Theor Appl Genet (1994) 88:56–64

M. Pérez de la Vega · L. E. Sáenz-de-Miera · R. W. Allard

Ecogeographical distribution and differential adaptedness of multilocus allelic associations in Spanish *Avena sativa* L.

Received: 18 February 1993 / Accepted: 23 July 1993

Abstract We determined the nine-locus isozyme genotype of 267 landrace accessions of *Avena sativa* from 31 provinces of Spain. Our results establish that level of genetic variability is usually high both within and among accessions of this heavily self-fertilizing hexaploid grass and that multilocus genetic structure differs in various ecogeographical regions of Spain. We concluded that selection favoring different multilocus genotypes in different environments was the main integrating force that shaped the internal genetic structure of local populations as well as the overall adaptive landscape of *A. sativa* in Spain. Implications in genetic resource conservation and utilization are discussed.

Key words *Avena sativa* · Genetic variability · Multivariate analysis · Multilocus associations · Genetic resources

Introduction

As interest in the conservation of plant genetic resources has increased over the plast several decades more and more reports of studies dealing with strategies for collecting, evaluating, managing and/or utilizing genetic variability have appeared in the litreature. In these studies genetic variability has usually been characterized on a single-locus basis (reviews in Goodman 1990; Marshall 1990; Shands 1990). However, it has become increasingly clear that epistatic interactions among loci at the two-locus, three-locus and higher-order levels, as well as interactions among alleles of the same locus (in polyploids), often have major effects on adaptedness and that it is important to take such interrelationships into account in genetic resource conservation and utilization (e.g., Evans 1984; Allard 1988, 1991; Pérez de la Vega et al. 1991b; García et al. 1991). Preliminary studies of a Spanish collection of hexaploid *Avena sativa* (Gómez et al. 1991; Pérez de la Vega et al. 1991a) have disclosed high levels of isozyme variability and the existence of multilocus associations of alleles in a Spanish collection of landraces of this species. However, as different regions of Spain are represented unequally in this collection, it proved difficult to obtain precise estimates of the ecogeographical distribution of genetic variability from the random sample of accessions that had been drawn from the collection. In the present study we have characterized isozyme variability, including multilocus associations, more adequately through studies of additional samples from 5 Spanish provinces (Badajoz, Castellón, Soria, Toledo, Zaragoza) that are especially well represented in the Spanish collection. Our results support those of the earlier studies in indicating that the level of genetic variability is high both within and among accessions of the collection. Our results also establish that specific single-locus and multilocus allelic combinations differ widely in frequency in different ecogeographical regions of Spain. This latter fact leads us to conclude that selection for different multilocus genotypes in different environments has been the main integrating force in shaping the internal genetic structure of local populations as well as the overall adaptive landscape of *A. sativa* in Spain.

Material and methods

The materials of this study were 267 accessions of the hexaploid cultivated oat, *Avena sativa* L. (2n = 6x = 42 chromosomes; homoeologous genomic formula AACCDD, Ansari and Thomas 1983) obtained from The Winter Cereal Collection, Germplasm Bank, National Institute of Agronomical Research, El Encín, Madrid. The accessions had been collected, mainly in 1944, at different localities within 37 provinces representing all climatic areas within continental Spain. According to available data the accessions are landraces, and each one should be the result of a single collecting action. With very few exceptions we have studied only 1 accession per locality.

Electrophoretic techniques, genetic nomenclature and descriptions of the enzyme systems are given in Gómez et al. (1991). A total of 3,355 seedlings (10–15 per accession) have been assayed electrophoretically at the present time, including the 490 seedlings (48 accessions) from Badajoz

Communicated by G. Wenzel

M. Pérez de la Vega (✉) · L. E. Sáenz-de-Miera
Department of Genetics Facultad de Biología, Universidad de León, E-24071 León, Spain

R. W. Allard
Department of Genetics, University of California, Davis, California 95616, USA

(BA), 205 seedlings (20 accessions) from Castellón (CS), 205 seedlings (20 accessions) from Soria (SO), 225 seedlings (22 accessions) from Toledo (TO) and 560 seedlings (38 accessions) from Zaragoza (ZA) scored in the present study. Each seedling was scored for the following eight enzyme systems to determine the multilocus enzyme genotype of the individual seedlings: glutamate oxaloacetate transaminase (GOT, EC 2.6.1.1), phosphoglucose isomerase (PGI, EC 5.3.1.9), leucine aminopeptidase (LAP, EC 3.4.11.1), peroxidase (PRX, EC 1.11.1.7), alcohol dehydrogenase (ADH, EC 1.1.1.1), esterase (EST, EC 3.1.1.2), phosphoglucose mutase (PGM, EC 2.7.5.1), malate dehydrogenase (MDH, EC 1.1.1.37). Nine zones of activity were analyzed (GOT2, PGI, LAP, PRX1, ADH, EST1, PGM, MDH1, MDH2). Seedlings of tetraploid *Avena barbata* of a known genotype were used as electrophoretic standards in gels. Locus and allelic designations of *A. sativa* follow those previously applied to *A. barbata* and the diploid *A. hirtula–A. wiestii* complex (Hutchinson et al. 1983; García et al. 1989). Studies of *A. barbata* have established that the homoeologous pairs of chromosomes in each of the two genomes of this heavily self-fertilizing tetraploid are homoallelic and homozygous for the same allele in both genomes for approximately one-third of the allozyme loci and also homozygous but heteroallelic (true-breeding for different alleles in the two genomes) for approximately two-thirds of the allozyme loci; also the heteroallelic homozygotes are often more widely adapted and more heterotic than their homoallelic counterparts (García et al. 1991).

These results with *A. barbata* lead us to two predictions concerning *A. sativa*: (1) that homoeologous pairs of chromosomes of *A. sativa*, which is also heavily self fertilizing, may sometimes be homoallelic for the same allele in all three genomes (e.g., genotypically *11 11 11*), sometimes heteroallelic but homozygous for two alleles (e.g., genotypically *11 11 22* or *11 22 22*) and sometimes heteroallelic but homozygous for three alleles (e.g., genotypically *11 22 33*); (2) that heteroallelic combinations may sometimes be more widely adapted and also have heterotic properties similar to those often attributed to true heterozygotes. Genotypes such as *11 11 22* and *11 22 22* often cannot be distinguished from one another with certainty through observation of electrophoretic gels. We therefore refer to variants in *A. sativa* as phenotypes rather than as genotypes and henceforth designate homoallelic combinations with a single-digit code [e.g., genotype *11 11 11* is coded *1*; two-allele combinations by a two-digit or a three-digit code (e.g., *12* if genotypes *11 11 22* and *11 22 22* cannot be distinguished unambiguously or *112* or *122* if genotypes *11 11 22* and *11 22 22* can be distinguished as a result of differences in banding intensity); and three-allele combinations by a three-digit code (e.g., genotype *11 22 33* is coded *123*)].

