*Proc. Nat. Acad. Sci. USA*
Vol. 69, No. 10, pp. 3043–3048, October 1972

# Evidence for Coadaptation in *Avena barbata*

**(electrophoresis/polymorphism/gametic phase disequilibrium/multiniche selection)**

R. W. ALLARD, G. R. BABBEL*, M. T. CLEGG†, AND A. L. KAHLER

Department of Genetics University of California, Davis, Calif. 95616

*Communicated by Th. Dobzhansky, August 10, 1972*

**ABSTRACT** **Two populations of *Avena barbata* have been sampled intensively and allelic frequencies have been determined on a fine grid for six enzyme loci. The distribution of alleles among microniches within the location occupied by one of these populations (CSA-1) was correlated with environment following patterns of macro- and microdifferentiation previously described in California. Multilocus analyses revealed striking gametic-phase disequilibrium for both linked and unlinked loci, and also that two balanced five-locus gametic types, one associated with mesic conditions and the other with more xeric conditions, are in substantial excess at the CSA-1 location. Evidence is presented that these two gametic types are maintained by selection acting on two coadapted combinations of alleles, held together by restriction of recombination resulting from linkage and/or inbreeding. Gametic-phase disequilibrium featuring different allelic complexes were found in each of three microniches in the second population (Geyserville), which occupies a site that is environmentally atypical for *A. barbata*.**

Recent studies of the slender wild oat species, *Avena barbata*, have shown that the distribution of alleles at five enzyme loci in this species is nonrandom and closely associated with environment on both macro- and microgeographical scales in California (1, 2). Populations that occupy xeric habitats are monomorphic for a specific combination of alleles, and populations that occupy mesic habitats are monomorphic for a specific balanced opposite set of alleles at these loci. Populations in intermediate habitats are usually polymorphic at all five loci; furthermore, the frequencies of "xeric" and "mesic" combinations of alleles over loci are highly correlated with environment, "xeric" combinations increasing as the habitat becomes increasingly xeric and vice versa. These observations suggest that natural selection acts on these loci as a unit, i.e., that they constitute what Dobzhansky (3) has termed coadapted gene complexes.

This paper reports a study of two polymorphic populations of *A. barbata* from which sufficiently large samples were taken to permit analysis on a multilocus basis. Striking correlations in allelic state over loci, including correlations among unlinked loci, were found in both populations. The results thus establish, at the level of the gene, that the genetic resources of these populations have been structured by natural selection into highly interacting, coadapted gene complexes.

* Present address: Department of Biology, University of South Florida, Tampa, Fla. 33620.
† Present address: Division of Biological and Medical Sciences, Brown University, Providence, R.I. 02912.

## METHODS AND MATERIALS

The first of the two populations of *A. barbata* selected for intensive sampling occupies the CSA-1 location of the previously studied Calistoga site (2, 4). This population was chosen for the following reasons: (*i*) the habitat of the CSA-1 location appears to be typical of intermediate habitats of the cool summer region of California; (*ii*) the CSA-1 population is highly polymorphic for the five enzyme loci [esterase loci $E_4$, $E_9$, $E_{10}$, phosphatase (EC 3.1.3.2) locus $P_5$ and anodal peroxidase (EC 1.11.1.7) locus $APX_5$] whose allelic distributions are correlated with environment on both macro- and microgeographical scales in California; (*iii*) *A. barbata* appears to have occupied this location for a long time; (*iv*) the location shows no sign of disturbance by man or domestic animals. Thus, CSA-1 appears to offer near optimal opportunity to detect all possible genotypic classes in a typical polymorphic population, which can be presumed to be in equilibrium.

Most of the plants of *A. barbata* in the CSA-1 location occur in a rectangular area about 8 m wide and 18 m long. This area, which contained about 50,000 individuals at the time of sampling in September, 1971, was divided into 21 plots, each 2.6 × 2.6 m, which were sampled by collection of about 100 seeds from the ground in each plot after seed shattering was complete (four plots at one edge of the location were too sparsely populated to be sampled). These seeds were germinated in soil in the greenhouse and 30-day-old seedlings were used as enzyme sources. Enzymes for four esterase loci, $E_1$, $E_4$, $E_9$, $E_{10}$, one phosphatase locus, $P_5$, and one anodal peroxidase locus, $APX_5$, were assayed by standard techniques (5). Three of these loci are linked ($P_5 \leftarrow 0.04 \pm 0.01 \rightarrow APX_5 \leftarrow 0.23 \pm 0.03 \rightarrow E_{10}$), whereas the $E_1$, $E_4$, and $E_9$ loci segregate independently of the three linked loci and each other (5). Following previous convention, the slower migrating allele at each locus is designated allele 1 and the faster migrating allele is designated allele 2.

