

## OUTCROSSING RATES AND HETEROZYGOSITY IN NATURAL POPULATIONS OF *HORDEUM SPONTANEUM* KOCH IN ISRAEL

A. H. D. BROWN\*, D. ZOHARY† and E. NEVO‡

\*CSIRO, Division of Plant Industry, P.O. Box 1600, Canberra A.C.T. 2601;

†Laboratory of Genetics, Hebrew University, Jerusalem, Israel; and

‡Institute of Evolution, Haifa University, Israel

Received 2.xii.77

### SUMMARY

The extent of cross-pollination was estimated quantitatively in 26 populations of wild barley (*H. spontaneum*) in Israel, using allozyme variation at 22 polymorphic loci. Individual population estimates varied from 0 to 9.6 per cent outcrossing. The overall average was 1.6 per cent, with a 95 per cent confidence range of 0.8-2.7 per cent. The evidence indicated that outcrossing was significantly higher in populations growing in the more mesic (2.1 per cent), than in the xeric regions (0.4 per cent). The average frequency of multilocus homozygosity for the 22 loci was 0.963 which was in agreement with the level expected under the observed high rate of self-fertilisation, indicating no heterozygous excess. However, the heterozygotes encountered, displayed a very high degree of multiple heterozygosity which arises from extensive gametic phase (or linkage) disequilibrium. In general, these results do not support the contention that the bulk of the genetic polymorphism found in *H. spontaneum* in Israel has arisen in recent times by periodic introgression from cultivated barley (*H. vulgare* L.). Rather they indicate that extensive genetic polymorphisms probably predates the domestication of the crop, and is maintained independently of introgression.

### 1. INTRODUCTION

It is now clearly recognised that reliable estimates of the parameters of the mating system are crucial to an understanding of the evolution of plant populations (Allard, 1975; Jain, 1975; Rick, 1976). We have undertaken a study of the genetic variation in natural populations of wild barley, *Hordeum spontaneum* Koch and thus a basic part of the study is an estimate of the level of outcrossing. The species is a highly successful colonising, self-fertilising annual of the Mediterranean basin and western Asia (Harlan and Zohary, 1966). Furthermore, because it is regarded as the sole progenitor of cultivated barley, with which it is fully interfertile, a knowledge of its genetic structure is of prime importance to an understanding of the evolution of barley. From the allozyme variation found at 28 loci in 28 natural populations in Israel, it is clear that these populations (i) are on average highly variable genetically, (ii) differ from one another in their level of variation, and (iii) show marked differences for the alleles they contain (Nevo, Brown and Zohary, 1978).

Wild barley has also been investigated for flavonoid patterns in comparison with lines of cultivated barley (Frost *et al.*, 1975; Frost and Holm, 1975). These workers found that *H. spontaneum* contains the three patterns (A, B and C) which occur commonly in various *H. vulgare* lines. In addition,

there occurs a somewhat rarer *S* pattern (15 per cent) which they postulate as the "true" primitive *H. spontaneum* flavonoid pattern because it is extremely rare in *H. vulgare*. They suggest that the high level of variation for flavonoids might now be present in *H. spontaneum* populations because of regular introgression from cultivated *H. vulgare*. The occurrence of such introgression in nature is well documented (Zohary, 1964, 1971). Frost and Holm (1975) are thus led to question whether "true" *H. spontaneum* can be found today.

However, there is an alternative to the introgression hypothesis which would account for the present high level of genetic variation within *H. spontaneum* (for both flavonoids and allozymes). The alternative hypothesis asserts that genetic variation is the primitive condition and that the various lines of cultivated barley represent samples drawn from the variable wild gene pool. The introgression hypothesis suggests that considerable levels of outcrossing would be prevalent in Israel in recent times, especially because the intermediate hybrid derivatives (*H. agrocithon*, *lagunculiforme*, and *intermedium*, Zohary, 1964) usually disappear in a few generations (Zohary, 1971). Furthermore, we might predict that since populations differ widely in their level of variation, they might also show marked concomitant variation in outcrossing rates, or have varied greatly in their frequency of recent proximity to cultivated barley. Here we report estimates of outcrossing in *H. spontaneum* which give little support to either of these two predictions of the introgression hypothesis because the rates are uniformly low. Instead, the results argue that populations of *H. spontaneum* represent considerable genetic resources which probably predate the evolution of cultivated forms.

## 2. MATERIALS AND METHODS

Individual spikes were collected at random from 28 populations of *H. spontaneum*. The locations of these populations were chosen to encompass the extraordinarily diverse ecological range in which the species occurs in Israel. Details of the sites of collection are given elsewhere (see Brown *et al.*, 1978). The populations can be grouped geographically into seven regions which possess broad climatic similarities: the Golan (population numbers 1-3), the Hule Valley and Upper Jordan (4-7), the Galilee and Carmel mountains (8-11), the central mountains (12-15), the Negev steppes (16-20), the lower Jordan Valley (21-23), and the Mediterranean coastal plain (24-28). The sites are numbered within regions approximately in a north to south direction.

