

Deterministische Rekombinationsdynamik mit Selektion

Diplomarbeit

von

Matthias Steinrücken

12. Juli 2006

Betreuer: Prof. Dr. Ellen Baake
Prof. Dr. Michael Baake



Inhaltsverzeichnis

Einleitung	1
1 Biologischer Hintergrund	3
1.1 Rekombination und Selektion	3
2 Grundlagen	7
2.1 Grundlegende Definitionen	7
2.2 Selektionsdynamik	9
2.2.1 Diskret	10
2.2.2 Kontinuierlich	11
2.3 Rekombinationsdynamik	13
2.3.1 Diskret	14
2.3.2 Kontinuierlich	15
2.4 Rekombinationsdynamik auf geordneten Partitionen	16
2.5 Rekombinations- und Selektionsdynamik	17
3 Kopplungsungleichgewichte	19
3.1 Geschichte der Kopplungsungleichgewichte	19
3.2 Zusammenhang der Kopplungsungleichgewichte	22
3.3 Kopplungsungleichgewichte auf vier Sites	23
3.4 Dynamik der zentralen Momente	25
4 Lösung der Rekombinations- und Selektionsdynamik	29
4.1 Lösung der Selektionsdynamik	29
4.2 Mathematische Modellierung der Rekombination	30
4.2.1 Projektoren	30
4.2.2 Rekombinatoren	31
4.2.3 Eigenschaften der Rekombinatoren	31
4.3 Lösung der Rekombinationsdynamik	32
4.4 Zusammenführen der Lösungen	37
4.4.1 Ein Gegenbeispiel	38
4.5 Lösung einfacher Spezialfälle	39
4.5.1 Ein neutraler und ein selektiver Site	39
4.5.2 Zwei neutrale und ein selektiver Site	43

5	Der Hitchhiking-Effekt	49
5.1	Effekt auf einen neutralen Site	49
5.2	Effekt auf Kopplungsungleichgewicht zwischen zwei neutralen Sites . . .	52
5.2.1	Das Szenario	53
5.2.2	Unsere Ergebnisse	56
6	Diskussion	63

Einleitung

Wie wird das genetische Material einer Population geformt?

Seit Beginn des Lebens auf diesem Planeten, in diesem Moment und auch weiterhin in der Zukunft wirkt die Evolution auf das genetische Material einer jeden Population.

Die evolutionären Kräfte sind vielfältig und teilweise schwer zu verstehen. Einige von ihnen lassen sich jedoch auf Wechselwirkungen zwischen wenigen Individuen oder sogar auf die Entwicklung von einzelnen Individuen zurückführen. In der vorliegenden Arbeit wollen wir aus dieser Vielfalt zwei Kräfte auswählen und Interaktion zwischen ihnen untersuchen.

Die Selektion sorgt dafür, dass sich Individuen mit vorteilhaften genetischen Anlagen gegenüber dem Rest der Population durchsetzen. Sie verursacht also eine Abnahme der Vielfalt im genetischen Material der Population.

Die Rekombination sorgt bei der Fortpflanzung für eine Neukombination der genetischen Informationen der Population. Es werden einerseits neue Kombinationen aus den 23 Chromosomen der Mutter und denen des Vaters gebildet. Andererseits kommt es aber auch zu einer Neukombination von genetischer Information innerhalb der homologen Chromosomen. Durch die Rekombination kommt es somit zu einer Erhöhung der genetischen Vielfalt der Population.

Die beiden Kräfte haben also auf den ersten Blick entgegengesetzte Auswirkungen auf das genetische Material. In dieser Arbeit wollen wir das auf den zweiten Blick kompliziertere Wechselspiel dieser beiden Kräfte genauer untersuchen.

Nachdem wir in Kapitel 1 in die biologischen Grundlagen der beiden Kräfte einführen, leiten wir in Kapitel 2 Modelle her, welche die Wirkung der Rekombination und der Selektion sowohl getrennt, als auch gleichzeitig beschreiben.

Daraufhin führen wir in Kapitel 3 Korrelationsfunktionen, sogenannte Kopplungsungleichgewichte, ein, welche sich, wie wir motivieren werden, für Untersuchungen zur Rekombinationsdynamik als nützlich erwiesen haben. Wir beabsichtigen, anschließend die exakte Lösung einer leicht eingeschränkten Rekombinations- und Selektionsdynamik, welche von Baake [1] vorgeschlagen wurde, zu rekapitulieren. Diese Lösung erwies sich jedoch als falsch, so dass wir in Kapitel 4, nachdem wir die richtigen Aspekte dieser Lösung aufzeigen und daraufhin anhand eines Beispiels die Gesamtlösung widerlegen, für zwei handhabbare Spezialfälle der Dynamik Lösungen herleiten.

Bei dem Wechselspiel zwischen Rekombination und Selektion kommt es zu dem Phänomen des sogenannten Hitchhikings. Dieses erklären wir in Kapitel 5 mittels unserer Lösungen der Dynamik. Anschließend reproduzieren wir Ergebnisse bezüglich des Hitchhiking-Effektes, welche Stephan [19] mit einem Modell der Rekombination und Selektion produziert hat, welche sich leicht unterscheidet von derjenigen, welche wir in Kapitel 4 lösen.

Die ursprüngliche Absicht im Reproduzieren der Ergebnisse war, dass Stephan ein Modell mit einer komplizierten Lösung verwendete, und wir hofften, diese Ergebnisse, mit dem einfacheren Ansatz von Baake, auf leichterem Wege produzieren zu können. Da sich dieser Ansatz jedoch als irreführend erwies, beschränkten wir uns darauf, auf einem ebenfalls komplizierten Wege zu zeigen, dass wir mit unserem Modell dieselben Ergebnisse erzielen.

1 Biologischer Hintergrund

Wir stellen nun die biologischen Zusammenhänge und Prozesse vor, welche wir im weiteren Verlauf dieser Arbeit mit Hilfe von mathematischen Modellen untersuchen wollen.

Diese Grundlagen sind in der Biologie anerkannt und werden in diversen Lehrbüchern behandelt. Eine umfangreiche Darstellung ist zum Beispiel bei Knippers [12] zu finden.

1.1 Rekombination und Selektion

Die Fortpflanzung diploider Organismen basiert auf der Paarbildung haploider Keimzellen, sogenannter *Gameten*, zweier unterschiedlicher Individuen. Wir gehen davon aus, dass die Partnerwahl zufällig geschieht und es keine Präferenzen gibt. Dies bedeutet, dass ein solcher Gamet zufällig einen Partner zugewiesen bekommt. Die Wahrscheinlichkeit für eine bestimmte Gametenpaarung hängt also nur von den relativen Gametenhäufigkeiten ab. Aus dieser Gametenpaarung entsteht ein Individuum mit einem vollständig diploiden Chromosomensatz.

Wir betrachten den Vorgang ohne Berücksichtigung der Geschlechtschromosomen, so dass wir davon ausgehen, dass das Geschlecht des Individuums zufällig bestimmt wird. Die Häufigkeiten der genetischen Anlagen im weiblichen und männlichen Teil der Population unterscheiden sich also nicht.

In der nun folgenden *Diplophase*, in welcher je zwei Gameten gepaart als Individuum vorliegen, prägen die Individuen abhängig von ihrem genetischen Material unterschiedliche Phänotypen aus. Dadurch sind unterschiedliche Individuen unterschiedlich gut für ihre Umwelt gerüstet oder sind anderen im direkten Konkurrenzkampf unter- beziehungsweise überlegen. Es erreichen deshalb von den Individuen mit vorteilhaften Genotypen mehr die Geschlechtsreife und pflanzen sich wieder fort, als von den Individuen, welche nachteiliges Erbgut besitzen. Bestimmte Individuen besitzen also eine größere *Fitness* als andere, welche in erster Linie durch ihren Genotyp bestimmt wird. Die Selektion sorgt somit dafür, dass benachteiligtes genetisches Material aus der Population verschwindet und der genetische Variantenreichtum der Population abnimmt. Dies wird in Abbildung 1.1 veranschaulicht.

Die haploiden Keimzellen, welche ein Individuum zur Fortpflanzung benötigt, entstehen aus den diploiden Zellen durch die sogenannte *Meiose*. Hierbei teilt sich die Zelle zweimal.

Vor der ersten meiotischen Teilung lagern sich die *homologen* Chromosomen gleichen Typs nebeneinander an. Dabei kommt es zu sogenannten *Crossover*-Ereignissen, bei denen Teile des Chromosoms ausgetauscht werden. Dadurch entstehen auf den Chromatiden neue Kombinationen von Allelen aus dem Erbgut beider Eltern. Diese Ereignisse

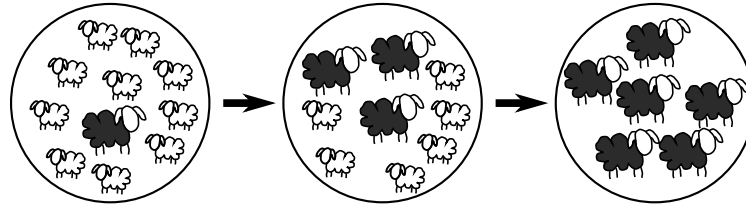


Abbildung 1.1: Zu Beginn besteht die Population aus überwiegend kleinen, weißen Individuen. Die von der Selektion bevorzugten größeren, schwarzen Individuen sind anfangs nur zu einem kleinen Teil vorhanden. Mit der Zeit verdrängen sie die weißen Individuen, bis die Population zum Schluss ausschließlich aus schwarzen besteht.

werden unter dem Begriff *intrachromosomale* Rekombination zusammengefasst.

Bei der ersten meiotischen Teilung trennen sich die homologen Chromosomen eines Individuums auf und bilden zwei Sätze bestehend aus je einem Chromosom eines Typs. Hierbei ist es dem Zufall überlassen, in welche der beiden Gruppen ein Chromosom fällt, weshalb sich die vom Vater und von der Mutter geerbten Chromosomen durchmischen. Dieses Phänomen wird *interchromosomale* Rekombination genannt.

Durch die in Abbildung 1.2 auf der nächsten Seite veranschaulichten Prozesse der inter- und intrachromosomalen Rekombination entstehen neue Kombinationen des bestehenden genetischen Materials, der Reichtum an Variation in der Population wird erhöht.

Bei der Fortpflanzung paaren sich die aus der Meiose entstandenen Gameten daraufhin mit den Gameten anderer Individuen. Sie bilden neue Individuen und gehen wieder in die Diplophase über.

Die zwei vorgestellten evolutionären Kräfte haben beide Auswirkungen auf den Reichtum an genetischer Variation in der Population. Die Selektion verringert den Variantenreichtum, welcher durch die Rekombination erhöht wird. Dieses Wechselspiel der beiden Kräfte soll in der vorliegenden Arbeit genauer untersucht werden.

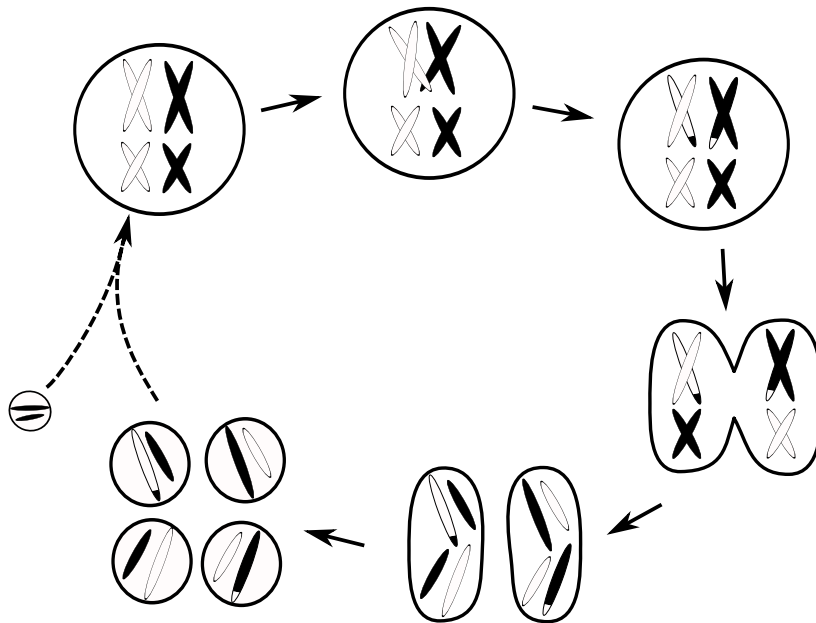


Abbildung 1.2: Schematische Darstellung der Meiose. Die Zelle hat mehrere diploide homologe Chromosomenpaare. Ein Paar bildet ein Crossover, wodurch es zu einem Austausch eines Endstückes kommt. Es wird intrachromosomal rekombiniert. Daraufhin bilden sich die vier haploiden Gameten mit *Ein-Chromatid-Chromosomen* aus. Bei der ersten Teilung werden die mütterlichen und väterlichen Chromosomen gemischt, es kommt daher ebenfalls zur Rekombination.

2 Grundlagen

2.1 Grundlegende Definitionen

Wir modellieren einen Gameten als eine lineare Abfolge von Sites oder Loci, welche wir mit der Menge $S = \{0, \dots, n\}$ indizieren. Es kann sich hierbei entweder um klassische Loci handeln, an denen unterschiedliche Allele stehen können, oder auf einer feineren Skala Sites, an denen Nukleotide stehen. Im Folgenden bezeichnen wir die Positionen als Sites und X_i ist die Menge der möglichen Allele, welche an Site i stehen können. Der Raum der möglichen Gameten oder Typen ist also $X = X_0 \times \dots \times X_n$ und jedes $x = (x_0 \dots x_n) \in X$ stellt einen möglichen Gameten dar.

Für die nachfolgenden Untersuchungen beschränken wir uns darauf, dass die Mengen X_i endlich sind, das heißt, es gibt nur einen endlichen Vorrat möglicher Allele an jedem Site, also auch nur einen endlichen Vorrat möglicher Typen. Ein großer Teil der hier präsentierten Ergebnisse lassen sich jedoch auf den Fall unendlich vieler Typen erweitern.

Eine Population von Gameten beschreiben wir wie folgt: Wir gehen zunächst davon aus, dass unsere Population groß genug ist und somit stochastische Schwankungen vernachlässigt werden können. Ihre Dynamik unter Rekombination und Selektion kann also im sogenannten *infinite population limit* (IPL) betrachtet werden.

Mit $p(x)$ bezeichnen wir die relative Häufigkeit der Gameten vom Typ x in der Population. Somit ist $\sum_{x \in X} p(x) = 1$ und $p(\cdot)$ ein gültiges Wahrscheinlichkeitsmaß auf X . Des Weiteren sei χ eine Zufallsvariable mit Werten in X . Die Wahrscheinlichkeit für eine Realisierung $x \in X$ sei gegeben durch $p(x)$, es gilt also $\mathbb{P}(\chi = x) = p(x)$. Somit ist χ die Zufallsvariable, welche ein zufälliges Ziehen mit Zurücklegen aus einer endlichen beziehungsweise Ziehen ohne Zurücklegen aus einer unendlichen Gametenpopulation beschreibt. Ein zufällig aus der Population gewählter Gamet ist mit Wahrscheinlichkeit $p(x)$ vom Typ x .

Zusätzlich zu den relativen Gametenhäufigkeiten definieren wir Häufigkeiten für Individuen, also für Paarungen von Gameten. Mit $p(x, y)$ bezeichnen wir die relative Häufigkeit der Individuen (x, y) . Es gilt die Marginalisierung auf die Gameten $p(x) = \sum_{y \in X} p(x, y)$.

Wie bereits erwähnt, ist die Funktion, welche die relativen Gametenhäufigkeiten beschreibt, ein Wahrscheinlichkeitsmaß über X , es gilt also $p \in \mathcal{P}(X)$. Motiviert durch die Produktstruktur der zugrundeliegenden Menge möglicher Gameten X führen wir die abkürzende Schreibweise $\mathcal{P}^\otimes := \mathcal{P}(X)$ ein. Da wir für die folgenden Untersuchungen annehmen, dass X endlich ist, ist der Raum der Wahrscheinlichkeitsmaße \mathcal{P}^\otimes gleich der Menge der Vektoren aus $[0, 1]^{|X|}$, für welche $\sum_{x \in X} p(x) = 1$ gilt.

Wir benötigen analoge Definitionen für die absoluten Häufigkeiten. $z(x)$ und $z(x, y)$

bezeichnen also die absoluten Gameten- respektive Individuenhäufigkeiten. Hier lässt sich ebenfalls marginalisieren: $z(x) = \sum_{y \in X} z(x, y)$. Die Funktion z , welche die absoluten Gametenhäufigkeiten angibt, kann analog zu p als ein positives Maß auf der Menge der möglichen Gameten X aufgefasst werden. Mit der analogen Kurzschreibweise $\mathcal{M}^\otimes := \mathcal{M}_+(X)$ gilt somit auch hier $z \in \mathcal{M}^\otimes$. Der Raum der Maße ist gleich der Menge der Vektoren aus $\mathbb{R}_{\geq 0}^{|X|}$, da X endlich ist. Mit $\|z\| = z(X) = \sum_{x \in X} z(x) = \sum_{x, y \in X} z(x, y)$ bezeichnen wir die Größe der Population. Es gilt

$$p(x) = \frac{z(x)}{\|z\|} \quad \text{und} \quad p(x, y) = \frac{z(x, y)}{\|z\|}.$$

Da die Häufigkeiten unter Umständen zeitlich variabel sind, schreiben wir an gegebener Stelle p_t und z_t statt p und z . Die zeitliche Ableitung $\frac{d}{dt}p_t$ schreiben wir der Übersichtlichkeit halber auch als Punkt über der Funktion: \dot{p}_t .

Um Gameten nicht vollständig, sondern nur an einer Teilmenge der Sites spezifizieren zu müssen, führen wir das Konzept der Zylindermengen ein. Dazu seien $j_1, \dots, j_k \in S$ und $j_1 < \dots < j_k$ wobei $k \leq n$. Dann definieren wir

$$\langle j_1, \dots, j_k \rangle := X_{\{0, \dots, j_1-1\}} \times \{x_{j_1}\} \times X_{\{j_1+1, \dots, j_2-1\}} \times \dots \times \{x_{j_k}\} \times X_{\{j_k+1, \dots, n\}} \quad (2.1)$$

als die Menge aller Typen, die an den Sites j_1, \dots, j_k die vorgegebenen Allele tragen und an den restlichen Sites beliebige. Hierbei handelt es sich um eine symbolische Notation, da wir nicht festlegen, welche Allele tatsächlich an den Sites vorhanden sind. Sie sind sozusagen beliebig, aber fest.

Weiter sei $J = \{j_1, \dots, j_k\}$. Damit definieren wir

$$\langle J \rangle := \langle j_1, \dots, j_k \rangle. \quad (2.2)$$

$\langle J \rangle$ ist also die Menge aller Typen, die an den Sites mit Index aus J festgelegte und an den anderen Sites beliebige Allele haben. $p(\langle J \rangle)$ ist dann die relative Häufigkeit der Individuen diesen Typs.

Ferner definieren wir

$$\tilde{p}(\langle J \rangle) := \prod_{j \in J} p(\langle j \rangle) = \prod_{j \in J} p(*, \dots, *, x_j, *, \dots, *) \quad (2.3)$$

als das Produkt der Marginalwahrscheinlichkeiten der durch $\langle J \rangle$ spezifizierten Allele.

Mit $\mathcal{C}_S(J)$ bezeichnen wir das Komplement der Menge J in S , also für $J \subset S$ gerade die Menge $S - J$. Da wir häufig das Komplement einer Menge in S benötigen werden, führen wir hierfür die abkürzende Schreibweise $\bar{J} := \mathcal{C}_S(J)$ ein.

Wir leiten nun Dynamiken her, welche die zeitliche Entwicklung einer Population unter Selektion oder Rekombination beschreiben. Diese Dynamiken sind in der Literatur weit verbreitet und eine umfangreiche Behandlung ist zum Beispiel bei Bürger [6] zu finden.

Im Folgenden geben wir keine Anfangsbedingungen an, sondern gehen davon aus, dass alle Anfangswertprobleme adäquat definierte Startverteilungen z_0 oder p_0 besitzen.

2.2 Selektionsdynamik

Wie wir eingangs bereits erwähnten, sind die Organismen in unserem Modell diploid, es vereinigen sich also zwei Gameten zu einem Individuum. Es hängt nun von der Fitness eines solchen Individuums ab, wieviele Individuen dieses Typs das fortpflanzungsfähige Alter erreichen und ihre zwei Gameten wieder in den Gametenpool geben. Die Größe der Subpopulation im nächsten Schritt beziehungsweise die Änderung der Größe ist somit proportional zur aktuellen Größe.

Unter dieser Sichtweise des Fortpflanzungsprozesses ist es in erster Linie sinnvoll, Reproduktionskoeffizienten nur für Individuen, also Paarungen von Gameten, zu definieren. Wir weisen deshalb einer Paarung von Gameten $(x, y) \in X^2$ einen Koeffizienten zu.

Im diskreten Fall ist dieser Koeffizient $w(x, y) \in \mathbb{R}_{\geq 0}$ die sogenannte *viability Fitness* des Individuums. Sie gibt an, um welchen Faktor sich die Zahl der Individuen des Typs (x, y) bis zur Paarungsreife ändert und bestimmt damit, wieviele Individuen für die nächste Generation von Gameten zur Verfügung stehen. Diese Koeffizienten fassen wir in der Matrix $W = (w(x, y))_{x, y \in X}$ zusammen.

Im kontinuierlichen Fall hingegen werden diese Koeffizienten $u(x, y) \in \mathbb{R}$ *Malthus'sche Fitness* genannt und entsprechen den Nettoreproduktionsraten der Individuen eines bestimmten Typs. Sie bilden die Matrix $U = (u(x, y))_{x, y \in X}$.

In unserem Reproduktionsmodell gilt somit für die zeitliche Dynamik der absoluten Individuenhäufigkeiten

a) in diskreter Zeit:

$$z_{t+1}(x, y) = w(x, y) z_t(x, y) \quad (2.4)$$

b) oder in kontinuierlicher Zeit:

$$\frac{d}{dt} z_t(x, y) = u(x, y) z_t(x, y). \quad (2.5)$$

Wir gehen davon aus, dass es keinen Unterschied in den Häufigkeiten zwischen den Geschlechtern gibt und die Paarungen zufällig stattfinden. Die Population befindet sich im sogenannten Hardy-Weinberg-Gleichgewicht und es gilt

$$p_t(x, y) = p_t(x) p_t(y) \quad \text{und} \quad z_t(x, y) = z_t(x) p_t(y) = z_t(x) \frac{z_t(y)}{\|z_t\|}. \quad (2.6)$$

Im Folgenden betrachten wir die zeitliche Entwicklung der Population unter diesem Modell der Reproduktion. Wir werden zeigen, dass auf der Ebene der absoluten Häufigkeiten z_t keine Selektion existiert, da hier die Teilpopulationen beliebig wachsen oder schrumpfen können, ohne sich gegenseitig zu beeinflussen. Aus dieser Reproduktionsdynamik wird eine Selektionsdynamik, wenn die Population normiert und die relativen Häufigkeiten p_t betrachtet werden. Diese Interaktion der Teilpopulationen kann als eine Limitierung der Ressourcen aufgefasst werden. Es gibt nur endlich viel Platz, den die Teilpopulationen entsprechend ihrer relativen Häufigkeiten gemeinsam ausfüllen. Das

Normieren der Population kann auch als Untersuchung der relativen Häufigkeiten der Subpopulationen anhand einer großen Stichprobe interpretiert werden, welche zufällig aus der Gesamtpopulation gezogen wird.

Wir behandeln nun die diskrete und die kontinuierliche Dynamik ausführlich und getrennt.

2.2.1 Diskret

Zunächst bestimmen wir ausgehend von der Dynamik der Individuen durch Marginalisierung die Dynamik der Gameten. Mit $(Wp_t)(x)$ bezeichnen wir im Folgenden die x -te Komponente des Vektors Wp_t .

Unter Verwendung der Individuendynamik (2.4) und des Hardy-Weinberg-Gleichgewichtes (2.6) erhalten wir

$$\begin{aligned} z_{t+1}(x) &= \sum_{y \in X} z_{t+1}(x, y) = \sum_{y \in X} w(x, y) z_t(x, y) \\ &= \sum_{y \in X} w(x, y) p_t(y) z_t(x) = (Wp_t)(x) z_t(x) \end{aligned} \quad (2.7)$$

für die Dynamik der absoluten Gametenhäufigkeiten.

