VARIATION FOR METRICAL CHARACTERS IN
DROSOPHILA POPULATIONS

I. GENETIC ANALYSIS

B. W. BARNES and M. J. KEARSEY

Department of Genetics, University of Birmingham

Received 29.i.69

1. INTRODUCTION

AN important problem of population genetics is to account for the widespread polymorphism found in natural populations (Harris, 1966; Lewontin and Hubby, 1966). Genetic polymorphism was originally defined by Ford (1940) in terms of discontinuous characters. However, this extensive polymorphism at individual loci manifests itself in variation in many metrical traits particularly those in which the non-additive genetic variance is small in comparison to the additive genetic variance.

The mechanisms responsible for single locus polymorphism may account, in part, for this variation. But for a particular metrical trait there is another level of complexity, in that selection has the opportunity to act on the character as a whole. We may recognise three types of selection acting on the phenotype, *i.e.* directional, disruptive and stabilising (Mather, 1953). Directional selection is not relevant in the present context since it does not lead to the maintenance of extensive genetic variation within a population. Disruptive selection, on the other hand, has been shown by Thoday and his colleagues (for a review see Thoday, 1963) to be a method capable of maintaining genetic variation and, furthermore, of establishing polymorphisms. However, the genetic consequences of stabilising selection and indeed its occurrence, are subject to dispute. Previously, models which give rise to stable equilibria under stabilising selection required that the loci controlling the character exhibit fairly substantial unidirectional dominance (Kojima, 1959; Lewontin, 1964). This is in contrast to the type of genetic variation shown by the characters under discussion. Nevertheless, recent theoretical work (Gale and Kearsey, 1968; Kearsey and Gale, 1968) has indicated that stabilising selection is able to maintain variation for characters which show no dominance if the effects of different loci are unequal. Stabilising selection for this type of character has been demonstrated in experimental populations of *Drosophila melanogaster* (Barnes, 1968).

At the present time little experimental evidence is available concerning the relationship between metrical characters and fitness in wild populations. With this in mind a programme was initiated to provide information on the following points.

1. The extent and nature of the genetic variation for a metrical character in a population recently isolated from the wild.
2. The mechanisms by which this variation is maintained.
3. The relationship between the type of selection involved and the genetic architecture shown by the character.

Drosophila melanogaster has been used for these studies because its suitability for genetic analysis by far outweighs the lack of knowledge about its ecology.

This first paper is designed to provide information relevant to point 1 above, and is concerned with the genetic analysis of sternopleural chaeta number.

2. MATERIALS AND METHODS

The wild population, "Texas", used in this study originated from 30 inseminated females caught at Austin in Texas in October 1965. The progeny from every female were maintained separately as mass cultures, in bottles, for several generations to allow screening for mites. The cage population was initiated in April 1966 from 600 flies—10 flies of each sex from every progeny group.

The population cage used has been described earlier by Barnes (1968). Essentially it consists of a perspex box with a gauze lid. Twenty tubes containing food medium are inserted through holes in the base of the box, and replaced in rotation so that any one tube remains in the cage for 17-18 days. The population size is approximately 2500.

Both adults and eggs are easily sampled from the cage. Adults are "trapped" in vials to which the access is restricted by a fine tube through a foam plug. Egg samples are taken by replacing a food vial, for a short time, with a vial containing a laying stick coated with a blackened starch mixture.

All experiments were carried out at 25° C. and the flies raised on normal yeasted oatmeal medium. The character scored was the number of sternopleural chaetæ per fly.

3. DIRECTIONAL SELECTION

In October 1966, progeny were raised in bottles from eggs sampled from this population. The mean and variance of chaeta number for each sex was as follows:

	Females	Males
Mean	18.40	17.67
Variance	3.13	2.94

From this sample four single pair matings were set up between flies selected for having a chaeta number equal to the mode of the population *i.e.* 18 hairs. Two of the full sib family groups provided the basis for lines selected for a high number of sternopleural chaetæ, ($H_{1.0}$; $H_{2.0}$), the remaining 2 being the basis for low selection ($L_{1.0}$; $L_{2.0}$).