Results

Table 1 presents: (1) the total number of phenotypes and alleles for nine isozyme zones (each zone is assumed to be coded by a single Mendelian locus) that were observed in the entire sample of 3,355 plants from 267 accessions; (2) the number of phenotypes observed at a frequency ≥ 0.01 (≥ 0.02 for esterases) and the frequency of each phenotypic class; (3) the mean number and the range in number of phenotypes per accession. Among the nine zones of activity examined two (GOT2 and PGI) were invariant in all plants for phenotypes *12* and *15*, respectively, wheras five zones (PGM, MDH1, MDH2, PRX1, LAP) were intermediate in variability (3–6 phenotypes) and two zones (ADH and EST1) were highly variable (17 and 33 phenotypes, 6 or 7 with frequencies ≥ 0.01 or ≥ 0.02, respectively). For each of the seven variable zones (loci) 1 phenotype was clearly predominant (present at a frequency > 0.50), 1 or 2 were moderately frequent, and 1 to several were infrequent but present at a frequency ≥ 0.01 (≥ 0.02 for esterases). Among the 33 phenotypes shown in Table 1, 8 are homoallelic, 21 are heteroallelic for two alleles, and four are heteroallelic for three alleles. The predominant phenotype was heteroallelic for seven of the nine loci and homoallelic for two loci (*Lap*, *Est1*). Among the 267 accessions analyzed only 41 (15%) were monomorphic for all nine isozyme systems, whereas 226 accessions (85%) were polymorphic for one or more of the seven variable systems. Note also that the two invariant systems (GOT2, PGI) are both

Table 1 Phenotypic frequencies over 267 accessions (3,355 plants)[a]

Isozyme system	Total observed[a] number of		Phenotypes and frequencies[b]							Mean[c]
	Phenotypes	Alleles								
GOT2	1	2	*12* 1.00							1.00
PGI	1	2	*15* 1.00							1.00
LAP	6	4	*2* 0.03	*23* 0.01	*3* 0.61	*34* 0.04	*4* 0.31			1.39 (1–4)
PRX1	5	3	*1* 0.13	*15* 0.56	*35* 0.01	*5* 0.29				1.61 (1–4)
ADH	17	8	*24* 0.40	*25* 0.03	*4* 0.54	*46* 0.01	*45* 0.01	*245* 0.01		1.54 (1–5)
EST1	33	10	*1* 0.02	*12* 0.05	*2* 0.61	*23* 0.09	*24* 0.03	*25* 0.06	*27* 0.05	1.87 (1–5)
PGM	3	3	*12* 0.59	*13* 0.40	*123* 0.01					1.37 (1–3)
MDH1	4	3	*113* 0.46	*133* 0.52	*134* 0.02					1.44 (1–3)
MDH2	4	4	*12* 0.01	*14* 0.38	*124* 0.61					1.42 (1–3)

[a] Total sample

[b] Phenotypes in frequency < 0.01 have been omitted; < 0.02 for esterases

[c] Mean (and range) in numbers of phenotypes per accession

heteroallelic. Thus, genetic diversity was high within a great majority of the accessions and also in the collection as a whole. Considered as a whole the collection behaved as a population in which, on average, each locus had 2 phenotypes at roughly similar frequencies: the predominant phenotype versus a second composite allele made up by summing the frequencies of all remaining phenotypes. When seedling data were pooled together to form a population of diallelic loci the polymorphic index of this population was 0.537.

Table 2 gives the distribution of the most-frequent single-locus phenotypes in the 5 province sample; loci *Got2* and *Pgi* are omitted from this table and all subsequent tables because phenotypes *12* and *15*, respectively, are fixed in all accessions. It can be seen from Table 2 that, aside from phenotype *2* of *Est1*, no single phenotype was most-frequent in all 5 provinces; e.g., phenotype *3* of *Lap* was the most-frequent phenotype in 4 provinces (CS, SO, TO, ZA) but phenotype *4* of *Lap* is clearly predominant in Badajoz (BA). Similarly, although phenotype *2* of *Est1* was the most-frequent phenotype in all provinces, the second most-frequent phenotype of *Est1* differed among the 5 provinces. Comparisons of Tables 1 and 2 show that the most-frequent single-locus phenotypes in the entire Spanish collection were usually also the most frequent in each of the 5 provinces. There were, however, notable exceptions; e.g., phenotypes *Lap 3*, *Pgm 12*, *Mdh1 133*, *Prx1 15* and *Mdh2 124* were the most-frequent phenotypes over the entire collection but the first 3 (*Lap 3*, *Pgm 12*, *Mdh1 133*) were not the most frequent in Badajoz, and the fourth (*Prx1 15*) and fifth (*Mdh2 124*) were not the most-frequent phenotypes in Soria. Clearly some single-locus phenotypes interact in higher-order associations, and these interactions differ from province to province.

