GENIC VARIATION IN MALE HAPLOIDS UNDER DETERMINISTIC SELECTION

P. PAMILO[1] AND R. H. CROZIER

School of Zoology, University of New South Wales, Kensington, N.S.W., Australia 2033

Manuscript received April 15, 1980
Revised copy received December 20, 1980

ABSTRACT

Genic variation in male haploids and male diploids was compared assuming constant fitnesses (derived from computer-generated random numbers) and infinite population size. Several models were studied, differing by the fitness correlation between the sexes (r_s) and genotypes (r_g), and by the intensity of selection as measured by the coefficient of variation (CV) of the fitness distribution. Genic variation was quantified using the proportion of polymorphic loci, P, the gene diversity at polymorphic loci, H_p, and the gene diversity over all loci, H_a. The two genetic systems were compared *via* variation ratios: variation in male haploidy/variation in male diploidy.——P and H_a were markedly lower for male-haploids than for male diploids, the variation ratios declining with increasing r_s, r_g and CV, but the two genetic systems were similar for H_p. Except for male diploids with $r_s = 1$, the two sexes had different equilibrium gene frequencies but the sample sizes required to detect such differences reliably were larger than usually possible in surveys of natural populations.——Data from natural populations fit the above trends qualitatively, but the variation ratios are much lower than those from our analyses, except that for H_p, which is higher when Drosophila is excluded. Also, the frequency distribution of most common alleles from electrophoretic data has a deficiency of intermediate frequencies compared to that from the computer-generated sets of fitnesses, possibly reflecting either the influence of stochastic processes shifting frequencies away from equilibrium or the involvement of alleles under selection-mutation balance.——While electrophoretic data suggest that Drosophila has unusually high levels of genic variation, unusually low levels of genic variation in male haploids compared with male diploids are not strongly indicated. However, if further data confirm male haploids as having low levels of genic variation, likely explanations are that the bulk of electrophoretically detected variation involves fixed-fitness balancing selection, selection-mutation balance involving slightly deleterious recessive alleles, new favorable male haploid alleles moving more rapidly to fixation than under male diploidy and thus carrying linked loci to fixation faster, or some combination of these possible factors.

IN more than 15% of animal species, most notably in the Hymenoptera, normal males are haploid and arise from unfertilized eggs, while females are diploid. In such male haploids, every locus follows the dynamics associated only with X-linked loci in species having both sexes diploid.

[1] Permanent address: Department of Genetics, University of Helsinki, P. Rautatiekatu 13, SF-00100 Helsinki 10, Finland.

One consequence of selection in male haploids is that the frequency of deleterious recessive alleles is expected to be lower than in male diploids due to their exposure to selection in the haploid males (SNELL 1932; CROZIER 1970). A second consequence concerns the expectation of balanced polymorphism, the conditions for which have been examined by numerous authors (BENNETT 1957, 1958; MANDEL 1959; HALDANE and JAYAKAR 1964; LI 1967) for the case of sex linkage and thus, by extension, for male haploidy. A principal finding has been that, when the fitness of a hemizygous male is assumed to equal that of the corresponding female homozygote, overdominance in the females is necessary, but not sufficient, for balanced polymorphism to occur. Under this assumption, the expectation of balanced polymorphism in male haploids is reduced compared to that of the diploid autosomal case (HARTL 1971). Examination of various regular selection models in which polymorphism stems from selection being of opposite direction in the two sexes leads to the same finding (LESTER and SELANDER 1979; PAMILO 1979).

CROZIER (1970, 1975 pp. 74–75) investigated the effects of reducing the correlation in fitnesses between the sexes by using random numbers to determine fitnesses. When he assumed that the two sexes of a male diploid population had identical fitnesses, but that males and females of a male haploid population were uncorrelated with respect to fitness, CROZIER found that the probability of balanced polymorphism is about the same in the two genetic systems. HALLIDAY (1978) obtained the same result.

All of these various studies have been restricted in scope. Thus, gene diversity at polymorphic loci has been studied for only three specific selection models, in which it showed a tendency to be reduced in male haploids compared to male diploids (PAMILO 1979). The effects of selection intensity and differing fitness correlations between the sexes have been previously largely overlooked. In this paper, we therefore investigate the population genetics of male haploids more closely, using computer-generated sets of fitnesses. Below, we examine deterministic selection models, varying the correlations in fitness between males and females and between homozygotes and heterozygotes, and also examine the effects of varying the intensity of selection.

MODELS

All of the models involve one locus with two alleles, constant fitnesses, and assume an infinite population size. The fitness of each genotype was calculated by summing four computer-generated random numbers, each uniformly distributed on (0,1). The resulting fitness distribution is pseudonormal, with an expected mean of 2. This method of generating fitness arrays was used primarily for convenience. The use of four numbers to derive each fitness enabled us to examine models differing in the level of fitness correlation between the sexes, in the intensity of selection, and in the deviation of the heterozygote fitness from that expected under strict additivity, as described below. We do not assume that the resulting fitness distributions reflect natural ones, about which there is considerable uncertainty (*e.g.*, KING 1972), although of course they may. But we do assume that the distributions are the same for male haploid genotypes as for male diploid ones.

