universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

A Two-Locus Two-Allele Migration-Selection Model

angestrebter akademischer Grad

Magistra der Naturwissenschaften
(Mag^{a}. rer. nat.)

Verfasser:	Ada Akerman
Matrikel-Nummer:	0403257
Studienrichtung:	A 405 Mathematik
Betreuer:	Ao. Univ.-Prof. Dr. Reinhard Bürger

Wien, im Jänner 2010

Abstract

In this work we investigate a population genetic model for the evolution of a quantitative trait in a geographically structured population. This trait is determined by two recombining diallelic loci. Selection on the trait is linear. Therefore, fitnesses are additive and there is no epistasis and no dominance. Moreover, we assume absence of genotype-environment interactions. For convenience, we assume most of the time that the population is distributed over two demes. Selection may act in different directions in the demes.

In Section 1.1, we introduce some important concepts and terms used in population genetics. There, we also state the scope of our model. In Section 1.2, we introduce our selection-recombination-migration model, (1.17). In Section 2, we investigate linkage equilibrium. In Section 3.1, we define the types of equilibria encountered and briefly introduce the basic concepts of stability analysis. In Section 3.2, we calculate equilibria of the model for some limiting cases and analyze their stability. Also, we find conditions for protection of alleles. In Section 4.1, we assume that selection is absent. In Section 4.2, we introduce selection as a small perturbation of the so-called weak-selection limit. In Section 5.1, we analyze the system without migration. Then, in Section 5.2, we introduce migration as a small perturbation of the system without migration.

Deutsche Zusammenfassung

In dieser Diplomarbeit untersuchen wir die Dynamik eines kontinuierlichen Merkmals, welches durch zwei diallelische Loci bestimmt ist, unter Selektion, Rekombination und Migration. Wir betrachten gerichtete Selektion. Diese ist additiv und es gibt keine Epistasis und keine Dominanz. Meistens gehen wir davon aus, dass die Population auf zwei Nischen aufgeteilt ist. Selektion kann je nach Nische in unterschiedliche Richtungen wirken.

Im Abschnitt 1.1 stellen wir wichtige populationsgenetische Begriffe und Konzepte vor. Dort treffen wir auch die Annahmen zu unserem Modell. Im Abschnitt 1.2 formulieren wir unser populationsgenetisches Modell (1.17). Aufgrund seiner Komplexizität lassen sich Ergebnisse nur in Spezialfällen explizit herleiten. Abschnitt 2 befasst sich mit

dem sogenannten Kopplungsgleichgewicht. In Abschnitt 3.1 befassen wir uns mit Gleichgewichtspunkten des Systems und deren Stabilitätsbedingungen. Dort finden wir auch Bedingungen für geschützte Allele. In Abschnitt 4.1 untersuchen wir das Modell wenn keine Selektion wirkt. Schwache Selektion wird dann im Abschnitt 4.2 als kleine Störung des sogenannten schwachen Selektionsgrenzwerts behandelt. Das Modell ohne Migration wird in Abschnitt 5.1 behandelt. In Abschnitt 5.2 wird Migration als kleine Störung des Systems ohne Migration betrachtet.

Danksagung

Ich bedanke mich bei Prof. Reinhard Bürger für seine Unterstützung und für seine Betreuung. Ein Dankeschön auch an Hannes Svardal, Hildegard Ücker und Stephan Peischl, dass sie mich in ihr sowieso schon kleines Zimmer aufgenommen haben. Letzterem danke ich auch für die hilfreichen Gespräche. Ich danke Sebastian Novak für seine Hilfe beim Programmieren.

Ich danke all meinen Freunden, die mit mir noch befreundet sind, obwohl ich kaum Zeit für sie gefunden habe. Ich danke meiner Familie für ihre Unterstützung und insbesondere für die Kaffeemaschine. Sie hat vieles erst möglich gemacht.

Meinem Freund Johannes danke ich für all seinen Optimismus und seine Unterstützung. Und dass das Leben mit ihm immer so ding ist.

Contents

Chapter 1

Basic Concepts and Models

In this chapter we give a brief overview of some important concepts used in population genetics and introduce the setting of the problem considered in this diploma thesis.

1.1 Introduction to Population Genetics

No matter whether single cell or complex multicellular organism, each cell contains the *Deoxyribonucleic Acid*, *DNA*, which is a macromolecule harboring the master plan of life. It is a double helix constituted by four types of nucleotide bases and a 2-deoxyribose-phosphate backbone, that stores the information for cell development and cell function [13, pp. 210–219].

During a complex process, called *gene expression*, regions with certain patterns of nucleic acid, so called *sequences*, are transcribed and translated by specific enzymes into peptides via certain *Ribonucleic Acids*, i.e., *mRNA*, or into biologically active Ribonucleic Acids, i.e., *rRNA*, *tRNA* or *snRNA* [6, p. 347]. They fulfill many crucial tasks in the cell (the detailed model of cell *transcription* and *translation* can be found for example in [6, pp. 324–349]). The regions of the DNA that are required for the production of RNA are called *genes*. This notion usually involves also regulatory elements of the DNA and both coding and non-coding sequences. A *locus* is the region of DNA where the gene is located, a particular sequence at the locus is called *allele*. The genetic constitution of a cell is termed its *genotype*. Sequences determining a particular gene can vary, e.g., due to mutations within the population. Therefore, we may distinguish between *diallelic*, *triallelic*, or *multiallelic* loci.

Differences in the coding sequence of a gene can lead to differences in its expression, altering the produced peptides and resulting in different *phenotypes*. This may have a

vital effect on an individual's *viability*, which is the probability to survive to reproductive age, and therefore affect the population composition. Therefore, it is important to study the allelic frequencies and to analyze whether over longer periods certain alleles will get *lost*, get *fixed*, or achieve a *polymorphic equilibrium*, i.e., coexistence of at least two alleles. A polymorphism is called *protected*, if both alleles are protected from eventual long-term loss, cf. [21]. In [15, pp. 210–212] an overview of the criteria that may lead to stable polymorphisms or to fixation of one allele are described.

A population can be formed by either of two types of cells, *procaryotic* or *eucaryotic*. Procaryotes are represented by bacteria. They reveal no cell nucleus, so the genomic DNA, which is in many cases circular, is found in the cytoplasm together with other extra-chromosomal circular DNA, namely plasmids. Usually procaryotes are *haploid*, i.e., only a single set of genomic DNA is present. Their reproduction is *asexual*, also known as *vegetative reproduction*, meaning that their genome is duplicated and the copy is inherited to the offspring. Thus, the offsprings are *clones*, i.e., they are identical copies of the parental cell. This is of course only the case if we ignore genetic forces altering the genetic information which we discuss in this chapter later on, i.e., mutations. Also, there is a possibility for bacteria to undergo *conjugation*, a form of sexual reproduction. We refer to [13, pp. 333–338] for more detailed literature.

Animals, fungi and plants are eucaryotes. Their cells posses a cell nucleus, where the DNA is stored in multiple compact units, the so called *chromosomes*. The number of chromosomes per cell depends on the species and can even differ due to malfunction within a species. Humans have a total of 46 chromosomes, consisting of 22 pairs of non-sex chromosomes, the *autosomes*, and two sex-determining chromosomes, the *genosomes*. Cells harboring pairs of chromosomes are called *diploid*. Many plants contain multiple copies of chromosomes and are therefore called *polyploid*. In this work we only consider the case that our population consists of diploid individuals.

Regarding a diallelic locus with the alleles A_1 and A_2 in a diploid organism, we encounter the following gene combinations: The genotypes A_1A_1 and A_2A_2 are called *homozygous* and A_1A_2 is called *heterozygous*. Of course one can extend this definition to more than two possible alleles: every diploid genotype where the same two alleles are present is called homozygous, and combinations of two different alleles are called heterozygous. For our purposes, the genotypes A_1A_2 and A_2A_1 need not to be distinguished, i.e., in our models there is no *position effect* [5, p. 3].

When the phenotype of the heterozygote A_1A_2 is the same as one of the homozygotes, say A_1A_1, allele A_1 is called *dominant* and A_2 is called *recessive* [5, p. 3]. An allele which prevents offspring from reaching the adult stage is called *lethal*; such individuals

are eliminated from the population [12, p. 200]. If each allele of the heterozygote A_1A_2 contributes only to a certain extent to the phenotype, we talk of *intermediate dominance*. In contrary, if the phenotypes of both alleles in A_1A_2 are able to manifest themselves independently of each other, we denote this phenomenon as *codominance* [13, pp. 47–59]. If the heterozygote has lower fitness than either homozygote, this constellation is called *heterozygous disadvantage* or *underdominance*. The opposite case is called *overdominance* [12, p. 286]. If more than one gene is responsible for a trait, we speak of *polygenes*. If one gene contributes to more than one trait, we speak of *pleiotropy*. *Epistasis* is encountered when a gene influences the expression of a gene at a different locus [6, p. 1332].

Eucaryotic chromosomes contain hundreds of linearly arranged genes, which are often inherited jointly to the offspring, i.e., they form *linkage groups*. Nevertheless during sexual reproduction, one can observe variation of genes along the chromosome. This is generated by a process called *recombination* or *crossing-over*. We will be concerned with recombination further below.

The basic principle of *sexual reproduction* is that considering the chromosomal pairs of the offspring, one member of each pair has been inherited from the maternal parent, the other from the paternal [5, p. 3]. Each of the parental individuals has to undergo *meiosis*, a complex process generating *gametes*, cells with only a simple set of parental chromosomes, i.e., they are haploid. The separation of the paired alleles from one other and their distribution to different gametes is called *segregation* [13, pp. 79–95]. By the fusion of two gametes a *zygote* is formed, which contains a complete diploid chromosome set. In this work, if the zygote reaches adultness, it represents a new adult individual, which again has the ability to originate gametes.

A population is said to be *monoecious* if every individual has both male and female sexual organs and, therefore, can be described by one set of genotypic frequencies. Otherwise, it is *dioecious* [6, p. 1330, p. 1340]. In this work we do not distinguish between the two sexes and therefore we assume that our population is constituted of monoecious individuals.

As already mentioned above, the variability among genotypes is not only caused by the pairing of homologous maternal and paternal chromosomes, but also by the process of recombination. It is a complex process and many models exist ([13, pp. 234–243]) but the basic pattern is that during meiosis parts of the homologous maternal and paternal chromosomes are exchanged and segregated afterwards, leading to a higher variability among gametes. In particular, there is the possibility that beneficial alleles of different ancestors are combined or that detrimental allele combinations are destroyed.

In our calculations below we will encounter the parameter c, denoting the *recombi-*

nation frequency or *recombination rate* between two loci. This is the probability that a recombination event occurs between them. The value of c usually depends on the distance between the two loci along the chromosome. Loci with $c = 0$ are called *completely linked* (and may be treated as a single locus, as they will never be separated during recombination) and loci with $c = \frac{1}{2}$ are called *unlinked*. The maximum value of $c = \frac{1}{2}$ typically occurs for loci on nonhomologous chromosomes, because then all four gametes are produced with equal frequency $\frac{1}{4}$. Thus, the recombination rate satisfies $0 \leq c \leq \frac{1}{2}$. If there is a statistical dependence among gametic frequencies of certain loci, these loci are said to be in *linkage disequilibrium* or *gametic phase disequilibrium* [5, pp. 21–22]. If the alleles of two loci are randomly associated, then the frequency of any gamete type is simply the product of the allele frequencies, i.e., they are in *linkage equilibrium*. For loci in linkage disequilibrium the gamete frequencies depart from this value. Therefore, another auxiliary measure is introduced which is called the *coefficient of linkage disequilibrium* or *linkage disequilibrium measure*, D; see (1.13).

The simplest rules during inheritance are given by *Mendelian inheritance* [13, pp. 27–37]. But not only recombination can effect the genetic constitution of an individual or a population. Other evolutionary forces that affect the genes and consequently a population are mutation, selection, and migration.

Every variation of the genetic constitution of a cell, not having its origin in the sexual propagation of the individual, is called *mutation* [13, pp. 465–521]. In this work, we will neglect the effect of mutations, i.e., we assume that different allelic types are present from the beginning. Also, mutation events are assumed to be very rare relatively to other evolutionary forces.

Selection occurs when individuals of different genotype leave different numbers of progeny because they differ in their probability to survive to reproductive age, in their mating success, or in the number of produced offsprings. The effect of selection is measured in terms of an individual's *fitness*, i.e., by the number of progeny contributed to the next generation, which we may call reproductive success [5, p. 5]. Different kinds of biological entities may vary in fitness, resulting in different levels of selection, i.e., genes (genic selection), individual organisms that differ in genotype or phenotype (individual selection), and populations within species. For our population genetic approach, the determination of the response to selection and not the process of selection itself is of interest [12, p. 251].

We will consider *quantitative traits*. These vary continuously and can be measured on a scale. Usually, they are polygenic. Three modes of selection can be distinguished for a quantitative trait: *directional*, *stabilizing* (also called *normalizing*), or *disruptive* (also called *diversifying*) selection. In the first case, one extreme phenotype is fittest and selec-

tion leads to its fixation (ignoring other evolutionary forces). If an intermediate phenotype is fittest, we speak of stabilizing selection [12, p. 270]. The final case is encountered, when the heterozygote is less fit than both homozygotes. Unless frequency dependent, this selection leads to fixation of one homozygote; which depends on the initial frequencies [15, p. 212]. Which genotype has the highest fitness under a given selection mode depends on the relation between phenotype and genotype. The relationship between phenotype and fitness can depend on the environment, since different environmental conditions can favor different phenotypes [12, p. 270]. Later, we will encounter this in our model, where the population will be distributed over different ecological niches, i.e., different environments.

In the following, we introduce a more detailed terminology for fitness, which will be used when analyzing our model in the subsequent chapters. The *relative fitness* is the absolute fitness value relative to that of some reference genotype, which is usually the fittest genotype, and often is chosen to have fitness 1. If for simplicity in the following chapters we speak of fitness, we refer to relative fitness rather than to absolute fitness values. The *mean fitness* is the average fitness of individuals in a population relative to the fittest genotype, i.e., the sum of the relative fitnesses of the genotypes, each weighted by its frequency in the population.

In this diploma thesis we consider traits under *linear directional selection*. The removal of disadvantageous alleles is the fundamental basis of adaptive evolution. An advantageous allele is initially very rare if it is a newly arisen mutation or if it was disadvantageous before an environmental change made it advantageous. An advantageous allele that increases from a very low frequency is said to *invade* a population [12, p. 274]. One case of directional selection is the linear one. This means that the fitness of the heterozygote is precisely intermediate between that of the two homozygotes, i.e., there is no dominance with respect to fitness [12, p. 275]. Directional selection in favor of the prevalent, advantageous homozygous genotype, is called *purifying selection* and leads to a reduction or elimination of disadvantageous alleles. The number of generations required for an advantageous allele to replace one that is disadvantageous depends on the initial allele frequencies, the *selection coefficient* (cf. (1.8)) and the degree of dominance [12, p. 276]. *Adaptive* or *fitness landscapes* offer a conceptual tool for representing the dynamics of allele frequencies, showing the relation between allele frequency and mean fitness within a generation [12, p. 287]. Another simplification we assume is that the strength of selection is time and frequency independent and acts equally on both sexes (although in this work it is not necessary to distinguish between the two sexes, as we assume that our population is monoecious).

Migration occurs when a population is divided into *subpopulations* which inhabit sep-

arate *ecological niches*, also called *demes*. Such populations are often referred to as *geographically* or *spatially structured*. In our model, adult individuals may be exchanged between the demes, where mating and reproduction takes place. In our case, migration occurs independently of the genetic background. In every niche, selection acts locally, i.e., its effect is dependent on the deme. In this work, we focus only on a *discrete-space setting*,in which there exists a finite number of demes. Before migration, every individual stays in the deme where it was born. Furthermore, *discrete time steps* are also assumed and in each time step the filial generation constitutes the new parental pool of individuals. Therefore, the generations are nonoverlapping.

Another assumption concerning the populations analyzed in this diploma thesis is that the size of the population and of each subpopulation is large enough to neglect the effects of *random genetic drift* (a description of this effect can be found in [5, p. 18]). Thus, the probabilistic variation of gene frequencies in a finite population due to random sampling of gametes or genes is ignored [5, p. 18].

The mating scheme has also to be taken into account. A natural assumption is *random mating*, which we will focus on in this work. This means that mating takes place without regard to ancestry or the genotype under consideration, which seems to occur frequently in nature. The opposite is called *assortative mating*. An illustrative example for random mating within humans is the inheritance of blood group but not the body height. Random mating conserves allele frequencies and, after one generation, genotypic frequencies. This important fact is known as the *Hardy-Weinberg law* [5, pp. 5–7].

Finally, we summarize the assumptions of the model considered in this diploma thesis:

- a diploid, monoecious population
- discrete, non-overlapping generations
- discrete space: the population is subdivided into $\Gamma \geq 1$ panmictic colonies, called demes
- random mating in every deme
- two diallelic loci with alleles A_1,A_2 and B_1,B_2
- no epistasis among loci
- no dominance
- local selection: in each deme selection acts through differential viabilities that are time and frequency independent

- after selection, recombination occurs and adults migrate
- the sexes are equivalent or the population is monoecious
- mutation is ignored
- infinitely large population size, therefore random genetic drift is ignored and frequencies can be identified with probabilities

1.2 The Model

In this section we introduce and formulate the basic model.

In a diploid organism we consider two loci each with two alleles. Alleles are denoted by A_1, A_2 at the first locus and by B_1, B_2 at the second locus. As we consider two loci, it is not sufficient to describe the joint dynamics in terms of allele frequencies, because nonrandom associations between alleles at different loci may occur (see Section 1.1). Nevertheless, the system can be described in terms of gametic frequencies because random mating is assumed and offsprings are in Hardy-Weinberg proportions ([5, p. 46]).

The frequencies of the four haploid gametes A_1B_1, A_1B_2, A_2B_1, and A_2B_2 in deme α are designated $x_{1,\alpha}$, $x_{2,\alpha}$, $x_{3,\alpha}$, and $x_{4,\alpha}$, respectively. Obviously, they fulfill

$$x_{1,\alpha} + x_{2,\alpha} + x_{3,\alpha} + x_{4,\alpha} = 1. \tag{1.1}$$

The set of gametes is denoted by $I = \{1, 2, 3, 4\}$, single gametes are denoted by letters $i, j \in I$, the set of all demes by G and single demes by $\alpha, \beta \in G$. Let the number of demes be $\Gamma = |G|$. When denoting a locus, we use the letters k and n. The allele frequencies of A_1, A_2, B_1 and B_2 in deme α are

$$p^{(1)}_{1,\alpha} = x_{1,\alpha} + x_{2,\alpha},\ p^{(1)}_{2,\alpha} = x_{3,\alpha} + x_{4,\alpha},\ p^{(2)}_{1,\alpha} = x_{1,\alpha} + x_{3,\alpha},\ p^{(2)}_{2,\alpha} = x_{2,\alpha} + x_{4,\alpha}. \tag{1.2}$$

Thus, the superscript denotes the locus. When convenient, the following simplified notation for allele frequencies will be used: $p^{(1)}_{\alpha} = p^{(1)}_{1,\alpha}$ and $p^{(2)}_{\alpha} = p^{(2)}_{1,\alpha}$. Obviously, $p^{(1)}_{\alpha}$ and $p^{(2)}_{\alpha}$ are sufficient to describe the allele frequencies because $p^{(1)}_{1,\alpha} + p^{(1)}_{2,\alpha} = 1$ and $p^{(2)}_{1,\alpha} + p^{(2)}_{2,\alpha} = 1$ as follows directly from (1.1).

If we want to emphasize the generation t of the population, we write $x_{i,\alpha}(t)$ for the gametic frequencies, and analogously for the allele and zygote frequencies. A prime,$'$, denotes the next generation. For example $x'_{1,\alpha}$ is the frequency of gamete A_1B_1 in deme α after one generation.

The *k-dimensional simplex* is denoted by

$$\Delta_{k+1} = \{z \in \mathbb{R}^{k+1} : z_i \geq 0 \text{ for every } i \in \{1, \ldots, k+1\}, \sum_i z_i = 1\} \tag{1.3}$$

and its Γ-fold Cartesian product by Δ_{k+1}^{Γ}. For a graphical representation of Δ_4, see Figure 1.1. In the following we will construct a dynamical system on Δ_4^{Γ}.

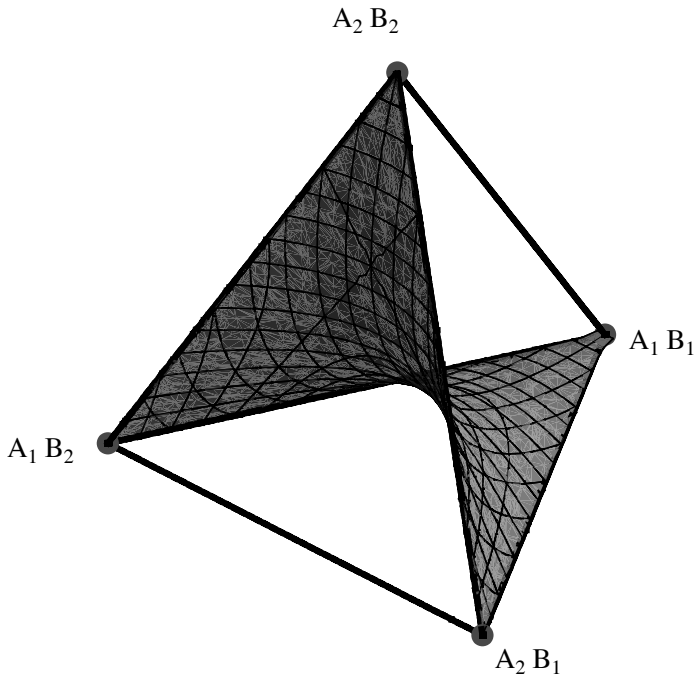

Figure 1.1: The three-dimensional simplex can be represented as a tetrahedron and represents the state space of the four gamete frequencies in a single deme. The vertices denote fixation of the labeled gametes. In the center of the simplex, all gametes have frequency $\frac{1}{4}$. The shaded two-dimensional surface is the linkage-equilibrium manifold, i.e., $D = 0$ (see below). Maximum linkage disequilibrium is assumed at the center of the edges connecting A_1B_2 with A_2B_1, and A_1B_1 with A_2B_2; it corresponds to the values $D = \pm\frac{1}{4}$.

