Theor Appl Genet (1993) 86:54–64

Isozyme variation in germplasm accessions of the wild oat *Avena sterilis* L.

T. D. Phillips*, J. P. Murphy, and M. M. Goodman

Department of Crop Science, North Carolina State University, Raleigh, NC 27695-7629, USA

Received May 4, 1992; Accepted September 19, 1992
Communicated by A. R. Hallauer

Summary. Optimal exploitation of crop genetic resources requires a knowledge of the range and structure of the variation present in the gene pool of interest. *Avena sterilis* L., the cultivated oat progenitor, contains a store of genetic diversity that is readily accessible to the oat breeder. The objectives of the present paper were: (1) to evaluate isozyme polymorphisms in a sample of *A. sterilis* accessions from the U.S. National Small Grains Collection, (2) to analyze the distribution of isozyme diversity across the geographic range of the accessions, (3) to classify the accessions into groups based on isozyme variation, and (4) to suggest strategies for efficient sampling of this germplasm collection. One thousand and five accessions from 23 countries and 679 collection sites were screened for variation using 23 enzyme systems. Due to limited information about the genetic relationship among individual members of families of isozymes in hexaploid oat species, data were recorded solely for band presence. The frequencies of bands in accessions from the various countries were used to calculate the probability of genotypic identity ($I_{x.y}$), the probability of a unique genotype ($U_{x.y}$), and an adjusted polymorphic index (H_x). Accessions from Turkey and Lebanon had the largest polymorphic index values, Turkish and Moroccan accessions displayed the greatest numbers of bands. Accessions from Iran, Turkey, Iraq, and Lebanon had the largest mean probabilities of containing unique genotypes. Based on isozyme data, Turkey appeared to represent the center of diversity in this germplasm collection. Band frequencies calculated among countries were used in a principal component analysis. Accessions from Israel and Morocco clustered together; accessions from Iran, Iraq, Turkey, and Ethiopia formed another group; and Algerian accessions formed an outlying group. Several isozyme bands had a regional distribution. These results suggested that choosing accessions from countries based on their groupings in the principal component analysis should secure a greater range of diversity than sampling from the collection at random. Cluster analyses based on Jaccard's distances calculated for all pairwise combinations of the 1005 accessions revealed six broad genetic groups of accessions. Groups 1 and 6 contained accessions from many countries and encompassed half of all accessions. Groups 2 and 4 were heavily populated by accessions from Israel and Morocco. Groups 3 and 5 were composed almost exclusively of accessions from Iran, Iraq, and Turkey. By selecting representative accessions from these six groups, oat breeders could most effectively sample the range of genetic variation in this *A. sterilis* collection.

Key words: Genetic diversity – Genetic distance – Introgression – Core collection

Introduction

Avena sterilis L., the wild-weedy hexaploid progenitor of the cultivated oat (*A. sativa* L.) constitutes a reservoir of genetic diversity that is readily accessible to the oat breeder. Both of these highly self-pollinating species are components of a hexaploid ($2n = 42$) gene pool and hybrids between them are fertile (McMullen et al.

* *Present address:* Agricultural Science Center-North, N222, University of Kentucky, Lexington, KY 40506-00914, USA
Correspondence to: T. D. Phillips

1982). The United States National Small Grains Collection (NSGC) houses approximately 6,000 accessions of *A. sterilis* and, except for data on a few readily measurable traits such as seed protein and oil concentrations (Rezai and Frey 1988) and the agronomic and morphological studies of Rezai (1977), there is little published information quantifying or categorizing genetic diversity in this *Avena* collection. In some instances, random sampling of *A. sterilis* accessions has been used to select donor parents that might contribute additional genetic variability for traits that are difficult to evaluate on field-grown accessions of this weak-strawed, seed-shattering wild species (Murphy and Frey 1984). In addition, the content of the collection is unbalanced in that disproportionate numbers of accessions have been collected in different countries across the geographic distribution of the species. For example, though geographically small, Israel is represented by almost 4500 accessions, or 75% of the collection; Turkey, while much larger, is represented by only 350 accessions, or 5.8% of the collection. An assessment of genetic diversity within this germplasm pool must also take into consideration the confounding implications of studies on the wild tetraploid *Avena barbata* Pott ex Link that revealed associations between allozyme variation and different environmental niches. Significant allozyme variation was observed between niches separated by distances as small as meters in addition to niches separated by hundreds of kilometers (Marshall and Allard 1970; Hamrick and Holden 1979; Kahler et al. 1980).

Because the collection and presevation of crop genetic resources have become the focus of considerable attention, a seeming plethora of variation has been assembled in some species. This germplasm needs to be evaluated to ensure the optimal utilization of these resources by plant breeding programs (Goodman 1990). Isozyme variability has been used frequently to characterize germplasm collections (Brown 1978; Goodman and Stuber 1983; Price and Kahler 1983; Bretting and Goodman 1989; Bretting et al. 1990; Souza and Sorrells 1991), and it should provide additional information on genetic diversity in the *A. sterilis* collection.

The objectives of this research were to: (1) evaluate isozyme polymorphisms in a representative sample of the *A. sterilis* accessions in the NSGC; (2) analyze the stratification of isozyme diversity across the range of distribution of these accessions, (3) systematically classify the accessions into groups based on isozyme variation and compare this with other methods (Rezai 1977), and (4) suggest strategies for sampling genetic variability found in this collection.

Materials and methods

The 1005 accessions of *A. sterilis* evaluated in this study were obtained from two sources. Nine hundred and sixty accessions were provided by Dr. H. E. Bockelman, USDA, ARS, National Small Grains Collection, Aberdeen, Idaho, USA. The remaining 45 accessions were provided by Dr. K. J. Frey, Department of Agronomy, Iowa State University, Ames, Iowa, USA. The accessions were selected as representative of the range of geographical origin of accessions maintained by the NSGC, based upon information contained in The Germplasm Resources Information Network. Only accessions with passport data that included country of origin and location of collection site within that country were chosen. A total of 23 countries and 679 collection sites were represented (Table 1).