The procedure for the estimation of genetic parameters (genotype frequencies and outcrossing rate) was design II of Brown (1975). This is a two-stage sampling procedure. First, a single seed from each spike was assayed for allozyme genotype. If this seed proved to be heterozygous at any of the marker loci, a further sample of four seeds was assayed to determine the maternal genotype. The procedure is optimally efficient in highly inbred populations. The 28 marker loci, determined on leaf or root samples from each 10-day old seedling were: *Acph-1*, -2, -3; *Adh-1*, -2; *Ald*; *Cat*; *Est-A*, -B, -C, -D, *Gdh*, *Got-1*, -2; *Mdh-1*, -2; *Nadh-1*, -2; *Pept-1*, -2; *Pepc*; *Pgi*; *Pgm*; 6 *Pgd-1*, -2; *To-1*, -2; *Gp*. Since polymorphism is required for outcrossing estimation the following relatively invariant loci

were not considered in this part of the study: *Ald*, *Cat*, *Nadhd-2*, *To-1*, and *Peptc*. The general protein (*Gp*) locus was also excluded as it may be partially encoded by non-nuclear DNA. Many of the remaining loci contribute information on outcrossing only in those few populations in which they were appreciably polymorphic (*i.e.* where the commonest allele had a frequency less than 0.95). The electrophoretic procedures and properties of the loci, the evidence for the genetic basis of the variation, and the estimates of the population frequencies of all alleles in all populations are given elsewhere (see Brown *et al.*, 1978). The aim of the laboratory routine was to assay all 28 loci on every individual. However, for various technical reasons the data are not complete. Generally we assumed that it was more efficient to assay additional families or sites rather than repeat materials already partially scored. Thus the statistical procedure for estimating outcrossing must take account of incomplete data. In addition, there were associations between the alleles of different loci in these populations (Brown *et al.*, 1977), so that estimates of outcrossing based on single loci would be highly correlated. Therefore we used a numerical, multilocus maximum likelihood procedure to estimate outcrossing as described in the Appendix.

### 3. RESULTS

Table 1 gives the following results for each of the 28 populations separately: the total number of families assayed ( $N$ ), the number of heterozygous maternal plants detected ( $h$ ), the observed number of genetically distinct outcrosses ( $O$ ), the number of polymorphic loci used in estimating outcrossing ( $P.L.$ ), the estimate of the maximum diversity parameter ( $H_e^{(0)}$ , defined in the Appendix) and its standard error, the maximum likelihood estimate of outcrossing and the lower and upper approximate confidence limits to this estimate at the 0.05 level of probability. These latter three values are expressed as percentages. The final column is explained below. Of the total of 1177 spikes included in this study, the first seed was heterozygous at any of the loci in only 43 (3.65 per cent) cases. The second samples from each of these families showed that 14 of the original heterozygotes must have arisen by outcrossing. The loci at which these outcrossing events were detected are given in table 2. This table shows that more than half the outcrosses (8/14) were multiply marked, and most (10/14) were detected at the well studied esterase loci (Kahler and Allard, 1970; Hvid and Nielsen, 1977).

The remaining 29 cases were seeds borne by heterozygous maternal plants. In addition to the published evidence in barley and other plants (see Brown *et al.*, 1974, 1975 and 1978 for discussion and references), the patterns observed in these segregating progeny arrays are evidence that the variation is genetic and conforms to Mendelian expectations. The number of segregating families found for each locus was, in decreasing order: *Est-B* (20); *Est-C* (15); *Nadhd-1* (7); *Acph-3* (6); *Acph-2* (5); *Adh-1*, *Pept-2* (3); *Adh-2*, *Est-A-D*, *Pgm* (2); *Cat*, *Got-2*, *Pept-1*, *Pgi*, *6 pgd-1* (1).

Theoretically the screening of only one seed per spike detects one half of maternal heterozygotes at a single locus. However, in practice a somewhat greater proportion (about 70 per cent) is recovered as discussed below. Except for two entirely monomorphic populations, the typical

TABLE 1

*Observed numbers of heterozygous families (h), outcrosses (O) in N spikes; Genetic diversity measures, and estimates of outcrossing t (see text for explanation)*

		Observed numbers			Diversity			Outcrossing (%)			Expected h ( $Q_x N$ )
		N	h	O	P.L.	$\hat{H}_e^{(0)}$	s.e.	$\hat{t}$	$t_{\min}$	$t_{\max}$	
A.	1 Hermon	30	0	0	0	0	—	—	—	—	—
	2 Shifon	49	0	0	4	0.63	0.03	0	0	9.5	0
	3 Afiq	31	0	0	7	0.90	0.04	0	0	10.2	0
B.	4 Tel Hay	50	1	0	6	0.80	0.03	0	0	7.5	0
	5 Rosh Pinna	41	0	0	7	0.57	0.09	0	0	12.3	0
	6 Gadot	49	0	1	9	0.96	0.01	2.1	0.1	9.0	3.0
	7 Tabigha	45	0	1	9	0.96	0.02	2.3	0.1	9.7	3.0
C.	8 Zefat	49	3	1	6	0.79	0.05	2.7	0.1	11.7	3.1
	9 Mt. Meron	48	0	0	0	0	—	—	—	—	—
	10 Ma'alot	50	1	0	6	0.62	0.06	0	0	9.8	0
	11 Damon	53	0	2	7	0.77	0.06	4.9	0.9	13.8	5.7
D.	12 Shechem	28	0	0	8	0.82	0.04	0	0	12.4	0
	13 Bar Giyyora	53	3	1	7	0.87	0.02	2.3	0.1	10.1	3.1
	14 Talpiyyot	27	6	2	8	0.99	0.01	9.6	1.9	25.9	6.7
	15 Eizariya	30	1	1	9	1.00	0.01	3.3	0.2	14.1	2.9
E.	16 Tel Shoqet	30	1	0	5	0.94	0.03	0	0	10.5	0
	17 Mashash	51	0	0	5	0.35	0.17	0	0	8.3	0
	18 Revivim	31	0	0	5	0.75	0.06	0	0	12.4	0
	19 Yeroham	32	0	0	6	0.89	0.04	0	0	10.0	0
	20 Sede Boqer	30	2	0	8	0.95	0.01	0	0	10.7	0
F.	21 Bet Shean	38	3	0	5	0.98	0.01	0	0	8.4	0
	22 Mehola	47	5	0	8	0.93	0.03	0	0	7.4	0
	23 Qilt	45	0	1	9	0.84	0.04	2.6	0.1	11.2	2.9
G.	24 Akhziv	48	0	1	8	0.79	0.03	2.6	0.1	11.2	2.9
	24 Atlit	49	0	2	9	0.92	0.02	4.4	0.2	11.2	5.8
	26 Caesarea	48	1	0	5	0.60	0.06	0	0	10.7	0
	27 Herzliyya	50	0	0	7	0.88	0.02	0	0	6.6	0
	28 Ashqelon	45	2	1	8	0.98	0.01	2.3	0.1	10.0	3.0

