

## HITCH-HIKING: AN ALTERNATIVE TO COADAPTATION FOR THE BARLEY AND SLENDER WILD OAT EXAMPLES

PHILIP W. HEDRICK and LARRY HOLDEN

*Division of Biological Sciences, University of Kansas, Lawrence, Kansas 66045*

Received 1.iii.79

### SUMMARY

The associations of alleles at different electrophoretic loci observed in barley and wild oats can be explained by hitch-hiking due to the mating system. In other words, selection at other loci in the genome may have resulted in gametic disequilibrium or a change in gametic disequilibrium at neutral loci in these highly self-fertilised plants. This explanation appears to be a simple alternative hypothesis to account for the change in gametic disequilibrium in barley and the presence of gametic disequilibrium in wild oats. However, it appears difficult to differentiate experimentally between the hitch hiking and the previously proposed coadaptation hypotheses.

### I. INTRODUCTION

IT has often been recognised that changes in gene frequency at a particular locus can be affected by the genetic background and in particular by selection at closely-linked loci (e.g. Lewontin, 1974; Nei, 1975). An example is the phenomenon of hitch-hiking (Kojima and Schaffer, 1967; Maynard Smith and Haigh, 1974; Ohta and Kimura, 1975; Thompson, 1977; Wagener and Cavalli-Sforza, 1975) in which neutral or nearly-neutral loci are carried along by a closely-linked locus undergoing selection. In fact, not only changes in gene frequency but changes in the association of alleles at different loci, gametic disequilibrium, may be caused by selection at a linked locus (Thomson, 1977). A number of authors have pointed out that inbreeding may affect the decay of associations of alleles at different loci (e.g. Bennet and Binet, 1956; Cockerham and Rawlings, 1967; Ohta and Cockerham, 1974). Recently, one of us (Hedrick, 1979) has shown using computer simulation that when a high proportion of the population reproduces by self-fertilisation, hitch-hiking effects can occur in quite complex patterns at neutral loci which are completely unlinked to a selected locus. Under many conditions, this partial-selfing hitch-hiking has a greater effect on gene frequency and gametic disequilibrium change than does linkage-induced hitch-hiking.

Allard and his co-workers (Allard *et al.*, 1972a and b; Clegg and Allard, 1972; Hamrick and Allard, 1972; Weir *et al.*, 1972, 1973) carried out detailed genetic studies in two highly-selfed plants, *Hordeum vulgare* (cultivated barley) and *Avena barbata* (slender wild oat). They found very high associations between alleles at different electrophoretic loci in both species. Except for a few cases, such gametic disequilibrium has not been found in outbreeding populations (Hedrick *et al.*, 1978). Furthermore, in two populations of *Hordeum* they observed that the amount of gametic disequilibrium actually increased over generations. Results such as these

have been used to argue for the existence of coadaptation in these highly self-fertilised species. For a summary and discussion of this work, see Hedrick *et al.* (1978).

We would like to present a somewhat different hypothesis to explain the data from *Avena* and *Hordeum*. This is an explanation based on hitch-hiking due to the mating system, *i.e.* where selection occurs at other loci in the genome and affects frequencies of the electrophoretic loci because of the high degree of self-fertilisation. Of course this is an alternative hypothesis to coadaptation and with the present information it appears unlikely that either hypothesis can be falsified.

## 2. MODEL

A three-locus, partial-selfing, discrete-generation model was used to simulate the change in gametic frequencies for the two sets of data. This involves a set of 36 simultaneous equations which give the frequencies of all the genotypic frequencies. The specific genotype frequencies are determined by the genotype frequencies in the previous generation, selection, recombination, and the proportion of self-fertilisation, and are a direct extension of the two-locus model (*e.g.* Holden, 1979).

