5-2013

Modeling the Genetic Consequences of Mutualism on Communities

Carrie E. Eaton
University of Tennessee - Knoxville, ceaton2@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Carrie E. Eaton entitled "Modeling the Genetic Consequences of Mutualism on Communities." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Mathematics.

<div align="right">

Sergey Gavrilets, Major Professor

</div>

We have read this dissertation and recommend its acceptance:

Suzanne M. Lenhart, Xia Chen, James A. Fordyce

<div align="right">

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

</div>

(Original signatures are on file with official student records.)

5-2013

Modeling the Genetic Consequences of Mutualism on Communities

Carrie E. Eaton

University of Tennessee - Knoxville, ceaton2@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Carrie E. Eaton entitled "Modeling the Genetic Consequences of Mutualism on Communities." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Mathematics.

Sergey Gavrilets, Major Professor

We have read this dissertation and recommend its acceptance:

Suzanne M. Lenhart, Xia Chen, James A. Fordyce

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Modeling the Genetic Consequences of Mutualism on Communities

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Carrie E. Eaton

May 2013

Dedication

To my family

Acknowledgments

Abstract

Three models of coevolutionary dynamics between mutualistically interacting species are developed. The first is a three loci, haploid model describing a general plant-pollinator system, such as *Greya* moth and its host plant. In this case, the system will always collapse to a single plant type and pollinator type. In a community with an mutant plant type, it is possible for a host-switch to occur, governed by the initial relative abundance plant type and the pollinator choosiness. In addition, genetic diversity can be maintained if the pollinator has no differential host preference, only adaptation to a host. Next, this model is extended to the case of the fig-fig wasp system, implementing a more complex life cycle of overlapping generations due to asynchronous flowering populations. In the fig system, extensive hybridization due to asynchronous flowering can maintain genetic diversity for thousands of generations, when pollinator choosiness is high. Therefore, mutualism can lead to low confidence trees in phylogenetic reconstructions affecting discordance among plant and pollinator phylogenetic trees. Lastly, the consequences of host-switching and other speciation events on coevolving phylogenies are explored through stochastic numerical simulations. The goal is to determine to what extent cophylogeny should be expected between mutualistic partners and what features of mutualistic webs can be explained by mutualistic coevolution alone.

Contents

List of Tables

List of Figures

xviii

Chapter 1

Introduction

It is well established through observation, experiment, and theory, that ecological interactions affect evolutionary outcomes between individuals and populations, *e. g.* [16, 30, 37]. The next step is to ask how evolution effects ecology. To what extent does the ecology explain evolution and interaction patterns at the community level? This paper attempts to address these questions with in the context of mutualistic interactions. Mutualistic effects on evolution are modeled and then investigated to determine the effect on community genetic and ecological structure. Below, established observational results, experimental results, and theory are explored and relevant terminology is introduced.

1.1   Mutualism and Coevolution

Coevolution, is the genetic change over time in a set of populations resulting from the interactions between those populations. Usually the interacting populations are different species, like plant-pollinator, predator-prey, or host-parasite [93]. There are different types of ecological interactions which can cause such genetic change, and are usually defined by their effect on **fitness** - the success of a population in propagating its genetic material [30]. **Mutualism** is a type of ecological interaction that positively affects the fitness of both interacting populations. Plant-pollinator interactions are

1

Chapter 1

Introduction

It is well established through observation, experiment, and theory, that ecological interactions affect evolutionary outcomes between individuals and populations, *e. g.* [16, 30, 37]. The next step is to ask how evolution effects ecology. To what extent does the ecology explain evolution and interaction patterns at the community level? This paper attempts to address these questions with in the context of mutualistic interactions. Mutualistic effects on evolution are modeled and then investigated to determine the effect on community genetic and ecological structure. Below, established observational results, experimental results, and theory are explored and relevant terminology is introduced.

1.1   Mutualism and Coevolution

Coevolution, is the genetic change over time in a set of populations resulting from the interactions between those populations. Usually the interacting populations are different species, like plant-pollinator, predator-prey, or host-parasite [93]. There are different types of ecological interactions which can cause such genetic change, and are usually defined by their effect on **fitness** - the success of a population in propagating its genetic material [30]. **Mutualism** is a type of ecological interaction that positively affects the fitness of both interacting populations. Plant-pollinator interactions are

1

usually considered mutualisms because the pollinator often receives a food source from the nectar, and the plant is able to reproduce due to pollen transfer that occurs during feeding. **Antagonism** is a case in which one population has a positive fitness change and the other has a negative fitness change. Parasite-host is one example of a highly studied type of antagonistic interaction [24].

An example of a mutualism that directly impacts reproductive success is the plant and pollinator relationship. By definition, a pollinator aids the plant in its reproductive success. In a mutualism, pollinators gain resources by either by gathering nectar or by exploiting a potential plant on which to oviposit. These systems are used as models because of the direct impact plants and pollinators have on each other's reproductive success [93]. **Diffuse mutualism** is sometimes used to describe the most common case of mutualism in which the relationship weakly or indirectly impacts reproductive success. In this case, the fitness of one population is weakly dependent on the mutualistic interaction because either the interaction has little effect or there are many other processes during the life cycle that contribute to the organism's fitness. In order to maximize the efficiency of a mutualistic interaction, traits on which this interaction depend evolve to match in each population [93]. Although this terminology seems to imply that there is an active component, the coevolutionary results of ecological interactions are derived via mechanisms such as mutation, genetic drift, natural selection, and recombination [33].

1.2 Mechanisms of Evolution

1.2.1 Mutation

A **gene** is a unit of genetic information, and the site of a gene is the **locus**. **Alleles** are the different biochemical forms of a gene, and **mutations** can alter which alleles are present at a locus [42]. Although often a rare event, typically having probability of occurring at a particular site on the order of 10^{-6}, mutation is a source for novel

genetic information. It allows genes flexibility to cope with and adapt to changing environments over time [42].

1.2.2   Genetic drift

Populations are subject to sampling bias, and as a result, the frequency of each allele type in that next generation may not equal the current ratio. This is called **genetic drift** and impacts small populations [42]. This effect is included in individual-based simulations that explicitly account for each individual's genetics. This approach is advantageous because it captures sampling bias and is often more biologically realistic, but it is computationally complex and irreducible to simpler forms, and it requires extensive analysis. One way to simplify this assumption is to include a stochastic component to describe how the genetics change because the sampling effect is random [84]. Several types of models assume infinite population size, but drift effects must be acknowledged when comparing model results to the biological systems.

1.2.3   Natural selection

A set of genes is called the **genotype**, and the expression of that genotype in the organism is called the **phenotype**. **Natural selection** acts on the phenotype (in turn acting on the genotype) to decrease the organisms with phenotypes in the next generation that are less fit [42]. For example, if a population of blue and red insects were found by a predator on a red surface, then we would expect that more blue are eaten than red, thus a higher proportion of red ants would result from selection. The proportion of blue insects that survive is the **fitness** of the blue phenotype, denoted by w. The proportion eaten or selected against is denoted by the parameter s. As a result of the change in proportions of the parental population, we then expect that more red insect will hatch and less blue insects will hatch in the next generation.

1.2.4 Recombination and segregation

In a **haploid** organism, each gene site has one allele and it determines the genotype of the organism. In a **diploid** system, each genotype has two alleles that together determine the genotype of the organism [42]. Consider the haploid case with two loci close to each other on the same chromosome. The chance that during chromosome sex these two alleles will be separated during recombination and only one inherited is called the **recombination rate**, r. Suppose that a haploid organism with allele **A** at locus 1 and allele **B** at locus 2 (so it has a genotype of **AB**) mates with a haploid organism with genotype **ab**. The offspring, also a haploid organism, will have a 1/2 chance of having genotype **AB** and a 1/2 chance of having genotype **ab** if recombination does not occur. Since the inheritance probability is independent of the recombination, the probability the offspring will have genotype **AB** given parents of genotypes **AB** and **ab** is $(1 - r)/2$. Likewise, if recombination does occur, then the offspring will have a 1/2 chance of having genotype **Ab** and a 1/2 chance of having genotype **aB**. Therefore, the probability an offspring has genotype **Ab** given parents with genotypes **AB** and **ab** is $r/2$.

If the loci are on separate chromosomes and therefore independently distributed (or independently **segregated**), then $r = 1/2$. Usually recombination rates of less than 1/2 indicate that the loci are physically close together on the strand of DNA. Low recombination rate may help certain genotypes to be maintained in the genetic pool longer because they are more often inherited together. This may slow evolution seeking to separate badly matched alleles (for example, if an organism had an allele at one locus that made it a predator, but possessed at a nearby locus an allele that caused an allergy to eating the prey). On the other hand, low recombination can make evolution proceed faster once it is on the right track, because good combinations of alleles will also be inherited [30].

1.2.5 Non-random mating

There are other mechanisms of evolution that can promote the formation of good combinations of alleles at loci. **Non-random mating** occurs when certain mating pairs form more often than others. One case of this is called **assortative mating** in which like individuals are more likely to mate. This can be an active choice or can result from a preference for particular mating conditions [30]. For example, if an insect has a genotype that causes it to prefer a particular plant on which to seek mates, it is highly likely that the mates it encounters will also have the same genotype at that preference loci. It is highly likely then, that their offspring will have the same plant preference. Now suppose that another locus determines the insects color and that during mating the insect's color does not match the plant's color, it is vulnerable to selection by predators. Thus, combinations of alleles at the two loci that both prefer and are color adapted to the same plant would be favored in this case. Then the system would evolve to eliminate genotypes that prefer one plant and are adapted to different plants [30].

1.3 Consequences of Evolution

Biodiversity can occur on a variety of levels, including at the genetic level. A **polymorphic locus** is a locus for which there are many allele forms available in the genetic pool, and it is a type of within-population genetic diversity. Different populations can have different allele forms available at their loci. This is an example of between-population genetic diversity [42]. As these and other genetic differences accumulate, these populations can divide into separate species, a process known as **speciation**. The exact point of transition to species is often a point of controversy. The end result is often that a certain amount of reproductive isolation between the species is present. In other words, the populations have accrued enough differences that they no longer, or rarely, mate with each other [30].

1.3.1 Speciation

The origin of species is often classified into three types based on the physical location of the populations under consideration, sympatric, parapatric, and allopatric. **Sympatric speciation** is when two sub-populations that diverge and speciate from a single population in one habitat together. **Parapatric speciation** is when speciation occurs where there are different habitats next to each other which are not completely mixed. The population diverges, and each adapt to the different habitats. **Allopatric speciation** occurs when the populations are separated completely from each other, and then each evolves independently along different trajectories [30]. These three cases can also be considered sub-cases of parapatric speciation with an index given for the frequency of population mixing, allopatric populations and sympatric populations being opposite ends of the extreme [30].

1.3.2 Coevolution

One-to-one coevolution is defined here as two interacting ecological partners evolving together. One-to-one coevolution characterizes coevolution the majority of the time between coevolutionary events. There are 4 major types of coevolutionary events: cospeciation, sorting, duplication, and host-switching, see Figure 1.1 [77].

In **cospeciation** or **lineage-tracking**, speciation in one class leads to speciation in the other class [77]. If this event alone was the only result of coevolution, then each of the resulting species stays in a one-to-one correspondence with its coevolutionary partner. This produces matching phylogenetic trees, *i.e.* display **cophylogeny**. An exception to this one-to-one matching rule would be during periods of transition where speciation has occurred in one partner and the other lags behind [22]. In **sorting**, a speciation event in one evolutionary partner occurs, but the other partner does not cospeciate or retain ecological connection to the new species [77]. Similar to sorting is **duplication**, in which a speciation event in one ecological partner occurs, but all ecological connections are maintained between the new species and the ancestral

partners [77]. **Host-switching** is where one ecological partner trades (**complete**) or expands (**incomplete**) its existing relationship for another partner [93]. Host-switching interrupts the formation of cophylogeny.
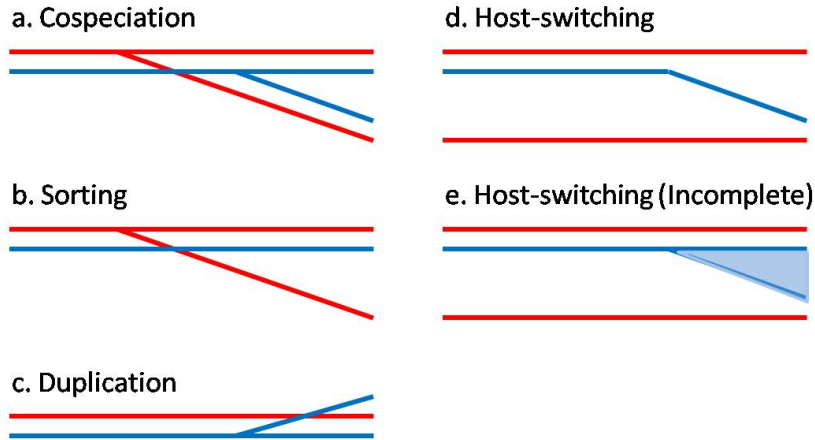
Figure 1.1: Schematic of the 4 major results of coevolution on phylogenies: cospeciation, sorting, duplication, and host-switching (complete and incomplete). The shaded area in the figure of incomplete host-switching indicates that the speciation event itself in the blue phylogentic tree is incomplete, so there is still exchange of genetic information between the species.

Cophylogeny is phylogenetic trees of coevolving groups taken together with their associations. In the mutualistic network, plants are connected to pollinators only and not to other plants directly (and vice versa) [77]. This type of network is represented by a **bipartite graph** in which only interactions between plants and pollinators are considered. A **graph** is a set of nodes (species) and edges (ecological associations) [7]. A graph **connected** if there is a path that connects every pair of nodes. Any graph that is not connected can be partitioned into **connected components** or **disjoint connected subgraphs** [4]. This will be important later as disjoint connected subgraphs of a network will act independently of each other over time.

1.4 Plant-Pollinator Coevolutionary Theory

Ecological relationships affect the evolution of communities. In particular, species diversity and the resulting interactions determine stability of complex community networks and also may be an important driving force behind speciation and thus biodiversity [19, 25, 46]. Interactions between two sets of species, *e.g.* host-parasite, plant-herbivore, and plant-pollinator, are well studied biologically [11, 13, 26, 45, 77, 97]. The majority of theoretical and experimental work in evolution as a result of species interactions has focused on antagonistic interactions, like predator-prey or host-parasite. This work demonstrated that antagonism promotes genetic diversity [16, 51, 59], but mutualism is another important way in which species can interact and affect coevolution.

Mutualisms between plant-pollinator, animal-plant, or host-symbiont are widely studied for their ecological value. With the exception of some model systems, little is known about how these mutualisms shape evolutionary trajectories and ecological networks [47]. Several model biological systems involving obligate mutualisms, like that of the fig-fig wasp and yucca-yucca moth, are studied to understand the results of mutualistic interactions [3, 14, 15, 44, 78]. The goal of this dissertation is to develop simple yet biologically realistic models of evolution as a result of obligate mutualisms. These models will help form intuition about how mutualism affects evolutionary trajectories.

An obligate mutualism is a specific type of mutualism in which each population needs to participate in the mutualistic interaction in order for the population to survive. The interaction is beneficial and necessary to both parties. Obligate mutualisms are considered rare, yet are the subject of intense investigation. There are several reasons for this. First, the populations often develop very specialized organs or traits in order to efficiently continue the mutualistic interaction; the populations coevolve. Second, because the interaction is necessary, there is very strong selection on these traits. Lastly, some investigations have focused on why obligate mutualisms

should persist when reproductive success depends so crucially on another species. The plant-pollinator cases considered here are obligate mutualisms in which the insect needs the plant as a larval food source and in return pollinates the plant [93]. Since the benefit provided by the insect and the plant are exclusive and therefore is necessary, it is an obligate mutualism. The fig-fig wasp system is a particularly interesting model system because of its high species diversity, boasting over 750 fig tree species and an estimated 1300 fig wasp species [71].

Theoretical studies of plant-pollinator obligate mutualisms showed how co-evolution could result in the allopatric or sympatric speciation in both species [22, 36, 58, 59]. The quantitative trait loci model in Kiester, *et al.*, showed that random genetic drift can move phenotypes in different directions in different populations. This sets the stage for allopatric speciation. This model assumes that each mutualistic partner has only one continuous, normally distributed trait with a fixed variance upon which the mutualistic interaction is based. An example is style length in plants and ovipositor length in insects. Fixed variance can limit the ability of bimodal distributions to form, as seen in fig plants. Kopp and Gavrilets [59] relax this fixed variance assumption by considering a multilocus model for a quantitative trait. This gives the ability of genetic variances to evolve independently of mean trait values. This allows for more equilibria due to the multilocus structure. As a result, coevolution between mutualists results in mean trait matching and can evolve towards a stable polymorphic equilibria at one locus. This is because fitness depends on success of matching and on some physiological optimum. The overall results for the mutualistic and antagonistic cases mirror a similar single-locus model of mimicry presented in Gavrilets and Hastings [32]. Cases of coevolution between a species and its mimic were modeled using a single-locus model. "Mutualism" is considered as Müllerian mimicry, and results in solely monomorphic stable equilibria. Notably, all polymorphic equilibria are unstable. In the adaptive dynamics model of Doebeli, *et al.*, sympatric speciation in one species is followed by secondary or cascade speciation in the other mutualistic partner species. The major drawback to adaptive dynamics

models is that there is no explicit underlying genetic basis upon which the fitness landscape is defined. This makes relating data to models difficult. In the gene-for-gene model in Gomulkiewicz, *et al.* [36], co-existence of multiple genotypes is mediated through fluctuating environments. In this system, the interaction itself fluxuates from being a positive interaction to a negative interaction which encompasses both mutualistic and antagonistic relationships. The two mutualism only models predict co-speciation, which over the long term would predict a matching cophylogeny.

Field studies confirm the generality of this one-to-one matching relationship [43]. However, the phylogenetic signal does not always predict congruent co-phylogenies between plant and pollinator [21, 44, 96]. There are also exceptions to the one-to-one rule in which two pollinator species can be found on the same plant [65, 71]. Theoretical work indicates that this is possible in periods of brief transience [22] after speciation in one ecological partner occurs, but before cospeciation in the other partner occurs. However, the roles of host-switching and hybridization must also be considered [98, 100]. Whether or not a simple mutualistic system can maintain diversity long term is explored. The Levene two-resource model with two fixed available resources predicts that genetic variation can be maintained, especially with niche based assortative mating [61]. To what extent will this result be modified when the resource is reciprocally evolving? Other aspects of the biology and ecology of model mutualistic systems such as asynchronous flowering times and availability of alternate host types, may also strongly influence coevolving phylogenies [93, 98].

In order to further understand the evolutionary outcomes of mutualistic relationships in general, an obligatory mutualistic relationship between a moth and plant species with a simple life cycle, Greya moth *Greya politella* and its host plants, *Lithophragma parviflorum* and *Huechera* (H-GM) is modeled. The plant has one diallelic locus describing state, such as chemical profile, and includes a parameter for overlapping generations in plants. The pollinator is described by two diallelic loci, one for preference for a plant and the other for larval adaptation for the plant. Included are parameters for selection, preference bias, and recombination between loci. Then the

H-GM model is expanded to the well-studied fig-fig wasp system (F-FW). The fig and fig wasp system is more complex in its dynamics due to an asynchronously flowering plant population. Therefore we consider more than one flowering plant population. In the case of H-GM, host switching in the system has been well recognized [93]. Yet in the case of F-FW only now is this system receiving recognition for the prevalence of host-switching and non-one-to-one plant-pollinator correspondences [98].

The H-GM and F-FW models give further insight to the types of coevolutionary events that arise as a consequence of mutualism. The next step is to examine how those coevolutionary events shape coevolving phylogenies. The nature of the coevolutionary relationship between figs and fig wasps is not as clear as overlaying congruent phylogenies. Not only are there violations to the one-to-one rule between phylogenies, but hybridization confuses the resolution of the phylogenies as well as rampant host-switching [49, 100]. Verbal models up to this point are careful to not assume perfect matching of traits in coevolutionary systems and acknowledge there may be other events happening on a larger geographic scale that influence local examinations [49, 50, 93, 100]. Still the expectation is that cophylogeny is the rule and that nearly related members of insects should be pollinating nearly related members of plants. Phylogenetic coevolution reconstructs phylogenetic relationships between coevolving partners and in doing so places penalities on events like host-switching when resolving these phylogenies [77]. Therefore, it is important to know how often host-switching happens and the effects of a higher or lower host-switch probability on coevolving phylongenies. Only then can we determine if patterns other than perfect cophylogeny are the exception or part of the rule.

A 2007 paper by Rezende *et al.* [83] suggested that evolutionary history should be included into models of network formation and maintenance. In 2009, Ings *et al.* made the same suggestion in their comprehensive review paper, that ecological networks needed to take evolutionary history under consideration [47]. Doing so provides a better understanding of the mechanisms behind patterns in networks, such as low connectedness. Also, an event like a host-switching occurs often between coevolving

phylogenetic trees can highly influence the extent to which overall cophylogeny should be the expectation. A stochastic model is developed here as to understand how phylogenies coevolve based on intuition gained from the H-GM and F-FW models.

Chapter 2

Modeling a simple mutualistic system: *Greya* moth and its host

2.1 *Greya* Moth and its Host (H-GM) Model

Consider the life cycle of the *Greya* moth and its host plant beginning at the larval stage. The larvae drop to the ground and overwinter in the soil, the adults emerging in the spring. Adult moths find mates on the host plant and copulate, then females deposit their eggs in flowers [36]. There is a large number of plants on which mating and oviposition, mathematically we assume infinitely many, and plants are pollinated by ovipositing females.

Assume that the mutualism is controlled by two major loci in the pollinating insect: insect adaptation to a plant type (locus A) and insect preference bias for a plant type (locus B), with recombination probability r. Assume also that the plant is haploid and that one locus controls the type of plant under investigation. For example, locus C could be interpreted as controlling chemical profile, and that chemical profile both attracts a certain type of insect and provides a chemical environment for the insect. The insect may or may not be optimally fit on this environment. See Table

for a full list of loci, variables and parameters under consideration in the H-GM model.

Table 2.1: List of H-GM model loci, variables, and parameters.

Loci under consideration		
Locus A	Adaptation in insect for plant	
Locus B	Preference in insect for plant	
Locus C	Chemical profile of plant	
Variables		**Range**
x_i	Frequency of type i insect	$[0, 1]$
y_m	Frequency of type m plant	$[0, 1]$
Variables		**Range**
s	Selection coefficient	$[0, 1]$
ϵ	Preference bias	$[0, 1]$
r	Recombination probability	$[0, 1/2]$
β	1/(Avg lifespan of plant)	$[0, 1]$

2.1.1   Dynamics of insects

To understand how insect genotype frequencies change over time, consider the change that occurs in each generation as a difference equation. Let x_i be the frequencies of insect genotypes in the pool of adults emerging from soil and let y_m be the frequencies of the plant types at that time. π_{im} is the preference of an insect with genotype i for a plant of type m. We will use indices i, j, k for insects and l, m, n for plants.

The frequency of adult insects of type i found on plants of type m is

$$x_{i,m} = \frac{\pi_{im} x_i}{\sum_i \pi_{im} x_i} \equiv P(i, m) x_i. \tag{2.1}$$

14

where

$$P(i, m) = \frac{\pi_{im}}{\sum_i \pi_{im} x_i}. \tag{2.2}$$

can be interpreted as the preference of insect i for plant m relative to the average preference of insects for this plant.

The proportion of the adult insects found on plants of type m is

$$c_m = \frac{y_m \sum_i \pi_{im} x_i}{\sum_m y_m \sum_i \pi_{im} x_i}, \tag{2.3}$$

the frequency of insects joining a particular mating pool m, normalized and weighted by plant frequency.

Assuming random mating on host, the frequency of mating pairs formed by females i and males j on a plant of type m is

$$M_{ij,m} = x_{i,m} x_{j,m}. \tag{2.4}$$

The frequency of eggs with genotype k produced by pairs (i, j) mating on a plant of type m is $M_{ij,m} R(i, j \to k)$, where R gives the corresponding offspring frequencies for a given set of parental genotypes.

We assume the contribution of each mating pool (*i.e.* plant type) to the overall offspring pool is equal to the proportion of insects that came to the pool. Thus, the proportion of eggs with genotype k and carried by mothers with genotype i in the whole set of eggs before oviposition is

$$E_{k,i} = \sum_m c_m \sum_j M_{ij,m} R(i, j \to k). \tag{2.5}$$

The frequency of eggs with genotype k oviposited on a plant of type n (by all females i) is

$$e_{k,n} = \frac{\sum_i E_{k,i} \pi_{in}}{\sum_k \sum_i E_{k,i} \pi_{in}}. \tag{2.6}$$

After selection on a plant of type n, the frequencies of larvae dropping to the soil from that plant are

$$e'_{k,n} = \frac{e_{k,n}W_{k,n}}{\overline{W}_n}, \tag{2.7}$$

where $W_{k,n}$ is fitness, *i.e.* viability, of genotype k on a plant of type n and $\overline{W}_n = \sum_k e_{k,n}W_{k,n}$ is the average fitness of the population on a plant of type n. We assume that the contribution of each larva pool (*i.e.* plant) to larvae is equal to the proportion of the eggs deposited on the plant. Note that variations on this assumption will be explored in a later model. The frequencies of the insect genotypes in the pool of adults emerging from soil in the next generation are

$$x'_k = \sum_n c_n e'_{k,n}. \tag{2.8}$$

The difference equation for insect genotypes in the next generation is then

$$x'_k = \sum_i \sum_j \sum_m \sum_n c_m c_n P(i,m) P(j,m) P(i,n) x_i x_j R(i,j \rightarrow k) \frac{W_{k,n}}{\overline{W}_n}. \tag{2.9}$$

2.1.2 Dynamics of plants

Now consider the genotype frequency change in plants. The probability that a female of type i goes to a plant of type m for mating and then to a plant of type n to lay the eggs is

$$\frac{\pi_{im}y_m}{\sum_m \pi_{im}y_m} \frac{\pi_{in}y_n}{\sum_n \pi_{in}y_n} = Q(i,m)y_m Q(i,n)y_n, \tag{2.10}$$

where

$$Q(i,m) = \frac{\pi_{im}}{\sum_m \pi_{im}y_m}. \tag{2.11}$$

By doing this the female pollinates a plant n by pollen from plant m. The term $Q(i,m)$ can be interpreted as the probability that an insect i visits a particular plant of type m. Thus, the frequency of plant l produced as a result of mating of plants m

and n is

$$y_{l,o} = \sum_i x_i Q(i,m) y_m Q(i,n) y_n S(m,n \to l) \equiv F_{mn} y_m y_n S(m,n \to l), \qquad (2.12)$$

where $S(m,n \to l)$ is the corresponding segregation probability (recall there is no recombination since we only consider one plant loci), and the term

$$F_{mn} = \sum_i x_i Q(i,m) Q(i,n) \qquad (2.13)$$

can be interpreted as fertility of mating pair m and n.

Finally, to account for the fact that the host plants are perennial we assume that only a proportion β of plants is replaced each generation by the offspring. Therefore β may also be interpreted as 1 over the average number of years in a plant lifespan. Then

$$y_l' = (1-\beta)y_l + \beta \sum_i \sum_m \sum_n x_i Q(i,m) Q(i,n) y_m y_n S(m,n \to l) \qquad (2.14)$$

$1-\beta$ is the proportion of plants that are perennial from the last generation of plants. This may also be interpreted as the proportion that were randomly wind pollinated (and thus have no change in frequency).

2.1.3 Plant resource dependent model

In creating our initial model, it is assumed the frequency of adult insect types in the next generation is proportional to the larva frequency types surviving on each plant and the frequency of egg types laid on the each plant. This assumes there is no plant resource limit or larval competition on plants. In a variation of that model, we assume the contributions of larva types remaining after viability selection are weighted in the next generation by the frequencies of the plant they are on. This changes Equation

17

to

$$x'_k = \sum_n y_n e'_{k,n}. \tag{2.15}$$

Then the equation for the next generation of adult insect is

$$x'_k = \sum_i \sum_j \sum_m \sum_n c_m y_n P(i,m) P(j,m) P(i,n) x_i x_j R(i,j \to k) \frac{W_{k,n}}{\overline{W}_n}, \tag{2.16}$$

which can be compared to Equation 2.9 above:

$$x'_k = \sum_i \sum_j \sum_m \sum_n c_m c_n P(i,m) P(j,m) P(i,n) x_i x_j R(i,j \to k) \frac{W_{k,n}}{\overline{W}_n}.$$

2.1.4 Model without re-assortment

To further simplify the above models, one of the multiple moth assortments to its host plant is eliminated. If multiple re-assortment merely shuffles insects but maintains the same frequencies on each plant, then mathematically we could delete one of the trips the female insect makes. In this model variation, instead of re-assorting to lay eggs, she lays eggs on the plant she mates on. In doing so, analytical tractability is improved without changing the qualitative behavior. The modification is made to the H-GM plant resource dependent model. This eliminates the c_m term, because females no longer have to re-assort to lay eggs:

$$x'_k = \sum_i \sum_j \sum_m y_m P(i,m) P(j,m) x_i x_j R(i,j \to k) \frac{W_{k,n}}{\overline{W}_n}. \tag{2.17}$$

2.2 H-GM Model Analytic and Numerical Results

2.2.1 Assumptions

For the following numerical and analytic analysis, we consider the case where each locus is diallelic. Let x_i $(i = 1,2,3,4)$ be the frequencies of four insect genotypes, **AB**, **Ab**, **aB** and **ab**, respectively, in the pool of adults emerging from soil. Let

y_m ($m = 1, 2$) be the frequencies of the two types of plants (**C** and **c**, respectively) during the insect mating period. A list of insect and plant genotypes matched up with indicies is in Table 2.2.

Table 2.2: A list of insect and plant genotypes in the H-GM model listed by index.

Insect Genotypes	
Index	Genotype
1	AB
2	Ab
3	aB
4	ab
Plant Genotypes	
Index	Genotype
1	C
2	c

Recall π_{im} is the relative preference of an insect with genotype i for a plant of type m. We consider the case where an insect having allele **A** encountering plant of type **C** will chose that plant with probability $\pi_{1,1} = \pi_{2,1} = (1 + \epsilon)/2$, but will chose a plant of type **c** with probability $\pi_{1,2} = \pi_{2,2} = (1 - \epsilon)/2$ (vice versa for an insect carrying the **c** allele). ϵ can be interpreted as the bias of an insect towards a particular plant choice and $0 \leq \epsilon \leq 1$. When $\epsilon = 0$, the insect has no preference and chooses whomever it encounters first. When $\epsilon = 1$, the insect will choose the matching profile every time and will never make a "mistake." Likewise, insect larvae born on a matching plant type will have higher fitness.

At the local adaptation loci an insect may have either allele **A** or **a**, which are best adapted to plant types **C** and **c**, respectively. We consider the case where a larvae laid on a matching plant to which it is adapted will have fitness $W = 1$, and a larva

developing on a non-matching plant will have fitness $W = 1 - s$. s is referred to as the selection coefficient and $0 \leq s \leq 1$.

2.2.2 Analytical Methods

The model introduced above is a non-linear discrete dynamical system. Time is described by generations, and each generation's genotype frequencies are calculated from the previous generation's. To predict the outcome of the system after many generations of evolution, equilibria and their stability are examined. Stable equilibria are of particular interest, because the system will evolve toward a stable equilibrium in the long term. Full analytical investigation of each model is presented in Appendix A.

Classic 1-1 co-evolution with one's evolutionary partner, host-switching and speciation or the maintenance of genetic variation are the primary points of interest. For classic co-evolution and host-switching to occur, fixation of each set of matching alleles (**A**, **B**, and **C** or **a**, **b**, and **c**) must be a stable equilibria for the system. Whether cospeciation or host-switching is taking place is inferred from initial conditions. Thus both numerical simulations and analytic work are performed to determine where the basin of attraction is for each fixed equilibrium when they are both stable. This means that one can determine how the system behaves long term based on the initial conditions. For the case of speciation, if the polymorphic state is stable, it is possible for both types of plants to co-exist in the system.

2.2.3 H-GM model results

Twenty equilibria emerge from the H-GM model, sixteen of which are listed in Table 2.3 and four additional biologically unrealistic equilibria, which are not listed. The full analysis proving the stability of the equilibria is presented in Appendix A.1.

The system is bistable, as illustrated through the analytic and numerical simulations. This means there are two steady states to which the system could evolve

Table 2.3: All equilibria are either always stable or always unstable for all biologically realistic parameters unless otherwise noted. **Note, numerical simulations indicate that this point seems to be unstable, however, this result is not shown analytically. z_3^* and $\bar{z}_3^*$, are $+/-$ solutions to a quadratic equation and are functions of r, s, and ϵ; the exact form is shown in Appendix A.1.

$y_1 = 1$	**C fixed**	
$x = [1, 0, 0, 0]$	AB fixed	stable
$x = [0, 1, 0, 0]$	Ab fixed	unstable
$x = [0, 0, 1, 0]$	aB fixed	unstable
$x = [0, 0, 0, 1]$	ab fixed	unstable
$x = \left[\frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, \quad \frac{-(1-\epsilon)(3-\epsilon)}{8\epsilon}, \quad 0, \quad 0 \right]$	biologically unrealistic	
$x = \left[0, \quad 0, \quad \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, \quad \frac{-(1-\epsilon)(3-\epsilon)}{8\epsilon} \right]$	biologically unrealistic	
$y_1 = 0$	**c fixed**	
$x = [1, 0, 0, 0]$	AB fixed	unstable
$x = [0, 1, 0, 0]$	Ab fixed	unstable
$x = [0, 0, 1, 0]$	aB fixed	unstable
$x = [0, 0, 0, 1]$	ab fixed	stable
$x = \left[\frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, \quad -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon}, \quad 0, \quad 0 \right]$	biologically unrealistic	
$x = \left[0, \quad 0, \quad \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, \quad \frac{-(1-\epsilon)(3-\epsilon)}{8\epsilon} \right]$	biologically unrealistic	
$y_1 = \frac{x_1 - x_2 + x_3 - x_4 + \epsilon}{2\epsilon}$	**Polymorphic**	
$x = [\frac{1}{2}, \frac{1}{2}, 0, 0]$	fr(C) = 1/2	unstable
$x = [0, 0, \frac{1}{2}, \frac{1}{2}]$	fr(C) = 1/2	unstable
$x = \left[\frac{1}{4} + \frac{1}{4}z_3^*, \quad \frac{1}{4} - \frac{1}{4}z_3^*, \quad \frac{1}{4} - \frac{1}{4}z_3^*, \quad \frac{1}{4} + \frac{1}{4}z_3^* \right]$	fr(A) = fr(B) = fr(C) = 1/2	unstable**
$x = \left[\frac{1}{4} + \frac{1}{4}\bar{z}_3^*, \quad \frac{1}{4} - \frac{1}{4}\bar{z}_3^*, \quad \frac{1}{4} - \frac{1}{4}\bar{z}_3^*, \quad \frac{1}{4} + \frac{1}{4}\bar{z}_3^* \right]$	biologically unrealistic	

depending on the initial conditions. More specifically, the system will always go to fixation for either the **A**, **B**, **C** alleles or the **a**, **b**, **c** alleles. This is illustrated in Figure 2.1 in which the trajectories of 100 numerical runs are plotted for the same set of parameter values, but with random starting initial conditions.
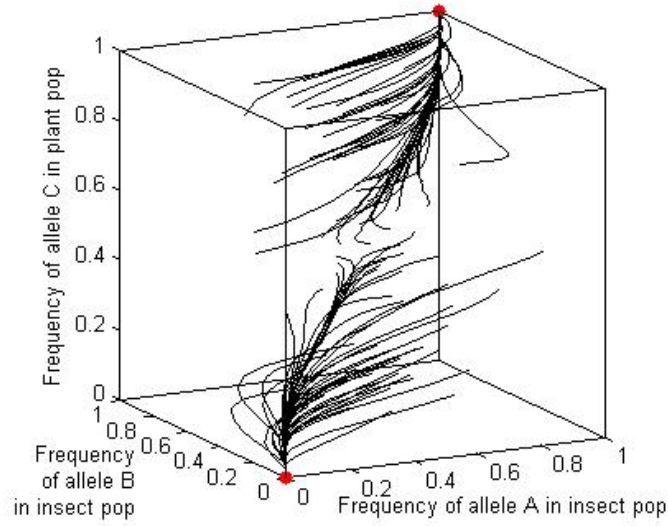
Figure 2.1: Phase portrait showing 100 trajectories of the H-GM model for random initial conditions. Parameter values are set to $\epsilon = 0.5$, $r = 0.2$, $s = 0.6$, and $\beta = 0.2$. Red dots indicate the value of the allele frequencies after 500 generations.

