

SPECIAL TRAIN ALGEBRAS ARISING IN GENETICS

by H. GONSHOR

(Received 18th January 1959)

(Received in revised form 22nd December 1959)

1. Introduction

Ever since Mendel promulgated his famous laws, probability theory and statistics have played an important role in the study of heredity (9). Etherington introduced some concepts of modern algebra when he showed how a non-associative algebra can be made to correspond to a given genetic system (1, 4). The fact that many of these algebras have common properties has led to their study from a purely abstract point of view (2, 3, 5, 6, 11, 12). Furthermore, the techniques of algebra give new ways of attacking problems in genetics such as that of stability.

We shall prove a stability theorem on special train algebras and apply it to several genetic systems.

2. A Stability Theorem for Special Train Algebras

For our purpose, the basis form of the definition will be most useful. A special train algebra is a commutative algebra for which there exists a basis a_0, a_1, \dots, a_n with a multiplication table of the following kind:

$$a_i a_j = \sum x_{ijk} a_k \text{ where}$$

$$x_{000} = 1, \quad (1)$$

$$\text{for } k < j, x_{0jk} = 0, \quad (2)$$

$$\text{for } i, j > 0 \text{ and } k \leq \max(i, j), x_{ijk} = 0, \quad (3)$$

and furthermore all powers of the ideal $\{a_1, a_2, \dots, a_n\}$ are ideals. (The powers I^r of an ideal I are defined inductively by $I^r = I^{r-1}I$).

In this section we shall require conditions (1), (2) and (3) only. It follows easily that $\{a_1, a_2, \dots, a_n\}$ is nilpotent and also that the map which sends each element into its coefficient of a_0 is the only non-trivial homomorphism of the algebra into the coefficient field. The x_{0jj} (abbreviated λ_j) will be called the train roots of the algebra. They are the characteristic roots of the operator which is multiplication by a_0 . Necessarily $\lambda_0 = 1$. (They include the principal train roots defined in (1), perhaps with repetitions, but it is important to note that they may also include $\frac{1}{2}$ even when this is not a principal train root.)

Theorem 2.1. *Every special train algebra which has no train root satisfying $2\lambda = 1$ has a unique non-zero idempotent.*

Proof. We shall construct the x_i inductively so that $\left(\sum_{i=0}^m x_i a_i\right)^2$ has the

form $\sum_{i=0}^m x_i a_i + \sum_{i=m+1}^n y_i a_i$. Since $x_{000} = 1$, x_0 must be 1 for a non-zero idempotent. Suppose x_0, x_1, \dots, x_n have been chosen. Then

$$\left(\sum_{i=0}^m x_i a_i + z a_{m+1} \right)^2 = \sum_{i=0}^n x_i a_i + \sum_{i=m+1}^r y_i a_i$$

for some y 's. This follows since from (2) and (3) the coefficient of a_i for $i \leq m$ in a product is independent of the coefficients of a_{m+1} in the factors. It remains only to equate the coefficients of a_{m+1} . Since from (3) $a_{m+1} a_i$ for $i \neq 0$ has coefficient of a_{m+1} equal to 0, the coefficient of a_{m+1} will have the form $2\lambda_{m+1}z + f(x_1, x_2, \dots, x_m)$. Since $2\lambda_{m+1} \neq 1$, the equation in z , $2\lambda_{m+1}z + f(x_1, x_2, \dots, x_m) = z$ has a unique solution which is taken as x_{m+1} . This proves the existence and uniqueness simultaneously.

Note. For a coefficient field of characteristic 2 the above condition is always satisfied. Otherwise, the condition may be stated in the form, no $\lambda = \frac{1}{2}$.

Similar theorems may be proved in cases where several of the λ 's are equal to $\frac{1}{2}$. The case where $\lambda_1 = \frac{1}{2}$ and all other λ 's are distinct from $\frac{1}{2}$ occurs frequently in genetics. By the same technique as before it is easy to show that non-zero idempotents exist only if $x_{001} = 0$ and in that case a one parameter family of idempotents exists. (In fact, x_1 may be chosen at will, and the other x 's are uniquely determined as before.) Similarly if r of the λ 's, and no more, are equal to $\frac{1}{2}$, it can happen that there is an r -parameter family of idempotents. This occurs in examples connected with multiple alleles and with linkage (3).

Henceforth, we shall consider algebras over the reals. We may then define $\sum_{m=0}^n x_{mi} a_m \rightarrow \sum_{m=0}^n x_m a_m \leftrightarrow x_{mi} \rightarrow x_m$ for all x . It is clear that convergence defined this way is independent of the basis used. In order to state the next theorem it will be convenient to define two concepts. The "weight" of an element is its image in the unique homomorphism of the algebra onto the reals. The "plenary powers" of an element A are defined as $A_0 = A, A_{n+1} = A_n^2$.