genetic structure of each population was homozygous polymorphism for four to nine loci with low heterozygosity.

This level of polymorphism was generally adequate to ensure a reasonable probability ( $H_e$ ) that if an outcross occurred, it would be detectable genetically. The lowest non-zero value of  $H_e$  was at Mashash and we therefore sampled two seeds per spike (Design IV in Brown, 1975) to improve precision. The standard errors of  $H_e$  were inversely related to the value of the estimate, so that estimates close to the maximal value ( $H_e = 1.0$ ) had very low errors. Of the 26 polymorphic sites, 20 gave estimates with standard errors less than or equal to 0.05. Thus this component of the error in estimating outcrossing is small in comparison with that arising from error in estimating the frequency of detectable outcrosses.

The estimates of outcrossing ( $t$ ) varied from zero (15 populations) to a maximum of 9.6 per cent at Talpiyyot. Although this later figure just falls within the confidence limits of most populations, it does seem to be an

TABLE 2

*Loci by which the fourteen outcrosses (O) were detected*

Site	Heterozygous loci
	<i>1-locus</i>
Atlit	<i>AcpH-2</i>
Gadot, Eizariya	<i>Est-B</i>
Talpiyyot, Akhziv, Atlit	<i>Nadhd-1</i>
	<i>2-loci</i>
Damon (both)	<i>Est-A, -C</i>
Tabigha, Ashqelon	<i>Est-B, -C</i>
Qilt	<i>Est-B, -D</i>
	<i>3-loci</i>
Zefat, Talpiyyot	<i>AcpH-2, Est-B, -C</i>
Bar Giyyora	<i>Est-A, -B, -C</i>

exceptional result. This population also exhibited a high frequency of heterozygous maternal plants and a unique alcohol dehydrogenase (*Adh-1*, *Adh-2*) complex polymorphism (Brown *et al.*, 1978). No other exceptional genetical variation at Talpiyyot was noted. This material will be tested for the presence of partial male sterility. In general, however, outcrossing is such a rare event that the sample size at each site was too small to demonstrate significant variation in outcrossing from site to site, unless it had been variation of the order found by Rick *et al.* (1977) in *Lycopersicon pimpinellifolium* (0-40 per cent), or Harding *et al.* (1974) in the *Lupinus nanus* group (0-100 per cent).

The Spearman rank correlation between outcrossing and the mean single-locus diversity (Nevo, *et al.*, 1978) over the 26 polymorphic populations was  $r_s = 0.61$ , significant at the 0.01 level. Thus there is a relationship between variation in outcrossing and variation in polymorphism. However it is probably indirect and it accounts for only 37 per cent of the observed variation in diversity.

An estimate of outcrossing and its confidence limits can be obtained for a group of sites by multiplying together the likelihood equations for each site in the group. Table 3 shows these estimates for the seven geographic regions. There was a tendency for the regions to differ, notably regions

TABLE 3

*Estimates of outcrossing rate for seven geographic regions, and various groupings*

Region	$\hat{t}$	$t_{\min}$	$t_{\max}$
A. Golan Heights	0	0	5.0
B. Hule, Upper Jordan Valley	1.3	0.2	4.1
C. Galilee and Carmel Mountains	2.8	0.8	6.9
D. Central Mountains	3.4	1.2	7.4
E. Negev Steppes	0	0	2.1
F. Lower Jordan Valley	0.9	0.1	3.9
G. Mediterranean Coastal Plain	2.0	0.7	4.5
E + F — "Xeric"	0.4	~0	1.8
A + B + C + D + G — "Mesic"	2.1	1.1	3.5
Overall	1.6	0.8	2.7

C and D had estimates of outcrossing which were higher than the confidence limit for region E.

The regions could themselves be divided into two sets, the first set (E, F) being the more xeric areas, and the remainder the more mesic areas (based on the Thornthwaite moisture index, see Amiran *et al.*, 1970). This was done because evidence from other highly self-pollinated grasses (*Avena barbata* in California, Marshall and Allard, 1970, and *Bromus mollis* in Australia, Brown *et al.*, 1974) indicated that more xeric sites tend to have lower outcrossing. The present populations of *H. spontaneum* gave evidence of a clear difference which repeated the pattern of these early studies. The mesic (subhumid) regions gave a mean outcrossing level about four times higher than that of the xeric regions. The mean of each set fell outside in 95 per cent confidence level around the alternative set.

