Theor Appl Genet (1993) 86:518–525

Storage-protein variation in wild emmer (*Triticum turgidum* ssp. *dicoccoides*) from Jordan and Turkey. II. Patterns of allele distribution

M. Ciaffi [1], D. Lafiandra [1], E. Porceddu [1], S. Benedettelli [2]

[1] Department of Agrobiology and Agrochemistry, University of Tuscia, 01100 Viterbo, Italy
[2] Agroforestry Institute, C.N.R., 05010 Porano (Terni), Italy

Received: 13 August 1992 / Accepted: 12 December 1992

Abstract. Genetic diversity in the seed storage-proteins encoded at the *Glu-A1*, *Glu-B1* and *Gli-B1/Glu-B3* loci was studied electrophoretically in 315 individuals belonging to nine populations of *T. dicoccoides* from Jordan and three from Turkey. The inter- and intra-population distribution of seed storage-protein alleles at the considered loci and its link with geographical factors were investigated. Population differentiation in seed storage-proteins was in some cases very high with very weak correlations with geographic distance. Greater gene differentiation was found within and between populations which were geographically very close in Jordan than between those from Jordan and Turkey. However the distribution of alleles appeared to be non random. Samples collected from populations at locations over 900 m above sea level were less polymorphic than those collected at lower altitudes (500–700 m), whereas the relative genetic differentiation between populations was greater between those collected at higher altitudes. Seed storage-protein differentiation was significantly correlated with the altitude of the collecting sites. Although it is difficult to point out the selective pressure of altitude per se, altitude can reflect an integration of several environmental parameters. The possible adaptive value of seed storage-proteins is discussed.

Key words: *Triticum turgidum* ssp. *dicoccoides* – Seed storage-protein polymorphism – Geographical pattern – Adaptive value

Introduction

Plant breeding has eroded the genetic variability among and within cultivars in many crop species, including wheat, thus making crop plants increasingly vulnerable to new disease and environmental stresses and reducing the possibility of further improvements. Restoration and enrichment of gene pools can be accomplished by utilizing the vast genetic variation found in the wild relatives of wheats (Feldman and Sears 1981). Consequently, knowledge of the amount and distribution of genetic variability in the wild species is an important pre-requisite for optimal use of this resource in plant breeding.

In recent years, electrophoretically discernible seed storage-proteins have been used to assess variation in cereal populations, landraces, and modern cultivars. A peculiar characteristic of storage proteins is their high level of heterogeneity which is strictly genetically determined. Since these proteins are direct gene products, relatively unaffected by environmental conditions, they can provide, either independently or complementary to other analyses, a reasonably accurate measure of genetic diversity and, in fact, they have been used in a wide array of studies, such as varietal identification (Bushuk and Zillman 1978), the evaluation of phylogenetic relationships (Fernandez-Calvin and Orellana 1990; Lafiandra et al. 1992), genome homologies (Kreis et al. 1985), the genetic structure of cultivated wheat germplasm (Lafiandra et al. 1990), and genetic diversity within and between populations (Nevo and Payne 1987).

Although seed storage-proteins might have additional functions other than providing amino acids for the germinating seed, the adaptive value of their interspecific (Ladizinsky 1983) and intraspecific variation (Levy et al. 1988) has not been clarified. Moreover, the neutralist-selectionist controversy on the nature of polymorphism in proteins has not been adequately resolved with regard to seed storage-proteins. This problem has both theoretical and practical implications. If neutral, then the abundant polymorphism in seed storage-proteins is predominantly

Communicated by A. F. Linskens

of interest only for identifying genetic markers which can be used as a basis for evolutionary studies. If, conversely, the existing polymorphism has adaptive value, then variation can be directly exploited to maximize sampling and breeding efficiency. A possible adaptive nature of seed storage-protein polymorphism in wild emmer was reported by Nevo and Payne (1987), who suggested that at least part of the variation in HMW-glutenin subunits could be accounted for by physical (climate and soil) and biotic (vegetation) variables.

The total polymorphism of seed storage-proteins at the *Glu-A1, Glu-B1* and *Gli-B1/Glu-B3* loci present in different *T. dicoccoides* accessions from Jordan and Turkey has been described in a previous paper (Ciaffi et al. 1993). In the present study we have investigated the inter- and intra-population distribution of seed storage-protein alleles at the above mentioned loci and examined whether this distribution is accountable in terms of geographical factors.

Materials and methods

Plant material

The samples used in the study came from the previously described nine populations of *T. dicoccoides* from Jordan and the three from Turkey (Ciaffi et al. 1993). Data on collecting sites are given in Table 1 together with the number of genotypes per populations. Collecting sites extended from 17 km northwest of Irbid in northern Jordan (population 1) to 14 km northwest of Madaba in central Jordan (population 9). Samples from populations 10 and 11 were collected in Turkish Kurdistan, near Siverek, whereas those from population 12 (Oguzeli) were collected close to the Syrian border.

