

# An Algorithmic Approach to Modelling the Co-Evolution of Parasites and Their Hosts

Charles J. Mode

Correspondence: Charles J. Mode, Department of Mathematics, Drexel University, Philadelphia PA 1, USA

Received: January 17, 2020 Accepted: February 19, 2020 Online Published: February 26, 2020

doi:10.5539/ijsp.v9n2p13 URL: <https://doi.org/10.5539/ijsp.v9n2p13>

## Abstract

This paper is a reformulation of the paper, Mode 1958 *Evolution* 12:158 - 165, which was written in terms of a deterministic paradigm, using differential equations. In this paper, however, the working paradigm will be stochastic, and from the mathematical point of view, it will be a stochastic process that may be viewed as a branching process within a branching process. In particular, it will be assumed that the population of host plants will evolve as a multitype branching process, and the pathogen, which grows on the leaves of the host in every generation of the host, will also be assumed to evolve as a multitype branching processes during each generation of the host. The contents of this paper, were motivated by problems in Agriculture in which Plant Pathologists and Plant Breeders work together to control the damage inflicted by a pathogen on a growing crop of a cultivar such as flax, wheat, and many other cultivars. The focus of attention in this paper is the development of algorithms that will guide the development of software to run Monte Carlo simulation experiments taking into account mutations in the host and pathogen. The writing of software to implement the algorithms developed in this paper would require a major effort, and is, therefore, beyond the scope of this paper.

**Keywords:** host, pathogen, genes for resistance in the host, genes to pathogenicity in the pathogen, mutations pathogenicity in the pathogen, mutations for resistance to the pathogen in the host, embedded deterministic model

## 1. A Stochastic Model Describing the Coevolution of an Obligate Parasite and Its Host

The model formulated in Mode (1958) was based on a system of differential equations that were used as a framework to describe the coevolution of an obligate parasite and its host. These equations were also used to suggest that the host and the parasite populations would coevolve to a state of equilibrium in which both the host and the pathogen coexist. During the 1950s, computing technology was very primitive when compared to technology that is available in the present era. But, nevertheless, some simple calculations were included in the (1958) paper to suggest that a host - pathogen system would indeed converge to a state of a mutual equilibrium in which the host and pathogen populations would coexist indefinitely. Two other papers on the dynamics of Host - Pathogen systems were also published. In Mode 1961, a deterministic formulation, which generalized the model in Mode 1958 was published. A stochastic model of a Host - Pathogen system was published in 1964. In this paper, the working paradigm belonged to a class of stochastic processes known as birth and death processes.

In this paper, however, the model that will be formulated to describe the coevolution of a host - parasite system will belong to a class of stochastic processes called branching processes. To illustrate the ideas, in spring of some year in the evolution of a host - pathogen system, let the symbols  $H_i$  for  $i = 1, 2, \dots, n$ , denote the number  $n > 0$  of host plants in the initial generation. For the sake of simplicity, it will be assumed that all host plants will be infected with the parasite or some other pathogen. It will also be assumed that the infection of a plant by the pathogen, starts with one spore which leads to a pustule that produces a spore, this spore in turn produces one or more spores and that this process continues until a leaf or leaves of a plant is fully covered by the pustules of the parasite. In the literature on stochastic processes, the type of multiplicative process of spores just described is known as a branching process.

Mathematically, a simple branching process may be described as follows. Let the random variable  $X_0 = 1$  denote the initial pustule on a leaf, and let the random variable  $\xi$ , which takes values in the set  $\mathbb{I} = \{0, 1, 2, \dots\}$  of nonnegative integers, denote the number of pustules produce by any initial pustule. It will be assumed that the random variable  $\xi$  has a Poisson distribution with positive parameter  $\lambda$  and probability density function

$$P[\xi = x] = f(x) = \exp[-\lambda] \frac{\lambda^x}{x!} \quad (1.1)$$

for  $x = 0, 1, 2, 3, \dots$ . It is well known that the expectation and variance of a random variable with a Poisson distribution are  $E[\xi] = \lambda$  and  $var[\xi] = \lambda$ . In the context of the model under consideration, the parameter  $\lambda$  may be interpreted as a measure of the virulence of the pathogen.

The infection process for any of the  $n$  plants will be formulated in terms of class of stochastic processes that are known

as a branching process. For the sake of simplicity, it will be assumed that an infection of any plant begins with one spore that which produces one pustule, Then, by assumption, the number of pustules produced by the first pustule is a random variable and is given by the equation

$$X_1 = \xi, \tag{1.2}$$

where  $\xi$  is a realization of the Poisson random variable. In general, let the random variable  $X_n$  denote the number of pustules on a plant in the  $n$ -th generation of the pathogen, and let  $(\xi_k | k = 1, 2, \dots)$  denote a sequence of independent random variables whose common distribution is that of the random variable  $\xi$ . Then, it follows that the number of pustules on a plant in generation 2 is

$$X_2 = \sum_{k=1}^{X_1} \xi_k. \tag{1.3}$$

And in general, this line of reasoning may be extended to conclude that

$$X_\nu = \sum_{k=1}^{X_{\nu-1}} \xi_k. \tag{1.4}$$

is the number of pustules on a plant for all generations  $\nu = 1, 2, \dots$  of the pathogen. For the case of flax rust, the generation time of the pathogen is about 10 days. In reality and in any computer simulation experiment, only some finite number  $m > 0$  of generations of the pathogen will be considered.

From this equation, it can be seen that the conditional expectation of  $X_\nu$ , given  $X_{\nu-1}$ , is

$$E [X_\nu | X_{\nu-1}] = X_{\nu-1} \lambda \tag{1.5}$$

for  $\nu \geq 1$ . Therefore, the unconditional expectation of the number of pustules in generation  $\nu \geq 1$  on any plant satisfies the recursive equation

$$E [X_\nu] = E [X_{\nu-1}] \lambda \tag{1.6}$$

for  $\nu \geq 1$ . The algorithms just described may be used to program a computer to simulate that number of pustules on any of the  $n$  plants under consideration at the end of a season. In any computer experiment, it would be necessary to repeat the infection process just described for each of the  $n$  plants under consideration.

For any host plant  $H_i$ , let the random variable  $Y_i$  denote the total number of pustules on the plant at the end of the growing season. Then, the total load of pustules on the  $n > 0$  plants under consideration is given by the equation

$$T = \sum_{i=1}^n Y_i. \tag{1.7}$$

The value of the random variable will influence that quantity and quality of the seeds produced by the plants. In extreme cases, the host plants may not be able to produce a sufficient number of viable seeds to produce the next generation of plants. Under such circumstances, the host-pathogen system would go extinct. But, even though the plants are infected with a pathogen, it may also happen that the plants produce a sufficient number of viable seeds to produce the next generation of plants. In such cases, the host - pathogen system would survive for another generation.

Given the a value of the random variable  $T$ , let  $S (\nu)$  denote the conditional probability that in any generation a plant in generation  $\nu \geq 1$  survives to produce plants in generation  $\nu + 1$ . If a plant is not infected by a pathogen, then it will be assumed that this conditional survival probability has the form

$$S (\nu) = \exp [-(\beta\nu)^2]. \tag{1.8}$$

If, however, a plant is infected with a pathogen, then it will be assumed that this survival probability has the form

$$S (\nu | T) = \exp [-(T\beta\nu)^2]. \tag{1.9}$$

Let  $n$  denote the number of plants at the end of a season. Then, it will be assumed that the number of plants that will survive and produce seeds in the next generation is a random variable  $W$  with a binomial distribution with sample size  $n$  and probability  $p$ . It is well known that the probability density function of the binomial distribution is

$$f_{BN} (x | n, p) = \binom{n}{x} p^x (1 - p)^{n-x} \tag{1.10}$$

for  $x = 0, 1, 2, \dots, n$ .

In what follows, the notation  $X \sim BN(n, p)$  will be used to denote that a random variable  $X$  has a binomial distribution sample size  $n$  and probability  $p$ . In any computer simulation experiment, a realization of the random variable  $W$  would be computed using the formula

$$W \sim BN(n, S(v | T)) , \tag{1.11}$$

and the realization of the random variable  $W$  would be the initial number of plants in the next generation. At this point in a simulation, the algorithms outlined above would be used to continue the Monte Carlo simulation experiment for another season.

As indicated in the title of this section, some obligate parasite was considered as the pathogen in the formulation above, but the model would could also be used for cases in which the pathogen may also grow in the soil or on media prepared in a laboratory.

