SPECIAL TRAIN ALGEBRAS ARISING IN GENETICS II

by HARRY GONSHOR

(Received 1st June 1964; revised MS. received 30th December 1964)

1. Introduction

We shall extend some of the results of (7) to the case of multiple alleles, our primary concern being that of polyploidy combined with multiple alleles. Generalisations often tend to make the computations more involved as is expected. Fortunately here, the attempt to generalise has led to a new method which not only handles the case of multiple alleles, but is an improvement over the method used in (7) for the special case of polyploidy with two alleles. This method which consists essentially of expressing certain elements of the algebra in a so-called " factored " form, gives greater insight into the structure of a polyploidy algebra, and avoids a great deal of the computation with binomial coefficients, e.g. see (7), p. 46.

In a subsequent paper we plan to consider mutations and overlapping of generations.

2. Characterisation of gametic multiple allelic algebras

If the gametic types are $D_1, D_2, \ldots, D_n$, then the multiplication table of this well-known algebra is: $D_iD_j = \frac{1}{2}(D_i + D_j)$. This is a special train algebra; in fact, the ideal consisting of all elements of zero weight has its square equal to 0.

Note that if c and d are any two elements of the algebra of weight one, then $cd = \frac{1}{2}(c+d)$. Let

$$c = \sum_{i=1}^{n} x_iD_i \text{ and } d = \sum_{i=1}^{n} y_iD_i \text{ where } \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = 1.$$

Then

$$cd = \sum_{i,j=1}^{n} x_iy_jD_iD_j = \tfrac{1}{2}\sum_{i,j=1}^{n} x_iy_j(D_i+D_j) = \tfrac{1}{2}\sum_{i=1}^{n} x_iD_i + \tfrac{1}{2}\sum_{j=1}^{n} y_jD_j = \tfrac{1}{2}c+\tfrac{1}{2}d.$$

It follows that the algebra has the following interesting property: Every linear mapping of the algebra into itself which preserves weight also preserves multiplication. We now show that the gametic multiple allelic algebras are characterised by this property.

Theorem 2.1. *Let A be a baric algebra with basis $(a_1, a_2, \ldots, a_n)$ where* *the a's have weight one and with the property that every linear mapping of the algebra into itself which preserves weight also preserves multiplication. Then* $a_i a_j = \frac{1}{2}(a_i + a_j)$.

Proof. Any linear mapping which preserves weight on the basis preserves weight on all of A. Hence the mapping which sends each of the a's into a_i for fixed i preserves weight. Thus it preserves multiplication. It follows that the image is closed under multiplication, i.e. $a_i^2 = x a_i$ for some scalar x. By comparing weights we see that x must be one.

Now consider the mapping which sends each a into $\frac{1}{2}(a_i + a_j)$. As in the first part of the proof $[\frac{1}{2}(a_i + a_j)]^2 = \frac{1}{2}(a_i + a_j)$. Hence

$$\tfrac{1}{4}(a_i^2 + 2a_i a_j + a_j^2) = \tfrac{1}{2}(a_i + a_j).$$

Using the fact that $a_i^2 = a_i$ and $a_j^2 = a_j$, this gives $a_i a_j = \frac{1}{2}(a_i + a_j)$. Q.E.D.

Note. The full strength of the fact that multiplication is preserved was not needed in the proof. We required only the consequence that the image is closed under multiplication.

3. Zygotic multiple allelic algebras

In this case the basis consists of the pairs $D_i D_j$ with multiplication table: $(D_i D_j)(D_k D_l) = \frac{1}{4}(D_i D_k + D_i D_l + D_j D_k + D_j D_l)$. According to (5) we may write $a_i = D_1 - D_i (i \neq 1)$ and use as a basis $D_1 D_1$, $D_1 a_i$, $a_i a_j$ where $i, j \neq 1$. We then have the table: $(D_1 D_1)^2 = D_1 D_1$, $(D_1 D_1)(D_1 a_i) = \frac{1}{2} D_1 a_i$, $(D_1 a_i)(D_1 a_j) = \frac{1}{4} a_i a_j$, $(a_i a_j)x = 0$ for all x. This is clearly a special train algebra.

