5

Evolution Algebras and Non-Mendelian Genetics

In this chapter, we shall apply evolution algebra theory to non-Mendelian genetics. In the first section, we give a brief reflection of how non-Mendelian genetics motivated the development evolution algebras. In section 2, we review 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 3, we use evolution algebras to study the heteroplasmy and homoplasmy of organelle populations, and show that concepts of algebraic transiency and algebraic persistency relate to biological transitory and stability, respectively. Coexistence of triplasmy in tissues of patients with sporadic mitochondrial disorders is studied as well. In section 4, we apply evolution algebra theory to the study of asexual progenies of *Phytophthora infestans*, an important agricultural pathogen.

5.1 History of General Genetic Algebras

There is a long history of recognizing algebraic structures and properties in Mendelian genetics. Mendel first exploited some symbols [30], 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. Serebrowsky [31] was the first to give an algebraic interpretation of the sign "$\times$," which indicated sexual reproduction, and to give a mathematical formulation of Mendel's laws. Glivenkov [32] 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 [33] also introduced a "symbolic multiplication" to express Mendel's laws. The systematic study of algebras occurring in genetics was due to I. M. H. Etherington. In his series of papers [34], he succeeded in giving a precise mathematical formulation of Mendel's laws in terms of nonassociative algebras. He pointed out that the nilpotent property is essential to these genetic algebras and formulated it in his definitions of train algebras and

baric algebras. 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 [35], Schafer [36], Holgate [37, 38], Hench [39], Reiser [40], Abraham [41], Lyubich [47], and Worz-Busekos [46]. It is worth mentioning two unpublished work in the field. One is Claude Shannon's Ph.D thesis submitted in 1940 (MIT) [43]. Shannon developed an algebraic method to predict the genetic makeup in future generations of a population starting with arbitrary frequencies. Particularly, the results for genetic algebras with three loci was quite interesting. The other one is Charles Cotterman's Ph.D thesis that was also submitted in 1940 (the Ohio State University) [44] [45]. Cotterman developed a similar system as Shannon did. He also put forward a concept of derivative genes, now called "identical by descent." During the early days in this area, it appeared that the general genetic algebras or broadly defined genetic algebras (by these term we mean any algebra that has 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 nonassociative algebras, such as Lie algebras, alternative algebras, or Jordan algebras. They possess some distinguished properties that lead to many interesting mathematical results. For example, baric algebras, which have nontrivial representations over the underlying field, and train algebras, whose coefficients of rank equations are only functions of the images under these representations, are new subjects for mathematicians. Until the 1980s, the most comprehensive reference in this area was Worz-Busekos' book [46]. More recent results, such as evolution theory in genetic algebras, can be found in Lyubich's book [47]. A good survey article is Reed's paper [48].

General genetic algebras are the product of interactions between biology and mathematics. Mendelian genetics offers a new subject to mathematics: general genetic algebras. The study of these algebras reveals the algebraic structures 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 structures and the corresponding genetic properties that makes this area so fascinating. However, after Baur [49] and Correns [50] first detected that chloroplast inheritance departed from Mendel's rules, and much later, mitochondrial gene inheritance were also identified in the same way, non-Mendelian inheritance of organelle genes became manifest with two features – uniparental inheritance and vegetative segregation. Non-Mendelian genetics is now a basic language of molecular geneticists. Logically, we can ask what new subject non-Mendelian genetics offers to mathematics, and what mathematics offers to understanding of non-Mendelian genetics. It is clear that non-Mendelian genetics introduces new mathematical challenges. When we try to formulate non-Mendelian genetics as algebras, we at

least need a new idea to formulate reproduction in non-Mendelian genetics as multiplication in algebras. Actually, "evolution algebras" [24] stems from this new idea.