## 2. Genetics of Host and Pathogen

The genetics of pathogenicity and host resistance to flax rust, *Melampsora lini* (Pers) Lev., has been reported in a pioneering works by Flor (1955 and 1956). Through a series of genetic studies, Flor has shown that the host and parasite possess complementary genetic systems. That is to say, any gene in the host for resistance acts if, and only if, there is a corresponding gene in the pathogen for avirulence. The genes for host resistance exist as a series of multiple alleles at five loci, designated as the K, L, M, N, and P in the genome of the host. There is one gene at the K locus, 11 at the L, six at the M, three at the N, and four at the P, for a total of twenty genes. The K, L, and M loci are inherited independently, but the N and P loci are linked with about 26 percent recombination. Unlike the genes for host resistance, the genes for pathogenicity in the pathogen exist at twenty-five separate loci.

Resistance of a host genotype to a particular pathogen genotype occurs whenever any allele in the host (at any one of the five multiple allelic loci) and its complementary gene for avirulence in the pathogen at any one of the twenty-five ( diallelic loci) are present simultaneously. The alleles for host resistance are all dominant or semi-dominant so that the heterozygote as well as the homozygote are resistant. The genes for avirulence in the pathogen are also dominant, with the exception of one locus where the homozygous recessive in combination with one or two doses of a gene in the host is necessary for resistance. Further details on the genetics flax resistance to flax rust may be found in Flor (1955) and (1956).

These complementary genetic systems of the host and parasite are illustrated in the model presented below for the simple case of one locus in both the host and parasite with two alleles at each locus. Let  $R$  and  $r$ , respectively, denote alleles for resistance and susceptibility to the pathogen in the host, and let  $A$  and  $a$  denote alleles for avirulence and virulence, respectively, in the pathogen. In this model, there are three genotypes of the host, namely  $RR$ ,  $Rr$  and  $rr$ . With respect to the pathogen there are also three genotypes:  $AA$ ,  $Aa$  and  $aa$ . The interactions of the three genotypes of the host with three genotypes of the pathogen are illustrated in the table below.

Table 2.1. Interactions of the Genotypes of the Host and Pathogen

<i>Genotypes</i>	<i>RR</i>	<i>Rr</i>	<i>rr</i>
<i>AA</i>	<i>RIS</i>	<i>RIS</i>	<i>SUS</i>
<i>Aa</i>	<i>RIS</i>	<i>RIS</i>	<i>SUS</i>
<i>aa</i>	<i>SUS</i>	<i>SUS</i>	<i>SUS</i>

Observe that the columns of the table represent the genotypes of the host, and the rows of the table represent the genotypes of the pathogen. The symbols in the body of the table represent the interactions of the genotypes of the host and pathogen. For example in Table 2.1, the symbol *RIS* denotes resistance of the host to the pathogen, and the symbol *SUS* denotes susceptibility of the host to the pathogen. By inspecting the first row of this table, it can be seen that according to the model under consideration, the allele  $R$  in the host for resistance to the pathogen is assumed to be dominant to the recessive allele  $r$  for host susceptibility to the pathogen. The dominance effect of the allele  $A$  is also illustrated in row 2 of the table. In row 3 of the table, the susceptibility of each of the three genotypes of the host to the virulent genotype  $aa$  of the pathogen is indicated.

The simple illustrative case illustrating the gene to gene relationship in the host and pathogen considered in this section could in principle be extended to the case of one locus with multiple alleles. But, such an extension would require a more complex notation, and, therefore, will not be considered in this section. In the next section, however, cases in which the gene for gene relationship in the host and the pathogen will be considered with respect to two or more loci with multiple alleles in the host and pathogen.

Research on the genetics of host - pathogen systems is still an active field of research, but the focus of attention of current research differs from that of Flor. Currently, quite a number of researchers are studying the genetics of host - pathogen systems by sequencing the genomes of the host and parasite. Among those researchers, who are sequencing the genomes of host and pathogens are Dr. Robert Brueggeman, Department of Plant Pathology, North Dakota State University. His current focus of attention is the sequencing the genomes of barley and one of its parasites, which is a fungus called net blotch. Two Ph.D. theses Boyle (2009) and Richards (2016) have also been devoted to research on net blotch.

Teams of researchers are working on sequencing the genomes of barley and net blotch. For example, the paper Wyatt et al. 2017 contains an account of the sequencing of the *Pyrenophora teres f. teres* Isolate 0-1, which the scientific name for net blotch. The paper Mascher et al. 2017 is devoted to an account of the sequencing of the barley genome and the paper by Tamang et al. 2019 is devoted to finding susceptibility/resistance to net blotch in the barley genome. In the paper Richards et al. 2016, there is an account of the fine mapping of the barley genome 6H net blotch susceptibility locus G3.. This brief review will provide the reader with some of the areas of genomic research that is currently in progress among scientists working on resistance of barley to net blotch, other plant species and pathogens.

### 3. Linkage of Genes in the Host for Resistance to Two Pathogens

It has been observed by plant geneticists during the past several decades that genes in the host for resistances to two or more pathogens are often linked, i.e. they are on the same chromosome in the host. An example of such linkage was reported in the paper Schaller and Briggs (1955) on genes for bunt resistance in wheat. Currently, one can find many reports on the internet on the location of genes for resistance in partially sequenced genomes of barley and wheat and their functions at the molecular level, but it is beyond the scope of this paper to review this more recent literature. The main thrust of this section is to focus on the formulation of a mathematical structure that accommodates linkage of genes for resistance in the host to two or more pathogens that may be used in computer simulation experiments.

In chapter 2 of the book Mode and Sleeman (2012), a mathematical structure for dealing with linkage of genes at two or more loci is described for diploid organisms. In that chapter a diploid genotype was represented by the symbol  $(x, y)$ , where  $x$  is the allele contributed by the maternal parent and  $y$  is the allele contributed by the paternal parent. For the case of two alleles denoted by 0 and 1, four genotypes may be identified; namely, (0, 0), (0, 1), (1, 0) and (1, 1). Observe that in genotype (0, 1) the maternal parent contributed the allele 0, but in genotype (1, 0) the maternal parent contributed the allele 1.

It should be noted that not all cereal species that are important in agriculture are diploids. A plant species is said to be a polyploid, if during its evolution two or more genomes have been incorporated to make a single genome. For example, durum wheat and bread wheat are polyploids. Durum wheat (*Triticum durum*) for example, has 14 pairs of chromosomes. It is thought that during its evolution, two genomes that were originally 7 pairs of chromosomes, were joined to make a species of 14 pairs of chromosomes. Bread wheat (*Triticum aestivum*) has 21 pairs of chromosomes, and it is thought that the genome of this species is made of the genomes of three species with genomes of 7 pairs of chromosomes. Polyploid species may not follow the same laws of inheritance than those of diploid species. Consequently, the linkage structure under consideration may not be applicable for polyploid species. Barley is an important species in agriculture, and it has a genome made up of 7 pairs of chromosomes. Moreover, this species follows the laws of inheritance for diploid species, and therefore the linkage structure that will be developed will be applicable to experiments with barley.

In chapter 2, section 2 of Mode and Sleeman (2012), formulas are derived that accommodate linkage at two or more loci that can be applied in computer simulation experiments involving numerical calculations. In this section, a revision of some formulas will be derived for the case of two autosomal linked loci. In this case, an arbitrary genotype of diploid individual with respect to two linked loci may be represented in the form

$$\frac{00}{11}, \quad (3.1)$$

where 00 represents the alleles at two loci contributed by the female parent. Similarly, the symbol 11 denoted the alleles contributed by the male parent.

An individual of this genotype may in turn produce four types of gametes; namely 00, 01, 10, and 11. Gametes 00 and 11 represent copies of the maternal and paternal gametes respectively; while 01 and 10 are recombination type gametes containing both maternal and paternal genes. In particular cases, the generic gametic symbols could be identified with particular haplotypes when the focus of attention is at the molecular level, or with phenotypes when the discussion is at the Mendelian level. It should also be noted that the Boolean notation under consideration is not phase dependent with respect to some specific genes at two loci to which attention is being directed, because in terms of Boolean indicators only parental origins of genes are under consideration.

The four types of gametes will be produced by a given genotype with certain probabilities depending on the probability

of recombination. Let  $\gamma(00)$ ,  $\gamma(01)$ ,  $\gamma(10)$ , and  $\gamma(11)$  denote the probabilities an arbitrary genotype produces gametes 00, 01, 10, and 11, respectively. In the two loci case, the linkage (gametic) distribution is the set

$$(\gamma(00), \gamma(01), \gamma(10), \gamma(11)) \tag{3.2}$$

of non-negative numbers whose sum is one. Because of the complementary nature of the meiotic mechanism, i.e., gametes are almost always produced in pairs, the mechanism of meiosis will be called balanced if the equations

$$\begin{aligned} \gamma(00) &= \gamma(11) \\ \gamma(10) &= \gamma(01) \end{aligned} \tag{3.3}$$

are satisfied.