Statistical analyses

The frequency of alleles at the *Glu-A1, Glu-B1* and *Gli-B1/Glu-B3* loci (Ciaffi et al. 1993) were calculated for each population. The expected heterozygosity at each locus and the average heterozygosity per population (He), the proportion of polymorphic loci (P), the average number of alleles per locus (A) and the average effective number of alleles per locus (Ne), were used to evaluate within-population genetic diversity (Nei 1987). The degree of genetic differentiation between populations was estimated using the following four parameters: genetic similarity coefficients (I) (Nei 1972); Chord distances (Cavalli Sforza and Edwards 1967); Roger's distances as modified by Wright (1978); Nei's genetic distances (D), calculated using the coefficient of genetic similarity (I) plus a constant ($D = -\ln I + 0.0005$). This transformation is indispensable for overcoming the complexity created by the high number of seed storage-protein alleles per locus present in *T. dicoccoides* which can result in a value of I equal to zero when the two populations compared displayed alternative alleles at the three considered loci. In this particular case the D value ($D = -\ln I$) is equal to infinity, but in the computer calculations it is recorded as zero.

Dendrograms, based on different estimates of between-population genetic differentiation, were obtained by complete link, flexible clustering, single link, and UPGMA, WPGMA, WPGMC clustering methods (Rohlf 1987). These were tested by co-phenetic matrix correlation during the reconstruction of a co-phenetic matrix based on tree matrices. Co-phenetic matrices were then compared with the original Chord, Roger's and Nei's matrices in order to establish the most significant one.

Associations between variability for seed storage-proteins at the considered loci and geographical factors were examined using Mantel's test (Mantel 1967) between the Chord genetic distance matrix and geographic distance matrices. Two matrices for geographical factors were computed: (1) the distance matrix between points relative to longitude and latitude; (2) the distance matrix between points relative to altitude.

Principal component analysis, performed on an allele frequency matrix, was used to study the relationships between seed storage-protein variation and the geographic distribution of the 12 *T. dicoccoides* populations.

Statistical analyses were carried out using NTSYS (Rohlf 1987) and STATGRAPHICS (STSC 1987) software.

Table 1. Geographical coordinates of collecting sites and number of genotypes per population in analyzed material

Population	Location	Lat (N)	Long (E)	Altitude (m)	No. of genotypes per population
Jordan					
1	17 km NW Irbid	32.37	35.54	650	32
2	Natifah	32.31	35.49	900	24
3	7 km SE Irbid	32.27	35.35	700	42
4	Zobia	32.24	35.47	900	11
5	Afana	32.23	35.41	900	35
6	Ibbein	32.18	35.47	700	32
7	Za'atri	32.04	35.45	1,000	25
8	Naor	31.40	35.40	600	22
9	Mabada	31.35	35.50	500	25
Turkey					
10	27 km NE Siverek	37.57	39.31	1,000	27
11	Herber 20 km NE Siverek	37.52	39.24	950	15
12	Oguzeli	36.58	37.32	650	25

Results

Geographical distribution of alleles and genetic variation within and between populations

Allele frequencies at the *Glu-A1*, *Glu-B1* and *Gli-B1/Glu-B3* loci in the 12 populations of wild emmer analysed are shown in Table 2. A wide range, from monomorphism to very high polymorphism, characterizes the seed storage-proteins at the considered loci. Gene diversity values (He or expected heterozygosity), estimated for each of the three loci, indicated that some populations were monomorphic at all the loci considered (e.g., population 4), others at two (population 11) or at only one of them (population 2). Some populations had relatively high levels of polymorphism, with three alleles at *Glu-A1* and

Table 2. Allele frequency and gene diversity at the *Glu-A1*, *Glu-B1* and *Gli-B1/Glu-B3* loci in 12 populations of *T. dicoccoides* from Jordan and Turkey

