

Polygene analysis

II. SELECTION

BY NEIL GILBERT

John Innes Institute, Bayfordbury, Hertford, Herts.

(Received 1 May 1961)

This paper continues and concludes a previous one (Gilbert, 1961).

CORRECTION

The suggestion on page 103 of the previous paper that the F_3 mean can provide extra information (when using the curved model) is wrong, since the F_3 mean is still linearly related to the F_1 and F_2 means ($2F_3 = 3F_2 - F_1$). This means that (1) the curved model is not sufficiently general and (2) a scaling test that compares the F_3 mean with the means of previous generations is no guarantee of additivity. This mistake does not affect the main argument; I apologize for my fallibility.

SELECTION

I shall consider the progress of selecting and selfing, starting from an F_2 progeny. The notation is that of the previous paper. A genotype homozygous for l 'good' genes, n 'bad', and heterozygous at m loci is referred to as (l, m, n) . Its frequency in F_r is

$$\frac{(l+m+n)!}{l!m!n!} u_r^l v_r^m w_r^n \quad (u+v+w=1),$$

and it experiences a selection pressure $\alpha(2\beta)^m \gamma^n$. The relative values of $\alpha:\beta:\gamma$ are constant, but their absolute values in each generation are adjusted so that

$$u\alpha + 2v\beta + w\gamma = 1.$$

This form of selection pressure is chosen for its algebraic tractability; it implies that the selection pressures at different loci are the same, and are independent of each other. Then consideration of the frequency with which (l, m, n) is derived in F_{r+1} from each individual in F_r leads to the relations

$$u_{r+1} = \alpha u_r + \frac{1}{2}\beta v_r$$

$$v_{r+1} = \beta v_r$$

$$w_{r+1} = \gamma w_r + \frac{1}{2}\beta v_r$$

The implication—that the distribution remains multinomial in succeeding generations—is true for the form of selection pressure adopted. Since in F_2 , $u = \frac{1}{2}v = w = \frac{1}{4}$, we find that, in F_{r+1} ,

$$u:v:w = \frac{\alpha^r - \beta^r}{\alpha - \beta} : 2\beta^{r-1} : \frac{\gamma^r - \beta^r}{\gamma - \beta}$$

these expressions being suitably multiplied to make their sum unity. This result can easily be checked by substitution in the above equations. It applies to *genotypic* selection from selfed progenies. Here I shall not consider other mating systems, nor the difficult problem of relating phenotypic to genotypic selection. At present, this result is useful only (1) for considering the possible importance, after selection, of the departures from additivity considered in the previous paper, and (2) as a theoretical basis for comparing responses to different rates of selection. (It is astonishing how rare are experiments to compare several different selection rates.) In the absence of selection, $\alpha = 2\beta = \gamma = 1$, so that $u = w = \frac{1}{2} - (\frac{1}{2})^{r+1}$ and $v = (\frac{1}{2})^r$.

If we adopt the curved model considered in the previous paper, the generation mean is approximately

$$\bar{F}_{r+1} = x^y [b + d_2 y + 4(1 - \frac{u-w}{2})^2 d_1 + 2v(h_1 - 2d_1 + h_2 y)],$$

where

$$y = 2 \frac{f+g}{f-g} (u-w).$$

This expression is obtained simply (if laboriously) by averaging the phenotype $[b + (l-n)d + mh] \theta^{l-n}$ over the multinomial distribution in l, m, n , and then substituting and approximating in θ in the manner of the previous paper. The expression is not symmetric in α and γ , so that upwards and downwards selection will give asymmetric results. If there is no selection,

$$\bar{F}_{r+1} = b + 4d_1 + (\frac{1}{2})^{r-1}(h_1 - 2d_1).$$

The recurrence relation $2\bar{F}_{r+1} = 3\bar{F}_r - \bar{F}_{r-1}$ then holds for the curved model, and the limiting value of \bar{F}_r (as inbreeding proceeds) is $b + 4d_1$ which is equal to $2\bar{F}_2 - \bar{F}_1$. Here d_1 is a 'curvature' term which does not appear in the additive theory. On this theory, therefore, $2\bar{F}_2 - \bar{F}_1$ would be a better predictor than $\frac{1}{2}(P_1 + P_2)$ of the average result of unselected inbreeding. Finally, the limit of upward selection ($u = 1, v = w = 0$) is $x^y(b + d_2 y)$, where y is now $2(f+g)/(f-g)$. This result could of course have been obtained directly from the original expression for the phenotype of $(f+g, 0, 0)$. It is necessary, for the argument of the next section, to note that y is a function of the degree of genetic dissimilarity between the two parents; it is independent of the genotype-phenotype relation.

SCALING TESTS

Giesbrecht (1961) has published some more sets of family means. To fit the expressions given on page 101 of the previous paper to six means, we must solve the quartic

$$\phi(x) = x^2(P_2 x^2 - 2B_2 x + F_1) - 2x(B_2 x^2 - 2F_2 x + B_1) + (F_1 x^2 - 2B_1 x + P_1) = 0$$