All isozyme analyses were conducted on seedlings tracing to one greenhouse-grown self-pollinated plant per accession. Details of various plant tissues, extraction buffers, gel and electrode buffers, and numerous enzyme staining protocols that were evaluated in the process of selecting the 23 most reproducible and consistent enzyme systems utilized in this study (Table 2) are contained in Phillips (1992). Preliminary studies showed that variation in isozyme phenotypes among plants tracing to a single self-pollinated plant was rare; so, for each accession, leaf blades and sheaths of three to five 14-day-old seedlings (150–250 mg) were macerated on ice with 100 µl of extraction buffer (Arulsekar and Parfitt 1986). Up to six paper

Table 1. Numbers of *A. sterilis* L. accessions, collection sites, and countries and islands of accession origin

Country	No. accessions	No. collection sites	Country	No. accessions	No. collection sites
Afghanistan	3	3	Morocco	103	70
Algeria	56	39	Pakistan	2	2
Cyprus	4	4	Portugal	4	3
Ethiopia	70	21	Sardinia	5	5
Greece	7	5	Sicily	12	12
Iran	115	91	Spain	49	34
Iraq	64	31	Syria	25	13
Israel	195	93	Tunisia	39	37
Italy	12	10	Turkey	167	160
Kenya	12	4	United Kingdom	1	1
Lebanon	39	24	Yugoslavia	2	2
Libya	19	15			
			Total	1005	679

Table 2. Enzyme systems investigated in an isozyme study of 1005 *A. sterilis* accessions

Abb.	Enzyme	Stain protocol source[a]	E.C. no.	Gel syst.[b]
ACP	Acid phosphatase	Stuber et al.	3.1.3.2	1
ACO	Aconitase	Stuber et al.	4.2.1.3	2
AK	Adenylate kinase	O'Malley et al.	2.7.4.3	2
ALD	Aldolase	Conkle et al.	4.1.2.13	1
AMP	Aminopeptidase	Stuber et al.	3.4.11.6	1
AMY	Amylase	Soltis and Soltis	3.2.1.1	2
CAT	Catalase	Conkle et al.	1.11.1.6	3
DIA	Diaphorase	Stuber et al.	1.6.2.2	3
ENP	Endopeptidase	Stuber et al.	3.4.11.-	1
EST1–5	Esterase	Stuber et al.	3.1.1.1	3
FLEST1, 2	Fluorescent esterase	Conkle et al.	3.1.1.2	2
GOT	Glutamate oxaloacetate transaminase	Stuber et al.	2.6.1.1	3
G2DH	Glycerate-2-dehydrogenase	O'Malley et al.	1.1.1.29	3
IDH	Isocitrate dehydrogenase	Stuber et al.	1.1.1.42	2
LAP	Leucine aminopeptidase	Conkle et al.	3.4.11.1	1
MDHs, f	Malate dehydrogenase	Stuber et al.	1.1.1.37	1
PER	Peroxidase	Conkle et al.	1.11.1.7	3
PGM	Phosphoglucomutase	Stuber et al.	2.7.5.1	3
PGD	6-Phosphogluconate dehydrogenase	Stuber et al.	1.1.1.44	1
PHI	Phosphohexose isomerase	Stuber et al.	5.3.1.9	2
SAD	Shikimate dehydrogenase	Stuber et al.	1.1.1.25	1
SOD	Superoxide dismutase	O'Malley et al.	1.15.1.1	3
TPI	Triose phosphate isomerase	Stuber et al.	5.3.1.1	3

[a] References: Conkle et al. (1982); O'Malley et al. (1980); Soltis and Soltis (1989); Stuber et al. (1988)
[b] Gel systems: 1, Morpholine-citrate, pH 6.1; 2, Tris-citrate, pH 6.2; 3, Tris-citrate, pH 8.0

wicks (15 × 3 mm, Whatman chromatography paper, no. 3MM) were saturated with extract and then used to load extract into gels precooled for 1 h in a 4 °C refrigerated cabinet. The gels (16 × 18 × 2 cm) were made of 10% hydrolyzed potato starch (w: volume of buffer) from Sigma Chemical Company (P.O. Box 14058, St. Louis, Mo. 63178, Cat. #S4501). After 15–20 min the loading wicks were removed from the gels. Constant voltage was applied to the gel systems as follows: the T8.0 gel ran at 280 V, T6.2 at 220 V, and M6.1 at 240 V. When front migration had progressed 10 cm from the origin, the gels were cut horizontally to yield six assayable slices per gel. Duplicate gels for each assay were run simultaneously to aid in scoring isozyme phenotypes. Enzyme activity staining protocols (Table 2) were adapted from O'Malley et al. (1980), Stuber et al. (1988), Soltis and Soltis (1989), and Conkle et al. (1982).

Because published information on the inheritance of isozyme polymorphism in hexaploid oat species was limited to esterase (Clegg and Allard 1972), diaphorase (Souza and Sorrells 1989) and peroxidase (Yen and Sadanaga 1977) when this study was initiated, and because our survey of 1005 accessions revealed band patterns additional to those published for esterase and diaphorase, isozyme phenotypes were recorded and assumed to be analogous to homozygous genotypes. Although only 23 enzyme systems were stained, three had more than one distinct zone of band activity (Table 2) based on migration distances (Kahler et al. 1980). Esterase gels were divided into five zones that approximated loci described for the hexaploid weedy *A. fatua* (Clegg and Allard 1972). In addition, fluorescent esterase and malate dehydrogenase had two zones each, giving a total of 29 zones of band activity across all 23 enzyme systems. Band patterns observed for each zone were diagrammed in Phillips (1992).

For each accession, an isozyme band presence/absence profile was recorded, consisting of a series of 134 designations of either 1s (presence) or 0s (absence) describing the two alternate states at each of the 134 positions at which scorable bands were observed over all 29 zones. For each enzyme zone, bands were numbered consecutively relative to migration distance – i.e., position '1' was nearest to the origin. Based upon the numbers of bands per zone, the scant available information on the genetics of isozymes in hexaploid oat, and the numbers of loci controlling activity in more thoroughly characterized species [primarily maize (*Zea mays* L., Stuber et al. (1988)], the 29 zones were divided into three groups based upon the relative number of loci possibly involved in expression of isozyme genotypes: putative one 'locus' (ESTI-5, G2DH, FLEST2, and AMY); putative two 'loci' (SOD, SAD, ACP, ALD, AK, CAT, ENP, FLESTI, TPI, PGM, AMP, DIA, GOT, IDH, and PER); and putative three 'loci' (LAP, PHI, ACO, MDHs, MDHf, and PGD).

For each locus class, Jaccard's distances (Sneath and Sokal 1973) were estimated between all pairs of the 1005 accessions using the following formula:

$$JD(I, J) = 1 - \frac{M}{(T_I + T_J) - M}$$

where M was the number of band matches between accessions I and J, T_I was the total number of bands in accession I, and T_J was the total number of bands in accession J. Standardization of Jaccard's distances between pairs of accessions was accomplished by calculating the distances based on each of the three putative locus groups separately and dividing by the standard deviation of the distances within the appropriate putative locus group. The three standardized distances were weighted according to the purported portion of genetic information contained in each distance estimate. Eight zones were assumed to be controlled by a single 'locus', 15 zones by two 'loci', and six zones by three 'loci'. The combined Jacard's distances between pairs of

accessions were calculated using these weightings (8:30:18) for the three separate estimates and the appropriate standard deviations. Because these distances had been adjusted for variances and the number of enzyme bands, they did not fall in the range of 0 to 1. To remedy this, the distances were ranged by dividing by the maximum value of all distances.