We examined three levels of fitness correlation between the sexes (Figure 1a). These levels were $r_s = 0$, 0.5 and 1, where r_s, the *sex correlation*, is the proportion of random numbers used

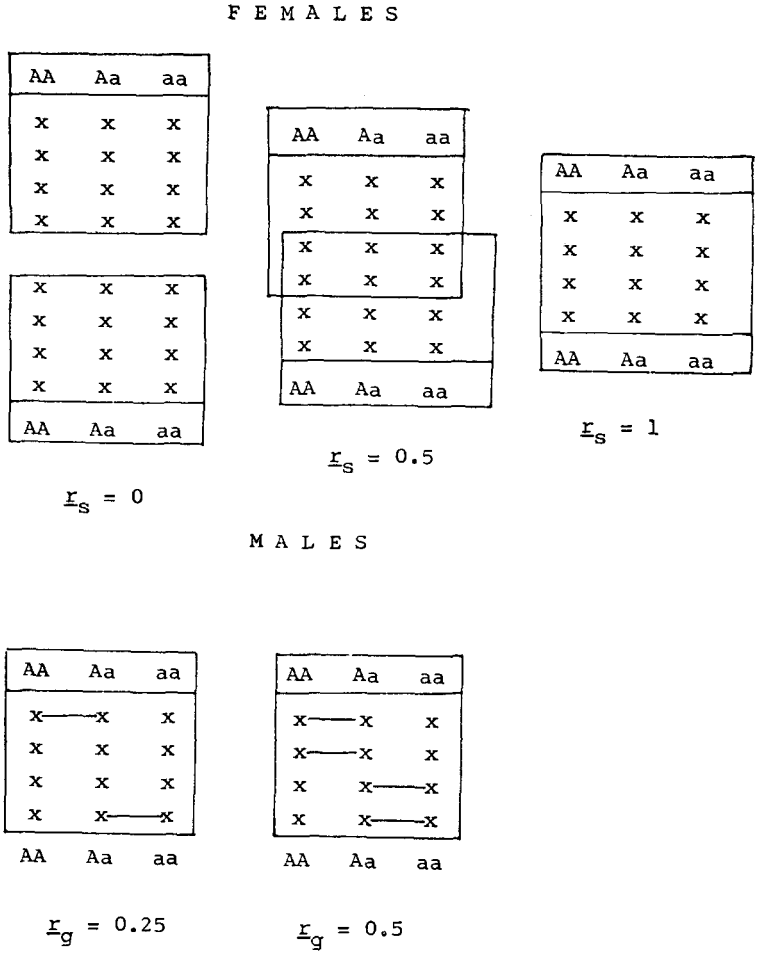

FIGURE 1.—Method of deriving fitnesses as sums of four random numbers, with (a) various sex correlations (r_g held at 0), and (b) various genotype correlations. The use of random numbers in common for the derivation of fitnesses is indicated either by overlapping blocks or by connecting cross bars.

in common when constructing the fitness arrays (APPENDIX 1). Under male haploidy, the heterozygote genotypes for the males were omitted.

We used the coefficient of variation (CV) of the fitness distribution as a measure of selection intensity. The unadjusted distribution has $CV = 0.29$ (APPENDIX 1); adding a constant to each fitness serves to decrease CV, with such a decrease indicating a reduction in the selective differentials between genotypes. The values 0.29, 0.15 and 0.01 were used for CV. The value 0.01 is commonly referred to as a possible value for selective differences in natural populations (*e.g.*, LEWONTIN 1974). It is probably the lowest level for which fitness differences could conceivably be detected, and we used it for our lowest CV value. For the highest value, we used the one of the unadjusted fitness distribution, $CV = 0.29$; the value 0.15 is intermediate between these extremes.

We used two methods (APPENDIX 2 and 3) to investigate the effects of varying the dependence of the heterozygote on the homozygote fitnesses. The common-number method in-

volved using a proportion of the random numbers in common in deriving the homozygote and heterozygote fitnesses (Figure 1b). The resulting genotype-correlation values (correlation between homozygote and heterozygote fitnesses) from this approach were $r_g = 0$, 0.25 and 0.5 (APPENDIX 2). In the averaging method, the heterozygote fitness was derived by assuming that it was likely to be intermediate between the homozygote fitnesses. It was given by modifying the mean of the homozygote fitnesses, using n random numbers (x_i), according to:

$$W_{Aa} = \frac{W_{AA} + W_{aa}}{2} + \sum_{i=1}^{n} (x_i - 0.5) \ .$$

This method was used with correlations $r_g = 0.25$, 0.41, 0.50 and 0.58 (corresponding to $n =$ 14, 4, 2 and 1, respectively; see APPENDIX 3). High values of n sometimes yielded negative fitnesses that were adjusted to zero. Under the common-number method, the heterozygote CV is the same as that of the homozygotes, and it is independent of r_g; under the averaging method it varies with n, according to:

$$CV = \frac{\sqrt{(n+2)}}{4\sqrt{3}} \ .$$

PROCEDURES

OWEN (1953, corrected by MANDEL 1971) and LI (1967) established the conditions for polymorphism and the equilibrium allele frequencies in the two sexes for the male diploid and male haploid cases, respectively. For each fitness array, we used the criteria established by OWEN and by LI to determine: (a) whether one or more stable internal equilibria occurred; (b) the equilibrium allele frequencies of the two sexes; (c) the average gene diversities over both sexes (assuming a 1:1 sex ratio; for male diploids, the sexes were weighted equally; whereas, for male haploids, females were given twice as much weight as males); and (d) the sample size required to be 50% certain of detecting any allele frequency differences between the sexes at the 5% level.

Each run computed 1,000 arrays of fitnesses and yielded the following measures of variation:

P, the proportion of fitness arrays giving stable, nontrivial equilibria,

H_p, the gene diversity at polymorphic loci, calculated as the mean gene diversity (expected heterozygosity) for those sets giving stable internal equilibria,

H_a, the mean gene diversity over all sets; this can also be derived as the product $P \times H_p$.

We repeated each run five times (yielding 5,000 arrays of fitnesses in all for each model), and we used the mean value from each group of five runs.