		Jordan									Turkey			
Sites		1	2	3	4	5	6	7	8	9	10	11	12	Total
Number of genotypes		32	24	42	11	35	32	25	22	25	27	15	25	315
Locus	Allele													Mean
Glu-A1	1	0	0	0	0	0	0	0	0	0	0.26	0	0	0.022
	2	0	0	0	0	0	0	0	0	0	0	0	0.16	0.013
	3	0.16	0	0.14	0	0	0.28	0	0.18	0	0	0	0	0.076
	4	0	0	0	0	0	0	0	0.09	0	0	0	0	0.006
	5	0	0	0	0	0	0.10	0	0	0	0	0	0	0.010
	6	0.08	0	0.07	0	0	0.10	0	0	0	0	0	0	0.025
	7	0	0	0	0	0	0.12	0	0	0	0	0	0.40	0.045
	8	0	0	0	0	0	0	0	0	0	0.18	0	0	0.016
	9	0	0	0	0	0	0	0	0	0	0.41	0	0.08	0.042
	10	0	0	0	0	0.20	0	0	0	0	0	0	0	0.022
	11	0	0.88	0	1	0.40	0	0.52	0	0	0.15	1	0	0.268
	12	0.20	0	0.38	0	0	0.40	0	0.25	0.60	0	0	0.12	0.188
	13	0.44	0.12	0.41	0	0	0	0	0.30	0.08	0	0	0.16	0.128
	14	0.12	0	0	0	0	0	0.08	0.18	0.32	0	0	0.08	0.063
	15	0	0	0	0	0.40	0	0.40	0	0	0	0	0	0.076
He		0.72	0.22	0.66	0	0.64	0.72	0.56	0.77	0.53	0.71	0	0.76	0.524
Glu-B1	1	0	0	0	0	0.40	0.50	0	0.23	0.24	0	0	0.12	0.140
	2	0	0	0	0	0.11	0	0	0.09	0	0	0	0	0.019
	3	0	0.88	0	1	0.31	0	0.30	0	0	0	0	0	0.160
	4	0	0	0	0	0.09	0	0.24	0	0	0	0	0	0.028
	5	0	0	0	0	0.09	0	0	0	0	0	0	0	0.010
	6	0	0	0	0	0	0	0	0	0	0	0.20	0	0.010
	7	0	0	0.17	0	0	0	0	0	0.08	0	0	0	0.028
	8	0.06	0	0	0	0	0.06	0	0.09	0	0	0	0.40	0.051
	9	0	0	0	0	0	0	0.34	0	0	0.26	0	0	0.049
	10	0.22	0.12	0.26	0	0	0.44	0	0.27	0.20	0.30	0	0	0.171
	11	0.36	0	0.31	0	0	0	0	0.18	0.28	0	0	0.08	0.119
	12	0	0	0	0	0	0	0	0	0	0.07	0	0	0.006
	13	0.22	0	0.26	0	0	0	0	0.14	0.20	0	0	0	0.084
	14	0	0	0	0	0	0	0	0	0	0	0.33	0	0.016
	15	0	0	0	0	0	0	0.12	0	0	0.30	0.47	0	0.057
	16	0	0	0	0	0	0	0	0	0	0	0	0.12	0.010
	17	0	0	0	0	0	0	0	0	0	0.07	0	0	0.006
	18	0.14	0	0	0	0	0	0	0	0	0	0	0.20	0.030
	19	0	0	0	0	0	0	0	0	0	0	0	0.08	0.006
He		0.75	0.22	0.74	0	0.71	0.55	0.72	0.80	0.77	0.73	0.63	0.76	0.615
Gli-B1/ Glu-B3	1	0	1	0	1	0.48	0	0.56	0	0	0.56	1	0	0.305
	2	0.44	0	0.40	0	0	0	0	0.27	0.16	0	0	0	0.130
	3	0	0	0.22	0	0	0.44	0	0.18	0	0	0	0	0.086
	4	0	0	0	0	0	0	0	0	0.20	0	0	0	0.035
	5	0	0	0	0	0	0	0	0	0	0	0	0.36	0.028
	6	0	0	0	0	0	0	0.32	0	0	0	0	0	0.025
	7	0	0	0.14	0	0	0	0	0	0	0	0	0	0.019
	8	0	0	0	0	0	0.37	0	0	0	0	0	0	0.038
	9	0	0	0	0	0.31	0	0	0	0	0	0	0	0.016
	10	0	0	0	0	0	0	0	0.18	0	0	0	0	0.013
	11	0	0	0	0	0	0.19	0	0	0.24	0	0	0	0.038
	12	0.16	0	0	0	0	0	0	0	0.12	0	0	0	0.025
	13	0.22	0	0.24	0	0	0	0	0	0.28	0	0	0	0.076
	14	0	0	0	0	0	0	0	0	0	0.33	0	0	0.028
	15	0	0	0	0	0	0	0.12	0	0	0.11	0	0	0.019
	16	0	0	0	0	0	0	0	0.23	0	0	0	0.40	0.048
	17	0	0	0	0	0	0	0	0	0	0	0	0.16	0.013
	18	0	0	0	0	0.21	0	0	0	0	0	0	0	0.022
	19	0.18	0	0	0	0	0	0	0.14	0	0	0	0.08	0.036
He		0.70	0	0.71	0	0.62	0.63	0.62	0.79	0.78	0.57	0	0.68	0.508

Table 3. Storage-protein diversity based on the *Glu-A1*, *Glu-B1* and *Gli-B1/GluB3* loci in 12 populations of *T. dicoccoides* in Jordan and Turkey

Population	Sample size (n)	Mean no. of alleles per locus (A)	Effective number of alleles (N_e)	Mean proportion of loci			Genic diversity (He)
				Polymorphic per population (P-5%)	Heterozygous per individual		
					Mean	SE	
Jordan							
1	32	4.67	3.61	1.00	0.021	0.015	0.723
2	24	1.67	1.17	0.67	0.000	0.000	0.147
3	42	4.00	3.37	1.00	0.000	0.000	0.703
4	11	1.00	1.00	0.00	0.000	0.000	0.000
5	35	3.67	2.92	1.00	0.000	0.000	0.657
6	32	3.67	2.72	1.00	0.000	0.000	0.633
7	25	3.33	2.72	1.00	0.013	0.013	0.633
8	22	5.33	4.69	1.00	0.015	0.015	0.787
9	25	4.33	3.31	1.00	0.000	0.000	6.697
Mean	248	3.52±0.46	2.83±0.39	0.85±0.11	0.005	0.0028	0.553±0.093
Turkey							
10	27	4.00	3.03	1.00	0.000	0.000	0.670
11	15	1.67	1.27	0.33	0.000	0.000	0.210
12	25	5.33	3.74	1.00	0.000	0.000	0.733
Mean	67	3.67±1.07	2.68±0.73	0.78±0.22	0.000	0.000	0.538±0.160
Total Mean	315	3.56±0.41	2.80±0.33	0.83±0.1	0.004	0.002	0.549±0.077