As stated in Section 1.1, the life cycle of the population is:

$$\boxed{\text{zygote} \xrightarrow{\text{selection}} \text{adult} \xrightarrow{\text{recombination}} \text{gamete} \xrightarrow[\text{and mating}]{\text{migration}} \text{zygote}}$$

To translate this cycle into the appropriate mathematical equations, we introduce the following notation:

Let the frequency and fitness of the genotype ij of a zygote (formed by the gametes i, j) in deme α be $x_{ij,\alpha}$ and $W_{ij,\alpha}$, respectively, and let $W_\alpha = (W_{ij,\alpha})$ denote the fitness matrix in deme α. Then $W_\alpha = W_\alpha^T \geq 0$ holds as for the genotype the position of alleles is not of relevance, i.e., W_α is symmetric (the superscript T indicates transposition), see also (1.10). Then the marginal fitness of gamete i and the mean fitness, respectively, in deme α (which are also referred to as local fitnesses values) are given by:

$$W_{i,\alpha} = \sum_j W_{ij,\alpha} x_{j,\alpha}, \tag{1.4a}$$

$$\overline{W}_\alpha = \sum_i W_{i,\alpha} x_{i,\alpha} \overset{(1.4a)}{=} \sum_{i,j} W_{ij,\alpha} x_{j,\alpha} x_{i,\alpha}, \tag{1.4b}$$

where sums without ranges indicate summation over all admissible indices, e.g., $\sum_i = \sum_{i \in I}$, $\sum_\alpha = \sum_{\alpha \in G}$.

In the following, we assume that selection acts on a quantitative trait that is controlled *additively* by two diallelic loci. Whenever we use the term additive fitnesses this means that there is no epistasis and no dominance. Let the contributions of alleles A_1, A_2, B_1, and B_2 to the *genotypic value* g of the trait be $-\frac{1}{2}\gamma_1$, $\frac{1}{2}\gamma_1$, $-\frac{1}{2}\gamma_2$, and $\frac{1}{2}\gamma_2$. We assume that these contributions are independent of the environment. In other words, we assume absence of genotype-environment interaction.

Because the alleles determine the genotypic value g purely additively, the contributions of the gametes A_1B_1, A_1B_2, A_2B_1, and A_2B_2 are $-\frac{1}{2}(\gamma_1 + \gamma_2)$, $-\frac{1}{2}(\gamma_1 - \gamma_2)$, $\frac{1}{2}(\gamma_1 - \gamma_2)$, $\frac{1}{2}(\gamma_1 + \gamma_2)$, respectively, and the resulting genotypic values are given by

$$\begin{array}{c} \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array} \begin{array}{c} \begin{array}{cccc} A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 \end{array} \\ \begin{pmatrix} -\gamma_1 - \gamma_2 & -\gamma_1 & -\gamma_2 & 0 \\ -\gamma_1 & -\gamma_1 + \gamma_2 & 0 & \gamma_2 \\ -\gamma_2 & 0 & \gamma_1 - \gamma_2 & \gamma_1 \\ 0 & \gamma_2 & \gamma_1 & \gamma_1 + \gamma_2 \end{pmatrix} \end{array}. \tag{1.5}$$

The contributions of the genotypic values of a single locus are:

$$\begin{array}{cccccc} A_1A_1 & A_1A_2 & A_2A_2 & \qquad & B_1B_1 & B_1B_2 & B_2B_2 \\ -\gamma_1 & 0 & \gamma_1 & & -\gamma_2 & 0 & \gamma_2 \end{array} \tag{1.6}$$

For definiteness we assume $\gamma_1 \geq \gamma_2 > 0$ and refer to these loci as *major* and *minor*, respectively. We call γ_1 and γ_2 the *effects of the loci*.

Throughout this diploma thesis, the trait is assumed to be under (linear) directional selection in every deme. Then, relative fitness of individuals with genotypic value g in deme α, designated $\Psi_\alpha(g)$, is given by

$$\Psi_\alpha(g) = 1 + s_\alpha g \tag{1.7}$$

or in matrix notation

$$\begin{array}{c} \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array} \begin{array}{c} \begin{array}{cccc} A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 \end{array} \\ \begin{pmatrix} 1 - s_\alpha(\gamma_1 + \gamma_2) & 1 - s_\alpha\gamma_1 & 1 - s_\alpha\gamma_2 & 1 \\ 1 - s_\alpha\gamma_1 & 1 - s_\alpha(\gamma_1 - \gamma_2) & 1 & 1 + s_\alpha\gamma_2 \\ 1 - s_\alpha\gamma_2 & 1 & 1 + s_\alpha(\gamma_1 - \gamma_2) & 1 + s_\alpha\gamma_1 \\ 1 & 1 + s_\alpha\gamma_2 & 1 + s_\alpha\gamma_1 & 1 + s_\alpha(\gamma_1 + \gamma_2) \end{pmatrix} \end{array}, \tag{1.8}$$

where s_α is the *selection coefficient* of deme α. It is the amount by which the fitness of one genotype is changed relative to the reference genotype and it measures the selective advantage of the fitter genotype, or the intensity of selection against the less fit genotype [12, p. 272]. Clearly, $|s_\alpha|$ must be sufficiently small so that $\Psi_\alpha(g) > 0$ for every α and every genotype. Of most interest are the cases where s_α changes sign between demes.

Of course one could also introduce different selection coefficients s_α for the first locus and t_α for the second locus:

$$\begin{array}{c} \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array} \begin{array}{c} \begin{array}{cccc} A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 \end{array} \\ \begin{pmatrix} 1 - s_\alpha\gamma_1 - t_\alpha\gamma_2 & 1 - s_\alpha\gamma_1 & 1 - t_\alpha\gamma_2 & 1 \\ 1 - s_\alpha\gamma_1 & 1 - s_\alpha\gamma_1 + t_\alpha\gamma_2 & 1 & 1 + t_\alpha\gamma_2 \\ 1 - t_\alpha\gamma_2 & 1 & 1 + s_\alpha\gamma_1 - t_\alpha\gamma_2 & 1 + s_\alpha\gamma_1 \\ 1 & 1 + t_\alpha\gamma_2 & 1 + s_\alpha\gamma_1 & 1 + s_\alpha\gamma_1 + t_\alpha\gamma_2 \end{pmatrix} \end{array}. \tag{1.9}$$

This is a more general model than (1.7). In fact it would suffice to chose only two parameters in each deme that affect fitness values as the four parameters $s_\alpha, t_\alpha, \gamma_1$, and γ_2 appear only in the combinations $s_\alpha\gamma_1$, and $s_\alpha\gamma_2$. Except in Section 5.1, for simplicity we will restrict our attention to (1.8).

From (1.5) it is easy to see that the following fitness values are equal in every deme:

$$W_{14,\alpha} = W_{23,\alpha} = W_{32,\alpha} = W_{41,\alpha}, \tag{1.10a}$$
$$W_{12,\alpha} = W_{21,\alpha}, \tag{1.10b}$$
$$W_{24,\alpha} = W_{42,\alpha}, \tag{1.10c}$$
$$W_{34,\alpha} = W_{43,\alpha}. \tag{1.10d}$$

Given the fitness matrix (1.8) for deme α, we can calculate the marginal fitnesses for the gametes and the mean fitness according to (1.4a) and (1.4b), respectively. This yields

$$\begin{aligned} W_{1,\alpha} &= 1 - s_\alpha[\gamma_1(x_{1,\alpha} + x_{2,\alpha}) + \gamma_2(x_{1,\alpha} + x_{3,\alpha})] \\ &= 1 - s_\alpha[\gamma_1 p_\alpha^{(1)} + \gamma_2 p_\alpha^{(2)}], \end{aligned} \tag{1.11a}$$
$$\begin{aligned} W_{2,\alpha} &= 1 - s_\alpha[\gamma_1(x_{1,\alpha} + x_{2,\alpha}) - \gamma_2(x_{2,\alpha} + x_{4,\alpha})] \\ &= 1 - s_\alpha[\gamma_1 p_\alpha^{(1)} - \gamma_2(1 - p_\alpha^{(2)})], \end{aligned} \tag{1.11b}$$
$$\begin{aligned} W_{3,\alpha} &= 1 - s_\alpha[-\gamma_1(x_{3,\alpha} + x_{4,\alpha}) + \gamma_2(x_{1,\alpha} + x_{3,\alpha})] \\ &= 1 - s_\alpha[-\gamma_1(1 - p_\alpha^{(1)}) + \gamma_2 p_\alpha^{(2)}], \end{aligned} \tag{1.11c}$$
$$\begin{aligned} W_{4,\alpha} &= 1 - s_\alpha[-\gamma_1(x_{3,\alpha} + x_{4,\alpha}) - \gamma_2(x_{2,\alpha} + x_{4,\alpha})] \\ &= 1 - s_\alpha[-\gamma_1(1 - p_\alpha^{(1)}) - \gamma_2(1 - p_\alpha^{(2)})], \end{aligned} \tag{1.11d}$$

and

$$\begin{aligned} \overline{W}_\alpha &= 1 - s_\alpha[(1 - 2(x_{3,\alpha} + x_{4,\alpha}))\gamma_1 + (1 - 2(x_{2,\alpha} + x_{4,\alpha}))\gamma_2] \\ &= 1 + s_\alpha[\gamma_1(1 - 2p_\alpha^{(1)}) + \gamma_2(1 - 2p_\alpha^{(2)})]. \end{aligned} \tag{1.12}$$

It is notable that the mean fitness is linear in the gamete frequencies (and allele frequencies) which is not the case under more general fitness schemes. Usually the mean fitness is a quadratic polynomial. The linearity follows from the assumption of additive fitnesses. This property will greatly simplify the analysis.

Let c denote the recombination rate between the two loci. As stated in Section 1.1, $c \in [0, 0.5]$. Let

$$D_\alpha = x_{1,\alpha}x_{4,\alpha} - x_{2,\alpha}x_{3,\alpha} \tag{1.13}$$

denote the measure of linkage disequilibrium in deme α. If for deme α the relation $D_\alpha = 0$ holds, we have linkage equilibrium in this deme, i.e., statistical independence of

the two loci. Therefore, for deme α we can define the two-dimensional *linkage equilibrium manifold*, also called *Wright manifold*, as

$$\begin{aligned}\Lambda_\alpha = \Big\{ & (x_{1,\alpha}, x_{2,\alpha}, x_{3,\alpha}, x_{4,\alpha})^T : \; x_{1,\alpha} = p^{(1)}_{1,\alpha} p^{(2)}_{1,\alpha}, \; x_{2,\alpha} = p^{(1)}_{1,\alpha} p^{(2)}_{2,\alpha}, \\ & x_{3,\alpha} = p^{(1)}_{2,\alpha} p^{(2)}_{1,\alpha}, \; x_{4,\alpha} = p^{(1)}_{2,\alpha} p^{(2)}_{2,\alpha} \Big\} \\ = \; & \big\{ (x_{1,\alpha}, x_{2,\alpha}, x_{3,\alpha}, x_{4,\alpha})^T : D_\alpha = 0 \big\} \subseteq \Delta_4. \end{aligned} \tag{1.14}$$

The *global linkage equilibrium manifold* is then defined as the Cartesian product of the Wright manifolds of every deme,

$$\Lambda = \Lambda_1 \times \cdots \times \Lambda_\Gamma \subseteq \Delta_4^\Gamma, \tag{1.15}$$

cf. ([2, p. 10]. In Figure 1.1, the shaded area represents the linkage equilibrium manifold in Δ_4.

Let $\eta_1 = \eta_4 = 1$, $\eta_2 = \eta_3 = -1$. Then the gamete frequencies in each deme after selection and recombination (i.e., meiosis) are given by [5, p. 47]:

$$x^*_{i,\alpha} = (x_{i,\alpha} W_{i,\alpha} - \eta_i c W_{14,\alpha} D_\alpha)/\overline{W}_\alpha \text{ for } i \in I, \alpha \in G. \tag{1.16}$$

Another good source for the derivation of (1.17a) with the list of gametes produced by each genotype can be found in [18]. Note that for every deme $W_{14,\alpha} = 1$, i.e., (1.16) reduces to:

$$x^*_{i,\alpha} = (x_{i,\alpha} W_{i,\alpha} - \eta_i c D_\alpha)/\overline{W}_\alpha \text{ for } i \in I, \alpha \in G. \tag{1.17a}$$

If no selection operates, (1.17a) becomes $x^*_{i,\alpha} = x_{i,\alpha} - \eta_i c D_\alpha$. The linkage disequilibrium in the next generation is then given by $D' = (1-c)D$. In this case it is obvious that $D_\alpha = 0$ implies $D'_\alpha = 0$, i.e., the Wright manifold is positively invariant. If $c > 0$, the linkage disequilibrium converges to zero as $t \to \infty$. On the other hand, if the population is subject to selection, then the Wright manifold is not invariant in general, and gamete frequencies need not converge to linkage equilibrium. For a mathematical introduction and a more detailed explanation of a system exhibiting only recombination, we refer to [5, pp. 21 – 24] or to [14, pp. 245 – 248].

After selection and recombination, migration occurs. Let $m_{\alpha\beta}$ be the probability that an adult individual in deme α migrated from deme β. Then the $\Gamma \times \Gamma$ *backward migration*

matrix $M = (m_{\alpha\beta})$ is *stochastic*, i.e., it satisfies

$$0 \leq m_{\alpha\beta} \leq 1 \text{ for every } \alpha, \beta \in G \text{ and } \sum_{\beta} m_{\alpha\beta} = 1 \text{ for every } \alpha.$$

Thus, row α of the backward migration matrix M specifies the distribution of the birthplace of individuals breeding in deme α [8, p. 17]. We assume that M is constant, as it is the case for *soft selection*, cf. [7, p. 12]. Soft selection occurs when population regulation acts within niches [21]. This means that there is no competition between the demes and that the deme sizes are not changed by selection. Let c_α and c^*_α denote the relative size of deme α before and after selection, respectively. Thus, in the case of soft selection $c_\alpha = c^*_\alpha$ holds in every deme.

The gametic frequencies after migration, hence in the next generation, are given by

$$x'_{i,\alpha} = \sum_{\beta} m_{\alpha\beta} x^*_{i,\beta}, \text{ for } i \in I, \alpha \in G. \tag{1.17b}$$

According to [2, p. 945], the processes of migration and recombination commute. We shall view (1.17) as a dynamical system on Δ_4^Γ, i.e., Δ_4^Γ is invariant under these equations. The two equations in (1.17) fully determine the dynamics of gamete frequencies in our population.

Of course, one can also define the dynamics via the *forward migration matrix*, $\tilde{M} = (\tilde{m}_{\alpha\beta})$, where $\tilde{m}_{\alpha\beta}$ is the probability that an individual in deme α migrates into deme β. Obviously, $\tilde{M}$ is stochastic. A straightforward calculation (where we assume that the number of individuals is preserved by migration) yields a simple transformation of the forward migration matrix into the backward migration matrix under soft selection:

$$m_{\alpha\beta} = \frac{c_\beta \tilde{m}_{\beta\alpha}}{\sum_\delta c_\delta \tilde{m}_{\delta\alpha}}.$$

Thus it suffices to work with the backward migration matrix.

An important special case of a migration model is the *Levene model* (see [17]). Here, individuals migrate independently of their origin. In this case, the migration rates are exactly the relative deme sizes, i.e., $m_{\alpha\beta} = c_\beta$ for all demes. Note that in the Levene model with soft selection, the gametic frequencies are equal within the demes after one round of migration: $x'_{i,\alpha} \overset{(1.17b)}{=} \sum_\beta m_{\alpha\beta} x^*_{i,\beta} = \sum_\beta c_\beta x^*_{i,\beta}$, thus independent of α. Consequently, it suffices to analyze (1.17) for deme-independent gamete frequencies x_i rather than $x_{i,\alpha}$ in the case of the Levene model. Thus, the state space is simply Δ_4. It is biologically reasonable to assume that migration rates are in $(0, \frac{1}{2})$. For two demes, we will assume

throughout this work that the Levene model is given by $m_{\beta\alpha} = m_{\alpha\beta} = \frac{1}{2}$. Then, the dynamical system for the Levene model is

$$x_i^{'} = \frac{x_i(W_{i,\alpha} + W_{i,\beta})/2 - \eta_i cD}{(\overline{W}_\alpha + \overline{W}_\beta)/2} \text{ for } i \in I. \tag{1.18}$$

To summarize, in the Levene model there is no population structure after the first round of migration. Only selection is dependent on the geographical structure and may be different within demes.

For most of the analysis, we will assume that M is *ergodic*, i.e., M is irreducible and aperiodic. A matrix M is *periodic* if $\exists d = d(\alpha) \in \mathbb{N}, d > 1$ called *period*, fulfilling $m_{\alpha\alpha}^{nd} > 0 \ \forall n \geq 1$ and $m_{\alpha\alpha}^{m} = 0$ if $m \neq nd$ (here, the superscript denotes the exponent). If such d does not exist, M is called *aperiodic*. A matrix M is *irreducible* if for every entry of $M = (m_{ij})$ there is a power k with $m_{ij}^k > 0$. If there is an exponent k such that $M^k > 0$, M is called *primitive*. For finite M, ergodicity is equivalent to primitivity ([11],[24]). This means that after k generations a descendant of an individual from any niche has a positive probability to be in every other niche. Thus, individuals can reach every deme starting from any deme in sufficiently large time.

Given irreducibility, aperiodicity is achieved if individuals have a positive probability of remaining in some deme, i.e., $m_{\alpha\alpha} > 0$ for some $\alpha \in G$, cf. [11]. Consequently for two demes, an ergodic backward migration matrix M is of the form

$$M = \begin{pmatrix} 1 - m_1 & m_1 \\ m_2 & 1 - m_2 \end{pmatrix}, \tag{1.19}$$

where $0 < m_1, m_2 < 1$. Note that $1 - m_1 = m_2$ leads to a special case of the Levene model because then the rate of migration is independent of the origin. In Section 3.2, we will consider only migration rates in the interval $(0, \frac{1}{2}]$ as this is biologically reasonable. Therefore, the Levene model holds when $m_1 = m_2 = \frac{1}{2}$.

Because the recombination rate is assumed positive, all gametes are generated in a randomly mating population after one round of mating if initially all alleles are present. If migration is ergodic, it takes at most $\Gamma - 1$ generations to introduce every allele in each deme. Therefore, it takes at most Γ generations to generate all gametes in every deme [3, p. 982].

Given ergodicity, there exists a unique *invariant distribution* (or *stationary distribution*) of the Markov chain defined by the transition matrix M [11], i.e., there exists a

principal left eigenvector $\mu \in \text{int}\Delta_\Gamma$ such that

$$\mu^T M = \mu^T$$

holds. The corresponding eigenvalue of M is 1, the principal right eigenvector is $(1,1)^T$ because of stochasticity of M.

For example, given two demes α and β, one can easily calculate that the invariant distribution $\mu = (\mu_\alpha, \mu_\beta) \in \mathbb{R}^2$ of an ergodic backward migration matrix M is given by

$$\mu_\alpha = \frac{m_{\beta\alpha}}{m_{\alpha\beta} + m_{\beta\alpha}} \text{ and } \mu_\beta = \frac{m_{\alpha\beta}}{m_{\alpha\beta} + m_{\beta\alpha}}. \tag{1.20}$$

For $z \in \mathbb{R}^\Gamma$ we define its norm as $\|z\| = \max_\alpha |z_\alpha|$. Let λ_1 denote the non-unit eigenvalue of M with largest modulus and $e = (1, \cdots, 1)^T \in \mathbb{R}^\Gamma$. Then the convergence theorem for ergodic matrices implies that for every κ with

$$|\lambda_1| < \kappa < 1, \tag{1.21}$$

we have

$$\|M^t z - e\mu^T z\| \leq C_z \kappa^t, \tag{1.22}$$

where C_z is a constant independent of the power t. If λ_1 is a simple eigenvalue (which it is for primitive matrices, cf. Theorem of Perron-Frobenius, [24]), we can take $\kappa = |\lambda_1|$.

Now we introduce the following vector notation which will be used later:

$$x_{(\alpha)} = (x_{1,\alpha}, x_{2,\alpha}, x_{3,\alpha}, x_{4,\alpha})^T \in \Delta_4, \tag{1.23a}$$
$$x_i = (x_{i,1}, \ldots, x_{i,\Gamma})^T \in \mathbb{R}^\Gamma, \tag{1.23b}$$
$$x = (x_{(1)}, \ldots, x_{(\Gamma)})^T \in \Delta_4^\Gamma, \tag{1.23c}$$
$$p_{(\alpha)} = (p_\alpha^{(1)}, p_\alpha^{(2)})^T \in \mathbb{R}^2, \tag{1.23d}$$
$$p^{(n)} = (p_1^{(n)}, \ldots, p_\Gamma^{(n)})^T \in \mathbb{R}^\Gamma, \tag{1.23e}$$
$$p = (p_{(1)}, \ldots, p_{(\Gamma)})^T \in \mathbb{R}^{2\Gamma}, \tag{1.23f}$$
$$D = (D_1, \ldots, D_\Gamma)^T \in \mathbb{R}^\Gamma, \tag{1.23g}$$
$$\eta = (1, -1, -1, 1)^T \in \mathbb{R}^4, \tag{1.23h}$$
$$W_i = (W_{i,1}, \cdots, W_{i,\Gamma})^T \in \mathbb{R}^\Gamma, \tag{1.23i}$$
$$1/\overline{W} = (1/\overline{W}_1, \cdots, 1/\overline{W}_\Gamma)^T \in \mathbb{R}^\Gamma, \tag{1.23j}$$

where $i \in I$, $n \in \{1, 2\}$, and $\alpha \in G$. Using (1.23) we rewrite (1.17):

$$x_i^{'} = M\frac{1}{\overline{W}}(x_i W_i - c\eta_i D). \tag{1.24}$$

The stationary distribution (1.20) is used to average gamete and allele frequencies:

$$X = (X_1, \ldots, X_4)^T \in \mathbb{R}^4 \text{ where } X_i = \mu^T x_i,\ i \in, I, \tag{1.25a}$$
$$P = (P_1, P_2)^T \in \mathbb{R}^2 \text{ where } P_n = \mu^T p^{(n)},\ n \in \{1, 2\}. \tag{1.25b}$$

Next, we define the *allele-frequency deviations* y and the *gamete-frequency deviations* q from the average frequencies X and Y:

$$y_\alpha^{(n)} = p_\alpha^{(n)} - P_n \in \mathbb{R}, \tag{1.26a}$$
$$y^{(n)} = p^{(n)} - P_n e \in \mathbb{R}^\Gamma, \tag{1.26b}$$
$$y = (y^{(1)}, y^{(2)})^T \in \mathbb{R}^{2\Gamma}, \tag{1.26c}$$
$$q_{i,\alpha} = x_{i,\alpha} - X_i \in \mathbb{R}, \tag{1.26d}$$
$$q_i = x_i - X_i e \in \mathbb{R}^\Gamma, \tag{1.26e}$$
$$q_{(\alpha)} = x_{(\alpha)} - X \in \mathbb{R}^4, \tag{1.26f}$$
$$q = (q_1, q_2, q_3, q_4)^T \in \mathbb{R}^{4\Gamma}, \tag{1.26g}$$

where $i \in I$, $n \in \{1, 2\}$, and $\alpha \in G$. By definition, q and y measure spatial heterogeneity or diversity, cf. [2, p. 947]. The distribution of gametes (of alleles) is spatially homogeneous, if $q = 0$ ($y = 0$).