The second population sampled occupies a previously studied site (1) near Geyserville, Calif. This population was chosen for intensive study because it is strikingly atypical of polymorphic populations in the following respects: (*i*) it occupies an open site with southwest exposure that is unusually xeric for the highly mesic north-coastal region; (*ii*) at this site alleles characteristic of the California warm summer region are in high frequency or fixed for three loci [$E_9^{(2)}$, $E_{10}^{(2)}$, $P_5^{(2)}$] and the reverse is the case for two other loci [$E_4^{(2)}$, *b* = gray lemma morph]; (*iii*) the population is fixed for allele $APX_5^{(3)}$, a rare allele in the cool summer region, and it is poly-

TABLE 1. *Frequencies of "mesic" alleles for five enzyme loci in the four subdivisions of location CSA-1*

| Subdivision | $N$ | Allelic frequencies | | | | | |
|---|---|---|---|---|---|---|---|
| | | $E_4^{(2)}$ | $E_9^{(1)}$ | $E_{10}^{(1)}$ | $P_5^{(1)}$ | $APX_5^{(2)}$ | Mean |
| *A* | 577 | .94 | .96 | .97 | .95 | .94 | .95 |
| *B* | 174 | .70 | .67 | .64 | .74 | .70 | .69 |
| *C* | 277 | .52 | .47 | .44 | .37 | .35 | .43 |
| *D* | 111 | .21 | .19 | .25 | .48 | .09 | .24 |
| Location | 1139 | .73 | .72 | .72 | .73 | .67 | .71 |

$N$ = number of individuals scored.

morphic for three alleles at the $E_1$ locus, including one allele [$E_1^{(3)}$] not found in any other population sampled in California.

The Geyserville population occupies 10 acres or more on a gently sloping west-facing hillside. The area sampled was 9 m wide by 63 m long, with long axis extending across the bottom of the hillside parallel to Highway 128. The site was divided into 21 plots, 3 m by 9 m, which were sampled by collection of 80–100 seeds from the ground after seed fall was complete (one plot was too sparsely populated for effective sampling). Seeds were germinated and seedlings were examined electrophoretically as for the Calistoga population.

## RESULTS

### CSA-1

All of the 17 plots at the CSA-1 location were found to be polymorphic for the same two alleles at the $E_4$, $E_9$, $E_{10}$, $P_5$, and $APX_5$ loci, and fixed for allele $E_1^{(2)}$, as reported in earlier studies (2, 4). However, substantial and significant differences in allelic frequencies were found among certain plots, indicating that the location is heterogeneous with respect to allelic frequencies. To divide the location into natural subdivisions according to allelic frequencies, genetic identity ($I$) values (6) were calculated for each plot and subjected to cluster analysis (7). This analysis indicated that the 17 plots fall into four internally homogeneous subdivisions, designated *A–D*, that differ significantly from each other in gene frequency. Allelic frequencies for each subdivision are given in Table 1, from which it can be seen that the $E_4^{(2)}$, $E_9^{(1)}$, $E_{10}^{(1)}$, $P_5^{(1)}$, and $APX_5^{(2)}$ (abbreviated 21112) alleles decrease and the 12221 alleles increase progressively in frequency, and at nearly identical rate, from subdivision *A–D*. Since the 21112 and 12221 alleles are characteristic of mesic and xeric habitats (1, 2) in California, the *A*, *B*, *C*, and *D* subdivisions were designated "most mesic," "intermediate mesic," "intermediate xeric," and "most xeric," respectively.

Upon discovery of heterogeneity in allelic frequencies within CSA-1, the location was examined again to determine whether it is visibly heterogeneous environmentally. This inspection revealed that the eight plots of subdivision *A* are situated at the bottom of a shallow swale, which constitutes a local drainage area, whereas the two plots of subdivision *D* are located on one side of the swale about 60 cm higher in elevation. The three plots of subdivision *B* and four plots of subdivision *C* are intermediate in elevation and also in soil type, which becomes progressively shallower and more rocky up the side of the swale. Thus, within a few feet there is a clear gradient with respect to mesic against xeric features of the environment that parallels the observed gradient in allelic frequencies.

