

Mit Mathematik und Computer auf Entdeckungs- reisen in der Evolutionsbiologie

Peter SCHUSTER ML (Wien, Österreich)

Mit 10 Abbildungen und 2 Tabellen



Zusammenfassung

Mathematische Modellierung hat einen Vorteil im Unterschied gegenüber experimentellen Techniken: Ein durch Beweis oder korrekt durchgeführte Computersimulation erhaltenes Resultat ist gültig und bedarf keiner Absicherung durch Wiederholung oder Verfeinerung der Messmethode. In Frage gestellt werden können lediglich die Voraussetzungen des Ansatzes oder die Interpretation der Ergebnisse. In der Vergangenheit wurde der Computer zumeist eingesetzt, um genauere Auskünfte über einen in groben Zügen bereits bekannten Sachverhalt zu bekommen. Die zumeist undurchschaubare Komplexität der meisten, heute aktuellen Probleme legt es nahe und die numerische Leistung der heutigen Rechner ermöglicht es, mit dem Computer auf Entdeckungsreisen zu gehen und Unbekanntes zu finden.

Der geschilderte Sachverhalt wird an Hand der Evolutionsbiologie illustriert. Den Anfang machen Beispiele aus der Geschichte, u. a. die durch den Mathematiker und Populationsgenetiker FISHER geglückte Vereinigung des Darwinischen Prinzips der natürlichen Auslese mit den Vererbungsregeln MENDELS, sie erfolgte zwanzig Jahre vor der Vollendung der synthetischen Theorie, und das Lotka-Volterra-Modell, das den Anfang der theoretischen Ökologie markiert. Alan TURING stellte Gleichungen für die chemische Musterbildung auf und vermutete, dass sie bei der entwicklungsbiologischen Morphogenese eine Rolle spielen. Die Entdeckung der großen Vielfalt und Variabilität der Muster in allen Bereichen der Biologie war dem Computereinsatz vorbehalten.

Eine reaktionskinetische Formulierung der Theorie der Evolution auf molekularer Ebene durch Manfred EIGEN gestattet eine Beschreibung der Rolle von Mutationen bei der Evolution von Molekülen, Viren und Bakterien. Erst durch Computereinsatz wurde es möglich, den durch zu häufige Mutationen bedingten Zusammenbruch der Vererbung zu entdecken, der große Ähnlichkeit mit einer Phasenumwandlung aufweist. Landschaften wurden schon 1930 von WRIGHT als Metapher zur Erklärung evolutionärer Vorgänge benutzt. Nur durch den Einsatz leistungsfähiger Computer konnten die von den Strukturen der Biomoleküle gebildeten Landschaften erforscht werden. Die Entdeckung der Ursachen für Neutralität und Streuung bei Evolutionsvorgängen folgte zwanglos daraus. Spieltheoretische Ansätze und kinetische Differentialgleichungen ermöglichten eine Diskussion der evolutionären Ursprünge von Altruismus und Kooperation.

Abstract

Mathematical modeling has one advantage over experimental techniques: a result obtained through proof or correctly performed computer simulation is valid and need not be reaffirmed by repeating or refining the method of measurement. The only things that can be questioned are the conditions of the approach and the interpretation of the findings. In the past, a computer was often used to obtain more precise information about generally known facts. The unintelligible complexity of most current problems demands this and the numeric capacities of today's computers allow one to go on a voyage of discovery with the computer in search of the unknown.

This is illustrated by evolutionary biology. We start with examples from history, including the successful union by the mathematician and population geneticist FISHER of DARWIN'S principles of natural selection with Mendelian inheritance which occurred twenty years before the synthetic theory was completed, and the Lotka-Volterra model which marked the beginning of theoretical ecology. Alan TURING presented equations for chemical pattern formation and assumed that they played a role in developmental biology's morphogenesis. The discovery of the large diversity and variability of the patterns in all areas of biology was reserved for the computer.

A reaction kinetics theory of molecular level evolution by Manfred EIGEN allows us to describe the role of mutation during the evolution of molecules, viruses and bacteria. Only with the aid of a computer is it possible to discover the breakdown in inheritance through too frequent mutation, which is very similar to phase transition. Landscapes were already used by WRIGHT in 1930 as a metaphor for explaining evolutionary processes. Only by using a high-performance computer could the landscapes formed by the structures of the bio-molecules be studied. The discovery of the causes of neutrality and diversification during evolutionary processes was the natural result. Applying game theory and kinetic differential equations made discussion on evolutionary origins of altruism and cooperation possible.

1. Welchen Dienst leistet die Mathematik bei der Naturbeschreibung?

Mit den Anfängen der neuzeitlichen Astronomie und Physik rückt die mathematische Naturbeschreibung ins Zentrum der wissenschaftlichen Forschung. Galileo GALILEIS berühmtes Zitat, „[...] Die Wissenschaft (Philosophie) steht in diesem großen Buch geschrieben, dem Universum, das unserem Blick ständig offen liegt. Aber das Buch ist nicht zu verstehen, wenn man nicht zuvor die Sprache erlernt und sich mit den Buchstaben vertraut gemacht hat, in denen es geschrieben ist. Es ist in der Sprache der Mathematik geschrieben, und deren Buchstaben sind Kreise, Dreiecke und andere geometrische Figuren, ohne die es dem Menschen unmöglich ist, ein einziges Wort davon zu verstehen; ohne diese irrt man in einem dunklen Labyrinth herum. [...]“¹ bringt dies unmissverständlich zum Ausdruck. Die Physik hat sich in der Folge ganz nach dem Wunsche, mit Hilfe der Mathematik im großen Buch zu lesen, entwickelt. Dies wird in einem zweiten, allgemein bekannten und fälschlich auch GALILEI zugeschriebenen Zitat (KLEINERT 2009) zur Forderung erhoben: „[...] Wer naturwissenschaftliche Fragen ohne Hilfe der Mathematik lösen will, unternimmt Undurchführbares. Man muss alles messen, was messbar ist, und (versuchen) messbar (zu) machen, was noch nicht messbar ist. [...]“ Empirie, Quantifizierung und Mathematisierung der Naturbeobachtung haben den Triumphzug der physikalischen Wissenschaften initiiert und überaus erfolgreich in den folgenden Jahrhunderten begleitet.

Mathematisierung einer Naturbeobachtung bedeutet immer Reduktion eines komplexen Geschehens, bedeutet beschränken auf das für das Erkennen von Zusammenhängen Wesentliche. Stellen wir uns vor, Isaac NEWTON (1686) wäre nicht von der Idee ausgegangen, ein möglichst einfaches, allgemein gültiges Fallgesetz zu finden, sondern hätte einen Katalog aufgestellt über das Fallen von Federn, Papierfliegern, Bumerangs, Fallschirmspringern, Äpfeln, und Steinen. Wir wissen nicht, was herausgekommen wäre – das Fallgesetz wäre es nicht gewesen. Ohne Zweifel sind physikalische Systeme wesentlich einfacher als biologische, aber Probleme bereiten sie auch. Ein gutes Beispiel ist die Wettervorhersage. Niemand wird bezweifeln, dass Wetter ein physikalisches Phänomen der Erdatmosphäre ist. Vor fünfzig Jahren entsprach dessen ungeachtet der tägliche Wetterbericht mehr einer, der Volksbelustigung dienenden Lotterie als einer ernststen wissenschaftlichen Prognose. Dies hat sich geändert, denn die Zahl der Messstationen wurde gewaltig erhöht, und verfeinerte Wettermodelle, die auf einer Weiterentwicklung der Mathematik und massivem Einsatz von Computern basieren, haben die Vorhersage entscheidend verbessert. Nichtsdestoweniger gibt es (zumindest zurzeit) nicht lösbare Probleme: Trotz einer perfekten Theorie für die Entstehung von Hurrikans ist es beispielsweise nicht möglich vorherzusagen, wann und wo ein Hurrikan entstehen wird. Die Physik der Atmosphäre gibt die Bedingungen vor: Eine Wasseroberflächentemperatur von 27 °C oder mehr und atmosphärische Druckdifferenzen sind für die Entstehung von Wirbelstürmen unerlässlich, aber Ort und Zeit der Ausbildung des Wirbelsturmes sind von unkontrollierbaren und unmessbaren kleinen Schwankungen abhängig, welche durch den Mechanismus der Bildung des Hurrikans verstärkt werden.

Die Mathematik leistet drei ganz wesentliche Beiträge zu den Erfolgen der Naturwissenschaften:

- Eine gelungene Beweisführung eines Theorems bietet – im Rahmen der Gültigkeit des verwendeten Modells – die endgültige, interpretationsfreie Antwort auf die gestellte Frage;

1 GALILEI 1896, p. 232. Transcript und möglichst wortgetreue deutsche Übersetzung von ‚Der Prüfer‘ 1623.

- eine empirisch bestätigte Quantifizierung eines Sachverhaltes erlaubt präzise Vorhersagen; und
- das mathematische Modell zeigt die Beschränkungen der Vorhersagbarkeit auf, wenn solche aus prinzipiellen Gründen gegeben sind.