If we let  $\rho$  be the probability of recombination, then it follows that

$$\gamma(10) + \gamma(01) = \rho \tag{3.4}$$

and

$$\gamma(00) + \gamma(11) = 1 - \rho . \tag{3.5}$$

But, if the mechanism of meiosis is balanced, then

$$\gamma(01) = \gamma(10) = \frac{1}{2}\rho \tag{3.6}$$

and

$$\gamma(00) = \gamma(11) = \frac{1}{2}(1 - \rho) . \tag{3.7}$$

When a objective of an investigation is to map the locations of two loci on the same chromosome expressed in terms of centimorgens (cM), then the pertinent values of  $\rho$  will satisfy the condition  $0 \leq \rho \leq \frac{1}{2}$ . The case of random assortment occurs when  $\rho = \frac{1}{2}$  so that the probability of each type of gamete is 1/4. In principle, however,  $\rho$  may be any value such that  $0 \leq \rho \leq 1$ . Note if  $0.5 < \rho \leq 1$ , then  $\gamma(01)$  and  $\gamma(10)$  are greater than  $\gamma(00)$  and  $\gamma(11)$ .

A problem of considerable theoretical importance in assessing the effects of linkage in populations is that of defining recombination probabilities and finding a relation between these recombination probabilities and the gametic distribution, when the number of loci under consideration is greater than two. An interesting step towards a solution of this problem was made by Schnell (1961), who observed that if one makes the transformation

$$\rho = \frac{1}{2}(1 - \lambda) , \tag{3.8}$$

or equivalently

$$\lambda = 1 - 2\rho , \tag{3.9}$$

then a certain orthogonality is introduced which leads to an extension to cases of an arbitrary number of linked loci. Observe that as  $\rho$  varies over the interval  $[0, 1]$ , the parameter  $\lambda$  varies over the interval  $[-1, 1]$ .

By using these equations, it can be seen that

$$\begin{aligned} \gamma(00) &= \frac{1}{4}(1 + \lambda) \\ \gamma(10) &= \frac{1}{4}(1 - \lambda) . \end{aligned} \tag{3.10}$$

These equations can perhaps be most easily comprehended in vector-matrix notation. Let

$$\begin{aligned} \gamma^T &= (\gamma(00), \gamma(10)) \\ \lambda^T &= (1, \lambda) \end{aligned} \tag{3.11}$$

be  $1 \times 2$  vectors, where the superscript  $T$  stands for transpose of a vector or matrix, and let

$$\mathbf{A}_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{3.12}$$

denote a  $2 \times 2$  matrix. Observe that the rows and columns of this matrix is orthogonal. Then, in vector-matrix notation the above equation may be written in the form

$$\gamma = \frac{1}{4} \mathbf{A}_2 \lambda . \tag{3.13}$$

This equation is of interest, because it expresses  $\gamma$  as a function of  $\lambda$ . In a computer simulation experiment, if the components of vector  $\lambda$  are assigned numerical values, then the vector  $\gamma$  is determined. From now on the symbol  $T$  will be used to denote the transpose of a matrix.

The following observations are very helpful in finding an extension to the case of an arbitrary number of linked loci. Firstly,

$$\mathbf{A}_2^T = \mathbf{A}_2 \tag{3.14}$$

so that  $\mathbf{A}_2$  transpose is  $\mathbf{A}_2$ . When a square matrix remains invariant under the operation of transposition, it is said to be symmetric. Secondly, observe that

$$\mathbf{A}_2^T \mathbf{A}_2 = \mathbf{A}_2^2 = 2\mathbf{I}_2 , \tag{3.15}$$

where  $\mathbf{I}_2$  is a  $2 \times 2$  identity matrix., which has the form

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.16}$$

Equation (3.15) expresses an orthogonality condition which, as we shall see, may be generalized to an arbitrary number of loci greater than two. Thirdly, if the symbols 00 and 10 are regarded as the vectors  $\xi_1^T = (0, 0)$  and  $\xi_2^T = (1, 0)$ , then the matrix  $\mathbf{A}_2$  may be represented in the form

$$\mathbf{A}_2 = (a_{ej}) = ((-1)^{\xi_i^T \xi_j}) , \tag{3.17}$$

for  $i, j = 1, 2$ . Lastly, we see the column vector  $\lambda$  may be represented in the form

$$\lambda = 2\mathbf{A}_2 \gamma \tag{3.18}$$

This equation could be taken as a definition of the column vector  $\lambda$ . In the next section, these results will be extended to the case of three linked loci.

#### 4. Linkage of Genes for Resistance in the Host to Three Pathogens

In this section in order to make the paper self contained, a revised version of section 2.5 in Mode and Sleeman 2012 will be presented that will provide a structure to consider genes in the host for resistance to three pathogens. A question that naturally arises is whether the above scheme considered in section 3 can be generalized to any arbitrary number of loci greater than two. As we shall see such a generalization is possible, but the manner in which the scheme may be generalized will not become clear until the three loci case is considered. In the three loci case, an arbitrary diploid genotype may be represented in the form

$$\frac{000}{111} , \tag{4.1}$$

where as before the zeros and ones represent genes contributed by maternal and paternal parents, respectively. This genotype is capable of generating  $2^3 = 8$  types of gametes containing various combinations of maternal and paternal genes.

The set of these eight types of gametes will be represented in the form

$$\mathbb{G} = (000, 100, 010, 110, 111, 011, 101, 001) , \tag{4.2}$$

and let the vector

$$(\gamma(\xi) \mid \xi \in \mathbb{G}) \tag{4.3}$$

denote the linkage distribution, *i.e.*, the set of non-negative numbers, whose sum is one, giving the probability that each type of gamete is produced by the meiotic process. Because the meiotic mechanism is assumed to be balanced, gametes are produced in pairs so that following symmetry or complementary conditions

$$\begin{aligned} \gamma(000) &= \gamma(111) \\ \gamma(100) &= \gamma(011) \\ \gamma(010) &= \gamma(101) \\ \gamma(110) &= \gamma(001) \end{aligned} \tag{4.4}$$

will be assumed. From these symmetry conditions, it follows that it will be sufficient to consider only four probabilities from the linkage distribution in setting up a correspondence with a set of recombination probabilities.

In the three loci case, a recombination probability may be associated with each pair of loci; namely, the pairs (1, 2), (1, 3) and (2, 3). Let  $\rho_{12}$ ,  $\rho_{13}$ , and  $\rho_{23}$ , denote the probability of recombination between the respective pairs of loci. With each recombination probability, we may associate a lambda parameter denoted by  $\lambda_{12}$ ,  $\lambda_{13}$  and  $\lambda_{23}$ . From the definitions of  $\rho_{12}$ ,  $\rho_{13}$ , and  $\rho_{23}$ , it follows that

$$\begin{aligned} \rho_{12} &= 2(\gamma(100) + \gamma(010)) \\ \rho_{13} &= 2(\gamma(100) + \gamma(110)) \\ \rho_{23} &= 2(\gamma(010) + \gamma(110)) \\ 1 - \rho_{12} &= 2(\gamma(000) + \gamma(110)) \end{aligned} \tag{4.5}$$

Observe that only four gametic probabilities appear on the right so it may be possible to solve four simultaneous linear equations. Also observe that the symmetry conditions were used in choosing the four gametic probabilities on the right.

