Theor Appl Genet (1990) 80:121–128

THEORETICAL AND APPLIED GENETICS

Worldwide pattern of multilocus structure in barley determined by discrete log-linear multivariate analyses

Qifa Zhang *, M. A. Saghai Maroof, ***, and R. W. Allard**

Department of Genetics, University of California, Davis, CA 95616, USA

Received January 17, 1990; Accepted January 22, 1990
Communicated by H. F. Linskens

Summary. Data from the electrophoretic assay for seven enzyme loci of 1,032 accessions of cultivated barley, *Hordeum vulgare* L., from the USDA world barley collection were analyzed for multilocus structure using discrete log-linear multivariate techniques. Three major steps were involved in the analysis: (i) identification and elimination of terms that have inconsequential effects in multilocus association; (ii) construction of a log-linear model that best describes the complete multilocus structure of the genetic system; and (iii) evaluation of each of the association terms included in the model. The results of analyses of two subsets of loci show that the multilocus genetic system of cultivated barley, including loci located on different chromosomes, is organized into hierarchically structured complexes of loci. Multilocus structure differs in various geographical regions of the world. The structure of barleys from Southwest Asia, the putative center of origin for cultivated barley, is intermediate for both subsets of loci. Differences increased progressively across the Eurasian-African landmasses in each direction with increasing distance from Southwest Asia, with the consequence that the barleys from West Europe, East Asia, and Ethiopia are maximally different from those of Southwest Asia and Middle South Asia.

Key words: Barley – Log-linear multivariate analyses – Multilocus associations – Enzyme loci

Introduction

It is well documented that barleys from different geographical areas often differ from each other in both single-locus genotype and multilocus combinations (e.g., Ward 1962; Brown et al. 1980; Kahler and Allard 1981; review in Allard 1988). In addition, studies of the experimental barley populations have shown that striking nonrandom multilocus associations of alleles develop over generations (Weir et al. 1972, 1974; Clegg et al. 1972), and that the same population develops different gene complexes when grown in different environments (Jana et al. 1989).

Associations among loci in populations have conventionally been assessed in terms of two-locus linkage disequilibrium measures, and most inferences about multilocus genetic structure have been based on estimates of two-locus parameters. However, experimental evidence (e.g., Clegg et al. 1972), as well as numerical studies (Clegg 1978), have established that additional complexities develop when more than two and especially when large numbers of loci are considered simultaneously (review in Allard 1988). Methods have been proposed to extend the two-locus measure to encompass multiple loci (e.g., Bennett 1954; Hill 1974, 1975). However, two major difficulties are encountered in such extensions. First, the number of parameters, and the number of terms included in each parameter, increase rapidly as the number of loci increases, quickly leading to unmanageable complexity. Second, the structure of the entire multilocus array cannot be characterized by examining the individual parameters, because the whole is not a summation of individual effects.

Analyses of associations among multiple genes using log-linear models and likelihood-ratio tests have proved to be a useful alternative in studying multilocus genetic

* Present address: Department of Agronomy, Huazhong Agricultural University, Wuhan, People's Republic of China
** Present address: Department of Crop and Soil Environmental Sciences, VPI & SU, Blacksburg, VA 24061, USA
*** To whom correspondence should be addressed

systems (Smouse 1974; Hill 1975). More recently, new statistical techniques for discrete multivariate analysis have become available and additional properties of test criteria have been characterized (Fienberg 1980). As a result, log-linear model likelihood-ratio tests have become even more useful for analyzing complex cross-classified data.

In this study, we have adapted these discrete multivariate techniques to the analysis of associations among alleles at different loci in multilocus genetic systems. We show that this technique has three main advantages: first, it allows the experimentalist to identify terms and eliminate from the analysis those that have inconsequential effects, thereby reducing the complexity of analysis to more manageable proportions; second, it allows the experimentalist to determine explicitly the complete multilocus structure for systems involving two to many loci; and third, it allows both qualitative and quantitative evaluations of each first-order, second-order, and higher-order association term involved in the system. The results of our analysis demonstrate that the multilocus genetic structure of the barley species is hierarchical in nature and that samples of barley from different geographical regions of the world differ markedly in multilocus structure.

Materials and methods

The materials of this study were 1,032 cultivated barley (*Hordeum vulgare* L.) accessions, chosen at random from the world barley collection maintained by the United States Department of Agriculture in Beltsville/MD. Fifty-two countries, which we have grouped into ten geographical regions to facilitate analysis, were represented in the sample.

Each accession was assayed for eight enzyme loci following methods described by Kahler and Allard (1970) and Kahler et al. (1981). The loci assayed were four esterase loci (*Est1*, *Est2*, *Est3*, and *Est4*), one acid phosphatase locus (*Acp1*), one glutamate oxidase transaminase (*Got1*), and two 6-phosphogluconate dehydrogenase loci (*6-Pgd1* and *6-Pgd2*). The *Est4* locus is located on chromosome 1, the *Est1*, *Est2*, and *Est3* loci form a tightly linked cluster on chromosome 3 ($Est2^{0.0023}$ $Est1^{0.0048}$ *Est3*), and the *Acp1* and *Got1* loci are located on chromosome 6; the chromosomal locations of *6-Pgd1* and *6-Pgd2* are not certain but there is strong evidence that they are located on chromosome 5 (Soliman and Allard 1989). Ten plants were assayed for each of the first 350 accessions. The results showed that the majority of the accessions was homozygous for one allele or another; hence, only 2 (occasionally 3) plants were assayed from the ramaining 682 accessions (a 3rd plant was assayed in rare cases when the banding pattern of the first 2 plants was not the same).

