

# Algebraic structure of non-Mendelian inheritance

Jianjun Paul Tian

Mathematical Biosciences Institute

The Ohio State University, Columbus OH 43210

## **Abstract**

Evolution algebra theory is used to study non-Mendelian inheritance, particularly organelle heredity and population genetics of *Phytophthora infestans*. We can not only explain a puzzling feature of establishment of homoplasmy from heteroplasmic cell population and the coexistence of mitochondrial triplasm, but also predict all mechanisms to form the homoplasmy of cell populations, which are hypothetical mechanisms in current mitochondrial disease research. The algebras also provide a way to easily find different genetically dynamic patterns from the complexity of the progenies of *Phytophthora infestans* which cause the late blight of potatoes and tomatoes. Certain suggestions to pathologists are made as well.

## **1 introduction**

In this article, We shall apply evolution algebra theory to the study of the non-Mendelian genetics. As Mendelian genetic models, non-Mendelian inheritance is a huge family in genetics. We so focus on two particular groups of genetic phenomena to see how evolution algebras to work for them. One is organelle population genetics, the other one is *Phytophthora infestans*

population genetics. Organelle here we mean chloroplasts and mitochondria. A puzzling feature of organelle inheritance is to establish homoplasmic cell populations from heteroplasmic cell populations over certain cell divisions. Because concepts of algebraic transiency and algebraic persistency catch the essences of biological transitory and biological stability respectively, evolution algebras can be used to explain this feature. The algebras could include any number of mutants as their generators, so modeling the triplasm (partial duplication of mt-DNAs, deletion of mt-DNAs and wild-type mt-DNAs) in some tissue of patients with sporadic mitochondrial disorders seems straightforward. Actually, algebras could explain all mechanisms behind the phenomena. We also study another type of uniparental inheritance about *Phytophthora infestans* which cause the late blight of potatoes and tomatoes. After we construct several relevant evolution algebras for the progeny populations of *Phytophthora infestans*, we could see different genetically dynamic patterns from the complexity of the progenies of *Phytophthora infestans*. We then predict the existence of intermediate transient races and the periodicity of the reproduction of biological stable races. Practically, we can suggest to plant workers to stop the spread of late blight disease in a right phase. Theoretically, we can use our algebras to provide information of *Phytophthora infestans* reproduction rates to Plant pathologists.

we first give a brief reflection of the history related to evolution algebras in the first section. In section 2, we will recall the basic biological components of non-Mendelian genetics and the inheritance of organelle genes. We also give a general algebraic formulation of non-Mendelian genetics in section 2. In section 3, we use evolution algebras to study the heteroplasmy and homoplasmy of organelle populations, and show that concepts of alge-

braic transiency and algebraic persistency catch the essences of biological transitory and stability respectively. Coexistence of triplasmmy in tissues of patients with sporadic mitochondrial disorders is also studied as well. In section 4, we apply evolution algebra theory to the study of asexual progenies of *Phytophthora infestans*.

## 2 History of general genetic algebras

In history, Mendel in his first paper [1] exploited some symbolism, which is quite algebraically suggestive, to express his genetic laws. In fact, it was later termed "Mendelian algebras" by several authors. In the 1920s and 1930s, general genetic algebras were introduced. Apparently, Serebrowsky [2] was the first to give an algebraic interpretation of the sign " $\times$ ", which indicated sexual reproduction, and to give a mathematical formulation of the Mendelian laws. Glivenkov [3] continued to work at this direction and introduced the so-called Mendelian algebras for diploid populations with one locus or two unlinked loci. Independently, Kostitzin [4] also introduced a "symbolic multiplication" to express the Mendelian laws. The systematic study of algebras occurring in genetics was due to I. M. H. Etherington. In his series of papers [5], he succeeded in giving a precise mathematical formulation of Mendel's laws in terms of non-associative algebras. He pointed out that the nilpotent property is essential to these genetic algebras and formulated it in his definition of train algebra and baric algebra. He also introduced the concept of commutative duplication by which the gametic algebra of a randomly mating population is associated with a zygotic algebra. Besides Etherington, fundamental contributions have been made by Gonshor [6], Schafer [7], Holgate [8, 9], Hench [10], Reiser [11], Abraham [12],

Lyubich [14], and Worz-Busekos [13]. During the early days in this area, it appeared the general genetic algebras or broadly defined genetic algebras, (by these term we mean all algebras or any algebra having been used in Mendelian genetics,) can be developed into a field of independent mathematical interest, because these algebras are in general not associative and do not belong to any of the well-known classes of non-associative algebras such as Lie algebra, alternative algebra, or Jordan algebra. They possess some distinguished properties that lead to many interesting mathematical results. For example, baric algebra, which has nontrivial representation over the underlying field, and train algebra, whose coefficients of rank equation are only functions of the image under this representation, are new objectives for mathematicians. Until 1980s, the most comprehensive reference in this area was book [13]. More recent results and direction, such as evolution in genetic algebras, can be found in the book [14]. A good survey article is Reed's [15].

General genetic algebras are the product of interaction between biology and mathematics. Mendelian genetics offers a new object to mathematics: general genetic algebras. The study of these algebras reveals the algebraic structure of Mendelian genetics, which always simplifies and shortens the way to understand genetic and evolutionary phenomena. Indeed, it is the interplay between the purely mathematical structure and the corresponding genetic properties that makes this area so fascinating. However, after Baur [16] and Correns [17] first detected that chloroplast inheritance departed from the Mendelian rules, and much later, mitochondrial gene inheritance were also identified in the same way, the non-Mendelian inheritance of organelle genes became manifest with two features — uniparental inheritance and vegetative segregation. Now, the non-Mendelian genetics is a

basic language of molecular geneticists. Logically, we could ask what the non-Mendelian genetics offer to mathematics. The answer is "evolution algebras".