To see more closely what effect the initial conditions have on the long-term dynamics, trajectory plots of allele frequencies over time are examined. Figure 2.2 shows the case in which the insect population, despite being nearly fixed for the **A** and **B** alleles in the population, evolves to a population fixed for **a** and **b** alleles. This results from the large initial proportion of **c** plants in the system. Therefore, this system has experienced a host-switch, where the insects evolve to adapt to the more plentiful resource type. Because **c** plants are the most available, the insects that prefer the **c** plants (**Ab** and **ab**) are going to have larger frequencies in the next

22

generation due to mating in higher frequencies. **Ab** will have a selective disadvantage on **c** plants, so the insect genotype that eventually prevails is **ab**.
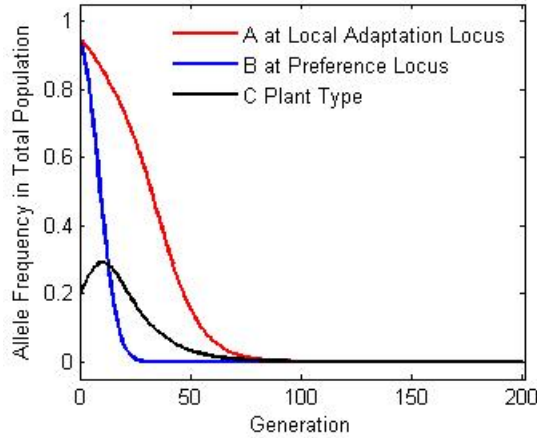
Figure 2.2: Parameter values $\epsilon = 0.2$, $r = 0.1$,$s = 0.1$, and $\beta = 0.2$. This simulation shows how the frequency of an alternate plant type drives host-switching.

In Figures 2.3a-d, the parameter sets are the same, but have different initial conditions. In Figures 2.3a and 2.3b, the plant and insect allele types that are most predominant are those which survive long term. This means a small amount of introduced genetic variation into either the plant or insect population will not interrupt the co-evolutionary congruence. However, a slightly higher proportion of **C** plants in the system will be enough to allow the insects to fix for the **A** and **B** allele and will instead induce the plants to follow that trajectory. This is shown in Figures 2.3c and 2.3d, where the plant population starts with a frequency only 0.1 less than its upper counterpart, but instead evolves to become fixed with only the **C** type of plant. In all cases, linkage disequilibrium goes to zero.

2.2.4   Effect of parameters on H-GM model

Initial conditions influence the final outcome of the bistable system, however varying strength of parameters will change the speed to which equilibria are attained and may also influence the basin of attraction for the stable equilibria. A discussion of

Figure 2.3: H-GM model simulations showing systems robust to some introduced variation, but that can also result in host-switching with a critical relative abundance of alternate host type. Parameter values for each of these graphs are the same set to $\epsilon = 0.5$, $r = 0.2$, $s = 0.6$, and $\beta = 0.2$. The blue and red lines are nearly overlapping in these cases. The graphs in each column (a,c and b,d) have the same initial insect frequencies, but the plant frequencies in the top and bottom graphs differ by only 0.1. This shows that the frequency of an alternate plant type heavily influences the long-term relationship.

the effects of parameters as well as special cases when parameters at the endpoints of their ranges is below.

Parameter ϵ. Recall that that ϵ can be interpreted as the bias of an insect towards a particular plant choice. Therefore, $\epsilon = 0$ implies that the insect has absolutely no preference for a particular plant and $\epsilon = 1$ implies that the insect has a strict preference for the matching plant type. In this case varying ϵ along points in the interior of its range $(0, 1)$ can change the outcome of the system as seen in Figure 2.4. This Figure illustrates how varying ϵ under a particular initial condition can affect the basin of attraction and therefore the final outcome.
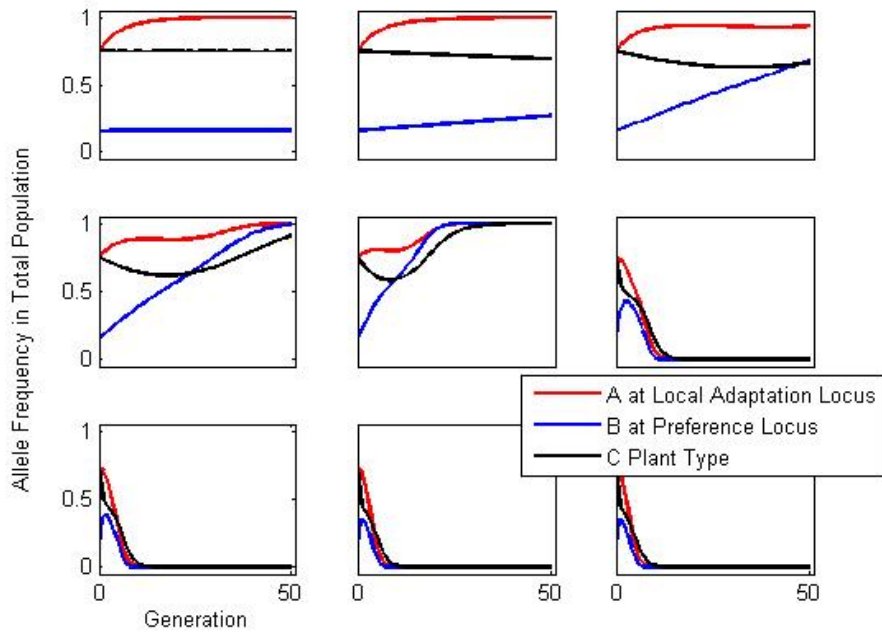
Figure 2.4: Simulations of the H-GM model with initial conditions $\vec{x}_0 = [0.1837, 0.8617, 0.0326, 0.3320]$ and $\vec{y}_0 = [0.7487, 0.2513]$ under the parameter set $r = 0.2698$, $s = 0.4896$, and $\beta = 0.4949$, with ϵ varied. From left to right and top to bottom, the values of ϵ are 0, 0.01, 0.05, 0.1, 0.25, 0.75, 0.9, 0.99, and 1.

In the case of $\epsilon = 0$, regardless of whether an insect has allele **A** or **a**, it will choose a plant of type **C** or **c** when presented with equal relative probability, $\pi = 1/2$. This results in random assortment on plants, which maintains genetic diversity of both plants and insects. The equilibria that emerge in this case are planes of equilibria:

$z = [x_1, 1 - x_1, 0, y_1]$ is stable only for $y_1 < \frac{1-s}{2-s}$, unstable elsewhere,

$z = \left[x_1, \frac{1}{s}((2-s)y_1 - x_1 s - (1-s)), \frac{(1-2y_1+y_1 s)x_1}{(2-s)y_1-(1-s)}, y_1 \right]$ is a stable manifold only for $\frac{1-s}{2-s} < y_1 < \frac{1}{2-s}$, unstable elsewhere, and

$z = [0, 0, x_3, y_1]$ is stable only for $y_1 > \frac{1}{2-s}$ is unstable elsewhere, as shown in 2.5.
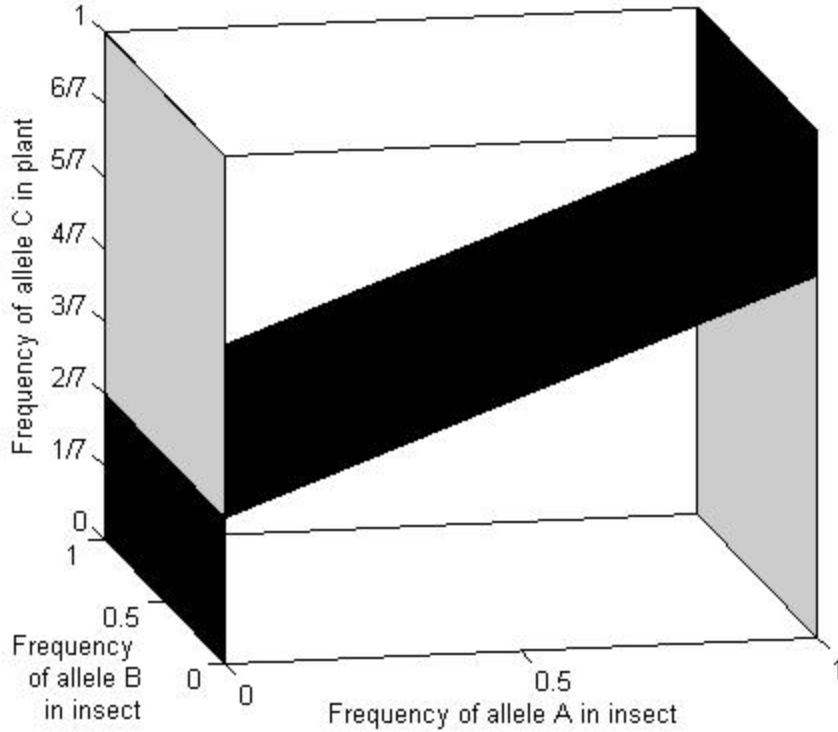
Figure 2.5: Phase portrait diagram showing unstable and stable planes of H-GM model under the parameter set to $\epsilon = 0$, $r = 0.2$, $s = 0.6$, and $\beta = 0.2$, and given that $x_1 x_4 - x_2 x_3 = 0$. There are three planes of equilibria here which intersect, leading to two transcritical bifurcations. These bifurcation lines are at $\{y_1 = \frac{1-s}{2-s} = \frac{2}{7}, x_1 = 0, x_1 x_4 - x_2 x_3 = 0\}$ and $\{y_1 = \frac{1}{2-s} = \frac{5}{7}, x_1 = 1, x_1 x_4 - x_2 x_3 = 0\}$. The unstable planes are indicated by gray shading, and the stable planes are given by black shading.

In the case of complete preference, where $\epsilon = 1$ and a non-matching plant type is never chosen, the system quickly settles to an all **AB** or **ab** insect types system, depending on the dominant plant. Separate populations are maintained in this case because the insects are assorting perfectly on their preferred plants and never mix

populations. The more frequent plant type has a higher mating probability, so will increase frequency until it becomes fixed in the population. The corresponding insect type that prefers and is adapted to the dominant plant type shortly follows.

Parameter s. When the selection coefficient is 0, lack of selection means the first adaptation locus is neutral and there is no penalty for choosing a non-matching plant. Insects with preference for the most abundant plant will have more abundant offspring, thereby driving the preference locus and the more abundant plant type to fixation.

In the case where $s = 1$, selection against non-adapted insects means the only surviving offspring leaving a plant **C** are **AB** and **Ab**. Likewise, the only surviving offspring to leave plant **c** are **aB** and **ab**. Those who have preference for and are adapted to the most abundant plant type will have a strong selective advantage as their offspring are likely to have the same qualities. This selective advantage will quickly drive the system to fixation for the most abundant plant type and the insects that prefer and are adapted to it.

Parameter r. r is the recombination rate. For $r = 0$, there is no recombination between the adaptation locus and the preference locus, like a "magic trait" locus [30]. All mixed genotypes, **Ab** and **aB**, quickly die out because they are selected against and cannot be recreated by the more fit **AB** and **ab** gene pool. Here, the most abundant plant again has the advantage in producing the most successful insects. Thus the system eventually fixes for the most abundant plant type and its associated best match insects. In the case of $r = 0.5$, assortment of the adaptation and the preference gene are independent. This works to increase the rate that **Ab** and **aB** genotype are replaced with **AB** and **ab** genotypes. As fixation is approached, retaining **AB** or **ab** genotypes is more difficult. As in the previous cases, we eventually see fixation of the most abundant plant type.

Parameter β. Recall that that β is the fraction of plant population in the next season made up by offspring produced by the insect pollination activity. $\beta = 1$ for annual plants without any secondary random pollination (like wind pollination). A

population where $\beta = 0$ would never have new offspring or all offspring would result from random wind pollination and thus would remain unchanged in composition. In this case varying β along points in the interior of its range $(0, 1)$ can change the outcome of the system as seen in Figure 2.6. It is a rather rare in occurrence for any random set of parameters and initial conditions. This means the basin boundary does not significantly shift as a result of varying ϵ. This Figure illustrates how varying ϵ under a particular initial condition can affect the basins of attraction and therefore the final outcome. Note that if one does remain in the same basin of attraction low values of β will be much slower to reach equilibrium, because only a small fraction of the population is being replaced by the population responding to the mutualistic interaction.
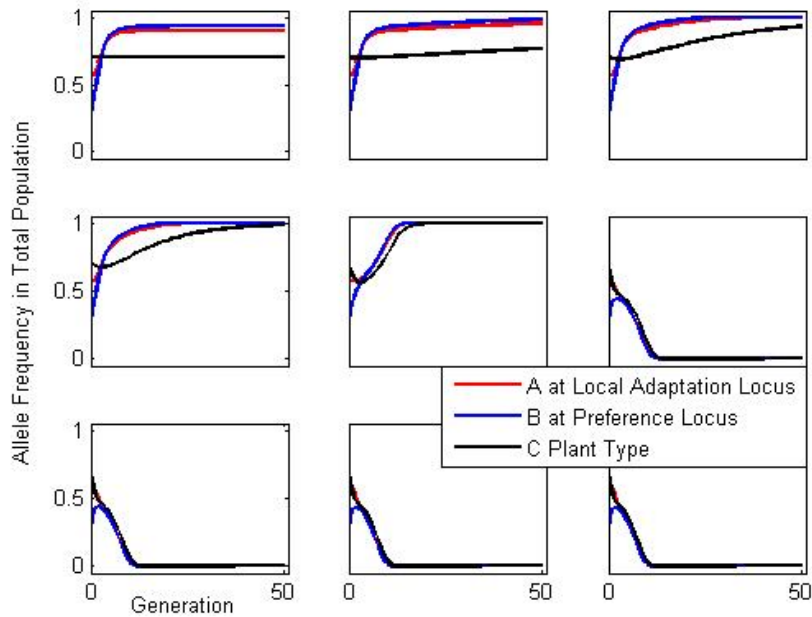
Figure 2.6: Simulations of the H-GM model with initial conditions $\vec{x}_0 = [0.1289, 0.4327, 0.1469, 0.2914]$ and $\vec{y}_0 = [0.6991, 0.3009]$ under the parameter set $r = 0.1702$, $s = 0.9597$, and $\beta = 0.5853$, with β varied. From left to right and top to bottom, the values of β are 0, 0.01, 0.05, 0.1, 0.5, 0.9, 0.95, 0.99, and 1.

2.2.5   Approximation of H-GM model

To gain more intuition into the model and its behavior, a weak selection approximation is used. Here, we assume that $\epsilon, s << r$ so that higher order terms are negligible. Below are the results for approximations of all models under consideration. Note that this assumption will result in $D = 0$, so the transformation that results in variables p_1, p_2, y and D will be reduced to only 3 variables. Recall that D=0 was satisfied in the full models' equilibria, so this simplification will maintain important aspects about the equilibria.

Performing the above approximation, the following equations result, where $q_1 = 1 - p_1$, $q_2 = 1 - p_2$, and $y_2 = 1 - y_1$.

$$
\begin{aligned}
\frac{dp_1}{dt} &= sp_1q_1(y_1 - y_2) \\
\frac{dp_2}{dt} &= 3\epsilon p_2 q_2(y_1 - y_2) \\
\frac{dy_1}{dt} &= 2\beta\epsilon y_1 y_2(p_2 - q_2)
\end{aligned}
\tag{2.18}
$$

Note: the rate of change in p_1, the frequency of the locus controlling adaptation, is proportional to the selection coefficient, the heterogeneity or variance at the adaptation locus, and the difference in plant type frequency. If there are more of plant type 1 than 2, then the frequency of those adapted to plant 1 will increase. If there are more of plant type 2 than 1, then the frequency of those adapted to plant type 1 will decrease (and so those adapted to plant type 2 will increase). Likewise, the rate of change in p_2, the frequency of the locus controlling plant preference, is proportional to the genetic variance of the plant preference locus and to the difference in plant types. This sets the stage for the preference and locus to go to fixation dependent on which type is the most dominant.

The rate of change in y_1, the locus controlling plant type, is proportional to the variance at that locus and to the difference between the frequency of those having a preference for plant type 1 and those having a preference for plant type 2. Therefore

relative plant type abundance is influenced by the relative abundance of pollinators that prefer it.

Theorem 2.2.1. *The solutions to the set of equations 2.18 are*

$$
\begin{aligned}
p_2(1 - p_2) &= A\left(y_1(1 - y_1)\right)^{\frac{3}{2\beta}} \\
p_2 &= B\left(\frac{p_1}{1 - p_1}\right)^{\frac{3\epsilon}{2s}}(y_1(1 - y_1))^{\frac{3}{4\beta}}.
\end{aligned}
$$

Proof. The two differential equations that depend on each other are

$$
\begin{aligned}
\frac{dp_2}{dt} &= 3\epsilon p_2 q_2 (y_1 - y_2) \\
\frac{dy_1}{dt} &= 2\beta\epsilon y_1 y_2 (p_2 - q_2),
\end{aligned}
$$

which, when divided, yield

$$
\frac{dp_2}{dy_1} = \frac{3p_2 q_2 (y_1 - y_2)}{2\beta y_1 y_2 (p_2 - q_2)}.
$$

Recall that $q_2 = 1 - p_2$ and $y_2 = 1 - y_1$. Then

$$
\frac{dp_2}{dy_1} = \frac{3p_2(1 - p_2)(2y_1 - 1)}{2\beta y_1(1 - y_1)(2p_2 - 1)}.
$$

Separate variables and integrate:

$$
\int \frac{2p_2 - 1}{p_2(1 - p_2)} dp_2 = \frac{3}{2\beta} \int \frac{2y_1 - 1}{y_1(1 - y_1)} dy_1.
$$

Integration using partial fractions:

$$
ln|p_2(1 - p_2)| = \frac{3}{2\beta} ln|y_1(1 - y_1)| + C_1,
$$

where C_1 is a constant of integration.

This implies that $p_2(1 - p_2) = A\left[y_1(1 - y_1)\right]^{\frac{3}{2\beta}}$, where $A = e^{C_1}$.

Now consider

$$\frac{dp_1}{dt} = sp_1q_1(y_1 - y_2)$$

$$\frac{dp_2}{dt} = 3\epsilon p_2q_2(y_1 - y_2).$$

Again, divide and separate by variables:

$$\int \frac{dp_1}{p_1(1-p_1)} = \frac{s}{3\epsilon}\int \frac{dp_2}{p_2(1-p_2)} \Rightarrow$$
$$ln|p_1| - ln|(1-p_1)| = \frac{s}{3\epsilon}(ln|p_2| - ln|(1-p_2)|) + C_2.$$

Recall from above $ln|p_2| + ln|(1-p_2)| = \frac{3}{2\beta}ln|y_1| + ln|(1-y_1)| + C_1$, so

$$2ln(p_2) = \frac{3\epsilon}{s}ln\left(\frac{p_1}{1-p_1}\right) + \frac{3}{2\beta}ln(y_1(1-y_1)) + C_3.$$

Thus, we conclude

$$p_2 = B\left(\frac{p_1}{1-p_1}\right)^{\frac{3\epsilon}{2s}}(y_1(1-y_1))^{\frac{3}{4\beta}}. \qquad \square$$

Phase Plane

Note that the adaptation locus does not influence the dynamics at the preference locus or the plant type relative abundance. Therefore, analysis of the solutions and phase portrait of p_2 versus y_1 reveals how the initial conditions determine long term dynamics. This phase portrait is illustrated in Figure 2.7 for a particular parameter set. It shows that the equilibrium point at $\left(\frac{1}{2}, \frac{1}{2}\right)$ is a saddle point. Also note that the boundary separating the basin of attraction for $(0,0)$ and $(1,1)$ is the stable manifold for the saddle point [90].
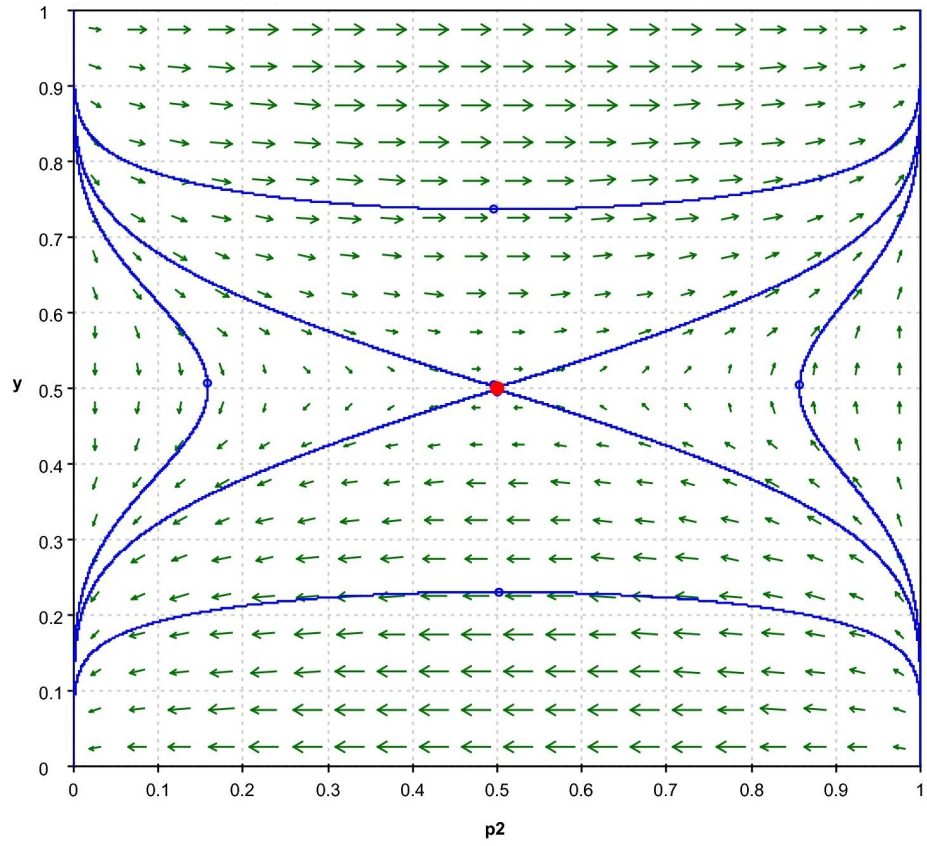
Figure 2.7: Phase portrait of p_2 vs y_1 for Equation 2.18 for $\epsilon = 0.1$ and $\beta = 0.2$.

Theorem 2.2.2. *The separatrices, which are the trajectories of the stable manifold of the equilibrium point at* $\left(\frac{1}{2}, \frac{1}{2}\right)$ *are*

$$4p_2(1 - p_2) = [4y_1(1 - y_1)]^{\frac{3}{2\beta}},$$

for $p_2 \in \left(0, \frac{1}{2}\right)$ *and* $p_2 \in \left(\frac{1}{2}, 1\right)$.

Proof. The solutions depicted in the phase portrait in Figure 2.7 are in the proof of the previous theorem:

$$p_2(1 - p_2) = A \left[y_1(1 - y_1)\right]^{\frac{3}{2\beta}}$$

.

Both trajectories end in $frac12, frac12$ as $t-> \inf$. With this condition we get that the basin boundary is:

$$4p_2(1 - p_2) = [4y_1(1 - y_1)]^{\frac{3}{2\beta}}$$

. □

The basin of attraction is influenced by β, but the ϵ cancels when finding the solution. Figure 2.4 does not satisfy the parameter assumptions of this approximation, so the approximation does not conflict with full numerical results. Note that the approximation loses the ability to predict the small changes in the basin boundary due to modifying ϵ.

Host-switching

A host-switch to a **C** (type 1) plant is defined as a trajectory with initial condition $p_2 > .5$ and $y_1 < .5$ that goes to the stable equilibrium at $(p_2, y_1) = (0, 0)$. In the example illustrated in Figure 2.7 where $\beta = 0.2$, the percentage of area in which this scenario happens is approximately 13.2 percent. This result is attained by integrating the area under the separatrix with the aforementioned limits. Similarly, since the

system is symmetrical, 13.2 percent of the initial condition area will result in a host switch to a **c** (type 2) plant (defining a host-switch to type 2 plant to be trajectory with initial condition $p_2 < .5$ and $y > .5$ that tend towards the stable equilibrium at $(p_2, y_1) = (1, 1)$).

2.3 H-GM Model Variations

2.3.1 H-GM plant resource dependent model

The full analytical results are shown in the Appendix. A total of twenty equilibria emerge, sixteen of which are listed in Table A.1.

This model variation is also bistable for the fixed states and the final solution depends on the initial condition. Numerical work confirms the instability of the polymorphic equilibrium whose stability matrix was rather intractable for confirming analytically. Figure 2.8 depicts the trajectories of 100 numerical runs plotted for the same set of parameter values, but with random starting initial conditions.

To see the effect of the initial conditions on the long-term dynamics, simulations shown in Figure 2.9 are compared and contrasted. In these figures, parameter sets are the same, but have different initial conditions. Figure 2.9c shows the host-switch case in which the insect population, despite being nearly fixed for the **A** and **B** alleles, evolves to a population fixed for **a** and **b**. This is because of the initial proportion of **c** plants in the system. Note the similarity to the original model.

Using the approximation techniques discussed above for the H-GM plant resource-dependent model, the following equations result:

$$
\begin{aligned}
\frac{dp_1}{dt} &= sp_1q_1(y_1 - y_2) \\
\frac{dp_2}{dt} &= 2\epsilon p_2 q_2(y_1 - y_2) \\
\frac{dy_1}{dt} &= 2\beta\epsilon y_1 y_2(p_2 - q_2).
\end{aligned}
\tag{2.19}
$$

Figure 2.8: Phase portrait showing 100 trajectories of the plant frequency dependent H-GM model for random initial conditions. Parameter values are set to $\epsilon = 0.5$, $r = 0.2$, $s = 0.6$, and $\beta = 0.2$. Red dots indicate the value of the allele frequencies after 500 generations.
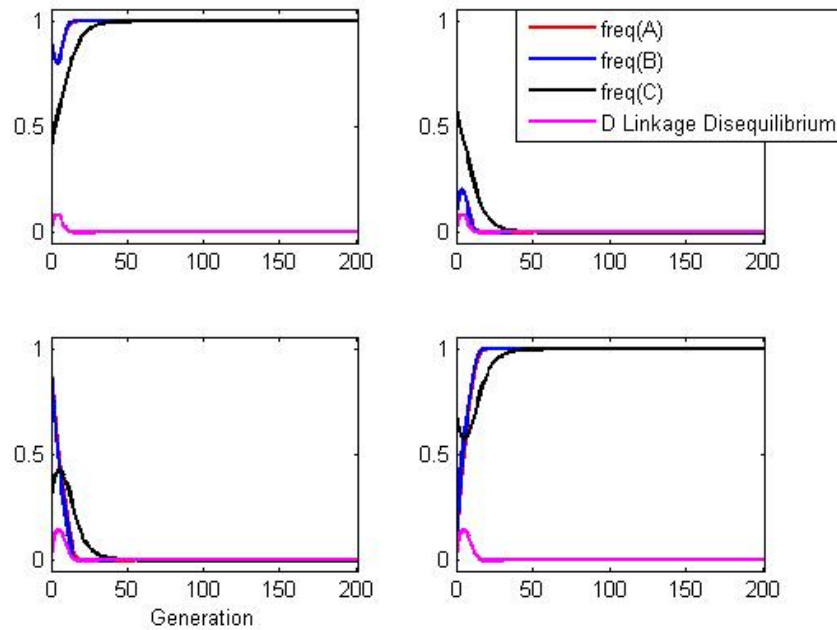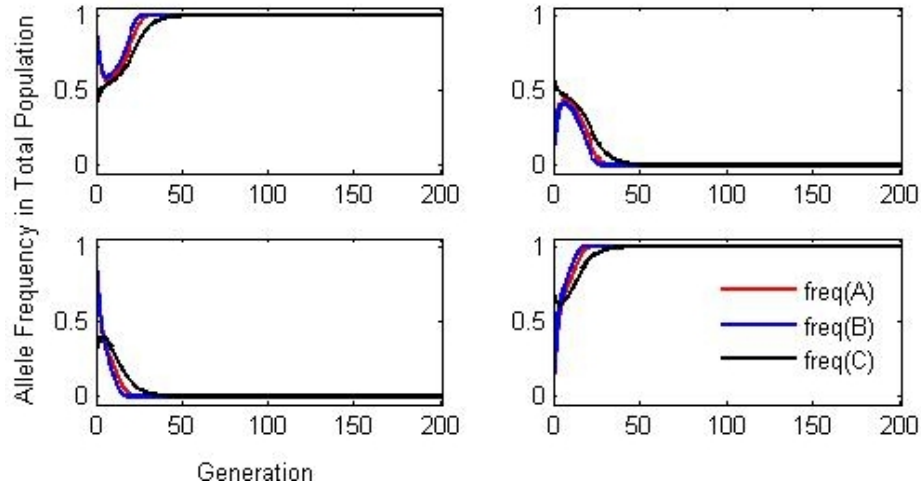
Figure 2.9: Plant-frequency dependent H-GM model simulations showing systems robust to some introduced variation, but that can also result in host-switching with a critical relative abundance of alternate host type. Parameter values for each of these graphs are the same set to $\epsilon = 0.5$, $r = 0.2$, $s = 0.6$, and $\beta = 0.2$. The blue and red lines are nearly overlapping in these cases. The graphs in each column (a,c and b,d) have the same initial insect frequencies, but the plant frequencies in the top and bottom graphs differ by only 0.1. This shows that the frequency of an alternate plant type heavily influences the long-term relationship.

This model is the same as 2.18. This explains why this variation of the full model has dynamics like that of the original H-GM model.

2.3.2   H-GM plant resource dependent model without re-assortment

In the above models, the qualitative dynamics are exactly the same, even down to the eigenvalues. Considered next is the model variation in which mating and egg-laying were done on the same plant without the re-assortment of females. Qualitative dynamics appear to be the same from numerical simulations, but the analytical advantages are not significant. Therefore, this model is less useful because this simplification requires a biologically less realistic assumption.

Using approximation techniques to the H-GM plant resource dependent model without reassortment, the following equations are derived:

$$\begin{aligned}
\frac{dp_1}{dt} &= sp_1q_1(y_1 - y_2) \\
\frac{dp_2}{dt} &= 2\epsilon p_2 q_2(y_1 - y_2) \\
\frac{dy_1}{dt} &= 2\beta\epsilon y_1 y_2(p_2 - q_2)
\end{aligned} \tag{2.20}$$

This model again reduces to 2.18. This explains why this variation of the full model also has dynamics like that of the original H-GM model.

2.4 Discussion

2.4.1 Discussion of the H-GM model

Analytic results show the cases of fixation of **AB** and **C** genotypes or **ab** and **c** genotypes are stable equilibria to which the system evolves long term for realistic parameter conditions. A few exceptions exist as a result of special cases discussed above. Interestingly, numerical simulations and analytical results from model approximations show host-switching as a possible outcome - where a system that has a majority of one type of insects evolves to exploit a dominant plant population of the other type. This agrees with biological observations about this system [36, 93] in which host-switching can occur rather easily. Whether or not a host-switch occurs is influenced mostly by the parameter β which controls how fast the plant population responds to matching insect abundance. It is less affected by the relative preference insects have for their matching plant, but this observation is not captured in the reduced model.

From analytic work and numerical simulations, like that in Figure 2.5, it is shown that a complete lack of preference leads to a long-term maintenance of genetic variation in the population. This is counterintuitive, because it has been conjectured

that cases of co-existence in these mutualistic systems are a result of strict preferences [57]. However, if the insects are choosing to oviposit randomly in plants, then the plant frequencies remain unchanged. This essentially converts the system to a two resource problem with only the insect evolving, and thus maintains genetic variability in the population.

Maintenance of genetic variation in this *Greya* moth system is more likely due to a Geographic Mosaic of Co-evolution [93] rather than a complete lack of pollinator preference. However, in many plant-pollinator systems, insects do not show significant correlation between adaptation and preference [1], and strict preferences can both facilitate and prevent speciation [30]. This lack of preference for host may contribute to observed instances of more than one species of wasp ovipositing on the same fig as well as explaining the large amount of hybridization among fig types. This will be revisited in models of the fig fig-wasp system.

2.4.2 Discussion of the H-GM model variations

It was shown that multiple reassortment merely shuffles insects, maintaining their frequencies on each plant. Therefore, deletion of one or more trips that the female insect would not make a qualitative difference and may be used to simplify our model. These models share the qualitative behavior of the original model despite the quantitative changes in model formulation. Unfortunately the variations considered did not make the stability analysis for the nontrivial equilibrium simpler, and thus there is no advantage to using one model over the other unless one of the model variations more closely matches the life-cycle of a different model organism under consideration.

2.4.3 Conclusion

Maintenance of genetic variation in individual populations are not possible in the H-GM system, except for in special cases of no relative preference ($\epsilon = 0$) or absolute

relative preference ($\epsilon = 1$). Instead, the system will always evolve to one fixed state or another for a particular matching allele set. Whether or not the coevolutionary result is to maintain status quo or to host-switch depends primarily on the relative abundance of alternate host type and how fast the plant population can evolve to insect pressure to host-switch. In the example considered in the approximation of the H-GM model in Figure 2.7, 26 percent of all initial conditions result in host-switch, which is rather high probability of occurance. Next, the H-GM model is expanded to the well-studied biological model system fig and fig wasp. The life cycle is different because fig plant populations do not all flower at the same time. This may affect the ability for the system to experience maintenance of genetic variation and host-switching.

Chapter 3

Modeling the Fig-Fig Wasp Mutualism

Results from Chapter 2 provide a baseline for understanding what life-cycle features and parameter values may be important in studying the effect of mutualistic interactions on coevolution. Next, a model is developed for the fig-fig wasp mutualism. In *Ficus-Courtella wardi* [96], the female enters a female phase fig, but in doing so loses her wings and dies within the fig after laying eggs in each ovule. During the interfloral phase of the fig, the fig wasp larvae mature then mate within the fig. When the fig reaches the male phase, the male fig wasps chew an opening for the females, and the females leave the fig carrying pollen. The plant genetics, however, become much more complicated. The male phase of the fig from which females emerge, laden with eggs, must overlap with the female phase (the start of a new floral generation of another fig) so that the fig wasp can pollinate the fig and oviposit her eggs. These overlapping generations are a major biological difference from the H-GM system.

Suppose networks of asynchronously flowering fig tree populations are closed. That is, the last flowering event pollinates the first population's next generation of fig trees. Consider a small pollination network of only two populations of figs that flower asynchronously. Note that this implies that fig wasps have only two generations in

the time that it takes for both fig populations to go through their male and female phases; see Figure 3.1.
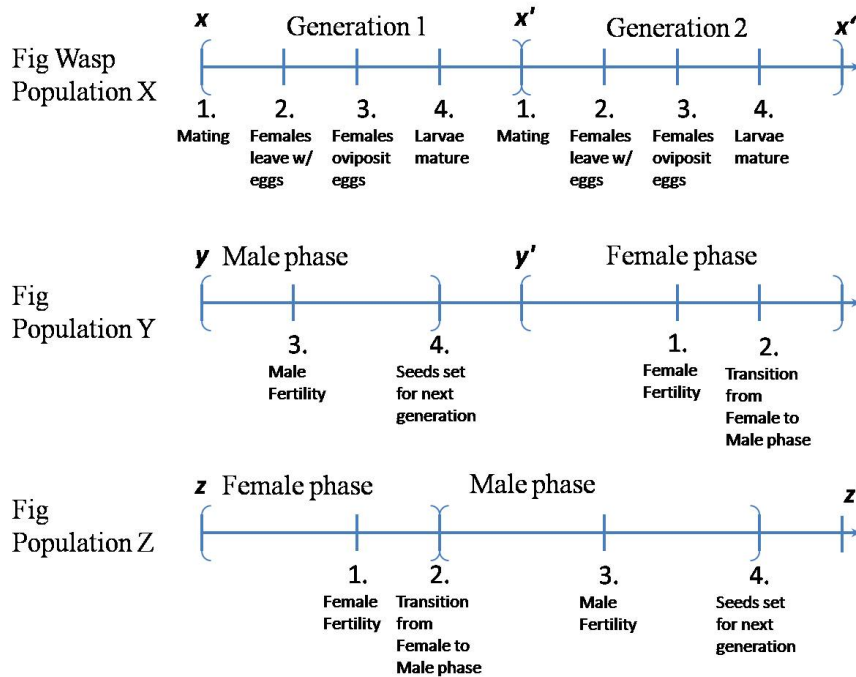
Figure 3.1: Figure showing the overlapping asynchronously flowering plant populations **Y** and **Z**, and how they match up with the insect generations.