Turning to the levels of heterozygosity found in the sampled natural populations, it is clear that heterozygosity is low as is expected under close inbreeding. The important question, however, is whether this observed level is in accordance with that expected in the absence of selection, but with the mating system taken into account. An exact analysis of this question would require precise knowledge of all maternal genotypes and their gametic associations of loci, as well as the recombination fraction between all polymorphic markers. Unfortunately these data are presently unavailable. The complexity of a full multilocus analysis can be gauged from the exact treatment in the case of two loci given by Weir and Cockerham (1973) in which a complete formulation of the genotypic transitions and equilibrium values requires eight descent measures.

In the multilocus case, an approximate treatment is derived from the following argument. Let  $[1 - Q_x]$  denote the expected proportion of multilocus homozygous mature plants at generation  $x$ . Then this proportion at generation  $x + 1$ , assuming no selection, is

$$[1 - Q_{x+1}] \approx [1 - Q_x][(1 - t) + t(1 - H_e)] + Q_x[(1 - t)p + t(1 - H_e)]$$

where  $H_e$  is the average probability that two uniting gametes from two different plants differ genetically at least at one locus, and  $p$  is the average probability that a seed derived from self-fertilisation of a heterozygous maternal plant will be homozygous. This recursive formula assumes that these probabilities are uniform within the two parental classes, homozygotes and heterozygotes which is not strictly true. The quantity  $p$  would vary across heterozygotes according to the number and kind of heterozygous marker loci because of linkage. For example, for single locus heterozygous maternal plants  $p = 1/2$ , whereas for plants heterozygous for two loci with recombination fraction  $r$  ( $0 < r < 0.5$ ),  $p = [1 - 2r + 2r^2]/2$ .

At equilibrium,  $Q_{x+1} = Q_x = Q$  and the solution is

$$Q \approx tH_e/[t + p(1 - t)]$$

This formula is strictly correct for one locus, two alleles.

An approximate estimate of  $p$  can be obtained from the observed distribution of the number of families segregating for a specified number of loci. This is shown in table 4. Because few heterozygotes were found at any one site, we used only the total distribution to compute the estimate. If the markers recombine independently, then the probability that a maternal



TABLE 4

*Distribution of segregating families for the number of loci at which segregation occurred*

Site	Number of families heterozygous for specified number of loci				
	1	2	3	4	5
Tel Hay	—	1	—	—	—
Zefat	—	1	2	—	—
Ma'alot	—	—	—	—	1
Bar Giyyora	1	1	1	—	—
Talpityyot	2	—	2	1	1
Eizariya	1	—	—	—	—
Tel Shoqet	1	—	—	—	—
Sede Boqer	1	—	—	1	—
Bet Shean	—	2	—	—	1
Mehola	1	3	—	1	—
Caesarea	—	1	—	—	—
Ashqelon	—	1	1	—	—
Total	7	10	6	3	3
Adjusted for tight linkage	11	13	2	3	—

plant heterozygous for  $i$  loci will remain undetected by sampling a single seed is  $p_i = (0.5)^i$ . For some loci (*Est-A*, *B*, and *C*; and *Adh-I*, -2) preliminary evidence indicates very close linkage. If the site data are reclassified to reduce the instances of multiple esterases or ADH to instances of single locus segregation, the distribution of adjusted totals is shown on the last line of table 4. For these values the expected total maternal heterozygosity  $\left( \sum_{i=1}^5 n_i/[1-(0.5)^i] \right)$ , where  $n_i$  families are observed to segregate at  $i$  loci) is 44.8. This figure agrees well with the heterozygosity observed in the progeny ( $O+h = 43$ , table 1) and suggests that from this viewpoint the populations are in approximate equilibrium, and further that about 15 or one-third of heterozygous maternal plants were undetected. The average value of  $p [= \sum n_i p_i / \sum n_i]$  was 0.3168.

Using this value of  $p$ , the values of  $Q$  multiplied by  $N$ , for each site were computed and are shown in the last column of table 1. Non-zero values only occur when  $O, \hat{t} > 0$ . These values are the expected number of maternal heterozygous plants included in the sample of given size under the estimated level of outcrossing, at equilibrium and assuming no selection. The total number was 42, which corresponds very well with the observed frequency in the progeny of 43 heterozygotes. We concluded that overall there is no evidence of excess heterozygosity over expected levels in these populations. However, there was a trend towards excess heterozygosity in the more xeric regions ( $E+F$ ), because 11 were found whereas the expectation was 2.9.

The data in table 4 demonstrates another striking feature about heterozygosity in *H. spontaneum* populations. This feature is the strong tendency toward multilocus heterozygosity. Such a phenomenon would be expected when linkage disequilibria are present. It is remarkable that when the observed frequency of heterozygosity is so low (3.65 per cent of mature plants), at least three families were found in which five loci were segregating

simultaneously. Allelic variation at these loci is obviously very far from being distributed over gametes at random.

#### 4. DISCUSSION

The overall level of outcrossing found in these 28 Israel populations of *Hordeum spontaneum* is estimated as 1.6 per cent, a very low rate. This estimate is probably subject to some downward bias because of genetic differentiation within local populations (Brown *et al.*, 1975). When such differentiation occurs, the pool of outcrossing pollen is no longer random across plants and outcrossing within genetically identical subpopulations is more common than expected. However, correction for this bias could only marginally increase the estimate which would still be regarded as far too low to permit massive introgression.