### **3 Non-Mendelian inheritance and its algebraic formulation**

#### **3.1 Mendelian vs non-Mendelian inheritance**

Following Birky's paper [18], Mendelian genetics has five components contrasting to non-Mendelian genetics:

- (1) During asexual reproduction, alleles of nuclear genes do not segregate: heterozygous cells produce heterozygous daughters. this is because all chromosomes in nuclear genomes are replicated once and only once in interphase and mitosis ensures that both daughter cells get one copy of each chromosome. In contrast, alleles of organelle genes in heteroplasmic cells segregate during mitotic as well as meiotic divisions to produce homoplasmic cells. This is because in the vegetative division of the organelles, some copies of the organelle genome can replicate more than others by chance or in response to selective pressures or intrinsic advantages in replication, and alleles can segregate by chance.
- (2) Alleles of a nuclear gene always segregate during meiosis, with half of the gametes receiving one allele and half the other. Alleles of organelle genes may or may not segregate during meiosis; the mechanisms are the same as for vegetative segregation.
- (3) Inheritance of nuclear genes is biparental. Organelle genes are often

inherited from only one parent, uniparental inheritance.

- (4) Alleles of different nuclear genes segregate independently. Organelle genes are nearly always on a single chromosome and recombination is often severely limited by uniparental inheritance or failure of organelles to fuse and exchange genomes.
- (5) Fertilization is random with respect to the genotype of the gametes. This is the only part of Mendel's model that applies to organelle as well as nuclear genes.

While most of heredity of nuclear genes obeys Mendel's laws; the inheritance of organelle is not Mendelian. We shall now to introduce the basic of organelle biology.

Cell organelles here we mean chloroplasts and mitochondria, which are sub-structural units in cells. Chloroplasts and mitochondria have endosymbiotic origin. They evolved from free-living prokaryotes. They are now integral parts of eukaryotic cells retaining only vestiges of their original genomes. Yet the genes encoded in these organelle are vital to their function as are the ones they have shed into the nucleus over the millennium. Bioenergetically, chloroplasts and mitochondria complement one another. Chloroplasts derive energy from light that is employed for splitting water and the production of molecular oxygen. The electrons produced from the splitting of water are used via the photosynthetic electron transport chain to drive photosynthetic phosphorylation. Ultimately, molecular  $CO_2$  is reduced by the protons and electrons derived from water and is converted into carbohydrates by the soluble enzymes of the chloroplast stroma. The mitochondrion, in contrast, catalyses the aerobic oxidation of reduced carbon

compounds via soluble enzymes of the tricarboxylic acid cycle found in its matrix. The electrons produced by the oxidation of reduced carbon compounds flow via the respiratory electron transport chain and drive oxidative phosphorylation. The electrons and protons derived from the oxidation of reduced carbon compounds convert molecular oxygen to water and  $CO_2$  is released as an oxidation product of the tricarboxylic acid cycle. In summary, the chloroplast reduces  $CO_2$  and splits water with the release of  $CO_2$ , while the mitochondrion oxidizes reduced carbon compounds with the formation of  $CO_2$  and water. However, chloroplasts and mitochondria are not simple energy-generating and utilizing systems. A vast array of other metabolic processes goes on within their confines as well, which are just as much key to the health and well-being of the cell as electron transport and energy generation. Genetically, mitochondrial and chloroplast (extra-nuclear) genomes are self-replicating units (but not physiologically) independent of the nuclear genome. Remarkably, the best way to think about chloroplast and mitochondrial gene inheritance is in terms of populations of organelle genes inside a single cell or cell line, subject to mutation, selection and random drift. Chloroplasts vary in size, shape, and number per cell. A typical flowering plant has 10 to 200 chloroplasts. For human, all cells contain many copies of mitochondrial genomes, on the order of thousands of molecules of mtDNA [19]. To other animals, the number of mitochondria is also huge. Therefore, it will be more appropriate if we treat the chloroplasts or mitochondria in the cell as if they are a single population inside a single organelle. This way we can take a perspective of population genetics and utilize the methods in population genetics when we think and study organelle inheritance. This is so-called intracellular population genetics of organelles.

Vegetative segregation is the most general characteristic of the inheritance of organelle genes, occurring in both mitochondria and chloroplasts in all individuals or clones of all eukaryotes. In other words, uniparental inheritance is a major means of genetic transmission. Some more knowledge will be introduced when we construct various evolution algebras in the next section.

### 3.2 Algebraic formulation of non-Mendelian inheritance

Let us consider a population of organelles in a cell or a cell clone, and suppose that there are  $n$  different genotypes in this organelle population. Denote these genotypes by  $g_1, g_2, \dots, g_n$ . By non-Mendelian inheritance component (3), the crossing of genotypes is impossible since it is uniparental inheritance. Mathematically, we take

$$g_i \cdot g_j = 0,$$

for  $i \neq j$ . By non-Mendelian inheritance component (2), alleles of organelle genes may or may not segregate during meiosis following vegetative segregation, so that the frequency of each gene in the next generation could be variant. From non-Mendelian inheritance component (4), intramolecular and intermolecular recombination within a lineage provides evidence that one organelle genotype could produce other different genotypes. Therefore, we mathematically define,

$$g_i^2 = \sum_{j=1}^n \alpha_{ij} g_j,$$

where  $\alpha_{ij}$  is positive number that can be interpreted as the rate of genotype  $g_j$  produced by genotype  $g_i$ . Now, we have the algebra defined by generators  $g_1, g_2, \dots, g_n$  which are subject to these relations.