5.2 Non-Mendelian Genetics and Its Algebraic Formulation

5.2.1 Some terms in population genetics

Before we discuss the mathematics of genetics, we need to acquaint ourselves with the necessary language from biology. DNA is a polymer and consists of a long chain of monomers called **nucleotides**. The DNA molecule is said to be a **polynucleotide**. Each nucleotide has three parts: a sugar, a nitrogen containing ring-structure called a **base**, and a phosphate group. DNA molecules have a very distinct and characteristic three-dimensional structure known as the double helix. It is the sequence of the bases in the DNA polynucleotide that encodes the genetic information. A **gene** is a unit of information and corresponds to a discrete segment of DNA that encodes the amino acid sequence of a polypeptide. In higher organisms, the genes are present on a series of extremely long DNA molecules called **chromosomes**. For example, in humans there are estimated 50–100,000 genes arranged on 23 chromosomes. Organisms with a double set of chromosomes are called **diploid organisms**. For example, humans are diploid. Organisms with one set of chromosomes are called haploid organisms. For instant, most fungi and a few algae are **haploid organisms**. The different variants of a gene are referred to as **alleles**. Biologists refer to individuals with two identical copies of a gene as being **homozygous**; and individuals with two different copies of the same gene as being **heterozygous**. Reproduction of organisms can take place by asexual or sexual processes. **Asexual reproduction** involves the production of a new individual(s) from cells or tissues of a preexisting organism. This process is common in plants and in many microorganisms. It can involve simple binary fission in unicellular microbes or the production of specialized asexual spores. Asexual reproduction allows some genetic changes in offspring by chance. **Sexual reproduction** differs, in that it involves fusion of cells (gametes) derived from each parent, to form a zygote. The genetic processes involved in the production of gametes also allow for some genetic changes from generation to generation. Sexual reproduction is limited to species that are diploid or have a period of their life cycle in the diploid state. The division of somatic cells is called **mitosis**; and the division of meiotic cells is called **meiosis**. **Prokaryote chromosomes** consist of a single DNA, which is usually circular, with only a small amount of associated protein. **Eukaryotes** have several linear chromosomes, and the DNA is tightly associated with large amounts of protein.

5.2.2 Mendelian vs. non-Mendelian genetics

Although most of heredity of nuclear genes obeys Mendel's laws, the inheritance of organelle is not Mendelian. Before we introduce the basic of organelle biology, we need review basic knowledge of Mendelian and non-Mendelian genetics.

Following Birky's paper [51], there are five aspects in comparison of Mendelian genetics and 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.

We now review the basic of organelle biology.

Cell organelles include chloroplasts and mitochondria, which are substructural units within cells. **Chloroplasts** and **mitochondria** of eukaryotes contain their own DNA genomes. These DNA genomes vary considerably in size but are usually circular. They probably represent primitive prokaryote organisms that were incorporated into early eukaryotes and have coevolved in a **symbiotic relationship**. The organelles have their own ribosomes and synthesize some of their own proteins, but others are encoded by nuclear genes. When all of the mitochondria DNA (mtDNA) within each cell becomes genetically homogeneous, we have **homoplasmic cells**; and when mutant mtDNA molecules coexist with original mtDNA, we have **heteroplasmic cells**. Evolutionarily, chloroplasts and mitochondria have **endosymbiotic origin**. They have 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 organelles are vital to their function as are the ones they have shed into the nucleus over the millennium. Bio-energetically, 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, catalyze 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 10–200 chloroplasts. All animal cells contain many copies of mitochondrial genomes, on the order of thousands of molecules of mtDNA [52]. Therefore, it is appropriate to treat the group of chloroplasts or mitochondria in a cell as a population. This way we can take a perspective of population genetics and utilize methods in population genetics to study organelle inheritance. This is **intracellular population genetics of organelles**.

Vegetative segregation is the most general characteristics 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. More knowledge will be introduced when we construct various evolution algebras in the next section.

5.2.3 Algebraic formulation of non-Mendelian genetics

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 $g_1, g_2, \ldots, g_n$. According to the point (3) in Subsection 5.2.2,

the crossing of genotypes is impossible since it is uniparental inheritance. Mathematically, we set

$$g_i \cdot g_j = 0,$$

for $i \neq j$. According to the point (2) in Subsection 5.2.2, alleles of organelle genes may or may not segregate during meiosis following vegetative segregation, so the frequency of each gene in the next generation could be variant. According to the point (4) in Subsection 5.2.2, intramolecular and intermolecular recombination within a lineage provides evidence that one organelle genotype could produce other different genotypes. Therefore, we can mathematically define,

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

where α_{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 $g_1, g_2, \ldots, g_n$, which are subject to these relations.

Obviously, this is a very general definition. But it is general enough to include all non-Mendelian inheritance phenomena. As an example, we will look at organelle heredity in the next section.