From these genotypic frequencies, gametic disequilibrium for alleles at two loci may be measured in several ways (see Hedrick *et al.*, 1978, for a more thorough discussion). One of the most common is

$$D = x_1 x_4 - x_2 x_3$$

where the frequency of gametes,  $A_1 B_1$ ,  $A_1 B_2$ ,  $A_2 B_1$  and  $A_2 B_2$  are  $x_1$ ,  $x_2$ ,  $x_3$  and  $x_4$ , respectively. In order to make the gametic disequilibrium parameter less sensitive to changes in gene frequency, the following measure can be used.

$$D' = D/D_{\max}$$

where  $D_{\max}$  is the magnitude of the maximum  $D$  possible for a given set of gene frequencies at the two loci. For three loci, the parameter below, which is the deviation when pairwise disequilibrium and single-locus gene frequencies are taken into account, can be used.

$$D_{ABC} = x_1 - p_1 D_{BC} - p_2 D_{AC} - p_3 D_{AB} - p_1 p_2 p_3$$

Here,  $D_{AB}$ ,  $D_{AC}$ , and  $D_{BC}$  are the pairwise disequilibrium between the loci and  $p_1$ ,  $p_2$  and  $p_3$  are the frequencies of alleles  $A_1$ ,  $B_1$  and  $C_1$ .

In these simulations, we have assumed that loci  $B$  and  $C$  are neutral and represent allozyme loci. Locus  $A$  is an unidentified locus which is undergoing differential selection. Loci  $B$  and  $C$  may be linked to each other depending upon which allozyme loci are under consideration. For these simulations, selected locus  $A$  is always assumed to be *unlinked* to the simulated allozyme loci.

## 3. RESULTS

### (i) *Hordeum (Barley)*

In the two experimental populations of barley started from crosses of collections from around the world, changes in gene frequency and gametic

disequilibrium have been observed for four esterase loci (Allard *et al.*, 1972b). We will concentrate on one of these populations, Composite Cross V (CCV), although a similar analysis could be carried out on the other population. In general, the gene frequency changes at these loci were rather small (an average of 0.086 during 26 generations) for the two most common alleles at each locus but there were sizeable changes in the values of gametic disequilibrium between pairs of loci. The pairwise gametic disequilibrium values for three linked esterase loci ( $E_A$ ,  $E_B$ ,  $E_C$ ) for all the generations examined are plotted in fig. 1 (solid lines and open circles).

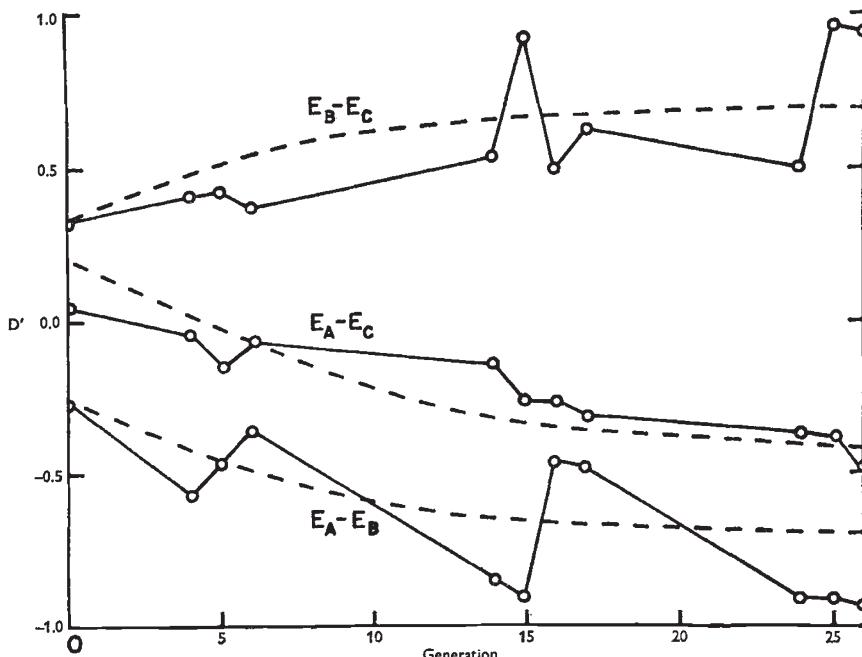


FIG. 1.—The observed changes in gametic disequilibrium for three barley esterase loci  $E_A$ ,  $E_B$  and  $E_C$  (solid lines) and simulations to mimic these changes (broken lines).

