BIOMETRICAL GENETICS WITH ONE OR TWO LOCI

I. THE CHOICE OF A SPECIFIC GENETIC MODEL

J. STEWART

Department of Genetics, University of Cambridge

Received 28.v.68

1. INTRODUCTION

GENETIC variation is said to be discontinuous if the individuals in a segregating generation fall into a number of discrete classes so that it is possible to classify each individual unambiguously. Such absolute classification can be applied to a quantitative character if, but only if, the effect of segregation at a single locus is very great compared with the effects of random environmental variation and of any residual genetic variation.

However, as the effect of the segregation at a single locus becomes smaller relative to the effect of environmental and other genetic variation, there comes a point when the segregating classes overlap and absolute classification of each individual is no longer possible. The genetic variation is now said to be continuous. The possibility that single locus segregation may still be recognised in this situation has rather generally been neglected, despite the fact that distributions not significantly departing from normality are consistent with single locus situations with heritabilities as high as 50 per cent.

The alternative biometrical methods that have been developed for the study of continuous variables (*e.g.* Fisher, 1918; Mather, 1949; Falconer, 1960) involve the assumption that the genetic variation is due to a relatively large number of genes, each of minute effect and essentially unidentifiable. This historical development has been very profitable, but has had one unfortunate consequence, for it has led to the growth of a general assumption that, if variation in a quantitative character is continuous, the genes responsible for this variation are essentially unidentifiable.

The consequence of this assumption is that, when presented with evidence that variation in a character of interest is continuous, research workers seldom ask the question whether there may in fact only be one or two loci segregating. If in any particular case this were true, then it might be possible to identify these loci critically, ascertain their linkage relations and perhaps investigate their physiological and biochemical effects. To neglect the possibility that continuous variation in a quantitative character might, in any particular case, be due to one or two identifiable genes is therefore to risk losing valuable information.

In fact, there is much evidence suggesting that genetic variation in a continuous quantitative character may often be controlled by relatively few genes (*e.g.* East, 1916; Payne, 1918; Sax, 1923; Wright, 1927; Leonard, Mann and Powers, 1957; Weber, 1959; Thoday, 1961; Wehrhahn and Allard, 1965; Spickett and Thoday, 1966; Law, 1967). Moreover, biometrical techniques for estimating the number of "effective factors" often suggest that the number of loci is not large (*e.g.* Jinks, 1954; Wehrhahn

and Allard, 1965; Mérat, 1967). In view of this it seemed desirable that we should have techniques that allow us to test whether the hypothesis of segregation at one or two loci is adequate to account for the variation in a quantitative character in particular experimental material. If such tests indicated that one or two loci could satisfactorily account for the results, subsequent experimentation could proceed on lines based on this hypothesis. If, on the other hand, the tests indicated that a large number of loci must be postulated, the more usual biometrical approach would have to be used.

FIG. 1.—The back-cross distributions expected on the basis of various genetic models. (a) Parental and F_1 distributions. (b) Single locus segregation. (c) Two equal additive loci, unlinked, in coupling phase. (d) Two unequal additive loci, linked in repulsion phase. (e) Many equal and additive loci in coupling phase.

Key: ϕ_{P_1} } Parental distributions.
 ϕ_{P_2} }
 ϕ_{F_1} F_1 distribution.
 ϕ_{BC_1} Back-cross distribution.
 μ_{P_1} } Parental means.
 μ_{P_2} }
 μ_{F_1} F_1 mean.
 μ_{BC_1} Back-cross mean.

It is the purpose of this paper to present techniques that could be used to this end. The basis of the method proposed involves comparison of the frequency distribution of a back-cross generation with those of the corresponding parental and F_1 generations. If the difference between the parental and F_1 distributions is due to a single locus, the back-cross distribution should not differ significantly from a composite distribution made up of 50 per cent. of the parental and 50 per cent. of the F_1 distribution. If the back-cross distribution does differ significantly from this composite distribution, the

form of the back-cross distribution itself can suggest an alternative genetic model whose fit to the data may be tested (fig. 1).