Populations growing in a wide range of environments were sampled for this study, with collections taken over 3 years. The populations proved to be highly distinct genetically. Yet there was only a weak relationship between outcrossing rate and level of diversity. There was evidence of a relation between environment and the level of outcrossing. When the data from the populations growing in more xeric regions were grouped, they were found to have a significantly lower outcrossing rate (0.4 per cent) than those in the more mesic regions (2.1 per cent). Whilst relatively this is about a four-fold difference, it is still variation within very low limits.

Therefore we conclude that recent hybridisation with *H. vulgare* can be ruled out as the major source of the allozyme diversity in *H. spontaneum* populations. This conclusion parallels that of Rick *et al.* (1977), derived from their extensive study of allozyme variation and outcrossing rates in *Lycopersicon pimpinellifolium*. Rick's previous studies of morphological polymorphisms in a limited sample of this species had suggested that *L. pimpinellifolium* had acquired these polymorphisms from recent introgression with domesticated strains. However, their later much more extensive data on allozymes led these authors to conclude that the morphological polymorphisms belonged to the wild tomato species, and that introgression was not their source. This question is of major importance to the evaluation of wild relatives as genetic resources for crop plants. If the genetic variation which these wild species currently contain is predominantly obtained from massive introgressions with cultivated forms, then the wild species might offer nothing new to the plant breeder which he cannot already obtain from primitive or modern cultivars. This is apparently the case in the weedy rye populations in California (Jain, 1977). Indeed such wild relatives may represent even poorer sources of the same variants because their genetic background is less domesticated. If on the other hand, the primitive cultivars represent samples (and probably very restricted samples) from their wild relatives, then the importance of wild relatives as genetic resources for breeding is confirmed. These are the two extremes in what is probably a continuum of situations as postulated by the co-evolutionary model of crops and their weeds by Harlan (1965).

It is interesting to compare the levels of outcrossing in wild barley, with those reported by Kahler *et al.* (1975) for various stages in the evolution of an experimental population of cultivated barley (composite cross V). Such comparisons must be tentative as their estimates were done under Cali-



fornian conditions using only the *Est-A*, *B* and *C* loci. Nevertheless they found that outcrossing in an early, intermediate and late generation, when grown in the same year, increased from 0.57 per cent, to 0.88 per cent to 1.24 per cent. This last estimate approaches that reported here for wild barley of 1.6 per cent. Presumably the lines entering the composite cross had previously been selected to varying degrees for reduced outcrossing and fidelity of breeding to type. Under the conditions prevailing in the composite cross bulk population, there was relaxed selection for pure-line fidelity and this population evolved toward a level of outcrossing typical of wild populations.

Because of the low level of outcrossing in *H. spontaneum*, heterozygosity is very rare. There was little indication that the observed level was higher than expected when correction was made for inbreeding. This result differs from those found for cultivated barley (Allard *et al.*, 1972), *L. pimpinellifolium* (Rick *et al.*, 1977) and *Avena barbata* in California (Marshall and Allard, 1970; Hamrick and Allard, 1972), but probably not in Israel (Kahler *et al.*, in preparation). Excess heterozygosity was not found in the tetraploid *Hordeum jubatum* (Babbel and Wain, 1977) in which allopolyploidy could substitute for heterozygosity in providing allozyme diversity. A study of *Bromus mollis* (Brown *et al.*, 1974) another tetraploid inbreeder, showed an intermediate result in that both heterozygous excess in xeric sites, and homeologous divergence occurred. Despite the many studies which indicate that heterozygous excess is a common feature of predominantly inbreeding plant populations, the present results suggest that it is not a universal feature.

The results do show another striking feature of heterozygosity in inbreeders, namely the tendency towards multilocus heterozygosity. Such correlated heterozygosity is a direct genetic consequence of occasional outcrossing in populations which are composed of several homozygous lines which are differentiated from each other at many loci. Bennett and Binet (1956) first noted the tendency toward the concentration of heterozygosity at neutral loci in fewer members than expected, in a population with a mating system of mixed selfing and random mating, even when the loci are unlinked. This concentration occurs without the development of linkage disequilibrium. However, the very high degree of concentration found in these populations, together with previous evidence of extensive disequilibrium (Brown *et al.*, 1977) suggest that either selection in favour of specific gametic arrays, or random drift and partial isolation, has magnified the Bennett and Binet effect greatly. It is likely that this phenomenon of concentrated heterozygosity gives to outcrosses an evolutionary significance far greater than the small absolute fraction of heterozygotes would suggest.

*Acknowledgments.*—We wish to thank Drs H. Doll, D. R. Marshall, R. N. Oram, and R. A. Richards for comments on the manuscript. This research was supported by a grant from the United States-Israel Binational Science Foundation.

