

## Genetic Algebras for Tetraploidy with Several Loci

Peter Fortini<sup>1</sup> and Richard Barakat<sup>2</sup>

<sup>1</sup> Department of Statistics, Rutgers University, New Brunswick, NJ 08903

<sup>2</sup> Division of Applied Sciences, Harvard University, Cambridge, MA 02138 and Bolt Beranek and Newman Inc., Cambridge, MA 02138, USA

**Abstract.** The algebra of gamete multiplication corresponding to a population of autotetraploids which differ at  $n$  loci, each with an arbitrary number of alleles, is proven to be a genetic algebra, irrespective of the mode of segregation of the  $n$  loci.

**Key words:** Genetic algebra – Polyploidy – Tetraploids.

### 1. Introduction

The purpose of this paper is to show that the algebra of gamete multiplication corresponding to a population of autotetraploid individuals which differ at  $n$  loci, each with an arbitrary number of alleles, is a genetic algebra. The result follows from an assertion, proved in Section 4, that certain linear combinations of the gamete distributions produced from different tetraploid genotypes vanish identically, together with a characterization of genetic algebras given by Gonshor (1973). The argument is valid whatever the mode of segregation of the  $n$  loci.

The gamete algebras for polyploid organisms have received extensive investigation. Gonshor (1965) and Holgate (1966) contain early contributions. More recently, Wörz-Busekros (1978) and Heuch (1978) have exhibited canonical bases for the genetic algebras connected with polyploid inheritance for arbitrary modes of segregation. All these results treat behavior at a single locus only. For two loci, Bennett (1954) and Crow (1954) obtained recursion relations for gamete frequencies without using genetic algebras.

### 2. Genetic Algebras

Nonassociative algebras were first introduced into genetics by Etherington (1939), and are the subject of a growing literature. They arise in the following way.

We characterize the mechanism of inheritance of an organism by a set  $\{G_1, G_2, \dots, G_N\}$  of distinguishable gametes, another set  $\{Z_{ij}: i, j = 1, \dots, N\}$  of zygote genotypes in which  $Z_{ij}$  is the genotype obtained by the fusion of gametes  $G_i$  and  $G_j$ , and by a law of gamete formation  $f$  which gives for each  $Z_{ij}$  the probability distribution of its gametic output.

An “algebra” for gamete multiplication arises when the gamete distribution  $f(Z_{ij})$  is used to define a multiplication of the initial gametes  $G_i$  and  $G_j$ , so that

$$G_i \times G_j = f(Z_{ij}) = \sum_{k=1}^N p_{ijk} G_k \tag{2.1}$$

in which  $p_{ijk}$  is the proportion of  $G_k$  gametes produced by zygote  $Z_{ij}$ . The multiplication of gametes can be extended in the obvious way to the entire vector space  $\mathfrak{G}$  of formal real or complex linear combinations of gametes by setting

$$\begin{aligned} \left( \sum_{i=1}^N g_i G_i \right) \times \left( \sum_{j=1}^N g'_j G_j \right) &= \sum_{i,j} g_i g'_j f(Z_{ij}) \\ &= \sum_{k=1}^N \left( \sum_{i,j} g_i g'_j p_{ijk} \right) G_k \end{aligned} \tag{2.2}$$

This multiplication of elements of  $\mathfrak{G}$  distributes over addition, is commutative, but is generally not associative.

This algebra of gamete multiplication is of interest primarily in the context of an infinite randomly mating population of individuals whose mechanism of inheritance is described by Eq. (2.1). If

$$\tilde{g}^{(0)} = \sum_{i=1}^N g_i^{(0)} G_i$$

represents in  $\mathfrak{G}$  the derived gamete distribution of the initial population, then the derived gamete distributions  $\tilde{g}^{(1)}, \tilde{g}^{(2)}, \dots$  of the population after 1, 2, ... generations of random mating are represented in  $\mathfrak{G}$  by the sequence of successive squares of  $\tilde{g}^{(0)}$ :

$$\begin{aligned} \tilde{g}^{(1)} &= \tilde{g}^{(0)} \times \tilde{g}^{(0)} \\ \tilde{g}^{(2)} &= \tilde{g}^{(1)} \times \tilde{g}^{(1)} \\ &\dots\dots\dots \\ \tilde{g}^{(r)} &= \tilde{g}^{(r-1)} \times \tilde{g}^{(r-1)} \end{aligned} \tag{2.3}$$