Die ersten beiden Beiträge zusammengenommen – Interpretationsfreiheit und Vorhersagbarkeit – ermöglichten und ermöglichen die mannigfaltigen, spektakulären Anwendungen der Naturwissenschaft in den verschiedenen technischen Disziplinen. Es war der Physik des 20. Jahrhunderts vorbehalten, natürliche Beschränkungen der Vorhersagbarkeit aufzuzeigen: Die Quantenmechanik erbrachte den Beweis, dass der Präzision der Beobachtungen grundsätzliche Grenzen gesetzt sind, und die komplexe Dynamik mancher natürlicher Systeme führt zu so großen Empfindlichkeiten gegenüber Anfangsbedingungen und unkontrollierbaren Schwankungen, dass genaue oder einigermaßen verlässliche Vorhersagen nur für eine beschränkte Zeit möglich sind.

In der zweiten Hälfte des vorigen Jahrhunderts erfuhr die auf den beiden Säulen Theorie und Experiment ruhende Naturbeschreibung eine Ergänzung durch die Entwicklung der elektronischen Rechenanlagen. Neben mathematischer Modellierung und Modellanalyse etablierte sich die Computersimulation als ein neues, mittlerweile unentbehrliches Verfahren zur Aufklärung komplexer Sachverhalte. Man kann davon sprechen, dass die Naturwissenschaft ein drittes Standbein erhalten hat (Abschnitt 4.).

2. Mathematik und Computerrechnen in der Biologie

Seit dem Beginn der Neuzeit entwickelten sich Mathematik und Physik, wie gesagt, gemeinsam und unter gegenseitiger Befruchtung. Ein stimulierendes Verhältnis zwischen Mathematik und Biologie im Allgemeinen, oder eines zwischen Mathematik und Evolutionsbiologie, gab es hingegen bis zur zweiten Hälfte des 20. Jahrhunderts nicht. Dabei hat es durchaus einige interessante Versuche von Mathematikern gegeben, biologische Fragestellungen zu analysieren, aber diese wurden von den Hauptrichtungen in der Biologie ignoriert oder belächelt. Einige ausgewählte Beispiele werden im nächsten Abschnitt beschrieben. Insbesondere fanden die Arbeiten Gregor MENDELS über die statistisch-mathematische Analyse der Daten zur Vererbung von Merkmalen lange Zeit keine Beachtung durch Evolutionsbiologen, obwohl diese einen Mechanismus der Vererbung dringend gebraucht hätten. Die Anfänge der Biostatistik – manchmal auch Biometrik genannt – datieren erst um die Wende vom 19. zum 20. Jahrhundert. Man darf in diesem Zusammenhang nicht vergessen, dass die Disziplin der mathematischen Statistik auch nicht älter ist (Abschnitt 3.3). DARWINS natürliche Auslese wurde mit den Mendelschen Vererbungsregeln durch die Begründer der Populationsgenetik, Ronald FISHER, John Burdon Sanderson HALDANE und Sewall WRIGHT, in einem mathematischen Modell zusammengeführt. Trotzdem diese erfolgreiche Vereinigung schon 1930 gelang, ließ die Fertigstellung der „Synthetischen Theorie der Evolution“ noch zwanzig Jahre auf sich warten. Bemerkenswert ist auch, dass die großen Werke der Biologie, z. B. Charles DARWINS *Origin of Species* (DARWIN 1859) und Ernst MAYRS *Growth of Biological Thought* (MAYR 1982), keine einzige mathematische Formel enthalten.

Die Entwicklung einer breiter angelegten theoretischen Biologie, welche über die Populationsgenetik hinausgeht, begann erst nach dem Zweiten Weltkrieg. Als einen Anfang kann

man Alan TURINGs Arbeiten zur chemischen Musterbildung (TURING 1952) ansehen. Zwanzig Jahre später formulierten Alfred GIERER und Hans MEINHARDT (GIERER und MEINHARDT 1972, MEINHARDT 1982, GORDON und BELLOUSOV 2006) ein molekularbiologisches Modell der biologischen Morphogenese, wobei anstelle einer mathematischen Analyse systematische Computeruntersuchungen in Form der numerischen Integration partieller Differentialgleichungen traten. Die Experimente von Sol SPIEGELMAN zur Evolution von RNA-Molekülen im Reagenzglas (MILLS et al. 1967, SPIEGELMAN 1971, JOYCE 2007) wurden etwa gleichzeitig durchgeführt mit Manfred EIGENS Entwicklung einer Theorie der Evolution auf molekularer Basis (EIGEN 1971, EIGEN und SCHUSTER 1977, 1978a, b). Die Modellierung molekularer Evolutionsprozesse begann von Anfang an mit einer Kombination mathematischer Analysen und numerischer Integration am Computer.

Im selben Jahr, in dem TURING seine Arbeit zur Morphogenese veröffentlichte, erschien Alan HODGKINS und Andrew HUXLEYS (1952) bahnbrechende Publikation zur Leitung elektrischer Impulse in den Axonen der Nervenzellen oder Neuronen. Die empirisch begründete und auf den Kenntnissen der elektrischen Leitung in mit Elektrolyt gefüllten Röhren aufbauende Hodgkin-Huxley-Gleichung ist eine nichtlineare partielle Differentialgleichung, welche es gestattet, durch numerische Integration die Gestalt und die Ausbreitungsgeschwindigkeit des propagierenden Pulses gleichzeitig zu bestimmen. Man kann heute mit Recht behaupten, dass die Hodgkin-Huxley-Gleichung am Beginn einer theoretischen Neurobiologie steht. Die Entwicklung der Neurobiologie war in jeder Hinsicht rasant und beeindruckend. (Als ein einführendes und umfassendes Lehrbuch sei das Werk KANDEL et al. 2000 genannt.) Neuronen sind insbesondere in den Gehirnen von Säugetieren zu stark verknüpften Netzwerken zusammengeschlossen. Reale neuronale Netzwerke sind aus zwei Gründen schwer zu modellieren:

- Die Zahl der Neuronen ist sehr groß und dies auch in verhältnismäßig einfachen Gehirnen; und
- die Schaltmuster, in welchen aktivierende und inhibierende Neuronen zusammenwirken, sind überaus komplex.

Ein stark vereinfachtes und auf die primitivsten *Input-Output*-Relationen reduziertes Neuron wurde von John HOPFIELD (1982) als Basis für die Untersuchung von diversen Netzwerkschaltungen eingeführt. Aus derartigen Hopfield-Netzwerken entwickelte sich die erfolgreiche Simulationsmethode der neuronalen Netzwerke.

Der berühmte Strukturvorschlag für die DNA (WATSON und CRICK 1953) und die ersten beiden Strukturaufklärungen von Proteinen (Myoglobin: KENDREW et al. 1960, Hämoglobin: PERUTZ et al. 1960) begründeten die Strukturbiologie und eröffneten ein neues Gebiet in der theoretischen Biologie: die Computer gestützte Berechnung von Biopolymerstrukturen. Die Strukturanalyse einsträngiger RNA-Moleküle hinkte etwas nach und nahm ihren Anfang mit den Strukturen der tRNAs (Sekundärstruktur: HOLLEY et al. 1965, 3D-Struktur: KIM et al. 1974, ROBERTUS et al. 1974). In Chemie und Physik wurden und werden quantenchemische *Ab-initio*-Rechenverfahren erfolgreich zur Strukturbestimmung verwendet (siehe z. B. RAMACHANDRAN et al. 2008). Für *Ab-initio*-Rechnungen sind die Biomoleküle (zumindest zurzeit) noch viel zu groß. Andere Methoden, welche entweder von empirischen, zumeist thermodynamischen Daten ausgehen und ebenso wie die *Ab-initio*-Verfahren die molekulare Struktur durch Minimieren der freien Energie bestimmen oder aus den Strukturen von sequenzverwandten Molekülen auf die Struktur des zu untersuchenden Moleküls schließen (MARTÍ-RENOM et al. 2000), sind erfolgreich.