RESULTS

1. *Sex correlation effects*: The effects of varying the degree of correlation between the sexes, r_s, can be seen by reading across the entries in Table 1. The chief findings when keeping $r_g = 0$ are (Table 1a): (a) Variability, in terms of P, H_p and H_a, is higher in male diploids than in male haploids for each value of r_s. (b) For both genetic systems, P decreases with increase in r_s, with this decrease always greater for male haploids. (c) The two genetic systems have about the same H_p when $r_s = 0$. Increasing r_s raises H_p for male diploids but lowers it for male haploids. (d) Because the average gene diversity, H_a, is strongly influenced by P, it decreases with increasing r_s for both genetic systems.

As OWEN (1953) noted, two stable internal equilibria are possible at male diploid loci. We found fitness arrays yielding two stable equilibria for the $r_s = 0$ case, but very rarely (for about 1% of the sets yielding polymorphisms).

TABLE 1

Probability of polymorphism, P, *expected gene diversities at polymorphic loci,* H_p, *and at all loci,* H_a

	P			H_p			H_a		
	$r_s = 0$	0.5	1	$r_s = 0$	0.5	1	$r_s = 0$	0.5	1
(a) Diploidy									
$CV = 0.01$	0.340±0.018	0.336±0.009	0.336±0.016	0.392±0.009	0.397±0.009	0.393±0.007	0.133±0.008	0.133±0.006	0.132±0.008
0.15	0.387±0.005	0.348±0.014	0.337±0.015	0.379±0.006	0.385±0.012	0.388±0.007	0.147±0.004	0.134±0.008	0.131±0.008
0.29	0.392±0.015	0.351±0.019	0.320±0.015	0.373±0.005	0.384±0.006	0.392±0.008	0.146±0.005	0.135±0.006	0.125±0.006
Male haploidy									
$CV = 0.01$	0.286±0.015	0.259±0.013	0.230±0.019	0.376±0.005	0.365±0.003	0.361±0.008	0.108±0.006	0.094±0.005	0.083±0.007
0.15	0.298±0.007	0.252±0.019	0.217±0.013	0.372±0.006	0.360±0.006	0.358±0.012	0.111±0.004	0.091±0.007	0.078±0.007
0.29	0.286±0.014	0.238±0.010	0.204±0.010	0.372±0.009	0.366±0.014	0.357±0.006	0.106±0.005	0.087±0.004	0.073±0.003
(b) Diploidy									
$r_g = 0$	0.392±0.015	0.351±0.019	0.320±0.015	0.373±0.005	0.384±0.006	0.392±0.008	0.146±0.005	0.135±0.006	0.125±0.006
0.25	0.372±0.010	0.336±0.008	0.308±0.004	0.367±0.004	0.379±0.007	0.381±0.006	0.136±0.004	0.127±0.003	0.117±0.000
0.50	0.296±0.012	0.267±0.017	0.252±0.007	0.354±0.005	0.366±0.006	0.364±0.008	0.103±0.004	0.098±0.007	0.092±0.004
Male haploidy									
$r_g = 0$	0.286±0.014	0.238±0.010	0.204±0.010	0.372±0.009	0.366±0.014	0.357±0.006	0.106±0.005	0.087±0.004	0.073±0.003
0.25	0.271±0.008	0.216±0.009	0.183±0.004	0.367±0.004	0.349±0.010	0.351±0.011	0.100±0.003	0.075±0.004	0.064±0.003
0.50	0.213±0.010	0.174±0.012	0.143±0.009	0.351±0.015	0.353±0.013	0.339±0.008	0.075±0.005	0.061±0.005	0.048±0.005

TABLE 1—Continued

	P			H_p			H_a		
	$r_s = 0$	0.5	1	$r_s = 0$	0.5	1	$r_s = 0$	0.5	1
(c) Diploidy									
$r'_g = 0.25$	0.442 ± 0.016	0.400 ± 0.009	0.384 ± 0.010	0.371 ± 0.005	0.394 ± 0.003	0.416 ± 0.006	0.164 ± 0.007	0.158 ± 0.004	0.160 ± 0.004
0.41	0.354 ± 0.009	0.329 ± 0.011	0.310 ± 0.012	0.357 ± 0.003	0.370 ± 0.003	0.382 ± 0.006	0.127 ± 0.004	0.122 ± 0.004	0.118 ± 0.005
0.50	0.312 ± 0.016	0.280 ± 0.007	0.254 ± 0.018	0.351 ± 0.005	0.353 ± 0.011	0.362 ± 0.009	0.110 ± 0.005	0.099 ± 0.004	0.092 ± 0.009
0.58	0.207 ± 0.008	0.179 ± 0.018	0.162 ± 0.010	0.342 ± 0.010	0.340 ± 0.012	0.351 ± 0.010	0.071 ± 0.004	0.061 ± 0.006	0.057 ± 0.004
Male haploidy									
$r'_g = 0.25$	0.347 ± 0.005	0.310 ± 0.005	0.278 ± 0.011	0.393 ± 0.005	0.380 ± 0.007	0.371 ± 0.008	0.136 ± 0.002	0.118 ± 0.003	0.103 ± 0.005
0.41	0.257 ± 0.011	0.218 ± 0.003	0.187 ± 0.015	0.363 ± 0.011	0.364 ± 0.007	0.354 ± 0.008	0.093 ± 0.006	0.079 ± 0.002	0.066 ± 0.006
0.50	0.216 ± 0.015	0.173 ± 0.012	0.137 ± 0.015	0.358 ± 0.010	0.348 ± 0.009	0.343 ± 0.006	0.077 ± 0.005	0.060 ± 0.003	0.047 ± 0.006
0.58	0.141 ± 0.009	0.105 ± 0.013	0.088 ± 0.005	0.346 ± 0.017	0.343 ± 0.014	0.326 ± 0.011	0.049 ± 0.004	0.036 ± 0.005	0.029 ± 0.003

Mean values and standard deviations are shown, as calculated from computer runs described in the text, with (a) no correlation between genotypes ($r_g = 0$), (b) fixed selection intensity ($CV = 0.29$), r_g adjusted by the common-number method, and (c) a fixed selection intensity in homozygotes only of $CV = 0.29$, r'_g adjusted by the averaging method.