This method is substantially similar in principle to the partitioning method of genetic analysis devised by Powers (1963), which has been widely applied, particularly in the field of plant breeding, for over 20 years. Although the method has sometimes been successful in detecting segregation at single loci (*i.e.* Leonard, Mann and Powers, 1957), it must be said that the "success-rate" overall has been disappointingly low (Powers, 1963). A possible reason for this is that the method has often been applied to relatively complex characters (*e.g.* height, yield) which have many possible biological causes of variation. It is thus perhaps not surprising that the genetical causes of variation in these characters are also relatively complex. This emphasis on relatively complex characters is very understandable, for the measurement of biologically more specific characters (*e.g.* the activity of a particular enzyme, the amount of a particular biochemical compound, some parameter of a specific element of a single physiological system) is generally much more time-consuming, and Powers' method requires a relatively large total number of individuals for its application. It is therefore an important feature of the methods described in this paper that some of the minor technical innovations suggested (particularly in the construction of frequency distributions and the testing of the difference between two frequency distributions) have the result that considerably fewer individual measurements are needed for the application of these methods than for the original form of this analysis suggested by Powers.

In addition, this paper serves as a necessary preliminary to further papers in which the detection and estimation of genetic linkage of loci affecting continuously variable characters is discussed. This represents an advance on Powers' method, since the detection of close linkage or even pleiotropy with a "marker" gene would provide convincing confirmatory evidence of the segregation of a single identifiable locus affecting a continuously variable character. The question of how many segregating loci are likely to be affecting a particular continuously variable character will be taken up again in the discussion.

2. METHODS

(i) *The construction of frequency distributions*

It will be necessary several times in the course of the analysis to construct the frequency distribution of a metric, $\phi(x)$, directly from a set of experimental observations which are a random sample of the distribution. The usual method of doing this is to draw a histogram. This method is unsatisfactory if the number of observations is less than about a thousand, as is demonstrated by the difficulty of deciding upon the class interval to be used. If the class interval is too small, the histogram becomes very "spiky" and it is difficult to judge the overall shape. On the other hand, if the class interval is too large, non-identical observations are treated as being indistinguishable and useful information is lost. The compromise selected is usually open to both these objections at once.

This difficulty is overcome here in the following way. The first step is to plot the cumulative frequency against the corresponding values of the metric (fig. 2). Each individual observation is represented by a point on

this graph, so that this representation contains all the information of the original data. (It is useful to add one "dummy" observation to the total number of observed individuals when calculating percentage cumulative frequency so that the plot is independent of which end one starts from.)

FIG. 2.—Cumulative frequency plot. Data are from simulated back-cross "A".

FIG. 3.—Expected frequency distribution, ϕ_E , and observed frequency distribution, ϕ_O , superimposed. For calculation of χ^2 test to fit, see text.

The second step is to draw a smooth curve through all the points on this graph (see fig. 2). This procedure is less subjective than it may appear, for two curves drawn independently by different people are always very alike. The curve could of course be fitted by the polynomial giving the least squares deviation, subject to the conditions that the curve should be monotonic increasing, but this would involve a large amount of computation, and since any subsequent conclusions are subjected to statistical tests which

are independent of any error introduced by the subjective drawing of a smooth curve, the simple subjective curve is not misleading.

The third and final step is to differentiate this smooth curve*. This can be conveniently done graphically by measuring the slope at a suitable number of evenly spaced points, including especially all points where the slope appears to reach a maximum or minimum. The graphical differentiation of fig. 2 is shown in fig. 5. This is the final representation of the frequency distribution, $\phi(x)$, of the original measurements.

(ii) *A test of the fit between two frequency distributions*

It will also be necessary in the course of this analysis to be able to test the significance of the difference between an observed frequency distribution, ϕ_O , and expected frequency distribution, ϕ_E . The method used here is as follows.

ϕ_O and ϕ_E are superimposed (fig. 3). The χ^2 test of fit will be most sensitive and specific if the class intervals are drawn at the points of intersection of ϕ_E and ϕ_O as shown in fig. 3.

