THE EFFECT OF SELECTION ON ESTERASE ALLOZYMES IN A BARLEY POPULATION[1]

R. W. ALLARD[2], A. L. KAHLER[2] AND B. S. WEIR[3,4]

University of California, Davis, California 95616

Manuscript received December 27, 1971
Revised copy received June 26, 1972

ABSTRACT

Changes in gene and genotypic frequencies at four esterase loci were monitored over 25 generations in Composite Cross V, an experimental population of barley, to obtain experimental evidence concerning the balance of forces responsible for: (1) the marked differences in allelic frequencies among barleys from different ecogeographical regions of the world; and (2) the extensive allelic variation found within local populations of barley. Analyses of the highly significant changes in allelic frequencies which occurred in CCV showed they were due to directional selection favoring particular alleles and not to mutation, migration or genetic drift. The results show that intense balancing selection, featuring consistent excesses of heterozygotes, also occurred in CCV. It is concluded that among the factors of neo-Darwinian evolution, natural selection plays the predominant role in determining the observed patterns of allelic variation in the barley species as a whole.

STUDIES of allozymes in the world collection of barley (*Hordeum vulgare* L.) maintained by the U. S. Department of Agriculture (KAHLER and ALLARD 1970; and in preparation) have shown that there is extensive allelic variation within many of the accessions of this collection, as well as pronounced differences in allelic frequencies among barleys from different ecogeographical regions of the world. According to the neo-Darwinian theory of evolution, the kind and amount of genetic variability within and among populations is governed mainly by the combined effects of mutation, selection, migration, population size and mating system. The present investigation was undertaken to determine the relative effects of these factors on the single-locus population dynamics of four esterase loci in Composite Cross V (CCV), an experimental population of barley. The results show that balancing selection with direction of selection different for different alleles is responsible for the changes in allelic frequencies which occurred over generations in this population, and for the maintenance of variability within the population. Extension of these results to the barley species as a whole leads us to the conclusion that the observed patterns of variability in this species are also due in large part to a combination of directional and balancing selection.

[1] Supported in part by grants from the National Institutes of Health (GM-10476) and from the National Science Foundation (GB 13213).

[2] Department of Genetics.

[3] Department of Agronomy and Range Science.

[4] Present address: Department of Mathematics, Massey University, Palmerston North, New Zealand.

MATERIALS AND METHODS

CCV was developed by the late H. V. HARLAN and his associates by intercrossing 30 barley varieties representing all of the major barley growing regions of the world (SUNESON 1956). In 1937 the parents were crossed in pairs and during the next three years the F_1 hybrids of each cycle were pair crossed to produce a single hybrid stock involving all 30 parents. This hybrid stock was planted in a plot in 1941 and allowed to reproduce by natural self pollination, giving the F_2 generation. The F_3 generation was grown in 1942, the F_4 in 1943, and so on, from random samples of seeds taken from the harvest of the previous year. The plot was managed according to usual agricultural practice and no conscious selection was practiced at any time. However, it should be noted that, in the 25 years that CCV has been grown at Davis, California, temperature, rainfall and many other factors of the environment have fluctuated sharply from year to year, and they have also fluctuated in longer cycles. Population size was approximately 15,000 in the earliest generations (F_2, F_3) but thereafter it was increased to more than 200,000 individuals/generation. We are indebted to Dr. C. A. SUNESON who, in 1958, made available to us stored seed of the F_3, F_{13}, and F_{18} generations of this population. The earliest generation for which viable seeds are still available is the F_4 generation.

The materials studied electrophoretically were obtained from seedlings grown from random samples of seeds taken from 10 different generations of the population, including three successive early generations (F_4–F_6), four successive intermediate (F_{14}–F_{17}) and three successive late (F_{24}–F_{26}) generations. The seeds were germinated at 22°C in a laboratory germinator. When the seedlings were 7 days old, plumules were excised above the coleoptile, crushed in a Petri dish, and the crude squeezate absorbed on a filter paper wick. Electrophoretic techniques followed those described in detail by KAHLER and ALLARD (1970). Four loci were assayed in the present study, Esterase loci EA, EB, EC and ED, henceforth designated A, B, C and D. The A, B, and C loci are tightly linked with each other ($B \leftarrow 0.0023 \pm 0.0007 \rightarrow A \leftarrow 0.0048 \pm 0.0008 \rightarrow C$), whereas the D locus segregates independently of these three loci. Among the 30 parents of CCV loci A and D are each represented by four alleles and loci B and C each by three alleles. Homozygotes are single banded and heterozygotes are double banded for loci A and B whereas homozygotes are double banded and heterozygotes are quadruple banded for loci C and D. Designation of these alleles is based on migrational distance in centimeters from the origin, e.g. $A^{0.2}$, $B^{1.6}$, $C^{4.4}$ (alleles producing double bands in homozygous condition are designated by the leading band). One of the four alleles at the *D* locus is a recessive null allele (D^N). All genotypes can be identified directly from starch gels, with the following exceptions (KAHLER and ALLARD 1970): (1) heterozygotes involving the recessive D^N allele, which cannot be distinguished from banded homozygotes; (2) heterozygotes involving the $A^{2.6}$ allele in genotypes which are either homozygous or heterozygous for the $B^{2.7}$ allele. The $B^{2.7}$ allele produces a heavy staining band which obscures the lighter staining $A^{2.6}$ band; hence $A^{0.2}A^{2.6}$, $A^{1.0}A^{2.6}$ and $A^{1.8}A^{2.6}$ heterozygotes cannot be distinguished from the corresponding $A^{0.2}$, $A^{1.0}$ or $A^{1.8}$ homozygotes in the presence of $B^{2.7}$. Heterozygotes involving these alleles are scored as homozygous for the alternative alleles; consequently these difficulties in classification, and the method of scoring they necessitate, lead to underestimation of heterozygosity and also to downward bias in estimates of the frequencies of the D^N and $A^{2.6}$ alleles. This is not a serious problem because the $A^{2.6}$ and D^N alleles are both low in frequency.

