

GENETIC ALGEBRAS ASSOCIATED WITH POLYPLOIDY

by P. HOLGATE

(Received 3rd June 1964; revised MS. received 4th May 1965)

1. Introduction

The relationship between certain non-associative algebras and the deterministic theory of population genetics was first investigated by Etherington (3)-(8), who defined the concepts of baric, train and special train algebras. Gonsior (10) dealt with, among other topics, algebras corresponding to autopolyploidy, on the assumption that chromosome segregation operated.† In this paper I discuss algebras corresponding to more general systems of inheritance among polyploids, which have been discussed without using algebras by Haldane (11), Geiringer (9), Moran (13) and Seyffert (16). These algebras are special cases of what I have defined as segregation algebras, and mixtures of them. All the algebras corresponding to a fixed ploidy have a relationship which I have called special isotopy. An example shows that algebras arise in other genetic systems which are not isotopic to segregation algebras.

2. Segregation algebras

As far as possible I have followed the notation of (10). The symbol D_a , ($a = 0, \dots, n$) stands for a gamete containing a dominant and $n - a$ recessive genes. For convenience the D_a will also be used to denote the natural basis elements of the genetic algebra. Those elements with non-negative coefficients which add up to unity can be taken to represent populations, and multiplication represents the formation of a filial generation by random mating between the populations corresponding to the factors. Consequently, from genetic reasoning, the multiplication table for chromosome segregation is

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

while for chromatid segregation it is

$$D_a D_b = \binom{4n}{n}^{-1} \sum_{i=0}^n \binom{2a+2b}{i} \binom{4n-2a-2b}{n-i} D_i$$

(see e.g. 13, pp. 29, 30).

This suggests the definition of an n -ploid segregation algebra of degree s

† Added in proof: Gonsior has since published a further paper, dealing largely with polyploidy, in these Proceedings, vol. 14.

(n, s integers), which will be denoted by $A(n, s)$, as an algebra with basis D_a , ($a = 0, \dots, n$) and multiplication table

$$D_a D_b = \binom{2sn}{n}^{-1} \sum_{i=0}^n \binom{sa+sb}{i} \binom{2sn-sa-sb}{n-i} D_i. \tag{1}$$

For $s = 1$ and $s = 2$ the algebra has a biological interpretation in terms of chromosome and chromatid segregation respectively.

For all s , it is convenient to transform to a canonical basis defined by

$$c_a = \sum_{i=0}^a (-1)^i \binom{a}{i} D_i$$

which has the inverse transformation

$$D_a = \sum_{i=0}^a (-1)^i \binom{a}{i} c_i.$$

In terms of the weight function of the baric algebra, c_0 has unit weight and the remaining c_i have zero weight (4, 5). In terms of the new basis, the multiplication table becomes

$$\begin{aligned} c_a c_b &= \left\{ \sum_{i=0}^a (-1)^i \binom{a}{i} D_i \right\} \left\{ \sum_{j=0}^b (-1)^j \binom{b}{j} D_j \right\} \\ &= \sum_{i=0}^a \sum_{j=0}^b (-1)^{i+j} \binom{a}{i} \binom{b}{j} D_a D_b. \end{aligned}$$

Writing $a + b = m$, substituting from (1), and using formulae of the type

$$\sum_{i+j=\text{const.}} \binom{a}{i} \binom{b}{j} = \binom{a+b}{i+j} \tag{2}$$

this becomes

$$\begin{aligned} &\binom{2sn}{n}^{-1} \sum_{k=0}^m (-1)^k \binom{m}{k} \sum_{u=0}^n \binom{sk}{u} \binom{2sn-sk}{n-u} \sum_{t=0}^u (-1)^t \binom{u}{t} c_t \\ &= \binom{2sn}{n}^{-1} \sum_{t=0}^n (-1)^t c_t \sum_{k=0}^m (-1)^k \binom{m}{k} \sum_{u=t}^n \binom{sk}{u} \binom{2sn-sk}{n-u} \binom{u}{t}. \end{aligned}$$

