

ALGEBRAIC STRUCTURE OF GENETIC INHERITANCE

MARY LYNN REED

ABSTRACT. In this paper we will explore the nonassociative algebraic structure that naturally occurs as genetic information gets passed down through the generations. While modern understanding of genetic inheritance initiated with the theories of Charles Darwin, it was the Augustinian monk Gregor Mendel who began to uncover the mathematical nature of the subject. In fact, the symbolism Mendel used to describe his first results (e.g., see his 1866 paper *Experiments in Plant-Hybridization* [30]) is quite algebraically suggestive. Seventy four years later, I.M.H. Etherington introduced the formal language of abstract algebra to the study of genetics in his series of seminal papers [9], [10], [11]. In this paper we will discuss the concepts of genetics that suggest the underlying algebraic structure of inheritance, and we will give a brief overview of the algebras which arise in genetics and some of their basic properties and relationships. With the popularity of biologically motivated mathematics continuing to rise, we offer this survey article as another example of the breadth of mathematics that has biological significance. The most comprehensive reference for the mathematical research done in this area (through 1980) is Wörz-Busekros [36].

1. GENETIC MOTIVATION

Before we discuss the mathematics of genetics, we need to acquaint ourselves with the necessary language from biology. A vague, but nevertheless informative, definition of a **gene** is simply a unit of hereditary information. The genetic code of an organism is carried on **chromosomes**. Each gene on a chromosome has different forms that it can take. These forms are called **alleles**. E.g., the gene which determines blood type in humans has three different alleles, A, B, and O. Since humans are **diploid** organisms (meaning we carry a double set of chromosomes – one from each parent), blood types are determined by two alleles. **Haploid** cells (or organisms) carry a single set of chromosomes. When diploid organisms reproduce, a process called **meiosis** produces **gametes** (sex cells) which carry a single set of chromosomes. When these gamete cells fuse (e.g., when sperm fertilizes egg), the result is a **zygote**, which is again a diploid cell, meaning it carries its hereditary information in a double set of chromosomes. When gametes fuse (or reproduce) to form zygotes a natural “multiplication” operation occurs.

1.1. Simple Mendelian Inheritance. As a natural first example, we consider simple Mendelian inheritance for a single gene with two alleles A and a . In this case, two gametes fusing (or reproducing) to form a zygote gives the multiplication table shown in Table 1, which in freshman biology class might be called a Punnett square.

Received by the editors August 1, 1996.

1991 *Mathematics Subject Classification.* Primary 17D92; Secondary 92-02.

©1997 American Mathematical Society

	A	a
A	AA	Aa
a	aA	aa

TABLE 1. Alleles passing from gametes to zygotes

	A	a
A	A	$\frac{1}{2}(A + a)$
a	$\frac{1}{2}(a + A)$	a

TABLE 2. Multiplication table of the gametic algebra for simple Mendelian inheritance

The zygotes AA and aa are called **homozygous**, since they carry two copies of the same allele. In this case, simple Mendelian inheritance means that there is no chance involved as to what genetic information will be inherited in the next generation; i.e., AA will pass on the allele A and aa will pass on a . However, the zygotes Aa and aA (which are equivalent) each carry two different alleles. These zygotes are called **heterozygous**. The rules of simple Mendelian inheritance indicate that the next generation will inherit either A or a with equal frequency. So, when two gametes reproduce, a multiplication is induced which indicates how the hereditary information will be passed down to the next generation. This multiplication is given by the following rules:

- (1) $A \times A = A,$
- (2) $A \times a = \frac{1}{2}A + \frac{1}{2}a,$
- (3) $a \times A = \frac{1}{2}a + \frac{1}{2}A,$
- (4) $a \times a = a.$

Rules (1) and (4) are expressions of the fact that if both gametes carry the same allele, then the offspring will inherit it. Rules (2) and (3) indicate that when gametes carrying A and a reproduce, half of the time the offspring will inherit A and the other half of the time it will inherit a . These rules are an algebraic representation of the rules of simple Mendelian inheritance. This multiplication table is shown in Table 2. We should point out that we are only concerning ourselves with **genotypes** (gene composition) and not **phenotypes** (gene expression). Hence we have made no mention of the dominant or recessive properties of our alleles.

Now that we've defined a multiplication on the symbols A and a we can mathematically define the two dimensional algebra over \mathbb{R} with basis $\{A, a\}$ and multiplication table as in Table 2. This algebra is called the *gametic algebra* for simple Mendelian inheritance with two alleles.

But gametic multiplication is just the beginning! In order for actual diploid cells (or organisms) to reproduce, they must first go through a reduction division process

	AA	Aa	aa
AA	AA	$\frac{1}{2}(AA + Aa)$	Aa
Aa	$\frac{1}{2}(AA + Aa)$	$\frac{1}{4}AA + \frac{1}{2}Aa + \frac{1}{4}aa$	$\frac{1}{2}(Aa + aa)$
aa	Aa	$\frac{1}{2}(Aa + aa)$	aa

TABLE 3. Multiplication table of the zygotic algebra for simple Mendelian inheritance

so that only one set of alleles is passed on. For humans this occurs when males produce sperm and females produce eggs. When reproduction occurs, the hereditary information is then passed on via the gametic multiplication we've already defined. Therefore, when two zygotes reproduce another multiplication operation is formed taking into consideration both the reduction division process and gametic multiplication. In our example of simple Mendelian inheritance for one gene with the two alleles A and a , zygotes have three possible genotypes: AA , aa , and Aa . Let's consider the case of two zygotes both with genotype Aa reproducing. The reduction division process splits the zygote and passes on one allele for reproduction. In the case of simple Mendelian inheritance the assumption is that both alleles will be passed on with equal frequency. Thus, half the time A gets passed on and half the time a does. We represent this with the "frequency distribution" $\frac{1}{2}A + \frac{1}{2}a$. Therefore, symbolically $Aa \times Aa$ becomes

$$\left(\frac{1}{2}A + \frac{1}{2}a\right) \times \left(\frac{1}{2}A + \frac{1}{2}a\right).$$

Formally multiplying these two expressions together results in

$$\frac{1}{4}AA + \frac{1}{2}Aa + \frac{1}{4}aa,$$

using the notion that $aA = Aa$. In this way, zygotic reproduction produces the multiplication table shown in Table 3. So we can define the three dimensional algebra over \mathbb{R} with basis $\{AA, Aa, aa\}$ and multiplication table as in Table 3. It is called the *zygotic algebra* for simple Mendelian inheritance with two alleles. The process of constructing a zygotic algebra from the original gametic algebra is called *commutative duplication of algebras*. We will discuss this process from a mathematical perspective later.

Now that we've seen how the gametic and zygotic algebras are formed in the most basic example, we shall begin to consider the mathematical (and indeed, algebraic) structure of such algebras.

1.2. The Nonassociativity of Inheritance. Depending on the "population" you are concerned with, a general element $\alpha A + \beta a$ of the gametic algebra which satisfies $\alpha, \beta \in \mathbb{R}$ with $0 \leq \alpha, \beta \leq 1$ and $\alpha + \beta = 1$ can represent a population, a single individual of a population, or a single gamete. In each case, the coefficients α and β signify the percentage of frequency of the associated allele. I.e., if the element represents a population, then α is the percentage of the population which carries

the allele A on the gene under consideration. Likewise, β is the percentage of the population which has the allele a .

For those elements of the gametic and zygotic algebras which represent populations, multiplication of two such elements represents random mating between the two populations. It seems logical that the order in which populations mate is significant. I.e., if population P mates with population Q and then the resulting population mates with R , the resulting population is not the same as the population resulting from P mating with the population obtained from mating Q and R originally. Symbolically, $(P \times Q) \times R$ is not equal to $P \times (Q \times R)$. So, we see that from a purely biological perspective, we should expect that the algebras which arise in genetics will not satisfy the associative property.

Now, if we study the multiplication tables of both the gametic and zygotic algebras for simple Mendelian inheritance, we will notice immediately that the algebras are commutative. From a biological perspective, if populations P and Q are mating, it makes no difference whether you say P mates with Q or Q mates with P ! However, as we should expect, these algebras do not satisfy the associative property. E.g., in the gametic algebra apply the rules of multiplication and the distributive property to see that $A \times (A \times a) = \frac{3}{4}A + \frac{1}{4}a$. However, $(A \times A) \times a = A \times a = \frac{1}{2}A + \frac{1}{2}a$. Hence, the associative property does not hold for the gametic algebra. The same is true for the zygotic algebra. In general, the algebras which arise in genetics are commutative but non-associative.

