THE ENUMERATION OF GENOTYPE-PHENOTYPE CORRESPONDENCES

J. H. BENNETT

Department of Genetics, Adelaide University

Received 18.i.57

THE distinction between an individual's genotype, which may be determined by reference to its breeding performance and possibly that of its relatives, and its phenotype, which refers only to its observable characteristics, is fundamental in genetics. Even when only one locus is concerned, many different genotype-phenotype relationships are possible, especially if there are more than two allelomorphs. In this article, we are concerned with a problem of enumeration which arises in the classification of the genotype-phenotype relationships at a single diploid locus.

FIG. 1.—The five two-allelomorph phenotype systems classified into four phenograms (after Cotterman, 1953).

There are $\frac{1}{2}m(m+1)$ possible diploid genotypes at a single locus with m allelomorphs. We shall suppose that each of these genotypes corresponds with one and only one phenotype but more than one genotype may correspond with the same phenotype. The $\frac{1}{2}m(m+1)$ genotypes may therefore correspond with any number of phenotypes from one to $\frac{1}{2}m(m+1)$. Following Cotterman (1953), such a genotype-phenotype correspondence will be called a *phenotype system* and two phenotype systems, each involving the same number of allelomorphs and phenotypes, will be said to be permutationally equivalent or to be images of the same *phenogram* if there exist one or more permutations of the gene symbols which transform one phenotype system into the other. Thus, with two allelomorphs and three genotypes, there are five phenotype systems. Cotterman (1953) has represented these phenotype systems as shown in fig. 1, where the genotypes are represented by circles, open and closed for homozygotes and heterozygotes respectively, and any two genotypes which are phenotypically

indistinguishable are joined by a line. Fig. 1 (i) represents the case of no dominance, figs. 1 (ii) and 1 (iii) the two cases of dominance, fig. 1 (iv) the case in which the two homozygotes are phenotypically indistinguishable from each other but both differ phenotypically from the heterozygote, and fig. 1 (v) the trivial case in which all the

FIG. 2.—The 203 three-allelomorph phenotype systems classified into 52 phenograms (after Cotterman, 1953).

genotypes correspond with the same phenotype. Clearly, the two phenotype systems represented by figs. 1 (ii) and 1 (iii) are images of the one phenogram and there are thus four distinct phenograms in this case.

With three allelomorphs, there are six genotypes and 203 phenotype systems, *i.e.* there are 203 different ways in which these six genotypes can be grouped so as to correspond with any number of phenotypes from one to six. Cotterman (1953) has shown that in this case there are 52 phenograms (see fig. 2). With four allelomorphs

FIG. 2.—continued

there are ten genotypes and 115,975 phenotype systems. The phenotype systems have not been described for this case but Cotterman conjectured that the phenograms would number about 7000. With five allelomorphs, there are fifteen genotypes and 1382,958,545 phenotype systems. Cotterman guessed that the phenotype systems would here belong to about 17,000,000 phenograms. The exact enumeration of these phenograms presents a rather intricate combinatorial problem, and has hitherto not been described.

Each m -allelomorph phenotype system may be represented in partitional notation as a partition of the total number $G = \frac{1}{2}m(m+1)$ of genotypes. For any such partition

$P = (p_1^{\pi_1} p_2^{\pi_2} \dots)$ of G , there are

$$\beta(P) = \frac{G!}{(p_1!)^{\pi_1} (p_2!)^{\pi_2} \dots \pi_1! \pi_2! \dots}$$

distinct phenotype systems. This is the number of ways in which G different genotypes can be grouped into $S_{\pi_i} = \phi$ different phenotypes according to the partition P . The total number $N_{m,\phi}$ of m -allelomorph phenotype systems in which the G genotypes correspond with exactly ϕ phenotypes is found by summing $\beta(P)$ over all partitions P with ϕ parts. The total number N_m of m -allelomorph phenotype systems is equal to $\beta(P)$ summed over all partitions P of the number G . Values of $N_{m,\phi}$ are given in table XXII of Fisher and Yates (1953), as $\Delta^{\phi} O^G / \phi!$, the leading ϕ -th difference of the G -th powers of the positive integers commencing at zero, for values of G from 2 to 25 and values of ϕ from 2 to G . Values of N_m can be obtained by summing $N_{m,\phi}$ over all values of ϕ from 1 to G .

The problem of the enumeration of the phenograms is much more difficult and belongs to a general class of enumerations which has been considered by Fisher (1950). A formula (phenotype system) is constructed by inserting appropriate "contents" into a finite number of "cells". Any two formulæ which can be derived from one another by a permutation of the basic group of permutations among the cells are said to belong to the same set (phenogram). We wish to enumerate the number of such sets. Every permutation h of the complete permutation group of m allelomorphs produces a corresponding permutation g of the $\frac{1}{2}m(m+1)$ genotypes. If the permutation g is of partitional type Q , then of the $\beta(P)$ phenotype systems which correspond to a given partition P of $\frac{1}{2}m(m+1)$, the number which are left unaltered by the permutation g is the bipartitional function $\phi(P, Q)$ (Fisher, 1950; Bennett, 1956). Summing over all partitions P , we obtain the total number of phenotype systems which are unaltered by the permutation g as

$$\sum_P \phi(P, Q) = e(Q).$$

$P \vdash \frac{1}{2}m(m+1)$

Since there are $m!$ permutations of the m allelomorphs, the total number of m -allelomorph phenograms must be