A clustering procedure [unweighted pair-group-method algorithm (Sneath and Sokal 1973)] was employed to produce a dendrograph of the 1005 accessions of *A. sterilis* based upon these standardized, weighted, and ranged Jaccard's distances. This dendograph was pruned arbitrarily at a level of between-group correlation of 0.40 to reveal 50 clusters of accessions. This level of pruning identified distinct clusters of accessions with between three and 53 members. Other levels of pruning produced clusters with excessively large or small numbers of accessions. Frequencies for each of the 134 possible bands were obtained for these 50 clusters and for the 11 countries each represented by more than 24 accessions.

Subsequently, these frequency data were utilized to estimate Hedrick's (1971, 1985) probability of genetic identity ($I_{x,y}$) between each of the 50 clusters of accessions and between the countries of accession origin. The probability of genetic identity was calculated as follows:

$$I_{x,y} = \frac{1}{B}\sum_{i=1}^{B}\left[\left(2\sum_{j=1}^{g_i} P_{ijx}P_{ijy}\right)\Big/\left(\sum_{j=1}^{g_i} P_{ijx}^2\right)\left(\sum_{j=1}^{g_i} P_{ijy}^2\right)\right]$$

where B was the number of band positions, P_{ijx} was the frequency of band j at position i in group x, and g_i was the number of genotypes (1 if the band was omni-present or never present in a group; 2 if both genotypes – presence and absence of the band – were observed) at position i. Genetic distances were obtained by taking the complement of the $I_{x,y}$ values and were utilized to construct a dendrograph of the 50 clusters of accessions using an unweighted pair-group algorithm. This analysis revealed six clusters of accessions (hereafter referred to as 'broad genetic groups') each containing between 69 and 502 members.

A second measure of among-population diversity, the probability of a unique genotype, focused on the presence or absence of specific genotypes averaged over all loci (Hedrick 1971). This measure, U, had two components:

$$U_{x,y} = \frac{1}{B}\sum_{i=1}^{B}\sum_{j=1}^{g_i}(P_{ijx}|P_{ijy} = 0)$$

and

$$U_{y,x} = \frac{1}{B}\sum_{i=1}^{B}\sum_{j=1}^{g_i}(P_{ijy}|P_{ijx} = 0)$$

The first component, $U_{x,y}$, was the mean probability of a band occurring in group x but not in y, while $U_{y,x}$ was the probability of a band occurring in group y but not in x.

Nei and Roychoudhury's (1974) adjusted polymorphic index was computed for each enzyme band in the countries of origin and the six broad genetic groups. The adjusted polymorphic index, or gene-diversity index, for band position i in group x, H_{ix}, was computed as follows:

$$H_{ix} = [n_x/(n_x - 1)]\left(1 - \sum_{j=1}^{g_i} P_{ijx}^2\right)$$

where n_x was the number of accessions in group x, P_{ijx} was the frequency of band j at position i in group x, and g_i was the number of genotypes at position i in group x. The mean gene diversity value over B band positions was computed as:

$$H_x = (1/B)\left(\sum_{i=1}^{B} H_{ix}\right)$$

where B was the number of band positions. The gene diversity value described intrapopulation diversity and was influenced both by the number of bands and their frequencies within the groups of accessions. A FORTRAN program for computing Jaccard's distance, Hedrick's probability of genotypic identity, the probability of a unique genotype, and Nei and Roychoudhury's adjusted polymorphic index, constructed by Drs. J. P. Murphy (North Carolina State University) and T. S. Cox (USDA-ARS, Kansas State University), was combined with parts of programs written by D. M. Rodgers (Murphy et al. 1986) and McCammon and Wenninger (1970) to estimate these statistics and to construct dendrographs (Lubbers et al. 1991). Principal component analyses based on covariance matrices of band frequencies were conducted for countries of accession origin and the six putative broad genetic groups (SAS 1985).

Results and discussion

The 29 isoenzymatic zones identified using 23 enzyme systems across 1005 accessions of *A. sterilis* together had bands at 134 positions ranging from one band either present or absent per zone (EST3, EST4, EST5), to variations in the presence of up to nine bands per zone (MDHf, MDHs) (Table 3). Nine of the one hun-

Table 3. Enzyme zones, numbers of bands observed per enzyme zone, and frequency of occurrence of bands in all 1005 accessions

Enzyme zone	No. bands	No. bands observed in			
		1–250 acc.	251–500 acc.	501–750 acc.	751–1005 acc.
ACO	8	4			4
ACP	5	3		1	1
AK	5	2	1		2
ALD	5	3			2
AMP	6	3		1	2
AMY	4	2	2		
CAT	5	3			2
DIA	4	1			3
ENP	5	3		1	1
EST-1	3	2		1	
-2	3	2		1	
-3	1		1		
-4	1		1		
-5	1		1		
FLEST-1	5	3	1		1
-2	3	2		1	
G2DH	2		1	1	
GOT	6	2			4
IDH	4	1			3
LAP	7	4		2	1
MDHs	9	5			4
MDHf	9	6			3
PER	2	1			1
PGD	6	3	2		1
PGM	5	2		2	1
PHI	8	5	1	1	1
SAD	4	1		1	2
SOD	3	1			2
TPI	5	3			2
Sum	134	67	11	13	43

dred and thirty four possible bands were present in all accessions, 56 bands were present in at least half of the accessions, and 78 bands were present in less than half the accessions. The considerable diversity that could be found by an extensive search through this sample of accessions was underscored by the fact that 54 of the bands occurred in 10% or less of all accessions. Furthermore, 20 bands occurred in 10 or fewer accessions (< 1%). Using Brown's (1978) classification system, 60% of bands were regarded as common (i.e., with a frequency > 10% in at least one country) and 40% were regarded as rare (i.e., never occurred at a frequency > 10% in any country). Forty-four percent were classified as both common and wide-spread (i.e., common in more than two countries). Thirty percent were classified as rare but widespread (i.e., rare, but found in more than one country), and 10% were classified as rare and localized (i.e., rare and only found in one country). These data suggested that over 70% of the isozymic variation contained in this sample of accessions could be uncovered by intense sampling of a large number of accessions from a few carefully chosen countries; however, the remainder would be difficult to uncover, particularly the 10% in the rare and localized category, without a much larger sampling of both accessions and countries. These data also forewarn of the ease with which many rare types could be lost through attrition of accessions, or through pursuing a policy such as maintenance of this collection by the propagation of accessions in bulk populations.