4. Sex linkage with multiple alleles

We denote the genotypes of the homogametic sex by pairs $D_i D_j$ and of the heterogametic sex by individual symbols D_i. The multiplication table is:

$$(D_i D_j)(D_k) = \tfrac{1}{4}(D_i D_k + D_j D_k + D_i + D_j),\ (D_i D_j)(D_k D_l) = (D_i)(D_j) = 0.$$

The table is obtained from the gametic case by a process analogous to duplication (5); cf. end of (4). However, caution is required since the process is more complicated; for example, change of basis does not work as simply here as in the case of duplication.

As a new basis we choose $D_1 D_1$, $D_1 a_i$, $a_i a_j$, D_1, a_i, where $i, j \neq 1$; and $D_1 a_i = D_1 D_1 - D_1 D_i$, $a_i a_j = D_1 D_1 - D_1 D_i - D_1 D_j + D_i D_j$ and $a_i = D_1 - D_i$. The multiplication table is:

$$(D_1 D_1)(D_1) = \tfrac{1}{2}(D_1 D_1 + D_1),\ (D_1 a_i)(D_1) = \tfrac{1}{4}(D_1 a_i + a_i),\ (a_i a_j)(D_1) = 0,$$

$$(D_1 D_1)(a_i) = \tfrac{1}{2} D_1 a_i,\ (D_1 a_i)(a_j) = \tfrac{1}{2} a_i a_j,\ (a_i a_j) a_k = 0,$$

and all other products are 0.

The space $\{D_1 D_1 + D_1,\ D_1 a_i,\ a_i a_j,\ a_i\}$ is an ideal of deficiency one which contains the product of any two elements in the original ring. Furthermore this ideal is a special train algebra. In it the ideal I of elements of zero weight is $\{D_1a_i, a_ia_j, a_i\}$, $I^2 = \{a_ia_j\}$, and $I^3 = 0$. Note that the powers of I are ideals in the original ring.

We feel that this notation is an improvement over the notation used in the special case treated in ((7), section 4), since it appears to make the structure of the multiplication table more lucid.

5. Polyploidy with multiple alleles

We shall first consider the gametic case. Assume n alleles $D_1, D_2, \ldots, D_n$ and r-ploidy. Thus the basis consists of all monomials in the D's of degree r and hence has $\binom{n+r-1}{r}$ elements. The multiplication table is defined as follows:

$$(x_1x_2, \ldots x_r)(x_{r+1}x_{r+2}\ldots x_{2r}) = \binom{2r}{r}^{-1} \Sigma\, x_{j_1}x_{j_2}\ldots x_{j_r}$$

where each x is one of the D's and the sets $j_1, \ldots, j_r$ run through all combinations of r integers from the set $1, 2, \ldots, 2r$.

Note that on the lines of the present notation D_i in ((7), section 6) would have been written D^iR^{n-i} and that the table there is obtained by collecting terms. We realise now that this has led to superfluous computation since the present form of the table leads to a simpler way of obtaining the structure of the algebra.

We attach significance to the symbol $\prod_{i=1}^{r}\left(\sum_{j=1}^{n} a_{ij}D_j\right)$ as follows. It is merely a convenient way of expressing the element $\sum_j a_{1j_1}a_{2j_2}\ldots a_{rj_r}D_{j_1}D_{j_2}\ldots D_{j_r}$ where $j_1, j_2, \ldots, j_r$ run through all r-tuples such that $1 \leqq j_i \leqq n$ for all i. This is of course what one would obtain by ordinary multiplication if the D's were regarded as ordinary indeterminates. However, a formal definition is needed here since the individual $\Sigma\, a_{ij}D_j$ have no meaning within the original algebra. For convenience we also allow the use of juxtaposition and the exponential notation in the obvious way. For example, c_s in ((7), section 6) may be written $D^{n-s}(D-R)^s$.

We next obtain the formula for the product of two elements expressed in " factored form ".

Theorem 5.1. $\left(\prod_{i=1}^{r}\sum_{j=1}^{n} a_{ij}D_j\right)\left(\prod_{i=r+1}^{2r}\sum_{j=1}^{n} a_{ij}D_j\right)$

$$= \binom{2r}{r}^{-1} \sum_k \left(\prod_{i=1}^{r}\sum_{j=1}^{n} a_{l_ij}\right)\left[\prod_{i=1}^{r}\sum_{j=1}^{n} a_{k_ij}D_j\right]$$

where the outside summation is taken over all subsets $(k_1, k_2, \ldots, k_r)$ of r integers from the set $1, 2, \ldots, 2r$ and where $(l_1, l_2, \ldots, l_r)$ is the complementary subset.