Table 2 Most-frequent single-locus phenotypes in the 5 provinces[a]

Code	Locus		BA	CS	SO	TO	ZA
A	*Lap*	*3*	0.14	0.89	0.98	0.70	0.76
		4	0.76	0.09	0.01	0.19	0.19
B	*Prx1*	*15*	0.87	0.39	0.21	0.55	0.62
		5	0.09	0.34	0.78	0.28	0.23
C	*Adh*	*4*	0.84	0.40	0.23	0.49	0.42
		24	0.11	0.60	0.73	0.45	0.44
D	*Est1*	*2*	0.48	0.49	0.76	0.81	0.64
		12					0.11
		23			0.16	0.10	
		27	0.23				
		29		0.10			
E	*Pgm*	*12*	0.12	0.88	0.97	0.60	0.70
		13	0.88	0.11	0.03	0.39	0.29
F	*Mdh1*	*113*	0.85	0.25	0.21	0.47	0.21
		133	0.15	0.75	0.74	0.53	0.79
G	*Mdh2*	*124*	0.88	0.69	0.05	0.32	0.57
		14	0.12	0.30	0.95	0.68	0.43

[a] BA, Badajoz; CS, Castellón; SO, Soria; TO, Toledo; ZA, Zaragoza

Distribution of multilocus associations

In all more than 400 multilocus associations were observed in the entire collection. Table 3 lists the 15 most-frequent multilocus associations and gives their frequencies in the entire collection and also in the 5-province set. The single most-frequent multilocus association in the entire collection and also in the 5-province set was phenotype *1* of Table 3 (*Lap 4*, *Prx1 15*, *Adh 4*, *Est1 2*, *Pgm 13*, *Mdh1 113*, *Mdh2 124*; coded A, B, C, D, E, F, G, left to right in Table 3). This phenotype made up nearly 11% of the entire collection (309/2,875 seedlings of 47 accessions). The frequencies of the remaining multilocus associations fell off rapidly thereafter to only 0.014 for the 15th most-frequent multilocus combination (Table 3). Also, each of 204 multilocus associations had its own unique genotype, i.e., among the more than 400 multilocus combinations 204 were observed only once among the 2,875 seedlings examined

Table 3 The 15 most frequent multilocus associations in the entire collection and the 5-province set

Phenotype	Locus and code							Whole collection			Five-province set			
	Lap A	*Prx1* B	*Adh* C	*Est1* D	*Pgm* E	*Mdh1* F	*Mdh2* G	Observed[a] frequency	Number of accessions	Number of provinces	Frequencies[b] Observed	Expected	Number of accessions	Number of provinces
1	4	15	4	2	13	113	124	0.107	47	18	0.092	0.113	29	4
2	3	5	24	2	12	133	14	0.056	33	11	0.081	0.093	25	5
3	3	15	24	2	12	133	14	0.046	22	8	0.050	0.204	16	4
4	4	15	4	27	13	113	124	0.043	22	4	0.073	0.015	21	3
5	4	15	4	23	13	113	124	0.037	18	9	0.032	0.016	11	3
6	4	15	4	25	13	113	124	0.030	15	7	0.028	0.012	10	3
7	3	15	4	2	12	133	14	0.030	13	8	0.039	0.268	9	4
8	3	5	24	2	12	133	124	0.023	16	9	0.026	0.129	8	3
9	3	1	4	2	12	113	124	0.022	8	7	0.010	0.044	2	2
10	3	5	4	2	12	133	124	0.018	12	8	0.004	0.169	5	1
11	3	5	4	2	12	133	14	0.018	17	11	0.017	0.121	9	4
12	3	15	24	2	12	133	124	0.017	14	6	0.017	0.283	10	3
13	3	5	24	2	12	113	14	0.016	9	6	0.017	0.072	1	1
14	3	15	4	2	12	133	124[c]	0.015	14	6	0.021	0.372	8	2
15	3	1	24	2	12	113	124	0.014	8	7	0.017	0.033	3	2

[a] Over a total of 2,875 plants and 235 accessions

[b] Over a total of 1,685 plants and 148 accessions. Expected frequencies were calculated from the single-locus frequencies in the five provinces

[c] The theoretically most-common phenotype on the basis of single-locus frequencies in both the whole collection and the 5-province set

($F = 1/2{,}875 = 0.003$ each). Only 41 of the 235 accessions were fixed for a single multilocus association; 27 different multilocus associations were observed in these 41 fixed accessions.

The theoretically most-frequent multilocus association, as calculated from the single-locus frequencies in the 5-province set, was phenotype 14 of Table 3 (*Lap 3, Prx1 15, Adh 4, Est1 2, Pgm 12, Mdh1 133, Mdh2 124*). Phenotype 14 was observed in only 44 seedlings ($F = 0.015$) of 14 accessions; i.e., it was present at a frequency of approximately one-seventh that of phenotype 1, the actual most-frequent phenotype.

Figure 1 shows that the frequencies of the 10 most-frequent multilocus phenotypes of Table 3 differ strikingly from province to province. As examples, phenotypes 1, 4, 5 and 6 were much more frequent in Badajoz than in CS, SO, TO or ZA, whereas phenotypes 2 and 8 were predominant, but phenotype 1 was absent, in SO. This suggests that different alleles in multilocus associations interact with each other and that the interactions differ in different ecogeographical regions. In the next section we apply log-linear multivariate techniques to the analysis of interactions among pairs, triplets and higher-order associations among three specific multilocus associations, designated Badajoz-1, Badajoz-2 and Soria, to test the statistical significance of such associations.