The specific discrete multivariate analysis adopted (Fienberg 1980) to assess multilocus associations in the sample of 1,032 accessions from the world barley collection is a two-step process: first, likelihood-ratio tests are employed in a series to identify and eliminate terms that have nonsignificant effects in each multilocus system under investigation; second, log-linear models are constructed to fit the data for the remaining terms in the system. Models were fit in a hierarchical manner such that a higher-order term was included only when lower-order terms failed to fit the data; also, when a higher-order term was included all of its lower-order relatives were included. The "best" fitting model was chosen on the basis that it included a minimal number of terms (i.e., was maximally parsimonious) and also provided a statistically acceptable fit to the data. Two screening methods were used in model selection, stepwise selection (Goodman 1971) and partial and marginal tests (Brown 1976).

Results

The total numbers of alleles observed were 5, 5, 4, and 4 for *Est1*, *Est2*, *Est3*, and *Est4*, respectively; the observed allelic frequencies and distributions for these four loci were very similar to those of Kahler and Allard (1981). The total numbers of alleles were 6, 2, 2, and 2 for *Acp1*, *6-Pgd1*, *6-Pgd2*, and *Got1*, respectively; population surveys of these four loci have not been carried out previously. Locus *Got1* was not included in subsequent analyses because it was nearly monomorphic for one allele. The number of seven-locus gametic combinations possible is very large with the numbers of alleles observed. Consequently, the data were reduced in two ways to bring the analysis into manageable proportions and also to make our results comparable with previous studies (e.g., Weir et al. 1972, 1974; Kahler and Allard 1981).

First, the seven loci were divided into two subsets to reduce the number of gametic combinations: the first subset included three loci, *Acp1*, *6-Pgd1*, and *6-Pgd2*, designated P, G and H, respectively; the four esterase loci *Est1*, *Est2*, *Est3*, and *Est4*, designated A, B, C, and D (which are the subject of many previous studies) formed the second subset. The names and designations of the loci will henceforth be used interchangeably to designate the chromosome segments marked by the seven enzyme loci. Second, each locus was reduced to the diallelic state following the convention used by Weir et al. (1972): one allele of each locus was designated allele 1 and all the other alleles were combined into a single "synthetic" allele designated allele 2. To be consistent with previous studies, alleles 1.8, 2.7, 5.4, and 6.4 respectively, are assigned as allele 1 for *Est1*, *Est2*, *Est3*, and *Est4*; alleles 1.8, 2.7, and 6.4 were present in frequency >0.5 in our sample and, as a result, allele 1 is the most frequent allele for three (*Est1*, *2*, and *4*) of the four esterase loci. Alleles 2.0, 2.3, and null which are, respectively, the most frequent alleles of the loci *Acp1*, *6-Pgd1*, and *6-Pgd2*, were designated allele 1 of these loci. For loci with a relatively large number of alleles, the convention of collapsing usually leads to underestimating linkage disequilibria (Zouros et al. 1977; Weir and Cockerham 1978).

Individuals were cross-classified and the data was tabulated in the form of a contingency table. An accession was assigned to category 1 (as possessing allele 1) for a given locus if the observed frequency of allele 1 in the

accession was larger than 0.5, and to category 2 (as possessing allele 2) if the frequency of allele 1 in the accession was smaller than 0.5.

Data of both subsets were then subjected to log-linear analyses to determine the patterns of multilocus structure in regional samples, as well as in the world sample. We illustrate the analytic procedures by describing in some detail the analysis of the first subset of marker loci (P, G, and H), for the world sample of 1,032 barley accessions.

Worldwide multilocus structure of barley

Loci Acp1, 6-Pdg1, 6-Pgd2. The contingency table obtained for these three loci is given in Table 1. Application of stepwise selection (Goodman 1971) and partial and marginal tests (Brown 1976) as screening procedures led to selection of the model [*PG*] [*H*] as best fitting the data, in the sense of being maximally parsimonious and also giving a statistically acceptable fit to the data. The resulting likelihood-ratio statistic, G^2, which is distributed approximately as χ^2, took the value of 2.44 with three degrees of freedom (probability = 0.4867); thus, the [*PG*] [*H*] model fits the data well. The log-linear form of the model is

$$\ln m_{ijk} = u + u_{P(i)} + u_{G(j)} + u_{H(k)} + u_{PG(ij)} \quad (1)$$

for all $i, j, k = 1, 2$, subject to the constraint that

$$\sum_i u_{P(i)} = \sum_j u_{G(j)} = \sum_k u_{H(k)} = \sum_{ij} u_{PG(ij)} = 0\,, \quad (2)$$