Chapter 2

Linkage Equilibrium

In this section we study some properties of linkage disequilibrium. In equation (1.13) we have introduced the linkage disequilibrium measure D_α in deme α. Equilibria for (1.17) with $D = 0$ exist, for example monomorphic equilibria, and they may be asymptotically stable. One has to be aware that this does not necessarily imply that the whole linkage equilibrium manifold is positively invariant. For arbitrary parameters, the linkage equilibrium manifold is in general not positively invariant. Trajectories converging to an asymptotically stable equilibrium with $D = 0$ may traverse the Wright manifold, and approach it again when reaching the equilibrium. In Figure 2.1, we give such an example (exceptionally, we used a migration rate larger than $\frac{1}{2}$, as in this case it was easier to show graphically that the trajectory is not on the linkage equilibrium manifold except for the initial point and end point).

In a large random mating panmictic population, stable linkage disequilibrium occurs only when there is epistasis (cf. Section 1.1) ([4, p. 215]). *Stable linkage disequilibrium* is encountered for example when there is a stable equilibrium of the system (1.17) where $D \neq 0$ holds. According to [19], a subdivided population can exhibit stable linkage disequilibrium in each subpopulation even if epistasis is absent. In the case of two subpopulations, Li and Nei ([19]) derived a necessary condition for linkage equilibrium. Li and Nei showed that linkage equilibrium is achieved if at least at one of the two loci under consideration the gene frequency is the same for the two populations. Their calculation showed that for (1.17), the linkage disequilibrium measure in the next generation in deme α is

$$D'_\alpha = m_{\alpha\alpha} D^*_\alpha + (1 - m_{\alpha\alpha}) D^*_\beta + m_{\alpha\alpha}(1 - m_{\alpha\alpha}) D^*_m, \tag{2.1}$$

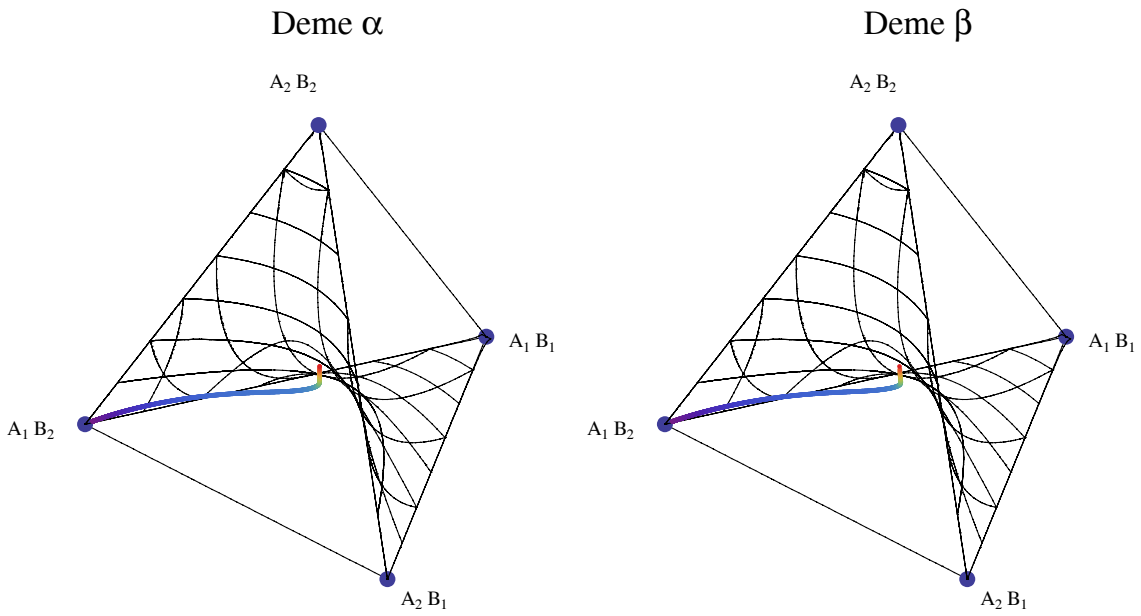

Figure 2.1: The linkage equilibrium manifold (meshed) is not positively invariant in general. In this figure a trajectory starting from the linkage equilibrium manifold is shown. The initial frequencies are $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})^T$ in both demes. The trajectory converges to the equilibrium where gamete A_1B_2 is fixed in both subpopulations, i.e., $D = 0$ holds here too. The linkage equilibrium manifold is not positively invariant in this case, as the trajectory leaves the manifold. (Parameters: $m_{\alpha\beta} = \frac{1}{2}$, $m_{\beta\alpha} = 0.7$, $\gamma_1 = 3$, $\gamma_1 = 1$, $s_\alpha = -0.1$, $s_\beta = 0.1$, $c = 0.01$, iteration steps: 1200)

where the following notation was used:

$$D^*_\alpha = x^*_{1,\alpha}x^*_{4,\alpha} - x^*_{2,\alpha}x^*_{3,\alpha}, \tag{2.2a}$$

$$D^*_m = (p^{(1)*}_\alpha - p^{(1)*}_\beta)(p^{(2)*}_\alpha - p^{(2)*}_\beta). \tag{2.2b}$$

The star, *, denotes the frequencies after selection and recombination, cf. (1.17a). Analogous equations can be obtained for deme β. From (2.1) it is seen that in the next generation the linkage disequilibrium measure in a subpopulation depends not only on the linkage disequilibrium in this population. It also depends on the linkage disequilibrium after selection and recombination in the other subpopulation and on the product, D^*_m, of gene-frequency differences among the populations after selection and recombination. Next, we state the Theorem postulated in [19, p. 176]:

Theorem 2.0.1 (Condition for linkage equilibrium)**.** *Assume a population subdivided in two demes α and β. If fitnesses are additive for both loci, a necessary condition for $D_\alpha = D_\beta = 0$ at equilibrium is $D^*_m = 0$, i.e.,*

$$(p^{(1)*}_\alpha - p^{(1)*}_\beta)(p^{(2)*}_\alpha - p^{(2)*}_\beta) = 0. \tag{2.3}$$

The above condition holds only if $(m_{\alpha\alpha} + m_{\beta\beta})/2 \neq \frac{1}{2}, 0 < m_{\alpha\alpha}, m_{\beta\beta} < 1$ and $c > 0$. Furthermore, the above condition is equivalent to

$$(p^{(1)}_\alpha - p^{(1)}_\beta)(p^{(2)}_\alpha - p^{(2)}_\beta) = 0. \tag{2.4}$$

Condition (2.4) *implies that for each of the two subpopulations to be in linkage equilibrium without epistasis, at least at one of the two loci the gene frequency must be the same in the two populations.*

Proof. The proof is found in [19, pp. 177–178]. □

For biologically reasonable migration rates, i.e., $m_{\alpha\beta} \in (0, 0.5]$, only $m_{\alpha\beta} = \frac{1}{2} = m_{\beta\alpha}$ corresponds to the excluded case in the proof. This corresponds to the Levene model. Recall that in the Levene model, after one round of migration the gamete frequencies are spatially homogeneous, i.e., $q = 0$. Consequently, the linkage disequilibrium measure is also deme independent and it suffices to write $D = 0$ rather than $D_\alpha = 0$ for every deme. In [20, pp. 204 – 207] it is proven that in the Levene model with multiallelic loci and additive fitnesses, generic convergence to the linkage equilibrium manifold occurs, thus $D \to 0$ for $t \to \infty$. In the Levene model the introduction of epistasis can lead to linkage disequilibrium ([25]).

Chapter 3

Equilibria and Stability

3.1 Introduction

In this section we study the existence of various types of equilibria and their stability properties.

We call $x \in \Delta_4^\Gamma$ an *equilibrium point* of the dynamical system (1.17) if

$$x' = (x'_{(1)}, \ldots, x'_{(\Gamma)}) = (x_{(1)}, \ldots, x_{(\Gamma)}) = x \tag{3.1}$$

holds. We distinguish two kinds of equilibria: *monomorphic* and *polymorphic* equilibria. The monomorphic equilibria of the system (1.17) are

$$\hat{x}_{i,\alpha} = 1 \text{ for every } \alpha \in G \text{ and one } i \in I. \tag{3.2}$$

The circumflex, $\widehat{\ }$, indicates an equilibrium state. In other words, there is only one gamete present. Because we neglect mutation, this is clearly an equilibrium. There are four such monomorphic equilibria and they always exist. [15, p. 215].

An equilibrium is called *fully polymorphic* if at every considered locus all alleles are segregating. Thus in our model, at a fully polymorphic equilibrium both alleles at both loci occur with positive frequencies. For $c > 0$, such equilibria are in the interior of Δ_4^Γ.

An equilibrium is called a *single-locus polymorphism* if at one locus an allele is fixed and at the other locus both alleles are segregating. Single-locus polymorphisms are found in the interior of the Cartesian product of two corresponding edges of Δ_4.

In general, it is difficult to explicitly determine fully polymorphic and single-locus polymorphic equilibria. Therefore, it is necessary to study the dynamics (1.17) in special cases, where results can be derived more easily or at all.

Until now, we have introduced the definition of an equilibrium point and specified the different kinds. The next question is, what kind of stability properties do equilibria of (1.17) exhibit? Is the equilibrium point *asymptotically stable* or *unstable*?

Let $U \subseteq \mathbb{R}^n$ with $U \subseteq U_{open}$, where U_{open} is an open set in $\mathbb{R}^n$. We consider

$$f : U \to U \tag{3.3}$$

such that f is also defined and continuously differentiable on U_{open}.

Note that the dynamical system (1.17) is also of the form

$$x^{'} = f(x), \tag{3.4}$$

where f is as in (3.3) and $U = \Delta_4^\Gamma$. An equilibrium $\widehat{x} \in U$ of f satisfies $f(\widehat{x}) = \widehat{x}$. We call $\widehat{x}$ *stable* if for some neighborhood V_1 of $\widehat{x}$ there exists a neighborhood V_2 with $\widehat{x} \in V_2 \subset V_1$ such that $f(\overline{V}_2) \subset V_2$ (where $\overline{V}_2$ denotes the closure of V_2). Therefore the iterates of $f^n(x) = f^{n-1}(f(x))$ for any starting point $x \in \overline{V}_2$ never leave V_2 (here, n denotes the number of iteration steps). We call $\widehat{x}$ *asymptotically stable* if $\widehat{x}$ is stable and every trajectory starting in V_2 converges to $\widehat{x}$ [16]. If no V_1 exists such that all trajectories converge to $\widehat{x}$, the equilibrium is said to be *unstable*. If every trajectory starting in the interior of U converges to $\widehat{x}$, we call $\widehat{x}$ *globally stable*. More complex dynamics may exist such as periodic orbits or limit cycles, but we will not introduce them here because apparently they do not occur in our model.

An important tool for determining stability properties of equilibria are *Lyapunov fuctions*. We give a definition for systems of difference equations of the form (3.3) and (3.4). Let $\widehat{x} \in U$ be an equilibrium of f. A global Lyapunov function of this system is a continuous mapping

$$L : U \to \mathbb{R} \tag{3.5}$$

such that

$$\Delta L = L(x^{'}) - L(x) \leq 0 \text{ for all } x \in U \tag{3.6}$$

holds [16, p. 10]. If L is defined only on a subset of $U \ni \widehat{x}$ then L is a local Lyapunov function. If equality holds only at equilibria, L is called a *strict Lyapunov function*. The definition of a Lyapunov function does not tell us how to find one. Indeed, not for every dynamical system one exists (for example, if there exist limit cycles). If one exists, it may be difficult to find one explicitly. But if a Lyapunov function is derived, statements about stability follow easily from:

Theorem 3.1.1 (Lyapunov's Theorem)**.** *Let $\widehat{x}$ be an equilibrium of a dynamical system f. If there exists a strict global/local Lyapunov function L, $\widehat{x}$ is globally/locally asymptotically stable.*

Proof. For the proof we refer to [16, Section 4.]. □

In this work, we will encounter Lyapunov functions in Section 4.2 in (4.22).

A special case of Lyapunov functions are *gradient systems*. A gradient system is a system of ordinary differential equations of the form

$$\dot{x} = -\nabla V(x), \tag{3.7}$$

where $x = (x_1, \ldots, x_k)$ is a function of time t and V is a sufficiently smooth real-valued function on $\mathbb{R}^k$ called the *potential function*. Then

$$\nabla V(x) = \left(\frac{\partial V}{\partial x_1}, \ldots, \frac{\partial V}{\partial x_k}\right)^T \tag{3.8}$$

is the gradient vector of first-order partial derivatives. As already indicated above, gradient systems have important properties. All equilibria can be obtained from the condition

$$\dot{V}(x) = 0. \tag{3.9}$$

Cycles or more complex equilibrium states can be excluded. Therefore, all trajectories converge to a set of equilibrium points. Furthermore, trajectories are perpendicular to the level surface of the potential function V [5, p. 42,p. 349–352]. In this work, we will encounter a gradient system in Section 4.2.

The standard procedure to determine local stability is analyzing the linearized system around an equilibrium. Given a mapping f defined as in (3.3) with equilibrium $\widehat{x}$, following *Taylor's Theorem*, one can rewrite the system

$$f(x) - \hat{x} = \underbrace{f(\hat{x}) - \hat{x}}_{=0} + J|_{x=\hat{x}}(x - \hat{x}) + \mathcal{O}((x - \hat{x})^2), \text{ as } x \to \widehat{x}, \tag{3.10}$$

which means we can approximate the dynamical system locally around an equilibrium point by its linearization. This linearization is given by the *Jacobian*, J. Then, the stability of the system is described by the *eigenvalues* λ of the Jacobian. The eigenvalues are calculated via the *characteristic polynomial*, $\phi(\lambda)$, of the linearized system J. These are exactly the zeros of the characteristic polynomial.

For difference equations, an equilibrium point is hyperbolic if no eigenvalue has modulus one, i.e., if none of the eigenvalues lies on the unit circle ($|\lambda| \neq 1$) [16]. A hyperbolic equilibrium is always an isolated equilibrium and in a compact set (for example Δ_4^Γ) there can be at most finitely many hyperbolic equilibria.

Although we mostly work with difference equations, in Section 4.2 and 5.2.1 we will also work with ordinary differential equations. Therefore, we next define hyperbolic equilibria for differential equations. An equilibrium of a system of differential equations is called *hyperbolic*, if the Jacobian matrix at that equilibrium point has no eigenvalues on the imaginary axis, i.e., no eigenvalues with zero real part ($|\lambda| \neq 0$).

3.2 Finding the Equilibria

In this section we find certain equilibria of the dynamical system (1.17) for a population inhabiting two demes, i.e., if $\Gamma = 2$. Throughout this section, we assume that migration rates are in the interval $(0, 0.5]$, $c > 0$ is fixed, and that selection is not absent. For the case when migration rates are zero we refer to Section 5.1. For the case when selection is absent we refer to Section 4.1. For simplicity, we abbreviate $m_{\alpha\beta} = m_1$ and $m_{\beta\alpha} = m_2$. Except for special cases, it is impossible to calculate all equilibria of (1.17) explicitly. This would require to solve polynomial equations of order higher than two.

We already know that the four monomorphic equilibria always exist, cf. (3.2). Obviously, in both subpopulations the same gamete is fixed there. In the following we will see that for certain cases the monomorphic equilibria are the only equilibria of (1.17). If selection is uniform (also called deme-independent), i.e., $s_\alpha = s_\beta$, Nagylaki has shown in [20, Theorem 5.1] the following result:

Theorem 3.2.1 (Equilibria for arbitrary uniform selection)**.** *Assume that equilibria of the zero-migration system* (1.17a) *are hyperbolic and spatially homogeneous. Then, equilibria and their stability properties are preserved under migration.*

Proof. The proof is found on page 210 in [20]. □

Assume that migration is absent and selection is uniform (for a more detailed analysis of this case, see Section 5.1). It is straightforward to calculate that the monomorphic equilibria of (1.17a) are hyperbolic if $s_\alpha = s_\beta \neq$ and $1 - c \neq w_{jj}$, where $W = (w_{ij})$ from (1.8). Then, in both demes the monomorphic equilibrium where the fittest gamete is present is globally stable. Thus, the requirements in Theorem 3.2.1 are met. The monomorhpic equilibrium of (1.17) where the fittest gamete is present exists under uniform selection and

the stability properties are the same as for (1.17a). The unstable, hyperbolic monomorphic equilibria of the zero-migration system are also unstable equilibria of (1.17).

Next, assume that $\text{sign}(s_\alpha) = \text{sign}(s_\beta)$. For a panmictic population in linkage equilibrium with directional selection, and partial or no dominance, Lyapunov functions for both loci are known from [3]:

$$\Delta p_{i,\alpha}^{(1)} \geq 0, \text{ and } \Delta p_{i,\alpha}^{(2)} \geq 0, \ \forall \alpha, \tag{3.11}$$

where $\Delta p_{i,\alpha}^{(n)} = p_{i,\alpha}^{(n)*} - p_{i,\alpha}^{(n)}$, n denotes the locus, and i denotes the fittest allele at each locus. Equality holds only at the monomorphic states. The Lyapunov functions assume their maxima at the monomorphic equilibrium where the fittest gamete is fixed. Using the same notation as above, we obtain the following result:

Theorem 3.2.2 (Lyapunov functions for $D = 0$)**.** *Assume selection acts in the same direction in both demes. Then, for the two loci the Lyapunov functions on the global linkage equilibrium manifold under* (1.17)*, are*

$$(m_2 p_{i,\alpha}^{(1)} + m_1 p_{i,\beta}^{(1)})' \geq m_2 p_{i,\alpha}^{(1)} + m_1 p_{i,\beta}^{(1)}, \tag{3.12a}$$

$$(m_2 p_{i,\alpha}^{(2)} + m_1 p_{i,\beta}^{(2)})' \geq m_2 p_{i,\alpha}^{(2)} + m_1 p_{i,\beta}^{(2)}. \tag{3.12b}$$

Proof. We will give an explicit proof for (3.12a). The remaining case follows analogously. As selection acts in the same direction in both demes, the fittest allele (denoted by i) at the first locus is the same for both subpopulations. Using (1.17b) and (3.11), we obtain

$$(m_2 p_{i,\alpha}^{(1)} + m_1 p_{i,\beta}^{(1)})' = m_2 p_{i,\alpha}^{(1)*} + m_1 p_{i,\beta}^{(1)*} \geq m_2 p_{i,\alpha}^{(1)} + m_1 p_{i,\beta}^{(1)}.$$

The functions obtain their maxima if and only if in both demes the fittest allele i is fixed at each locus. □

The Lyapunov functions (3.12) are useful when the linkage equilibrium manifold is invariant. We have seen in Section 2 that this is not the case in general. Obviously, the linkage equilibrium manifold is invariant for the marginal one-locus systems. Thus, in the marginal one-locus systems the fittest gamete, which is the same for both demes, is asymptotically stable. No single-locus polymorphisms exist. From 3.2.2, it follows that for the full system (1.17) the monomorphic equilibria formed by gametes which are not the fittest gamete are unstable.

For finding single-locus polymorphisms it suffices to analyze (1.17) restricted to the marginal one-locus systems. We will use an important result from [22, pp. 130–132]. There, the conditions for protected polymorphisms at one diallelic locus and two demes if there is no dominance are given. Fitnesses are rescaled such that the heterozygote has fitness 1 (which is allowed as we assume soft selection) and the homozygotes have rescaled fitness $1 - \sigma_k$ and $1 + \sigma_k$, respectively, where $k = \alpha, \beta$ denotes the deme. It is straightforward to calculate σ_k for the marginal one-locus systems of (1.17):

$$\sigma_k = \frac{s_k\gamma_2}{1 - s_k\gamma_1} \quad (A_1B_1, A_1B_2), \tag{3.13a}$$

$$\sigma_k = \frac{s_k\gamma_2}{1 + s_k\gamma_1} \quad (A_2B_1, A_2B_2), \tag{3.13b}$$

$$\sigma_k = \frac{s_k\gamma_1}{1 - s_k\gamma_2} \quad (A_1B_1, A_2B_1), \tag{3.13c}$$

$$\sigma_k = \frac{s_k\gamma_1}{1 + s_k\gamma_2} \quad (A_1B_2, A_2B_2), \tag{3.13d}$$

where $k = \alpha, \beta$ and the brackets denote the marginal one-locus system. Note that if $s_\alpha s_\beta < 0$ holds, $\sigma_\alpha\sigma_\beta < 0$ holds too. We define

$$\kappa = \frac{m_1}{\sigma_\alpha} + \frac{m_2}{\sigma_\beta}. \tag{3.14}$$

Theorem 3.2.3 (Conditions for protection)**.** *Define*

$$\Omega_+ = \{(\sigma_\alpha, \sigma_\beta) : \sigma_\alpha\sigma_\beta < 0 \text{ and } |\kappa| < 1\}, \tag{3.15a}$$

$$\Omega_0 = \{(\sigma_\alpha, \sigma_\beta) : \sigma_\alpha\sigma_\beta < 0 \text{ and } \kappa > 1\}, \tag{3.15b}$$

$$\Omega_1 = \{(\sigma_\alpha, \sigma_\beta) : \sigma_\alpha\sigma_\beta < 0 \text{ and } \kappa < -1\}. \tag{3.15c}$$

Then, in region Ω_+ there is a protected polymorphism, and in Ω_0 and Ω_1 one of the two alleles is protected. The existence of a single-locus polymorphism implies that it is protected.

It is straightforward to see that a protected single-locus polymorphism exists only if the monomorphic equilibria are unstable. Figure 3.1 shows the areas where one allele or both alleles are protected. Note that only selection coefficients with changing sign have the potential for a protected polymorphism. Also, note that Nagylaki's result do not imply uniqueness of single-locus polymorphisms if they exist. In fact, solving the marginal one-locus system (3.17) for polymorphic equilibria leads to solving a quadratic equation. Thus, multiple single-locus polymorphisms cannot be excluded in general.