Stratification by accession country of origin

To preclude artifactual results from insufficient sample sizes, this analysis was based on band frequencies in the 11 countries represented by 25 or more accessions, retaining 925 of the 1005 accessions (Table 1). Nei's adjusted polymorphic index over the 925 accessions was 0.148 (Table 4). This value was close to the mean (0.141) within-population heterozygosity reported for 35 populations of self-pollinated species by Bretting and Goodman (1989).

Because accessions originating in Israel comprise 75% of the NSGC *A. sterilis* collection, it is worthwhile

Table 4. Isozyme bands which were heavily weighted in a principal component analysis of band frequencies for *A. sterilis* accessions from 11 countries, the component(s) for which bands were weighted heavily, band frequencies, numbers of accessions, numbers of bands, and Nei and Roychoudhury's adjusted polymorphic index

Band	Principal component(s) with eigenvector coefficient > 0.10 or < −0.10	Frequency of band in:											
		All accessions	Iran	Iraq	Turk	Eth	Alg	Syr	Tun	Moro	Iber	Isr	Leb
ACP2	3	0.53	0.66	0.53	0.43	0.61	0.91	0.56	0.31	0.46	0.58	0.44	0.72
AK5	3	0.33	0.13	0.41	0.26	0.16	0.09	0.08	0.56	0.35	0.56	0.50	0.39
ALD1	2	0.05	0.00	0.00	0.00	0.00	0.02	0.24	0.00	0.13	0.00	0.07	0.31
ALD2	2	0.95	1.00	1.00	1.00	1.00	0.98	0.76	1.00	0.86	1.00	0.93	0.69
ALD4	2	0.95	1.00	1.00	1.00	1.00	0.98	0.76	1.00	0.86	1.00	0.93	0.69
AMP2	1, 3	0.73	1.00	0.98	0.81	1.00	0.48	0.48	0.54	0.75	0.48	0.61	0.46
AMP3	1	0.18	0.00	0.00	0.14	0.00	0.21	0.16	0.41	0.23	0.19	0.30	0.41
AMY1	1, 3	0.34	0.61	0.39	0.59	0.26	0.77	0.12	0.21	0.09	0.14	0.06	0.41
AMY2	3	0.33	0.26	0.36	0.29	0.34	0.21	0.00	0.46	0.30	0.58	0.34	0.39
AMY3	1, 2	0.23	0.10	0.14	0.12	0.23	0.02	0.60	0.26	0.43	0.17	0.40	0.21
DIA2	2	0.85	0.92	0.94	0.93	0.44	0.64	0.92	0.90	0.96	0.81	0.91	0.85
ENP4	3	0.73	0.95	0.72	0.78	0.76	0.91	0.64	0.56	0.50	0.65	0.72	0.74
EST1-3	3	0.75	0.59	0.72	0.76	0.97	0.91	0.80	0.77	0.78	0.50	0.76	0.77
EST2-3	2	0.18	0.14	0.25	0.19	0.06	0.02	0.40	0.21	0.10	0.10	0.27	0.31
FLEST1–3	3	0.30	0.32	0.44	0.25	0.21	0.16	0.24	0.28	0.46	0.42	0.25	0.28
FLEST2–3	2, 3	0.61	0.78	0.34	0.62	0.53	0.04	0.76	0.36	0.88	0.75	0.67	0.67
IDH6	1	0.19	0.33	0.34	0.27	0.40	0.02	0.24	0.26	0.00	0.00	0.07	0.10
LAP1	1, 3	0.21	0.48	0.41	0.43	0.19	0.05	0.16	0.23	0.05	0.02	0.02	0.03
LAP3	1, 2, 3	0.66	0.37	0.47	0.43	0.60	0.79	0.44	0.82	0.83	0.87	0.90	0.67
LAP5	1, 2	0.74	1.00	0.98	0.80	1.00	0.32	0.92	0.51	0.76	0.48	0.61	0.54
MDHf3	1	0.07	0.28	0.27	0.07	0.07	0.04	0.00	0.03	0.02	0.02	0.00	0.00
MDHf4	1	0.91	0.62	0.80	0.82	0.97	1.00	0.96	0.97	0.99	0.98	1.00	1.00
MDHf6	1	0.25	0.92	0.96	0.36	0.09	0.09	0.04	0.05	0.03	0.02	0.00	0.05
PER2	2	0.17	0.04	0.06	0.19	0.04	0.43	0.08	0.05	0.19	0.33	0.26	0.10
PGD1	3	0.42	0.52	0.38	0.48	0.44	0.25	0.36	0.33	0.46	0.50	0.41	0.28
PGD3	3	0.44	0.52	0.38	0.51	0.44	0.27	0.36	0.41	0.48	0.50	0.42	0.28
PGM5	3	0.56	0.33	0.55	0.54	0.66	0.25	0.44	0.59	0.61	0.71	0.68	0.59
SAD1	1, 2, 3	0.71	0.84	0.75	0.82	0.89	0.80	0.00	0.74	0.51	0.98	0.73	0.21
SAD4	1	0.90	0.72	0.86	0.71	0.93	0.96	1.00	0.95	0.97	1.00	0.99	1.00
No. accessions		997	115	64	167	70	56	25	39	103	52	195	39
No. bands		134	107	102	116	104	100	99	105	111	101	109	103
Polymorphic index		.148	.135	.146	.152	.106	.102	.139	.148	.136	.135	.128	.151

Table 5. Hedrick's probability of a unique genotype, averaged over 134 isozyme bands, for combinations of the 11 countries each represented by more than 24 accessions of *A. sterilis*