### Gametic analysis

To obtain quantitative measures of the extent of interactions between pairs of loci, we have calculated values of the gametic phase disequilibrium parameter, $D$, defined as the determinant of the gametic matrix, for each of the 10 locus pairs within each of the four subdivisions of the CSA-1 location. Since $D$ depends on gene frequencies, we have expressed it relative to the maximum value it can take for the observed gene frequency for each locus pair, i.e., in terms of the relative gametic phase disequilibrium parameter, $D'$, [$D' \in (-1.0, 1.0)$], (8). We have also computed correlation coefficients, defined as $r = D/[p_1(1 - p_1)p_2(1 - p_2)]^{1/2}$, where $p_1$ and $p_2$ are allelic frequencies for loci *A* and *B*, because this measure of disequilibrium is less sensitive than $D'$ to sampling errors when gene frequencies are very different for the two loci of a pair. Mean absolute values of $D'$ and $r$ were large for each of the 10 locus pairs, and for each of the four subdivisions (Table 2), indicating strong interactions between all pairs of loci. Further, the sign of $D'$ and $r$ is negative for the $E_4$–$E_9$, $E_4$–$E_{10}$, $E_4$–$P_5$, $E_9$–$APX_5$, $E_{10}$–$APX_5$, and $P_5$–$APX_5$ pairs and positive for the $E_4$–$APX_5$, $E_9$–$E_{10}$, $E_9$–$P_5$, and $E_{10}$–$P_5$ pairs. Thus, the departures from random association are in the direction expected if 21112 and 12221 represent correlated five-locus allelic complexes.

Table 2 also shows that absolute values of $D'$ and $r$ are not always higher for linked than for unlinked pairs of loci, indicating that selection is not equally intense for all loci. $D'$ and $r$ values are generally low for combinations of $P_5$ with the three esterase loci, especially in the most xeric subdivision. This suggests that epistatic interactions of this locus with the esterase loci are weaker than those of the other loci.

TABLE 2. *$D'$ values and correlation coefficients ($r$, in parentheses) for locus pairs in the four subdivisions of CSA-1*

| Two-locus pairs | Subdivision | | | | |
|---|---|---|---|---|---|
| | *A* | *B* | *C* | *D* | Total |
| $E_4$, $E_9$ | −.68(−.58) | −.50(−.46) | −.59(−.53) | −.52(−.49) | −.69(−.68) |
| $E_4$, $E_{10}$ | −.30(−.61) | −.55(−.48) | −.50(−.43) | −.46(−.41) | −.65(−.64) |
| $E_4$, $P_5$ | −.73(−.67) | −.02(−.02) | −.35(−.26) | .03(.01) | −.44(−.44) |
| $E_4$, $APX_5$ | .73(.72) | .05(.05) | .50(.35) | .50(.30) | .66(.58) |
| $E_9$, $E_{10}$ | .75(.67) | .51(.48) | .58(.55) | .51(.42) | .70(.70) |
| $E_9$, $P_5$ | .50(.46) | .18(.16) | .42(.34) | −.11(−.05) | .49(.48) |
| $E_9$, $APX_5$ | −.63(−.53) | −.20(−.18) | −.49(−.38) | −.29(−.19) | −.66(−.60) |
| $E_{10}$, $P_5$ | .67(.55) | .40(.32) | .50(.42) | −.09(−.05) | .56(.54) |
| $E_{10}$, $APX_5$ | −.71(−.54) | −.37(−.32) | −.55(−.45) | −.57(−.31) | −.70(−.64) |
| $P_5$, $APX_5$ | −.90(−.83) | −.84(−.77) | −.74(−.71) | −.81(−.26) | −.88(−.78) |
| Mean $\lvert D'\rvert$, $\lvert r\rvert$ | .71(.62) | .36(.32) | .52(.44) | .39(.25) | .64(.61) |
| $N$ | 577 | 174 | 277 | 111 | 1218* |

$N$ = number of individuals scored.

* Includes 79 individuals collected at random over the entire location.

To determine if selection favors specific combinations of alleles over all five loci, we have also compared observed and expected five-locus gametic frequencies in each of the four subdivisions. Observed frequencies of five-locus gametes were computed from the sums of the appropriate genotypic frequencies, following the procedure of Clegg *et al.* (9), and expected frequencies were calculated as the products of the observed single-locus allelic frequencies. Table 3 gives observed frequencies and relative deviations from expected frequencies for the eight most frequent (among 32 possible) five-locus gametic types. Two gametic types predominate, the 12221 and 21112 types characteristic of "xeric" and "mesic" habitats, respectively. Relative frequencies of the gametic types are not constant over the CSA-1 location. The 21112 gametic type, which is in very high frequency (0.915) in subdivision *A*, falls regularly in frequency with increasing xerism to 0.019 in subdivision *D*. Conversely, the 12221 gametic type increases progressively in frequency from 0.017 in *A* to 0.319 in *D*. In subdivision *D*, a third gametic type, 12211, which differs from 12221 only at the $P_5$ locus, is also in high frequency. These two gametic types (12221, 12211) combined account for 62.4% of all gametic types in the most xeric subdivision. This apparent lack of preference for the $P_5^{(1)}$ and $P_5^{(2)}$ alleles is consistent with the lower $D'$ and $r$ values, discussed above, in suggesting lesser selection at this locus in the most xeric subdivision. The data of Table 3 also show that the two predominating five-locus gametic types were present in great excess over expectations (38.0 and 11.2%), whereas the six other more frequent types showed only small excesses or were less frequent than expected. There was a deficiency of nearly all of the remaining 24 infrequent gametic types, the cumulative deficiency being 55% below expectation. Analysis of gametic frequencies thus indicates strong gametic phase disequilibrium involving all five loci (including the $P_5$ locus in the most mesic subdivision).