Durch Normieren resultiert daraus die diskrete Dynamik der relativen Gametenhäufigkeiten:

$$\begin{aligned} p_{t+1}(x) &= \frac{z_{t+1}(x)}{\|z_{t+1}\|} = \frac{(Wp_t)(x)}{\sum_{y \in X} (Wp_t)(y) z_t(y)} z_t(x) = \\ &= \frac{(Wp_t)(x)}{\sum_{y \in X} (Wp_t)(y) z_t(y) \frac{1}{\|z_t\|}} \frac{1}{\|z_t\|} z_t(x) = \frac{(Wp_t)(x)}{\langle p_t, Wp_t \rangle} p_t(x). \end{aligned} \quad (2.8)$$

Im letzten Schritt schreiben wir aus Gründen der Übersichtlichkeit die Summe über y als Skalarprodukt.

Es fällt auf, dass der Bruch zeitabhängig ist. Der Zähler hängt linear und der Nenner quadratisch von p_t ab. Die Dynamik wird dadurch sehr komplex. Wir vereinfachen sie deshalb dahingehend, dass wir bestimmte Bedingungen an die Fitnesskoeffizienten auf der Individuenebene stellen und dadurch einen Satz Fitnesskoeffizienten erhalten, der auf Gametenebene wirkt. Wir schränken die Koeffizienten darauf ein, dass es keine *Dominanz* geben darf. Dies entspricht dem Phänomen der *intermediären Merkmalsausprägung* und bedeutet, dass sich die Fitness eines Individuums zu gleichen Teilen und unabhängig aus den Werten der beteiligten Gameten zusammensetzt.

Im diskreten Fall ist dies möglich, wenn sich die viability Fitness aller Individuen als das geometrische Mittel von zwei Werten darstellen lässt, die jeweils nur von einem der beteiligten Gameten abhängen, wenn also

$$w(x, y) = \sqrt{\alpha(x)^2 \alpha(y)^2} \quad \forall x, y \in X \quad (2.9)$$

gilt. Die Quadrierung der α lässt sich in dieser Definition zwar auflösen, wir schreiben sie dennoch hin, damit die Bildung des geometrischen Mittels erkennbar bleibt.

Dies setzen wir nun in den Zähler und den Nenner von (2.8) ein und vereinfachen die Ausdrücke, so dass für den Zähler

$$(Wp_t)(x) = \sum_{y \in X} w(x, y) p_t(y) = \sum_{y \in X} \sqrt{\alpha(x)^2} \sqrt{\alpha(y)^2} p_t(y) = \alpha(x) \langle \alpha, p_t \rangle \quad (2.10)$$

gilt, wobei

$$\alpha := (\alpha(x_1), \dots, \alpha(x_{|X|}))^T. \quad (2.11)$$

Unter Verwendung von (2.10) lässt sich der Nenner zu

$$\langle p_t, Wp_t \rangle = \sum_{x \in X} (Wp_t)(x) p_t(x) = \sum_{x \in X} \langle \alpha, p_t \rangle \alpha(x) p_t(x) = \langle \alpha, p_t \rangle^2 \quad (2.12)$$

vereinfachen.

Nach Einsetzen von (2.10) und (2.12) in die Dynamik (2.8) und Kürzen ergibt sich

$$p_{t+1}(x) = \frac{(Wp_t)(x)}{\langle p_t, Wp_t \rangle} p_t(x) = \frac{\alpha(x) \langle \alpha, p_t \rangle}{\langle \alpha, p_t \rangle^2} p_t(x) = \frac{\alpha(x)}{\langle \alpha, p_t \rangle} p_t(x). \quad (2.13)$$

Im Vergleich zu (2.8) steht nun im Zähler eine Konstante, welche für jeden Gameten eine viability Fitness auf Gametenebene darstellt. Im Nenner steht die mittlere viability Fitness der Gametenpopulation, welche nur linear von den relativen Gametenhäufigkeiten abhängt. Der Faktor vor der relativen Gametenhäufigkeit ist zwar immer noch zeitabhängig, jedoch wesentlich unkomplizierter als dies in (2.8) der Fall war.

2.2.2 Kontinuierlich

Diese Rechnungen lassen sich analog auch im Kontinuierlichen durchführen. Allerdings vertauschen wir die letzten beide Schritte, so dass wir erst die intermediäre Merkmalsausprägung zum Vereinfachen nutzen, und daraufhin die Population auf 1 normieren. Dies liefert zwar das gleiche Ergebnis, wir erhalten aber als Zwischenergebnis eine Reproduktionsdynamik für die absoluten Gametenhäufigkeiten, welche wir im weiteren Verlauf benötigen.

Wir beginnen zunächst wieder damit, aus der Dynamik für die Individuen (2.5) durch Marginalisierung und unter Verwendung der Hardy-Weinberg-Proportionen (2.6) eine Dynamik auf der Gametenebene herzuleiten:

$$\begin{aligned} \dot{z}_t(x) &= \sum_{y \in X} \dot{z}_t(x, y) = \sum_{y \in X} u(x, y) z_t(x, y) \\ &= \sum_{y \in X} u(x, y) p_t(y) z_t(x) \\ &= (Up_t)(x) z_t(x). \end{aligned} \quad (2.14)$$

Im kontinuierlichen Fall kann die Reproduktionsdynamik auf der Gametenebene betrachtet werden, wenn sich die Malthus'sche Fitness eines Individuums als das arithmetische Mittel aus zwei jeweils nur von einem Gameten abhängigen Werten darstellen lässt, wenn es also Werte $f(x)$ gibt, so dass

$$u(x, y) = \frac{1}{2}(2f(x) + 2f(y)) \quad \forall x, y \in X \quad (2.15)$$

gilt. Auch in dieser Definition gilt, dass die Faktoren zwar gekürzt werden können, wir sie jedoch hinschreiben, um auf das arithmetische Mittel hinzuweisen.

Dies setzen wir in Gleichung (2.14) ein, um

$$\begin{aligned} \dot{z}_t(x) &= \sum_{y \in X} u(x, y) p_t(y) z_t(x) \\ &= \sum_{y \in X} \frac{1}{2}(2f(x) + 2f(y)) p_t(y) z_t(x) \\ &= f(x) z_t(x) + \langle f, p_t \rangle z_t(x) \end{aligned} \quad (2.16)$$

für die Gametendynamik zu erhalten, wobei wir die Normierung der Wahrscheinlichkeiten sowie die Abkürzung

$$f := (f(x_1), \dots, f(x_{|X|}))^T \quad (2.17)$$

verwenden.

Dies normieren wir, um die Dynamik der relativen Gametenhäufigkeiten zu erhalten

$$\begin{aligned} \frac{d}{dt} p_t(x) &= \frac{d}{dt} \frac{z_t(x)}{\|z_t\|} = \frac{\dot{z}_t(x) \|z_t\| - z_t(x) \|\dot{z}_t\|}{\|z_t\|^2} \\ &= \frac{1}{\|z_t\|} (f(x) z_t(x) + \langle f, p_t \rangle z_t(x)) - \frac{p_t(x)}{\|z_t\|} \sum_{y \in X} \dot{z}_t(y) \\ &= f(x) p_t(x) + \langle f, p_t \rangle p_t(x) - \frac{p_t(x)}{\|z_t\|} \sum_{y \in X} (f(y) z_t(y) + \langle f, p_t \rangle z_t(y)) \\ &= f(x) p_t(x) + \langle f, p_t \rangle p_t(x) - \langle f, p_t \rangle p_t(x) - \langle f, p_t \rangle p_t(x) \\ &= f(x) p_t(x) - \langle f, p_t \rangle p_t(x), \end{aligned} \quad (2.18)$$

wobei wir nacheinander die Quotientenregel und die Gametendynamik (2.16) verwenden und an gegebener Stelle relative Häufigkeiten einsetzen.

Nun gibt es erneut zwei Vorfaktoren für p_t . Einen konstanten Faktor, der die Malthus'sche Fitness eines Gameten angibt, und einen linear von p_t abhängigen, der für die Erhaltung der Normierung sorgt. Es ist also unter der Annahme einer intermediären Merkmalsausprägung im kontinuierlichen Fall ebenfalls ausreichend, die Selektionsdynamik nur auf der Gametenebene zu betrachten.

Wir wollen uns nun erneut der Reproduktionsdynamik auf Gametenebene zuwenden. Nach den Vereinfachungen stehen auf der rechten Seite der Differentialgleichung (2.16)

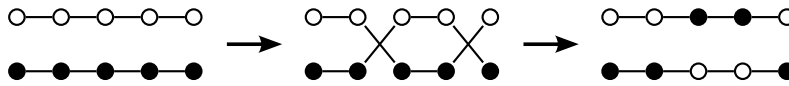


Abbildung 2.1: Schematische Darstellung des intrachromosomalen Rekombinationsprozesses zwischen zwei Gameten

zwei Summanden. Jeder dieser Summanden besteht aus einem Vorfaktor und der absoluten Gametenhäufigkeit $z_t(x)$. Bei dem ersten Summanden ist dieser Vorfaktor die zu dem Typen gehörende Malthus'sche Fitness und bei dem zweiten ein von dem Typen unabhängiger Faktor, der als mittlere Fitness der Population interpretiert werden kann. Diese mittlere Fitness ist zu jedem Zeitpunkt für alle Typen gleich.

Streichen wir diesen zweiten Summanden, so verringert sich die Anzahl der Summanden in der Rechnung (2.18), wir gelangen jedoch zu dem gleichen Ergebnis. Erhöhen wir die Raten der unterschiedlichen Typen also um den gleichen Wert, so hat dies keine Auswirkungen auf die normierte Dynamik, so dass wir, wenn wir auf eine spätere Normierung der Dynamik abzielen,

$$\dot{z}_t = f(x) z_t(x) \quad (2.19)$$

als Reproduktionsdynamik auf der Ebene der absoluten Häufigkeiten ansehen können.

Wir sind in diesem Abschnitt von der Reproduktionsdynamik auf der Ebene der absoluten Individuenhäufigkeiten aus gestartet und haben für den kontinuierlichen sowie für den diskreten Fall mit biologisch motivierten Einschränkungen argumentiert, um in beiden Fällen eine Selektionsdynamik auf der Gametenebene zu erhalten.

2.3 Rekombinationsdynamik

Nachdem wir im vorangegangenen Abschnitt Gleichungen für die Dynamik der Selektion in unserem Modell hergeleitet haben, wollen wir uns in diesem Abschnitt der Rekombination zuwenden. Wir betrachten also die Dynamik zunächst ausschließlich unter Rekombination.

Die in Abschnitt 1.1 geschilderte inter- und intrachromosomale Rekombination, welche bei der meiotischen Teilung stattfindet, modellieren wir direkt auf der Gametenebene, da sich die Gametenhäufigkeiten in der Diplophase nicht ändern. Somit werden die zwei Gameten, aus denen ein Individuum entstanden ist, durch Rekombination zu zwei neuen Gameten. Ein Gamet setzt sich dadurch zusammen, dass er an einer gegebenen Teilmenge der Sites die Allele des einen Eltergameten und an den verbliebenen Sites die Allele des anderen Elter erbt. Der zweite Gamet erhält die Allele der jeweils komplementären Sitemengen. Abbildung 2.1 illustriert dies. Da lediglich neue Kombinationen der schon bestehenden Allele gebildet werden, sind die marginalen Allelhäufigkeiten zeitlich konstant.

Die Rekombination findet zwischen den beiden Gameten statt, welche sich vor der Meiose zufällig getroffen haben. Da wir zunächst ohne Selektion modellieren und später

von einer intermediären Merkmalsausprägung ausgehen, ändert sich in der Diplophase nichts an den Hardy-Weinberg-Verhältnissen (2.6), weshalb die Paarbildung in unserem Modell der Rekombination zufällig geschieht.

2.3.1 Diskret

Die relative Häufigkeit eines bestimmten Gametentyps in der nächsten Generation ergibt sich also als die Summe über alle möglichen Teilmengen R der Sites S . Pro Summand gibt es einen Faktor, der die Wahrscheinlichkeit dafür angibt, dass die Rekombination gerade diese Teilmenge mit ihrer Komplementärmenge rekombiniert und einen Faktor, der angibt, wie wahrscheinlich es ist, dass bei einem solchen Rekombinationsereignis der gewünschte Gamet entsteht, nämlich die Wahrscheinlichkeit, dass ein Gamet der Population an den Sites aus R die richtigen Allele stehen hat, multipliziert mit derselben Wahrscheinlichkeit für die Sites aus \bar{R} . Das diskrete dynamische System genügt also der Gleichung

$$p_{t+1}(x) = \sum_{R \subset S} r(R) p_t(x_{r_1}, *, x_{r_2}, \dots, *, x_{r_{|R|}}) p_t(*, x_{\bar{r}_1}, *, \dots, x_{\bar{r}_{|\bar{R}|}}, *), \quad (2.20)$$

wobei $r_i \in R$ und $\bar{r}_i \in \bar{R}$ gilt. Wir schreiben hier $p_t(x_1, \dots, x_n) := p_t(x)$ für die relative Gametenhäufigkeit. $r(R)$ ist hier die Wahrscheinlichkeit, dass ein Rekombinationsereignis stattfindet, welches die Sitemenge R von der Menge \bar{R} trennt. Es geschieht immer mindestens ein Rekombinationsereignis, und sei es nur das leere Ereignis $r(\emptyset)$, also gilt

$$\sum_{R \subset S} r(R) = 1. \quad (2.21)$$

Durch die Wahrscheinlichkeiten $r(\cdot)$ können wir wählen, ob inter- oder intrachromosomale Rekombination modelliert werden soll. Die Wahrscheinlichkeiten für die nicht leeren Rekombinationsereignisse müssen dann sehr niedrig respektive sehr hoch gewählt werden. Es sind sogar beliebige Mischungen möglich.

Gleichung (2.20) kann mit der in Abschnitt 2.1 eingeführten Zylindermengenschreibweise umgeschrieben werden zu

$$p_{t+1}(\langle S \rangle) = \sum_{R \subset S} r(R) p_t(\langle R \rangle) p_t(\langle \bar{R} \rangle). \quad (2.22)$$

Durch Marginalisieren von Gleichung (2.22) kann auch eine Dynamik für nicht vollständig spezifizierte Gameten erhalten werden:

$$p_{t+1}(\langle J \rangle) = \sum_{R \subset J} r_J(R) p_t(\langle R \rangle) p_t(\langle \mathcal{C}_J(R) \rangle). \quad (2.23)$$

Die Marginalisierung geschieht durch die Zylindermengenschreibweise für die relativen Gametenhäufigkeiten automatisch. Als Rekombinationswahrscheinlichkeiten muss ein

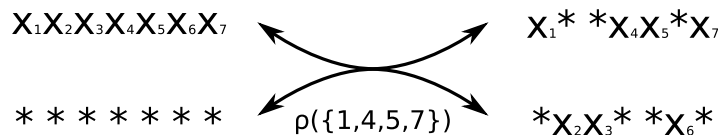


Abbildung 2.2: In dem dargestellten Beispiel trennt ein Rekombinationsereignis die Sites der Menge $\{1, 4, 5, 7\}$ von der Menge $\{2, 3, 6\}$. Dies geschieht mit Rate $\rho(\{1, 4, 5, 7\})$. Dass die Pfeile in beide Richtungen gehen, bedeutet, dass Typen der Form x sowohl entstehen als auch verschwinden können.

neuer Satz $r_J(R)$ verwendet werden. Hierbei handelt es sich um die Marginale der $r(R)$ und es gilt

$$r_J(R) = \sum_{K \subset (\bar{J} \cup R)} r(K). \quad (2.24)$$

Diese diskrete Rekombinationsdynamik ist weit verbreitet und wurde in der Literatur eingehend untersucht. In Kapitel 3 werden einige dieser Untersuchungen genauer erörtert. Die Ergebnisse, welche wir in Kapitel 5 reproduzieren wollen, wurden ebenfalls mit einem diskreten Modell der Rekombinations- und Selektionsdynamik gewonnen.

2.3.2 Kontinuierlich

Die Änderung der relativen Häufigkeit eines Gametentyps setzt sich aus zwei Teilen zusammen. Jede mögliche Trennung von Teilmengen der Sites geschieht mit einer bestimmten Rate. Dies bedeutet zum einen, dass die Häufigkeit des Typs mit dieser Rate abnimmt, da ein solches Rekombinationsereignis unter Beteiligung des betrachteten Typs diesen zerstört. Andererseits kann durch ein solches Ereignis aber auch ein Typ der betrachteten Form entstehen, wenn sich zwei Partner treffen, die unter der gegebenen Rekombination den gewünschten Typ ergeben. In Abbildung 2.2 ist dieses Verhalten für ein Beispiel dargestellt. Da alle möglichen Trennungen der Teilmengen zu der Dynamik des untersuchten Typs beitragen können, muss über alle Teilmengen summiert werden, so dass die Bilanzgleichung

$$\dot{p}_t(x) = \sum_{R \subset S} \rho(R) \left(p_t(x_{r_1}, *, x_{r_2}, \dots, *, x_{r_{|R|}}) p_t(*, x_{\bar{r}_1}, *, \dots, x_{\bar{r}_{|\bar{R}|}}, *) - p_t(x) \right) \quad (2.25)$$

lautet. Es gilt wie schon im diskreten Fall $r_i \in R$ und $\bar{r}_i \in \bar{R}$. Weiter ist $\rho(R)$ die Rate des Rekombinationsereignisses, welches die Sites in R von den Sites in \bar{R} trennt.

Durch geeignete Wahl der Rekombinationsraten lässt sich hier ebenfalls sowohl intra- als auch interchromosomale Rekombination untersuchen. Im Gegensatz zum diskreten Modell wird freie Neukombination und damit interchromosomale Rekombination dadurch erreicht, dass die Rekombinationsraten gegen unendlich gehen. Kleine Rekombinationsraten entsprechen aber ebenfalls intrachromosomaler Rekombination.

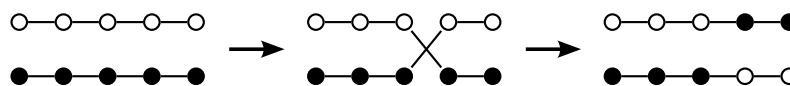


Abbildung 2.3: Ein Rekombinationsereignis, bei dem es nur zu einem Crossover kommt, also lediglich einem Austausch der Enden stattfindet.

Gleichung (2.25) lässt sich ebenfalls besser in der Zylindermengenschreibweise notieren als

$$\dot{p}_t(x) = \sum_{RCS} \varrho(R) \left(p_t(\langle R \rangle) p_t(\langle \bar{R} \rangle) - p_t(x) \right). \quad (2.26)$$

In der Literatur wurde die kontinuierliche Rekombinationsdynamik weitestgehend unbeachtet gelassen. Sie dient jedoch als Grundlage für die in dieser Arbeit zentrale Dynamik, welche wir im nächsten Abschnitt erläutern.

2.4 Rekombinationsdynamik auf geordneten Partitionen

In [1] stellt Baake eine modifizierte Version der Dynamik (2.26) vor. Es werden nur solche Terme betrachtet, welche einem Rekombinationsereignis entsprechen, das die Sites in eine geordnete Partition der Größe zwei zerlegt. Durch diese Einschränkung lässt sich die resultierende Dynamik besser lösen.

Es gibt also einen Term für jedes $\alpha \in \{\frac{1}{2}, \dots, \frac{2n-1}{2}\} =: L$ und die Mengen, in welche die Sites zerlegt werden, sind $R = \{0, \dots, \lfloor \alpha \rfloor\}$ und $\bar{R} = \{\lceil \alpha \rceil, \dots, n\}$. Die Rate eines solchen Ereignisses bezeichnen wir mit ϱ_α . Alle anderen Raten werden auf null gesetzt.

Diese Ereignisse entsprechen einzelnen Crossoverereignissen. Es wird argumentiert, dass bei den anderen Rekombinationsereignissen zwei oder mehr Crossover zu einem Zeitpunkt auftreten müssen und die Raten dafür sehr gering sind, weshalb diese Terme in einer ersten Näherung vernachlässigt werden können. Abbildung 2.1 stellt ein vernachlässigbares Ereignis dar, wohingegen auf Abbildung 2.3 ein relevantes zu sehen ist.

Die reduzierte Dynamik ergibt sich somit zu

$$\dot{p}_t(x) = \sum_{\alpha \in L} \varrho_\alpha \left(p_t(x_0, \dots, x_{\lfloor \alpha \rfloor}, *, \dots, *) p_t(*, \dots, *, x_{\lceil \alpha \rceil}, \dots, x_n) - p_t(x) \right). \quad (2.27)$$

Baake führt in ihrer Arbeit Rekombinationsoperatoren, sogenannte *Rekombinatoren* R_α ein, um diese Dynamik besser bearbeiten zu können. Wir gehen in Abschnitt 4.2.2 genauer auf die Rekombinatoren ein und definieren an dieser Stelle lediglich

$$(R_\alpha z)(x) := \frac{1}{\|z\|} z(x_0, \dots, x_{\lfloor \alpha \rfloor}, *, \dots, *) p(*, \dots, *, x_{\lceil \alpha \rceil}, \dots, x_n), \quad (2.28)$$

so dass wir Gleichung (2.27) zu

$$\dot{p}_t(x) = \sum_{\alpha \in L} \varrho_\alpha \left((R_\alpha p_t)(x) - p_t(x) \right) \quad (2.29)$$

umschreiben können.

Durch die adäquat definierten Rekombinatoren erhält diese Dynamik sowohl die Norm eines Wahrscheinlichkeitsmaßes $p \in \mathcal{P}^\otimes$ als auch die eines beliebigen positiven Maßes $z \in \mathcal{M}^\otimes$.

Diese modifizierte Dynamik wird in Kapitel 4 eingehend behandelt und gelöst.

2.5 Rekombinations- und Selektionsdynamik

Wir modellieren Rekombination und Selektion als zwei unabhängig voneinander wirkende Kräfte. Dies muss in der Natur nicht zwingend gegeben sein, hat aber den Vorteil, dass die Dynamik übersichtlich bleibt.

Außerdem können Wechselwirkungen zwischen diesen beiden Kräften nach Argumentation analog zu der in [10] in erster Näherung in der Regel vernachlässigt werden.

Bei parallel wirkender Rekombination und Reproduktion ist die Bilanzgleichung für die zeitliche Entwicklung der absoluten Häufigkeiten lediglich die Summe von (2.19) und (2.29) und ergibt sich zu

$$\dot{z}_t = \left(F + \sum_{\alpha \in L} \varrho_\alpha (R_\alpha - \mathbb{1}) \right) z_t. \quad (2.30)$$

Diese Dynamik erhält nicht die Norm des Maßes z_t .

Um die normerhaltende Dynamik für die Wahrscheinlichkeitsmaße aus \mathcal{P}^\otimes zu erhalten, können wir die beiden Teile der Dynamik (2.30) getrennt normieren oder direkt Gleichung (2.18) und (2.29) aufsummieren, um so zu der normerhaltenden Rekombinations- und Selektionsdynamik

$$\dot{p}_t = \left(F - \bar{f}_t \mathbb{1} + \sum_{\alpha \in L} \varrho_\alpha (R_\alpha - \mathbb{1}) \right) p_t \quad (2.31)$$

zu gelangen, wobei wir mit

$$\bar{f}_t := \frac{\|F z_t\|}{\|z_t\|} = \langle f, p_t \rangle \quad (2.32)$$

die mittlere Fitness der Population bezeichnen. Wir sind hier zur übersichtlicheren Vektorschreibweise der Differentialgleichungen übergegangen.

Diese Dynamik lösen wir in Kapitel 4 und untersuchen in Kapitel 5 einige Aspekte und Besonderheiten von parallel wirkender Rekombination und Selektion.

3 Kopplungsungleichgewichte

In der Literatur haben sich für Untersuchungen der Rekombinationsdynamik bestimmte Funktionen etabliert, welche Korrelationen zwischen den Allelhäufigkeiten der Sites beschreiben. Diese *Kopplungsungleichgewichte*, oder englisch *linkage disequilibria* (LDE), wurden von diversen Autoren unterschiedlich definiert, was zu Verwechslungen führte. Die folgende Darstellung soll die unterschiedlichen Definitionen und die Beziehungen unter ihnen aufzeigen. Sie ist zu großen Teilen an eine ähnliche Darstellung von Bürger [6, S. 183–186] angelehnt, hat jedoch einen anderen Fokus und enthält einige Ergänzungen.

Wie schon eingangs erwähnt, wurde die diskrete Rekombinationsdynamik (2.22) in der Literatur am Intensivsten untersucht. Daher basieren die nachstehenden Untersuchungen größtenteils auf dieser Dynamik. Ferner wurden zum Teil diploide Organismen betrachtet, so dass wir im Folgenden die Kopplungsungleichgewichte auf haploide Organismen anpassen, sofern dies nötig ist.