By substituting the  $\lambda$ 's for the  $\rho$ 's in these equations, it can be shown that

$$\begin{aligned} \frac{1}{4}(1 - \lambda_{12}) &= \gamma(100) + \gamma(010) \\ \frac{1}{4}(1 - \lambda_{13}) &= \gamma(100) + \gamma(110) \\ \frac{1}{4}(1 - \lambda_{23}) &= \gamma(010) + \gamma(110) \\ \frac{1}{4}(1 + \lambda_{12}) &= \gamma(000) + \gamma(110) \end{aligned} \tag{4.6}$$

By solving these four equations for  $\gamma(000)$ ,  $\gamma(100)$ ,  $\gamma(010)$  and  $\gamma(110)$ , it can be seen that

$$\begin{aligned} \gamma(000) &= \frac{1}{8}(1 + \lambda_{13} + \lambda_{23} + \lambda_{12}) \\ \gamma(100) &= \frac{1}{8}(1 - \lambda_{13} + \lambda_{23} - \lambda_{12}) \\ \gamma(010) &= \frac{1}{8}(1 + \lambda_{13} - \lambda_{23} - \lambda_{12}) \\ \gamma(110) &= \frac{1}{8}(1 - \lambda_{13} - \lambda_{23} + \lambda_{12}) \end{aligned} \tag{4.7}$$

Just as in the two loci case, these equations may be most easily comprehended if they are cast in vector-matrix notation. Let

$$\gamma^T = (\gamma(000), \gamma(100), \gamma(010), \gamma(110)) \tag{4.8}$$

and

$$\lambda^T = (1, \lambda_{13}, \lambda_{23}, \lambda_{12}) \tag{4.9}$$

denote row vectors, and let

$$\mathbf{A}_3 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \tag{4.10}$$

denote a  $4 \times 4$  matrix. Then the equations may be written in the compact form

$$\gamma = \frac{1}{2^3} \mathbf{A}_3 \lambda \tag{4.11}$$

It should also be noted that the matrix  $\mathbf{A}_2$  is related to the matrix  $\mathbf{A}_3$  by the simple recursion formula

$$\mathbf{A}_3 = \begin{bmatrix} \mathbf{A}_2 & \mathbf{A}_2 \\ \mathbf{A}_2 & -\mathbf{A}_2 \end{bmatrix} \tag{4.12}$$

From this recursive equation, it can be seen that

$$\mathbf{A}_3^T = \begin{bmatrix} \mathbf{A}_2^T & \mathbf{A}_2^T \\ \mathbf{A}_2^T & -\mathbf{A}_2^T \end{bmatrix} = \begin{bmatrix} \mathbf{A}_2 & \mathbf{A}_2 \\ \mathbf{A}_2 & -\mathbf{A}_2 \end{bmatrix} = \mathbf{A}_3 \tag{4.13}$$

because  $\mathbf{A}_2$  is symmetric. Furthermore, from this equation, it follows that

$$\begin{aligned} \mathbf{A}_3^T \mathbf{A}_3 &= \mathbf{A}_3^2 = \begin{bmatrix} 2\mathbf{A}_2^2 & \mathbf{0} \\ \mathbf{0} & 2\mathbf{A}_2^2 \end{bmatrix} = 2 \begin{bmatrix} 2\mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & 2\mathbf{I}_2 \end{bmatrix} \\ &= 2^2 \mathbf{I}_4 \end{aligned} \tag{4.14}$$

where  $\mathbf{0}$  is a  $2 \times 2$  zero matrix and  $\mathbf{I}_4$  is an identity matrix of order 4.

We thus see that orthogonality condition carries over to the three loci case. Moreover, let  $\xi_1^T = (0, 0, 0)$ ,  $\xi_2^T = (1, 0, 0)$ ,  $\xi_3^T = (0, 1, 0)$  and  $\xi_4^T = (1, 1, 0)$  denote row vectors. Then, by inspection, it can be seen that

$$\mathbf{A}_3 = (a_{ij} = (-1)^{\xi_i^T \xi_j}) \tag{4.15}$$

for all  $i, j = 1, 2, 3, 4$ . Just as in the case of two loci, it also follows, by using the above results and solving for the vector  $\lambda$ , that the equation

$$\lambda = 2\mathbf{A}_3 \gamma, \tag{4.16}$$

expresses the vector  $\lambda$  as a linear function of the vector  $\gamma$ , given the matrix  $\mathbf{A}_3$ .

At this juncture it is important to note that the ordering of the gametic symbols

$$(000, 100, 010, 110) \tag{4.17}$$

played a basic role in extending the observations made in the case two loci to the case of three loci. Furthermore, the linear relation (4.16), connecting the vectors  $\lambda$  and  $\gamma$ , suggests that to extend the results for the case of three loci to cases of four or more loci, it would be prudent to simplify the notation by abandoning subscripts on the  $\lambda$  and  $\rho$  parameters and replacing them by a function notation of the form  $\lambda(\xi)$  and  $\rho(\xi)$ , where  $\xi$  is an arbitrary gametic symbol containing 0's and 1's.

Given this function notation, the matrix  $\mathbf{A}_3$  and the ordering of the elements of the gametic vector  $\gamma$ , equation (4.2) leads to an automatic ordering of the elements of the vector  $\lambda$ . This observation suggests that extensions of equation (4.1) to cases of four or more loci could be used to define the vector  $\lambda$  for an arbitrary number of loci. Then, given any  $\xi \in \mathbb{G}$ , a corresponding recombination probability may be determined by the equation

$$\lambda(\xi) = 1 - 2\rho(\xi). \tag{4.18}$$

By way of an illustrative example, suppose  $\xi = 000$ , indicating that in gametes of this type there was no genetic recombination. For the case of three loci under consideration, it was observed that  $\lambda(\xi) = 1$ , which implies  $\rho(\xi) = 0$ , indicating with gametes of type  $\xi = 000$  recombination occurs with probability 0, which is consistent with our intuition. In this paper, however, there will be no attempt to extend the results presented in this section to the cases of four or more loci, because it seems likely that demand for such cases will be very limited. If, however, a reader is interested in considering four or more linked loci, chapter 2 of the book Mode and Sleeman 2012 may be consulted for an account of working with four or more linked loci.

It would be of interest to present an example of data such that the probabilities of recombination for the case of three linked loci could be estimated, but no attempt to construct such an example will be undertaken. For readers who are interested in three point test for linkage, the book by Liu 1998 may be consulted.

### 5. Mutations in the Host and Pathogen

It is well known that mutations do occur in plant pathogens. For example, a variety of durum or bread wheat may have been grown extensively in an area for many years, because it's resistant to stem rust, a plant pathogen. However, it may happen that there is a mutation in the pathogen, stem rust, so that a variety of wheat that was resistant to the pathogen is no longer resistant, due to a mutation in the pathogen. As the science of genomics continues to develop, it may become possible to identify the location in the genome of the pathogen where the mutation occurred and characterize it at the molecular level. In this section, however, mutations in the host and pathogen will be described and analyzed mathematically with a view towards providing a framework for studying the interactions of hosts and pathogens as well as their coevolution.

Let  $A$  denote a gene in the pathogen for avirulence, and let  $a$  denote an allele for virulence. Similarly, let  $R$  let denote a gene in the host for resistance to the pathogen. and  $r$  denote an allele for susceptibility in the host . It will be assumed that the interactions of the host and pathogen follow those described in Table 2.1 in section 2. Let  $\mu_{12}$  denote the probability per generation that gene  $A$  in the pathogen mutates to gene  $a$  for virulence. Similarly, let  $\nu_{12}$  denote the probability per generation that gene  $R$  for resistance in the host mutates to gene  $r$ , which confers susceptibility of the host to the pathogen. If gene  $a$  in the pathogen mutates to gene  $A$  for avirulence, then, by definition, a back mutation has occurred.

The set of possible mutations in the pathogen may be represented in a two by two matrix

$$\mathbb{P} = \begin{bmatrix} \nu_{11} & \nu_{12} \\ \nu_{21} & \nu_{22} \end{bmatrix}. \tag{5.1}$$

In the first row of this matrix  $\nu_{11} = 1 - \nu_{12}$  so that if  $\nu_{12} = 10^{-9}$ , then  $\nu_{11}$ , which is the probability a mutation does not occur, is close to 1. Row 2 of this matrix, takes into account back mutations, which may occur with probability  $\nu_{21} > 0$ . In computer simulation experiments, the row of the matrix  $\mathbb{P}$  are assigned numbers such that the rows of this matrix sum to 1. Similarly, mutations in the host may be represented by the matrix

$$\mathbb{H} = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix}, \tag{5.2}$$

row sum to 1. Observe that  $\nu_{12}$  is the probability per generation that gene  $R$  in the host mutates to gene  $r$ . The probabilities in row 2 of the matrix  $\mathbb{H}$  take into account that back mutations may occur in the host.

It seems likely that before wheats, barley, flax and other plants that were domesticated, they were outbreeders so that in an evolving population random mating was prevalent, because winds and perhaps insects would distribute pollen randomly among plants. According to this view, as these species of plants were domesticated they evolved to a state in which they reproduced by self fertilization. In this situation, the anatomy of a flowering plant evolved in such a way that the pollen fertilizes only its eggs in same flower. Consequently, when formulating models describing the evolution of these species of plants, it will be necessary to take into account not only random mating but also reproduction by self fertilization. in models dealing with the joint evolution of the host and pathogen. It will be assumed that the pathogen reproduces asexually. That is individuals of type  $A$  produce only individuals of type  $A$  if a mutation does not occur. Similar remarks hold for individuals of type  $a$  of the pathogen.