Ten flies of each sex were scored from every line. The lowest female from $L_{1.0}$ was crossed to the lowest male from $L_{2.0}$, while the lowest female from $L_{2.0}$ was crossed to the lowest male from $L_{1.0}$. Similar crosses were made between the flies with high hair counts in $H_{1.0}$ and $H_{2.0}$. Two less extreme selections and a mass mating of unselected individuals were also set up to guard against failures. From among the successful cultures only the most extreme selection from each reciprocal was used to provide parents for the next generation, the selected parents used in a mating being related as double first cousins. This system of selection with double first cousin

matings was continued throughout the experiment but lower selection pressures and random mating were used on occasions to prevent the complete loss of the lines.

The results of the selection experiment are shown in fig. 1, chaeta number at any generation being averaged over sexes and reciprocals.

In the low direction the hair count dropped fairly linearly for about 12 generations and then plateaued at about 12 hairs. At this point the lines became very difficult to maintain and during generations 17 through 31 they

FIG. 1.—Frequency distribution of chaeta number in the Texas population, and the response to directional selection.

were randomly mated with selection only every 2 or 3 generations. It can be seen that relaxing selection resulted in the hair count rising to about 13 hairs. Subsequently these lines have been maintained by selecting as parents only those flies with 12 hairs or less.

Selection for high chaeta number also resulted in a linear response for the first 18 generations followed by an irregular plateau between 44 and 46 hairs. Generations 28, 29, 30 and 35 were produced without selection, and generations 30 and 35 were maintained at 18° C. These factors are responsible for the marked reduction in chaeta number at these generations. Selection has subsequently failed to regain the earlier high counts and the lines have apparently reached a new plateau at approximately 42 hairs.

The realised heritability as estimated from those generations over which the response to selection was approximately linear (*i.e.* generations 1-16 low; 1-18 high) is 0.8141 ± 0.0180 , as compared to that calculated from the regression of offspring on midparent, using the original population, of 0.5931 ± 0.1688 . However, these estimates are not significantly different.

4. GENETIC ANALYSIS

From flies extracted at generation 34 the high and low selection lines together with their reciprocal F_1 , F_2 and backcross progenies were raised and the females scored. In all, 14 generations were produced with 12 replicate cultures per generation, the whole experiment being in a single randomised block.

A comparison of the reciprocal F_2 and backcross generations fails to demonstrate any significant sex-linked differences between the lines. The absence of a sex-linked component was unexpected in view of the marked difference in chaeta number between the lines, but has been confirmed by subsequent experiments. For this reason the data from reciprocal crosses have been pooled and the generation means thus obtained are given in table 1.

TABLE 1

Mean sternopleural chaeta number for generations derived from the high and low selection lines

Generation	Mean	Number Scored	Model			
			m	$[d]$	$[h]$	$[j]$
P_1	37.5833 ± 0.6352	120	1	1	.	.
P_2	13.5333 ± 0.1944	90	1	-1	.	.
F_1	20.4149 ± 0.1552	188	1	.	1	.
F_2	22.0625 ± 0.4175	240	1	.	0.5	.
B_1	27.8638 ± 0.4803	470	1	0.5	0.5	0.5
B_2	17.5418 ± 0.2982	371	1	-0.5	0.5	-0.5

The genetic model fitted to these data involved the parameters m , $[d]$, $[h]$ and $[j]$. Here $[d]$ is the balance of homozygous deviations, $[h]$ is the sum of heterozygous deviations, while $[j]$ is the balance of digenic homozygote $\times$ heterozygote interaction deviations. All these are estimated deviations from m , the mean of all possible homozygous lines that might be produced from P_1 and P_2 . This is the $F \infty$ metric discussed by Van der Veen (1959).