2. GAMETIC AND ZYGOTIC ALGEBRAS

So far, we've only considered the case of simple Mendelian inheritance; i.e., all alleles occur with equal frequency. In many genetic situations, this law does not hold. E.g., gene mutation or recombination both result in different inheritance rules. The gametic and zygotic algebras we discussed in the previous section corresponded to the very specific example of simple Mendelian inheritance for a single gene with two alleles. We now give more general definitions for gametic and zygotic algebras. Suppose now we have a random mating population with n distinct gametes. Call them a_1, \dots, a_n . These gametes could differ at one or more genetic loci. Then consider these n gametes as basis elements of an n -dimensional real vector space. Multiplication is defined by

$$a_i a_j = \sum_{k=1}^n \gamma_{ijk} a_k,$$

such that

$$(1) \quad 0 \leq \gamma_{ijk} \leq 1, \quad i, j, k = 1, \dots, n.$$

$$(2) \quad \sum_{k=1}^n \gamma_{ijk} = 1, \quad i, j = 1, \dots, n.$$

$$(3) \quad \gamma_{ijk} = \gamma_{jik} \quad i, j, k = 1, \dots, n.$$

The resulting algebra \mathfrak{G} is called an n -dimensional *gametic algebra*.

For the zygotic algebra we consider pairs of the n gametes, $a_{ij} = a_i a_j$ with the understanding that $a_{ij} = a_{ji}$, so without loss we only consider a_{ij} with $i \leq j$. Then random mating of zygotes a_{ij} and a_{pq} will yield zygotes a_{ks} with a certain probability; call it $\gamma_{ij,pq,ks}$. This defines zygotic multiplication,

$$a_{ij} a_{pq} = \sum_{k \leq s} \gamma_{ij,pq,ks} a_{ks},$$

such that

$$\begin{aligned} (1) \quad & 0 \leq \gamma_{ij,pq,ks} \leq 1, \\ (2) \quad & \sum_{k,s=1}^n \gamma_{ij,pq,ks} = 1, \\ (3) \quad & \gamma_{ij,pq,ks} = \gamma_{pq,ij,ks}, \end{aligned}$$

where in each case $i \leq j$, $p \leq q$, and $k \leq s$. The resulting algebra \mathfrak{Z} is the *zygotic algebra*. We note that the zygotic algebra can be constructed from the gametic algebra through a process called commutative duplication [10], which was originally introduced by Etherington in the general setting of a (not necessarily commutative nor associative) linear algebra. Using this process, one can calculate the zygotic multiplication constants from the gametic multiplication constants in the following way:

$$\gamma_{ij,pq,ks} = \begin{cases} \gamma_{ijk}\gamma_{pqs} + \gamma_{ijs}\gamma_{pqk}, & \text{for } k < s, \\ \gamma_{ijk}\gamma_{pqs}, & \text{for } k = s. \end{cases}$$

In modern terms, commutative duplication can be realized using tensor products. For any commutative algebra A , tensor it with itself (in the sense of vector spaces) to form $A \otimes A$. Then, commutative duplication can be achieved via the quotient $(A \otimes A)/I$, where I is the subspace generated by elements of the form $x \otimes y - y \otimes x$. This quotient space is, in fact, a commutative algebra, where multiplication is defined by $(a, b)(c, d) = (ab, cd)$. H. Gonsior [14] first gave this as a basis-free definition of commutative duplication of an algebra.

In addition, beginning with a zygotic algebra \mathfrak{Z} , commutative duplication produces another algebra \mathfrak{C} with genetic relevance, which is generally referred to as the *copular algebra*. The genetic significance of this algebra is that its elements, which are unordered pairs of zygotes, represent the mating types of a population.

3. STRUCTURE AND PROPERTIES OF ALGEBRAS IN GENETICS

Mathematically, the algebras that arise in genetics (via gametic, zygotic, or copular algebras) are very interesting structures. They are generally commutative but nonassociative, yet they are not necessarily Lie, Jordan, or alternative algebras. In addition, many of the algebraic properties of these structures have genetic significance. Indeed, it is the interplay between the purely mathematical structure and the corresponding genetic properties that makes this subject so fascinating. We turn our attention now from the motivating genetics to the more formal mathematical study of the underlying algebraic structure. Again, we draw attention to the comprehensive reference Wörz-Busekros [36].

The most general definition of an algebra which could have genetic significance is that of an *algebra with genetic realization*. An algebra with genetic realization is an algebra A over the real numbers \mathbb{R} which has a basis $\{a_1, \dots, a_n\}$ and a multiplication table

$$a_i a_j = \sum_{k=1}^n \gamma_{ijk} a_k,$$

such that $0 \leq \gamma_{ijk} \leq 1$ for all i, j, k and

$$\sum_{k=1}^n \gamma_{ijk} = 1,$$

for $i, j = 1, \dots, n$. Such a basis is called the *natural basis* for A .

It is easy to see that our earlier examples of gametic and zygotic algebras for simple Mendelian inheritance, as well as the general gametic and zygotic algebras, are all algebras with genetic realization. In a general algebra A with genetic realization, an element x in A represents a population, or a gene pool for a population, if its expression as a linear combination of the basis elements a_1, \dots, a_n ,

$$x = \xi_1 a_1 + \xi_2 a_2 + \dots + \xi_n a_n,$$

satisfies $0 \leq \xi_i \leq 1$ for all $i = 1, \dots, n$ and $\sum_{i=1}^n \xi_i = 1$. Then ξ_i is the percentage of the population x which carries the allele a_i .

The class of all algebras with genetic realization is too large to say much about. However, since all gametic algebras (and their commutative duplicates) satisfy the definition, it provides a solid framework for what constitutes an algebra with genetic significance.

3.1. Baric Algebras. For strictly mathematical purposes, it is unnecessary to restrict the field our algebras are defined over to be the real numbers. Hence, we will work over an arbitrary field k when appropriate. As we have seen, algebras with genetic realization are not necessarily associative algebras. However, they do belong to a rather special class of nonassociative algebras. A general nonassociative algebra need not possess a matrix representation. Yet, algebras with genetic realization do. In fact, they possess the simplest possible matrix representation – a scalar representation.

An algebra A over a field k is called a *baric algebra* if it admits a non-trivial algebra homomorphism $\omega: A \rightarrow k$. In other words, a baric algebra is an algebra with a one-dimensional representation. The homomorphism ω is called the *weight function* (or *baric function*).

Proposition 3.1. *Let A be an n -dimensional algebra with genetic realization over \mathbb{R} . Then A is a baric algebra.*

Proof. Let $\{a_1, \dots, a_n\}$ denote a natural basis for A . Define $\omega: A \rightarrow \mathbb{R}$ by $\omega(a_i) = 1$ for $i = 1, \dots, n$ and then extend linearly onto A . I.e., if $x = \sum_{i=1}^n \xi_i a_i$, then $\omega(x) = \sum_{i=1}^n \xi_i \omega(a_i)$. Hence, $\omega(x) = \sum_{i=1}^n \xi_i$. Then we need only show that ω is a homomorphism.

Let $x = \sum_{i=1}^n \xi_i a_i$ and $y = \sum_{j=1}^n \beta_j a_j$. Then

$$\begin{aligned} xy &= \sum_{i=1}^n \xi_i a_i \sum_{j=1}^n \beta_j a_j \\ &= \sum_{i=1}^n \xi_i \left(\sum_{j=1}^n \beta_j a_i a_j \right) \\ &= \sum_{i=1}^n \xi_i \left(\sum_{j=1}^n \beta_j \left(\sum_{k=1}^n \gamma_{ijk} a_k \right) \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \xi_i \beta_j \gamma_{ijk} a_k. \end{aligned}$$

Then apply ω to get that

$$\begin{aligned} \omega(xy) &= \sum_{k=1}^n \left(\sum_{i=1}^n \xi_i \right) \left(\sum_{j=1}^n \beta_j \right) \gamma_{ijk} \\ &= \left(\sum_{i=1}^n \xi_i \right) \left(\sum_{j=1}^n \beta_j \right) \sum_{k=1}^n \gamma_{ijk} \\ &= \left(\sum_{i=1}^n \xi_i \right) \left(\sum_{j=1}^n \beta_j \right), \end{aligned}$$

since $\sum_{k=1}^n \gamma_{ijk} = 1$. But, then $\omega(xy) = \omega(x)\omega(y)$. Therefore, ω is a homomorphism and A is a baric algebra. \square

From a strictly mathematical perspective, an interesting question to ask about baric algebras is whether or not their weight functions are uniquely determined. The following example (as found in [36]) shows that in general, they are not.

Example 3.1. Let $A = \langle a_1, a_2, a_3 \rangle_{\mathbb{R}}$ be a commutative 3-dimensional algebra with the multiplication table below.

	a_1	a_2	a_3
a_1	$a_1 + a_2$	a_2	a_2
a_2	a_2	a_2	a_2
a_3	a_2	a_2	$a_2 + a_3$

Then define $\omega_1: A \rightarrow \mathbb{R}$ via $\omega_1(a_1) = 1$ and $\omega_1(a_2) = \omega_1(a_3) = 0$. And define $\omega_2: A \rightarrow \mathbb{R}$ via $\omega_2(a_3) = 1$, while $\omega_2(a_1) = \omega_2(a_2) = 0$. It is easy to see that $\omega_1 \neq \omega_2$, and it is a simple verification that they both define homomorphisms.