It is obviously that this is a very general definition. But it is general enough to include all non-Mendelian inheritance phenomena. We have developed a general theory for this type of algebras. We will study concrete examples in the rest of the paper.

## 4 Algebras of organelle population genetics

### 4.1 Heteroplasmy and homoplasmy

Organelle population geneticists are usually concerned a special case that there are two different phenotypes or genotypes: homoplasmic and heteroplasmic. Let us denote the heteroplasmic cell by  $g_0$ , and the two different type of homoplasmy by  $g_1$  and  $g_2$  respectively. Just suppose  $g_1$  and  $g_2$  are mutant and wild-type respectively. From the inheritance of organelles we know that heteroplasmic parent can produce both heteroplasmic progeny and homoplasmic progeny, and homoplasmic parent can only produce homoplasmic progeny with the same type if any other mutation is not considered for the moment. The following Figure 1 shows the Wright-Fisher model for organelle genes.

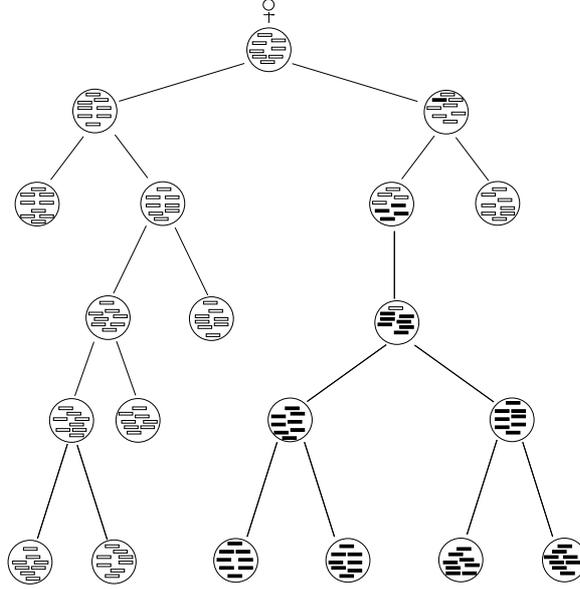


Figure 1. Wright-Fisher model for organelle genes

Therefore, we have the following reproduction relations.

$$g_0^2 = \pi g_0 + \alpha g_1 + \beta g_2, \quad (1)$$

$$g_1^2 = g_1, \quad (2)$$

$$g_2^2 = g_2; \quad (3)$$

and for  $i \neq j$ ,  $i, j = 0, 1, 2$ ,

$$g_i \cdot g_j = 0; \quad (4)$$

where  $\pi$ ,  $\alpha$ ,  $\beta$  are all positive real numbers. Actually, these numbers can be taken as the segregation rates of corresponding types. For any specific example, we can determine these coefficients by combinatorics or modified Wright-Fisher model.

Thus, we have an evolution algebra, denoted by  $A_h$ , generated by  $g_0$ ,  $g_1$  and  $g_2$  and subject to the above defining relations 1 - 4.

By our knowledge of evolution algebras, algebraic generator  $g_0$  is transient;  $g_1$  and  $g_2$  are persistent. Because  $g_1$  and  $g_2$  do not communicate, we have two simple subalgebras of  $A_h$  generated by  $g_1$  and  $g_2$  respectively. Biologically,  $g_0$  is transitory as N. W. Gillham pointed out [20];  $g_1$  and  $g_2$  are of stable homoplasmic cell states. By transitory, biologists mean that the cells of transitory are not stable, they are just transient phases, and they will disappear eventually after certain cell generations. This property is imitated by algebraic transiency. By biological stability, we mean it is not changeable over a period of time, and it is kept the same from generation to generation. This property is imitated by algebraic persistency.

The puzzling feature of organelle heredity is that heteroplasmic cells eventually disappear and the homoplasmic progenies are observed. The underlying biological mechanisms are still unknown. Actually, it is a intensive research field currently since it is related to aging and many other diseases caused by mitochondrial mutations [21],[22]. However, we could mathematically understand this phenomenon from the theory of evolution algebras. Because  $g_0$  is transient,  $g_1$  and  $g_2$  are persistent, we can eventually have two simple subalgebras of  $A_h$ . These two subalgebras are of zero-th in the hierarchy of this evolution algebra, and thus they are stable. The subalgebra generated by  $g_1$  is homoplasmic and mutant; the subalgebra generated by  $g_2$  is homoplasmic and wild-type. Moreover, the mean time  $T_h$  to reach these homoplasmic progeny is given

$$T_h = \frac{1}{1 - \pi}.$$

If, now, we consider a mutant to be lost, say gene  $g_2$  will be lost, we

have the following several ways to model this phenomenon. The algebraic generator set still is  $\{g_0, g_1, g_2\}$ .

Firstly, we think that  $g_2$  disappears in a dramatic way, that is

$$g_2^2 = 0.$$

Other defining relations are 1, 2 and 4. Thus, the evolution algebra we get here is different from  $A_h$ . It has one non-trivial simple subalgebra that is corresponding to homoplasmic progeny generated by  $g_1$ .