2. *Selection intensity effects*: For male diploids with complete sex correlation but no intrasex fitness correlation, $E(P) = 1/3$, irrespective of CV, because whether or not an equilibrium occurs is determined wholly by whether or not there is overdominance. The numerical results coincide with this expectation (Table 1a). At lower values of r_s, the point estimates of P increase slightly with CV for male diploids and decrease for male haploids, but the changes are slight compared with the confidence limits.

No consistent effect of CV on H_p or H_a is apparent in Table 1a.

3. *Genotype-correlation effects*: The two methods of adjusting the genotype correlation yield similar results (Table 1b, c). P decreases with increasing r_g, as expected from the fact that the strength and likelihood of overdominance declines as the genotype correlation rises. Similarly, H_p falls with increasing r_g. Due to its dependence on P and H_p, H_a also declines with increasing r_g.

4. *Comparing the genetic systems*: Comparisons between the levels of variation found for male haploidy and male diploidy are conveniently expressed in terms of the *variation ratio*: R_v = (variation in male haploidy)/(variation in male diploidy), where variation can be measured in terms of P, H_p, or H_a.

The variation ratios are usually less than one (Table 2) because, as discussed above, male haploids have less genetic variation than male diploids under our models. However, Table 2 also shows that the variation ratios for P, H_p and H_a differ significantly and also have different sensitivities to changes in CV, r_s and r_g.

In terms of P, male haploids have 84.1% of the variation of male diploids when selection is weak ($CV = 0.01$) and fitnesses are not correlated ($r_s = r_g = 0$). However, the variation ratio declines with increasing selection intensity and also with increasing fitness correlation, so that, for $CV = 0.29$ and $r_s = 1$ (keeping $r_g = 0$), male haploids have only 63.8% as many polymorphic loci as male diploids (Table 2). Increasing r_g leads to a further fall in the variation ratio.

The variation ratio for H_p is generally close to one, indicating that male hap-

TABLE 2

Variation ratios calculated from Table 1 (ratios of variation in male haploids to variation in male diploids)

		P			H_p			H_a		
	$r_s = 0$	0.5	1	0	0.5	1	0	0.5	1
$r_g = 0$ $CV = 0.01$	0.841	0.771	0.685	0.959	0.919	0.919	0.812	0.707	0.629
0.15	0.770	0.724	0.644	0.982	0.935	0.923	0.755	0.679	0.595
0.29	0.730	0.678	0.638	0.997	0.953	0.911	0.726	0.644	0.584
$CV = 0.29$ $r_g = 0$	0.730	0.678	0.638	0.997	0.953	0.911	0.726	0.644	0.584
0.25	0.728	0.643	0.594	1.000	0.921	0.921	0.735	0.591	0.547
0.50	0.720	0.652	0.567	0.992	0.964	0.931	0.728	0.622	0.521
$CV = 0.29$ $r'_g = 0.25$	0.784	0.776	0.724	1.059	0.965	0.892	0.829	0.749	0.644
0.41	0.726	0.663	0.603	1.017	0.984	0.927	0.732	0.648	0.559
0.50	0.692	0.618	0.539	1.020	0.986	0.948	0.700	0.604	0.511
0.58	0.681	0.587	0.543	1.012	1.009	0.929	0.690	0.596	0.501

r_g denotes adjustment of genotype correlation by the common-number and r'_g by the averaging methods.

loids and male diploids have about the same gene diversity at polymorphic loci. The ratio falls with increasing r_s, reflecting the fact that H_p increases with r_s under male diploidy, but decreases with r_s under male haploidy (Table 1). The effects of selection intensity and genotype correlation are less clear, although R_v for H_p seems to increase with CV for low values of r_s.

Changes in H_a are dominated by changes in P (Table 1), and this is also reflected in its variation ratio (Table 2).

5. *Differences between the sexes*: The equilibrium allele frequencies are always different in males and females under selection in male haploids; this also holds for male diploids when there are different selective conditions in the two sexes. An observed difference in allele frequencies might therefore lead to inferences about the mode of selection operating, although other factors, such as finite population size, can also result in such differences.

In our analyses, females in a male haploid population are consistently more variable (in terms of gene diversity) than the males. This difference results because the females receive genes from two parental gene pools undergoing different selection. Under male diploidy, however, any systematic bias between the sexes with respect to gene diversity is neither expected nor detected.

We measured the allele-frequency difference between the sexes in terms of the sample size needed to detect it. Thus, for a fitness array yielding a stable polymorphism, N is the sample size required to be 50% certain of detecting the allele frequency difference at the 5% significance level (SOKAL and ROHLF 1969 pp. 607–610).

The differences in allele frequencies between the sexes are, as expected, greatest when the likelihood of opposing selection pressures is greatest ($r_s = 0$ and $CV = 0.29$). Decreasing the strength of selection or increasing the sex correlation diminishes the allele-frequency difference and hence leads to larger required sample sizes to detect it. Thus, with $CV = 0.01$ and $r_s = 0$, N was over 8,000 in 99.5% of the polymorphic cases for both genetic systems. With CV held at 0.29, the median N fell in the range of 128 to 256 for both genetic systems when $r_s = 0$, and in the range of 256 to 512 for $r_s = 0.5$. For $r_s = 1$, no differences between the sexes are possible in male diploids, but such differences can occur in male haploids. The median N in this case fell in the range of 512 to 1024.

Increasing the genotype correlation has no clear effect on N.