Even though the above example shows that not all baric algebras have a unique weight function, many of them do. In order to exhibit at least a sufficient condition for a baric algebra to have a unique weight function, we must first discuss the issue of powers in a nonassociative algebra.

3.2. Principal and Plenary Powers. In a commutative, nonassociative algebra, there are several ways to define and interpret the powers of an element. There are two main types of powers which have genetic significance. Let x be an element of a commutative nonassociative algebra A . The *principal powers* are defined to be x, x^2, x^3, \dots , where $x^i = x^{i-1}x$. If A is an algebra with genetic realization and an element P represents a population, then each element P^i of the sequence of principal powers represents a population which resulted from the previous population P^{i-1} mating back with the original population P . On the other hand, the *plenary powers* $x, x^{[2]}, x^{[3]}, \dots$ are defined as $x^{[i]} = x^{[i-1]}x^{[i-1]}$. When P is an element representing a population, the sequence of plenary powers contains the successive generations resulting from random mating within the population, beginning with P . I.e., $P^{[2]}$ is the result of the population P mating within itself, and $P^{[3]}$ is the result of the

population $P^{[2]}$ mating within itself. Both the principal and plenary powers are of biological as well as mathematical interest.

The following proposition provides a sufficient condition for a baric algebra to have a unique weight function. Our proof follows [36].

Proposition 3.2. *Let A be a baric algebra over a field k with weight function ω . If $N = \text{Ker } \omega$ is nil (i.e., all elements of N are nilpotent), then ω is uniquely determined.*

Proof. Let $\varphi: A \rightarrow k$ be any non-trivial homomorphism. Then our hypothesis is that every $x \in N$ is nilpotent, meaning there exists a positive integer r such that the principal power $x^r = 0$. Then $\varphi(x^r) = \varphi(0) = 0$. So, $\varphi(x)^r = 0$, and since $\varphi(x) \in k$, this implies $\varphi(x) = 0$.

Then let $y \in A \setminus N$. So $\omega(y) \neq 0$. Then consider the element $\frac{y^2}{\omega(y)} - y$. Since ω is a homomorphism, it is easy to see that $\frac{y^2}{\omega(y)} - y \in N$. But by our above argument, $\varphi(\frac{y^2}{\omega(y)} - y) = 0$. And this leads to $\varphi(y) \left(\frac{\varphi(y)}{\omega(y)} - 1 \right) = 0$. Thus, either $\varphi(y) = 0$ or $\varphi(y) = \omega(y)$. If $\varphi(y) = 0$, then we've shown $\varphi = 0$, which contradicts the assumption that φ is non-trivial.

Hence, $\varphi(y) = \omega(y)$ for every $y \in A \setminus N$ and $\varphi(x) = 0$ for every $x \in N$. Therefore, $\varphi = \omega$ and the baric function is unique. \square

3.3. Idempotents. A non-zero element e in an algebra which satisfies the relationship $e^2 = e$ is called an *idempotent*. In addition to their mathematical importance, idempotents also have genetic significance. If a population P satisfies the equation $P^2 = P$, this means that genetic equilibrium has been achieved after one generation of random mating within the population P . I.e., the population P^2 has the same genetic pool as the initial population P .

Mathematically, the existence of an idempotent in an algebra provides a direct sum decomposition of the algebra. Hence, idempotents play a crucial role in describing the general structure of an algebra. Let A be a baric algebra over a field k . The existence of an idempotent in A is not guaranteed. However, if A does contain an idempotent e , then $\omega(e) = \omega(e^2) = \omega(e)\omega(e)$, so either $\omega(e) = 0$ or $\omega(e) = 1$.

Proposition 3.3. *Let A be a baric algebra over a field k with weight function ω . Suppose A contains an idempotent e such that $\omega(e) = 1$. Then,*

$$A = ke \oplus \text{Ker } \omega.$$

Proof. Let $N = \text{Ker } \omega = \{x \in A \mid \omega(x) = 0\}$. By the first homomorphism theorem, $A/N \cong k$. N is a two sided ideal with codimension one in A . Since $\omega(e) = 1$, $ke \cap N = 0$.

Let x be an arbitrary element of A . Then $x - \omega(x)e$ is in N , since $\omega(e) = 1$. And

$$x = \omega(x)e + (x - \omega(x)e).$$

Hence, $A = ke \oplus N$. \square

3.4. Train Algebras. For any element x in an associative algebra, there is a normalized polynomial m_x which annihilates x and has minimal degree. This is called the minimal polynomial of x . For a nonassociative algebra A , a little care must be taken to define polynomials in terms of principal powers (see [36]), but for every element x there still exists a minimal polynomial m_x .

Let A denote a finite dimensional, commutative nonassociative algebra over a field k . Let $\{a_1, \dots, a_n\}$ be a basis of A . It turns out that not only do minimal polynomials of single elements exist, but there also exists a polynomial in principal powers that annihilates all elements of A . The polynomial of minimal degree which does this is called the *rank polynomial* of A . Denote it

$$(5) \quad f(x) = x^r + \theta_1 x^{r-1} + \theta_2 x^{r-2} + \dots + \theta_{r-1} x,$$

where θ_p is a homogeneous polynomial of degree p in the coordinates ξ_i of the generic element $x = \sum_{i=1}^n \xi_i a_i$.

For many of the algebras which arise in genetics, the polynomials θ_p depend only on the baric weight function values of the ξ_i and not on the coefficients ξ_i themselves. The algebras that satisfy this property are known as *train algebras*. While we have defined the rank polynomial from a purely mathematical point of view, it does have genetic relevance. Since the rank polynomial annihilates all elements of the algebra, by setting the polynomial equal to zero we obtain an equation (called the *rank equation*) which gives a recurrence relation among the principal powers. If the rank polynomial is of degree r , then this recurrence relation indicates how the genetic pool of the r -th generation (in the sequence of principal powers) is related to the genetic pools of previous generations.

Definition 3.1. Let A be a baric algebra with weight function ω and rank polynomial (5). A is called a *train algebra of rank r* if the coefficients θ_p of the rank polynomial of A are functions of $\omega(x)$.

Let A be a train algebra and suppose L is an extension of k which contains the splitting field of the rank polynomial f . Then in L , f splits into linear factors

$$f(x) = x(x - \lambda_0 \omega(x))(x - \lambda_1 \omega(x)) \cdots (x - \lambda_{r-2} \omega(x)).$$

The elements $\lambda_0, \lambda_1, \dots, \lambda_{r-2}$ of L are called the *principal train roots* of A . It is easy to see that one of the principal train roots of A must be 1. (Apply the rank polynomial to an element of weight 1 and then apply ω to the resulting polynomial equation.)

The following are easy observations in the light of these definitions.

Proposition 3.4. Let A be a train algebra of rank r with weight function $\omega: A \rightarrow k$. Then every element in $N = \text{Ker } \omega$ is nilpotent of an index not greater than r .

Proof. Since the θ_p of (5) are homogeneous polynomials of degree p and they are functions of $\omega(x)$, then $\theta_p = \beta_p \omega^p(x)$, where $\beta_p \in k$. Hence, the rank polynomial of A can be written

$$f(x) = x^r + \beta_1 \omega(x) x^{r-1} + \beta_2 \omega^2(x) x^{r-2} + \dots + \beta_{r-1} \omega^{r-1}(x) x.$$

And we know this polynomial annihilates every element x of A . Hence if $x \in N$, then $\omega(x) = 0$, so the rank equation $f(x) = 0$ becomes $x^r = 0$. Therefore, x is nilpotent of index less than or equal to r . \square

Corollary 3.1. A train algebra has a uniquely determined weight function.

Proof. Let A be a train algebra with weight function ω . By Proposition 3.4, $N = \text{Ker } \omega$ is nil. Then Proposition 3.2 implies ω is unique. \square

In Proposition 3.4, we proved that the subalgebra $N = \text{Ker } \omega$ of a train algebra A is always nil, meaning all of its elements are nilpotent. A subalgebra B is said

to be *nilpotent* if there exists a power of the subalgebra B^m which is zero. This is a stronger condition than having all elements in the subalgebra be nilpotent. For a general train algebra, the subalgebra N is not always nilpotent. Mathematically, this stronger condition motivates the following definition.

Definition 3.2. A baric algebra A with weight function ω is called a *special train algebra* if $N = \text{Ker } \omega$ is nilpotent and all of the principal power subalgebras N^i of A defined inductively by $N^1 = N, N^i = N^{i-1}N$ for $i = 2, 3, \dots$, are ideals of A .

Notice that this definition implies that the sequence of descending ideals

$$A \supset N = N^1 \supset N^2 \supset N^3 \supset \dots \supset N^r \supset N^{r+1} = \langle 0 \rangle,$$

terminates.