## 5. APPENDIX

### *Maximum likelihood estimation of outcrossing and its confidence limits in predominantly inbred populations*

Allard, Kahler and Weir (1972) give a method to estimate the extent of outcrossing in predominantly self-pollinated plant populations using

multiallelic codominant variation at a single marker locus. The method described here is a straightforward extension and approximation of their method for multilocus data with the following advantages. First, it derives a single estimate based on all the segregating loci simultaneously. In this way it avoids the difficulties of the variable levels of correlation between estimates on each locus when they are obtained separately. The correlation reflects the linkage disequilibria which are commonly found in inbreeding populations. The sampling error belonging to an average of such estimates is especially difficult to formulate. In this regard it is important to distinguish between the increased probability of detecting an outcrossing event genetically obtained by scoring more loci, compared with the increased precision for estimating rates of outcrossing which arises from scoring more zygotes. Second, the method provides estimates of the confidence limits of the estimates, rather than their standard errors. This is useful because under the basically binomial structure of the model, estimates of zero outcrossing would yield zero standard errors. Furthermore the sampling distribution of very low estimates is highly skewed unless sample size is very large, and this skewness obviates the use of normal distribution theory for testing and computing limits.

(i) *Simplest case*

Consider a sample of  $N$  spikes from each of which a single seed is assayed for its multilocus (allozyme) genotype. Suppose first that no heterozygotes are encountered and that the data are complete. The  $N$  spikes are then classified into  $k\{k \leq N\}$  different multilocus homozygous genotypes distinguishable at one or more loci. Obviously, when  $k = 1$ , that is complete monomorphism, no estimate of outcrossing is possible. In polymorphic cases, suppose the frequencies of the multilocus homozygous genotypes in the population are  $G_1, G_2, \dots, G_k, G_{k+1}, G_{k+2}, \dots$ , a potentially infinite vector ( $\sum G_i = 1$ ). We define the diversity parameter  $H_e = [1 - \sum G_i^2]$  which is analogous to one minus the Simpson (1949) index of diversity. Suppose that all ovules have a uniform probability of outcrossing  $t$ , and of self-fertilisation of  $(1-t)$ , irrespective of genotype. Then the likelihood of the present sample is

$$L = \prod_i (1 - \bar{G}_i t)^{g_i N}, \quad \text{where } \bar{G}_i = 1 - G_i$$

and the  $g_i$  are the frequencies of the  $k$  genotypes in the sample. If  $t$  is small such that  $t^2, t^3$  can be neglected, and the  $G_i$  are estimated accurately by their sample values, then using the Maclaurin expansion

$$\begin{aligned} L &\simeq \prod_i (1 - G_i \bar{G}_i t)^N \simeq [1 - t \sum G_i \bar{G}_i]^N \\ &= (1 - H_e t)^N \end{aligned}$$

Our sample estimate of  $H_e$  is

$$\hat{H}_e = \left[ 1 - \sum_i^k g_i^2 \right] N / (N - 1)$$

The approximate sampling variance of this estimate is given by Simpson as  $4[\sum g_i^3 - (\sum g_i^2)^2]/N$ . The estimate of  $t$  is obviously zero; more interest

surrounds the value of the upper confidence limit to  $t$ ; that is how large could the true population value of  $t$  be, without an outcross being detected genetically at some specified level of probability. Strictly, the answer to this question involves the sampling variation of  $H_e$  and  $t$ , and their covariation. As an approximation to the minimum confidence limit, we here assume that the value of  $H_e$  is known exactly. The upper confidence limit at the 0.05 level is obtained as a solution to the equation

$$0.05 = (1 - H_e t_{\max})^N$$

(ii) *Incomplete data; no heterozygosis*

Suppose that of the total  $N$  spikes assayed,  $N_0$  are completely classified at all polymorphic loci,  $N_1$  lack data at one particular critical locus,  $N_2$  lack data on a second critical locus, etc., in a hierarchical arrangement ( $\sum N_i = N$ ). We then assume that each layer in the hierarchy constitutes an independent sample from the population, and form the likelihood equation

$$L = (1 - H_e^{(0)}t)^{N_0} (1 - H_e^{(1)}t)^{N_1} (1 - H_e^{(2)}t)^{N_2} \dots$$

Again the maximum likelihood estimate of outcrossing ( $t$ ) is zero, and the upper confidence limit is derived numerically by finding that value of  $t$  for which  $L = 0.05$ . Before this can be done however, the values of  $H_e^{(i)}$  and  $N_i$  must be computed. To keep the complexity of the computations within reasonable limits we used the following approximate procedure. The quantity  $H_e^{(i)}$  represents the probability that the ovule from a randomly chosen maternal parent will, when outcrossed to a random pollen grain from the population, be heterozygous at any one of the (P.L. -  $i$ ) loci and thus genetically detectable as an outcross. The data of all  $N$  zygotes we used to compute this quantity for all relevant values of  $i$  by simply asking which of the  $N(N-1)/2$  possible matings would yield at least one heterozygous marker.  $H_e^{(i)}$  is then the sum of probabilities of all such matings. Thus the lack of datum at a locus is treated conceptually as a recessive homozygous null for that locus. In many cases, the lack of a score for one locus was not critical for the diagnosis of any outcross from all other known genotypes because the maternal plant in question carried a distinctive allele at another locus. In such cases, the lack of a score at an effectively redundant locus could be ignored and those mothers could be considered as scored at a greater number of loci in the hierarchy.