Let e_i be the number of individuals in ϕ_E in the class interval i , and o_i the corresponding number in ϕ_O . If $e_i < 5$ for any i , then the class i should be merged with an adjacent class. Then the test of fit is given by the following χ^2 statistic:

$$\chi^2_{l-1} = \sum_{i=1}^{i=l} \frac{d_i^2}{e_i}$$

where $d_i = o_i - e_i$; l = total number of classes.†

(iii) *The construction of parental and F_1 frequency distributions*

The frequency distributions of the two parental strains, (ϕ_{P_1} and ϕ_{P_2}) and of the F_1 , (ϕ_{F_1}) are constructed from the data as described in section (i) above. These distributions are then tested for fit to a normal distribution of the same mean and variance by the method of section (ii). If the χ^2 for fit is significantly large, the experimental observations are subjected to any simple transformation (e.g. logarithm, arc sin, etc.) which are suggested both by the nature of the experimental measurements and by the form of the deviation between the observed and the expected normal distribution. (Any such transformation adopted must of course be applied consistently to all measurements of the character.) If no such transformation is available, then the observed distribution constructed by the method of section (i) must be used as the best estimate of ϕ . However, it will usually be found that the data do fit a normal distribution, with or without simple transformation.

When the data fit a normal distribution it is preferable to replace the observed distributions by the normal distributions with the same mean and

* See note on p. 224.

† It has been kindly pointed out to me by Professor C. A. B. Smith and Dr R. C. Elston among others, that the choice of class-boundaries in this method is "data-guided" and hence is not statistically unimpeachable. The method thus gives a maximum rather than an unbiased estimate of the significance of the difference between two distributions and this should be borne in mind in its application. In so far as this paper is concerned, it is only the opposite error (e.g. a failure to detect existing evidence that a distribution is non-normal, or that a backcross distribution cannot be accounted for by single-locus segregation) which would give seriously misleading results.

variance, since there are general *a priori* reasons for expecting these distributions to be normal. If two or more of ϕ_{P_1} , ϕ_{P_2} , and ϕ_{F_1} are replaced by normal distributions, the variance ratios can be tested for deviation from unity by Fisher's F test (Fisher, 1954). If two or more of the variances are not significantly different by this test they can each be replaced by the pooled estimate of the common variance.

(iv) *The test of a single locus model*

The test of a single locus model is now quite straightforward. The appropriate parental and F_1 frequency distributions are constructed as described in section (iii). If the difference between the parental and F_1 distributions is due to a single Mendelian locus, assuming that the environment has been kept constant over the three generations and that there is no viability disturbance, the back-cross should consist of statistically equal numbers of parental and F_1 types. The "expected" back-cross distribution on a single locus model is therefore given by the composite distribution made up of 50 per cent. of the parental distribution and 50 per cent. of the " F_1 " distribution. The observed back-cross can be tested against this expected back-cross distribution as described in section (ii) above. If the fit is satisfactory, a single locus is sufficient to account for the facts and there is no evidence requiring us to postulate more complex models. It is, however, useful to have an idea of the extent to which other models could account for the data. The alternative models to be tested here are those with n equal and additive loci, where n varies from 2 to infinity. It is assumed that all the alleles increasing the character are grouped in one parent, and all the alleles decreasing the character are grouped in the other parent. It is also assumed that the dominance ratios are identical at all loci. If the back-cross data do fit the one-locus model, the fit of these particular alternative models is a function only of

$$d = \frac{\mu_{P_1} - \mu_{F_1}}{\sigma}$$

where μ_{P_1} and μ_{F_1} are the parental and F_1 means respectively, and σ is the common standard deviation of the parental and F_1 distributions. The graph of $F_{r,s}$ against d for values of $n = 1, 5, 10$, and infinity is given in fig. 4. This graph has been constructed from the variances expected in the back-cross generation on the basis of each particular model. Hence $F_{r,s}$ is Fishers' F-statistic for variance ratio, where r = number of back-cross individuals - 1, and s = number of parental + number of F_1 individuals - 2. Figure 4 thus enables calculation of the probability that a certain n -locus model also fits the data. Of course, if one is considering alternative models with a number of unequal loci, or with loci linked in coupling, the distinctions become correspondingly less certain.