DISCUSSION

Our results indicate that a lower level of genic variation is to be expected in male haploid than in male diploid populations, to the extent that such variation stems from fixed-fitness, single-niche, balancing selection. This finding agrees with those earlier studies that suggested that there should be fewer polymorphisms in male haploids (HARTL 1971; PAMILO 1979; LESTER and SELANDER 1979). Our present results also agree with those of CROZIER (1970, 1975) and HALLIDAY (1978), who compared the cases of $r_s = 0$ in male haploids and $r_s = 1$ in male diploids and found that the difference in P in favor of the latter model is not very great. CROZIER's and HALLIDAY's results thus spring from the cases

compared and *not* from the use of random numbers (as suggested by LESTER and SELANDER 1979). However, it is more appropriate to compare cases with the same degree of sex correlation in the two genetic systems; that we have done here.

Although our results may indicate qualitatively the trends expected in natural populations for selected loci with constant fitnesses, the actual numerical values that we derived can be used only with great caution in quantitative prediction of the levels in real populations. For one thing, the actual distribution of fitnesses for new mutations is unknown and may differ significantly from the artificial ones we used. Furthermore, robust estimates of mutation rate and of the length of time the population had been at the present size would be necessary before precise predictions or analyses could be carried out.

Extrapolation from the biallelic case examined here to the multiallelic one would be even more difficult. Thus, overdominance is both a sufficient and a necessary condition for polymorphism at a biallelic male diploid locus, but in a multi-allelic system neither pairwise heterosis nor even total heterosis is either sufficient or necessary for the establishment of an equilibrium maintaining all alleles (LEWONTIN, GINZBURG and TULJAPURKAR 1978), and the volume of the fitness space leading to stable equilibria is very limited. On these grounds, LEWONTIN, GINZBURG and TULJAPURKAR suggested that multiple-niche selection is a more plausible selective mechanism than heterosis for the selective maintenance of multi-allelic equilibria.

With the above qualifications in mind, it is now appropriate to examine the trends from our analyses in greater detail. The most significant result is that the greatest difference between male haploidy and male diploidy is in the probability of polymorphism. The differences between the two systems in gene diversity at polymorphic loci are not great, and in some cases H_p for male haploids actually exceeds that for male diploids.

The trends from Tables 1 and 2 agree with those observable in electrophoretic data from 47 Hymenopteran and 56 bisexual male diploid species to the extent that, on average, Hymenoptera have less genetic variation than the others, and this difference is mainly due to a reduced proportion of polymorphic loci on the part of the male haploids (Table 3).

The variation ratios for P, and hence also H_a, are much lower for the electrophoretic data than for our simulations. This result also holds true for the variation ratio of H_p when comparing male haploids with Drosophila or the total male diploid data set, but H_p for male haploids is greater than that for male diploids when Drosophila is excluded. As NEVO (1978) notes, Drosophila species have the highest H_a values known for bisexual animals; this seems to result from high H_p values. Whether or not this uniqueness of Drosophila will be maintained when more male-diploid insect groups are surveyed by a broader array of investigators is of course open to question.

Other important quantities examined include allele frequency differences between sexes and the distribution of allele frequencies over loci. According to our analyses, the sample sizes required to detect allele frequency differences between

TABLE 3

Observed genetic variation in natural insect populations, together with variation ratios comparing hymenopterons with other insects

	Number of species	Mean number of loci	$\overline{P}$	$\overline{H}_p$	$\overline{H}_a$
Hymenoptera	47	15	0.163±0.109	0.230±0.115	0.036±0.026
Male diploid insects (excluding Drosophila)	13	19.3	0.437±0.178	0.175±0.156	0.092±0.092
Drosophila	43	21	0.431±0.130	0.331±0.100	0.140±0.053
Variation ratios when compared to:					
Male diploids excl. Drosophila			0.373	1.314	0.391
Drosophila			0.378	0.695	0.257
All male diploid insects			0.377	0.780	0.279

The data for male diploid insects come from NEVO (1978), while the hymenopteran information was drawn from SNYDER (1974), METCALF, MARLIN and WHITT (1975), HALLIDAY (1978), PAMILO *et al.* (1978), PAMILO, VARVIO-AHO and PEKKARINEN (1978), SHAUMAR, ROJAS-ROUSSE and PASTEUR (1978), WARD (1978) and LESTER and SELANDER (1979); only populations for which ten or more loci were surveyed are included.

sexes are generally too large to achieve in practice in most studies of natural populations.

A marked departure between our numerical results and those from electrophoretic studies occurs in the case of the distribution of allele frequencies (Figure 2). In our results, this distribution is nearly uniform, only slightly affected by r_s, r_g and CV; whereas, the electrophoretic data yield one with a comparative excess of extreme allele frequencies. This departure could stem from a number of cases. Thus, if most polymorphisms in natural populations are maintained by selection, the observed difference between the electrophoretic results and our simulation results could indicate that our method of obtaining fitnesses differs from that used in nature. While we naturally accept the need for caution in extrapolating results based on computer-generated fitness arrays, it must also be born in mind that polymorphisms in natural populations can result from many causes other than constant-fitness balancing selection. Therefore, we shall now consider other possible explanations of the observed differences between male haploids and male diploids on the assumption that these differences are, in fact, real.