A computer experiment that would mimic that rise of virulent mutations in the pathogens and their impact on populations of domestic wheats, barley and other crops would be of practical interest, because they are actually observed in such populations. Bread and durum wheats are polyploids, but with respect to one locus in which genes for resistance or susceptibility to a pathogen are located, the genetic laws governing these traits are similar to those of diploids. The laws governing the resistance to pathogens in barley follow those of a diploid population. Consider a host pathogen system in which pathogen is haploid and reproduces asexually. It will be assumed that the pathogen has two alleles  $A$  for avirulence and  $a$  for virulence. As is section 2, let  $R$  denote a allele in the host for resistance to the pathogen and let  $r$  denote at the same locus for susceptibility to the pathogen. In this model, the host population would consist of three genotypes  $RR$ ,  $Rr$  and  $rr$ . In the absence of mutation, when a population reproduces either by random mating or selfing, individuals of genotype  $RR$  would produce offspring only of genotype  $RR$ . But, is such population individuals of genotype  $Rr$  would produce offspring with genotypes  $RR$ ,  $Rr$  and  $rr$  with corresponding probabilities 0.25, 0.50, and 0.25. Similarly, in the absence of mutation, individuals of genotype  $rr$  would produce offspring only of genotype  $rr$ .

Mutations are rare events and thus need to be taken into account in a model according the laws of evolution of a stochastic processes. For example, with regard to the pathogen population, let  $M$  denote the number of individuals of genotype  $A$  in the population in some generation , and let  $m$  denote the number of mutations from genotype  $A$  to the virulent genotype  $a$ . If these mutations occur independently, then it follows that the random variable  $m$  has a binomial distribution with the probability density function

$$P [m = x] = f(x) = \binom{M}{x} \mu_{12}^x (1 - \mu_{12})^{M-x} \tag{5.3}$$

for  $x = 0, 1, 2, \dots, M$ . When  $M$  is large, then the probability density function in 5.3 may be approximated by a Poisson distribution with parameter  $M\mu_{12}$ .

As a first step in formulating a model in a host population with respect to one locus, consider a population of plants that are homozygous for a gene  $R$  for resistance to some pathogen. In this case, all plants in a population would be of genotype  $RR$ . It will be assumed that the population reproduces by selfing and that mutations may occur. Specifically, it will be assumed that allele  $R$  may mutate to allele  $r$  with probability  $\nu_{12}$  per generation, and that allele  $r$  may back mutate to allele

$R$  with probability  $\nu_{21}$  per generation, see 5.2. Given a population of individuals of genotype  $RR$ , it will be necessary to describe the set of events that may occur when mutation occurs in the evolution of a population. In particular, under the assumption that mutations may occur, in what follows formulas will be derived for the distribution of the three genotypes  $RR$ ,  $Rr$  and  $rr$  among the offspring of parents with genotypes  $RR$ ,  $Rr$  and  $rr$ .

Observe that under the assumption that copies of allele  $R$  mutate independently, it follows that

$$(1 - \mu_{12}) \times (1 - \mu_{12}) = (1 - \mu_{12})^2 \tag{5.4}$$

is the probability per generation that neither of the copies of the allele  $R$  in the genotype  $RR$  mutate to allele  $r$ . Suppose that among the offspring of a plant of genotype  $RR$  an offspring of genotype  $Rr$  is found, indicating that a mutation of the form  $R \rightarrow r$  has occurred, This mutation can occur in two ways. Suppose the left allele in the genotype  $RR$  does not mutate with probability  $(1 - \mu_{12})$ , but the right allele does mutate with probability  $\mu_{12}$ . In this case, the probability of the occurrence of the mutant genotype  $Rr$  is  $(1 - \mu_{12})\mu_{12}$ . It can be seen by the same line of reasoning, that the probability that that left allele in the genotype  $RR$  mutates but the right allele dose not mutate is  $\mu_{12}(1 - \mu_{12})$ . The two mutational events just described are disjoint. Therefore, the probability of finding an offspring of genotype  $Rr$  among the offspring of an individual of genotype  $RR$  is

$$(1 - \mu_{12})\mu_{12} + \mu_{12}(1 - \mu_{12}) = 2\mu_{12}(1 - \mu_{12}). \tag{5.5}$$

The genotype  $rr$  may also be occur among the offspring of individuals of genotype  $RR$ . In this case, both the left and right alleles in the genotype  $RR$  have mutated to the allele  $r$ . The probability of this rare event is

$$\mu_{12}^2. \tag{5.6}$$

The next step in formulating the model of mutation under consideration is to derive formulas for the probabilities of finding mutant genotypes among the offspring of individuals of genotype  $Rr$ . The probability that neither of the alleles in the genotype mutate is

$$(1 - \mu_{12})(1 - \mu_{21}). \tag{5.7}$$

The probability that the right allele  $r$  in the genotype  $Rr$  back mutates to allele  $R$  is  $\nu_{21}$ . Therefore, the probability of finding an individual of genotype  $RR$  among the offspring of an individual of genotype  $Rr$  is

$$(1 - \mu_{12})\mu_{21}. \tag{5.8}$$

Similarly, the probability of finding an individual of genotype  $rr$  among the offspring of an individuals of genotype  $Rr$  is

$$\mu_{12}(1 - \mu_{21}). \tag{5.9}$$

The last step in the process of deriving formulas for finding mutant genotypes among the offspring of individuals of genotype  $rr$ . The probability that there are no mutant offspring among the offspring of an individual of genotype  $rr$  is

$$(1 - \mu_{21})^2. \tag{5.10}$$

The probability of finding an individual of genotype  $Rr$  among the offspring on an individual of genotype  $rr$  is

$$2\mu_{21}(1 - \mu_{21}). \tag{5.11}$$

Note that the argument used to derive this formula is the same as that used in deriving formula 5.5. Finally, the probability of finding an offspring of genotype  $RR$  among the offspring of an individual of genotype  $rr$  is

$$\mu_{21}^2. \tag{5.12}$$

The next step towards reaching the goal stated above is to derive a formula the probabilities of finding genotypes  $RR$ ,  $Rr$  and  $rr$  among the offspring an individual of the genotype  $RR$ . Let  $TOR_{RR}$  denote the number of offspring on an individual of genotype  $RR$ . From 5.4, 5.5 and 5.6, it can be seen that the total probability of finding non-mutant as well as mutant genotypes among the offspring of an individual of genotype  $RR$  is

$$P[TOR_{RR}] = (1 - \mu_{12})^2 + 2\mu_{21}(1 - \mu_{21}) + \mu_{12}^2 \tag{5.13}$$

As an aid in deriving formulas for the probabilities of finding mutant offspring  $RR$ ,  $Rr$  and  $rr$  among the offspring of genotype  $RR$  under the assumption mutation, a random variable  $X_{RR}$  will be introduced to indicate the genotype of an offspring of genotype  $RR$ . Therefore, the conditional probability of finding an offspring of genotype  $RR$  among those of genotype  $RR$  is

$$P[X_{RR} = RR | TOT_{RR}] = \frac{(1 - \mu_{12})^2}{(1 - \mu_{12})^2 + 2\mu_{21}(1 - \mu_{21}) + \mu_{12}^2} \tag{5.14}$$

Similarly,

$$P[X_{RR} = Rr | TOT_{RR}] = \frac{2\mu_{21}(1 - \mu_{21})}{(1 - \mu_{12})^2 + 2\mu_{21}(1 - \mu_{21}) + \mu_{12}^2} \tag{5.15}$$

and

$$P[X_{RR} = rr | TOT_{RR}] = \frac{\mu_{12}^2}{(1 - \mu_{12})^2 + 2\mu_{21}(1 - \mu_{21}) + \mu_{12}^2} . \tag{5.16}$$

Let  $X_{Rr}$  denote a random variable indicating the genotype of an offspring of an parental individual of genotype  $Rr$ . Observe that

$$TOT_{Rr} = (1 - \nu_{12})(1 - \nu_{21}) + (1 - \nu_{12})\nu_{21} + (1 - \nu_{21})^2 . \tag{5.17}$$

Therefore, from 5.8 it follows that

$$P[X_{Rr} = RR | TOT_{Rr}] = \frac{(1 - \mu_{12})\mu_{21}}{(1 - \mu_{12})(1 - \mu_{21}) + (1 - \mu_{12})\mu_{21} + (1 - \mu_{21})^2} . \tag{5.18}$$

Similarly, from 5.9 it can be seen that

$$P[X_{Rr} = Rr | TOT_{Rr}] = \frac{2\mu_{21}(1 - \mu_{21})}{(1 - \mu_{12})(1 - \mu_{21}) + (1 - \mu_{12})\mu_{21} + (1 - \mu_{21})^2} \tag{5.18}$$

and from 5.10 it can be see that

$$P[X_{Rr} = rr | TOT_{Rr}] = \frac{\mu_{12}(1 - \mu_{21})}{(1 - \mu_{12})(1 - \mu_{21}) + (1 - \mu_{12})\mu_{21} + (1 - \mu_{21})^2} \tag{5.19}$$

