeclus four cluster solution of the haplotypes (Fig. 4e) are identical to the results of the 10 populations structure (Fig. 4a); similarly the three clusters solution (Fig. 4f) is roughly similar to the three populations in one region (Fig. 4b), whereas the two clusters solution (Fig. 4g) is closer to the nine populations into three regions when combining the within and among percentages of the molecular variance (65% + 19%) 84 and 16% among regions (Fig. 4d) but is dissimilar with the population structure of nine populations in one region (Fig. 4c). This raises the following question: can one assume that some sort of clustering can be found that would predict population structure from the pattern of the haplotypes without knowledge of their distribution?

Finally, we fully concur with view of Sasanuma et al. (2002) that for a continued accessibility of genetic resources that the best way is *in situ* conservation. In the case of *T. urartu*, the area of greatest genetic diversity is found in the area north-west Syria–south Turkey (Fig. 1 6, B, C; see also Table 2) and although important, it is the kind of variation exhibited in the different populations in the total area of the

species that is of significance for tapping genes useful for the improvement of the wheat crop, including of course conferring disease resistance.

Acknowledgments This manuscript benefited from the careful comments of Dr. Stephen Molnar, Agriculture & Agri-Food Canada, Ottawa.

References

- Baum BR, Bailey LG (2004) The origin of the A genome donor of wheats (*Triticum*: Poaceae)—a perspective based on the sequence variation of the 5S DNA gene units. *Genet Resour Crop Evol* 51:183–196
- Dvorak J, di Terlizzi P, Zhang HB, Resta P (1993) The evolution of polyploid wheats, identification of the A genome donor species. *Genome* 36:21–31
- Excoffier L (2000) Analysis of population subdivision. In: Balding D, Bishop M, Canning C (eds) *Handbook of statistical genetics*. Wiley
- Excoffier L, Lischer H (2010) ARLEQUIN Ver 3.5.1.2. A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland. Accessed Dec 2010. <http://cmpg.unibe.ch/software/arlequin35>
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Gower JC (1971) A general coefficient of similarity and some of its properties. *Biometrics* 27:671–857
- Hamrick JL, Linhart YB, Mitton JB (1979) Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. *Ann Rev Ecol Syst* 10:173–200
- Hartl DL, Clark AG (1997) *Principles of population genetics*, 3rd edn. Sinauer Associates, Sunderland
- Hollingsworth PM, Ennos RA (2004) Neighbour joining trees, dominant markers and population genetic structure. *Heredity* 92:490–498
- Huff DR, Peakall R, Smouse PE (1993) RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloeë dactyloides* (Nutt.) Engelm.]. *Theor Appl Genet* 86:927–934
- Lewontin RC (1972) The apportionment of human diversity. *Evol Biol* 6:381–398
- Maguire TL, Peakall R, Saenger P (2002) Comparative analysis of genetic diversity in the mangrove species *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae) detected by AFLPs and SSRs. *Theor Appl Genet* 104:388–398
- Mizumoto K, Hirose S, Nakamura C, Takumi S (2002) Nuclear and chloroplast genome genetic diversity in the wild einkorn wheat, *Triticum urartu*, revealed by AFLP and SSLP analyses. *Heredity* 137:208–214
- Moghaddam M, Ehdai B, Waines JG (2000) Genetic diversity in populations of wild diploid wheat *Triticum urartu* Tum. ex Gandil. revealed by isozyme markers. *Genet Resour Crop Evol* 47:323–334

- Nei M (1972) Genetic distance between populations. *Amer Nat* 106:283–292
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Nat Acad Sci USA* 70(part 1): 3321–3323
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590
- Nei M (1987) *Molecular evolutionary genetics*. Columbia Univ Press, New York
- Nevo E (1978) Genetic variation in natural populations: pattern and theory. *Theor Popul Biol* 13:121–177
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6: 288–295. Version 6.4 available at <http://www.anu.edu.au/Bozo/GenALEX/>
- Peakall R, Smouse PE, Huff DR (1995) Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloë dactyloides*. *Mol Ecol* 4:135–147
- Rohlf FJ (2000) NTSYS-pc. Numerical taxonomy and multivariate analysis system. Ver 2.1. User guide. Exeter Software. Setauket, NY
- Rouse MN, Jin Y (2011) Stem rust resistance in A-genome diploid relatives of wheat. *Plant Dis* 95:941–944
- Sarle WS (1982) Cluster analysis by least squares. In: SAS users group international conference proceedings, vol SUG17, pp 651–653
- Sarle WS, An-Hsiang K (1993) The MODECLUS procedure. SAS Technical Report P-256. SAS Institute Inc, Cary
- SAS Institute (2004) SAS/STAT® user's guide. Ver 9.1, SAS Institute Inc., Cary
- Sasanuma T, Chabane K, Endo TR, Valkoun J (2002) Genetic diversity of wheat wild relatives in the Near East detected by AFLP. *Euphytica* 127:81–93
- Smith-Huerta NL (1986) Isozymic diversity in three allotetraploid *Clarkia* species and their putative diploid progenitors. *J Hered* 77:349–354
- Smith-Huerta NL, Huerta AJ, Barnhart D, Waines JG (1989) Genetic diversity in wild diploid wheats *Triticum monococcum* var. *boeoticum* and *T. urartu* (Poaceae). *Theor Appl Genet* 78:260–264
- Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull* 38:1409–1438
- Vierling RA, Nguyen HT (1992) Use of RAPD markers to determine the genetic diversity of diploid wheat genotypes. *Theor Appl Genet* 84:835–838
- Yaghoobi-Saray J (1979) An electrophoretic analysis of genetic variation within and between populations of five species in *Triticum-Aegilops* complex. PhD Thesis, University of California, Davis.
- Yeh FC, Boyle T (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian J Bot* 129:157–163