From  $\tilde{g}^{(r)}$  we can compute the genotype distribution of the population after  $r$  generations as

$$\sum_{i,j} g_i^{(r)} g_j^{(r)} Z_{ij}$$

An algebra  $\mathfrak{G}$  of gamete multiplication is a *genetic algebra* if a basis  $\{\gamma_0, \gamma_1, \dots, \gamma_{N-1}\}$  can be found for it with respect to which the multiplication satisfies

$$\begin{aligned} \gamma_i \times \gamma_j &= \sum_k c_{ijk} \gamma_k \\ c_{000} &= 1 \\ c_{0jk} &= 0 \quad \text{if } k < j \\ c_{ijk} &= 0 \quad \text{if } i, j > 0 \text{ and } k \leq \max(i, j) \end{aligned} \tag{2.4}$$

Such a basis is termed a canonical basis of  $\mathfrak{G}$ .

Equation (2.3), expressed in terms of the gamete frequencies  $g_i^{(r)}$ , is a quadratic recurrence relation. If, however,  $\mathfrak{G}$  is a genetic algebra, study of the recurrence relation is greatly simplified, for it is possible to find a vector of polynomial functions of the gene frequencies

$$\tilde{h}^{(r)} = P_1|g_1^{(r)}, \dots, g_N^{(r)}|, \dots, P_M|g_1^{(r)}, \dots, g_N^{(r)}|$$

which obeys a linear recurrence relation (Holgate, 1967).

Notice from Eq. (2.4) that the subspace  $\mathfrak{C} = \text{span}(\gamma_1, \dots, \gamma_{N-1})$  of  $\mathfrak{G}$  is closed under multiplication by any element of  $\mathfrak{G}$ , and thus in algebraic terms is an ideal of  $\mathfrak{G}$ . It is not difficult to see that  $\mathfrak{C}$  can always be taken to be the subspace of contrasts (as the term is used in statistics) of gamete types. Examining Eq. (2.2), if either factor on the left is a contrast ( $\sum g_i = 0$  or  $\sum g'_i = 0$ ), then so is the resultant on the right, since

$$\begin{aligned} \sum_k \left( \sum_{i,j} g_i g'_j p_{ijk} \right) &= \sum_{i,j} g_i g'_j \sum_k p_{ijk} \\ &= \left( \sum_i g_i \right) \left( \sum_j g'_j \right) \end{aligned}$$

It also follows from Eq. (2.4) that a genetic product containing sufficiently many ( $n$  or more) factors from  $\mathfrak{C}$  vanishes however associated. Thus if  $\mathfrak{G}$  is genetic, the ideal of contrasts is algebraically nilpotent. A converse due to Gonshor (1973), characterizing an algebra  $\mathfrak{G}$  as genetic in terms of the vanishing of products containing enough factors from  $\mathfrak{C}$ , will be applied in Section 4.

### 3. Single Locus Case

We introduce the algebras of gamete multiplication for autotetraploid organisms by considering first the process of gamete formation at a single locus,  $A$ . A typical genotype is  $A_1 A_2 A_3 A_4$ . Following Fisher (1947), let  $\alpha$  be the probability that the two alleles of the (diploid) gamete derive from the same chromosome of the zygote. Then possible gametes  $A_i A_j$  ( $i \neq j$ ) are produced in proportions  $(1 - \alpha)/6$ , while gametes  $A_i A_i$  are produced in proportions  $\alpha/4$  each. We shall write this

$$\begin{aligned} f(A_1 A_2 A_3 A_4) &= \frac{(1 - \alpha)}{6} (A_1 A_2 + A_1 A_3 + A_1 A_4 + A_2 A_3 + A_2 A_4 + A_3 A_4) \\ &\quad + \frac{\alpha}{4} (A_1 A_1 + A_2 A_2 + A_3 A_3 + A_4 A_4) \end{aligned} \tag{3.1}$$

The parameter  $\alpha$  specifies the mode of segregation at locus  $A$ .  $\alpha = 0$  corresponds to chromosome segregation;  $\alpha = \frac{1}{2}$  to chromatid segregation.