Theorem 2.2. *The sequence of plenary powers of an element of weight 1 in a special train algebra whose train roots other than $\lambda_0 = 1$ all have absolute value less than $\frac{1}{2}$ tends to an idempotent.*

Proof. Suppose $A_i = \sum_{m=0}^n x_{mi} a_m$. It is required to show that $x_{mi} \rightarrow x_m$ where x_m is defined as in Theorem 2.1. It is clear that $x_{0i} = 1$ for all i . Assuming $x_{1i} \rightarrow x_1, x_{2i} \rightarrow x_2, \dots, x_{mi} \rightarrow x_m$, we shall show that $x_{m+1, i} \rightarrow x_{m+1}$. As in the proof of Theorem 2.1, $x_{m+1, i+1} = 2\lambda_{m+1}x_{m+1, i} + f(x_{1i}, \dots, x_{mi})$ where f is a function depending only on the multiplication table of the algebra. Since f is a quadratic function $f(x_{1i}, \dots, x_{mi}) \rightarrow f(x_1, \dots, x_m)$. The result now reduces to the following theorem in analysis. Let (a_n) and (b_n) be two sequences satisfying

$a_{n+1} = \lambda a_n + b_n$ for some λ with $|\lambda| < 1$ and let $b_n \rightarrow b$; then (a_n) converges to the limit a determined by $a = \lambda a + b$.

Write $a_n = a + u_n$, $b_n = b + y_n$. We have $u_{n+1} = \lambda u_n + y_n$, where $|\lambda| < 1$ and $y_n \rightarrow 0$, and have to prove that $u_n \rightarrow 0$. Since

$$u_{n+1} = \lambda^n u_1 + \lambda^{n-1} y_1 + \lambda^{n-2} y_2 + \dots + y_n,$$

if we define $z_n = \lambda^{n-1} y_1 + \lambda^{n-2} y_2 + \dots + y_n$, it will suffice to show that $z_n \rightarrow 0$. In this form the result is contained in Knopp ((10), p. 76, 9(b)).

The theorem may be extended to cases where several of the λ 's are equal to $\frac{1}{2}$. If λ_1 only = $\frac{1}{2}$, then the theorem is still valid if $x_{001} = 0$; otherwise the sequence x_{1i} is an arithmetic progression which certainly does not converge.

3. Mutation Algebras

If the simplest case is considered with mutation rates r and s , the gametic algebra has basis D and R and the multiplication table:

$$D^2 = (1-r)D + rR, \quad R^2 = sD + (1-s)R,$$

$$DR = \frac{1}{2}\{(1-r)D + rR\} + \frac{1}{2}\{sD + (1-s)R\} = \frac{1}{2}(1-r+s)D + \frac{1}{2}(1-s+r)R.$$

It can be verified by direct computation that if a and b are chosen as a basis where $a = D$ and $b = D - R$, the table takes on the simplified form:

$$a^2 = a - rb, \quad ab = \frac{1}{2}(1-r-s)b, \quad b^2 = 0.$$

Thus the algebra is a special train algebra. (Note that the basis is independent of r and s .) Theorems 2.1 and 2.2 can therefore be applied. By theorem 2.1 there is a unique non-zero idempotent unless $r+s = 0$. By theorem 2.2 the plenary powers of an element of weight 1 approach this idempotent for $0 < r+s < 2$.

Although the algebra is defined for all real r and s , the corresponding genetic system has meaning only for $0 \leq r \leq 1$, $0 \leq s \leq 1$. Since it is known that in the case $r = s = 0$ (no mutations) stability is achieved in the second generation, this proves that the gene distribution approaches a stable distribution unless possibly when $r = s = 1$. This case is clearly unstable since $D^2 = R$ and $R^2 = D$. More generally, the plenary powers of $\alpha D + \beta R$, where $\alpha + \beta = 1$, oscillate between $\alpha D + \beta R$ and $\beta D + \alpha R$, only the distribution $\frac{1}{2}D + \frac{1}{2}R$ being one of unstable equilibrium.

The zygotic algebra has the table

$$D^2 = (1-r)^2 D + 2(1-r)rH + r^2 R, \quad R^2 = s^2 D + 2(1-s)sH + (1-s)^2 R,$$

$$DH = \frac{1}{2}(1-r)(1-r+s)D + \frac{1}{2}[1 + (1-2r)(r-s)]H + \frac{1}{2}r(1+r-s)R,$$

$$RH = \frac{1}{2}s(1-r+s)D + \frac{1}{2}[1 + (1-2s)(s-r)]H + \frac{1}{2}(1-s)(1+r-s)R,$$

$$DR = (1-r)sD + (1-r-s+2rs)H + (1-s)rR,$$

$$H^2 = \frac{1}{4}(1-r+s)^2 D + \frac{1}{2}(1-r+s)(1+r-s)H + \frac{1}{4}(1+r-s)^2 R.$$

If a , b , and c are chosen as a basis where $a = D$, $b = D - H$, and

$c = D - 2H + R$ we obtain:

$$\begin{aligned} a^2 &= a - 2rb + r^2c, & ab &= \frac{1}{2}(1-r-s)(b-rc), \\ b^2 &= \frac{1}{4}(1-r-s)^2c, & ac &= bc = c^2 = 0. \end{aligned}$$

Thus this is also a special train algebra.