3.1 Fig-fig Wasp Model

3.1.1 Dynamics of insects

The frequency of adult insects of type i found on plants of type m is denoted by $x_{i,m}$ *. Let y_m and z_m be the frequencies of the types of plants during the insect mating period in male and female phase during even during odd generations, respectively, and female and male phase during even generations, respectively. Again, indices

---

*Note: This model starts with the part of the insect life-cycle after larval selection. So if we assume all loci under consideration are diallelic, then we have 8 insect types to consider. We also now have 4 types of plants, 2 types in each population. This contrasts the H-GM model in which we started the life cycle with the females in the air assorting to plants to mate. In the fig wasp, most of the life cycle takes place within one fig (so 8 cases between the two fig types currently harboring a population), and when the female leaves, she leaves with eggs (so 16 cases).

i, j, k are used for insects and l, m, n for plants. Assuming random mating on host, the frequency of mating pairs formed by females i and males j on a plant of type m is

$$M_{ij,m} = x_{i,m} x_{j,m}. \tag{3.1}$$

Therefore, the frequency of eggs with genotype k produced by pairs (i, j) mating on a plant of type m is $M_{ij,m} R(i, j \to k)$, where R gives the corresponding recombination frequencies for the genotypes in this system.

Females leave the fig carrying eggs[†]. Assume each fig has a carrying capacity of wasps that it can support. Therefore, the frequency of females leaving a particular fig, *i.e.* the contribution of each mating pool to offspring, is equal to the proportion of the fig type in that system. This is a reasonable assumption since there is a limited resource inside the fig which leads to larval competition[98]. Females leave the male phase figs and bring pollen to female phase figs. The proportion of eggs with genotype k and carried by mothers with genotype i in the whole set of eggs before oviposition is

$$E_{k,i} = \sum_m y_m \sum_j M_{ij,m} R(i, j \to k). \tag{3.2}$$

Females then search for a suitable female phase fig in which to oviposit the eggs. The frequency of eggs with which genotype k oviposited on a plant of type n is

$$e_{k,n} = \frac{\sum_i E_{k,i} \pi_{in}}{\sum_k \sum_i E_{k,i} \pi_{in}} = \sum_i E_{k,i} P_y(i, n), \tag{3.3}$$

where

$$P_y(i, m) = \frac{\pi_{im}}{\sum_i \pi_{im} \sum_m y_m x_{i,m}}. \tag{3.4}$$

After selection on a plant of type n, the frequencies of larvae in the plant are

$$x'_{k,n} = e'_{k,n} = \frac{e_{k,n} W_{k,n}}{\overline{W}_n}. \tag{3.5}$$

[†]This is the only time females are in the air, leaving the host plant. This is unlike the original *Greya* moths which required they mate and lay eggs on different plants, assorting multiple times.

Recall it is assumed that the percentage contribution of each larva pool (*i.e.* plant) to the total larvae population is equal to the proportion of each plant type. Therefore the final equation for insects going from the Y population of plants to the Z population of plants is

$$x'_{k,n} = \sum_i \sum_j \sum_m y_m P_y(i,n) x_{i,m} x_{j,m} R(i,j \to k) \frac{W_{k,n}}{W_n}. \tag{3.6}$$

Two plant populations need to be fertilized consecutively to complete a full plant generation of both male and female phase plant [‡]. Therefore, two generations of insects are modeled consecutively. The second generation is calculated in the same way above except the plant population in the male phase is Z, so where y was used, now z is used.

$$x''_{k,n} = \sum_i \sum_j \sum_m z_m P_z(i,n) x'_{i,m} x'_{j,m} R(i,j \to k) \frac{W_{k,n}}{W_n}. \tag{3.7}$$

3.1.2   Dynamics of fig trees

The probability that a female of type i goes to a plant of type n to lay the eggs is

$$\frac{\pi_{in} z_n}{\sum_n \pi_{in} z_n} = Q_z(i,n) z_n, \tag{3.8}$$

where

$$Q_z(i,n) = \frac{\pi_{in}}{\sum_n \pi_{in} z_n}. \tag{3.9}$$

The female insect pollinates a plant n in population Z by pollen from plant m in population y with probability

$$F_{m,n} = \sum_i y_m x_{i,m} Q_z(i,n) z_n \tag{3.10}$$

[‡]To model the second plant population, one more insect generation needs to be modeled. This second insect generation will depend on z, as Z is now the plant population that is in the male phase.

which can be interpreted as mating probability of male plant type m from the Y population of fig trees and female plant type n from the z population of fig trees. Thus, the frequency of plant l produced as a result of mating of plants m and n is $F_{mn}S(m, n \to l)$, where $S(m, n \to l)$ is the corresponding segregation probability.

Assume flowering time is paternally inherited, *i.e.* the offspring of these matings from the Y population to the Z population will contribute offspring only to the Y flowering population. So the frequency of offspring is

$$y_{l,o} = \sum_m \sum_n F_{mn} S(m, n \to l) = \sum_i \sum_m \sum_n y_m z_n x_{i,m} Q_z(i,n) S(m, n \to l). \quad (3.11)$$

Finally, to account for the fact that figs last more than one generation, let only a proportion β of plants be replaced each generation. Then

$$y_l' = (1 - \beta) y_l + \beta \sum_i \sum_m \sum_n y_m z_n x_{i,m} Q_y(i,n) S(m, n \to l). \quad (3.12)$$

The model developed thus far models Y pollinating Z. Modeling the next generation of fig wasp pollination, when Z pollinating Y, leads to a final equation of

$$z_l' = (1 - \beta) z_l + \beta \sum_i \sum_m \sum_n z_m y_n' x_{i,m}' Q_z(i,n) S(m, n \to l). \quad (3.13)$$

3.2 Fig-Fig Wasp Model Results

The long term behavior of the F-FW model is the same as the H-GM model for most parameter values, except those discussed below. See the phase portrait in Figure 3.2. The F-FW model is bistable, where the system will evolve to completely **AB**, **C** or **ab**, **c** depending on initial conditions. It can also experience some robustness to a moderate amount of introduced variation, or can experience host-switching given a high enough relative abundance of alternative plant type (See Figure 3.4 to see a close-up of simulations in the first 50 generations).
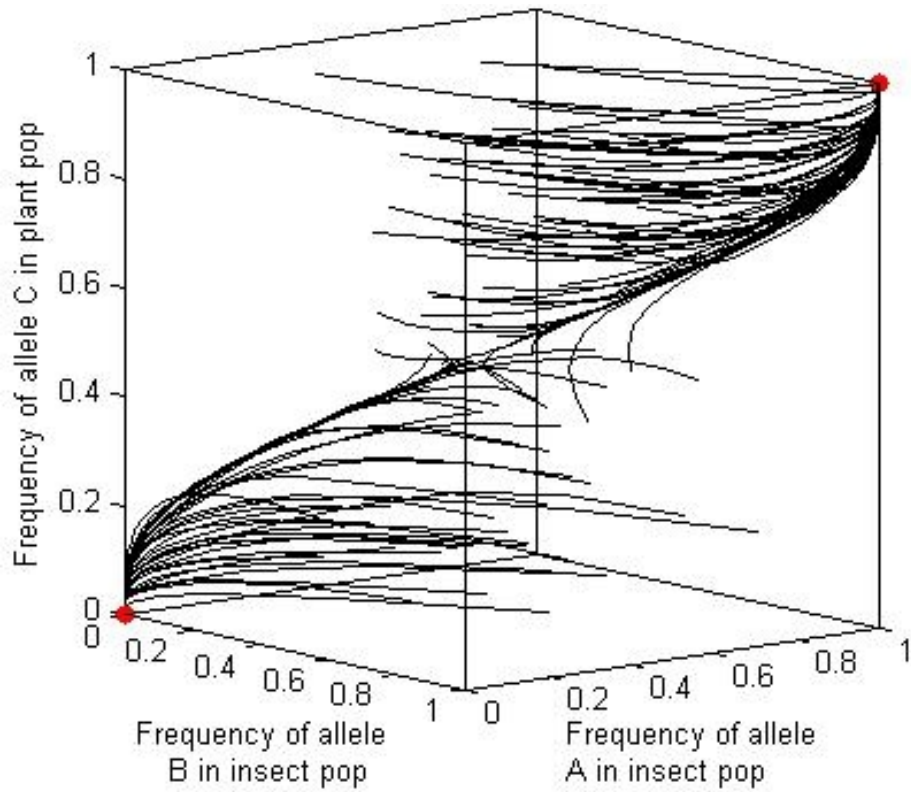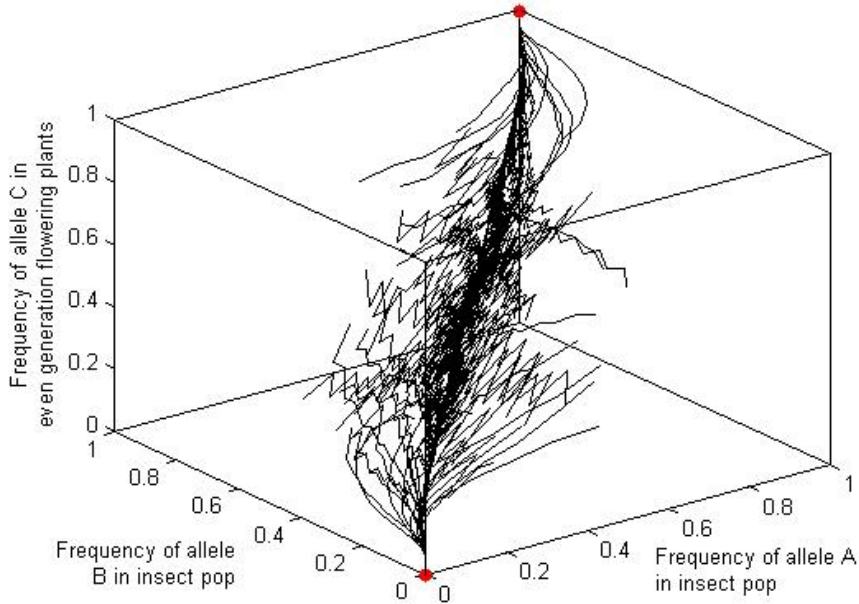
Figure 3.2: Phase portrait showing 100 trajectories of the F-FW model for random initial conditions. Parameter values are set to $\epsilon = 0.5$, $r = 0.2$, $s = 0.6$, and $\beta = 0.2$. Red dots indicate the value of the allele frequencies after 2000 generations.

3.2.1 The effect of parameters in the F-FW model

Parameter ϵ. Like the H-GM model, numerical simulations indicate that for $\epsilon = 0$, corresponding to no preference bias for like plant types, the system will maintain genetic diversity. For intermediate values of ϵ, the system is bistable for the fixation of either capital or lower-case alleles. Furthermore, host-switching as well as classic 1-1 coevolution can occur as seen in Figure 3.3.

The transient behavior of the F-FW model differs from the H-GM model with respect to the preference parameter, ϵ as seen in Figure 3.4. If the odd generation flowering plants are different in composition from the even generation flowering plants, then the insect genetics will oscillate between the two until the odd and even generation flowing plants are genetically similar in composition. For values of ϵ nearer to 1, corresponding to very specific preference, the population eventually goes to a fixed state. This is achieved only after several hundreds to thousands of generations of
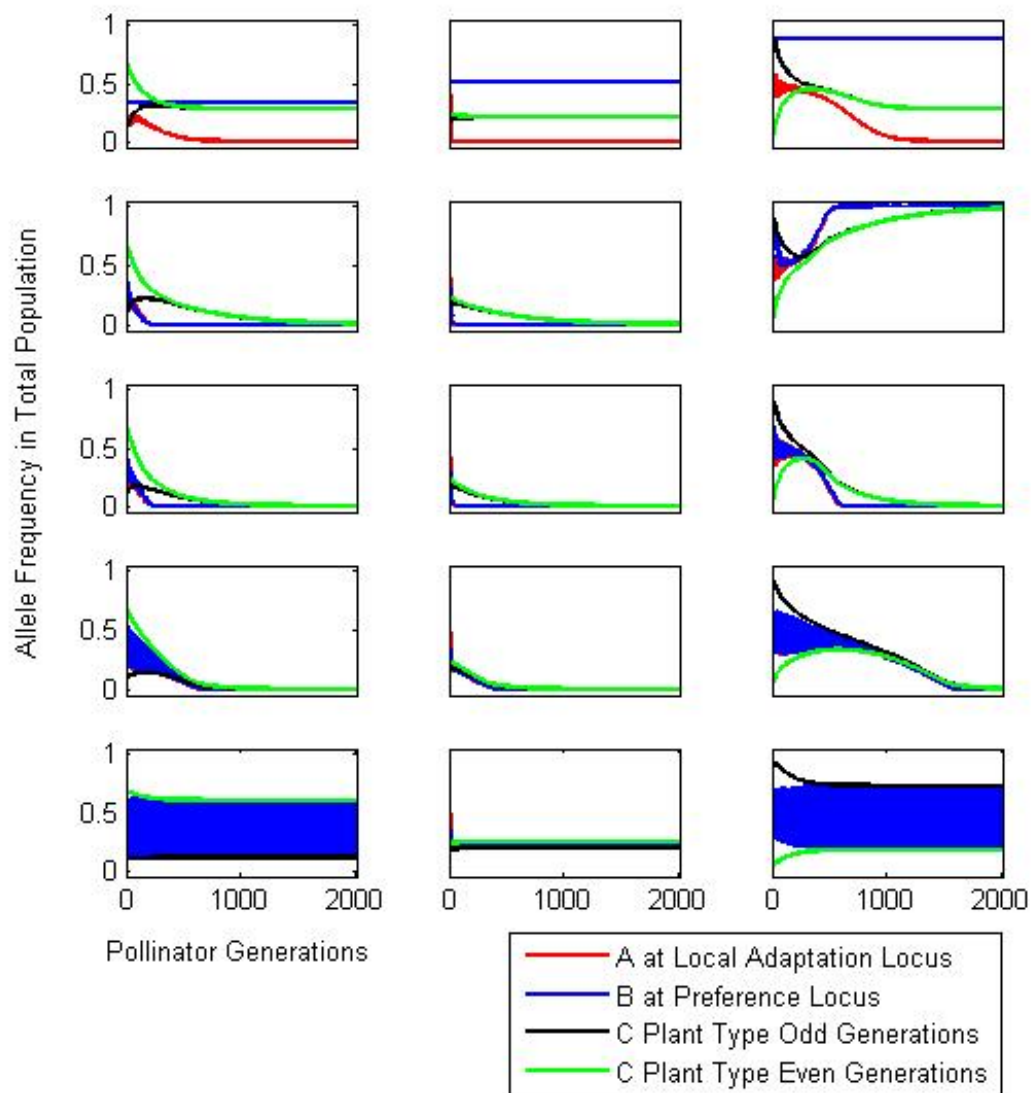
Figure 3.3: Possible trajectories of the F-FW model over 2000 total pollinator generations. The simulations in each column have the same set of initial conditions, and values for r, s, and β (0.2, 0.6, and 0.02, respectively). Each row varies ϵ from 0 (top row) to 1 (bottom row) in increments of 0.25. Note that the solid blue areas are actually very rapid oscillations of both the red and blue solutions. For better resolution of this, see Figure 3.4

oscillating insect populations affected by the plant populations' plant type frequencies. At complete pollinator preference ($\epsilon = 1$), this period-two oscillation is stable, with the genetic frequencies of the pollinator populations forever oscillating between what is optimal for each flowering population. Note that these results are only attainable in the F-FW model with asynchronous flowering, not the H-GM model. This result is discussed later in context of the resulting fig and fig wasp phylogenetic trees.

The parameter, ϵ, also affects the time to fixation and basin of attraction that divides the two stable fixed points, like in the H-GM model. This is illustrated in Figure 3.5, where the final outcome changes as ϵ is increased.

Parameter s. The selection coefficient, s, does not impact the stability of equilibria, except in special cases (*e.g.* when there is no selection), but does affect the time to equilibrium, and can shift the basin of attraction for the stable equilibria, like in the H-GM model as seen in Figure 3.6.

Parameter r. The probability of recombination or recombination rate, r, between the two insect loci under consideration also affects the basin of attraction and the time to equilibrium as seen in Figure 3.7. For $r = 0$, as in the H-GM model, **Ab** and **aB** genotypes are quickly removed as the system goes to a fixed state. In the special case of $r = 0.5$, favorable genotypes can be produced quickly initially, but the high recombination rate can also break apart favorable combinations. In both special cases only the **AB, C, C** and **ab, c, c** states are possible for insect and the two asynchronously flowering plant populations.

Parameter β. The contribution of seeds to the next generation, β, impacts the trajectories that divides the basin of attraction for the two fixed points and influences the time to equilibrium. We see this in FigureFig:FIG b basin. If $\beta = 0$, then there is no fixation of a single genotype because the plant populations cannot evolve, whereas if $\beta = 1$ evolution is very rapid as plants can respond more quickly.
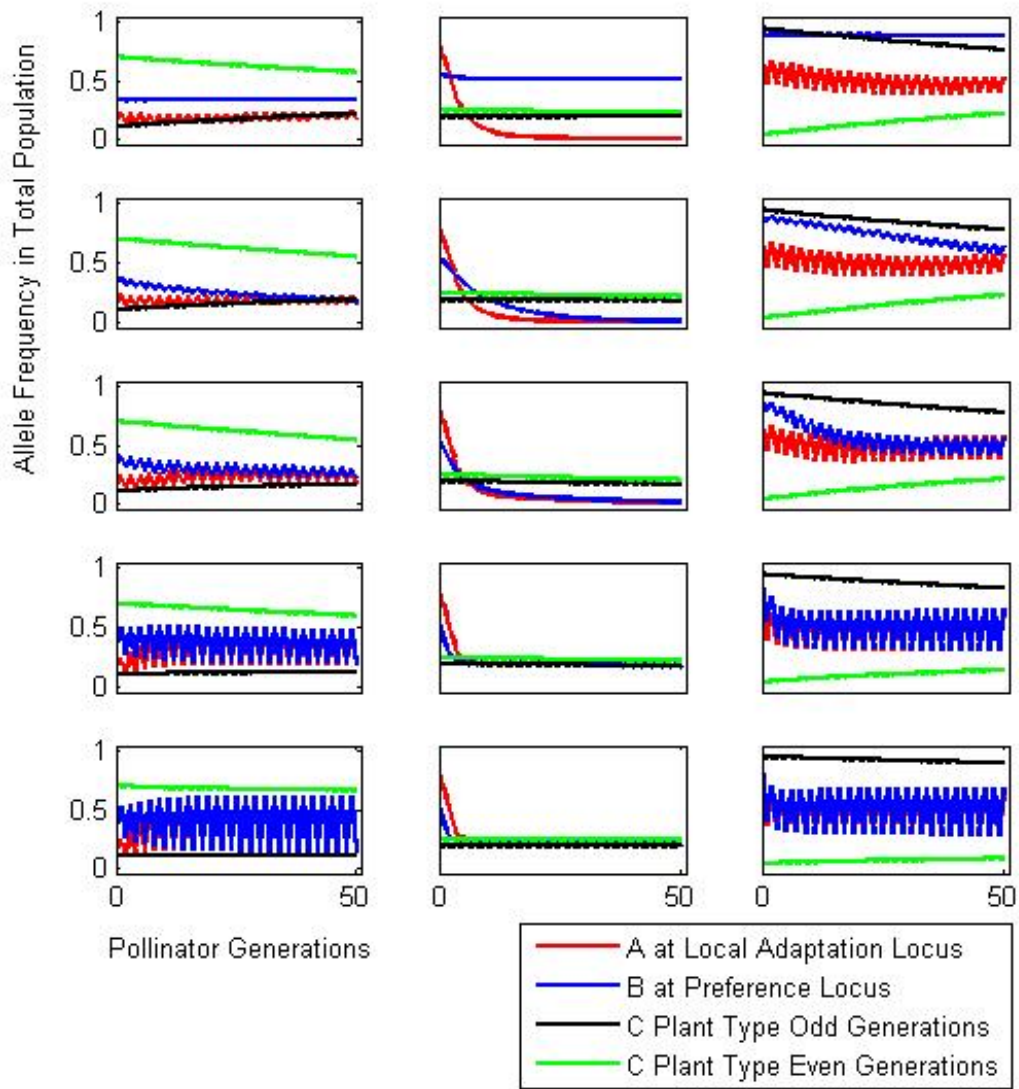
Figure 3.4: These are the same F-FW model simulations as shown in Figure 3.3, but zoomed in over the first 50 generations to show the oscillatory behavior.

48
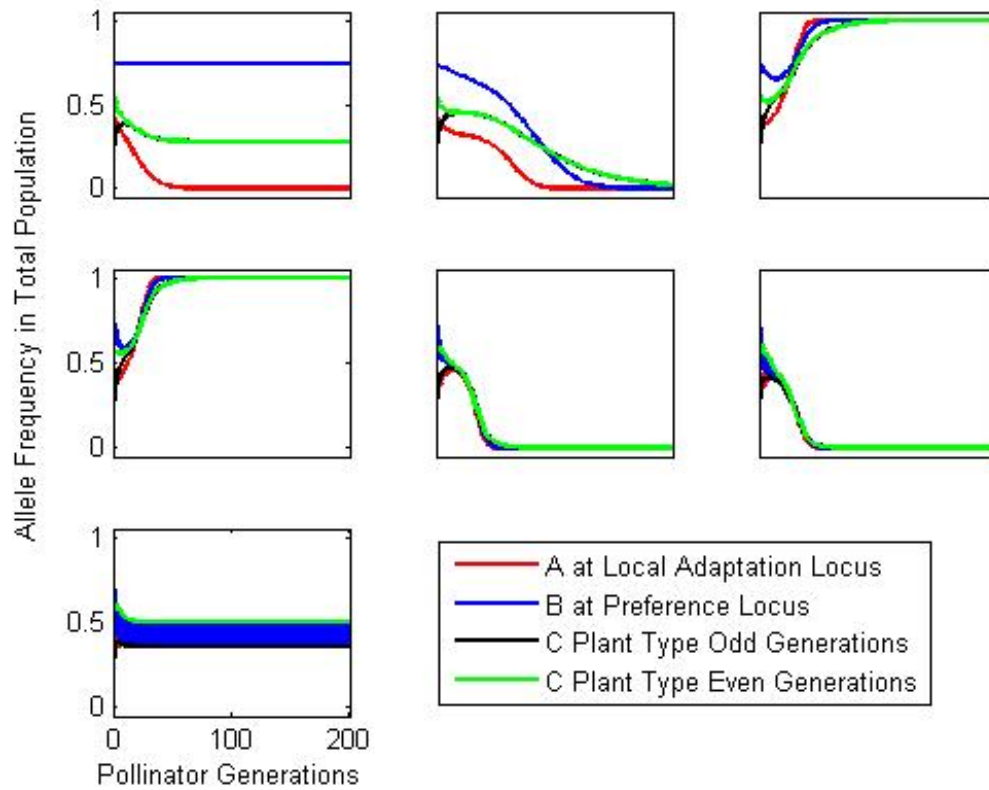
Figure 3.5: Simulations all with the same parameters and same initial conditions, but varying ϵ. Parameters are $s = 0.3899, r = 0.5909, b = 0.4594$ and ϵ values moving from left to right then top down are $0, .1, .25, .5, .75,$ and 1. Initial conditions for all runs are $y0 = 0.2180, z0 = 0.5716,$ and $x_0 = [0.0446, 0.4091; 0.2026, 0.0370; 0.7390, 0.3169; 0.0138, 0.2369]$ This illustrates that varying the relative preference only, ϵ, can affect final outcome, *i.e.* affects the basin of attraction.

49
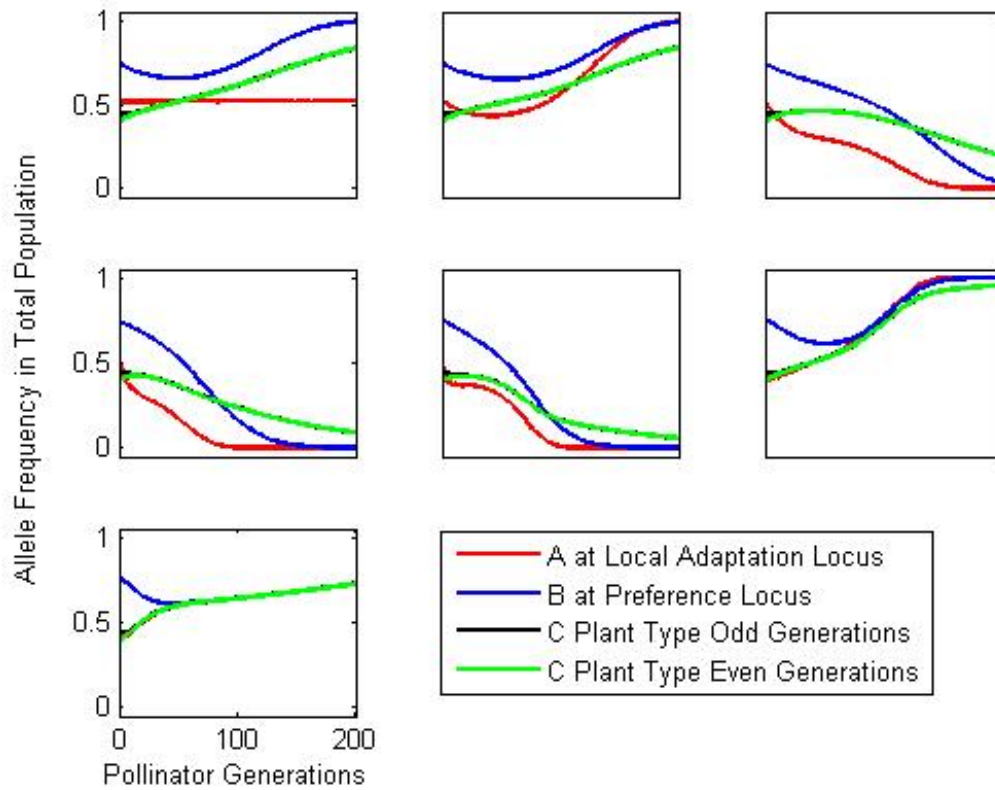
Figure 3.6: Simulations all with the same parameters and same initial conditions, but varying s. Parameters are $r = 0.6393, b = 0.2554;, er = 0.0887$ and s values moving from left to right then top down are $0, .1, .25, .5, .75,$ and 1. Initial conditions for all runs are $y0 = 0.4425, z0 = 0.3934,$ and $x_0 = [0.3447, 0.3096; 0.2404, 0.1510; 0.38980.4383; 0.0251, 0.1011]$ This illustrates that varying the selection coefficent only, s, can affect final outcome, *i.e.* affects the basin of attraction.
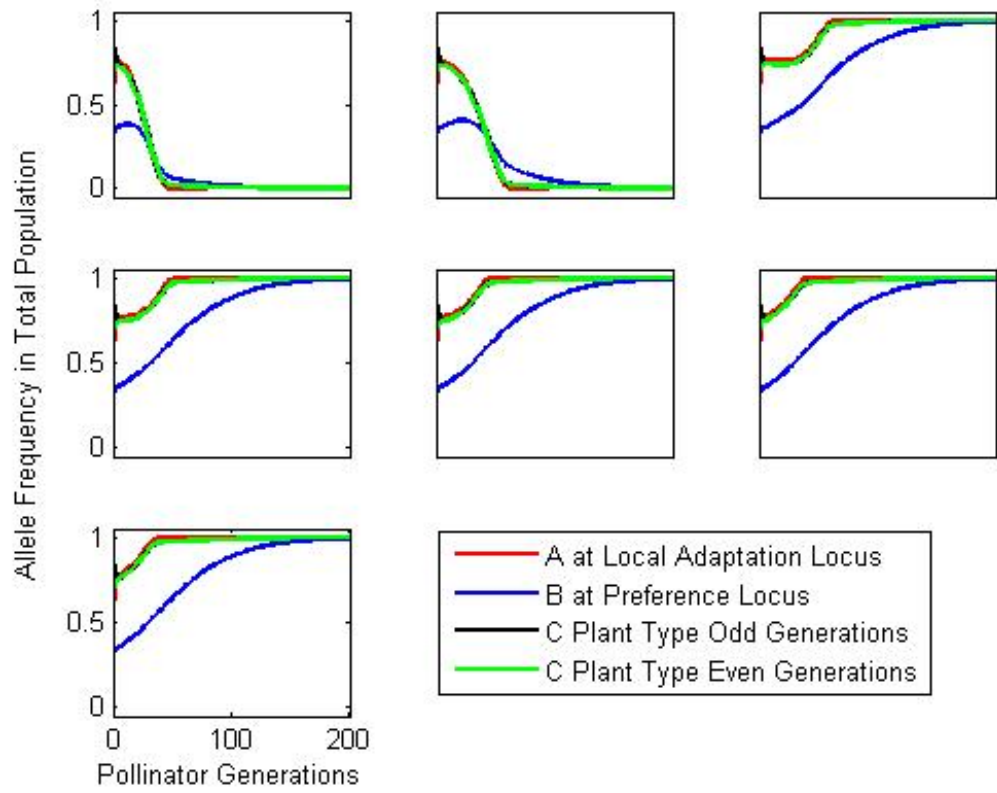
Figure 3.7: Simulations all with the same parameters and same initial conditions, but varying r. Parameters are $s = 0.9521, b = 0.9759, er = 0.0309$ and r values moving from left to right then top down are $0, .1, .2, .25, .3, .4$ and $.5$. Initial conditions for all runs are $y_0 = 0.9044, z_0 = 0.6804$, and $x_0 = [0.2029, 0.4459; 0.3545, 0.3447; 0.0998, 0.0012; 0.3428, 0.2081]$ This illustrates that the recombination rate affects the basin of attraction.
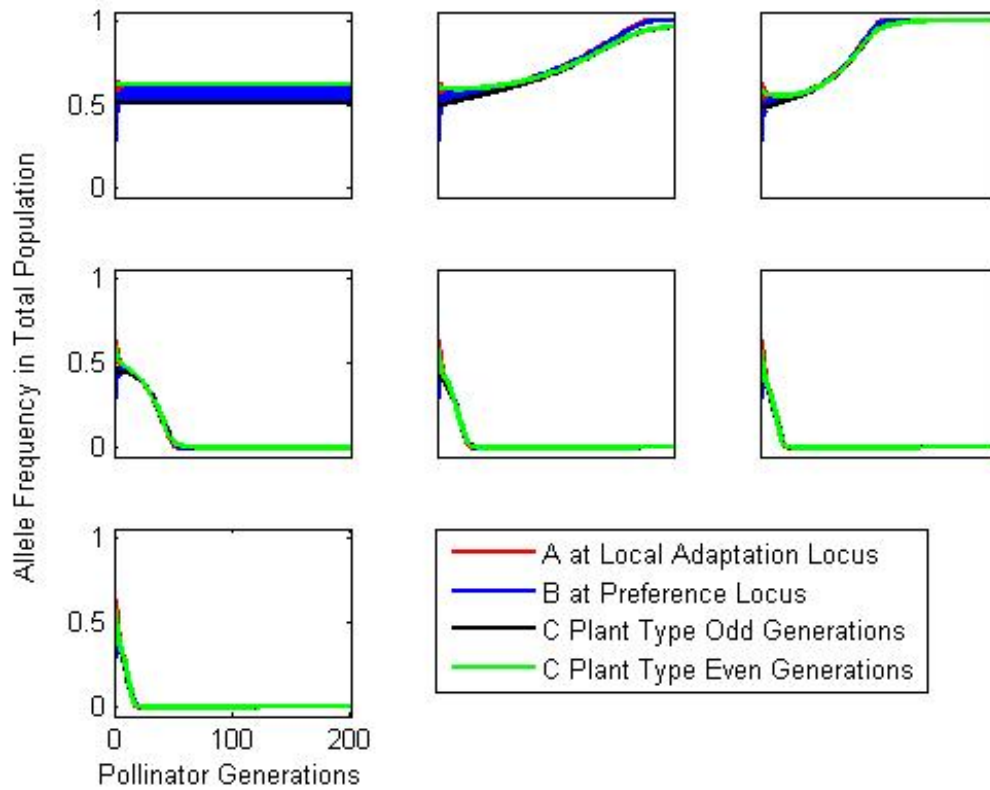
Figure 3.8: Simulations all with the same parameters and same initial conditions, but varying β. Parameters are $s = 0.7267, r = 0.5158, er = 0.7906$ and β values moving from left to right then top down are $0, .1, .2, .25, .3, .4$ and $.5$. Initial conditions for all runs are $y_0 = 0.5100, z_0 = 0.6149$, and $x_0 = [0.1178, 0.2585; 0.3906, 0.3604; 0.0302, 0.0349; 0.4614, 0.3461]$ This illustrates that β affects the basin of attraction.

3.3 Discussion of Fig-Fig Wasp Model Results

The major difference in behavior between the F-FW model and the H-GM model is in the transient behavior and response to the preference bias parameter, ϵ. Even under intermediate values of ϵ, oscillatory insect and differing plant frequencies between flowering populations can persist for thousands of insect generations. A typical fig wasp lineage may rely on 8-10 different flowering fig populations in order to maintain its viability. In this case, it is expected that the transient behavior seen in the two population case will last even longer, perhaps tens of thousands of generations under realistic parameter conditions. This may be extremely important in clarifying why so many fig phylogenies are extremely difficult to resolve, pointing to extensive hybridization and confusion when determining historical relationships with pollinators [100].

Looking at larger evolutionary time scales, reproduction of three major co-evolutionary phylogenetic patterns in the fig-fig wasp system are observed: robustness of the 1-1 co-evolutionary relations between plant and pollinator, host-switching of pollinator to another fig type, and extensive hybridization among figs leading to violations in the 1-1 relationship between fig and fig-wasp. Initial plant type distributions and magnitude of preference bias, ϵ, will determine which of these scanerios is observed. A study of fig and fig wasp species pairs in the Ogasawara Islands revealed that *F. nishimurae* and the Higashidaira type selected their own host fig odors significantly more often, while wasps of *F. boninsimae* did not show specific preference for a particular fig odor [103]. This preference bias parameter does vary *in situ* and may explain some major variations in evolutionary trajectories in different fig-fig wasp systems.

All parameters can also change outcomes of the model, primarily by moving the trajectory that divides the basin of attraction between the fixed point solutions, and by affecting the time to reach equilibrium. This differs from the H-GM model in that the H-GM model only ϵ and β could affect the basin of attraction.

3.4 Conclusion

Asymptotic behavior in the F-FW model remains unchanged. There is no maintenance of genetic variation (except for the cases of $\epsilon = 0$ or $\epsilon = 1$), but there is the ability to host-switch. However, the F-FW model differs markedly in results from the H-GM model in transient behavior. Genetic variation in the short term can be maintained much longer in the F-FW model with up to tens of thousands of insect generations of incomplete host-switch observed due to the asynchronous flowering in plants delaying fixation of only one type. The effect of incomplete and complete host-switch on the coevolution of phylogenetic trees is considered in the next section.

Chapter 4

Effect of Mutualisms on Coevolving Phylogenies

4.1 Introduction

Until now, the focus has been a 2 plant types possible in one or two population. This allowed the exploration of a the relative likelihood that host-switching, maintenance of genetic variation, or maintenance of the specialized relationship. It was established that reciprocal evolution will always encourage traits to match, and small asymmetries in relative plant and pollinator abundances will cause strict fixation of only one type of each in the system. Therefore, maintenance of multiple types is unlikely to persist unless other mechanisms are present. For example, opposing forces exerted by multiple pollinators, plant partners as seen in the fig-fig wasp system, or a genetic basis that allows for many phenotypes, like a quantitative trait model. It is possible that understanding the contribution of mutualistic relationships to patterns of biodiversity lays within exploring the entire mutualistic network [7, 26, 92], more specifically a quantitative trait-driven mutualistic network based on both evolution and ecology [7, 18, 47, 83].