The special train algebra was Etherington's first attempt to specialize the subject of "genetic algebras" down to a manageable size. The goal was to find a definition that would be general enough to encompass most of the algebras which arise in genetics, but specific enough to study the intricate details of the structure mathematically. However, in [10], Etherington pointed out the fatal flaw of his own definition. Although the fundamental gametic algebras are special train algebras, their commutative duplicates (the zygotic algebras) may fail to be special train algebras. Hence, a more general definition was still needed.

3.5. Genetic Algebras. R. D. Schafer [32], less than a decade after Etherington's first papers appeared, was the first to provide a candidate for the definition of a *genetic algebra*.

Before stating the definition we should say a few words about transformation algebras. Suppose A is any commutative, nonassociative algebra over a field k . For $x \in A$, let $R_x: A \rightarrow A$ indicate right multiplication by x . Since A is commutative, right multiplication by x is equivalent to left multiplication by x . Then the *transformation algebra* $T(A)$ of A is the algebra consisting of all polynomials of right multiplication operators on A , as well as the identity operator on A . Hence any transformation T in the transformation algebra of A can be represented as $T = \alpha I + f(R_{x_1}, \dots, R_{x_s})$, where $\alpha \in k$, I is the identity on A , and f is a polynomial.

Definition 3.3 (Schafer [32]). Let A be a commutative baric algebra over a field k with weight function ω . Then A is a *genetic algebra* if the characteristic function $|\lambda I - T|$ of any transformation $T = \alpha I + f(R_{x_1}, \dots, R_{x_s})$, in the transformation algebra of A , is a function of $\omega(x_1), \dots, \omega(x_s)$.

It should be noted that, unlike Etherington, Schafer studied these algebras purely for their mathematical interest and, in fact, specifically stated in [32] that he could offer no insight into their genetic significance. However, by applying the above definition to the transformation $T = R_x$, Schafer [32] proved that all of his genetic algebras are also train algebras. In the same paper he also proved that every special train algebra is a genetic algebra and that genetic algebras are closed under commutative duplication. These facts alone signify that Schafer's notion of a genetic algebra is worthy of its name. However, two decades later, P. Holgate [24] demonstrated that Schafer's definition did, in fact, have relevance genetically. Using tools from Lie algebra theory, Holgate provided some alternative characterizations of Schafer's genetic algebras which illuminated their genetic significance. It turns out that the algebras which naturally occur in genetics will satisfy Schafer's condition

as long as there is a certain amount of symmetry in the genetic system. And as we saw in our simple Mendelian inheritance examples, the most basic genetic systems do contain a substantial amount of symmetry.

Although Schafer's definition succeeded in filling the gap between train algebras and special train algebras, the impracticality of verifying the definition was still a problem. For a given genetic situation, it is, in general, quite difficult to verify that its gametic (or zygotic) algebra satisfies Schafer's condition on the characteristic function of *every* transformation in the transformation algebra. For this reason, another definition was sought. In 1971, twenty-two years after Schafer's original paper [32] appeared, H. Gonshor [14] gave a definition which is equivalent to Schafer's, but has the benefit of utility.

Definition 3.4 (Gonshor [14]). Let A be a commutative finite dimensional algebra. Then A is a *genetic algebra* if there exists a basis $\{a_0, a_1, \dots, a_n\}$ with multiplication table

$$a_i a_j = \sum_{k=0}^n \lambda_{ijk} a_k,$$

such that the multiplication constants satisfy the following:

$$\begin{aligned} \lambda_{000} &= 1, \\ \lambda_{0jk} &= 0, \quad \text{for } k < j, \\ \lambda_{ijk} &= 0, \quad \text{for } i, j > 0 \text{ and } k \leq \max(i, j). \end{aligned}$$

Such a basis is called a *canonical basis* of A .

In Section 5 we will encounter a few of the genetic applications that benefit from the verifiability of Gonshor's definition.

Up to this point, our definitions all fall nicely into a linear implication diagram (see Figure 1). Yet there are other classes of algebras with genetic relevance which do not fit in between any of our previous definitions. Those are the Bernstein and k -th order Bernstein algebras which we'll consider in the next section.

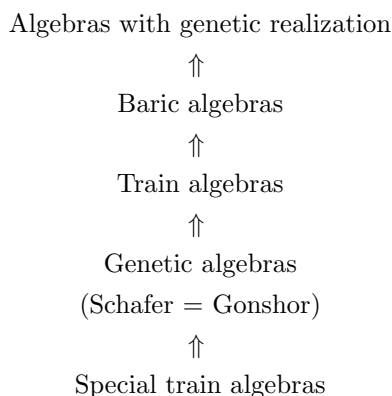


FIGURE 1. Implication diagram for nonassociative algebras with genetic significance

4. BERNSTEIN AND k -TH ORDER BERNSTEIN ALGEBRAS

Etherington's [9] definition of a train algebra specifies a condition on the sequence of principal powers of a baric algebra. From a mathematical point of view, the principal powers are much more tractable than their cousins, the plenary powers. However, from a geneticist's point of view, the sequence of plenary powers is of greater interest since the plenary powers more accurately model the way most populations reproduce. This is the viewpoint that motivates the classes of Bernstein and k -th order Bernstein algebras. In the early 1920's, S. Bernstein [2], [3], [4] studied a quadratic evolutionary operator Ψ , which mapped the simplex of genetic frequency distributions

$$\Delta_n = \{ (x_0, \dots, x_n) \mid x_i > 0, \sum x_i = 1 \}$$

into itself and represented the passage of one generation to the next. Bernstein set out to classify all such operators which satisfied the condition $\Psi^2 = \Psi$. This condition later came to be known as Bernstein's stationarity principle. The condition $\Psi^2 = \Psi$ indicates that the population is in equilibrium after one generation. Evolutionary operators satisfying this condition are idempotent evolutionary operators. After Bernstein's initial results (which included a complete classification for $n = 2$), the problem was later studied by Lyubich [27] and then, in 1975, by Philip Holgate [25].

Holgate's approach began by extending the quadratic evolutionary operator Ψ linearly to an entire real vector space V with basis $\{a_0, \dots, a_n\}$. Then by defining a multiplication on V via

$$xy = \frac{1}{2}\{\Psi(x+y) - \Psi(x) - \Psi(y)\},$$

V becomes a commutative algebra over \mathbb{R} . Then we define a map $\omega: V \rightarrow \mathbb{R}$ by setting $\omega(a_i) = 1$ for all $0 \leq i \leq n$ and extending this definition linearly onto the vector space V . Recalling that Bernstein's evolutionary operator is quadratic, it follows that $\omega(\Psi(x)) = \omega^2(x)$ for all $x \in V$. From this property and the linearity of ω , one can directly calculate that ω is a homomorphism. I.e.,

$$\begin{aligned} \omega(xy) &= \omega\left(\frac{1}{2}\{\Psi(x+y) - \Psi(x) - \Psi(y)\}\right), \\ &= \frac{1}{2}\{\omega(\Psi(x+y)) - \omega(\Psi(x)) - \omega(\Psi(y))\}, \\ &= \frac{1}{2}\{\omega^2(x+y) - \omega^2(x) - \omega^2(y)\}, \\ &= \frac{1}{2}\{(\omega(x) + \omega(y))^2 - \omega^2(x) - \omega^2(y)\}, \\ &= \omega(x)\omega(y). \end{aligned}$$

And hence V is a baric algebra with weight function ω .

On the entire vector space V , Bernstein's principle becomes $\Psi^2(x) = \omega^2(x)\Psi(x)$. And recalling that the operator $\Psi(x)$ represents the genetic evolution of a population x to the next generation, i.e., the first plenary power $x^{[2]}$, we obtain Holgate's definition of a Bernstein algebra.

Definition 4.1. Let A be a finite-dimensional, commutative baric algebra over a field k (with characteristic of k not equal to 2), and let ω denote the weight function.

Then A is called a *Bernstein algebra* if the plenary powers of any element $x \in A$ satisfy

$$x^{[3]} = \omega^2(x)x^{[2]}.$$

We note that Holgate's original definition was for a baric algebra over the complex numbers, but today, the general definition above (using a field k of characteristic not 2) is more commonly given.

Since a Bernstein algebra is a baric algebra with a condition on the plenary powers, whereas a train algebra required a condition on the principal powers, this new class of algebras does not fit nicely between any of our previous definitions. However, as Holgate proved in his seminal paper [25], certain Bernstein algebras do contain subalgebras which are special train algebras. In addition, it would seem that the class of Bernstein algebras, perhaps more so than any of the other classes discussed so far, provides a fertile meeting ground between the mathematics and the genetics. From Bernstein's original problem and Holgate's development of the definition of a Bernstein algebra, the genetic relevance is quite clear. Those elements of a Bernstein algebra with baric weight 1 satisfy the equation $x^{[3]} = x^{[2]}$; i.e., they reach genetic equilibrium after one generation of random mating within the population. Therefore, complete knowledge of the structure of these abstract nonassociative algebras would provide a great deal of genetically significant information. On the other hand, mathematically speaking, the challenge to classify these algebras is a rich open problem. Here is a small sample of what is known about the structure of a general Bernstein algebra.