For the case of finding genotypes  $RR$ ,  $Rr$  and  $rr$  among the offspring of an individual of genotype  $rr$ , the total probability of a mutation is

$$TOT_{rr} = \mu_{21}^2 + 2\mu_{21}(1 - \mu_{21}) + (1 - \mu_{21})^2 \tag{5.20}$$

Therefore,

$$P[X_{rr} = RR | TOT_{rr}] = \frac{\mu_{21}^2}{\mu_{21}^2 + 2\mu_{21}(1 - \mu_{21}) + (1 - \mu_{21})^2} . \tag{5.21}$$

By using the same line of reasoning, it follows that

$$P[X_{rr} = Rr | TOT_{rr}] = \frac{2\mu_{21}(1 - \mu_{21})}{\mu_{21}^2 + 2\mu_{21}(1 - \mu_{21}) + (1 - \mu_{21})^2} \tag{5.22}$$

and

$$P[X_{rr} = rr | TOT_{rr}] = \frac{(1 - \mu_{21})^2}{\mu_{21}^2 + 2\mu_{21}(1 - \mu_{21}) + (1 - \mu_{21})^2} . \tag{5.23}$$

The expressions in the formulas derived above will be interpreted as elements in three dimensional  $1 \times 3$  vectors that are defined below.

$$\begin{aligned} p_{RR} &= (P[X_{RR} = RR | TOT_{RR}], P[X_{RR} = Rr | TOT_{RR}], P[X_{RR} = rr | TOT_{RR}]) \\ p_{Rr} &= (P[X_{Rr} = RR | TOT_{Rr}], P[X_{Rr} = Rr | TOT_{Rr}], P[X_{Rr} = rr | TOT_{Rr}]) \\ p_{rr} &= (P[X_{rr} = RR | TOT_{rr}], P[X_{rr} = Rr | TOT_{rr}], P[X_{rr} = rr | TOT_{rr}]) \end{aligned} \tag{5.24}$$

In the next section., it will be shown that these three vectors play an essential role in Monte Carlo simulation experiments designed to study the occurrence of virulent genotypes in the pathogen and genes for susceptibilities in the host that arise the host and pathogen populations by the process of mutation.

### 6. Evolution of the Host Population as a Multitype Branching Processes

In this section, the formulation of a multitype branching process evolving on a time scale of discrete generations will be given along with a description of algorithms to compute a sample of Monte Carlo realizations of such a multitype branching processes. The class of multitype branching processes considered in this section, is an extension of the one type branching process described in section 1. To illustrate the algorithms underlying a Monte Carlo simulation procedures to simulate a sample of realizations of this stochastic process, for the sake of simplicity, attention will be focused on the case of  $m = 3$  types, which will be referred as genotypes. Let  $\mathbb{G} = \{1, 2, 3\}$  denote the set of three genotypes. For the case of the host with one locus with two alleles  $R$  and  $r$ , these three genotypes would be  $RR, Rr$  and  $rr$  as illustrated in section 2.

The number of offspring will be characterized in terms of random variables  $N_\nu$  for  $\nu \in \mathbb{G}$ , taking values in the set  $I = \{n \mid n = 0, 1, 2, 3, \dots\}$  of non-negative integers. The probability density functions of these random variables will be denoted by

$$P[N_\nu] = g_\nu(n) \text{ for } n \in I \tag{6.1}$$

and  $\nu \in \mathbb{G}$ . The expected value of a random variable  $N_\nu$  is

$$E[N_\nu] = \lambda_\nu \geq 0 \tag{6.2}$$

and may be interpreted as the average number of offspring contributed to the next generation by each genotype  $\nu \in \mathbb{G}$ .

As in the foregoing sections of this paper, the probability density functions for the random variables  $N_\nu$ ,  $\nu \in \mathbb{G}$ , will be chosen as the simple Poisson densities

$$g_\nu(n) = \exp[-\lambda_\nu] \frac{\lambda_\nu^n}{n!} \tag{6.3}$$

for  $n \in \mathbb{N}$  and  $\nu \in \mathbb{G}$ . It is easy to show that for the density in (6.3),  $E[N_\nu] = \lambda_\nu > 0$  so that the measure of reproductive success  $\lambda_\nu$  is the parameter for a Poisson density for each genotype. In the experiments reported in this paper, a decision was made to consider only the special case of Poisson distributions for the random variables  $N_\nu$  for  $\nu = 1, 2, 3$ . It is well known for this simple distribution that the expectation and variance both equal the parameter  $\lambda$ .

A multitype branching process in discrete time may be defined as follows. In generation  $t$ , where  $t = 0, 1, 2, 3, \dots$ , let the random function  $X_i(t)$ , taking values in the set  $\mathbb{N}$ , denote the number of individuals of genotype  $i \in \mathbb{G}$  in generation  $t$ , and let

$$\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t)) \tag{6.4}$$

denote a vector of these random functions. For  $t = 0$ , an experimenter needs to assign initial values in the set  $\mathbb{N}$  to each of the elements in the vector (6.4). Let the  $1 \times 3$  vector  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3})$  denote the total number of offspring each genotype produced by genotype  $i \in \mathbb{G}$  in any generation. For each genotype  $i$ , let

$$Y_{i1}, Y_{i2}, \dots \tag{6.5}$$

denote a sequence of conditionally independent random variables, given  $X_i(t)$  the number of individuals of genotype  $i$  in generation  $t$ . Each of the random variables in this sequence has a Poisson distribution with parameter  $\lambda_i$ . Then, the number of individuals in the population at time  $t + 1$  of genotype  $i$  is

$$X_i(t + 1) = \sum_{\nu=1}^{X_i(t)} Y_{i\nu} \tag{6.6}$$

for  $i = 1, 2, 3$ .

The next step in the formulation of the model is to take into account mutation. To include mutations in the model, it will be necessary to introduce the multinomial distribution. The probability density function of the multinomial distribution for the case of three dimensions is

$$f(z_1, z_2, z_3) = \frac{n!}{z_1!z_2!z_3!} p_1^{z_1} p_2^{z_2} p_3^{z_3} \tag{6.7}$$

where

$$z_1 + z_2 + z_3 = n \tag{6.8}$$

and

$$p_1 + p_2 + p_3 = 1 \tag{6.9}$$

and  $p_i > 0$  for  $i = 1, 2, 3$ .

Let  $p_{RR}$  denote the  $1 \times 3$  row vector defined in 5.24. Then for an individual of genotype  $RR$ , when mutations occur the number of offspring of each of the three genotypes,  $RR$ ,  $Rr$  and  $rr$ , is given by the  $1 \times 3$  row vector  $OFF_{RR}$ , which is a realization of a multinomial distribution with probability vector  $p_{RR}$  and sample size  $n = X_i(t + 1)$ , with genotype  $i = RR$ . For the case of genotype  $Rr$ , the number of offspring of each genotype is given by the  $1 \times 3$  vector  $OFF_{Rr}$ , which is a realization of a multinomial distribution with probability vector  $p_{Rr}$  and sample size  $n = X_i(t + 1)$  with  $i = Rr$ . Lastly, for the case of genotype  $rr$  the number of offspring of the three genotypes is given by the  $1 \times 3$  vector  $OFF_{rr}$ , which is a realization of a multinomial distribution with a  $1 \times 3$  probability vector  $p_{rr}$  and sample size  $n = X_i(t + 1)$  with  $i = rr$ . An algorithm for simulating a realization of a multinomial random vector may be found in the paper by Mode and Gallop 2008.