$$n_m = \frac{1}{m!} \sum_Q e(Q)$$

where the summation is taken over all partitions Q corresponding to the $m!$ permutations of the $\frac{1}{2}m(m+1)$ genotypes induced by all possible permutations of the m allelomorphs. Since there are

$$\alpha(R) = \frac{m!}{\rho_1! \rho_2! \dots r_1^{\rho_1} r_2^{\rho_2} \dots}$$

permutations of the m allelomorphs of partition type

$$R = (r_1^{\rho_1} r_2^{\rho_2} \dots),$$

it follows that

$$n_m = \frac{1}{m!} \sum_{R/m} \alpha(R) e(Q(R))$$

where the summation is taken over all partitions R of the number m . Values of $e(Q)$ can readily be calculated using the following operational procedure (Bennett, 1956):—for every separation of the partition Q , construct a differential operator by forming the product of (i) $(D+f)$ for each separate with only one part and with f factors (excluding unity) in that part, and (ii) for each separate with more than one part, the sum of a^{s-1} for all factors a excluding unity which are common to the s parts in that separate. These operators are summed over all possible separations of the partition Q and are then allowed to operate on $F(x) = e^{e^x-1}$ when they yield $e(Q)$. Since

$$D^n F(x) = e(1^n) = \sum_r \Delta^r O^n / r!,$$

this operational procedure enables us to calculate $e(Q)$ very readily if we make use of published tables of $\Delta^r O^n / r!$ (Fisher and Yates, *loc. cit.*).

The calculation of n_m for $m = 3, 4, 5$ proceeds as shown in tables 1, 2, 3 respectively. The six permutations of three allelomorphs A, B, and C, may be classified according to the partitions of the number three as shown in the first column of table 1. The partition (1^3) represents the identity which leaves the allelomorphs unchanged, (21) represents an interchange of pairs of allelomorphs of which there are three possibilities, (AB), (BC) and (CA), and (3) represents a cyclic interchange of three allelomorphs of which there are two possibilities, (ABC) and (ACB). The permutations of the six genotypes brought about by these permutations of the allelomorphs are represented in the second column of table 1. Thus if the three allelomorphs are left unchanged, so will the six genotypes be unchanged and this corresponds with the identity (1^6) . An interchange of two allelomorphs corresponding to the partition (21) will lead to two possible interchanges of pairs of genotypes which may be represented by the

partition $(2^2 1^2)$. Thus, the interchange of allelomorphs A and B will lead to the interchange of genotypes AA and BB and also of AC and BC whilst the remaining genotypes AB and CC will be unchanged.

TABLE 1
Calculation of the number of phenograms with three allelomorphs

R	Q	$e(Q)$	$\alpha(R)$	$\alpha(R) e(Q)$
(1^3)	(1^6)	203	1	203
(21)	$(2^2 1^2)$	31	3	93
(3)	(3^2)	8	2	16

$$n_3 = \frac{1}{3!} S \alpha(R) e(Q(R)) = 52$$

TABLE 2
Calculation of the number of phenograms with four allelomorphs

R	Q	$e(Q)$	$\alpha(R)$	$\alpha(R) e(Q)$
(1^4)	(1^{10})	115975	1	115975
(21^2)	$(2^3 1^4)$	2119	6	12714
(2^2)	$(2^4 1^2)$	999	3	2997
(31)	$(3^3 1)$	73	8	584
(4)	$(4^2 2)$	55	6	330

$$n_4 = \frac{1}{4!} S \alpha(R) e(Q(R)) = 5525$$

TABLE 3
Calculation of the number of phenograms with five allelomorphs

R	Q	$e(Q)$	$\alpha(R)$	$\alpha(R) e(Q)$
(1^5)	(1^{15})	1382,958545	1	1382,958545
(21^3)	$(2^4 1^7)$	1,748,249	10	17,482,490
$(2^2 1)$	$(2^6 1^3)$	214,157	15	3,212,355
(31^2)	$(3^4 1^3)$	4622	20	92,440
(32)	$(63^2 21)$	386	20	7720
(41)	$(4^3 21)$	783	30	23,490
(5)	(5^3)	70	24	1680

$$n_5 = \frac{1}{5!} S \alpha(R) e(Q(R)) = 11,698,156$$

Finally, a cyclic interchange of the three allelomorphs corresponding to the partition (3) will lead to two cyclic interchanges of three genotypes corresponding to the partition (3^2) . For example, the cyclic interchange of the three allelomorphs corresponding to (ABC) will lead to the cyclic interchange of the genotypes AA, BB and CC and also of AB, BC and CA.

The numbers of phenograms with four and five allelomorphs are 5525 and 11,698,156 respectively, rather less than conjectured by Cotterman (1953), but of the same order of magnitude. The number of phenograms increases very rapidly with the increasing number of allelomorphs, an indication of the great variability inherent in a multi-allelomorphic system.

SUMMARY

A general method is described for the enumeration and classification of the genotype-phenotype relationships possible with a single diploid locus and any number of allelomorphs.

REFERENCES

- BENNETT, J. H. 1956. Partitions in more than one dimension. *J. Roy. Stat. Soc.*, B, 18, 104-112.
- COTTERMAN, C. W. 1953. Regular two-allele and three-allele phenotype systems. *Amer. J. Hum. Genet.*, 5, 193-235.
- FISHER, R. A. 1950. A class of enumerations of importance in genetics. *P.R.S.*, B, 136, 509-520.
- FISHER, R. A., AND YATES, F. 1953. *Statistical Tables for Biological, Agricultural and Medical Research*. Edinburgh : Oliver and Boyd.