Other modes of balancing selection that have been suggested to be important in maintaining genic variation in natural populations include multiple-niche selection, density dependency, and frequency dependency. SNYDER (1974) and SYLVESTER (1976) argued, on the basis of the expected homeostatic effects imposed by social life in honey bees and ants, that the resulting relatively fine-grained nature of the environment to such insects explains their relatively depauperate nature. However, this apparent trend has persisted after a wide, although not exhaustive, range of Hymenoptera was surveyed electrophoretically, and we agree with METCALF, MARLIN and WHITT (1975) and LESTER and SELANDER (1979) that multiple-niche selection differences seem an unlikely

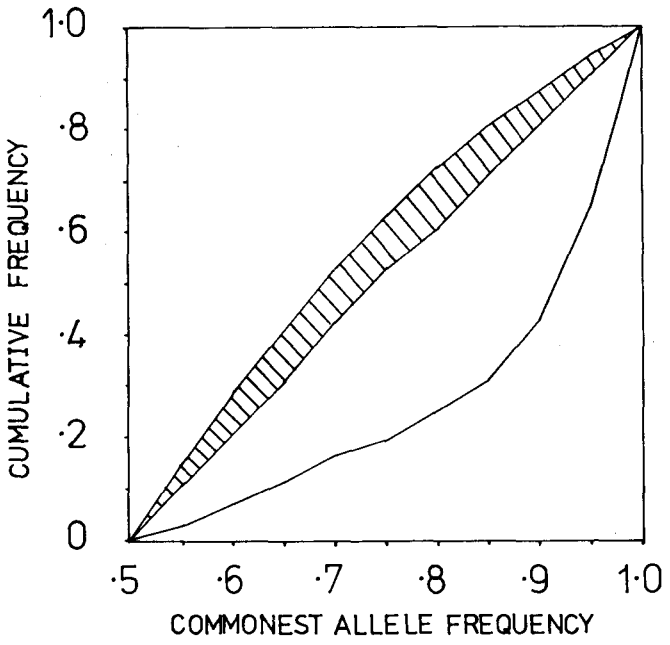

FIGURE 2.—Cumulative frequency distributions for the most common allele in male haploids at polymorphic loci. The distributions obtained from the numerical analyses of this paper (made with various values of r_s, r_g and CV) fall within the hatched area, while that derived from observed allele frequencies of natural Hymenopteran populations (using a criterion of the most common allele frequency being less than 0.99 to define polymorphism) falls below the lower solid curve.

hypothesis, because this explanation would require Hymenoptera to average narrower niches than other insects, which seems unlikely. We also find no *a priori* grounds for supposing that density dependency and frequency dependency operate less effectively in male haploids than in male diploids.

A likely explanation for the observed excess of extreme allele frequencies in natural populations (Figure 2) is that this excess either reflects the influence of stochastic processes shifting frequencies away from equilibrium (see KIMURA 1955) or results from a significant fraction of loci being in mutation-selection balance, involving rare alleles that are slightly deleterious compared with the allele in highest frequency. If the deleterious alleles are semidominant to the common one, then the algebraic treatment reduces to one of gametic selection, and no difference in genic variation level is expected between male haploid and male diploid populations of the same effective size. The situation is different if the deleterious alleles are recessive. Consider the charge-state model number IV of OHTA and KIMURA (1975). In this model, alleles mutate from one to another at a rate $v/2$, and the fitness of all genotypes not involving at least one common allele is $1 - s$. Given a frequency of the optimum common allele as x_0, the fitness of any deleterious allele in a male diploid population is $1 - s(1 - x_0)$ (OHTA and KIMURA 1975), but is easily shown to be $1 - s(1 - 2x_0/3)$ in male haploids.

OHTA and KIMURA give the expected equilibrium allele frequencies for male diploids, and we used modified versions of their equations to determine these frequencies in male haploids, as shown in Table 4. The variation ratios obtained at various values of s/v are also tabulated, and these fall as the strength of selection increases relative to the mutation rate. A similar result follows from applying CROZIER's (1979) equation 21. The variation ratio trends suggest that at highly polymorphic loci (with selection weak relative to mutation), male haploid and male diploid loci will have similar gene diversities, but the proportion of loci detected as being polymorphic will fall more rapidly for male haploid than for male diploid loci (because of both sampling difficulties and an increased tendency of the lower-frequency male haploid alleles to be lost through random genetic drift). The model of slightly recessive deleterious alleles thus leads to results similar to those predicted by our deterministic models, and to those observed in natural populations.

CROZIER (1976, 1979) and MAYO (1976) pointed out that the reduced effective population size, N_e, of male haploids furnishes a possible explanation for their lower genic variability through reduced variation for effectively selectively neutral alleles. But there are complications to a simple invocation of N_e. Thus, the actual relationship between N, the actual number of individuals, and N_e varies differently with sex ratio under the two genetic systems (CROZIER 1976, 1979), and the female-biased sex ratios of many social and parasitic Hymenoptera reduces the strength of any population size effect (relative to male diploid populations of similar actual size) for these groups. Furthermore, examination of relevant formulas for expected heterozygosity, such as equation 7.2.4 of CROW and KIMURA (1970), shows that, even assuming a 50:50 sex ratio, the relative

TABLE 4

Allele frequencies and gene diversities under the charge-state model of OHTA *and* KIMURA *(1975) involving a type allele* A_0 *and slightly deleterious recessive mutations* $A_{\pm i}$

s/v		A_0	$A_{\pm 1}$	$A_{\pm 2}$	$A_{\pm 3}$	Gene diversity	Variation ratio
2	mh	0.500	0.167	0.056	0.019	0.687	0.937
	md	0.445	0.171	0.066	0.025	0.733	
4	mh	0.651	0.137	0.029	0.006	0.537	0.866
	md	0.573	0.156	0.042	0.011	0.620	
8	mh	0.774	0.098	0.012	0.002	0.381	0.762
	md	0.682	0.129	0.024	0.005	0.500	
20	mh	0.885	0.054	0.003	0.000	0.211	0.588
	md	0.790	0.093	0.011	0.001	0.359	
40	mh	0.936	0.031	0.001	0.000	0.122	0.452
	md	0.849	0.070	0.006	0.000	0.270	
80	mh	0.966	0.017	0.000	0.000	0.066	0.330
	md	0.891	0.051	0.003	0.000	0.200	

v is the mutation rate per gamete per generation and s is the selection coefficient against the genotypes not involving the allele A_0. The terms mh and md refer to male haploid and male diploid genetic systems, respectively.

effect of male haploidy depends critically on the quantity $N_e u$. For moderately small populations, male haploidy could reduce the genic variation by one quarter, due to effectively neutral alleles, but for very large populations the effect would probably be negligible. Therefore, we feel that, while a simple role for the effect of male haploidy on N_e cannot be categorically excluded, variations in N_e due to actual population size differences are potentially much more important. The social forms might then be expected to have much less variability, because population size in these is chiefly a matter of the queens and males present, but the data currently available are mainly from social forms that do not permit a valid comparison with nonsocial ones.