The following example illustrates the principles of this approximate method. In this example we will ignore the bias term  $N/(N-1)$ . Consider the following data:

Maternal plant genotype	Locus				Observed number
	1	2	3	4	
1	<i>BB</i>	<i>CC</i>	<i>AA</i>	?	10
2	<i>BB</i>	<i>CC</i>	<i>BB</i>	?	10
3	<i>BB</i>	<i>CC</i>	<i>BB</i>	<i>CC</i>	10
4	<i>BB</i>	<i>CC</i>	<i>BB</i>	<i>AA</i>	10
5	<i>DD</i>	<i>BB</i>	?	<i>AA</i>	10

Then  $H_e^{(0)} = 1 - 5 \times (0.2)^2 - 2 \times 0.2 \times 0.4 = 0.64$ , where the third term arises from the inability to detect crosses between genotype 2 and genotypes 3 or 4.

$$H_e^{(1)} = 1 - 2 \times (0.2)^2 - (0.6)^2 = 0.56$$

At first it would seem that  $N_0 = 20$ ,  $N_1 = 20$  and  $N_2 = 10$  (genotype 5). However, the scoring of genotype 5 for loci 2, 3 or 4 is redundant. Also the scoring of genotype 1 for locus 4 is redundant. Therefore  $N_0 = 40$  and  $N_1 = 10$  and we need not consider the 2 locus level. Thus the approximate likelihood equation is

$$L \simeq (1 - 0.64t)^{40} (1 - 0.56t)^{10}$$

A more exact, but generally much more laborious equation can be derived by considering each maternal genotype separately, and the probability that an outcross would be detected in each case.

$$L = (1 - 0.8t)^{10} (1 - 0.4t)^{10} (1 - 0.6t)^{10} (1 - 0.6t)^{10} (1 - 0.8t)^{10}$$

The value of  $t$  for which  $L = 0.05$  in the approximate equation is 0.093 and in the exact case 0.091. Using unbiased estimates of  $H_e^{(0)} = 0.6531$  and  $H_e^{(1)} = 0.5714$  the solution for  $t_{\max}$  ( $= 0.091$ ) is in much closer agreement. Since our data contained up to 30 different maternal genotypes in most populations, and in most instances few data were missing, and hierarchies were readily formed, we used the approximate method. For the two cases in which the data were most incomplete (Caesarea and Ma'alot), the exact likelihood formulae were computed. The approximate  $t_{\max}$  values as shown in table 1 were then substituted into the exact formulae and the values for  $L$  obtained were 0.038 and 0.047 (instead of 0.05), indicating only a slight loss of information from use of the approximate method.

### (iii) *Heterozygous maternal plants present*

Suppose that among the  $N$  genotypes assayed,  $h$  showed a heterozygous marker. On subsequent additional assay *all* these individual seeds proved to belong to segregating families and were not newly arising outcrosses. The estimate of outcrossing is still zero. The only modification to the simplest case above is that the likelihood equation is now

$$L = (1 - H_e t)^{N-h}$$

where  $H_e$  is estimated as

$$\hat{H}_e = \left[ 1 - \sum_i^k g_i^2 \right] N^2 / [N(N-1) - h/2]$$

The multilocus gametes of the heterozygote are included in the estimate of the population  $H_e$ . A problem arises in the gametic typing of seedlings heterozygous for two or more loci. Thus should  $1^A 1^B$ ;  $2^A 2^B$  be regarded as arising from the gametes  $1^A 2^A + 1^B 2^B$  or  $1^A 2^B + 1^B 2^A$ ? Without a large amount of progeny testing this cannot be answered unequivocally. Our rule was to choose if possible that gametic combination which was already known to be most frequent in the homozygous maternal combinations. This would slightly bias downwards our estimates of  $H_e$  and thus slightly increase our maximum confidence limits.

One other problem arises from the undetected heterozygous maternal plants whose probability structure for progeny arrays differs from a homozygous maternal plant. Thus the probability that the single-locus heterozygote of multilocus genotype 12 (*i.e.* genotype  $1A_1B$ ,  $2A_2A$ ,  $3A_3A$ ,  $4B_4B$  . . .) will yield a homozygous seed is  $[1-t(1-g_1-g_2)]/2$  compared with that for the homozygous multilocus genotype 11 of  $[1-t(1-g_1)]$ . Since  $1 > g_1 + g_2$ , both probabilities are never increasing functions of  $t$ . Therefore their behaviour is similar in  $L$  and furthermore since undetected heterozygous maternal plants are rare, we consider the error from this source to be negligible. The progeny of detected heterozygotes at the second round of sampling could not be incorporated in  $L$  because estimates of recombination fractions and parental gametic arrangements were lacking. The loss of statistical information in their omission is negligible.

#### (iv) *Outcrosses detected*

Suppose that of the  $N$  seeds originally assayed,  $o_1$  were heterozygotes and by subsequent testing were found to be newly arising outcrosses on multilocus homozygous maternal genotype 11,  $o_2$  on genotype 22, etc., and that  $O = \sum o_i$ , then the approximate likelihood equation

$$L = \frac{N!}{O!(N-O)!} t^O (1-H_e t)^{N-O} \prod (1-g_i)^{o_i}$$

The estimate(s) of  $H_e$  are obtained as described above. If the data are complete, then the maximum likelihood estimate of  $t$  is given by

$$\hat{t} = O/(NH_e)$$

If the data are incomplete then  $L$  must be extended as described above and the value of  $t$  for which  $L$  is maximised is most easily found numerically. The 0.05 confidence limits for  $t$  are determined as the values of  $t$  for which  $L = 0.05$ .

#### (v) *General case*

The most general case obtains when the genotypic determinations are incomplete, and both heterozygous maternal plants and outcrosses are encountered. The principles we used to handle these complexities are as given in the separate sections above, and solutions were generally obtainable numerically on a programmable calculator.