To facilitate scoring, a standard genotype, developed by 40 generations of enforced selfing, and invariant in banding pattern, was included in each gel. Also all gels were read independently by two or more persons, including one of the authors (ALK), to ascertain objectivity in scoring. Since results were identical in all cases it seems unlikely that scoring errors or biases could have had more than inconsequential effects in this study.

Many alleles in addition to those represented in CCV have been found at each of these four loci in the world collection of barley. At present, the numbers of alleles (including those represented in CCV) that have been established by segregation tests are, respectively, 5, 6, 5, and 6 for loci A, B, C, and D. Consistent and repeatable differences in migrational distances obtained in tissue samples composited from two seedlings indicate that the total number of alleles at each

of these loci is considerably larger yet, perhaps 2- or 3-fold larger, in the 13,000 items in the world barley collection. Thus spontaneous mutations affecting electrophoretic mobility have not only occurred at these loci but have also become established in populations. However, in a study of mutation rates now in progress, more than 80,000 individuals, representing more than 160,000 possible mutational events per locus, have been examined to date for mutation at these loci. Since no mutants have been found, these four loci do not appear to be unusually mutable.

RESULTS

The parents of CCV: Each of the parents of CCV is an entry in the world barley collection maintained by the U. S. Department of Agriculture. The introduction of items into the collection is on the basis of a small sample of seeds and thereafter each entry is maintained by growing a short row when seed supply is nearly exhausted (Personal communication, Dr. J. C. CRADDOCK, Crops Research Division, Agricultural Research Service, U. S. Department of Agriculture). Thus, due to founder effect at the time of introduction into the collection, and severe genetic drift arising from recurring drastic reductions in population size, the entries in the collection are expected to be less variable genetically than the populations from which they were derived. To determine the extent of allozyme variability within and among the 30 parents of CCV, 54 or more individuals of each parent stock were assayed electrophoretically (the average number assayed was 152 per parent for loci A, B and C and 90 for locus D). The results have been presented in detail elsewhere (KAHLER and ALLARD 1970) but they are given here in summary form because of their relevance to the present study. Twelve of the 30 parents were monomorphic for one or the other of the three A locus alleles; however, 12 were polymorphic with two A locus alleles present and six were polymorphic with three alleles present. At the B locus 21 of the 30 parents were monomorphic and the majority were fixed for the $B^{2.7}$ allele; this is expected because this allele is very frequent on a worldwide basis. However, nearly one-third of the parents were polymorphic for 2 or 3 alleles at this locus. The C and D loci were the most extensively polymorphic; two-thirds or more of the parents were polymorphic, including many polymorphic for three alleles and three polymorphic for four alleles. Considering all four loci simultaneously only three of the parents were monomorphic at all four loci, whereas six were polymorphic for one locus, six for two loci, six for three loci and nine for all four loci. The polymorphic parents included heterozygotes in approximately the proportions expected in a species which is about 99% self pollinated. It is therefore apparent that many of the parents of CCV are extensively polymorphic. This result is not consistent with adaptive neutrality, since adaptively neutral alleles are expected to become fixed very rapidly in such small populations. Maintenance of such polymorphisms by migration also seems unlikely; each parent is remarkably uniform and distinctive morphologically so that the rare hybrids which occur between entries in the world collection are easily recognized and rogued. The extensive polymorphism observed appears, however, to be consistent with certain types of balancing selection to be considered later.

Gene frequencies and percent heterozygosity in CCV: Tables 1, 2, 3 and 4 give

TABLE 1

*Allelic frequencies and percent heterozygosity at the A locus**

Generation	Number of plants	$A^{0.2}$	$A^{1.0}$	$A^{1.8}$	$A^{2.6}$	Percent heterozygosity
Initial†	4569	.1061	.3484	.5149	.0306	—
4	1390	.1072	.3482	.5277	.0169	7.12
5	1486	.1003	.3640	.5185	.0172	4.31
6	1006	.0944	.3613	.5224	.0219	3.29
14	1931	.1142	.3729	.5046	.0083	2.44
15	2843	.0922	.3765	.5222	.0091	2.78
16	2369	.1097	.3624	.5241	.0038	1.48
17	2461	.0941	.3393	.5626	.0040	0.65
24	4802	.0800	.3004	.6087	.0109	2.50
25	3967	.0651	.3113	.6042	.0194	1.61
26	3083	.0478	.3258	.6098	.0166	1.30

* Standard errors < 0.01.
† Allelic frequencies inferred from those of the 30 parents.

gene frequencies for Loci A, B, C and D, respectively, in each of 10 generations of CCV, together with the size of sample from which each frequency was estimated. Since viable seeds of the initial (F_1) generation are no longer available, gene frequencies in that generation were inferred from the gene frequencies in the 30 parents of CCV, assuming no selection took place during the intercrossing phase of the development of the population. Gene frequencies for the other generations were computed directly from observed genotypic frequencies (genotypic frequencies are not reported since they can be computed from data of Tables 1–4 and the Fixation Indices given in Table 5). The number of individuals assayed

TABLE 2

*Allelic frequencies and percent heterozygosity at the B locus**

Generation	Number of plants	$B^{1.6}$	$B^{2.7}$	$B^{3.9}$	Percent heterozygosity
Initial†	4569	.0656	.9050	.0294	—
4	1279	.0606	.9042	.0352	3.83
5	1486	.0491	.9189	.0320	1.34
6	1006	.0905	.8857	.0238	0.00
14	1928	.0604	.8955	.0441	1.87
15	2843	.0769	.8417	.0814	4.37
16	2369	.0656	.9057	.0287	1.14
17	2461	.0508	.9118	.0374	0.08
24	4587	.1521	.7771	.0708	2.10
25	3967	.1592	.7724	.0684	1.06
26	3083	.1758	.7408	.0834	0.64

* Standard errors < 0.01.
† Allelic frequencies inferred from those of the 30 parents.