On applying the formula

$$\binom{sk}{u} \binom{u}{t} = \binom{sk}{t} \binom{sk-t}{u-t}$$

this reduces to

$$\begin{aligned} &\binom{2sn}{n}^{-1} \sum_{t=0}^n (-1)^t c_t \sum_{k=0}^m (-1)^k \binom{m}{k} \binom{sk}{t} \sum_{u=t}^n \binom{sk-t}{u-t} \binom{2sn-sk}{n-u} \\ &= \binom{2sn}{n}^{-1} \sum_{t=0}^n \binom{2sn-t}{n-t} (-1)^t c_t \sum_{k=0}^m (-1)^k \binom{m}{k} \binom{sk}{t} \end{aligned}$$

the last step following from a formula like (2). The coefficient of c_t contains a factor

$$S_{m,t} = \sum_{k=0}^m (-1)^k \binom{m}{k} \binom{sk}{t}.$$

The factorial power $(sk)^{(t)}$ can be written as a linear combination of $k^{(j)}$ for $j = 0, \dots, t$. However

$$\sum_{k=0}^m (-1)^k \binom{m}{k} k^{(j)} = 0 \quad (0 \leq j \leq m)$$

which implies $S_{m,t} = 0$ for $t < m$. In particular, the whole expression for $c_a c_b$ vanishes if $a + b = m > n$. These results can be expressed by saying that if the multiplication table is

$$c_a c_b = \sum_{t=0}^n \gamma_{abt} c_t$$

then γ_{abt} depends on a and b only through $a + b = m$ (and can be denoted by γ_{mt}), and furthermore is zero for $a + b = m > t$. Also, for $m = t$,

$$S_{m,m} = \frac{1}{m!} \sum_{k=0}^m (-1)^k \binom{m}{k} (ks)^{(m)}$$

and $(ks)^{(m)} = s^m k^{(m)} +$ terms involving factorial powers of k with indices less than m . Hence

$$\begin{aligned} S_{m,m} &= \frac{s^m}{m!} \sum_{k=0}^m (-1)^k \binom{m}{k} k^{(m)} \\ &= (-s)^m \end{aligned}$$

and the coefficient of c_m in $c_a c_b$ with $a + b = m$ is

$$\binom{2sn}{n}^{-1} \binom{2sn-m}{n-m} s^m = n^{(m)} s^m / (2sn)^{(m)}. \tag{4}$$

The facts that $\gamma_{mt} = 0$, $t < m$ and $\gamma_{m,m} \neq 0$ show, on comparison with the basis definitions, (6) and (10), that $A(n, s)$ is a special train algebra. It is obvious that every power of the ideal $\{c_1, \dots, c_n\}$ will have the form $\{c_k, \dots, c_n\}$ for some k and will consequently be itself an ideal. The train roots, (10) of $A(n, s)$ are given by (4) for $m = 0, \dots, n$. In particular $\lambda_0 = 1$, $\lambda_1 = \frac{1}{2}$ and $\lambda_2 = \frac{1}{2} \left(1 - \frac{1}{n}\right) \left(2 - \frac{1}{ns}\right)^{-1}$. λ_1 is not a principal train root, as defined in (4) but the remaining λ_i are.

3. Plenary sequences

The right principal powers of an element x in a non-associative algebra, or simply the principal powers in the commutative case, are defined by $x^1 = x$, $x^{n+1} = x^n x$. The plenary powers are defined by $x^{[1]} = x$, $x^{[n+1]} = x^{[n]} x^{[n]}$. A baric algebra is one which admits a non-trivial homomorphism into its

coefficient field, $x \rightarrow w(x)$. $w(x)$ is called the weight of x , and in the present case the coefficient field is the real numbers. A train algebra is a baric algebra in which the coefficients of the rank equation of the principal powers of each element x , insofar as they depend on x , do so only through $w(x)$. Equivalently, the rank equation of every element of unit weight has constant coefficients, and the roots of the corresponding scalar polynomial are the principal train roots of the algebra. It may happen that the sequence of plenary powers of each element of unit weight satisfies a recurrence equation with constant coefficients. In this case the plenary powers are said to form a train, the recurrence equation is called the plenary train equation, its degree the plenary rank, and the roots of the corresponding scalar polynomial equation the plenary train roots of the algebra.