3.1 Geschichte der Kopplungsungleichgewichte

Bereits 1944 zeigte Geiringer [9], dass die relativen Gametenhäufigkeiten durch die Rekombinationsdynamik (2.22) gegen das Produkt der marginalen Allelwahrscheinlichkeiten konvergieren.

$$p_t(\langle S \rangle) \xrightarrow{t \rightarrow \infty} \tilde{p}(\langle S \rangle) = \prod_{i=0}^n p(*, \dots, *, x_i, *, \dots, *). \quad (3.1)$$

Diese Marginalwahrscheinlichkeiten sind von Anfang an gegeben und unter der Dynamik (2.22) zeitlich konstant.

Die Kopplungsungleichgewichte wurden von verschiedenen Autoren unterschiedlich definiert, sind aber stets ein Maß der Abweichung von diesem Gleichgewichtszustand, in welchem die marginalen Allelwahrscheinlichkeiten entkoppelt sind. Hier bedeutet entkoppeln, dass die Wahrscheinlichkeit für ein bestimmtes Allel an einem Site unabhängig davon ist, welche Allele an den anderen Sites sind und nur durch die relative Häufigkeit der Allele an dem untersuchten Site bestimmt ist. Die Kopplungsungleichgewichte sind also Funktionen, welche die Stärke der Korrelation zwischen den Allelhäufigkeiten an den verschiedenen Sites angeben. Der entkoppelte Zustand wird auch *Kopplungsgleichgewicht* genannt.

Unter diesem Gesichtspunkt sind die direkte Differenz der Wahrscheinlichkeit des aktuellen Zustands und der entkoppelten Wahrscheinlichkeit

$$\theta(\langle J \rangle) := p(\langle J \rangle) - \tilde{p}(\langle J \rangle) \quad (3.2)$$

intuitiv naheliegende Kopplungsungleichgewichte.

Diese Kopplungsungleichgewichte werden von Lyubich in [14, S. 257] mit $E_K(p)$ bezeichnet und als Zwischenstufe eingeführt, auf der weitere Definitionen aufbauen.

Im Jahre 1954 führte Bennett [3] eine Transformation ein, welche die Typhäufigkeiten in sogenannte *Hauptkomponenten* überführt. Diese Transformation wird so gewählt, dass sie die nichtlineare Dynamik der Gametenhäufigkeiten (2.22) in eine entkoppelte lineare Dynamik der Hauptkomponenten umwandelt. Die Hauptkomponenten zerfallen also unabhängig und geometrisch und werden im Gleichgewicht ebenfalls null.

Bennett legte lediglich für Systeme aus 1, 2, 3, 4 und 6 Loci explizit die Berechnungsformeln dar, und umriss grob, wie diese auf allgemeine Fälle auszudehnen seien. Seine Untersuchungen beschränkten sich auf die diskrete Dynamik (2.22), allerdings behauptete er, dass die Hauptkomponenten für die kontinuierliche Rekombinationsdynamik (2.26) nach dem gleichen Schema berechnet werden können.

Bennetts Ideen für die diskrete Dynamik wurden von Dawson [8] erneut aufgegriffen. Er gab an, dass sich die von Bennett eingeführten Hauptkomponenten für ein System mit beliebig vielen Sites als

$$b(\langle J \rangle) = \sum_{F \in \Phi(J)} c_F \cdot \left(\prod_{A \in F} p(\langle A \rangle) \right), \quad (3.3)$$

darstellen lassen, wobei $\Phi(J)$ die Menge alle Partitionen von J ist.

Jedoch gab er keine expliziten Formeln für die c_F an, sondern nur, wie diese sich rekursiv berechnen lassen. Sind vier oder mehr Sites beteiligt, so hängen diese Konstanten zusätzlich noch von den Rekombinationswahrscheinlichkeiten ab.

Lewontin und Kojima [13] erwähnten den Begriff Kopplungsungleichgewicht zum ersten Mal. Die Autoren untersuchten ein 2-Site-/2-Allel-System, weshalb sie sowohl Bennetts Hauptkomponenten als auch die θ verwendeten, da diese für zwei beteiligte Sites identisch sind. Jedoch sind sie die ersten, die diese Abweichung von Gleichgewichtszustand mit dem Begriff Kopplungsungleichgewicht bezeichneten.

Das nächste Kopplungsungleichgewicht wurde von Slatkin [18] im Jahre 1972 eingeführt. Er untersuchte die zentralen Momente eines Produktes von Indikatorvariablen, welche je Site angeben, ob ein bestimmtes Allel vorhanden ist oder nicht.

Zunächst zeigt er, dass

$$\mathbb{E}(\mathbb{1}_{\{\chi_j = x_j\}}) = p(\langle j \rangle) \quad \text{und} \quad \mathbb{E}\left(\prod_{j \in J} \mathbb{1}_{\{\chi_j = x_j\}}\right) = p(\langle J \rangle)$$

für $j \in S$ und $J \subset S$ gilt.

Slatkin definierte daraufhin die zentralen Momente dieses Produktes von Indikatorvariablen als Kopplungsungleichgewichte. Genauer definiert er

$$\Delta(\langle J \rangle) := \mathbb{E}\left(\prod_{j \in J} (\mathbb{1}_{\{\chi_j = x_j\}} - p(\langle j \rangle))\right) \quad (3.4)$$

für $J \subset S$.

Dies kann ausmultipliziert und zu

$$\Delta(\langle J \rangle) = \sum_{RCJ} (-1)^{|\mathcal{C}_J(R)|} p(\langle R \rangle) \tilde{p}(\langle \mathcal{C}_J(R) \rangle) \quad (3.5a)$$

$$= \sum_{RCJ} (-1)^{|\mathcal{C}_J(R)|} \theta(\langle R \rangle) \tilde{p}(\langle \mathcal{C}_J(R) \rangle) \quad (3.5b)$$

vereinfacht werden. Details zu diesen Rechnungen finden sich in [17].

Die zentralen Momente verhalten sich unter der diskreten Rekombinationsdynamik (2.22) wie folgt:

$$\Delta_{t+1}(\langle J \rangle) = \sum_{ACJ} r_J(A) \Delta_t(\langle A \rangle) \Delta_t(\langle \mathcal{C}_J(A) \rangle). \quad (3.6)$$

Ein Beweis hierzu findet sich in Abschnitt 3.4.

Die Kumulanten des Produktes der Indikatorvariablen $\mathbb{1}_{\{x_j=x_j\}}$ können ebenfalls als ein Maß der Abweichung von der Unabhängigkeit der Sites auf Gametenebene betrachtet werden. Schreibt man diese in derselben Form wie die zentralen Momente auf, so ergibt sich

$$\Gamma(\langle J \rangle) := \sum_{F \in \Phi(J)} (-1)^{(|F|-1)} (|F|-1)! \prod_{A \in F} p(\langle A \rangle). \quad (3.7)$$

Dies hat im Wesentlichen die Form, die Dawson für Bennetts Hauptkomponenten angab, wenn $c_F = (-1)^{(|F|-1)} (|F|-1)!$ gewählt wird. Über Untersuchungen zur Rekombinationsdynamik der Kumulanten auf Gametenebene ist nichts Genaueres bekannt.

Ein weiteres Kopplungsungleichgewicht auf Gametenebene wurde 1991 von Nagylaki [16] eingeführt. In seinen Untersuchungen betrachtet Nagylaki die Dynamik der Population unter Rekombination und Selektion, hier seien die Kopplungsungleichgewichte aber nur für den Fall ohne Selektion dargestellt.

Zunächst einmal lässt sich die Rekombinationsdynamik (2.22) zu

$$p_{t+1}(\langle J \rangle) = p_t(\langle J \rangle) - \sum_{RCJ} r_J(R) \left(p_t(\langle J \rangle) - p_t(\langle R \rangle) p_t(\langle \mathcal{C}_J(R) \rangle) \right) \quad (3.8)$$

umschreiben. In einem Gleichgewichtszustand gilt $p_{t+1}(\langle J \rangle) = p_t(\langle J \rangle)$, also muss die Summe gleich null sein. Dies motivierte Nagylaki dazu, diese Summe als Kopplungsungleichgewicht

$$\Theta(\langle J \rangle) := \sum_{RCJ} r_J(R) \left(p_t(\langle J \rangle) - p_t(\langle R \rangle) p_t(\langle \mathcal{C}_J(R) \rangle) \right) \quad (3.9)$$

zu definieren.

Baake [1] betrachtete 2003, wie in Abschnitt 2.4 bereits erwähnt wurde, einen Spezialfall der kontinuierlichen Rekombinationsdynamik, in welchem seltene Mehrfachüberkreuzungen vernachlässigt werden. Sie führte die Kopplungsungleichgewichtsoperatoren

$$T_G(p_t) := \sum_{H \supseteq G} (-1)^{|H-K|} R_G(p_t) \quad (3.10)$$

ein, welche die Rekombinationsdynamik (2.29) entkoppeln und linearisieren. Mit

$$R_H := \bigcirc_{\alpha \in H} R_\alpha. \quad (3.11)$$

wird die Hintereinanderausführung der kommutierenden Operatoren R_α für $\alpha \in H$ bezeichnet. In Abschnitt 4.3 gehen wir noch genauer auf diese Kopplungsungleichgewichte ein.

3.2 Zusammenhang der Kopplungsungleichgewichte

Wir wollen nun einige Eigenschaften der Kopplungsungleichgewichte herausstellen und aufzeigen, weshalb sie für die Untersuchungen der jeweiligen Autoren von Interesse waren. Des Weiteren gehen wir an den entsprechenden Stellen auf Unterschiede und Gemeinsamkeiten ein.

Alle bisher eingeführten Kopplungsungleichgewichte werden im entkoppelten Gleichgewichtszustand null. Jedoch linearisieren und entkoppeln lediglich die von Bennett eingeführten Hauptkomponenten und die von Baake eingeführten T_G die Dynamik und zerfallen geometrisch respektive exponentiell. Die zentralen Momente unterliegen derselben Dynamik wie die Gametenhäufigkeiten, die Transformation erhält also die nichtlineare Rekombinationsdynamik. Für die Kumulanten und Θ ist keine geschlossene Dynamik unter Rekombination bekannt, jedoch behauptet Bürger [6, S. 57], dass es möglich ist, für die Kumulanten geometrisch zerfallende obere Schranken anzugeben. Dasselbe wird von Nagylaki [16] für die Θ behauptet. Alle Kopplungsungleichgewichte konvergieren also unter der Rekombinationsdynamik gegen null. Weitere Eigenschaften der θ sind nicht bekannt.

Die Transformationen in die zentralen Momente, Kumulanten, Hauptkomponenten und T_G sind invertierbar. Es sind also jeweils komplette Variablensätze, welche zum Spezifizieren einer Population ausreichen. Die Variablensätze lassen sich durch Herausrechnen gewisser Abhängigkeiten reduzieren, genauere Untersuchungen hierzu sind lediglich bei Baake [1] für die T_G zu finden. Innerhalb der θ und Θ gibt es solche Abhängigkeiten wohl auch, allerdings ist hier noch weniger bekannt als für die vorherigen.

Die einfachste Struktur haben die θ . Sie bestehen aus jeweils zwei Summanden und zwar der sich zeitlich verändernden Gametenwahrscheinlichkeit und dem konstanten Produkt der marginalen Allelwahrscheinlichkeiten. Die zentralen Momente haben mehrere Summanden, allerdings ist jeder Summand ein Produkt aus maximal einer sich zeitlich verändernden Teilgametenwahrscheinlichkeit multipliziert mit zeitlich konstanten Marginalwahrscheinlichkeiten. Bei den Θ besteht jeder Summand aus einem Produkt von zwei Wahrscheinlichkeiten. Hier können aber beide Faktoren zeitlich veränderliche Wahrscheinlichkeiten sein, so dass es für mehr als drei Sites nichtlineare Summanden gibt. Bei den Hauptkomponenten und Kumulanten kommen alle möglichen Kombinationen von Wahrscheinlichkeiten als Summanden vor. Für jede Partition der Menge der betrachteten Sites gibt es einen Summanden, wie aus Gleichung (3.7) und (3.3) leicht ersichtlich ist. Lediglich die Faktoren vor den Summanden unterscheiden sich bei die-

sen beiden Kopplungsungleichgewichten. Bei den T_G gibt es nur die Summanden für geordnete Partitionen und die Faktoren sind $+1$ oder -1 .

Interessant ist, dass sich bei bestimmten Kopplungsungleichgewichten für wenige betrachtete Sites dieselben Ausdrücke ergeben, weshalb sie in der Literatur auch verwechselt wurden. Aufgrund ihrer völlig anderen Struktur gilt dies für die Θ nicht. Sie ähneln auf keiner Ebene den anderen Kopplungsungleichgewichten, da bei ihnen immer die Rekombinationsparameter in den Summanden stehen.

Für zwei Sites liefern allerdings die Hauptkomponenten, zentralen Momente, Kumulanten, T_G und θ jeweils dieselben Ausdrücke, da die oben beschriebenen Unterschiede hier noch nicht sichtbar sind. Bei drei betrachteten Sites liefern die θ und T_G abweichende Ausdrücke, die zentralen Momente, Hauptkomponenten und Kumulanten sind aber noch gleich. Ab vier Sites treten die Unterschiede heraus, und alle Kopplungsungleichgewichte liefern verschiedene Ausdrücke.

3.3 Kopplungsungleichgewichte auf vier Sites

Im vorangegangenen Abschnitt erwähnten wir bereits, dass alle Kopplungsungleichgewichte in der vierten Ordnung unterschiedliche Ausdrücke liefern. Deshalb seien hier nochmal die Kopplungsungleichgewichte für vier Sites explizit aufgeschrieben und auf die sichtbaren Unterschiede an geeigneter Stelle aufmerksam gemacht.

Ohne Beschränkung der Allgemeinheit bezeichnen wir hier die Indizes aus der vierelementigen Menge J mit 1, 2, 3 und 4. Des Weiteren bedienen wir uns im Folgenden der abkürzenden Schreibweise $\langle J \rangle$ anstelle von $p(\langle J \rangle)$. Es ergibt sich für

a) die direkte Differenz

$$\theta(\langle 1, 2, 3, 4 \rangle) = \langle 1, 2, 3, 4 \rangle - \langle 1 \rangle \langle 2 \rangle \langle 3 \rangle \langle 4 \rangle, \quad (3.12)$$

b) die Hauptkomponenten

$$\begin{aligned} \text{b}(\langle 1, 2, 3, 4 \rangle) &= \langle 1, 2, 3, 4 \rangle \\ &- \langle 1 \rangle \langle 2, 3, 4 \rangle - \langle 2 \rangle \langle 1, 3, 4 \rangle - \langle 3 \rangle \langle 1, 2, 4 \rangle - \langle 4 \rangle \langle 1, 2, 3 \rangle \\ &+ (1 + \alpha) \langle 1, 2 \rangle \langle 3 \rangle \langle 4 \rangle + (1 + \beta) \langle 1, 3 \rangle \langle 2 \rangle \langle 4 \rangle + (1 + \gamma) \langle 1, 4 \rangle \langle 2 \rangle \langle 3 \rangle \\ &+ (1 + \gamma) \langle 2, 3 \rangle \langle 1 \rangle \langle 4 \rangle + (1 + \beta) \langle 2, 4 \rangle \langle 1 \rangle \langle 3 \rangle + (1 + \alpha) \langle 3, 4 \rangle \langle 1 \rangle \langle 2 \rangle \\ &- \alpha \langle 1, 2 \rangle \langle 3, 4 \rangle - \beta \langle 1, 3 \rangle \langle 2, 4 \rangle - \gamma \langle 1, 4 \rangle \langle 2, 3 \rangle \\ &- (3 + \alpha + \beta + \gamma) \langle 1 \rangle \langle 2 \rangle \langle 3 \rangle \langle 4 \rangle, \end{aligned} \quad (3.13)$$

wobei

$$\begin{aligned} \alpha &:= \frac{r_{\{1,2,3,4\}}(\{1,2\})}{r_{\{1,2\}}(\{1,2\}) r_{\{3,4\}}(\{3,4\}) - r_{\{1,2,3,4\}}(\{1,2,3,4\})} \\ \beta &:= \frac{r_{\{1,2,3,4\}}(\{1,3\})}{r_{\{1,3\}}(\{1,3\}) r_{\{2,4\}}(\{2,4\}) - r_{\{1,2,3,4\}}(\{1,2,3,4\})} \end{aligned}$$

$$\gamma := \frac{r_{\{1,2,3,4\}}(\{1,4\})}{r_{\{1,4\}}(\{1,4\})r_{\{2,3\}}(\{2,3\}) - r_{\{1,2,3,4\}}(\{1,2,3,4\})},$$

c) die zentralen Momente

$$\begin{aligned} \Delta(\langle 1, 2, 3, 4 \rangle) &= \langle 1, 2, 3, 4 \rangle \\ &\quad - \langle 1 \rangle \langle 2, 3, 4 \rangle - \langle 2 \rangle \langle 1, 3, 4 \rangle - \langle 3 \rangle \langle 1, 2, 4 \rangle - \langle 4 \rangle \langle 1, 2, 3 \rangle \\ &\quad + \langle 1, 2 \rangle \langle 3 \rangle \langle 4 \rangle + \langle 1, 3 \rangle \langle 2 \rangle \langle 4 \rangle + \langle 1, 4 \rangle \langle 2 \rangle \langle 3 \rangle \\ &\quad + \langle 2, 3 \rangle \langle 1 \rangle \langle 4 \rangle + \langle 2, 4 \rangle \langle 1 \rangle \langle 3 \rangle + \langle 3, 4 \rangle \langle 1 \rangle \langle 2 \rangle \\ &\quad - 3 \langle 1 \rangle \langle 2 \rangle \langle 3 \rangle \langle 4 \rangle, \end{aligned} \tag{3.14}$$

d) die Kumulanten

$$\begin{aligned} \Gamma(\langle 1, 2, 3, 4 \rangle) &= \langle 1, 2, 3, 4 \rangle \\ &\quad - \langle 1 \rangle \langle 2, 3, 4 \rangle - \langle 2 \rangle \langle 1, 3, 4 \rangle - \langle 3 \rangle \langle 1, 2, 4 \rangle - \langle 4 \rangle \langle 1, 2, 3 \rangle \\ &\quad + 2 \langle 1, 2 \rangle \langle 3 \rangle \langle 4 \rangle + 2 \langle 1, 3 \rangle \langle 2 \rangle \langle 4 \rangle + 2 \langle 1, 4 \rangle \langle 2 \rangle \langle 3 \rangle \\ &\quad + 2 \langle 2, 3 \rangle \langle 1 \rangle \langle 4 \rangle + 2 \langle 2, 4 \rangle \langle 1 \rangle \langle 3 \rangle + 2 \langle 3, 4 \rangle \langle 1 \rangle \langle 2 \rangle \\ &\quad - \langle 1, 2 \rangle \langle 3, 4 \rangle - \langle 1, 3 \rangle \langle 2, 4 \rangle - \langle 1, 4 \rangle \langle 2, 3 \rangle \\ &\quad - 6 \langle 1 \rangle \langle 2 \rangle \langle 3 \rangle \langle 4 \rangle, \end{aligned} \tag{3.15}$$

e) Nagylakis Θ

$$\begin{aligned} \Theta(\langle 1, 2, 3, 4 \rangle) &= (1 - r(\{1, 2, 3, 4\}) - r(\emptyset)) \langle 1, 2, 3, 4 \rangle \\ &\quad - (r(\{1\}) + r(\{2, 3, 4\})) \langle 1 \rangle \langle 2, 3, 4 \rangle \\ &\quad - (r(\{2\}) + r(\{1, 3, 4\})) \langle 2 \rangle \langle 1, 3, 4 \rangle \\ &\quad - (r(\{3\}) + r(\{1, 2, 4\})) \langle 3 \rangle \langle 1, 2, 4 \rangle \\ &\quad - (r(\{4\}) + r(\{1, 2, 3\})) \langle 4 \rangle \langle 1, 2, 3 \rangle \\ &\quad - (r(\{1, 2\}) + r(\{3, 4\})) \langle 1, 2 \rangle \langle 3, 4 \rangle \\ &\quad - (r(\{1, 3\}) + r(\{2, 4\})) \langle 1, 3 \rangle \langle 2, 4 \rangle \\ &\quad - (r(\{1, 4\}) + r(\{2, 3\})) \langle 1, 4 \rangle \langle 2, 3 \rangle, \end{aligned} \tag{3.16}$$

wobei wir $r(\cdot)$ für $r_{\{1,2,3,4\}}(\cdot)$ verwenden

f) und für Baakes T_G

$$\begin{aligned} T_{\emptyset}(p)(\langle 1, 2, 3, 4 \rangle) &= \langle 1, 2, 3, 4 \rangle \\ &\quad - \langle 1 \rangle \langle 2, 3, 4 \rangle - \langle 1, 2, 3 \rangle \langle 4 \rangle \\ &\quad + \langle 1, 2 \rangle \langle 3 \rangle \langle 4 \rangle + \langle 1 \rangle \langle 2, 3 \rangle \langle 4 \rangle + \langle 1 \rangle \langle 2 \rangle \langle 3, 4 \rangle \\ &\quad - \langle 1, 2 \rangle \langle 3, 4 \rangle \\ &\quad - \langle 1 \rangle \langle 2 \rangle \langle 3 \rangle \langle 4 \rangle. \end{aligned} \tag{3.17}$$

Hier lässt sich gut verdeutlichen, dass es bei den Hauptkomponenten b und bei den Kumulanten Γ für jede mögliche Partition der vierelementigen Menge J einen Summanden gibt, sich die Vorfaktoren jedoch unterscheiden. Bei den zentralen Momenten Δ sind keine nichtlinearen Terme vorhanden und die Θ bestehen lediglich aus Summanden, welche aus zwei Wahrscheinlichkeitsfaktoren zusammengesetzt sind. Die T_G besitzen lediglich Summanden für geordnete Partitionen und θ besteht aus nur zwei Summanden.

Es ist ebenfalls gut erkennbar, dass die θ , Δ , Γ und T_G im Gegensatz zu den Θ und b unabhängig von den Rekombinationsparametern sind.

Bei jedem der Kopplungsungleichgewichte ist zu beobachten, dass es unter Berücksichtigung der Vorfaktoren gleich viele positive wie negative Summanden gibt. Im Grenzfall gilt $p(\cdot) = \tilde{p}(\cdot)$, weshalb sie sich gegenseitig aufheben, da alle Summanden bis auf die Koeffizienten identisch sind. Das Kopplungsungleichgewicht verschwindet. Diese Beobachtung gilt jedoch für Θ nicht.

Mit der genannten Beweisidee ist es möglich, anhand der allgemeinen Formeln für die Kopplungsungleichgewichte zu beweisen, dass diese im Grenzfall gleich Null sind.

3.4 Dynamik der zentralen Momente

Wir wollen nun den in Abschnitt 3.1 angekündigten Beweis für die Dynamik der zentralen Momente liefern. In [7, S. 39] wird dieser Beweis lediglich für zwei Allele pro Site geführt. Diese Beweisidee wollen wir im Folgenden auf den Fall beliebig aber endlich vieler Allele erweitern.

Satz 3.1 *Die zentralen Momente (3.5) verhalten sich unter der diskreten Rekombinationsdynamik (2.22) genauso wie die Gametenhäufigkeiten. Es gilt*

$$\Delta_{t+1}(\langle J \rangle) = \sum_{A \subset J} r_J(A) \Delta_t(\langle A \rangle) \Delta_t(\langle \mathcal{C}_J(A) \rangle). \quad (3.18)$$

Beweis: Zunächst zeigen wir, dass sich die Gametenhäufigkeiten durch die zentralen Momente ausdrücken lassen. Für $A \subset S$ gilt

$$\begin{aligned} & \sum_{G \subset A} \tilde{p}(\langle \mathcal{C}_A(G) \rangle) \Delta(\langle G \rangle) \\ &= \sum_{G \subset A} \tilde{p}(\langle \mathcal{C}_A(G) \rangle) \left(\sum_{K \subset G} (-1)^{|\mathcal{C}_G(K)|} p(\langle K \rangle) \tilde{p}(\langle \mathcal{C}_G(K) \rangle) \right) \\ &= \sum_{G \subset A} \sum_{K \subset G} (-1)^{|\mathcal{C}_G(K)|} p(\langle K \rangle) \tilde{p}(\langle \mathcal{C}_A(K) \rangle) \\ &= \sum_{K \subset A} \sum_{K \subset G \subset A} (-1)^{|\mathcal{C}_G(K)|} p(\langle K \rangle) \tilde{p}(\langle \mathcal{C}_A(K) \rangle) \\ &= \sum_{K \subset A} p(\langle K \rangle) \tilde{p}(\langle \mathcal{C}_A(K) \rangle) \sum_{K \subset G \subset A} (-1)^{|\mathcal{C}_G(K)|} \\ &= p(\langle A \rangle). \end{aligned} \quad (3.19)$$

Hierbei benutzen wir die Definition von Δ (3.5a) und die Identität

$$\tilde{p}(\langle \mathcal{C}_A(G) \rangle) \tilde{p}(\langle \mathcal{C}_G(K) \rangle) = \tilde{p}(\langle \mathcal{C}_A(K) \rangle). \quad (3.20)$$

Daraufhin verwenden wir die Tatsache, dass wir die Summe $\sum_{H \subset L \subset S}$ auf zwei verschiedene Arten auseinander ziehen können. Im letzten Schritt streichen wir die Summe $\sum_{K \subset G \subset A} (-1)^{|\mathcal{C}_G(K)|}$. Berge [4, S. 83] zeigt in seinem Beweis zur Möbiusinversion, unter welchen Bedingungen diese Summe gestrichen werden kann.