Mean probability of drawing a genotype from country	That does not occur in country:											Avg. for all countries
	Alg	Eth	Iran	Iraq	Leb	Syr	Tun	Turk	Mor	Isr	Iber	
Algeria	0.000	0.016	0.014	0.006	0.003	0.013	0.002	0.001	0.003	0.003	0.004	0.006
Ethiopia	0.003	0.000	0.001	0.001	0.005	0.013	0.003	0.002	0.004	0.002	0.006	0.004
Iran	0.012	0.006	0.000	0.005	0.021	0.024	0.010	0.003	0.011	0.023	0.017	0.013
Iraq	0.010	0.005	0.002	0.000	0.012	0.018	0.009	0.002	0.008	0.014	0.013	0.009
Lebanon	0.004	0.025	0.021	0.016	0.000	0.011	0.012	0.010	0.002	0.001	0.012	0.011
Syria	0.007	0.015	0.019	0.011	0.005	0.000	0.012	0.011	0.004	0.003	0.012	0.010
Tunisia	0.005	0.019	0.017	0.010	0.005	0.016	0.000	0.001	0.007	0.003	0.006	0.009
Turkey	0.011	0.010	0.008	0.007	0.012	0.021	0.008	0.000	0.008	0.010	0.012	0.011
Morocco	0.009	0.016	0.016	0.013	0.007	0.011	0.010	0.009	0.000	0.003	0.009	0.010
Israel	0.006	0.015	0.014	0.008	0.005	0.014	0.005	0.005	0.003	0.000	0.005	0.008
Iberia	0.005	0.014	0.014	0.007	0.005	0.017	0.003	0.003	0.003	0.002	0.000	0.007
Any other country	0.007	0.014	0.013	0.008	0.008	0.016	0.007	0.005	0.005	0.006	0.010	

to utilize Israel as a benchmark for comparisons with other countries. Among the countries or regions represented in this study by the largest numbers of accessions, within-country diversity was greatest for Turkey, followed in decreasing order by Lebanon, Tunisia, Iraq, Syria, Morocco, Iran, Iberia (Spain and Portugal combined), and Israel. Joint consideration of the number of representative accessions, the number of bands observed, and the polymorphic indices, indicated that accessions originating in several countries – including Iran, Iraq, Turkey, Morocco, Spain and Portugal, and perhaps Tunisia and Lebanon – would be as fruitful a source of isozyme variation in this collection as accessions from Israel.

The range in $I_{x,y}$, Hedrick's probability of genotypic identity, between pairs of these 11 countries was 0.942 to 0.990 (data not shown). The country or region pairs that had the most similar genotypic arrays were Israel-Morocco, Israel-Tunisia, Israel-Iberia, and Turkey-Iraq. Turkey had the highest overall genotypic identities with other countries (range 0.973–0.986), reflecting its high level of isozyme diversity.

The mean probability of discovering a unique genotype ($U_{x,y}$) in country x, not represented in country y, was largest (Table 5) for accessions from Iran (0.013), Turkey (0.011), and Lebanon (0.011). The converse ($U_{y,x}$), the probability of lacking a genotype found in another country, was smallest for Turkey (0.005), Morocco (0.005), Israel (0.006), Tunisia (0.007), and Algeria (0.007). Iranian accessions contained the highest frequency of unique genotypes relative to Israeli accessions, and the probability of finding a unique variant in Turkish accessions, not found in Israeli accessions, was twice the reciprocal probability. The ideal

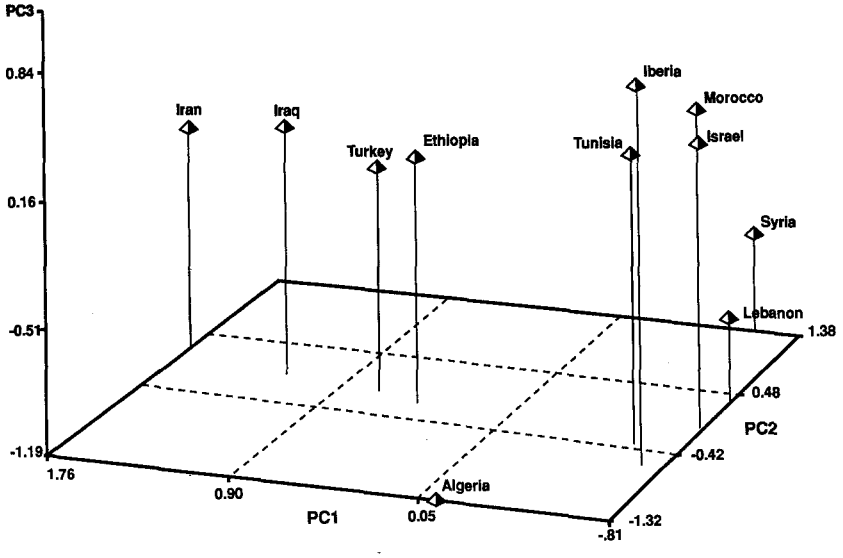


Fig. 1. Plot of the first three principal component scores for 11 countries represented by more than 24 accessions of *A. sterilis*, based on the covariance matrix of isozyme band frequencies among 11 countries. The first component accounted for 34% of the total variance; the second and third accounted for 18 and 15%, respectively

from the perspective of a plant breeder wishing to screen a germplasm collection is to identify a country or region with a high mean $U_{x.y}$ and a low mean $U_{y.x}$ as these accessions would contain the most comprehensive sample of the diversity in the germplasm collection. For this data set, Turkey best fits these criteria.

In his analysis of morphological and agronomic traits in 457 accessions of *A. sterilis*, Rezai (1977) concluded that Israel was the center of diversity for accessions of the species in the NSGC. Nevertheless, he found that accessions from Iran, Iraq, and Turkey had the most variation for the vegetative traits measured. In addition, Turkey had the greatest within-country variance components for five of the fifteen traits measured. Approximately one-half of Rezai's accessions were examined in this study; joint consideration of his findings and our data on isozyme variation suggested that Turkish accessions may also represent a center of diversity in this germplasm collection.

Clustering of countries based upon principal component analysis (Fig. 1) reflected relationships previously mentioned for genotypic identities ($I_{x.y}$). Turkey, Iraq, Iran, and Ethiopia were separated from the seven other countries on the first principal component which accounted for 34% of the variation in band frequencies. Israel, Morocco, Tunisia, and Iberia clustered together, as did Lebanon and Syria. Algeria was the outlier in this analysis. It was somewhat unexpected that clustering would occur between countries at the eastern and western end of the Mediterranean basin, yet Rezai (1977) also reported clustering of accessions from Israel and Morocco. Rezai and Frey (1988) subsequently defined the eastern and western Mediterranean as two distinct regions pertaining to panicle and seed characteristics when Israeli accessions were excluded from the analysis. The distinction was less evident when Israeli accessions were included, because the means of four of the six distinguishing traits were almost identical for Israeli and Moroccan accessions.