The function  $f$  in Eq. (3.1), which gives the law of gamete formation for this one-locus system, has for its domain the set of all possible genotypes, and for its range the space of probability distributions over (or formal linear combinations of) gamete types. By extension  $f$  is a linear function from the vector space of all formal linear combinations of (tetraploid) zygotes to that of formal linear combinations of (diploid) gametes.

Consider the set of formal linear combinations of genotypes which may be represented in "factorized" form as

$$\begin{aligned}
 & A_1 A_1 A_1 A_1 \\
 & A_1 A_1 A_1 (A_1 - A_i) \\
 & A_1 A_1 (A_1 - A_i)(A_1 - A_j) \\
 & A_1 (A_1 - A_i)(A_1 - A_j)(A_1 - A_k) \\
 & (A_1 - A_i)(A_1 - A_j)(A_1 - A_k)(A_1 - A_l)
 \end{aligned} \tag{3.2}$$

where the possible alleles for locus  $A$  are  $A_1, A_2, \dots, A_m$ . Each expression in Eq. (3.2) becomes, upon purely formal expansion, a linear combination of genotypes with coefficients  $\pm 1$ . The combinations will be linearly independent, and so are a basis of the space of zygote genotypes, if we restrict the subscripts  $i, j, k$ , and  $l$  to satisfy  $2 \leq i \leq j \leq k \leq l \leq m$ .

Using Eq. (3.1), one can verify by straightforward calculation that the representation of  $f$  with respect to this basis is

$$f[A_1 A_1 A_1 A_1] = A_1 A_1 \tag{3.3a}$$

$$\begin{aligned}
 f[A_1 A_1 A_1 (A_1 - A_i)] &= f(A_1 A_1 A_1 A_1) - f(A_1 A_1 A_1 A_i) \\
 &= \frac{1}{2} A_1 (A_1 - A_i) - \frac{\alpha}{4} (A_1 - A_i)(A_1 - A_i)
 \end{aligned} \tag{3.3b}$$

$$f[A_1 A_1 (A_1 - A_i)(A_1 - A_j)] = \frac{(1 - \alpha)}{6} (A_1 - A_i)(A_1 - A_j) \tag{3.3c}$$

$$f[A_1 (A_1 - A_i)(A_1 - A_j)(A_1 - A_k)] = 0 \tag{3.3d}$$

$$f[(A_1 - A_i)(A_1 - A_j)(A_1 - A_k)(A_1 - A_l)] = 0 \tag{3.3e}$$

It can be seen that basis elements having the form of the last two lines of Eq. (3.2) are in the kernel of the function  $f$ .

Equation (3.3) implies that the algebra for gamete multiplication of a one-locus autotetraploid organism is a genetic algebra provided no mutation or selection effects occur. A canonical basis is obtained by setting  $\gamma_0 = A_1 A_1$ ,  $\gamma_i = A_1 (A_1 - A_{i+1})$ ,  $i = 1, \dots, m - 1$ , and  $\gamma_m, \dots, \gamma_N$  (where  $N = m(m + 1)/2$  is the total number of possible gamete types) equal to the basis elements  $(A_1 - A_i)(A_1 - A_j)$  taken in any order. The genetic product of two basis elements  $\gamma_i \times \gamma_j$  is in this case simply  $f$  applied to their juxtaposition.

#### 4. Multilocus Case

We now suppose for an autotetraploid organism that  $n$  loci  $A, B, C, \dots$  are in question, each having two or more possible alleles. In order to show that the corresponding gametic algebra is genetic, we derive formulas analogous to those of Eq. (3.3), relating to the segregation behavior at one locus,  $A$ , given that at the others. A combinatorial argument then shows that the multilocus algebra satisfies Gonshor's (1973) characterization of genetic algebras. This approach proves that the algebra is genetic without actually exhibiting a canonical basis for it.