Wir setzen nun Gleichung (3.19) in die diskrete Dynamik (2.23) ein, um

$$\begin{aligned} p_{t+1}(\langle R \rangle) &= \sum_{A \subset R} r_R(A) p_t(\langle A \rangle) p_t(\langle \mathcal{C}_R(A) \rangle) \\ &= \sum_{A \subset R} r_R(A) \left(\sum_{B \subset A} \tilde{p}(\langle \mathcal{C}_A(B) \rangle) \Delta_t(\langle B \rangle) \right) \\ &\quad \cdot \left(\sum_{C \subset \mathcal{C}_R(A)} \tilde{p}(\langle \mathcal{C}_{\mathcal{C}_R(A)}(C) \rangle) \Delta_t(\langle C \rangle) \right) \\ &= \sum_{A \subset R} r_R(A) \sum_{B \subset A} \sum_{C \subset \mathcal{C}_R(A)} \tilde{p}(\langle \mathcal{C}_A(B) \rangle) \Delta_t(\langle B \rangle) \\ &\quad \cdot \tilde{p}(\langle \mathcal{C}_{\mathcal{C}_R(A)}(C) \rangle) \Delta_t(\langle C \rangle) \\ &= \sum_{A \subset R} r_R(A) \sum_{H \subset R} \tilde{p}(\langle \mathcal{C}_A(H \cap A) \rangle) \Delta_t(\langle H \cap A \rangle) \\ &\quad \cdot \tilde{p}(\langle \mathcal{C}_{\mathcal{C}_R(A)}(H \setminus A) \rangle) \Delta_t(\langle H \setminus A \rangle) \\ &= \sum_{A \subset R} r_R(A) \sum_{H \subset R} \tilde{p}(\langle \mathcal{C}_R(H) \rangle) \Delta_t(\langle H \cap A \rangle) \Delta_t(\langle H \setminus A \rangle) \end{aligned} \quad (3.21)$$

zu erhalten. Nach dem Ausklammern fassen wir die Summen zusammen, um anschließend die Identität

$$\tilde{p}(\langle \mathcal{C}_A(H \cap A) \rangle) \tilde{p}(\langle \mathcal{C}_{\mathcal{C}_R(A)}(H \setminus A) \rangle) = \tilde{p}(\langle \mathcal{C}_R(H) \rangle) \quad (3.22)$$

zu verwenden. Aus Gründen der besseren Übersicht schreiben wir einen Punkt unter die Menge, über die summiert wird.

Diese Gleichung setzen wir in die Dynamik der zentralen Momente (3.5b) ein, um durch Ausmultiplizieren

$$\begin{aligned}
\Delta_{t+1}(\langle J \rangle) &= \sum_{R \subset J} (-1)^{|\mathcal{C}_J(R)|} \tilde{p}(\langle \mathcal{C}_J(R) \rangle) p_{t+1}(\langle R \rangle) \\
&= \sum_{R \subset J} (-1)^{|\mathcal{C}_J(R)|} \tilde{p}(\langle \mathcal{C}_J(R) \rangle) \sum_{A \subset R} r_R(A) \\
&\quad \cdot \sum_{H \subset R} \tilde{p}(\langle \mathcal{C}_R(H) \rangle) \Delta_t(\langle H \setminus A \rangle) \Delta_t(\langle H \cap A \rangle) \\
&= \sum_{R \subset J} \sum_{H \subset R} (-1)^{|\mathcal{C}_J(R)|} \tilde{p}(\langle \mathcal{C}_J(H) \rangle) \\
&\quad \cdot \sum_{A \subset R} r_R(A) \Delta_t(\langle H \setminus A \rangle) \Delta_t(\langle H \cap A \rangle) \tag{3.23} \\
&= \sum_{H \subset J} \sum_{H \subset R \subset J} (-1)^{|\mathcal{C}_J(R)|} \tilde{p}(\langle \mathcal{C}_J(H) \rangle) \\
&\quad \cdot \sum_{A \subset R} r_R(A) \Delta_t(\langle H \setminus A \rangle) \Delta_t(\langle H \cap A \rangle) \\
&= \sum_{A \subset J} r_J(A) \Delta_t(\mathcal{C}_J(A)) \Delta_t(\langle A \rangle)
\end{aligned}$$

zu erhalten. Wir verwenden dabei die Identität

$$\tilde{p}(\langle \mathcal{C}_J(R) \rangle) \tilde{p}(\langle \mathcal{C}_R(H) \rangle) = \tilde{p}(\langle \mathcal{C}_J(H) \rangle) \tag{3.24}$$

und vertauschen die Doppelsumme. Durch Streichen der Summe ergibt sich anschließend die behauptete diskrete Rekombinationsdynamik (3.18) der zentralen Momente. \square

4 Lösung der Rekombinations- und Selektionsdynamik

Wir möchten im Folgenden eine Lösung für die in Abschnitt 2.5 eingeführte Rekombinations- und Selektionsdynamik der relativen Gametenhäufigkeiten

$$\dot{p}_t = \left(F - \bar{f}_t \mathbb{1} + \sum_{\alpha \in L} \varrho_\alpha (R_\alpha - \mathbb{1}) \right) p_t \quad (4.1)$$

herleiten. Wie bereits in Abschnitt 2.5, bezeichnet

$$\bar{f}_t = \frac{\|Fz_t\|}{\|z_t\|} = \langle f, p_t \rangle \quad (4.2)$$

die mittlere Fitness der Population.

Zu diesem Zwecke leiten wir zunächst eine Lösung für die nicht normierte Form der Dynamik

$$\dot{z}_t = \left(F + \sum_{\alpha \in L} \varrho_\alpha (R_\alpha - \mathbb{1}) \right) z_t \quad (4.3)$$

her. Dies lässt sich unkomplizierter bewerkstelligen, da beim Betrachten von Reproduktion anstelle von Selektion die Nichtlinearität der mittleren Populationsfitness fehlt. Wir gehen zunächst getrennt auf die Lösung der Reproduktionsdynamik ein. Danach leiten wir die umfangreichere Lösung der Rekombinationsdynamik her. Wir versuchen daraufhin diese mit Ersterem nach einer Idee von Baake [1] zusammenzuführen, zeigen jedoch, dass dies nicht möglich ist. Wir sind also nicht in der Lage eine allgemeine Lösung der Rekombinations- und Reproduktionsdynamik anzugeben. Wir werden aber daraufhin Lösungen für Spezialfälle der gekoppelten Dynamik herleiten, welche wir in Kapitel 5 noch benötigen werden.

Durch Normieren der Lösung der Dynamik für die absoluten Gametenhäufigkeiten (4.3) erhalten wir eine Lösung für die Dynamik der relativen Gametenhäufigkeiten (4.1).

4.1 Lösung der Selektionsdynamik

Da das direkte Lösen der Selektionsdynamik (2.18) aufgrund der Nichtlinearität kompliziert ist, leiten wir indirekt aus der Lösung der Reproduktionsdynamik (2.19) die gewünschte Lösung her.

Wie bereits erwähnt, entstammt die Nichtlinearität in der Selektionsgleichung der Tatsache, dass wir die Population auf 1 normieren. Verzichten wir auf die Normierung der Differentialgleichung, betrachten wir also die Reproduktionsdynamik (2.19), so sind wir ohne weiteres in der Lage, eine Lösung anzugeben. Sie setzt sich zusammen aus der Halbgruppe, welche durch den als Generator aufgefassten Reproduktionsoperator F erzeugt wird

$$\Psi_t := e^{tF} \quad (4.4)$$

und der entsprechenden Anfangsbedingung z_0 . Durch Anwenden der Halbgruppe auf die Anfangsbedingung

$$z_t := \Psi_t(z_0) = e^{tF} z_0 \quad (4.5)$$

erhalten wir die Lösung. Die Berechnung von e^{tF} ist kein Problem, da es sich bei F um eine Diagonalmatrix handelt.

Indem wir nun diese Lösung der unnormierten Gleichung nachträglich normieren

$$p_t := \frac{z_t}{\|z_t\|} = \frac{\Psi_t(z_0)}{\|\Psi_t(z_0)\|}, \quad (4.6)$$

erhalten wir auf einfachem Wege die Lösung der Selektionsdynamik (2.18). Dieses nachträgliche Normieren der Lösung ist in der Literatur auch als *Thompson-Trick* [20] bekannt. Es ermöglicht uns, die Nichtlinearität in Gleichung (4.1) beim Lösen der Dynamik zu umgehen.

4.2 Mathematische Modellierung der Rekombination

Um die in Kapitel 2.4 angedeuteten Rekombinatoren adäquat modellieren zu können, benötigen wir die Definition der *Projektoren* $\pi_{<\alpha}$ und $\pi_{>\alpha}$. Mit den Projektoren sind wir dann in der Lage die Rekombinatoren zu definieren.

4.2.1 Projektoren

Die Projektoren entsprechen der Projektion eines Gameten auf die Allele an den Sites vor beziehungsweise hinter dem Link α , also einem Schneiden des Gameten an α . Die Definition für $\pi_{<\alpha}$ lautet

$$\pi_{<\alpha} : X \rightarrow X_{<\alpha} \quad (4.7)$$

$$(x_0, \dots, x_n) \mapsto (x_0, \dots, x_{[\alpha]}), \quad (4.8)$$

wobei $X_{<\alpha} = X_0 \times \dots \times X_{[\alpha]}$ gilt.

Da diese Projektion nicht injektiv ist, definieren wir die Umkehrabbildung $\pi_{<\alpha}^{-1}$ als die Menge aller Urbilder, die unter der Abbildung einen gegebenen Wert haben. Für $x_{<\alpha} \in X_{<\alpha}$ ist

$$\pi_{<\alpha}^{-1}(x_{<\alpha}) := \{x \in X \mid \pi_{<\alpha}(x) = x_{<\alpha}\}. \quad (4.9)$$

Die Projektion induziert auf der Menge der endlichen positiven Maße über X , \mathcal{M}^\otimes den sogenannten *pullback* von $\mathcal{M}^\otimes(X)$ nach $\mathcal{M}^\otimes(X_{<\alpha})$. Dieser ist definiert als

$$\pi_{<\alpha}.z := z \circ \pi_{<\alpha}^{-1} \quad (4.10)$$

für alle $z \in \mathcal{M}^\otimes$. Gleichzeitig entspricht dies der Marginalisierung einer Population von Gameten auf die Sites vor dem Link α . $\pi_{<\alpha}.z(x_{<\alpha})$ ist also die Häufigkeit der Gameten in der Population, welche vor Link α die Allele $x_{<\alpha}$ haben.

Die Definitionen für $\pi_{>\alpha}$ laufen analog.

4.2.2 Rekombinatoren

In Kapitel 2 wurde der Rekombinationsprozess als eine zufällige Paarbildung mit anschließendem Austausch von Erbinformation vor beziehungsweise hinter einer gegebenen Position eingeführt. Dieser Vorgang soll nun durch die sogenannten Rekombinatoren mathematisch modelliert werden.

Die Häufigkeit eines Gameten nach der Rekombination ergibt sich als die marginale Häufigkeit des Teilgameten vor Link α multipliziert mit der Marginalwahrscheinlichkeit, also der relativen Häufigkeit des Teilgameten hinter Link α . Mit Hilfe der Projektoren definieren wir den Rekombinator auf $z \in \mathcal{M}^\otimes$ als

$$R_\alpha(z) := \frac{1}{\|z\|} (\pi_{<\alpha}.z) \otimes (\pi_{>\alpha}.z), \quad (4.11)$$

wobei wir hier mit \otimes das Tensorprodukt bezeichnen.

Mit dieser Definition gilt komponentenweise für $x = (x_{<\alpha}, x_{>\alpha}) \in X$

$$R_\alpha(z)(x) = \frac{1}{\|z\|} \cdot (\pi_{<\alpha}.z)(x_{<\alpha}) \cdot (\pi_{>\alpha}.z)(x_{>\alpha}) \quad (4.12)$$

$$= z(x_0, \dots, x_{[\alpha]}, *, \dots, *) \cdot p(*, \dots, *, x_{[\alpha]}, \dots, x_n), \quad (4.13)$$

wobei $x_{<\alpha} = (x_0, \dots, x_{[\alpha]})$ und $x_{>\alpha} = (x_{[\alpha]}, \dots, x_n)$ ist. Der Rekombinator führt also die gewünschte Operation durch, nämlich das Rekombinieren der gesamten Population an Link α .

4.2.3 Eigenschaften der Rekombinatoren

Die Rekombinatoren besitzen diverse Eigenschaften, welche für die Lösung der Dynamik benötigt werden. Die Beweise für die Eigenschaften finden sich bei Baake [1]. An dieser Stelle wollen wir nochmals hervorheben, dass wir uns in den folgenden Untersuchungen auf positive Maße $z \in \mathcal{M}^\otimes = \mathcal{M}_+(X)$ beschränken. Während die vorangegangenen Definitionen noch für beliebige Maße gelten, so lassen sich die folgenden Eigenschaften nur auf positiven beweisen.

Proposition 2 in [1] besagt, dass die Rekombinatoren idempotent sind und miteinander kommutieren. Das heißt, für $z \in \mathcal{M}^\otimes$ und $\alpha, \beta \in L$ gilt

$$R_\alpha(R_\alpha(z)) = R_\alpha(z) \quad (4.14)$$

und

$$R_\alpha(R_\beta(z)) = R_\beta(R_\alpha(z)). \quad (4.15)$$

Da die Reihenfolge der Ausführung irrelevant ist, erlaubt uns Eigenschaft (4.15) nun,

$$R_G := \bigcirc_{\alpha \in G} R_\alpha \quad (4.16)$$

als Verkettung der Rekombinatoren zu definieren.

Des Weiteren kann aus Proposition 4 in [1] abgeleitet werden, dass sich die Rekombinatoren linear auf bestimmten Konvex-Kombinationen verhalten. Genauer gilt für $\lambda \in [0, 1]$ und $z \in \mathcal{M}^\otimes$

$$R_\alpha(\lambda z + R_\alpha((1 - \lambda)z)) = R_\alpha(z). \quad (4.17)$$

Ferner zeigt Baake mit Fakt 2 in [1], dass die Rekombinatoren die Norm des Maßes erhalten. Es gilt also

$$\|R_\alpha(z)\| = \|z\| \quad (4.18)$$

für alle $\alpha \in L$ und $z \in \mathcal{M}^\otimes$.

4.3 Lösung der Rekombinationsdynamik

Wir wollen nun mit Hilfe dieser Eigenschaften die Rekombinationsdynamik (2.29) lösen.

Zunächst beschränken wir uns auf die sogenannte *Ein-Punkt-Rekombinationsgleichung*, welche die Dynamik der Population beschreibt, wenn die Rekombination nur an einem Link α wirkt. Sie lautet

$$\dot{z}_t = \varrho_\alpha (R_\alpha - \mathbb{1}) z_t. \quad (4.19)$$

Lemma 4.1 *Die Ein-Punkt-Rekombinationsgleichung wird durch die Halbgruppe Φ_t^α mit*

$$\Phi_t^\alpha := e^{-\varrho_\alpha t} \mathbb{1} + (1 - e^{-\varrho_\alpha t}) R_\alpha \quad (4.20)$$

angewendet auf die Anfangsbedingung z_0 gelöst.

Beweis: Wir verifizieren die Lösung $z_t = \Phi_t^\alpha(z_0)$, indem wir zeigen, dass sie der Differentialgleichung (4.19) genügt. Differenzieren ergibt

$$\begin{aligned} \frac{d}{dt} \Phi_t^\alpha(z_0) &= \frac{d}{dt} (e^{-\varrho_\alpha t} \mathbb{1} + (1 - e^{-\varrho_\alpha t}) R_\alpha)(z_0) \\ &= -\varrho_\alpha e^{-\varrho_\alpha t} z_0 + \varrho_\alpha e^{-\varrho_\alpha t} R_\alpha(z_0) \\ &= \varrho_\alpha (e^{-\varrho_\alpha t} R_\alpha(z_0) - e^{-\varrho_\alpha t} z_0) \\ &= \varrho_\alpha (R_\alpha(z_0) - (e^{-\varrho_\alpha t} z_0 + (1 - e^{-\varrho_\alpha t}) R_\alpha(z_0))) \\ &= \varrho_\alpha (R_\alpha(e^{-\varrho_\alpha t} z_0 + (1 - e^{-\varrho_\alpha t}) R_\alpha(z_0)) - (e^{-\varrho_\alpha t} z_0 + (1 - e^{-\varrho_\alpha t}) R_\alpha(z_0))) \\ &= \varrho_\alpha (R_\alpha - \mathbb{1})(e^{-\varrho_\alpha t} \mathbb{1} + (1 - e^{-\varrho_\alpha t}) R_\alpha)(z_0) \\ &= \varrho_\alpha (R_\alpha - \mathbb{1}) \Phi_t^\alpha(z_0), \end{aligned} \quad (4.21)$$

wobei wir im vierten Schritt (4.17) verwendet haben. Aufgrund der Halbgruppeneigenschaft $\Phi_0^\alpha = \mathbb{1}$ ist die Anfangsbedingung erfüllt. \square

Wir zeigen nun, dass die Halbgruppen Φ_t^α ebenso wie die Rekombinatoren miteinander kommutieren, da wir diese Eigenschaft zur Lösung der kompletten Rekombinationsdynamik benötigen.

Satz 4.1 *Sei Φ_t^α wie in (4.20) definiert, dann gilt*

$$\Phi_t^\alpha \circ \Phi_s^\beta = \Phi_s^\beta \circ \Phi_t^\alpha \quad (4.22)$$

für alle $\alpha, \beta \in L$ und $t, s \in \mathbb{R}_{\geq 0}$.

Der Beweis für das Kommutieren der Halbgruppen baut auf dem Kommutieren einer Halbgruppe mit einem Rekombinator auf. Daher wollen wir zunächst das nachfolgende Lemma beweisen.

Lemma 4.2 *Die Rekombinatoren kommutieren mit den Rekombinationshalbgruppen. Es gilt*

$$R_\alpha \circ \Phi_s^\beta = \Phi_s^\beta \circ R_\alpha \quad (4.23)$$

für $\alpha, \beta \in L$ und $s \in \mathbb{R}_{\geq 0}$.

Beweis: Wir beschränken uns an dieser Stelle auf den Beweis des Falles $\alpha < \beta$; der Fall $\alpha = \beta$ folgt unmittelbar aus (4.17), und der Fall $\alpha > \beta$ ist analog zu $\alpha < \beta$.

Aus der Definition des Rekombinators (4.11) und der Halbgruppe (4.20) ergibt sich

$$\begin{aligned} & R_\alpha \circ \Phi_s^\beta(z) \\ &= \frac{1}{\|\Phi_s^\beta(z)\|} \left(e^{-\varrho\beta s} \pi_{<\alpha} \cdot z + (1 - e^{-\varrho\beta s}) \pi_{<\alpha} \cdot R_\beta(z) \right) \\ & \quad \otimes \left(e^{-\varrho\beta s} \pi_{>\alpha} \cdot z + (1 - e^{-\varrho\beta s}) \pi_{>\alpha} \cdot R_\beta(z) \right) \\ &= e^{-2\varrho\beta s} R_\alpha(z) + (1 - e^{-\varrho\beta s})^2 R_{\{\alpha, \beta\}}(z) \\ & \quad + \frac{1}{\|z\|} e^{-\varrho\beta s} (1 - e^{-\varrho\beta s}) \left((\pi_{<\alpha} \cdot R_\beta(z)) \otimes (\pi_{>\alpha} \cdot z) + (\pi_{<\alpha} \cdot z) \otimes (\pi_{>\alpha} \cdot R_\beta(z)) \right) \\ &= e^{-2\varrho\beta s} R_\alpha(z) + (1 - e^{-\varrho\beta s})^2 R_{\{\alpha, \beta\}}(z) \\ & \quad + \frac{1}{\|z\|} e^{-\varrho\beta s} (1 - e^{-\varrho\beta s}) \left((\pi_{<\alpha} \cdot z) \otimes (\pi_{>\alpha} \cdot z) + (\pi_{<\alpha} \cdot R_\beta(z)) \otimes (\pi_{>\alpha} \cdot R_\beta(z)) \right) \\ &= e^{-2\varrho\beta s} R_\alpha(z) + (1 - 2e^{-\varrho\beta s} + e^{-2\varrho\beta s}) R_{\{\alpha, \beta\}}(z) \\ & \quad + (e^{-\varrho\beta s} - e^{-2\varrho\beta s}) (R_\alpha(z) + R_{\{\alpha, \beta\}}(z)) \\ &= e^{-\varrho\beta} R_\alpha(z) + (1 - e^{-\varrho\beta}) R_{\{\alpha, \beta\}}(z) \\ &= \Phi_s^\beta \circ R_\alpha(z) \end{aligned} \quad (4.24)$$

für alle $z \in \mathcal{M}^\otimes$. Dabei verwenden wir neben elementaren Umformungen die Normerhaltung der Rekombinatoren sowie die Tatsache, dass die Projektion eines Gameten auf

den Teil vor Link α unabhängig davon ist, ob hinter Link α rekombiniert wird oder nicht, also $\pi_{<\alpha}.R_\beta(z) = \pi_{<\alpha}.z$ für $\alpha < \beta$. \square

Beweis von Satz 4.1: Wir erhalten

$$\begin{aligned}
\Phi_t^\alpha \circ \Phi_s^\beta(z) &= e^{-\varrho_\alpha t} (\Phi_s^\beta(z)) + (1 - e^{-\varrho_\alpha t}) R_\alpha (\Phi_s^\beta(z)) \\
&= e^{-\varrho_\alpha t} (\Phi_s^\beta(z)) + (1 - e^{-\varrho_\alpha t}) \Phi_s^\beta(R_\alpha(z)) \\
&= e^{-\varrho_\alpha t} (e^{-\varrho_\beta s} z + (1 - e^{-\varrho_\beta s}) R_\beta(z)) \\
&\quad + (1 - e^{-\varrho_\alpha t}) (e^{-\varrho_\beta s} R_\alpha(z) + (1 - e^{-\varrho_\beta s}) R_\beta(R_\alpha(z))) \\
&= e^{-\varrho_\alpha t} e^{-\varrho_\beta s} z + (1 - e^{-\varrho_\beta s}) e^{-\varrho_\alpha t} R_\beta(z) \\
&\quad + (1 - e^{-\varrho_\alpha t}) e^{-\varrho_\beta s} R_\alpha(z) + (1 - e^{-\varrho_\alpha t}) (1 - e^{-\varrho_\beta s}) R_{\{\beta,\alpha\}}(z)
\end{aligned} \tag{4.25}$$

für alle $z \in \mathcal{M}^\otimes$, wobei wir die Definition der Halbgruppen und Lemma 4.2 anwenden, um den Ausdruck vereinfachen zu können.

Die letzte Zeile ist symmetrisch in (α, t) und (β, s) , das heißt, wir können die beiden Paare jeweils durch das andere ersetzen und erhalten denselben Ausdruck. Diese Substitution können wir also auch im ersten Ausdruck vornehmen und erhalten so Behauptung (4.22). \square

Wir müssen zum Lösen der Rekombinationsdifferentialgleichung die zeitliche Ableitung einer Kombination von Halbgruppen bilden. Daher leiten wir nun eine Art *Produktregel* für kommutierende Halbgruppen her.

Satz 4.2 Seien Υ_t^i für $1 \leq i \leq n$ Halbgruppen und es gelte $\Upsilon_t^i \circ \Upsilon_s^j = \Upsilon_s^j \circ \Upsilon_t^i$ für alle $i, j \in \{1, \dots, n\}$ und $t, s \in \mathbb{R}_{\geq 0}$, das heißt die Halbgruppen kommutieren zu verschiedenen Zeiten. Dann gilt

$$\frac{d}{dt} \bigcirc_{i=1}^n \Upsilon_t^i = \sum_{i=1}^n \left(\frac{d}{dt} \Upsilon_t^i \right) \circ \bigcirc_{\substack{j=1 \\ j \neq i}}^n \Upsilon_t^j. \tag{4.26}$$

Die verkettete Ausführung der kommutierenden Halbgruppen wird also gemäß einer Produktregel differenziert.