TABLE 3

*Allelic frequencies and percent heterozygosity at the C locus**

Generation	Number of plants	$C^{4.4}$	$C^{4.9}$	$C^{5.4}$	Percent heterozygosity
Initial†	4569	.1437	.2474	.6089	—
4	1234	.0332	.2646	.7022	9.24
5	1486	.0495	.3008	.6497	5.72
6	1006	.0740	.2868	.6392	1.89
14	1651	.0336	.2035	.7629	7.20
15	2843	.0823	.2812	.6365	2.33
16	2369	.0500	.3263	.6237	3.17
17	2461	.0772	.3074	.6154	0.61
24	4397	.1023	.3157	.5820	7.62
25	3967	.2113	.3079	.4808	1.48
26	3083	.2791	.2538	.4671	1.56

* Standard errors < 0.01.
† Allelic frequencies inferred from those of the 30 parents.

was large (from 1006 to 4802) in each of the three early (F_4–F_6), four intermediate (F_{14}–F_{17}) and three late (F_{24}–F_{26}) generations that were monitored. Standard errors of allelic frequencies are small (<0.01) so that changes of 0.02 or larger in frequencies are significant.

The data of Tables 1–4 show that highly significant changes in allelic frequencies occurred in a number of single-generation transitions. In many cases (e.g. allele $C^{5.4}$ in transition from generation 14 to 15) changes in allelic frequency >0.10, i.e. 10 standard errors or larger, occurred. In addition it is clear that longer-term changes also took place. Thus at locus A the 1.8 allele increased gradually at the expense of 0.2, at locus B alleles 1.6 and 3.9 increased at the expense of 2.7, while at loci C and D, the $C^{4.4}$ and $D^{6.4}$ alleles increased markedly at the expense of alleles $C^{5.4}$ and D^{N}. Also allelic frequencies in generations F_{14}–

TABLE 4

*Allelic frequencies and percent heterozygosity at the D locus**

Generation	Number of plants	$D^{6.4}$	$D^{6.5}$	$D^{6.6}$	D^{N}	Percent heterozygosity
Initial†	2698	.5830	.0843	.1730	.1597	—
5	1486	.4401	.0027	.1568	.4004	0.40
6	1006	.5238	.0318	.1730	.2714	0.79
16	2369	.5047	.0025	.2277	.2651	0.29
17	2458	.5388	.0867	.1804	.1941	0.85
25	3946	.6263	.0038	.0833	.2866	1.09
26	3082	.7279	.0513	.1663	.0545	1.39

* Standard errors < 0.01.
† Allelic frequencies inferred from those of the 30 parents.

F_{17} were not always intermediate to those in the early and late generations, probably reflecting longer-term reversals in environment. However, attempts to relate either single-generation or longer-term changes in allelic frequencies to rainfall, temperature, or other factors of the environment revealed no consistent associations.

Tables 1 to 4 also show that the proportion of heterozygotes decreased rapidly in early generations (F_4–F_6), as expected in a population in which there is more than 99% of self fertilization. However, the proportion of heterozygotes thereafter showed no further consistent decreases but fluctuated about apparent equilibrium values. Estimation of the proportion of heterozygotes for the D locus is complicated by the recessive D^N allele, which causes estimates of heterozygosity to be biased downward. This bias, which is expected to be larger in generations when D^N was frequent than in generations when its frequency was lower, probably contributes to the erratic results for the D locus (Table 4).

Amount of outcrossing: Before the above changes in gene and genotypic frequencies can be discussed quantitatively, precise information is required about the mating system of CCV. In particular, an estimate is required of the extent of self fertilization (s) versus random outcrossing ($t = 1 - s$). The mating system was assessed by harvesting 100 random spikes from each of the F_4–F_6, F_{14}–F_{17} and F_{24}–F_{26} generations, grown in three different years. This sampling was done prior to harvesting the population and so is independent of the samples from which allelic frequencies were estimated. Sixteen seedlings grown from the seeds of each spike were then assayed electrophorectially. If more than 3 heterozygotes and/or two different homozygotes were observed in any family for a particular locus, the maternal parent was judged to be heterozygous and the family was not used in estimating outcrossing for the locus in question. The parameter t and its standard error were estimated from the data obtained using the methods given in Appendix A. The estimates of t obtained were homogeneous over loci, generations, and also over years. They also agree closely with previous estimates made using morphological polymorphisms, and they are in accord with general experience at Davis, California, which is that a consistent low level of outcrossing occurs in barley. Hence, it is appropriate to use the mean estimate of $t = 0.0057 \pm 0.0014$ to characterize the mating system of CCV.

Inbreeding coefficients and fixation indices: For an inbreeding population with neutral alleles at each locus, the amount of heterozygosity in any generation is directly related to the Inbreeding Coefficient, F. Values of F in generation n for a population which is initially non inbred and practices an amount t of random outcrossing and $s = (1 - t)$ of selfing, are given by

$$F^n = \frac{s}{1+t}\left(1 - \left(\frac{s}{2}\right)^n\right) . \qquad (1)$$

Values of F for CCV computed from equation (1), assuming that $F = 0$ at the end of the intercrossing phase and that $t = 0.0057$ in all generations, are given in Table 5.