4. Sex Linkage

We shall use the subscripts 1 and 2 for the haploid (usually male) and diploid sex respectively. The basis elements, representing genotypes, will thus be called D_1, R_1, D_2, H_2, R_2 . Since members of the same sex do not produce offspring it is convenient to define the product of any two basis elements with the same subscript as zero. The rest of the table is:

$$\begin{aligned} D_1D_2 &= \frac{1}{2}(D_1 + D_2), & D_1R_2 &= \frac{1}{2}(R_1 + H_2), & D_1H_2 &= \frac{1}{4}(D_1 + R_1 + D_2 + H_2), \\ R_1R_2 &= \frac{1}{2}(R_1 + R_2), & R_1D_2 &= \frac{1}{2}(D_1 + H_2), & R_1H_2 &= \frac{1}{4}(D_1 + R_1 + H_2 + R_2). \end{aligned}$$

The algebra is certainly not a special train algebra. In fact, the existence of a basis consisting entirely of nilpotent elements makes it clear that no non-trivial homomorphism into the reals exists (i.e. the algebra is not baric in the terminology of (1)).

A simplification can be effected by using as a basis $D_1, a = D_1 + D_2, b = D_1 - R_1, c = D_2 - R_2, d = D_2 - 2H_2 + R_2$. The table then becomes

$$\begin{aligned} D_1^2 &= 0, & aD_1 &= \frac{1}{2}a, & bD_1 &= 0, & cD_1 &= \frac{1}{2}b + \frac{1}{4}c + \frac{1}{4}d, & dD_1 &= 0, \\ a^2 &= a, & ab &= \frac{1}{4}c + \frac{1}{4}d, & ac &= \frac{1}{2}b + \frac{1}{4}c + \frac{1}{4}d, & ad &= 0, & b^2 &= 0, & bc &= \frac{1}{2}d, \\ bd &= 0, & c^2 &= 0, & cd &= 0, & d^2 &= 0. \end{aligned}$$

It follows immediately from the table that the subspace $\{a, b, c, d\}$ is an ideal containing the product of any two elements in the original ring. This ideal is the set of all elements for which the sums of the coefficients of the original basis elements of subscripts one and two are equal. (The genetic significance of the ideal is that it contains all the elements which represent populations.) Furthermore, the ideal is a special train algebra. This can be more easily seen if $b + c$ is used as a basis element instead of c . It is interesting to note that the powers of the ideal $\{b, c, d\}$, $(\{b, c, d\}^2 = d, \{b, c, d\}^3 = 0)$ are ideals even in the original ring.

The ideas in § 2 may be used to show that the elements of the form $a + yb + yc + (y^2 + y)d$ represent all the non-zero idempotents and that the plenary powers of an element of weight one approach an idempotent. The elements A which are non-zero and satisfy $(A^2)^2 = A^2$ are precisely those of the form $a + yb + yc + xd$ where x and y are arbitrary.

The genetic significance of these results can be made clear by expressing them in terms of the original basis. By letting $y = -\beta$ and $\alpha = 1 - \beta$ we find that the idempotents are exactly all those elements $\alpha D_1 + \beta R_1 + \alpha^2 D_2 + 2\alpha\beta H_2 + \beta^2 R_2$ with $\alpha + \beta = 1$.

It clearly follows that the elements which satisfy $(A^2)^2 = A^2$ are precisely those of the form

$$\alpha D_1 + \beta R_1 + (\alpha^2 + \gamma)D_2 + (2\alpha\beta - 2\gamma)H_2 + (\beta^2 + \gamma)R_2 \text{ where } \gamma = x - (y^2 + y).$$

A more intuitive feeling for this result can be grasped if it is noted that $\gamma(D_2 - 2H_2 + R_2)$ is the most general element which can be added to a genotype distribution which will not change the distribution of gametes produced. For a more classical approach see (9).

5. Sex Linkage with Mutations

Using the same style as before, the table is:

$$D_1 D_2 = \frac{1}{2}[(1-r)D_1 + rR_1 + (1-r)^2 D_2 + 2r(1-r)H_2 + r^2 R_2],$$

$$R_1 R_2 = \frac{1}{2}[sD_1 + (1-s)R_1 + s^2 D_2 + 2s(1-s)H_2 + (1-s)^2 R_2],$$

$$R_1 D_2 = \frac{1}{2}[(1-r)D_1 + rR_1 + s(1-r)D_2 + \{rs + (1-r)(1-s)\}H_2 + r(1-s)R_2],$$

$$D_1 R_2 = \frac{1}{2}[sD_1 + (1-s)R_1 + s(1-r)D_2 + \{rs + (1-r)(1-s)\}H_2 + r(1-s)R_2],$$

$$D_1 H_2 = \frac{1}{4}[(1-r+s)D_1 + (1+r-s)R_1 + (1-r)(1-r+s)D_2 + \{1 + (1-2r)(r-s)\}H_2 + r(1+r-s)R_2],$$

$$R_1 H_2 = \frac{1}{4}[(1-r+s)D_1 + (1+r-s)R_1 + s(1-r+s)D_2 + \{1 + (1-2s)(s-r)\}H_2 + (1-s)(1+r-s)R_2].$$