It will be helpful to arrange the realization of the three  $1 \times 3$  vectors in the form of a  $3 \times 3$  matrix of the form

$$\begin{bmatrix} off_{11} & off_{12} & off_{13} \\ off_{21} & off_{22} & off_{23} \\ off_{31} & off_{32} & off_{33} \end{bmatrix} \tag{6.10}$$

when it is necessary to compute the number of individuals of each genotype when mutations occur in a population. For example, when mutations are taken into account, the number of individuals of genotypes  $RR$  is at time  $t + 1$

$$X_{RR}(t + 1) = \sum_{j=1}^3 off_{j1} . \tag{6.11}$$

Similarly, when mutations are taken into account the numbers of individuals in the population of genotypes  $Rr$  and  $rr$  are given by the sums

$$X_{Rr}(t + 1) = \sum_{j=1}^3 off_{j2} \tag{6.12}$$

and

$$X_{rr}(t + 1) = \sum_{j=1}^3 off_{j3} \tag{6.13}$$

**7. Embedding a Deterministic Model in a Stochastic Process**

Let  $Z_1, Z_2, \dots$  denote a sequence of random variables taking values in the set of real numbers  $R = (-\infty, \infty)$ , and suppose the expectation  $E[Z_i^2]$  is finite for every  $i = 1, 2, \dots$ . It is of interest to find the best estimator of the random variable  $Z_{n+1}$ , given the random variable  $Z_n$ . Let  $\widehat{Z}_{n+1}$  denote this estimator. Then it is well known that the conditional expectation

$$\widehat{Z}_{n+1} = E[Z_{n+1} | Z_n] \tag{7.1}$$

minimizes the expectation

$$E\left[(Z_{n+1} - \widehat{Z}_{n+1})^2\right] . \tag{7.2}$$

Thus  $\widehat{Z}_{n+1}$  in 7.1 is the best estimator of  $Z_{n+1}$  in the sense of least squares. In what follows, conditional expectations of the form 7.1 will be used extensively .

The objective of this section is to describe a procedure for embedding a deterministic model in a stochastic process. As a first step in describing the derivation of the embedded model, consider the Galton-Watson process defined by equation 1.4 in section 1 by the equation

$$X_\nu = \sum_{k=1}^{X_{\nu-1}} \xi_k \tag{7.3}$$

for  $\nu = 1, 2, \dots$ . In this case, for a given  $\nu$

$$E[X_\nu | X_{\nu-1}] = E[X_{\nu-1}]\lambda \tag{7.4}$$

because for each  $k$  it is assumed that the random variable  $\xi_k$  has a Poisson distribution with expectation  $\lambda$ . As indicated above, let

$$\widehat{X}_i = \widehat{X}_{i-1}\lambda \tag{7.5}$$

for  $\nu = 1, 2, \dots$ .

For  $\nu = 1$  equation 7.5 takes the form

$$\widehat{X}_1 = E [X_0] \lambda , \tag{7.6}$$

where  $E [X_0] = X_0$ , the initial size of the population, which is an assigned number. Thus,

$$\widehat{X}_1 = X_0 \lambda . \tag{7.7}$$

is known. From equations 7.5 and 7.7, it follows that

$$\widehat{X}_2 = E [X_2 | X_1] = \widehat{X}_1 \lambda . \tag{7.8}$$

By using the process just described, it can be shown that

$$\widehat{X}_k = \widehat{X}_{k-1} \lambda \tag{7.9}$$

for  $k \geq 1$ . By definition, equation 7.9 is the deterministic model embedded in the Galton-Watson process.

For the case of a multitype branching process described in section 6, the basic equation of the process is

$$X_i(t + 1) = \sum_{\nu=1}^{X_i(t)} Y_{i\nu} \tag{7.10}$$

for genotypes  $i = 1, 2$  and  $3$ , where given  $X_i(t), Y_{i1}, Y_{i2}, \dots$  is a sequence of conditionally independent Poisson random variables for expectation  $\lambda_i$  for each genotype  $i = 1, 2, 3$ . For each genotype  $i$ , it can be shown that

$$E [X_i(t + 1) | X_i(t)] = X_i(t) \lambda_i . \tag{7.11}$$

Therefore, just as in the case of a one type branching process, the recursive equation

$$\widehat{X}_i(t + 1) = \widehat{X}_i(t) \lambda_i \tag{7.12}$$

will be used to calculate estimates of the sample functions for each genotype  $i = 1, 2, 3$  and  $t = 0, 1, 2, \dots$ .

To take mutation into account in the embedded deterministic model, it will be necessary the use the modified matrix

$$\begin{bmatrix} \widehat{off}_{11} & \widehat{off}_{12} & \widehat{off}_{13} \\ \widehat{off}_{21} & \widehat{off}_{22} & \widehat{off}_{23} \\ \widehat{off}_{31} & \widehat{off}_{32} & \widehat{off}_{33} \end{bmatrix} , \tag{7.13}$$

which was defined in section 6 in equation 6.10. The symbol  $\widehat{off}$  indicated that the elements of the matrix have been calculated using estimates of the sample functions of the process. In section 6, the elements of the matrix in 7.13 were calculated as realizations of multinomial random vectors. But, in the embedded deterministic model the rows of the desired matrix, will be calculated as the mean or expectation of a vector with a multinomial distribution.

Consequently, the modified version of the matrix in 7.13 has the form

$$\begin{bmatrix} \widehat{X}_1 \widehat{off}_{11} & \widehat{X}_1 \widehat{off}_{12} & \widehat{X}_1 \widehat{off}_{13} \\ \widehat{X}_2 \widehat{off}_{21} & \widehat{X}_2 \widehat{off}_{22} & \widehat{X}_2 \widehat{off}_{23} \\ \widehat{X}_3 \widehat{off}_{31} & \widehat{X}_3 \widehat{off}_{32} & \widehat{X}_3 \widehat{off}_{33} \end{bmatrix} \tag{7.14}$$

In this equation,  $\widehat{X}_1$  is actually  $\widehat{X}_1(t + 1)$ , but for the sake of simplicity, the symbol  $(t + 1)$  was not shown in matrix 7.14. Given this matrix, the estimates of the number of each of the three genotypes at time  $t + 1$  are as follows:

$$\begin{aligned} \widehat{X}_1(t + 1) &= \sum_{\nu=1}^3 \widehat{X}_\nu \widehat{off}_{\nu 1} \\ \widehat{X}_2(t + 1) &= \sum_{\nu=1}^3 \widehat{X}_\nu \widehat{off}_{\nu 2} \\ \widehat{X}_3(t + 1) &= \sum_{\nu=1}^3 \widehat{X}_\nu \widehat{off}_{\nu 3} \end{aligned} \tag{7.15}$$

### 8. Evolution of the Pathogen Population as a Multitype Branching Process

For the sake of simplicity, it will be assumed that the pathogen is a haploid so that each individual of the pathogen population has only one copy of a gene at each locus. In what follows the genotype of an individual in the pathogen population will be denoted by the symbols  $A$  and  $a$ , which denote avirulence and virulence respectively. According to the mutation matrix in 5.1, gene  $A$  of the pathogen mutates to gene  $a$  with probability  $\nu_{12}$  per generation, and gene  $a$  mutates back to gene  $A$  with probability  $\nu_{21}$  per generation.

At time  $t$  let the random variable  $W_1(t)$  denote the number of individuals of genotype  $A$  in the pathogen population, and let the random variable  $W_2(t)$  denote the number of individuals of genotype  $a$  in the pathogen population at time  $t$ . Let  $\xi_j(i)$  for  $j = 1, 2, \dots$  denote a sequence of conditionally independent Poisson random variables with parameter  $\lambda_i$ , where  $i = 1$  or  $i = 2$  denotes the genotype of an individual in the pathogen population. Then the evolution of the number of individuals of genotype  $1 = A$  in the pathogen population is governed by the recursive equation

$$W_1(t + 1) = \sum_{j=1}^{W_1(t)} \xi_j(1) , \tag{8.1}$$

for  $t = 0, 1, 2, \dots$ . If  $W_1(t) = 0$ , then  $W(t + 1) = 0$  for all  $t$ . Similarly, the evolution of the number of individuals in the pathogen population of genotype  $2 = a$  has the form

$$W_2(t + 1) = \sum_{j=1}^{W_2(t)} \xi_j(2) , \tag{8.2}$$

for  $t = 0, 1, 2, \dots$ . To initialize the recursive procedures in equations 8.1 and 8.2, the number of individuals of each genotype in the initial population,  $W_1(0)$  and  $W_2(0)$ , must be specified by an experimenter. For each  $j$  the expectation of the Poisson random variables are  $E[\xi_j(1)] = \lambda_1$  and  $E[\xi_j(2)] = \lambda_2$ . Because genotype  $2$  is  $a$ , the gene for virulence, the values of  $\lambda_1$  and  $\lambda_2$  chosen by an experimenter must satisfy the inequality  $\lambda_1 < \lambda_2$  to take into account that virulent genotypes of the pathogen will have many more offspring per generation than an avirulent genotype.