In (5) Etherington showed that train algebras of ranks 1, 2 and 3 are necessarily special train algebras, and that their plenary train roots are simply related to their principal train roots. In (6) he illustrated the use of the method of annulling polynomials to obtain the plenary train equation of the algebra corresponding to three linked loci in haploid gametes. Reiersøl (14) proved that in algebras corresponding to n linked loci in haploids, plenary powers form a train, and his constructive proof shows how they can be computed by recursion with respect to n .

To investigate this problem for segregation algebras, let the operator E be defined by $E x^{[k]} = x^{[k+1]}$ and consider a typical element of $A(n, s)$, of unit weight

$$x = c_0 + u_1 c_1 + \dots + u_n c_n.$$

If

$$E^k x = x^{[k]} = v_0 + v_1 c_1 + \dots + v_n c_n$$

each v is a polynomial of degree k in u_1, \dots, u_n . However in view of the conditions on the constants in the multiplication table of $A(n, s)$ given below equation (3), v_j will only contain terms for which the sum of the indices of the u_i , counted according to their multiplicity, does not exceed j . That is, the coefficient of c_j will contain at most $\gamma(j, 2^k)$ distinct terms where $\gamma(j, q)$ is the number of compositions of j having at most q parts. Clearly $\gamma(j, q)$ does not increase with q for $q \geq j$. Now consider the operator

$$E^m + l_1 E^{m-1} + \dots + l_m$$

which may be thought of as acting on the coefficients in the vector representation of x . Its effect can be written as

$$\left(\sum_{j=0}^m l_j E^{m-j} \right) u_i = \sum_{j=0}^m l_j \sum q_s v_s.$$

From the remarks above, the number of distinct products of powers of the u 's, when the right-hand side is written out in terms of them, does not increase for $2^m > k$. Hence by taking m sufficiently large, a set of $l_j (j = 1, \dots, m)$ can be found to make the right-hand side zero. The operator on the left is then

an annulling operator for u_i . Annulling polynomials can be found in this way for all the $u_i (i = 1, \dots, n)$. For segregation algebras $(E-1)$ is always necessary to annul u_1 , and this also annuls the component c_0 . Hence the L.C.M. of the operators found in this way annuls every element of the algebra, thus proving that in the segregation algebra $A(n, s)$, for all n, s , the plenary powers form a train.†

4. Particular examples

Gonshor has shown in (10) that for chromosome segregation the multiplication table for the canonical basis has the particularly simple form

$$c_a c_b = \binom{2n}{n}^{-1} \binom{2n-m}{n-m} c_m.$$

Since tetraploidy and hexaploidy occur frequently in nature, the multiplication tables for pure chromatid segregation are recorded here. For tetraploidy,

$$c_0^2 = c_0, c_0 c_1 = \frac{1}{2}c_1 - \frac{1}{8}c_2, c_0 c_2 = c_1^2 = \frac{1}{7}c_2$$

and the remaining products zero. For hexaploidy,

$$c_0^2 = c_0, c_0 c_1 = \frac{1}{2}c_1 - \frac{1}{22}c_2, c_0 c_2 = c_1^2 = \frac{2}{11}c_2 + \frac{1}{55}c_3$$

$$c_0 c_3 = c_1 c_2 = \frac{2}{55}c_3$$

and remaining products zero.