Gli-B1, as in population 5; others displayed very high levels of polymorphism, with five alleles at *Glu-A1* and *Gli-B1*, and six in *Glu-B1*, as in population 8, the most polymorphic population in this study.

Out of the 53 alleles detected at the *Glu-A1, Glu-B1* and *Gli-B1/Glu-B3* loci, 24 occurred only in the Jordanian material (six for *Glu-A1*, six for *Glu-B1* and 12 for *Gli-B1*) and 13 were present only in the Turkish material (four for *Glu-A1*, six for *Glu-B1* and three for *Gli-B1*); the remaining 16 had common origins (Table 2).

Allele distribution varied among the populations studied. To assess the various kinds of allele distribution we followed the classification proposed by Marshall and Brown (1975). Some alleles were rare (never occurring with a frequency $\geq 10\%$) and unique to specific populations, like allele 4 at *Glu-A1* and alleles 5, 12, 17 and 19 at *Glu-B1* (Table 2). Other alleles, classified as common (at least one sample with a frequency $\geq 10\%$), were also detected in only one population (eight alleles in Jordanian and nine in Turkish material). On the other hand some common alleles were widespread across the 12 locations, like alleles 11, 12, 13, and 14 at *Glu-A1*, alleles 1, 8, 10, 11 and 15 at *Glu-B1*, and allele 1 at *Gli-B1*. Others were widespread only in populations from Jordan, like allele 3 at *Glu-A1*, alleles 3 and 13 at *Glu-B1*, and alleles 2, 3 and 13 at *Gli-B1*. Finally, some alleles were sporadic (common occurrence in two populations), such as allele 15 at *Glu-A1* and allele 4 at *Glu-B1*, which were found in populations from close locations in Jordan, alleles 2 and 7 at *Glu-B1* and alleles 11 and 12 at *Gli-B1*, which occurred in geographically distant sites in Jordan, and such as allele 7 of *Glu-A1*, alleles 9 and 18 of *Glu-B1*, and allele 16 of *Gli-B1*, which were common to Jordanian and Turkish populations (Table 2).

The mean number of alleles per locus (A), the effective number of alleles (Ne), polymorphism (P−5%) and genic diversity (He) in the 12 populations of wild emmer were equal to 3.56, 2.80, 0.83 and 0.549, respectively (Table 3). These estimates of genetic diversity were slightly higher than those reported by Nevo and Payne (1987) for HMW-glutenin subunits in 11 wild emmer populations; this is probably due to the greater numbers of protein loci and genotypes per population considered in the present study.

Estimates computed by considering materials from Jordan and Turkey as two megapopulations indicated the absence of statistical differences between the countries for genetic diversity at the considered loci. The mean number of alleles per locus (A) was 3.52 and 3.67 for the two groups and the effective number of alleles (Ne) was equal to 2.83 and 2.68; the proportion of polymorphic loci was 0.85 versus 0.78, and the genic diversity (He) 0.553 and 0.538, respectively.

The mean of Wright's fixation index (F, Wright 1965) was equal to 1.0 in most populations, except in populations 1, 7, and 8, which had observed mean-heterozygos-

Table 4. Coefficients of genetic similarity (I), based on the *Glu-A1*, *Glu-B1*, *Gli-B1/Glu-B3* loci, between 12 populations of *T. dicoccoides* from Jordan and Turkey

Location	2	3	4	5	6	7	8	9	10	11	12
1	0.056	0.864	0.000	0.000	0.246	0.009	0.740	0.620	0.064	0.000	0.214
2	------	0.055	0.992	0.687	0.036	0.746	0.056	0.023	0.457	0.761	0.013
3	----------	----------	0.000	0.000	0.478	0.000	0.743	0.667	0.084	0.000	0.161
4	----------	----------	----------	0.686	0.000	0.744	0.000	0.000	0.413	0.750	0.000
5	----------	----------	----------	----------	0.219	0.697	0.113	0.099	0.329	0.569	0.053
6	----------	----------	----------	----------	----------	0.000	0.634	0.522	0.146	0.000	0.209
7	----------	----------	----------	----------	----------	----------	0.017	0.024	0.431	0.689	0.007
8	----------	----------	----------	----------	----------	----------	----------	0.612	0.100	0.000	0.369
9	----------	----------	----------	----------	----------	----------	----------	----------	0.061	0.000	0.183
10	----------	----------	----------	----------	----------	----------	----------	----------	----------	0.559	0.000
11	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	0.000

Mean 0.387; Range 0.000–0.992

ity levels per individual (H) equal to 0.021, 0.013 and 0.015, respectively. Thus the mean of F is 0.996, in agreement with the value reported by Nevo et al. (1988) and Nevo and Beiles (1989) in their allozyme variation studies. These values of H and F were expected since *T. dicoccoides* is predominantly self pollinated, with only about 1% outcrossing as reported by Golemberg (1986).