Again the table is simplified by using D_1, a, b, c, d as a basis.

$$D_1^2 = 0, aD_1 = \frac{1}{2}a - \frac{1}{2}rb - \frac{1}{2}rc - \frac{1}{2}r(1-r)d, bD_1 = 0,$$

$$cD_1 = \frac{1}{2}(1-r-s)b + \frac{1}{4}(1-r-s)c + \frac{1}{4}(1-2r)(1-r-s)d, dD_1 = 0,$$

$$a^2 = a - rb - rc - r(1-r)d, ab = \frac{1}{4}(1-r-s)c + \frac{1}{4}(1-2r)(1-r-s)d,$$

$$ac = \frac{1}{2}(1-r-s)b + \frac{1}{4}(1-r-s)c + \frac{1}{4}(1-2r)(1-r-s)d, ad = 0$$

$$b^2 = 0, bc = \frac{1}{2}(1-r-s)^2 d, bd = 0, c^2 = 0, cd = 0, d^2 = 0.$$

Again $\{a, b, c, d\}$ is an ideal containing the product of any two elements in the ring. Furthermore, the ideal is a special train algebra, containing a unique non-zero idempotent unless $r + s = 0$. As in the case of no sex linkage, the genotype distributions approach a stable distribution unless $r = s = 1$.

6. Polyploidy

We shall first consider the gametic n -ploidy algebra. The types will be denoted by D_n, D_{n-1}, \dots, D_0 where D_i stands for the type which has i dominant genes and $n-i$ recessive genes. The multiplication table is:

$$D_a D_b = \binom{2n}{n}^{-1} \sum_{i=0}^n \binom{a+b}{i} \binom{2n-a-b}{n-i} D_i$$

where $\binom{n}{r}$ stands for the usual binomial coefficient. (It is convenient to use

the symbol $\binom{n}{r}$ even when $r < 0$ or $r > n$ in which case we define $\binom{n}{r} = 0$. It is then easily seen that the identity $\binom{n}{r} + \binom{n}{r-1} = \binom{n+1}{r}$ is valid in general.)

Lemma 6.1.

$$\sum_{r=0}^s (-1)^r \binom{s}{r} P_{2n-r} = \binom{n}{s} \binom{2n}{s}^{-1} \left[\sum_{r=0}^s (-1)^r \binom{s}{r} D_{n-r} \right] \text{ for } 0 \leq s \leq n$$

$$= 0 \text{ for } n < s \leq 2n,$$

where P_t is defined as $\binom{2n}{n}^{-1} \sum_{i=0}^n \binom{t}{i} \binom{2n-t}{n-i} D_i$ for $0 \leq t \leq 2n$.

($P_t = D_a D_b$ for any a, b satisfying $a + b = t$.)

Proof. By definition

$$\sum_{r=0}^s (-1)^r \binom{s}{r} P_{2n-r} = \binom{2n}{n}^{-1} \sum_{r=0}^s \sum_{i=0}^n (-1)^r \binom{s}{r} \binom{2n-r}{i} \binom{r}{n-i} D_i.$$

Using the fact that

$$\binom{s}{r} \binom{r}{p} = \binom{s}{p} \binom{s-p}{r-p},$$

the latter expression becomes

$$\binom{2n}{n}^{-1} \sum_{r=0}^s \sum_{i=0}^n (-1)^r \binom{2n-r}{i} \binom{s}{n-i} \binom{s-(n-i)}{r-(n-i)} D_i. \dots\dots\dots(1)$$

From the identity $\binom{n}{i} - \binom{n-1}{i} = \binom{n-1}{i-1}$ it may be obtained by induction on s that $\sum_{t=0}^s (-1)^t \binom{s}{t} \binom{n-t}{i} = \binom{n-s}{i-s}$. Replacing n by $n+1$ and s by $s-(n-i)$ this becomes $\sum_{t=0}^{s-(n-i)} (-1)^t \binom{s-(n-i)}{t} \binom{n+i-t}{i} = \binom{2n-s}{n-s}$. We now let $t = r-(n-i)$ and sum over r . The relationship becomes:

$$\sum_{r=n-i}^s (-1)^{r-(n-i)} \binom{s-(n-i)}{r-(n-i)} \binom{2n-r}{i} = \binom{2n-s}{n-s}.$$