Die Anwendung der von SANGER (SANGER et al. 1977) und GILBERT (MAXAM und GILBERT 1977) entwickelten Techniken für die DNA-Sequenzierung führte zu einer Revolution in der molekularen Genetik und hatte entscheidenden Einfluss auf die Evolutionsbiologie: DNA-Sequenzen konnten rasch und kostengünstig bestimmt werden, ganze Genome wurden zugänglich, und die Rekonstruktion genetischer Stammbäume aus den Sequenzdaten heute lebender Organismen wurde auf eine neue und solide Basis gestellt. Gleichzeitig waren die Computerwissenschaftler vor die Herausforderung gestellt, ungeheure Datenmengen in Sequenzvergleichen rasch und verlässlich zu verarbeiten. Die ersten Algorithmen basierten auf dynamischer Programmierung (SMITH und WATERMAN 1981 sowie NEEDLEMAN und WUNSCH 1970) und sind in ihrem Ablauf gut zu verstehen. Sie waren aber sehr bald nicht mehr ausreichend und wurden von immer schnelleren und trickreicheren Verfahren abgelöst (MOUNT 2004). Ein neues Fach, die Bioinformatik, war entstanden. Die rasante Zunahme des in den Datenbanken zur Verfügung stehenden Sequenzmaterials hat sich bis heute fortgesetzt, und es ist kein Ende abzusehen. Neue, noch effizienter und preisgünstiger arbeitende Techniken der DNA-Sequenzanalyse wurden entwickelt, und die Zahl der vollständig sequenzierten Genome steigt weiterhin dramatisch an.

Auf die Entschlüsselung der Genome folgte die Aufklärung der Gesamtheit aller Proteine, die Entschlüsselung des Proteoms (PHIZICKY et al. 2003, NEET und LEE 2002 und YARMUSH und JAYARAMAN 2002). Die Funktionen der Gene und Proteine im Gesamtkontext einer Zelle oder eines Organismus können nur durch Analyse der Prozesse verständlich gemacht werden. Die Zahlen der molekularen *Player* in Zellen erreichen in den einfachsten Fällen viele Tausende. Es ist die Notwendigkeit entstanden, die Dynamik von genetischen und metabolischen Netzwerken mit vielen Tausend Knoten und noch mehr Wechselwirkungen zu analysieren. Hinzu kommen noch die überaus komplexen und verschachtelten Netzwerke der Zellkommunikation durch molekulare Signale. Eine Modellierung derartiger, wahrhaft monströser dynamischer Gebilde ist – wenn überhaupt – nur mit Hilfe von Computersimulation möglich. Das neue Gebiet der theoretischen Systembiologie widmet sich mit großem Enthusiasmus dieser Aufgabe, an deren Ende der simulierte menschliche Organismus steht. Sollte diesem Vorhaben ein echter Erfolg beschieden sein, so sind seine medizinischen Anwendungsmöglichkeiten gewaltig.

3. Einige für die Evolutionsbiologie wichtige Beispiele

In diesem Abschnitt sollen einige historische Anwendungen der Mathematik in der Biologie Erwähnung finden, welche wesentliche Beiträge zur Basis der heutigen theoretischen Evolutionsbiologie geleistet haben. Um die logischen Zusammenhänge der einzelnen Beispiele besser herausarbeiten zu können, wird auf die genaue historische Reihenfolge verzichtet.

3.1 Leonardo Da Pisa und Hasenzählen

Das erste Beispiel ist im Mittelalter angesiedelt. Leonardo DA PISA – auch FIBONACCI, *Filius Bonacci*, genannt – behandelt in seinem Buch *Liber Abbaci*, mit dem er das von den Arabern übermittelte dekadische Zahlensystem der Inder nach Europa brachte, unter anderem ein Zahlenbeispiel für Wachstum (siehe z. B. DUNLAP 1997, S. 35–49). Eine Reihe von Zahlen entsteht dadurch, dass jedes Element durch Summieren der beiden vorangehenden Zahlen gebildet wird:

$$F_n = F_{n-1} + F_{n-2}, n = 2, 3, \dots, F_0 = 0, F_1 = 1. \quad [1]$$

Die dadurch entstehende Reihe ist: 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ... FIBONACCI gibt eine einfache biologische Deutung: Die einzelnen Zahlen stellen Paare von Hasen – je ein Weibchen und ein Männchen – dar. Die erwachsenen Hasenpärchen bringen in jedem Zeitabschnitt ein Paar Junge – wieder ein Weibchen und ein Männchen – zur Welt. Die jungen Hasen sind nicht sofort zeugungsfähig und müssen noch einen Zeitabschnitt mit dem Jungewarten warten. Dann wird noch angenommen, dass die Hasen nicht altern: Alle Hasen leben ewig und bleiben ohne Einschränkung fruchtbar. Wir beginnen mit einem Paar junger Hasen, $F_1 = 1$, und daher gibt es nach dem ersten Zeitabschnitt noch keine Jungen, $F_2 = 1$. Im nächsten Zeitabschnitt bekommen die alten Hasen Nachwuchs, und zusammen mit den bereits vorhandenen Hasen haben wir nunmehr, $F_3 = 1 + 1 = 2$ Hasenpärchen. Im folgenden Zeitabschnitt gibt es wieder nur bei den alten Hasen Nachwuchs, da die jungen Hasen erst erwachsen werden müssen: $F_4 = 2 + 1 = 3$, und weiter setzt sich die Reihe fort wie oben angegeben. Die Fibonacci-Reihe ist ein Beispiel für exponentielles Wachstum in diskreter Form, d. h. in ganzen Zahlen ausgedrückt. In der Tat kann man die Reihe in einem gegebenen Intervall sehr genau durch eine obere und untere Exponentialfunktion beschränken:

$$f_{\text{upper}}(n) = \exp[0,44259(n-1)] \geq F_n \geq f_{\text{lower}}(n) = \exp[0,500917(n-2)], 1 \leq n \leq 10. \quad [2]$$

Der Begriff der Exponentialfunktion wurde allerdings lange nach FIBONACCI eingeführt und geht auf den Schweizer Mathematiker Leonhard EULER zurück, der sich als erster mit unendlichen Reihen und ihren Grenzwerten befasste (EULER 1748).

Der britische Nationalökonom und Begründer der Demographie Thomas Robert MALTHUS (1798) hat als erster das Wachstumsgesetz von Populationen formuliert und benutzte dazu eine geometrische Reihe: Ein Paar hätte zwei Kinder und diese jeweils zwei Nachkommen, die Population wächst von Generation zu Generation gemäß der Reihe 2, 4, 8, 16, 32, 64, 128, ...

Für kontinuierliche Zeit gleichbedeutend mit überlappenden Generationen oder dem Wegfall eines Zeitgebers für die Vermehrung wird aus der geometrischen Reihe eine Exponentialfunktion. Mit einer konstanten (relativen) Wachstumsrate r nimmt das Malthussche Wachstumsmodell die folgende Gestalt an:

$$P(t) = P(0) \exp(rt); r \dots \text{Wachstumsrate oder Malthus-Parameter, } t \dots \text{Zeit.} \quad [3]$$

Unkontrolliertes Wachstum führt zu Populationsgrößen, welche auf Grund der limitierten Nahrungsmittelproduktion nicht mehr ernährt werden können. Seine pessimistische Weltsicht, welche bereits an der Wende zum 19. Jahrhundert die Notwendigkeit von Geburtenkontrolle nahelegte, ist seinem Hauptwerk *An Essay on the Principle of Population* (MALTHUS 1798) entnommen, er schreibt: „Epidemics, pestilence and plague advance in terrific array, and sweep off their thousands and ten thousands. Should success be still incomplete, gigantic famine stalks in the rear, and with one mighty blow, levels the population with the food of the world.“² Der Einfluss von Robert MALTHUS auf die nach ihm kommenden Generationen war ungeheuer groß: Die Väter der natürlichen Auslese, Charles DARWIN und Alfred WALLACE, entnahmen dem Wachstumsmodell, dass der Großteil der Nachkommen in tierischen Populationen vor dem Erreichen der Geschlechtsreife sterben muss und daher nur die „Tüchtigsten“ Nachkom-

2 MALTHUS 1798, S. 61. Ende von Kapitel VII.

men haben können. Die Marxisten führten ins Treffen, dass durch den wissenschaftlichen Fortschritt die Nahrungsmittelproduktion nahezu beliebig steigerbar wäre, und kritisierten die Malthusschen Vorstellungen als eine bourgeoise Ideologie, um das Elend der Arbeiterklasse zu prolongieren. Das 20. Jahrhundert hat letzten Endes MALTHUS recht gegeben: Zwar wurde die Nahrungsmittelproduktion in einem fast unvorstellbarem Ausmaß gesteigert, aber das Wachstum der Weltbevölkerung hat noch mehr zugenommen, und das Problem lässt sich eben nur durch Geburtenkontrolle lösen.