in which, m_{ijk} is the expected number of accessions carrying the i^{th} allele of locus *P*, the j^{th} allele of locus *G*, and the k^{th} allele of the locus *H*. Goodness-of-fit to the model was further checked by examining standardized residuals. Asymptotically, a standardized residual [SR = (obs − exp)/$\sqrt{\text{exp}}$] follows a normal distribution with zero mean and unit variance (Dr. W. O. Johnson, personal communication). Thus, observed absolute values of an SR of 1.960, 2.576, or 3.291 indicate statistically significant departures from expectation for the model specified at probability levels of 0.05, 0.01, or 0.001, respectively. The expected value for each cell, from which the SR is derived, is listed in Table 1. As can be seen, all the SRs are smaller than or equal to 1.0 in absolute value, which also indicates that the [*PG*] [*H*] model fits the data well.

Table 1. Observed number, expected number, and standardized residual (top, middle, and bottom rows, respectively) for each cell under the model [*PG*] [*H*]

		6-Pgd1			
		1 *6-Pgd2*		2 *6-Pgd2*	
		1	2	1	2
Acp1	1	573	66	68	11
		572.2	66.3	70.8	8.2
		0.0	0.6	−0.3	1.0
	2	190	17	94	13
		185.5	21.5	95.9	11.1
		0.3	−1.0	−0.2	0.6

Upon the acceptance of the model, the *u*-terms were estimated. Terms such as $u_{P(i)}$ measure the main effect of each locus, i.e., provide comparisons of allelic frequencies at each locus, whereas terms such as $u_{PG(ij)}$ provide estimates of associations between alleles at two different loci. There are two estimated *u*'s for each single-locus effect (e.g., locus 1) related to each other by $u_{1(1)} = -u_{1(2)}$. There are four estimated *u*'s for each pairwise association (e.g., loci 1 and 2) related by $u_{12(11)} = -u_{12(12)} = -u_{12(21)} = u_{12(22)}$ because of constraint (2). Thus, the estimated *u*'s of a particular order differ only in algebraic signs. To relate the *u* terms to the traditional two-locus measures, *r* (Wright 1933) and *D* (Geiringer 1944), let p_i be the frequency of the most frequent allele of the i^{th} locus and p_{ij} the frequency of gametic type *ij* (allele *i* of the first locus and allele *j* of the second locus). It can be shown that $u_{12(11)} = \frac{1}{4} ln(1 + D/p_{12}\, p_{21}) = \frac{1}{4} ln(1 + r\, p_1 \sqrt{(1-p_1)\, p_2\, (1-p_2)/p_{12}\, p_{21}})$, and that $du/dD > 0$ and $du/dr > 0$. Thus, $u_{12(11)}$ is a monotonically increasing function of *D* and *r*, and $u_{12(11)} = 0$ only if $D = 0$ or $r = 0$. The large sample distribution of a standardized *u*, e.g., $u_{P(i)}$, under the assumption $u_{P(i)} = 0$, is $n(0, 1)$. The estimated standardized *u*'s, therefore, allow statistical inferences concerning the effect of each term.

According to the [*PG*] [*H*] model, identified above as giving the "best" fit to the three-locus data set, locus *Acp1* (*P*) is associated with locus *6-Pgd1* (*G*), and locus *6-Pgd2* (*H*) is independent of both loci *Acp1* and *6-Pdg1*. The absolute value of the estimated standardized *u*'s for the *PG* association is 8.489*** ($P < 0.001$), indicating a highly significant association effect; the algebraic signs (not listed) indicate further that allele 1 of *Acp1* is associated with allele 1 of *6-Pgd1*, and vice versa. It can also be calculated from the data of Table 1 that the observed numbers of gametic types 11 (allele 1 of *Acp1* and allele 1 of *6-Pdg1*), 12, 21, and 22 are 639, 79, 207, and 107, respectively, whereas under the hypothesis of no pairwise associations the expected numbers are 589, 129, 257, and 57. This also shows that gametic types 11 and 22 are much in excess and that gametic types 12 and 21 are in substantial deficiency. A χ^2 test of this association gives $\chi^2\,[3] = 78.882$, $P < 0.001$, in parallel with analysis of absolute values of the standardized *u*'s.

The four esterase loci. The model that best fits the data for these four loci is the saturated model, [*ABCD*], under which the expected and observed numbers for each cell are equal and all the standardized residuals are zero. The

Table 2. The absolute value for each of the estimated standardized u's for all the association effects under the model *ABCD*

Effect	Absolute u-value	Effect	Absolute u-value
AB	6.386***	*ABC*	1.484
AC	4.062***	*ABD*	0.600
AD	0.517	*ACD*	2.815**
BC	1.808	*BCD*	0.586
BD	2.097*		
CD	3.823***	*ABCD*	3.479***

*, **, *** Significant at probability levels of 0.05, 0.01, and 0.001, respectively

log-linear form of the model is

$$\begin{aligned}\ln m_{ijkl} = {} & u + u_{A(i)} + u_{B(j)} + u_{C(k)} + u_{D(l)} + u_{AB(ij)} + u_{AC(ik)} \\ & + u_{AD(il)} + u_{BC(jk)} + u_{BD(jl)} + u_{CD(kl)} \\ & + u_{ABC(ijk)} + u_{ABD(ijl)} + u_{ACD(ikl)} + u_{BCD(jkl)} \\ & + u_{ABCD(ijkl)}\end{aligned} \tag{3}$$

for all $i, j, k, l = 1, 2$.