4.2   Ecological Network Theory

As ecology is the study of community processes, network theory gives the quantitative tools for describing the community interactions [76]. Ecological network theory has focused for decades primarily on food webs, experiencing an exponential increase in publications over the past four decades. Studies of mutualistic webs and antagonistic webs have been slower to emerge, but are now experiencing similar trends over the past two decades [47]. Jordano's 1987 paper [52] was the seminal paper on mutualistic networks which led this field of study. Since then, mutualistic interactions have been suggested as driving the "architecture of biodiversity" [7].

Community-level studies of mutualisms are on the rise giving insight into ecological complexity, but this has also strengthened the argument that more work needs to be done [7, 47]. A review of network stability concluded plant-pollinator networks appeared the most fragile of any type of ecological network[80]. This network perspective has added insight to the nature of the structure of mutualistic webs, but has also added a new tool for ecosystem restoration [8, 28, 53, 80, 95].

Due to the increase in field studies and data collection, mathematical network theory is also advancing, now addressing questions particular to the structure of bipartite networks, *eg.* [39, 63, 91]. Bipartite, or two-mode, networks are communities which are divided into two sets of nodes and relationships are only possible between them. A recent review of advances and insights into mutualistic networks is now presented.

4.2.1   Properties of Mutualistic Networks

Although mutualistic webs encompass several functional types, such as plant-pollinator and plant-frugivore webs, several climates, of various sizes, several commonalities in network structure persist [7, 94]. A recent review of mutualistic networks by Vazquez [94] noted seven general properties. I will use a similar

framework below to introduce mutualistic network theory while expanding on observations by Vazquez.

Low connectance: Only a few of the potential interspecific interactions occur.

From the ecological perspective, one might consider the example of body size [18]. Interaction predictions based on body-size similarity alone might produce an over-prediction of interactions, *e.g.* every like-sized insect pollinator interacting with the exact same plants. On the other hand these apparent "missing links" may actually be a result of biological constraints to matching [20, 47, 75]. Constraints like this might occur as a result of coevolving with other organisms throughout the species history [7].

From a network perspective, consider each animal species as a set of nodes and each plant species as a set of nodes. Then define an interaction matrix, A, with each entry A_{ij} as either 1 or 0 depending on whether or not there occurs an interaction between animal species, i, and plant species, j. Then the matrix of real mutualistic systems is much more sparse, *i.e.* more zeros, than would be expected by trait-matching alone. There have been some attempts at ascertaining the role of evolution in determining links. For example Rezende *et al.* reconstructed the phylogenies behind several plant-pollinator networks and found that in many cases relatedness was a predictor of interaction [83], and termed this phylogenetic signal.

Connectance is a way of measuring how sparse this interaction matrix is as a measure of the total number of links relative to the total connections possible in a network, $C = \frac{\sum\sum A_{i,j}}{m_J * n_I}$. A review by Jordano [52] of several mutualistic webs indicated that connectance is on average around 30%, with higher connectance in smaller networks. Also it is conjectured that connectance must decrease as the size of the network increases to maintain stability [9, 47, 67, 68]. Robert May suggested that complex ecological networks are stable if $i(SC)^{1/2} < 1$, where S is the total number of species and i is the mean interaction strength.

Node size asymmetry: An imbalance in the number of species of plants and animals in the network.

Meta-analyses consistently find more pollinators observed than plants, though the difference in species number, or relative number of nodes for plants versus pollinators seems to vary depending on the type of mutualism. The ratio of plant to animal nodes ranged from 1:2 to 1:6 with plant-pollinators webs at 1:4 [12, 41, 94]. Sampling effects, however, should not be ignored [76, 94]. Since most plant-pollinator studies rely on the observations of plant visitors, this may give us a skewed view of the node number. For example, it is possible in observing 2 plant species that 5 pollinator species visit. However, this is a sampling of a piece of the network, and may not mean that the entire network has the ratio 2:5. In fact, those 5 pollinator species, if one is a generalist, may be the same species observed if a 3rd plant were included, and in this case, the ratio would be 3:5. This may also skew connectance levels [47]. Research that explores the properties of mutualistic networks of varying sizes and random network sampling may be useful in determining how sampling effects may skew observations.

Heterogenous: Most species have few links, few have many links.

Ecologically, this means a few generalists and several specialists. In network theory terms, the binary connectivity matrix, A, when summed by row or column indicate the number of partners or degree, k, a particular animal or plant species has. In networks with more specialists and a few generalists, these degree distributions to be highly skew left, but with a long tail, [53]. The term heterogeneous describes networks in which most species have few interactions, and a few have more interactions than would be expected by chance [2, 6, 7]. Jordano also used the term heterogeneous in his 1987 paper to more specifically describe that the variance in degree was higher than the average degree [52].

Many real networks other than mutualistic webs seem to share this property of heterogeneity and are referred to as scale-free. Scale-free means that the cumulative degree distribution probability function, $P(k)$, follows a power-law (*i.e.* $P(k) \propto e^k$). Preferential attachment is the current mathematical explanation for the mechanisms behind these distributions [5, 81, 87]. In this explanation, new nodes at each time step interact with other nodes preferentially with respect to the degree. This is affectionately known as the "rich-get-richer" process. However, it is important to understand what biological mechanisms could drive a preferential attachment-like behavior in mutualistic networks.

The structure of the network, such as the larger number of animal species versus plant species, may also promote a natural skew for plant species to have more generalists than expected by chance. Instead of a power-law connectivity, a truncated power-law (*i.e.* $P(k) \propto k^\gamma e^{k/kc}$) has also been suggested as an alternative best-fit. Jordano examined 29 plant-pollinator networks and 24 plant-frugivore networks and determined that the best fit for pollinators was either truncated power-law or power-law, but for pollinated plants it was truncated power-law. Here the power-law exponent was on average 1.23 for pollinators and .84 for plants [53]. In plant-seed networks the best-fit was a truncated power-law. It is possible that a preferential attachment-like process in combination with size asymmetry could produce the truncated effect on the power-law distribution.

Unfortunately, even some of these data sets may also be influenced by inconsistencies in sampling processes. A popular complaint is that sampling may be influenced by abundance [88, 89]. In this case, the worry is that some low-abundance pollinators may be mistaken for specialist pollinators because they are only observed on one plant by chance. This can be extended to low in relative visitation frequency, not just as a result of low population size, but including variations in visitation due to flowering times, emergence times of pollinators or the availability of other more preferred partners. In fact, longer-term community studies which explicitly control for such sampling effects, found high "turnover" in plant-animal mutualistic interactions.

This means visitations changed depending on various conditions, including seasonality and flowering time [76]. Simulations studies by Morales [72] suggest that spatial effects of searching and interactions may affect also decreasing connectance. Therefore observations for a small time period might conclude specialist designation when in fact, that pollinator may visit a few different plants throughout the entire year.

In addition, one should also consider the grouping of nodes in a study before making comparisons between networks. One extensive community-level study by Inoue, *et al.* [48] used in a review by Bascompte and Jordano [7] to illustrate the power-law connectivity distribution was not data taken at the species association level, but rather at the family or sub-family level. It is important to understand how this might affect the extremely high specificity suggested by some studies **??**.

Weak and asymmetric dependences: Most links are weak, few are strong. Strength in a link in one direction of an interaction is often coupled by a weak link in the opposing direction.

Mutualism strength or the dependence of a plant species on an animal species, d_{ij}^P is (estimated as) the relative frequency of visits to plant species i by the frugivore or pollinator species j. Likewise, d_{ji}^A, the dependence of an animal species on a plant species, is (estimated as) the relative frequency of visits pollinator j makes to a particular plant species i [9, 52]. The distributions of these dependences are highly skew right in mutualistic communities [9], meaning that most interactions are weak. In food webs, weak interactions are recognized for their ability to buffer perturbations through the entire community contributing stability [7, 9, 47, 67].

In addition to many weak links and few strong links, strength in one direction of an interaction is often coupled by weak dependence in the other direction. This is called dependence asymmetry [9, 88, 95]. For example, ants are more dependent on the plants than the reverse, but plants tend to be more dependent on pollinator or frugivores than vice versa [27]. Even though the directionality of asymmetry is not always the same, the property of dependence asymmetry is universal and likely

results from differential fitness benefits between mutualistic partners [27]. Bascompte *et al.* suggested based on a population dynamics model, that communities are stable if the product of mutual dependences is less than the products of the average intraspecific competition coefficients divided by the product of the number of animal and plant species [9]. This means that to stability (especially in large networks) is best maintained by minimizing the product of mutual dependences, $d_{ij}^P d_{ji}^A$, which is to make them asymmetrical. Thus stable communities will have dependence asymmetry.

One measure of this asymmetry is as follows:

$$AS(i,j) = \|d_{ij}^P - d_{ji}^A\|/max(d_{ij}^P, d_{ji}^A)[9].$$

Note that if plants and animals have similar strength in dependence in both directions, then AS will be near 0, but if they are highly asymmetric, then AS will be near 1. The skew left of these distributions suggests high asymmetry in dependence, and further simulation results of null models suggests that these asymmetry index distributions in nature are a direct result of the skewed distribution of dependence values [9].

Nested: specialists interact with subsets of the species with which generalists interact.

Bascompte *et al.*, 2003, described the nestedness of mutualistic networks with the following analogy: "if we rank plants from the most specialized to the least specialized, we find that the set of animals a plant interacts with are contained in a larger set, which in turn is contained in a larger set, and so on, as in nested Chinese boxes," [8].

There are several theories as to why one should observe nestedness in mutualistic networks, and in fact up to 95% of all real world networks exhibit this feature [86]. For example, it is possible that nestedness is merely a sampling artifact of relative abundance. Null models have shown that if one observes many species, it is more likely to observe the rare event of a less abundant species [47, 62]. At the very least, it is reasonable to assume that sampling and relative abundance may contribute

to patterns, but other mechanisms may also be at play. Several have suggested that weak and asymmetric link distributions result in nestedness as generalists cores are connected weakly to many species which in turn are connected asymmetrically strong (*i.e.* specialists), *eg.* [6, 7, 88]. Ings, *et al.* in their review paper in 2009, suggested determining the mechanisms for nestedness was an open question for theoreticists [47], but it seems reasonable to assume that that the mechanism that produces many weak and asymmetric links may also produce nestedness by default. Medan *et al.*, 2007, showed that the isocline of perfect nestedness and cumulative degree distributions were fundamentally related, geometrically through derivates of the degree distributions, and both approached a truncated-power law for perfect nested bipartite networks [69].

A nested system describes a system with stability and cohesiveness. The generalist core interacting with a large number of specialists means that these core groups have small degrees of separation and specialists may not have to put all their eggs in one basket, but could adapt if there were an extinction event [7, 74]. Also generalists, pulled by many different pollinator needs are less likely to experience large shifts in phenotype or population size changes that would discourage specialists, assisting in the persistence of specialists [7, 8, 35]. In fact analysis has suggested that generalists also can have weak links to each other giving further cohesiveness to the network [35]. The latter observation has ecosystem management implications as rare or specialist species may need management of also its more common or generalist partners [7, 34]. But it also is presented by Bascompte *et al.* as a result of nestedness may actually hold a clue for determining a mechanistic explanation for the emergence of nestedness: a result of the stability in relative phenotype or abundance of generalist core groups. Ecological theory, statistical meta-analysis, and network theory support the notion that if a mechanism produces a cumulative degree distribution that is truncated-power law, it likely also produce nestedness [7, 35, 69, 75].

Modular or compartmentalized: contain subgroups of species that interact amongst themselves mostly and with others much less.

The example above of a generalist core group and its specialists is also an example of modularity in mutualistic networks. "A module is a group of species that interact strongly among themselves, but very little with species belonging to other modules," [6]. Detection algorithms for modularity have improved dramatically and have allowed meta-analysis of modularity in real networks [47, 63, 73].

Trait complementarity and convergence could explain the presences of modularity in mutualistic networks, *eg.* [7, 27, 53, 62, 92]. Furthermore, Rezende, *et al.* found that phenotypic complementarity in combination with similar evolutionary history produced modularity and nestedness. They argue phenotype in combination with evolutionary history may play a large role in determining the structure of mutualistic networks and should be integrated into a single, more broad framework [82]. In essence, the debate in the ecological network community about whether ecological interactions or simply species abundance and sampling play a role in creating observed interaction patterns should not be considered mutually exclusive, but should be expanded as pieces of a complex puzzle of factors including evolutionary history [7, 9, 18, 47, 82]. An essay series in PLoS called "Highlighting fundamental, unifying challenges in biology," included a piece called "Evolution, Interactions, and Biological Networks," extending this challenge outside the realm of mutalistic network theory and to biological network theory in general [101].

4.2.2 Mathematical Modeling of Mutualistic Networks

Modeling in food webs has long been the leading edge of ecological network theory [47], but interest in mutualistic network modeling has emerged as well. More community level interaction data has become available and these networks are being mined for patterns using emerging tools in network theory. At the same time there has been a simultaneous rising interest in community approaches to inform ecosystem

management and restoration [6, 80]. The scientific community is interested in the intrinsic structure of networks to anticipate effects due to species loss or climate change.

Heterogeneous networks, like mutualistic networks are robust to random node removal [2]. Recall this is a result of cohesive modules containing a generalist core and many weak and asymmetric connections. Fonseca and Bascompte [27] used existing real networks to overlay onto heterogeneous landscapes and manipulate the metapopulation landscape to look at network changes. They concluded that real networks are more robust than would be expected by chance. Memmott *et al.* simulated cextinctions on real pollinator-plant networks, concluding that those networks with truncated power-law distributions in connectivity were most robust [53, 70]. Simulated coextinctions on real pollination networks revealed that removing specialized species preferentially has little impact on the stability of the whole system until a substantial fraction of the nodes are removed [83]. A model that uses statistical properties to generate a Boolean framework for network assembly simulation concluded that assembly time to stable networks was relatively fast for mutualistic networks in comparison to food webs [17]. However, meta-analysis of ecological network stability through simulated extinctions discovered that while real ecological networks were robust, plant-pollinator webs were still the most fragile of the group [80].

The debate between abundance and geography versus trait complementarity as explanations for network features as well as the field's background in food web models has also driven current modeling studies. Assembly models have therefore focused on incorporating species abundance and geographic availability. Lockwood, 1997, tested how robust these assembled networks of Lotka-Volterra population dynamic assembled networks were to the invasion of these new nodes [64]. Santamaria and Rodríguez-Gironés, 2007, built a model based on trait matching and threshold levels to trait matching in an effort to reproduce "missing links" or the low connectances levels observed in mutualistic webs. Stang *et al.* [88, 89]. Meta-analysis suggests

64

flowering time dynamics are more important than phylogenetic signal in predicting interactions, though both were important [94].

Recall Albert and Barabasi determined preferential attachment mechanisms were responsible for power law-connectivity distributions [2]. Guimarães *et al.*, 2007, built a model based on this preferential attachment for bipartite networks. They concluded that for bipartite networks, this could explain both power-law distributions and truncated power-law distributions. The differential growth rate between node groups determined the exact distribution. If node group B grew slower than A, then $P(k_A)$ was best fit by power law. If B grew faster than A, then $P(k_B)$ was best-fit by a truncated power-law [41]. Thus preferential attachment can explain both of the connectivity distributions observed in mutualistic networks, though why preferential attachment should exist as an assembly mechanism is less clear [41, 53].

In an article in *Nature*, Saavedra *et al.*, [85], adopted food web assembly models and then applied this to business-consumer mutualistic networks. They also showed that it was better at producing desired degree distributions, nestedness, and modularity patterns than earlier models presented by Santamaria and Rodríguez-Gironés [86] and Guimarães *et al.* [40]. In this rule based stochastic model, partners were chosen randomly from a pool and assigned links based on an exponential probability of cooperation. This was found to be a function of current link number, the number of nodes in the group, and a reward. Another model meta-analysis including food web models and mutualistic models by Pires, 2011, concluded, like Rezende, 2007, that evolutionary history, more specifically, hierarchy may play a role in the structure of mutualistic webs [79, 82]. Rezende had simulated various hierarchal structures independently than based interaction matrices off of resulting traits and their complementarity. He found that matching hierarchal structures lead to nestedness of mutualistic networks.

To date, mutualistic models which incorporate evolution by looking at the network assembly process are lacking in the realism of their genetic basis (if any is included at all). Statistical mechanics models which employ hundreds of species interacting

with either have yet to incorporate bipartite structure [19, 19]. Some have purposely excluded a genetic basis to create null models based solely on maximizing entropy [102], which is an observation in some ant and pollinator mutualistic systems [12]. Other assembly models assume some random probability of a trait being affected by its partners [40], and no standard evolutionary notion of fitness has been incorporated.

However, the addition of trait coevolution in the assembly process may complement some of this earlier work in null modeling based on observations of statistical mining. For example, the equilibrium state which maximizes the discrete Shannon entropy, $H = -\sum p_i \ln p_i$ [29], means that the pressures, p_i, (inverse of the number of links [102] or interaction frequency [12]) either are very weak to minimize p_i or very strong to minimize $\ln p_i$, *i.e.* heterogeneous networks are stable. Ecologically speaking, pollinators seek to maximize entropy by forming network structures that promote more information reliability, *i.e.* less likelihood of strong perturbations. This means complexes of generalists that have pressures from many directions have phenotypes that remain approximately fixed, and specialists adhere themselves to these complexes so that they are less likely to lose their partners for which they are so specialized. Perhaps this assembly process indicates an evolutionary mechanism for the formation of heterogeneous networks. Despite many calls for ecological and evolutionary history to be integrated into the field of ecological network theory, relatively little has been done [6, 47, 80, 94].

4.2.3 Cophylogeny: Coevolution and Phylogeny

One advantage of incorporating evolution into network assembly models is that one can explore open questions in cophylogeny, phylogenies and the connections between coevolving groups. The expectation is that matching phylogenies are the rule in cophylogenies and other patterns are the exception. The nature of the coevolutionary relationship between figs and fig wasps is not as clear as overlaying congruent phylogenies. Not only are there violations to the one-to-one rule between

phylogenies, but also presence of hybridization that confuses the resolution of the phylogenies as well as rampant host-switching [49, 100]. However, analysis on real mutualistic networks indicates that many have a positive phylogenetic signal [83].

Current phylogenetic tree reconstruction efforts between coevolving partners often places penalties on events such as host-switching when resolving these phylogenies [77]. Little is known about how often these events are expected to happen. Verbal models are careful to not assume perfect matching of traits or phylogenies in coevolutionary systems and acknowledge there may be other events happening on a larger geographic scale that influence local examinations [49, 50, 93, 100]. A recent paper attempted to merge evolution with ecological interaction by simulating independently simulated phylogenies then creating associations based on character matching [83]. This null model investigation concluded that the amount of phylogenetic hierarchy, to what degree phylogenetic trees experience speciation within a main branch versus in many branches, could play a role in determining the observed nestedness of connections in mutualistic webs. However, this approach added little to the understanding of how continuous interaction might shape coevolving trees and ecological networks. The results of coevolving phylogenies on connectivity patterns, timing of speciation events, frequency of apparent host switches, etc. is a territory that remains to be explored [47].

Next, stochastic simulation model is developed to understand how mutualistic interactions shape cophylogeny and mutualistic networks. Plant-pollinator systems are used to compare and explain results, but the model is general enough to apply to other mutualistic webs with little modification. These resulting bipartite coevolving networks may provide insight into the formation of observed ecological network patterns and cophylogeny patterns as a result of ecological interactions.

4.3   Model Description

4.3.1   Coevolution of mutualism trait

A model is developed for the coevolution of a trait involved in a mutualistic interaction. Then a stochastic model is introduced which will couple the deterministic trait coevolution model with stochastic processes like speciation and extinction and will allow for evolution of networks based on historical associations and the fitness of each connection.

Let each mutualism characteristic be a 1-D quantitative trait. An example in the fig-fig wasp systems is the style length of the fig population and ovipositor length of the fig wasp population [47, 54, 99]. Each initial plant and pollinator has a starting mean trait value of 0 (this could mean some distance from a reference length like 10 mm). The trait value for each species, i, in plants is denoted x_i at an arbitrary generation and x_i' in the next generation. The trait value for each species, j, in pollinators is denoted y_j at an arbitrary generation and y_j' in the next generation. The fitness depends on the relative frequency of visits between any two coevolutionary partners as it affects pollination rate and food gathering. So fitness is a function of encounter rate and the visitation preference based on matching of traits [7, 18, 58, 89].

4.3.2   KLS Model

A model of the coevolution of a quantitative trait between one plant and one pollinator was done by Kiester, *et al.,* [58] and is presented in this section to lay the foundation for the models that follow. The KLS model was intended for one plant and one pollinator population only. After this section, an extension of this model is introduced that will incorporate an entire set of available partners with which to interact.

Plant model

Let $\Psi(x|y)$ be the relative preference for a plant with phenotype x by a pollinator with phenotype y. Given two interacting populations i and j, the probability of a visit to plant x_i given pollinator y_j is determined by,

$$p_i^*(x_i|y_j) = \frac{p_i(x_i)\Psi(x_i|y_j)}{\int p_i(x)\Psi(x|y_j)dx}. \tag{4.1}$$

Define $\Psi^*(x_i|y_j)$ so that $p^*(x_i|y_j) = \Psi^*(x_i|y_j)p(x_i)$. Then Ψ^* can be interpreted as the relative contribution of y_j pollinator to the x_i plant population, discussed above:

$$\Psi_i^*(x_i|y_j) = \Psi(x_i|y_j)/\int p_i(x_i)\Psi(x_i|y_j)dx_i. \tag{4.2}$$

Assume an infinite population or that all pollinator types will visit the same number of plants so that contribution to the next generation only depends on plant type frequency and not the number or density of plant type. Then, the total relative fitness for a particular plant population x_i depends on the contribution of each pollinator type multiplied by the relative abundance of pollinator types.

$$w_x(x) = \int \Psi_i^*(x|y)p_y(y)dy. \tag{4.3}$$

Then the mean phenotypic value of population i in the next generation is

$$\overline{x}' = \frac{1}{\overline{w}_x}\int xp_x(x)w_x(x)dx, \tag{4.4}$$

where $\overline{w}_x = \int p_x(x)w_x(x)dx = 1$ is the average fitness of plant population i.

Pollinator model

The derivation of the KLS pollinator model is not a mirror of the plant derivation. Instead, Kiester, Lande and Schemske posit that the "relative fitnesses of pollinator phenotypes are proportional to the total frequency of plant that they visit," [58],

69

thereby producing a fitness equation of

$$w_y(y) = \int p_x(x)p_x^*(x|y)dx. \tag{4.5}$$

This formulation of the fitness equation is awkward because the probability of a visit to a plant with trait value x by a given pollinator of trait value y is already defined as $p_x^*(x|y)$, in which the plant trait distribution is already accounted for. However, visits here are a probability distribution so using $p_x^*(x|y)$ alone would result in equal fitness values across pollinator phenotypes if the additional $p_x(x)$ was not inserted.

As in plants, the pollinator phenotype in the next generation is calculated by

$$y' = \frac{1}{\overline{w}_y} \int y p_y(y) w_y(y) dy. \tag{4.6}$$

Analysis

Selection differentials are used to describe change in mean trait value, or $S_x = \overline{x}' - \overline{x}$ and $S_y = \overline{y}' - \overline{y}$. For specialist pollinators an absolute preference function for Ψ is employed. This assumes pollinators prefer a certain trait value, regardless of its distribution in the population is defined as

$$\Psi(x|y) = e^{-(x-y)^2/2v^2}. \tag{4.7}$$

This form for a matching function was used in Kiester *et al.* [58], but the need for matching between corolla length and fruit cross width in many systems was also noted by Jordano's 1987 review on pollinator networks [52].

Plants and pollinators are assumed distributed normally with mean $\overline{x}$ and $\overline{y}$, respectively, and with variance σ_x^2 and σ_y^2, respectively. In this case, the selection

differential derived for plants can be described by the following equation,

$$\frac{S_x}{\sigma_x^2} = \frac{\bar{y} - \bar{x}}{v^2 + \sigma_x^2} \tag{4.8a}$$

$$\frac{S_y}{\sigma_y} = \frac{\bar{x} - \bar{y}}{\sigma_x^2 + \left(v^2 + \sigma_x^2\right)\left(2v^2 + \sigma_x^2\right)/\sigma_x^2}. \tag{4.8b}$$

This is equivalent to

$$\bar{x}' = (1 - \alpha)\bar{x} + \alpha\bar{y} \tag{4.9a}$$

$$\bar{y}' = (1 - \beta)\bar{y} + \beta\bar{x}, \tag{4.9b}$$

where $\alpha = \sigma_x^2/\left(v^2 + \sigma_x^2\right)$ and $\beta = \sigma_x^2\sigma_y^2/\left(\sigma_x^2\sigma_y^2 + \left(v^2 + \sigma_x^2\right)\left(2v^2 + \sigma_x^2\right)\right)$.

Note: The line of equilibrium at $\bar{x} = \bar{y}$ is stable. Kiester *et al.* note in their model between one plant and one pollinator that this stable line of equilibria could mean that geographically isolated populations could evolve to different trait values, thus providing a mechanism for allopatric speciation [58]. However, the total trait distribution of all possible interaction partners needs to be considered in networks. Developing these interaction probabilities as this has been recognized as important in pollination networks [89], but is not addressed in models.

4.3.3 Kiester, Lande and Schemske Extension and Symmetric Fitness Modification

The equations for the mean trait value in the next generation are derived following Kiester *et al.* and extends the 1-1 plant-pollinator case to m plant and n pollinators interacting. Consider plant species i pollinator species j. The interaction or connectivity matrix, A, is an $m \times n$ matrix such that $A(i, j) = 1$ if population i and j interact and $A(i, j) = 0$ if they do not. J is the set of all plant species that interact with pollinator species j and is of size m_J. Likewise, I is the set of all

pollinator populations that interact with plant species i and is of size n_I. Note that $0 \le m_J \le m$ and $0 \le n_I \le n$. Also note $A(i,j) = 1$ if and only if $i \in J$ or $j \in I$.

Define $p_i(x)$ as the trait probability distribution function of the ith population of plants. The probability of picking an individual with phenotype x from all interactors of j, J, is determined by

$$p_J(x) = \frac{1}{m_J} \sum_{i \in J} p_i(x). \tag{4.10}$$

Analogously, define $p_j(y)$ as the probability distribution function of the jth population of pollinators. The probability of picking an individual with phenotype y from all interactors of i, I, is determined by

$$p_I(y) = \frac{1}{n_I} \sum_{j \in I} p_j(y). \tag{4.11}$$

Refer to Table 4.1 for a list of all functions used in this model and the models below.

The probability of a visit to a plant of phenotype x from the ith population given phenotype y pollinator from population j and the set of its interactors J is determined by,

$$p_J^*(x_i|y_j) = \frac{p_J(x_i)\Psi(x_i|y_j)}{\int p_J(x)\Psi(x_i|y_j)dx_i}. \tag{4.12}$$

Define $\Psi^*(x_i|y_j)$ so that $p^*(x_i|y_j) = \Psi^*(x_i|y_j)p(x_i)$.

$$\Psi_J^*(x_i|y_j) = \Psi(x_i|y_j) / \int p_J(x_i)\Psi(x_i|y_j)dx_i. \tag{4.13}$$

Then Ψ^* can be interpreted as the relative contribution of y_j pollinator to the x_i plant population relative to all of the other interacting populations this partner services.

Plant model

Assume an infinite population or that all pollinator types will visit the same number of plants so that contribution to the next generation only depends on plant type

Table 4.1: List of Coevolving Phylogenies models' variables and functions.

Variables		
x		Trait value for mutualistic character in plant
y		Trait value for mutualistic character in insect
$\bar{x}_i$		Mean trait value for mutualistic character in plant population i
$\bar{y}_j$		Mean trait value for mutualistic character in insect population j
I		The set of all insect populations, $j \in I$, that interact with plant population i
J		The set of all plant populations, $i \in J$, that interact with insect population j
Probability Distribution Functions		
$p_i(x)$		Frequency of trait value x in plant population i
$p_j(y)$		Frequency of trait value y in insect population j
$p_J(x)$	$\dfrac{1}{m_J} \sum\limits_{i \in J} p_i(x)$	Frequency of trait value x among all plant interactors with insect j
$p_I(y)$	$\dfrac{1}{n_I} \sum\limits_{j \in I} p_j(y)$	Frequency of trait value y among all plant interactors with plant i
Functions		
$\Psi(x\|y)$		Relative preference for plant phenotype x given insect phenotype y
$p_i^*(x_i\|y_j)$	$\dfrac{p_i(x_i)\Psi(x_i\|y_j)}{m_J \int p_J(x_i)\Psi(x_i\|y_j)dx_i}$	Frequency of visits to a plant with phenotype x in population i given a pollinator y_j
$p_j^*(x_i\|y_j)$	$\dfrac{p_j(y_j)\Psi(x_i\|y_j)}{n_I \int p_I(y_j)\Psi(x_i\|y_j)dy_j}$	Frequency of visits to a plant with phenotype x in population i given a pollinator y_j
$p_J^*(x_i\|y_j)$	$\sum\limits_{i \in J} p*_i(x_i\|y_j)$	Frequency of visits by pollinator y_j given a plant x_i
$p_I^*(x_i\|y_j)$	$\sum\limits_{j \in I} p*_j(x_i\|y_j)$	Frequency of visits to a plant with phenotype x_i given a pollinator y_j
$\Psi_i^*(x\|y)$	$\dfrac{\Psi(x_i\|y_j)}{m_J \int p_J(x_i)\Psi(x_i\|y_j)dx_i}$	Relative contribution of x_i plant to the y_j pollinator)
$\Psi_j^*(x\|y)$	$\dfrac{\Psi(x_i\|y_j)}{n_I \int p_I(y)\Psi(x_i\|y_j)dy_j}$	Relative contribution of y_j pollinator to the x_i plant pollination visits)

frequency and not the number or density of plant type. Then, the total relative fitness for a particular phenotype in plant population i depends on the contribution of each pollinator type multiplied by the relative abundance of pollinator types.

$$w_i(x_i) = \int \Psi_i^*(x_i|y_j)p_I(y_j)dy_j. \tag{4.14}$$

Then the mean phenotypic value of population i in the next generation is

$$\overline{x}_i' = \frac{1}{\overline{w}_i} \int x_i p_i(x_i)w_i(x_i)dx_i, \tag{4.15}$$

where $\overline{w}_i = \int p_i(x_i)w_i(x_i)dx_i$ is the average fitness of plant population i.

Pollinator Model

The pollinator's fitness depends on the relative frequency of plant visits it makes. Three variations of this fitness function are explored. The first two are considered below and the third is a stochastic model considered in a later section. The first version is a direct extension of the model developed in Kiester, Lande, and Schemske [58] and so is referred to as the KLS Extension:

$$w_j(y_j) = \int p_J^*(x_i|y_j)p_J(x_i)dx_i. \tag{4.16}$$

The second definition for the fitness, as done in Chapter 3 for the explicit genetic model, calculates the appropriately scaled conditional probability on the choice of plant by any particular insect. This model, will be referred to as the Symmetric Fitness Coevolver (SFC) Model since the pollinator fitness calculation is symmetric to that of the plant fitness.

$$w_j(y_j) = \int \Psi_j^*(x_i|y_j)p_J(x_i)dx_i. \tag{4.17}$$

In both cases, the mean of the pollinator population j in the next generation is

$$\bar{y}_j' = \frac{1}{\bar{w}_j} \int y_j p_j(x_j) w_j(y_j) dy_j, \tag{4.18}$$

where $\bar{w}_j = \int p_j(y_j) w_j(y_j) dy_j$ is the average fitness of pollinator population j.

4.4 Analysis

Assume the following relative preference for a plant with phenotype x by a pollinator with phenotype y.

$$\Psi(x|y) = e^{\frac{-(x-y)^2}{2v^2}}. \tag{4.19}$$

v can be interpreted as a tolerance for matching of traits. Assume that traits in each population are normally distributed with the constant variance, so

$$p_i(x_i) = e^{\frac{-(x_i-\bar{x}_i)^2}{2\sigma_x^2}}, \tag{4.20a}$$

$$p_j(y_j) = e^{\frac{-\left(y_j-\bar{y}_j\right)^2}{2\sigma_y^2}}. \tag{4.20b}$$

Approximate the probability distribution of a group of interactors by a normal distribution with a mean that is the average of the group's means.

$$p_J(x_i) = e^{\frac{-(x_i-\bar{x}_J)^2}{2\sigma_J^2}}, \tag{4.21a}$$

$$p_I(y_j) = e^{\frac{-\left(y_j-\bar{y}_I\right)^2}{2\sigma_I^2}}, \tag{4.21b}$$

where

$$\bar{x}_J = \frac{1}{n_J} \sum \bar{x}_i = \frac{A_{i,j} y_{n,j}}{\sum_{j=1}^{N_x} A_{i,j}}, \tag{4.22a}$$

$$\bar{y}_I = \frac{1}{m_I} \sum \bar{y}_I = \frac{A_{i,j} x_{n,i}}{\sum_{i=1}^{N_y} A_{i,j}}. \tag{4.22b}$$

Note that populations in a particular interactor group have similar mean to that of its coevolutionary partner.

4.4.1 KLS Extension

Under the assumptions above, and assuming that $p_J(x)$ is normal with mean $\overline{x}_J$ and standard deviation σ_x and $p_J(x)$ is normal with mean $\overline{y}_I$ and standard deviation σ_y, equivalent equations from Kiester, Lande and Schemeske should and do emerge [58]. The only difference is the mean of all partners is replaces the mean of one partner.

$$\overline{x}'_i = \frac{v^2\overline{x}_i + \sigma_x^2\overline{y}_I}{v^2 + \sigma_x^2} = (1 - \alpha)\,\overline{x}_i + \alpha\overline{y}_I, \tag{4.23a}$$

$$\overline{y}'_j = \frac{v^2\overline{y}_j + \sigma_y^2\overline{x}_J}{v^2 + \sigma_y^2} = (1 - \beta)\,\overline{y}_j + \beta\overline{x}_J, \tag{4.23b}$$

where $\alpha = \sigma_x^2/\left(v^2 + \sigma_x^2\right)$ and $\beta = \sigma_x^2\sigma_y^2/\left(\sigma_x^2\sigma_y^2 + \left(v^2 + \sigma_x^2\right)\left(2v^2 + \sigma_x^2\right)\right)$.

The α and β terms determine how important selection for trait matching are for the plant and pollinator, respectively. If they are close to 0, then the trait matching with its mutualistic partners applies less selective force, and so the next generation will have trait value likely more similar to its parents. Note that when $\sigma_x^2 > \sigma_y^2$, $\alpha > \beta$. In this case, plants responds less quickly to insect evolution.

4.4.2 SFC Model

In the Kiester, Lande, and Schemske paper, they use the same conditional preference function, and try to modify the form to fit both uses. The Symmetric Fitness Coevolver modification requires that each species has the same functional effect on each others evolution.

Under these new assumptions, and assuming that $p_J(x)$ is normal with mean $\overline{x}_J$ and standard deviation σ_x and $p_J(x)$ is normal with mean $\overline{y}_I$ and standard deviation σ_y, the result is the same dynamic fitness equation as in the KLS Extension (Equations

4.9a and 4.9b), but with more symmetry in the α and β parameters. The parameters are now $\alpha = \sigma_x^2/\left(v^2 + \sigma_x^2\right)$ and $\beta = \sigma_y^2/\left(v^2 + \sigma_y^2\right)$.