Beweis: Wir wollen die Formel mittels vollständiger Induktion beweisen. Dazu zeigen wir zum *Induktionsanfang*, dass zwei kommutierende Halbgruppen nach diesem Schema abgeleitet werden können. Wir erhalten nach der Definition der Ableitung und Einsetzen

einer geschickten Null

$$\begin{aligned}
\frac{d}{dt} \Upsilon_t^i \circ \Upsilon_t^j &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} (\Upsilon_{t+\epsilon}^i \circ \Upsilon_{t+\epsilon}^j - \Upsilon_t^i \circ \Upsilon_t^j) \\
&= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} (\Upsilon_{t+\epsilon}^i \circ \Upsilon_{t+\epsilon}^j + \Upsilon_t^i \circ \Upsilon_{t+\epsilon}^j - \Upsilon_t^i \circ \Upsilon_{t+\epsilon}^j - \Upsilon_t^i \circ \Upsilon_t^j) \\
&= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} (\Upsilon_{t+\epsilon}^i \circ \Upsilon_{t+\epsilon}^j + \Upsilon_{t+\epsilon}^j \circ \Upsilon_t^i - \Upsilon_t^i \circ \Upsilon_{t+\epsilon}^j - \Upsilon_t^j \circ \Upsilon_t^i) \\
&= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \left((\Upsilon_{t+\epsilon}^i - \Upsilon_t^i) \circ \Upsilon_{t+\epsilon}^j + (\Upsilon_{t+\epsilon}^j - \Upsilon_t^j) \circ \Upsilon_t^i \right) \\
&= \left(\frac{d}{dt} \Upsilon_t^i \right) \circ \Upsilon_t^j + \left(\frac{d}{dt} \Upsilon_t^j \right) \circ \Upsilon_t^i.
\end{aligned} \tag{4.27}$$

Hierbei benutzen wir im dritten Schritt, dass die Operatoren zu unterschiedlichen Zeiten kommutieren. Anschließend wird die triviale Rechts-Distributivität sowie die rechtssteigige Differenzierbarkeit verwendet.

Die Idee dieser Rechnung lässt sich für den *Induktionsschritt* erweitern, so dass wir

$$\begin{aligned}
\frac{d}{dt} \bigcirc_{i=1}^n \Upsilon_t^i &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \left(\Upsilon_{t+\epsilon}^1 \circ \bigcirc_{i=2}^n \Upsilon_{t+\epsilon}^i - \Upsilon_t^1 \circ \bigcirc_{i=2}^n \Upsilon_t^i \right) \\
&= \left(\frac{d}{dt} \Upsilon_t^1 \right) \circ \bigcirc_{i=2}^n \Upsilon_t^i + \left(\frac{d}{dt} \bigcirc_{i=2}^n \Upsilon_t^i \right) \circ \Upsilon_t^1 \\
&= \sum_{i=1}^n \left(\frac{d}{dt} \Upsilon_t^i \right) \circ \bigcirc_{\substack{j=1 \\ j \neq i}}^n \Upsilon_t^j
\end{aligned} \tag{4.28}$$

erhalten. Wir verwenden lediglich die Eigenschaft, dass aus der Verkettung von Halbgruppen wieder eine Halbgruppe entsteht. \square

Mit Hilfe von Satz 4.1 und Satz 4.2 sind wir nun in der Lage, die Rekombinationsgleichung zu lösen.

Satz 4.3 *Die Lösung der Rekombinationsdynamik (2.29) lautet*

$$z_t = \Phi_t(z_0) := \bigcirc_{\alpha \in L} \Phi_t^\alpha(z_0) \tag{4.29}$$

und ist somit die Verkettung der Lösungen der beteiligten Ein-Punkt-Gleichungen.

Beweis: Wir verifizieren die Lösung, indem wir zeigen, dass sie der Differentialgleichung

genügt. Wir erhalten

$$\begin{aligned}
\frac{d}{dt}\Phi_t(z_0) &= \frac{d}{dt} \bigcirc_{\alpha \in L} \Phi_t^\alpha(z_0) \\
&= \sum_{\alpha \in L} \left(\frac{d}{dt} \Phi_t^\alpha \right) \circ \bigcirc_{\beta \in L \setminus \{\alpha\}} \Phi_t^\beta(z_0) \\
&= \sum_{\alpha \in L} \varrho_\alpha (R_\alpha - \mathbb{1}) \Phi_t^\alpha \circ \bigcirc_{\beta \in L \setminus \{\alpha\}} \Phi_t^\beta(z_0) \\
&= \sum_{\alpha \in L} \varrho_\alpha (R_\alpha - \mathbb{1}) \circ \bigcirc_{\beta \in L} \Phi_t^\beta(z_0) \\
&= \sum_{\alpha \in L} \varrho_\alpha (R_\alpha - \mathbb{1}) \circ \Phi_t(z_0),
\end{aligned} \tag{4.30}$$

indem wir auf die kommutierenden Halbgruppen die *Produktregel* 4.2 anwenden und danach die Ableitung der Ein-Punkt-Gleichungen einsetzen. Da, wie schon erwähnt, die Verkettung von Halbgruppen wieder eine Halbgruppe ist, ist die Anfangsbedingung $\Phi_0(z_0) = \mathbb{1} z_0 = z_0$ trivialerweise erfüllt. \square

Durch Ausmultiplizieren lässt sich die Lösung in die Form

$$z_t = \bigcirc_{\alpha \in L} \Phi_t^\alpha(z_0) = \sum_{G \subset L} a_G(t) R_G(z_0) \tag{4.31}$$

bringen, wobei für die Koeffizienten

$$a_G(t) := \prod_{\alpha \in \bar{G}} e^{-\varrho_\alpha t} \prod_{\beta \in G} (1 - e^{-\varrho_\beta t}) = \sum_{H \subset G} (-1)^{|G-H|} e^{-\sum_{\alpha \in H} \varrho_\alpha t} \tag{4.32}$$

gilt. Sie können interpretiert werden als die Wahrscheinlichkeit, dass zum Zeitpunkt t an den Links in G mindestens ein und an den restlichen Links noch kein Rekombinationsereignis stattgefunden hat. Mit $\bar{G} = \mathcal{C}_L(G)$ bezeichnen wir an dieser Stelle das Komplement von G in L .

Die Lösung gilt insbesondere für Anfangswerte $p_0 \in \mathcal{P}^\otimes$. Da die Rekombinatoren normerhaltend sind, führt die Halbgruppe nicht aus diesem Raum heraus, so dass

$$p_t = \Phi_t(p_0) = \sum_{G \subset L} a_G(t) R_G(p_0) \tag{4.33}$$

die normerhaltende Rekombinationsdynamik (2.29) löst.

An dieser Form der Lösung lässt sich außerdem erkennen, dass sie eine Linearkombination aller Maße darstellt, welche sich aus den möglichen Rekombinatoren angewendet auf p_0 ergeben. Die Koeffizienten der Linearkombination enthalten Exponentialfunktionen, welche unterschiedliche Zerfallsraten haben. Baake gruppiert in [1] die unterschiedlichen Summanden der Linearkombination entsprechend ihrer Koeffizienten und erhält so eine Transformation der Gametenhäufigkeiten

$$T_G := \sum_{H \supset G} (-1)^{|H-G|} R_H, \tag{4.34}$$

welche die Rekombinationsdynamik (2.29) zu exponentiell zerfallenden Größen entkoppelt.

4.4 Zusammenführen der Lösungen

Nun wollen wir untersuchen, inwiefern sich die Lösung der Selektionsdynamik (2.19) mit der Lösung der Rekombinationsdynamik (2.29) verbinden lässt, um eine Lösung der gekoppelten Dynamik (4.3) zu erhalten.

Für allgemeine Reproduktionsoperatoren F ist dies nicht möglich. Baake [1] behauptet jedoch, dass sich die Lösungen verbinden lassen, wenn wir uns auf eine biologisch relevante Teilmenge der allgemeinen Reproduktionsoperatoren beschränken. Diese biologisch relevante Teilmenge besteht aus Reproduktionsoperatoren, welche das Prinzip der sogenannten *additiven Fitness* beschreiben.

Additive Fitness bedeutet an dieser Stelle, dass sich der Wert der Malthus'schen Fitness eines Gameten aus jeweils voneinander unabhängigen Fitnesswerten für die einzelnen Allele additiv zusammensetzt. Mathematisch beschreiben lässt sich dies, indem wir zunächst für jeden Site einen Operator

$$B_i = \mathbb{1} \otimes \dots \otimes \mathbb{1} \otimes b_i \otimes \mathbb{1} \otimes \dots \otimes \mathbb{1} \quad (4.35)$$

definieren, welcher ausschließlich den Allelen an Site i mittels des Operators b_i Werte zuweist und alle anderen neutral lässt. Diese Operatoren summieren wir dann auf, um

$$F = \sum_{i=0}^n B_i \quad (4.36)$$

als Reproduktionsoperator zu erhalten, welcher additive Fitness beschreibt.

In Lemma 5 behauptet Baake [1], dass die Rekombinatoren mit streng positiven beschränkten linearen Operatoren auf \mathcal{M}^{\otimes} kommutieren, welche eine komplette Tensorproduktstruktur besitzen. Diese Struktur besitzt ein Operator O , wenn er sich als $O = o_0 \otimes \dots \otimes o_n$ darstellen lässt, wenn sich also die Wirkung auf einen Gameten multiplikativ aus Werten zusammensetzt, welche für jeden einzelnen Site spezifisch und unabhängig von den anderen Sites sind. Die eben beschriebenen Operatoren B_i besitzen eine komplette Tensorproduktstruktur und fallen daher in diese Kategorie.

Beschränken wir uns also auf Reproduktionsoperatoren, die additive Fitness modellieren, so ließe sich aus diesem Lemma insbesondere folgende Behauptung ableiten:

Behauptung 4.1 *Die Rekombinatoren kommutieren mit der Reproduktionshalbgruppe. Es gilt*

$$R_\alpha \circ \Psi_s = \Psi_s \circ R_\alpha. \quad (4.37)$$

Dies würde bedeuten, dass die Reproduktion unabhängig von der Rekombination wirkt. Baake zeigt basierend auf diesem Lemma, dass die Verkettung der zugehörigen Halbgruppen

$$z_t = \Psi_t \circ \bigcirc_{\alpha \in L} \Phi_t^\alpha(z_0) = e^{tF} \sum_{GCL} a_G(t) R_G(z_0) = \sum_{GCL} a_G(t) R_G(e^{tF} z_0) \quad (4.38)$$

das Anfangswertproblem (4.3) löst.

Unter der Annahme, dass der Rekombinator mit der Reproduktionshalbgruppe kommutiert, lässt sich zeigen, dass die Halbgruppen ebenfalls vertauschen. Mit Hilfe dieser Eigenschaft ließe sich wieder die Produktregel 4.2 für kommutierende Halbgruppen anwenden, um zu zeigen, dass (4.38) der Differentialgleichung (4.3) genügt.

Es stellte sich jedoch im Laufe dieser Diplomarbeit heraus, dass Behauptung 4.1 in dieser Form falsch ist, was wir im Folgenden näher erläutern wollen.

4.4.1 Ein Gegenbeispiel

Behauptung 4.1 und damit auch Lemma 5 aus [1], auf welchem Baakes und der hier skizzierte Beweis beruhen, lässt sich anhand des folgenden Gegenbeispiels widerlegen:

Beispiel: Wir untersuchen nun ein 2-Site-/2-Allel-System. Die Allelmengen an den zwei Sites $S = \{0, 1\}$ sind $X_0 = \{0, 1\}$ und $\{0, 1\}$. Der Vektor der Gametenhäufigkeiten ergibt sich zu $z_t = (z_t(0, 0), z_t(0, 1), z_t(1, 0), z_t(1, 1))^T$. In diesem Szenario gibt es nur einen Link $L = \{\frac{1}{2}\}$, weshalb wir den einzigen Rekombinator mit $R := R_{\frac{1}{2}}$ bezeichnen.

Die Allele am ersten Site tragen nichts zur Malthus'schen Fitness des Gameten bei, sie sind also neutral. Am zweiten Site hingegen besitzt Allel 0 eine Malthus'sche Fitness von 0 und Allel 1 eine Fitness von λ . Damit ist Allel 1 gegenüber Allel 0 im Vorteil. Es ergibt sich für den Reproduktionsoperator

$$F = \mathbf{1} \otimes \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda \end{pmatrix}. \quad (4.39)$$

Da nur ein Site zur Fitness beiträgt, passt dieser Operator überdies in das Schema der additiven Fitness. Es folgt für die Reproduktionshalbgruppe

$$\Psi_s = e^{sF} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & e^{s\lambda} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{s\lambda} \end{pmatrix}. \quad (4.40)$$

Nun rechnen wir explizit die Verkettung der beiden Operatoren angewendet auf $z = (0.5, 0, 0, 0.5)^T$ aus, wobei wir

a) einmal zunächst den Rekombinator anwenden

$$\Psi_s(R(z)) = e^{sF} R \begin{pmatrix} 0.5 \\ 0 \\ 0 \\ 0.5 \end{pmatrix} = e^{sF} \begin{pmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{pmatrix} = \begin{pmatrix} 0.25 \\ 0.25e^{s\lambda} \\ 0.25 \\ 0.25e^{s\lambda} \end{pmatrix} \quad (4.41)$$

b) und beim anderen Mal mit der Reproduktionshalbgruppe beginnen

$$R(\Psi_s(z)) = R \begin{pmatrix} 0.5 \\ e^{sF} 0 \\ 0 \\ 0.5 \end{pmatrix} = R \begin{pmatrix} 0.5 \\ 0 \\ 0 \\ 0.5e^{s\lambda} \end{pmatrix} = \frac{1}{\|\Psi_s(z)\|} \begin{pmatrix} 0.25 \\ 0.25e^{s\lambda} \\ 0.25e^{s\lambda} \\ 0.25e^{2s\lambda} \end{pmatrix}. \quad (4.42)$$

Der Vergleich von (4.41) und (4.42) zeigt, dass diese für allgemeine λ und s nicht identisch sind. Die beiden Operatoren kommutieren also nicht.

Aus diesem Beispiel sehen wir, dass die Rekombinatoren im Allgemeinen nicht mit der Reproduktionshalbgruppe vertauschen. Das bedeutet, dass es Reproduktionshalbgruppen Ψ_s gibt, für die

$$R_\alpha \circ \Psi_s \neq \Psi_s \circ R_\alpha \quad (4.43)$$

gilt.

In dem Beweis zu Lemma 5 in [1] gibt es ein Problem bei der Konstruktion der Normhomogenität. Es wird implizit verwendet, dass $\|Fz\| = \|FR_\alpha(z)\|$ gilt, was jedoch im Allgemeinen nicht der Fall ist.

4.5 Lösung einfacher Spezialfälle

Wir haben gezeigt, dass die allgemeine Lösung der Rekombinations- und Selektionsdynamik, welche Baake in [1] herleitete, falsch ist. Es scheint, als ließe sich die Lösung auch nicht auf einfachem Wege retten.

Dennoch ist in einigen Spezialfällen, welche für diese Arbeit relevant sind, eine explizite Lösung möglich. Im Folgenden wollen wir diese Lösungen vorstellen.

4.5.1 Ein neutraler und ein selektiver Site

Das bereits in Abschnitt 4.4.1 geschilderte Szenario eines neutralen und eines selektiven Sites mit jeweils zwei Allelen lässt sich analytisch lösen.

Die Malthus'sche Fitness der beiden Allele am selektiven Site ist, wie im Beispiel aus Abschnitt 4.4.1, 0 für Allel 0 und λ für Allel 1, die Fitness ist also erneut additiv. Da wir die Selektion ebenfalls am zweiten Site wirken lassen, ergeben sich die gleichen Operatoren R und F und der gleiche Raum möglicher Gameten $X = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Analog zum Rekombinator bezeichnen wir die einzige Rekombinationsrate mit $\varrho := \varrho_{\frac{1}{2}}$.

Bevor wir die Reproduktions- und Rekombinationsdynamik (4.3) für diesen Fall konkret aufschreiben, möchten wir zunächst ein bestimmtes Kopplungsungleichgewicht näher untersuchen.

Nach der allgemeinen Definition (4.34)

$$T_G = \sum_{H \supset G} (-1)^{|H-G|} R_H \quad (4.44)$$

ergeben sich in diesem Fall zwei verschiedene Kopplungsungleichgewichtsoperatoren, und zwar

$$T_{\emptyset} = \mathbb{1} - R \quad \text{und} \quad T_{\{\frac{1}{2}\}} = R. \quad (4.45)$$

Die allgemeine Dynamik (4.3) lässt sich also für den 2-Site-/2-Allel-Fall spezifizieren und mit Hilfe von T_{\emptyset} zu

$$\begin{aligned} \dot{z}_t &= Fz_t + \sum_{\alpha \in L} \varrho_{\alpha} (R_{\alpha} - \mathbb{1}) z_t \\ &= Fz_t + \varrho (R - \mathbb{1}) z_t \\ &= Fz_t - \varrho T_{\emptyset} (z_t) \end{aligned} \quad (4.46)$$

umschreiben. Im Folgenden bezeichnen wir $z_t(x_0, x_1)$ zur besseren Übersicht mit $(x_0, x_1)_t$.

Die Dynamik (4.46) lässt sich weiter vereinfachen, indem wir zeigen, dass die Einträge des Vektors $T_{\emptyset}(z_t)$ bis auf ihr Vorzeichen identisch sind. Über die Beziehung

$$\|z_t\| = (0, 0)_t + (0, 1)_t + (1, 0)_t + (1, 1)_t \quad (4.47)$$

lässt sich $T_{\emptyset}(z_t)(0, 0)$ zu

$$\begin{aligned} T_{\emptyset}(z_t)(0, 0) &= (\mathbb{1} - R)(z_t)(0, 0) \\ &= (0, 0)_t - \frac{1}{\|z_t\|} (0, *)_t (*, 0)_t \\ &= (0, 0)_t - \frac{1}{\|z_t\|} ((0, 0)_t + (0, 1)_t)((0, 0)_t + (1, 0)_t) \\ &= (0, 0)_t - \frac{1}{\|z_t\|} ((0, 0)_t^2 + (0, 0)_t(1, 0)_t + (0, 1)_t(0, 0)_t + (0, 1)_t(1, 0)_t) \\ &= (0, 0)_t - \frac{1}{\|z_t\|} ((0, 0)_t \|z_t\| - (0, 0)_t(1, 1)_t + (0, 1)_t(1, 0)_t) \\ &= \frac{(0, 0)_t(1, 1)_t - (0, 1)_t(1, 0)_t}{\|z_t\|} \end{aligned} \quad (4.48)$$

umschreiben. Definieren wir nun

$$h_t := \frac{(0, 1)_t(1, 0)_t - (0, 0)_t(1, 1)_t}{\|z_t\|}, \quad (4.49)$$

so können wir mit Hilfe von (4.48) und analogen Rechnungen die Beziehungen

$$T_{\emptyset}(z_t)(0, 0) = -h_t = T_{\emptyset}(z_t)(1, 1) \quad (4.50)$$

und

$$T_{\emptyset}(z_t)(0, 1) = h_t = T_{\emptyset}(z_t)(1, 0) \quad (4.51)$$

zeigen, so dass sich die Dynamik (4.46) zu

$$\begin{aligned}\dot{z}_t &= Fz_t - \varrho T_{\emptyset}(z_t) \\ &= Fz_t + \varrho h_t \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}\end{aligned}\quad (4.52)$$

umschreiben lässt.

Einsetzen von F ergibt nun die folgenden vier nichtlinearen gekoppelten Differentialgleichungen:

$$\begin{pmatrix} \dot{0,0} \\ \dot{0,1} \\ \dot{1,0} \\ \dot{1,1} \end{pmatrix}_t = \begin{pmatrix} 0 \\ \lambda(0,1)_t \\ 0 \\ \lambda(1,1)_t \end{pmatrix} + \varrho h_t \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}.\quad (4.53)$$

Die Dynamik der Marginalhäufigkeiten am selektierten Site lässt sich daraus direkt herleiten. Es ergeben sich die Differentialgleichungen

$$\frac{d}{dt}(*,0)_t = 0 \quad \text{und} \quad \frac{d}{dt}(*,1)_t = \lambda(*,1)_t\quad (4.54)$$

mit den Lösungen

$$(*,0)_t = (*,0)_0 \quad \text{und} \quad (*,1)_t = (*,1)_0 e^{\lambda t}.\quad (4.55)$$

Für die Gesamtgröße der Population folgt

$$\begin{aligned}\|z_t\| &= (*,*)_t = (*,0)_t + (*,1)_t \\ &= (*,0)_0 + (*,1)_0 e^{\lambda t}.\end{aligned}\quad (4.56)$$

Setzen wir h_t als bekannt voraus, so können die Differentialgleichungen (4.53) zu

$$\begin{aligned}(0,0)_t &= (0,0)_0 + \varrho \int_0^t h_{\tau} d\tau \\ (0,1)_t &= (0,1)_0 - \varrho e^{\lambda t} \int_0^t e^{-\lambda \tau} h_{\tau} d\tau \\ (1,0)_t &= (1,0)_0 - \varrho \int_0^t h_{\tau} d\tau \\ (1,1)_t &= (1,1)_0 + \varrho e^{\lambda t} \int_0^t e^{-\lambda \tau} h_{\tau} d\tau\end{aligned}\quad (4.57)$$

gelöst werden. Dies ergibt sich aus der Formel für homogene beziehungsweise inhomogene lineare Differentialgleichungen erster Ordnung, welche zum Beispiel bei Walter [21, S. 25] zu finden ist, kann aber auch durch Nachrechnen verifiziert werden.

Eine explizite Form für h_t erhalten wir nun, indem wir zunächst die Differentialgleichung, welcher h_t genügt, herleiten und diese dann lösen. Damit lässt sich dann auch eine explizite Form für die Integrale finden, welche in den Lösungen (4.57) vorkommen.

Durch Einsetzen der Dynamiken (4.53) in die Definition von h_t und Ableiten nach der Quotientenregel zeigen wir, dass die Dynamik von h_t durch die Differentialgleichung

$$\begin{aligned}
\dot{h}_t &= \frac{d}{dt} \frac{(0,1)_t (1,0)_t - (0,0)_t (1,1)_t}{\|z_t\|} \\
&= \frac{\frac{d}{dt} \left((0,1)_t (1,0)_t - (0,0)_t (1,1)_t \right)}{\|z_t\|} - \frac{\left((0,1)_t (1,0)_t - (0,0)_t (1,1)_t \right) \frac{d}{dt} \|z_t\|}{\|z_t\|^2} \\
&= \frac{\dot{(0,1)}_t (1,0)_t + (0,1)_t \dot{(1,0)}_t - \dot{(0,0)}_t (1,1)_t - (0,0)_t \dot{(1,1)}_t}{\|z_t\|} - \frac{\|\dot{z}_t\|}{\|z_t\|} h_t \\
&= \frac{(\lambda (0,1)_t - \varrho h_t) (1,0)_t - (0,1)_t \varrho h_t - \varrho h_t (1,1)_t - (0,0)_t (\lambda (1,1)_t + \varrho h_t)}{\|z_t\|} - \frac{\|\dot{z}_t\|}{\|z_t\|} h_t \\
&= \frac{\lambda \left((0,1)_t (1,0)_t - (0,0)_t (1,1)_t \right) - \varrho h_t \left((0,0)_t + (0,1)_t + (1,0)_t + (1,1)_t \right)}{\|z_t\|} - \frac{\|\dot{z}_t\|}{\|z_t\|} h_t \\
&= \left(\lambda - \varrho - \frac{\|\dot{z}_t\|}{\|z_t\|} \right) h_t
\end{aligned} \tag{4.58}$$

beschrieben wird.

Wie wir bereits wissen, ist die Rate des exponentiellen Abfalls der Kopplungsgleichgewichte T_G durch $-\sum_{\alpha \in \bar{G}} \varrho_\alpha$ gegeben. Der Faktor $-\varrho$ in (4.58) ist also interpretierbar als die Wirkung der Rekombination auf h_t . Der Term $\left(\lambda - \frac{\|\dot{z}_t\|}{\|z_t\|} \right)$ ist somit auf die Reproduktion zurückzuführen.

Die Lösung dieser Differentialgleichung ist

$$\begin{aligned}
h_t &= e^{(\lambda - \varrho)t} \frac{\|z_0\|}{\|z_t\|} h_0 \\
&= \frac{\left((0,1)_0 (1,0)_0 - (0,0)_0 (1,1)_0 \right) e^{(\lambda - \varrho)t}}{(0,0)_0 + (1,0)_0 + \left((0,1)_0 + (1,1)_0 \right) e^{\lambda t}} \\
&= \frac{\left((0,1)_0 (1,0)_0 - (0,0)_0 (1,1)_0 \right) e^{-\varrho t}}{\left((0,0)_0 + (1,0)_0 \right) e^{-\lambda t} + (0,1)_0 + (1,1)_0},
\end{aligned} \tag{4.59}$$

wie sich durch Ableiten verifizieren lässt.