(v) *The choice of an alternative genetic model*

If the simple one-locus model does not fit the data satisfactorily, the shape of the observed backcross distribution (ϕ_{BC_1}) in relation to the "expected" distribution ($\frac{1}{2}\phi_{P_1} + \frac{1}{2}\phi_{F_1}$) will usually suggest not only the reasons for the failure of the one-locus model but also an alternative genetic

model that may be tested. The general procedure here is to subtract from ϕ_{BC} , a distribution proportional to $(\phi_{P_1} + \phi_{F_1})$ until part of the distribution

FIG. 4.—Values for $F_{n,m}$ for alternative genetic models if data fit 1-locus model, plotted as a function of

$$d = \frac{\mu_{P_1} - \mu_{F_1}}{\sigma}$$

where μ_{P_1} = parental mean.

μ_{F_1} = F_1 mean.

σ = common standard deviation of parental and F_1 distributions.

n = number of back-cross individuals.

m = number of parental individuals + number of F_1 individuals - 2.

is reduced to zero. The remainder of the distribution then represents the "recombinant" classes; it is usually possible to estimate the number, means, variances and frequencies of these recombinant classes directly from

this remainder. If the means of the two "recombinant" classes in a two-locus model are asymmetric with respect to the means of the parental and F_1 distributions, this indicates that there is non-allelic interaction between the two loci. This procedure is probably not very meaningful if more than two loci, and hence more than two "recombinant" classes, are postulated. Even in this case it can be salutary to consider the effect of shifting the parental and/or F_1 distribution by one or two standard errors of their mean.

With this the discussion of the choice of a specific genetic model is concluded. If no one or two-locus models can be constructed to fit the data, then the analysis ends at this point with the conclusions that more than two loci are segregating.

3. RESULTS

The techniques developed in the methods section above will now be illustrated by application to some data simulated by appropriate use of random numbers, kindly prepared for me by Professor J. M. Thoday. Each of the distributions he provided consisted of 40 individual values. These distributions were analysed by the methods presented in this paper without knowledge of the genetic model Professor Thoday had built into the distributions. The degree to which the conclusions fit the model built into the distribution provide a test of the efficiency of these methods.

(a) Construction of parental and F_1 distributions

The comparisons between the simulated parental and F_1 distributions, and the theoretical distributions from which they were drawn are given in table 1. None of the simulated distributions is significantly different, in

TABLE 1
Comparisons between simulated parental and F_1 distributions, and the theoretical distributions from which they were drawn
Each simulated distribution consists of 40 individuals

Distribution		Theoretical basis of simulation	Calculated	P
P_1	Mean	0.00	0.05 ± 0.16	> 0.7
	Variance	1.00	1.14	> 0.1
	Normality	(Normal)	$\chi^2_2 = 2.4$	> 0.3
F_1	Mean	2.40	2.46 ± 0.16	> 0.7
Back-crosses A-D	Variance	1.00	1.04	> 0.2
	Normality	(Normal)	$\chi^2_3 = 2.8$	> 0.3
F_1	Mean	1.20	1.19 ± 0.16	> 0.9
Back-crosses E-F	Variance	1.00	0.95	> 0.2
	Normality	(Normal)	$\chi^2_2 = 1.7$	> 0.3

mean, variance or normality, from the corresponding "theoretical" distributions. Each of the simulated parental and F_1 distributions has therefore been replaced by normal distributions with the same calculated mean and pooled estimate of variance.

(b) *Test of single-locus model*

The observed back-cross distributions, and that expected on the basis of a single locus model, are drawn superimposed for each of the six simulated back-cross distributions A-F (fig. 5). The values of χ^2 for fit between the observed and expected distributions are given in table 2.

FIG. 5.—Simulated back-cross distributions A-F: in each case the distribution expected on single-locus model is superimposed.

Distribution A clearly fits the one-locus model. Indeed, the form of deviation from expectation is only that the distribution is *too* bimodal.

Since the distribution A fits the one-locus model, it is possible to test the fit of certain alternative models using fig. 4. The value of

$$d = \frac{\mu_{P_1} - \mu_{F_1}}{\sigma}$$

for distribution A is

$$\frac{2.46 - 0.05}{1.09} = 2.21.$$

From fig. 4 it is possible to read off the F-ratio for the two-locus model with $d = 2.2$; this gives a value of 1.4. There were 40 back-cross individuals and 80 parental + F_1 individuals. The test-statistic is then $F_{39,78} = 1.4$; $P = 0.05$. Thus the two-locus model would fit the data significantly less well.