### 6. REFERENCES

- ALLARD, R. W. 1975. The mating system and microevolution. *Genetics*, 79s, 115-126.  
 ALLARD, R. W., KAHLER, A. L., AND WEIR, B. S. 1972. The effect of selection on esterase allozymes in a barley population. *Genetics*, 72, 489-503.  
 AMIRAN, D. H. K., ELSTER, J., GILEAD, M., ROSENAN, N., KADMON, K., AND PARAN, U., (Eds.). 1970. *Atlas of Israel*. Elsevier, Amsterdam.  
 BABEL, G. R., AND WAIN, R. P. 1977. Genetic structure of *Hordeum jubatum*. I. Outcrossing rates and heterozygosity levels. *Can. J. Genet. Cyt.*, 19, 143-152.  
 BENNETT, J. H., AND BINET, F. E. 1956. Association between Mendelian factors with mixed selfing and random mating. *Heredity*, 10, 51-55.  
 BROWN, A. H. D. 1975. Efficient experimental designs for the estimation of genetic parameters in plant populations. *Biometrics*, 31, 145-160.

- BROWN, A. H. D., MARSHALL, D. R., AND ALBRECHT, L. 1974. The maintenance of alcohol dehydrogenase polymorphism in *Bromus mollis* L. *Aust. J. Biol. Sci.*, 27, 545-59.
- BROWN, A. H. D., MATHESON, A. C., AND ELDRIDGE, K. G. 1975. Estimation of the mating system of *Eucalyptus obliqua* L. Herit using allozyme polymorphisms. *Aust. J. Bot.*, 23, 931-49.
- BROWN, A. H. D., NEVO, E., AND ZOHARY, D. 1977. Association of alleles at esterase loci in wild barley *Hordeum spontaneum*. *Nature*, 268, 430-431.
- BROWN, A. H. D., NEVO, E., ZOHARY, D., AND DAGAN, O. 1978. Genetic variation in natural populations of wild barley (*Hordeum spontaneum*). In preparation.
- FRÖST, S., AND HOLM, G. 1975. Variation of flavonoid patterns in *Hordeum spontaneum* and *H. agriocrithon*. *Hereditas*, 80, 167-172.
- FRÖST, S., HOLM, G., AND ASKER, S. 1975. Flavonoid patterns and the phylogeny of barley. *Hereditas*, 79, 133-142.
- HAMRICK, J. L., AND ALLARD, R. W. 1972. Microgeographical variation in allozyme frequencies in *Avena barbata*. *Proc. Nat. Acad. Sci. U.S.A.*, 69 (8), 2100-2104.
- HARDING, J., MAKINEN, C. B., AND ELLIOT, M. H. 1974. Genetics of *Lupinus*. VII: Outcrossing autofertility, and variability in natural populations of the nanus group. *Taxon*, 23, (5/6), 729-738.
- HARLAN, J. R. 1965. The possible role of weed races in the evolution of cultivated plants *Euphytica*, 14, 173-176.
- HARLAN, J. R., AND ZOHARY, D. 1966. Distribution of wild wheats and barley. *Science*, 153, 1074-1080.
- HVID, S., AND NIELSEN, G. 1977. Esterase isoenzyme variants in barley. *Hereditas*, 87, 155-162.
- JAIN, S. K. 1975. Population structure and the effects of breeding system. In *Crop Genetic Resources for Today and Tomorrow*, IPB2, eds. O. H. Frankel and J. G. Hawkes, 15-36.
- JAIN, S. K. 1977. Genetic diversity of weedy rye populations in California. *Crop. Sci.*, 17, 480-482.
- KAHLER, A. L., AND ALLARD, R. W. 1970. Genetics of isozyme variants in barley. 1. Esterases. *Crop. Sci.*, 10, 444-448.
- KAHLER, A. L., CLEGG, M. T., AND ALLARD, R. W. 1975. Evolutionary changes in the mating system of an experimental population of barley (*Hordeum vulgare* L.). *Proc. Nat. Acad. Sci. U.S.A.*, 72, 943-946.
- MARSHALL, D. R., AND ALLARD, R. W. 1970. Maintenance of isozyme polymorphisms in natural populations of *Avena barbata*. *Genetics*, 66, 393-399.
- NEVO, E., BROWN, A. H. D., AND ZOHARY, D. 1978. Genetic diversity in the wild progenitor of barley in Israel. In preparation.
- RICK, C. M. 1976. Natural variability in wild species of *Lycopersicon* and its bearing on tomato breeding. *Genet. Agr.*, 30, 249-259.
- RICK, C. M., FOBES, J. F., AND HOLLE, M. 1977. Genetic variation in *Lycopersicon pimpinellifolium*: evidence of evolutionary change in mating systems. *Plant Syst. and Evol.*, 127, 139-170.
- SIMPSON, E. H. 1949. Measurement of diversity. *Nature*, 163, 688.
- WEIR, B. S., AND COCKERHAM, C. C. 1973. Mixed self and random mating at two loci. *Genet. Res.*, 21, 247-262.
- ZOHARY, D. 1964. Spontaneous brittle six-row barleys, their nature and origin. Barley Genetics I. *Proc. 1st Intern. Barley Genet. Symp., Wageningen*, 1963, pp. 27-31.
- ZOHARY, D. 1971. The fate of natural "hybrid swarms" between *Hordeum spontaneum* and *H. vulgare*. Barley Genetics II. *Proc. 2nd Intern. Barley Genet. Sump.*, pp. 63-64.