LESTER and SELANDER (1979) favor the male haploid effect on N_e as an explanation of the apparently reduced genic variation of male haploids, but emphasize its action through effects on loci with selection coefficients fluctuating randomly and on alleles undergoing hitchhiking. As OHTA (1972) notes, the chief effect of random fluctuations in selection intensity is to render the alleles concerned more nearly effectively neutral, so that this case essentially devolves to one of effective neutrality. The effect of male haploidy on the dynamics of hitchhiking is two-fold: first, there is a direct effect from N_e reduction, and second, there is an effective tightening of linkage because of the lack of recombination in the males. However, as we have argued above, the direct effect of N_e is unclear, and probably negligible, and a consideration of recombination rates also renders an effect on hitchhiking questionable. Thus, Drosophila includes the most heterozygous bisexual animals known and, yet, has no male crossing over! Furthermore, linkage will have a significant effect only on relatively tightly linked loci, and the total map length is likely to represent a useful estimate of the average tightness of linkage between neighboring loci. *Drosophila melanogaster* has a total map length of 289 (ALTMAN and DITTMER 1972) and the genetically best-known Hymenopteran, *Nasonia vitripennis*, one of 324 (SAUL, SAUL and BECKER 1967; ALTMAN and DITTMER 1972). Another parasitoid, phylogenetically distant from Nasonia, is *Bracon hebetor*, for which a total map length cannot yet be calculated, but has one linkage group 380 units long (WHITING 1961)—perhaps the longest for any animal. There is thus no reason at present to suggest that linkage is tighter between adjacent loci in Hymenoptera than in organisms such as Drosophila, although we stress that the data are sparse and it is unfortunate that no linkage maps are available for any of the Hymenoptera that have been surveyed electrophoretically, including honey bees.

Even if there is no increase in hitchhiking under male haploidy through effects on N_e and linkage, an increased effect of hitchhiking is still possible. HARTL (1972) showed that, for alleles with additive fitness effects under weak selection in panmictic populations, substitution occurs one-third more rapidly in male haploids than in male diploids. Assuming that new advantageous alleles arise as frequently in male haploids as in male diploids, the faster movement to fixation in male haploids would be expected to result in a similar more rapid fixation of linked alleles. However, the magnitude of any hitchhiking effect depends on the yet unknown rate of occurrence of such new favorable alleles.

We thank J. JAMES, A. STARK, J. SVED, S. VARVIO-AHO and two anonymous reviewers for comments on earlier drafts of this manuscript, and G. B. SAUL and W. ROTHENBUHLER for helpful correspondence. P. P.'s stay in Australia was made possible by a grant under the Australian/European Scheme from the Australian Education Department, and R. H. C.'s research is supported by grants from the Australian Research Grants Committee and the Ian Potter Foundation.

LITERATURE CITED

ALTMAN, P. L. and D. S. DITTMER, 1972 *Biology Data Book, Volume I*, 2nd ed. Fed. Amer. Soc. Exp. Biol., Bethesda, Maryland.

BENNETT, J. H., 1957 Selectively balanced polymorphism at a sex-linked locus. Nature **180:** 1363–1364. ——, 1958 The existence and stability of selectively balanced polymorphism at a sex-linked locus. Aust. J. Biol. Sci. **11:** 598–602.

CROW, J. F. and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper and Row, New York.

CROZIER, R. H., 1970 On the potential for genetic variability in haplo-diploidy. Genetica **41:** 551–556. ——, 1975 *Animal Cytogenetics 3 Insecta 7 Hymenoptera*. Gebrueder Borntraeger, Berlin. ——, 1976 Counter-intuitive property of effective population size. Nature **262:** 384. ——, 1979 Genetics of sociality. pp. 223–286. In: *Social Insects, Volume I*. Edited by H. R. HERMANN. Academic Press, New York.

HALDANE, J. B. S. and S. D. JAYAKAR, 1964 Equilibria under natural selection at a sex-linked locus. J. Genetics **59:** 29–36.

HALLIDAY, R. B., 1978 Genetic studies of meat ants (*Iridomyrmex purpureus*). Unpublished Ph.D. dissertation, University of Adelaide, Australia.

HARTL, D. L., 1971 Some aspects of natural selection in arrhenotokous populations. Amer. Zoologist **11:** 309–325. ——, 1972 A fundamental theorem of natural selection for sex linkage or arrhenotoky. Am. Naturalist **106:** 516–524.

KIMURA, M., 1955 Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harbor Symp. Quant. Biol. **20:** 33–53.

KING, J. L., 1972 The role of mutation in evolution. pp. 69–100. In: *Darwinian, Neo-Darwinian, and Non-Darwinian Evolution*. Edited by L. M. LE CAM, J. NEYMAN and E. L. SCOTT. Proc. 6th Berkeley Symp. Math. Stat. Prob., Vol V. Univ. California Press. Berkeley and Los Angeles.