We were able to model these changes by assuming simple directional selection at a locus *unlinked* to the allozyme loci. Changes resembling the observed values were found for the initial gametic arrays indicated as  $E_A-E_B$ ,  $E_A-E_C$ , and  $E_B-E_C$  in table 1. By using these initial values we have approximated the behaviour of both  $D$  and  $D'$  quite well (as shown in fig 1) although other gametic arrays also satisfactorily mimic the changes in  $D$  and  $D'$ . In these simulations, the reported recombination values were used, *i.e.* recombination between  $E_A-E_B$ ,  $E_A-E_C$  and  $E_B-E_C$  were 0.023, 0.0048, and 0.0061, respectively, and selfing was assumed to be the estimated value of 0.9943. Since the population was started from a series of crosses, a random distribution of gametes into zygotes was assumed in the initial generation. For the locus undergoing selection, there was a 20 per cent selection difference between the homozygotes, and the heterozygote was assumed intermediate. There was no gene frequency change at the neutral

TABLE 1

The three initial gamete frequency arrays used to mimic the changes in the barley population for loci pairs  $E_A-E_B$ ,  $E_A-E_C$ , and  $E_B-E_C$  and the initial gametic array used to mimic the changes in the *Avena* population. Below are given the pairwise D and D' values and  $D_{ABC}$  for these arrays

Gamete	Initial frequency			<i>Avena</i>	
	<i>Hordeum</i>		$E_B-E_C$		
	$E_A-E_B$	$E_A-E_C$			
$A_1B_1C_1$	0.0	0.0	0.35	0.5	
$A_1B_1C_2$	0.3	0.2	0.0	0.0	
$A_1B_2C_1$	0.3	0.2	0.0	0.0	
$A_1B_2C_2$	0.0	0.0	0.3	0.0	
$A_2B_1C_1$	0.25	0.3	0.0	0.0	
$A_2B_1C_2$	0.0	0.0	0.2	0.0	
$A_2B_2C_1$	0.0	0.0	0.15	0.0	
$A_2B_2C_2$	0.15	0.3	0.0	0.5	
$D_{AB}$	-0.03	0.0	-0.0075	0.25	
$D_{AC}$	-0.03	0.0	0.025	0.25	
$D_{BC}$	-0.0525	0.05	0.075	0.25	
$D_{ABC}$	-0.1170	-0.12	0.1125	0.0	
$D'_{AB}$	-0.167	0.0	-0.048	1.0	
$D'_{AC}$	-0.167	0.0	0.143	1.0	
$D'_{BC}$	-0.259	0.20	0.333	1.0	

loci for the  $E_A-E_C$  simulation and changes of approximately 0.02 for alleles in the other runs. We have not tried to mimic simultaneously the change in gene frequencies and gametic disequilibrium for all four esterase loci, but have concentrated on an illustration of how the most critical observations, the increase in gametic disequilibrium over time, can occur between neutral loci. Of course, greater selection, linkage of the selected locus, other gametic arrays, and different initial zygotic proportions will often increase the generation of disequilibrium and change in gene frequencies so that the values generated by these specific simulations are much below the maximum possible (Hedrick, 1979).

From the similarity of the observed values and the simulation, it appears that changes such as those observed in this population could be the result of hitch-hiking of allozyme loci when there is directional selection at other loci in the genome. Since the experimental populations of barley represent the gene pools of a world-wide collection, it would be very unlikely that the original population was adapted to the Davis, California, environment. Directional selection against one or more unfavourable alleles in the population would be an obvious and simple expectation. Such selection can easily explain the observed changes in gametic disequilibrium and gene frequency without the need to invoke coadaptation.