These parameters were estimated by weighted least squares, using as weights the reciprocal of the variances of the generation means, (Cavalli, 1952; Jinks, 1956), and proved to be as follows:

$$\begin{aligned}\hat{m} &= 25.2556 \pm 0.2755 \\ [\hat{d}] &= 11.7741 \pm 0.2944 \\ [\hat{h}] &= -4.9067 \pm 0.3366 \\ [\hat{j}] &= -1.3639 \pm 0.5563\end{aligned}$$

The χ^2 for 2 d.f. testing the adequacy of this model equals 4.28 ($P = 20\% - 10\%$) i.e. the above model is sufficient to explain the data.

Jinks and Morley Jones (1958) have defined $[d]$ as $r_d \Sigma d$ where r_d is a measure of the degree of association of genes of increasing effect in the parents. Since P_1 and P_2 are extreme selection lines it can safely be assumed that $r_d = 1.0$ and hence $[d] = \Sigma d$. Similarly, $[h]$ is defined as Σh . Thus in this case the potence ratio $\frac{[h]}{[d]}$ is a minimal estimate of average dominance, any ambidirectional dominance causing it to be an underestimate. This ratio $\left(\frac{[h]}{[d]}\right) = -0.4167 \pm 0.0226$.

The variance of $\frac{[h]}{[d]}$ is estimated as follows

if

$$T = \frac{[h]}{[d]}$$

then

$$V(T) = \left(\frac{\partial T}{\partial [h]}\right)^2 V([h]) + \left(\frac{\partial T}{\partial [d]}\right)^2 V([d]) + 2 \left(\frac{\partial T}{\partial [h]}\right) \left(\frac{\partial T}{\partial [d]}\right) \text{Cov}([h][d])$$

There is, then, considerable directional dominance for this character in the Texas population, while epistasis is a relatively minor component. Information on average dominance is available also from the variance within families of these generations. In the absence of linkage, estimates of dominance from these second-degree statistics are not reduced by ambidirectional dominance. However, if genes with opposing dominance properties are linked then these estimates, like the potency ratio, will be reduced also.

The average variance within families for every generation is shown in table 2. The variation within P_2 and the reciprocal F_1 's was homogeneous and considerably less than that for P_1 . For this reason two estimates of environmental variation within families (E_1 and E'_1) have been used and these are appropriately weighted into the expectations of the segregating generations.

TABLE 2

Average variance of sternopleural chaeta number within families (For explanation see text)

Generation	Variance	Model				
		D	H	F	E_1	E'_1
P_1 (high)	10.9130	.	.	.	1.0	.
P_2 (low), F_1	4.0992	.	.	.	.	1.0
F_2	42.0514	0.5	0.25	.	0.25	0.75
B_1	44.9106	0.25	0.25	-0.5	0.50	0.50
B_2	7.9052	0.25	0.25	0.5	.	1.0

$$\hat{D} = 62.5742$$

$$\hat{H} = 19.8468$$

$$\hat{F} = -33.5985$$

$$\text{Dominance ratio} = \frac{\sqrt{\hat{H}}}{\sqrt{\hat{D}}} = -0.5632$$

$$\hat{E}_1 = 10.9130$$

$$\hat{E}'_1 = 4.0992$$

$$r_{d,h} = \frac{F}{\sqrt{D.H}} = -0.9534$$

The genetic components $D(= \Sigma d^2)$, $H(= \Sigma h^2)$ and $F(= \Sigma dh)$ are as defined by Mather (1949). The magnitude of the dominance ratio together with the large difference in the backcross variances confirms the conclusion obtained from the first-degree statistics, *i.e.* directional dominance for low chaeta number. Although the dominance ratio is numerically larger than the potency ratio no reliable standard error is available for testing the significance of the difference. However, should this difference prove to be real it

would indicate that at some loci at least there is dominance for genes of increasing effect. Indeed, Wolstenholme and Thoday (1963) have shown that in a polymorphic population of *Drosophila melanogaster* (Thoday and Boam, 1959) two genes on the third chromosome are fully dominant for increasing chaeta number.

Further information on dominance can be obtained from the ratio $\frac{F}{\sqrt{DH}}$. This is an estimate of the correlation between d_i and h_i and in this case has a value very close to -1 (table 2). Thus d and h are related, the negative correlation being due to the fact that decreasing genes are dominant.