TABLE 2
Test fit of observed back-cross distributions to those expected on single-locus model

Back-cross	χ^2	P
A	$\chi^2_4 = 2.54$	> 0.6
B	$\chi^2_2 = 5.4$	0.07
C	$\chi^2_2 = 7.9$	< 0.02
D	$\chi^2_2 = 10.6$	< 0.01
E	$\chi^2_2 = 1.4$	> 0.5
F	$\chi^2_2 = 2.0$	> 0.3

(c) *The choice of alternative models*

Distribution B does not fit the one-locus model well ($P = 0.07$). The reason for this is suggested if ≈ 80 per cent. of the expected one-locus distribution ($\frac{1}{2}\phi_{P_1} + \frac{1}{2}\phi_{F_1}$) is subtracted from the observed back-cross distribution (fig. 6). A clear single central peak remains. This suggests that there are two equal recombinant classes, making up 20 per cent. of the total distribution.

FIG. 6.—Residual distribution from simulated back-cross B after subtraction of 80 per cent. of the composite distribution ($\frac{1}{2}\phi_{P_1} + \frac{1}{2}\phi_{F_1}$).

This implies that there are two loci of equal and additive effect linked in coupling 20 map units apart. Standard errors are calculated as follows: there is 0 per cent. recombination (*i.e.* a single locus situation) with probability 0.07, which corresponds to 1.8 standard errors. The recombination frequency may thus be given as $20 \pm 20/1.8 = 20\% \pm 11\%$. On the basis of 20 per cent. recombination there are 4 individuals in each of the recombinant classes: taking the variance as equal to the parental and F_2 variance, this gives a standard error of the mean of each of the recombinant classes of

$$\sqrt{\frac{1.07}{4}} \approx 0.5,$$

corresponding to an effect of each of the two loci of 1.2 ± 0.5 .

Distribution C is significantly different from that expected on a single-locus model ($P < 0.02$). The reason (fig. 5) is clearly that there are some significant central peaks intermediate between the parental and F_1 distributions. Subtracting the composite ($\frac{1}{2}\phi_{P_1} + \frac{1}{2}\phi_{F_1}$) distribution gives an estimate of 42 per cent. for these central recombinant peaks. The standard error on this 42 per cent. (representing a number of recombinants, 17, which is approximately binomially distributed) is ± 8 per cent. There is thus no significant indication that the two loci are linked. The fact that there are two different central peaks with symmetrically placed means of 0.8 and 1.6, suggests that there are two loci segregating with additive but unequal effects, *i.e.* 1.6 and 0.8 units respectively. Assuming a variance equal to that of the parental and F_1 classes, the standard errors on each of these effects are

$$\pm \sqrt{\frac{1.07}{8.5}} = \pm 0.4.$$

The difference between the effects of the two loci is thus

$$0.8 \pm 0.6, t_8 = 1.3 \quad P \approx 0.2,$$

i.e. not significant although suggestive.

Distribution D is also significantly different from that expected on a single-locus model ($P < 0.01$). From fig. 5 it is apparent that the reason for this significant deviation is the presence of a large central peak. By a process similar to that conducted above for distributions B and C, it is concluded that two loci are segregating the effect of each being 1.2 ± 0.3 , the map distance between them being $50\% \pm 8\%$.

The difference in mean between the parental and F_1 distribution is smaller for back-cross distributions E and F although the environmental variance remains the same (table 1). In these cases, neither of the back-cross distributions is significantly different from that expected on the basis of a single locus model (table 2). It is therefore possible to use fig. 4 to test the fit of certain alternative models as was done for distribution A. In this case

$$d = \frac{1.19 - 0.05}{1.00} = 1.14.$$

From fig. 4 a value of $d = 1.14$ gives values of the F-ratio for certain alternative models as follows:

$$2 \text{ loci } F_{39,78} = 1.2 \quad P = 0.2$$

$$4 \text{ loci } F_{39,78} = 1.3 \quad P = 0.1$$

$$10 \text{ loci } F_{39,78} = 1.5 \quad P < 0.05$$

4. DISCUSSION

The conclusions reached by analysis of the data, and the theoretical basis on which the data were simulated, are compared in table 3.

The analysis has been successful in that within the stated limits of error the conclusions drawn agree with the basis on which the data was simulated. The methods developed in this paper therefore seem to be valid. It remains now to discuss the necessary conditions for the application of these methods.