Secondly, we consider that  $g_2$  gradually mutates back to  $g_1$ , that is

$$g_2^2 = \eta g_1 + \rho g_2,$$

where  $\eta$  is not zero and could be 1. And other defining relations are 1, 2 and 4. Although we eventually have one simple subalgebra by this relations, the evolution path is different.

Thirdly, we consider that  $g_2$  always keeps heteroplasmic property, that is

$$g_2^2 = \eta g_0 + \rho g_2.$$

The other defining relations are still 1, 2 and 4. Eventually, we have homoplasmic progenies which all are  $g_1$ . That is the only simple subalgebra generated by  $g_1$ .

In conclusion, here we have four different evolution algebras derived from the study of homoplasmy. They are not the same in the skeletons. Therefore, their dynamics, which is actually genetic evolution processes, are different. However, we need to look for what is the biological evidences for defining these different algebras. In Ling et al [22], several hypothetical mechanisms

were put forward for establishment of homoplasmy, and their hypothetical mechanisms are actually corresponding to four different algebraic structures above.

## 4.2 Coexistence of triplasm

In mitochondrial genetics, if we consider different genotypes of mutants instead of just two different phenotypes of homoplasmy and heteroplasmy, we will have higher dimensional algebras that contain more genetic information. Recently, in Tang et al [23], it studied the dynamic relationship among wild-type and rearranged mtDNAs.

Large-scale rearrangements of human mitochondrial DNA (including partial duplications and deletion) are found to be associated with a number of human disorders, including Kearns-Sayre syndrome, progressive external ophthalmoplegia, Pearson's syndrome, and some sporadic myopathies. Each patient usually harbors a heteroplasmic population of wild-type mitochondrial genomes (wt-mtDNA) together with a population of a specific partially deleted genome ( $\Delta$ -mtDNA) in clinically affected tissues. These patients also harbor a third mtDNA species, a partial duplication (dup-mtDNA), as well. To study the dynamic relationship among these genotypes, authors of paper [23] cultured cell lines from two patients. After a long-term (6 months, 210-240 cell divisions) culture of homoplasmic dup-mtDNAs from one patient, they found the culture contained about 80% dup-mtDNA, 10% wt-mtDNA, and 10%  $\Delta$ -mtDNA. After a long-term culture of the heteroplasmic which contains wt-mtDNA and  $\Delta$ -mtDNA from the same patient, they did not find any new cell species, although there were the fluctuations of percentages of these two cell populations. From this same patient, after

cultured two year of  $\Delta$ -mtDNA cell line, they did not find any new cell species. Now, let's formulate this genetic dynamics as an algebra.

Denote triplasmic cell population by  $g_0$  that contain dup-mtDNA, wt-mtDNA and  $\Delta$ -mtDNA, denote heteroplasmy that contains dup-mtDNA and wt-mtDNA by  $g_1$ , heteroplasmy that contains dup-mtDNA and  $\Delta$ -mtDNA by  $g_2$ , heteroplasmy which contains wt-mtDNA and  $\Delta$ -mtDNA by  $g_3$ , and homoplasmy dup-mtDNA by  $g_4$ , homoplasmy wt-mtDNA by  $g_5$ , homoplasmy  $\Delta$ -mtDNA by  $g_6$ . According to the genetic dynamic relations described above, we set algebraic defining relations as follows:

$$g_0^2 = \beta_{00}g_0 + \beta_{01}g_1 + \beta_{02}g_2 + \beta_{03}g_3,$$

$$g_1^2 = \beta_{14}g_4 + \beta_{15}g_5,$$

$$g_2^2 = \beta_{24}g_4 + \beta_{26}g_6,$$

$$g_3^2 = \beta_{35}g_5 + \beta_{36}g_6,$$

$$g_4^2 = \beta_{44}g_4 + \beta_{45}g_5 + \beta_{46}g_6,$$

$$g_5^2 = \beta_{54}g_4 + \beta_{56}g_6,$$

$$g_6^2 = \beta_{64}g_4 + \beta_{65}g_5,$$

and for  $i \neq j$ ,  $i, j = 0, 1, \dots, 6$ ,

$$g_i \cdot g_j = 0.$$

And the generator set is  $\{g_0, g_1, \dots, g_6\}$ . This algebra has three levels of a hierarchy. On the 0-th level, it has one simple subalgebra generated by  $g_4$ ,  $g_5$  and  $g_6$ . These three generators are algebraic persistent. Biologically, they consist of the genotypes that can be observed, and genetically stable.

On the 1-st level, it has three subalgebras; each of them has dimension 1. On the 2-nd level, there is one subalgebra generated by  $g_0$ . Generators on the 1-st and 2-nd levels are all algebraic transient. They are unobservable biologically.