TABLE 5

*Theoretical inbreeding coefficients and observed fixation indices**

Generation	Theoretical F	Fixation indices $\hat{F}$							
		Locus A			Locus B	Locus C			Locus D
		1.8, 1.0	1.8, 0.2	1.0, 0.2	2.7, 1.6	5.4, 4.9	5.4, 4.4	4.9, 4.4	6.4, 6.5
4	.8672	.8865	.8297	.9168	.7788	.7950	.6700	.9539	
5	.9283	.9234	.8895	.9921	.8298	.8760	.8848	.9548	.9919
6	.9586	.9447	.9346	.9527	1	.9512	1	.9766	.9842
14	.9885	.9505	.9707	.9832	.9030	.8166	.7520	.8229	
15	.9886	.9473	.9467	.9815	.9782	.9479	.9832	.9620	
16	.9887	.9811	.9645	.9590	.9333	.9365	.9526	1	.9942
17	.9887	.9904	.9926	.9695	.9950	.9839	1	1	.9945
24	.9887	.9470	.9718	.9542	.9415	.8669	.7919	.9613	
25	.9887	.9705	.9654	.9712	.9699	.9753	.9826	.9666	.9767
26	.9887	.9796	.9546	.9691	.9831	.9658	.9863	.9725	.9649
Mean $(F-\hat{F})$		.0260	.0361	.0132	.0468	.0666	.0778	.0241	.0031

* Alleles $A^{0.2}$ and $A^{2.6}$, alleles $B^{1.6}$ and $B^{3.9}$, and alleles $D^{6.5}$, $D^{6.6}$, D^{N} have been combined.

If factors other than inbreeding (such as selection) affect the relationship between gene and genotypic frequencies, their combined effects can be represented by the Fixation Index, denoted by $\hat{F}^{(n)}$. The Fixation Index is defined (WEIR 1970) for the i^{th} and j^{th} alleles at a locus by

$$\hat{F}_{ij} = 1 - f_{ij}/p_i p_j, \tag{2}$$

where $2f_{ij}$ is the observed frequency of the heterozygote and p_i and p_j are allelic frequencies. Values of the Fixation Index for CCV are also given in Table 5.

Values of the Fixation Index are generally lower than those of the Inbreeding Coefficient for each locus, indicating an excess of heterozygotes. Most of the exceptions ($\hat{F}^n > F$) occurred either in generation 17, in which generation the proportion of heterozygotes was low for all loci, or in the early and intermediate generations for locus D, when heterozygosity was underestimated due to the high frequency of the recessive D^N allele. Although values of $\hat{F}^n$ increased substantially for all loci from the early (F_4–F_6) to the intermediate (F_{14}–F_{17}) generations, no further increases occurred from the intermediate to the late (F_{24}–F_{26}) generations. Apparently, therefore, CCV had not reached inbreeding equilibrium by generation 6 but had done so by generation 14, or earlier. All stages of the inbreeding process, including the apparent inbreeding equilibrium state, feature an excess of heterozygotes. Thus some force or forces other than mating system affect the relationship between gene and genotypic frequencies. In the next section we assume that mating system and selection are the forces responsible for the changes in gene and genotypic frequencies which occurred and estimate the selective values necessary to produce the observed results.

Estimation of selective values: Selective values for the individual genotypes were estimated from census data on genotypic frequencies in successive generations using the methods given in Appendix B, which are based on Model II of WORKMAN and JAIN (1966). This model, which assumes that census data are taken each generation soon after mating and that no selection occurs between mating and scoring, is appropriate because germination was excellent and survival to time of assay was high (about 99% of seeds produced assayable 7-day-old seedlings). Estimates of selective values (relative to the selective value of the most frequent homozygote taken as unity), and their standard errors, are given for the A locus in Table 6. Selective values of the homozygotes tend to be lower than unity (mean = 0.95, range 0.73–1.08); averaged over generations for each genotype, they reflect the increase which occurred in the frequency of $A^{1.8}A^{1.8}$ ($\hat{w} = 1$), the slight decrease in frequency of $A^{1.0}A^{1.0}$ ($\hat{w} = 0.99$) and the larger decrease which occurred for the combined $A^{0.2}A^{0.2}$, $A^{2.6}A^{2.6}$ genotypes ($\hat{w} = 0.90$). The low frequencies of the heterozygotes led to high standard errors and, as expected, estimates of their selective values were more erratic than those of homozygotes. However, the majority of the estimates exceed unity and their average (1.73) indicates substantial net reproductive advantage of heterozygotes over homozygotes. As noted by ANDERSON (1969), estimates of selective values made from successive pairs of generations are not independent since they are based on a common set of observations. This has the effect of causing an extreme value in one interval to be followed by a value at the opposite extreme in the next interval

TABLE 6

*Selection estimates and standard deviations for the A locus with 0.57% outcrossing**