If A is a Bernstein algebra, then it must contain at least one idempotent element. Just let x be some element of A with baric weight 1 (i.e., $\omega(x) = 1$) and take $e = x^2$. Then the definition of a Bernstein algebra implies that $e^2 = e$. Now, let $Z = \text{Ker } \omega$. Then Holgate [25] showed that Z^2 is an ideal of A , and so the following proposition holds.

Proposition 4.1 (Holgate [25]). *Let A be a Bernstein algebra with weight function ω and $Z = \text{Ker } \omega$. If $Z^3 = 0$, then A is a special train algebra.*

Mathematically, one of the most useful facts about the structure of a Bernstein algebra is the direct sum decomposition it possesses. Without proof, we offer the result below.

Proposition 4.2 (Holgate [25]). *Let A be a Bernstein algebra with weight function ω . Let e denote an idempotent of A . Let $Z = \text{Ker } \omega$. Let $U = \{x \in Z \mid ex = \frac{1}{2}x\}$ and $V = \{x \in Z \mid ex = 0\}$. Then*

$$A = ke \oplus U \oplus V.$$

Another remarkable (and extremely useful) fact about a Bernstein algebra is the classification of all of its idempotents.

Proposition 4.3 (Holgate [25]). *Let A be a Bernstein algebra and let e denote one idempotent in A . Then all idempotents of A have the form $e + u + u^2$, where $u \in U$ and $U = \{x \in Z \mid ex = \frac{1}{2}x\}$.*

There is much more known about the structure of Bernstein algebras, and we hope this brief sampling has enticed the reader to learn more.

A very natural generalization of a Bernstein algebra arises when we alter the condition on the plenary powers to reflect not equilibrium after one generation of

mating, but instead after k generations of mating. The resulting class of algebras is called the k -th order Bernstein algebras. The definition of a k -th order Bernstein algebra was initially given by Abraham [1], and a good reference for the basic properties they possess is Hentzel, Peresi, and Holgate [21]. We will just state the definition here.

Definition 4.2. Let A be a finite-dimensional, commutative baric algebra over a field K (with characteristic of K not equal to 2), and let ω denote the weight function. Then A is called a *k -th order Bernstein algebra* if the plenary powers of any element $x \in A$ satisfy

$$x^{[k+2]} = \omega(x)^{2^k} x^{[k+1]}.$$

5. APPLICATIONS

Now that we have outlined the basic structure of the nonassociative algebras which arise from the process of genetic inheritance, we turn our attention to some actual genetic applications. How does our knowledge of the algebraic structure apply to real genetic situations?

In [11], Etherington observed that geneticists already used some of these basic algebraic notions without explicitly recognizing them. And so he set out to formalize those algebraic notions that geneticists were already utilizing and to hopefully, via abstract algebra, advance the symbolism to a point where it could simplify the geneticists' methods of dealing with certain problems.

The branch of genetics we are concerned with is called population genetics, i.e., the study of how populations evolve through the generations. For a given mating system, the ultimate goal of population genetics is to determine the distribution of genetic types in the n -th filial generation based on the genetic types present in the original population as well as to determine the equilibrium distribution, if one exists.

In the examples which follow we will discuss the two major types of application of the genetic algebra theory to genetics: (1) using the algebra as formalization of traditional methods and (2) using the algebra to advance the study of genetics.

5.1. Linearization in Genetic Algebras. The first application we will consider is actually an important result in genetic algebra theory, but we consider it as an application because it illustrates the way the algebraic theory can contribute directly to genetics. As we mentioned in Section 4, the evolutionary operator which represents the passage of genetic frequencies from one generation to the next is a quadratic operator — not a linear one. Hence, matrix methods are not immediately useful to population geneticists. However, in 1930, a geneticist named Haldane [20] described a procedure which, in some cases, enabled this quadratic evolutionary operator to be represented by a linear transformation (on a higher dimensional space). Following up on this idea and implementing the mathematical theory, Holgate [22] proved that for any mating system that forms a genetic algebra, the quadratic evolutionary operator can be represented by a linear transformation on a higher dimensional space. There is also a systematic procedure for calculating the matrix associated to this linear transformation.

Suppose A is a genetic algebra with weight function $\omega: A \rightarrow \mathbb{R}$. Let $E: A \rightarrow A$, which sends a population p to p^2 , be the evolution map. If the population mates randomly within its generation, successive generations are given by the plenary

powers $p^{[n]} = p^{[n-1]}p^{[n-1]}$. The main application problem in population genetics is to determine these successive generations in terms of an initial population, i.e., to find an equation (or recurrence relation) which determines the n -th plenary power from the initial population. After proving that this linearization process (sometimes referred to as Haldane linearization) works for all genetic algebras, Holgate [22] used the process to prove that if a genetic algebra contains an idempotent element, then the plenary powers satisfy a train equation (i.e., a recurrence relation). Hence, the linearization process is a crucial tool in the algebraic theory for extracting genetically significant information from the algebras that model genetic inheritance systems.

We now describe the basic scenario of the Haldane linearization process. Let H denote the set $\omega^{-1}(1)$ of elements in A of baric weight one. The set H contains all the population states in the genetic algebra. Written in algebraic language, the Haldane linearization process says that we can imbed H into a finite dimensional vector space V such that the following diagram commutes:

$$\begin{array}{ccc} H & \xrightarrow{\phi} & V \\ \downarrow E & & \downarrow L \\ \tilde{H} & \xrightarrow{\phi} & \tilde{V} \end{array}$$

where $\phi: H \rightarrow V$ is the imbedding, $E: H \rightarrow H$ is the quadratic evolutionary map, and $L: V \rightarrow V$ is a linear map. After Holgate [22] proved this result for genetic algebras, McHale and Ringwood [29] found necessary and sufficient conditions for the result to hold in a general baric algebra.

We now move on to consider three different genetic inheritance systems and some of the algebraic theory that can be applied in each case. The three examples, self fertilization, autopolyploidy, and sex linked inheritance, give a representative sampling of how the algebraic theory is used to model and study real genetic inheritance systems.

5.2. Self Fertilization. In [11], Etherington considered several different mating systems, including self fertilization, and used the algebraic formalization of the corresponding inheritance rules in each situation to compute the distribution of the n -th filial generation and the equilibrium distribution. We will illustrate this type of application here in the case of self fertilization.

Self fertilization occurs mainly in the plant world. The scenario we will consider is the pure selfing of diploids at one genetic locus. I.e., we'll start with three zygotes AA , Aa , and aa and the zygotic distribution

$$P = \alpha AA + \beta Aa + \gamma aa,$$

where α, β , and γ are the proportions of the population with zygotes AA , Aa , and aa , respectively. Hence, $\alpha + \beta + \gamma = 1$. So in terms of algebras, this population P is an element of the zygotic algebra for simple Mendelian inheritance with two alleles. However, we do not want to study what happens when P mates randomly with other populations. Instead we will focus on the resulting filial generations after repeated self fertilization, starting with P . Using the rules of the zygotic algebra

(found in Table 3), the first filial generation F_1 after self fertilization will be

$$\begin{aligned} F_1 &= \alpha(AA \times AA) + \beta(Aa \times Aa) + \gamma(aa \times aa) \\ &= \alpha AA + \beta \left(\frac{1}{4} AA + \frac{1}{2} Aa + \frac{1}{4} aa \right) + \gamma aa \\ &= \left(\alpha + \frac{1}{4} \beta \right) AA + \frac{1}{2} \beta Aa + \left(\frac{1}{4} \beta + \gamma \right) aa. \end{aligned}$$

The first question we would like to answer is: What is the n -th filial generation

$$F_n = \alpha_n AA + \beta_n Aa + \gamma_n aa$$

through self fertilization? We'll follow Etherington's clever method, as found in [11]. Let u_n denote the population increase from the F_{n-1} generation to the n -th filial generation F_n . So, $u_1 = F_1 - P$, $u_2 = F_2 - F_1$, etc. Using the F_1 distribution we calculated above, we see that

$$\begin{aligned} u_1 &= \frac{1}{4} \beta AA - \frac{1}{2} \beta Aa + \frac{1}{4} \beta aa \\ &= \frac{1}{2} \beta \left(\frac{1}{2} AA - Aa + \frac{1}{2} aa \right). \end{aligned}$$

Again using the rules of the zygotic algebra from Table 3, we calculate the second filial generation $F_2 = F_1 \times F_1$, which is

$$F_2 = \left(\alpha + \frac{3}{8} \beta \right) AA + \frac{1}{4} \beta Aa + \left(\frac{3}{8} \beta + \gamma \right) aa.$$