Total gene diversity (Ht) across the three loci in the 12 populations was equal to 0.8715 (ranging from 0.8598 for *Glu-A1* to 0.8904 for *Glu-B1*), with the within- and between-population components accounting for about 68% and 32% of the total, respectively. The average relative differentiation among populations was, in fact, Gst = 0.3151 (ranging from 0.2729 for *Glu-B1* to 0.3417 for *Gli-B1/Glu-B3*). This means that a substantial amount of gene differentiation exists among populations, although lower than that estimated by Nevo and Payne (1987) for HMW-glutenin subunits in wild emmer from Israel.

A comparison of the material analysed from the two different countries produced similar results, whereas comparisons relative to the collecting sites indicated that total genetic variance (Ht) was much higher in the populations collected below 700 m than those collected above 900 m (0.8327 vs 0.6466) and that genetic variance within populations (Hs) followed the same trend (0.7036 vs 0.4590). In contrast, the two measures of genetic differentiation among populations (Dst and Gst) showed a higher separation of demes within populations collected at high elevations (Dst = 0.1821 and Gst = 0.2774) compared with those collected at lower altitudes (Dst = 0.1293 and Gst = 0.1542). This result is probably due mainly to a reduced genetic variability within the populations from the mountainous areas at altitudes over 900 m, as shown by the estimates of genetic diversity (Table 3).

Genetic distances

Coefficients of genetic similarity (I) (based on the normalized identity, Nei 1972) at the *Glu-A1, Glu-B1* and *Gli-B1/Glu-B3* loci were calculated for all possible paired comparisons between the 12 populations. The estimates, reported in Table 4, can be used only for comparing the 12 populations studied here for seed storage-proteins at the considered loci, and not as a general estimate of genetic similarity, as was done on a multilocus allozymic basis for wild emmer (Nevo and Beiles 1989). The mean of I for HMW-glutenin subunits and for *Gli-B1/Glu-B3*-encoded proteins was 0.387 (range 0.000–0.992) in our study. Populations with I = 0 (e.g., populations 1–4) displayed alternative alleles at the three considered loci. In contrast, populations with I = 0.992 (e.g., 2–4) shared mostly the same range of allelic variation.

The estimates of genetic similarity between populations were independent of the collecting site. Some geographically-close populations were very different in their seed storage-protein structure at the considered loci (e.g., populations 1–2, 7.6 km apart, I = 0.056, or populations 6–7, 21.3 km apart, which had no alleles in common). In contrast some geographically-distant populations were very similar in their seed storage-proteins (e.g., populations 1–8, 108.6 km apart, I = 0.740). An extreme example was that of populations 2 (Natifah in Jordan) and 11 (Herber in Turkey), separated by 672,4 km, for which I was very high (0.761) compared with the lower values obtained for population 2 and the other Jordanian populations. Many other such dramatic contrasts occurred and are given in Table 4.

Dendrograms, based on genetic distance matrices and obtained by different clustering methods, were tested for the significance of the clustering method. The highest coefficient of co-phenetic correlation (R = 0.975) was ob-

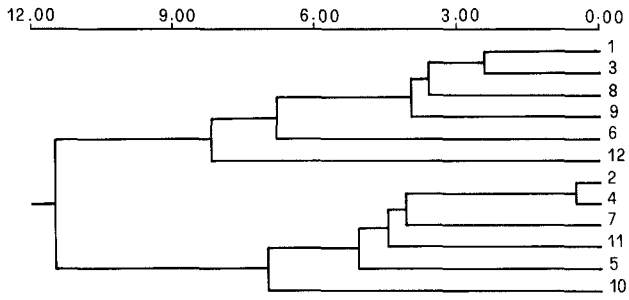

Fig. 1. UPGMA dendrogram, based on the Chord genetic distance matrix, of the Jordanian and Turkish populations of *T. dicoccoides*

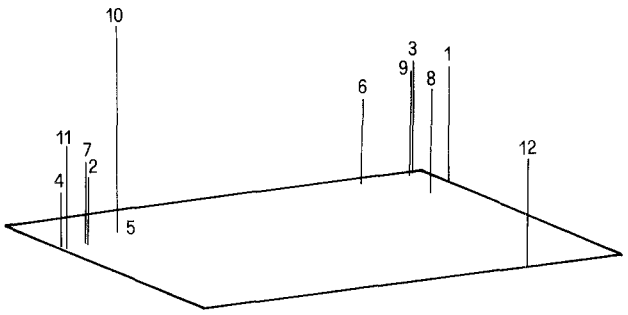

Fig. 2. Three-dimensional principal components diagram of the Jordanian and Turkish populations of *T. dicoccoides*

tained for the UPGMA clustering algorithm based on the Chord distance matrix, the results of which are shown in Fig. 1, whereas the highest values of the co-phenetic correlation coefficient calculated on Nei's and Roger's distances were lower than 0.8 (data not shown). It is worth noting that values of co-phenetic correlation above 0.9 represent a very good fit of a cluster to the data (Rohlf and Fisher 1986).