Generations	Genotypes					
	$A^{1.8}A^{1.8}$	$A^{1.0}A^{1.0}$	$A^{0.2}A^{0.2}$	$A^{1.8}A^{1.0}$	$A^{1.8}A^{0.2}$	$A^{1.0}A^{0.2}$
4/5	1	1.08	0.99	1.34	1.19	0.05
		(0.07)	(0.16)	(0.30)	(0.36)	(0.20)
5/6	1	0.97	0.96	1.30	1.09	10.47
		(0.07)	(0.18)	(0.39)	(0.49)	(12.06)
14/15	1	0.97	0.79	1.98	2.76	0.93
		(0.04)	(0.10)	(0.45)	(1.31)	(0.83)
15/16	1	0.96	1.08	0.48	1.25	4.10
		(0.04)	(0.13)	(0.19)	(0.57)	(2.66)
16/17	1	0.87	0.81	0.38	0.08	0.91
		(0.04)	(0.10)	(0.36)	(0.26)	(0.60)
24/25	1	1.05	0.92	0.94	1.91	0.98
		(0.04)	(0.09)	(0.21)	(0.79)	(0.58)
25/26	1	1.04	0.73	1.05	1.77	1.40
		(0.04)	(0.10)	(0.35)	(0.78)	(1.03)
Mean	1	0.99	0.90	1.07	1.44	2.69
		(0.02)	(0.05)	(0.13)	(0.27)	(1.78)

Mean of all homozygotes = 0.96
Mean of all heterozygotes = 1.73

* Alleles $A^{0.2}$ and $A^{2.6}$ have been combined. Standard deviations are in parenthesis.

TABLE 7

Average selection estimates and standard deviations (in parentheses) for Esterase loci A, B, C and D with 0.57% outcrossing

	Allele			Genotype					
Locus	1	2	3	11	22	33	12	13	23
A	1.8	1.0	0.2†	1.00	0.99	0.90	1.07	1.44	2.69
					(0.02)	(0.05)	(0.13)	(0.27)	(1.78)
B	2.7	1.6†		1.00	1.08		1.13		
					(0.05)		(0.16)		
C	5.4	4.9	4.4	1.00	1.17**	1.90**	1.15	0.83	1.74
					(0.03)	(0.14)	(0.13)	(0.31)	(0.80)
D	6.4	6.5†		1.00	0.73**		2.38		
					(0.03)		(0.77)		

† Alleles are combined at the A, B and D loci as follows: $A^{0.2} + A^{2.6}$ = allele 3, $B^{1.6} + B^{3.9}$ = allele 2, $D^{6.5} + D^{6.6} + D^{N}$ = allele 2.

** Significant departure from 1.00, $P < 0.01$.

as, for example, seems to have been the case for genotype $A^{0.2}A^{1.0}$ in the transitions from generations 4 to 5 and 5 to 6.

Results for loci B, C and D were similar in pattern and consequently only mean selective values are reported for these loci (Table 7). At locus B, the $B^{2.7}$ allele is present in high frequency in all generations. Thus all genotypes other than the $B^{2.7}B^{2.7}$ homozygote are infrequent with the result that estimates of their selective values are subject to large sampling errors and, as expected, quite variable. Nevertheless, averaged over generations, the selective values reflect the decrease which occurred in the $B^{2.7}B^{2.7}$ homozygote ($\hat{\bar{w}} = 1$) and the increase in frequency of the combined $B^{1.6}B^{1.6}$ and $B^{3.9}B^{3.9}$ homozygotes ($\hat{\bar{w}} = 1.08$). There is also indication that, on the average, heterozygotes ($\hat{\bar{w}} = 1.13$) have a selective advantage over all homozygotes. For the C locus the most conspicuous features of the results are high average selective values of the $C^{4.4}C^{4.4}$ homozygote ($\hat{\bar{w}} = 1.90$) and the $C^{4.4}C^{4.9}$ heterozygote ($\hat{\bar{w}} = 1.74$), and the low selective value of the $C^{4.4}C^{5.4}$ heterozygote ($\hat{\bar{w}} = 0.83$). In estimating selective values for the D locus the three least frequent alleles ($D^{6.5}$, $D^{6.6}$ and D^{N}) were combined which, together with inability to distinguish D^{N} heterozygotes, reduced the number of genotypic classes to three: $D^{6.4}D^{6.4}$ (including also $D^{6.4}D^{N}$); $D^{6.5}D^{6.5}$ (including also $D^{6.6}D^{6.6}$, $D^{N}D^{N}$, $D^{6.5}D^{6.6}$, $D^{6.5}D^{N}$, $D^{6.6}D^{N}$); and $D^{6.4}D^{6.5}$ (including also $D^{6.4}D^{6.6}$). The net effect of misclassification of D^{N} heterozygotes and the reduction in number of classes is to underestimate the frequency of heterozygotes and hence also their selective values. Despite this bias the excess of heterozygotes was very large for this locus which was reflected in consistently higher selective values of the heterozygotes relative to the superior homozygote ($D^{6.4}D^{6.4}$), and a very high mean selective value ($\hat{\bar{w}} = 2.38$).

DISCUSSION

The structure of this experiment allows it to be stated with confidence that the

observed genetic changes in CCV were not due to genetic drift, mutation or migration. Genetic drift can be eliminated on the basis that population size was far too large in all generations for sampling accidents to have had any measureable effect on allelic frequencies over the 25 generation interval involved. Mutation can be eliminated on two counts: first the mutation-rate study mentioned earlier shows that mutation rates at these loci are too low to have affected the short-term dynamics of the population; second, even though these loci are known from studies of worldwide variability in barley to be capable of mutating to many allelic forms, no alleles not present in the parents were found in any generation. Migration can be eliminated on this same basis: no alleles not present originally were found in any generation although such alleles are unlikely to have escaped detection had the isolation in which the population was grown broken down and migration from outside populations of barley taken place. Thus, among the established evolutionary forces, this leaves selection and mating system as the ones responsible for the changes in gene and genotypic frequencies which occurred in CCV.