### (ii) *Avena* (*wild oats*)

A portion of a population of the slender wild oat, *Avena barbata*, on a hillside has been analysed in detail with regard to the gene frequency and gametic disequilibrium for five allozyme loci (Allard *et al.*, 1972a). In

the most mesic areas (A), the mean frequency of the most common alleles at these five loci was 0.95 while in the xeric areas (D), these alleles have an average frequency of only 0.24. The gametic disequilibrium for these two sites is quite high and generally similar in value. The average  $D'$  within site for areas A and D, and the total  $D'$  value, are given in Table 2 for the three linked locus pairs and the unlinked pairs. Except for the locus pair 27 map units apart ( $E_{10}-P_5$ ), gametic disequilibrium is greater with tighter linkage.

*Avena barbata* was "introduced into California less than 200 years ago during the mission period (1769-1823)" (Jain and Marshall, 1967). Assuming that this hillside population is 200 years old or less, we can

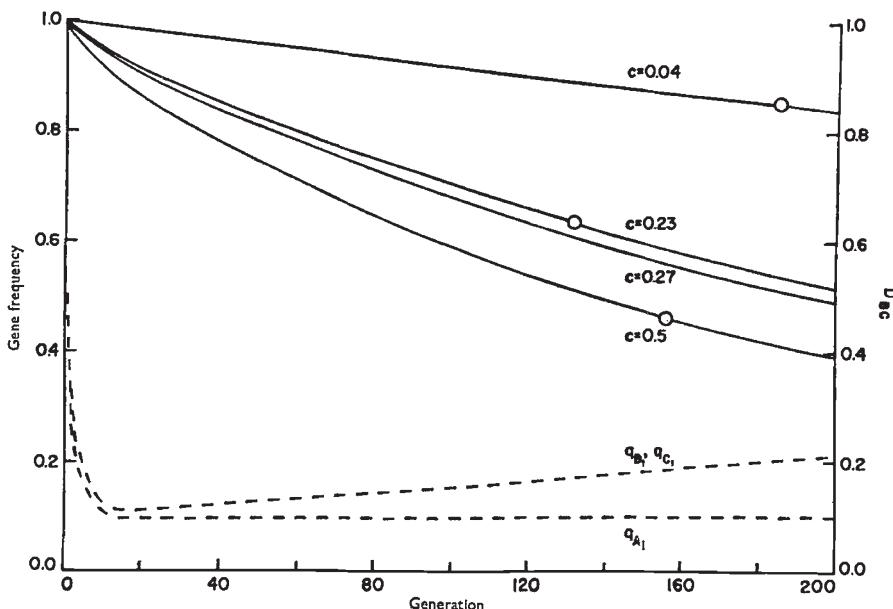


FIG. 2.—The changes in gene frequency (broken lines) and within-niche gametic disequilibrium (solid lines) for a two-niche simulation model, used to mimic possible changes in a population of *Avena barbata*. The open circles indicate the values of  $D'$  observed in the *Avena* population.

hypothesise how the gametic disequilibrium values could have resulted between neutral loci as a result of hitch-hiking. Let us assume that the hillside was colonised originally by equal numbers of genotypes  $A_1B_1C_1/A_1B_1C_1$  and  $A_2B_2C_2/A_2B_2C_2$  (initial gamete array in last column of table 1) and that the area is divided into two different niches where the fitnesses of the genotypes at a selected locus  $A_1A_1$ ,  $A_1A_2$  and  $A_2A_2$  are 1.5, 1.25, and 1.0 in one niche, and 1.0, 1.25 and 1.5 in the second niche. Assume also that there is limited migration between niches ( $m = 0.05$ ) and there is 98 per cent self-fertilisation (Hamrick and Allard, 1972).