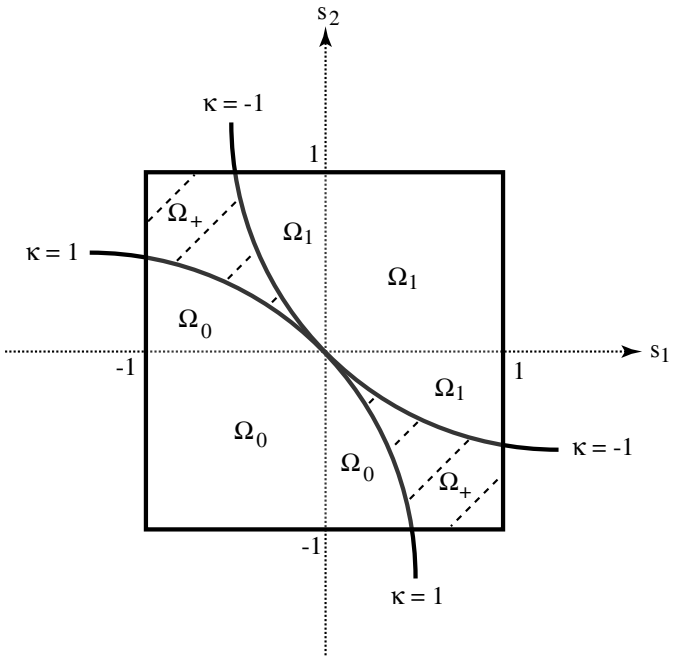

Figure 3.1: With no dominance, in region Ω_+ there is a protected polymorphism. In regions Ω_0 and Ω_1, one of the two alleles is protected. In region Ω_0, the allele which is the fittest if $s_\alpha, s_\beta < 0$ is protected. In region Ω_1, the allele which is the fittest if $s_\alpha, s_\beta > 0$ is protected. Note that the regions where selection acts uniformly is in accordance to Theorem 3.2.1. [22, Figure 4.2]

$(A_1B_1,\ A_1B_1)$	$(A_1B_1,\ A_1B_2)$	$(A_1B_2,\ A_1B_2)$
$1-s_\alpha(\gamma_1+\gamma_2)$	$1-s_\alpha\gamma_1$	$1-s_\alpha(\gamma_1-\gamma_2)$

Table 3.1: Fitness values of genotypes formed by A_1B_1 and A_1B_2

3.2.1 Single-locus Polyorphisms

Finding single-locus polymorphisms of (1.17) is possible as then the system is reduced to two equations. For single-locus polymorphisms, there are four possibilities which gametes can segregate:

$$A_1B_1 \text{ and } A_1B_2,\ A_2B_1 \text{ and } A_2B_2,$$
$$A_1B_1 \text{ and } A_2B_1,\ A_1B_2 \text{ and } A_2B_2.$$

Single-locus polymorphisms are on the Wright manifold and thus satisfy $D_1 = D_2 = 0$. There, it follows from (1.17a) that $x'_{(\alpha)}$ is independent of the recombination rate c. The marginal dynamics is a one-locus selection-migration model. Consequently, if all subpopulations are formed of, for instance, A_1B_1 and A_1B_2 only, then in each simplex this edge is invariant. Thus, the system (1.17) simplifies a lot. It is reduced to a system of two equations describing the allele frequencies:

$$p_{i,\alpha}^{(n)\prime} = (1-m_1)\frac{p_{i,\alpha}^{(n)}W_{i,\alpha}}{\overline{W}_\alpha} + m_1\frac{p_{i,\beta}^{(n)}W_{i,\beta}}{\overline{W}_\beta}, \tag{3.17a}$$

$$p_{i,\beta}^{(n)\prime} = m_2\frac{p_{i,\alpha}^{(n)}W_{i,\alpha}}{\overline{W}_\alpha} + (1-m_2)\frac{p_{i,\beta}^{(n)}W_{i,\beta}}{\overline{W}_\beta}, \tag{3.17b}$$

where n denotes the locus and i the allele at locus n.

In the Levene model, (3.17) becomes a single equation:

$$p_i^{(n)\prime} = \frac{p_i^{(n)}}{2}\left(W_{i,\alpha}/\overline{W}_\alpha + W_{i,\beta}/\overline{W}_\beta\right). \tag{3.18}$$

Two subpopulations formed by A_1B_1 and A_1B_2

We consider two subpopulations which consist of the gametes A_1B_1 and A_1B_2. Consequently, in every deme $p_{1,\alpha}^{(1)} = x_{1,\alpha} + x_{2,\alpha} = 1$, $p_{1,\alpha}^{(2)} = x_{1,\alpha}$, and $p_{2,\alpha}^{(2)} = x_{2,\alpha}$ hold. Therefore, it suffices to describe the gamete frequencies by $p_\alpha = p_{1,\alpha}^{(2)}$ and $p_\beta = p_{1,\beta}^{(2)}$.

The fitness values of the relevant genotypes are summarized in Table 3.1. Recall that

for positive fitnesses

$$\frac{1}{\gamma_1} > s_\alpha,\ \frac{1}{\gamma_1+\gamma_2} > s_\alpha,\ \frac{1}{\gamma_1-\gamma_2} > s_\alpha, \tag{3.19}$$

must hold. The marginal and mean fitnesses are:

$$W_{1,\alpha} = 1 - s_\alpha(\gamma_1 + \gamma_2 p_\alpha), \tag{3.20a}$$

$$W_{2,\alpha} = 1 - s_\alpha(\gamma_1 - \gamma_2(1-p_\alpha)), \tag{3.20b}$$

$$\overline{W}_\alpha = 1 - s_\alpha(\gamma_1 - \gamma_2(1-2p_\alpha)). \tag{3.20c}$$

As $D_\alpha = 0$ holds for every deme and following (3.17), (1.17) turns into:

$$p_\alpha^{'} = (1-m_1)\frac{p_\alpha(1-s_\alpha(\gamma_1+\gamma_2 p_\alpha))}{1-s_\alpha(\gamma_1-\gamma_2(1-2p_\alpha))} + m_1\frac{p_\beta(1-s_\beta(\gamma_1+\gamma_2 p_\beta))}{1-s_\beta(\gamma_1-\gamma_2(1-2p_\beta))}, \tag{3.21a}$$

$$p_\beta^{'} = m_2\frac{p_\alpha(1-s_\alpha(\gamma_1+\gamma_2 p_\alpha))}{1-s_\alpha(\gamma_1-\gamma_2(1-2p_\alpha))} + (1-m_2)\frac{p_\beta(1-s_\beta(\gamma_1+\gamma_2 p_\beta))}{1-s_\beta(\gamma_1-\gamma_2(1-2p_\beta))}. \tag{3.21b}$$

As the denominators in (3.21) must not be zero, $p_\alpha = \frac{1}{2}(1-\frac{1}{\gamma_2}(\gamma_1-\frac{1}{s_\alpha}))$ can never hold in any deme.

In general, internal equilibria cannot be expressed in simple form because they are the solutions of a polynomial equation of order greater or equal three. Therefore, we follow Theorem 3.2.3 to obtain conditions for protection. Using (3.13a), we calculate that the region of a protected polymorphism is

$$\Omega_+ = \left\{(s_\alpha, s_\beta) : s_\alpha s_\beta < 0 \text{ and } \left|m_1(-\gamma_1+\frac{1}{s_\alpha}) + m_2(-\gamma_1+\frac{1}{s_\beta})\right| < \gamma_2\right\}. \tag{3.22}$$

The regions where B_1 and B_2 are protected are

$$\Omega_0 = \left\{(s_\alpha, s_\beta) : m_1(-\gamma_1+\frac{1}{s_\alpha}) + m_2(-\gamma_1+\frac{1}{s_\beta}) > \gamma_2\right\}, \tag{3.23a}$$

$$\Omega_1 = \left\{(s_\alpha, s_\beta) : m_1(-\gamma_1+\frac{1}{s_\alpha}) + m_2(-\gamma_1+\frac{1}{s_\beta}) < -\gamma_2\right\}, \tag{3.23b}$$

respectively. Therefore, a single-locus polymorphism involving A_1B_1 and A_1B_2 which is protected within its marginal single-locus polymorphism exists if and only if $(s_\alpha, s_\beta) \in \Omega_+$, cf. (3.22). This single-locus polymorphism is not spatially homogeneous as the following consideration shows:

If $1 - m_1 \neq m_2$, then equilibrium solutions of (3.21) for $q = 0$ are obtained (due to stochasticity of M) when the fractions $\frac{W_{1,\alpha}}{\overline{W}_\alpha}$ and $\frac{W_{1,\beta}}{\overline{W}_\beta}$ take the value one. This is the case

$(A_2B_1,\ A_2B_1)$	$(A_2B_1,\ A_2B_2)$	$(A_2B_2,\ A_2B_2)$
$1+s_\alpha(\gamma_1-\gamma_2)$	$1+s_\alpha\gamma_1$	$1+s_\alpha(\gamma_1+\gamma_2)$

Table 3.2: Fitness values of genotypes formed by A_2B_1 and A_2B_2

only if $p=1$, which is a monomorphic equilibrium. Therefore, if $1-m_1 \neq m_2$ no spatially homogeneous single-locus polymorphisms exists.

If $1-m_1=m_2$, i.e., in the Levene model, the system is given by (3.18). Explicitly, equilibria must fulfill:

$$2=\frac{1-s_\alpha(\gamma_1+\gamma_2 p)}{1-s_\alpha(\gamma_1-\gamma_2(1-2p))}+\frac{1-s_\beta(\gamma_1+\gamma_2 p)}{1-s_\beta(\gamma_1-\gamma_2(1-2p))}. \tag{3.24}$$

The only internal solution of (3.24) is:

$$\widehat{p}=\frac{1}{2}\left(1-\frac{\gamma_1}{\gamma_2}+\frac{1}{2\gamma_2 s_\beta}+\frac{1}{2\gamma_2 s_\alpha}\right). \tag{3.25}$$

The single-locus polymorphism given by (3.25) must satisfy $\widehat{p}\in(0,1)$. A necessary and sufficient condition for this is

$$2(\gamma_1-\gamma_2)<\frac{1}{s_\alpha}+\frac{1}{s_\beta}<2(\gamma_1+\gamma_2). \tag{3.26}$$

It is easy to see that (3.26) is a special case of (3.22). If $s_\alpha s_\beta<0$ holds, inequality (3.26) can be satisfied because it does not contradict (3.19). For the marginal one-locus system, for every choice of parameters at most one single-locus polymorphism $\widehat{p}$ can exist in the Levene model [22].

It is worth noticing that in both cases the equilibrium frequency of the single-locus polymorphism and its existence depend on the effect of the non segregating locus (i.e., γ_1), where allele A_1 is already fixed in both subpopulations from the beginning.

Two subpopulations formed by A_2B_1 and A_2B_2

The analysis of the marginal one-locus system of the two subpopulations formed by the gametes A_2B_1 and A_2B_2 is obtained analogously to the first case. We use the same notation for the allele frequencies as in the case above. The fitness values of possible genotypes are given in Table 3.2. Using Table 3.2 and (3.17), it is straightforward to calculate the two equations describing the dynamics. As $\overline{W}_\alpha \neq 0 \neq \overline{W}_\beta$, $p_\alpha=\frac{1}{2}(1+\frac{\gamma_1}{\gamma_2}+\frac{1}{\gamma_2 s_\alpha})$ must not hold in any deme.

Recall Theorem 3.2.3 and (3.13b). A single-locus polymorphism exists if the parame-

$(A_1B_1,\ A_1B_1)$	$(A_1B_1,\ A_2B_1)$	$(A_2B_1,\ A_2B_1)$
$1-s_\alpha(\gamma_1+\gamma_2)$	$1-s_\alpha\gamma_2$	$1+s_\alpha(\gamma_1-\gamma_2)$

Table 3.3: Fitness values of genotypes formed by A_1B_1 and A_2B_1

ters are in region

$$\Omega_+ = \left\{(s_\alpha, s_\beta) : s_\alpha s_\beta < 0 \text{ and } \left|m_1(\gamma_1 + \frac{1}{s_\alpha}) + m_2(\gamma_1 + \frac{1}{s_\beta})\right| < \gamma_2\right\}. \tag{3.27}$$

The regions where B_1 and B_2 are protected are

$$\Omega_0 = \left\{(s_\alpha, s_\beta) : m_1(\gamma_1 + \frac{1}{s_\alpha}) + m_2(\gamma_1 + \frac{1}{s_\beta}) > \gamma_2\right\}, \tag{3.28a}$$

$$\Omega_1 = \left\{(s_\alpha, s_\beta) : m_1(\gamma_1 + \frac{1}{s_\alpha}) + m_2(\gamma_1 + \frac{1}{s_\beta}) < -\gamma_2\right\}, \tag{3.28b}$$

respectively. If $1 - m_1 \neq m_2$ and $q = 0$, no single-locus polymorphism exists.

In the Levene model, we obtain as a special case of (3.27) for the frequency of B_1 at the single-locus polymorphism:

$$\widehat{p} = \frac{1}{2}\left(1 + \frac{\gamma_1}{\gamma_2} + \frac{1}{2\gamma_2 s_\beta} + \frac{1}{2\gamma_2 s_\alpha}\right). \tag{3.29}$$

The single-locus polymorphism exists if and only if (3.26) holds. It is the same parameter region as if the two subpopulations are formed by gametes A_1B_1 and A_1B_2 only. The conditions for existence of a single-locus polymorphism when the alleles B_1, B_2 segregate are independent of the allele which is fixed from the beginning at the first locus, i.e., A_1 or A_2. Still, the equilibrium frequencies of the single-locus polymorphisms are different for the two cases.

Two subpopulations formed by A_1B_1 and A_2B_1

Next, we consider two subpopulations formed by the gametes A_1B_1 and A_2B_1. Here, at the second locus B_1 is fixed and at the first locus the alleles A_1 and A_2 segregate. As $p^{(2)}_{1,\alpha} = x_{1,\alpha} + x_{3,\alpha} = 1$, $p^{(1)}_{1,\alpha} = x_{1,\alpha}$, and $p^{(1)}_{2,\alpha} = x_{3,\alpha}$ hold, it suffices to describe the dynamics by $p^{(1)}_{1,\alpha} = p_\alpha$ and $p^{(1)}_{1,\beta} = p_\beta$.

The fitnesses of the genotypes are summarized in Table 3.3. Using (3.17) and 3.3 one can easily derive the equations describing the dynamics. To avoid singularities, $p_\alpha = \frac{1}{2}(1 - \frac{\gamma_2}{\gamma_1} + \frac{1}{s_\alpha\gamma_1})$ must not hold in both demes.

$(A_1B_2,\ A_1B_2)$	$(A_1B_2,\ A_2B_2)$	$(A_2B_2,\ A_2B_2)$
$1 - s_\alpha(\gamma_1 - \gamma_2)$	$1 + s_\alpha\gamma_2$	$1 + s_\alpha(\gamma_1 + \gamma_2)$

Table 3.4: Fitness values of genotypes formed by A_1B_2 and A_2B_2

Recall Theorem 3.2.3 and (3.13c). The region of a protected polymorphism is

$$\Omega_+ = \left\{(s_\alpha, s_\beta) : s_\alpha s_\beta < 0 \text{ and } \left|m_1(-\gamma_2 + \frac{1}{s_\alpha}) + m_2(-\gamma_2 + \frac{1}{s_\beta})\right| < \gamma_1\right\}. \tag{3.30}$$

The regions where A_1 and A_2 are protected are

$$\Omega_0 = \left\{(s_\alpha, s_\beta) : m_1(-\gamma_2 + \frac{1}{s_\alpha}) + m_2(-\gamma_2 + \frac{1}{s_\beta}) > \gamma_1\right\}, \tag{3.31a}$$

$$\Omega_1 = \left\{(s_\alpha, s_\beta) : m_1(-\gamma_2 + \frac{1}{s_\alpha}) + m_2(-\gamma_2 + \frac{1}{s_\beta}) < -\gamma_1\right\}, \tag{3.31b}$$

respectively. If $1 - m_1 \neq m_2$ and $q = 0$ hold, no single-locus polymorphism exists.

If $m_1 = \frac{1}{2} = m_2$, the frequency of A_1 at a single-locus polymorphism becomes

$$\widehat{p} = \frac{1}{2}\left(1 - \frac{\gamma_2}{\gamma_1} + \frac{1}{2\gamma_1 s_\beta} + \frac{1}{2\gamma_1 s_\alpha}\right). \tag{3.32}$$

The necessary and sufficient condition that $\widehat{p} \in (0, 1)$ is

$$2(\gamma_2 - \gamma_1) < \frac{1}{s_\alpha} + \frac{1}{s_\beta} < 2(\gamma_1 + \gamma_2). \tag{3.33}$$

Note that also in this case the equilibrium frequency at the first locus, (3.32), and the condition for existence, (3.33), depend on the allele which is already fixed in both subpopulations from the beginning.

Two subpopulations formed by A_1B_2 and A_2B_2

Next, assume that two subpopulations are formed only by gametes A_1B_2 and A_2B_2. We use the same notation as in the case above. The fitness values of the genotypes are given in Table 3.4. Using 3.4 and (3.17), the dynamics of the system is easily derived. There, the denominators must not be zero and consequently, $p_\alpha = \frac{1}{2}(1 + \frac{\gamma_2}{\gamma_1} + \frac{1}{\gamma_1 s_\alpha})$ must not hold in any deme.

Recall Theorem 3.2.3 and (3.13d). A protected polymorphism exists if

$$\Omega_+ = \left\{(s_\alpha, s_\beta) : s_\alpha s_\beta < 0 \text{ and } \left|m_1(\gamma_2 + \frac{1}{s_\alpha}) + m_2(\gamma_2 + \frac{1}{s_\beta})\right| < \gamma_1\right\} \tag{3.34}$$

holds.

The regions where A_1 and A_2 are protected are

$$\Omega_0 = \left\{(s_\alpha, s_\beta) : m_1(\gamma_2 + \frac{1}{s_\alpha}) + m_2(\gamma_2 + \frac{1}{s_\beta}) > \gamma_1\right\}, \tag{3.35a}$$

$$\Omega_1 = \left\{(s_\alpha, s_\beta) : m_1(\gamma_2 + \frac{1}{s_\alpha}) + m_2(\gamma_2 + \frac{1}{s_\beta}) < -\gamma_1\right\}, \tag{3.35b}$$

respectively. If $1 - m_1 \neq m_2$, no single-locus polymorphisms exist where $q = 0$.

If $m_1 = m_2 = \frac{1}{2}$, an easy calculation yields that

$$\widehat{p} = \frac{1}{2}\left(1 + \frac{\gamma_2}{\gamma_1} + \frac{1}{2\gamma_1 s_\beta} + \frac{a}{2\gamma_1 s_\alpha}\right) \tag{3.36}$$

is the frequency of A_1 at a single-locus polymorphism. If (3.33) holds then $\widehat{p} \in (0, 1)$. This is the same necessary and sufficient condition as for the existence of a single-locus polymorphism in two subpopulations formed of A_1B_1 and A_2B_1 and $q = 0$.

In summary, we have obtained that deme-dependent selection harbors the potential of maintaining single-locus polymorphisms.

3.2.2 Stability in the Levene model

In this section we will briefly study the protection of alleles in the Levene model. For simplicity, we restrict our statements about stability to the marginal one-locus systems. For two diallelic loci with no dominance, the stability behavior may be found in [4, pp. 224-226].

Also, Nagylaki has shown that in the Levene model with two demes the geometric mean fitness

$$\tilde{W} = \prod_\alpha \overline{W}_\alpha^{\frac{1}{2}} \tag{3.37}$$

is a Lyapunov function, cf. [20, p. 216]. This follows from the fact that (i) the mean fitnesses $\overline{W}_\alpha$ depend only on the allele frequencies, cf. (1.4b), that (ii) in the Levene model the allele frequencies are deme independent after one life cycle, and that (iii) the allele frequencies in the next generation are independent of the recombination rate. As we

have a Lyapunov function the existence of a protected polymorphism implies the existence of a single-locus polymorphism. It is stable within its marginal one-locus system ([20]). Complex behavior like periodic orbits can be excluded.

From Section 3.2.1, we recall that two monomorphic equilibria of (3.17) exist, i.e., $p_\alpha = 1\ \forall\alpha$, or $p_\alpha = 0\ \forall\alpha$. As in the Levene case the dynamics reduce to one equation, (1.18), the stability of an equilibrium is determined by the derivative of this mapping evaluated at the equilibrium. Consequently, if we abuse notation and call $f(p) = p'$, where $\widehat{p}$ is an equilibrium of $f(p)$, then $\widehat{p}$ is stable if $|f'(\widehat{p})| < 1$, where the prime, $'$, denotes the first derivative of f (cf. 3.1). Taking the derivative of (3.24) gives the following condition for a stable monomorphic equilibrium where $\widehat{p}_\alpha = 1$ (A_1B_1 fixed) in both demes:

$$-2 < \frac{1-\gamma_1 s_\alpha}{1-(\gamma_1+\gamma_1)s_\alpha} + \frac{1-\gamma_1 s_\beta}{1-(\gamma_1+\gamma_2)s_\beta} < 2. \tag{3.38}$$

It is straightforward to calculate that (3.38) is equivalent to:

$$\frac{1}{s_\alpha} + \frac{1}{s_\beta} > 2(\gamma_1+\gamma_2). \tag{3.39}$$

If (3.39) holds the monomorphic equilibrium $\widehat{p}_\alpha = 1\ \forall\alpha$ is stable. This corresponds to the region Ω_0.

Next we analyze the stability of the monomorphic equilibrium where A_1B_2 is fixed, i.e., $\widehat{p}_\alpha = 0\ \forall\alpha$. The condition that this equilibrium is stable is

$$\frac{1}{s_\alpha} + \frac{1}{s_\beta} < 2(\gamma_1-\gamma_2). \tag{3.40}$$

This corresponds to the region Ω_1 in (3.14).

In the Levene model we have obtained an explicit expression for a single-locus polymorphism, cf. (3.25), and the necessary and sufficient condition for its existence, cf. (3.26). The eigenvalue of (3.18) at the single-locus polymorphism is in the unit circle if

$$s_\alpha s_\beta(3+\gamma_2^2-\gamma_1^2+\gamma_1(s_\alpha+s_\beta)) < 2(s_\alpha^2+s_\beta^2). \tag{3.41}$$

Drawing conclusions for the stability of the single-locus polymorphism via (3.41) is not easy to do. Still, we have found conditions for protected single-locus polymorphisms in Section 3.2.1. The Lyapunov function (3.37) implies its stability in Ω_+, cf. (3.22).