Let the possible alleles for locus  $A$  be denoted  $A_i$  or  $a_i$ . Also let  $W, X, Y, Z, U$  and  $V$  represent combinations of alleles from loci  $B, C, \dots$ , each specifying the makeup of a chromosome for the  $n - 1$  loci other than  $A$ .

We call a *unit contrast* the difference between two gamete types which differ only in the substitution of one allele for another in one chromosome. A typical unit contrast for locus  $A$ , represented also in “factorized” form, is

$$A_1W \cdot A_2X - A_1W \cdot a_2X = A_1W \cdot (A_2 - a_2)X$$

Unit contrasts for other loci  $B, C, \dots$  are defined similarly.

The following lemma permits us to restrict attention to unit contrasts in evaluating genetic products involving contrasts.

**Lemma 4.1.** *For the gametic algebra  $\mathfrak{G}$  of an autotetraploid organism, the ideal of contrasts  $\mathfrak{C}$  is spanned by unit contrasts.*

*Proof.* It is clear that  $\mathfrak{C}$  is spanned by all formal differences between pairs of gamete types. Any one gamete type, however, can be transformed into any other by at most  $2n$  one-allele substitutions, two for each locus. Hence the difference between two gamete types is the sum of at most  $2n$  unit contrasts.

A series of three propositions give the results of certain products of unit contrasts. Suppose that, ignoring  $A$ , the mode of gamete formation of loci  $B, C, \dots$  can be expressed as

$$f(W \cdot X \cdot Y \cdot Z) = \sum_{U,V} p(U, V)U \cdot V \tag{4.1}$$

where the sum is over unordered pairs of chromosomes  $U$  and  $V$ , and  $p(U, V)$  is the probability that the gamete  $U \cdot V$  is produced from a zygote of genotype  $W \cdot X \cdot Y \cdot Z$ .  $p(U, V)$  also depends on  $W, X, Y$ , and  $Z$ , but we have suppressed this dependence in the notation. The dots in  $W \cdot X \cdot Y \cdot Z$  and in  $U \cdot V$  serve to separate the contents of homologous chromosomes.

A typical zygote, described at all  $n$  loci, is  $A_1W \cdot A_2X \cdot A_3Y \cdot A_4Z$ . For notational convenience, the subscripts  $i = 1, 2, 3, 4$  serve to distinguish not only between possibly different alleles at locus  $A$ , but also to indicate place. The distribution of gametes produced from this zygote may be written

$$f(A_1W \cdot A_2X \cdot A_3Y \cdot A_4Z) = \sum_{U,V} \sum_{i,j=1}^4 p(U, V; i, j)A_iU \cdot A_jV \tag{4.2}$$

where  $p(U, V; i, j)$  is the probability that a gamete produced from zygote  $A_1W \cdot A_2X \cdot A_3Y \cdot A_4Z$  will be of type  $A_iU \cdot A_jV$ . Notice that

$$\sum_{i,j=1}^4 p(U, V; i, j) = p(U, V)$$

$$\sum_{U,V} p(U, V; i, j) = \begin{cases} (1 - \alpha)/12 & \text{if } i \neq j \\ \alpha/4 & \text{if } i = j \end{cases}$$

where  $\alpha$  is as in Eq. (3.1).  $(1 - \alpha)/12$  rather than  $(1 - \alpha)/6$  appears above since the gamete types  $A_iU \cdot A_jV$  and  $A_jU \cdot A_iV$ , while identical in genetic content, represent different arrangements of the genetic material and so receive separate accounting.