Using these values, we obtain the gene frequencies (the gene frequency in the other niche is the complement of the one given) and within niche  $D'$  values given in fig. 2 for a time course of 200 generations. The simulated gene frequencies are of the same magnitude as those observed at the sites A

TABLE 2

*Gametic disequilibrium values for loci in *Avena barbata* and similar values for a simulation with neutral loci.*

	c	D' <sup>1</sup>		D'	
		within site	total	100	200
Avena loci	.04	.86	.88		
	.23	.64	.70		
	.27	.29	.56		
	.5	.46	.62		
Simulation		Generation	Generation		
		100	200	100	200
	.04	.919	.843	.956	.895
	.23	.712	.521	.845	.680
	.27	.688	.491	.832	.660
	.5	.600	.392	.785	.594

and D. For example, the difference in gene frequency between the two simulated niches is 0.68 after 100 generations, compared to the observed difference between sites A and D of 0.71. The simulated within-niche D' values and total D' values are given in Table 2 for generations 100 and 200 for the four levels of linkage. These simulated values bracket the observed values except for the  $E_{10}-P_5$  ( $c = 0.27$ ) locus pair and the total D' for the  $P_5-APX_5$  ( $c = 0.04$ ) locus pair where the simulated D' values are even higher than those observed. The open circles in fig. 2 indicate when between generations 100 and 200 the observed within-niche D' value occurs in the simulation for the other three map distances. Of course, greater selection, linkage of the selected locus, reduced migration between niches, and younger population age could make the simulated effects more pronounced.

As was the case in the barley example, it is again obvious that the observed gene frequencies and gametic disequilibrium values could be the result of hitch-hiking of the allozyme loci undergoing directional selection. Since there is no temporal information on gene and gametic frequencies in *Avena*, it is difficult to determine the cause of the present-day gametic arrays. Overall the data are consistent with a model like that suggested above, which assumed differential selection at a locus different from the allozyme loci examined.

#### 4. DISCUSSION

From these simulations, it appears that hitch-hiking resulting from partial-selfing is a realistic and simple alternative to explain the gene frequencies and gametic disequilibrium in the barley and wild oats examples. Perhaps the most critical assumptions for these simulations are the initial gametic arrays and the amount of selection. Concerning the initial genetic arrays in the barley example, B. Weir (pers. comm.) has suggested that "the present CCV populations all descend from a very small number of founder individuals." This will of course lead to the presence of two-way, three-way or higher-order gametic disequilibrium and the possibility of hitch-hiking effects. In the wild oats, we need only assume that the particular hillside studied was founded primarily with the two genotypes, since simulations with different initial frequencies of these two genotypes make little difference in the simulation results. This is likely, since the two homozygotes predominate throughout the region. Concerning the

amount of selection, Allard and Jain (1962) observed large directional changes for a number of quantitative traits in the barley population. Hamrick and Allard (1975) have demonstrated the presence of large phenotypic differences between *Avena* from the xeric and mesic sites in a common garden experiment. Furthermore, Hedrick (1979) has shown that if the selective differences are split between several loci there is an even greater synergistic effect on hitch-hiking. Overall then, even the assumptions concerning the initial gametic arrays and amount of selection, the least conservative of the assumptions, are not unlikely.

Clegg *et al.* (1978), in a recent paper on the experimental barley populations, have indicated their current belief which appears entirely consistent with our model. They state that "What is measured at the marker loci is the selective flux transmitted throughout the genome by the correlational structure of the entire multilocus distribution. We can conclude that there is substantial selection, but we can isolate neither the phenotypic structures nor the genetic loci causally related to the selection."

*Acknowledgments.*—We wish to thank Tony Brown, Mike Clegg, Jim Hamrick and Bruce Weir for their comments on this manuscript.