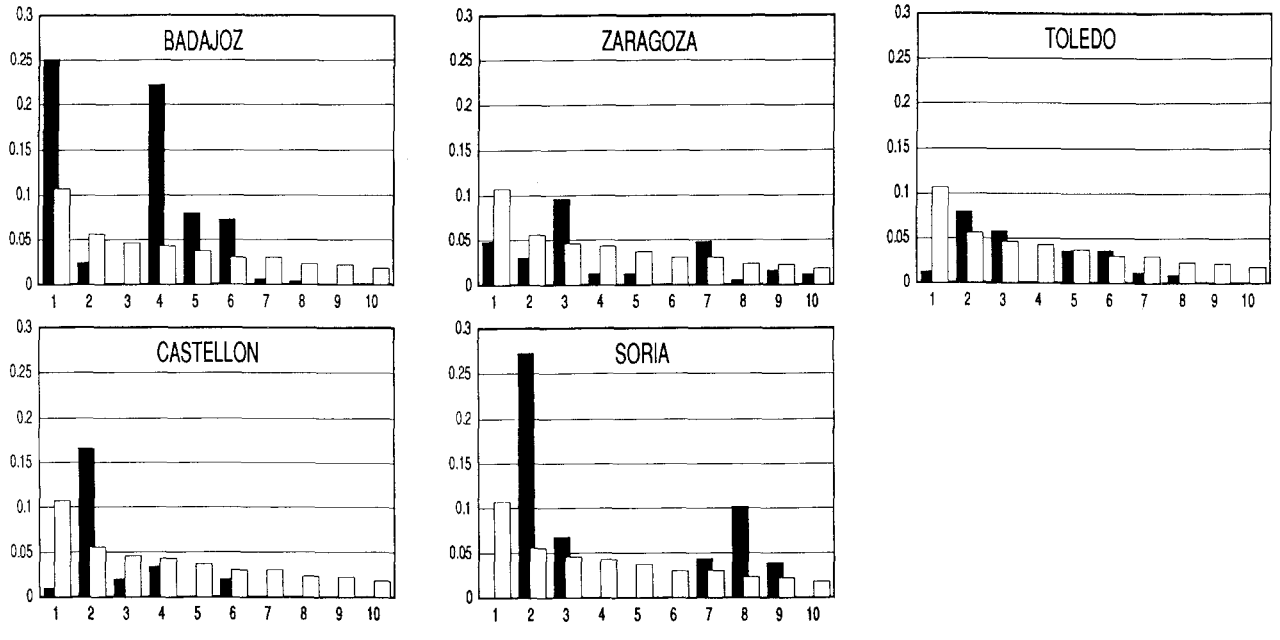
Multivariate analysis

The specific multivariate analysis (Fienberg 1980) used in the study of multilocus associations is the one previously used in studies of multilocus associations in barley (Zhang et al. 1990; Saghai Maroof et al. 1990) and *A. barbata* (Pérez de la Vega et al. 1991b). The seven polymorphic loci with frequencies ≥ 0.01 yield a total of 39 single-locus phenotypes. Hence, the number of seven-locus associations is more than an order of magnitude larger than the number of plants classified (1,685) leaving no possibility of analyzing all of the enzyme systems or phenotypes simultaneously. Accordingly, to limit the number of classes to a manageable level we: (1) grouped the loci into sets of six and (2) reduced each locus to diallelic state by designating the predominant phenotype of each locus as allele *1* and summing the frequencies of all other phenotypes of that locus to form a second composite allele that we designated allele *2*. In this notation *111111* indicates a plant with the predominant allele (allele *1*) in each of six loci and *222222*, a plant with the composite allele (allele *2*) in all six loci. Thus, for six loci there are $2^6 = 64$ classes (1,685 plants over 64 classes gives 26 plants expected on average per class assuming that all alleles are equally frequent and all associations are at random).

The multivariate analysis is a two-step process: first, likelihood ratio tests are employed in a series to eliminate terms with statistically insignificant effects; second, log-linear models are constructed to fit the data to the remaining terms in the system. Models are fitted in a hierarchical manner such that a higher-order term is included only when lower-order terms fail to fit the data. Also, when a higher-order term is included all of its lower-order terms are also included. The best-fitting model is then chosen on the basis that it is maximally parsimonious, i.e., it includes a minimal number of terms and also provides a statistically acceptable fit to the data. We used two screening methods in model selection, stepwise selection (Goodman 1971) and marginal tests (Brown 1976).

The designations of the seven variable loci were abbreviated A, B, C, D, E, F and G, as in Table 2 and 3. The loci were grouped into the Badajoz-1, Badajoz-2 and Soria sets of six loci each by including or excluding EST1 or PRX1 and by assigning single-locus phenotypes to either allele *1* or allele *2*. Individuals were cross classified and the data of each set were tabulated in the form of contingency tables and then subjected to log-linear analyses to determine the two-locus, three-locus and higher-order multilocus structure of each set.

Fig. 1 Frequencies of the 10 most-frequent multilocus associations in each of the 5 provinces (*filled columns*) and in the whole collection (*empty columns*)



The first set of six loci, designated Badajoz-1, included as selected phenotypes (designated *1*) each single-locus phenotype present at the highest frequency in Badajoz province (*Lap 4, Prx1 15, Adh 4, Pgm 13, Mdh1 113* and *Mdh2 124*); the allelic designation of this set is *111111*, and the allelic designation of the alternative composite allelic set is *222222*. *Est1* was not included because phenotype *2* was the most-frequent phenotype in all of the provinces and also because the second allele was a composite of many phenotypes and hence was exceptionally heterogeneous. The frequency of the *111111* configuration in Badajoz province was 69.2% (339/490 plants); its overall frequency in Spain was much lower, 24.9%, which is equivalent to the summation of the frequencies of phenotypes 1, 4, 5 and 6 of Table 3 (= 0.2170) plus those of a few other phenotypes too infrequent to be included in Table 3. The six-locus *111111* association was present in 39/48 (81%) of the accessions from Badajoz (fixed or clearly predominant in 33/48 = 69% accessions) and also present in 42 additional accessions from 18 of the 37 provinces represented in the Spanish collection. The *111111* association is clearly the most successful association in Badajoz province, and it is also reasonably well-adapted in a number of other regions of Spain. The best-fitting model in the sense of being maximally parsimonious and also giving a satistically acceptable fit to the

Table 4 Observed and expected numbers and standardized residuals for each cell under the best-fitting model for the Badajoz-1 set

				LAP1			
				1		2	
				PRX1		PRX1	
				1	2	1	2
MDH2 1	MDH1 1	PGM 1	ADH 1	418	4	42	6
				415.2	7.0	44.0	3.8
				0.1	−1.1	−0.3	1.1
			ADH 2	22	4	6	7
				21.5	4.2	7.3	6.0
				0.1	−0.1	−0.5	0.4
		PGM 2	ADH 1	6	3	12	18
				5.5	1.5	11.5	20.6
				0.2	1.1	0.1	−0.6
			ADH 2	1	2	25	35
				3.9	1.2	23.1	34.8
				−1.4	0.6	0.4	0.0
	MDH1 2	PGM 1	ADH 1	2	2	17	17
				1.6	2.1	16.7	17.5
				0.2	−0.1	0.1	−0.1
			ADH 2	3	1	18	13
				2.2	2.0	19.4	11.3
				0.5	−0.7	−0.3	0.5
		PGM 2	ADH 1	2	5	73	40
				5.7	3.4	71.8	39.1
				−1.5	0.8	0.1	0.1
			ADH 2	10	14	77	74
				8.3	13.6	76.2	76.9
				0.6	0.1	0.1	−0.3
MDH2 2	MDH1 1	PGM 1	ADH 1	6	1	19	10
				7.2	0.9	18.6	9.2
				−0.4	0.1	0.1	0.2
			ADH 2	2	0	0	7
				0.8	0	0.3	7.8
				1.0	0.0	−0.4	−0.3
		PGM 2	ADH 1	2	2	3	2
				3.8	0.9	2.3	2.0
				−0.9	0.9	0.4	0.0
			ADH 2	12	0	20	34
				11.1	0.3	19.9	34.8
				0.3	−0.3	0.0	−0.1
	MDH1 2	PGM 1	ADH 1	2	1	10	23
				0.1	1.8	12.5	21.6
				2.5	−0.5	−0.7	0.3
			ADH 2	6	5	27	16
				8.2	3.9	13.1	28.7
				−0.8	0.5	0.8	−0.3
		PGM 2	ADH 1	2	1	87	48
				0.9	1.5	85.6	50.1
				0.9	−0.3	0.2	−0.3
			ADH 2	2	0	101	255
				1.8	0.8	103.7	251.7
				0.1	−0.7	−0.3	0.2