If we have more information about the reproduction rates  $\beta_{ij}$ , we could quantitatively compute certain relevant quantities. For example, let's set

$$\begin{aligned}\beta_{00} = \beta_{01} = \beta_{02} = \beta_{03} &= \frac{1}{4}, \\ \beta_{14} = \beta_{15} &= \frac{1}{2}, \\ \beta_{24} = \beta_{26} &= \frac{1}{2}, \\ \beta_{35} = \beta_{36} &= \frac{1}{2}, \\ \beta_{44} = \frac{5}{6}, \beta_{45} = \beta_{46} &= \frac{1}{12}, \\ \beta_{54} = \frac{2}{3}, \beta_{56} &= \frac{1}{3}, \\ \beta_{64} = \frac{2}{3}, \beta_{65} &= \frac{1}{3}.\end{aligned}$$

We then can compute the long-term frequencies of each genotype in the culture. Actually, the limit of the evolution operator will give the answer. Suppose we start with a transient genotype  $g_0$ , then we have a starting vector  $v_0 = (1, 0, \dots, 0)'$ . As time goes to infinity, we have

$$\lim_{n \rightarrow \infty} L^n v_0 = (0, \dots, 0, 0.80, 0.10, 0.10)'.$$

Therefore, to this patient, we can see the algebraic structure of his mitochondrial genetic dynamics. Besides the experimental results, we could reproduce by our algebraic model, we could predict that there are several

transient phases. These transient phases are algebraic transient generators of the algebra. They are important for medical treatments. If we could have drug to stop the transitions during the transient phases of mitochondrial mutations, we could help these disorder patients.

## **5 Algebraic structures of asexual progenies of *Phytophthora infestans***

In this section, we shall apply evolution algebra theory to the study of algebraic structure of asexual progenies of *Phytophthora infestans* based on experimental results in paper [24]. The basic biology of *Phytophthora infestans* and related experiments are first briefly introduced. Then we will construct evolution algebras for each race of *Phytophthora infestans*. Most of our biological materials is taken from paper [24]and [25].

### **5.1 The basic biology of *Phytophthora infestans***

Oomycetes are a group of organisms in a kingdom separate from the true fungi, plants, or animals. They are included in the Kingdom Protocista or Chromista. This group of organisms is characterized by the absence of chitin in the cell walls (true fungi contain chitin), zoospores with heterokont flagella (one whiplash, one tinsel) borne in sporangia, diploid nuclei in vegetative cells, and sexual reproduction via antheridia and oogonia [25]. The genus *Phytophthora* contains some species including *P. infestans* that are heterothallic (A1 and A2 mating types) and some that are homothallic. The Chromista organism *P. infestans* (Mont.) de Bary, the cause of potato and tomato late blight, is the most important foliar and tuber pathogen of

potato worldwide. The Irish Potato Famine is a well-known result of these early epidemics. Tomato late blight was detected sometime later and has also been a persistent problem. Most scientists agree that the center of origin of *P. infestans* is in the highlands of central Mexico and that this region has been the ultimate source for all known migrations. It was the only location where both mating types of *P. infestans* were found prior to the 1980s. Outside Mexico *P. infestans* populations were dominated by a single clonal lineage that are confined to asexual reproduction [26]. Sexual reproduction of *P. infestans*, associated with genetic recombination during meiosis in the antheridium or the oogonium, is a major mechanism of genetic variation in this diploid organism. However, other mechanism of genetic variability may have a significant role in creating new variants of this pathogen. Mutation, mitotic recombination and parasexual recombination are the most common mechanism of genetic variability in the absence of sexual reproduction [27]. The most important aspect of genetic variability in plant pathogens is the variability in pathogenicity and virulence toward the host. Virulence variability in *P. infestans* populations is recognized as a major reason for failure of race specific genes for resistance in cultivated potato management strategy. The race concept as applied to *P. infestans* refers to possession of certain virulence factors. Isolates sharing the same virulence factors are considered to be a race that can be distinguished from other races possessing other groups of virulence factors. Characterization of isolates to different races is based on their interaction with major genes for resistance in potato. So far 11 major genes for resistance have been identified in *Solanum* spp [28].

In paper [24], five parental isolates of *P. infestans*, PI-105, PI-191, PI-52, PI-126 and PI-1, collected from Minnesota and North Dakota in 1994 to 1996, were chosen to represent different race structures. Single zoospore

progenies were generated from each of the parental strains by inducing zoospore production. The proportion of zoospores that developed into vegetative colonies varied from 2 to 50% depending on the parental isolate. The parental isolate PI-1 produced very small zoospores and the percent recovery of colonies was very low. Other parental isolates produced large sized zoospores and showed higher levels of developed colonies. In total, 102 single zoospore isolates were recovered, 20 isolates from isolate PI-105, 29 isolates from PI-191, 28 isolates from PI-52, 14 isolates from PI-126, and 11 isolates from PI-1. These single zoospore demonstrated different levels of variability for virulence. Although some single zoospore isolates showed the same virulence as their parental isolate, others showed lower or higher virulence than the isolate from which they were derived. Single zoospore isolates derived from PI-1 (11 isolates) were identical in virulence to their parental isolate. Single zoospore isolates derived from isolate PI-191 (29 isolates) showed low levels of variability for virulence compared with their parental isolate; 73% of these isolates (21 isolates) retained the same virulence pattern as their parent. Four isolates gained additional virulence to R8 and R9. One isolate had additional virulence to R9 which was stable. The other two showed lower virulence compared with the parental isolate. Six races were identified from the single zoospore isolates of the parental isolate PI-191.

Single zoospore isolates derived from isolate PI-126 showed higher levels of variability for virulence. Three isolates in this series gained virulence to both R8 and R9, three isolates gained additional virulence to R8, six isolates gained additional virulence to R9, and only two isolates retained the same virulence level of the parental isolate. Four races were identified within this series of isolates.