3.2 Eine hypothetische Mathematik der Darwinschen Evolutionstheorie

Charles DARWINS *Origin of Species* (DARWIN 1859) ist, wie schon erwähnt, ebenso wie viele andere grundlegende Werke der Biologie frei von mathematischen Ausdrücken. Hätte DARWIN mathematische Gleichungen zur Modellierung der natürlichen Auslese verwendet, wie hätten diese wohl ausgesehen? Ohne Reduktion auf das für die Selektion Wesentliche und Abstraktion von entbehrlichen Komplikationen im Sinne von GALILEI und NEWTON kann man nicht weit kommen. Die vielleicht wichtigste Vereinfachung ist die Trennung von systemimmanenten Evolutionsprozessen und Einflüssen der Umwelt. Dies führt sogleich auf eine Grundfrage: Bis zur zweiten Hälfte des 20. Jahrhunderts wurde immer wieder angezweifelt, dass Evolution in einer konstanten Umgebung möglich wäre. Die Erfolge der Theorie der Selbstorganisation in Physik und Chemie und die Evolutionsexperimente mit Bakterien (ELENA und LENSKI 2003) und Molekülen (WATTS und SCHWARZ 1997, JOYCE 2007) haben diese Frage entschieden. Dies bedeutet nicht, dass die Veränderungen in der Umwelt keinen Einfluss auf die ablaufenden Evolutionsprozesse hätten, aber Selektion und Evolution gibt es auch unter konstanten Bedingungen. Einen einfachen Vergleich erlauben die Fallgesetze: Der Luftwiderstand hat sehr wohl einen Einfluss auf das Fallen der Körper, aber sie fallen auch im Vakuum – und dann alle gleich schnell.

Die Reproduktion von Organismen – auch von Viren und den relativ einfachen Bakterien – ist ein komplexer Vorgang, und bereits die Replikation, das Kopieren, von Nukleinsäuremolekülen ist ein aus sehr vielen Einzelschritten bestehender Prozess (BIEBRICHER et al. 1983, siehe auch Abschnitt 5.2). Wären die Detailkenntnisse dieser Vorgänge für eine quantitative Beschreibung der Evolutionsvorgänge unentbehrlich, dann stellte sich jeder Versuch einer Mathematisierung als ebenso schwierig heraus wie das Vorhaben, Verallgemeinerungen ohne einen Leitfaden aus einer Fülle von minutiösen Detailbeobachtungen herzuleiten. Den richtigen Leitfaden zu finden, war Charles DARWINS geniale Leistung bei der Erarbeitung seiner Evolutionstheorie. Heute können wir die notwendigen physikalischen Voraussetzungen hinzufügen:

- Ein Energie- und/oder Materialfluss muss das evolvierende System davon abhalten, den Gleichgewichtszustand zu erreichen; und
- die Reproduktion muss unter Bedingungen ablaufen, unter welchen die Bruttokinetik der Vermehrung durch eine exponentielle Wachstumsfunktion beschrieben werden kann.

Die Flussbedingung ist auf dem Planeten Erde im Prinzip leicht zu erfüllen: Das eingestrahelte Sonnenlicht hat eine Temperatur von 5780 K, wie sie der auf Oberfläche der Sonne herrscht, die Erde strahlt Licht mit einer Temperatur von 300 K ab, und dieser Temperaturunterschied kann im Sinne eines Carnotschen Kreisprozesses zur Leistung von Arbeit aller Art verwendet werden (CARNOT 1824). Beispielsweise nutzen die Pflanzen die in den von der Sonne kommenden Photonen vorhandene Strahlungsenergie zur Erzeugung chemischer Energie in Form

der Photosynthese. Exponentielles Wachstum wird in der Natur durch den Prozess der Reproduktion implementiert. Organismen testen, ob genügend Nahrung für die Vermehrung vorhanden ist, und Bakterien, aber auch Schleimpilze oder andere Organismen gehen in Dauerformen – Sporen – über, wenn das Nahrungsangebot für eine erfolgreiche Teilung zu gering ist.

Exponentielles Wachstum kann durch eine einfache Differentialgleichung ausgedrückt werden: Alle vorhandenen Individuen X haben Nachkommen, und daher ist der Zuwachs in der Population proportional zur Populationsgröße. Die Zahl der Individuen bezeichnen wir mit der Variablen $x(t) = [X]$, die Proportionalitätskonstante ist der Malthussche Wachstumsparameter, und für unbeschränktes Wachstum ergibt sich

$$\frac{dx}{dt} = rx, x(t) = x(0)e^{rt}. \quad [4]$$

Bereits der Mathematiker Pierre François VERHULST zeigte, wie man exponentielles Wachstum beschränken kann (VERHULST 1838, 1845):

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{C}\right), x(t) = \frac{x(0)C}{x(0) + (C - x(0))e^{-rt}}. \quad [5]$$

Die Konstante C ist die Maximalzahl der Individuen, welche im System ernährt werden können. Jede exponentiell wachsende Population wird diese Größe erreichen, der Parameter r gibt an, wie rasch der Grenzwert erreicht wird. Das Wachstum geht in eine Sättigung über, welche durch C – die Tragfähigkeit – im Englischen *Carrying Capacity* – des Ökosystems, bestimmt wird (Abb. 1).

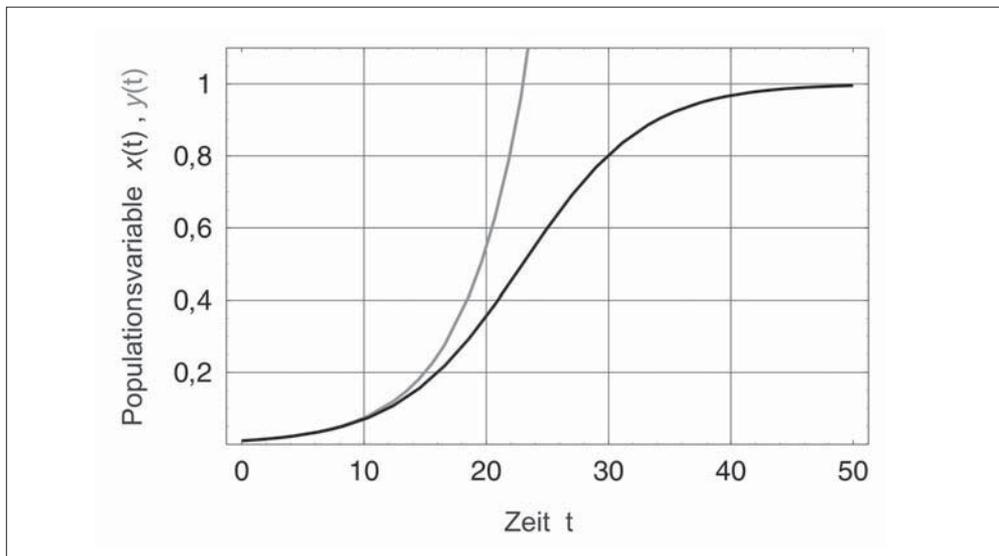


Abb. 1 Exponentielles Wachstum und endliche Tragfähigkeit eines Ökosystems. Die Gleichung nach Pierre François VERHULST beschreibt exponentielles Wachstum in einer Umwelt mit beschränkter Tragfähigkeit („*Carrying capacity*“) C . Verglichen sind hier unbeschränktes exponentielles Wachstum, $y(t)$ (grau), mit beschränktem Wachstum $x(t)$ (schwarz). Parameterwahl: $r = 0,2$, $C = 1$ und $x(0) = y(0) = 0,02$.

Das Verhulstische Modell wird oft als logistische Gleichung bezeichnet, und es ist anzunehmen, dass Charles DARWIN die Verhulstische Gleichung nicht kannte. Mit einem kleinen Trick schreiben wir die Gleichung so um, dass sie für die Beschreibung von Selektionsvorgängen geeignet ist:

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{C} \right) = x \left(r - \frac{1}{C} \varphi(t) \right) \text{ mit } \varphi(t) = rx. \quad [6]$$

Die Funktion $\varphi(t)$ beschreibt die Reproduktionsgeschwindigkeit der Population, die im speziellen Fall nur aus einer einzigen Spezies besteht. Es ist naheliegend, die Variable x durch Normierung dimensionsfrei zu machen, und wir ersetzen zu diesem Zweck x durch x/C . Von hier ist es nur mehr ein kleiner Schritt bis zur Selektion. Wir nehmen an, dass die Population n exponentiell wachsende Varianten in Konkurrenz aufweist: X_1, X_2, \dots, X_n . Die Variablen $x_i = [X_i]$ beschreiben die Zusammensetzung der Population und sind dementsprechend normiert: $\sum_{i=1}^n x_i = 1$. Für die Entwicklung der Population ergibt sich daraus die kinetische Gleichung:

$$\frac{dx}{dt} = x_i (f_i - \varphi(t)) = x_i F_i(t) \text{ mit } \varphi(t) = \sum_{j=1}^n f_j x_j, \quad [7]$$

wobei die Fitnesswerte der einzelnen Varianten mit $f_i (i = 1, \dots, n)$ bezeichnet wurden und $\varphi(t)$ die mittlere Fitness der Population darstellt. Durch eine einfache Rechnung erhalten wir die Zeitabhängigkeit dieser Funktion

$$\frac{d}{dt} \varphi(t) = \sum_{j=1}^n f_j^2 x_j - \left(\sum_{j=1}^n f_j x_j \right)^2 = \text{var}\{f\} \geq 0. \quad [8]$$