(obtained by eliminating b, d_1, d_2, h_1, h_2). In no less than five of Giesbrecht's six examples, this equation has no real root near $x = 1$. The diploid inheritance of the characters concerned cannot be doubted. The curved model therefore cannot be fitted to Giesbrecht's figures. In the first paragraph we saw that the curved model

imposed a relation between the F_3 mean and the F_1 and F_2 means, and concluded that the model is insufficiently general. The same conclusion must obviously be drawn here. Now it is perfectly simple to generalize the model further, perhaps by including a multiplicative dominance term so that the phenotype of (l, m, n) becomes $[b + (l - n)d + mh] \theta^{l-n} \psi^m$. But there are then too many parameters to be fitted to a limited set of data; statistically speaking, the situation is hopelessly fluid. Now this unfortunate result is not due merely to the unavoidable errors of estimation of family means, nor is it entirely due to the vagaries of the genotypic-phenotypic relation; to some extent it is intrinsic in the genetic relationship between inbred parents and their offspring. To see this, we consider a hypothetical case wherein the family means are known exactly and the three scaling tests

$$P_2 - 2B_2 + F_1 = B_2 - 2F_2 + B_1 = F_1 - 2B_1 + P_1 = 0$$

are satisfied, so that $x = 1$ is a solution of $\phi(x) = 0$. But

$$\frac{1}{4} \frac{d\phi}{dx} = x(P_2 x^2 - 2B_2 x + F_1) - (B_2 x^2 - 2F_2 x + B_1),$$

so that the scaling tests ensure, not only that $\phi = 0$ at $x = 1$, but also that $d\phi/dx = 0$. Consequently, any arbitrary value of x in the neighbourhood of 1 will fit the data. I take the $D \times J$ 1938 data used by Mather (1949) as an example, because they most nearly satisfy the scaling tests; other data that satisfy these tests will give similar results. Take arbitrary values 0.9, 1, 1.1 for x , we find:

x	0.9	1	1.1	
b	7.86	7.31	7.09	(± 0.06)
d_1	-0.03	0.10	0.16	(± 0.05)
d_2	1.71	0.91	0.23	(± 0.03)
h_1	-1.17	-0.90	-0.79	(± 0.05)
h_2	-0.05	0.05	0.11	(± 0.07)
$\Sigma(\text{obs.} - \text{exp.})^2$	0.0019	0.0019	0.0012	

The figures in brackets are the standard errors of the various parameters. $\Sigma(\text{obs.} - \text{exp.})^2$, the sum of squares of deviations of the family means from their expectations, is quoted here merely to indicate that—as is easily verified—the different values of x give (roughly) equally good fits. It may be objected that the additive and multiplicative terms in the model are compensating each other. This is true as far as the family means are concerned. But when we extrapolate outside the range from P_1 to P_2 , the consequences are quite different. Taking arbitrary (but plausible) values for y , the limit of selective advance becomes:

$x =$	0.9	1	1.1	
	$x^y(b + d_2 y)$			
$y = 2$	9.13	9.13	9.13	($= P_1$)
3	9.46	10.04	10.35	
4	9.64	10.95	11.72	
5	9.68	11.87	13.26	

We must conclude that the scaling tests do not guarantee additivity; they guarantee an indeterminate situation in which additivity is one solution. Of course, it is always possible to draw an infinite number of curves through a finite number of fixed points, but if these points are collinear the curves must usually appear improbably tortuous in comparison with the straight line. The present case is peculiar in that there is no apparent reason to prefer the additive solution, for the alternative models seem perfectly reasonable. One can of course invoke Occam's Razor to argue that, if the system admits an additive interpretation, it should be additive. But there is no guarantee that the genes will agree.

It follows, then, that predictions of absolute advance under selection—based on polygene analysis—cannot be relied on. (They may, of course, chance to be accurate in some cases; but we do not know which cases these will be.) This practical failure of polygene analysis does not invalidate the theoretical side of polygene work. Such ideas as that of polygene balance (Mather, 1943) are very valuable. But the practical difficulties are legion. In the previous paper it was shown that great care is needed in the choice of scale, before the additive analysis can be used. Worse remains behind; for it now appears that the assumption of additive gene effects would still be only an assumption, even if the scaling tests were perfectly satisfied. As soon as we investigate a simple non-additive model, the situation becomes indeterminate. For some things (e.g. the heritability) this may not matter much; for others (response to selection) it clearly does. I have not found a way of resolving this difficulty. The example above also shows that dominance and curvature are to some extent confounded. For practical analysis of quantitative inheritance, it seems wiser to adopt simpler methods, e.g. heritability, combining ability, offspring-parent regression, or the use of marker genes. It is particularly unfortunate that polygene analysis should sometimes be advocated as an aid to practical breeding work. Of course, the simpler methods mentioned also involve an additive hypothesis, and so might be equally sensitive to failure of additivity. But they might not; for the unpretentious model underlying the 'heritability' of an outbreeding population is almost a statistical convention in comparison with the detailed genetic assumptions involved in the polygene analysis of inbred parents. On the other hand, elaboration of the 'heritability' model to include terms for dominance, epistacy, etc., is likely to suffer, once again, from the kind of curvature discussed in this paper. (It is doubtful whether such elaboration is very helpful or meaningful, anyway.) This paper shows that polygene analysis has an intrinsic theoretical weakness. The curved model of gene action, being more flexible than the strictly additive model, shows up this weakness admirably; but neither the additive nor the curved model is sufficiently flexible to describe real life.

SUMMARY

A modest theory of polygene selection is presented. Recently published data show that the curved model of gene action (presented in a previous paper) is insufficiently general. The curved model does, however, show that the scaling tests (used in the additive type of polygene analysis) do not guarantee additivity; they guaran-

tee a state of indeterminacy, in which additivity is one of a range of reasonable possibilities. These different possibilities give entirely different predictions of selective advance. The failure (in practice) of polygene analysis does not reduce the value of polygene concepts.

I wish to thank the Referee for his valuable suggestions.

REFERENCES

- GIESBRECHT, J. (1961). The inheritance of ear height in *Zea mays*. *Canad. J. Genet. Cytol.* **3**, 26–33.
- GILBERT, N. (1961). Polygene analysis. *Genet. Res.* **2**, 96–105.
- MATHER, K. (1943). Polygenic inheritance and natural selection. *Biol. Rev.* **18**, 32–64.
- MATHER, K. (1949). *Biometrical Genetics*. London: Methuen.