Proof. $\left(\prod_{i=1}^{r}\sum_{j=1}^{n} a_{ij}D_j\right)\left(\prod_{i=r+1}^{2r}\sum_{j=1}^{n} a_{ij}D_j\right)$

$$= (\Sigma a_{1j_1}\dots a_{rj_r}D_{j_1}\dots D_{j_r})(\Sigma a_{(r+1)j_{r+1}}\dots a_{(2r)j_{2r}}D_{j_{r+1}}\dots D_{j_{2r}})$$
$$= \Sigma a_{1j_1}a_{2j_2}\dots a_{(2r)j_{2r}}(D_{j_1}D_{j_2}\dots D_{j_r})(D_{j_{r+1}}D_{j_{r+2}}\dots D_{j_{2r}})$$

where $j_1, j_2, \dots, j_{2r}$ run through all sequences such that $1 \leqq j_i \leqq n$ for all i. This in turn is

$$(\Sigma a_{1j_1}a_{2j_2}\dots a_{(2r)j_{2r}})\left[\binom{2r}{r}^{-1} \Sigma D_{j_{k_1}}D_{j_{k_2}}\dots D_{j_{k_r}}\right]$$

where the inside summation is taken over all subsets $\{k_1, k_2, \dots, k_r\}$ of r integers from the set 1, 2, ..., $2r$. We now interchange the order of summation and fix $k_1, k_2, \dots, k_r$. Let $l_1, l_2, \dots, l_r$ be the complementary set. The summation becomes

$$\binom{2r}{r}^{-1} \Sigma a_{l_1j_{l_1}}\dots a_{l_rj_{l_r}}a_{k_1j_{k_1}}\dots a_{k_rj_{k_r}}D_{j_{k_1}}D_{j_{k_2}}\dots D_{j_{k_r}}$$
$$= \binom{2r}{r}^{-1} (\Sigma a_{l_1j_{l_1}}\dots a_{l_rj_{l_r}})(\Sigma a_{k_1j_{k_1}}\dots a_{k_rj_{k_r}}D_{j_{k_1}}\dots D_{j_{k_r}}).$$

The right factor is exactly $\prod_{i=1}^{r}\sum_{j=1}^{n} a_{k_ij}D_j$ by our notation and the left factor is $\prod_{i=1}^{r}\sum_{j=1}^{n} a_{l_ij}$ by ordinary algebra. This proves the result.

The rough idea is that multiplication of elements expressed in "factored form" resembles multiplication of the original basis elements except that for each "factor" not chosen we must multiply by the sum of its coefficients.

We now specialise the above result to elements of the form

$$D_1^{i_1}(D_1-D_2)^{i_2}\dots(D_1-D_n)^{i_n}$$

where $\sum_{j=1}^{n} i_j = r$. Note that these elements form a basis to the polyploidy algebra. (The proof is identical to the case of ordinary algebra.)

Theorem 5.2. $[D_1^{i_1}(D_1-D_2)^{i_2}\dots(D_1-D_n)^{i_n}][D_1^{j_1}(D_1-D_2)^{j_2}\dots(D_1-D_n)^{j_n}]$

$$= \binom{2r}{r}^{-1}\binom{i_1+j_1}{r} D_1^{i_1+j_1-r}(D_1-D_2)^{i_2+j_2}\dots(D_1-D_n)^{i_n+j_n} \textit{ if } i_1+j_1 \geqq r$$
$$= 0 \textit{ if } i_1+j_1 < r.$$

Proof. Since the sum of the coefficients of $D_1 - D_i$ is zero, the only non-vanishing terms in the product are those containing all the $(D_1 - D_i)$'s in the factors. If $i_1+j_1 < r$ there are no such terms.

If $i_1+j_1 \geqq r$ there are $\binom{i_1+j_1}{r}$ such terms, this being the number of ways of choosing the (i_1+j_1-r) D_1's to bring the number of factors up to r. Of

https://doi.org/10.1017/S0013091500009020 Published online by Cambridge University Press

course, the sum of the coefficients of each omitted D_1 is one. Thus we have $\binom{i_1+j_1}{r}$ terms each of which is

$$\binom{2r}{r}^{-1} D_1^{i_1+j_1-r}(D_1-D_2)^{i_2+j_2}\dots(D_1-D_n)^{i_n+j_n}.$$

This proves the result.