Then,

$$\begin{aligned} u_2 &= \frac{1}{8} AA - \frac{1}{4} \beta Aa + \frac{1}{8} \beta aa \\ &= \frac{1}{4} \beta \left(\frac{1}{2} AA - Aa + \frac{1}{2} aa \right). \end{aligned}$$

If we continued to calculate in this fashion, we would find that

$$u_n = \frac{1}{2^n} \beta \left(\frac{1}{2} AA - Aa + \frac{1}{2} aa \right).$$

And then the total increase in n generations will be

$$\begin{aligned} u_1 + u_2 + u_3 + \cdots + u_n &= \left(\frac{1}{2} AA - Aa + \frac{1}{2} aa \right) \left(\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} \right) \beta \\ &= \beta \left(1 - \frac{1}{2^n} \right) \left(\frac{1}{2} AA - Aa + \frac{1}{2} aa \right). \end{aligned}$$

Hence,

$$\begin{aligned} F_n &= \alpha AA + \beta Aa + \gamma aa + u_n \\ &= \left(\alpha + \frac{1}{2} \beta - \frac{1}{2^{n+1}} \beta \right) AA + \frac{1}{2^n} \beta Aa + \left(\frac{1}{2} \beta + \gamma - \frac{1}{2^{n+1}} \beta \right) aa. \end{aligned}$$

From here it is easy to identify the equilibrium distribution. Examining F_n we see that as n increases, F_n approaches the equilibrium distribution of $(\alpha + \frac{1}{2} \beta) AA + (\gamma + \frac{1}{2} \beta) aa$. Notice that in the equilibrium distribution the heterozygote Aa does not appear. Thus, repeated self fertilization will kill off the heterozygotes. This

procedure for finding F_n is an example of the first type of application, i.e., using the algebra as a formalization of traditional methods.

A more modern use of the genetic algebra theory to self fertilization can be found in Holgate [26]. In this paper Holgate moves beyond the algebraic formalization to try to apply some of the key ideas of genetic algebras to self fertilizing mating systems. He studies mating systems that are capable of both self fertilization and random mating. He defines a compatibility condition between the self fertilization in the algebra and random mating and then proves that when this compatibility condition is satisfied, Haldane linearization can be extended to these algebras. He also exhibits three genetic situations whose zygotic algebras satisfy his compatibility condition: (1) a single diploid locus with multiple alleles, (2) polyploidy, with chromosome or chromatid segregation or a mixture of the two, and (3) k independently segregating loci with no distinction between the partitions of genes between chromosomes.

In summary, self fertilization provides us with an example of how historically the application of genetic algebra theory has been applied to a real genetic inheritance system. At first the algebra is used as formalism for the traditional approach, and as time goes by the algebraic investigation becomes more sophisticated (e.g., the Haldane linearization process). Hopefully, the end result is to provide new insight into real genetic problems.

5.3. Autopolyploidy. While humans and many of the higher animal species are diploid organisms (i.e., carrying two sets of chromosomes), cells in many plant species and some invertebrates may contain more than two sets of chromosomes. Such cells or organisms are, in general, called **polyploid**. Specifically, a cell which contained three sets of chromosomes would be called a **triploid** and a cell which contained four sets of chromosomes would be called a **tetraploid**. An **autopolyploid** is a polyploid that carries more than two identical haploid sets of chromosomes from the same species. This is one area of genetic investigation where the theory of genetic algebras is decidedly advantageous over traditional methods. Before algebras were introduced to the theory of genetics, the study of autopolyploid individuals involved rigorous computation. Haldane [20] was the first to initiate such investigations on autopolyploids with an even number of chromosome sets. The genetic algebra history of autopolyploidy again begins with Etherington [9]. He described the simplest possible example, the gametic algebra for autotetraploid individuals at one gene locus with two possible alleles (with pure chromosome segregation). The gametic types are AA, Aa , and aa . The multiplication table appears in Table 4. This gametic algebra turns out to be a special train algebra with principal train roots $0, 1, \frac{1}{6}$.

After Etherington, Gonshor [12], [13], [14] continued studying autopolyploidy taking into account populations with mutation and multiple alleles. In [13], he proved that both the gametic and zygotic algebras for polyploidy with multiple alleles are special train algebras. And in [14], he proved that polyploidy multiple allelic algebras with mutation are genetic algebras. It was in this same paper that Gonshor provided his definition of the genetic algebra and proved it was equivalent to Schafer's. In fact, it was the polyploid algebras that Gonshor used to demonstrate the utility of his new definition. The following is Gonshor's concluding paragraph of this important paper [14].

	AA	Aa	aa
AA	AA	$\frac{1}{2}AA + \frac{1}{2}Aa$	$\frac{1}{6}AA + \frac{2}{3}Aa + \frac{1}{6}aa$
Aa	$\frac{1}{2}AA + \frac{1}{2}Aa$	$\frac{1}{6}AA + \frac{2}{3}Aa + \frac{1}{6}aa$	$\frac{1}{2}Aa + \frac{1}{2}aa$
aa	$\frac{1}{6}AA + \frac{2}{3}Aa + \frac{1}{6}aa$	$\frac{1}{2}Aa + \frac{1}{2}aa$	aa

TABLE 4. Multiplication table of the gametic algebra for autotetraploids

We have emphasized throughout the use of mappings. Although an explicit computation of the multiplication tables would be highly involved we have seen that at least theoretically the structures of the algebras can be visualized. If we were forced to use the original basis elements which correspond to genotypes the table would truly become messy. But this is essentially what happens in classical population genetics. Thus we have seen examples how the technique of genetic algebras can handle problems which are difficult to handle by classical methods.

And so the study of polyploidy serves as a bridge between the superficial and the productive applications of nonassociative algebras to real genetics.

5.4. Sex Linked Inheritance. The final application we will consider, in more detail than the others, is sex linked genetic inheritance. Modelling sex linked inheritance with algebras involves overcoming the obstacle of asymmetry in the genetic inheritance rules. Inheritance which is not sex linked is symmetrical with respect to the sexes of the organisms, while sex linked inheritance is not. Our main focus in this section will be to carefully examine how the basic algebraic model must be altered in order to compensate for this lack of symmetry in the genetic inheritance system.

Again, it was Etherington [11] who began the study of this application with the simplest possible case. Consider a diploid population of male and female organisms. The males will be heterogametic (carrying chromosomes X and Y) while the females are homogametic (carrying two X chromosomes). Etherington began by considering a single gene difference on the X chromosome. So the gametic types for females are A and a , while the gametic types for males on the X chromosome are A and a , and we will indicate the gametic type on the Y chromosome with the letter Y . Assuming simple Mendelian segregation occurs in both sexes, gametic multiplication is shown in Table 5.

The possible female zygotes are AA , Aa , and aa , and the possible male zygotes are AY and aY . From the table of gametic multiplication we can then obtain a zygotic multiplication table. E.g., we calculate

$$AA \times AY = A\left(\frac{1}{2}A + \frac{1}{2}Y\right) = \frac{1}{2}AA + \frac{1}{2}AY.$$

Since only opposite sex mating will produce offspring, we have many products in the zygotic algebra that will be set equal to zero. E.g., the following would represent a

	A	a	Y
A	A	$\frac{1}{2}A + \frac{1}{2}a$	$\frac{1}{2}A + \frac{1}{2}Y$
a	$\frac{1}{2}A + \frac{1}{2}a$	a	$\frac{1}{2}a + \frac{1}{2}Y$

TABLE 5. Gametic multiplication for sex linked inheritance

	AA	Aa	aa	AY	aY
AA	0	0	0	$\frac{1}{2}AA + \frac{1}{2}AY$	$\frac{1}{2}Aa + \frac{1}{2}AY$
Aa		0	0	$\frac{1}{4}AA + \frac{1}{4}Aa + \frac{1}{4}AY + \frac{1}{4}aY$	$\frac{1}{4}Aa + \frac{1}{4}aa + \frac{1}{4}AY + \frac{1}{4}aY$
aa			0	$\frac{1}{2}Aa + \frac{1}{2}aY$	$\frac{1}{2}aa + \frac{1}{2}aY$
AY				0	0
aY					0

TABLE 6. Multiplication table for zygotic algebra for sex linked inheritance

product of two male zygotes and so they are all set equal to zero:

$$\begin{aligned} AY \times AY &= 0, \\ aY \times aY &= 0, \\ AY \times aY &= 0. \end{aligned}$$

Similarly, products of two female zygotes are zero. The zygotic multiplication table is shown in Table 6. For the sake of space only half of the table is given, but we note that the blanks in the table can be filled in by using commutativity. E.g., it still makes no difference whether we write $AA \times AY$ or $AY \times AA$, since both describe the mating of female zygote AA and male zygote AY . The algebra with vector space basis $\{AA, Aa, aa, AY, aY\}$ and multiplication table as in Table 6 is called the *zygotic algebra for sex linked inheritance for two alleles with simple Mendelian segregation rates*.