The pattern that emerges from the dendrogram shown in Fig. 1 has no simple explanation in relation to the geographic localization of the populations considered. However, it is interesting to note the presence of two subclusters, one formed by populations 1, 3, 8, 9, 6, 12, collected at altitudes below 700 m, and the other by populations 2, 4, 7, 11, 5, 10, collected at sites above 900 m. The distribution was confirmed by Mantel's test; genetic distances were in fact significantly correlated with the altitude difference matrix ($R = 0.7829$), whereas they did not show direct association with longitude and latitude ($R = 0.1369$). This shows that when altitude distances are small, genetic distances also tend to be small, whereas when geographic distances (km) are small the corresponding genetic distances are randomly distributed, indicating the lack of a clear relationship between the location of origin of a population and its gene frequency.

Principal components analysis

To ascertain which alleles were involved in the distribution of the 12 populations obtained by the UPGMA dendrogram, a principal components analysis was carried out on the frequency matrix of the 53 alleles detected at the considered loci. The results indicated that the first six principal components accounted for about 80% of the total variation, with the first three explaining respectively 22.03%, 17.38% and 13.17% of total variation. The distribution obtained by plotting the first three eigenvectors calculated for the 12 populations produced a clear separation (Fig. 2), similar to that obtained by the UPGMA dendrogram based on the Chord distance matrix.

The correlation coefficients between the principal components and the original variables (Table 5) indicated that the first principal component contained the contrast between the alleles from low and high altitudes. It was, in fact, positively correlated with the *Glu-B1-13*, *Gli-B1-2* and *Gli-B1-13* alleles, which were widespread in the Jordanian populations collected at elevations below 700 m, and with the *Glu-A1-12*, *Glu-A1-13* and *Glu-B1-11* alleles which were also present in population 12 (Oguzeli, Turkey), and negatively correlated with the frequencies of the *Glu-A1-11*, *Glu-B1-3* and *Gli-B1-1*, which were very common in populations collected at elevations above 900 m. The second principal component was negatively correlated with the frequencies of alleles that were mainly present in population 12, whereas the third was positively correlated with the frequencies of alleles peculiar to population 10. These two populations are, in fact,

Table 5. Association between the first ten principal components and the original variables[a]

Principal components	Original variables (alleles)
1	*Glu-A1-11* (−0.869), *Glu-A1-12* (+0.840), *Glu-A1-13* (+0.709) *Glu-B1-3* (−0.691), *Glu-B1-11* (+0.815), *Glu-B1-13* (+0.766) *Gli-B1-1* (−0.942), *Gli-B1-2* (+0.703), *Gli-B1-13* (+0.694)
2	*Glu-A1-2* (−0.912), *Glu-A1-7* (−0.839) *Glu-B1-8* (−0.879), *Glu-B1-16* (−0.912), *Glu-B1-19* (0.912) *Gli-B1-5* (−0.912), *Gli-B1-16* (−0.801), *Gli-B1-17* (−0.912)
3	*Glu-A1-1* (+0.737), *Glu-A1-8* (+0.737) *Glu-B1-12* (+0.737) *Gli-B1-14* (+0.737)
4	*Glu-A1-3* (−0.691), *Glu-A1-5* (−0.841) *Glu-B1-1* (−0.792), *Glu-B1-10* (−0.695) *Gli-B1-3* (−0.758), *Gli-B1-8* (−0.841)
5	*Glu-A1-11* (+0.935), *Glu-B1-3* (+0.690), *Gli-B1-1* (+0.867)
6	
7	*Glu-B1-4* (+0.685), *Gli-B1-7* (0.693), *Gli-B1-8* (+0.689)
8	*Gli-B1-4* (+0.717), *Gli-B1-11* (0.719)
9	*Glu-A1-11* (−0.686), *Glu-B1-3* (−0.986), *Gli-B1-1* (−0.691)
10	*Glu-B1-7* (+0.817)

[a] The close association was determined considering the coefficient of correlation ($R > 0.6835$ corresponding to $P < 0.01$ with 11 *df*

separated along the second and third principal component axes, respectively (Fig. 2). The fourth principal component accounted for the frequencies of alleles widespread in low-altitude populations (*Glu-A1-3, Glu-B1-1, Glu-B1-10* and *Gli-B1-3*) or peculiar to one of them (*Glu-A1-5* and *Gli-B1-8*). Information given by the remaining components was in some cases additional to that already provided by the previous ones (principal components 5 and 9) and in others relative to some alleles that were unique to specific populations.