LESTER, L. J. and R. K. SELANDER, 1979 Population genetics of haplodiploid insects. Genetics **92:** 1329–1343.

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

LEWONTIN, R. C., L. R. GINZBURG and S. D. TULJAPURKAR, 1978 Heterosis as an explanation for large amounts of genic polymorphism. Genetics **88:** 149–170.

LI, C. C., 1967 The maximization of average fitness by natural selection for a sex-linked locus. Proc. Natl. Acad. Sci. U.S. **57:** 1260–1261.

MANDEL, S. P. H., 1959 Stable equilibrium at a sex-linked locus. Nature **183:** 1347–1348. ——, 1971 OWEN's model of a genetical system with differential viability between the sexes. Heredity **26:** 49–63.

MAYO, O., 1976 Neutral alleles at X-linked loci. A cautionary note. Hum. Heredity **26:** 49–63.

METCALF, R. A., J. C. MARLIN and G. S. WHITT, 1975 Low levels of genetic heterozygosity in Hymenoptera. Nature **257:** 792–794.

NEVO, E., 1978 Genetic variation in natural populations: Patterns and theory. Theoret. Pop. Biol. **13:** 121–177.

OHTA, T., 1972 Fixation probability of a mutant influenced by random fluctuation of selection intensity. Genet. Research **19**: 33–38.

OHTA, T. and M. KIMURA, 1975 Theoretical analysis of electrophoretically detectable polymorphisms: models of very slightly deleterious mutations. Am. Naturalist **109**: 137–145.

OWEN, A. R. G., 1953 A genetical system admitting of two distinct stable equilibria under natural selection. Heredity **7**: 97–102.

PAMILO, P., 1979 Genic variation at sex-linked loci: Quantification of regular selection models. Hereditas **91**: 129–133.

PAMILO, P., R. ROSENGREN, K. VEPSÄLÄINEN, S. VARVIO-AHO and B. PISARSKI, 1978 Population genetics of *Formica* ants I. Patterns of enzyme gene variation. Hereditas **89**: 233–248.

PAMILO, P., S. VARVIO-AHO and A. PEKKARINEN, 1978 Low enzyme gene variability in Hymenoptera as a consequence of haplodiploidy. Hereditas **88**: 93–99.

SAUL, G. B., S. W. SAUL and S. BECKER, 1967 Linkage in Mormoniella. Genetics **57**: 369–384.

SHAUMAR, N., D. ROJAS-ROUSSE and N. PASTEUR, 1978 Allozyme polymorphism in the parasitic hymenoptera *Diadromus pulchellus* WSM. (Ichneumonidae). Genet. Research **32**: 47–54.

SNELL, G. D., 1932 The role of male parthenogenesis in the evolution of the social Hymenoptera. Am. Naturalist **66**: 381–384.

SNYDER, T. P., 1974 Lack of allozymic variability in three bee species. Evolution **28**: 687–689.

SOKAL, R. R. and F. J. ROHLF, 1969 *Biometry*. W. H. Freeman and Company, San Francisco.

SYLVESTER, H. A., 1976 Allozyme variation in honeybees (*Apis mellifera* L.). Unpublished Ph.D. dissertation, University of California, Davis.

WARD, P. S., 1978 Genetic variation, colony structure and social behaviour in the *Rhytidoponera impressa* group, a species complex of ponerine ants. Unpublished Ph.D. dissertation, University of Sydney, Sydney, Australia.

WHITING, A. R., 1961 Genetics of *Habrobracon*. Adv. Genet. **10**: 295–348.

Corresponding editor: W. W. ANDERSON

APPENDIX

Let x_i, y_i and z_i be random numbers having uniform distributions between 0 and 1. Then $E(x_i) = 0.5$ and $E(x_i^2) = 1/3$.

1. *Adjustment of the fitness distribution CV*

Let the fitness of a genotype be

$$X = \sum_{i=1}^{4} x_i \ ,$$

then, $E(X) = 4E(x_i) = 2$

$$D^2(X) = 4[E(x_i^2) - E^2(x_i)] = 1/3$$

and hence,

$$CV = D(X)/E(X) = 0.29.$$

The magnitude of the CV can be adjusted by adding a constant, say A, to X, so that

$$CV = D(X)/[E(X) + A] \ .$$

If the mean fitness is adjusted to 1.0, approximately 95% of all fitness values fall within the ranges of 0.42 to 1.58, 0.70 to 1.30 and 0.98 to 1.02, corresponding to the values 0.29, 0.15 and 0.01 used for CV in this study.

2. *Adjusting the genotype correlation by the common-number method*

Let $y_1, \ldots, y_n$ be the random numbers used in common in deriving homozygote and heterozygote fitnesses and let

$X = \sum_{i=1}^{n} y_i + \sum_{i=1}^{4-n} x_i$ be the fitness of the homozygote

and

$Z = \sum_{i=1}^{n} y_i + \sum_{i=1}^{4-n} z_i$ be the fitness of the heterozygote.

Then,

$E(Z) = E(X) = 2, D^2(Z) = D^2(X) = 1/3$, and $CV(Z) = CV(X) = 0.29$
$\mathrm{COV}(XZ) = n/12$
$r_{XZ} = r_g = n/4$.

3. *Genotype-correlation adjustment using the averaging method*

Let X and Y be the fitnesses of the two homozygotes, given as a sum of four random numbers, and Z is the fitness of the heterozygote, given as explained in the text. Then

$$E(Z) = 2$$
$$D^2(Z) = (n+2)/12$$
$$CV(Z) = \sqrt{(n+2)}/4\sqrt{3}$$
$$\mathrm{COV}(XZ) = \mathrm{COV}(YZ) = 1/6$$

and

$$r_{XZ} = r_{YZ} = r'_g = 1/\sqrt{(n+2)}.$$