4.4.3   Equilibria

For a particular set of interactions between plants and pollinators, described by interaction matrix, A, the trait values at equilibria for both the KLS Extension and the SFM are described by

$$\bar{x}_i = \bar{y}_I, \ \text{for } i = 1, .., m, \tag{4.24a}$$

$$\bar{y}_j = \bar{x}_J, \ \text{for } j = 1, .., n. \tag{4.24b}$$

If the entire network is a connected graph, then this implies that $\bar{x}_1 = ... = \bar{x}_m = \bar{y}_1 = ... = \bar{y}_n$, forming a line of equilibria. If the entire network is not connected, then mean trait values will converge within each disjoint connected subgraph, but independently between subgraphs. Kiester, et $al.$, [58], discusses how two or more geographically independent populations might allopatrically speciate as a result of independent evolution along a line of equilibria. This is extended to two or more disjoint subgraphs or metapopulations which vary independently due to high specificity in trait matching.

Theorem 4.4.1. *For the dynamical system in Equations 4.23a and 4.23b, the line of equilibria in Equations 4.24b in a connected network are stable.*

Proof. The dynamical system of equations in 4.23a and 4.23b can be rewritten in vector form using the interaction matrix, A where $\mathbf{n^{-1}}$ is an array with 1/the number of interactors each plant population i has and $\mathbf{m^{-1}}$ is 1/number of interactors that each pollinator population j has:

$$\bar{\mathbf{x}}' = (1 - \alpha)\bar{\mathbf{x}} + \alpha\left(A\bar{\mathbf{y}}\right) \bullet \mathbf{m^{-1}} \tag{4.25a}$$

$$\bar{\mathbf{y}}' = (1 - \beta)\bar{\mathbf{y}} + \beta\left(\bar{\mathbf{x}}A\right) \bullet \mathbf{n^{-1}}. \tag{4.25b}$$

Note that for both models $0 < \alpha, \beta < 1$. Since it is expected that the network is completely connected at this equilibria because all phenotypes are the same, we assume A is a $m \times n$ ones matrix. All entries of $\mathbf{m}^{-1} = 1/n$, because there are n pollinator populations and $\mathbf{n}^{-1} = 1/m$, because there are m plant populations. Therefore, the dynamical system is

$$\bar{\mathbf{x}}' = (1 - \alpha)\bar{\mathbf{x}} + \frac{\alpha}{n}\left(\bar{y}_1 + ... + \bar{y}_n\right) \tag{4.26a}$$

$$\bar{\mathbf{y}}' = (1 - \beta)\bar{\mathbf{y}} + \frac{\beta}{m}\left(\bar{x}_1 + ... + \bar{x}_m\right). \tag{4.26b}$$

This is a linear difference equation system, $\mathbf{z}' = L\mathbf{z}$ where L can be described as a block matrix:

$$\begin{bmatrix} (1 - \alpha)\, I_m & \frac{\alpha}{n} N_{m \times n} \\ \\ \frac{\beta}{m} N_{n \times m} & (1 - \beta)\, I_n \end{bmatrix},$$

where $L2$ and $L3$ are some constant coefficient interaction matrix with dimensions indicated by the subscripts.

To find the eigenvalues, we solve $\det(L - \lambda I_{n+m}) = 0$. Note that the left hand side is also a block matrix:

$$\begin{bmatrix} (1 - \alpha - \lambda)\, I_m & \frac{\alpha}{n} N_{m \times n} \\ \\ \frac{\beta}{m} N_{n \times m} & (1 - \beta - \lambda)\, I_n \end{bmatrix},$$

We can use the Schur complement of the block matrix to rewrite the determinant of this block matrix as either form below:

$$\det\left(L - \lambda I_{n+m}\right) = \det\left((1 - \alpha - \lambda)\,I_m\right)\det\left((1 - \beta - \lambda)\,I_n - \frac{\beta}{m}N_{n\times m}\frac{I_n}{1 - \alpha - \lambda}\frac{\alpha}{n}N_{m\times n}\right),$$

(4.27a)

$$\det\left(L - \lambda I_{n+m}\right) = \det\left((1 - \beta - \lambda)\,I_n\right)\det\left((1 - \alpha - \lambda)\,I_m - \frac{\alpha}{n}N_{m\times n}\frac{I_m}{1 - \beta - \lambda}\frac{\beta}{m}N_{n\times m}\right).$$

(4.27b)

Two cases are considered, $m \geq n$ and $m < n$, respectively. Each case is the same proof, but with different forms of the Schur complement. Using the first form, we can rewrite the Schur complement as

$$(1 - \alpha - \lambda)^{m-n}\det\left((1 - \alpha - \lambda)(1 - \beta - \lambda)\,I_n - \frac{\alpha\beta}{n}N_{n\times n}\right).$$

(4.28)

Note that the eigenvalue $1 - \alpha$ has multiplicity $m - n$. Solving the second determinant in the Schur complement for $= 0$ reveals the other $2n$ eigenvalues. This determinant can be rewitten in the form

$$
\begin{vmatrix}
(1 - \alpha - \lambda)(1 - \beta - \lambda) - \alpha\beta & ((1 - \alpha - \lambda)(1 - \beta - \lambda) - \alpha\beta)\,N_{1\times n-1} \\[2ex]
-(1 - \alpha - \lambda)(1 - \beta - \lambda)\,N_{n-1\times 1} & (1 - \alpha - \lambda)(1 - \beta - \lambda)\,I_{n-1}
\end{vmatrix}
$$
$$
= ((1 - \alpha - \lambda)(1 - \beta - \lambda) - \alpha\beta)
\begin{vmatrix}
1 & N_{1\times n-1} \\[2ex]
-(1 - \alpha - \lambda)(1 - \beta - \lambda)\,N_{n-1\times 1} & (1 - \alpha - \lambda)(1 - \beta - \lambda)\,I_{n-1}
\end{vmatrix},
$$

revealing another set of eigenvalues, at 1 and $1 - \alpha - \beta$.

The remaining determinant can be written as the sum of n determinants of size $n-1$:

$$\left|\left(1-\alpha-\lambda\right)\left(1-\beta-\lambda\right)I_{n-1}\right|$$

$$+\quad\left(1-\alpha-\lambda\right)\left(1-\beta-\lambda\right)$$

$$\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & \left(1-\alpha-\lambda\right)\left(1-\beta-\lambda\right) & 0 & \cdots & 0 \\ 0 & 0 & \left(1-\alpha-\lambda\right)\left(1-\beta-\lambda\right) & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \left(1-\alpha-\lambda\right)\left(1-\beta-\lambda\right) \end{vmatrix}$$

$$-\quad\left(1-\alpha-\lambda\right)\left(1-\beta-\lambda\right)$$

$$\begin{vmatrix} & 1 & 1 & 1 & \cdots & 1 \\ \left(1-\alpha-\lambda\right)\left(1-\beta-\lambda\right) & 0 & 0 & \cdots & 0 \\ & 0 & 0 & \left(1-\alpha-\lambda\right)\left(1-\beta-\lambda\right) & & \vdots \\ & \vdots & \vdots & & \ddots & 0 \\ & 0 & 0 & \cdots & 0 & \left(1-\alpha-\lambda\right)\left(1-\beta-\lambda\right) \end{vmatrix}$$

$$+\quad\ldots.$$

The latter $n-1$ determinants are all I_{n-2} with a column of zeros inserted in each column and row of ones at the top, and with alternating signs. This means these can

be rewritten:

$$
\begin{aligned}
& |(1 - \alpha - \lambda)(1 - \beta - \lambda) I_{n-1}| \\
& \quad + (1 - \alpha - \lambda)(1 - \beta - \lambda) \sum |(1 - \alpha - \lambda)(1 - \beta - \lambda) I_{n-2}| \\
& = \; n * (1 - \alpha - \lambda)^{n-1} (1 - \beta - \lambda)^{n-1}. \qquad\qquad\qquad \square
\end{aligned}
$$

Therefore, the other $2(n-1)$ eigenvalues are $1 - \alpha$ and $1 - \beta$ with multiplicity $n - 1$ each. In summary, the eigenvalues for the $m \geq n$ case are $\lambda = 1$, $1 - \alpha - \beta$, $1 - \alpha$ with multiplicity $m - 1$, and $1 - \beta$ with multiplicity $n - 1$. Through the same argument, but using the second form of the Schur complement, we find the same eigenvalues apply to the $m < n$ case as well. Because one eigenvalue is 1 and $|\lambda| < 1$ for all others, the line of equilibria is stable.

To illustrate how connected networks and disjoint networks differ in equilibrium solutions, consider the following cases.

Connected network example

Suppose there are 3 plants and 2 pollinators such that plant populations 1 and 2 are connected to pollinator population 1 and plant populations 2 and 3 are connected to pollinator population 2. The network graph looks like:
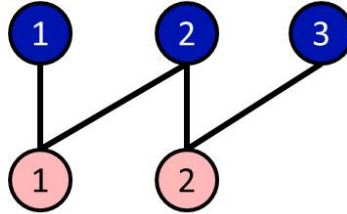
Figure 4.1: Graph of interacting plant pollinator populations in a connected network. The blue nodes are plant nodes and the red nodes are pollinator populations.

Then associated interaction matrix is a 3×2 matrix:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$$

The network equilibrium equations in Equation 4.24b are:

$$\bar{x}_1 = \bar{y}_1, \tag{4.29a}$$

$$\bar{x}_2 = \frac{\bar{y}_1 + \bar{y}_2}{2}, \tag{4.29b}$$

$$\bar{x}_3 = \bar{y}_2, \tag{4.29c}$$

$$\bar{y}_1 = \frac{\bar{x}_1 + \bar{x}_2}{2}, \tag{4.29d}$$

$$\bar{y}_2 = \frac{\bar{x}_1 + \bar{x}_2}{2}. \tag{4.29e}$$

This implies that $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = \bar{y}_1 = \bar{y}_2$. Therefore, without stochasticity, mean trait values of all plants and all pollinators converge to a single value.

Disjoint connected networks example
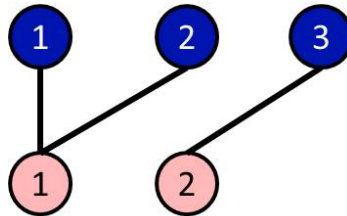
Figure 4.2: Graph of interacting plant pollinator populations in a network comprised of disjoint connected subgraphs.

Then associated interaction matrix is a 3×2 matrix:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The network equilibrium equations in Equation 4.24b are:

$$\bar{x}_1 = \bar{y}_1, \tag{4.30a}$$

$$\bar{x}_2 = \bar{y}_1, \tag{4.30b}$$

$$\bar{x}_3 = \bar{y}_2, \tag{4.30c}$$

$$\bar{y}_1 = \frac{\bar{x}_1 + \bar{x}_2}{2}, \tag{4.30d}$$

$$\bar{y}_2 = \bar{x}_3. \tag{4.30e}$$

This implies that $\bar{x}_1 = \bar{x}_2 = \bar{y}_1$ and $\bar{x}_3 = \bar{y}_2$. Therefore, without stochasticity, average trait values in each subgraph evolve to a single value. Since the subgraphs are disjoint, each can evolve to a different value. This extends the original results of Kiester *et al.* [58], but geographic isolation was considered necessary for divergent evolutionary drift to occur. Next, ecological network independence will be examined through a quantitative trait value mismatch such as ovule and ovipositor length or flowering time versus emergence time is sufficient.

The effects of choosiness

In the case of no preference or lack of choosiness, characterized by $v \to \infty$, and in the absence of stochasticity, we examine the resulting mean phenotypes in the next

generation in Equations 4.23a and 4.23b. In this case, $\alpha \to 0$ and $\beta \to 0$, so

$$\bar{x}'_i = \bar{y}_I \tag{4.31}$$

$$\bar{y}'_j = \bar{x}_J. \tag{4.32}$$

This means that in the absence of choosiness or preference, phenotypes will tend to evolve to match the mean phenotype of all of the cooperating partners. They will show no bias towards any particular partner more strongly than any other.

In the case of a highly choosy partner, where preference $v \to 0$, $\alpha \to 1$ and $\beta \to 1$. The resulting equations for the next generation phenotype in the absence of stochasticity are:

$$\bar{x}'_i = \bar{x}_i \tag{4.33a}$$

$$\bar{y}'_j = \bar{y}_j. \tag{4.33b}$$

Under a highly choosy system, the next generation phenotypes will be near the current phenotype, with little ability to evolve towards the phenotype of all available resources. In the absence of any stochasticity in maintaining connections, this might be fine, but as long as there is some probability in losing connection if partner phenotypes drift or evolve away, high choosiness could also result in network instability.

4.5   Stochastic Simulation Models

4.5.1   Connections Between Species

Interactions between plant population i and pollinator population j are tracked via a connectivity matrix, A. Mutualistic connections are maintained, created, and lost with some probability proportional to the new trait distance between all of the mutualistic candidate partners. Each entry in the connectivity matrix in the next

generation, A', is either 0 for not connected or 1 for connected.

$$P\left(A'\left(i,j\right)=1\right)=\left(1-c\right)m(i,j)+cA_{i,j} \tag{4.34}$$

where $m(i,j)$ is a probability function that describes how likely species i and j will interact in the absence of previous association for phenotype matching between plant and pollinator.

The parameter c affects how strongly having a current connection determines a connection in the next generation. For example, when partners are interacting, there maybe several traits involved in the interaction, and we are only measuring the matching of one. So a decrease in the matching level in one trait may not be enough reason to justify a full loss of association. If $c=0$, then historical associations do not matter. Under this scenario, even if $m(i,j)=.9998$, the probability that they would maintain their connection for 5000 generations is only about 0.37. If $c=1$, then historical connection matters only, and mean trait matching between a plant and pollinator population does not, so any association that is inherited after speciation continues forever. In this case, if the initial conditions of the simulation are a connected network of plants and pollinators, then only those connections will inherit and never be lost. This results in a well connected network after any period of time. If initial conditions were a disjoint network, then it would remain a disjoint network due to the inability to create new connections other than what is inherited. It has been theorized that historical coevolution may be responsible for "missing links" observed in pollination webs [7].

If historical connection plays an intermediate role (i.e. c is some intermediate, between 0 and 1), then connections can persist for a long period of time, but can evolve based on how well beneficial the connection is. In our prior example, for $c=0$, the probability of persistence of a desirable connection between closely matching partners for 5000 generations was only 0.37. For $c=0.5$, this is increased to 0.67, and new connections are formed approximately 50% of the time that a partner of that

match is available. For an even higher relative importance of historical connection, $c = .9$, peristance probability in this example increased to approximately 0.90, but the probability of making a new connection with a similarly matched partner is decreased to 10%. A value of c near 1 may be an appropriately justifiable expectation if we think that associations with partners are based on multiple traits and/or that mutualistic webs are likely to persist in a constant environment.

4.5.2   Species Birth-death

Species birth is a random event for both plants and pollinators and is a random event with probability b_x and b_y, respectively. The lineage of parent species is copied as the evolutionary history of the new species and the trait value of the new species is the parental trait value plus some random variation from the parental. When a species birth happens, the column (or row) of lineage connections is duplicated and addended to the end of the connectivity matrix. Now the mean phenotypic value of each population can be modified independently by stochastic forces like drift.

Species death happens when a species cannot maintain its connections. If at any time m_J (or n_I) = 0, species j or i is removed from the connectivity matrix and the list of extant species. However, a delay of several generations for some systems may be an appropriate modification if loss of connections is frequent as discussed in the previous section. The biological reasoning is a plant species like fig would not go extinct after 1 generation without reproduction. Pollinator species would then be given a reasonable chance to (re-)establish the mutualistic relationship.

At each update, the average trait value of each species is derived from its current trait value and the average trait value of all of the species to which it is connected, reflecting selective pressures on the mutualistic trait. In addition, it experiences stochastic fluctuation, which reflects possible environmental stochasticity and drift. This fluctuation, ξ_x and ξ_y, is a normal random variable with mean 0 and standard deviation V_x and V_y, respectively. Multiple external and internal stochastic influences

act on the quantitative trait (*e. g.* environment and mutation), each with small effect and in random direction with respect the trait. Therefore, the sum of these effects on the quantitative trait is approximated by a normal distribution.

Recall that b_x and b_y are the probability of speciation/species birth in the plant population and the pollinator population, respectively. When speciation occurs, the new population has mean trait value equal to the parent population, plus some random fluctuation, ξ_s. ξ_s is a normal random variable with mean 0 and standard deviation V_s.

4.5.3   KLS Extension and Symmetric Fitness Model

Accounting for stochastic effects due to environmental factors or drift, we calculate the phenotype in the next generation as

$$\bar{x}'_i = (1 - \alpha)\, \bar{x}_i + \alpha \bar{y}_I + \xi_x, \tag{4.35a}$$

$$\bar{y}'_j = (1 - \beta)\, \bar{y}_j + \beta \bar{x}_J + \xi_y. \tag{4.35b}$$

To determine mean partner trait values, we now incorporate the evolving interaction matrix. For this SFC model, we use the absolute trait preference function used in [58] as the matching function which helps determine connectivity:

$$m(i, j) = e^{-(\bar{x}_{n+1,i} - \bar{y}_{n+1,j})^2 / 2\theta^2}, \tag{4.36}$$

where θ is some tolerance for phenotype matching between plant and pollinator.

4.5.4   Stochastic Asymmetric Coevolving Network Model

In previous analysis, a trait matching function determined the connectivity matrix. For ease of analysis, it was also assumed that the fitness of species is independent of the other species interacting with its partners. These assumptions have been used in food web construction models, though not for quantitative traits.

87

Fitness is now assumed to be dependent on other plants and pollinators and will also determine entries of the connectivity matrix. This is computationally more complex, but more accurately emulates how fitness affects the evolution of partner interactions. Plants and insects may desire partners that are not sharing their resources with too many other populations. Thus, it makes sense to consider how many suitors your partner has when trying to decide whether a particular relationship is worth pursuing [12].

In the new connectivity matrix, each entry in the connectivity matrix in the next generation, A', is either 0 for not connected or 1 for connected based on a probability determined by both historical connection and average contribution of the candidate plant population to the pollinator's fitness. This introduces asymmetry as the matching function is defined as the pollinator's perceived fitness, not the plants:

$$m(i,j) = w_j(i|j)/\bar{w}_j. \tag{4.37}$$

Recall the prior discussion about the ecological interpretation for the process of maximizing network entropy. The interaction matrix will evolve so that maximum resources are being obtained by insects. The question is to what extent will this result in a network that does this by balancing generalist relationships with many weak dependences with fragile specialist relationships that could be disrupted by the evolutionary pressures on traits by any other connections or external pressures.

Under these new assumptions, the following equations are for the next generation of plant population i,

$$\bar{x}'_i = (1-\alpha)\,\bar{y}_I + \alpha\bar{x}_i + \gamma_1\left(\bar{y}_I\frac{1}{n_I}\sum_{j\in I}\bar{x}_J\right), \tag{4.38a}$$

$$\bar{y}'_j = (1-\beta)\,\bar{x}_J + \beta\bar{y}_j + \gamma_2\left(\bar{x}_J - \frac{1}{m_J}\sum_{i\in J}\bar{y}_I\right), \tag{4.38b}$$

where $\alpha = \frac{\sigma_I^2 + \sigma_I v + v^2}{\sigma_I^2 + \sigma_I v + v^2 + \sigma_x v}$, $\gamma_1 = \frac{\sigma_I \sigma_x}{\sigma_I^2 + \sigma_I v + v^2 + \sigma_x v}$, $\beta = \frac{\sigma_J^2 + \sigma_J v + v^2}{\sigma_J^2 + \sigma_J v + v^2 + \sigma_y v}$ and $\gamma_2 = \frac{\sigma_J \sigma_y}{\sigma_J^2 + \sigma_J v + v^2 + \sigma_y v}$.

4.6 Numerical Results

The following are simulation results for the full stochastic network. Initial conditions, unless otherwise specified, are one plant and one pollinator species, each with trait value 0 and connected to each other. Unless otherwise noted, all simulations have parameter values, $T = 400$ (number of iterations), $b_x = b_y = 0.01$ (probability of speciation/species birth so number of generations is approximately $10^6 bT = 4$ million), $v^2 = 1$, $\theta_x^2 = \theta_y^2 = .1$, $V_x = V_y = .05$, and $V_s = .2$.

The stability of mutualistic networks increased as c, the relative importance of historical connection, increased. Since $c = 1$ would result in a completely connected network with no possibility of losing or forming new connections, we set c only close to this upper limit at $c = 0.99$ for the default choice.

4.6.1 KLS Extension

As expected, simulations using c parameter values close to one were less likely to suffer from full extinction. For $c = .5$, 54% of all runs survived to $T = 400$ iterations. At $c = .99$, 99% of runs survived to $T = 400$ iterations. Those that survived were extremely well-connected.

Connectance is a measure of the total number of links relative to the total connection possible in a network,

$$C = \frac{\sum \sum A_{i,j}}{m_J * n_I}.$$

For the default parameter set, connectance level in simulated networks is much larger than the 30% seen in meta-analysis [52]. For $c = .5$, average connectance was

Table 4.2: List of simulation models and parameters.

Stochastic Model Equations	KLS Extension	SFC Model	Asymmetric Fitness Model
Dynamic Equations			
$\bar{x}_i' = (1-\alpha)\,\bar{y}_I + \alpha\bar{x}_i$ $+\gamma_1\left(\bar{y}_I \frac{1}{n_I}\sum_{j\in I}\bar{x}_J\right) + \xi_x$	$\alpha = \frac{\sigma_x^2}{v^2+\sigma_x^2}$ $\gamma_1 = 0$	$\alpha = \frac{\sigma_x^2}{v^2+\sigma_x^2}$ $\gamma_1 = 0$	$\alpha = \frac{\sigma_I^2+\sigma_I v+v^2}{\sigma_I^2+\sigma_I v+v^2+\sigma_x v}$ $\gamma_1 = \frac{\sigma_I \sigma_x}{\sigma_I^2+\sigma_I v+v^2+\sigma_x v}$
$\bar{y}_j' = (1-\beta)\,\bar{x}_J + \beta\bar{y}_j$ $+\gamma_2\left(\bar{x}_J - \frac{1}{m_J}\sum_{i\in J}\bar{y}_I\right) + \xi_y$	$\beta = \frac{\sigma_x^2\sigma_y^2}{\sigma_x^2\sigma_y^2+(v^2+\sigma_x^2)(2v^2+\sigma_x^2)}$ $\gamma_2 = 0$	$\beta = \frac{\sigma_y^2}{v^2+\sigma_y^2}$ $\gamma_2 = 0$	$\beta = \frac{\sigma_J^2+\sigma_J v+v^2}{\sigma_J^2+\sigma_J v+v^2+\sigma_y v}$ $\gamma_2 = \frac{\sigma_J \sigma_y}{\sigma_J^2+\sigma_J v+v^2+\sigma_y v}$
Assumptions about $p_I(y)$ and $p_J(x)$			
$p_I(y)\ N(\bar{y}_I,\sigma_I)$	$\sigma_I = \sigma_y$		$\sigma_I = \sigma_y/n_I$
$p_J(x)\ N(\bar{x}_J,\sigma_J)$	$\sigma_J = \sigma_x$		$\sigma_I = \sigma_y/n_I$
Interaction Matrix Entries			
$P(A'(i,j)=1) = (1-c)m(i,j)+cA(i,j)$	$m(i,j) = e^{-\left(\bar{x}_i'-\bar{y}_j'\right)^2/2\theta^2}$		$m(i,j) = w_j(i|j)/\bar{w}_j$
Default Parameter set			
$T = 400$	Number of iterations		
$b_x = b_y = 0.01$	Probability of speciation, plant and pollinator, respectively		
Note: the number of generations simulated is approximately $10^6 bT = 4$ million			
$v^2 = 1$	Mutualism trait matching tolerance		
$\theta_x^2 = \theta_y^2 = .1$	Variance of mutualism trait in plant and pollinator population, respectively		
$V_x = V_y = .1$	Standard deviation of stochastic drift of trait in each iteration, $\xi\ N(0,V)$		
$V_s = .5$	Standard deviation of stochastic change in trait value due to speciation event, $\xi_s\ N(0,V_s)$		
$c = .99$	Proportion of connection probability based on historical association		

approximately 90%, and for $c = .99$, average connectance was approximately 91%. The results of typical simulation is illustrated in Figure 4.3.
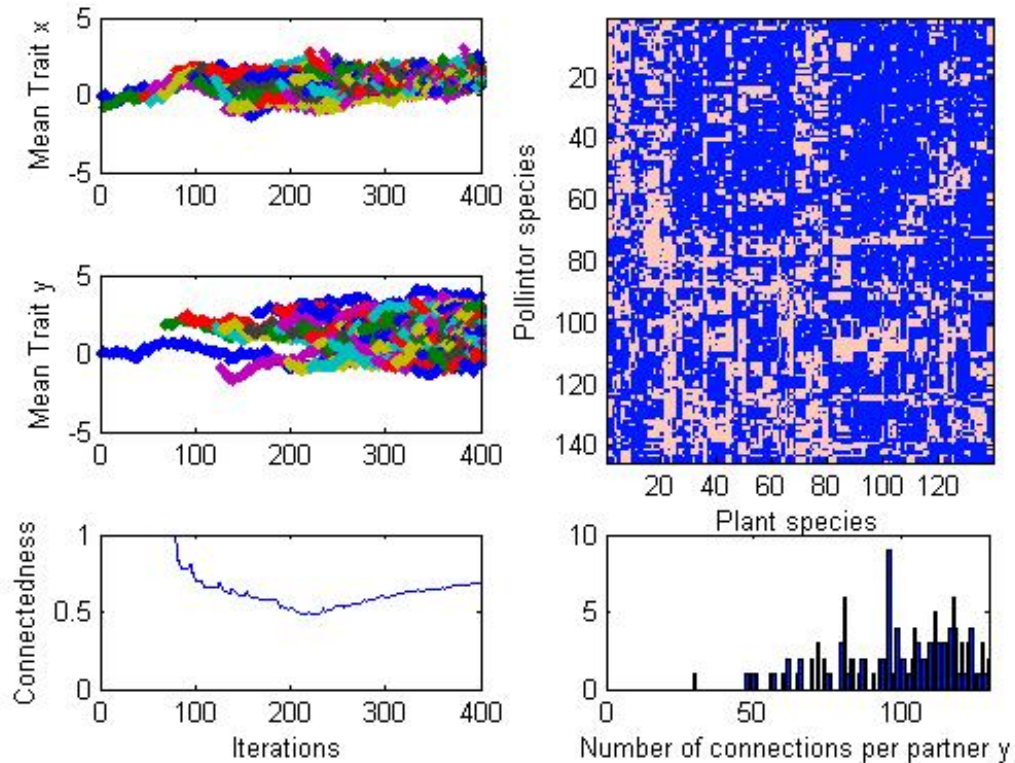
Figure 4.3: Typical simulation of an evolving KLS plant-pollinator network in the default parameter set (see Table 4.2) that results in a well-connected network with connectance level, C = 68.47%. The simulations in the upper left indicate value of the mutualistic trait under consideration for each species in plant (top, x) and pollinator (bottom, y). The right is an illustration of which species is connected to which, dark squares indicating connected plant and pollinator and light squares indicating no interaction. Bottom left indicated connectance level over time, and bottom right indicates the frequency of pollinator connections. The skew left behavior of the connectivity histogram indicates a high number of generalists and no specialist.

In the default simulation parameters, with equal generation to generation variation in plants and insects, $V_y = V_x$, we are not likely to get an extinction scenario since both plants and pollinators evolve with the same restriction on drift. Thus extremely well-connected networks are produced. If the variation in plants is increased so that

$V_y \leq V_x = V_s$, the qualitative outcome does not change, but the number of network extinction events increases.

In the well-connected networks observed, removal of a node means the network will still be well-connected. Therefore, it is unlikely disjoint bipartite subgraphs will emerge by stochasticity alone. In Figure 4.4, the initial condition is two sets of one plant and one pollinator species connected to each other, so that the initial network is a disjoint subgraph. If the trait value of a plant does not drift too close to that of a member pollinator in a disjoint network to start a new association, then these disjoint subgraphs can be maintained. As associations begin, the whole newly connected network will evolve toward a single trait value.

In general, the simulations produce overconnectedness, with an average of approximately 90% where real plant-pollinator systems are observed to have only a 30% connectance level [52]. This also results in skew left connectivity distributions because of the large number of generalists. We now examine how symmetry in fitness definitions affects connectivity.

4.6.2   SFC Model

Recall that the SFC model is constructed so the fitness definitions are now symmetric. Since this symmetry can lead to faster and more close trait matching in plants and pollinators, this leads to networks with extremely high connectivity, near 99.8%. We also get increased stability, with no observed network extinction. A typical run is shown in Figure 4.5.

Although the mathematical formulation of this model seems more realistic, the higher connectivity and the skew left connectivity distribution for reasonable parameter choices is not. A more biologically realistic stochastic Asymmetric Network Model is now investigated.
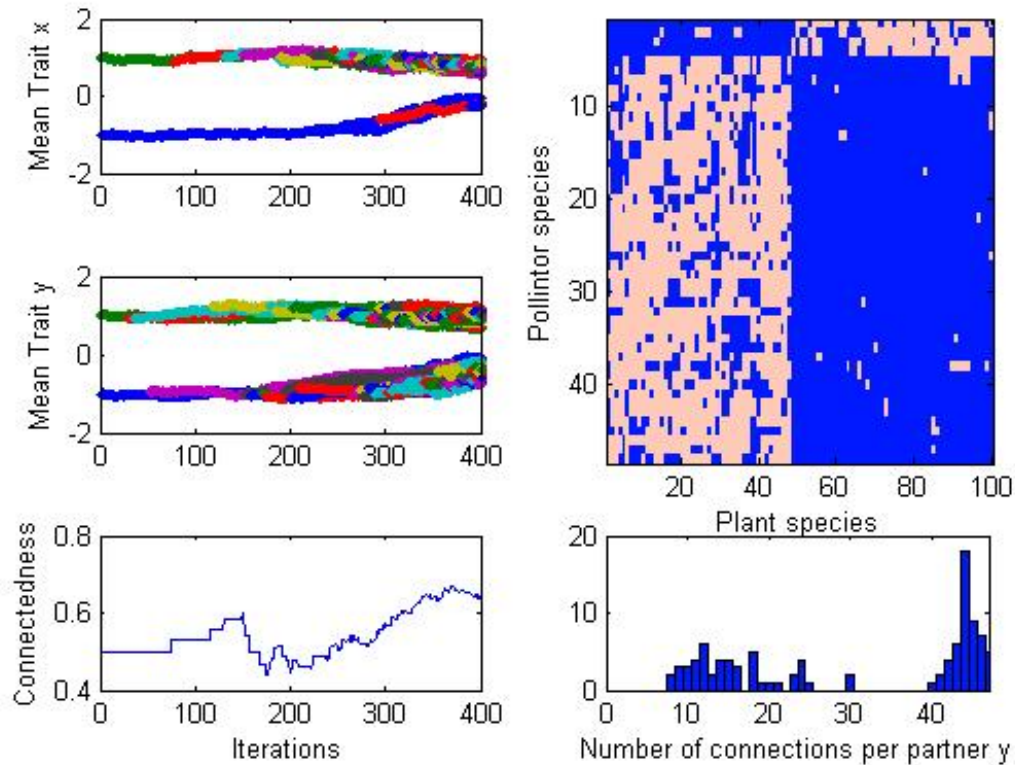
Figure 4.4: Simulation of an evolving KLS plant-pollinator network with a disjoint bipartite network as the initial condition. Parameter set is the default set as in Table 4.2, except to allow two disjoint networks to remain disjoint for longer, I reduced the drift by an order so that $V_x = V_y = 0.01$ and $V_s = 0.05$.
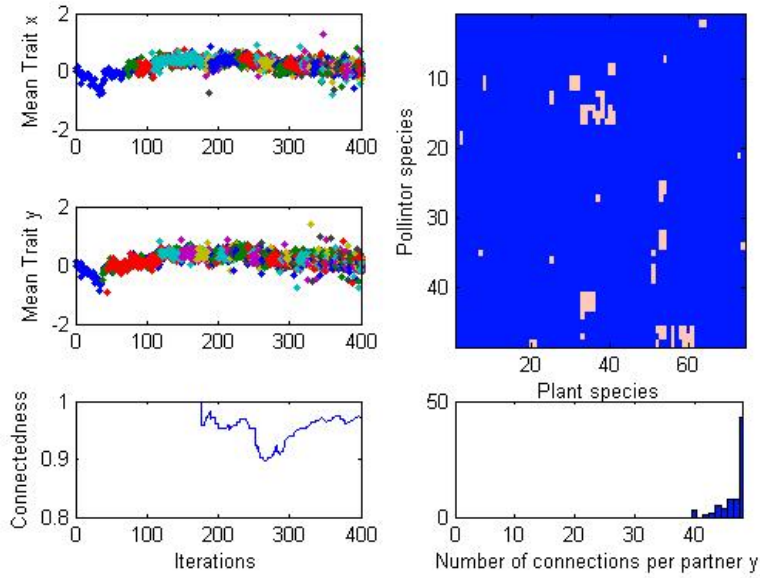
Figure 4.5: Typical simulation of an evolving SFC plant-pollinator network under the default parameter set (Table 4.2) that results in a well-connected network with connectance level, C = 97.3%

4.6.3 Stochastic Asymmetric Coevolving Network Model

The following results are for the Stochastic Asymmetric Coevolving Network (ACM) model. For the purpose of implementing these simulations, $\sigma_J^2 = \sigma_x^2/m_J$ and $\sigma_I^2 = \sigma_y^2/n_I$ as might be predicted by the central limit theorem had partners been picked randomly from a probability distribution of means.

Cophylogeny

Four major coevolutionary events are highlighted when reconstructing cophylogenies: cospeciation, sorting, duplication, and host-switching [77]. Figure 4.6 illustrates a small network in early stages of development. At iteration 18, the first speciation event occurs in the plant lineage. We will call this plant species 2. It is a duplication event, but by the 100th iteration, it forms a 1-1 connection with pollinator species 3, which arose in iteration 68. This begins as a duplication event but results in a host-switch

from the plant perspective and a sorting event from the pollinator perspective. Note that this simulation model does not impose cospeciation, but cascade-like cospeciation events do occur by chance as one can see from the timing of speciation events that result in pollinator species 2 and plant species 3 at iterations 29 and 65, respectively.
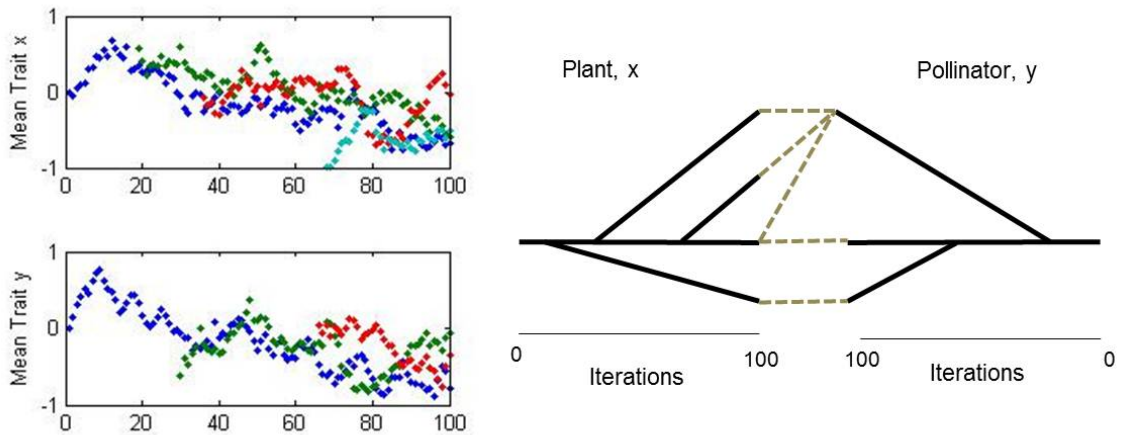
Figure 4.6: Left: Zoom of the first 100 iterations of a simulated cophylogeny. Color indicates speciation order with blue as the initial connected species, then green, red, and light blue (if applicable). Right: Associated untangled cophylogenetic network.

Figure 4.7 shows the evolution of a cophylogenetic network over 300 iterations. Both patterns of modularity and phylogenetic conservation as well as one-to-one specificity in concordance with phylogenetic signal are present at various points of time in the network. This is one outcome, where as Figure 4.6 illustrates how mutualistic processes can result in host-switching and associations that are unexpected based on phylogenetic signal alone. Therefore patterns other than perfect concordance with phylogenetic signal alone are part of the rule, not an exception.