## 5. REFERENCES

ALLARD, R. W., BABBEL, G. R., CLEGG, M. T., AND KAHLER, A. L. 1972a. Evidence for co-adaptation in *Avena barbata*. *Proc. Nat. Acad. Sci.*, 69, 3043-3048.

ALLARD, R. W., AND JAIN, S. K. 1962. Population studies in predominantly self-pollinated species. II. Analysis of quantitative genetic changes in a bulk-hybrid population of barley. *Evolution*, 16, 90-101.

ALLARD, R. W., KAHLER, A. L., AND WEIR, B. S. 1972b. The effect of selection on esterase allozymes in a barley population. *Genetics*, 72, 489-503.

BENNETT, J. H. 1954. On the theory of random mating. *Ann. Eugenics*, 18, 311-317.

BENNETT, J. H., AND BINET, F. E. 1956. Association between Mendelian factors with mixed selfing and random mating. *Heredity*, 44, 5-13.

CLEGG, M. T., AND ALLARD, R. W. 1972. Patterns of genetic differentiation in the slender wild oat species *Avena barbata*. *Proc. Nat. Acad. Sci.*, 69, 1820-1824.

CLEGG, M. T., KAHLER, A. L., AND ALLARD, R. W. 1978. Estimation of life cycle components of selection in an experimental plant population. *Genetics*, 89, 765-792.

COCKERHAM, C. C., AND RAWLINGS, J. D. 1967. Apparent heterosis of a neutral gene with inbreeding. *Ciencia E. Cultura*, 19, 89-94.

HAMRICK, J. L., AND ALLARD, R. W. 1972. Microgeographical variation in allozyme frequencies in *Avena barbata*. *Proc. Nat. Acad. Sci.*, 69, 2100-2104.

HAMRICK, J. L., AND ALLARD, R. W. 1975. Correlations between quantitative characters and enzyme genotypes in *Avena barbata*. *Evolution*, 29, 438-442.

HEDRICK, P. W. 1979. Hitchhiking: A comparison of linkage and partial selfing. *Genetics* (in press).

HEDRICK, P. W., JAIN, S., AND HOLDEN, L. 1978. Multilocus systems in evolution. *Evolutionary Biology*, 11, 101-184.

HOLDEN, L. R. 1979. New properties of the two-locus partial-selfing model with selection. *Genetics* (in press).

JAIN, S. K., AND MARSHALL, D. R. 1967. Population studies in predominantly self-pollinating species. X. Variation in natural populations of *Avena fatua* and *A. barbata*. *Amer. Natur.*, 101, 19-33.

KOJIMA, K., AND SCHAFER, H. E. 1967. Survival process of linked mutant genes. *Evolution*, 21, 518-531.

LEWONTIN, R. C. 1974. *The Genetic Basis of Evolutionary Change*. Columbia, New York.

MAYNARD SMITH, J., AND HAIGH, J. 1974. The hitch-hiking effect of a favorable gene. *Genet. Res.*, 23, 23-35.

NEI, M. 1975. *Molecular Population Genetics and Evolution*. North Holland Publ. Co., Amsterdam.

OHTA, T., AND COCKERHAM, C. C. 1974. Detrimental genes with partial selfing and effects on a neutral locus. *Genet. Res.*, 23, 191-200.

OHTA, T., AND KIMURA, M. 1975. The effect of selected linked locus on heterozygosity of neutral alleles (the hitch-hiking effect). *Genet. Res.*, 25, 313-326.

THOMSON, G. 1977. The effect of a selected locus on linked neutral loci. *Genetics*, 85, 753-788.

WAGENER, D. K., AND CAVALLI-SFORZA, L. L. 1975. Ethnic variation in genetic disease: Possible roles of hitchhiking and epistasis. *Amer. J. Hum. Genet.*, 27, 348-364.

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

WEIR, B. S., AND COCKERHAM, C. CLARK. 1973. Mixed self and random mating at two loci *Genet. Res.*, 21, 247-262.