Um die Lösungen der Gametendynamiken angeben zu können, müssen wir über h_t und $e^{-\lambda t} h_t$ integrieren. Genauer gesagt integrieren wir mit Hilfe von *Mathematica* Integrale der Form

$$I(a, b, A, B; t) := \int_0^t \frac{e^{-A\tau}}{a + be^{-B\tau}} d\tau \tag{4.60}$$

und erhalten als Lösung

$$I(a, b, A, B; t) = \frac{1}{b(B-A)} \left(e^{(B-A)t} {}_2F_1\left(1, 1 - \frac{A}{B}, 2 - \frac{A}{B}, -\frac{a}{b}e^{Bt}\right) - {}_2F_1\left(1, 1 - \frac{A}{B}, 2 - \frac{A}{B}, -\frac{a}{b}\right) \right) \quad (4.61)$$

eine Kombination von hypergeometrischen Funktionen. Diese Lösung kann durch Ableiten verifiziert werden.

Integrale dieser Form sind also prinzipiell lösbar und es können konkrete Werte ausgerechnet werden, wir schreiben im Folgenden aus Gründen der Übersichtlichkeit jedoch weiterhin an gegebener Stelle lediglich das Integral.

Schlussendlich erhalten wir den Vektor der relativen Gametenhäufigkeiten $p_t = \frac{z_t}{\|z_t\|}$, indem wir durch die Norm teilen.

4.5.2 Zwei neutrale und ein selektiver Site

Nun wollen wir uns der Lösung des Szenarios mit drei Sites, wovon einer selektiv ist, zuwenden. Wir beschränken uns wieder auf zwei mögliche Allele pro Site, so dass $S = \{0, 1, 2\}$ und $X_1 = X_2 = X_3 = \{0, 1\}$ gilt. Die Menge der möglichen Gameten ergibt sich somit zu $X = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$.

In diesem Szenario gibt es zwei mögliche Links zum Rekombinieren, somit ist $L = \{\frac{1}{2}, \frac{3}{2}\}$. Aus Anschauungsgründen bezeichnen wir die Rekombinationsraten als $\varrho_l := \varrho_{\frac{1}{2}}$ und $\varrho_r := \varrho_{\frac{3}{2}}$, da die Rekombination zum einen *links* von und zum anderen *rechts* des mittleren Sites 1 stattfindet.

Der Selektionsvorteil λ kann nun an drei verschiedenen Sites wirken. Aus Symmetriegründen sind die Fälle, dass der Vorteil an einem der äußeren Sites 0 oder 2 wirkt, gleich zu behandeln. Wir stellen im Folgenden die Lösung für Selektion an Site 0 dar. Danach wenden wir uns der strukturell von den beiden vorherigen Fällen abweichenden Selektion in der Mitte an Site 1 zu. Da es in jedem Fall nur einen selektiven Site gibt, handelt es sich erneut um additive Selektion.

Wir führen hier ebenfalls die kürzere Notation $(x_1, x_2, x_3)_t := z_t(x_1, x_2, x_3)$ ein. Darüberhinaus betrachten wir nicht direkt die Dynamik der Gametenhäufigkeiten z_t , sondern die Dynamik der möglichen Marginalisierungen auf das Allel 0.

Genauer gesagt bedienen wir uns der invertierbaren linearen Abbildung

$$M = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad (4.62)$$

welche den Vektor $q = (0, 1)^T$ zu

$$Mq = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0+1 \\ 0 \end{pmatrix} = \begin{pmatrix} * \\ 0 \end{pmatrix} \quad (4.63)$$

transformiert. Mit dem Tensorprodukt erweitern wir die Abbildung M auf drei Allele. Wir betrachten also im Folgenden die zeitliche Dynamik des Vektors

$$y_t := (M \otimes M \otimes M) z_t = \begin{pmatrix} *, *, * \\ *, *, 0 \\ *, 0, * \\ *, 0, 0 \\ 0, *, * \\ 0, *, 0 \\ 0, 0, * \\ 0, 0, 0 \end{pmatrix}_t, \quad (4.64)$$

welcher aufgrund der Invertierbarkeit von M die Dynamik von z_t bestimmt.

Selektion an Site 0

Basierend auf der Dynamik der Gametenhäufigkeiten (4.3) leiten wir Ausdrücke für die zeitlichen Ableitungen der unterschiedlichen Komponenten von y_t her und lösen diese.

Es lässt sich zeigen, dass

$$\frac{d}{dt} (0, *, *)_t = 0 \quad (4.65)$$

und

$$\frac{d}{dt} (1, *, *)_t = \lambda (1, *, *)_t \quad (4.66)$$

gilt. Daraus folgt für die fünfte Komponente von y_t

$$(0, *, *)_t = (0, *, *)_0. \quad (4.67)$$

Weiterhin ergibt sich mit Hilfe der Lösung von (4.66)

$$(1, *, *)_t = e^{t\lambda} (1, *, *)_0 \quad (4.68)$$

für die erste Komponente

$$(*, *, *)_t = (0, *, *)_t + (1, *, *)_t = (0, *, *)_0 + e^{t\lambda} (1, *, *)_0 = \|z_t\|. \quad (4.69)$$

Um Lösungen für die weiteren Komponenten zu erhalten, benötigen wir ähnlich wie in Abschnitt 4.5.1 den zeitlichen Verlauf bestimmter Kopplungsungleichgewichte. Zunächst definieren wir

$$h_t(l) := -T_{\{\frac{3}{2}\}}(z_t)(0, 0, *) = \frac{1}{\|z_t\|} \left((0, *, *)_t (*, 0, *)_t - (0, 0, *)_t \|z_t\| \right) \quad (4.70)$$

und

$$h_t(l, r) := -T_{\emptyset}(z_t)(0, *, 0) = \frac{1}{\|z_t\|} \left((0, *, *)_t (*, *, 0)_t - (0, *, 0)_t \|z_t\| \right). \quad (4.71)$$

Es lässt sich zeigen, dass sich die Differentialgleichung der sechsten Komponente von y_t zu

$$\frac{d}{dt} (0, *, 0)_t = (\varrho_l + \varrho_r) \left(\frac{1}{\|z_t\|} (0, *, *)_t (*, *, 0)_t - (0, *, 0)_t \right) \quad (4.72)$$

ergibt. Dies ähnelt der Differentialgleichung für die Komponente $(0, 0)$ in (4.53) und wenn die Häufigkeiten

$$\begin{aligned} (0, *, 0)_t &\stackrel{\wedge}{=} (0, 0)_t \\ (0, *, 1)_t &\stackrel{\wedge}{=} (0, 1)_t \\ (1, *, 0)_t &\stackrel{\wedge}{=} (1, 0)_t \\ (1, *, 1)_t &\stackrel{\wedge}{=} (1, 1)_t \end{aligned} \quad (4.73)$$

miteinander identifiziert werden, lässt sich zeigen, dass sie den gleichen algebraischen Zusammenhängen und zeitlichen Dynamiken unterliegen. ϱ entspricht dabei $(\varrho_l + \varrho_r)$ und h_t entspricht $h_t(l, r)$. Der mittlere Site kann also ignoriert und die Rekombinationsraten aufaddiert werden.

Somit lässt sich analog zu dem Vorgehen in Kapitel 4.5.1 zeigen, dass die Lösung von (4.72) durch

$$(0, *, 0)_t = (0, *, 0)_0 + (\varrho_l + \varrho_r) \int_0^t h_\tau(l, r) d\tau \quad (4.74)$$

gegeben ist.

Eine ähnliche Argumentation wie für die sechste Komponente von y_t lässt sich auch für die siebte Komponente durchführen, so dass sich hier die Lösung zu

$$(0, 0, *)_t = (0, 0, *)_0 + \varrho_l \int_0^t h_\tau(l) d\tau \quad (4.75)$$

ergibt. Der rechte Site wird vernachlässigt, so dass nur die Rekombinationsrate ϱ_l und das Kopplungsungleichgewicht $h_t(l)$ relevant sind.

Mit diesen Lösungen und den Definitionen von $h_t(l)$ (4.70) sowie $h_t(l, r)$ (4.71) lassen sich nun auch die zweite und dritte Komponente von y_t auf bekannte Größen zurückführen. Durch elementare Umformungen der Definitionen lässt sich zeigen, dass

$$(*, *, 0)_t = \frac{h_t(l, r) \|z_t\| + (0, *, 0)_t \|z_t\|}{(0, *, *)_t} \quad (4.76)$$

und

$$(*, 0, *)_t = \frac{h_t(l) \|z_t\| + (0, 0, *)_t \|z_t\|}{(0, *, *)_t} \quad (4.77)$$

gilt.

Die Transformation $(M \otimes M \otimes M)$ entkoppelt die acht Komponenten von z_t soweit, dass sechs Komponenten sukzessive gelöst werden können. Für die restlichen beiden

Komponenten gilt dies nicht, so dass die Differentialgleichungen für diese gekoppelt betrachtet werden müssen. Aus der Gametendynamik (4.3) ergibt sich also

$$\frac{d}{dt} \begin{pmatrix} *, 0, 0 \\ 0, 0, 0 \end{pmatrix}_t = \begin{pmatrix} -(\varrho_l + \varrho_r) & \frac{\varrho_l}{\|z_t\|} (0, *, *)_t \\ -\lambda & \lambda - \varrho_l \end{pmatrix} \begin{pmatrix} *, 0, 0 \\ 0, 0, 0 \end{pmatrix}_t + \begin{pmatrix} \frac{\varrho_r}{\|z_t\|} (*, 0, *)_t (*, *, 0)_t \\ \frac{\varrho_r}{\|z_t\|} (0, 0, *)_t (*, *, 0)_t \end{pmatrix} \quad (4.78)$$

für die vierte und achte Komponente von y_t .

Da der Ausdruck $\frac{\varrho_l}{\|z_t\|} (0, *, *)_t$ zeitabhängig ist, ist die Differentialgleichung (4.78) nicht ohne weiteres lösbar und muss zunächst in der allgemeinen Form

$$\dot{x}_t = A_t x_t + b_t \quad (4.79)$$

einer inhomogenen linearen gekoppelten Differentialgleichung betrachtet werden. Da in unserem Fall $[A_t, A_s] \neq 0$ ist, also die Matrix A_t für unterschiedliche Zeiten nicht mit sich selbst kommutiert, muss die Lösung der homogenen Differentialgleichung $\dot{x}_t = A_t x_t$ in der allgemeinen Form des zeitgeordneten Exponentials $OE[A]_{t,t_0}$ betrachtet werden, für welches wir die Definition

$$OE[A]_{t,t_0} = \mathbb{1} + \sum_{i=1}^{\infty} \int_{t_0}^t A_{\tau_1} \int_{t_0}^{\tau_1} A_{\tau_2} \dots \int_{t_0}^{\tau_{n-1}} A_{\tau_n} d\tau_n \dots d\tau_2 d\tau_1 \quad (4.80)$$

verwenden.

Anhand dieser Definition lässt sich für den Fall $[A_t, A_s] = 0$ zeigen, dass

$$OE[A]_{t,t_0} = e^{\int_{t_0}^t A_{\tau} d\tau} \quad (4.81)$$

gilt. Für einen mit sich selbst kommutierenden Operator ist diese allgemeine Lösung also gleich der gemeinhin bekannteren Lösung für lineare Differentialgleichungen.

Weiterhin lässt sich mit Hilfe der Flusseigenschaft $OE[A]_{t,r} OE[A]_{r,s} = OE[A]_{t,s}$ und dem allgemeinen Satz zur Wronski-Determinante zeigen, dass die Beziehung

$$OE[A]_{t_0,t}^{-1} = OE[A]_{t,t_0} \quad (4.82)$$

gilt und der Ausdruck auf der rechten Seite sowohl Links- als auch Rechtsinverses ist. Details zur Wronski-Determinante finden sich bei Walter [21].

Unter Zuhilfenahme der Eigenschaften von (4.80) lässt sich zeigen, dass inhomogene gekoppelte lineare Differentialgleichungen der Form (4.79) durch

$$x_t = OE[A]_{t,t_0} x_0 + OE[A]_{t,t_0} \int_0^t OE[A]_{0,\tau} b_{\tau} d\tau \quad (4.83)$$

gelöst werden. Da für $t \neq s$ in unserem Fall $[A_t, A_s] \neq 0$ gilt, müssen wir diese allgemeine Form verwenden.

Durch Einsetzen der entsprechenden Ausdrücke aus der Differentialgleichung (4.78) in die Lösung (4.83) ist es möglich, einen Ausdruck für die zeitliche Entwicklung der beiden verbliebenen Komponenten

$$\begin{pmatrix} *, 0, 0 \\ 0, 0, 0 \end{pmatrix}_t \quad (4.84)$$

zu erhalten.

Somit haben wir Lösungen für alle acht Komponenten von y_t , falls die Selektion am Site 0 wirkt, und wenden uns nun der Selektion an Site 1 zu.

Selektion an Site 1

Wirkt die Selektion am mittleren Site, so entkoppelt die Transformation $(M \otimes M \otimes M)$ die Dynamik der Gametenhäufigkeiten soweit, dass wir jede Komponente von y_t einzeln betrachten können und nicht auf ein zeitgeordnetes Exponential zurückgreifen müssen. Der restliche Lösungsweg ist weitestgehend analog zu dem der Selektion an Site 0.

Wir beginnen wieder damit, die Lösungen für die Marginalhäufigkeiten am selektiven Site

$$(*, 0, *)_t = (*, 0, *)_0 \quad (4.85)$$

sowie

$$(*, 1, *)_t = e^{t\lambda} (*, 1, *)_0 \quad (4.86)$$

und darauf aufbauend die Lösung für die Dynamik der Populationsgröße

$$(*, *, *)_t = (*, 0, *)_0 + e^{t\lambda} (*, 1, *)_0 = \|z_t\| \quad (4.87)$$

herzuleiten.

Mit Hilfe der Definition für $h_t(l)$ und der zusätzlichen Definition

$$h_t(r) := -T_{\{\frac{1}{2}\}}(z_t) (*, 0, 0) = \frac{1}{\|z_t\|} \left((*, 0, *)_t (*, *, 0)_t - (*, 0, 0)_t \|z_t\| \right) \quad (4.88)$$

können wir erneut die Lösung für die Marginalisierungen auf den selektiven und einen neutralen Site angeben. Es gilt

$$(0, 0, *)_t = (0, 0, *)_0 + \int_0^t h_\tau(l) d\tau \quad (4.89)$$

und

$$(*, 0, 0)_t = (*, 0, 0)_0 + \int_0^t h_\tau(r) d\tau. \quad (4.90)$$

Aus den Definitionen der Kopplungsungleichgewichte folgt wieder

$$(0, *, *)_t = \frac{h_t(l) \|z_t\| + (0, 0, *)_t \|z_t\|}{(*, 0, *)_t} \quad (4.91)$$

sowie

$$(*, *, 0)_t = \frac{h_t(r) \|z_t\| + (*, 0, 0)_t \|z_t\|}{(*, 0, *)_t} \quad (4.92)$$

für die zeitlichen Dynamiken der neutralen Marginalen.

Wir sind nun in der Lage die beiden verbleibenden Größen, welche im vorherigen Fall nur gekoppelt betrachtet werden konnten, getrennt zu lösen.

Die Häufigkeit des Gameten $(0, 0, 0)$ gehorcht der Differentialgleichung

$$\frac{d}{dt} (0, 0, 0)_t = -(\varrho_l + \varrho_r) (0, 0, 0)_t + \underbrace{\frac{1}{\|z_t\|} (\varrho_l (0, *, *)_t (*, 0, 0)_t + \varrho_r (0, 0, *)_t (*, *, 0)_t)}_{:=g_t}, \quad (4.93)$$

welche mit Hilfe der allgemeinen Formel für lineare inhomogene Differentialgleichungen aus dem Buch von Walter [21, S. 25] zu

$$(0, 0, 0)_t = e^{-t(\varrho_l + \varrho_r)} \left((0, 0, 0)_0 + \int_0^t e^{\tau(\varrho_l + \varrho_r)} g_\tau d\tau \right) \quad (4.94)$$

gelöst werden kann.

Die Differentialgleichung für die Marginalisierung auf die beiden neutralen Allele lautet

$$\begin{aligned} \frac{d}{dt} (0, *, 0)_t &= \lambda \left((0, *, 0)_t - (0, 0, 0)_t \right) + (\varrho_l + \varrho_r) \left(\frac{1}{\|z_t\|} (0, *, *)_t (*, *, 0)_t - (0, *, 0)_t \right) \\ &= (\lambda - (\varrho_l + \varrho_r)) (0, *, 0)_t - \lambda (0, 0, 0)_t + \underbrace{\frac{\varrho_l + \varrho_r}{\|z_t\|} (0, *, *)_t (*, *, 0)_t}_{\tilde{g}_t} \end{aligned} \quad (4.95)$$

und ist ebenfalls linear und inhomogen. Sie lässt sich mit

$$(0, *, 0)_t = e^{t(\lambda - (\varrho_l + \varrho_r))} \left((0, *, 0)_0 + \int_0^t e^{\tau(\varrho_l + \varrho_r - \lambda)} \tilde{g}_\tau d\tau \right) \quad (4.96)$$

lösen.

Wir haben für jede der acht Komponenten von y_t Lösungen hergeleitet, sowohl im Falle der Selektion am mittleren Site als auch am linken und damit aus Symmetriegründen ebenfalls am rechten Site. Wir sind also in der Lage y_t und damit z_t sowie $p_t = \frac{z_t}{\|z_t\|}$ in jedem der drei Fälle zu berechnen. Allerdings beinhalten die Lösungen zeitgeordnete Exponentiale und sie erfordern das Integrieren über hypergeometrische Funktionen, so dass nur schwer von expliziten Lösungen gesprochen werden kann.

Nachdem sich die von Baake [1] vorgeschlagene Lösung als falsch erwies, leiteten wir für zwei Spezialfälle explizite Lösungen her, da wir diese Spezialfälle für die Untersuchungen im nächsten Kapitel benötigen.

Wir überzeugten uns von der Richtigkeit der Lösung des 2-Site-/2-Allel-Systems (4.57), indem wir die Differentialgleichung (4.46) von *Mathematica* mit Hilfe des Befehls *NDSolve* [22, S. 103] approximieren ließen. Da die Differenz zu der exakten Lösung im Rahmen der numerischen Genauigkeit lag, ist die hergeleitete Lösung richtig.

5 Der Hitchhiking-Effekt

Wir wollen nun den sogenannten *Hitchhiking-Effekt* mittels unseres Modells der Rekombinations- und Selektionsdynamik und den zugehörigen Lösungen beziehungsweise deren numerischen Approximationen untersuchen.

Der Hitchhiking-Effekt wird bei zwei bezüglich Rekombination stark gekoppelten Sites beobachtet, von denen einer selektiv und einer neutral ist. Befindet sich die Population im Kopplungsungleichgewicht und es sind bestimmte Allele am neutralen Site überrepräsentiert gekoppelt mit den selektiv bevorteilten Allelen am selektiven Site, so kommt es zu einem Anstieg der marginalen Häufigkeit dieser Allele, obwohl sie direkt keinen Selektionsvorteil gegenüber den anderen Allelen am neutralen Site besitzen. Dieser Anstieg resultiert aus der starken Kopplung mit den selektiv bevorteilten Allelen, welche die neutralen Allele „per Anhalter“ mitnehmen. Untersuchungen zu diesem Thema finden sich zum Beispiel bei Maynard Smith [15], Kaplan [11] und Braverman [5], wobei der Begriff *Hitchhiking-Effekt* zuerst bei Maynard Smith auftauchte.

Im Folgenden geben wir zunächst einige Ergebnisse an, die diesen Effekt genauer aufzeigen sollen.

Darüberhinaus war das Ziel dieser Arbeit ursprünglich, in dem von Stephan [19] präsentierten Szenario die vom Autor gewonnenen Ergebnisse bezüglich des Hitchhiking-Effektes auf das Kopplungsungleichgewicht zwischen neutralen Sites mit Hilfe der von Baake [1] vorgeschlagenen allgemeinen und expliziten Lösung zu reproduzieren. Da sich diese allgemeine Lösung im Laufe dieser Arbeit jedoch als falsch erwies, reproduzieren wir die Ergebnisse mit einer numerischen Approximation der Rekombinations- und Selektionsdynamik (4.1), welche wir auf dieses Szenario spezialisieren.

5.1 Effekt auf einen neutralen Site

Die Grundzüge des Effektes lassen sich in dem in Abschnitt 4.5.1 eingeführten 2-Site-/2-Allel-System aufzeigen. Wie bereits erwähnt, besteht dieses System aus zwei Sites $S = \{0, 1\}$, an denen jeweils zwei Allele $X_0 = \{0, 1\} = X_1$ stehen können. An Site 1 wirkt die Selektion, wobei Allel 0 neutral ist und Allel 1 einen Vorteil von λ hat, was zu dem Reproduktionsoperator $F = \text{diag}(0, \lambda, 0, \lambda)$ führt. Die einzig vorkommende Rekombinationsrate bezeichnen wir erneut mit ϱ .

Der Effekt tritt nur auf, wenn die Häufigkeiten bestimmter Kombinationen von Allelen häufiger sind als das Produkt der Marginalwahrscheinlichkeiten, die Population sich also

im Kopplungsungleichgewicht befindet. Deshalb wählen wir

$$z_0 = \begin{pmatrix} z_0(0,0) \\ z_0(0,1) \\ z_0(1,0) \\ z_0(1,1) \end{pmatrix} = \begin{pmatrix} 0.4 \\ 0.1 \\ 0.1 \\ 0.4 \end{pmatrix} \quad (5.1)$$

als Anfangsbedingung. Somit ist der Wert des Kopplungsungleichgewichtes

$$T_{\varnothing} \left(\frac{z_0}{\|z_0\|} \right) (1,1) = 0.25 - 0.04 = 0.19. \quad (5.2)$$

Es fällt auf, dass Allel 1 an Site 0 häufiger mit Allel 1 an Site 1 zusammen vorkommt als es der Fall wäre, wenn die marginalen Allelhäufigkeiten von den Allelhäufigkeiten des jeweils anderen Sites unabhängig wären. Genauer gesagt ist die auf Allel 1 an Site 1 bedingte Wahrscheinlichkeit $p_0((1,*)|(1,1)) = 0.8$ von Allel 1 an Site 0 größer als die bedingte Wahrscheinlichkeit $p_0((0,*)|(1,1)) = 0.2$ von Allel 0. Da Allel 1 an Site 1 den Selektionsvorteil hat, erwarten wir bei starker Kopplung bezüglich Rekombination, dass die Marginalhäufigkeit von Allel 1 an Site 0 ansteigt. Je schwächer die Kopplung wird, desto schwächer wird dies geschehen.

Durch Variieren des Parameters ϱ können wir nun die beiden Sites unterschiedlich stark aneinander koppeln. Ein kleiner Wert entspricht einer starken Kopplung. Für $\varrho = 0$ verhalten sich die beiden Sites bezüglich Rekombination und Selektion, als wären sie ein Site. Wird der Wert größer, so lässt die Stärke der Kopplung nach. Der Einfluss des selektiven auf den neutralen Site wird aufgrund der Struktur der Dynamik nur im Limes $\varrho \rightarrow \infty$ verschwinden, jedoch spielt er ab einer gewissen Größenordnung praktisch keine Rolle mehr.

Um das Verhalten der Marginalhäufigkeit an Site 0 zu quantifizieren, bedienen wir uns der in Abschnitt 4.5.1 hergeleiteten Lösung der Rekombinations- und Selektionsdynamik für ein 2-Site-/2-Allel-System mit einem selektiven Site. Die Ergebnisse sind in Abbildung 5.1 auf der nächsten Seite zu sehen. Die Zeit entwickelt sich entlang der x -Achse und auf der z -Achse ist die Marginalwahrscheinlichkeit des Allels 1 an Site 0 $p_t(1,*)$ aufgetragen. Die Stärke der Kopplung, also der Parameter ϱ , wird entlang der y -Achse variiert. Da das Szenario der zwei Sites vollkommen symmetrisch ist, bedeutet ein negativer Wert für die Rekombinationsrate ϱ , dass in die Lösung der Absolutbetrag von ϱ eingesetzt wird.