Die erhaltene Aussage ist von allgemeiner Gültigkeit: Die Zeitabhängigkeit der mittleren Fitness $\varphi(t)$ ist gegeben durch die Varianz der Fitnesswerte, und eine Varianz ist in der Statistik *per definitionem* eine nicht negative Größe. Mit anderen Worten, die mittlere Fitness einer Population kann nicht abnehmen, und dies ist bereits das Darwinsche Prinzip ausgedrückt in möglichst knapper Form. Der Faktor $F_i(t) = f_i - \varphi(t)$ in der kinetischen Gleichung beinhaltet bereits Charles DARWIN'S Prinzip der *Selection of the Fittest*. Das Vorzeichen entscheidet, ob $x_i(t)$ wächst oder abnimmt. Die mittlere Fitness legt die Latte und entscheidet, wer im Spiel bleibt und wer ausscheidet. Das stete Anwachsen der mittleren Fitness entspricht dem Höherlegen der Latte bei einem Hochsprungwettbewerb und bewirkt, dass immer mehr Hochspringer oder Varianten ausscheiden. Das Endergebnis, der Springer, welcher die größte Höhe gemeistert hat, oder die Variante mit der größten Zahl an Nachkommen bleibt letztendlich (als Sieger) über. Für unterschiedliche Fitnesswerte der einzelnen Varianten ist die Lösung des Selektionsproblems eindeutig.

Einige Schlussfolgerungen folgen unmittelbar aus der mathematischen Formulierung des Darwinschen Prinzips:

- Evolution ist ein Vorgang auf Populationsebene, und daher ist jede einzelne Variante oder Spezies im Zusammenhang mit den anderen existierenden Varianten zu sehen.
- Die Konkurrenz zwischen den Varianten führt zur Selektion der Variante mit der höchsten Fitness. Die mittlere Fitness einer Population ist eine nicht abnehmende Funktion der Zeit und bildet somit ein Optimierungsprinzip.
- Exponentielles Wachstum ist eine Voraussetzung für Darwinsche Selektion. Bei subexponentiellem Wachstum, z. B. bei parabolischem, linearem oder logarithmischem Wachstum,

wird im Allgemeinen nicht selektiert, und mehrere Varianten verbleiben auch im Grenzwert langer Zeiten in der Population.

Obwohl das Darwinsche Prinzip der natürlichen Auslese sehr klar formuliert ist und unschwierig in ein mathematisches Gewand gehüllt werden kann, stieß es sowohl in der britischen Gesellschaft der 19. Jahrhunderts als auch bei den Fachkollegen auf einigen Widerstand. Das gesellschaftlich-religiöse Problem hatte in erster Linie damit zu tun, dass die Darwinsche Evolutionstheorie den Menschen seiner Funktion als „Krone der Schöpfung“ entthronte. Die Kritik der Biologen zielte u. a. auf den Optimierungsaspekt ab: Wodurch lässt sich die Annahme belegen, dass die in der Natur beobachtbaren Arten und Individuen tatsächlich optimiert sind. Trotz vieler klärender Fortschritte erweist sich diese Frage auch 150 Jahre nach dem Erscheinen von DARWINS *Origin of Species* als ein nicht triviales Problem, zu dem wir später noch zurückkehren werden (Abschnitt 5.2.1).

3.3 Gregor Mendels Erbsenstatistik

Im dritten Beispiel tritt ein typisches Merkmal biologischer Gesetzmäßigkeiten zu Tage: Viele Zusammenhänge können nicht im Einzelexperiment erkannt werden, und es bedarf einer Vielzahl von Wiederholungen unter kontrollierten Bedingungen, um die Regelmäßigkeiten herauszuarbeiten. Der Augustinermönch Gregor Johann MENDEL (Erbsen: MENDEL 1866 und Habichtskraut: MENDEL 1870) machte als Erster gezielte Versuche zur Vererbung von Merkmalen in Pflanzen an der Gartenerbse *Pisum sativum*, entdeckte und begründete mit seinen Arbeiten die Genetik. In Experimenten vermied er durch kontrollierte Befruchtungstechnik unerwünschte Fremdbestäubung. MENDEL untersuchte sieben verschiedene Merkmale und fand für zwei dominant/rezessive Allelpaare³, Form der Früchte und Farbe des Endosperms, die ausgezählten Verhältnisse $5474/1850 = 2,96$ und $6022/2001 = 3,01$, woraus er korrekterweise auf das idealisierte Verhältnis 3:1 schloss. Die erhaltenen Verhältnisse bei den anderen fünf Merkmalen lagen zwischen 2,84 und 3,16. MENDELS Experimente involvierten rund 29000 Pflanzen und bilden ein wahres Musterbeispiel für die bei statistischen Gesetzen notwendige Vorgangsweise. Aus der Auswertung von MENDELS Beobachtungen resultieren drei Gesetze:

- Das Uniformitätsgesetz: Die Nachkommen von homozygoten, rein- oder gleicherbigen Individuen sind untereinander gleich.
- Das Spaltungsgesetz: Die Nachkommen einer Kreuzung mischerbiger Individuen sind nicht gleichförmig, sondern spalten ihr Erscheinungsbild in einem bestimmten Zahlenverhältnis auf.
- Das Gesetz der freien Kombinierbarkeit der Gene: Im Fall zweier oder mehrerer Merkmale gilt die freie Kombinierbarkeit, d. h., die Merkmale werden unabhängig voneinander kombiniert.

Das Gesetz gilt nur – wie wir heute wissen –, wenn die Gene für die Merkmale auf verschiedenen Chromosomen liegen, was bei den von MENDEL untersuchten Erbsen der Fall war.

3 Unter einem *Allel* versteht man eine Variante eines Gens. Beispielsweise produziert ein Allel (A) runde und ein anderes (a) runzelige Erbsenfrüchte. Jedes Individuum erhält ein Allel vom Vater- und ein Allel von der Mutterpflanze. Im Fall eines dominant/rezessiven Merkmals bilden AA, Aa und aA runde und nur aa runzelige Erbsenfrüchte aus, wodurch das Verhältnis 3:1 erklärt ist.

Betrachten wir die Ergebnisse Gregor MENDELS mit den Kenntnissen der modernen Molekulargenetik, so ergeben sich einfache Erklärungen aus dem molekularen Mechanismus der Meiose, der durch *Crossing-over* zur Rekombination der Erbanlagen führt (ALBERTS et al. 2008). Gleichzeitig werden die Mendelschen Gesetze als Regeln erkannt, welche nur im Grenzfall unabhängiger Gene gültig sind. Dies ist im Allgemeinen gewährleistet, wenn die Gene auf verschiedenen Chromosomen situiert sind. Wenn sie auf demselben Chromosom liegen, kann freie Kombinierbarkeit nur dann erfüllt sein, wenn die Genorte auf der DNA-Sequenz hinreichend weit voneinander entfernt sind. Das molekularbiologische Wissen liefert nicht nur eine mechanistische Erklärung der Mendelschen Gesetze, sondern sagt auch die empirisch beobachteten Abweichungen korrekt voraus.

Die Bedeutung der Befunde Gregor MENDELS wurde von seinen Zeitgenossen nicht erkannt, was wegen der mangelnden Mathematikausbildung der Biologen nicht verwundern sollte: Obwohl die Anwendung statistischer Methoden bis in das 17. Jahrhundert zurückreicht, ist die mathematische Statistik eine relativ junge Wissenschaft. Sie nahm ihre Anfänge um die Wende vom 19. zum 20. Jahrhundert, und einer der Väter dieser Disziplin, Karl PEARSON, wurde erst im Jahre 1911 der Vorstand des *Departments of Applied Statistics* am *University College London*, des weltweit ersten Departments für Statistik. Aus dieser historischen Perspektive heraus kann man verstehen, dass MENDELS Arbeiten fast 40 Jahre warten mussten, um in ihrer grundlegenden Bedeutung erkannt zu werden.