A closer inspection of the distribution of seed storage-protein alleles at the considered loci, showed a strong trend of clustering both at a local (within a population) and at a regional level (comparison between populations). For example, out of the 47 genotypes that carried one of the *Glu-A1-13, Glu-B1-11* and *Gli-B1-2* alleles, 62% was characterized by a cluster of all three alleles and, in general, by the same gliadin and glutenin patterns. Additionally, out of the 104 genotypes with at least one of the *Glu-A1-11, Glu-B1-3* and *Gli-B1-1* alleles, 42% showed the same storage-protein pattern with all three alleles again clustered. These, and other examples of multilocus associations not indicated here, suggest that the distribution of seed storage-protein alleles could be affected both by single and multilocus structure.

Discussion

The analysis of storage-protein variation in wild relatives of *Triticum* species and in particular in wild emmer *T. dicoccoides*, the wild progenitor of all polyploid cultivated wheats, is important for two major reasons. First, practically, for evaluating the potential genetic resources for technologically useful characteristics, in terms of the type and amount of storage proteins. Second, for evolutionary studies; as the level of polymorphism is very high, storage proteins can contribute to the analysis of the evolutionary forces responsible for population genetic structure and differentiation.

The high levels of seed storage-protein diversity, found in our and other studies (Nevo and Payne 1987; Levy et al. 1988), contrast with the relatively lower levels of isozyme diversity present in wild emmer. The mean number of alleles per locus (A) in *T. dicoccoides* from Israel (Nevo et al. 1982) and from Turkey (Nevo et al. 1988) were 1.33 and 1.22 for 50 and 48 loci, respectively. In contrast, in our study the values of A for four seed storage-protein loci in wild emmer from Jordan and Turkey were 3.52 and 3.56, respectively. A number of reasons can be proposed to explain the differences between the isozyme and the seed storage-protein results. First, seed storage-proteins are not randomly distributed since they are encoded by multigene families which provide additional opportunities for polymorphism. A second factor may depend on the lower level of functional constraints that seed storage-proteins have in comparison to isozymes.

Results from the present study support the hypothesis that seed storage-proteins are, at least partly, adaptive, and so may experience some sort of divergent selection. Population differentiation, as clearly indicated by the analysis of genetic distances, was, in some cases, very high, with very weak correlations to geographic distances. We have found substantially more gene differentiation within- and between-populations, sometimes geographically very close, in Jordan than between wild emmer in Jordan and Turkey. Of the total genetic diversity of *T. dicoccoides*, 68% existed within populations, 32% between populations, whereas only 3% was present between Jordan and Turkey when treated as two mega-populations. The altitude of collecting sites seems to be an important criterion: populations collected at locations above 900 m were less polymorphic than those collected at lower altitudes (500–700 m), whereas the relative genetic differentiation between populations increased in those collected at higher altitudes. Reduction of variability and higher genetic differentiation among populations could be due to the severe conditions prevailing at higher altitudes. The geographical patterns of seed storage-proteins and their correlations with altitude suggest that they may be due in part to natural selection, and cannot be accounted for using migration and genetic drift as major explanatory models. The fact that nearby populations are very different in seed storage-protein content, whereas in contrast distant populations are very similar, indicates that migration plays little or no role in population genetic differentiation of these proteins. Neither can genetic drift be easily invoked to explain allele distribution. If seed storage-proteins were neutral, their spatial pattern would be expected to display marked interpopulation differentiation independent of the environmental factors. However, seed storage-protein differentiation was highly correlated with the altitude of the collecting site; although it is difficult to point out the selective pressure of altitude per se, it might be that altitude reflects an integration of several parameters such as rainfall, temperature and radiation. The sporadic and localized distribution of some alleles of wild emmer could be determined by other physical and biotic factors which we have not considered.

These results, however, provide no conclusive proof of the direct adaptiveness of seed storage-protein polymorphism. Another class of models may be invoked to explain the high diversity of seed storage-proteins found in *T. dicoccoides* on the assumption that while the variation at these loci may be selectively neutral it is linked to loci which are targets of selection (Thompson 1977). These models are especially relevant in annual self-fertilizing species, such as *Avena*, *Hordeum* and *Triticum*

where an extensive multilocus association beyond the physical linkage groups is present, because inbreeding would intensify any associations between loci (Ohta and Cockerham 1974).

Jaradat and Humeid (1990) studying the variation for morphological and physiological traits in a wild emmer collection from Jordan, from which part of the material utilized in the present analysis was derived, showed that two distinct forms, resembling the 'grassy' and 'robust' types (Poyarkova 1988), were present. The first, collected from mountainous areas of Jordan, at altitudes between 900 and 1,350 m, was characterized by short, slender plants with short spikes and fewer tiller per plants and by late heading and a longer filling period. The second, characterized by tall, early-maturing plants which have large spikes, with a high number of spikelets per spike and a large number of tillers per plant, came from warmer sites at lower altitudes. These findings and the high correlation detected between the genetic differentiation in seed storage-proteins and the altitude of the collecting sites, suggest a possible association between seed storage-protein polymorphism and phenotypical polymorphism. More detailed and extensive studies, including samples from additional ecogeographical distribution areas of *T. dicoccoides*, would allow clarification of this point.