(i) *Number of loci*

The methods of this paper can only be fully carried out if the bulk of genetic variation in a back-cross is due to the segregation of two loci at most. (This is of course the situation they have been designed to detect.) This

TABLE 3
Comparison between basis for simulation and conclusions from analysis

Back-cross distribution	Theoretical basis for simulation	Heritability	Conclusion from analysis
A	Single locus Effect 2.4	0.59	Single locus (not two loci, P = 0.07) Effect 2.4 ± 0.2
B	Two loci Additive effects 1.2 1.2 Linked in coupling Recombination: 20%	0.53	Two loci (not one, P = 0.07) Additive effects 1.2 ± 0.5 1.2 ± 0.5 Linked in coupling Recombination: $20\% \pm 11\%$
C	Two loci Additive effects 1.2 1.2 Linked in coupling Recombination: 50%	0.44	Two loci (not one, P < 0.02) Additive effects 1.6 ± 0.4 0.8 ± 0.4 Linked in coupling Recombination: $42\% \pm 8\%$
D	Two loci Additive effects 1.6 0.8 Linked in coupling Recombination: 50%	0.42	Two loci (not one, P < 0.01) Additive effects 1.2 ± 0.3 1.2 ± 0.3 Linked in coupling Recombination: $50\% \pm 8\%$
E	Two loci Additive effects 0.6 0.6 Linked in coupling Recombination: 50%	0.26	<div style="display: flex; align-items: center;"> <div style="font-size: 3em; margin-right: 10px;">}</div> <div> No evidence against single locus Possibly 2 loci P = 0.2 Probably < 4 loci P = 0.1 less than 10 loci P < 0.05 Effect of single locus: 1.14 ± 0.23 </div> </div>
F	Single locus Effect 1.2	0.26	

condition is not as restrictive as it may at first sight appear, since it is relatively easy to reduce the number of genes segregating in any particular back-cross by repeated back-crossing to an inbred line (Wehrhahn and Allard, 1965), and other methods based on character analysis (see Thoday, 1967) may also achieve the same effect.

(ii) *Heritability*

In general, it appears from table 3 that the segregation of a locus will be detected if the effect of the locus is approximately one standard deviation of residual variance, but not if it is much less than this. In order to detect single-locus segregation this means a value of

$$d = \frac{\mu_{P_1} - \mu_{F_1}}{\sigma}$$

of about 1 corresponding to a heritability arising from segregation of this

locus in the back-cross of 0.3; for a two-locus situation the corresponding values are $d \simeq 2$, $h^2 \simeq 0.4$. If there is dominance, *i.e.*, the F_1 mean is not half-way between the two parents, the values of d will not be the same for each of the two back-crosses. In this case it is clearly more informative to observe the back-cross to the relatively recessive parent, for which the value of d will be greater.

(iii) *Numbers of individuals*

As mentioned in the section on methods, error in the estimates of the means of the parental and F_1 distributions can cause significant differences in the conclusions drawn. This factor becomes troublesome if the standard error of the parental or F_1 mean is greater than about 10 per cent. of the difference between their means. For a value of $d \simeq 1$, this condition reduces to

$$\frac{\text{Standard deviation}}{\text{Standard error of mean}} \geq 10,$$

so that the number of P_1 and F_1 individuals required is $> 10^2 = 100$. For a value of $d \simeq 2$ the corresponding number of P_1 and F_1 individuals is 25.

A suitable number of back-cross individuals is the sum of the parental and F_1 numbers; in the minimum case just given this is 50. In this case, if the data fit a single locus model, and $d = 1$, the probability that the data also fit the extreme alternative of a many-locus model is given from fig. 4 as $F_{50.50} = 1.4$, $P = 0.05$. This is clearly the limiting case for which the analysis is meaningful.

(iv) *Metric*

Finally, it is important that the metric measured should be a genuinely continuous variable. If the metric is only semi-quantitative (*e.g.* there are only, say, 10 discrete phenotypic classes), the data are effectively in the form of a histogram, entailing a consequent loss of information (see above). When this is so, the above conditions become more stringent.