Isolates derived from the parental isolate PI-52 were highly variable for

virulence. The overall trend in this series of isolates was toward lower virulence relative to the parental isolate. The total number of races identified from this parental isolate is 12.

The single zoospore progeny isolates derived from isolate PI-105 were highly variable for virulence. In this series of isolates there was a tendency for reduced virulence of the single zoospore isolates compared with their parent. 13 races were identified from this set of isolates.

## 5.2 Algebras of progenies of *Phytophthora infestans*

In order to mathematically understand the complexity of structure of progenies of *P. infestans*, we assume that there are 11 loci in genome of *P. infestans* corresponding to the resistant genes, or 11 phenotypes corresponding to the resistant genes, denote by  $\{c_1, c_2, \dots, c_{11}\}$ , and if  $c_j$  functions (is expressed), the progeny resists gene  $R_j$ . Any non-repeated combination of these  $c_j$  could forms a race mathematically. So, we can have 2048 races. For simplicity, we just record a virulence part of a race by  $E_i$ , the complement part is avirulence. For example,  $E_i = \{c_2, c_3, c_5, c_8, c_{10}\}$  represents race type  $c_2c_3c_5c_8c_{10}/c_1c_4c_6c_7c_9c_{11}$ . Take these 2048 races as generators set, we then have a free algebra over real number field  $R$ . Since reproduction of zoospore progeny is asexual reproduction, the generating relations among races are evolution algebra types. That is,

$$E_i^2 = \sum p_{ij} E_j,$$

and if  $i \neq j$

$$E_i \cdot E_j = 0,$$

where  $p_{ij}$  are non-negative numbers. If we interpret  $p_{ij}$  as frequency,

we have  $\sum p_{ij} = 1$ . If we have enough biological information about the generating relations among the races or within one race, we could write the detailed algebraic relations.

For example, let's look at the race PI-126P and its progenies. PI-126P has race type  $E_1 = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_{10}, c_{11}\}$ . It has four different type of progenies:

$$\{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_{10}, c_{11}\} = E_2,$$

$$\{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_9, c_{10}, c_{11}\} = E_3,$$

$$\{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}\} = E_4,$$

and  $E_1$  itself. Actually, these four types of progenies are those that are biologically stable, and we could eventually observe as their outcomes of asexual reproduction. These four types of progenies, as generators algebraically, are persistent elements. There could have many transient generators that produce biologically unstable progenies. These progenies serve as intermediate transient generations, and produces stable progenies. However, the simple evolution algebra without intermediate transient generations that we could construct for race PI-126P seems to have the following defining relations:

$$E_1^2 = p_1 E_2 + q_1 E_3,$$

$$E_2^2 = p_2 E_1 + q_2 E_4,$$

$$E_3^2 = p_3 E_1 + q_3 E_4,$$

$$E_4^2 = r_1 E_1 + r_0 E_4;$$

and if  $i \neq j$ ,

$$E_i \cdot E_j = 0.$$

If we know the frequency  $P_j$  of the  $j$ th race in the population as in paper [24], we could easily set above coefficients. For example, suppose all coefficients have the same value, 0.5, then the algebra generated by PI-126P is a simple evolution algebra. Biologically, this simple evolution algebra means that each race can reproduce other races within the population. We can also compute that the period of each generator, for each race, is 2. This means to reproduce any race itself at least needs to generations later. Eventually, the frequency of races  $E_1$ ,  $E_2$ ,  $E_3$  and  $E_4$  in the whole population are  $\frac{1}{3}$ ,  $\frac{1}{6}$ ,  $\frac{1}{6}$  and  $\frac{1}{3}$  respectively. This can be done by computing

$$\lim_n L^n(E_1),$$

where  $L$  is the evolution operator of the simple algebra.

Now, let's assume there exists an intermediate transient generation, therefore there exists a transient race,  $E_5$ , in the developing process of progeny population of PI-126P. We just assume  $E_5$  is  $\{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_{10}\}$ . Usually, it is very difficult to observe the transient generation biologically. Our evolution algebra is now generated by  $E_1$ ,  $E_2$ ,  $E_3$ ,  $E_4$  and  $E_5$ . Thus the defining relations are given

$$\begin{aligned}
E_1^2 &= p_1 E_2 + q_1 E_3, \\
E_2^2 &= p_2 E_1 + q_2 E_4 + r_2 E_5, \\
E_3^2 &= p_3 E_1 + q_3 E_4, \\
E_4^2 &= r_1 E_1 + r_0 E_4, \\
E_5^2 &= 0
\end{aligned}$$

and if  $i \neq j$ ,

$$E_i \cdot E_j = 0.$$

We could verify that this evolution algebra has a simple subalgebra, which is the algebra in the above example. We can also claim that intermediate transient races will extinct, and they are not biologically stable, when the parental race is within its progeny population. Mathematically, these intermediate transient races are nilpotent elements.

The progeny population of PI-52P displays a distinct algebraic feature.