For the other marginal one-locus systems analogous results are obtained. First, consider the marginal one-locus system introduced in Section 3.2.1. The monomorphic equi-

librium where A_2B_1 is fixed is asymptotically stable if

$$\frac{1}{s_\alpha} + \frac{1}{s_\beta} > 2(\gamma_2 - \gamma_1). \tag{3.42}$$

This corresponds to Ω_0.The monomorphic equilibrium where A_2B_2 is fixed in both subpopulations is stable if

$$\frac{1}{s_\alpha} + \frac{1}{s_\beta} < -2(\gamma_1 + \gamma_2). \tag{3.43}$$

Next, recall the marginal one-locus system in Section 3.2.1. The monomorphic equilibrium where A_1B_1 is fixed in both subpopulations is stable if (3.39) holds (region Ω_0). The monomorphic equilibrium where A_2B_1 is fixed is stable if

$$\frac{1}{s_\alpha} + \frac{1}{s_\beta} < 2(\gamma_2 - \gamma_1). \tag{3.44}$$

holds. This corresponds to Ω_1.

Next, recall the marginal one-locus system in Section 3.2.1. The monomorphic equilibrium where A_1B_2 is fixed in both subpopulations is stable if

$$\frac{1}{s_\alpha} + \frac{1}{s_\beta} > 2(\gamma_1 - \gamma_2) \tag{3.45}$$

holds, i.e., in Ω_0. The monomorphic equilibrium where A_2B_2 is fixed in both subpopulations is stable if (3.43) holds.

3.2.3 Full polymorphisms

In the following, we restrict attention to the Levene model. We can expect that finding an internal (thus, fully polymorphic) equilibrium is more involved than finding a single-locus polymorphism as no equation in (1.17) can be omitted.

For uniform selection we have already seen that no full polymorphisms exist, cf. Theorem 3.2.1. In the following assume that the signs of the selection coefficients differ in the two demes.

In the Levene model, results for full polymorphisms can be obtained. Recall that in in the Levene model the Wright manifold is globally attracting if all equilibria are hyperbolic. Thus, internal equilibria are in linkage equilibrium. Also, recall that $q = 0$ holds trivially. For two diallelic loci and no dominance, Bürger has derived in [4, pp. 224–226] the equilibria and their stability properties of the system. He has shown that a line of internal polymorphisms exists if and only if in addition there is no genotype-

environment interaction. Hence, this line of internal equilibria exists nongenerically. This manifold is stable and globally attracting. In our case, generically no internal equilibria exist in the Levene model. A property is called *generic* if it holds on an open dense set of full measure [2, p. 957].

We will explicitly calculate the line of equilibria of (1.18) for the nongeneric set of parameters $s_\alpha = s = -s_\beta$, and $\gamma_1 = \gamma_2 = \gamma$. As $D = 0$ and $q = 0$ hold, it follows from (1.1) that

$$x_2 = \frac{x_1(1 - x_1 - x_3)}{x_1 + x_3}, \tag{3.46a}$$

$$x_4 = \frac{x_3(1 - x_1 - x_3)}{x_1 + x_3}, \tag{3.46b}$$

must hold. Solving (1.18) for internal equilibria fulfilling (3.46), one obtains the following solutions. They must be of the form

$$x_1 = \frac{1}{2}(1 - x_2 - x_3). \tag{3.47}$$

Because of (3.46) the coordinates of the equilibria are of the form:

$$(x_1, x_2, x_3, x_4)^T = (\sqrt{x_3} - x_3, (1 - \sqrt{x_3})^2, x_3, \sqrt{x_3} - x_3)^T, \tag{3.48a}$$

or

$$(x_1, x_2, x_3, x_4)^T = (\sqrt{x_2} - x_2, x_2, (1 - \sqrt{x_2})^2, \sqrt{x_2} - x_2)^T. \tag{3.48b}$$

The vector $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})^T$ fulfills (3.48). Note that the equilibria found in (3.48) can never be hyperbolic as they form a manifold.

Assume (3.48). We will calculate explicitly that $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})^T$, which is part of this manifold, is not hyperbolic. The linearization of (1.18) in $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})^T$ gives the characteristic polynomial

$$\psi(\lambda) = \lambda^2(-1 + \lambda)^2. \tag{3.49}$$

The eigenvalues are the zeros of (3.49). It is straightforward to calculate that the eigenvalues are 0 and 1, both with multiplicity 2. Obviously, this equilibrium is not hyperbolic as it has two eigenvalues equal to one.

For nonuniform selection in the general migration model the existence of internal equilibria still remain an open question.

Chapter 4

No or Weak Selection

We will now study the dynamics of the model (1.17) in the case where migration and recombination are the dominant evolutionary forces and selection is absent or weak. In Section 4.1, we recapitulate results about migration and recombination in the absence of selection. Then, there exists a globally attracting manifold on which global linkage equilibrium holds. Additionally, allele frequencies are identical across demes at equilibria. In Section 4.2, selection will be weak relative to migration and recombination. Results will be presented which show that the dynamics can be derived as a perturbation of the so called weak-selection limit. Furthermore, every equilibrium of the perturbation has the same stability properties as in the unperturbed system.

4.1 No Selection

In this section we briefly study our model with migration and recombination in the absence of selection. Consequently, all fitness values are 1 and (1.17) reduces to

$$x'_{i,\alpha} = \sum_{\beta} m_{\alpha\beta}(x_{i,\alpha} - \eta_i c D_\alpha). \tag{4.1}$$

Using the notation from (1.23), the dynamical system reads

$$x'_i = M(x_i - \eta_i c D). \tag{4.2}$$

It is obvious that for the allele frequencies

$$p_i^{(j)\prime} = M p_i^{(j)}, \text{ for } i, j \in \{1, 2\} \tag{4.3}$$

holds, as this follows from (1.1). From (1.25) and (1.20) one can easily derive that the allele frequencies are constant within the generations, i.e.,

$$p_i^{(j)} = p_i^{(j)}(t) = p_i^{(j)}(0). \tag{4.4}$$

Recall the definitions (1.20) and (1.25). From [2, pp. 947 – 955] the following theorem is known:

Theorem 4.1.1 (Dynamics without selection)**.** *Suppose that* (4.1) *holds and M is ergodic. Then the manifold*

$$\Psi_0 = \{x \in \Delta_4^\Gamma : x_i = \widehat{X}_i e\} = \{p \in \Delta_4^\Gamma : D = 0 \text{ and } q = 0\} \tag{4.5}$$

is invariant under (4.1) *and globally attracting at a uniform geometric rate. Moreover, every point on Ψ_0 is an equilibrium point. Therefore, linkage equilibrium and spatial homogeneity are quickly approached under recombination and ergodic migration.*

Proof. A proof for the general multiallelic multilocus migration-selection model is given in [2, pp. 950 – 954]. Of course, our two-locus model is just a special case and the proof also applies. □

Note that an equivalent representation of Ψ_0 is

$$\begin{aligned}\Psi_0 = \big\{x \in \Delta_4^\Gamma : {} & x_{1,\alpha} = P_1P_2, x_{2,\alpha} = P_1(1-P_2),\\ & x_{3,\alpha} = (1-P_1)P_2, x_{4,\alpha} = (1-P_1)(1-P_2)\ \forall\alpha, \text{ where } q = 0\big\}.\end{aligned} \tag{4.6}$$

Generically, the rate of approach is $\max\{|\lambda_1|, 1-c\}$, where λ_1 denotes the non-unit eigenvalue of M with largest modulus and c denotes the recombination rate, cf. [2, p. 949].

The equilibria in (4.5) cannot be hyperbolic because they are not isolated. As $q = 0$ and $D = 0$ hold, every equilibrium must satisfy (3.46). Linearizing (1.17) for $s_\alpha = s_\beta = 0$ at such equilibria yields the following eigenvalues: 1 with multiplicity 3, $1-m_{\alpha\beta}-m_{\beta\alpha}$ with multiplicity 3, and the simple eigenvalues $1-c$, and $(1-c)(1-m_{\alpha\beta}-m_{\beta\alpha})$. Thus, in the direction of the equilibrium manifold (4.5) the eigenvalues are 1. All other eigenvalues are smaller than one in modulus and all trajectories are attracted by the equilibrium manifold.

Figure 4.1 displays the trajectories from numerical iterations of the recursion (1.17) for a set of initial conditions. In both demes they to the equilibrium point $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})^T$. Obviously, $D = 0$ and $q = 0$ holds there.

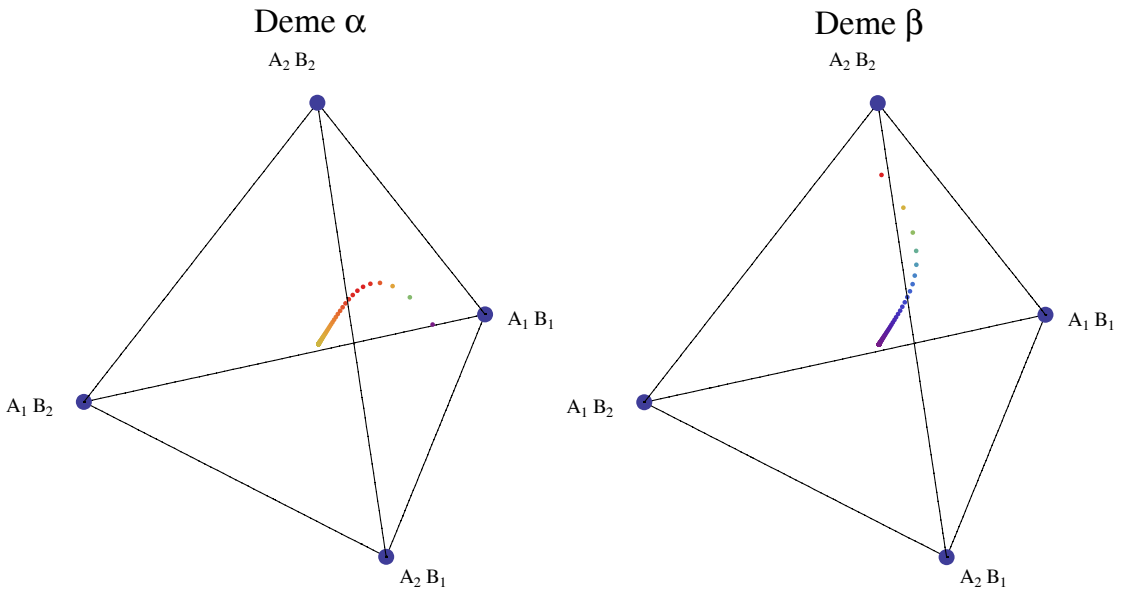

Figure 4.1: Trajectories for the model (1.17) with recombination and migration but no selection. Simulation parameters are: $m_{\alpha\beta} = 0.2$, $m_{\alpha\beta} = 0.2$, $c = 0.1$ (in both demes). Selection is absent, i.e., $s_\alpha = s_\beta = 0$. Initial frequencies in deme α and β are $(0.80, 0.10, 0.05, 0.05)^T$ and $(0.05, 0.05, 0.10, 0.80)^T$, respectively. The number of iteration steps is 123.

4.2 Weak Selection

In this section we investigate the case when selection is weak relative to migration and recombination. Bürger has shown previously in [2] that in this case the dynamics is obtained as a perturbation of the system (4.1). To this aim, let us write the fitnesses in deme $\alpha \in G$ as

$$w_{ij,\alpha} = 1 + \epsilon r_{ij,\alpha}, \tag{4.7}$$

where $\epsilon \geq 0$ is sufficiently small and $|r_{ij,\alpha}| \leq 1$. The case $\epsilon = 0$ corresponds to (4.1). The $r_{ij,\alpha}$ are fixed and denote the selection coefficient rates of genotype ij in deme α. Migration and recombination rates are assumed fixed, so that compared to them fitness differences are small. We define the selection coefficients of gamete $i \in I$ and of the entire population as

$$r_{i,\alpha}(x_\alpha) = \sum_j r_{ij,\alpha} x_{j,\alpha} \text{ and } \bar{r}_\alpha(x_\alpha) = \sum_{i,j} r_{ij,\alpha} x_{i,\alpha} x_{j,\alpha}. \tag{4.8}$$

From (1.1) and (4.7) we deduce the fitness of gamete i and the mean fitness:

$$w_{i,\alpha} = 1 + \epsilon r_{i,\alpha}(x_\alpha) \text{ and } \bar{w}_\alpha(x_\alpha) = 1 + \epsilon \bar{r}_\alpha(x_\alpha). \tag{4.9}$$

Next, we introduce the vector of allele frequencies in deme α at the two loci and the vector of average allele frequencies at both loci,

$$\rho_\alpha = (p_\alpha^{(1)}, p_\alpha^{(2)})^T \in \mathbb{R}^2, \tag{4.10a}$$

$$\pi = (P_1, P_2)^T \in \mathbb{R}^2, \tag{4.10b}$$

where P_1, P_2 are defined as in (1.25). The idea behind this approach is that for strong migration and weak selection, the population differentiation is weak and that selection acts approximately with averaged fitnesses throughout the entire population. From (4.4) it follows that without selection the average allele frequencies, π, are time independent. This is not the case here as selection is present. Instead of x, we will use π, D, and q (cf. (1.23) and (1.26)) to analyze (1.17). Recall the definition of the linkage-equilibrium manifold in deme α, Λ_α, (1.14). Then, we can derive from (4.8) the selection coefficients of gamete i, allele i_n at locus n, and of the entire population in deme α on Λ_α:

$$r_{i,\alpha} = r_{i,\alpha}(\rho_\alpha) = \sum_j r_{ij,\alpha} \prod_k p_{j_k,\alpha}^{(k)}, \tag{4.11a}$$

$$r_{i_n,\alpha}^{(n)} = r_{i_n,\alpha}^{(n)}(\rho_\alpha) = \sum_{i|i_n} r_{i,\alpha}(\rho_\alpha) \prod_{k:k\neq n} p_{i_k,\alpha}^{(k)}, \tag{4.11b}$$

$$\bar{r}_\alpha(\rho_\alpha) = \bar{r}_\alpha(\rho_\alpha) = \sum_i r_{i,\alpha}(\rho_\alpha) \prod_k p_{i_k,\alpha}^{(k)}. \tag{4.11c}$$

The sum $\sum_{i|i_n}$ runs over all multi-indices $i \in \{11, 12, 21, 22\}$ with the components at locus n fixed as i_n. Now we specialize to two diallelic loci and change notation. Subsequently, we use indices $i \in \{1, 2, 3, 4\}$ instead of $i \in \{11, 12, 21, 22\}$ for the gametes. Then (4.11a) and (4.11b) become

$$\begin{aligned} r_{i,\alpha} = {} & r_{i1,\alpha} p_\alpha^{(1)} p_\alpha^{(2)} + r_{i2,\alpha} p_\alpha^{(1)} (1 - p_\alpha^{(2)}) \\ & + r_{i3,\alpha}(1 - p_\alpha^{(1)}) p_\alpha^{(2)} + r_{i4,\alpha}(1 - p_\alpha^{(1)})(1 - p_\alpha^{(2)}), \end{aligned} \tag{4.12a}$$

$$r_{1,\alpha}^{(1)} = r_{1,\alpha} p_\alpha^{(2)} + r_{2,\alpha}(1 - p_\alpha^{(2)}), \tag{4.12b}$$

$$r_{1,\alpha}^{(2)} = r_{1,\alpha} p_\alpha^{(1)} + r_{3,\alpha}(1 - p_\alpha^{(1)}). \tag{4.12c}$$

As in (1.20), let μ denote the principal left eigenvector of the migration matrix M. We introduce the average selection coefficients of genotype ij, gamete i, allele i_n at locus n, and of the entire population:

$$\omega_{ij} = \sum_\alpha \mu_\alpha r_{ij,\alpha}, \tag{4.13a}$$

$$\omega_i(\pi) = \sum_j \omega_{ij} \prod_k P_{j_k}^{(k)} = \sum_\alpha \mu_\alpha r_{i,\alpha}(\pi), \tag{4.13b}$$

$$\omega_{i_n}^{(n)}(\pi) = \sum_{i|i_n} \omega_i(\pi) \prod_{k:k\neq n} P_{i_k}^{(k)} = \sum_\alpha \mu_\alpha r_{i_n,\alpha}^{(n)}(\pi), \tag{4.13c}$$

$$\bar\omega(\pi) = \sum_i \omega_i(\pi) \prod_k P_{i_k}^{(k)} = \sum_\alpha \mu_\alpha \bar r_\alpha(\pi). \tag{4.13d}$$

For $\bar\omega$, we obtain the alternative representation

$$\bar\omega(\pi) = \sum_n \sum_{i_n} \omega_{i_n}^{(n)} P_{i_n}^{(n)} = \sum_{i,j} \omega_{ij} (\prod_n P_{i_n}^{(n)})(\prod_k P_{j_k}^{(k)}). \tag{4.14}$$

A simple calculation shows that

$$\frac{d\bar\omega(\pi)}{dP_{i_n}^{(n)}} = 2\omega_{i_n}^{(n)}(\pi) \tag{4.15}$$

holds, where in the differentiation the $P_{i_n}^{(n)}$ are used as independent variables. For two loci and two alleles, we obtain from (4.13c)

$$\begin{aligned}
\omega_1^{(1)} =& P_1[P_2^2\omega_{11} + 2P_2(1-P_2)\omega_{12} + (1-P_2)^2\omega_{22}] \\
&+ (1-P_1)[P_2^2\omega_{13} + 2P_2(1-P_2)\omega_{14} + (1-P_2)^2\omega_{24}], && (4.16a)\\
\omega_2^{(1)} =& P_1[P_2^2\omega_{13} + 2P_2(1-P_2)\omega_{14} + (1-P_2)^2\omega_{24}] \\
&+ (1-P_1)[P_2^2\omega_{33} + 2P_2(1-P_2)\omega_{34} + (1-P_2)^2\omega_{44}], && (4.16b)\\
\omega_1^{(2)} =& P_2[P_1^2\omega_{11} + 2P_1(1-P_1)\omega_{13} + (1-P_1)^2\omega_{33}] \\
&+ (1-P_2)[P_1^2\omega_{12} + 2P_1(1-P_1)\omega_{14} + (1-P_1)^2\omega_{34}], && (4.16c)\\
\omega_2^{(2)} =& P_2[P_1^2\omega_{21} + 2P_1(1-P_1)\omega_{23} + (1-P_1)^2\omega_{43}] \\
&+ (1-P_2)[P_1^2\omega_{22} + 2P_1(1-P_1)\omega_{24} + (1-P_1)^2\omega_{44}]. && (4.16d)
\end{aligned}$$

From (4.13d) we obtain

$$\bar{\omega}(\pi) = P_1\omega_1^{(1)} + (1-P_1)\omega_2^{(1)} \tag{4.17a}$$
$$= P_2\omega_1^{(2)} + (1-P_2)\omega_2^{(2)}. \tag{4.17b}$$

To state the first theorem we introduce the *weak-selection limit* of (1.17) on Δ_4^Γ:

$$\frac{dP_n}{dt} = P_n[\omega_1^{(n)}(\pi) - \bar{\omega}(\pi)] \text{ for } n = 1, 2, \tag{4.18a}$$
$$D = 0, q = 0. \tag{4.18b}$$

For two loci, we obtain from (4.17)

$$\omega_1^{(n)} - \bar{\omega} = (1-P_n)\left(\omega_1^{(n)} - \omega_2^{(n)}\right) \text{ for } n = 1, 2. \tag{4.19}$$

Therefore, (4.18a) becomes

$$\frac{dP_n}{dt} = P_n(1-P_n)\left(\omega_1^{(n)} - \omega_2^{(n)}\right) \text{ for } n = 1, 2. \tag{4.20}$$

In view of the following theorem, it is more convenient to consider (4.18) as a system of ordinary differential equations on Δ_4^Γ instead of (4.18a) on $\mathbb{R}^2$. For practical applications on the other hand, it will be much more convenient to first study (4.18a) on $\mathbb{R}^2$.

The differential equation (4.18a) is a *Svirezhev-Shashahani gradient* with potential function $\bar{\omega}$. In Section 3.1, we have introduced the definition of a gradient system. An equation of the form (4.18a) is a gradient system with respect to the *Shahshahani inner product*, which is given by

$$\langle \xi, \zeta \rangle_x = \sum_{i=1}^{k} \frac{1}{x_i}\xi_i\zeta_i \tag{4.21}$$

for tangent vectors $\xi, \zeta \in T_x\Delta_k$, i.e., the tangent space in $x \in \Delta_k$. For more details about Svirezhev-Shahsahani gradients we refer to [14, pp. 257–261], [5, p. 42], or [1, pp. 37–43].

Because of

$$\frac{d\bar{\omega}}{dt} = 2\sum_n \sum_{i_n} P_{i_n}^{(n)}[\omega_{i_n}^{(n)}(\pi) - \bar{\omega}(\pi)]^2 \geq 0, \tag{4.22}$$

$\bar{\omega}$ increases strictly along nonconstant solutions of (4.18a). Hence, $\bar{\omega}$ is a Lyapunov function for the system (4.18a), cf. Section 3.1. In the following, we will need the assumption that all equilibria of (4.18a) are hyperbolic. Hyperbolicity is a generic property for systems of the form (4.18a) [23].

Theorem 4.2.1 (Dynamics under weak selection)**.** *Suppose that* (1.17) *holds, and that the backward migration matrix M is ergodic and fixed. Let the linearized system of* (4.18a) *be hyperbolic, the recombination rate c fixed, and $\epsilon > 0$ sufficiently small.*

(a) The set of equilibria $\Xi_0 \subset \Delta_4^\Gamma$ of (4.18a) *contains only isolated points, as does the set of equilibria $\Xi_\epsilon \subset \Delta_4^\Gamma$ of the dynamical system* (1.17)*. As $\epsilon \to 0$, each equilibrium in Ξ_ϵ converges to the corresponding equilibrium in Ξ_0.*

(b) In the neighborhood of each equilibrium in Ξ_0, there exists exactly one equilibrium point in Ξ_ϵ. The stability of each equilibrium in Ξ_ϵ is the same as that of the corresponding equilibrium in Ξ_0. Thus, each pair is either asymptotically stable or unstable.