Since $\binom{s-(n-i)}{r-(n-i)} = 0$ for $r < n-i$ in view of our convention, the summation may be considered as going from 0 to s . Expression (1) now simplifies to

$$\binom{2n}{n}^{-1} \sum_{i=0}^n (-1)^{n-i} \binom{s}{n-i} \binom{2n-s}{n-s} D_i = \binom{2n}{n}^{-1} \sum_{i=0}^n (-1)^i \binom{s}{i} \binom{2n-s}{n-s} D_{n-i}.$$

Since $\binom{2n}{n}\binom{n}{s} = \binom{2n}{s}\binom{2n-s}{n-s}$ this becomes

$$\binom{2n}{s}^{-1} \binom{n}{s} \sum_{i=0}^n (-1)^i \binom{s}{i} D_{n-i}.$$

(Note the convenience of using the symbol $\binom{n}{r}$ for all integers n and r . Thus the proof is valid even when $s > n$ in which case the last line reduces to 0.)

Lemma 6.2. If c_s is defined as $\sum_{i=0}^s (-1)^i \binom{s}{i} D_{n-i}$ for $0 \leq s \leq n$, then

$$\begin{aligned} c_i c_j &= \binom{2n}{i+j}^{-1} \binom{n}{i+j} c_{i+j} \text{ for } i+j \leq n \\ &= 0 \text{ for } i+j > n. \end{aligned}$$

Proof. By comparison with the binomial theorem it is easily seen that

$$\left[\sum_{i=0}^s (-1)^i \binom{s}{i} D_{n-i} \right] \left[\sum_{i=0}^t (-1)^i \binom{t}{i} D_{n-i} \right] = \sum_{i=0}^{s+t} (-1)^i \binom{s+t}{i} P_{2n-i}$$

where P_t is defined as in Lemma 6.1. The result now follows from Lemma 6.1.

As an immediate consequence of Lemma 6.2 we obtain the following:

Theorem 6.1. *The gametic n -ploidy algebra is a special train algebra with train roots $\binom{2n}{i}^{-1} \binom{n}{i}$ for $0 \leq i \leq n$. Since $\lambda_0 = 1$, $\lambda_1 = \frac{1}{2}$ and all other λ 's are less than $\frac{1}{2}$, the remarks in § 2 may be applied to this case.*

When $n = 2$ this reduces to the well-known case of the tetraploidy gametic algebra with table

$$c_0^2 = c_0, c_0 c_1 = \frac{1}{2} c_1, c_0 c_2 = c_1^2 = \frac{1}{8} c_2, c_1 c_2 = c_2^2 = 0 \text{ (1)}.$$

By applying the technique of duplication (5) we can now consider the zygotic algebra. We shall identify $c_i c_j$ with fixed $i+j$. This corresponds to identifying all $D_a D_b$ with fixed $a+b$ which in turn corresponds genetically to disregarding the nature of the contributions of each gamete to the zygote. It will be convenient to use the symbol d_i to stand for any product $c_a c_b$ with $a+b = i$. d_0, d_1, \dots, d_{2n} is then a basis of the algebra.

Theorem 6.2. *The zygotic n -ploidy algebra is a special train algebra with basis elements $d_i, 0 \leq i \leq 2n$ and multiplication table:*

$$d_i d_j = \binom{2n}{i}^{-1} \binom{2n}{j}^{-1} \binom{n}{i} \binom{n}{j} d_{i+j} \text{ for } i, j \leq n,$$

$$d_i d_j = 0 \text{ otherwise.}$$

Proof. The table follows immediately from the process of duplication

and implies the special train property. The remarks in § 2 remain valid in the zygotic case.

When $n = 2$ the table becomes $d_0^2 = d_0$, $d_0d_1 = \frac{1}{2}d_1$, $d_1^2 = \frac{1}{2}d_2$, $d_0d_2 = \frac{1}{6}d_2$, $d_1d_2 = \frac{1}{12}d_3$, $d_2^2 = \frac{1}{36}d_4$ with all other products zero.

Lemma 6.3.

$$\sum_{i=0}^n x_i D_{n-i} = \sum_{r=0}^n (-1)^r \left[\sum_{s=r}^n \binom{s}{r} x_s \right] c_r.$$

Proof. By using the definition $c_r = \sum_{i=0}^r (-1)^i \binom{r}{i} D_{n-i}$ we obtain for the right-hand side

$$\sum_{r=0}^n \sum_{s=r}^n \sum_{i=0}^r (-1)^i (-1)^r \binom{s}{r} \binom{r}{i} x_s D_{n-i}.$$

The coefficient of $x_s D_{n-i}$ is

$$(-1)^i \sum_{r=i}^s (-1)^r \binom{s}{r} \binom{r}{i}.$$

Now

$$\sum_{r=i}^s (-1)^r \binom{s}{r} \binom{r}{i} = \sum_{r=i}^s (-1)^r \binom{s}{i} \binom{s-i}{r-i} = \binom{s}{i} \sum_{r=i}^s (-1)^r \binom{s-i}{r-i} = (-1)^i \delta_{is}$$

since the expression within the summation sign is simply an expansion of $(-1)^i (1-1)^{s-i}$. Thus the coefficient is δ_{is} . Hence the coefficient of D_{n-i} is x_i . This completes the proof.