This model suggests that: (i) allelic state at any given locus depends on allelic states at each of the other three loci; (ii) the association between any two loci cannot be specified without knowing the associations of the other two loci; and (iii) the relationship of any three loci is influenced by the fourth locus. The strong interdependence among these four loci further suggests that caution is necessary in interpreting the individual terms included within the [*ABCD*] model. Nevertheless, the relative importance of a particular term can be assessed by examining the estimated u-terms. The absolute values for each of the standardized u-terms under this model are listed in Table 2. The u terms for four among the six pairwise association effects are significant; the largest is the association between *A* and *B* (standardized u equal to 6.386***), followed by *A* and *C* (4.062***), *C* and *D* (3.823***), and *B* and *D* (2.097*). Among the third-order associations, only the effect of *ACD* is significant. In addition, the absolute value of the standardized u for the fourth-order interaction (3.479***) term is highly significant.

Combined analysis of the two subsets

A combined analysis of the two subsets was performed by selecting a five-locus data set (loci *A*, *C*, *D*, *P*, *H*) from the seven-locus data to reduce the number of gametic combinations to manageable proportions. The selection was result-guided. In the first subset, the strong association between *Acp1* (*P*) and *6-Pgd1* (*G*) indicates that allelic state at either of these two loci can be predicted in large part from information about allelic state of the other locus, and vice versa. Thus, loci *Acp1* (*P*) and *6-Pgd2* (*H*) were chosen from this subset. Similarly, the highly significant effects of *AB*, *ABCD*, and the significant effect of *BD*, suggested the choice of the loci *A*, *C*, and *D* from the second set. The "best" fitting model for the five-locus set is [*CD*] [*DH*] [*ADP*] [*ACPH*], for which G^2 is 13.37 with ten degrees of freedom, yielding a probability of 0.2037. The observed and expected numbers and the standardized residual for each cell are given in Table 3. Only 2 among the 32 standardized residuals (approximately 6%) are larger than 1.96, which supports selection of this model (the log-linear form of this model is not given).

Table 3. Observed and expected numbers and standardized residual (top, middle, and bottom rows, respectively) for each cell under the model [*CD*] [*DH*] [*ADP*] [*ACPH*]

				Acp1			
				1		2	
				6-Pgd2		*6-Pgd2*	
				1	2	1	2
Est1 1	*Est3* 1	*Est4* 1		51	0	30	4
				51.1	0.9	28.2	3.6
				0.0	−0.9	0.3	0.2
			2	13	1	4	0
				12.9	0.1	5.8	0.4
				0.0	2.7**	−0.7	−0.6
	Est3 2	*Est4* 1		103	15	54	12
				101.9	15.1	56.6	11.6
				0.1	0.0	−0.3	0.1
			2	53	4	27	2
				54.1	3.9	24.4	2.4
				−0.2	0.1	0.5	−0.3
2	*Est3* 1	*Est4* 1		214	33	32	0
				210.9	32.3	36.1	0.8
				0.2	0.1	−0.7	−0.9
			2	33	2	19	1
				36.1	2.7	14.9	0.2
				−0.5	−0.4	1.1	2.1*
	Est3 2	*Est4* 1		124	19	68	8
				128.0	18.8	63.3	7.8
				−0.4	0.1	0.6	0.1
			2	50	3	50	3
				46.0	3.2	54.7	3.2
				0.6	−0.1	−0.6	−0.1

*, ** Significant at probability levels of 0.05 and 0.01, respectively

It can be seen from the model that pairwise associations between *C* and *D* and between *D* and *H* are consequential, but that the two pairs are independent of each other and independent of all other association terms incorporated in the model. A third-order association exists among *A*, *D*, and *P*, but the association among these three loci is not influenced by the other terms in the model. The same is the case for the fourth-order associa-

tion. The absolute values for the estimated standardized u's for all of the associations in the model are given in Table 4. Among the second-order association terms, the association between *C* and *D* is highly significant (4.543***); allele 1 of *Est3* and allele 1 of *Est4* are associated. The absolute u of the *DH* association is significant (2.560*) and the algebraic signs show that allele 1 of *Est4* and allele 2 of *6-Pgd2* go together.