3.4 Ronald Fishers Fundamentaltheorem der natürlichen Auslese

Eine bahnbrechende Leistung von Ronald FISHER in der Evolutionsbiologie ist die Tatsache, dass es ihm gelang, die Darwinsche Selektion und die Mendelsche Vererbung in einem mathematischen Formalismus zusammenzuführen und dabei zu zeigen, dass (in einfachen Fällen) ein Optimierungsprinzip auch für die sexuelle Reproduktion gilt (FISHER 1930). In einer Population seien n verschiedene Allele, (A_1, A_2, \dots, A_n) mit den Häufigkeiten x_1, x_2, \dots, x_n für ein bestimmtes Gen vorhanden. Ein diploider Organismus trägt zwei Allele von jedem Gen, beispielsweise $A_i A_j$. Fitnesswerte werden den Allelkombinationen zugeordnet: $a_{ij} = f(A_i A_j)$. Da in der konventionellen Genetik davon ausgegangen wird, dass die genetisch bedingten Eigenschaften eines Individuums unabhängig davon sind, ob das Allel eines autosomalen⁴ Gens vom Vater oder von der Mutter vererbt wurde. Dadurch gilt für die Fitness: $f(A_i A_j) = f(A_j A_i)$ und dementsprechend $a_{ij} = a_{ji}$; die Matrix der Fitnesskoeffizienten $A = \{a_{ij}\}$ ist symmetrisch. Ohne weitere Probleme lässt sich eine Differentialgleichung für die zeitliche Entwicklung der Allelverteilung eines Genortes in der Population angeben:

$$\frac{dx_i}{dt} = x_i \left(\sum_{j=1}^n a_{ij} x_j - \varphi(t) \right) \text{ mit } \varphi(t) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_j x_i \text{ und } \sum_{i=1}^n x_i = 1. \quad [9]$$

Ebenso wie im vorhin besprochenen Fall der asexuellen Vermehrung kann gezeigt werden, dass die mittlere Fitness $\varphi(t)$ eine nicht abnehmende Funktion der Zeit ist:

$$\frac{d}{dt} \varphi(t) = 2 \text{var}\{a\} \geq 0. \quad [10]$$

4 Mit dem Begriff *Autosom* bezeichnet man alle Chromosomen mit Ausnahme der Geschlechtschromosomen, die beim Menschen die Konfigurationen XX (weiblich) und XY (männlich) aufweisen.

Bei der Herleitung dieses Extremalprinzips wird die Symmetrie der Matrix A benutzt. Die Fishersche Selektionsgleichung ist ebenso wie die vorher besprochene Selektionsgleichung für den asexuellen Fall eine Replikatorgleichung (SCHUSTER und SIGMUND 1983). Zum Unterschied von der eindeutigen Lösung im asexuellen Fall kann die Fishersche Selektionsgleichung auch mehrere stabile stationäre Zustände aufweisen. Dies ist am Beispiel zweier Allele einfach zu illustrieren: Wir nehmen an, dass die Heterozygote A_1A_2 geringere Fitness als die beiden Homozygoten aufweist, dann gilt $a_{12} < (a_{11}, a_{22})$, und jede der beiden Homozygoten, A_1A_1 oder A_2A_2 kann selektiert werden. Je nach den Anfangsbedingungen wird die Population einem der beiden stabilen Zustände, entweder $\lim_{t \rightarrow \infty} x_1(t) = 1, \lim_{t \rightarrow \infty} x_2(t) = 0$ oder $\lim_{t \rightarrow \infty} x_1(t) = 0, \lim_{t \rightarrow \infty} x_2(t) = 1$ zustreben. Für mehr als zwei Allele kann es dementsprechend mehr Langzeitlösungen für die Population geben. Im Allgemeinen – keine Symmetrie der Matrix A – ist die Dynamik von Replikatorgleichungen nicht so einfach; es können Oszillationen der Variablen und deterministisches Chaos auftreten (siehe Abschnitt 5.3).

Ronald FISHER nannte das oberhalb skizzierte Extremalprinzip das Fundamentaltheorem der natürlichen Auslese und erntete für diesen Universalanspruch einige Kritik. In der Tat ist das Theorem in der Formulierung über die Selektionsgleichung nur unter sehr eingeschränkten Bedingungen gültig. Im Fall mehrerer Genorte gilt es nur unter der unrealistischen Annahme unabhängiger Genwirkungen. In einer späteren Arbeit (FISHER 1941) formuliert er einschränkend „[...] the increase in average fitness of the population ascribable to a change in gene frequency [...] is equal to the genetic variance in fitness“, und dieser reduzierte Anspruch wird in neuerer Zeit von den meisten Populationsgenetikern akzeptiert (OKASHA 2008).

3.5 Alfred Lotka und Vito Volterra mit Hasen und Füchsen

Die von Alfred LOTKA (LOTKA 1925) und Vito VOLTERRA (VOLTERRA 1926) unabhängig vorgeschlagenen nichtlinearen Differentialgleichungen stellen einen ersten Versuch dar, die Wechselwirkung interagierender Populationen zweier Arten in einem Ökosystem zu beschreiben. Eine Art bildet die Beute, beispielsweise Hasen (X), während die zweite Art eine ausschließlich von diesen Beutetieren lebende Räuberpopulation darstellt, zum Beispiel eine Population von Füchsen (Y). Die Zahlen der Hasen, $x = [X]$, und Füchse, $y = [Y]$, sind die Variablen der Gleichungen:

$$\frac{dx}{dt} = (\alpha - \beta y)x \quad \text{und} \quad \frac{dy}{dt} = -(\gamma - \delta x)y. \quad [11]$$

Die vier griechischen Buchstaben stellen die Parameter des Systems dar und bedeuten im Einzelnen, den Reproduktionsparameter der Hasen (α), den Parameter für die Abnahme der Hasenpopulation als Folge des Gefressenwerdens durch Füchse (β), den Parameter für die Abnahme der Fuchspopulation durch Verhungern (γ) und den Parameter für die Zunahme der Fuchspopulation durch das Beutemachen an der Hasenpopulation (δ). Charakteristisch für die Lösungskurven des Lotka-Volterra-Systems sind Oszillationen (Abb. 2). Allerdings sind diese Oszillationen nicht stabil, und die Amplitude ist abhängig von den Anfangsbedingungen $[x(0), y(0)]$. Das Lotka-Volterra-Modell war zu seiner Zeit ein ganz wichtiger Beitrag zum Verstehen der Wechselwirkungen zwischen den Populationen zweier Spezies in einem Ökosystem. Heute wäre es in seiner ursprünglichen Form nicht mehr publizierbar, da es eine strukturelle Instabilität aufweist. Der Attraktor ist ein Zentrum, und dies hat zur Konsequenz, dass die Oszillationen gegenüber einer Veränderung der Amplitude instabil sind (HOFBAUER und SIGMUND 1998).