Selection in CCV took two forms. First, the observed change in gene and genotypic frequencies show that directional selection, in which certain homozygotes were favored over other homozygotes, took place at all four loci. The selective values indicate that this directional selection was intense. Thus in 10 of the 24 single-generation transitions for which selective values could be calculated the better or best homozygote had a net reproduction advantage of more than 50% over the poorer or poorest homozygote. As expected in a population grown in a temporally heterogeneous environment, the same homozygote was not always favored in different generations. Nevertheless, long-term trends in allelic frequencies are discernible and they indicate that the $C^{4.4}C^{4.4}$ and $D^{6.4}D^{6.4}$ homozygotes were much superior, and that the $A^{1.8}A^{1.8}$ and $B^{1.6}B^{1.6}$ homozygotes were moderately superior, on the average, to the other homozygotes at these loci. The selective values give a quantitative measure of this superiority. Thus the average selective value of the $C^{4.4}C^{4.4}$ homozygote, averaged over seven single-generation transitions, was 1.90 and that of the next best homozygote ($C^{4.9}C^{4.9}$) was 1.17, indicating a mean reproductive advantage of $1.90/1.17 = 1.62$ or 62%. Parallel values for the other three loci are: $D^{6.4}D^{6.4} = 1/0.73 = 1.37$; $B^{1.6}B^{1.6} = 1.08/1 = 1.08$ and $A^{1.8}A^{1.8} = 1/0.99 = 1.01$. The reproductive advantages of the best homozygote over the poorest homozygote were 11, 8, 90 and 37%, respectively, for the A, B, C and D loci.

An important feature of the selective values is that they differ substantially for the A, B and C loci, indicating that selection operated differentially on these three very tightly linked loci. These loci are ordered $B \leftarrow 0.0023 \rightarrow A \leftarrow 0.0048 \rightarrow C$. This suggests that the A locus itself (WEIR, ALLARD and KAHLER 1972), and not genes linked to it, affects reproductive capacity: the chromosome segment involved is too small to carry many additional genes.

The second form of selection which occurred in CCV is balancing selection featuring an excess of heterozygotes over expectations based on considerations of mating system alone. It is possible that the observed departures reflect the effects

of single-locus overdominance. However, multilocus analyses of these data show that specific positive and negative epistatic interactions of a type that can lead to excesses of heterozygotes occurred between particular alleles at different loci in CCV (WEIR, ALLARD and KAHLER 1972; CLEGG *et al.* 1972). There is also evidence from competition studies in CCV that reproductive capacity is often higher when individuals compete with genotypes other than their own and that such competitive interactions can also lead to a balancing type of selection that may produce excess heterozygosity (ALLARD and ADAMS 1969; SCHUTZ and USANIS 1969). Hence it seems likely that the surplus of heterozygotes observed in CCV has multiple and complex causes.

These results lead us to the conclusion that the observed patterns of enzymatic variation in CCV result from balancing selection with strong selection intensities and with the direction of selection different for various alleles. This same hypothesis also appears adequate to explain the main features of variation in the barley species as a whole. The high degree of polymorphism found within local populations of barley is consistent with balancing selection of the intensity found in CCV. The marked differentiation in allelic frequencies between populations from different ecogeographical regions is expected if selection favors particular alleles at some places and other alleles at other places. There can be no doubt that migration takes place frequently between populations of barley, both inadvertently and also through deliberate introductions by man. However, it is well known that even a low rate of migration is sufficient to nearly equalize gene frequencies among populations unless selection strongly favors alternative alleles in different populations. Furthermore, migration rates high enough to maintain the high levels of polymorphism observed within populations would almost certainly reduce differences between populations to low levels, even in the face of strong directional selection in different directions in different populations. Thus, observed patterns of variability in the barley species are consistent with the evidence from CCV that some combination of directional and balancing selection plays a much more important role than mutation, migration or genetic drift in determining the present-day genetic makeup of the barley species.

No other hypothesis appears to be in accord with all of the observations. The neutrality hypothesis (KIMURA 1968) predicts large amounts of genetic variability within large populations such as CCV, provided many mutational sites are assumed within each gene. However, this hypothesis does not predict the observed excesses of heterozygotes in CCV. The existence of extensive polymorphism within accessions of the world barley collection is also at variance with the neutrality hypothesis. Observed mutation rates are too low to support more than trivial levels of polymorphism in such small populations. At the same time changes in allelic frequencies which occurred in CCV (a large population) were far too large to be accounted for by mutation and the drift of neutral alleles.

We wish to thank Mrs. SHARON PAGLIUSO and Mr. FARHAD MONADJEM for their capable technical assistance in the electrophoretic laboratory and field laboratory, respectively.

LITERATURE CITED

ALLARD, R. W. and J. ADAMS, 1969 Population studies in predominantly self-pollinated species. XIII. Intergenotypic competition and population structure in barley and wheat. Am. Naturalist **103**: 621–645.

ANDERSON, WYATT W., 1969 Selection in experimental populations. I. Lethal genes. Genetics **62**: 653–672.

CLEGG, M. T., R. W. ALLARD and A. L. KAHLER, 1972 Is the gene the unit of selection? Evidence from two experimental plant populations. Proc. Nat. Acad. Sci., in press.

KAHLER, A. L. and R. W. ALLARD, 1970 Genetics of isozyme variants in barley. I. Esterases. Crop Science **10**: 444–448. ——, 1972 Worldwide patterns of variation of esterase allozymes in barley (*Hordeum vulgare* and *H. spontaneum*). (In preparation).

KIMURA, M., 1968 Evolutionary rate at the molecular level. Nature **217**: 624–626.

SCHUTZ, W. M. and S. A. USANIS, 1969 Inter-genotypic interactions in plant populations. II. Maintenance of allelic polymorphisms with frequency-dependent selection and mixed selfing and random mating. Genetics **61**: 875–891.

SUNESON, C. A., 1956 An evolutionary plant breeding method. Agron. J. **48**: 188–191.