data for the Badajoz-1 set was [ABCG] [ACEG] [AEFG] [BCEF] [BEFG] [CEFG] [ABF] [ACF]. According to this model the six loci are associated through a series of interlocking fourth and lower-order interactions. The value of the resulting G^2 statistic, which is distributed approximately as a χ^2, was 23.02 with 17 degrees of freedom and a probability of 0.1485; thus the model fits the data well. The contingency table comparing the observed and the expected individuals according to the selected model is given in Table 4, which also includes the standardized residuals [SR = (observed − expected)/(expected)$^{1/2}$]. Asymptotically, standardized residuals follow a normal distribution with zero mean and unit variance. Thus, observed absolute values of an SR of 1.96, 2.58, or 3.29 indicate statistically significant departures from the expectation for the model specified, the null hypothesis being tested at probability levels of 0.05, 0.01 or 0.001, respectively. Only 1 of the 64 values of Table 4 is significant (SR = 2.5), which supports the G^2 statistic in indicating that the selected model fits the data well.

Lower-order terms in the models are nested within the higher-order associations, and they must consequently be interpreted with caution. Nevertheless, inferences about the nature and intensity of interactions can be made from the standardized *u* values (Table 5). As an example, *11* and *22* combinations are generally in excess in the two-locus associations of the Badajoz-1 set; thus, for the AF pair (*u* value = 6.545***) observed versus expected numbers (assuming random association) of gametic types *11*:*12*:*21*:*22* are 485:58:246:896 and 235:307:496:647, respectively ($\chi^2_{[1]}$ = 688.4, $P < 0.001$). Under the null hypothesis of random association, standardized residuals (SR) for AF gametic classes *11* and *22* are 16.3 and 9.8 respectively, whereas they are −14.2 and −11.2 for gemetic classes *12* and *21*, respectively; thus the *11* and *22* classes are clearly in significant excess ($P < 0.001$) and the *12* and *21* classes are clearly in significant deficiency ($P < 0.001$). Standardized *u* values indicate that significant associations also occur between all other pairs, AB, AE, AG, BF, BG, CE, CF, EF and FG. Even the BG pair (lowest significant *u* value, 2.139*) deviated significantly from random association; observed *11*:*12*:*21*:*22* numbers were 734:290:245:416 and expected numbers were 429:595:277:384 ($\chi^2_{[1]}$ = 379.3, $P < 0.001$). Algebraic signs (not reported) indicate that the *11* and *22* gametic configurations are consistently in excess and that the *12* and *21* configurations are consistently in deficiency. Thus, the associations always show excesses in the direction of the preferred Badajoz-1 and non-Badajoz-1 configurations but deficiencies in the combinations involving mixtures of the two types.

The 2 most intense four-locus associations are AEFG (5.143***) and BCEF (5.156***). Algebraic signs of the *u* values, even after allowing for all lower-order associations, indicate there are excesses in the four-locus allelic associations *1111*, *1122*, *1212*, *1221*, *2112*, *2121*, *2211*, *2222* and deficiencies in the corresponding associations. As examples, expected versus observed numbers in AEFG were 41 versus 448 for *1111* and 217 versus 491 for *2222*, and in BCEF they were 99 versus 485 for *1111* and 102 versus 343 for *2222*. Results were similar for the other significant four-locus terms of Table 5 except for CGEF in which *1111* and *2222* are in deficiency.

Table 5 Standardized absolute *u* values for the statistically significant terms (*A* LAP, *B* PRX1, *C* ADH, *D* EST1, *E* PGM, *F* MDH1, *G* MDH2)

Term	Badajoz-1	Badajoz-2	Soria
AB	4.054***		
AC			2.197*
AD		7.225***	
AE	5.733***	6.699***	7.769***
AF	6.545***	7.613***	4.841***
AG	3.690***	5.236***	4.039***
BC			2.401*
BE			2.724**
BF	4.186***		4.236***
BG	2.139*		3.727***
CE	4.819***	9.891***	6.627***
CF	3.433***	6.189***	
CG		5.609***	3.068**
DF		4.387***	
EF	4.368***	5.054***	4.693***
EG		2.009*	2.670**
FG	3.495***	4.162***	1.994*
ABC	2.088*		
ABF	4.162***		3.201**
ACE	2.289*		
ACF	2.840**	4.335***	3.764***
AEG			2.920**
AFG			1.979*
BCF	2.734**		2.183*
BEF	2.360*		
BEG	3.882***		2.215*
BFG			4.240***
CEF	5.351***	6.318***	
CFG			3.320***
EFG	3.474***	6.192***	
ABCE			3.354***
ABCG	3.049***		
ABEG			3.181**
ACEG	2.740**		
AEFG	5.143***	7.915***	5.472***
BCEF	5.156***		3.405***
BEFG	2.834**		
CEFG	3.166**		

*, **, ***, $P \leq 0.05$, 0.01 and 0.001, respectively

Thus, with occasional exceptions, associations featuring all 4 Badajoz or all 4 non-Badajoz phenotypes, or pairs such as *1122*, were in excess. The existence of statistically significant fourth-order interactions involving all 6 phenotypes of the model (ABCEFG) implies that two- and three-locus interactions involving these phenotypes are nested within the higher-order associations. The fourth-order associations of this model include all of the 15 second-order associations, although only 10 are statistically significant, and 17 of the 20 third-order associations, of which only 7 are significant. Among the 3 third-order associations not included in four-locus associations, ABF and ACF are included in the model and are statistically significant, whereas ABE is not included in the model and is not satistically significant.