5. SUMMARY

1. Methods are described which provide a practical and reasonably economical means of testing whether segregation at one or two loci will account adequately for continuous quantitative variation in any particular back-cross. These methods are based on the comparison of parental, F_1 and back-cross distributions.

2. The validity of these methods is tested empirically by application to some simulated data.

3. The minimum requirements for the application of these methods are in summary as follows: there should be two inbred strains; the characters to be measured should be continuous variables with a heritability of at least 0.3 in one of the back-crosses; and it should be possible to measure a total of 100 individuals from three successive generations.

4. These conditions may be satisfied by a wide variety of organisms used for biological research, and are not restricted to those usually regarded as genetically "well worked".

Acknowledgments.—The ideas set out in this paper have evolved by a process of criticism and discussion with many people. To all of these I am most grateful. I would particularly like to mention Professor J. M. Thoday, who also provided the simulated data.

6. REFERENCES

- EAST, E. M. 1916. Inheritance in crosses between *Nicotiana Langsdorfi* and *Nicotiana Alata*. *Genetics*, 1, 311-333.
- FISHER, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinburgh*, 52 (2), 399-433.
- FISHER, R. A. 1954. *Statistical Methods for Research Workers*, 12th Edition. Oliver & Boyd, Edinburgh.
- JINKS, J. L. 1954. The analysis of continuous variation in a diallel cross of *Nicotiana Rustica* varieties. *Genetics*, 39, 767-788.
- LAW, C. N. 1967. The location of Genetic factors controlling a number of quantitative characters in wheat. *Genetics*, 56, 445-461.
- LEONARD, W. H., MANN, H. O., AND POWERS, L. 1957. Partitioning method of genetic analysis applied to plant-height inheritance in barley. *Colorado Agr. Mech. College Agr. Exp. St. Tech. Bull.*, 60.
- MATHER, K. 1949. *Biometrical Genetics*. Methuen, London.
- MÉRAT, P. 1967. L'étude des distributions de fréquences en vue de déceler la présence de gènes à effet important. *Ann. Biol. anim. Biophys.*, 7, 489-493.
- PAYNE, F. 1918. An experiment to test the nature of the variations on which selection acts. *Indiana Univ. Stud.*, 5, 1-45.
- POWERS, L. 1963. The partitioning method of genetic analysis and some aspects of its application to plant breeding. In *Statistical Genetics and Plant*, Ed. Hanson, W. D., and Robinson, H. F., pp. 280-318. National Academy of Sciences, Washington.
- SAX, K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics*, 8, 552-560.
- SPICKETT, S. G., AND THODAY, J. M. 1966. Regular responses to selection. 3. Lateraction between located polygenes. *Genet. Res.*, 1, 96-121.
- THODAY, J. M. 1961. Location of polygenes. *Nature*, 191, 368-370.
- THODAY, J. M. 1967. Genes in the study of continuous variation. *Cienca e cultura*, 19, 54-63.
- WEBER, E. 1959. The genetical analysis of characters with continuous variability on a Mendelian basis. I. Monohybrid segregation. *Genetics*, 44, 1131-1139.
- WEHRHAHN, C., AND ALLARD, R. W. 1965. The detection and measurement of the effects of individual genes involved in the inheritance of a quantitative character in wheat. *Genetics*, 51, 109-119.
- WRIGHT, S. 1927. The effects in combination of the major color factors of the guinea pig. *Genetics*, 12, 530-569.

Note added in proof (see page 215).—I am grateful to Dr. A. R. G. Owen for the suggestion that both the fitting of a smooth curve to the cumulative plot, and the subsequent differentiation, could well be done objectively by the methods of numerical analysis. Using the "cumulative number of individuals" as the abscissa to take advantage of the equally-spaced intervals, and fitting a quadratic equation to seven consecutive points by least-squares, the following useful mid-point formula is obtained:

$$\phi_{x_0} = \frac{28}{3x_3 + 2x_2 + x_1 - x_{-1} - 2x_{-2} - 3x_{-3}},$$

where ϕ_{x_0} is the required frequency at the midpoint x_0 , and $x_{-3}, x_{-2}, \dots, x_3$ are the seven consecutive values of x . This formula works well in practice, and has the additional advantage that the frequency curve can be constructed directly from the experimental observations without the intermediate stages of plotting, drawing and measuring the cumulative curve.