This lemma gives an explicit formula for change of basis. In the special case where $x_i = 1$ and all other x 's are 0 the lemma becomes

$$D_{n-i} = \sum_{r=0}^i (-1)^r \binom{i}{r} c_r.$$

In other words the relationship expressing the D 's in terms of the c 's is the same as the relationship expressing the c 's in terms of the D 's. For example in the simplest case where $c_0 = D_2$, $c_1 = D_2 - D_1$ and $c_2 = D_2 - 2D_1 + D_0$ we have $D_2 = c_0$, $D_1 = c_0 - c_1$, and $D_0 = c_0 - 2c_1 + c_2$.

The special case where $x_i \geq 0$ for all i and $\sum_{i=0}^n x_i = 1$ corresponds to a distribution of gametic types. $\left(\sum_{i=0}^n x_i D_{n-i} \right)$ corresponds to the distribution in which x_i is the fraction of types which have exactly i recessive genes.) If a package of i genes is chosen at random from one individual of the population, the probability of all genes being recessive is

$$\binom{n}{i}^{-1} \sum_{s=i}^n \binom{s}{i} x_s.$$

If the probability is denoted by y_i , then it follows from Lemma 6.3 that

$$\sum_{i=0}^n x_i D_{n-i} = \sum_{i=0}^n (-1)^i \binom{n}{i} y_i c_i.$$

Theorem 6.3. *A necessary and sufficient condition for a non-zero element to be idempotent is that $y_i = y_1^i$ for all i satisfying $0 \leq i \leq n$. Otherwise no plenary power of the element is idempotent.*

Proof. We first prove the sufficiency of the first statement.

Since $\left[\sum_{i=0}^n (-1)^i \binom{n}{i} y_i c_i \right]^2 = \sum_{i=0}^n \sum_{j=0}^n (-1)^{i+j} \binom{n}{i} \binom{n}{j} y_i y_j c_i c_j$

and

$$c_i c_j = \binom{2n}{i+j}^{-1} \binom{n}{i+j} c_{i+j},$$

the coefficient of c_k in the product becomes

$$\sum_{i=0}^k (-1)^k \binom{n}{i} \binom{n}{k-i} y_i y_{k-i} \binom{2n}{k}^{-1} \binom{n}{k}.$$

The use of the fact that

$$\sum_{i=0}^k \binom{n}{i} \binom{n}{k-i} = \binom{2n}{k}$$

completes the proof. (The latter is verified by comparing the coefficient of x^k in $(1+x)^{2n}$ and in $(1+x)^n(1+x)^n$.) The necessity follows from the fact that the value of y_1 determines the idempotent uniquely.

The last part follows from a careful examination of the proof of Theorem 2.2. The basic idea is that if, in a recursion relation of the form $A_{n+1} = \lambda A_n + B$ for $\lambda \neq 0$, $A_1 \neq A_2$, then $A_n \neq A_{n+1}$ for all n .

If the idempotent $\sum_{i=0}^n (-1)^i \binom{n}{i} y_1^i c_i$ is expressed in terms of D 's, the coefficient of D_{n-j} becomes $\binom{n}{j} y_1^j (1-y_1)^{n-j}$. Thus the corresponding genotype distribution is a binomial distribution as seems intuitively reasonable.

Analogous results apply to the zygotic algebra. Let us use the symbol D_i to stand for any product $D_a D_b$ with $a+b=i$. No confusion should arise if the zygotic and gametic algebras are not discussed simultaneously. $D_{2n}, D_{2n-1}, \dots, D_0$ is a basis of the algebra where D_i stands for the type which has i dominant genes and $2n-i$ recessive genes. It is easily seen that the D 's and the d 's bear the same relation to each other as the D 's and the c 's of the gametic algebra provided that n is replaced by $2n$. Thus we have

$$d_s = \sum_{i=0}^s (-1)^i \binom{s}{i} D_{2n-i}$$

E.M.S.—D

as well as the analogue of Lemma 6.3. Furthermore we have:

Theorem 6.4. *A necessary and sufficient condition for a non-zero element to be an idempotent is that $y_i = y_1^i$ for all i satisfying $0 \leq i \leq 2n$. If $y_i = y_1^i$ for all i satisfying $0 \leq i \leq n$ then the element satisfies $(A^2)^2 = A^2$. Otherwise no plenary power of the element is idempotent.*

Proof. As in the proof of Theorem 6.3 we consider

$$\left[\sum_{i=0}^{2n} (-1)^i \binom{2n}{i} y_i d_i \right]^2 = \sum_{i=0}^{2n} \sum_{j=0}^{2n} (-1)^{i+j} \binom{2n}{i} \binom{2n}{j} y_i y_j d_i d_j.$$