Acknowledgements. This research was supported by C.N.R., Italy; Special Grant R.A.I.S.A., sub-project 2, paper no. 761.

References

Bushuk W, Zillmann RR (1978) Wheat cultivar identification by gliadin electrophoregrams. I. Apparatus, method, and nomenclature. Can J Plant Sci 58:505–515

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analyses: models and estimation procedures. Evolution 32:550–570

Ciaffi M, Lafiandra D, Porceddu E, Benedettelli S (1993) Storage protein variation in wild emmer (*Triticum turgidum* ssp *dicoccoides*) from Jordan and Turkey. I. Electrophoretic characterization of genotypes. Theor Appl Genet

Feldman M, Sears ER (1981) The wild gene resources of wheat. Sci Am 244:102–112

Fernandez-Calvin B, Orellana J (1990) High-molecular-weight glutenin subunit variation in the *Sitopsis* section of *Aegilops*. Implication for the origin of the B genome of wheat. Heredity 65:455–463

Golemberg EM (1986) Multilocus structures in plant populations: population and genetic dynamics of *Triticum dicoccoides*. PhD thesis, State University of New York at Stony Brook

Jaradat AA, Humeid BO (1990) Morphological variation in *Triticum dicoccoides* from Jordan. In: Srivastava JP, Damania AB (eds) Wheat genetic resources: meeting diverse needs. Aleppo, Syria, pp 215–222

Kreis M, Shewry PR, Forde BG, Miflin BJ (1985) Structure and evolution of seed storage-proteins and their genes with particular reference to those of wheat, barley and rye. Oxford Surveys Plant Mol Cell Biol 2:253–317

Ladizinsky G (1983) Study of evolutionary problems by means of seed proteins electrophoresis. In: Gottschalk W, Muller HP, Nijhoff M (eds) Seed proteins, biochemistry, genetics, nutritive value. Junk, The Hague Boston London, pp 481–498

Lafiandra D, Benedettelli S, Margiotta B, Spagnoletti-Zeuli PL, Porceddu E (1990) Seed storage-proteins and wheat genetic resources. In: In: Srivastava JP, Damania AB (eds) Wheat genetic resources: meeting diverse needs. Aleppo, Syria, pp 73–87

Lafiandra D, Masci S, D'Ovidio R, Tanzarella OA, Porceddu E, Margiotta N (1992) Relationship between the D genome of hexaploid wheats (AABBDD) and *Ae. squarrosa* as deduced by seed storage-proteins and molecular-marker analyses. Hereditas 116:233–238

Levy AA, Galili G, Feldman M (1988) Polymorphism and genetic control of high-molecular-weight glutenin subunits in wild tetraploid wheat *Triticum turgidum* var. *dicoccoides*. Heredity 61:63–72

Mantel NA (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Marshall DR, Brown AHD (1975). Optimum sampling strategies in genetic conservatin. In: Frankel OH, Hawkes JG (eds) Crop genetic resources for today and tomorrow. Cambridge University Press, Cambridge, pp 53–70

Nei M (1972) Genetic distances between populations. Am Nat 106:283–292

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nevo E, Payne PI (1987) Wheat storage proteins: diversity of HMW-glutenin subunits in wild emmer from Israel. 1. Geographical patterns and ecological predictability. Theor Appl Genet 74:827–836

Nevo E, Beiles A (1989) Genetic diversity of wild emmer wheat in Israel and Turkey. Structure, evolution, and application in breeding. Theor Appl Genet 77:421–455

Nevo E, Golemberg E, Beiles A, Brown ADH, Zohary D (1982) Genetic diversity and environmental associations of wild wheat, *T. dicoccoides* in Israel. Theor Appl Genet 62:241–254

Nevo E, Beiles A, Kaplan D (1988) Genetic diversity and environmental associations of wild emmer wheat in Turkey. Heredity 61:31–45

Ohta T, Cockerham CC (1974) Detrimental genes with partial selfing and effects on a neutral locus. Genet Res 23:191–200

Poyarkova H (1988) Morphology, geography and infraspecific taxonomics of *Triticum dicoccoides* Korn. A retrospective of 80 years of research. Euphytica 38:11–23

Rohlf FJ (1987) NTSYS-pc, Numerical Taxonomy and Multivariate Analysis System for the IBM PC microcomputer and compatibles, Exeter Software, Setauket NY, USA

Rohlf FJ, Fisher DL (1968) Test for hierarchical structure in random data sets. Systematic Zool 17:407–412

STATGRAPHICS, (STSC) (1987) Statistical Graphics System by Statistical Graphics Corporation, User's Guide, version 3 ed. STSC Inc. NY, USA

Thompson G (1977) The effect of a selected locus on linked neutral loci. Genetics 85:753–788

Wright S (1965) The interpretation of population structure by F-statistics with special regard to system of matings. Evolution 19:395–420

Wright S (1978) Evolution and the genetics of populations, vol. 4. Variability within and among natural populations. University of Chicago Press, Chicago, pp 220–246