**Analysis of genotypic frequencies**

Comparisons of observed two- and five-locus genotypic numbers with genotypic numbers computed from products of single-locus genotypic frequencies, give another and in some ways more informative picture of the genetic structure of the CSA-1 population. Comparisons were made first within each of the four subdivisions; however, since the pattern of departures from expectations was nearly the same for each subdivision, data were pooled and the analysis was made over the entire location.

TABLE 3. *Observed five-locus gametic frequencies (%) and relative deviations from expected frequencies (in parentheses) for the eight (among 32) most frequent types in CSA-1*

| Gametic type* | Subdivision *A* | *B* | *C* | *D* | Location total |
|---|---|---|---|---|---|
| 21112 | 91.5(13.7) | 39.9(24.4) | 16.9(15.5) | 1.9(1.8) | 56.7(38.0) |
| 12221 | 1.7(1.7) | 4.1(3.8) | 27.9(22.0) | 31.9(9.2) | 11.3(11.2) |
| 12211 | 0.1(0.1) | 1.5(0.8) | 3.2(-0.2) | 30.5(9.4) | 4.0(3.5) |
| 11112 | 0.7(-4.2) | 4.8(-1.8) | 2.7(1.4) | 0.6(0.4) | 1.8(-5.1) |
| 21121 | 0.7(0.4) | 5.8(3.4) | 8.8(4.5) | 5.7(5.3) | 4.0(0.7) |
| 21221 | 0.0(0.0) | 4.6(3.3) | 5.6(0.1) | 1.6(0.2) | 2.2(1.0) |
| 12212 | 0.4(0.4) | 10.7(8.9) | 1.5(-0.3) | 0.3(-1.8) | 2.2(1.2) |
| 22221 | 0.2(0.2) | 3.9(3.3) | 6.5(0.2) | 2.0(-3.9) | 2.5(2.0) |
| *N* | 577 | 174 | 277 | 111 | 1218† |

*N* = number of individuals scored.

* $E_4$, $E_9$, $E_{10}$, $P_5$, $APX_5$.

† Includes 79 individuals collected at random over the entire location.

TABLE 4. *Observed frequencies (%) of representative two-locus genotypes and relative deviations from products of single-locus genotypic frequencies (in parentheses) for location CSA-1*

| Two-locus pairs | Genotype 11,11 | 11,22 | 22,11 | 22,22 | 11,12 | 12,11 | 12,22 | 22,12 | 12,12 |
|---|---|---|---|---|---|---|---|---|---|
| $E_4$,$E_9$ | 4.2 | 64.0 | 19.1 | 5.5 | 0.6 | 1.6 | 1.0 | 1.2 | 2.8 |
| $\chi^2$=973 | (-12.9) | (15.5) | (12.7) | (-12.8) | (-2.6) | (0.2) | (-2.8) | (0.0) | (2.6) |
| $E_4$,$P_5$ | 9.6 | 14.1 | 59.2 | 10.2 | 1.1 | 1.3 | 1.6 | 1.1 | 1.6 |
| $\chi^2$=394 | (-7.8) | (7.7) | (9.8) | (-8.1) | (0.2) | (-1.9) | (0.5) | (-1.6) | (1.5) |
| $E_{10}$,$P_5$ | 60.7 | 8.5 | 8.5 | 16.7 | 1.1 | .9 | .7 | 1.1 | 1.8 |
| $\chi^2$=616 | (11.4) | (-9.7) | (-10.0) | (9.9) | (-1.7) | (-1.5) | (-0.1) | (0.0) | (1.7) |
| $P_5$,$APX_5$ | 5.9 | 63.1 | 24.5 | 1.1 | 1.1 | 1.3 | .9 | .4 | 1.7 |
| $\chi^2$=1072 | (-16.3) | (17.5) | (16.3) | (-15.8) | (-1.2) | (0.1) | (-1.6) | (-0.4) | (1.6) |

*N* = number of individuals examined.
$\chi^2[7] \geq 24.3$, $P \geq 0.001$.