Using the formula

$$d_i d_j = \binom{2n}{i}^{-1} \binom{2n}{j}^{-1} \binom{n}{i} \binom{n}{j} d_{i+j},$$

the coefficient of d_k in the product becomes

$$\begin{aligned} \sum_{i=0}^k (-1)^k \binom{2n}{i} \binom{2n}{k-i} y_i y_{k-i} \binom{2n}{i}^{-1} \binom{2n}{k-i}^{-1} \binom{n}{i} \binom{n}{k-i} \\ = \sum_{i=0}^k (-1)^k \binom{n}{i} \binom{n}{k-i} y_1^k. \end{aligned}$$

The rest of the proof is similar to the gametic case except for the consideration of the property $(A^2)^2 = A^2$.

Since the coefficient of d_i in a product does not depend on the coefficient of d_j for $j > i$, the condition $y_i = y_1^i$ for $0 \leq i \leq n$ will at least guarantee that the coefficient of d_i for $i \leq n$ will be the same for A and A^2 . Since $d_i d_j = 0$ for $i > n$ it follows that A^2 depends only on the coefficients of d_i for $i \leq n$ in A . Hence $A^2 = (A^2)^2$.

As before it may be noted that

$$\sum_{i=0}^{2n} (-1)^i \binom{2n}{i} y_1^i d_i = \sum_{j=0}^{2n} \binom{2n}{j} y_1^j (1 - y_1)^{2n-j} D_{2n-j}.$$

For a more classical discussion of polyploidy see (7) and (8).

7. Polyploidy with Mutations

Instead of discussing the product directly it will be convenient to regard the product as being obtained from the product with no mutations by a linear mapping which we call the mutation mapping. To be more explicit $D_a D_b = T(D_a \times D_b)$ where $D_a \times D_b$ is the product as defined in § 6 and T is the linear mapping satisfying

$$T(D_i) = \left[\sum_{u+v=i} \binom{i}{u} \binom{n-i}{v} (1-r)^u r^{i-u} (1-s)^{n-i-v} s^v \right] D_i.$$

The mapping is obtained by elementary reasoning in probability. Since the method applies equally well to the gametic and zygotic algebra we simply use n .

(Our previous notation would require $2n$ for the zygotic case.) Note that the coefficient of D_i is the coefficient of x^i in the expansion of

$$[(1-r)x+r]^i[sx+(1-s)]^{n-i}.$$

This formulation will be useful in the sequel. Let

$$[(1-r)x+r]^i[sx+(1-s)]^{n-i} = \sum_{t=0}^n a_{it}x^t.$$

Writing $T(D_i) = \sum_{u=0}^n a_{iu}D_u$ it follows from Lemma 6.3 that

$$T(D_i) = \sum_{v=0}^n (-1)^v \left[\sum_{u=0}^{n-v} \binom{n-v}{v} a_{iu}c_v \right].$$

Now consider the expansion $[(1-r)x+r(x-1)]^i[sx+(1-s)(x-1)]^{n-i}$. This is $\sum_{u=0}^n a_{iu}x^u(x-1)^{n-u}$. The coefficient of x^{n-v} is $(-1)^v \sum_{u=0}^{n-v} \binom{n-u}{v} a_{iu}$ which is the coefficient of c_v in $T(D_i)$. Thus by simplifying the above expression we find that the coefficient of c_v in $T(D_i)$ is the coefficient of x^{n-v} in $(x-r)^i [x-(1-s)]^{n-i}$. We call this coefficient b_i .

It follows that the coefficient of c_v in $T(c_w)$ is $\sum_{i=0}^w (-1)^i \binom{w}{i} b_{n-i}$. Now consider $[x-r]^{n-w}[(x-r)-\{x-(1-s)\}]^w$. Expanding the second term by the binomial theorem we may write this as $\sum_{i=0}^w \binom{w}{i} (x-r)^{n-i} [x-(1-s)]^i$. Thus the coefficient of x^{n-v} is $\sum_{i=0}^w (-1)^i \binom{w}{i} b_{n-i}$. This shows that the coefficient of c_v in $T(c_w)$ is the coefficient of x^{n-v} in $[x-r]^{n-w}[(x-r)-\{x-(1-s)\}]^w$ or $(x-r)^{n-w}(1-r-s)^w$. This finally proves the important result:

$$T(c_w) = (1-r-s)^w \sum_{v=w}^n (-1)^{v-w} \binom{n-w}{v-w} r^{v-w} c_v,$$

i.e.

$$T(c_w) = (1-r-s)^w \left[c_w - \binom{n-w}{1} r c_{w+1} + \binom{n-w}{2} r^2 c_{w+2} \dots \right].$$

Now that we have found T explicitly in terms of the c 's we can write down the multiplication table.