Esterase locus (*Est1*) was included in the second set of loci (Badajoz-2) because several esterase phenotypes in addition to phenotype *2* (the most frequent phenotype, $F = 0.48$) were well represented in this province; *Est1 27*, present at a frequency of 0.23 in the province of Badajoz (Table 2), was designated as

esterase phenotype *1* in the model. PRX1 was excluded to keep the number of loci in the set at six; all *Est1 27* plants also were *Prx1 15*, which had the advantage of predicting the peroxidase phenotype exactly. The chosen phenotype of the Badajoz-2 set was *Lap 4, Adh 4, Est1 27* (also *Prx1 15*), *Pgm 13, Mdh1 113, Mdh2 124*, coded A, C, D (= B), E, F, G. This association is the fourth most-frequent association of Table 3. Among the 2,875 plants of the Spanish collection only 125 (4.35%) had this phenotype (109 plants from 17 accessions of Badajoz and 16 plants from 5 accessions from 3 different provinces). The best-fitting model for this set was [AEFG] [ACF] [CEF] [AD] [CG] [DF], $G^2 = 48.99$, $P = 0.1092$ with 38 *df*. Only two SR values in the contingency table (not shown) were significant ($P < 0.05$). Thus, interactions are fewer and less intense than in the Badajoz-1 set, probably due to the exceptional heterogeneity of EST1. Nevertheless, some terms, [AEFG] and [ACF], are common with those of Badajoz-1, or are included in higher-order term, e.g., [CEF] in [CEFG], of Badajoz-1. *Est1* (D) participates only in 2 second-order interactions, 1 with *Lap* (A) and 1 with *Mdh1* (F).

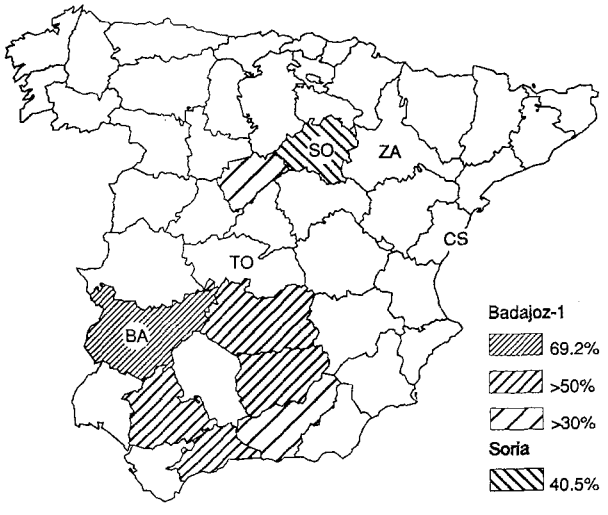
The Badajoz-2 model included fewer two-locus and fewer higher-order associations (only 12 out of 15 in total, Table 5) than the Badajoz-1 model. The 3 two-locus associations not included in the Badajoz-2 model included the esterases (CD, DE, DG). It is thus clear that the polymorphism of the esterases lumped into composite allele *2* introduces "noise" that masks the extent of interactions in the Badajoz-2 set. Regardless, the overall pattern is similar to that of the Badajoz-1 set: gametic classes including two to four *1* or *2* alleles are in excess relative to expectations calculated on the assumption that the associations are random.

The third six-locus set included the most-frequent phenotypes, designated *1*, and a composite of all other phenotypes, designated *2*, of the province of Soria. The selected six-locus set included *Lap 3, Prx1 5, Adh 24, Pgm 12, Mdh1 133, Mdh2 14* (coded A, B, C, E, F, G). This six-locus association was present in 83/205 plants (40.5%) from 12/20 accessions from Soria, wheras its frequency over all of Spain was only 7.48%. However, this six-locus association was relatively widespread (it was observed in 29 additional accessions from 13 provinces). The best-fitting model was [ABCE] [ABEG] [AEFG] [BCEF] [ABF] [ACF] [BFG] [CFG], $G^2 = 30.6$, $P = 0.806$. For this model only one SR was significant ($P < 0.05$). There were excesses of the *11* and *22* combinations in all of the 13 significant two-locus associations and an excess of the *1111* and *2222* combinations in all of the significant four-locus associations. Thirteen three-locus associations were included in the four-locus terms of the model; 4 additional three-locus terms (ABF, ACF, BFG, CFG) were directly included in the model, wheras 3 remaining three-locus terms (ACG, BCG, CEF) were not significant. Thus, 17 three-locus terms, of which 8 were significant (Table 5), were included in the model.

Comparisons of the Badajoz-1, Badajoz-2 and Soria models revealed that the term AEFG is the only four-locus term that is significant in all three models (Table 5). This term was also present in the previous multilocus analysis of the Spanish oat collection (Pérez de la Vega et al. 1991a). Analysis of this four-locus set revealed that the model which best fit the data was the saturated model [AEFG], under which model the expected and observed numbers for each cell do not differ significantly and all standardized residuals take values of zero. This result indicates that the allelic state at any one locus depends on the allelic state at each of three other loci and also that the relationship among any three loci is influenced by the fourth locus.

When the data of only those provinces represented by several populations were taken into account (some provinces are represented in the collection by only 1 or 2 accessions), the geographical distribution of the multilocus associations discussed above showed a clear regional pattern. The Badajoz-1 set was present at frequencies of 0.69 in Badajoz, at frequencies higher than 0.50 in provinces near Badajoz and at frequencies of approximately 0.30 in the provinces of Granada and Segovia (Fig. 2). The six-locus Soria set is infrequent in all provinces other than Soria (Fig. 2). The distributional pattern for the four-locus Badajoz set [AEFG] is somewhat different: its frequency is 0.73 in Badajoz, 0.47 in the southwest part of Spain, between 0.08 and 0.10 in Castellón, Toledo and Zaragoza, 0.23 in Cataluña, about 0.29 in the Northern Meseta (plateau) including Madrid, but absent in Soria (Fig. 3). The four-locus Soria set [AEFG] is predominant in Soria (0.71), moderately frequent in Castellón and Zaragoza (0.28), Cataluña (0.23) and in the Meseta (0.30), but almost absent in Badajoz and Southwestern Spain (including Ciudad Real) (Fig. 3). Another relevant point is that the number of plants with the Badajoz four-locus and Badajoz-1 six-locus sets are nearly identical in those provinces where these two sets are well represented. In Badajoz 339/359 (94%) of plants with the Badajoz four-locus set also carried the six-locus Badajoz-1 set; in Zaragoza and the northern half of the Meseta corresponding numbers were 49 and 46 and 83 and 83, respectively. Correspondences were less close for the Soria four-locus and six-locus sets; in Soria, of 145 plants with the four-locus set, 83 had the six-locus set; in Zaragoza and the Meseta corresponding numbers were 155 and 17 and 188 and 47, respectively. Clearly, the Badajoz-1 six-locus association is strongly