WEIR, B. S., 1970 Equilibria under inbreeding and selection. Genetics **63**: 371–378.

WEIR, B. S., R. W. ALLARD and A. L. KAHLER, 1972 Analysis of complex allozyme polymorphisms in a barley population. Genetics **72**: 505–523.

WORKMAN, P. L. and S. K. JAIN, 1966 Zygotic selection under mixed random mating and self-fertilization: theory and problems of estimation. Genetics **54**: 159–171.

APPENDIX A

ESTIMATION OF OUTCROSSING

Estimation of t: With k alleles at a locus, suppose that over all families from parents homozygous for the i^{th} allele, N_i individuals are observed. If b_i of these are also homozygous for the i^{th} allele, the likelihood function is

$$L = \begin{bmatrix} N_i \\ b_i \end{bmatrix} (1-tq_i)^{b_i} (tq_i)^{N_i - b_i} , \tag{1}$$

where t is the unknown outcrossing rate and q_i is one minus the population frequency of the i^{th} allele.

Differentiating to obtain the maximum likelihood estimate of t:

$$\frac{d \log L}{dt} = \sum_{i=1}^{k} \frac{N_i - b_i}{t} - \sum_{i=1}^{k} \frac{b_i q_i}{1-tq_i} , \tag{2}$$

so that the estimate $\hat{t}$ satisfies

$$\hat{t} \sum_{i=1}^{k} \frac{b_i q_i}{1-\hat{t}q_i} = \sum_{i=1}^{k} [N_i - b_i] . \tag{3}$$

Equation (3) must in general be solved numerically, and it was in the present case, but if t is small enough for tq_i to be negligible, an adequate estimate of t is given by

$$\hat{t} = \sum_{i=1}^{k} (N_i - b_i) / \sum_{i=1}^{k} b_i q_i . \tag{4}$$

Variance of estimates: Standard errors of the estimates of t were obtained from large sample variance theory (see e.g. CRAMER 1946, equation 27.7.3). The coefficients in equation (3) involve

b_i and q_i, for which only estimates are available. We write these estimates just as b_i and q_i. The b_i are independent and binomially distributed:

$$b_i \sim B\,(N_i, 1-tq_i) \quad \text{var}\,(b_i) = N_i\,tq_i\,\,(1-tq_i) \tag{5}$$

The estimates of q_i are obtained from a different sample, of size M, so that q_i and b_i are independent but the q_i are correlated among themselves. In the sample suppose the relative frequency of the homozygote for the i^{th} allele is f_{ii}, and the relative frequency of the heterozygote for the i^{th} and j^{th} alleles is $2f_{ij}$. Then

$$\begin{aligned}\text{var}\,(f_{ii}) &= f_{ii}\,(1-f_{ii})/M\\ \text{cov}\,(f_{ii}, 2f_{ij}) &= -\,2f_{ii}\,f_{ij}/M.\end{aligned}$$

The estimate of the frequency of the i^{th} allele, p_i, is thus

$$p_i = \sum_{j=1}^{k} f_{ij} \text{ and } q_i = 1 - p_i,$$

so that

$$\begin{aligned}\text{var}\,(q_i) &= \text{var}\,(p_i) = p_i(1-p_i)/M - (p_i - f_{ii})/2M\\ \text{cov}\,(q_i, q_j) &= cov\,(p_i, p_j) = -\,p_ip_j/M + f_{ij}/2M\end{aligned} \tag{6}$$

The variance of the estimate of the outcrossing parameter is thus a function of the variances and covariances of the b_i and q_i. To the order of accuracy of the variances [obtained by substituting estimates of b_i and q_i into (5) and (6)] we have

$$\begin{aligned}\text{var}\,(\hat{t}) = &\sum_{i=1}^{k} \left[\frac{d\hat{t}}{db_i}\right]^2 \text{var}\,(b_i)\\ &+ \sum_{i=1}^{k} \left[\frac{d\hat{t}}{dq_i}\right]^2 \text{var}\,(q_i)\\ &+ \sum_{i=1}^{k}\sum_{j=1}^{k} \left[\frac{d\hat{t}}{dq_i}\right]\left[\frac{d\hat{t}}{dq_i}\right] \text{cov}\,(q_i, q_j).\end{aligned} \tag{7}$$

CRAMÉR, H., 1946 *Mathematical methods of statistics*. Princeton.

APPENDIX B

ESTIMATION OF SELECTIVE VALUES

Estimation: Denoting the frequency of the homozygote for the i^{th} among k alleles in generation n by ${f^n}_{ii}$, the frequency of the heterozygote for the i^{th} and j^{th} alleles in generation n by ${2f_{ij}}^n$, the selective values of homozygotes and heterozygotes by w_{ii} and w_{ij}, respectively, and the amount of outcrossing in generation n by t, expected genotypic frequencies in generation $n+1$ are given by the following transition equations:

$$\begin{aligned}f_{ii}^{n+1} &= \frac{s}{\bar{w}}\,(w_{ii}f_{ii}^n + \frac{1}{2}\sum_{\substack{j=1\\ j\neq i}}^{k} w_{ij}f_{ij}^n) + \frac{t}{\bar{w}^2}\,(\sum_{j=1}^{k} w_{jj}f_{ij}^n)^2, \quad i = 1,2,..,k\\ f_{ij}^{n+1} &= \frac{s}{2\bar{w}}\,(w_{ij}f_{ij}^n) + \frac{t}{\bar{w}^2}\,(\sum_{m=1}^{k} w_{im}f_{im}^n)\,(\sum_{l=1}^{k} w_{lj}f_{lj}^n), \quad \begin{matrix} i\neq j\\ i,j=1,2,..,k\end{matrix}\end{aligned} \tag{1}$$

where dummy suffices m and l range over integers 1 to k. In these transition equations the frequency of each genotype in the n^{th} generation has been weighted by its selective value w, and $\bar{w} = \sum\sum_{ij} w_{ij}{f^n}_{ij}$ is the mean fitness for that generation.