The main features of the two-locus results (representative cases given in Table 4) can be summarized as follows: (*i*) highly significant excesses or deficiencies occurred in all homozygous classes for each of the 10 pairwise combinations of loci; (*ii*) the 11,22 and 22,11 homozygotes were in excess for locus pairs $E_4$–$E_9$, $E_4$–$E_{10}$, $E_4$–$P_5$, $E_9$–$APX_5$, $E_{10}$–$APX_5$, and $P_5$–$APX_5$, whereas the opposite was the case for the $E_4$–$APX_5$, $E_9$–$E_{10}$, $E_9$–$P_5$, and $E_{10}$–$P_5$, locus pairs; (*iii*) the extent of the excesses or deficiencies was remarkably consistent for all locus pairs (11–17%), except for the three pairs involving $P_5$ and the esterase loci ($E_4$–$P_5$, $E_9$–$P_5$, $E_{10}$–$P_5$) where the excesses and deficiencies were smaller (about 9–11%); (*iv*) deviations were largest for the tightly linked $P_5$–$APX_5$ pair but no larger in general for the loosely linked pairs ($E_{10}$–$APX_5$, $E_{10}$–$P_5$) than for the unlinked pairs ($E_4$–$E_9$, $E_4$–$P_5$, $E_4$–$APX_5$, $E_9$–$E_{10}$, $E_9$–$P_5$, $E_9$–$APX_5$). Thus, the results of the genotypic analysis parallel those of the gametic analysis and lead to the same conclusions.

The data of Table 4 bring out one additional feature not apparent from the gametic analysis, i.e., that correlated inheritance of multi-locus units causes homozygosity and heterozygosity to be concentrated in relatively few individuals. Thus, for each of the 10 two-locus combinations there is an excess of double homozygotes and double heterozygotes and a deficiency of genotypes that are homozygous for one locus and heterozygous for the other. Summed over all 10 two-locus pairs, excesses of double homozygotes and double heterozygotes are 206 (16.9%) and 207 (17.0%), respectively, and there is a corresponding deficiency of 413 (33.9%) among the mixed homozygous–heterozygous genotypes.

Considering all five loci simultaneously, there are $3^5 = 243$ possible genotypes, among which 32, 80, 80, 40, 10, and 1 are homozygous at 5, 4, 3, 2, 1, and 0 loci, respectively. Table 5 gives observed and expected numbers for each of these six classes, as well as observed and expected numbers of the

TABLE 5. *Comparison of observed numbers with expected numbers computed as products of single-locus genotypic frequencies, and deviations for each of the homozygous and heterozygous classes of genotypes, and for the eight most frequent genotypes in Population CSA-1*

| Genotype or Genotypic Class | Observed | Expected | Deviation |
|---|---|---|---|
| 21112/21112 | 663 | 190 | 473 |
| 12221/12221 | 124 | 2 | 122 |
| 12211/12211 | 46 | 5 | 41 |
| 22112/22112 | 17 | 71 | -54 |
| 21121/21121 | 58 | 34 | 24 |
| 21221/21221 | 21 | 13 | 8 |
| 12212/12212 | 24 | 9 | 15 |
| 22221/22221 | 21 | 5 | 16 |
| Remaining (24) 5-locus homozygotes | 127 | 659 | -532 |
| Sum of: | | | |
| 5-locus homozygotes (32) | 1101 | 988 | 113 |
| 4-locus homozygotes (80) | 50 | 211 | -161 |
| 3-locus homozygotes (80) | 27 | 18 | 9 |
| 2-locus homozygotes (40) | 21 | 0.7 | 20 |
| 1-locus homozygotes (10) | 13 | 0.02 | 13 |
| 5-locus heterozygote (1) | 6 | 0.0001 | 6 |

$N$ = 1218 individuals.

eight most frequent genotypes in the population. The most striking feature of the results are the great excesses of the 21112/21112 and 12221/12221 homozygotes, indicating these two homozygotes have great selective advantage over all other homozygotes. The 12211/12211 homozygote, which differs from the 12221/12221 homozygote only at the $P_5$ locus, was also present in large excess. Analysis of this excess showed that it was largest in the two more mesic subdivisions, providing further evidence that substitution of the $P_5^{(1)}$ for the $P_5^{(2)}$ allele has a small effect on fitness in xeric habitats. In contrast, the large deficiency of the 22112/22112 homozygote, which differs from the 21112/21112 homozygote only at the $E_9$ locus, suggests that substitution of the $E_9^{(2)}$ for the $E_9^{(1)}$ allele leads to serious reduction in fitness. There was a modest excess of the other four more frequent genotypes in the population, which differ from the 21112/21112 or 12221/12221 homozygotes at only one or two loci. In contrast, there was a deficiency of nearly all of the remaining 24 homozygotes, most of which represent more extreme recombinants relative to the two most favored types. Summed over these 24 genotypes, observed and expected numbers were 127 and 659, representing a total deficiency of 532 individuals, or 43.7% relative to total population size of 1218. It is therefore apparent that 21112/21112 and 12221/12221 are highly favored homozygotes and that homozygotes representing increasing departures from the allelic constitution of these two most favored types lead to increasing inferiority in fitness.

The effect of the inheritance of the correlated five-locus blocks of genes on concentrating homozygosity and heterozygosity in relatively few individuals can also be seen from Table 5: there is a great excess of quintuple homozygotes and multiple heterozygotes (especially quadruple heterozygotes and the quintuple heterozygote), whereas single heterozygotes are correspondingly deficient.