Fig. 2 Geographical distribution of the six-locus associations of Badajoz-1 and Soria. See Table 2 for the observations of provinces



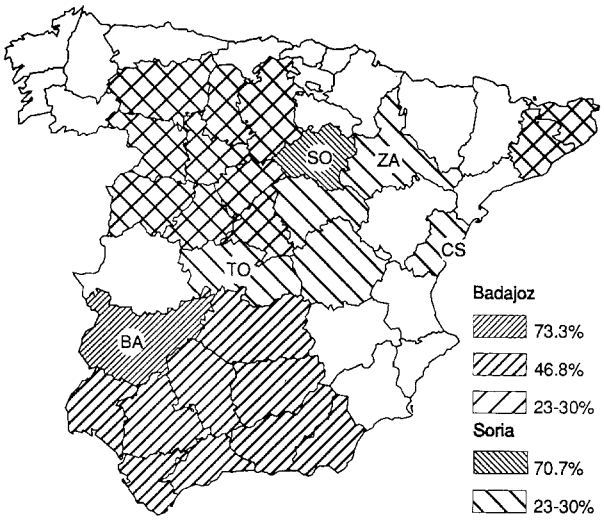


Fig. 3 Geographical distribution of the four-locus associations

influenced by the four-locus association, but the Soria six-locus association is less influenced by the four-locus association.

Discussion

The results of this study indicate that levels of genetic variability are relatively high both within and among Spanish landraces of *A. sativa* and also that the multilocus genetic structure of the landraces differs in different ecogeographical regions of Spain. Each of the nine enzyme loci exmined was genetically variable. Much of the observed genetic variability was attributable to intralocus allelic diversity. *A. sativa* forms 21 pairs of bivalents with great regularity at meiosis indicating that it is fully diploidized and that genetic exchanges do not occur between corresponding chromosomes of the homoeologous genomes. Sometimes each of the three pairs of homoeologous loci carries the same allele (say allele *1*) of a given locus; such individuals are genotypically *11 11 11*, i.e., they are homoallelic and also homozygous (*A. sativa* is heavily self fertilized and homozygous at virtually all loci). Other loci are heteroallelic, sometimes for two different alleles (genotypically *11 11 22* or *11 22 22*) and sometimes for three diffrent alleles (genotypically *11 22 33*). The basic units of allelic function in *A. sativa* are thus nonsegregating sets of three homoeologous pairs of alleles (six alleles) stabilized by preferential pairing within each of the three genomes. Among the 33 such sets that were present at a frequency ≥ 0.01 (≥ 0.02 for *Est1*) in the Spanish landraces (Table 1) 8 (24%) were homoallelic whereas the 25 remaining sets were heteroallelic for two different alleles (21 sets, 64%), or for three different alleles (4 sets, 12%).

Two of nine loci examined were monomorphic throughout the entire sample of 3,355 plants (locus *Got2* for phenotype *12* and locus *Pgi* for phenotype *15*). Thus, all of the genetic variability at these two loci is due to intralocus allelic diversity. The remaining seven loci were polymorphic for 1 or more homoallelic and/or heteroallelic phenotypes; consequently part of the genetic variability of these seven loci results from intralocus allelic diversity and part from polymorphism for diffrent single-locus homoallelic and/or heteroallelic nonsegregating sets of homoeologous pairs of alleles. The frequencies of different single-locus phenotypes differed widely in different ecogeographical regions, often in correlation with environmental variables. As an example, among the 170 plants bearing allele 7 of *Est1*, 121 were found in the province of Badajoz (6 in homoallelic 7 plants and 115 in heteroallelic *27* plants, most of them (93%) as part of the Badajoz multilocus associations) and 47 were found in 5 other provinces (including 4 provinces well represented in the collection). Thus, allele 7 provides a clear example of an allele associated with a geographical area and also with a multilocus association; its distribution suggests that this allele contributes to the adaptedness of the Badajoz complex in the mild climate of Badajoz and that the *27* heteroallelic six-allele combination is favored over the 7 homoallelic six-allele combination. In this connection we also note that the most-frequent single-locus phenotype was heteroallelic for six of the nine loci (*Got2 12, Pgi 15, Prx1 15, Pgm 12, Mdh1 133, Mdh2 124*) and homoallelic for only three loci (*Lap 3, Adh 4, Est 2*); this implies that some intralocus interactions among different alleles of loci *Got2, Pgi, Prx1, Pgm, Mdh1* and *Mdh2* are adaptively favorable.

The multilocus data provide evidence that interlocus (epistatic) interactions among alleles of diffrent loci often have large effects on adaptedness and hence on the multilocus structure of population genotypes. The first indication that this might be the case in *A. sativa* came from large differences between observed multilocus frequencies and the expected frequencies calculated assuming that associations between loci form at random (the null hypothesis). Discrete log-linear multivariate analyses provided direct evidence that at least six of the nine loci of this study are organized into complex multilocus associations held together by overlapping two- to four-locus interactions. Thus, in common with other inbreeding species, e.g., barley (Zhang et al. 1990) and *Avena barbata* (Clegg and Allard 1972; Pérez de la Vega et al. 1991b) the genomes of the Spanish landraces of *A. sativa* are hierarchically structured: some loci interact in pairs, and some pairs interact with other loci or with other pairs to form third-, fourth- and increasingly complex higher-order associations. The frequencies of different mutilocus associations vary widely in different ecogeographical regions in correlation with environmental conditions, e.g., the Badajoz multilocus associations are present at the highest frequency in the warmer climates of southern Spain, while the Soria associations are most frequent in the colder areas (mean temparatures and number of days of frost are 16.8 °C and 10.7 days in Badajoz and 10.5 °C and 92.1 days in Soria). On the other hand, the distribution pattern of multilocus associations does not fit possible historical or commercial patterns. To test the correlation between multilocus associations and climate, the 5 provinces were divided into two climatic classes, the warmer one with an average annual temperature higher than 16 °C (including Badajoz and Castellón) and the colder one with an average temperatures lower than or equal to 15 °C (including