For an observed set of genotypic frequencies in generation n, the above equations give the expected genotypic frequencies in generation $n+1$. By equating these expected values to observed frequencies in generation $n+1$, a set of equations is obtained which can be solved to give maximum likelihood estimates of the selective values. Because there is one more selective value than independent genotypic frequencies in each generation, we set w_{11} (corresponding to the more frequent homozygote) equal to one. The maximum likelihood estimates are then:

$$\hat{w}_{ii} = \frac{f_{11}^n V_i}{f_{ii}^n V_1} \qquad i=1,2,...,k.$$

$$\hat{w}_{ij} = \frac{2f_{11}^n U_{ij}}{f_{ij}^n V_1} \qquad i,j=1,2,...,k.\ i\neq j. \tag{2}$$

where

$$V_i = 2f_{ii}^{n+1} - p_i^{n+1}(s+2tp_i^{n+1}) \qquad i=1,2,...,k.$$

$$U_{ij} = f_{ij}^{n+1} - tp_i^{n+1}p_j^{n+1} \qquad i,j=1,2,...,k.\ i\neq j.$$

$$p_i^{n+1} = \sum_{j=1}^{k} f_{ij}^{n+1} \qquad i=1,2,...,k.$$

Note that w_{ij} is not estimable when $f_{ij}^n = 0$.

Variance of estimates: As emphasized by LORENZ (1970) the above estimates are functions of three sets of random variables (genotypic frequencies in generation n and generation $n+1$, and the outcrossing parameter) for which estimates only are available. The estimates of genotypic frequencies are distributed multinomially. For a sample of size N_n in generation n we write both estimates and true values as f_{ii}^n, $2f_{ij}^n$

$$\text{var}(f_{ii}^n) = f_{ii}^n(1-f_{ii}^n)/N_n$$

$$\text{var}(f_{ij}^n) = 1/4\ \text{var}\ (2f_{ij}^n) = f_{ij}^n(1-2f_{ij}^n)/2N_n$$

$$\text{cov}(f_{ii}^n, f_{ij}^n) = 1/2\ \text{cov}\ (f_{ii}^n, 2f_{ij}^n) = -f_{ii}^n f_{ij}^n/N_n$$

$$\text{cov}(f_{ij}^n, f_{pq}^n) = 1/4\ \text{cov}\ (2f_{ij}^n, 2f_{pq}^n) = -f_{ij}^n f_{pq}^n/N_n$$

Estimates in generation n are independent of those in generation $n+1$. An estimate of the outcrossing parameter and its variance is also available (Appendix A). This variance takes proper account of the fact that the estimate of t is not independent of genotypic frequencies in generation n. To the order of accuracy of the estimates of the variances of the genotypic frequencies and the outcrossing parameter, for any selection value w:

$$\text{var}\ (\hat{w}) = \sum_{pquv}\sum\sum\sum \left(\frac{d\hat{w}}{df_{pq}^n}\right)\left(\frac{d\hat{w}}{df_{uv}^n}\right)\ \text{cov}(f_{pq}^n, f_{uv}^n)$$

$$+ \sum_{pquv}\sum\sum\sum \left(\frac{d\hat{w}}{df_{pq}^{n+1}}\right)\left(\frac{d\hat{w}}{df_{uv}^{n+1}}\right)\ \text{cov}(f_{pq}^{n+1}, f_{uv}^{n+1})$$

$$+ \left(\frac{d\hat{w}}{dt}\right)^2 \text{var}(t).$$

The sums are over all genotypic frequencies, so that $p,q,u,v = 1,2,...,k$. When $p=u$, $q=v$ (or $p=v$, $q=u$) we have the variance terms. For example, the first sum in the two allele case becomes:

$$\left(\frac{d\hat{w}}{df_{11}^n}\right)^2 \text{var}(f_{11}^n) + \left(\frac{d\hat{w}}{df_{22}^n}\right)^2 \text{var}(f_{22}^n) + \left(\frac{d\hat{w}}{df_{12}^n}\right)^2 \text{var}(f_{12}^n)$$

$$+2\left(\frac{d\hat{w}}{df_{11}^n}\right)\left(\frac{d\hat{w}}{df_{22}^n}\right)\operatorname{cov}(f_{11}^n, f_{22}^n)+2\left(\frac{d\hat{w}}{df_{11}^n}\right)\left(\frac{d\hat{w}}{df_{12}^n}\right)\operatorname{cov}(f_{11}^n, f_{12}^n)$$
$$+2\left(\frac{d\hat{w}}{df_{22}^n}\right)\left(\frac{d\hat{w}}{df_{12}^n}\right)\operatorname{cov}(f_{22}^n, f_{12}^n).$$

In applying the above theory to CCV the computations of selective values and their variances were made in three ways: (1) on the basis of the 36 individual single-locus, single-generation estimates of t; (2) using the overall estimate of t but ignoring its variance; and (3) using the overall estimate of t and its variance. Since all three sets of selection values and standard errors were the same to two decimal place accuracy, the effect of ignoring the stochastic nature of the outcrossing parameter appears to be trivial in practice.

LORENZ, R. J., 1970 On the determination of relative fitness from frequencies of genotypes in subsequent generations. II, III. Statistical problems in estimation of fitness. Theoretical and Applied Genetics **40**: 76–87.