A formulation of the foundations of genetics and evolution

Brian Edward Bahr*

16299 Dakota Shores Dr., Park Rapids, MN 56470, United States

A B S T R A C T

This paper proposes a formulation of theories of the foundations of genetics and evolution that can be used to mathematically simulate phenotype expression, reproduction, mutation, and natural selection. It will be shown that Mendelian inheritance can be mathematically simulated with expressions involving matrices and that these expressions can also simulate phenomena that are modifications to Mendel's basic principles, like alleles that give rise to quantitative effects and traits that are the expression of multiple alleles and/or multiple genetic loci.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Similar to the way that Newton's formulation of the laws of motion can be used to mathematically simulate the trajectory of objects under the influence of forces, this paper proposes a formulation of the foundations of genetics and evolution that can be used to mathematically simulate phenotype expression, reproduction, mutation, and natural selection. This is not a new model of these phenomena but a mathematical representation of an organism with matrices that are acted on by functions designed to have the same effect on the representation as the biological processes listed above have on true organisms. Accordingly, each organism is represented by its own matrices and each matrix is operated on separately, which means that simulating a population of any significant size demands a computer.

Simulating these biological processes on paper is not as simple as using an equation that models them; however, a well-written computer program can make a simulation that is nearly as simple to operate. The main advantage of this formulation, though, is that we can observe the effects that each biological process has on the genotype and/or phenotype as a whole, which as we will see has several benefits when simulating natural selection. We will also see that, with this formulation, we are not constrained to modeling non-overlapping generations, nor are we constrained to using fitness values that are constant over time when we are simulating natural selection.

The majority of this paper will involve exploring the effects of each function and what each function can simulate. Each section in which a function is introduced will be followed by an example

of the effects of that function on an organism's representation, culminating in an example simulation of a small population involving every function that has been presented. Before the particular functions are presented, though, some notation needs to be introduced.

1.1. Notation

This formulation uses square diagonal matrices with entries from $\mathbb{Z}_n$ (the ring of integers *modulo n*) where n will depend on the complexity of an organism's phenotype expression (as we will see later). And the functions involved will operate on these matrices with matrix addition and multiplication (there will also be one action involving a calculation of the trace of a matrix).

All variables in italics used in this paper represent integers, so it will be automatically assumed and not specified that they are integers whenever a new variable is introduced; likewise, all matrices will be represented by boldface variables and will not necessarily be specified as matrices when they are introduced.

We will begin by distinguishing between two different matrix types.

The first type of matrix, the *genotype matrix*, will be used to represent the genotype of an organism; in particular, each position on the diagonal of a genotype matrix will represent one allele from that organism's genotype (and genotype matrices can either be used to represent an organism's total genotype or a section of it).

The second type of matrix we will call a *phenotype matrix*. The phenotype matrix will initially be constructed from operations on an organism's genotype matrices but, as we will see, its entries can also be altered by other functions. Accordingly, each entry along the diagonal of a phenotype matrix will represent one phenotypic trait that is either the expression of the organism's genotype, the

* Tel.: +1 507 272 3874.
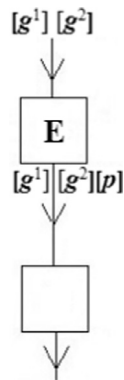*E-mail address:* brianedwardbahr@gmail.com
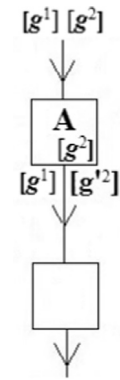
Fig. 1. Expression gate.
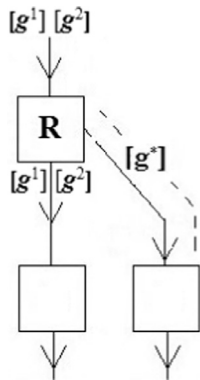

Fig. 2. Reproduction gate.

Fig. 3. Alteration gate.

result of environmental influences, or some combination of the two.

The other matrices (which will be presented in later sections) exist in the *environment space* which is a system of *paths* and *gates*. Paths are simply the trajectory an organism must follow between gates (like a wire in an electric circuit); and gates contain functions that act on an organism's matrices and also determine which matrices leave on which paths.

Now, when a matrix is added to or multiplied with an organism's matrix, this invariably produces a new matrix, so according to the definition of gates there are three things that can happen to this new matrix: either it leaves the gate it was created in on the same path as the matrices that entered the gate; it follows a different path than the matrices that entered the gate; or it does not leave the gate. Likewise, the matrices that entered the gate can either leave on the same path, follow different paths, or be prevented from leaving the gate.

Thus we will singularize three different types of gates: *expression gates, reproduction gates*, and *alteration gates*. In an expression gate, an organism's genotype matrices are operated on to create a phenotype matrix that leaves the gate it was created in on the same path as the organism's genotype matrices (Fig. 1). In a reproduction gate, matrices are generated from operations on the organism's genotype matrices which then follow a different path than the organism's matrices (Fig. 2). And in alteration gates, one of the organism's matrices will be operated on to produce a matrix that leaves the gate on the same path as the organism's other matrices; however, the particular matrix that was operated on will not leave the gate (Fig. 3).

Lastly, we will make the definition that an organism is any set of matrices that simultaneously enter or leave the same gate so that a matrix produced in a gate either becomes included in the set of the organism's matrices, or it becomes included in the set of

a new organism's matrices. This means that expression gates create a phenotype matrix which becomes a part of the organism; reproduction gates leave the original organism's genotype matrices unchanged but create matrices for a new organism; and alteration gates replace one of the organism's matrices with a new matrix.

In this manner, the action of having an organism enter an expression gate will be used in this formulation to simulate the biological phenomenon of phenotype expression; the action of having an organism enter a reproduction gate will be used to simulate the biological phenomenon of reproduction; and the action of having an organism enter an alteration gate will be used to simulate the biological phenomenon of mutation.

One final type of gate will be included to represent natural selection. A *selection gate* will contain a function that assesses the value of a certain entry in the organism's phenotype matrix, and then uses that value to determine whether the organism leaves the gate; and the path leaving a natural selection gate will always lead to a reproduction gate or another natural selection gate. So the action of having an organism enter a selection gate will be used to simulate natural selection.

We can see from the above definitions that an organism might, for example, be two genotype matrices and a phenotype matrix that simultaneously follow a path to a gate and then leave that vertex together on another path to simultaneously enter another gate, etc.

Now, a true organism really has a genotype for each cell in its body and a set of genotype and phenotype matrices could conceivably be made for each cell in an organism, but for most cases, we probably only need to distinguish between an organism's *germ-line matrices* and *somatic matrices* (matrices representing the genotype and phenotype of cells that contribute and do not contribute to gametes respectively). This distinction will come into play in the reproduction and alteration actions.

From the definition of an organism it is also clear that a *population* of organisms must contain a collection of paths and gates for each individual organism since two organisms cannot simultaneously enter the same gate. Thus a population can be represented by a collection of paths through the environment space that sets of matrices follow (which is why a population of any significant size demands a computer). An example involving a small population will be simulated in Section 6.1.

1.2. Constructing genotype matrices

One reason for choosing to use diagonal matrices in this formulation is so that we can use a mathematical operation that will act on the entire set of alleles in an organism's genotype, but will act on any two alleles if and only if they are *interactive alleles* (alleles—from different genetic loci or from the same genetic