(c) Every solution $x(t)$ of (1.17) *converges to one of the equilibrium points in Ξ_ϵ.*

Proof. The above theorem is a special case with $n = 1, 2$, $I_n = 2$, $L = 2$ of Theorem 4.3 in [2, pp. 958–960]. Therefore, its proof also applies to our model. □

In summary, the importance of Theorem 4.2.1 is that the exact dynamics for weak selection can be perceived as a perturbation of the weak-selection limit. This is much easier to study because of linkage equilibrium and deme-independent allele frequencies. Under weak selection, the exact dynamics quickly lead to *quasi-linkage equilibrium*, and *spatial quasi-homogeneity*, cf. [2, p. 962, Remark 4.9]. The quasi-linkage equilibrium manifold Ψ_ϵ is an invariant manifold close to Ψ_0, cf. (4.5), and $\Xi_\epsilon \subseteq \Psi_\epsilon$ holds. From Theorem 4.2.1 it follows that the dynamics on Ψ_ϵ can be described by a small perturbation of the system without selection, (4.1).

For the case of no epistasis and weak selection relative to other evolutionary forces, we can formulate a result that even excludes polymorphic equilibria:

Theorem 4.2.2 (Weak selection and no epistasis)**.** *If selection is nonepistatic and sufficiently weak relative to migration and recombination, then generically, there exists a globally asymptotically stable monomorphic state if single-locus fitnesses are additive and there is no dominance at both loci. In particular, no polymorphism is maintained.*

Proof. This Theorem is part of Proposition 2.1 in [3, pp. 987–988] and applies to our case. □

4.2.1 Analyzing the Weak Selection Limit

In this subsection we apply the theory of Section 4.2 to our general model (1.17). We shall gain insight into existence of equilibria and their stability by assuming weak selection.

As fitness values are of the form (4.7), we rewrite the fitness matrix (1.8) as follows

$$\begin{array}{c} \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array} \begin{array}{c} \begin{array}{cccc} A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 \end{array} \\ \begin{pmatrix} 1-\epsilon s_\alpha(\gamma_1+\gamma_2) & 1-\epsilon s_\alpha\gamma_1 & 1-\epsilon s_\alpha\gamma_2 & 1 \\ 1-\epsilon s_\alpha\gamma_1 & 1-\epsilon s_\alpha(\gamma_1-\gamma_2) & 1 & 1+\epsilon s_\alpha\gamma_2 \\ 1-\epsilon s_\alpha\gamma_2 & 1 & 1+\epsilon s_\alpha(\gamma_1-\gamma_2) & 1+\epsilon s_\alpha\gamma_1 \\ 1 & 1+\epsilon s_\alpha\gamma_2 & 1+\epsilon s_\alpha\gamma_1 & 1+\epsilon s_\alpha(\gamma_1+\gamma_2) \end{pmatrix} \end{array}. \tag{4.23}$$

From (4.12) and (4.23) we calculate the marginal and mean selection coefficients in deme α:

$$r_{1,\alpha} = -s_\alpha[\gamma_1 p_\alpha^{(1)} + \gamma_2 p_\alpha^{(2)}], \tag{4.24a}$$
$$r_{2,\alpha} = s_\alpha[\gamma_2(1-p_\alpha^{(2)}) - \gamma_1 p_\alpha^{(1)}], \tag{4.24b}$$
$$r_{3,\alpha} = s_\alpha[\gamma_1(1-p_\alpha^{(1)}) - \gamma_2 p_\alpha^{(2)}], \tag{4.24c}$$
$$r_{4,\alpha} = s_\alpha[\gamma_1(1-p_\alpha^{(1)}) + \gamma_2(1-p_\alpha^{(2)})], \tag{4.24d}$$
$$\bar{r}_\alpha = s_\alpha[\gamma_1(1-2p_\alpha^{(1)}) + \gamma_2(1-2p_\alpha^{(2)})]. \tag{4.24e}$$

For simplicity, we consider in the following only two demes and we write $s_\alpha = s_1$, $s_\beta = s_2$, $m_{\alpha\beta} = m_1$, $m_{\beta\alpha} = m_1$, $p_\alpha^{(1)} = p_1^{(1)}$, $p_\beta^{(1)} = p_2^{(1)}$, $p_\alpha^{(2)} = p_1^{(2)}$, and $p_\beta^{(2)} = p_2^{(2)}$. Using (1.20) and (4.10), we derive the average allele frequencies:

$$P_1 = \frac{m_2}{m_1+m_2} p_1^{(1)} + \frac{m_1}{m_1+m_2} p_2^{(1)}, \tag{4.25a}$$
$$P_2 = \frac{m_2}{m_1+m_2} p_1^{(2)} + \frac{m_1}{m_1+m_2} p_2^{(2)}. \tag{4.25b}$$

The average selection coefficients are derived according to (4.16) using (4.25):

$$\omega_1^{(1)} = \frac{[\gamma_2(1-2P_2) - \gamma_1 P_1](m_2 s_1 + m_1 s_2)}{m_1+m_2}, \tag{4.26a}$$
$$\omega_2^{(1)} = \frac{[\gamma_2(1-2P_2) + \gamma_1(1-P_1)](m_2 s_1 + m_1 s_2)}{m_1+m_2}, \tag{4.26b}$$
$$\omega_1^{(2)} = \frac{[\gamma_1(1-2P_1) - \gamma_2 P_2](m_2 s_1 + m_1 s_2)}{m_1+m_2}, \tag{4.26c}$$
$$\omega_2^{(2)} = \frac{[\gamma_1(1-2P_1) + \gamma_2(1-P_2)](m_2 s_1 + m_1 s_2)}{m_1+m_2}. \tag{4.26d}$$

The equations (4.17) transforms to

$$\bar{\omega} = \frac{[\gamma_1(1-2P_1)+\gamma_2(1-2P_2)](m_2s_1+m_1s_2)}{m_1+m_2}. \tag{4.27}$$

To calculate the weak-selection limit we need to compute the differences $\omega_1^{(1)} - \omega_2^{(1)}$, $\omega_1^{(2)} - \omega_2^{(2)}$, which are:

$$\omega_1^{(1)} - \omega_2^{(1)} = -\frac{\gamma_1(m_2s_1+m_1s_2)}{m_1+m_2}, \tag{4.28a}$$

$$\omega_1^{(2)} - \omega_2^{(2)} = -\frac{\gamma_2(m_2s_1+m_1s_2)}{m_1+m_2}. \tag{4.28b}$$

Consequently, we have obtained the following two differential equations describing the weak-selection limit (4.20) for the additive diallelic two-locus model with no dominance:

$$\frac{dP_1}{dt} = -P_1(1-P_1)\frac{\gamma_1(m_2s_1+m_1s_2)}{m_1+m_2}, \tag{4.29a}$$

$$\frac{dP_2}{dt} = -P_2(1-P_2)\frac{\gamma_2(m_2s_1+m_1s_2)}{m_1+m_2}, \tag{4.29b}$$

where $D = 0$ and $q = 0$. We note that (4.29) is a decoupled system of two ordinary differential equations, which is easy to study. As P_1, P_2 are frequencies and thus nonnegative, the signs of the frequency changes in (4.29) depend only on the sign of $m_2s_1 + m_1s_2$. This factor is the same for $\dot{P}_1$ and $\dot{P}_2$.

We can use (4.29) to calculate the time derivative of the mean fitness, cf. (4.27):

$$\frac{d\bar{\omega}}{dt} = \frac{2(m_2s_1+m_1s_2)^2}{(m_1+m_2)^2}[\gamma_1^2(1-P_1)P_1 + \gamma_2^2(1-P_2)P_2], \tag{4.30}$$

which shows directly that

$$\frac{d\bar{\omega}}{dt} \geq 0 \tag{4.31}$$

holds, cf. (4.22). No periodic orbits or other complex dynamics can occur. According to Theorem 4.2.1, equilibria for (4.29) will allow us to predict equilibria of the dynamical system (1.17). Equilibria of the dynamical system (4.29) must fulfill the conditions

$$\frac{dP_1}{dt} = 0 \text{ and } \frac{dP_2}{dt} = 0. \tag{4.32}$$

Generically, (i.e., for the open dense set of parameters $s_1, s_2, \gamma_1, \gamma_2, m_1, m_2$ where $m_1s_2 + m_2s_1 \neq 0$), the first equation in (4.29) is zero only if:

(a) $P_1 = 0$, i.e., $p^{(1)} = 0$ (A_2 is fixed in both populations), or

(b) $P_1 = 1$, i.e., $p^{(1)} = 1$ (A_1 is fixed in both populations),

and the second equation in (4.29) is zero only if:

(c) $P_2 = 0$, i.e., $p^{(2)} = 0$ (B_2 is fixed in both populations), or

(d) $P_2 = 1$, i.e., $p^{(2)} = 1$ (B_1 is fixed in both populations).

Combining these conditions we obtain that generically the equilibium solutions of the weak-selection limit (4.29) are exactly the monomorphic equilibria with $q = 0$, cf. (3.2). This is in accordance with Theorem 4.2.2. In the generic case, the mean fitness and its time derivative is not degenerate and $\frac{d\bar{\omega}}{dt} = 0$ holds only at the equilibria; otherwise, $\frac{d\bar{\omega}}{dt} > 0$ holds. As $\bar{\omega}$ is a Lyapunov function, we can derive stability results by considering the sign of

$$m_2 s_1 + m_1 s_2. \tag{4.33}$$

Recall that in the generic case the monomorphic equilibria are hyperbolic [23]. We have to distinguish the following cases (where $m_1, m_2 \in (0, \frac{1}{2}]$):

1. Selection acts in the same direction in both demes: If $s_1, s_2 > 0$ it follows that $\frac{dP_1}{dt} < 0$ and $\frac{dP_2}{dt} < 0$ for $P_1, P_2 \in (0, 1)$. Consequently, the equilibria where A_1B_1, A_1B_2, or A_2B_1 are fixed are unstable as any perturbation of these points will lead to fixation of the gamete A_2B_2. Therefore, the equilibrium where A_2B_2 is fixed is globally asymptotically stable. If $s_1, s_2 < 0$ it follows that $\frac{dP_1}{dt} > 0$ and $\frac{dP_2}{dt} > 0$ for $P_1, P_2 \in (0, 1)$. Consequently, the equilibria where A_1B_2, A_2B_1, or A_2B_2 are fixed are unstable. The monomorphic equilibrium where A_1B_1 is fixed is globally asymptotically stable.

2. Selection acts in different directions: $s_1 > 0 > s_2$ or $s_1 < 0 < s_2$. If $s_1 m_2 < -s_2 m_1$, then $\frac{dP_1}{dt} > 0$ and $\frac{dP_2}{dt} > 0$, i.e., the monomorphic equilibrium where A_1B_1 is fixed is globally asymptotically stable. If $s_1 m_2 > -s_2 m_1$, then $\frac{dP_1}{dt} < 0$ and $\frac{dP_2}{dt} < 0$, i.e., the equilibrium where A_2B_2 is fixed is globally asymptotically stable. We analyze the case $s_1 m_2 = -s_2 m_1$ below.

To sum up, we have proved that generically, there are only monomorphic equilibria. As the mean fitness is a Lyapunov function, we can exclude complex dynamic behavior. We obtained that there is always one globally asymptotically stable monomorphic state. Stability depends on the sign of (4.33).

As $\gamma_1 \geq \gamma_2 > 0$, the nongeneric case, when (4.32) holds, is

$$s_1 m_2 = -s_2 m_1. \tag{4.34}$$

This can only hold if the nonzero selection coefficients change signs between the two demes. Obviously, (4.34) holds in the symmetric case when $s_1 = -s_2$ and $m_1 = m_2$. Note that any deviation of these parameters eventually leads to the dynamics in the generic case. If (4.34) holds, the mean fitness and its time derivative, cf. (4.27) and (4.30), are always zero. In this degenerate case, any combination of P_1, P_2 is an equilibrium and thus the whole manifold $D = 0, q = 0$ consists of equilibria.

Chapter 5

No or Weak Migration

Here, we study the dynamics of the full system (1.17) in another limiting case. First, we analyze the model without migration. Then, we introduce weak migration as a perturbation of the system with no migration. Finally, we investigate (1.17) under weak migration and weak selection.

5.1 No Migration

In this section we assume that no migration between demes occurs. Therefore, the two subpopulations turn into two independent panmictic populations. In each deme the dynamics of the gamete frequencies is then described by an additive diallelic two-locus selection model. For a panmictic population with no epistasis and partial dominance (and in particular no dominance, i.e., additivity) the gamete with the highest fitness (if there exists exactly one) becomes fixed and polymorphic equilibria do not exist. Each population exhibits the following modified life cycle:

$$\boxed{\text{zygote} \xrightarrow{selection} \text{adult} \xrightarrow{recombination} \text{gamete} \xrightarrow{mating} \text{zygote}}$$

5.1.1 The Additive Two-Locus Two-Allele Model

As no migration between demes occurs, the migration rates $m_{\alpha\beta}$ for $\alpha \neq \beta$ are set to zero. We can consequently neglect the geographical structure of the population and we omit deme-specific subscripts in all formulas. Therefore, the dynamical system reduces to a selection-recombination model

$$x_i^{'} = x_i \frac{W_i - \eta_i c D}{\overline{W}}, \tag{5.1}$$

in analogy to (1.17a). The description of the additive two-locus two-allele model follows [5, pp. 46–50]. We construct the fitness matrix W of our system by first defining the fitness values for each locus. To do so, we assume the fitness values of the single-locus genotypes to be

$$\begin{array}{cccccc} A_1A_1 & A_1A_2 & A_2A_2 & B_1B_1 & B_1B_2 & B_2B_2 \\ \frac{1}{2}-s\gamma_1 & \frac{1}{2} & \frac{1}{2}+s\gamma_1 & \frac{1}{2}-t\gamma_2 & \frac{1}{2} & \frac{1}{2}+t\gamma_2, \end{array} \tag{5.2}$$

where s,t denote the selection coefficients and $|s|, |t|$ are sufficiently small. Note that this model is exactly (1.9). Within a deme the strength of selection acting on the two loci may be different and we have linear directional selection at both loci. As in (1.9) only the parameter combinations $s\gamma_1$ and $t\gamma_2$ occur, we can rewrite the matrix as

$$\begin{array}{c} \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array} \begin{array}{c} \begin{array}{cccc} A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 \end{array} \\ \begin{pmatrix} 1-(\sigma_1+\sigma_2) & 1-\sigma_1 & 1-\sigma_2 & 1 \\ 1-\sigma_1 & 1-\sigma_1+\sigma_2 & 1 & 1+\sigma_2 \\ 1-\sigma_2 & 1 & 1+\sigma_1-\sigma_2 & 1+\sigma_1 \\ 1 & 1+\sigma_2 & 1+\sigma_1 & 1+\sigma_1+\sigma_2 \end{pmatrix} \end{array}, \tag{5.3}$$

where $\sigma_1 = s\gamma_1$ and $\sigma_2 = t\gamma_2$. The mean fitnesses $\bar{a}$ and $\bar{b}$ of the single loci are

$$\bar{a} = \frac{1}{2} - \sigma_1(2p_1^{(1)} - 1), \tag{5.4a}$$

$$\bar{b} = \frac{1}{2} - \sigma_2(2p_1^{(2)} - 1). \tag{5.4b}$$

Subsequently, we may express $\overline{W}$ as

$$\overline{W} = \bar{a} + \bar{b}. \tag{5.5}$$

We observe that $\overline{W}$ is linear in $p^{(1)}$ and $p^{(2)}$. The additive diallelic two-locus model is one of the few classes of two-locus models, in which mean fitness is nondecreasing. Therefore, the dynamics is rather simple and several aspects can be inferred from one-locus theory.

Theorem 5.1.1 ([9])**.** *Assume that fitnesses in a diallelic two-locus selection model are additive. Then, the following holds:*

(i) The mean fitness is nondecreasing, i.e., $\overline{W}(p_1^{(1)'}, p_1^{(2)'}) \geq \overline{W}(p^{(1)}, p^{(2)})$ *and* $\overline{W}(p^{(1)'}, p^{(2)'}) = \overline{W}(p^{(1)}, p^{(2)})$ *if and only if* $p^{(1)'} = p^{(1)}$ *and* $p^{(2)'} = p^{(2)}$.

(ii) If there is a fittest gamete, it gets fixed, i.e., the corresponding monomorphic eigenvalue is globally asymptotically stable.

Proof. Recall equation (5.1).

(i) The gene-frequency changes

$$\Delta p^{(k)} = p^{(k)'} - p^{(k)} \text{ for } k = 1, 2 \tag{5.6}$$

are independent of c and they are the same as in the special case $c = 0$. The mean fitness changes in the same way as if $c = 0$, and so our problem reduces to a one-locus four-allele system, where the mean fitness is nondecreasing (*The Fundamental Theorem of Natural Selection*, [5, pp. 33–35]). At equilibrium, $\Delta p^{(k)} = 0$ holds for $k = 1, 2$. Consequently, $\Delta \overline{W} = \overline{W}(p^{(1)'}, p^{(2)'}) - \overline{W}(p^{(1)}, p^{(2)}) = 0$ holds at equilibrium, i.e., the mean fitness remains constant at equilibrium.

(ii) Next, we show that the mean fitness is maximized at one monomorphic state. As (5.5) holds, the mean fitness is maximized if and only if $\overline{a}$ and $\overline{b}$ are maximized. From (5.4), we observe that $\overline{W}$ is maximized if and only if the fittest gamete is present. Therefore, there is only one stable equilibrium which coincides with a vertex, i.e., the fittest double homozygote gets fixed (if there exists exactly one), and there is no interior equilibrium point.

□

Recall that, generically, every equilibrium of (5.1) is hyperbolic in the absence of epistasis [2, p. 971]. In Figure 5.1 examples of trajectories converging to the fittest gamete are shown.

5.1.2 Extending the diallelic two-locus model to two demes

In this section we will apply the model and the results from Section 5.1.1 to a population inhabiting two demes which are completely isolated, i.e., there is no migration between them. Note that selection may act in different directions in both demes. The dynamics of the gamete frequencies in the two demes is independent of each other, i.e., both populations are treated as two independent panmictic populations.

Although the two subpopulations are independent of each other, the dynamic behavior of the overall population is given on $\Delta_4 \times \Delta_4$. In Theorem 5.1.1 we have shown that for directional selection in a panmictic population where there is a fittest gamete, this

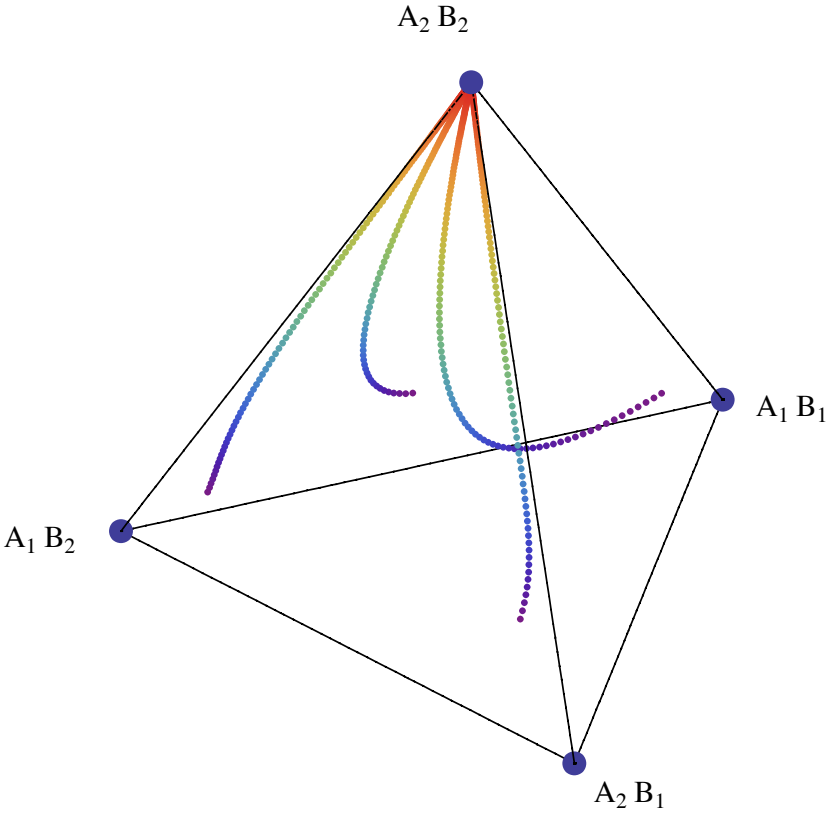

Figure 5.1: Trajectories for the diallelic two-locus model without migration for several initial frequencies. The different initial frequencies of A_1B_1, A_1B_2, A_2B_1, and A_2B_2 are $(0.05, 0.1, 0.7, 0.15)^T$, $(0.8, 0.05, 0.05, 0.1)^T$,$(0.05, 0.8, 0.05, 0.1)^T$, and $(0.3, 0.4, 0.05, 0.25)^T$. The fittest gamete is A_2B_2 and in Theorem 5.1.1 we have shown that this gamete becomes always fixed ($x_4 = 1$). (Parameters: $s = t = 0.05$, $c = 0.1$, $\gamma_1 = \gamma_2 = 1$, 3000 iteration steps)

gamete will get fixed after sufficiently long time. Thus, there is one monomorphic equilibrium which is globally asymptotically stable. In addition, there are the three other monomorphic equilibria which are all unstable.