Soria, Toledo and Zaragoza). When the four-locus sets were tested, a clear accordance between the expected phenotype-temperature associations and the results observed by the multivariate analysis was obtained. Thus, for instance, when the four-locus Badajoz set [AEFG] was analyzed together with the temparature [T] the best-fitting model was [AEFG] [AEGT] [EFT], $G^2 = 11.36$, $P = 0.078$. It shows that (1) the association among the four loci is not broken up when temperature is included in the analysis; (2) there is a fourth-order association (AEGT, $u = 3.03$) between three loci and the temperature and (3) the fourth loci (F) not included in this term is included in a third-order one interacting with *Pgm* [EFT]. In the terms [AEGT] and [EFT] the respective multilocus associations Badajoz *111* and *11* were associated in excess with the warmer temperature class and the complementary non-Badajoz *222* and *22* with the colder temperature class.

Genetic linkage among loci is generally considered to be necessary for the development and maintenance of multilocus associations. However, in barley (Allard 1988; Zhang et al. 1990) and in *A. barbata* (Pérez de la Vega et al. 1991b) the closest associations are those among loci located on different chromosomes rather than associations among linked loci on the same chromosome. The mating system of *A. sativa*, in common with that of other heavily selfing species, is favorable for the development of patchwork patterns of epistatic combinations of alleles that provide for high local adaptedness. The one percent of within-population outcrossing that occurs in *A. sativa* leads to heterozygosity, segregation and recombination that produces the novel interlocus allelic combinations upon which evolutionary change depends; it also allows favorable new mutant and migrant alleles to be rapidly incorporated into the population genotype. The 99% of selfing causes all loci, whether located on the same or different chromosome, to behave as if they are linked with crossover values ≤ 0.01, thus restricting recombination sufficiently to increase the likelihood that favorable interlocus combinations, including combination with relatively small selection advantage, will be integrated into local populations before they are broken up by segregation and recombination (Allard 1975).

The results of the present study support the notion that information concering the adaptive properties of specific alleles, specific within-locus (intralocus) associations and specific interlocus (epistatic) multilocus allelic combinations can be useful in developing strategies for collecting, evaluating, managing and utilizing genetic variability.

Acknowledgements This research was supported by grants from the Spanish National Institute for Agronomical Research (INIA 7689) and the DGICYT (PB88-4015).

References

Allard RW (1975) The mating system and microevolution. Genetics 79:115–126

Allard RW (1988) Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenitors. J Hered 79:225–238

Allard RW (1991) Predictive methods for germplasm identification. In: Stalker HT, Murphy JP (eds) Plant breeding in the 1990s. CAB Int, Wallingford, UK, pp 119–146

Ansari N, Thomas H (1983) A study of homoeologous relationships in the cultivated oat *Avena sativa* ($2n = 4x = 42$). Theor Appl Genet 66:303–305

Brown MB (1976) Screening effects in multidimensional contingency tables. Appl Stat 25:37–46

Clegg MT, Allard, RW (1972) Patterns of genetic differentiation in the Slender Wild Oat species, *A. barbata*. Proc Natl Acad Sci USA 69:1820–1824

Evans LT (1984) Physiological aspects of varietal improvement. In: Gustafson JP (ed) Gene manipulation in plant improvement. Plenum Press, New York, pp 21–46

Fienberg SE (1980) The analysis of cross-classified categorial data. MIT Press, Cambridge, Mass.

García P, Vences FJ, Pérez de la Vega M, Allard RW (1989) Allelic and genotypic composition of ancestral Spanish and colonial Californian gene pools of *Avena barbata*: Evolutionary implications. Genetics 122:687–694

García P, Morris MI, Sáenz-de-Miera LE, Allard RW, Pérez de la Vega M, Ladizinsky G (1991) Genetic diversity and adaptedness in tetraploid *Avena barbata* and its diploid ancestors *Avena hirtula* and *Avena wiestii*. Proc Natl Acad Sci USA 88:1207–1211

Gómez C, Sáenz-de-Miera LE, Pérez de la Vega M (1991) Estimation of the isozymatic genetic variability of a Spanish *Avena sativa* germplasm collection. Plant Breed 106:293–300

Goodman LA (1971) The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classification. Technometrics 13:33–61

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

Hutchinson ES, Hakim-Elahi A, Miller RD, Allard RW (1983) The genetics of the diploidized tetraploid *Avena barbata*: acid phosphatase, esterase, leucin aminopeptidase, peroxidase and 6-phosphogluconate dehydrogenase loci. J Hered 74:325–330

Marshall DR (1990) Crop genetic resources: current and emerging issues. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding, and genetic resources. Sinauer Assoc, Sunderland, Mass., pp 367–388

Pérez de la Vega M, Gómez C, Cabo D, Vences FJ, Sáenz-de-Miera LE (1991a) Genetic variability of a Spanish *Avena* germplasm collection. Vort Pflanzenzuecht 20:22–27

Pérez de la Vega M, García P, Allard RW (1991b) Multilocus genetic structure of ancestral Spanish and colonial Californian populations of *Avena barbata*. Proc Natl Acad Sci USA 88:1202–1206

Saghai Maroof MA, Allard RW, Zhang Q (1990) Genetic diversity and ecogeographical differentiation among ribosomal DNA alleles in wild and cultivated barley. Proc Natl Acad Sci USA 87:8486–8490

Shands HL (1990) Plant genetic resources conservation: the role of the gene bank in delivering useful genetic materials to the research scientist. J Hered 81:7–10

Zhang Q, Saghai Maroof MA, Allard RW (1990) Woldwide pattern of multilocus structure in barley determined by discrete log-linear multivariate analyses. Theor Appl Genet 80:121–128