The product in  $\mathfrak{G}$  of a gamete type  $A_1W \cdot A_2X$  and a unit contrast in locus  $A$ ,  $A_3Y \cdot (A_4 - a_4)Z$ , is given by the gamete distribution

$$f[A_1W \cdot A_2X \cdot A_3Y \cdot (A_4 - a_4)Z] = f(A_1W \cdot A_2X \cdot A_3Y \cdot A_4Z) - f(A_1W \cdot A_2X \cdot A_3Y \cdot a_4Z)$$

This is the difference between two expressions of the form Eq. (4.2), identical but that where  $A_4$  appears in the first,  $a_4$  appears in the second. In the difference, all terms  $A_iU \cdot A_jV$  in Eq. (4.2) cancel except those for which  $i$  or  $j = 4$ . Hence

$$f[A_1W \cdot A_2X \cdot A_3Y \cdot (A_4 - a_4)Z] = \sum_{U,V} \left\{ \sum_{i \neq 4} p(U, V; i, 4) A_iU \cdot (A_4 - a_4)V + \sum_{i \neq 4} p(U, V; 4, i) (A_4 - a_4)U \cdot A_iV + p(U, V; 4, 4) [A_4U \cdot A_4V - a_4U \cdot a_4V] \right\}$$

Expanding the last term using the identity

$$A_4U \cdot A_4V - a_4U \cdot a_4V = A_4U \cdot (A_4 - a_4)V + (A_4 - a_4)U \cdot A_4V - (A_4 - a_4)U \cdot (A_4 - a_4)V$$

and rearranging the final result we obtain an analogue of Eq. (3.3b). We record it as

**Proposition 4.2.**

$$f[A_1W \cdot A_2X \cdot A_3Y \cdot (A_4 - a_4)Z] = \sum_{U,V} \left\{ \sum_{i=1}^4 p(U, V; i, 4) A_iU \cdot (A_4 - a_4)V + \sum_{i=1}^4 p(U, V; 4, i) (A_4 - a_4)U \cdot A_iV - p(U, V; 4, 4) (A_4 - a_4)U \cdot (A_4 - a_4)V \right\} \tag{4.3}$$

The product either of two unit contrasts in locus  $A$  or of a gamete type and the linear combination  $(A_3 - a_3)Y \cdot (A_4 - a_4)Z$  of four gamete types, is

$$\begin{aligned} & f(A_1W \cdot A_2X \cdot (A_3 - a_3)Y \cdot (A_4 - a_4)Z) \\ &= f(A_1W \cdot A_2X \cdot A_3Y \cdot (A_4 - a_4)Z) \\ & \quad - f(A_1W \cdot A_2X \cdot a_3Y \cdot (A_4 - a_4)Z) \end{aligned} \tag{4.4}$$

Upon substituting Eq. (4.3) into Eq. (4.4), we obtain after cancellation of identical terms and formal factorization

**Proposition 4.3.**

$$\begin{aligned} & f[A_1W \cdot A_2X \cdot (A_3 - a_3)Y \cdot (A_4 - a_4)Z] \\ &= \sum_{U,V} \{ p(U, V; 3, 4) [(A_3 - a_3)U \cdot (A_4 - a_4)V] \\ & \quad + p(U, V; 4, 3) [(A_4 - a_4)U \cdot (A_3 - a_3)V] \} \end{aligned} \tag{4.5}$$

By further subtracting Eq. (4.5) (which contains no subscript  $i = 1$  or  $2$ ) from itself, we obtain

**Proposition 4.4.** *For any mode of gamete formation in an autotetraploid organism described by Eq. (4.2)*

$$\begin{aligned}
 f[A_1W \cdot (A_2 - a_2)X \cdot (A_3 - a_3)Y \cdot (A_4 - a_4)Z] &= 0 \\
 f[(A_1 - a_1)W \cdot (A_2 - a_2)X \cdot (A_3 - a_3)Y \cdot (A_4 - a_4)Z] &= 0 \quad (4.6)
 \end{aligned}$$

That is to say, the gamete distributions derived from these linear combinations of zygote genotypes vanish identically.

Proposition 4.2 to 4.4 give us expressions for all genetic products of elements of  $\mathfrak{G}$  having the forms

$$\begin{aligned}
 &A_1W \cdot A_2X \\
 &A_1W \cdot (A_2 - a_2)X \\
 &(A_1 - a_1)W \cdot A_2X \\
 &(A_1 - a_1)W \cdot (A_2 - a_2)X
 \end{aligned}$$

The noteworthy feature is that whenever the difference appears in either factor, it also appears in each term of the product. Also if the total number of  $(A_i - a_i)$  in the factors is 3 or 4, the result vanishes. The following argument deduces that  $\mathfrak{G}$  is thus a genetic algebra.