If we now look at two isolated subpopulations, the number of equilibria increases greatly. Every combination of equilibria in deme α and deme β is an equilibrium of the extended system:

$$x'_{i,\alpha} = x_{i,\alpha}\frac{W_{i,\alpha} - \eta_i c D_\alpha}{\overline{W}_\alpha}, \tag{5.7a}$$

$$x'_{i,\beta} = x_{i,\beta}\frac{W_{i,\beta} - \eta_i c D_\beta}{\overline{W}_\beta}. \tag{5.7b}$$

Note that no equilibria except the combinations mentioned above exist. Thus, there are no internal equilibria or equilibria in the interior of the edges of the simplices. Although restricted to a subpopulation, every equilibrium is monomorphic, this is not necessarily true for the whole population. In the state space Δ_4^2, there are four momomorphic equilibria:

$$\widehat{x}^1 = (1,0,0,0,1,0,0,0)^T, \tag{5.8a}$$

$$\widehat{x}^2 = (0,1,0,0,0,1,0,0)^T, \tag{5.8b}$$

$$\widehat{x}^3 = (0,0,1,0,0,0,1,0)^T, \tag{5.8c}$$

$$\widehat{x}^4 = (0,0,0,1,0,0,0,1)^T. \tag{5.8d}$$

There are eight equilibria that are single-locus polymorphisms, four where at the first locus one allele is fixed and where at the second locus two alleles are segregating:

$$\widehat{x}^5 = (1,0,0,0,0,1,0,0)^T, \tag{5.9a}$$

$$\widehat{x}^6 = (0,1,0,0,1,0,0,0)^T, \tag{5.9b}$$

$$\widehat{x}^7 = (0,0,1,0,0,0,0,1)^T, \tag{5.9c}$$

$$\widehat{x}^8 = (0,0,0,1,0,0,1,0)^T. \tag{5.9d}$$

At the other four single-locus polymorphism, one allele is fixed at the second locus and at the first locus two alleles are segregating:

$$\widehat{x}^9 = (1,0,0,0,0,0,1,0)^T \tag{5.10a}$$

$$\widehat{x}^{10} = (0,0,1,0,1,0,0,0)^T, \tag{5.10b}$$

$$\widehat{x}^{11} = (0,1,0,0,0,0,0,1)^T, \tag{5.10c}$$

$$\widehat{x}^{12} = (0,0,0,1,0,1,0,0)^T. \tag{5.10d}$$

Finally, there are four fully polymorphic equilibria:

$$\widehat{x}^{13} = (1,0,0,0,0,0,0,1)^T, \tag{5.11a}$$
$$\widehat{x}^{14} = (0,0,0,1,1,0,0,0)^T, \tag{5.11b}$$
$$\widehat{x}^{15} = (0,1,0,0,0,0,1,0)^T, \tag{5.11c}$$
$$\widehat{x}^{16} = (0,0,1,0,0,1,0,0)^T. \tag{5.11d}$$

Note that of these 16 invariant points, only one is asymptotically stable. This is the equilibrium which is constituted by the two gametes that are the best in one of the demes each. All other equilibria are unstable. In the following, we denote the set of equilibria of (5.7) by Σ_0.

Figure 5.2 shows one example of the dynamical behavior of the gamete frequencies of a population inhabiting two demes under our model. The initial gamete frequencies in deme α and β are $x_{(\alpha)} = (0.2, 0.6, 0.05, 0.15)^T$ and $x_{(\beta)} = (0.3, 0.35, 0.25, 0.1)^T$, respectively. In deme α the gamete A_2B_2 is fittest, whereas in deme β the gamete A_1B_1 is fittest. Due to Theorem 5.1.1, these gametes become fixed, i.e., $\widehat{x}_{(\alpha)} = (0,0,0,1)^T$ and $\widehat{x}_{(\beta)} = (1,0,0,0)^T$ are globally stable in the first deme and in the second deme, respectively. Therefore, the overall gamete frequencies at the equilibrium suffice $\widehat{x} = (0,0,0,1,1,0,0,0)^T$. Of course, one could extend the two-allele two-locus model to more than two demes. This would follow the same ideas as stated above.

5.2 Weak Migration

In this section we will introduce migration between two demes as a perturbation of the selection-recombination model (5.7) for two isolated subpopulations. We will consider only the interesting case where the selection coefficients have different signs. The two cases where selection acts in the same direction in both demes are treated in Theorem 3.2.1. Unless otherwise mentioned, throughout this section we assume that $s_\alpha > 0$ and $s_\beta < 0$ holds.

We call migration weak if there exist constants $a_{\alpha\beta} > 0$ such that the migration rates can be written as

$$m_{\alpha\beta} = \delta_{\alpha\beta} + \epsilon a_{\alpha\beta}, \tag{5.12}$$

where $\epsilon > 0$ is a small number and $\delta_{\alpha\beta}$ is the Kronecker delta, i.e., $\delta_{\alpha\alpha} = 1$ and $\delta_{\alpha\beta} = 0$

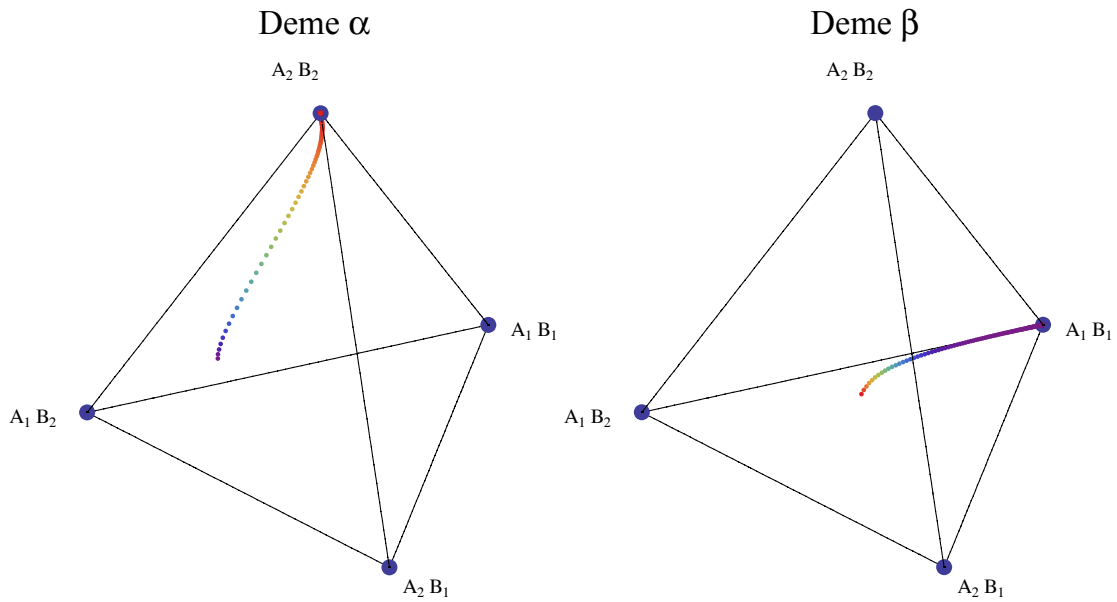

Figure 5.2: Trajectories for the diallelic two-locus model in two demes without migration. The parameters used are: $\gamma_1 = 3$, $\gamma_2 = 1$, $s_\alpha = 0.05$, $s_\beta = -0.05$, $r = 0.1$. Selection coefficients are the same for both loci, i.e., $s_j = t_j$ for $j = \alpha, \beta$. We used 1000 iteration steps.

for $\alpha \neq \beta$. The strength of migration is measured by $\epsilon \geq 0$. If $\epsilon = 0$, no migration occurs and the dynamics for two demes is given by (5.7). From the stochasticity of M it follows that

$$a_{\alpha\beta} > 0 \text{ for every and } \beta \neq \alpha, \sum_{\beta} a_{\alpha\beta} = 0 \text{ for every } \alpha \tag{5.13}$$

must hold. For two demes the backward migration matrix M can be written as

$$M = \begin{pmatrix} 1 - \epsilon a_1 & \epsilon a_1 \\ \epsilon a_1 & 1 - \epsilon a_2 \end{pmatrix}, \tag{5.14}$$

where $a_1 = a_{\alpha\beta}$, $a_2 = a_{\beta\alpha}$. As $a_1, a_2 > 0$, the migration matrix M is irreducible.

The following theorem is a special case (no epistasis) of Theorem 5.4. in [2, p. 971]:

Theorem 5.2.1 (Dynamics under weak migration)**.** *Recall that the full dynamics of our selection-recombination-migration model is given by* (1.17)*. Suppose that fitnesses are additive and given by* (1.8) *such that every genotype has a positive fitness and that every equilibrium of* (5.7) *is hyperbolic. Let c be fixed and $\epsilon > 0$ be sufficiently small. Then for every combination (W, M) of the fitness* (1.8) *and the migration matrix* (5.14) *the following holds:*

(i) The set of equilibria $\Sigma_0 \subset \Delta_4^2$ *of* (5.7) *contains only isolated points. The set of equilibria* $\Sigma_{(W,M)} \subset \Delta_4^2$ *of* (1.17) *also contains only isolated points. As* $\epsilon \to 0$ *in* (5.12)*, each equilibrium in* $\Sigma_{(W,M)}$ *converges to the corresponding equilibrium in* Σ_0*.*

(ii) In the neighborhood of each asymptotically stable equilibrium in Σ_0*, there exists exactly one equilibrium point in* $\Sigma_{(W,M)}$ *and it is also asymptotically stable. In the neighborhood of each unstable boundary equilibrium in* Σ_0*, there exists at most one equilibrium point in* $\Sigma_{(W,M)}$*. If it exists, it is unstable.*

(i) Every solution $x(t)$ *of* (1.17) *converges to one of the equilibria in* $\Sigma_{(W,M)}$*.*

Proof. The proof is found in [2, p. 971]. □

Note that in contrast to the case of weak selection (4.2), unstable boundary equilibria can leave the state space under weak migration. For $\epsilon = 0$, all equilibria are in Λ_0, i.e., in every deme there is linkage equilibrium. Due to the assumption $s_\alpha > 0$ and $s_\beta < 0$, the full polymorphism $\widehat{x}_{14}$ is globally asymptotically stable.

Recall the set of equilibria Σ_0 of (5.7). Using Theorem 5.2.1 we will now derive the dynamics of (1.17) under weak migration between two demes if selection is non-uniform. Assume (5.14) holds. The four monomorphic equilibria in Σ_0, (5.8), are not affected by migration and are also in the set $\Sigma_{(W,M)}$ of equilibria of (1.17) for arbitrarily strong migration. Also, their stability properties are preserved under weak migration.

Next, we analyze the behavior of the eight single-locus polymorphisms in Σ_0, i.e., (5.9) and (5.10). We analyze a population formed by A_1B_1 and A_1B_2 only. Then, there are two single-locus polymorphisms $\widehat{x}^5$ and $\widehat{x}^6$ in the total population. Recall that $s_\alpha > 0$ and $s_\beta < 0$ hold. For this subsystem, A_1B_2 is fittest in deme α and A_1B_1 is fittest in deme β. Theorem 5.1.1 implies that the equilibrium $\widehat{x}^6$ is asymptotically stable and $\widehat{x}^5$ is unstable restricted to the marginal one-locus system. Following Theorem 5.2.1, for $\epsilon > 0$ small there exists an asymptotically stable equilibrium of (1.17) in the neighborhood of $\widehat{x}^6$. In the neighborhood of the unstable equilibrium $\widehat{x}^5$ there could exist one unstable equilibrium of (1.17). To obtain explicit results, we assume that small migration rates ϵa_1 and ϵa_2 will introduce a perturbation of the equilibria. Thus, the corresponding equilibrium $\widehat{x}_\epsilon^6$ in $\Sigma_{(W,M)}$ of the perturbed system (1.17) must be of the form:

$$\widehat{x}_\epsilon^6 = (\epsilon\widehat{y}, 1 - \epsilon\widehat{y}, 0, 0, 1 - \epsilon\widehat{z}, \epsilon\widehat{z}, 0, 0)^T, \tag{5.15}$$

where $\epsilon > 0$ is small. We can approximate the right-hand side of the dynamics (1.17) by

a Taylor series of first order in ϵ. This leads to a much simpler linear system:

$$\epsilon y' = \epsilon \frac{a_1(1-(\gamma_1+\gamma_2)s_\alpha)+y(1-\gamma_1 s_\alpha)}{1-(\gamma_1+\gamma_2)s_\alpha}, \tag{5.16a}$$

$$\epsilon z' = \epsilon \frac{a_2(1-(\gamma_1-\gamma_2)s_\beta)+z(1-\gamma_1 s_\beta)}{1-(\gamma_1+\gamma_2)s_\beta}. \tag{5.16b}$$

One easily obtains the equilibrium of (5.16):

$$\widehat{y} = \frac{a_1(1-(\gamma_1-\gamma_2)s_\alpha)}{\gamma_2 s_\alpha} \text{ and } \widehat{z} = \frac{-a_2(1-(\gamma_1+\gamma_2)s_\beta)}{\gamma_2 s_\beta}. \tag{5.17}$$

As $s_\alpha > 0$, $s_\beta < 0$, $a_1 > 0$, $a_2 > 0$, $\gamma_1 \geq \gamma_2 > 0$, and because we assume $W > 0$ it follows that $\widehat{y}$ and $\widehat{z}$ are positive. For existence, $\widehat{y}$ and $\widehat{z}$ must fulfill:

$$a_1\left(1+\frac{1}{\gamma_2 s_\alpha}-\frac{\gamma_1}{\gamma_2}\right) \leq 1 \text{ and } a_2\left(1-\frac{1}{\gamma_2 s_\beta}+\frac{\gamma_1}{\gamma_2}\right) \leq 1. \tag{5.18}$$

Thus, if this is fulfilled, which is the case if migration rates are sufficiently small relative to other evolutionary forces, the equilibrium (5.16) is asymptotically stable restricted to the marginal one-locus system.

Following the same notation as above, we analyze the perturbation of the equilibrium $\widehat{x}^5$ under weak migration. For $\widehat{x}^5_\epsilon = (1-\epsilon\widehat{y}, \epsilon\widehat{y}, 0, 0, \epsilon\widehat{z}, 1-\epsilon\widehat{z}, 0, 0)^T$ we derive that

$$\widehat{y} = -a_1 \frac{1-(\gamma_1+\gamma_2)s_\alpha}{\gamma_2 s_\alpha} \text{ and } \widehat{z} = a_2 \frac{1+(\gamma_1-\gamma_2)s_\beta}{\gamma_2 s_\beta} \tag{5.19}$$

must hold at equilibrium. Obviously, for our parameters $\widehat{y} < 0, \widehat{z} < 0$ hold, thus this equilibrium does not exist anymore for weak migration in the state space Δ^2_4.

For other single-locus polymorphisms in Σ_0 the results yield analogous dynamics: Given a population consisting only of A_2B_1 and A_2B_2, the perturbation of the unstable equilibrium $\widehat{x}^7$ in Σ_0 leaves the state space Δ^2_4:

$$\widehat{y} = -a_1 \frac{1+(\gamma_1-\gamma_2)s_\alpha}{\gamma_2 s_\alpha} < 0 \text{ and } \widehat{z} = a_2 \frac{1+(\gamma_1+\gamma_2)s_\beta}{\gamma_2 s_\beta} < 0, \tag{5.20}$$

for $\widehat{x}^7_\epsilon = (0, 0, 1-\epsilon\widehat{y}, \epsilon\widehat{y}, 0, 0, \epsilon\widehat{z}, 1-\epsilon\widehat{z})^T$.

The perturbation $\widehat{x}^8_\epsilon = ((0, 0, \epsilon\widehat{y}, 1-\epsilon\widehat{y}, 0, 0, 1-\epsilon\widehat{z}, \epsilon\widehat{z})^T$ of the asymptotically stable equilibrium $\widehat{x}^8$ of the marginal one-locus system is in the state space if

$$\widehat{y} = a_1 \frac{1+(\gamma_1+\gamma_2)s_\alpha}{\gamma_2 s_\alpha} \leq 1 \text{ and } \widehat{z} = -a_2 \frac{1+(\gamma_1-\gamma_2)s_\beta}{\gamma_2 s_\beta} \leq 1. \tag{5.21}$$

Given a population consisting only of A_1B_1 and A_2B_1, the unstable single-locus polymorphism $\widehat{x}^9$ leaves the state space under weak migration as

$$\widehat{y} = -a_1\frac{1-(\gamma_1+\gamma_2)s_\alpha}{\gamma_1 s_\alpha} < 0 \text{ and } \widehat{z} = a_2\frac{1+(\gamma_1-\gamma_2)s_\beta}{\gamma_1 s_\beta} < 0, \tag{5.22}$$

where $\widehat{x}^9_\epsilon = (1-\epsilon\widehat{y}, 0, \epsilon\widehat{y}, 0, \epsilon\widehat{z}, 0, 1-\epsilon\widehat{z}, 0)^T$.

The perturbation of the asymptotically stable equilibrium $\widehat{x}^{10}$ of the marginal one-locus system, $(\epsilon\widehat{y}, 0, 1-\epsilon\widehat{y}, 0, 1-\epsilon\widehat{z}, 0, \epsilon\widehat{z}, 0)^T$, exists and is asymptotically stable in the marginal system if

$$\widehat{y} = a_1\frac{1+(\gamma_1-\gamma_2)s_\alpha}{\gamma_1 s_\alpha} \leq 1 \text{ and } \widehat{z} = -a_2\frac{1-(\gamma_1+\gamma_2)s_\beta}{\gamma_1 s_\beta} \leq 1. \tag{5.23}$$

Given a population consisting only of A_1B_2 and A_2B_2, the unstable single-locus polymorphism $\widehat{x}^{11}$ leaves the state space under weak migration as

$$\widehat{y} = -a_1\frac{1-(\gamma_1-\gamma_2)s_\alpha}{\gamma_1 s_\alpha} < 0 \text{ and } \widehat{z} = a_2\frac{1+(\gamma_1+\gamma_2)s_\beta}{\gamma_1 s_\beta} < 0, \tag{5.24}$$

if $\widehat{x}^{11}_\epsilon = (0, 1-\epsilon\widehat{y}, 0, \epsilon\widehat{y}, 0, \epsilon\widehat{z}, 0, 1-\epsilon\widehat{z})^T$.

The perturbation of the asymptotically stable equilibrium $\widehat{x}^{12}$ of this marginal one-locus system, $(0, \epsilon\widehat{y}, 0, 1-\epsilon\widehat{y}, 0, 1-\epsilon\widehat{z}, 0, \epsilon\widehat{z})^T$, exists and is asymptotically stable if

$$\widehat{y} = a_1\frac{1+(\gamma_1+\gamma_2)s_\alpha}{\gamma_1 s_\alpha} \leq 1 \text{ and } \widehat{z} = -a_2\frac{1-(\gamma_1-\gamma_2)s_\beta}{\gamma_1 s_\beta} \leq 1. \tag{5.25}$$

Next, we analyze the dynamics under weak migration for the fully polymorphic equilibria in Σ_0. First, we analyze the asymptotically stable equilibrium $\widehat{x}^{14}$ of (5.7). The perturbed equilibrium must be of the form:

$$\widehat{x}^{14}_\epsilon = (\epsilon\widehat{y}_1, \epsilon\widehat{y}_2, \epsilon\widehat{y}_3, 1-\epsilon(\widehat{y}_1+\widehat{y}_2+\widehat{y}_3), 1-\epsilon(\widehat{z}_1+\widehat{z}_2+\widehat{z}_3), \epsilon\widehat{z}_3, \epsilon\widehat{z}_2, \epsilon\widehat{z}_1)^T. \tag{5.26}$$

Note that this implies that for $\epsilon > 0$, there is weak linkage disequilibrium within each deme, i.e., $D = \mathcal{O}(\epsilon)$ as $D_\alpha = \mathcal{O}(\epsilon y_1)$ and $D_\beta = \mathcal{O}(\epsilon z_1)$.

Solving (1.17) under weak migration where we omit terms of order $\mathcal{O}(\epsilon^2)$, we obtain

that at equilibrium

$$\begin{aligned}
\widehat{y}_1 &= a_1\frac{1+(\gamma_1+\gamma_2)s_\alpha}{c+(\gamma_1+\gamma_2)s_\alpha}, & \widehat{z}_1 &= a_2\frac{1-(\gamma_1+\gamma_2)s_\beta}{c-(\gamma_1+\gamma_2)s_\beta},\\
\widehat{y}_2 &= ca_1\frac{1+(\gamma_1+\gamma_2)s_\alpha}{\gamma_1 s_\alpha(c+(\gamma_1+\gamma_2)s_\alpha)}, & \widehat{z}_2 &= -ca_2\frac{1-(\gamma_1+\gamma_2)s_\beta}{\gamma_1 s_\beta(c-(\gamma_1+\gamma_2)s_\beta)},\\
\widehat{y}_3 &= ca_1\frac{1+(\gamma_1+\gamma_2)s_\alpha}{\gamma_2 s_\alpha(c+(\gamma_1+\gamma_2)s_\alpha)}, & \widehat{z}_3 &= -ca_2\frac{1-(\gamma_1+\gamma_2)s_\beta}{\gamma_2 s_\beta(c-(\gamma_1+\gamma_2)s_\beta)},
\end{aligned}$$

must hold. As expected, recombination plays a role now. For $c = 0$, $\widehat{y}_2 = \widehat{y}_3 = \widehat{z}_2 = \widehat{z}_3 = 0$ holds. Next, assume $c > 0$. As fitnesses are positive and for our choice of parameters, it follows that all coordinates of $\widehat{x}^{14}_\epsilon$ are greater or equal zero. Note that $\widehat{y}_i < 1$ and $\widehat{z}_i < 1$ for $i = 1, 2, 3$ such that $\widehat{x}^{14}_\epsilon$ is internal in Δ^2_4. In the symmetric scenario, where $a_1 = a_2$, $0 < s_\alpha = s = -s_\beta$, $\gamma_1 = \gamma_2$, the population is spatially homogeneous, i.e., $q = 0$. Then, also the frequencies of gametes A_1B_2 and A_2B_1 are equal. As $\widehat{x}^{14}$ is globally asymptotically stable for (5.7), Theorem 5.2.1 implies that $\widehat{x}^{14}_\epsilon$ is globally asymptotically stable. Furthermore, we have shown that this equilibrium is in linkage disequilibrium of order $\mathcal{O}(\epsilon)$.

Next, we analyze the behavior of the three unstable fully polymorphic equilibria in Σ_0 under weak migration. The perturbation of $\widehat{x}^{13}$ is of the form

$$\widehat{x}^{13}_\epsilon = (1-\epsilon(\widehat{y}_1+\widehat{y}_2+\widehat{y}_3), \epsilon\widehat{y}_3, \epsilon\widehat{y}_2, \epsilon\widehat{y}_1, \epsilon\widehat{z}_1, \epsilon\widehat{z}_2, \epsilon\widehat{z}_3, 1-\epsilon(\widehat{z}_1+\widehat{z}_2+\widehat{z}_3))^T. \tag{5.27}$$

$\widehat{x}^{13}_\epsilon$ is only an equilibrium of (1.17) under weak migration if

$$\begin{aligned}
\widehat{y}_1 &= a_1\frac{1-(\gamma_1+\gamma_2)s_\alpha}{c-(\gamma_1+\gamma_2)s_\alpha}, & \widehat{z}_1 &= a_2\frac{1+(\gamma_1-\gamma_2)s_\beta}{c+(\gamma_1+\gamma_2)s_\beta},\\
\widehat{y}_2 &= -ca_1\frac{1-(\gamma_1+\gamma_2)s_\alpha}{\gamma_1 s_\alpha(c-(\gamma_1+\gamma_2)s_\alpha)}, & \widehat{z}_2 &= ca_2\frac{1+(\gamma_1+\gamma_2)s_\beta}{\gamma_1 s_\beta(c+(\gamma_1+\gamma_2)s_\beta)},\\
\widehat{y}_3 &= -ca_1\frac{1-(\gamma_1+\gamma_2)s_\alpha}{\gamma_2 s_\alpha(c-(\gamma_1+\gamma_2)s_\alpha)}, & \widehat{z}_3 &= ca_2\frac{1+(\gamma_1+\gamma_2)s_\beta}{\gamma_2 s_\beta(c+(\gamma_1+\gamma_2)s_\beta)}.
\end{aligned}$$

There is no choice of parameters such that all $\widehat{y}_i$ or all $\widehat{z}_i$ are positive. Thus, this equilibrium does not exist under weak migration.