It is also of interest to compare observed with expected levels of heterozygosity computed from single-locus allelic frequencies, assuming no selection and inbreeding equilibrium. Expected frequencies of heterozygotes under these assumptions are $2pq(1 - F_e)$, where $F_e = s/1 + t$. Previous estimates (2, 4) show that the proportion of outcrossing $t$ ($=1 - s$, where $s$ = the proportion of self fertilization) $\leq 0.02$ in CSA-1, so that $F_e \geq 0.96$. Observed numbers of heterozygotes significantly exceeded expected numbers ($\chi^2_{[1]} > 15$, $P < 0.001$) for each of the five enzyme loci (observed:expected = 56,19; 65,19; 41,19; 48,19; 39,21 for $E_4$, $E_9$, $E_{10}$, $P_5$, $APX_5$, respectively). Also fixation indices (10), which measure realized inbreeding, were three or more standard errors smaller than $F_e$ for each locus. Thus, there was a significant excess of heterozygotes at each locus, indicating that selection favors heterozygotes. This result is in accord with previous studies of CSA-1, as well as studies of other populations of *A. barbata* (refs. 2, 4, Clegg and Allard, unpublished results). However, observed and expected numbers of specific heterozygotes among the 211 possible types are too small in the sample studied to permit analyses of heterosis on a multilocus basis; consequently, it is not known whether the observed excesses of heterozygotes result from single-locus and/or multilocus heterotic effects.

**Geyserville**

The Geyserville population is polymorphic for four of the six enzyme loci scored: locus $E_1$ is polymorphic for three alleles ($E_1^{(1)}$, $E_1^{(2)}$, $E_1^{(3)}$) and loci $E_4$, $E_9$, and $E_{10}$ for the same alleles as the CSA-1 population. Cluster analysis of genetic identity values for each of the 20 sampling plots divided the site into three subdivisions, labeled $A$, $B$, and $C$, which differed significantly in gene frequencies. The $A$ and $B$ subdivisions included, respectively, 5 and 6 contiguous plots located in the middle of the site, whereas the $C$ subdivision included 7 plots at one end and 2 plots at the opposite end of the site. We were

TABLE 6. *Allelic frequencies in the three subdivisions of the Geyserville site*

| Subdivision | $N$ | Locus and allele $E_4^{(1)}$ | $E_9^{(2)}$ | $E_{10}^{(2)}$ | $E_1^{(3)}$ | $E_1^{(2)}$ |
|---|---|---|---|---|---|---|
| *A* | 298 | .04 | .54 | .87 | .19 | .81 |
| *B* | 333 | .53 | .88 | .98 | .33 | .53 |
| *C* | 698 | .18 | .83 | .99 | .66 | .30 |
| Total | 1327 | .24 | .78 | .96 | .47 | .47 |

unable to correlate the differences in allelic frequencies with discernible features of the physical or biotic environment of the three subdivisions.

Allelic frequencies for each of the three subdivisions and for the entire site are given in Table 6. The predominant alleles for the polymorphic loci are the "xeric" $E_9^{(2)}$, $E_{10}^{(2)}$ and the "mesic" $E_1^{(2,3)}$, $E_4^{(2)}$, and *b* (gray lemma) alleles. The site is fixed for the "xeric" $P_5^{(2)}$ and the "mesic" $APX_5^{(3)}$ alleles. A distinct pattern of heterogeneity exists among the three subdivisions with the major differences being the higher frequencies of alleles $E_{10}^{(1)}$ and $E_1^{(2)}$ in subdivision *A*, and the lower frequencies of alleles $E_9^{(2)}$, $E_4^{(2)}$, and $E_1^{(2)}$ in subdivisions *A*, *B*, and *C*, respectively. Thus, the site tends to be somewhat more "mesic" than "xeric" with respect to allelic frequencies. Allelic frequencies thus are consistent with our subjective appraisal that the site is unusual environmentally.

The results of analyses of two-locus gametic frequencies, similar to those made for CSA-1, can be summarized as follows: (*i*) significant gametic phase disequilibrium was found for each of the six two-locus combinations within each of the three subdivisions, and also over the entire population; however, $D'$ and $r$ values were generally smaller than for CSA-1 ($\bar{D}'$ and $\bar{r}$ for the entire population were 0.22 and 0.32, compared to 0.64 and 0.61, respectively for CSA-1); (*ii*) there was a consistent excess of double homozygotes and double heterozygotes and a consistent deficiency of single homozygotes; (*iii*) in general, the two-locus combinations favored were not the same ones favored at Calistoga.