Connectance

Recall that survival probability is related to its connectance and the relative importance of historical connection, c. We showed that the highest values of c would produce the most stability for a connection over time. Since the loss of all
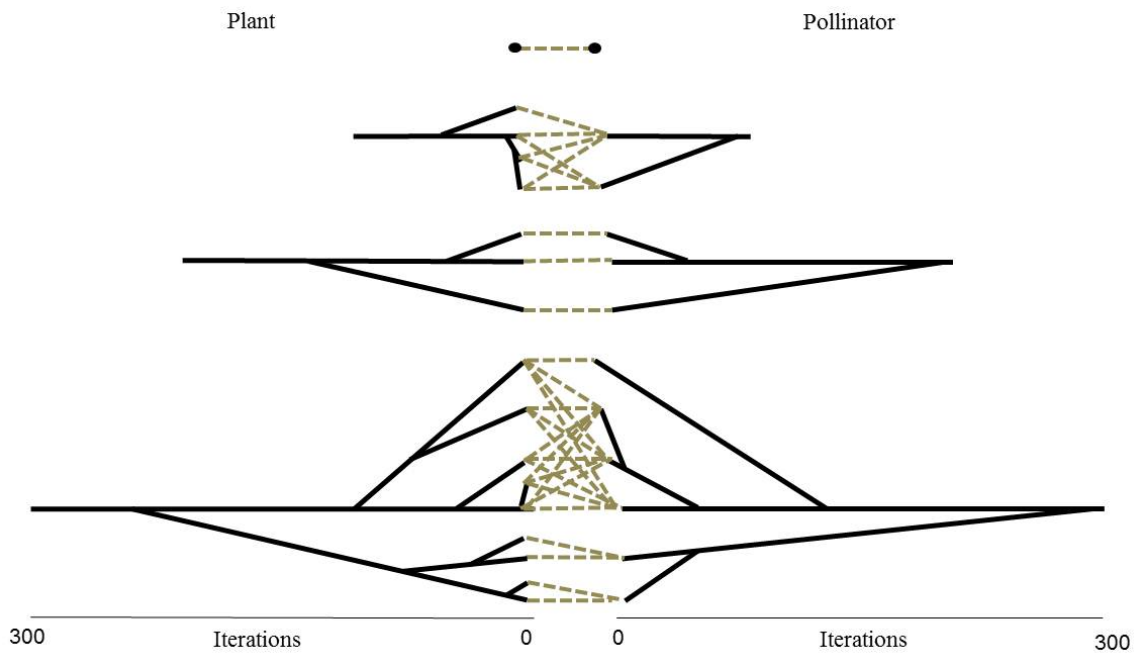
Figure 4.7: The following are snapshots of a cophylogenetic network over time at iteration 0, 100, 200, and 300. Notice how at iteration 200 there is a 1-1 correspondence in accordance with phylogenetic signal, and snapshots at 100 and 300 show modularity and phylogenetic conservation without the 1-1 correspondence.

connections for a species results in the immediate extinction of a species, it might be guessed that higher values of c confer more stability on networks. It was confirmed this mathematical intuition via a bar graph representing survival probabilities for a networks under various parameter conditions in Figure 4.8.
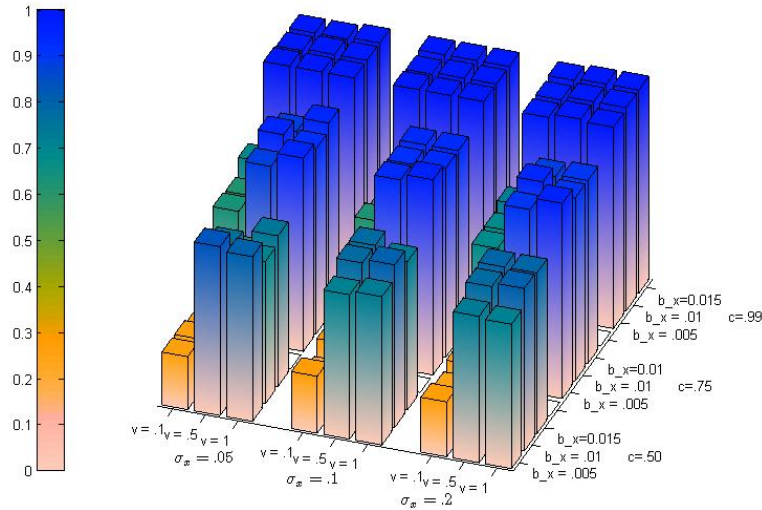
Figure 4.8: Bar graph of the probability of network survival levels after $T = 400$ iterations of 500 simulations of the stochastic ACN model under 81 parameter set combinations, varying 4 parameters over 3 values each. Those parameters not shown in the figure are set at the default parameter set.

Recall that a review by Jordano [52] of several mutualistic webs indicated that connectance is on average around 30%, with higher connectance in smaller networks. Results indicate an average connectance level in stochastic simulations of approximately .3175 or 34.75% (Figure 4.9) for all networks surviving until the end of the simulation time length with the default parameter set. To look at the effects of parameter choice on connectance level, more extensive simulations under 81 variations of the parameter set as shown in Figure 4.10 show average connectivity levels ranging from .3117 to .4003. Less connectance is generally observed for smaller drift of plant phenotype versus pollinator phenotype. This could be due to less environmental pressure on plants or smaller population sizes or inbreeding in pollinators, like in

fig-wasp. In this case, randomly walking specialist plants would take longer to explore the phenotype space and intersect with another pollinator who could exploit it. Alternatively, less fluctation in plant phenotype could mean less risk for specialists.
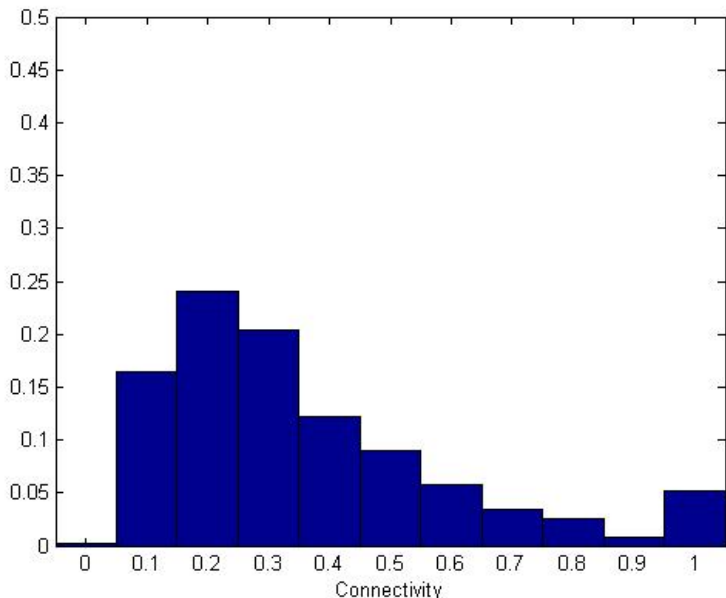
Figure 4.9: Histogram of connectance levels of surviving networks after $T = 400$ iterations of 500 simulations of the stochastic ACN model under the default parameter set. Of the 500 simulations run, 241 survived to 400 iterations. Average connectance was calculated at .3481.

Several models were fit to the relationship between connectance levels and species number from simulations generated, see Table 4.3. The best fit model was an inverse power-law model with exponent of 0.5. Results displayed in Figure 4.11 also indicate an inverse power-law relationship between connectance and network size, also reflected in a meta-analysis of field studies [52]. Those some studies suggest a stronger relationship with exponent 2 [7].

Recall that the connectivity matrix, A, is a function of mostly historical connection, but also a probability based on the fitness effect of the connection. Bascompte, *et al.* observed that connectance must decrease as the size of the network increases to maintain stability [9]. A general feature of these simulations is that
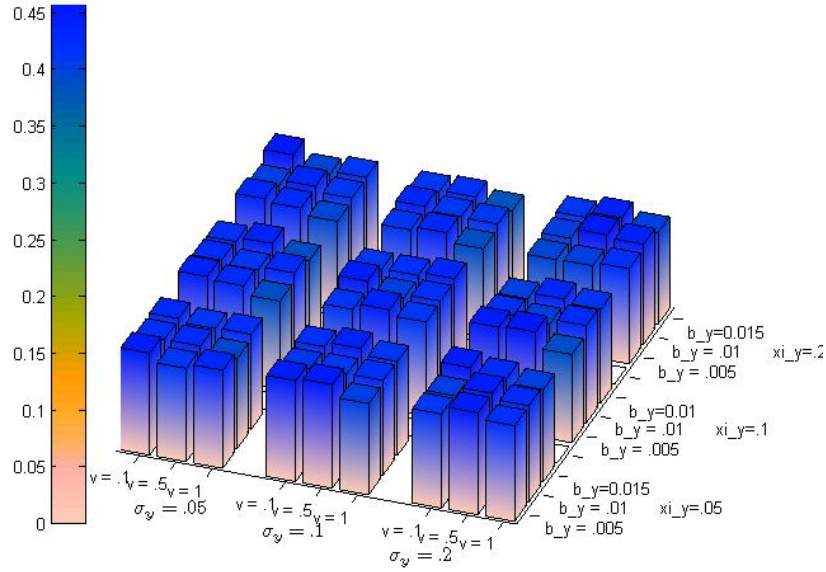
Figure 4.10: Bar graph of average connectance levels after $T = 400$ iterations of 500 simulations of the stochastic ACN model under 81 parameter set combinations, varying 4 parameters over 3 values each. Those parameters not shown in the figure are set at the default parameter set.

Table 4.3: List of best fit models for connectance versus number of species for 50 runs of the Asymmetric Coevolving Network model.

Equation type	Best-fit model	r^2
Linear	$C = -0.0019N + 0.4708$	0.2763
Quadratic	$C = 0.00001N^2 - 0.0054N + 0.5945$	0.3756
Exponential	$C = 0.4573e^{-0.008N}$	0.4725
Power	$C = 2.5257N^-0.583$	0.5607
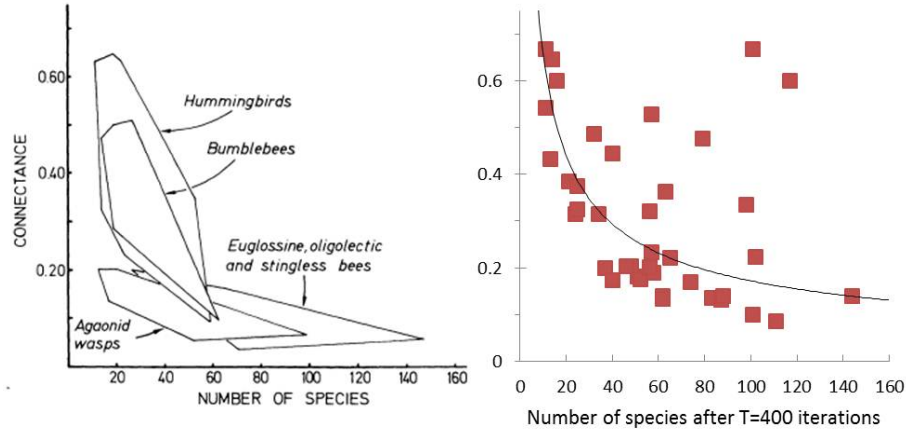
Figure 4.11: A visual comparison between connectance values versus network size for typical plant-pollinator webs and simulations from the AFC model with the best fit line from Table 4.3. The left is from a meta-analysis of field data [52]. The right includes data points from a run of 50 simulations that fell into network size range and connectance levels of those inventoried by that meta-analysis.

connectance level decreases over time as networks grow. See the simulation in Figure 4.12 for a typical graph of connectance level over time.

4.6.4 Node Asymmetry

One observation is that animals tend to have more species then their mutualistic plant partners. For 500 simulations, an index was calculated for node asymmetry:

$$AN = \frac{\text{the number of extant plants - animals}}{\text{the total number of extant plants and animals}}.$$

A network with no node asymmetry should have on average $AN = 0$ and should be negative if there are more plants on average than animals. Recall pollination networks experience a ratio of 1:4 plants vs pollinators, $AS = -0.6$. For the default parameter set which assumes equal speciation rates for both plants and pollinators, the average index of node asymmetry is .0235 over a batch of 500 simulations. Therefore, increasing the speciation rate of pollinators from .01 probability per iteration to .015, increases the average index of node asymmetry to -0.5112 and decreases average

100
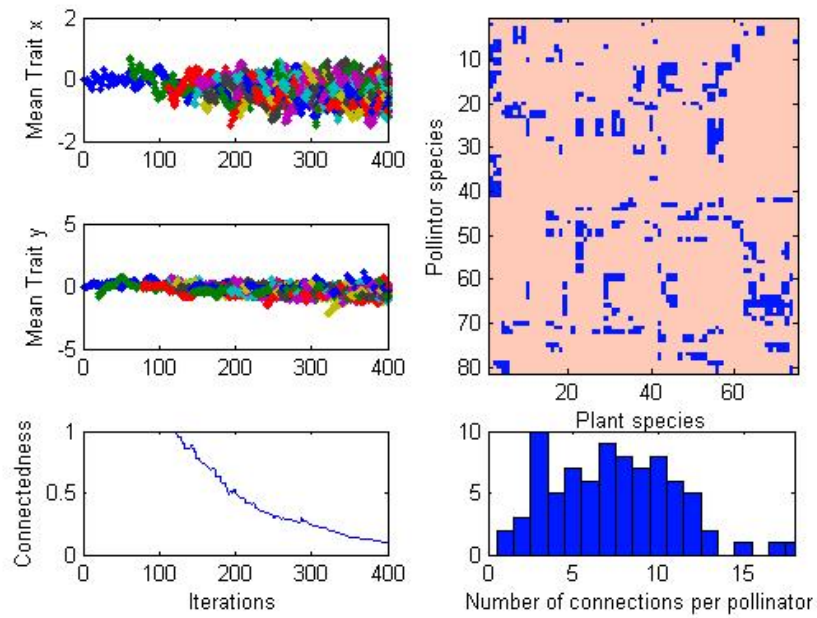
Figure 4.12: A typical simulation of the Asymmetric Coevolving Network model under the default parameter set in Table 4.2. The upper left figures show the evolution of the mutualistic trait under consideration and the upper right shows the interaction matrix. The lower left is a graph of the network connectance level over time, and the lower right is a connectivity distribution showing the number of connections per pollinator.

connectance to .2594. A histogram of node asymmetry index is provided in Figure 4.13.
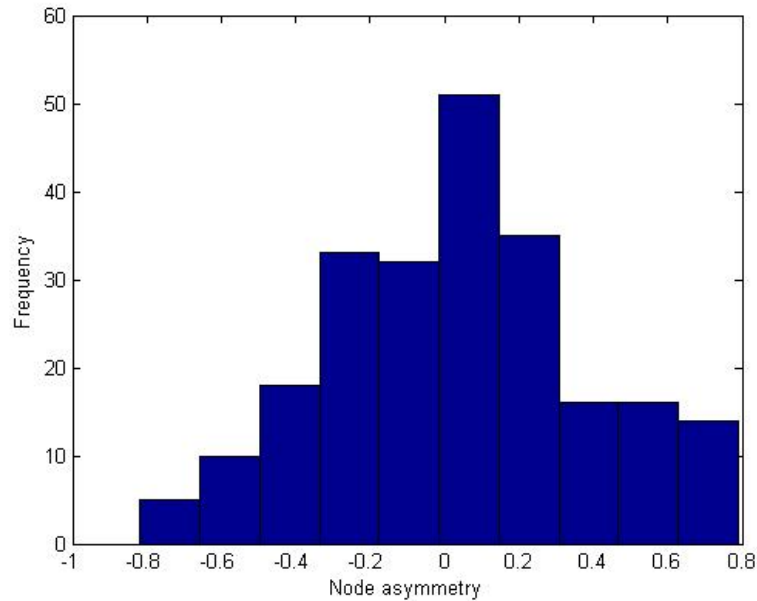
Figure 4.13: Histogram of connectance levels of surviving networks after $T = 400$ iterations of 500 simulations of the stochastic ACN model under the default parameter set, but with $b_y = 0.015$. Average node asymmetry index is at -0.5112 showing a ratio of approximately 1:3 plants:animals.

An argument for differential rates of speciation is not unreasonable, especially since differences in dispersal ability for plants and animals could result in higher rates of speciation in animals that can more readily encounter and exploit new habitat [41, 92]. It has also been found that differential growth rate (and differential limiting size) promote power-law truncations best fit models for connectivity distributions [41]. A second parameter set, termed "differential speciation rate parameter set," will be investigated extensively as an alternative to the default parameter set. It is the default parameter set with a modified pollinator speciation rate 50% higher than the plant rate (*i.e.* at 0.015).

Link distributions

Frequency distributions of the number of interactions per pollinators in some studies suggest that there are many specialist pollinators and with a long tail of generalist pollinators [7, 48]. Simulations which include connectivity distributions are usually near normal or skew right, but do not show the extent of specialist pollinators as predicted in some communities. A sample of 46 simulation runs in the default parameter set yielded 45 surviving networks, 60% of which had a greater or equal number of pollinators than plants. The result is a possible abundance of specialist pollinators. In these networks, only 24% had pollinators visiting one plant species as the most common strategy, but the rest were all either normal or skew right. In the networks where the number of plant species were greater, only 8% of networks had a specialist pollinator. In these cases, this strategy was just as common as pollinating a few plant species. 60% of these networks were skew left. Figure 4.14 is a histogram representing the connectivity of pollinators.

The effect of parameters on the frequency of specialist pollinators was also explored (see Figure 4.15). The average proportion of specialists in each of 81 parameter sets ranged from 0.0278 to 0.3251. For each value of c, variation was small in the proportion of pollinator specialists in the network. However, the runs in which $c = 0.99$ had significantly more specialists than for lower values of c ($p = 0.000$). Therefore, coevolutionary history promotes specialization. Specialization may also be promoted by size asymmetry which can be exaggerated though sampling. One network's plants were sampled at rate of 18%, but still interacted with %77 of the pollinator species in the full network.

Recall that a feature of mutualistic networks is heterogeneity in link connection distributions. This is described by the large number of low connectivity members and long tail of highly connected members. Power-law distributions were suggested as an alternative to exponential connectivity distributions for some networks, but in mutualistic networks, truncated power-law is the most common best-fit of the three
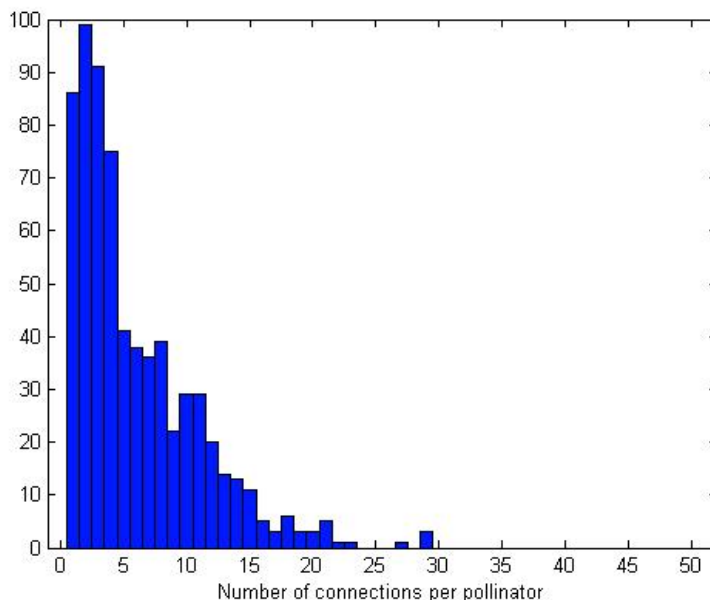
Figure 4.14: Histogram of the number of connections pollinators have from 45 simulations surviving to $T = 400$ out of 50 simulations of the ACN model under the default parameter set.

scenarios [53]. Exponential, power-law and truncated power-law models were fit to 16 sets of data (4 networks for plants and for pollinators in each default parameter set analyzed at T=400). In each case, truncated power-law was the best fit. See Figure 4.16 for illustrative examples of how the truncated power-law best-fits our generated networks. See [53] to compare this with results on real networks.

Although power-law fits to plant-pollinator networks are not the best-fit, they can still convey the relative abundance of specialists. The best power-law fit to these networks has an average exponent of 1.23 ± 0.04 for pollinators and 0.84 ± 0.04 for plants [53]. Table 4.4 shows average exponents for power-law fits for simulated networks of the ACN model. These networks do capture the higher power law exponent for pollinator connectivity over plant connectivity. The fits shown in the bottom of Figure 4.16 have exponents of 0.5396 and 0.3303 for pollinator and plants, respectively. However the smaller values of γ in simulated networks indicated less pscecialism is expected under the ACN model than is observed in field studies. One
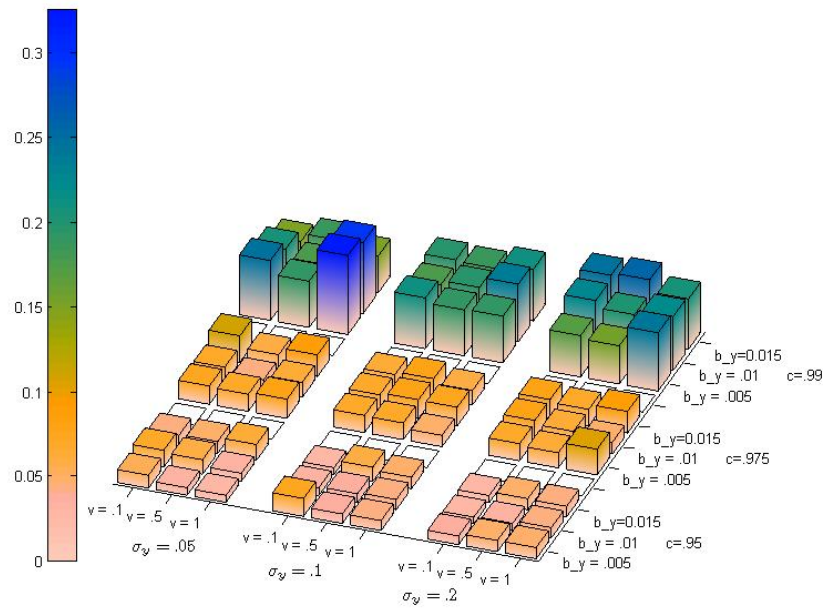
Figure 4.15: Bar graph of relative abundance of specialists after $T = 400$ iterations of 500 simulations of the stochastic ACN model under 81 parameter set combinations, varying 4 parameters over 3 values each. Those parameters not shown in the figure are set at the default parameter set.
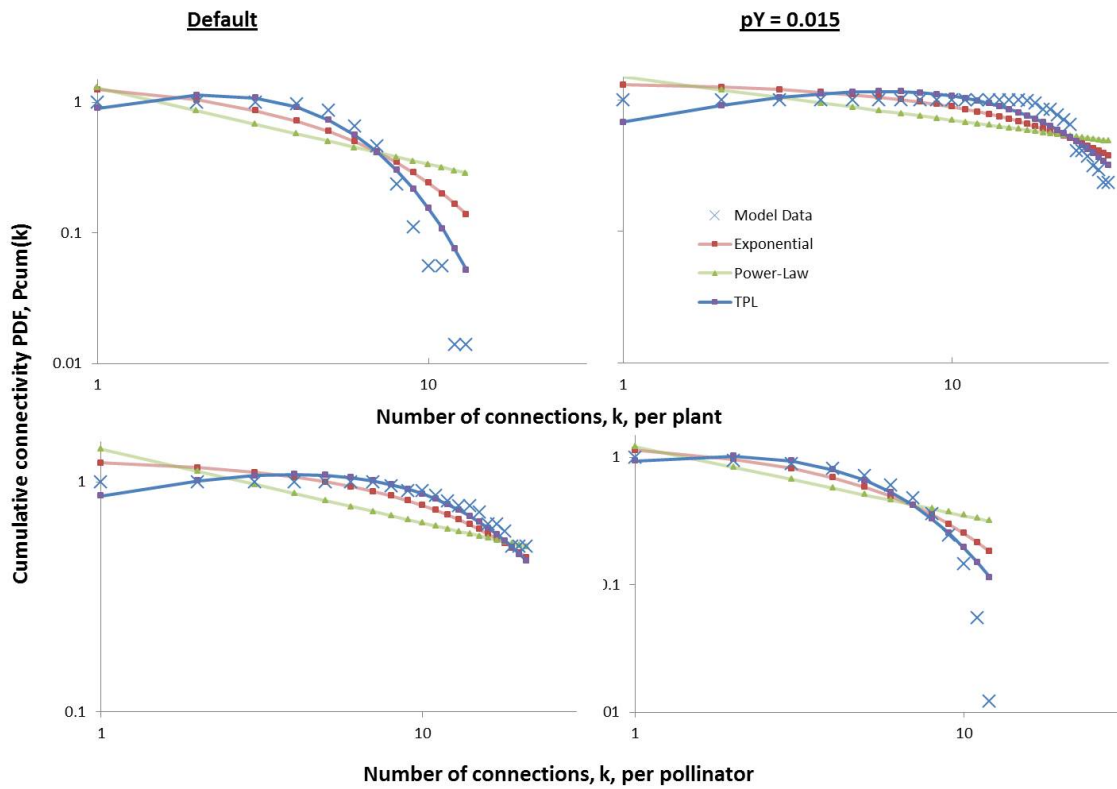
Figure 4.16: Example best fit models to simulated bipartite networks. The top is a bipartite network simulated using the default parameter set and the bottom is a network simulated using the differential speciation rate set. In all cases shown above, the truncated power-law was the best-fit.

contributing factor is the role of sampling. For the same simulated community, 18% of all plants were sampled randomly. Despite only a small fraction of plants under consideration, 77% of all pollintors interact with these plants. Power-law fits to this new sub-community yielded exponents of 1.4718 for pollinators and 0.2837 for plants (see Table 4.4). This example shows how dramatically sampling may affect network properties.

Table 4.4: List of exponents of the power-law model fit for simulated networks. Eight plant and pollinator connectivity distributions are represented here from four simulated networks in each parameter set. Power-law model used: $p(k) \propto k^{-\gamma}$.

Parameter set	Animal or Plant	$\bar{\gamma} \pm$ SE	$\gamma_\mathbf{A} \stackrel{-}{=} \gamma_\mathbf{B} \pm$ SE
Default	Animal	0.435 ± 0.076	-0.068 ± 0.107
	Plant	0.503 ± 0.039	
DSR	Animal	0.619 ± 0.029	0.2779 ± 0.023
	Plant	0.341 ± 0.011	
Description	**Animal or Plant**	γ	$\gamma_\mathbf{A} - \gamma_\mathbf{B}$
DSR, full	Animal	0.5396	0.2093
	Plant	.3303	
DSR, sample	Animal	1.4718	1.1881
	Plant	0.2837	

Dependence distributions describe the connectivity of one's connections. Dependence is defined as the sum of the proportions of partners your partners have or as a relative frequency of visitation. Recall the review of Jordano, 1987 [52], in which the dependence histograms from real mutualistic communities where often skew right for both plants and pollinators. Simulation results confirm that this same pattern emerges, including a more pronounced peak of weak animal dependences on plants. See Figures 4.17 and 4.18 for average results over 100 simulations. In addition, Figure 4.17 shows that quantitative measures of dependence via relative interaction frequency
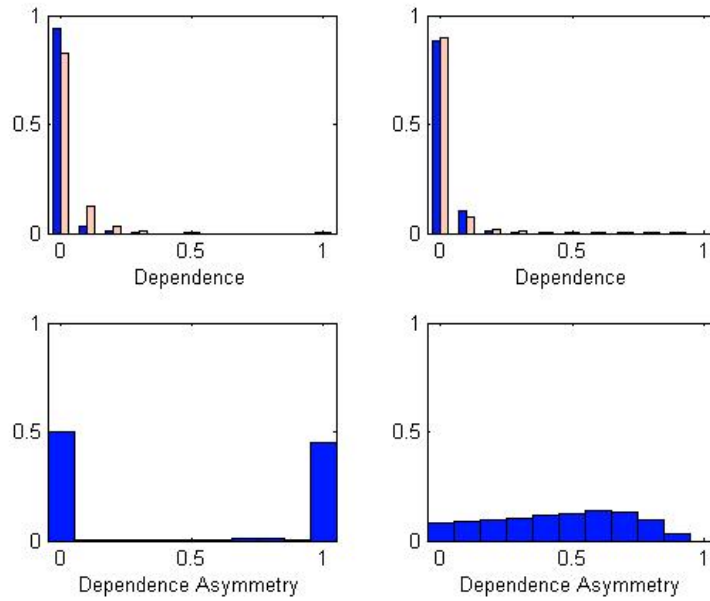
Figure 4.17: Dependence histograms resulting from the combined result of 500 simulations of the ACN model under the default parameter set. The left shows dependence as calculated from the binary interaction matrix and the right set uses frequency of visitation as a quantitative measure of dependence. Histograms at the top show dependence of plants on animals (dark) and animals on plants (light). Histograms at the bottom show the dependence asymmetry.

when speciation rates are equal can quite difference than measures of dependence via a binary interaction matrix. This does not happen on average when speciation rates are different, see Figure 4.18.

4.6.5 Dependence asymmetry, nestedness and modularity

Dependence asymmetry between plants and pollinators occurs primarily when strong relative size asymmetry is present, but also agrees with prior observations that high frequency of weak dependence values and their asymmetry occur when one dependence is large [9]. The bottom histograms in Figures 4.17 and 4.18 show the distribution of dependence asymmetry using the same method of Bascompte *et al.*, 2006 [9]. A value near one indicates strong asymmetry *i.e.* that strong dependence in one direction is accompanied by weak dependence in the other direction. Therefore
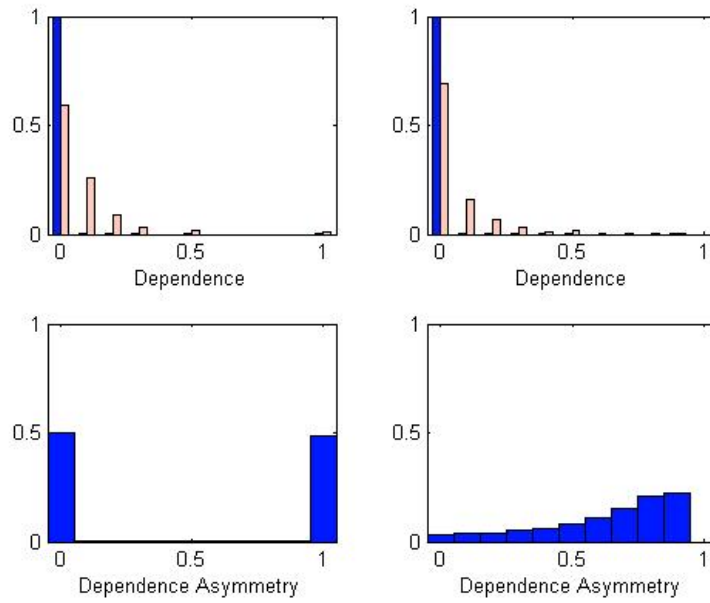
Figure 4.18: Dependence histograms resulting from the combined result of 100 simulations of the ACN model under the differential speciation rate parameter set. The left shows dependence as calculated from the binary interaction matrix and the right set uses frequency of visitation as a quantitative measure of dependence. Histograms at the top show dependence of plants on animals (dark) and animals on plants (light). Histograms at the bottom show the dependence asymmetry.

the skew left nature of the asymmetry index indicates that simulated networks often have attachment of specialists to generalists. This pattern is more pronounced in the case of the differentiated speciation rate as seen in Figure 4.18 and can be compared to asymmetry histograms of real mutualistic networks in Bascompte, 2009 [6].

Recall that heterogeneity coupled with this pattern of generalists attached to specialists may also be related to nestedness and modularity. Some examples of interaction matrices ordered according to their connectivity are in Figure 4.19. These can be compared to the figures of nested mutualistic networks in Figure ?? [6]. Modularity and nestedness are likely mechanistically related, and in these cophylogenetic networks is considered the result of phylogenetic conservation [7]. In the simulations of the ACN model presented here, modularity is the result of both the importance of historical association in addition to the trait matching necessary to confer fitness onto one's partner to maintain connections. Figure 4.20 is a simulation which shows this modularity strongly.

4.7 Discussion

The incorporation of populations with 1-dimensional distributions of quantitative traits and ecological interactions throughout their evolutionary history seems to produce a relatively realistic mutualistic network. The ASN model has few mechanisms at play, but all are biologically realistic. Therefore this should provide intuition into why certain network patterns emerge as a result of coevolution.

The ASN simulations replicated all patterns of cophylogenetic "events" discussed early on: cospeciation, sorting, duplication, and host-switching. Duplication events are how all speciation events are defined in this model. Within a small number of iterations, these duplicated lines can result in sorting, cospeciation or host-switching, as the lineage evolves toward connections with partners that can provide the most attention or relative frequency of visits. For example, in our model, "cospeciation" starts as a duplication event with one species splits and sharing all partners. Because
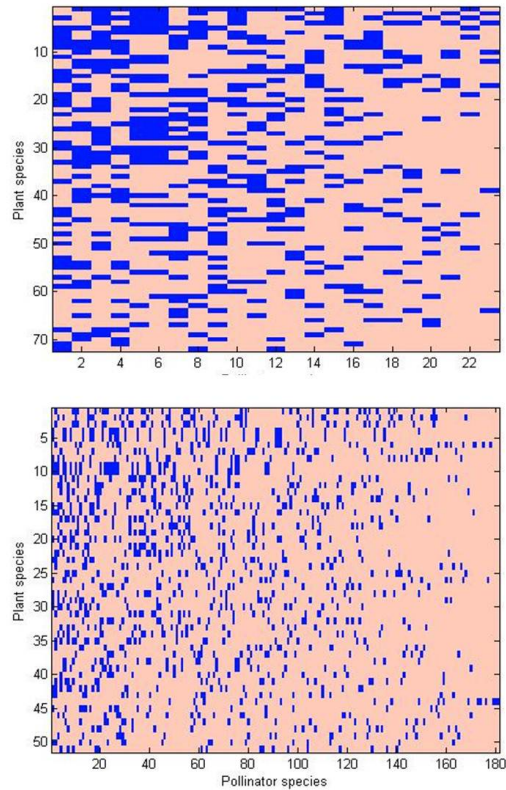
Figure 4.19: Simulation showing nestedness of interactions. The top figure is an example under the default parameter set and the bottom is under the differential speciation rate set.

relative fitness contributions are lowered when partners are shared, a novel partner species can be favored. Therefore, cospeciation can happen merely as a result of chance that two partners speciate at similar times, but when this does occur, it is maintained preferentially over other chance speciation events as are favored by selection.

We can also observe the evolution of cophylogenies with both specificity and modularity. All of these patterns are consistent with meta-analysis of phylogenetic reconstruction efforts [65, 66].

The models presented here only consider the effect of 1-dimensional traits. In random walk theory, the probability that in one dimension, a random walk will eventually return back to its point of origin is 1. The probability that the difference
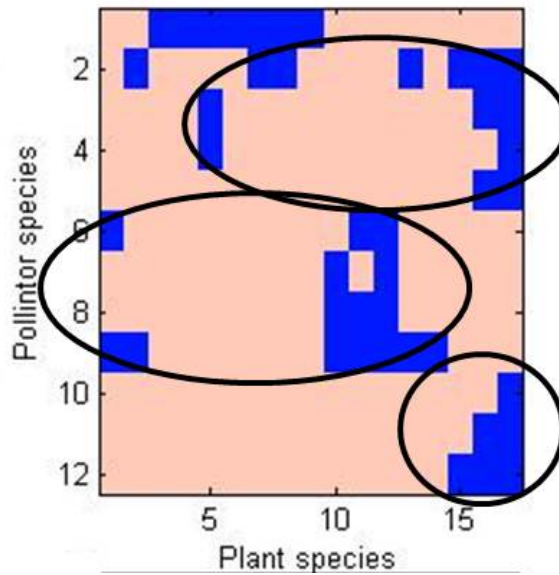
Figure 4.20: Simulation of the ACN model under the default parameter set with a strong clustering effect resulting.

between any two random walks will be zero at some point is 1 [38]. Regardless of the matching tolerance, eventually any two disjoint subgraphs will evolve to one connected group. This is due to each subgraph making an independent random walk along a line through stochastic fluctuations in the model. Therefore complete ecological isolation, for even a short time, could be enough to allow populations to diverge and/or reproductively isolate before connecting again.