Wie erwartet ist zu beobachten, dass sich für eine Rekombinationsrate von 0 die beiden Sites bezüglich der wirkenden Kräfte wie ein Site verhalten. Das Allel 1 an Site 0 wird von dem Allel mit Selektionsvorteil mitgezogen, so dass die Marginalwahrscheinlichkeit asymptotisch gegen die anfängliche Marginalwahrscheinlichkeit bedingt auf Allel 1 an Site 1 $p_0((1,*)|(1,1)) = 0.8$ strebt. Diese bedingte Marginalwahrscheinlichkeit ist zeitlich konstant, da wegen $\varrho = 0$ keine Rekombination stattfindet.

Erhöhen wir nun die Rekombinationsrate, nimmt die Kopplung des neutralen Sites mit dem selektiven ab und der durch die Kopplung bedingte Vorteil des Allels 1 an Site 0 verringert sich. Die Marginalwahrscheinlichkeit strebt nicht mehr gegen das für $\varrho = 0$

Marginalwahrscheinlichkeit des Allels 1 am neutralen Site

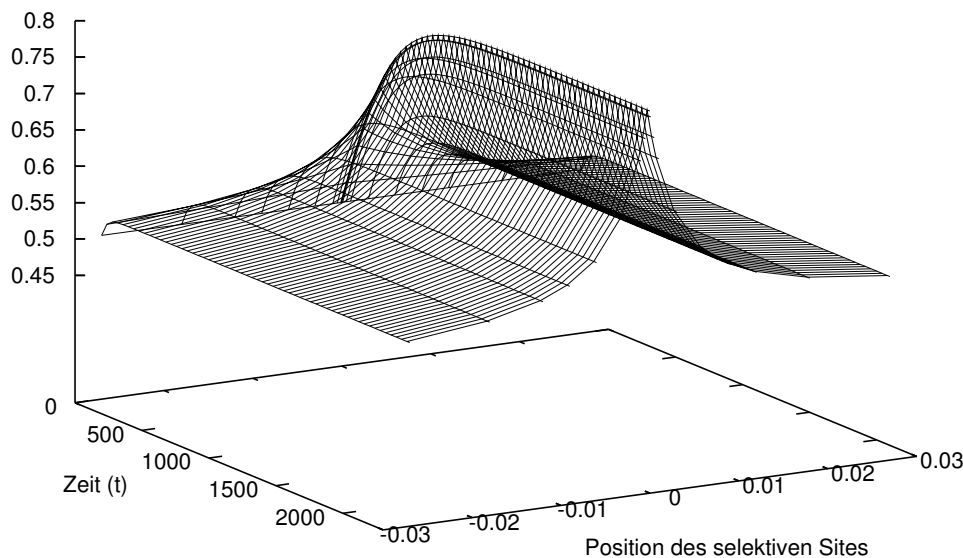


Abbildung 5.1: Auf der z -Achse ist die Marginalwahrscheinlichkeit des Allels 1 am neutralen Site aufgetragen, welche sich mit der Zeit entlang der x -Achse entwickelt. In Richtung der y -Achse wird die Rekombinationsrate ρ und dadurch der *Abstand* zwischen dem selektiven und dem neutralen Site variiert.

mögliche Maximum von 0.8, sondern nähert sich asymptotisch einem geringeren Wert an.

Mit zunehmender Rekombinationsrate zwischen den beiden Sites verringert sich das von der Marginalwahrscheinlichkeit angestrebte Maximum. Diese Abnahme geschieht bei geringen Werten für ρ relativ schnell, für größere Werte nimmt das Maximum zunehmend langsamer ab. Für große ρ nähert es sich der Anfangsmarginalwahrscheinlichkeit des Allels 1 $p_0(1, *) = 0.5$, bedingt auf die gesamte Population, an. Außer der Selektion an Site 1 wirken keine Kräfte, welche die Marginalwahrscheinlichkeit verringern können, weshalb die asymptotische Schranke keinen geringeren Wert als 0.5 annehmen kann.

Wir haben in einem kleinen Szenario, welches analytisch behandelt werden konnte, aufgezeigt, wie der Hitchhiking-Effekt auf die Allelhäufigkeiten eines neutralen Sites in der Nachbarschaft eines selektiven Sites wirkt. Je nachdem, wie stark die Rekombination zwischen den Sites ist, bewirkt der Hitchhiking-Effekt, dass sich bestimmte Allele gegenüber anderen in der Population durchsetzen, obwohl diese für sich betrachtet keinen selektiven Vorteil besitzen. Der Hitchhiking-Effekt bewirkt also eine Abnahme der ge-

netischen Variabilität in der unmittelbaren genetischen Umgebung selektiver Sites. Mit zunehmender Entfernung lässt dieser Effekt jedoch nach.

Der Hitchhiking-Effekt lässt sich ebenfalls in Abbildung 1.1 auf Seite 4 beobachten. Gehen wir davon aus, dass die Erbanlagen für Farbe und Größe an zwei getrennten Sites liegen, so handelt es sich hier um ein 2-Site-Szenario. Die beiden Allele für schwarz und weiß an dem Farbsite sind neutral bezüglich Selektion. An dem anderen Site gibt es zwei Allele, welche die Größe des Individuums bestimmen. Das Allel, welches ein Individuum größer werden lässt, gibt diesem einen Selektionsvorteil.

Es ist nun zu beobachten, dass das Allel für die Farbe weiß zu Beginn die Population dominiert. Allerdings ist das Allel für die Farbe schwarz stark mit dem Allel für die vorteilhafte Größe gekoppelt. Die Selektion wirkt zwar nur auf die Größe eines Individuums, jedoch besitzt am Ende die gesamte Population das Allel für schwarz. Das neutrale Allel für die Farbe schwarz wurde von der vorteilhaften Größe „per Anhalter“ mitgenommen.

5.2 Effekt auf Kopplungsungleichgewicht zwischen zwei neutralen Sites

Nachdem wir gezeigt haben, wie sich der Hitchhiking-Effekt auf die Gametenhäufigkeiten auswirkt, wollen wir nun die Auswirkungen auf das Kopplungsungleichgewicht zwischen zwei neutralen Sites untersuchen.

Ähnliche Untersuchungen wurden bereits von Stephan [19] durchgeführt. Die Autoren benutzten zu diesem Zwecke den Formalismus von Barton und Turelli.

Barton und Turelli leiten in ihrer Arbeit [2] eine Lösung der Rekombinationsdynamik (2.22) mit zusätzlich wirkender Selektion her. Sie untersuchen also die diskrete Dynamik, in welcher Rekombinationen von beliebigen Teilmengen der Sites erlaubt sind. Sie leiten keine exakte Lösung her, sondern lediglich eine Approximation.

Stephan [19] benutzt zwar diesen Formalismus, welcher ihm erlaubt, beliebig zu rekombinieren, beschränkt sich jedoch bei seinen Untersuchungen auf Rekombinationsereignisse, welche die Sites in zwei zusammenhängende Gruppen teilen, also solche, wie sie in Abschnitt 2.4 beschrieben und in Abbildung 2.3 zu sehen sind.

Da wir uns beim Herleiten der Lösung für die Rekombinationsdynamik ebenfalls auf solche Rekombinationsereignisse beschränken, unterscheidet sich die Dynamik 4.1 von der in [19] untersuchten Dynamik nur darin, dass sie kontinuierlich und nicht diskret ist.

Aufgrund dieser Ähnlichkeit der Dynamiken und der Tatsache, dass Stephan lediglich eine Approximation der Lösung der diskreten Dynamik untersuchte, beabsichtigten wir, einen Teil der Ergebnisse aus [19] mit der exakten Lösung der kontinuierlichen Dynamik von Baake [1] zu reproduzieren. Wir benutzten zunächst die Variante der vermeintlichen Lösung (4.38), welche erst rekombiniert und dann selektiert. Da unsere ersten Ergebnisse denen von Stephan stark ähnelten, gingen wir davon aus, dass die leichten Abweichungen auf den Unterschied zwischen diskreten und kontinuierlichen Dynamiken zurückzuführen sei. Es stellte sich jedoch heraus, dass die vermeintlich exakte Lösung falsch ist, so dass die ersten Ergebnisse nicht verwendet werden können.

Wir lösten die Dynamik 4.1 daraufhin für den in Abschnitt 4.5.2 vorgestellten und hier benötigten Spezialfall exakt. Allerdings ist die analytische Lösung für den Fall mit drei Sites in einer Form, in der das Ausrechnen konkreter Werte nicht möglich ist, so dass wir zum Reproduzieren der Ergebnisse von Stephan eine numerische Approximation der Differentialgleichung durch *Mathematica* verwendeten.

5.2.1 Das Szenario

Das Szenario, in welchem Stephan die Auswirkung des Hitchhiking-Effektes auf das Kopplungsungleichgewicht zwischen zwei neutralen Sites untersuchte, haben wir bereits in Abschnitt 4.5.2 vorgestellt. Es gibt drei Sites $S = \{0, 1, 2\}$ und an jedem dieser Sites sind zwei Allele möglich $X_0 = X_1 = X_2 = \{0, 1\}$. Zwischen den Sites gibt es zwei Links $S = \{\frac{1}{2}, \frac{3}{2}\}$, an denen die Rekombination mit den Raten $\varrho_l := \varrho_{\frac{1}{2}}$ und $\varrho_r := \varrho_{\frac{3}{2}}$ wirkt.

Den Selektionsvorteil λ besitzt das Allel 1 gegenüber Allel 0, welches lediglich eine Nettoreproduktionsrate von 0 hat. Dieser Selektionsvorteil wirkt in den drei strukturell zu unterscheidenden Szenarien an jeweils einem anderen Site. Im Folgenden bezeichnen wir den *linken* neutralen Site, also den mit dem kleinsten Index aus der Menge S , mit L, den *rechten* neutralen Site mit R und den Site, an dem die Selektion wirkt, mit S. Abbildung 5.2 zeigt die drei möglichen linearen Anordnungen SLR, LSR, und LRS der drei Sites auf.

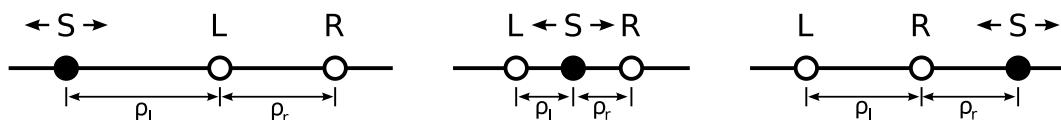


Abbildung 5.2: Der selektive Site (gefüllter Kreis) kann links, in der Mitte oder rechts von den beiden neutralen Sites (ungefüllter Kreis) sein. Die Rekombinationsraten sind an den entsprechenden Stellen eingezeichnet.

Über die Rekombinationsraten ϱ_l und ϱ_r lässt sich der *Abstand* zwischen den Sites angeben. Ist die Rekombinationsrate zwischen zwei Sites groß, so gehen wir davon aus, dass auf dem Genom ein großer Abstand zwischen den Sites ist. Bei einer kleinen Rekombinationsrate gehen wir von einem kleinen Abstand aus, und für eine Rekombinationsrate von 0 liegen die Sites sogar *übereinander*.

Wir wählen die Rekombinationsraten so, dass der *Abstand* zwischen den beiden neutralen Sites L und R in allen drei Szenarien gleich 0.0002 ist. Für SLR bedeutet dies, dass $\varrho_r = 0.0002$ gewählt wird, wohingegen bei LRS $\varrho_l = 0.0002$ zu setzen ist. Im LSR-Szenario erreichen wir einen gleichbleibenden *Abstand*, indem immer $\varrho_l + \varrho_r = 0.0002$ gilt. Hierbei sind $\varrho_l, \varrho_r > 0$.

Der verbleibende Freiheitsgrad bestimmt nun die Position von S relativ zu den neutralen Sites. Um diese Position entlang des Freiheitsgrades variieren zu können, führen wir den Parameter s ein, welcher die Werte der Rekombinationsraten und die Form des Reproduktionsoperators bestimmt.

Für die von s abhängigen Rekombinations- und Selektionsparameter gilt

$$\varrho_l(s) = \begin{cases} -s - 0.0001 & \text{für } s \in (-\infty, -0.0001) \\ s + 0.0001 & \text{für } s \in [-0.0001, 0.0001], \\ 0.0002 & \text{sonst} \end{cases} \quad (5.3)$$

$$\varrho_r(s) = \begin{cases} 0.0002 & \text{für } s \in (-\infty, -0.0001) \\ -s + 0.0001 & \text{für } s \in [-0.0001, 0.0001] \\ s - 0.0001 & \text{sonst} \end{cases} \quad (5.4)$$

und

$$F(s) = \begin{cases} F = \text{diag}(0, \lambda) \otimes \mathbb{1} \otimes \mathbb{1} & \text{für } s \in (-\infty, -0.0001) \\ F = \mathbb{1} \otimes \text{diag}(0, \lambda) \otimes \mathbb{1} & \text{für } s \in [-0.0001, 0.0001] \\ F = \mathbb{1} \otimes \mathbb{1} \otimes \text{diag}(0, \lambda) & \text{sonst} \end{cases} \quad (5.5)$$

Ist s kleiner als -0.0001 , so wirkt die Selektion an Site 0 und wir befinden uns in Szenario SLR. ϱ_r ist also gleich 0.0002 und für den Reproduktionsoperator gilt $F = \text{diag}(0, \lambda) \otimes \mathbb{1} \otimes \mathbb{1}$. ϱ_l setzen wir zu $-0.0001 - s$. Es ist somit immer positiv.

Für s aus dem Intervall $[-0.0001, 0.0001]$ befindet sich der selektive Site an Position 1, also zwischen den beiden neutralen. Wir befinden uns somit im LSR-Szenario, so dass sich der Reproduktionsoperator zu $F = \mathbb{1} \otimes \text{diag}(0, \lambda) \otimes \mathbb{1}$ ergibt. Die Rekombinationsraten setzen wir zu $\varrho_l = s + 0.0001$ und $\varrho_r = -s + 0.0001$. Es ist also gewährleistet, dass die Rekombinationsraten stets positiv sind und $\varrho_l + \varrho_r = 0.0002$ gilt.

Falls s größer als 0.0001 ist, so befinden wir uns im LRS-Szenario und es ergibt sich $F = \mathbb{1} \otimes \mathbb{1} \otimes \text{diag}(0, \lambda)$. Für die Rekombinationsraten gilt Ähnliches wie im SLR-Fall, lediglich die Rollen sind vertauscht. Es gilt $\varrho_l = 0.0002$ und $\varrho_r = s - 0.0001$.

Die obigen Definitionen und Fallunterscheidungen sorgen dafür, dass sich der selektive Site mit dem Parameter s linear verschieben lässt. Für stark negatives s ist er auf der *linken* Seite. Mit wachsendem s wird der selektive Site näher an den *linken* neutralen Site geschoben, bis er mit diesem für $s = -0.0001$ zusammenfällt. Für $s = 0$ befindet er sich genau in der Mitte, bis er schließlich für $s = 0.0001$ mit dem rechten Site zusammenfällt. Wächst s nun weiter, so liegt der selektive Site auf der *rechten* Seite und wird zunehmend stark von den beiden neutralen entkoppelt. Dieses Verschieben des selektiven Sites wird in Abbildung 5.3 veranschaulicht.

Wir betrachten im Folgenden das Kopplungsungleichgewicht zwischen dem Allel 1 an Site 0 und dem Allel 1 an Site 1. Es hängt also erneut von s ab, welches Kopplungsungleichgewicht wir genau betrachten müssen, weshalb wir

$$LDE(s) = \begin{cases} T_{\{\frac{1}{2}\}}(z_t)(*, 1, 1) & \text{für } s \in (-\infty, -0.0001) \\ T_{\emptyset}(z_t)(1, *, 1) & \text{für } s \in [-0.0001, 0.0001] \\ T_{\{\frac{3}{2}\}}(z_t)(1, 1, *) & \text{sonst} \end{cases} \quad (5.6)$$

erneut mit Hilfe einer Fallunterscheidung definieren.

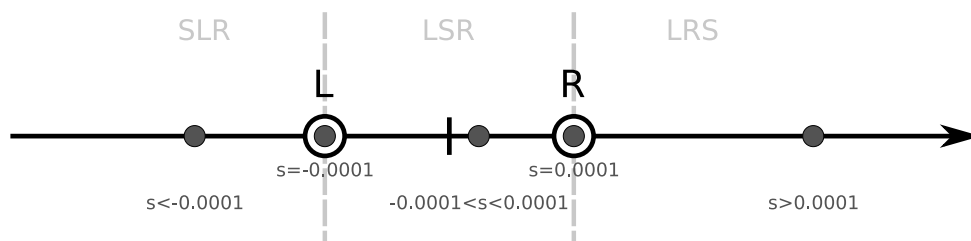


Abbildung 5.3: Der Parameter s verschiebt den selektiven Site (grauer Punkt) entlang der Achse relativ zu den beiden neutralen Sites L und R (ungefüllte Kreise). Es sind unterschiedliche Positionen des selektiven Sites für unterschiedliche Werte von s dargestellt.

Wir wollen nun untersuchen, wie sich die Position einer selektiven bevorzugten Mutation relativ zu zwei neutralen Sites auf deren Kopplungsungleichgewicht auswirkt. Zu diesem Zweck haben wir den Parameter s eingeführt, der einerseits die Position des selektiv wirkenden Sites bestimmt, andererseits aber auch dafür sorgt, dass wir mit $LDE(s)$ das richtige Kopplungsungleichgewicht betrachten.

Die Tatsache, dass wir die Auswirkungen einer selektiv bevorzugten Mutation untersuchen wollen, modellieren wir mit Hilfe der Anfangsbedingungen. Wir gehen davon aus, dass das bevorzugte Allel 1, welches den Vorteil λ besitzt, in der Anfangspopulation zu einem sehr geringen Anteil vorkommt und der Großteil der Population zu Beginn am selektiven Site das Allel 0 hat. Die vorteilhafte Mutation hat also gerade erst stattgefunden. Ohne Beschränkung der Allgemeinheit legen wir fest, dass der Vorteil zu Beginn bei den Gameten vorhanden ist, welche an den beiden neutralen Sites das Allel 1 besitzen. Da die Subpopulation mit dem Allel 0 am selektiven Site zu Beginn nahezu die gesamte Population ausmacht, bestimmt sie effektiv das Gesamtkopplungsungleichgewicht. Wir wählen die Allelhäufigkeiten in dieser Subpopulation so, dass sich das zu untersuchende Kopplungsungleichgewicht zu 0.0242 ergibt.

In den drei verschiedenen Szenarien müssen die drei unterschiedlichen Anfangsbedingungen

$$z_0 = \begin{pmatrix} 0, 0, 0 \\ 0, 0, 1 \\ 0, 1, 0 \\ 0, 1, 1 \\ 1, 0, 0 \\ 1, 0, 1 \\ 1, 1, 0 \\ 1, 1, 1 \end{pmatrix}_0 = \begin{pmatrix} 0.18 \\ 0.2 \\ 0.23 \\ 0.38995 \\ 0 \\ 0 \\ 0 \\ 0.00005 \end{pmatrix}, \begin{pmatrix} 0.18 \\ 0.2 \\ 0 \\ 0 \\ 0.23 \\ 0.38995 \\ 0 \\ 0.00005 \end{pmatrix} \text{ oder } \begin{pmatrix} 0.18 \\ 0 \\ 0.2 \\ 0 \\ 0.23 \\ 0 \\ 0.38995 \\ 0.00005 \end{pmatrix} \quad (5.7)$$

SLR LSR LRS

gewählt werden.

Nun sind wir in der Lage mit dem Parameter s die Anfangsbedingungen passend zu den Parametern der Dynamik zu wählen. s stellt ebenfalls sicher, dass wir mit $LDE(s)$ das richtige Kopplungsungleichgewicht untersuchen.

5.2.2 Unsere Ergebnisse

Kopplungsungleichgewicht zwischen den neutralen Sites (LDE)

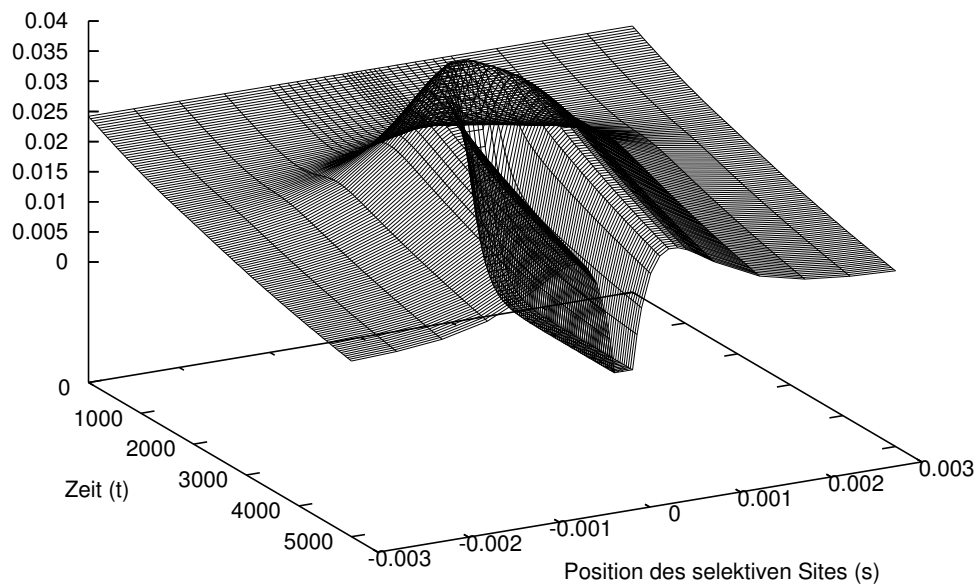


Abbildung 5.4: Dargestellt ist die zeitliche Entwicklung des Kopplungsungleichgewichtes zwischen den neutralen Sites. Diese wird beeinflusst von dem Parameter s , der entlang der y -Achse variiert.

In den Abbildungen 5.4 und 5.5 auf der nächsten Seite ist die zeitliche Entwicklung des Kopplungsungleichgewichtes für unterschiedliche Werte von s dargestellt. Die Werte wurden durch eine numerische Approximation der Rekombinations- und Selektionsdynamik mit Hilfe von *Mathematica* berechnet. Die Zeit entwickelt sich entlang der x -Achse und der Wert des Kopplungsungleichgewichtes ist auf der z -Achse aufgetragen. Die Entwicklung des Kopplungsungleichgewichtes wird durch die Position des selektiven Sites, also durch den Parameter s bestimmt, welchen wir entlang der y -Achse variiieren. Wie schon erwähnt, entspricht ein Wert von 0 genau der Mitte zwischen den beiden neutralen Sites und ein Wert von -0.0001 beziehungsweise 0.0001 dem bezüglich Rekombination untrennbaren Koppeln des selektiven mit dem *linken* beziehungsweise *rechten* neutralen

Kopplungsungleichgewicht zwischen den neutralen Sites (LDE)

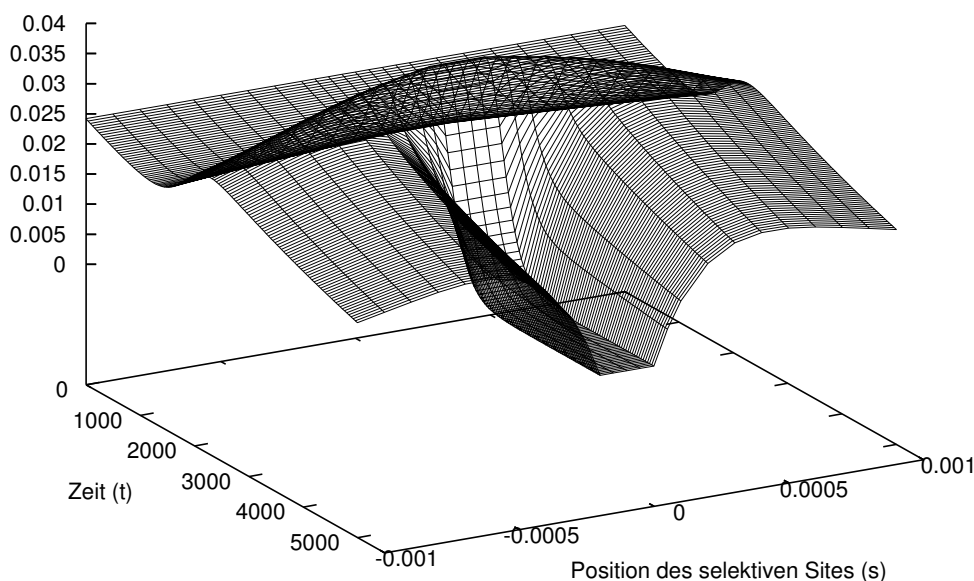


Abbildung 5.5: Es handelt sich hier um dieselbe zeitliche Entwicklung wie in Abbildung 5.4 auf der vorherigen Seite, wir betrachten lediglich einen kleineren Bereich für den Parameter s .