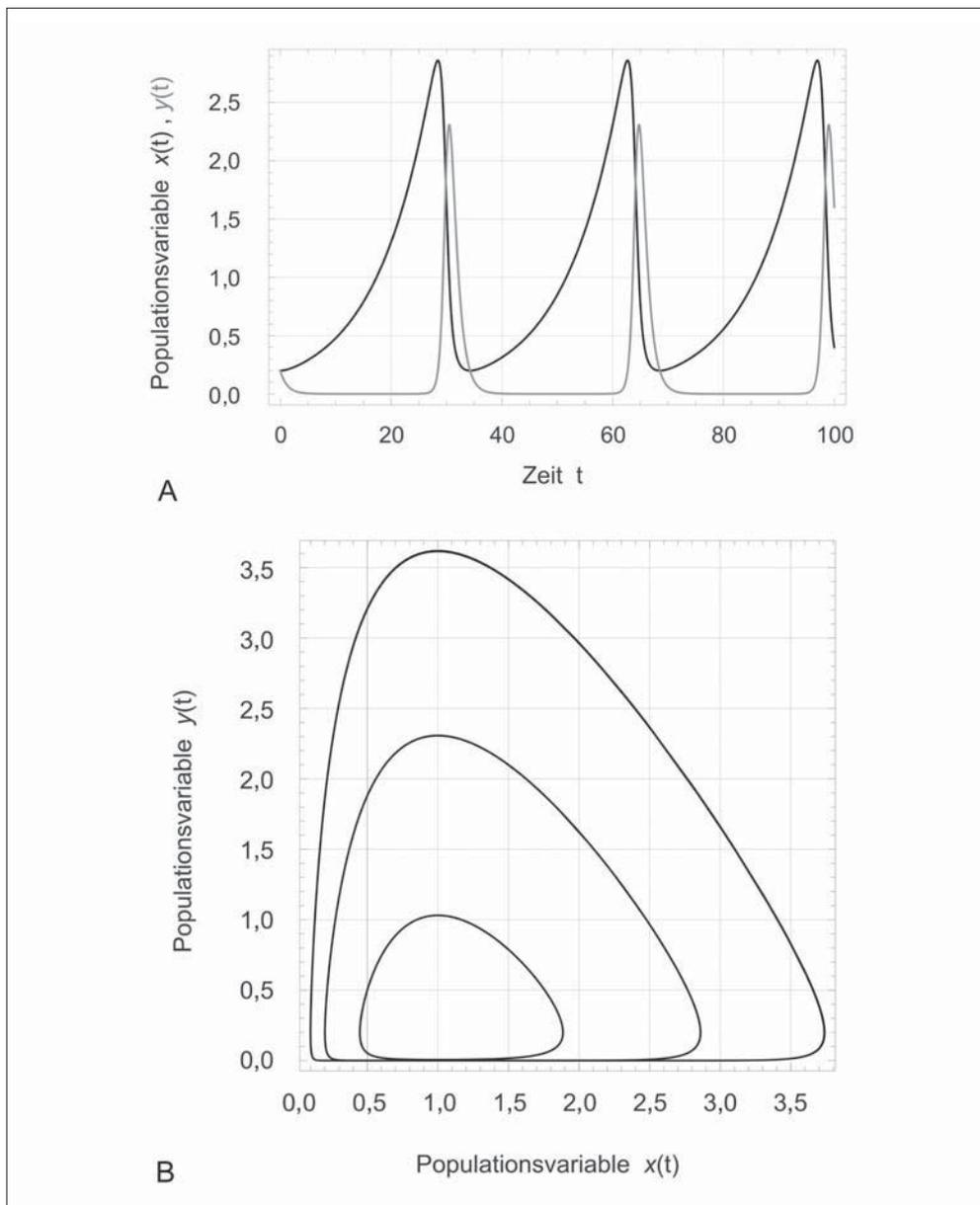


Abb. 2 Lösungskurven des Räuber-Beute-Ökosystems nach Alfred LOTKA und Vito VOLTERRA. Das von Alfred LOTKA und Vito VOLTERRA entwickelte Modell für ein Ökosystems mit einer Räuber- und einer Beutespezies, zum Beispiel Füchse und Hasen, führt zu oszillierenden Lösungen. Das Ergebnis einer numerischen Integration der Differentialgleichungen ist im oberen Teil der Abbildung gezeigt: (A) Die Hasenpopulation (schwarz) wächst, bis sie einen hinreichend großen Überschuss an Nahrung für die Füchse bietet. Diese vermehren sich dann sehr rasch (grau) und dezimieren die Hasen so stark, dass sie selbst wenig später an den Rand des Aussterbens kommen. In (B) zeigen wir drei Kurven einer parametrischen Darstellung der Lösungen: $y[x(t)]$, welche sich nur in den Anfangsbedingungen unterscheiden. Die Kurven werden mit fortschreitender Zeit t im Gegenuhrzeigersinn durchlaufen. Die Lösungen der Lotka-Volterra-Gleichung sind instabil gegenüber Änderungen in den Anfangsbedingungen, sie zeigen nur marginale

4. Das Zeitalter des Computers

In den 1960er Jahren brach eine neue Periode in den Naturwissenschaften an. Neben den traditionellen Säulen des wissenschaftlichen Frage- und Antwortspiels, Vorhersage durch mathematisierte Theorie und Überprüfung durch das Experiment, kam die Computersimulation als drittes Standbein hinzu. Waren die Leistungsfähigkeiten der Computer in der Anfangszeit noch zu gering, um eine echte Alternative zu mathematischer Analyse und empirischer Untersuchung zu bilden, so änderte sich die Situation in den letzten beiden Jahrzehnten des 20. Jahrhunderts durch die enormen Steigerungen von Computerleistung und Speicherkapazität. Die numerische Modellierung ist heute ein vollwertiger Partner von Theorie und Experiment. Die Modellierung hat mehrere Aufgaben übernommen, und es ist naheliegend, ein paar Klassen von Computermodellen zu unterscheiden:

- Computerrechnungen zur Kontrolle oder Erweiterung von mathematischen Modellen;
- Modelle zur Berechnung von Ergebnissen, die nicht analytisch erhältlich sind; und
- exploratorische Modelle.

Die Rolle und die Bedeutung von Modellen der ersten Klasse sind naheliegend: Die mathematische Beweisführung ist oft so schwierig, dass eine Rückversicherung gewonnener Resultate durch numerische Berechnung sehr hilfreich ist. Ein berühmt gewordenes Beispiel ist das Vierfarbenproblem bei der Aufteilung einer Ebene in Flächen, wie es beispielsweise bei einer Landkarte mit verschiedenen gefärbten Ländern der Fall ist (GONTHIER 2008): Eine Landkarte ist so zu bemalen, dass aneinander grenzende Länder verschiedene Farben aufweisen, und es ist zu beweisen, dass niemals mehr als vier Farben dazu nötig sind. Trotz der Einfachheit der Fragestellung war der Beweis nur in einer Arbeit mit umfangreichem Einsatz eines Computers zur erschöpfenden Falldiskussion möglich (APPEL und HAKEN 1977).

Mathematische Formulierungen von Sachverhalten besitzen oft einen Komplexitätsgrad, der eine rigorose mathematische Analyse sehr schwierig bis unmöglich macht. Als Illustrationsbeispiel sei die Erarbeitung von Lösungen der Schrödingergleichung für Moleküle in der theoretischen Chemie genannt. Paul DIRAC behauptete in seinem berühmten und oft kritisierten Statement: „[...] The underlying physical laws necessary for a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble [...]“ (DIRAC 1929). Wenn dies von vielen Chemikern auch als eine unnötig provokative Anmaßung eines Physikers interpretiert wurde, so bietet die Schrödingergleichung dessen ungeachtet heute einen wertvollen Zugang zur Berechnung der Eigenschaften bekannter und unbekannter Moleküle. Über Jahrzehnte hin entwickelten die theoretischen Chemiker Verfahren für numerische Näherungen zu den Lösungen der Schrödingergleichung und erreichten schließlich in den 1980er Jahren Genauigkeiten, welche mit den experimentellen Daten in der Chemie, insbesondere in der Molekülspektroskopie, konkurrieren konnten. Im Jahre 1998 wurde schließlich der Nobelpreis für Chemie an Walter KOHN und John POPLE vergeben, womit das neue Fach

Stabilität. Parameterwahl: $\alpha = 0,1$, $\beta = 0,5$, $\gamma = 1,0$ und $\delta = 1,0$; Anfangsbedingungen, obere Lösungskurven: $x(0) = y(0) = 0,2$; Anfangsbedingungen, parametrische Darstellung: äußere Kurve $x(0) = y(0) = 0,5$, mittlere Kurve $x(0) = y(0) = 0,2$ und innere Kurve $x(0) = y(0) = 0,1$. Durch Verfeinerungen des Modells erhält man asymptotisch stabile Lösungen (HOFBAUER und SIGMUND 1998).

„Computational Chemistry“ erstmals eine weltweite Beachtung außerhalb der Quantenchemie fand (RAMACHANDRAN et al. 2008).

Eine zweite Gruppe von Modellen, welche unter den zweiten Punkt fallen, bilden die Algorithmen: Ein Sachverhalt kann zwar nicht in eine lösbare Gleichung gebracht werden, aber dennoch ist eine iterative, rekursive oder anders aufgebaute schrittweise Annäherung an eine Lösung möglich. Algorithmen können zur exakten Lösung konvergieren oder eine Approximation berechnen. Die dynamische Programmierung ist ein oft gebrauchtes Beispiel für eine algorithmische Vorgangsweise. Diese Technik benutzende Methoden werden u. a. für die Berechnungen von optimalen Alignments von DNA-, RNA- oder Proteinsequenzen ebenso wie für die Vorhersage von RNA-Sekundärstrukturen verwendet.

Exploratorische Computermodelle schließlich erfüllen einen anderen, einen dritten Zweck. Sie gehen dem mechanistischen oder mathematischen Verstehen von Sachverhalten voraus und versuchen, durch die Erarbeitung von einfach zu modellierenden und analysierbaren Zusammenhängen die Möglichkeiten einer rigorosen Problembehandlung vorzubereiten. Im Bereich der Evolutionsforschung werden u. a. spezielle Computermodelle, die Prozesse durch die Belegung von Speicherplatz in der CPU darstellen, dazu verwendet, Evolutionsdynamik zu simulieren und zu analysieren (Abschnitt 5.1).

Spezielle Beispiele exploratorischer Modelle sind Computerspiele nach vorgegebenen Regeln. Das berühmteste Beispiel dieser Art ist ein zweidimensionaler zellulärer Automat,⁵ John CONWAYS *Game of Life*⁶ (GARDNER 1967): Auf einem im Prinzip unendlichen Spielbrett mit Schachbrettmuster befinden sich einzelne Partikel, die jeweils ein quadratisches Feld belegen. Ausgehend von einer Anfangskonfiguration werden neue Konfigurationen, Generation für Generation, nach vorgegebenen einfachen Regeln berechnet. Diese Regeln betreffen Änderungen in der Belegung der Felder: Ein belegtes Feld kann in der nächsten Generation belegt bleiben oder leer werden, und umgekehrt kann ein leeres Feld in der nächsten Generation belegt werden oder leer bleiben: Je nach der Belegungsdichte in der Umgebung eines Feldes überlebt die Belegung und bleibt in der nächsten Generation bestehen, oder das Feld wird frei (Abb. 3). Aus der Anfangskonfiguration evolviert die Belegung der Felder nach diesen einfachen Regeln. In den meisten Fällen endet ein Lauf nach einigen Generationen entweder durch Aussterben oder durch Erreichen einer stabilen, d. h. sich nicht mehr weiter ändernden Endkonfiguration. Manchmal kommt es auch zu periodischen Zuständen, und selten bilden sich erstaunlich komplexe Muster in Raum und Zeit heraus. Derartige Muster täuschen mitunter gezielte Handlungen⁷ vor, und diese und andere erstaunliche Phänomene waren Anlass für die Bezeichnung *Game of Life*. Andere Beispiele für die Anwendung von Spielen auf viele Fragestellungen in der Evolutionsbiologie finden sich in der Monographie von SIGMUND (1995).