5.3 Algebras of Organelle Population Genetics

5.3.1 Heteroplasmy and homoplasmy

Organelle population geneticists are usually concerned about a special case where 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 homoplasmic cells 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 parents can produce both heteroplasmic progeny and homoplasmic progeny, and homoplasmic parents can only produce homoplasmic progeny with the same type where mutation is not considered for the moment. Figure 5.1 shows the Wright-Fisher model for organelle genes.

Therefore, we have the following reproductive relations.

$$g_0^2 = \pi g_0 + \alpha g_1 + \beta g_2, \tag{5.1}$$

$$g_1^2 = g_1, \tag{5.2}$$

$$g_2^2 = g_2; \tag{5.3}$$

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

$$g_i \cdot g_j = 0; \tag{5.4}$$

where π, α, β are all positive real numbers. Actually, these numbers can be taken as the segregation rates of corresponding types. For any specific

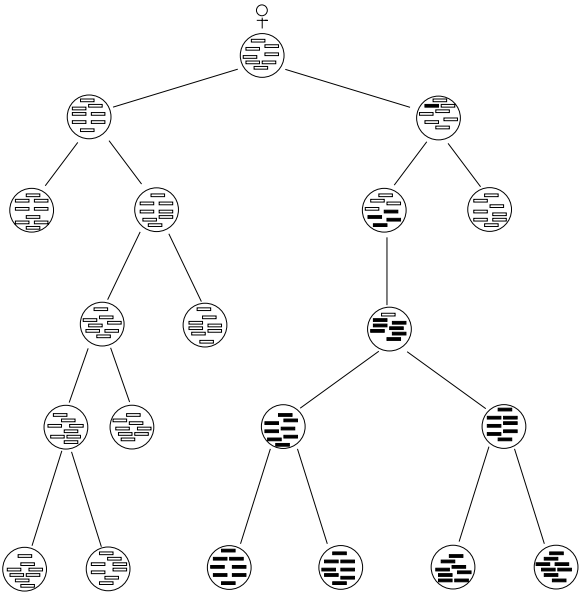

Fig. 5.1. Wright-Fisher model for organelle genes

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 (5.1)–(5.4).

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 [53]; 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 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 the 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 [54], [55]. However, by the theory of evolution algebras we could mathematically understand this phenomenon. Because g_0 is transient, g_1 and g_2 are persistent, by evolution algebra theory 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 by

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

If we now 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 is still $\{g_0, g_1, g_2\}$.

First, we think that g_2 disappears in a dramatic way, that is

$$g_2^2 = 0.$$

Other defining relations are (5.1), (5.2), and (5.4). Thus, the evolution algebra we get here is different from A_h. It has one nontrivial simple subalgebra that is corresponding to homoplasmic progeny generated by g_1.

Second, we consider that g_2 gradually mutates back to g_1, that is

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

where η is not zero and could be 1. And other defining relations are (5.1), (5.2), and (5.4). Although we eventually have one simple subalgebra by these relations, the evolution path is different.

Third, we consider that g_2 always keeps heteroplasmic property, that is

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

Other defining relations are still (5.1), (5.2), and (5.4). Eventually, we have homoplasmic progenies that are all g_1. That is the only simple subalgebra generated by g_1.

In conclusion, we have four different evolution algebras derived from the study of homoplasmy. They are not the same in skeletons. Therefore, their dynamics, which are actually genetic evolution processes, are different. However, it seems that we need to look for the biological evidences for defining these different algebras. In Ling et al. [55], several hypothetical mechanisms were put forward for the establishment of homoplasmy. These hypothetical mechanisms are actually corresponding to four different algebraic structures above.

5.3.2 Coexistence of triplasmy

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. [56], it studied the dynamical 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 (Δ-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 [56] 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% Δ-mtDNA. After a long-term culture of the heteroplasmic that contains wt-mtDNA and Δ-mtDNA from the same patient, they did not find any new cell species, although there were fluctuations of percentages of these two cell populations. From this same patient, after culturing Δ-mtDNA cell line for two years, they did not find any new cell species. Now, let us formulate this genetic dynamics as an algebra.