Site. Abbildung 5.4 zeigt das Verhalten für Werte von s aus dem Intervall $[-0.003, 0.003]$. In Abbildung 5.5 sind dieselben Werte des Kopplungsungleichgewichtes für das kleinere Intervall $[-0.001, 0.001]$ dargestellt, um das Verhalten in diesem Bereich besser beobachten zu können.

Es fällt auf, dass die Dynamik des Kopplungsungleichgewichtes ebenfalls symmetrisch ist und zwar um die Position $s = 0$ in der Mitte der beiden neutralen Sites. Im Gegensatz zum 2-Site-System ist dies aber nicht trivial. Stephan beweist diese Symmetrie für seine diskrete Dynamik. Es ist sicherlich auch möglich, diese Symmetrie für die kontinuierliche Dynamik des Kopplungsungleichgewichtes zu beweisen, diese Betrachtungen stellen wir im Rahmen dieser Arbeit jedoch nicht an.

Weiterhin ist zu beobachten, dass der zeitliche Verlauf des Kopplungsungleichgewichtes für alle $s \in [-0.0001, 0.0001]$, also genau für den Bereich LSR, gleich ist. Befindet sich der selektive Site zwischen den beiden neutralen Sites, so ist die genaue Position für das Verhalten des Kopplungsungleichgewichtes irrelevant. Diese Unabhängigkeit von der exakten Position beweist Stephan für den zeitdiskreten Fall ebenfalls. Da sich unsere Beobachtungen mit denen von Stephan weitestgehend decken, gehen wir davon aus, dass

dieser Beweis in unserem Modell ebenfalls möglich ist

Um die Dynamik des Kopplungsungleichgewichtes genauer betrachten zu können, sind in Abbildung 5.6 sechs Kurven der zeitlichen Entwicklung für unterschiedliche Werte von s zu sehen. In diesen Diagrammen ist nicht nur der Verlauf des Kopplungsungleichgewichtes, sondern auch die relative Größe der selektiv bevorzugten Subpopulation sowie die Kopplungsungleichgewichte zwischen den neutralen Sites in der selektiv bevorzugten und selektiv benachteiligten Subpopulation dargestellt.

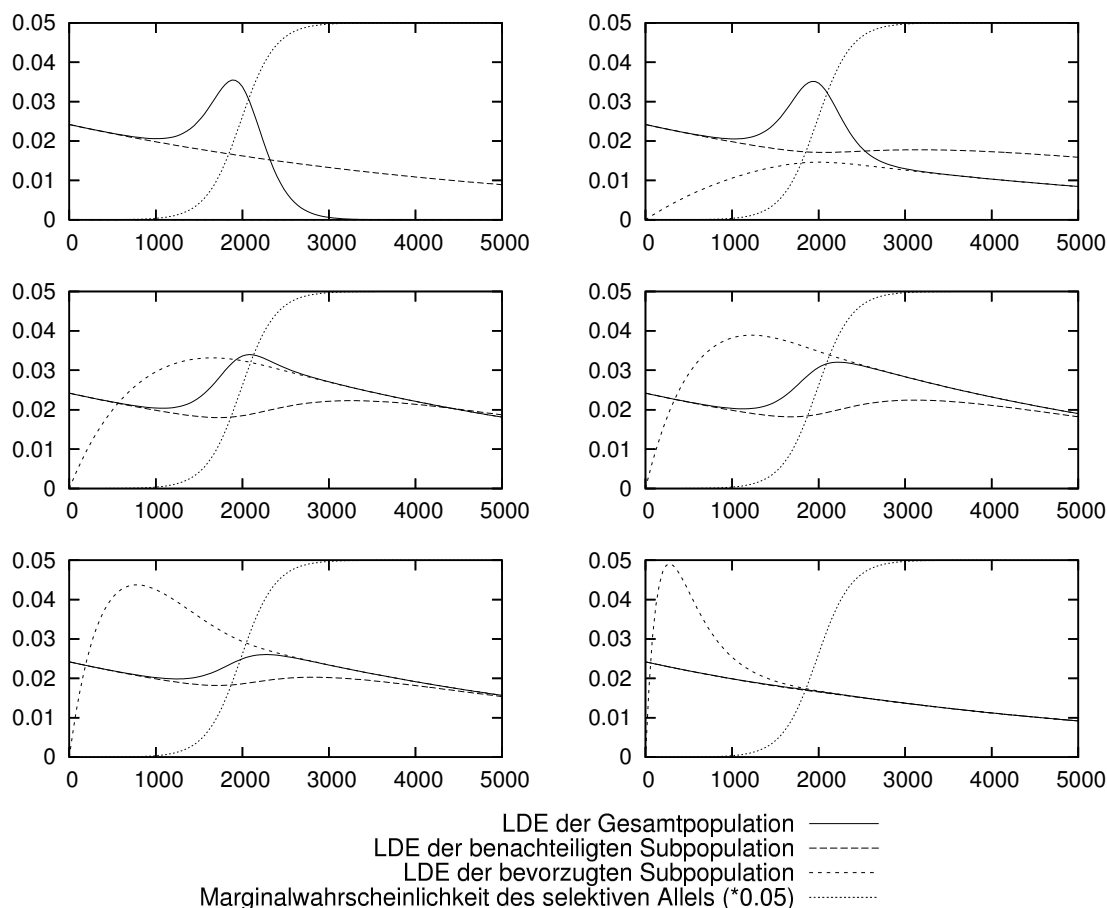


Abbildung 5.6: Für unterschiedliche Werte von s ist zusätzlich zum Verlauf des Kopplungsungleichgewichtes (LDE) der Gesamtpopulation der Verlauf der Kopplungsungleichgewichte der Teilpopulationen mit beziehungsweise ohne Selektionsvorteil dargestellt. Außerdem die Größe der Teilpopulation mit dem Vorteil. Die Werte von s sind von links-oben nach rechts-unten 0.0001, 0.00018, 0.0004, 0.0006, 0.001 und 0.003.

Wir wollen nun anhand der Größe und des Kopplungsungleichgewichtes der Teilpopulationen das Verhalten des Kopplungsungleichgewichtes der gesamten Population erklären.

Es lässt sich beobachten, dass die zeitliche Entwicklung der relativen Größen der Subpopulationen von s unabhängig ist. Zu Beginn besteht die Gesamtpopulation fast ausschließlich aus selektiv benachteiligten Typen. Ab $t = 1000$ wird der Anteil der bevorzugten Teilpopulation zunehmend größer. Bedingt durch die im Verhältnis zur Rekombination starke Selektion steigt die relative Größe daraufhin sprunghaft an und nach einer relativ kurzen Übergangsphase besteht die Gesamtpopulation für $t = 3000$ fast ausschließlich aus selektiv bevorzugten Gameten. Ab diesem Zeitpunkt wirkt effektiv auf die Gesamtpopulation keine Selektion mehr, da alle den Selektionsvorteil besitzen und dadurch keiner den anderen überlegen ist.

Die Verhältnisse der Subpopulationen sind unabhängig von s , die Kopplungsungleichgewichte in den Teilpopulationen jedoch nicht.

Das Kopplungsungleichgewicht in der bevorzugten Population steigt zunächst an. Die Stärke und Dauer dieses Anstiegs ist durch die Entfernung des selektiven Sites von dem äußeren neutralen Site bestimmt. Für $s = 0.0001$ beträgt diese Entfernung null, weshalb es hier einen Anstieg mit Rate 0, sprich keinen wirklichen Anstieg, gibt. Je größer s ist, je weiter also der selektive Site von dem äußeren neutralen entfernt wird, desto steiler wird der Anstieg und desto höher ist das Maximum, welches durchlaufen wird. Mit zunehmendem s verstreicht weniger Zeit, bis das Maximum erreicht wird.

Dieser Anstieg des Kopplungsungleichgewichtes der zu Beginn homogenen Teilpopulation ist dadurch bedingt, dass hier aufgrund der selektiv benachteiligten Teilpopulation, welche in der Anfangsphase der Dynamik noch die Mehrheit in der Gesamtpopulation ausmacht, durch Rekombination neue Typen entstehen und die Homogenität gestört wird. Die starken Abhängigkeiten in den Marginalwahrscheinlichkeiten der Allele, welche in der zu Beginn homogenen Teilpopulation keinen Einfluss auf das Kopplungsungleichgewicht haben, bewirken in der zunehmend inhomogenen Teilpopulation eine starke Abweichung vom Kopplungsungleichgewicht.

Wie bereits erwähnt, ist dieser Anstieg für kleine s langsam und flach, so dass das Maximum dadurch bedingt wird, dass die selektiv bevorzugte Teilpopulation einen immer größer werdenden Anteil an der Gesamtpopulation hat. Die Homogenität kann durch die Rekombination nicht mehr gestört werden, da der Anteil der störenden Typen an der Population geringer wird. Nachdem sich dieser Zustand eingestellt hat, ist nur noch der durch die Rekombination bedingte exponentielle Abfall zu beobachten.

Wird s größer, so ist das Maximum und der darauffolgende Abfall auf eine andere Ursache zurückzuführen, da das Maximum erreicht wird, bevor die bevorzugte Teilpopulation einen signifikanten Teil der Gesamtpopulation ausmacht. Wie bereits erwähnt, ist der Anstieg des Kopplungsungleichgewichtes in der Subpopulation mit dem vorteilhaften Allel durch eine Störung der Homogenität bedingt. Diese Störung sorgt aber nur in der Anfangsphase für einen relativ starken Anstieg des Kopplungsungleichgewichtes. Die *eindringenden* Typen etablieren sich mit der Zeit, so dass sich das Kopplungsungleichgewicht der verhältnismäßig kleineren, bevorzugten Subpopulation nach Durchschreiten eines Maximums dem Kopplungsungleichgewicht der größeren Teilpopulation von oben annähert.

Das Überschießen des Kopplungsungleichgewichtes lässt sich, wie in Abbildung 5.7 aufgezeigt wird, auch ohne Mitwirken von Selektion beobachten.

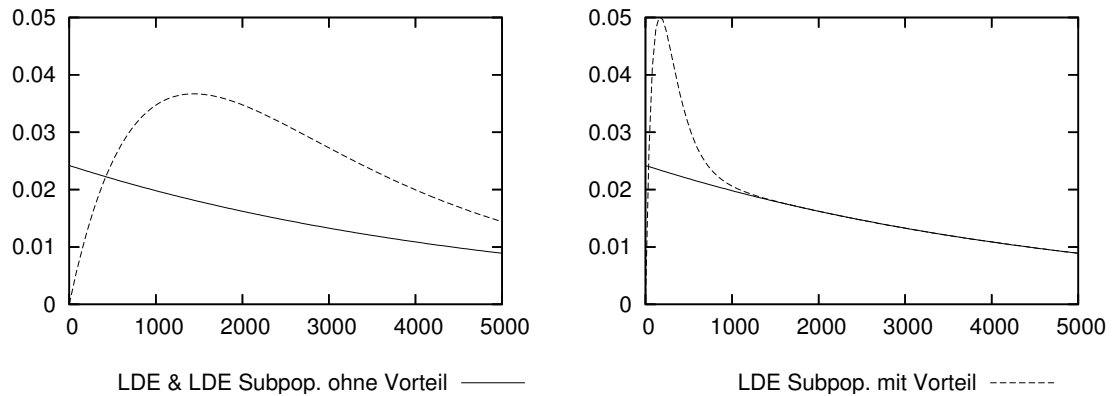


Abbildung 5.7: Auf der y -Achse sind die Kopplungsungleichgewichte (LDE) der Subpopulationen abgebildet, welche sich mit der Zeit entlang der x -Achse entwickeln. Die Parameter der Dynamik sind dieselben wie in Stephans Szenario, lediglich λ ist gleich Null gesetzt. Es gibt also keine Selektion. Die linke Grafik wurde mit $s = 0.0005$ und die rechte mit $s = 0.005$ erstellt.

Die beiden beschriebenen Kräfte bestimmen das Verhalten des Kopplungsungleichgewichtes in der selektiv bevorzugten Teilpopulation. Wie wir aufgezeigt haben, liegt mehr Gewicht auf der einen oder der anderen Kraft, je nachdem wie weit der selektive Site von den neutralen entfernt ist.

Das Kopplungsungleichgewicht der benachteiligten Subpopulation entwickelt sich wesentlich träger, da die Änderungen der Gametenhäufigkeiten im Verhältnis zur Gesamtgröße der Subpopulation nicht so stark sind wie in der bevorzugten Teilpopulation. Zu Beginn ist der durch Rekombination bedingte exponentielle Abfall zu beobachten. Im Laufe der Zeit wird ein immer größerer Teil der Population durch die selektiv bevorteilte Subpopulation ausgemacht, so dass sich das Kopplungsungleichgewicht der benachteiligten Population aufgrund der Rekombination langsam dem Wert der bevorzugten Teilpopulation annähert. Nach Erreichen dieses Wertes ist erneut ein exponentielles Abklingen zu beobachten.

Das Kopplungsungleichgewicht der gesamten Population setzt sich nun aus denen der Teilpopulation zusammen. Für jeden Wert von s ist zu beobachten, dass der Wert zunächst dem der benachteiligten Population entspricht. Während der Übergangsphase, in welcher die selektiv bevorzugte Subpopulation die Oberhand gewinnt, geht auch der Wert des Kopplungsungleichgewichtes der Gesamtpopulation über auf den Wert der bevorzugten Teilpopulation. Diese Beobachtung ist klar, da der Wert des Kopplungsungleichgewichtes der Gesamtpopulation außerhalb der Übergangsphase immer dem der Subpopulation entspricht, welche die Gesamtpopulation dominiert.

Für $s = 0.003$ ist diese Übergangsphase jedoch effektiv nicht zu erkennen, da das Kopplungsungleichgewicht der bevorzugten Subpopulation bereits vorher denselben Wert wie das der benachteiligten angenommen hat.

Für die Werte $s = 0.001$ und $s = 0.0006$ ist der Wert des Kopplungsungleichgewichtes in der bevorzugten Population während und am Ende der Übergangsphase größer als der Wert in der benachteiligten, so dass das Gesamtkopplungsungleichgewicht während der Übergangsphase langsam ansteigt, bis es mit dem der bevorzugten Subpopulation zusammenfällt.

Setzen wir $s = 0.0004$, so ist zu beobachten, dass am Ende der Übergangsphase der Wert des Kopplungsungleichgewichtes ebenfalls mit dem der bevorzugten Subpopulation zusammenfällt, jedoch übersteigt er diesen zwischenzeitlich sogar.

Dieses Übersteigen kann dadurch erklärt werden, dass die Teilpopulationen strukturell unterschiedlich sind. Zu Beginn ist die selektiv bevorzugte Subpopulation homogen. Die große Rekombinationsrate zwischen den neutralen Sites und dem selektiven in den vorherigen Fällen sorgt jedoch dafür, dass die Subpopulation in der Übergangsphase schon ausreichend inhomogen ist, um einen glatten Übergang zwischen den beiden Kopplungsungleichgewichten zu gewährleisten. Wird s und damit die Rekombinationsrate kleiner, so kann bis zur Übergangsphase keine ausreichende Inhomogenität hergestellt werden und das Gesamtkopplungsungleichgewicht liegt während der Übergangsphase nicht mehr zwischen den Kopplungsungleichgewichten der Teilpopulationen, sondern darüber.

Dieser Effekt wird mit abnehmender Entfernung des selektiven Sites für $s = 0.00018$ noch größer, und ist schließlich für $s = 0.0001$ am stärksten ausgeprägt.

Das Kopplungsungleichgewicht der selektiv bevorzugten Population steigt wie bereits erwähnt mit einer immer geringer werdenden Rate an, bis es für $s = 0.0001$ sogar bei einem Wert von 0 bleibt. Das Gesamtkopplungsungleichgewicht nimmt nichtsdestotrotz in der Übergangsphase einen weit über den beiden Teilkopplungsungleichgewichten liegenden Maximalwert an, fällt dann aber wieder auf den Wert der bevorzugten Population, also 0, zurück.

Da die hier beschriebenen Verhaltensweisen für unterschiedliche Werte von s stetig ineinander übergehen, ergeben sich Abbildung 5.4 auf Seite 56 und Abbildung 5.5 auf Seite 57 als Gesamtbild. Der Hitchhiking-Effekt bewirkt also zunächst ein Ansteigen des Kopplungsungleichgewichtes zwischen den beiden neutralen Sites, gefolgt von einem Rückgang. Liegt der selektive Site zwischen den beiden neutralen, so erfolgt ein rapider Abfall auf einen Wert von 0. Ist der selektive Site in unmittelbarer Nachbarschaft, so fällt das Kopplungsungleichgewicht ebenfalls schnell ab, jedoch nicht ganz auf 0. Der verbleibende Rest verschwindet dann wieder durch die Rekombination exponentiell. Mit zunehmender Entfernung beginnt dieser exponentielle Abfall bereits an dem eingenommenen Maximum, welches immer geringer wird, bis es schließlich für große s ganz verschwindet und die einzige sichtbar wirkende Kraft die Rekombination zwischen den beiden neutralen Sites ist.

6 Diskussion

Wir haben in Kapitel 2 ein mathematisches Modell entwickelt, welches die zeitliche Entwicklung des genetischen Materials einer Population beschreibt, auf welches die in Kapitel 1 vorgestellten evolutionären Kräfte Selektion und Rekombination wirken. Mit diesem Modell waren wir in der Lage Anfangswertprobleme zu formulieren, die sowohl das getrennte als auch das gleichzeitige Wirken der Kräfte beschreiben.

Daraufhin widmeten wir uns zunächst der Motivation von Korrelationsfunktionen zwischen den Sites eines Genoms, den Kopplungsungleichgewichten. Hier gaben wir einen kurzen Überblick über vergangene Arbeiten, in denen die Kopplungsungleichgewichte bereits behandelt wurden, und zeigten einige ihrer Eigenschaften auf, aufgrund derer sie für Untersuchungen zur Rekombinationsdynamik von Interesse sind.

Nach diesem kleinen Exkurs wendeten wir uns in Kapitel 4 der Lösung der in Kapitel 2 hergeleiteten Differentialgleichungen für die Selektion und Rekombination zu. Wir leiteten explizite Lösungen für die Dynamiken her, in denen entweder nur Selektion oder nur Rekombination wirken.

Ursprünglich beabsichtigten wir daraufhin aufzuzeigen, wie sich diese Teillösungen, laut Baake [1], zu einer Lösung der Dynamik verbinden lassen, in welcher Rekombination und Selektion zur gleichen Zeit wirken. Es stellte sich jedoch heraus, dass sich die Dynamiken nicht nach dem beschriebenen Vorgehen verbinden lassen, und wir führten ein Gegenbeispiel an, um dies zu verdeutlichen.

Im Anschluss daran leiteten wir neue Lösungen für zwei Spezialfälle der Rekombinations- und Selektionsdynamik her, welche wir verwendeten, um zum einen die grundsätzliche Idee des Hitchhiking-Effektes zu veranschaulichen und zum anderen Ergebnisse von Stephan [19] teilweise zu reproduzieren. Sie untersuchten die Auswirkungen des Hitchhiking-Effektes auf das Kopplungsungleichgewicht zwischen zwei neutralen Sites.

Da die explizite Lösung der Dynamiken nicht verwendet werden konnte, um konkrete Werte zu berechnen, verwendeten wir zum Reproduzieren eine numerische Approximation der Dynamik.

Es gelang uns, die gleichen grundlegenden Beobachtungen von Stephan in unserem Modell anzustellen und diese in Ansätzen zu erklären. Allerdings bewies Stephan einige der beobachteten Eigenschaften der Dynamik, was in unserem Modell ebenfalls möglich scheint, da wir die Eigenschaften ebenfalls beobachteten. Es steht jedoch noch aus, hier konkrete Beweise anzugeben.

Danksagung

Ich danke meinen beiden Betreuern Ellen und Michael Baake, die mir nicht nur dieses Thema ermöglichten, sondern mir ebenfalls mit Rat und Tat zur Seite standen, wenn ich auf Probleme stieß.

Des Weiteren möchte ich den aktuellen, ehemaligen und zukünftigen Mitgliedern der Arbeitsgruppe Biomathematik für die freundliche Arbeitsatmosphäre bedanken. Ein besonderer Dank gilt hier Inke Hildebrandt, die mich beim Einarbeiten in das Thema tatkräftig unterstützt hat.

Weiterhin bedanke ich mich bei Peter Husemann und Marcel Martin, die einen großen Teil zum optischen Gelingen dieser Arbeit beigetragen haben.

Ferner danke ich Melanie Darmawan, Dries Guth, Thiemo Hustedt, Marcel Martin und Anja Steinrücken für das Korrekturlesen von Vorabversionen und viele nützliche Tipps.

Ganz besonders danke ich meiner Freundin, Melanie Darmawan, der es gefallen wird, dass sie hier von zwei Kommata eingerahmt erwähnt wird, für ihre Unterstützung bei der Anfertigung dieser Arbeit. Nachdem ich den ein oder anderen langen Abend mit dieser Arbeit zugebracht habe, und normalsterbliche Pizzabuden schon längst die Nahrungsausgabe verweigerten, hatte sie immer noch eine Packung Schnellnudeln für meine hungrigen Magen im Schrank.

Ein letzter Dank geht an meine Familie und Freunde, die mich auf meinem Weg begleitet und auf diese Weise ihren Beitrag zu dieser Arbeit geleistet haben.

Es sei an dieser Stelle auch noch Tuna, Peter, Greebo und Wanda, sowie den Schafen gedankt.

Literaturverzeichnis

- [1] M. Baake und E. Baake. *An exactly solved model for mutation, recombination and selection. Can. J Math*, 55:Seiten 3–41, 2003.
- [2] N. H. Barton und M. Turelli. *Natural and sexual selection on many loci. Genetics*, 127(1):Seiten 229–255, Jan 1991.
- [3] J. H. Bennett. *On the theory of random mating. Ann Eugen*, 18(4):Seiten 311–317, Mar 1954.
- [4] C. Berge. *Principles of Combinatorics*. Academic Press, 1971. ISBN 0120897504.
- [5] J. M. Braverman, R. R. Hudson, N. L. Kaplan, C. H. Langley und W. Stephan. *The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics*, 140(2):Seiten 783–796, Jun 1995.
- [6] R. Bürger. *The Mathematical Theory of Selection, Recombination, and Mutation*. John Wiley & Sons, 2000. ISBN 0471986534.
- [7] F. B. Christiansen. *Population Genetics of Multiple Loci*. John Wiley & Sons, 2000. ISBN 0471979791.
- [8] K. J. Dawson. *The decay of linkage disequilibrium under random union of gametes: how to calculate Bennett's principal components. Theor Popul Biol*, 58(1):Seiten 1–20, Aug 2000.
- [9] H. Geiringer. *On the Probability Theory of Linkage in Mendelian Heredity. Annals of Math. Stat.*, 15:Seiten 25–57, 1944.
- [10] J. Hofbauer. *The selection mutation equation. J Math Biol*, 23(1):Seiten 41–53, 1985.
- [11] N. L. Kaplan, R. R. Hudson und C. H. Langley. *The "hitchhiking effect" revisited. Genetics*, 123(4):Seiten 887–899, Dec 1989.
- [12] R. Knippers. *Molekulare Genetik*. Thieme, Stuttgart, 2006. ISBN 3134770091.
- [13] R. C. Lewontin und K.-I. Kojima. *The Evolutionary Dynamics of Complex Polymorphisms. Evolution*, 14:Seiten 458–472, 1960.
- [14] Y. I. Lyubich. *Mathematical Structures in Population Genetics*. Springer, Berlin, 1998. ISBN 3540533370.

- [15] J. Maynard Smith und J. Haigh. *The hitch-hiking effect of a favourable gene*. *Genet. Res., Camb.*, 23:Seiten 23–35, 1974.
- [16] T. Nagylaki. *Introduction to Theoretical Population Genetics*. Springer, 1991. ISBN 0387533443.
- [17] T. Nagylaki. *The evolution of multilocus systems under weak selection*. *Genetics*, 134(2):Seiten 627–647, Jun 1993.
- [18] M. Slatkin. *On treating the chromosome as the unit of selection*. *Genetics*, 72(1):Seiten 157–168, Sep 1972.
- [19] W. Stephan, Y. S. Song und C. H. Langley. *The hitchhiking effect on linkage disequilibrium between linked neutral sites*. *Genetics*, 172:Seiten 2647–2663, Feb 2006. doi:10.1534/genetics.105.050179.
- [20] C. J. Thompson und J. L. McBride. *On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules*. *Math Biosci.*, 21:Seiten 127–142, 1974.
- [21] W. Walter. *Gewöhnliche Differentialgleichungen*. Springer, Berlin, 2000. ISBN 3540676422.
- [22] S. Wolfram. *The Mathematica Book*. Wolfram Media, 1999. ISBN 1579550045.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig bearbeitet und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Bielefeld, den 12. Juli 2006