If we assume this relationship is linear, then

$$h_i = a + bd_i$$

where

b = regression of h on d

$$= \frac{F}{D}$$

When $d_i = 0$, it is reasonable to assume that $h_i = 0$ also, and thus,

$$\begin{aligned} h_i &= bd_i \\ &= -0.5369d_i \end{aligned}$$

i.e. the dominance ratio is -0.5369 , which agrees well with that shown in table 2.

The above reasoning must be treated with some caution, however, since the number of effective segregating factors in the F_2 and backcross generations is bound to be low because of linkage alone.

Non-allelic interaction of the type detected from the generation means, namely $[j]$, will affect the second-degree statistics, but to a negligible amount. It will not, however, affect the overall conclusion of directional dominance.

Previous investigations on crosses between long-inbred lines of *Drosophila* have revealed little dominance variation for this character (Cooke and

TABLE 3

Mean number of sternopleural chaetae and per cent. survival from larva to adult of the Texas population (T), 6 C/L and the reciprocal crosses between them

Cross	Chaeta number	Survival (%)
T $\times$ T	17.66 $\pm$ 0.2279	87.4
T $\times$ 6 C/L	22.76 $\pm$ 0.2977	89.0
6 C/L $\times$ T	22.54 $\pm$ 0.2683	90.4
6 C/L $\times$ 6 C/L	33.22 $\pm$ 0.2755	54.0
F_1	22.65 $\pm$ 0.2004	—
Midparent	25.48 $\pm$ 0.1787	—

Mather, 1962; Hill, 1964; Killick, personal communication). In view of these findings it was felt necessary to confirm that the observed dominance had not been produced either during the selection programme or as a result of natural selection for flies with lower chaeta numbers in the F_1 and segregating generations.

To this end flies from the original population were crossed to a tester line 6 C/L. This line was chosen because of its relatively extreme hair count and

because there is normally no potency when it is crossed to other inbred lines. Fixed numbers of larvae were collected from these crosses as well as from 6 C/L and the Texas population. The progeny raised were scored both for chaeta number and for viability from larva to adult. The results are shown in table 3.

As the F_1 deviates significantly from the midparent it is evident that the dominant genes are present in the original population. Furthermore, the yield of the F_1 is greater than either parent, and hence the deviation of the F_1 from the midparent in chaeta number is not due to selection.

5. DISCUSSION

Although the phenotypic variance of sternopleural chaeta number in the base population is small, less in fact than that for many inbred lines with a comparable mean hair count, selection has revealed extensive potential genetic variation. It must be emphasised that we have not selected for alleles which occur in the base population at low frequencies. Both high and low selections were started from four individuals, and thus the lowest possible frequency of any allele is 0.125.

It is not possible at this stage to obtain a reliable estimate of the number of loci involved. On the simplifying assumption of equal gene effects at all loci one can obtain a minimal estimate from the parameters $[d]$ and D (Mather, 1949). Thus if k_1 is the number of loci, then $[d] = k_1 d$ and $D = k_1 d^2$ and $k_1 = \frac{[d]^2}{D} \approx 2$. Since only one generation of recombination has occurred in the production of the F_2 and backcross generations, this suggests that no more than two chromosomes are involved. Since it has been shown that there is no detectable difference between X chromosomes extracted from the high and low selection lines, it is reasonable to expect that this estimate is simply a measure of the contributions of chromosomes II and III. A preliminary whole chromosome assay has confirmed this and further suggests that by far the major component is on chromosome III. However, there was evidence from this assay that the high lines at least had regressed by some 8-10 hairs during the construction of the substitution lines, and these lines are being reconstructed. It is hoped to have more information on these in a later paper.

With the particular selection scheme used, the linear response to selection over many generations must mean that several loci are involved. Furthermore with the values of $[d]$ and $[h]$ as estimated it is not possible for the original population to have a mean of 18.40 and a variance of 3.13 if only two or three loci are segregating.