All other second-order association terms are nested in the higher-order associations and, consequently, they cannot be interpreted independently. The effect of the association *ADP*, which has an absolute standardized u-value of 3.407***, is the largest among the third-order terms. The signs of the estimated u's (not listed) indicate that, after allowing for all lower-order associations among the loci *A*, *D*, and *P*, there are still deficiencies of the triplets 111 (allele 1 of *A*, *D*, and *P*), 122, 212, and 221 and excesses of the remaining triplets, which can only be accounted for by this third-order association. All the remaining third-order terms are nested in the fourth-order association. The absolute standardized u for the fourth-order term *ACPH* is 2.089*. By the same token, when all of the lower-order associations among loci *A*, *C*, *P*, and *H* were partitioned out, the quadruplets 1111, 1122, 1221, 2112, 2121, 2211, and 2222 were still slightly in excess, and the remaining eight quadruplets were in deficiency. Thus, pairwise associations (e.g., *DH*) and also third- as well as fourth-order associations exist among the loci of the two subsets. Further, very strong associations were found between *P* and *G* within the first subset (standardized $u=8.489$***), and between *A* and *B* ($u=6.386$***) and *ABCD* ($u=3.478$***) within the second subset. Thus, the membership of loci *P* and *A* in the *ACPH* complex indicates that loci *G* and *B* are also members of this complex. Hence the complex includes at least six loci, *ABCPGH*.

Table 4. The absolute value for each of the estimated standardized u's for all the association effects under the model [*CD*] [*DH*] [*ADP*] [*ACPH*]

Effect	Absolute u-value	Effect	Absolute u-value
AC	2.230*	*ACP*	3.382***
AD	0.950	*ACH*	0.602
AP	2.876**	*ADP*	3.407***
AH	0.045	*APH*	2.696***
CD	4.543***	*CPH*	0.023
CP	1.333		
CH	2.594**		
DP	2.112*	*ACPH*	2.089*
DH	2.560*		
PH	0.156		

*, **, *** Significant at probability levels of 0.05, 0.01 and 0.001, respectively

Table 5. The selected model for each of the seven regions for loci *P*, *G*, and *H*

Region	Model	G^2	*df*	*P*
N. Europe	[*PG*] [*H*]	0.84	3	0.8394
S. Europe	[*PG*] [*H*]	5.87	3	0.1180
S. W. Asia	[*PG*] [*PH*]	4.47	2	0.1071
M. S. Asia	[*PG*] [*GH*]	1.78	2	0.4104
E. Asia	[*PG*] [*GH*]	3.69	2	0.1584
Ethiopia	[*PGH*]	0.00	0	
N. America	[*P*] [*G*] [*H*]	2.88	4	0.5789

Multilocus structure within different geographical regions

Our sample included 70 or more accessions from seven of the geographical regions: North Europe (Finland, Sweden, Denmark, England, Scotland, Ireland, Poland, Belgium, Norway, and Germany), South Europe (France, Portugal, Spain, Czechoslovakia, Switzerland, Italy, Austria, Hungary, Romania, Yugoslavia and Greece); Southwest Asia (Cyprus, Turkey, Iraq, Israel, Saudi Arabia, Syria, Yemen, and Jordan); Middle South Asia (Iran, Afghanistan, India, and Nepal); East Asia (China, Korea, and Japan), North America (USA and Canada) and the country Ethiopia. Sample sizes ≥ 70 were large enough to justify regional analyses within the *P*, *G*, *H* and *A*, *B*, *C*, *D* subsets, but not combined analyses over both subsets (seven loci). Consequently, analyses of multilocus structure within regions were carried out only within subsets.

Loci Acp1, 6-Pgd1, and 6-Pgd2

The model selected for these three loci for each of the seven regions is given in Table 5, from which it can be seen that five different models are required to fit the seven data sets. The model [*PG*] [*H*] fits accessions from both North Europe and South Europe; thus, loci *Acp1* (*P*) and *6-Pgd1* (*G*) are associated and locus *6-Pgd2* (*H*) is independent of *P* and *G* in these two regions. In both regions, algebraic signs of the estimated standardized u's (not given) indicate that alleles 1 of *Acp1* and *6-Pgd1* go together and that non-1 alleles of *Acp1* and *6-Pdg1* go together, i.e., that gametic types 11 and 22 are in excess and gametic types 12 and 21 are in deficiency.

The Southwest Asian accessions fit the model [*PG*] [*PH*], indicating pairwise associations between *Acp1* and *6-Pgd1* and between *Acp1* and *6-Pgd2*. The absolute standardized u-value for *PG* is 2.559**, and algebraic signs indicate that allele 1 of *Acp1* is associated with allele 1 of *6-Pgd1*. The absolute standardized u-value for *PH* is 1.603 (NS), which indicates that the *PH* association, although necessary for fitting the model, is nevertheless not statistically singificant in itself. Thus, the difference between the model for Southwest Asia [*PG*]

[*PH*] and that for Central and South Europe [*PG*] [*H*] is quantitative rather than qualitative.

The model [*PG*] [*GH*] fits both Middle South Asian and East Asian accessions, showing that *Acp1* and *6-Pgd1* and also *6-Pgd1* and *6-Pgd2* are associated. The absolute standardized *u*-values are 1.966* and 2.438* for *PG* and *GH*, respectively, in Middle South Asia, and 2.581* and 2.347* in East Asia. In both regions, allele 1 of *Acp1* goes with allele 1 of *6-Pgd1* and allele 1 of *6-Pgd1* goes with allele 1 of *6-Pgd2*.