For the tetraploid algebras, of rank 3, the plenary rank equations may be obtained from the results in (5). They are, for chromosome and chromatid segregation respectively,

$$x[x-1][x-\frac{1}{3}] = 0, x[x-1][x-\frac{2}{7}] = 0.$$

For hexaploidy with chromosome segregation, let a typical element of unit weight be

$$x = c_0 + u_1 c_1 + u_2 c_2 + u_3 c_3.$$

Then

$$Ex = x^{[2]} = c_0 + u_1 c_1 + (\frac{2}{5}u_2 + \frac{1}{5}u_1^2)c_2 + (\frac{1}{10}u_3 + \frac{1}{10}u_1 u_2)c_3,$$

or

$$Eu_1 = u_1, Eu_2 = \frac{2}{5}u_2 + \frac{1}{5}u_1^2, Eu_3 = \frac{1}{10}u_3 + \frac{1}{10}u_1 u_2.$$

It is clear that $(E-1)$ annuls u_1 . Recursive computation leads to

$$E^2 u_1 = u_1, E^2 u_2 = \frac{4}{25}u_2 + \frac{7}{25}u_1^2, E^2 u_3 = \frac{1}{100}u_3 + \frac{1}{20}u_1 u_2 + \frac{1}{50}u_1^3$$

and hence

$$(E^2 + l_1 E + l_2)u_2 = (\frac{4}{25} + \frac{2}{5}l_1 + l_2)u_2 + (\frac{7}{25} + \frac{1}{5}l_1)u_1^2.$$

From this it can be seen that the values $l_1 = -\frac{7}{5}, l_2 = \frac{2}{5}$ make the right-hand side zero, hence $E^2 - \frac{7}{5}E + \frac{2}{5} = (E-1)(E-\frac{2}{5})$ annuls u_2 . For economy in

† Added in proof: I have since obtained a more general result on these lines, to be published shortly in the *Journal of the London Math. Soc.*

degree consider now the operator $(E-l')(E-1)(E-\frac{2}{5})$. Further recursion leads to

$$(E-l')(E-1)(E-\frac{2}{5})u_3 = (\frac{27}{1000} - \frac{27}{100}l')u_3 - (\frac{9}{1000} - \frac{9}{100}l')u_1u_2 + (\frac{1}{500} - \frac{1}{50}l')u_1^3$$

and hence for $l' = \frac{1}{10}$, u_3 is annulled. The plenary rank equation is therefore

$$x[x-1][x-\frac{2}{5}][x-\frac{1}{10}] = 0.$$

For chromatid segregation, the computation needed is rather more protracted, and the plenary rank equation is

$$x[x-1][x-\frac{4}{11}][x-\frac{4}{55}][x-\frac{16}{605}] = 0.$$

In all the above cases, the plenary train roots are multiples of the principal train roots and products of pairs of them.

5. Duplication

Zygotic algebras are derived from gametic algebras by the process of commutative duplication (7). The ordinary commutative duplicates of segregation algebras are not special train algebras. For instance on duplicating $A(2, 1)$ and taking as basis elements d_{ab} the isomorphs of $c_a c_b$, the multiplication table is

	d_{00}	d_{01}	d_{02}	d_{11}
d_{00}	d_{00}	$\frac{1}{2}d_{01}$	$\frac{1}{6}d_{02}$	$\frac{1}{6}d_{02}$
d_{01}		$\frac{1}{4}d_{11}$	$\frac{1}{12}d_{12}$	$\frac{1}{12}d_{12}$
d_{02}			$\frac{1}{36}d_{22}$	$\frac{1}{36}d_{22}$
d_{11}				$\frac{1}{36}d_{22}$

and remaining products zero. The square of the nil ideal is $\{d_{11}, d_{12}, d_{22}\}$ but it is not an ideal since e.g. $d_{00}d_{11} = \frac{1}{6}d_{02}$. However, Gonshor (10), adopts the device of identifying all elements d_{ab} with $a+b = m$ and the algebra derived in this way, which could be called the reduced commutative duplicate, is a special train algebra. Since the multiplication table of the reduced commutative duplicate of any segregation algebra has the form

$$d_m d_n = k d_{m+n} + \text{terms in } d_j, j > m+n$$

with $k \neq 0$, it is a special train algebra. In passing from the gametic to the zygotic algebra, the reduced duplicate is clearly what is required in the polyploid case. This does not hold true in passing from the zygotic to the copular algebra, which cannot be represented by a special train algebra.