If there is more than one quantitative trait involved in the mutualism, the probability of such disjoint subgraphs reconnecting could decrease. Two independent traits, representing a 2-dimensional random walk would return with probability 1. For n independent traits, $n > 2$, the probability of return decreases [10, 38]. This also supports the notion that the c parameter should be close to 1 as c represents the importance of historical connection. It is likely multiple traits are involved in the mutualism meaning matching in any one trait is not predicted to enable a new association. Using the analogy of intersecting random walks may also help us to understand why some systems maintain one-to-one connections much longer than we

might expect. Random walks with small steps on average will take a long time to explore phenotype space in search of a more fit interaction.

Relative importance of historical connection is also significant to maintain stability and to mimic results of field meta-analysis. This suggests that specialization actually confers stability to the network since many traits are involved in a mutualism. Although biologically non-intuitive, mathematically, this makes sense because interactors are much less likely to host-switch due to the availability of alternate resources that match in terms of just one trait, as seen in prior models in this paper. However, it is also partially an artifact of extinction that is determined only by a loss of connections instead of also by random environmental influence. Testing whether or not these networks are robust to sudden and random node loss could help us understand the fragility of plant-pollinator systems. Testing this loss on varying types of realistically evolved networks with varying configurations could also improve predictive efforts.

Power-law connectivity scaling associated with "rich get richer" arguments [5] are observed in these distributions, but are formed here with out any direct preferential node attachment imposed. Instead, there is an underlying biological mechanism in which species with many connections are more likely to experience a partner speciation event in any particular time step. Species with many connections will likely gain more nodes in any particular interaction since these new species inherit parental connections. This is also likely to contribute to patterns of nestedness and modularity seen in mutualistic webs. Furthermore, the cumulative probability distributions of mutualistic webs are even best-fit by truncated power-law models, both in real networks and in the networks simulated here [53].

Connectance levels simulated are similar to that of meta-analysis results [52], both on average at the end of simulations and over time as networks grow. Dependence and connectivity distributions vary, but are all skew left as predicted by meta-analysis and other numerical investigations [7, 9]. Dependence distributions seem to agree, but connectivity distributions show less specialization by pollinators than

113

is indicated by some studies. There are multiple explanations for this, the most plausible is connectivity distributions may vary depending on community. The extreme example given in [7] may not be reflective of all communities. Another is that the penalty for sharing partners may need to be weighted higher. The best indicator of specialization in simulated networks is high relative importance of historical connectedness. Recall another interpretation for a large c is that multiple traits are involved in the mutualism. Lastly, one has to consider the effect of sampling on network properties. This currently remains an open question. In the case of specialist as indicated by the exponent of power-law fits to connectivity distributions. It is possible that sampling only a portion of the community in combination with observing only a fraction of interactions due to time-scale or rare interactions would increase specialism noted in field studies.

It would be interesting to compare the evolution of mutualistic networks with that of evolving antagonistic networks. A synthesis of the two would likely give the best insight into systems like fig-fig wasp, since that mutualistic system is parasitized by cheater non-pollinators. Overall, these models as a whole show that the role of ecology is important in determining evolution and the role of evolution is important in shaping ecological networks. This is a network of its own, in which each piece alone may not give the whole story.

Chapter 5

Summary and Future Directions

In summary, the following points are made throughout this paper

- Adding a reciprocally evolving environment to a two-loci, diallelic resource user model results in the instability of the polymorphic equilibria. This means that although fixed resources can promote genetic biodiversity as percentages of resource and user match, biological environments will reciprocally evolve to meet the need of the dominant resource user. Although this means that most plant-pollinators relationships are robust to any small amount of introduced variation, this process leads to rapid fixation of one type and so may promote a loss of biodiversity.

- The same process that promotes fixation in mutualistic systems, also can promote rapid host-switch. This is driven primarily by the relative abundance in alternate-host type. As long as pollinators are not strict in their preference, any variation enabling adaptation to a more abundant host can be exploited.

- The time to fixation in mutualistic systems is influenced by the choosiness of the pollinator. If pollinators can visit even just two plant populations, as in the fig-fig wasp system modeled, the time to fixation can be lengthened considerably. This can lead to extensive hybridization among plant phylogenies in closely

related species. If one considers the effect of visiting multiple plant populations, a lag in time to reproductive maturity for fig-trees, and environmental variation, these processes could promote the maintenance of biodiversity.

- In the models discussed in the three points above, the models predicts two stable equilibria at the fixed genotypes. However, the biologically interesting consequences deal with more than just stability of equilibria. The second point is only elucidated from the understanding of the phase-portrait and the differences in long term dynamics due to initial conditions. The third is a result of investigating numerically the time to fixation.

- If historical connectedness plays no role in future mutualistic associations, then networks will go extinct rapidly. If historical connectedness plays the only role in future mutualistic associations, then completely connected networks will always result.

- The relative importance of historical connectedness is a proxy for the importance of the evolutionary energy devoted to specialized traits. One might surmise then that high levels of this relative importance would lead to higher rates of specialism in networks, which is true. However, one might also infer that this specialism may make these networks less stable, which is not true. Networks have the highest survival probability when the relative importance of historical connectedness is close to one, indicating that the probability of losing or gaining a new partner based on the anticipated fitness gain on one trait is very low. This helps network stability as the network is not continuously reorganized based on the whim of pursuing potentially better partners. This explanation is analogous also to the idea that information reliability or Shannon entropy is maximized in complex networks [29, 102].

- Even with a high relative importance of historical connectedness, resulting simulated networks do not display perfect cophylogenetic matching. This agrees

116

with cophylogenetic analysis of fig and fig wasp trees [55]. It is part of the rule, and not the exception, that host switches through out a long lineage history have occurred and have contributed to this mismatch. Also is it part of the rule that multiple connections and many crossings among interacting species are present in the tips, and that this becomes more common in lineages that have interacted over a long period of time [49].

- A network whose connections evolved over time based on fitness of relationships can produce many of the phenomena observed in real networks. Shannon entropy is maximized through the reliability of partners. There are few generalists and many specialists, confirmed by emerging truncated power-law distributions of node connectivities. The emergent networks are biodiverse and stable, despite low overall connectivity. Modularity and some nestedness can occur. This is a mechanistic approach to network assembly that employs the fundamentals of ecology and evolution. Niche trait-models and models that maximize Shannon entropy may result in good models[17, 29, 85], but they do not explain how mutualistic networks evolve over time through natural principles.

- Sampling of real ecological networks may dramatically affect observed results for important network features, such as the decay parameter on the truncated power-law curves for connectivity distributions. This amplifies the potential issue with creating networks based on mined patterns. It is possible that models based on observations from sampling networks may lead to conclusions different than for simulations of full networks. This becomes important as simulated networks are used as experimentally to understand how climate change or biodiversity loss may affect ecosystems.

The simplest models of mutualism with a genetic basis between plant and pollinator illustrate both robust one-to-one associations as well as host switching. Relative abundance of alternate plant type is the primary driver in these simple

117

systems. When pollinators utilize a network of plants because of seasonality or flowering times, time to fixation of a single plant and pollinator increases. Variation is maintained long-term, especially for increased choosiness. Therefore, pollinator specificity contributes biodiversity. It would be a natural extension to increase the number of loci under consideration and/or network these models in space with varying backgrounds. This might more realistically approximate mutualistic networks and give rise to longer maintenance of biodiversity.

The effect of mutualistic interactions on the formation of cophylogenic and mutualistic web patterns was next explored. To date cophylogeny models are limited to reconstruction and web models have focused on assembly rather than evolution [47]. Using ecological and evolutionary principles of mutualistic interactions on a quantitative trait, we simulated evolved phylogenies and networks. These principles gave rise to many similar properties as observed in real mutualistic webs, but with realistic biological mechanisms. The most important feature in these models were that associations were based on relative fitness and historical connectedness. Relative fitness takes into consideration what proportion of visits any partner population comprises. Historical connectedness is how important prior association in determining a future connection. Increased importance of historical connectedness is associated with increased network stability and increased specialism. Therefore stability and specialism are intertwined mechanistically. Simulations also reveal the complexity of cophylogeny. Mutualism with high importance of historical association does not imply perfect cophylogenetic matching.

The importance of this historical connectedness, c warrants further investigation. Alternatives to this paramter include describing as a function of the number of consecutive years associated or modeling trait matching explicitly by considering many more traits involved in the mutualism. Another fruitful avenue of this modeling effort would be in the analysis of simulated networks. For example, many simulated networks are tested for robustness to species extinction. An R-package exists for bipartite networks that can do automated connectivity distribution best fits and

calculate nestedness and modularity [23]. One could use these established tools to be able to compare results across papers. Also the automation provided by a code conversion to R would enable the data-mining of large numbers of generated networks as seen in recent PNAS, Nature, and Science papers [9, 17, 85]. Once networks are simulated on a large scale with automated data mining techniques, one can answer questions about the role of sampling effects on realistically evolved mutualistic webs and observed webs.

Lastly, the role of mutualistic webs within larger webs which include antagonistic interactions and population dynamics are also possible future avenues for investigations. However, as each of these complexities is added to the model, it is more difficult to discern which underlying mechanisms are contributing to overall patterns. Therefore, this work is fundamental in connecting biological mechanisms to results from large complex simulated networks with varying population and ecological dynamics. Knowing the underlying mechanisms increases understanding of biodiverse network stability and therefore informs policy on conservation measures [80].

Bibliography

Bibliography

[1] Agosta, S. J. (2006). On ecological fitting, plant-insect associations, herbivore host-shifts, and host plant selection. *Oikos*, 114:556–565. 38

[2] Albert, R. and Barabási, A.-L. (2000). Topology of evolving networks: local events and universality. *Phys. Rev. Lett.*, 85:5234–5237. 58, 64, 65

[3] Althoff, D. M., Segraves, K. A., Leebens-Mack, J., and Pellmyr, O. (2006). Patterns of speciation in the yucca moths: Parallel species radiations within the tegeticula yuccasella species complex. *Systematic Biology*, 55(3):398–410. 8

[4] Asratian, A. S., Denley, T. M. J., and Haggkvist, R. (1998). *Bipartite graphs and their applications*. Cambridge University Press. 7

[5] Barabási, A.-L. and Albert, R. (286). Emergence of scaling in random networks. *Science*, 286:509–512. 59, 113

[6] Bascompte, J. (2009). Mutualistic networks. *Frontiers in Ecology and the Environment*, 7(8):429–436. 58, 62, 63, 64, 66, 110

[7] Bascompte, J. and Jordano, P. (2007). Plant-animal mututalistic networks: the architecture of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 38:567–593. 7, 55, 56, 57, 58, 60, 62, 63, 68, 85, 98, 103, 110, 113, 114

[8] Bascompte, J., Jordano, P., Melián, and Olesen, J. M. (2003). The nested assembly of plant-animal mutualistic networks. *Proc. Natl. Acad. of Sci. USA*, 100:9383–9387. 56, 61, 62

[9] Bascompte, J., Jordano, P., and Olesen, J. M. (2006). Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science*, 312:431–433. 57, 60, 61, 63, 98, 108, 113, 119

[10] Beichelt, F. (2006). *Stochastic Process in Science, Engineering and Finance.* Taylor Francis Group. 112

[11] Benkman, C. W. (1999). The selection mosiac and diversifying coevolution between crossbills and lodgepole pine. *American Naturalist*, 153:S75–S91. 8

[12] Bluthgen, N., Menzel, F., Hovestadt, T., Fiala, B., and Bluthgen, N. (2007). Specialization, constraints, and conflicting interests in mutualistic networks. *Current Biology*, 17:341–346. 58, 66, 88

[13] Brodie, E. D., I. and Brodie, E., J. (1999). Predator-prey arms races. *BioScience*, 49:557–568. 8

[14] Bronstein, J. L. (1988). Mutualism, antagonism, and fig-pollinator interaction. *Ecology*, 69:1298–1302. 8

[15] Bronstein, J. L. and Hossaert-McKey, M. (1996). Variation in reproductive success within a subtropical fig/pollinator mutualism. *Ecology*, 23:433–446. 8

[16] Buckling, A. and Rainey, P. B. (2002). The role of parasites in sympatric and allopartic host diversification. *Nature*, 420:496–499. 1, 8

[17] Campbell, C., Yang, S., Albert, R., and Shea, K. (2011). A network model for plant-pollinator community assembly. *Proceedings of the National Academy of Science*, 108:197–202. 64, 117, 119

[18] Cattin, M., Bersier, L., Banasek-Richter, C., Baltensperger, R., and Gabriel, J. (2004). Phylogenetic constraints and adaptation explain food-web structure. *Nature*, 427:835–839. 55, 57, 63, 68

[19] Christensen, K., di Collobiano, S. A., Hall, M., and Jensen, H. J. (2002). Tangled nature: a model of evolutionary ecology. *Journal of Theoretical Biology*, 216:73–84. 8, 66

[20] Clauset, A., Moore, C., and Newman, M. E. J. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453:98–101. 57

[21] Cook, J. and Rasplus, J. (2003). Mututalists with attitude. *Trends in Ecology and Evolution*, 18:241–248. 10

[22] Doebeli, M. and Dieckmann, U. (2000). Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *American Naturalist*, 156:S77–S101. 6, 9, 10

[23] Dormann, C. F., Gruber, B., M., D., Fruend, J., Iriondo, J., Strauss, R., Vazquez, D., Bluthgen, N., Clauset, A., Strauss, R., and Rodriguez-Girones, M. (2011). bipartite: Visualising bipartite networks and calculating some (ecological) indices. 119

[24] Drickamer, L., Vessey, S., and Jakob, E. M. (2002). *Animal Behavior*. McGraw Hill. 2

[25] Emerson, B. C. and Kolm, N. (2005). Species diversity can drive speciation. *Nature*, 434:1015–1017. 8

[26] Erlich, P. R. and Raven, P. (1964). Butterflies and plants: a study in coevolution. *Evolution*, 18:586–608. 8, 55

[27] Fonseca, C. R. and Ganade, G. (1996). Asymmetries, compartments and null interactions in an amazonian ant-plant community. *Journal of Animal Ecology*, 65:339–347. 60, 61, 63, 64

[28] Fortuna, M. A. and Bascompte, J. (2006). Habitat loss and the structure of plant–animal mutualistic networks. *Ecology Letters*, 9:278–283. 56

[29] Frigg, R. and Werndl, C. (2011). Entropy - a guide for the perplexed. In Probabilities in Physics. Oxford University Press. 66, 116, 117

[30] Gavrilets, S. (2004). *Fitness landscapes and the origin of species*. Princeton University Press. 1, 4, 5, 6, 27, 38

[31] Gavrilets, S. (2006). The maynard smith model of sympatric speciation. *Journal of Theoretical Biology*, 239:172–182. 132

[32] Gavrilets, S. and Hastings, A. (1998). Coevolutionary chase in two-species systems with applications to mimicry. *Journal of Theoretical Biology*, 191:415–427. 9

[33] Gavrilets, S. and Vose, A. (2005). Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Science*, 102:18040–18045. 2

[34] Gibson, R. H., Nelson, I. L., Hamlett, B. J., and Memmott, J. (2006). Pollinator webs, plant communities and the conservation of rare plants: arable weeds as a case study. *Journal of Applied Ecology*, 43:246–257. 62

[35] Gilarranz, L. J., Pastor, J. M., and Galeano, J. (2012). The architecture of weighted mutualistic networks. *Oikos*, 121:1154–1162. 62

[36] Gomulkiewicz, R., Nuismer, S. L., and Thompson, J. N. (2003). Coevolution in variable mutualisms. *American Naturalist*, 162:S81 – S93. 9, 10, 13, 37

[37] Grant, P. R. and Grant, B. R. (2008). *How and Why Species Multiply*. Princeton University Press. 1

[38] Grinstead, C. M. and Snell, J. L. (1997). *Introduction to Probability*. American Mathematical Society. 112

[39] Guillaume, J. and Latapy, M. (2004). Bipartite structure of all complex networks. *Information Processing Letters*, 90:215–222. 56

[40] Guimarães, P. R., Rico-Gray, V., Oliveira, P. S., Izzo, T. J., dos Reis, S. F., and Thompson, J. N. (2007). Interaction intimacy affects structure and coevolutionary dynamics in mutualistic networks. *Current Biology*, 17:1797–1803. 65, 66

[41] Guimarães Jr., P. R. adn Machado, G., de Aguiar, M. A. M., Jordano, P., Bascompte, J., Pinheiro, A., and Furtado dos Reis, S. (2007). Build-up mechanisms determining the topology of mutualistic networks. *Journal of Theoretical Biology*, 249:181–189. 58, 65, 102

[42] Hartl, D. L. and Clark, A. G. (1997). *Principles of Population Genetics*. Sinauer Associates, Inc. 2, 3, 4, 5, 133

[43] Herre, E. A. (1996). An overview of studies on a community of panamanian figs. *Journal of Biogeography*, 23:593–607. 10

[44] Herre, E. A., Jander, K. C., and Machado, C. A. (2008). Evolutionary ecology of figs and their associates: Recent progress and outstanding puzzles. *Annual Review of Ecological Evolutionary Systematics*, 39:439 – 458. 8, 10

[45] Hoberg, E. P., Brooks, D. R., and Seigel-Causey, D. (1997). Host-parasite co-speciation: History, principles, and prospects. In Clayton, D. H. and Moore, J., editors, *Host-parasite evolution: General principles and avian models*, chapter Host-parasite co-speciation: History, principles, and prospects, pages 212–235. Oxford University Press. 8

[46] Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press. 8

[47] Ings, T. C., Montoya, J. M., Bascompte, J., Bluthgen, N., Brown, L., Dormann, C. F., Edwards, F., Figueroa, D., Jacob, U., Jones, J. I., Lauridsen, R. B., Ledger, M. E., Lewis, H. M., Olesen, J. M., Frank van Veen, F. J., and Woodward, G. (2009). Ecological networks - beyond food webs. *Journal of Animal Ecology*, 78:253–269. 8, 11, 55, 56, 57, 58, 60, 61, 62, 63, 66, 67, 68, 118

[48] Inoue, T., Kato, M., Kakutani, T., Suka, T., and Itino, T. (1990). Insect-flower relationship in the temperate decidouous forest of kibue, kyoto: an overview of the flowering phenology and the seasonal pattern of insect visits. *Contributions from the Biological Laboratory, Kyoto University*, 27:377–463. 60, 103

[49] Jackson, A. P., Machado, C. A., Robbins, N., and Herre, E. A. (2008). Multi-locus phylogenetic analysis of neotropical figs does not support co-speciation with pollinators: The importance of systematic scale in fig/wasp cophylogenetic studies. *Symbiosis*, 45:57 – 72. 11, 67, 117

[50] Janzen, D. H. (1980). When is it coevolution? *Evolution*, 34:611–612. 11, 67

[51] Jayakar, S. D. (1970). A mathematical model for interaction of gene frequencies in a parasite and its host. *Journal of Theoretical Biology*, 1:140–164. 8

[52] Jordano, P. (1987). Patterns of mutualistic interactions in pollination and seed dispersal: Connectance, dependence asymmetries, and coevolution. *The American Naturalist*, 129:657–677. xviii, 56, 57, 58, 60, 70, 89, 92, 97, 98, 100, 107, 113

[53] Jordano, P., Bascompte, J., and Olesen, J. M. (2003). Invariant properties in coevolutionary networks of plant–animal interactions. *Ecology Letters*, 6:69–81. 56, 58, 59, 63, 64, 65, 104, 113

[54] Jousselin, E., Rasplus, J., and Kjellberg, F. (2003). Convergence and coevolution in a mutualism: Evidence from a molecular phylogeny of *Ficus*. *Evolution*, 57:1255–1269. 68

[55] Jousselin, E., van Noort, S., Rasplus, J. Y., and Greeff, J. M. (2006). Patterns of diversification of afrotropical otiteselline fig wasps: phylogenetic study reveals a double radiation across host figs and conservatism of host association. *Journal of Evolutionary Biology*, 19(1):253–266. 117

[56] Karlin, S. and Fledman, M. W. (1970). Linkage and selection: two-locus symmetric viability model. *Theoretical Population Biology*, 1:39–71. 132

[57] Kerdelhuè, C. (1997). Active pollination of *Ficus sur* by two sympatric fig wasp species in west africa. *Biotropica*, 29:69–75. 38

[58] Kiester, A. R., Lande, R., and Schemske, D. W. (1984). Models of coevolution in plants and their pollinators. *American Naturalist*, 124:220–243. 9, 68, 69, 70, 71, 74, 76, 77, 83, 87

[59] Kopp, M. and Gavrilets, S. (2006). Multilocus genetics and the coevolution of quantitative traits. *Evolution*, 60:1321–1336. 8, 9

[60] Kot, M. (2001). *Elements of Mathematical Ecology*. Cambridge University Press. 133

[61] Levene, H. (1953). Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*, 87(836):331. 10

[62] Lewinsohn, T., Inácio, P., Jordano, P., Bascompte, J., and Olesen, J. M. (2006). Structure in plant-animal assemblages. *Oikos*, 113:174–184. 61, 63

[63] Lind, P. G., González, M. C., and Herrmann, H. J. (2005). Cycles and clustering in bipartite networks. *Physical Review*, 72:056127. 56, 63

[64] Lockvvood, J. L., Powell, R. D., Nott, M. P., and L., P. S. (1997). Assembling ecological communities in time and space. *Oikos*, 80:549–553. 64

[65] Machado, C. A., Herre, E. A., McCafferty, S., and Bermingham, E. (1996). Molecular phylogenies of fig pollinating an non-pollinating wasps and the implications for the origin and evolution of the fig-fig wasp mutualism. *Journal of Biogeography*, 23:531–542. 10, 111

[66] Machado, C. A., Robbins, N., Gilbert, T. P., and Herre, E. A. (2005). Critical review of host-specificity and its coevolutionary implications in the fig/fig-wasp mutualism. *Proceedings of the National Academy of Science*, 102:6558–6565. 111

[67] May (1973). Qualitative stability in model ecosystems. *Ecology*, 54:638–641. 57, 60

[68] May, R. (1972). Will a large complex system be stable? *Nature*, 238:413–414. 57

[69] Medan, D., Perazzo, R. P. J., Devoto, M., Burgos, E., Zimmermann, M., and Delbue, A. M. (2007). Analysis and assembling of network structure in mutualistic systems. *Journal of Theoretical Biology*, 246:510–521. 62

[70] Memmott, J., Waser, N. M., and Price, M. V. (2004). Tolerance of pollination networks to species extinctions. *Proc. R. Soc. Lond. B*, 271:2605–2611. 64

[71] Michaloud, G., Carriére, S., and Kobbi, M. (1996). Exceptions to the one:one relationship between african fig trees and other fig wasp pollinators: possible evolutionary scenarios. *Journal of Biogeography*, 23:513–520. 9, 10

[72] Morales, J. M. and Vazquez, D. P. (2008). The effect of space in plant-animal mutualistic networks: insights from a simulation study. *Oikos*, 117:1362–1370. 60

[73] Newman, M. E. J. (2012). Communities, modules and large-scale structure in networks. *Nature Physics*, 8:25–31. 63

[74] Olesen, J. M., Bascompte, J., Dupont, Y. L., and Jordano, P. (2006). The smallest of all worlds: Pollination networks. *Journal of Theoretical Biology*, 240:270–276. 62

[75] Olesen, J. M., Bascompte, J., Dupont, Y. L., and Jordano, P. (2007). The modularity of pollination networks. *Proceedings of the National Academy of Science*, 104(50):19891–19896. 57, 62

[76] Olesen, J. M., Bascompte, J., Elberling, H., and Jordano, P. (2008). Temporal dynamics in a pollination network. *Ecology*, 89:1573–1582. 56, 58, 60

[77] Page, R. D. M., editor (2003). *Tangled Trees: Phylogeny, Cospeciation, and Coevolution.* University of Chicago Press. 6, 7, 8, 11, 67, 94

[78] Pellmyr, O. and Segraves, K. A. (2003). Pollinator divergence within an obligate mutualism: Two yucca moth species (lepidoptera; prodoxidae: *Tegeticlua*) on the joshua tree (*Yucca breevifolia*; agavaceae). *Annals of the Entomological Society*, 96:716–722. 8

[79] Pires, M. M., Prado, P. I., and Guimarães, P. R. (2011). Do food web models reproduce the structure of mutualistic networks? *PLoS ONE*, 6:e27280. 65

[80] Pocock, M. J. O. (2012). The robustness and restoration of a network of ecological networks. *Science*, 335:973. 56, 64, 66, 119

[81] Price, D. J. (1965). Networks of scientific papers. *Science*, 149:510–515. 59

[82] Rezende, E. L., Jordano, P., and Bascompte, J. (2007a). Effects of phenotypic complementarity and phylogeny on the nested structure of mutualistic networks. *Oikos*, 116:1919–1929. 63, 65

[83] Rezende, E. L., Lavabre, J. E., Guimarães, P. R., Jordano, P., and Bascompte, J. (2007b). Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature*, 448:925–928. 11, 55, 57, 64, 67

[84] Rice, S. (2004). *Evolutionary Theory.* Sinauer Associates, Inc. 3

[85] Saavedra, S., Reed-Tsochas, F., and Uzzi, B. (2009). A simple model of bipartite cooperation for ecological and organizational networks. *Nature*, 457:463–466. 65, 117, 119

[86] Santamaria, L. and Rodriguez-Girones, M. A. (2007). Linkage rules for plant-pollinator networks: trait complementarity or exploitation barriers? *PLoS Biology*, 5:e31. 61, 65

[87] Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42:3–4. 59

[88] Stang, M., Klinkhamer, P., and van der Meijden, E. (2007). Asymmetric specialization and extinction risk in plant-flower visitor webs: A matter ofmorphology or abundance? *Oecologia*, 151(3):442–453. 59, 60, 62, 64

[89] Stang, M., Klinkhamer, P. G. L., Waser, N. M., Stang, I., and van der Meijden, E. (2009). Size-specific interaction patterns and size matching in a plant–pollinator interaction web. *Annals of Botony*, 103:1459–1469. 59, 64, 68, 71

[90] Strogatz, S. H. (1994). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering*. Perseus Publishing, Cambridge, MA. 31

[91] Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410:268–276. 56

[92] Thompson, J. N. (1994). *The coevolutionary process*. University of Chicago Press. 55, 63, 102

[93] Thompson, J. N. (2005). *The geographic mosiac of coevolution*. The University of Chicago Press. 1, 2, 7, 9, 10, 11, 37, 38, 67

[94] Vazquez, D. P., Bluthgen, N., Cagnolo, L., and Chacoff, N. P. (2009). Uniting pattern and process in plant–animal mutualistic networks: a review. *Annals of Botony*, 103:1445–1457. 56, 58, 65, 66

[95] Vazquez, D. P., Morris, W. F., and Jordano, P. (2005). Interaction frequency as a surrogate for the total effect of animal mutualists on plants. *Ecology Letters*, 8:1088–1094. 56, 60

[96] Wade, M. J. (2007). The co-evolutionary genetics of ecological communities. *Nature Reviews Genetics*, 8:185 – 195. 10, 40

[97] Waser, N. M. and Ollerton, J., editors (2006). *Plant-Pollinator Interactions: From Specialization to Generalization.* University of Chicago Press. 8

[98] Weiblen, G. D. (2002). How to be a fig wasp. *Annual Review of Entomology*, 47:299–330. 10, 11, 42

[99] Weiblen, G. D. (2004). Correlated evolution in fig pollination. *Syst. Biol.*, 53:128–139. 68

[100] Weiblen, G. D. and Bush, G. L. (2002). Speciation in fig pollinators and parasites. *Molecular Ecology*, 11:1573–1578. 10, 11, 53, 67

[101] Weitz, J. S., Benfrey, P. N., and Wingreen, N. S. (2007). Evolution, interactions, and biological networks. *PLoS Biology*, 5:e11. 63

[102] Williams, R. J. (2011). Biology, methodology or chance? the degree distributions of bipartite ecological networks. *PLoS ONE*, 6:e17645. 66, 116

[103] Yokoyama, J. (2003). Cospeciation of figs and fig-wasps: a case study of endemic species pairs in the ogasawara islands. *Population Ecology*, 45:249 – 256. 53

Appendix A

A.1 Analysis of Equilibria for H-GM Model

In order to more easily find equilibria analytically and characterize the stability, we apply a linear transformation to the insect frequencies [31, 56]. We let

$$u_1 = x_1 + x_2 + x_3 + x_4 \tag{A.1}$$

$$u_2 = x_1 - x_2 + x_3 - x_4 \tag{A.2}$$

$$u_3 = x_1 + x_2 - x_3 - x_4 \tag{A.3}$$

$$u_4 = x_1 - x_2 - x_3 + x_4 \tag{A.4}$$

be the new variables that describe the frequencies of the insect genotypes. Because $u_1 = x_1 + x_2 + x_3 + x_4 = 1$, we can immediately replace u_1 in Equation A.1 by 1 and reduce our system from four variables, to only three, u_2, u_3, and u_4. Because x_1 is the frequency of genotype **AB** and x_2 is the frequency of genotype **Ab**, $x_1 + x_2$ is the frequency of the allele **A**. This is a helpful observation because when we interpret our results, we can talk about the evolution at each loci. The frequency of allele **A** will be denoted by p_1, and the frequency of allele **B** will be denoted by p_2 . Notice that $(1 + u_3)/2 = x_1 + x_2 = p_1$ so u_3 is directly related to the frequency of allele **A**, and $(1 + u_2)/2 = x_1 + x_2 = p_1$ so u_2 is directly related to the frequency of allele **B**. Another helpful measure of genetic structure is linkage disequilibrium, defined as

$D = x_1x_4 - x_2x_3$. This measure indicates whether or not there is any bias toward particular genotypes [42]. In terms of the new variables, $D = (u_4 - u_2u_3)/4$.

The genetic composition of the plant population, in Equation 2.14, is described by y_1 and y_2. Since $y_1 + y_2 = 1$ we can reduce the equations describing the plant allele frequency dynamics down to only one variable by the substitution $y_2 = 1 - y_1$. Therefore, our biological model system depends on only four dynamic variables: u_2, u_3, u_4, and y_1, which we have renamed vector z for the purposes of finding equilibria and determining their stability.

The full model is a non-linear discrete dynamical system. For an equilibrium to be stable, it must satisfy the condition that $|\lambda_i| < 1$ for all eigenvalues, λ_i, otherwise the equilibrium is unstable [60].

Recall the range of parameters: $\epsilon \in [0, 1]$ is the bias of an insect towards a particular plant choice, $r \in [0, 1/2]$ is the probability of insect loci recombination, $s \in [0, 1]$ is the selection coefficient, and $\beta \in [0, 1]$ is the proportion of plants is replaced each generation.

For the analysis below, when the result yields an inconclusive eigenvalue of 1 or -1, these special cases will be examined using numerical simulations.

A.1.1 $z = [1, 1, 1, 1]$, AB and C fixed, stable

Eigenvalues are

$$\begin{aligned}
\lambda_1 &= \frac{1 - \epsilon}{(1 + \epsilon)^2}, \\
\lambda_2 &= 1 - 2\frac{\beta\epsilon}{1 + \epsilon}, \\
\lambda_3 &= 1 - s, \text{ and} \\
\lambda_4 &= \frac{(1 - \epsilon)(1 - s)(1 - r)}{(1 + \epsilon)^2}.
\end{aligned}$$

$\lambda_1 \in (-1, 1)$ because each term is positive, the numerator is less than 1, and the denominator is greater than 1 ($\epsilon \neq 0$). When $\epsilon = 0$, $\lambda_1 = 1$. When ϵ or β equal

0, $\lambda_2 = 1$. For all other biologically realistic parameter values, $\lambda_2 \in (-1, 1)$ because $2\frac{\beta\epsilon}{1+\epsilon} \in (0, 1]$. $\lambda_3 \in (-1, 1)$ for $s \neq 0$, in which case, $\lambda_3 = 1$. $\lambda_4 = 1$ when ϵ, s, and r all equal 0. For all other biologically realistic parameter values, each term is positive, the numerator is less than 1, and the denominator is greater than 1, so $|\lambda_3| < 1$. Because $|\lambda_i| < 1$ for $i = 1, 2, 3, 4$, this equilibria is **stable**.

A.1.2 $\mathbf{z} = [-1, 1, -1, 1]$, Ab and C fixed, unstable

Eigenvalues are

$$\lambda_1 = \frac{1+\epsilon}{(1-\epsilon)^2},$$
$$\lambda_2 = 1 + 2\frac{\beta\epsilon}{1-\epsilon},$$
$$\lambda_3 = 1 - s, \text{ and}$$
$$\lambda_4 = \frac{(1+\epsilon)(1-s)(1-r)}{(1-\epsilon)^2}.$$

Note that $\lambda_2 > 1$, for $\beta, \epsilon \neq 0$. This means that this equilibrium will be **unstable**.

A.1.3 $\mathbf{z} = [1, -1, -1, 1]$, aB and C fixed, unstable

Eigenvalues are

$$\lambda_1 = \frac{1-\epsilon}{(1+\epsilon)^2},$$
$$\lambda_2 = 1 - 2\frac{\beta\epsilon}{1+\epsilon},$$
$$\lambda_3 = (1-s)^{-1}, \text{ and}$$
$$\lambda_4 = \frac{(1-\epsilon)(1-r)}{(1+\epsilon)^2(1-s)}.$$

Since $\lambda_3 > 1$ for $0 < s < 1$ this equilibrium is **unstable**.

A.1.4 $z = [-1, -1, -1, 1]$, ab and C fixed, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= \frac{1 + \epsilon}{(1 - \epsilon)^2}, \\
\lambda_2 &= 1 + 2\frac{\beta\epsilon}{1 - \epsilon}, \\
\lambda_3 &= (1 - s)^{-1}, \text{ and} \\
\lambda_4 &= \frac{(1 + \epsilon)(1 - r)}{(1 - \epsilon)^2(1 - s)}.
\end{aligned}
$$

Thus, $\lambda_3 = (1 - s)^{-1} > 1$ for $s \neq 0$, therefore this equilibrium is **unstable**.

A.1.5 $z = [-\frac{3+\epsilon^2}{4\epsilon}, 1, -\frac{3+\epsilon^2}{4\epsilon}, 1]$, biologically unrealistic

This equilibria corresponds to the insect frequencies

$$
x = \left[\begin{array}{cccc} \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, & \frac{-(1-\epsilon)(3-\epsilon)}{8\epsilon}, & 0, & 0 \end{array} \right].
$$

One can see that $x_2 = -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon} < 0$ for $0 < \epsilon < 1$. That means that **this equilibrium is not biologically meaningful**. When $\epsilon = 0$, this equilibria does not exist, and when $\epsilon = 1$, this is the same as the insect frequency genotypes $x = [1, 0, 0, 0]$.

A.1.6 $z = [-\frac{3+\epsilon^2}{4\epsilon}, -1, \frac{3+\epsilon^2}{4\epsilon}, 1]$, biologically unrealistic

This equilibrium corresponds to the insect genotype frequencies

$$
x = \left[\begin{array}{cccc} 0, & 0, & \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, & \frac{-(1-\epsilon)(3-\epsilon)}{8\epsilon} \end{array} \right].
$$

One can see that $x_4 = -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon} < 0$ for $0 < \epsilon < 1$. That means that **this equilibrium is not biologically meaningful**. When $\epsilon = 0$, this equilibria does not exist, and when $\epsilon = 1$, this is the same as the insect frequency genotypes $x = [0, 0, 1, 0]$.

There are two additional equilibria in which plant **C** is fixed. However, the equilibria solutions for the insects are too long to write out. Results of testing at several parameter sets indicate that they are not biologically realistic.

A.1.7 $z = [1, 1, 1, 0]$, AB and c fixed, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 + 2\frac{\beta\,\epsilon}{1 - \epsilon}, \\
\lambda_2 &= \frac{1 + \epsilon}{(1 - \epsilon)^2}, \\
\lambda_3 &= (1 - s)^{-1}, \text{ and} \\
\lambda_4 &= \frac{(1 + \epsilon)\,(1 - r)}{(1 - \epsilon)^2\,(1 - s)}.
\end{aligned}
$$

Since $\lambda_3 = (1 - s)^{-1} > 1$, for $0 < s < 1$, this equilibrium is **unstable**.