Quite different features were found for the accessions from North America and Ethiopia. The model for the North American accessions, [*P*] [*G*] [*H*], is one of complete independence among these three loci (Table 5), whereas the situation is the opposite for the Ethiopian accessions, which fit the model of a third-order association, [*PGH*]. The absolute standardized *u*'s under this model are 2.308* for *PG*, 0.070 for *PH* and 2.234* for *GH*, respectively, indicating significant associations between *PG* and *GH*. The absolute standardized *u* for the third-order term [*PGH*] is 1.545 (NS). This again indicates that the difference between the model [*PGH*] and that for Middle South Asia and East Asia [*PG*] [*GH*] is quantitative rather than qualitative.

The four esterase loci

The best fitting models for *Est1*, *Est2*, *Est3*, and *Est4* for each of the seven geographical regions are given in Table 6. The model for North European accessions is [*ABD*] [*AC*] [*BCD*]. The estimated absolute standardized *u* for *AC* is 2.201*; allele 1 of *Est1* is associated with allele 1 of *Est3*. The absolute standardized *u* for *ABD* is 2.517*; the individual *u* values indicate that, after allowing for all the pairwise associations, there are still excesses of gametic types 112 (1-allele of *A*, 1-allele of *B*, and non-1-alleles of *D*), 121, 211, and 222, and deficiencies of the remaining types. The absolute *u* for *BCD* is 3.258**; gametic types 111, 122, 212, and 221 are in excess and the remaining types are in deficiency. The model for South Europe is one of three pairwise associations, [*AB*] [*BC*] [*CD*]. The absolute standardized *u*'s for *AB*, *BC*, and *CD* are 1.927, 1.842, and 2.044*, respectively; thus, all three associations (allele 1 of *A* associated with non-1-alleles of *B*, allele 1 of *B* with allele 1 of *C*, and allele 1 of *C* with allele 1 of *D*, and vice versa) are weak. The data of Southwest Asia fit the model [*AB*] [*AD*] [*C*]. The absolute standardized *u*'s are 2.844** and 4.208*** for *AB* and *AD*, respectively (allele 1 of *Est1* associated with non-1-alleles of *Est2*, and allele 1 of *Est1* with allele 1 of *Est4*). The model for Middle South Asian accessions is [*AB*] [*C*] [*D*]. The absolute standardized *u* for *AB* is 4.459*** (allele 1 of *A* associated with non-1-alleles of *B*). The model [*ABD*] [*AC*] fits the data for the East Asian accessions. The absolute standardized *u* for *AC* is 2.302* (allele 1 of *A* associated with non-1-alleles of *C*). The absolute value for the three-way *ABD* association is 2.851**. The structure for these three loci is such that, after partitioning out all the pairwise associations, the gametic types 111, 122, 212, and 221 are in deficiency and the other types are in excess. There are four pairwise associations in the model that fits the Ethiopian accessions. The absolute standardized *u* for *BC* (1.758) is not significant, although it is necessary to include it in the model for a good fit. The absolute standardized *u*'s for the *AB*, *AC*, and *AD* are 2.357*, 2.449*, and 3.135**, respectively (allele 1 of *A* goes with non-1-alleles of *B*, allele 1 of *A* with non-1-alleles of *C* and allele 1 of *A* with non-1-alleles of *D*). The model of complete independence among these four loci [*A*] [*B*] [*C*] [*D*] fits the North American accessions.

It should be noted that, although the associations between locus pairs *AC* and *AD* appear to be the same in Table 6, the direction of association, as shown by the algebraic signs of these terms, is different in different geographical regions; as examples, allele 1 of locus *A* is associated with allele 1 of locus *C* in North Europe, with non-1-alleles of *C* in East Asia and Ethiopia, and with allele 1 of *D* in North Europe and Southwest Asia, but with non-1-alleles of *D* in Ethiopia.

Table 6. The selected model for each of the seven regions for the loci *A*, *B*, *C*, and *D*

Region	Model	G^2	*df*	*P*
N. Europe	[*ABD*] [*BCD*] [*AC*]	2.70	3	0.4409
S. Europe	[*AB*] [*BC*] [*CD*]	8.17	8	0.4175
S. W. Asia	[*AB*] [*AD*] [*C*]	13.92	9	0.1254
M. S. Asia	[*D*] [*AB*] [*C*]	12.31	10	0.2649
E. Asia	[*ABD*] [***AC***][a]	9.54	6	0.1455
Ethiopia	[*AB*] [***AC***] [***AD***] [*BC*]	10.66	7	0.1542
N. America	[*A*] [*B*] [*C*] [*D*]	15.31	11	0.1688

[a] Boldface terms indicate that allelic associations are in reverse direction in some regions. For the *AC* association, gametic types 11 and 22 are in excess in N. Europe but in deficiency in Ethiopia and E. Asia; for the *AD* association, 11 and 22 are in excess in S. W. Asia and N. Europe, but in deficiency in Ethiopia