6. Mixture of algebras

Suppose two algebras A^0 and A are defined on the same vector space, with products formed according to "circle multiplication" $a \circ b$ and "dot multiplication" $a \cdot b$. Then the algebra on the same space with multiplication defined by

$$a \times b = \alpha a \circ b + (1-\alpha) a \cdot b \quad (0 \leq \alpha \leq 1) \tag{5}$$

will be called a mixture of the two given algebras. Certain modes of inheritance that occur in nature can be described in terms of mixtures of chromosome and chromatid segregation, although this is unfortunately not possible when the probability of non-disjunction exceeds $1/7$ (see 13, p. 30). The relationship (5) is preserved under changes in the basis of the vector space. It is clear from the form of the multiplication tables that a mixture of special train algebras is again a special train algebra with train roots "mixed" in the same proportions as the algebras. For an algebra representing a mixture of α chromatid segregation and $(1-\alpha)$ chromosome segregation, the dominant non-unit principal train root is $\frac{1}{8} - \frac{1}{4}\alpha$ for tetraploidy, and $\frac{1}{5} - \frac{1}{5}\alpha$ for hexaploidy. It is implicit in (10), § 2, that the rate of convergence of a sequence of plenary powers, corresponding to repeated random mating, is ultimately twice the largest non-unit principal train root, i.e. $\frac{1}{4} - \frac{1}{2}\alpha$ and $\frac{2}{5} - \frac{2}{5}\alpha$ in the above cases.

7. Isotopy and special isotopy

The elements of an algebra A can be taken as vectors of coefficients of the basis elements, and to each element $x = \sum x_i c_i \equiv (x_0, \dots, x_n)$ a matrix R_x can be made to correspond, defined by $y \cdot x = yR_x$. The properties of A can be usefully studied in terms of the transformation algebra generated by the R_x and I (12, 1, 2) and this fact has been used in genetic algebras by Schafer (15). If the multiplication table of the algebra is (3), the matrix R_x corresponding to (x_0, \dots, x_n) has (i, t) th element $\sum_{j=0}^n \gamma_{ijt} x_j$. If for another algebra (circle multiplication) $y \circ x = yR_x^0$ and non-singular matrices can be found such that

$$R_x^0 = PR_xQW \tag{6}$$

the algebras are defined by Albert (1) to be isotopic. An equivalent definition by Bruck (2) is

$$y \circ x = (yP \cdot xQ)W.$$

Isotopy is an equivalence relation, and so are the specialisations of it obtained by requiring any set of P, Q, W to be unit matrices. For instance if $W = I$ the relation of principal isotopy is obtained (1). The genetic algebras studied here are isotopic with $P = Q = I$, and I have defined this relation as special isotopy. It is preserved if each algebra is transformed into an equivalent one by the same transformation, as can be seen simply by substitutions like those in § 11 of (1). It reduces to the fact that multiplication in A^0 is obtained by carrying out multiplication according to the rules of A , followed by post-multiplication by a non-singular matrix. The mutation algebras introduced by Gonshor (10) are in general special isotopes of those not involving mutation, and consequently of each other. The exceptions occur because for certain values of the mutation rates, the mutation mapping may be singular. In view of the conditions on the γ_{ab} , obtained for segregation algebras, the matrices