Sixteen enzyme systems with a total of 29 bands were heavily weighted in the eigenvectors of the first three principal components (Table 4). These 29 bands exhibited significant among-country variation (data not shown). A considerable portion of the genome probably contributed to the differentiation among countries, as 55% of the enzyme systems studied, 63% of putative 'loci', and 22% of the total bands observed, are included in Table 4. Examination of patterns of distribution of the 29 bands identified as important in the principal component analysis aided interpretation of the clustering in Fig. 1. Iran, Iraq, and Turkey had band frequency differences of less than 0.21 for ten of the thirteen most important bands in the first principal component. Israel, Lebanon, and Syria had similar frequencies for six of the thirteen bands. Only five of the thirteen bands had similar frequencies over accessions originating in Algeria, Morocco, Tunisia, and Iberia. The observed tendency for Israeli and Moroccan accessions to cluster can be explained by their similarity in frequencies for 12 of the 13 important bands. Regional distribution of certain bands was observed for four of the six enzyme zones which were influential in separating the 11 countries. Both AMP2 and AMP3 were at extreme frequencies in accessions from Iran, Iraq, Turkey, and Ethiopia; other regions had these bands at intermediate frequencies. Three LAP bands (1, 3, and 5), and two MDHf bands (3 and 6) also contributed to the separation of some or all of the Iran-Iraq-Turkey accessions from other countries. Band IDH6 was more common in the western range of the species (Iran, Iraq, Turkey, and Ethiopia) than in the eastern range. Band AMY3 was more common in Israel, Syria, and Morocco than in other countries. These data seemed to indicate that isozyme variation was not randomly distributed across the geographic range of the species. Regional partitioning of several isozyme bands probably reflected correlations among particular isozyme genotypes and certain environmental features.

These data indicated that random sampling of accessions in the NSGC would be an inefficient approach if the goal was to maximize isozyme variation in the resulting sample. Such a strategy would result in a sample comprised of 75% Israeli accessions. While the Israeli accessions contain substantial variation based on the stratification revealed in this analysis, a more efficient approach would be to choose equal numbers of accessions from the countries with: (1) large polymorphic indices (Table 4), (2) high probabilities of containing unique bands, (3) low probabilities of lacking bands (Table 5), and (4) from countries representing the clusters revealed in Fig. 1. On a macrogeographic scale, Turkey, Iran, Iraq, and Ethiopia formed one cluster of countries whereas the Mediterranean basin – particularly Israel, Morocco, Tunisia, and Iberia – formed a second cluster. Algerian accessions appeared to form a third cluster.

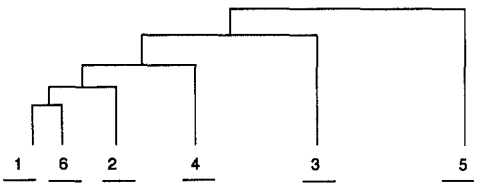
While selection of accessions from several countries would be most appropriate, our data illustrated that Turkey was a possible center of diversity for isozyme variation in this collection. It is noteworthy that Turkey was represented by 160 collection sites, the largest for any country in this study; and this may have been very significant given the amply demonstrated association between ecological and isozyme variation in *Avena* and other species (Hamrick and Holden 1979; Nevo et al. 1979; Kahler et al. 1980; Nevo and Beiles 1989; Souza and Sorrells 1989). The detailed information (e.g., latitude, longitude, altitude, minimum and maximum temperatures, moisture, etc.) necessary for an analysis of potential ecological and isozyme variation association was unavailable for these accessions.

Nevertheless, it would be important for any sampling strategy to maximize the number of collection sites within countries whenever that information is available.

Stratification of accessions disregarding provenance information

By defining accessions from a country as the initial population, one can inadvertently consolidate broad geographical areas that may contain dissimilar environments and widely differentiated populations. Alternatively, clustering that combines accessions from disparate countries may be based on similarities in environmental features at collection sites, assuming that some isozyme variability is associated with adaptive value. Rezai (1977) found large within-country variation for several traits in *A. sterilis*, and his analysis revealed ten groups of accessions, many of which contained accessions from several countries.

The dendrograph based on Jaccard's distances between all pairs of 1005 accessions revealed 50 clusters of genetically similar accessions when pruned at 0.40.



	1	6	2	4	3	5
Algeria (56)	48	0	6	0	1	1
Ethiopia (70)	61	3	3	2	0	1
Iberia (52)	31	0	8	13	0	0
Iran (115)	35	13	0	1	26	40
Iraq (64)	30	4	3	0	15	12
Israel (195)	87	10	58	35	5	0
Lebanon (39)	8	12	9	7	3	0
Morocco (103)	61	10	22	10	0	0
Syria (25)	10	2	4	7	2	0
Tunisia (39)	16	2	14	6	1	0
Turkey (167)	72	31	18	12	15	19
Others (80)	43	10	11	13	1	2
Total (1005)	502	97	156	106	69	75

Fig. 2. Dendrograph based on a matrix of genetic distances ($1 - I_{x,y}$) among six broad genetic groups of *A. sterilis* accessions, with total numbers of accessions in each group and numbers from each country in each group

Table 6. Isozyme bands which were heavily weighted in a principal component analysis of band frequencies for six broad genetic groups of *A. sterilis*, the component(s) for which the bands were weighted heavily, band frequencies, numbers of bands, and Nei and Roychoudhury's adjusted polymorphic index

Band	Principal component(s) with eigenvector coefficient > 0.10 or < -0.10	Frequency in:						
		All accessions	Genetic group					
			1	2	3	4	5	6
ACP2	1	0.53	0.61	0.32	0.28	0.45	0.84	0.50
AK4	2	0.89	0.92	0.96	0.41	0.97	0.93	0.86
AMP2	1, 3	0.73	0.85	0.25	0.96	0.49	1.00	0.78
AMP3	1, 3	0.18	0.05	0.65	0.00	0.36	0.00	0.18
AMY1	1	0.33	0.30	0.19	0.65	0.18	0.64	0.45
ENP3	1, 2, 3	0.13	0.03	0.01	0.01	1.00	0.03	0.03
ENP4	1, 3	0.73	0.84	0.74	0.77	0.00	0.87	0.76
ENP5	1, 2, 3	0.87	0.97	0.99	0.99	0.00	0.97	0.97
G2DH1	2	0.46	0.50	0.46	0.10	0.58	0.60	0.28
G2DH2	2	0.54	0.50	0.54	0.90	0.43	0.40	0.72
IDH6	2, 3	0.19	0.19	0.05	0.74	0.09	0.17	0.12
LAP1	1, 2	0.21	0.17	0.03	0.06	0.05	0.93	0.40
LAP3	1	0.65	0.68	0.85	0.73	0.84	0.12	0.37
LAP5	1, 3	0.74	0.82	0.33	0.97	0.58	0.97	0.83
MDHf2	1	0.05	0.00	0.00	0.00	0.00	0.52	0.06
MDHf4	1, 2	0.91	0.96	0.99	0.99	1.00	0.11	0.92
MDHf6	1, 3	0.25	0.16	0.03	0.83	0.07	0.97	0.28
MDHf8	1, 2	0.93	0.98	0.99	1.00	1.00	0.27	0.93
PGD3	2	0.44	0.44	0.51	0.68	0.39	0.28	0.32
PHI5	2, 3	0.63	0.65	0.74	0.15	0.57	0.81	0.57
PGM5	2	0.55	0.58	0.66	0.25	0.69	0.44	0.42
SAD2	2	0.07	0.03	0.01	0.45	0.06	0.01	0.14
SAD4	1	0.89	0.92	0.99	0.97	0.98	0.41	0.83
No. accessions		1005	502	156	69	106	75	97
No. bands		134	129	114	107	106	101	119
Polymorphic index		.148	.132	.128	.117	.126	.126	.166