Note that Lemma 6.2 in (7) is a special case.

As an immediate consequence of Theorem 5.2 we obtain the following.

Theorem 5.3. *Gametic polyploidy multiple allelic algebras are special train algebras.*

The ideal I of elements of zero weight is the subspace generated by all the new basis elements other than D_1^r. I^k is the space generated by all basis elements with at least k factors of the form D_1-D_i for some i. The train roots are $\binom{2r}{r}^{-1}\binom{m}{r}$ where $2r \geqq m \geqq r$; $\binom{2r}{r}^{-1}\binom{m}{r}$ is repeated $\binom{2r-m+n-2}{2r-m}$ times. (This is the number of monomials of degree $2r-m$ in $n-1$ variables.)

By means of the factored form it is possible to give a very simple proof of Lemma 6.3 in (7). This states essentially the self-reciprocal relationship between the old and new bases. Since as a vector space the algebra is isomorphic to the linear space of polynomials in n variables of homogeneous degree r, it suffices to prove this for ordinary algebra. But $D_i = D_1-(D_1-D_i)$, hence

$$D_1^{s_1}D_2^{s_2}\dots D_n^{s_n} = D_1^{s_1}[D_1-(D_1-D_2)]^{s_2}\dots[D_1-(D_1-D_n)]^{s_n}.$$

This makes the self-reciprocal property evident.

By applying the technique of duplication (5) we can now consider the zygotic algebra. We identify all pairs

$$[D_1^{i_1}(D_1-D_2)^{i_2}\dots(D_1-D_n)^{i_n},\ D_1^{j_1}(D_1-D_2)^{j_2}\dots(D_1-D_n)^{j_n}]$$

with fixed i_k+j_k for all k. Thus we may use the symbol

$$D_1^{i_1}(D_1-D_2)^{i_2}\dots(D_1-D_n)^{i_n}$$

to stand for any pair for which the sum of the exponents of D_1-D_k is i_k for all k. This leads to the following theorem:

Theorem 5.4. *Zygotic polyploidy multiple allelic algebras are special train algebras.*

In fact, the table has the form

$$[D_1^{i_1}(D_1-D_2)^{i_2}\dots(D_1-D_n)^{i_n}][D_1^{j_1}(D_1-D_2)^{j_2}\dots(D_1-D_n)^{j_n}]$$
$$= \binom{i_1}{r}\binom{j_1}{r}\binom{2r}{r}^{-2} D_1^{i_1+j_1-2r}(D_1-D_2)^{i_2+j_2}\dots(D_1-D_n)^{i_n+j_n}$$

for $i_1, j_1 \geqq r$; $= 0$ otherwise.

Theorems 6.3 and 6.4 in (7) can also be generalised to the case of multiple alleles. Since no essentially new technique is involved it will suffice to state what happens in the gametic case. If $y_2, y_3, \ldots, y_n$ are arbitrary, then a typical idempotent has the form

$$\sum_{i_1+i_2\ldots i_n=r}(-1)^{r-i_1}\frac{r!}{i_1!i_2!\ldots i_n!}y_2^{i_2}y_3^{i_3}\ldots y_n^{i_n}D_1^{i_1}(D_1-D_2)^{i_2}\ldots(D_1-D_n)^{i_n}.$$

In terms of the original basis this is expressed as

$$\frac{r!}{j_1!j_2!\ldots j_n!}y_2^{j_2}\ldots y_n^{j_n}(1-y_2-y_3\ldots-y_n)^{j_1}D_1^{j_1}D_2^{j_2}\ldots D_n^{j_n}.$$

The form of this last expression leads to a result which can be expressed in the language of classical genetics, namely that if the genotype distribution is a multinomial distribution, then it is in equilibrium. (Note that some of the y's or $1-y_2-y_3\ldots-y_n$ may be negative. This is because in the algebraic theory we are working with the whole algebra, not only with elements which represent distributions.)

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. Gonshor, Special train algebras arising in genetics. *Proc. Edin. Math. Soc.* (2) **12** (1960), 41-53.

(8) R. Raffin, Axiomatisation des algèbres génétiques. *Acad. Roy. Belgique Bull. Cl. Sci.* (5) **37** (1951), 359-366.

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

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

Rutgers University
New Brunswick
New Jersey