R_x can be written in the form

$$\begin{aligned}
 & \begin{bmatrix} \gamma_{00}x_0 & \gamma_{01}x_0 + \gamma_{11}x_1 & \gamma_{02}x_0 + \gamma_{12}x_1 + \gamma_{22}x_2 & \dots \\ 0 & \gamma_{11}x_0 & \gamma_{12}x_0 + \gamma_{22}x_1 & \dots \\ 0 & 0 & \gamma_{22}x_0 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \\
 = & \begin{bmatrix} x_0 & x_1 & x_2 & \dots \\ 0 & x_0 & x_1 & \dots \\ 0 & 0 & x_0 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} \gamma_{00} & \gamma_{01} & \gamma_{02} & \dots \\ 0 & \gamma_{11} & \gamma_{12} & \dots \\ 0 & 0 & \gamma_{22} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}
 \end{aligned}$$

Since the γ_{mm} are non-zero, it follows from this and from (5) that all segregation algebras and all mixtures of them are specially isotopic.

However, it is possible for algebras to arise corresponding to other genetic systems which are not even isotopic to a segregation algebra. Consider the following algebras of order 3: (i) the zygotic algebra for diploid chromosome segregation, and (ii) any gametic algebra for tetraploidy. With the canonical bases, the element whose vector representation is (x, y, z) corresponds in (i) and (ii) respectively to the matrices

$$\begin{bmatrix} x & \frac{1}{2}y & 0 \\ 0 & \frac{1}{2}x & \frac{1}{4}y \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} x & \frac{1}{2}y & kz \\ 0 & \frac{1}{2}x & kz \\ 0 & 0 & kz \end{bmatrix}$$

which cannot be transformed into each other by an equation such as (6).

I am grateful to the referee for a number of helpful suggestions.

REFERENCES

(1) A. A. ALBERT, Non-associative algebras. I. Fundamental concepts and isotopy, *Ann. Math.*, **43** (1943), 685-707.
 (2) R. H. BRUCK, Some results in the theory of linear non-associative algebras, *Trans. Amer. Math. Soc.*, **56** (1944), 141-199.
 (3) I. M. H. ETHERINGTON, On non-associative combinations, *Proc. Roy. Soc. Edinburgh*, **59** (1939), 153-162.
 (4) I. M. H. ETHERINGTON, Genetic algebras, *Proc. Roy. Soc. Edinburgh*, **59** (1939), 242-258.
 (5) I. M. H. ETHERINGTON, Commutative train algebras of ranks 2 and 3, *J. London Math. Soc.*, **15** (1940), 136-149.
 (6) I. M. H. ETHERINGTON, Special train algebras, *Quart. J. Math.* **12** (1941), 1-8.

(7) I. M. H. ETHERINGTON, Duplication of linear algebras, *Proc. Edinburgh Math. Soc.* (2), **6** (1941), 222-230.

(8) I. M. H. ETHERINGTON, Non-associative algebra and the symbolism of genetics, *Proc. Roy. Soc. Edinburgh B*, **61** (1941), 24-42.

(9) H. GEIRINGER, Chromatid segregation of tetraploids and hexaploids, *Genetics*, **34** (1949), 665-684.

(10) H. GONSHOR, Special train algebras arising in genetics, *Proc. Edinburgh Math. Soc.* (2), **12** (1960), 41-53.

(11) J. B. S. HALDANE, Theoretical genetics of autoployploids, *J. Genetics*, **22** (1930), 359-372.

(12) N. JACOBSON, A note on non-associative algebras, *Duke Math. J.*, **3** (1937), 544-548.

(13) P. A. P. MORAN, *Statistical Processes of Evolutionary Theory* (Oxford, 1962).

(14) O. REIERSØL. Genetic algebras studied recursively and by means of differential operators. *Math. Scand.*, **10** (1962), 25-44.

(15) R. D. SCHAFER, Structure of genetic algebras, *Amer. J. Math.*, **71** (1949), 121-135.

(16) W. SEYFFERT, Theoretische Untersuchungen über die Zusammensetzung tetrasomer Populationen. I. Panmixie, *Biom. Zeit.*, **2** (1960), 1-44.

BIOMETRICS SECTION

THE NATURE CONSERVANCY

19 BELGRAVE SQUARE, LONDON, S.W.1