The perturbation of $\widehat{x}^{15}$ is of the form

$$\widehat{x}^{15}_\epsilon = (\epsilon\widehat{y}_1, 1-\epsilon(\widehat{y}_1+\widehat{y}_2+\widehat{y}_3), \epsilon\widehat{y}_2, \epsilon\widehat{y}_3, \epsilon\widehat{z}_1, \epsilon\widehat{z}_2, 1-\epsilon(\widehat{z}_1+\widehat{z}_2+\widehat{z}_3), \epsilon\widehat{z}_3)^T. \tag{5.28}$$

(5.28) is an equilibrium of (1.17) under weak migration if

$$\widehat{y}_1 = ca_1\frac{1-(\gamma_1-\gamma_2)s_\alpha}{\gamma_2 s_\alpha(c-(\gamma_1-\gamma_2)s_\alpha)}, \qquad \widehat{z}_1 = ca_2\frac{1+(\gamma_1-\gamma_2)s_\beta}{\gamma_1 s_\beta(c+(\gamma_1-\gamma_2)s_\beta)},$$
$$\widehat{y}_2 = a_1\frac{1-(\gamma_1-\gamma_2)s_\alpha}{c-(\gamma_1-\gamma_2)s_\alpha}, \qquad \widehat{z}_2 = a_2\frac{1+(\gamma_1-\gamma_2)s_\beta}{c+(\gamma_1-\gamma_2)s_\beta},$$
$$\widehat{y}_3 = -ca_1\frac{1-(\gamma_1-\gamma_2)s_\alpha}{\gamma_1 s_\alpha(c-(\gamma_1-\gamma_2)s_\alpha)}, \qquad \widehat{z}_3 = -ca_2\frac{1+(\gamma_1-\gamma_2)s_\beta}{\gamma_2 s_\beta(c+(\gamma_1-\gamma_2)s_\beta)},$$

holds. There exists no choice of parameters such that $\widehat{y}_i > 0$ and $\widehat{z}_i > 0$ for every $i = 1, 2, 3$. Therefore, this equilibrium does not exist under weak migration.

The perturbation of $\widehat{x}^{16}$ is

$$\widehat{x}^{16}_\epsilon = (\epsilon\widehat{y}_1, \epsilon\widehat{y}_2, 1-\epsilon(\widehat{y}_1+\widehat{y}_2+\widehat{y}_3), \epsilon\widehat{y}_3, \epsilon\widehat{z}_1, 1-\epsilon(\widehat{z}_1+\widehat{z}_2+\widehat{z}_3), \epsilon\widehat{z}_2, \epsilon\widehat{z}_3)^T. \tag{5.29}$$

(5.28) is an equilibrium of (1.17) under weak migration if

$$\widehat{y}_1 = ca_1\frac{1+(\gamma_1-\gamma_2)s_\alpha}{\gamma_1 s_\alpha(c+(\gamma_1-\gamma_2)s_\alpha)}, \qquad \widehat{z}_1 = -ca_2\frac{1-(\gamma_1-\gamma_2)s_\beta}{\gamma_1 s_\beta(c-(\gamma_1-\gamma_2)s_\beta)},$$
$$\widehat{y}_2 = a_1\frac{1+(\gamma_1-\gamma_2)s_\alpha}{c+(\gamma_1-\gamma_2)s_\alpha}, \qquad \widehat{z}_2 = a_2\frac{1-(\gamma_1-\gamma_2)s_\beta}{c-(\gamma_1-\gamma_2)s_\beta},$$
$$\widehat{y}_3 = -ca_1\frac{1+(\gamma_1-\gamma_2)s_\alpha}{\gamma_2 s_\alpha(c+(\gamma_1-\gamma_2)s_\alpha)}, \qquad \widehat{z}_3 = -ca_2\frac{1-(\gamma_1-\gamma_2)s_\beta}{\gamma_2 s_\beta(c-(\gamma_1-\gamma_2)s_\beta)}.$$

Again, there is no choice of parameters such that every $\widehat{y}_i > 0$ and $\widehat{z}_i > 0$ for $i = 1, 2, 3$ and this equilibrium does not exist under weak migration.

5.2.1 Weak Selection and Weak Migration

In this section we analyze (1.17) under weak migration and weak selection. Thus, recall the definitions (4.7) and (5.12). The recombination rate is arbitrary but fixed. As already discussed in Section 1.2, in the absence of selection and migration the linkage equilibrium manifold Λ_0 is invariant and globally attracting at a uniform geometric rate. For weak migration and weak selection, i.e., ϵ sufficiently small, the theory of normally hyperbolic manifolds implies the existence of a smooth invariant manifold Λ_ϵ close to Λ_0 [2, p. 968]. Also, Λ_ϵ is attracting at a geometric rate. From [2, p. 968] we know that on Λ_ϵ, linkage disequilibria are of order ϵ and change very slowly. Therefore, Λ_ϵ can be called the *quasi-linkage equilibrium manifold* under weak migration and weak selection. Bürger ([2, p.

968]) has derived the limiting dynamics on Δ_4^Γ:

$$\frac{dp_{i_n,\alpha}^{(n)}}{dt} = p_{i_n,\alpha}^{(n)}\left[r_{i_n,\alpha}^{(n)}(\rho_\alpha) - \bar{r}_\alpha(\rho_\alpha)\right] + \sum_\beta a_{\alpha\beta} p_{i_n,\beta}^{(n)}, \tag{5.30a}$$

$$D = 0, \tag{5.30b}$$

where $r_{i_n,\alpha}^{(n)}$ and $\bar{r}_\alpha$ are the marginal and mean selection coefficients of the subpopulation in deme α when selection is weak, cf. (4.8). From [2, p. 969] the following result is known:

Theorem 5.2.2 (Dynamics under weak migration and weak selection)**.** *Assume that migration and selection are both weak, i.e., that* (5.12) *and* (4.7) *hold. Let the backward migration matrix M and the recombination rate c be fixed, and $\epsilon > 0$ be sufficiently small.*

(a) Every solution $x(t)$ of the full dynamics (1.17) *converges to the quasi-linkage equilibrium manifold Δ_ϵ.*

(b) In the neighborhood of each hyperbolic, asymptotically stable equilibrium of (4.18a)*, there exists exactly one equilibrium point of* (1.17)*, and it is too asymptotically stable. In the neighborhood of each hyperbolic, unstable internal equilibrium of* (4.18a)*, there exists exactly one equilibrium point of* (1.17)*, and it is also unstable. In the neighborhood of each hyperbolic, unstable boundary equilibrium of* (4.18a)*, there exists at most one equilibrium point of* (1.17)*, and if it exists, it is unstable.*

Proof. For the proof we refer to [2, p. 969]. □

Now we specialize to two demes. In the additive diallelic two-locus case, (5.30) becomes:

$$\frac{dp_\alpha^{(1)}}{dt} = -s_\alpha\gamma_1 p_\alpha^{(1)}(1-p_\alpha^{(1)}) - a_1(p_\alpha^{(1)} - p_\beta^{(1)}), \tag{5.31a}$$

$$\frac{dp_\beta^{(1)}}{dt} = -s_\beta\gamma_1 p_\beta^{(1)}(1-p_\beta^{(1)}) - a_2(p_\beta^{(1)} - p_\alpha^{(1)}), \tag{5.31b}$$

$$\frac{dp_\alpha^{(2)}}{dt} = -s_\alpha\gamma_2 p_\alpha^{(2)}(1-p_\alpha^{(2)}) - a_1(p_\alpha^{(2)} - p_\beta^{(2)}), \tag{5.31c}$$

$$\frac{dp_\beta^{(2)}}{dt} = -s_\beta\gamma_2 p_\beta^{(2)}(1-p_\beta^{(2)}) - a_2(p_\beta^{(2)} - p_\alpha^{(2)}). \tag{5.31d}$$

Note that the ordinary differential equations in (5.31) are uncoupled for the two loci. Obviously, the monomorphic equilibria are equilibria of (5.31). Beside these spatially

homogeneous monomorphic equilibria, other equilibria also exist. The allele frequencies at the first locus remain unchanged under (5.31) if

$$p_\alpha^{(1)} = \frac{1}{2}\left(1 + 2\frac{a_1}{\gamma_1 s_\alpha} + \sqrt{1 - \frac{4a_1a_2}{\gamma_1^2 s_\alpha s_\beta}}\right), \tag{5.32a}$$

$$p_\beta^{(1)} = \frac{1}{2}\left(1 + 2\frac{a_2}{\gamma_1 s_\beta} - \sqrt{1 - \frac{4a_1a_2}{\gamma_1^2 s_\alpha s_\beta}}\right), \tag{5.32b}$$

or if

$$p_\alpha^{(1)} = \frac{1}{2}\left(1 + 2\frac{a_1}{\gamma_1 s_\alpha} - \sqrt{1 - \frac{4a_1a_2}{\gamma_1^2 s_\alpha s_\beta}}\right), \tag{5.33a}$$

$$p_\beta^{(1)} = \frac{1}{2}\left(1 + 2\frac{a_2}{\gamma_1 s_\beta} + \sqrt{1 - \frac{4a_1a_2}{\gamma_1^2 s_\alpha s_\beta}}\right). \tag{5.33b}$$

An easy calculation shows that if $s_\alpha > 0 > s_\beta$ holds, (5.32) does not give a valid equilibrium. However, for $p_\alpha^{(1)}, p_\beta^{(1)}$ defined by (5.33), $0 < p_\alpha^{(1)} < p_\beta^{(1)} < 1$ holds, if

$$\left|\frac{a_1}{s_\alpha} + \frac{a_2}{s_\beta}\right| < \gamma_1. \tag{5.34}$$

Analogously, if $s_\alpha < 0 < s_\beta$ holds, (5.33) does not give a valid equilibrium. Thus, for $p_\alpha^{(1)}, p_\beta^{(1)}$ defined in (5.32), $0 < p_\beta^{(1)} < p_\alpha^{(1)} < 1$ if (5.34) holds.

The allele frequencies at the second locus remain unchanged under (5.31) if

$$p_\alpha^{(2)} = \frac{1}{2}\left(1 + 2\frac{a_1}{\gamma_2 s_\alpha} + \sqrt{1 - \frac{4a_1a_2}{\gamma_2^2 s_\alpha s_\beta}}\right), \tag{5.35a}$$

$$p_\beta^{(2)} = \frac{1}{2}\left(1 + 2\frac{a_2}{\gamma_2 s_\beta} - \sqrt{1 - \frac{4a_1a_2}{\gamma_2^2 s_\alpha s_\beta}}\right), \tag{5.35b}$$

or if

$$p_\alpha^{(2)} = \frac{1}{2}\left(1 + 2\frac{a_1}{\gamma_2 s_\alpha} - \sqrt{1 - \frac{4a_1a_2}{\gamma_2^2 s_\alpha s_\beta}}\right), \tag{5.36a}$$

$$p_\beta^{(2)} = \frac{1}{2}\left(1 + 2\frac{a_2}{\gamma_2 s_\beta} + \sqrt{1 - \frac{4a_1a_2}{\gamma_2^2 s_\alpha s_\beta}}\right). \tag{5.36b}$$

If $s_\alpha > 0 > s_\beta$ holds, (5.35) does not give a valid equilibrium. Then, given (5.36),

$0 < p_\alpha^{(2)} < p_\beta^{(2)} < 1$ if

$$\left| \frac{a_1}{s_\alpha} + \frac{a_2}{s_\beta} \right| < \gamma_2. \tag{5.37}$$

If $s_\alpha < 0 < s_\beta$ holds, (5.36) does not give a valid equilibrium. Then, given (5.35), $0 < p_\beta^{(2)} < p_\alpha^{(2)} < 1$ if (5.37) holds. The results for each single locus are consistent with [10]. Set

$$\frac{a_1}{\gamma_k s_\alpha} + \frac{a_2}{\gamma_k s_\beta} = \kappa \tag{5.38}$$

for locus $k = 1, 2$. Then, for each locus the regions of protection are as given by Theorem 3.2.3 and Figure 3.1. The conditions for a protected polymorphism are given by the combinations of the condition for protection of the two marginal systems.

To summarize, besides the monomorphic equilibria there exist the following polymorphisms:

(i) If $s_\alpha > 0 > s_\beta$, there exist four single-locus polymorphisms: First, assume that A_1 or A_2 is fixed at the first locus, i.e., $p_\alpha^{(1)} = p_\beta^{(1)} = 1$ or $p_\alpha^{(1)} = p_\beta^{(1)} = 0$ hold. Then, (5.37) must hold for existence of a protected single-locus polymorphism. Then, the polymorphic frequencies at the second locus are given by (5.36).

If B_1 or B_2 are fixed, i.e., $p_\alpha^{(2)} = p_\beta^{(2)} = 1$ or $p_\alpha^{(2)} = p_\beta^{(2)} = 0$ hold, (5.34) must hold for a protected single-locus polymorphism. Then, the frequencies at the second locus are given by (5.33).

If (5.34) and (5.37) hold there exists a full polymorphism. There, the frequencies are given by (5.33) and (5.36).

(ii) If $s_\alpha < 0 < s_\beta$ there exist four single-locus polymorphisms: First, $p_\alpha^{(1)} = p_\beta^{(1)} = 1$ or $p_\alpha^{(1)} = p_\beta^{(1)} = 0$ hold. For the existence of a single-locus polymorphism, (5.37) must hold. Then, the frequencies at the second locus are given by (5.35).

If $p_\alpha^{(2)} = p_\beta^{(2)} = 1$ or $p_\alpha^{(2)} = p_\beta^{(2)} = 0$ hold, (5.34) must be satisfied for existence of the single-locus polymorphism. Then, the frequencies at the second locus are given by (5.32).

If (5.34) and (5.37) hold, there exists a full polymorphism. There, the frequencies are given by (5.32) and (5.35).

Chapter 6

Summary

We introduced the fundamental population genetic concepts in Section 1.1, and the basic model (1.8), especially the evolutionary dynamics (1.17), in Section 1.2. We derived several results for the diallelic two-locus model with no dominance and no epistasis. Most of the analysis assumes migration between two demes. Under uniform selection, the dynamical behavior of (1.17) is well understood (Theorem 3.2.1). Only non-uniform selection harbors the potential for maintaining single-locus polymorphisms and full polymorphisms. In the following, we will summarize the results for the most interesting case, where $\text{sign}(s_\alpha) \neq \text{sign}(s_\beta)$.

In Section 2, we discussed the concept of linkage equilibrium, (1.13). In general, not every equilibrium of (1.17) must be in linkage equilibrium. In Theorem 2.0.1, we stated the necessary condition for global linkage equilibrium, (2.4). In the Levene model, the linkage equilibrium manifold (1.15) is globally attracting.

In Section 3.1, we introduced basic concepts about equilibria of (1.17) and their stability. The monomorphic equilibria (3.2) of (1.17) exist always. In Section 3.2, conditions for protected alleles in the marginal one-locus systems were derived, from Theorem 3.2.3.

In the Levene model, the dynamics of (1.17) is well understood, cf. Sections 3.2.2 and 3.2.3. A Lyapunov function (3.37) is known in this case. For the marginal one-locus systems, protected single-locus polymorphisms can exist if (3.26) or (3.33) hold. Also, if they exist they are uniquely determined. Internal equilibria of (1.17) in the Levene model exist only non-generically and they form a line which is globally attracting. We calculated this manifold of equilibria in (3.48) for $s_\alpha = s = -s_\beta$, and $\gamma_1 = \gamma_2 = \gamma$.

If selection is absent, the dynamics of (1.17) reduces to (4.1) and is fully understood. Then, the manifold (4.5) is globally attracting. Every point on this manifold is an equilibrium of (4.1) (cf. Theorem 4.1.1). If selection is weak, the dynamics of (1.17) is introduced

is described by the weak-selection limit (4.29) and is well understood (Theorem 4.2.1). (4.29) is a system of ordinary differential equations on (4.5). Generically, only monomorphic equilibria of (4.29) exist. The equilibrium representing the gamete with the highest (average) fitness is globally asymptotically stable. In the nongeneric case, given by (4.34), every point in (4.5) is invariant.

The dynamics for two subpopulations when there is no migration, (5.7), is well understood if equilibria of (5.7) are hyperbolic. Generically, this is the case for non-zero selection (Section 3.2). Also, the assumption of additivity, i.e., no epistasis and no dominance, was necessary. Then, there is a globally asymptotically stable full polymorphism which is not internal, cf. Section 5.1.2. The three other full polymorphisms, the monomorphic equilibria, and the single-locus polymorphisms are unstable, which is a direct consequence of Theorem 5.1.1. Restricted to the marginal one-locus systems, there is one asymptotically stable single-locus polymorphisms and one which is unstable per system each. Introducing weak migration, (5.12), the equilibria of (1.17) are obtained as a perturbation of (5.7), cf. 5.2.1. Under weak migration, the monomorphic equilibria, the stable full polymorphism, the marginally stable single-locus polymorphisms, and their stability properties are preserved (cf. Section 5.2, for example, (5.17)). The unstable full polymorphisms and single-locus polymorphisms leave the state space under the perturbation (for example, (5.19)). Moreover, we have shown that the globally asymptotically stable full polymorphism is in linkage disequilibrium.

For (1.17) under weak migration (5.12) and weak selection (4.7), we obtained that the monomorphic equilibria and their stability are preserved (Theorem 5.2.2). If (5.34) holds there exists a protected single-locus polymorphism where the first locus is polymorphic. If (5.37) holds, there exists a protected single-locus polymorphism where the second locus is polymorphic. If both (5.34) and (5.37) hold, there exists a protected single-locus polymorphism.

Bibliography

[1] E. Akin. *The Geometry of Population Genetics.* Springer, Berlin, 1979.

[2] R. Bürger. Multilocus selection in subdivided populations I. Convergence properties for weak or strong migration. *J. Math. Biol. 58, 939–978, 2009.*

[3] R. Bürger. Multilocus selection in subdivided populations II. Maintenance of polymorphism under weak or strong migration. *J. Math. Biol. 58, 979–997, 2009.*

[4] R. Bürger. Polymorphism in the two-locus Levene model with nonepistatic directional selection. *Theor. Popul. Biol. 76, 214–228, 2009.*

[5] R. Bürger. *The Mathematical Theory of Selection, Recombination, and Mutation.* Wiley, Chichester, 2000.

[6] N. Campbell. *Biologie.* Spektrum, Heidelberg, 1997.

[7] F. Christiansen. Hard and soft selection in a subdivided population. *Amer. Natur. 109, 11–16, 1975.*

[8] F. Christiansen and M. Feldman. Subdivided populations: a review of the one-and two-locus deterministic theory. *Theor. Popul. Biol. 7, 13, 1975.*

[9] W. Ewens. With additive fitness, the mean fitness increases. *Nature 221, 1076, 1969.*

[10] E. Eyland. Moran's island migration model. *Genetics 69, 399, 1971.*

[11] W. Feller. *An Introduction to Probability Theory and Its Applications.* Wiley, New York, 1968.

[12] D. Futuyma. *Evolution.* Sinauer, Sunderland, 2005.

[13] W. Hennig. *Genetik.* Springer, Berlin, 2002.

[14] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, 1998.

[15] S. Karlin and J. McGregor. Polymorphisms for genetic and ecological systems with weak coupling. *Theor. Popul. Biol. 3, 210, 1972.*

[16] J. LaSalle. *Stability theory for difference equations*. 1977.

[17] H. Levene. Genetic equilibrium when more than one ecological niche is available. *Amer. Natur. 87, 331, 1953.*

[18] R. Lewontin and K. Kojima. The evolutionary dynamics of complex polymorphisms. *Evolution 14, 458–472, 1960.*

[19] W. Li and M. Nei. Stable linkage disequilibrium without epistasis in subdivided populations. *Theor. Popul. Biol. 6, 173, 1974.*

[20] T. Nagylaki. Evolution under the multilocus Levene model without epistasis. *Theor. Popul. Biol. 76, 197–213, 2009.*

[21] T. Nagylaki. *Introduction to Theoretical Population Genetics*. Springer, Berlin, 1992.

[22] T. Nagylaki. *The Dynamics of Migration-Selection Models*. Springer, Berlin, 2008.

[23] T. Nagylaki, J. Hofbauer, and P. Brunovsky. Convergence of multilocus systems under weak epistasis or weak selection. *J. Math. Biol. 38, 103–133, 1999.*

[24] E. Seneta. *Non-Negative Matrices and Markov Chains*. Springer, Berlin, 2006.

[25] T. Wiehe and M. Slatkin. Epistatic selection in a multi-locus levene model and implications for linkage disequilibrium. *Theor. Popul. Biol. 53, 75–84, 1998.*

Curriculum Vitae

Peronal Data

Name:	Ada Akerman
Date of Birth:	30^{th} of September, 1984
Place of Birth:	Zadar, Croatia
Nationality:	Austria

Education

High school diploma:	HBLVA für chemische Industrie, 1170 Vienna, 14^{th} of June, 2004
Diploma Studies:	Mathematics, University of Vienna, fall term 2004 – fall term 2009
Bachelor Studies:	Food- and Biotechnology, University of Natural Resources and Applied Life Sciences, Vienna, since fall term 2005
Exchange Semester:	Swiss Federal Institute of Technology in Zurich (ETH), Switzerland, summer term 2009

Subject-specific Work Experience

Health Center:	Porvoo, Finland, Lab for Chemical Analytics, summer 2001
Aventis Behring:	Vienna, Lab for Quality Control, summer 2003
Medical University:	Vienna, Institute for Cancer Research, Molecular Epidemiology group, part time technician, August 2004 – October 2006; temporary personnel, 2006 – 2007
Baxter BioScience:	Orth a.d. Donau, Austria, Lab for Virology, summer 2008
Tutor for Mathematics:	University of Natural Resources and Applied Life Sciences, Vienna, 2007 – 2010