While modelling the phenomenon of sex linked inheritance in this way seems rather straight-forward, there are significant differences in the structure of these algebras when compared to genetic inheritance which is symmetric with respect to sex. The major structural difference is that this zygotic algebra is not the commutative duplicate of a gametic algebra. E.g., $AA \times AA \neq AA$ as it would be in the commutative duplicate. For the same reason, the zygotic algebra for sex linked inheritance is not a baric algebra. To overcome this apparent complication,

Etherington [11] introduced the idea of treating the male and female components of a population separately. Subsequently, Holgate [23] formalized this concept with the introduction of his sex differentiation algebra and a generalization of baric algebras called dibaric algebras. Following the modern notation of Wörz-Busekros [36], we introduce Holgate's definitions below.

Definition 5.1. Let $S = \langle w, m \rangle_{\mathbb{R}}$ denote a two dimensional commutative algebra over \mathbb{R} with multiplication table

$$\begin{aligned} w^2 = m^2 &= 0, \\ wm &= \frac{1}{2}(w + m). \end{aligned}$$

Then S is called the *sex differentiation algebra*.

It is clear that $S^2 = \langle w + m \rangle_{\mathbb{R}}$ is an ideal of S which, in fact, is isomorphic with the field \mathbb{R} . And so the algebra S^2 is a baric algebra. From this we can define Holgate's generalization of a baric algebra.

Definition 5.2. An algebra will be called *dibaric* if it admits a homomorphism onto the sex differentiation algebra S .

The purpose of this definition is the following quick result of Holgate's [23].

Proposition 5.1. *If an algebra A is dibaric, then A^2 is baric.*

Proof. Since A is dibaric, there exists an onto homomorphism $\phi: A \rightarrow S$. Thus, $\phi(A) = S$, which by the homomorphism property implies $\phi(A^2) = S^2$. However, we've already noted that S^2 is isomorphic with \mathbb{R} . Therefore, A^2 is a baric algebra since $\phi: A^2 \rightarrow \mathbb{R}$ is an onto homomorphism. \square

With these notions in place we now introduce the general definition of a zygotic algebra for sex linked inheritance with more than two gametic types. Let a_0, a_1, \dots, a_n be the different gametic types on the X chromosome. There are $(n+1)^2$ female genotypes $a_i a_j$, which we denote by a_{ij} for all $i, j = 0, \dots, n$. Inherent in this notation $a_i a_j$ is that the a_i is of maternal origin and a_j is of paternal origin. Male genotypes are given by $a_i Y$ for $i = 0, \dots, n$. We assume that the female genotype a_{ij} produces the gamete a_k with probability γ_{ijk} , while the male genotype $a_i Y$ produces a_k with probability γ_{ik} . In the absence of selection and if male genotypes produce X and Y gametes with the same frequency, then

$$\begin{aligned} \sum_{k=0}^n \gamma_{ijk} &= 1, \\ \sum_{k=0}^n \gamma_{ik} &= \frac{1}{2}. \end{aligned}$$

Let \mathcal{Z} be the real vector space with basis

$$\{ a_{ij}, a_i Y \mid i, j = 0, \dots, n \}$$

of dimension $(n+1)^2 + n + 1 = n^2 + 3n + 2$. Give \mathcal{Z} the multiplication

$$\begin{aligned} a_{ij} \cdot a_l Y &= a_l Y \cdot a_{ij} = \sum_{k,r=0}^n \gamma_{ijk} \gamma_{lra} a_{lr} + \frac{1}{2} \sum_{k=0}^n \gamma_{ijk} a_k Y, \\ a_{ij} \cdot a_l p &= a_l Y \cdot a_p Y = 0, \end{aligned}$$

for $i, j, l, p = 0, \dots, n$. Then \mathcal{Z} is the *zygotic algebra for asymmetric sex linked inheritance with gametic types* a_0, \dots, a_n . \mathcal{Z} is a commutative and nonassociative algebra. \mathcal{Z} is not a baric algebra since certain basis elements are nilpotent of index two. However, the zygotic algebra \mathcal{Z} is dibaric via the homomorphism $\phi: \mathcal{Z} \rightarrow S$ where $\phi(a_{ij}) = w$ and $\phi(a_i Y) = m$ for $i, j = 0, \dots, n$. Then by Proposition 5.1, \mathcal{Z}^2 is a baric algebra.

From here one approach is to apply the development of baric algebras to \mathcal{Z}^2 in order to determine structural information about \mathcal{Z} . This is the line of thought that was pursued by Holgate [23]. However, the dimension of \mathcal{Z}^2 depends on the segregation rates of the multiplication in \mathcal{Z} (i.e., the numbers γ_{ijk} and γ_{ik}). A fruitful alternative is to follow Gonshor [12], [13], [15] and Wörz-Busekros [34], [35], [36], who utilize a baric ideal of \mathcal{Z} for the same purpose. The following proposition can be found in Wörz-Busekros [36] with a proof appearing in [34].

Proposition 5.2. *The set*

$$\mathcal{L} = \left\{ \sum_{i,j=0}^n \alpha_{ij} a_{ij} + \sum_{i=0}^n \alpha_i a_i Y \mid \sum_{i,j=0}^n \alpha_{ij} = \sum_{i=0}^n \alpha_i \right\}$$

is an ideal of \mathcal{Z} which contains \mathcal{Z}^2 . Furthermore, \mathcal{L} is a baric algebra with weight homomorphism $\varphi: \mathcal{L} \rightarrow \mathbb{R}$ given by

$$\varphi \left(\sum_{i,j=0}^n \alpha_{ij} a_{ij} + \sum_{i=0}^n \alpha_i a_i Y \right) = \sum_{i=0}^n \alpha_i.$$

Genetically, concentrating on the subalgebra \mathcal{L} will give us all of the actual populations in the zygotic algebra \mathcal{Z} . I.e., \mathcal{L} contains all the elements of \mathcal{Z} which correspond to actual populations, whereas \mathcal{Z}^2 only contained those populations resulting from one generation of random mating. We notice also that the dimension of \mathcal{L} is $n^2 + 3n + 1$, regardless of the segregation rates in \mathcal{Z} .

Next we turn our attention to Gonshor [15], who introduced the concept of sex linked duplication. This concept followed Gonshor's [14] own basis-free approach to commutative duplication.

Let A be a baric algebra with weight homomorphism ω . We will define the *sex linked duplicate* of (A, ω) to be an algebra C with underlying vector space $A \times A \oplus A$. We note that Gonshor uses the notation $A \times A$ to represent the vector space $A \otimes A / I$, where I is the subspace generated by elements of the form $a \otimes b - b \otimes a$. The multiplication defined on $A \times A \oplus A$ is given by

$$(a \times b \oplus c)(d \times e \oplus f) = \frac{1}{2}(ab \times f \oplus \omega(f)ab) + \frac{1}{2}(de \times c \oplus \omega(c)de),$$

where ab denotes the product of a and b in A . Gonshor [15] showed that this operation is well defined. Now, identifying $a \times b$ with $a \times b \oplus 0$ and c with $0 \oplus c$, we see that

$$\begin{aligned} (a \times b)(c \times d) &= 0, \\ ab &= 0, \\ (a \times b)c = c(a \times b) &= \frac{1}{2}(ab \times c + \omega(c)ab), \end{aligned}$$

for all $a, b, c, d \in A$. By examining these equalities we begin to see the sex linked algebraic structure of \mathcal{Z} appear. In addition, the algebra C that we have defined

on $A \times A \oplus A$ is not baric, while C^2 is. We note that every $x \in C$ can be written uniquely as $x_1 + x_2$ where $x_1 \in A \times A$ and $x_2 \in A$. In C we now define two linear functions ω' and ω'' . If $x = a \times b + c$, define $\omega'(x) = \omega(c)$ and $\omega''(x) = \omega(a)\omega(b)$, then extend by linearity to all of C . Then let D be the subset of C consisting of all $x \in C$ such that $\omega'(x) = \omega''(x)$. It turns out that $C^2 \subset D$, D is a subalgebra of C , and D is a baric algebra. Therefore, C^2 is baric as well. Gonshor [15] goes on to show that if A is a genetic algebra, then D is also a genetic algebra by again utilizing his definition of a genetic algebra. Finally, we state the equivalence of this construction to our original sex linked zygotic algebra \mathcal{Z} .