Theorem 7.1. *The multiplication table of the n -ploidy gametic mutation algebra is*

$$c_i c_j = \binom{2n}{i+j}^{-1} \binom{n}{i+j} (1-r-s)^{i+j} \left[c_{i+j} - \binom{n-i-j}{1} r c_{i+j+1} + \binom{n-i-j}{2} r^2 c_{i+j+2} + \dots \right] \text{ for } i+j \leq n,$$

= 0 otherwise.

E.M.S.—D 2

A similar result is obtained for the zygotic algebra.

Theorem 7.2. *The multiplication table of the n-ploidy zygotic mutation algebra is*

$$d_i d_j = \binom{2n}{i}^{-1} \binom{2n}{j}^{-1} \binom{n}{i} \binom{n}{j} (1-r-s)^{i+j}$$

$$\left[d_{i+j} - \binom{n-i-j}{1} r d_{i+j+1} + \binom{n-i-j}{2} r^2 d_{i+j+2} - \dots \right] \text{ for } i, j \leq n,$$

= 0 otherwise.

The tables in § 3 may be seen to be a special case of this. When $n = 2$ the table in the zygotic case becomes

$$d_0^2 = d_0 - 4rd_1 + 6r^2d_2 - 4r^3d_3 + r^4d_4,$$

$$d_0d_1 = \frac{1}{2}(1-r-s)(d_1 - 3rd_2 + 3r^2d_3 - r^3d_4), \quad d_1^2 = \frac{1}{4}(1-r-s)^2(d_2 - 2rd_3 + r^2d_4),$$

$$d_0d_2 = \frac{1}{6}(1-r-s)^2(d_2 - 2rd_3 + r^2d_4), \quad d_1d_2 = \frac{1}{12}(1-r-s)^3(d_3 - rd_4),$$

$$d^2 = \frac{1}{36}(1-r-s)^4d_4 \text{ with all other products zero.}$$

Corollary 1. The n -ploidy gametic and zygotic mutation algebras are special train algebras.

Since the train roots are $\binom{2n}{i}^{-1} \binom{n}{i} (1-r-s)^i$, the results in § 2 apply.

A unique non-zero idempotent exists if $0 < r+s \leq 2$. Furthermore, the genotype distributions approach a stable distribution unless $r = s = 1$.

The case $r+s = 1$ is incredibly simple. In fact, if $e_0, c_1, c_2, \dots, c_n$ is chosen as a basis where $e_0 = c_0 - \binom{n}{1}rc_1 + \binom{n}{2}r^2c_2 + \dots + (-1)^n r^n c_n$, we obtain for the multiplication table that $e_0^2 = e_0$ and that all other products of basis elements are zero! From this follows immediately

Corollary 2. If $r+s = 1$, stability is obtained in one generation.

It is easily seen in this case that in terms of the D 's the idempotent has the form $\sum_{i=0}^n \binom{n}{i} s^i r^{n-i} D_i$. Furthermore, these results are valid in the zygotic case as well.

In conclusion, the author wishes to thank the referee for many invaluable suggestions.

REFERENCES

(1) I. M. H. ETHERINGTON, Genetic algebras, *Proc. Roy. Soc. Edin.*, **59** (1939), 242-258.
 (2) I. M. H. ETHERINGTON, Commutative train algebras of ranks 2 and 3, *J. London Math. Soc.*, **15** (1940), 136-149; **20** (1945), 238.
 (3) I. M. H. ETHERINGTON, Special train algebras, *Quart. J. Math.* (Oxford), **12** (1941), 1-8.

- (4) I. M. H. ETHERINGTON, Non-associative algebra and the symbolism of genetics, *Proc. Roy. Soc. Edin. (B)*, **61** (1941), 24-42.
- (5) I. M. H. ETHERINGTON, Duplication of linear algebras, *Proc. Edin. Math. Soc. (2)*, **6** (1941), 222-230.
- (6) I. M. H. ETHERINGTON, Non-commutative train algebras of ranks 2 and 3, *Proc. London Math. Soc. (2)*, **52** (1951), 241-252.
- (7) H. GEIRINGER, On some mathematical problems arising in the development of Mendelian genetics, *J. Amer. Stat. Assoc.*, **44** (1949), 526-547.
- (8) J. B. S. HALDANE, Theoretical genetics of autopolyploids, *J. Genetics*, **22** (1930), 359-372.
- (9) H. S. JENNINGS, The numerical results of diverse systems of breeding, *Genetics*, **1** (1916), 53-89.
- (10) K. KNOPP, *Theory and Application of Infinite Series* (Blackie), 1928.
- (11) R. RAFFIN, Axiomatisation des algèbres génétiques, *Acad. Roy. Belgique Bull. Cl. Sci. (5)*, **37** (1951), 359-366.
- (12) R. D. SCHAFER, Structure of genetic algebras, *Amer. J. Math.*, **71** (1949), 121-135.

MATHEMATICS DEPARTMENT
 RUTGERS UNIVERSITY
 NEW BRUNSWICK, N.J.
 U.S.A.