A.1.8 $z = [-1, 1, -1, 0]$, Ab and c fixed, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 - 2\frac{\beta\epsilon}{1 + \epsilon}, \\
\lambda_2 &= \frac{1 - \epsilon}{(1 + \epsilon)^2}, \\
\lambda_3 &= (1 - s)^{-1}, \text{ and} \\
\lambda_4 &= \frac{(1 - \epsilon)\,(1 - r)}{(1 + \epsilon)^2\,(1 - s)}.
\end{aligned}
$$

Since $\lambda_3 >$ for $0 < s < 1$, this equilibrium is **unstable**.

A.1.9 $\mathbf{z} = [1, -1, -1, 0]$, aB and c fixed, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 + 2\frac{\beta\epsilon}{1 - \epsilon}, \\
\lambda_2 &= \frac{1 + \epsilon}{(1 - \epsilon)^2}, \\
\lambda_3 &= 1 - s, \text{ and} \\
\lambda_4 &= \frac{(1 + \epsilon)(1 - s)(1 - r)}{(1 - \epsilon)^2}.
\end{aligned}
$$

Note $\lambda_2 > 1$ for $0 < \epsilon < 1$ because the numerator is greater than 1 and the denominator is less than 1. Therefore, this equilibrium is **unstable**.

A.1.10 $\mathbf{z} = [-1, -1, -1, 0]$, ab and c fixed, stable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 - 2\frac{\beta\epsilon}{1 + \epsilon}, \\
\lambda_2 &= \frac{1 - \epsilon}{(1 + \epsilon)^2}, \\
\lambda_3 &= 1 - s, \text{ and} \\
\lambda_4 &= \frac{(1 - \epsilon)(1 - s)(1 - r)}{(1 + \epsilon)^2}.
\end{aligned}
$$

When ϵ or β equal 0, $\lambda_1 = 1$. For all other biologically realistic parameter values, $\lambda_1 \in (-1, 1)$ because $2\frac{\beta\epsilon}{1+\epsilon} \in (0, 1]$. $\lambda_2 \in (-1, 1)$ because each term is positive, the numerator is less than 1, and the denominator is greater than 1 ($\epsilon \neq 0$). When $\epsilon = 0$, $\lambda_1 = 1$. $\lambda_3 \in (-1, 1)$ for $s \neq 0$, in which case, $\lambda_3 = 1$. $\lambda_4 = \frac{(1-\epsilon)(1-s)(1-r)}{(1+\epsilon)^2}$ will be 1 when ϵ, s, and r all equal 0. For all other biologically realistic parameter values, each term is positive, the numerator is less than 1, and the denominator is greater than 1, so $|\lambda_3| < 1$. Because $|\lambda_i| < 1$ for $i = 1, 2, 3, 4$, this equilibria is **stable**.

137

A.1.11 $\;\; \mathbf{z} = [\frac{\mathbf{3}+\boldsymbol{\epsilon}^2}{\mathbf{4}\boldsymbol{\epsilon}}, \mathbf{1}, \frac{\mathbf{3}+\boldsymbol{\epsilon}^2}{\mathbf{4}\boldsymbol{\epsilon}}, \mathbf{0}]$, biologically unrealistic

This equilibria corresponds to the insect frequencies

$$x = \left[\;\; \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, \;\; -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon}, \;\; 0, \;\; 0 \;\; \right].$$

One can see that $x_2 = -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon} < 0$ for $0 < \epsilon < 1$. That means that **this equilibrium is not biologically meaningful**. When $\epsilon = 0$, this equilibria does not exist, and when $\epsilon = 1$, this is the same as the insect frequency genotypes $x = [1, 0, 0, 0]$.

A.1.12 $\;\; \mathbf{z} = [\frac{\mathbf{3}+\boldsymbol{\epsilon}^2}{\mathbf{4}\boldsymbol{\epsilon}}, -\mathbf{1}, -\frac{\mathbf{3}+\boldsymbol{\epsilon}^2}{\mathbf{4}\boldsymbol{\epsilon}}, \mathbf{0}]$, biologically unrealistic

This equilibrium corresponds to the insect genotype frequencies

$$x = \left[\;\; 0, \;\; 0, \;\; \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon} \;\; -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon} \;\; \right].$$

One can see that $x_4 = -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon} < 0$ for $0 < \epsilon < 1$. That means that **this equilibrium is not biologically meaningful**. When $\epsilon = 0$, this equilibria does not exist, and when $\epsilon = 1$, this is the same as the insect frequency genotypes $x = [0, 0, 1, 0]$.

There are two additional equilibria in which plant **c** is fixed. However, the equilibria solutions for the insects are too long to write out. Results of testing at several parameter sets indicate that they are not biologically realistic.

A.1.13 $\mathbf{z} = [0, 1, 0, 1/2]$, half Ab, half AB, half C, half c, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 - \frac{1}{2}\beta\epsilon^2 + \frac{1}{2}\sqrt{\beta\epsilon^2(6 + \epsilon^2\beta - 2\epsilon^4)}, \\
\lambda_2 &= 1 - \frac{1}{2}\beta\epsilon^2 - \frac{1}{2}\sqrt{\beta\epsilon^2(6 + \epsilon^2\beta - 2\epsilon^4)}, \\
\lambda_3 &= \frac{1}{4(1-s)}\left((s^2 - 2s + 2)(2 - r + r\epsilon^2)\right. \\
&\qquad \left. -\sqrt{\epsilon^2 s^2(s^2 - 4s + 8)(1 + \epsilon^2)^2 + r^2(1 - \epsilon^2)(s^2 - 2s + 2)^2}\right), \text{ and} \\
\lambda_4 &= \frac{1}{4(1-s)}\left((s^2 - 2s + 2)(2 - r + r\epsilon^2)\right. \\
&\qquad \left. +\sqrt{\epsilon^2 s^2(s^2 - 4s + 8)(1 + \epsilon^2)^2 + r^2(1 - \epsilon^2)(s^2 - 2s + 2)^2}\right).
\end{aligned}
$$

Note that $\lambda_1 > 1$ if

$$
\beta\epsilon^2 < \sqrt{\beta\epsilon^2(6 + \epsilon^2\beta - 2\epsilon^4)},
$$

or

$$
\beta^2\epsilon^4 < \beta\epsilon^2(6 + \epsilon^2\beta - 2\epsilon^4),
$$

which is true, because $0 < \beta\epsilon^2(6 - 2\epsilon^4)$. Therefore, this equilibria is **unstable**.

A.1.14   z $= [0, -1, 0, 1/2]$, half aB, half ab, half C, half c, unstable

Eigenvalues are

$$\lambda_1 = 1 - \frac{1}{2}\beta\epsilon^2 + \frac{1}{2}\sqrt{\beta\epsilon^2(6 + \epsilon^2\beta - 2\epsilon^4)},$$

$$\lambda_2 = 1 - \frac{1}{2}\beta\epsilon^2 - \frac{1}{2}\sqrt{\beta\epsilon^2(6 + \epsilon^2\beta - 2\epsilon^4)},$$

$$\lambda_3 = \frac{1}{4(1-s)}\left((s^2 - 2s + 2)(2 - r + r\epsilon^2)\right.$$
$$\left. - \sqrt{\epsilon^2 s^2(s^2 - 4s + 8)(1 + \epsilon^2)^2 + r^2(1 - \epsilon^2)(s^2 - 2s + 2)^2}\right), \text{ and}$$

$$\lambda_4 = \frac{1}{4(1-s)}\left((s^2 - 2s + 2)(2 - r + r\epsilon^2)\right.$$
$$\left. + \sqrt{\epsilon^2 s^2(s^2 - 4s + 8)(1 + \epsilon^2)^2 + r^2(1 - \epsilon^2)(s^2 - 2s + 2)^2}\right).$$

Note that $\lambda_1 > 1$ if $\beta\epsilon^2 < \sqrt{\beta\epsilon^2(6 + \epsilon^2\beta - 2\epsilon^4)}$, or $\beta^2\epsilon^4 < \beta\epsilon^2(6 + \epsilon^2\beta - 2\epsilon^4)$. Which is true, because $0 < \beta\epsilon^2(6 - 2\epsilon^4)$. Therefore, this equilibria is **unstable**.

A.1.15   z $= [0, 0, z_3^*, 1/2]$, half A, half B, fr(AB)=fr(ab), fr(Ab)=fr(aB), half C

This point is equivalent to the insect gene frequencies of

$$\overrightarrow{x} = \left[\begin{array}{cccc} \frac{1}{4} + \frac{1}{4}z_3^*, & \frac{1}{4} - \frac{1}{4}z_3^*, & \frac{1}{4} - \frac{1}{4}z_3^*, & \frac{1}{4} + \frac{1}{4}z_3^* \end{array}\right],$$

so note that $0 \le z_3^* \le 1$ for this equilibrium to exist and be biologically realistic. z_3^* is the solution to $z_3^2 + 2Bz_3 - 1 = 0$ that satisfies this constraint, where $B = \frac{r(2-s)(1-\epsilon^2)}{\epsilon s(\epsilon^2+1)}$. Since the candidate equilibria solutions for z_3 are $-B \pm \sqrt{B^2 + 1}$, we note that because $B < 0$ for our considered parameter space, $-B - \sqrt{B^2 + 1} < 0$, so this solution does not satisfy the constraint. However, $B < \sqrt{B^2 + 1}$ implies that $-B + \sqrt{B^2 + 1} > 0$, so in this case, we need just to ensure that $-B + \sqrt{B^2 + 1} < 1$. Solving this

constraint for B, we find that for this equilibrium to exist, $B > 0$, which is always true for the parameter range under consideration, so this equilibria always exists, and $z_3^* = -B + \sqrt{B^2 + 1}$. We denote the conjugate, biologically unrealistic solution by $\overline{z_3^*}$.

The eigenvalues at our non-trivial equilibrium can be solved for, but most are extremely unwieldy. We can easily write one of them, $\lambda_3 = \frac{-(s\epsilon(1+\epsilon^2))^2 + 4(5-4s)(1-r(1-\epsilon^2))}{[2(2-s)+z_3^*\epsilon s(1+\epsilon^2)]^2}$. λ_3 is easily shown as always positive, however it is not always greater than one. Therefore analytic results are inconclusive and we rely on numerical simulations to complement our findings.

A.2   Special Cases of the H-GM Model

Here $D = x_1 x_4 - x_2 x_3$.

Special case where $\epsilon = 0$

The stability here is determined by numerical simulations as eigenvalue analysis alone was inconclusive because some eigenvalues are 1.

A.2.1 $\mathbf{z} = \left[\mathbf{x_1}, \mathbf{1 - x_1}, \mathbf{0}, \mathbf{y_1}\right]$

If $D = 0$, All **AB** or **Ab** insects, completely adapted to dominant plant type, stable only for $y_1 < \frac{1-s}{2-s}$, unstable elsewhere.

A.2.2 $\mathbf{z} = \left[\mathbf{x_1}, \frac{\mathbf{1}}{\mathbf{s}}((\mathbf{2-s})\mathbf{y_1} - \mathbf{x_1 s} - (\mathbf{1-s})), \frac{(\mathbf{1-2y_1+y_1 s})\mathbf{x_1}}{(\mathbf{2-s})\mathbf{y_1} - (\mathbf{1-s})}, \mathbf{y_1}\right]$

If $D = 0$, completely adapted, stable manifold only for $\frac{1-s}{2-s} < y_1 < \frac{1}{2-s}$, unstable elsewhere.

A.2.3 $\mathbf{z} = [\mathbf{0}, \mathbf{0}, \mathbf{x_3}, \mathbf{y_1}]$

All **aB** or **ab** insects, completely adapted to dominant plant type, stable only for $y_1 > \frac{1}{2-s}$, unstable elsewhere. $\epsilon = 1$

A.2.4 $\mathbf{z} = [\mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1}]$, AB fixed, C fixed, D=0, stable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 0, \\
\lambda_2 &= 1 - \beta, \\
\lambda_3 &= 1 - s, \text{ and} \\
\lambda_4 &= 0.
\end{aligned}
$$

Because $|\lambda_i| < 1$ for $i = 1, 2, 3, 4$, and $s, \beta \epsilon (0, 1)$, this equilibria is **stable**.

A.2.5 $\mathbf{z} = [\mathbf{1}, -\mathbf{1}, -\mathbf{1}, \mathbf{1}]$, aB fixed, C fixed, D=0, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 0, \\
\lambda_2 &= 1 - \beta, \\
\lambda_3 &= 0, \text{ and} \\
\lambda_4 &= \frac{1}{1 - s}.
\end{aligned}
$$

Note $\lambda_4 = \frac{1}{1-s} > 1$ for all $s \in (0, 1)$, so this equilibrium is **unstable**.

A.2.6 $\ z = [-1, 1, -1, 1]$, Ab fixed, c fixed, D=0, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 0, \\
\lambda_2 &= 1 - \beta, \\
\lambda_3 &= \frac{1}{1-s}, \text{ and} \\
\lambda_4 &= 0.
\end{aligned}
$$

Since $\lambda_3 = \frac{1}{1-s} > 1$ for all $s \in (0, 1)$, this equilibrium is **unstable**.

A.2.7 $\ z = [-1, -1, 1, 1]$, ab fixed, c fixed, D=0, stable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 0, \\
\lambda_2 &= 1 - \beta, \\
\lambda_3 &= 0, \text{ and} \\
\lambda_4 &= 1 - s.
\end{aligned}
$$

Because $|\lambda_i| < 1$ for $i = 1, 2, 3, 4$, and $s, \beta \in (0, 1)$, this equilibria is **stable**.

A.2.8 $z = [0, 0, 1, \frac{1}{2}]$, half AB, half ab, half C, half c, D=0.25, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= -\frac{1}{2}\beta + 1 + \frac{1}{2}\sqrt{\beta^2 + 4\beta}, \\
\lambda_2 &= -\frac{1}{2}\beta + 1 - \frac{1}{2}\sqrt{\beta^2 + 4\beta}, \\
\lambda_3 &= 1 - s, \text{ and} \\
\lambda_4 &= 1 - s.
\end{aligned}
$$

Since $\beta < \sqrt{\beta^2 + 4\beta}$, $\lambda_1 = -\frac{1}{2}\beta + 1 + \frac{1}{2}\sqrt{\beta^2 + 4\beta} > 1$ for all $\beta \in (0,1)$, this equilibrium is **unstable**.

A.2.9 $z = [0, 0, -1, \frac{1}{2}]$, half aB, half Ab, half C, half c, D=-0.25, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 0, \\
\lambda_2 &= \frac{4}{1-\beta}, \\
\lambda_3 &= \frac{1}{1-s}, \text{ and} \\
\lambda_4 &= \frac{1}{1-s}.
\end{aligned}
$$

Since $\lambda_2 > 1$ for all $\beta \epsilon (0,1)$, this equilibrium is **unstable**.

A.2.10 $z = [0, 1, 0, \frac{1}{2}]$, half AB, half Ab, half C, half c, D=0, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= \frac{-1}{2(1-s)}\sqrt{s^4 - 4s^3 + 8s^2 - 8s + 4 + s(2-s)^2}, \\
\lambda_2 &= \frac{1}{2(1-s)}\sqrt{s^4 - 4s^3 + 8s^2 - 8s + 4 + s(2-s)^2}, \\
\lambda_3 &= -\frac{1}{2}\beta + 1 + \frac{1}{2}\sqrt{\beta^2 + 4\beta}, \text{ and} \\
\lambda_4 &= -\frac{1}{2}\beta + 1 - \frac{1}{2}\sqrt{\beta^2 + 4\beta}.
\end{aligned}
$$

Since $\beta < \sqrt{\beta^2 + 4\beta}$, $\lambda_3 => 1$ for all $\beta \in (0,1)$, so this equilibrium is **unstable**.

A.2.11 $z = [0, -1, 0, \frac{1}{2}]$, half AB, half ab, half C, half c, D=0, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= \frac{1}{(1-s)}, \\
\lambda_2 &= 1 - s, \\
\lambda_3 &= -\frac{1}{2}\beta + 1 + \frac{1}{2}\sqrt{\beta^2 + 4\beta}, \text{ and} \\
\lambda_4 &= -\frac{1}{2}\beta + 1 - \frac{1}{2}\sqrt{\beta^2 + 4\beta}.
\end{aligned}
$$

Since $\lambda_1 > 1$ for all $s \in (0,1)$, this equilibrium is **unstable**.

A.3 Analysis of Equilibria for H-GM Model with Host Frequency Dependence

Recall the range of our parameters: $\epsilon \in [0, 1]$ is the bias of an insect towards a particular plant choice, $r \in [0, 1/2]$ is the probability of insect loci recombination, $s \in [0, 1]$ is the selection coefficient, and $\beta \in [0, 1]$ is the proportion of plants is replaced each generation.

For the analysis below, when the result yields an inconclusive eigenvalue of 1 or -1, these special cases will be examined using numerical simulations.

A.3.1 z = [1, 1, 1, 1], AB and C fixed, stable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 - 2\frac{\beta\epsilon}{1 + \epsilon}, \\
\lambda_2 &= \frac{1 - \epsilon}{(1 + \epsilon)^2}, \\
\lambda_3 &= 1 - s, \text{ and} \\
\lambda_4 &= \frac{(1 - \epsilon)(1 - s)(1 - r)}{(1 + \epsilon)^2}.
\end{aligned}
$$

When ϵ or β equal 0, $\lambda_1 = 1$. For all other biologically realistic parameter values, $\lambda_1 \in (-1, 1)$ because $2\frac{\beta\epsilon}{1+\epsilon} \in (0, 1]$. $\lambda_2 \in (-1, 1)$ because each term is positive, the numerator is less than 1, and the denominator is greater than 1 ($\epsilon \neq 0$). When $\epsilon = 0$, $\lambda_2 = 1$. $\lambda_3 = 1 - s \in (-1, 1)$ for $s \neq 0$, in which case, $\lambda_3 = 1$. $\lambda_4 = 1$ when ϵ, s, and r all equal 0. For all other biologically realistic parameter values, each term is positive, the numerator is less than 1, and the denominator is greater than 1, so $|\lambda_3| < 1$. Because $|\lambda_i| < 1$ for $i = 1, 2, 3, 4$, this equilibria is **stable**.

$y_1 = 1$	C fixed	
$x = [1, 0, 0, 0]$	AB fixed	stable
$x = [0, 1, 0, 0]$	Ab fixed	unstable
$x = [0, 0, 1, 0]$	aB fixed	unstable
$x = [0, 0, 0, 1]$	ab fixed	unstable
$x = \left[\frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, \quad \frac{-(1-\epsilon)(3-\epsilon)}{8\epsilon}, \quad 0, \quad 0 \right]$	biologically unrealistic	
$x = \left[0, \quad 0, \quad \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, \quad \frac{-(1-\epsilon)(3-\epsilon)}{8\epsilon} \right]$	biologically unrealistic	
$y_1 = 0$	**c fixed**	
$x = [1, 0, 0, 0]$	AB fixed	unstable
$x = [0, 1, 0, 0]$	Ab fixed	unstable
$x = [0, 0, 1, 0]$	aB fixed	unstable
$x = [0, 0, 0, 1]$	ab fixed	stable
$x = \left[\frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, \quad \frac{-(1-\epsilon)(3-\epsilon)}{8\epsilon}, \quad 0, \quad 0 \right]$	biologically unrealistic	
$x = \left[0, \quad 0, \quad \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, \quad \frac{-(1-\epsilon)(3-\epsilon)}{8\epsilon} \right]$	biologically unrealistic	
$y_1 = \frac{x_1 - x_2 + x_3 - x_4 + \epsilon}{2\epsilon}$	**Polymorphic**	
$x = [\frac{1}{2}, \frac{1}{2}, 0, 0]$	fr(C) = 1/2	unstable
$x = [0, 0, \frac{1}{2}, \frac{1}{2}]$	fr(C) = 1/2	unstable
$x = \left[\frac{1}{4} + \frac{1}{4}z_3^*, \quad \frac{1}{4} - \frac{1}{4}z_3^*, \quad \frac{1}{4} - \frac{1}{4}z_3^*, \quad \frac{1}{4} + \frac{1}{4}z_3^* \right]$	fr(C) = 1/2	unstable**
$x = \left[\frac{1}{4} + \frac{1}{4}\bar{z}_3^*, \quad \frac{1}{4} - \frac{1}{4}\bar{z}_3^*, \quad \frac{1}{4} - \frac{1}{4}\bar{z}_3^*, \quad \frac{1}{4} + \frac{1}{4}\bar{z}_3^* \right]$	biologically unrealistic	

Table A.1: All equilibria are either always stable or always unstable for all biologically realistic parameters unless otherwise noted. **Note, numerical simulations indicate that this point seems to be unstable, however, this result is not shown analytically. z_3^* and $\bar{z}_3^*$, are $+/-$ solutions to a quadratic equation and are functions of r, s, and ϵ.

A.3.2 $z = [-1, 1, -1, 1]$, Ab and C fixed, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 + 2\frac{\beta\epsilon}{1-\epsilon}, \\
\lambda_2 &= \frac{1+\epsilon}{(1-\epsilon)^2}, \\
\lambda_3 &= 1 - s, \text{ and} \\
\lambda_4 &= \frac{(1+\epsilon)(1-s)(1-r)}{(1-\epsilon)^2}
\end{aligned}
$$

Note that $\lambda_1 > 1$, for $\beta, \epsilon \neq 0$. This means that this equilibrium will be **unstable**.

A.3.3 $z = [1, -1, -1, 1]$, aB and C fixed, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 - 2\frac{\beta\epsilon}{1+\epsilon}, \\
\lambda_2 &= \frac{1-\epsilon}{(1+\epsilon)^2}, \\
\lambda_3 &= (1-s)^{-1}, \text{ and} \\
\lambda_4 &= \frac{(1-\epsilon)(1-r)}{(1+\epsilon)^2(1-s)}.
\end{aligned}
$$

Since $\lambda_3 > 1$ for $0 < s < 1$, this equilibrium is **unstable**.

A.3.4 $z = [-1, -1, 1, 1]$, ab and C fixed, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 + 2\frac{\beta\epsilon}{1-\epsilon}, \\
\lambda_2 &= \frac{1+\epsilon}{(1-\epsilon)^2}, \\
\lambda_3 &= (1-s)^{-1}, \text{ and} \\
\lambda_4 &= \frac{(1+\epsilon)(1-r)}{(1-\epsilon)^2(1-s)}.
\end{aligned}
$$

Since $\lambda_3 = (1-s)^{-1} > 1$ for $s \neq 0$, this equilibrium is **unstable**.

A.3.5 $z = [-\frac{3+\epsilon^2}{4\epsilon}, 1, -\frac{3+\epsilon^2}{4\epsilon}, 1]$, biologically unrealistic

This equilibria corresponds to the insect frequencies

$$
x = \left[\begin{array}{cccc} \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, & \frac{-(1-\epsilon)(3-\epsilon)}{8\epsilon}, & 0, & 0 \end{array} \right].
$$

One can see that

$$
x_2 = -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon} < 0
$$

for $0 < \epsilon < 1$. That means that **this equilibrium is not biologically meaningful**. When $\epsilon = 0$, this equilibria does not exist, and when $\epsilon = 1$, this is the same as the insect frequency genotypes $x = [1, 0, 0, 0]$.

A.3.6 $z = [-\frac{3+\epsilon^2}{4\epsilon}, -1, \frac{3+\epsilon^2}{4\epsilon}, 1]$, biologically unrealistic

This equilibrium corresponds to the insect genotype frequencies

$$
x = \left[\begin{array}{cccc} 0, & 0, & \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, & \frac{-(1-\epsilon)(3-\epsilon)}{8\epsilon} \end{array} \right].
$$

One can see that

$$x_4 = -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon} < 0$$

for $0 < \epsilon < 1$. That means that **this equilibrium is not biologically meaningful**. When $\epsilon = 0$, this equilibria does not exist, and when $\epsilon = 1$, this is the same as the insect frequency genotypes $x = [0, 0, 1, 0]$.

There are two additional equilibria in which plant C is fixed. However, the equilibria solutions for the insects are extremely unwieldy. Numerical results of testing at several parameter sets indicate that they are not biologically realistic.

A.3.7 $\quad$ z $= [1, 1, 1, 0]$, AB and c fixed, unstable

Eigenvalues are

$$\lambda_1 = \frac{1+\epsilon}{(1-\epsilon)^2},$$

$$\lambda_2 = 1 + 2\frac{\beta\epsilon}{1-\epsilon},$$

$$\lambda_3 = (1-s)^{-1}, \text{ and}$$

$$\lambda_4 = \frac{(1+\epsilon)(1-r)}{(1-\epsilon)^2(1-s)}.$$

Since $\lambda_3 = (1-s)^{-1} > 1$, for $0 < s < 1$, the equilibrium will be **unstable**.

A.3.8 $\quad$ z $= [-1, 1, -1, 0]$, Ab and c fixed, unstable

Eigenvalues are

$$\lambda_1 = 1 - 2\frac{\beta\epsilon}{1+\epsilon},$$

$$\lambda_2 = \frac{1-\epsilon}{(1+\epsilon)^2},$$

$$\lambda_3 = (1-s)^{-1}, \text{ and}$$

$$\lambda_4 = \frac{(1-\epsilon)(1-r)}{(1+\epsilon)^2(1-s)}$$

Since $\lambda_{3>1}$ for $0 < s < 1$, this equilibrium will be **unstable**.

A.3.9   z $= [1, -1, -1, 0]$, aB and c fixed, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 + 2\frac{\beta\epsilon}{1-\epsilon}, \\
\lambda_2 &= \frac{1+\epsilon}{(1-\epsilon)^2}, \\
\lambda_3 &= 1 - s, \text{ and} \\
\lambda_4 &= \frac{(1+\epsilon)(1-s)(1-r)}{(1-\epsilon)^2}.
\end{aligned}
$$

Note that $\lambda_2 > 1$ for $0 < \epsilon < 1$, because the numerator is greater than 1 and the denominator is less than 1. This implies the equilibrium is **unstable**.

A.3.10   z $= [-1, -1, 1, 0]$, ab and c fixed, stable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 - 2\frac{\beta\epsilon}{1+\epsilon}, \\
\lambda_2 &= \frac{1-\epsilon}{(1+\epsilon)^2}, \\
\lambda_3 &= 1 - s, \text{ and} \\
\lambda_4 &= \frac{(1-\epsilon)(1-s)(1-r)}{(1+\epsilon)^2}.
\end{aligned}
$$

When ϵ or β equal 0, $\lambda_1 = 1 - 2\frac{\beta\epsilon}{1+\epsilon}$. For all other biologically realistic parameter values, $\lambda_1 \in (-1, 1)$ because $2\frac{\beta\epsilon}{1+\epsilon} \in (0, 1]$. $\lambda_2 \in (-1, 1)$ because each term is positive, the numerator is less than 1, and the denominator is greater than 1 ($\epsilon \neq 0$). When $\epsilon = 0$, $\lambda_2 = 1$. $\lambda_3 = 1 - s \in (-1, 1)$ for $s \neq 0$, in which case, $\lambda_3 = 1$. $\lambda_4 = \frac{(1-\epsilon)(1-s)(1-r)}{(1+\epsilon)^2}$ will be 1 when ϵ, s, and r all equal 0. For all other biologically realistic parameter

values, each term is positive, the numerator is less than 1, and the denominator is greater than 1, so $|\lambda_4| < 1$. Because $|\lambda_i| < 1$ for $i = 1, 2, 3, 4$, this equilibria is **stable**.

A.3.11 $\mathbf{z = [\frac{3+\epsilon^2}{4\epsilon}, 1, \frac{3+\epsilon^2}{4\epsilon}, 0]}$, biologically unrealistic

This equilibria corresponds to the insect frequencies

$$x = \left[\begin{array}{cccc} \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon}, & -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon}, & 0, & 0 \end{array} \right].$$

One can see that

$$x_2 = -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon} < 0$$

for $0 < \epsilon < 1$. That means that **this equilibrium is not biologically meaningful**. When $\epsilon = 0$, this equilibria does not exist, and when $\epsilon = 1$, this is the same as the insect frequency genotypes $x = [1, 0, 0, 0]$.

A.3.12 $\mathbf{z = [\frac{3+\epsilon^2}{4\epsilon}, -1, -\frac{3+\epsilon^2}{4\epsilon}, 0]}$, biologically unrealistic

This equilibrium corresponds to the insect genotype frequencies

$$x = \left[\begin{array}{cccc} 0, & 0, & \frac{(\epsilon+3)(1+\epsilon)}{8\epsilon} & -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon} \end{array} \right].$$

One can see that

$$x_4 = -\frac{(1-\epsilon)(3-\epsilon)}{8\epsilon} < 0$$

for $0 < \epsilon < 1$. That means that **this equilibrium is not biologically meaningful**. When $\epsilon = 0$, this equilibria does not exist, and when $\epsilon = 1$, this is the same as the insect frequency genotypes $x = [0, 0, 1, 0]$.

There are two additional equilibria in which plant c is fixed. However, the equilibria solutions for the insects are too unwieldy to display here. Numerical results of testing at several parameter sets indicate that they are not biologically realistic.

A.3.13 $z = [0, 1, 0, 1/2]$, half Ab, half 1/2 AB, half C, half c, unstable

Eigenvalues are

$$
\begin{aligned}
\lambda_1 &= 1 - \frac{1}{4}\epsilon^2(2\beta + 1 + \epsilon^2) \\
&\quad + \frac{\epsilon}{4}\sqrt{\epsilon^2(1 + \epsilon^2)^2 - 4\beta\epsilon^2(1 - \beta) + 12\beta(2 - \epsilon^4)}, \\
\lambda_2 &= 1 - \frac{1}{4}\epsilon^2(2\beta + 1 + \epsilon^2) \\
&\quad - \frac{\epsilon}{4}\sqrt{\epsilon^2(1 + \epsilon^2)^2 - 4\beta\epsilon^2(1 - \beta) + 12\beta(2 - \epsilon^4)}, \\
\lambda_3 &= \frac{1}{4(1 - s)}\left((s^2 - 2s + 2)(2 - r + r\epsilon^2) \right. \\
&\quad \left. - \sqrt{\epsilon^2 s^2(2 - s)^2(1 + \epsilon^2)^2 + r^2(1 - \epsilon^2)(s^2 - 2s + 2)^2}\right), \\
\lambda_4 &= \frac{1}{4(1 - s)}\left((s^2 - 2s + 2)(2 - r + r\epsilon^2) \right. \\
&\quad \left. + \sqrt{\epsilon^2 s^2(2 - s)^2(1 + \epsilon^2)^2 + r^2(1 - \epsilon^2)(s^2 - 2s + 2)^2}\right).
\end{aligned}
$$

Note $\lambda_1 > 1$ if,

$$
\epsilon(2\beta + 1 + \epsilon^2) < \sqrt{\epsilon^2(1 + \epsilon^2)^2 - 4\beta\epsilon^2(1 - \beta) + 12\beta(2 - \epsilon^4)}
$$

or

$$
\epsilon^2(2\beta + 1 + \epsilon^2)^2 < \epsilon^2(1 + \epsilon^2)^2 - 4\beta\epsilon^2(1 - \beta) + 12\beta(2 - \epsilon^4)
$$

is required. Since $0 < 8\beta(3 - 2\epsilon^4 - \epsilon^2)$ for $\beta > 0$, this condition is satisfied. Therefore, this equilibria is **unstable**.

A.3.14 $z = [0, -1, 0, 1/2]$, half aB, half ab, half C, half c, unstable

Eigenvalues are

$$\lambda_1 = 1 - \frac{1}{4}\epsilon^2(2\beta + 1 + \epsilon^2)$$
$$+ \frac{\epsilon}{4}\sqrt{\epsilon^2(1 + \epsilon^2)^2 - 4\beta\epsilon^2(1 - \beta) + 12\beta(2 - \epsilon^4)},$$

$$\lambda_2 = 1 - \frac{1}{4}\epsilon^2(2\beta + 1 + \epsilon^2)$$
$$- \frac{\epsilon}{4}\sqrt{\epsilon^2(1 + \epsilon^2)^2 - 4\beta\epsilon^2(1 - \beta) + 12\beta(2 - \epsilon^4)},$$

$$\lambda_3 = \frac{1}{4(1 - s)}\left((s^2 - 2s + 2)(2 - r + r\epsilon^2)\right.$$
$$\left. - \sqrt{\epsilon^2 s^2(2 - s)^2(1 + \epsilon^2)^2 + r^2(1 - \epsilon^2)(s^2 - 2s + 2)^2}\right),$$

$$\lambda_4 = \frac{1}{4(1 - s)}\left((s^2 - 2s + 2)(2 - r + r\epsilon^2)\right.$$
$$\left. + \sqrt{\epsilon^2 s^2(2 - s)^2(1 + \epsilon^2)^2 + r^2(1 - \epsilon^2)(s^2 - 2s + 2)^2}\right).$$

Note $\lambda_1 > 1$ if

$$\epsilon(2\beta + 1 + \epsilon^2) < \sqrt{\epsilon^2(1 + \epsilon^2)^2 - 4\beta\epsilon^2(1 - \beta) + 12\beta(2 - \epsilon^4)}$$

or

$$\epsilon^2(2\beta + 1 + \epsilon^2)^2 < \epsilon^2(1 + \epsilon^2)^2 - 4\beta\epsilon^2(1 - \beta) + 12\beta(2 - \epsilon^4).$$

This is true, because $0 < 8\beta(3 - 2\epsilon^4 - \epsilon^2)$ for $\beta > 0$. Therefore, this equilibria is **unstable**.

A.3.15 $z = [0, 0, z_3^*, 1/2]$, half A, half B, fr(AB)=fr(ab), and equal C and c

This point is equivalent to the insect gene frequencies of

$$\vec{x} = \left[\begin{array}{cccc} \frac{1}{4} + \frac{1}{4}z_3^*, & \frac{1}{4} - \frac{1}{4}z_3^*, & \frac{1}{4} - \frac{1}{4}z_3^*, & \frac{1}{4} + \frac{1}{4}z_3^* \end{array} \right],$$

so note that $0 \leq z_3^* \leq 1$ for this equilibrium to exist and be biologically realistic. z_3^* is the solution to $z_3^2 + 2Bz_3 - 1 = 0$ that satisfies this constraint, where $B = \frac{r(2-s)(1-\epsilon^2)}{\epsilon s(\epsilon^2+1)}$. Since the candidate equilibria solutions for z_3 are $-B \pm \sqrt{B^2 + 1}$, we note that because $B < 0$ for our considered parameter space, $-B - \sqrt{B^2 + 1} < 0$, so this solution does not satisfy the constraint. However, $B < \sqrt{B^2 + 1}$ implies that $-B + \sqrt{B^2 + 1} > 0$, so in this case, we need just to ensure that $-B + \sqrt{B^2 + 1} < 1$. Solving this constraint for B, we find that for this equilibrium to exist, $B > 0$, which is always true for the parameter range under consideration, so this equilibria always exists, and $z_3^* = -B + \sqrt{B^2 + 1}$. We denote the conjugate, biologically unrealistic solution by $\overline{z_3^*}$.

The eigenvalues at our non-trivial equilibrium can be solved for, but most are extremely unwieldy. The tractable one is

$$\lambda_3 = \frac{4(2-s)^2(1 - r + r\epsilon^2) - \epsilon^2 s^2 (1 + \epsilon^2)^2}{[2(2-s) + z_3^* \epsilon s(1 + \epsilon^2)]^2}.$$

λ_3 is easily shown < 1 for $0 \leq z_3^* \leq 1$, so analysis of this eigenvalue alone is inconclusive for determining the stability.

Vita

Carrie Diaz Eaton was born in 1981 in Warwick, Rhode Island, and graduated high school from Canton High School, in Canton, Mass. in 1998. She was on the Massachusetts A team of the American Regional Mathematics League when they won first place at PennState in 1998. She earned an Bachelor of Arts in Mathematics with a minor in Zoology in 2002 and a Master of Arts in Interdisciplinary Mathematics in 2004, both from the University of Maine and both with theses in computational neurobiology. In 2012, she graduated with a Doctor of Philosophy in Mathematics with a concentration in Theoretical Ecology and Evolution from University of Tennessee, Knoxville. She is now a faculty member at Unity College and lives in Hampden, ME with her husband and two children.