It is possible, again with certain simplifying assumptions to approach the problem of estimating the number of effective factors by using information from the base population rather than from the F_2 and backcross generations. If the base population is randomly mating and in linkage equilibrium then its mean ($\bar{P}$) and additive genetic variance (V_A) can be expressed as follows:

$$\bar{P} = m + \sum_{i=1}^{k_1} [(u_i - v_i)d_i + 2u_i v_i h_i]$$

$$V_A = \sum_{i=1}^{k_1} [2u_i v_i \{d_i + h_i(v_i - u_i)\}^2]$$

where u_i = the frequency of the increasing allele at the i th locus

$$(i = 1, 2, \dots k_2).$$

We need to make the following assumptions also :

1. $h_i = bd_i$ for all loci. (There is some evidence presented above that this may be so)
2. d_i is a constant for all loci (d)
3. u_i is a constant for all loci (u).

Thus,

$$\bar{P} = m + (u-v)k_2d + 2uvbk_2d$$

on substituting $v = (1-u)$

$$\bar{P} = m - 2bk_2du^2 + 2k_2d(1+b)u - k_2d$$

and as $k_2d = [d]$ then

$$\bar{P} = m - 2b[d]u^2 + 2[d](1+b)u - [d]$$

Since $\bar{P}$, $[d]$, m and b are known, the quadratic expression in u can be solved. For the present data $\bar{P}(\varnothing\varnothing) = 18.40$, $[d] = 11.7741$, $m = 25.2556$ and b (from the potence ratio) = -0.4167 and thus u is estimated as 0.2956.

Similarly, it may be shown that

$$V_A = 2uvk_2d^2[1+b^2-4uvb^2-2(u-v)b]$$

The value of u estimated above and b can be substituted into this formula, and with a value of V_A , k_2d^2 can be estimated.

$V_A = h^2V_p$, where $V_p = 3.13$ and h^2 = the narrow heritability. Two values of h^2 have been presented earlier, but only that obtained from the selection data (0.8141) will be used here because it is known with greater precision and because it will give the lowest estimate of k_2 . Thus

$$V_A = 2.5471.$$

and therefore $k_2d^2 = 8.8858$.

Since

$$[d] = k_2d$$

$$k_2 = \frac{[d]^2}{k_2d^2} \approx 16.$$

As a result of multiplying up the stocks for the "generation mean experiment", both the high and low lines have regressed (*cf.* $\bar{P}_1$ and $\bar{P}_2$ with the high and low selections in fig. 1). Thus $[d]$ is an underestimate of the potential difference between the high and low lines, and severely reduces the estimate of k_2 .

Although we have made several assumptions, none of which we can justify, this exercise is by no means useless. Consider what would happen if certain assumptions fail.

Firstly inequality of gene effects or dominance deviations would result in k_2 being underestimated (Mather, 1949).

Secondly, consider what happens if we relax the assumption that $u_i = a$ constant for all i . Although it has not been possible to examine the failure of

this assumption in general, some special cases considered numerically suggest that k may be over-estimated but to a negligible extent.

Thirdly, the larger V_A , the smaller would be the estimate of k . The maximum possible value of V_A from these data is obtained by putting $V_A = V_p$, i.e. assume the heritability (h^2) is unity. This produces an estimate of the number of loci (k_3), where

$$k_3 = h^2 k_2 \approx 13.$$

As a minimum estimate this is still large.

Finally there is no evidence that this population is anything but randomly mating in respect of sternopleural chaeta number (Birley, personal communication).

On these arguments, therefore, if the population really is in linkage equilibrium there are at least 16 loci controlling chaeta number which are polymorphic. However, it may be more reasonable to assume that far fewer loci are segregating, but that these loci are in considerable negative disequilibrium as a result of selection. It is known that the low and in particular the high selection lines regress when artificial selection is relaxed, but further evidence of the nature and intensity of selection will be presented in the second paper in this series.