Discussion

We have applied a discrete log-linear multivariate technique to data of seven enzyme loci from a worldwide sample of 1,032 barley accessions in an attempt to estimate precisely the extent of multilocus association in the barley species. The analysis showed that four esterase loci (coded *A*, *B*, *C*, and *D*) are organized into complex multilocus associations that are held together by a number of two- and three-locus interactions, and by the four-locus interaction among the marker loci (and/or by other loci

located within the chromosome segments marked by the marker loci; however, see Allard 1988). Three additional loci, *Acp1*, *6-Pgd1*, and *6-Pgd2* (coded *P*, *G*, and *H*), which had not previously been subjected to population genetic analysis, were also found to be structured on a worldwide basis: loci *P* and *G* are associated, but locus *H* is independent of both *P* and *G*. Combined analysis of the *ABCD* and *PGH* subsets of loci showed that multilocus associations of various orders of complexity structure the seven loci into a complex that includes at least six of the loci. The barley genome is thus hierarchically structured: many loci interact in pairs and become structured into duplexes; duplexes frequently interact with other loci and other duplexes to form triplexes or quadriplexes, which merge here and there in the genome to form higher-ordered complexes of loci.

Our results established that the marker loci of both the *PGH* and *ABCD* subsets are organized into multilocus sets within six of the seven geographical regions investigated, and that multilocus structure is often not the same in the different regions. Marker loci *P* and *G* are associated in the same way in all six regions (11 and 22 gametic associations were consistently in excess and the 12 and 21 gametic associations were consistently in deficiency). However, other association terms of the *PGH* subset often differ from region to region. Among European barley accessions, the model is [*PG*] [*H*], i.e., locus *H* is independent of locus *P* and *G*. The model for Southwest Asia can be viewed as either [*PG*] [*H*] or [*PG*] [*PH*] because the *PH* association is not, in itself, significant in that region. However, the *GH* association is significant in Middle South Asia and East Asia, and in both regions the association is such that gametic types 11 and 22 are in excess and 12 and 21 are in deficiency. In Ethiopia there is a weak three-locus association.

It should be pointed out that "allele 2" of five of the seven loci assayed is a "synthetic" allele made up by lumping two or more non-1-alleles. Lumping alleles in this way may cause loss of information from infrequent alleles and as a result, the 2-alleles of these loci may differ from region to region. This is unlikely, however, to cause serious problems in our analyses or interpretations because: (i) worldwide, the most frequent allele at each of the five loci is present in frequency ≥ 0.50, whereas the frequency of the next most frequent allele of each locus is much lower ($<$ one-half the frequency of allele 1 in four of the five cases). This situation holds for nearly every locus in every region, which suggests that our dichotomization of alleles at each locus is into widely adapted "wild type" versus locally adapted "minor" alleles; and (ii) our interpretation of the association is limited to the relationship of the 1-allele with non-1-alleles of the loci studied.

Among the four esterase loci, loci *A* and *B* were related through the two-locus *AB* associations in four regions (Southwest Asia, South Europe, Middle South Asia, and Ethiopia) and through the three-locus *ABD* association in two regions (North Europe and East Asia). The association between loci *A* and *B* was in the same direction in all of the six regions (the 12 and 21 gametic types in excess and 22 and 11 gametic types in deficiency). Loci *A* and *C* were unassociated in three regions (Southwest Asia, Middle South Asia, and South Europe) and associated through the two-locus *AC* term in the other three regions. However, the association was in different directions in these latter regions: gametic types 11 and 22 were predominant in North Europe, whereas gametic types 12 and 21 were in excess in East Asia and Ethiopia. Loci *A* and *D* were unassociated in two regions (South Europe, Middle South Asia), associated through the second-order *AD* term in two regions (Southwest Asia, Ethiopia), and through the third-order *ABD* term in the two remaining regions (North Europe and East Asia). Allelic combinations 11 and 22 of the *AD* association were in excess in North Europe and Southwest Asia but in deficiency in Ethiopia.

The loci of both the *PGH* and *ABCD* subsets were found to be unassociated in the North American sample. The barleys of North America are either recent introductions or they are derived from recent introductions from various parts of the world, and they differ in multilocus structure in different ecogeographical regions of the continent. The random sample of the present experiment, which included a few accessions from each of many regions, was a conglomerate of accessions adapted to many different habitats and it is not surprising that the sample studied had no structure of its own.

The general pattern that emerges is that the multilocus structure of Southwest Asian accessions is intermediate for both the *PGH* and *ABCD* subsets, and that the differences from the Southwest Asian structure increase progressively across the Eurasian-African landmass in each direction with increasing distance from Southwest Asia. This is consistent with the proposition that barley was first domesticated in Southwest Asia and later spread to other regions of the world (Harlan and Zohary 1966).

It is clear from our results that Ethiopian and East Asian barleys differ from the barleys of the rest of the world in both gametic combinations (data not shown) and in multilocus structure. this is also consistent with previous findings from studies of many characteristics of barley (Ward 1962; Harlan 1979). Such differences have been presumed to be due to a combination of various factors such as historical patterns of migration, geographical isolation, and both natural and artificial selection leading to local adaptation (Ward 1962; Harlan 1979).