Observed three-locus gametic frequencies for loci $E_4$, $E_9$, and $E_{10}$ and relative deviations from products of single-locus allelic frequencies for the 4 most common (among 8) gametic types are given in Table 7. Neither the standard mesic (21112) nor standard xeric (12221) gametic types occur at Geyserville because the population is fixed for alleles $P_5^{(2)}$ and $APX_5^{(3)}$. However, the $E_4^{(1)}$, $E_9^{(2)}$, $E_{10}^{(2)}$ gametic type, which preserves the standard xeric allelic configuration for these loci, is predominant in subdivision *B* (46.8%) and it is the second most common gametic type over the entire site. Also the 222 gametic type, which differs from the standard xeric gametic type only at $E_4$ is by far the most frequent (55.9%) at Geyserville. In contrast, the standard mesic gametic type (211) is rare (1.8%), and the type that is most similar to it (212) occurs in only moderate frequency (16.9%). Thus, the standard xeric configuration of alleles is less modified at Geyserville than the standard mesic configuration, and by this criterion the site is more xeric than mesic.

An analysis of four-locus ($E_4$, $E_9$, $E_{10}$, $E_1$) genotypic frequencies at the Geyserville site shows that genotypes present in high frequency differ in the three subdivisions and that a different genotype is present in highest frequency in each subdivision (2222/2222, 1222/1222, 2223/2223 in *A*, *B*, and *C*, respectively). This result, together with the results reported in Tables 6 and 7, establishes that the sampling area at Geyserville is occupied by three distinct populations, each with its unique genetic structure. The pattern of microdifferentiation at Geyserville thus differs drastically from that of CSA-1, which is heterogeneous with respect to allelic frequencies but homogeneous with respect to organization at the multilocus level. The genotypic frequency data also establish that multiple homozygotes and multiple heterozygotes are in excess and single heterozygotes are in deficiency within subdivisions *A*, *B*, and *C*. This result provides additional evidence that substantial and unique multilocus organization has developed in each subdivision.

TABLE 7. *Observed three-locus gametic frequencies (%) and relative deviations from products of single-locus allelic frequencies (in parentheses) for the 4 most common (among 8) gametic types at Geyserville*

| Gametic type* | Subdivision A | Subdivision B | Subdivision C | Site total |
|---|---|---|---|---|
| 211 | 7.4(1.6) | 0.0(-0.1) | 0.3(0.1) | 1.8(1.1) |
| 122 | 2.9(0.9) | 46.8(1.1) | 14.4(-0.5) | 19.9(2.2) |
| 212 | 37.6(-0.9) | 6.0(0.5) | 13.2(-0.5) | 16.9(0.6) |
| 222 | 45.1(0.4) | 39.9(-0.9) | 68.2(1.0) | 55.9(-1.0) |

* $E_4$, $E_9$, $E_{10}$.

As with CSA-1, it is of interest to compare observed heterozygosity with values expected, assuming no selection. The proportion of outcrossing at this site has been estimated to be 0.02 (Clegg and Allard, unpublished data), so that $F_e$ = 0.96. Fixation indices for loci $E_4$, $E_9$, $E_{10}$, and $E_1$ are all significantly ($P < 0.001$) smaller than $F_e$, and observed numbers of heterozygotes significantly ($P < 0.001$) exceed expected numbers (observed:expected = 98:57, 67:39, 39:12, 12:4, respectively). The selective advantage of heterozygotes over homozygotes at Geyserville is therefore about equal to that observed in CSA-1.

In summary, the three populations at the Geyserville site have some features in common with the general pattern found throughout California (exemplified by CSA-1). However, they also differ in many ways, and the differences might be predicted to occur in an atypical environment featuring local xerism in a highly mesic geographical region.

## DISCUSSION

Dobzhansky developed the concept of coadaptation more than two decades ago to account for differences in the adaptive properties of chromosomes that have different gene arrangements due to inversions. He recognized that the suppression of recombination that attends heterozygosis for inversions would guard concordant gene complexes from dissipation by recombination, and he postulated that inversions binding together favorable combinations of alleles would be favored by selection. Subsequently, it has been shown theoretically and by numerical methods that selection in combination with suppression of recombination due to linkage, inbreeding, and/or restriction of population size can also lead to the buildup of nonrandom associations among alleles at different loci (gametic phase disequilibrium) so that favorable groups of genes are selected as coadapted units (11–15). Experimental evidence pertaining to correlated complexes has been difficult to obtain for various technical reasons, including the large samples that must be examined when several loci are considered simultaneously. Recently, however, Clegg *et al.* (9) have demonstrated the development of striking correlations in allelic state over four enzyme loci, including correlations among unlinked loci, in two experimental populations of *Hordeum vulgare* L. The present study of natural populations of *A. barbata* provides a second experimental demonstration,

also at the level of the gene, that selection operating on coadapted allelic complexes is an important determinant of the genetic structure of populations.