The next step in the formulation of the pathogen process is to take into account that mutations may occur in populations avirulent individuals that are of genotype  $A$ . Mutations in individuals of the virulent genotype  $a$  will also be included in the formulation. According to the  $2 \times 2$  matrix of mutation probabilities in 5.1 for the pathogen, the mutation probabilities for genotype  $A$  are  $\nu_{11}$  and  $\nu_{12}$ , and those for genotype  $a$  are  $\nu_{21}$  and  $\nu_{22}$ . Let  $OA(t) = (n_{A1}(t), n_{a1}(t))$  denote a  $1 \times 2$  vector of the numbers of genotypes  $A$  and  $a$  that are offspring of genotype  $A$  at time  $t + 1$ . Then, by assumption, the vector  $OA(t + 1)$  has a multinomial distribution with sample size  $W_1(t + 1)$  and probability vector  $p_A = (\nu_{11}, \nu_{12})$  at time  $t + 1$  in the evolution of the pathogen process. Similarly, let the  $2 \times 1$  vector  $Oa(t + 1) = (n_{A2}(t + 1), n_{a2}(t + 1))$  denote the number of offspring of genotype  $a$  with genotypes  $A$  and  $a$ . By assumption this vector also has a multinomial distribution with sample size  $W_2(t)$  and probability vector  $p_a = (\nu_{21}, \nu_{22})$ .

To complete the derivation of the process for the pathogen in any generation, it will be helpful for arrange the two vectors just derived in the  $2 \times 2$  matrix

$$\begin{bmatrix} n_{A1}(t + 1) & n_{a1}(t + 1) \\ n_{A2}(t + 1) & n_{a2}(t + 1) \end{bmatrix} . \tag{8.3}$$

From this matrix it can be seen that, when mutation is taken into account, the number of individuals of genotype  $A$  in the population at time  $t + 1$  is

$$W_1(t + 1) = \sum_{j=1}^2 n_{Aj}(t + 1) , \tag{8.4}$$

and the number of individuals of genotype  $a$  in the population at time  $t + 1$  is

$$W_2(t + 1) = \sum_{j=1}^2 n_{aj}(t + 1) \tag{8.5}$$

By using procedures similar to those in section 7, a deterministic model may be embedded in the pathogen process under consideration. For example, the recursive deterministic model corresponding to equation 8.1 is

$$\widehat{W}_1(t + 1) = \widehat{W}_1(t) \lambda_1 \tag{8.6}$$

for  $t = 0, 1, 2, \dots$ . Similarly, the recursive deterministic model corresponding to equation 8.2 is

$$\widehat{W}_2(t + 1) = \widehat{W}_2(t) \lambda_2 . \tag{8.7}$$

To include mutation in the embedded deterministic model, the probabilities in the mutation matrix

$$\begin{bmatrix} \nu_{11} & \nu_{12} \\ \nu_{21} & \nu_{22} \end{bmatrix} \tag{8.8}$$

will play a fundamental role. The estimated expectation matrix for the numbers of mutations for the embedded deterministic model is

$$\begin{bmatrix} \widehat{W}_1(t + 1) \nu_{11} & \widehat{W}_1(t + 1) \nu_{12} \\ \widehat{W}_2(t + 1) \nu_{21} & \widehat{W}_2(t + 1) \nu_{22} \end{bmatrix} . \tag{8.9}$$

Therefore, it follows that the estimated number of individuals of genotype 1 in the pathogen population at time  $t + 1$  is

$$\widehat{W}_1(t + 1) = \sum_{k=1}^2 \widehat{W}_k(t + 1) \nu_{k1} \tag{8.10}$$

Similarly, the estimated number of individuals of genotype 2 in the pathogen population at time  $t + 1$  is

$$\widehat{W}_2(t + 1) = \sum_{k=1}^2 \widehat{W}_k(t + 1) \nu_{k2} . \tag{8.11}$$

### 9. Generation Times of Small Grains, Other Cultivars, Their Pathogens and Balanced Polymorphisms

An essential ingredient of an evolutionary model of small grains, such as wheat and barley, is their generation times. By definition, a generation time, is the time from the germination of seeds to the time that the mature plants produce seeds. For those varieties of wheat and barley that are planted in the spring in northern temperate regions of the earth, the generations times of these species are in the range of 90 for 100 days. The generation times for varieties of flax in northern temperate regions also are in the range of 90 to 100 days. When these species are grown in the southern temperate regions of the earth, the generation times of the species under consideration are essentially the same.

The generation times of the pathogens, which vary among pathogens, are usually much shorter than those of the host. In the computer experiments reported in this paper, it will be assumed that the generation time of a pathogen is 10 days. Under this assumption, for every generation of the host, there are about 10 generations of the pathogen that will be simulated. Mutations in the host and pathogen often occur during the reproduction process in connection with the copying of *DNA* in both the host and pathogen when cells divide. Because for every generation of the host, there will be 10 generations of the pathogen, it is more likely that during every growing season, there many more mutations in the pathogen than the host.

After the seeds sprout in the hosts under consideration, there is a rapid increase in biomass, in the forms of leaves, stems and structures connected with reproduction, such as heads that contain the seeds in wheat and barley and structures that contain the seeds in flax. To model the process of plant growth after germination would technically be a very difficult. Consequently, no attempt will be made to model plant growth in this paper. But, it will be tacitly assumed that plant grow does occur in computer simulation experiments and provides a medium on which a pathogen grows and damages the plant. The procedures described in section 1 will be used in all computer simulation experiments to quantify the damage the pathogen does to the host in each generation of a computer experiment.

The term, balanced polymorphisms, is used in evolutionary biology. From the mathematical point of view, it refers to the structure that results form the convergence of a Host-Pathogen model to a limit. For the case of a deterministic model the so called balanced polymorphism will be the constant structure that is the limit of the solution of deterministic equations, describing the evolution of a host-pathogen system as  $t \rightarrow \infty$  when it exists. For the case of a stochastic model describing the evolution of a host-parasite population, in some formulations, such as a Markov processes, the model will converge in distribution to what is called a quasi-stationary distribution. If a reader is interested of examples in which a stochastic model converges to a quasi-stationary distribution, the book by Mode and Sleeman (2012) may be consulted. For a so called quasi - stationary distribution, the mean and variance of the distribution remains constant for all  $t$  after convergence. In the case of a stochastic model, the quasi-station distribution may be referred to as a balanced polymorphism. The models described in the foregoing sections are Markov processes.

## References

- Flor, H. H. (1955). Host-parasite interaction in flax rust-its genetics and other implications. *Phytopath.*, 45, 680-685.
- Flor, H. H. (1956). *The complementary genic systems in flax and flax rust*. In *Advances in Genetics*, New York: Academic Press, Inc. [https://doi.org/10.1016/S0065-2660\(08\)60498-8](https://doi.org/10.1016/S0065-2660(08)60498-8)
- Liu, B. H. (1998). *Statistical Genomics, Linkage, Mapping and QTL Analysis*. CRC Press, Boca Raton and New York.
- Liu, Z., Holmes, D., Faris, J. D., Chao, S., Brueggeman, R. S., Edwards, M. C., & Friesen, T. L. (2015). *QTL mapping reveals effector-triggered susceptibility underlying the barley-Pyrenophora teres f. teres interaction*. *Molecular Plant Pathology*.
- Mascher, M., Gundlach, H., & Himmelbach, A. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544: 427-433.
- Mode, C. J. (1958). A mathematical model for the co-evolution of obligate parasites and their hosts. *Evolution*, 12, 158-165. <https://doi.org/10.2307/2406026>
- Mode, C. J. (1961). A Generalized Model of a Host-Pathogen System. *Biometrics*, 17, 386-403. <https://doi.org/10.2307/2527833>
- Mode, C. J. (1964). A Stochastic Model of the Dynamics of Host-Pathogen Systems With Mutation. *The Bulletin of Mathematical Biophysics*, 26, 205-233. <https://doi.org/10.1007/BF02479043>
- Mode, C. J., & Gallop, R. J. (2008). A review of Monte Carlo simulation methods as they apply to selection and mutation in Wright-Fisher models of evolutionary genetics. *Mathematical Biosciences*, 211, 205-225.
- Mode, C. J., & Sleeman, C. K. (2012). *Stochastic Processes in Genetics and Evolution*. World Scientific, New Jersey, London, Singapore, Beijing, and Hong Kong. <https://doi.org/10.1142/8159>
- O' BOYLE, P. D. (2009). *Genetic characterization and linkage mapping of barley and net blotch resistance genes*. Ph. D. thesis, Department of Crop and Soil Sciences, Virginia Polytechnic Institute and State University.
- Richards, J. K. (2016). *Genetic and molecular characterization of host resistance and susceptibility to Pyrenophora F. teres in Hordeum vulgare*. Ph. D. thesis, Department of Plant Pathology, North Dakota State University.
- Richards, J., Chao, S., Friesen, T., & Brueggeman, R. (2016). Fine mapping of the barley chromosome 6H net form net blotch susceptibility locus. *G3: Genes, Genomes, Genetics*, 6(7), 1809-1818. <https://doi.org/10.1534/g3.116.028902>

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).