How do the data from this population compare with those from other studies on sternopleural chaeta number? As was stated earlier, most crosses between long-established inbred lines show little or no significant potency. These lines tend to be relatively uniform in chaeta number, and thus one might well expect the F_1 to lie outside the parental range if they all show directional dominance for low chaetæ, but this is seldom found. Diallel analyses on such lines detect dominance but suggest its magnitude is small (Hill, 1964; Killick, personal communication). This would appear to be in marked contrast to this population. However, Robertson (personal communication) has confirmed that when extreme selection lines from his Kaduna population are crossed, the F_1 again deviates from the midparent by a value almost identical to that described here.

The absence of dominance in the inbred line crosses is not simply a function of their phenotypic similarity as the inbred line 6 C/L (derived from Thoday's *vg* 6 strain) has a relatively high chaeta number (33.2) and shows no significant potency in crosses with other less extreme inbred lines. The high selection line derived from the Texas population, on the other hand, shows a significant negative potency ratio, approximately -0.4 , when crossed to the same inbred lines.

6. SUMMARY

1. The variation of sternopleural chaeta number in a population of *Drosophila melanogaster* recently isolated from the wild has been examined.

2. The population has a small phenotypic variance, but high and low selection lines have revealed extensive genetic variation. The data suggest that there may be as many as 16 loci involved. Furthermore, the frequencies of the alleles at these loci cannot be less than 0.125.

3. The average dominance level at these loci is around 45 per cent., and is in the direction of low chaeta number.

Acknowledgments.—We are indebted to Professor J. L. Jinks and Dr J. S. Gale for helpful discussions, and to Miss P. Wilkinson for valuable technical assistance. This work was supported by the Agricultural Research Council.

7. REFERENCES

- BARNES, B. W. 1968. Stabilising selection in *Drosophila melanogaster*. *Heredity*, 23, 433-442.
- CAVALLI, L. L. 1952. An analysis of linkage in quantitative inheritance. *Quantitative Inheritance*. H.M.S.O., London, 135-144.
- COOKE, P., AND MATHER, K. 1962. Estimating the components of continuous variation. II. Genetical. *Heredity*, 17, 211-236.
- FORD, E. B. 1940. Polymorphism and taxonomy in *The New Systematics* (ed. Julian Huxley). Clarendon Press, Oxford, 493-513.
- GALE, J. S., AND KEARSEY, M. J. 1968. Stable equilibria under stabilising selection in the absence of dominance. *Heredity*, 23, 553-561.
- HARRIS, H. 1966. Enzyme polymorphisms in man. *Proc. Roy. Soc. B.*, 164, 298-310.
- HILL, J. 1964. Effects of correlated gene distributions in the analysis of diallel crosses. *Heredity*, 19, 27-46.
- JINKS, J. L. 1956. The F_2 and backcross generations from a set of diallel crosses. *Heredity*, 10, 1-30.
- JINKS, J. L., AND MORLEY JONES, R. 1958. Estimation of the components of heterosis. *Genetics*, 43, 223-234.
- KEARSEY, M. J., AND GALE, J. S. 1968. Stabilising selection in the absence of dominance: An additional note. *Heredity*, 23, 617-620.
- KOJIMA, K. 1959. Stable equilibria for the optimum model. *Proc. Nat. Acad. Sci.*, 45, 989-993.
- LEWONTIN, R. C. 1964. The interaction of selection and linkage. II. Optimum models. *Genetics*, 50, 757-782.
- LEWONTIN, R. C., AND HUBBY, J. L. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, 54, 595-609.
- MATHER, K. 1949. *Biometrical Genetics*. Methuen, London.
- MATHER, K. 1953. The genetical structure of populations. *Symp. Soc. Expl. Biol.*, 7, 66-95.
- THODAY, J. M. 1963. Effects of selection for genetic diversity. *Proc. Xth Int. Congr. Genetics*, 3, 533-540.
- THODAY, J. M., AND BOAM, T. B. 1959. Effects of disruptive selection. II. Polymorphism and divergence without isolation. *Heredity*, 13, 205-218.
- VAN DER VEEN, J. H. 1959. Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. *Genetica*, XXX, 201-232.
- WOLSTENHOLME, D. R., AND THODAY, J. M. 1963. Effects of disruptive selection. VII. A third chromosome polymorphism. *Heredity*, 18, 413-431.