There are 12 races in the progeny population of PI-52P, and the parental race is not in the population. We name these races as follows. According to paper [24]:  $E_0 = \{c_3, c_4, c_7, c_8, c_{10}, c_{11}\}$  which is parental race, and the progenies are:

$$\begin{aligned}
E_1 &= \{c_3, c_7, c_{10}, c_{11}\}, \\
E_2 &= \{c_{10}, c_{11}\}, \\
E_3 &= \{c_1, c_3, c_7, c_{10}, c_{11}\}, \\
E_4 &= \{c_3, c_{10}, c_{11}\}, \\
E_5 &= \{c_1, c_2, c_3, c_{10}, c_{11}\}, \\
E_6 &= \{c_2, c_4, c_{10}, c_{11}\}, \\
E_7 &= \{c_1, c_{10}, c_{11}\}, \\
E_8 &= \{c_7, c_{11}\}, \\
E_9 &= \{c_7, c_{10}, c_{11}\}, \\
E_{10} &= \{c_3, c_4, c_7, c_{10}, c_{11}\}, \\
E_{11} &= \{c_1, c_3, c_4, c_7, c_{10}, c_{11}\}, \\
E_{12} &= \{c_2, c_3, c_4, c_{10}, c_{11}\}.
\end{aligned}$$

Thus, our evolution algebra is generated by  $E_0, E_1, \dots, E_{12}$ . As to the defining relations, we need the detailed biological information, such as the frequency of each race in progeny population. However,  $E_0$  must be a transient generator, an intermediate transient race in the progeny population; while all other generators must be persistent generators, biologically stable races that can be observed in experiments. For illustration, we give the defining relation set as follows:

$$E_0^2 = \sum_{i=1}^{12} \frac{1}{12} E_i,$$

$$E_1^2 = \frac{1}{2} E_1 + \frac{1}{2} E_2,$$

for  $2 \leq j \leq 11$ ,

$$E_j^2 = \frac{1}{3} E_{j-1} + \frac{1}{3} E_j + \frac{1}{3} E_{j+1},$$

and for  $j = 12$ ,

$$E_{12}^2 = \frac{1}{2} E_{11} + \frac{1}{2} E_{12};$$

and if  $i \neq j$ ,

$$E_i \cdot E_j = 0.$$

This algebra is not simple. But it has a simple subalgebra generated by  $\{E_1, E_2, \dots, E_{12}\}$ . We know that this subalgebra forms a progeny population of parental race PI-52P. This subalgebra is aperiodic, which means biologically, each race in progeny population could reproduce itself in the next generation. By computing

$$\lim_n L^n(E_0),$$

we get that in the progeny population, frequency of parental race  $E_0$  is 0, frequency of race  $E_1$  and  $E_{12}$  is 5.88%, frequency of race  $E_2, E_3, \dots, E_{11}$  is 8.82%. This is the asymptotic behavior of the evolution operator.

Now let us add some intermediate transient races, biological unstable races, into the population. Suppose we have two such races,  $E_\alpha$  and  $E_\beta$ . Theoretically, there are many ways to build an evolution algebra with these

two transient generators based on the above algebra with biology information; each way will carry different biological evolution information. Here, let's choose the following way to construct our evolution algebra.

The generator set now is  $\{E_\alpha, E_\beta, E_0, E_1, \dots, E_{12}\}$ . The defining relation set is taken as

$$\begin{aligned} E_0^2 &= pE_\alpha + qE_\beta, \\ E_\alpha^2 &= \sum_{i=1}^{12} \frac{1}{12} E_i, \\ E_\beta^2 &= \sum_{i=1}^{12} \frac{1}{12} E_i, \\ E_1^2 &= \frac{1}{2} E_1 + \frac{1}{2} E_2, \end{aligned}$$

for  $2 \leq j \leq 11$ ,

$$E_j^2 = \frac{1}{3} E_{j-1} + \frac{1}{3} E_j + \frac{1}{3} E_{j+1},$$

and for  $j = 12$

$$E_{12}^2 = \frac{1}{2} E_{11} + \frac{1}{2} E_{12};$$

and if  $i \neq j$ ,

$$E_i \cdot E_j = 0.$$

Although this new algebra is not simple, it has a simple subalgebra that forms progeny population. Two unstable races, mathematically not necessarily nilpotent, will eventually disappear through producing other races. Whatever the values of  $p$  and  $q$  are, we eventually get the same frequency of each race in the population as that in the above simple algebra, except that  $E_\alpha$  and  $E_\beta$  have 0 frequency.

There is a trivial simple algebra generated by race PI-1P. If we denote PI-1P by  $E_{-1}$ , the progeny population is generated by  $E_{-1}$  which is subject to  $E_{-1}^2 = E_{-1}$ .

In paper [24], there are 5 different parental races in Minnesota and North Dakota from 1994 to 1996. If we want to study the whole structure of *P. infestans* population in Minnesota and North Dakota, we need to construct a big algebra which is reproduced by 5 parental races, PI-105P, PI-191P, PI-52P, PI-126P and PI-1P. This algebra will have 5 simple subalgebras which corresponds to the progeny subpopulations produced by 5 parental races. We also need to compute the frequency of each progeny subpopulation. This way, we encode the complexity of structure of progenies of *P. infestans* into an algebra.

Let's summarize what the evolution algebras could provide to plant pathologists theoretically.

- (1) Evolution algebra theory can predict the existence of intermediate transient races. Intermediate transient races correspond to algebraic transient elements. They are biologically unstable, and will extinct or disappear by producing other races after a short period of time. If we can catch the intermediate transient races that do not extinct but disappear through producing other new races, and remove or kill them, we will easily stop the spread of late blight disease.
- (2) Evolution algebra theory states that biologically stable races correspond to algebraic persistent elements. It predicts the periodicity of reproduction of stable races. This is helpful to understand the speed of spread of plant diseases.
- (3) Evolution algebra theory re-recovers the existence of progeny subpop-

ulation. Furthermore, because these progeny subpopulations correspond to simple subalgebras, each race in the same subpopulation shares the same dynamics of reproduction and spreading. Evolution algebras are therefore helpful to simplify the complexity of progeny population structure.