Results for the CSA-1 location, which appears to be a typical polymorphic population of the cool summer region of California, strongly support earlier circumstantial evidence that $E_4^{(1)}$, $E_9^{(2)}$, $E_{10}^{(2)}$, $P_5^{(2)}$, $APX_5^{(1)}$, and $E_4^{(2)}$, $E_9^{(1)}$, $E_{10}^{(1)}$, $P_5^{(1)}$, and $APX_5^{(2)}$, respectively, are favorable interacting allelic complexes inherited as units in the great majority of habitats occupied by *A. barbata* in California (1, 2). In the CSA-1 population, the 12221/12221 and 21112/21112 homozygotes are predominant among 243 possible genotypes and both occur in much higher frequency than predicted on the basis of expectations calculated as products of single-locus genotypic frequencies. These two homozygotes are highly favored by selection and the excess frequencies in which they occur leads to the great excesses in which the balanced opposite 12221 and 21112 gametic types occur in the population. The inheritance of these two allelic complexes as units in turn leads to excesses in the frequencies of multiple homozygotes and heterozygotes, and to deficiencies in the frequencies of single heterozygotes.

An overall excess of heterozygotes was also observed in the CSA-1 population, but the low expected and observed numbers of the 211 different heterozygous genotypes preclude analysis on a multilocus basis. Hence, it is not possible to determine from the present data whether the observed excesses result from single-locus heterosis, from favorable heterozygous interactions among particular combinations of alleles at two-, three-, or four-loci, or to heterozygous interactions that endow the five-locus 12221/21112 heterozygote with superior fitness.

Viewed in terms of Wright's (16) concept of adaptive topographies, populations that occupy intermediate habitats (such as CSA-1) feature two sharp adaptive peaks corresponding to the 12221/12221 and 21112/21112 homozygotes (each with adjoining foothills representing closely related genotypes) and one or more additional peaks representing the favored heterozygote or heterozygotes. The relative height of the two peaks representing the homozygotes is correlated with environment. The more mesic the environment the higher the peak corresponding to 21112/21112 and vice versa, until in extreme mesic and extreme xeric environments fixation occurs and only a single monolithic peak remains.

In intermediate habitats, correlated inheritance is incomplete with the result that all of the 32 gametic types are expected to occur in large populations (29 were found in our sample of $N = 1218$ individuals from CSA-1). Thus, "leakage" in the system provides free genetic variability for quick response to temporal changes in the environment and for opportunistic colonization of unusual habitats, such as those at the Geyserville site. This population, which occupies a locally xeric site in the highly mesic north coastal region of California, features some combinations of alleles characteristic of typical habitats occupied by *A. barbata*, but also some unique combinations of alleles.

This study also establishes that spatial differentiation occurs on an extremely small scale in *A. barbata* and that selection operates in different directions in different microniches. Since it is known that migration (gene flow) occurs among spatially isolated populations (1, 2), it presumably also occurs among contiguous microniches. The population biology of *A. barbata* thus appears to provide opportunity for complex interactions between multiniche selection and migration that in theory could contribute to excesses of heterozygotes and provide numerous possibilities for stable polymorphism (17).

This work was supported in part by NIH Grant GM 10476 and NSF Grant GB 13213. G. R. B. was a postdoctoral fellow on NIH Training Grant GM 00701.

1. Clegg, M. T. & Allard, R. W. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1820–1824.
2. Hamrick, J. L. & Allard, R. W. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 2100–2104.
3. Dobzhansky, Th. (1970) *Genetics of the Evolutionary Process* (Columbia University Press, New York).
4. Marshall, D. R. & Allard, R. W. (1970) *Genetics* **66**, 393–399.
5. Marshall, D. R. & Allard, R. W. (1969) *J. Hered.* **60**, 17–19.
6. Nei, M. (1972) *Amer. Natur.* **106**, 283–292.
7. Sokal, R. R. & Sneath, P. H. A. (1963) *Principles of Numerical Taxonomy* (W. H. Freeman and Co., San Francisco).
8. Lewontin, R. C. (1965) *Genetics* **49**, 49–67.
9. Clegg, M. T., Allard, R. W. & Kahler, A. L. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 2474–2478.
10. Brown, A. H. D. (1970) *Genetics* **41**, 399–406.
11. Lewontin, R. C. & Kojima, K. (1960) *Evolution* **14**, 458–472.
12. Jain, S. K. & Allard, R. W. (1966) *Genetics* **53**, 633–659.
13. Sved, J. A. (1968) *Genetics* **59**, 543–563.
14. Karlin, S. & Feldman, M. W. (1970) *Theor. Pop. Biol.* **1**, 39–71.
15. Franklin, I. & Lewontin, R. C. (1970) *Genetics* **65**, 707–734.
16. Wright, S. (1931) *Genetics* **16**, 97–159.
17. Karlin, S. & McGregor, J. (1972) *Theor. Pop. Biol.* **3**, 186–210.