Let  $\mathfrak{C}^{(s)}$  be the set of linear combinations of right products  $((x_1 \times x_2) \times \dots) \times x_m$  of elements of  $\mathfrak{G}$  having at least  $s$  factors from  $\mathfrak{C}$ . Theorem 3.1, part 2, of Gonshor (1973) states that  $\mathfrak{G}$  is genetic if and only if  $\mathfrak{C}^{(s)} = 0$  for some  $s$ . We show that the  $n$  locus autotetraploid algebra satisfies this condition with  $s = 2n + 1$ .

**Theorem 4.5.** *The algebra of gamete multiplication for an autotetraploid organism with  $n$  loci is a genetic algebra for any mode of segregation of the loci.*

*Proof.* Consider the right gametic product

$$((x_1 \times x_2) \times \dots) \times x_m \quad (4.7)$$

of  $m$  elements  $x_i$  of  $\mathfrak{G}$  of which  $2n + 1$  or more are in  $\mathfrak{C}$ . Write each  $x_i$  contained in  $\mathfrak{C}$  as a linear combination of unit contrasts. Also, write each  $x_i$  not in  $\mathfrak{C}$  as a linear combination of gamete types. Using distributivity, we can expand Eq. (4.7) as a linear combination of right products

$$((y_1 \times y_2) \times \dots) \times y_m \quad (4.8)$$

where  $y_i$  is a unit contrast if  $x_i$  is in  $\mathfrak{C}$  and a gamete type if not. Thus, at least  $2n + 1$  factors of Eq. (4.8) are unit contrasts.

It suffices to show that Eq. (4.8) vanishes. Since there are only  $n$  loci, at least three in any set of  $2n + 1$  or more unit contrasts must be unit contrasts for some one locus. Supposing it to be  $A$ , we have that at least three  $y_i$  are unit contrasts in locus  $A$ . Again using distributivity, Eq. (4.8) can be further expanded in terms of right products of  $m$  factors  $z_i$  each of the form  $A_1W \cdot A_2X$  or  $A_1W \cdot (A_2 - a_2)X$ , where at least three factors are of the latter type. (Note that for these three,  $z_i = y_i$ ). Evaluating these terms using Propositions 4.2, 4.3, and 4.4, the expression vanishes.

*Acknowledgement.* We wish to thank the reviewers for their helpful comments.

## References

- Bennett, J. H.: Panmixia with tetrasomic and hexasomic inheritance. *Genetics* **39**, 150–158 (1954)
- Crow, J. F.: Random mating with linkage in polysomics. *Amer. Nat.* **88**, 431–434 (1954)
- Etherington, I. M. H.: Genetic algebras. *Proc. Roy. Soc. Edin.* **59**, 242–258 (1939)
- Fisher, R. A.: The theory of linkage in polysomic inheritance. *Philos. Trans. Roy. Soc. B* **233**, 55–87 (1947)
- Gonshor, H.: Special train algebras arising in genetics II. *Proc. Edinburgh Math. Soc.* **14**, 333–338 (1965)
- Gonshor, H.: Contributions to genetic algebras II. *Proc. Edinburgh Math. Soc.* **18**, 273–279 (1973)
- Heuch, I.: The genetic algebra for polyploidy with an arbitrary amount of double reduction. *J. Math. Biol.* **6**, 343–352 (1978)
- Holgate, P.: Genetic algebras associated with polyploidy. *Proc. Edinburgh Math. Soc.* **15**, 1–9 (1966)
- Holgate, P.: Sequences of powers in genetic algebras. *J. London Math. Soc.* **42**, 489–496 (1967)
- Wörz-Busekros, A.: Polyploidy with an arbitrary mixture of chromosome and chromatid segregation. *J. Math. Biol.* **6**, 353–365 (1978)

Received February 20/Revised November 1, 1979