Table 7. Hedrick's probability of a unique genotype based on presence/absence of 134 isozyme bands, for all combinations of six broad genetic groups of *A. sterilis*

Mean probability of drawing a genotype from genetic group	That does not occur in genetic group:						Avg. for all groups
	1	2	3	4	5	6	
1	0.000	0.001	0.005	0.023	0.007	0.002	0.008
2	0.000	0.000	0.011	0.022	0.017	0.001	0.010
3	0.000	0.002	0.000	0.025	0.006	0.003	0.007
4	0.000	0.001	0.007	0.000	0.013	0.001	0.004
5	0.001	0.008	0.019	0.044	0.000	0.005	0.015
6	0.002	0.005	0.010	0.027	0.018	0.000	0.012
Any other group	0.001	0.003	0.010	0.028	0.012	0.002	

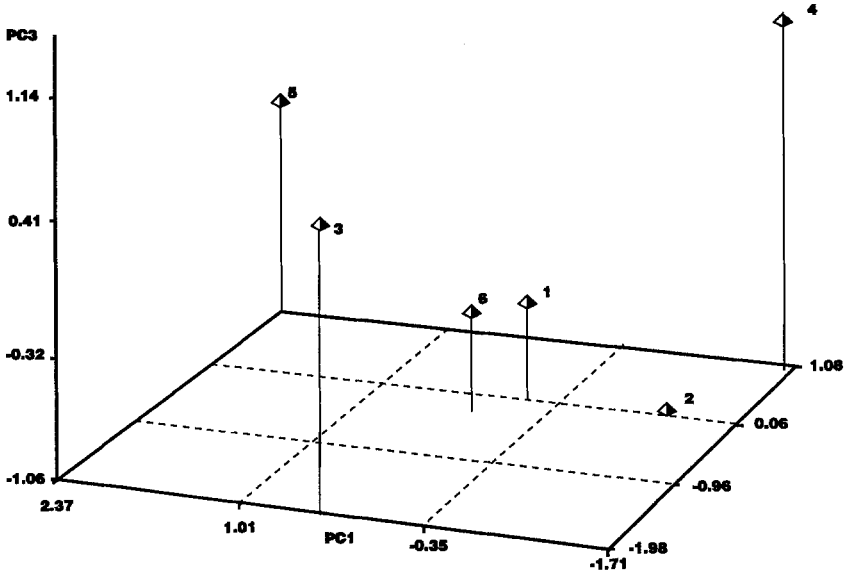


Fig. 3. Plot of the first three principal component scores for six broad genetic groups of *A. sterilis*, based on the covariance matrix of isozyme band frequencies among six groups. The first component accounted for 46% of the total variance; the second and third accounted for 28 and 15%, respectively

The mean number of accessions per cluster was 20 and ranged between three and 53 accessions (Phillips 1992). To summarize the information in this dendrograph, an additional cluster analysis based upon the complement of the mean probability of genetic identity ($1 - I_{x,y}$) between these 50 clusters revealed five broad groups of accessions containing a total of 908 accessions (Fig. 2). The remaining 97 accessions not included in any of the five clusters were assigned to a sixth group.

Genetic group 1 contained half of the accessions in the study and 129 of the 134 bands (Table 6). It included over 80% of the Ethiopian and Algerian accessions and over 40% of the Israeli, Moroccan, Iraqi, Iberian, and Turkish accessions (Fig. 2). It could justifiably be classified as a 'broad' genetic group based upon its low mean $U_{y.x}$ (Table 7). Group 1 probably should be the source of a considerable portion of a stratified sampling of this germplasm collection based upon isozyme variation. Each of the other five groups also should be sampled; although they may not be as inclusive as group 1 in overall genetic content, their divergent frequency arrays may reflect valuable adaptational variation. Group 6 was comprised of only 97 accessions and contained the second largest number of bands (Table 6). Its large polymorphic index reflected its contingent of accessions that failed to cluster with any of the other five groups. It also contained unique bands not found in group 1 and had a low mean $U_{y.x}$. Group 6 contained seven of the 12 Kenyan accessions, 12 of the 39 Lebanese accessions, and 32 of the 167 Turkish accessions (Fig. 2).

The principal component plot based upon genotypic frequencies in each of the six groups illustrated the divergence between groups 3 and 5 versus groups 2 and 4 (Fig. 3). Although this reflected the previous separation of country clusters Turkey-Iraq-Iran from Israel-Western Mediterranean, the inherent deficiency of a simple grouping of accessions by country of origin was

demonstrated by the divergences between groups 3 and 5 between groups 2 and 4. Group 4 – containing 18% of the Israeli, 10% of the Moroccan, and 25% of the Iberian accessions – is notable for having the highest mean $U_{y.x}$ of all the groups (Table 7). It is separated from group 2, also dominated by Israeli and Western Mediterranean accessions, on the basis of differences for the ENP and AMP enzyme systems (Table 6). In contrast, groups 3 and 5 – which are distinctly composed of accessions from Turkey, Iraq, and Iran – are separated by differences for nine of the enzyme systems (Table 6). Groups 1 and 6 plot near each other, reflecting the larger degree of similarity in band frequencies between these groups.

Devising a strategy for sampling A. sterilis

One of the goals of this study was to devise a procedure to sample the genetic variation contained in the NSGC of *A. sterilis*. A systematic, representative sample of the 1005 *A. sterilis* accessions was identified by selecting 24 equidistantly spaced accessions across each of the six broad genetic groups identified in the cluster analysis. These 144 accessions contained 85% of all band patterns observed in the 29 enzyme zones over all 1005 *A. sterilis* accessions. An additional 28 accessions, identified as possessing the remaining rare band patterns lacking in the initial set of 144 accessions, were added to give a set of 172 accessions that comprehensively represented the range of isozyme diversity found in this study. This collection is not the only 'core collection' which could have been constructed from these 1005 accessions; however, it does contain all isozyme bands and the accessions come from throughout the geographic range of the species.