- (4) Evolution algebra theory provides a way to compute the frequency of each race in progeny population given the reproduction rates, which are algebra structural constants. Practically, these frequencies can be measured, and therefore the reproduction rates could be computed by formulae in evolution algebras. Therefore, evolution algebras will be a helpful tool to study many aspects of asexual reproduction process, like that of Oomycetes, Phytophthora.

**Acknowledgement 1.** *I would like to thank Professor Jean-Michel Morel for his strong encouragement to write the application of evolution algebras. I would also like to thank Professor C. William Birky Jr. for his explanation of uniparental inheritance by emails. I would also like to thank Professor Avner Friedman for his valuable suggestions. The author would also like to acknowledge the support from the National Science Foundation by agreement No.0112050.*

## References

- [1] Mendel, G., *Experiments in plant-hybridization*, Classic Papers in Genetics, pages 1-20, J. A. Peter editor, Prentice-Hall Inc. 1959.
- [2] Serebrowsky, A., *On the properties of the Mendelian equations*, Doklady A.N.SSSR. **2**, 33-36, 1934 (in Russian).

- [3] Glivenkov, V., *Algebra Mendelienne comptes rendus* (Doklady) de l'Acad. des Sci. de l'URSS **4**, (13), 385-386, 1936 (in Russian).
- [4] Kostitzin, V.A., *Sur les coefficients mendeliens d'heredite*, Comptes rendus de l'Acad. des Sci. **206**, 883-885, 1938 (in French).
- [5] Etherington, I.M.H., *Non-associative algebra and the symbolism of genetics*, Proc. Roy. Soc. Edinburgh B **61**, 24-42, 1941.
- [6] Gonshor, H., *Contributions to genetic algebras*, Proc. Edinburgh Math. Soc (2), 273-279, 1973.
- [7] Schafer, R.D., *An introduction to non-associative algebras*, Acad. Press, New York, 1966.
- [8] Holgate, P., *Sequences of powers in genetic algebras*, J. London Math., **42**, 489-496, 1967.
- [9] Holgate, P., *Selfing in gentic algebras*, J. Math. Biology, **6**, 197-206, 1978.
- [10] Hench, I., *Sequences in genetic algebras for overlapping generations*, Proc. Edinburgh Math. Soc. (2) **18**, 19-29, 1972.
- [11] Reiser, O., *Genetic algebras studied recursively and by means of differential operators*, Math. Scand. **10**, 25-44, 1962.
- [12] Abraham, V.M., *Linearising quadratic transformations in genetic algebras*, Thesis, Univ. of London, 1976.
- [13] Worz-Busekros, A., *Algebras in Genetics*, Lecture Notes in Biomath. 36, Springer-Verlag, Berlin, 1980.

- [14] Lyubich, Y.I., *Mathematical Structures in Population Genetics*, Springer-Verlag, New York, 1992.
- [15] Reed, M.L., *Algebraic structure of genetic inheritance*, Bull. of AMS, 34, (2), 107-130, 1997.
- [16] Baur, E., *Zeit. Vererbungsl.* **1**, 330-351, 1909.
- [17] Correns, C., *Zeit. Vererbungsl.* **1**, 291-329, 1909.
- [18] Birky, C.W.Jr., *The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models*, *Annu. Rev. Genet.* **35**, 125-148, 2001.
- [19] Birky, C.W.Jr., *Inheritance of mitochondrial mutations*, *Mitochondrial DNA Mutations and Aging, Disease and Cancer*, K.K. Singh, edit, Spring, 1998
- [20] Gillham, N.W., *Organelle Genes and Genomes*, Oxford University Press, 1994
- [21] Emmerson, C.F., Brown, G.K. and Poulton, J., *Synthesis of mitochondrial DNA in permeabilised human cultured cells*, *Nucleic Acids Res.* 29, **2**, 2001.
- [22] Ling F. and Shibata, T., *Mhr1p-dependent concatemeric mitochondrial DNA formation for generating yeast mitochondrial homoplasmic cells*, *Mol. Biol. Cell*, vol.15, 310-322, Jan. 2004.
- [23] Tang, Y., Manfredi, G., Hirano, M. and Schon, E.A., *Maintenance of human rearranged mitochondrial DNAs in long-term transmitochondrial cell lines*, *Mol. Biol. Cell*, vol.11, 2349-2358, Jul. 2000

- [24] Samen, F.M.A., Secor, G.A., and Gudmestad, N.C., *Variability in virulence among asexual progenies of Phytophthora infestans*, *Phytopathology*, **93**, 293-304, 2003.
- [25] Fry, W.E. and Goodwin, S.B., —textitRe-emergence of potato and tomato late blight in the United States, *Plant Disease*, 1349-1357, dec. 1997.
- [26] Goodwin, S.B., Cohen, B.A., and Fry, W.E., *Panglobal distribution of a single clonal lineage of the Irish potato famine fungus*, *Proc. Natl. Acad. Sci. USA*, **91**, 11591-11595, 1994.
- [27] Goodwin, S.B., *The population genetics of Phytophthora*, *Phytopathology*, **87**, 462-473, 1997.
- [28] Flor, H.H., *Current status of the gene for gene concept*, *Annu. Rev. Phytopathol.* **9**, 275-296, 1971.