Agricultural history shows that exchanges of crop germ plasm due to human activities occurred much more

frequently among Southwest Asia, Middle South Asia, and Europe than between these regions and East Asia and Ethiopia. Our results show that barleys from Southwest Asia, Middle South Asia, and Europe are more similar to each other in multilocus structure than they are with barleys from the other two regions. However, the differences in multilocus structure among barleys of Southwest Asia, Middle South Asia, North Europe, and South Europe are also substantial (Tables 5 and 6). Within the *ABCD* subset, e.g., barleys differ among these regions by several association terms. Assuming that barleys of Middle South Asia and Europe were derived from Southwest Asian barley, the associations that are present in the European barleys but not in Southwest Asian barley are novel. Examples of such novel multilocus associations are *BC* and *CD* in South Europe and *ABD*, *BCD*, and *AC* in North Europe. Based on neutral models, which assume historical migrations, genetic drift, and the effects of sampling, it is expected that the novel associations that developed in the derivative populations would be more intense for the linked than for unlinked loci. Thus, for the four esterase loci it is expected that associations would be closest among the very tightly linked *A*, *B*, and *C* loci and less intense between locus *D* and loci *A*, *B*, and *C*, i.e., that the intensity of association would be ordered $AB \simeq AC \simeq BC > AD \simeq BD \simeq CD$. By the same argument, the largest three-way association is expected to be among the tightly linked *ABC* loci. This, however, was not the case for either the South European sample in which *CD* is significant but *AC* is not significant, nor in the North European sample in which the *ABC* association is not significant but the associations of *ABD* and *BCD* are significant. This provides additional support for invoking selection as the major factor in explanation of the differences in the observed multilocus structures among these regions.

Acknowledgements. The authors thank Dr. P. Smouse for his helpful comments on an earlier version of this manuscript. This work was supported in part by National Science Foundation Grant BSR-10869 and by U.S. Public Health Service Grant (NIHGM-32429). M.A.S.M. acknowledges contributions from the College of Agriculture and Life Sciences, VPI & SU.

References

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

Bennett JH (1954) On the theory of random mating. Ann Eugen 18:311–317

Brown AHD, Feldman M, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics 96:523–536

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

Clegg MT (1978) Dynamics of correlated genetic system. II. Simulation studies of chromosomal segments under selection. Theor Popul Biol 13:1–23

Clegg MT, Allard RW, Kahler AL (1972) Is the gene the unit of selection? Evidence from two experimental plant populations. Proc Natl Acad Sci USA 69:2474–2479

Fienberg SE (1980) The analysis of cross-classified categorical data. MIT Press, Cambridge/MA

Geiringer H (1944) On the probability theory of linkage in Mendelian heredity. Ann Math Stat 15:27–57

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

Harlan JR (1979) On the origin of barley. In: Barley: origin, botany, culture, winter hardiness, genetics, utilization, pests. USDA-SEA Agricultural Handbook No. 338, Washington/DC, pp 10–36

Harlan JR, Zohary D (1966) Distribution of wild wheats and barley. Science 153:1074–1080

Hill WG (1974) Disequilibrium among several linked neutral genes in finite populations. I. Mean changes in disequilibrium. Theor Popul Biol 5:366–392

Hill WG (1975) Tests for association of gene frequencies at several loci in random mating diploid populations. Biometrics 31:881–888

Jana S, Zhang Q, Saghai Maroof MA (1989) Influence of environments on the development of multivariate structures in a barley composite cross at three locations. Genome 32:40–45

Kahler AL, Allard RW (1970) Genetics of isozyme variants in barley. I. Esterases. Crop Sci 10:444–448

Kahler AL, Allard RW (1981) Worldwide patterns of genetic variation among four esterase loci in barley (*Hordeum vulgare* L.). Theor Appl Genet 59:101–111

Kahler AL, Heath-Pagliuso S, Allard RW (1981) Genetics of isozyme variants in barley. II. 6-phosphogluconate dehydrogenase, glutamate oxalate transaminase, and acid phosphatase. Crop Sci 21:536–540

Soliman KM, Allard RW (1989) Chromosome locations of additional barley enzyme loci identified using wheat-barley addition lines. Plant Breed 102:177–181

Smouse P (1974) Likelihood analysis of recombinational disequilibrium and multiple-locus genetic frequencies. Genetics 76:557–565

Ward DJ (1962) Some evolutionary aspects of certain morphologic characters in a world collection of barley. USDA Tech Bull 1276:1–112

Weir BS, Allard RW, Kahler AL (1972) Analysis of complex allozyme polymorphisms in a barley population. Genetics 72:505–523

Weir BS, Allard RW, Kahler AL (1974) Further analysis of complex allozyme polymorphisms in a barley population. Genetics 78:911–919

Weir BS, Cockerham CC (1978) Testing hypothesis about linkage disequilibrium with multiple alleles. Genetics 88:633–642

Wright S (1933) Inbreeding and recombination. Proc Natl Acad Sci USA 19:420–433

Zouros E, Golding GB, Mackay TFC (1977) The effect of combining alleles on detecting linkage disequilibrium. Genetics 85:545–556