The six broad genetic groups of *A. sterilis* determined by isozyme analysis in this study, together with Rezai's (1977) clusters based on morphological and agronomic data, can serve as sources of parental material in oat breeding studies. By selecting accessions from each cluster, some clusters may be identified as superior contributors of useful variation to cultivated oat when compared to the collection as a whole. This information could then be used to determine additional sources (accessions) to exploit. In a similar manner, perhaps new collecting expeditions would be warranted if certain regions are identified as being rich in useful alleles.

Finally, it should be re-emphasized that this data set was based upon evaluation of progeny from a single self-pollinated plant per accession. This sampling procedure was used in order to maximize the numbers of accessions and collection sites investigated. Within-accession phenotypic variation for seed characteristics was sometimes observed among seeds received from the NSGC. The relative amount of isozyme variation within and between accessions warrants further investigation. A list of the 172 'core' accessions and a list of all 1005 accessions evaluated and the putative broad genetic group to which they belong are available from the senior author.

Acknowledgements. We acknowledge the technical assistance provided by Drs. Ed Souza and Mark Sorrells at the initiation of this work. The research reported in this publication was funded by the North Carolina Agricultural Research Service, Raleigh, North Carolina and the Quaker Oats Company, Chicago, Illinois, USA. The use of trade names in this publication does not imply endorsement by the N. C. Agricultural Research Service nor criticism of similar ones not mentioned.

References

Arulsekar S, Parfitt DE (1986) Isozyme analysis procedures for stone fruits, almonds, grape, walnut, pistachio, and fig. Hort Sci 21:928–933

Bretting PK, Goodman MM (1989) Genetic variation in crop plants and management of germplasm collections. In: Stalker HT, Chapman C (eds) IBPGR training course: Lecture series 2. Scientific management of germplasm: characterization, evaluation and enhancement. International Board for Plant Genetic Resources, Rome, pp 41–54

Bretting PK, Goodman MM, Stuber CW (1990) Isozymatic variation in Guatemalan races of maize. Am J Bot 77:211–225

Brown AHD (1978) Isozymes, plant population genetic structure and genetic conservation. Theor Appl Genet 52:145–157

Clegg MT, Allard RW (1972) The genetics of electrophoretic variants in *Avena*. II. The esterase E_1, E_2, E_4, E_5, E_6 and anodal peroxidase APX_4 loci in *A. fatua*. J Hered 64:3–6

Conkle MT, Hodgskiss PD, Nunnally LB, Hunter SC (1982) Starch gel electrophoresis of conifer seeds: a laboratory manual. Gen Tech Rept PSW-64, Pacific Southwest Forest and Range Experiment Station, US For Serv, Berkeley, California

Goodman MM (1990) Genetic and germ plasm stocks worth conserving. J Hered 81:11–16

Goodman MM, Stuber CW (1983) Races of maize. VI. Isozyme variation among races of maize in Bolivia. Maydica 28:169–187

Hamrick JL, Holden LR (1979) Influence of microhabitat heterogeneity on gene frequency distribution and gametic phase disequilibrium in *Avena barbata*. Evolution 33:521–533

Hedrick PW (1971) A new approach to measuring genetic similarity. Evolution 25:276–280

Hedrick PW (1985) Genetics of populations. Jones and Bartlett, Boston

Kahler AL, Allard RW, Krzakowa M, Wehrhahn CF, Nevo E (1980) Associations between isozyme phenotypes and environment in the slender wild oat (*Avena barbata*) in Israel. Theor Appl Genet 56:31–47

Lubbers EL, Gill KS, Cox TS, Gill BS (1991) Variation of molecular markers among geographically diverse accessions of *Triticum tauchii*. Genome 34:354–361

Marshall DR, Allard RW (1970) Isozyme polymorphisms in natural populations of *Avena fatua* and *A. barbata*. Heredity 25:373–382

McCammon RB, Wenninger G (1970) The dendrograph. Kansas State Geol Surv Computer Contrib 48:1–27

McMullen MS, Phillips RL, Stuthman DD (1982) Meiotic irregularities in *Avena sativa* L. × *A. sterilis* L. hybrids and breeding implications. Crop Sci 22:890–892

Murphy JP, Cox TS, Rodgers DM (1986) Cluster analysis of red winter wheat cultivars based upon coefficients of parentage. Crop Sci 26:672–676

Murphy JP, Frey KJ (1984) Comparisons of oat populations developed by interspecific hybridization. Crop Sci 24:531–536

Nei M, Roychoudhury AK (1974) Sampling variance of heterozygosity and genetic distance. Genetics 76:379–390

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, Zohary D, Brown AHD, Haber M (1979) Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum*, in Israel. Evolution 33:815–833

O'Malley D, Wheeler NC, Guries RP (1980) A manual for starch gel electrophoresis. Staff paper series #11. Department of Natural Resources, University of Wisconsin-Madison

Phillips TD (1992) Evaluation of genetic diversity in the National Small Grains Collection of wild oat (*Avena sterilis* L.) germplasm based on isozyme polymorphisms. PhD thesis, Department of Crop Science, North Carolina State University, Raleigh

Price SC, Kahler AL (1983) Oats (*Avena* spp.). In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding, part B. Elsevier Scientific Publ. Co., Amsterdam, Netherlands, pp 103–127

Rezai A (1977) Variation for some agronomic traits in the world collection of wild oats (*Avena sterilis* L.). PhD thesis, Department of Agronomy, Iowa State Univ, Ames (Univ. Microfilms No 78-05968)

Rezai A, Frey KJ (1988) Variation in relation to geographical distribution of wild oats – seed traits. Euphytica 39:113–118

SAS (1985) SAS users guide: Statistics, version 5 edn. SAS Institute, Inc, Cary, North Carolina

Sneath PHA, Sokal RR (1973) Numerical taxonomy. The principles and practice of numerical classification. W. H. Freeman and Co, San Francisco, California

Soltis DE, Soltis PS (1989) Isozymes in plant biology. Dioscorides Press, Portland, Oregon

Souza E, Sorrells ME (1989) Inheritance and frequency of a null allele for diaphorase activity in North American oat cultivars. J Hered 80:501–503

Souza E, Sorrells ME (1991) Relationships among 70 North American oat germplasms: I. Cluster analysis using qualitative characters. Crop Sci 31:605–612

Stuber CW, Wendel JF, Goodman MM, Smith JSC (1988) Techniques and scoring procedures for starch gel electrophoresis of enzymes from maize (*Zea mays* L.). N.C. Agric Res Tech Bull 286, pp 1–95

Yen ST, Sadanaga K (1977) Inheritance of leaf peroxidase in oats. Can J Genet Cytol 19:303–312