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

$$\begin{aligned}
g_0^2 &= \beta_{00}g_0 + \beta_{01}g_1 + \beta_{02}g_2 + \beta_{03}g_3,\\
g_1^2 &= \beta_{14}g_4 + \beta_{15}g_5,\\
g_2^2 &= \beta_{24}g_4 + \beta_{26}g_6,\\
g_3^2 &= \beta_{35}g_5 + \beta_{36}g_6,\\
g_4^2 &= \beta_{44}g_4 + \beta_{45}g_5 + \beta_{46}g_6,\\
g_5^2 &= \beta_{54}g_4 + \beta_{56}g_6,\\
g_6^2 &= \beta_{64}g_4 + \beta_{65}g_5,
\end{aligned}$$

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

$$g_i \cdot g_j = 0.$$

And the generator set is $\{g_0, g_1, \ldots, g_6\}$. This algebra has three levels of hierarchy. On the *0th* 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 *1st* level, it has three subalgebras; each of them is of dimension 1. On the *2nd*

level, there is one subalgebra generated by g_0. Generators on the *1st* and *2nd* levels are all algebraic transient. They are unobservable biologically.

If we have more information about the reproduction rates β_{ij}, we could quantitatively compute certain relevant quantities. For example, let us set

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

Then we 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 $v_0 = (1, 0, \ldots, 0)'$. As time goes to infinity, we have

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

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 drugs to stop the transitions during the transient phases of mitochondrial mutations, we could help these disorder patients.

5.4 Algebraic Structures of Asexual Progenies of *Phytophthora infestans*

In this section, we shall apply evolution algebra theory to the study of algebraic structures of asexual progenies of *Phytophthora infestans* based on experimental results in Fry and Goodwin [57]. 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 Fry and Goodwin [57] and [58].

5.4.1 Basic biology of *Phytophthora infestans*

Oomycetes are a group of organisms in a kingdom separated from the true fungi, plants, or animals. They are included in the Kingdom Protoctista 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 [58]. 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 [59]. 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 [60]. 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 [61].

In paper [57], five parental isolates of *P. infestans*, PI-105, PI-191, PI-52, PI-126, and PI-1, collected from Minnesota and North Dakota in 1994–1996, were chosen to represent different race structures. Single zoospore progenies were generated from each of the parental strains by inducing asexual 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. Thirteen races were identified from this set of isolates.

5.4.2 Algebras of progenies of *Phytophthora infestans*

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 $\{c_1, c_2, \ldots, c_{11}\}$, and if c_j functions (is expressed), the progeny resists gene R_j. Any nonrepeated combination of these c_j could form 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 the real number field R. Since reproduction of zoospore progeny is asexual, the generating relations among races are types of evolution algebras. 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 nonnegative 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 us 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 biologically stable, and we could eventually observe them as outcomes of asexual reproduction. These four types of progenies, as generators algebraically, are persistent elements. There could have been many transient generators that produce biologically unstable progenies. These unstable progenies serve as intermediate transient generations, and produces stable progenies. A simple evolution algebra without intermediate transient generations that we could construct for race PI-126P may 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 jth race in the population as in paper [57], we could easily set the 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 two generations. Eventually, frequencies of races E_1, E_2, E_3, and E_4 in the 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 us assume that 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 that 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. The defining relations we choose 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 can verify that this evolution algebra has a simple subalgebra, which is just constructed above. We also claim that intermediate transient races will extinct, and they are not biologically stable. Mathematically, these intermediate transient races are nilpotent elements.

The progeny population of PI-52P shows 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 [57]: 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, ..., 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 relations below:

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

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

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 $\{E_1, E_2, \ldots, E_{12}\}$. 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, frequencies of races E_1 and E_{12} both are 5.88%, frequencies of races E_2, E_3, $\ldots$, E_{11} all are 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_α and E_β. 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 us choose the following way to construct our evolution algebra.

The generator set is $\{E_\alpha, E_\beta, E_0, E_1, \ldots, E_{12}\}$. The set of defining relations is taken as

$$E_0^2 = pE_\alpha + qE_\beta,$$

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

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

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

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 simple algebra above, except that E_α and E_β both 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 [57], there are five 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 that is reproduced by 5 parental races, PI-105P, PI-191P, PI-52P, PI-126P, and PI-1P. This algebra will have five simple subalgebras, which corresponds to the progeny subpopulations produced by five 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 us summarize what evolution algebras can 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 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 says 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 can rerecover the existence of progeny subpopulation. 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 reproduction rates, which are algebra structural constants. Practically, these frequencies can be measured, and therefore 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.