Proposition 5.3. *The zygotic algebra \mathcal{Z} for sex linked inheritance and the sex linked duplicate $A \times A \oplus A$ of (A, ω) are isomorphic.*

This basis-free approach to sex linked inheritance has proven to be a worthy path to travel. In addition to Gonshor, Wörz-Busekros [35] also followed this line of inquiry and obtained results concerning the genetically significant sequence of plenary powers. If $x \in \mathcal{L}$ with $\varphi(x) = 1$, we let

$$x^{[t]} = \sum_{i,j=0}^n \alpha_{ij}^{[t]} a_{ij} + \sum_{i=0}^n \alpha_i^{[t]} a_i Y$$

be the representation of the t -th plenary power of x . Wörz-Busekros found that the coordinates of $x^{[t]}$ satisfy the recurrence relations

$$\begin{aligned} \alpha_{kr}^{[t+1]} &= \left(\sum_{i,j=0}^n \alpha_{ij}^{[t]} \gamma_{ijk} \right) \left(\sum_{l=0}^n \alpha_l^{[t]} 2\gamma_{lr} \right), \\ \alpha_k^{[t+1]} &= \sum_{i,j=0}^n \alpha_{ij}^{[t]} \gamma_{ijk}, \end{aligned}$$

where $k, r = 0, \dots, n$. These recurrence relations also lead to conditions for an element of \mathcal{L} to be an idempotent, which, as we discussed earlier, also has genetic significance.

The application of the genetic algebra theory to the genetics of sex linked inheritance provides us with a good example of how versatile the algebraic theory can be. As we saw, even the asymmetric nature of sex linked inheritance can be modelled effectively with algebras.

CONCLUSION

As we began to see in the last section on applications, describing different variations of genetic inheritance systems and studying the population genetics of each seem to be a never ending exercise. The examples we considered in this paper indicate that utilizing the theory of genetic algebras is usually a worthwhile endeavor. Some genetic systems clearly benefit more than others from the algebraic theory. But even those systems that are just as easily considered via traditional methods can be “cleaned up” with the formal language of abstract algebra. And so we conclude that this area of research activity is full of possibilities for the future, not only for mathematicians, but also for geneticists seeking a more systematic and powerful way to model real genetics.

Non-associative algebra, in general, is currently a very active field of mathematical research. However, in comparison with the body of literature of other classes

of nonassociative algebras (e.g, Lie algebras or Jordan algebras), the study of the algebras associated with the problems of genetic inheritance is still in its infancy. With that said, I should add that there is current research taking place on most of the algebras introduced in this survey article. A random sampling of current references might include Wörz-Busekros [37]; Walcher [33]; Guzzo [18], [19]; Costa and Guzzo [7], [8]; Hentzel, Peresi, and Holgate [21]; Peresi [31]; Cortés [6]; Martínez [28]; Burgueño, Neuberger, and Suazo [5]; González and Martínez [16]; and González, Martínez, and Vicente [17]. This list is by no means complete, but it does provide a starting point for some of the current work being done in the field. It is interesting to note there are very few American authors among this list. For whatever reason, these “genetic algebras” are not widely discussed or studied presently by American mathematicians. Hopefully, this article will open an avenue for future discussion and research into this fascinating class of nonassociative algebras and their relationship to the science of genetic inheritance.

REFERENCES

- [1] V.M. Abraham. Linearizing quadratic transformations in genetic algebras. *Proc. London Math. Soc. (3)*, 40:346–363, 1980. MR **82c**:92013a
- [2] S. Bernstein. Demonstration mathématique de la loi d’hérédité de Mendel. *Comptes Rendus Acad. Sci. Paris*, 177:528–531, 1923.
- [3] ———. Principe de stationarité et généralisation de la loi de Mendel. *Comptes Rendus Acad. Sci. Paris*, 177:581–584, 1923.
- [4] ———. Solution of a mathematical problem connected with the theory of heredity. *Ann. Sci. de l’Ukraine*, 1:83–114, 1924. (Russian).
- [5] Burgueño C., M. Neuberger, and A. Suazo. Totally orthogonal Bernstein algebras. *Arch. Math.*, 56:349–351, 1991. MR **92f**:17042
- [6] T. Cortes. Modular Bernstein algebras. *J. of Algebra*, 163:191–206, 1994. MR **95d**:17038
- [7] R. Costa and H. Guzzo Jr. Indecomposable baric algebras. *Lin. Alg. and its Applications*, 183:223–236, 1993. MR **94a**:17023
- [8] ———. Indecomposable baric algebras II. *Lin. Alg. and its Applications*, 196:233–242, 1994. MR **95e**:17030
- [9] I.M.H. Etherington. Genetic algebras. *Proc. Roy. Soc. Edinburgh*, 59:242–258, 1939. MR **1**:99e
- [10] ———. Duplication of linear algebras. *Proc. Edinburgh Math. Soc. (2)*, 6:222–230, 1941. MR **3**:103b
- [11] ———. Non-associative algebra and the symbolism of genetics. *Proc. Roy. Soc. Edinburgh*, 61:24–42, 1941. MR **2**:237e
- [12] H. Gonshor. Special train algebras arising in genetics. *Proc. Edinburgh Math. Soc. (2)*, 12:41–53, 1960. MR **23**:A1680
- [13] ———. Special train algebras arising in genetics II. *Proc. Edinburgh Math. Soc. (2)*, 14:333–338, 1965. MR **33**:2428
- [14] ———. Contributions to genetic algebras. *Proc. Edinburgh Math. Soc. (2)*, 17:289–298, 1971. MR **46**:1371
- [15] ———. Contributions to genetic algebras II. *Proc. Edinburgh Math. Soc. (2)*, 18:273–279, 1973. MR **48**:3522
- [16] S. González and C. Martínez. Idempotent elements in a Bernstein algebra. *J. London Math. Soc. (2)*, 42:430–436, 1990. MR **91m**:17048
- [17] S. González, C. Martínez, and P. Vicente. Idempotent elements in a 2nd-order Bernstein algebra. *Comm. Alg.* 22(2):595–609, 1994. MR **94m**:17033
- [18] H. Guzzo Jr. Embedding nil algebras in train algebras. *Proc. Edinburgh Math. Soc.*, 37:463–470, 1994. MR **95h**:17043
- [19] ———. The Peirce decomposition for commutative train algebras. *Comm. Alg.*, 22(14):5745–5757, 1994. MR **95h**:17042
- [20] J.B.S. Haldane. Theoretical genetics of auto-polyploids. *J. Genetics*, 22:359–372, 1930.

- [21] I.R. Hentzel, L.A. Peresi, and P. Holgate. On k -th order Bernstein algebras and stability at the $k + 1$ generation in polyploids. *IMA J. of Math. Appl. in Med. & Biol.*, 7:33–40, 1990. MR **91k**:17039
- [22] P. Holgate. Sequences of powers in genetic algebras. *J. London Math. Soc.*, 42:489–496, 1967. MR **36**:1499
- [23] ———. Genetic algebras associated with sex linkage. *Proc. Edinburgh Math. Soc. (2)*, 17:113–120, 1970. MR **46**:6858
- [24] ———. Characterisations of genetic algebras. *J. London Math. Soc. (2)*, 6:169–174, 1972. MR **47**:3479
- [25] ———. Genetic algebras satisfying Bernstein’s stationarity principle. *J. London Math. Soc. (2)*, 9:613–623, 1975. MR **57**:5175
- [26] ———. Selfing in genetic algebras. *J. Math. Biology*, 6:197–206, 1978. MR **83b**:92037
- [27] Y.I. Lyubich. Basic concepts and theorems of the evolutionary genetics of free populations. *Russian Mathematical Surveys*, 26(5):51–123, 1971. MR **56**:4906
- [28] C. Martinez. Isomorphisms of Bernstein algebras. *J. of Algebra*, 160:419–423, 1993. MR **94i**:17037
- [29] D. McHale and G.A. Ringwood. Haldane linearisation of baric algebras. *J. London Math. Soc. (2)*, 28:17–26, 1983. MR **84f**:17012
- [30] G. Mendel. Experiments in Plant-Hybridization. In James A. Peters, editor, *Classic Papers in Genetics*, pages 1–20. Prentice-Hall, Inc., 1959.
- [31] L. Peresi. On baric algebras with prescribed automorphisms. *Lin. Alg. and its Applications*, 78:163–185, 1986. MR **87i**:17034
- [32] R.D. Schafer. Structure of genetic algebras. *American J. of Mathematics*, 71:121–135, 1949. MR **10**:350a
- [33] S. Walcher. On Bernstein algebras which are train algebras. *Proc. Edinburgh Math. Soc.*, 35:159–166, 1992. MR **92m**:17055
- [34] A. Wörz-Busekros. The zygotic algebra for sex-linkage. *J. Math. Biol.*, 1:37–46, 1974. MR **51**:8194
- [35] ———. The zygotic algebra for sex-linkage. II. *J. Math. Biol.*, 2:359–371, 1975. MR **53**:13339
- [36] ———. *Algebras in Genetics*. Lecture Notes in Biomathematics, vol. 36, Springer-Verlag, New York, 1980. MR **82e**:92033
- [37] ———. Bernstein algebras. *Arch. Math.*, 48:388–398, 1987. MR **88d**:17024

DEPARTMENT OF MATHEMATICS, PHILADELPHIA COLLEGE OF PHARMACY AND SCIENCE, PHILADELPHIA, PENNSYLVANIA 19104

Current address: National Security Agency, Ft. George G. Meade, Maryland 20755

E-mail address: mlreedphd@aol.com