5 Als *zelluläre Automaten* versteht man räumlich diskrete dynamische Systeme, in welchen die Entwicklung der einzelnen Zellen durch Regeln beschrieben wird. Die Regeln sind so gestaltet, dass der Zustand einer Zelle zum Zeitpunkt $t+1$ vom eigenen Zellzustand und von den Zellzuständen in einer vorgegebenen Nachbarschaft zum Zeitpunkt t abhängt.

6 *Game of Life* studiert man am besten durch Simulation auf dem eigenen Computer. Eine umfangreiche englische Dokumentation mit Anleitungen ist zu finden bei: <http://www.bitstorm.org/gameoflife>. Eine Webpage mit Erklärungen und Beispielen in deutscher Sprache findet sich unter: <http://www.mathematische-basteleien.de/gameoflife.htm>.

7 Ein Beispiel für vermeintlich zielorientiertes Handeln ist die *Gosper Glider Gun*.

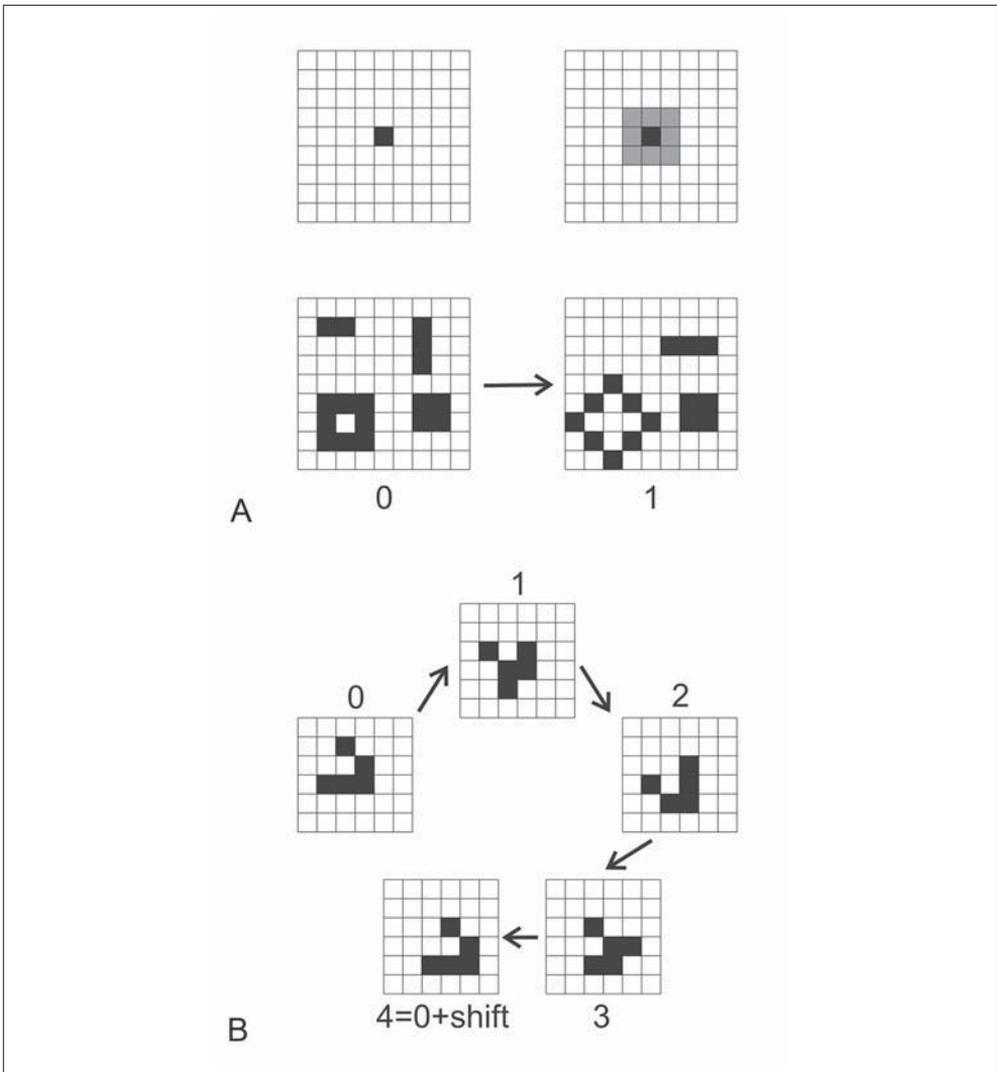


Abb. 3 Die Regeln von John CONWAYS ‚Game-of-Life‘. Im ‚Game-of-Life‘ sind besetzte oder aktive Felder auf einem unendlichen Schachbrett gekennzeichnet (schwarz). (A) Für jedes Feld wird eine Umgebung mit acht Nachbarn definiert (grau, rechtes oberes Bild). Einfache Regeln bestimmen die nächste Generation: (i) Jedes aktive Feld mit einem oder keinem aktiven Nachbarn stirbt und wird inaktiv (weiß), (ii) ein aktives Feld mit zwei oder drei aktiven Nachbarn überlebt und bleibt aktiv, (iii) ein aktives Feld mit vier oder mehr Nachbarn stirbt wegen Überbevölkerung, und (iv) jedes inaktive Feld mit drei Nachbarn wird aktiv. Die beiden Bilder in der Mitte zeigen vier Muster in aufeinanderfolgenden Generationen: Muster 1 (zwei benachbarte Felder, oben links) stirbt aus, Muster 2 (drei vertikale Felder, oben rechts) oszilliert zwischen drei vertikalen und drei horizontalen Feldern, Muster 3 (die gesamte Umgebung eines inaktiven Feldes, unten links) dehnt sich aus und geht schließlich in einen etwas komplizierteren Oszillator über (nicht gezeigt) und Muster 4 (ein Quadrat von vier aktiven Feldern, unten rechts) bleibt unverändert. (B) Die untere Bilderserie von fünf Bildern zeigt schließlich einen Wanderpattern („Glider“), nach vier Generationen entsteht ein Bild, das abgesehen von einer Verschiebung um je ein Feld nach rechts und nach unten mit der Anfangskonfiguration identisch ist.

Zelluläre Automaten dienen aber nicht nur zur Implementierung exploratorischer Modelle, sondern sind ganz allgemein zur Durchführung von Computersimulationen geeignet. Da sie ihrer diskreten Natur nach dem Rechnen mit Computern sehr entgegenkommen, haben sie viele Anwendungen in Simulationen von komplexen Systemen gefunden. Zelluläre Automaten sind zumeist einfach zu programmieren und auch für die Behandlung von nicht linearen dynamischen Systemen in Raum und Zeit einsetzbar (WEIMAR 1997). Sie werden bisweilen auch verwendet, wenn die entsprechenden mathematischen Zusammenhänge in Form von partiellen Differentialgleichungen bekannt sind. Die Lösungen werden dann in der Regel qualitativ richtig wiedergegeben (als *caveat* siehe HUBERMAN und GLANCE 1993). Benötigt man allerdings verlässliche quantitative Aussagen, ist die direkte numerische Integration vorzuziehen, da sie genaue Fehlerabschätzungen erlaubt. Eine wichtige Anwendung der zellulären Automaten ist fehlertolerantes Rechnen, das mit Hilfe probabilistischer Ansätze analysiert werden kann (GÁCS 2001). Die diskrete Natur der Beschreibung macht einfache zelluläre Automaten auch für eine rigorose mathematische Analyse ihrer Eigenschaften zugänglich. Die Literatur zu diesem Gebiet ist recht umfangreich, und es sei nur ein anspruchsvolles Buch zur Illustration der Vielseitigkeit von zellulären Automaten genannt (ILACHINSKI 2001).

5. Evolutionsforschung am Computer

Die gewaltige Zunahme an Computerleistung in den vergangenen fünfzig Jahren bietet der Erforschung von Evolutionsvorgängen ungeahnte neue Möglichkeiten. Wir greifen hier drei Ansätze heraus, für die der Computereinsatz zu echten Entdeckungen geführt hat. Im ersten Beispiel wird Konkurrenz direkt durch die Belegung von Speicherplätzen in der *Central Processing Unit* (CPU) eines Computers simuliert. Aus dem einfachen Computerspiel *Core War*⁸ sind eine Reihe von Simulationsprogrammen für *Artificial Life* und *Digital Evolution* hervorgegangen (O'NEILL 2003, WILKE und ADAMI 2002). Im zweiten Teil stellen wir Entdeckungen mit dem Computer im Bereich der kinetischen Modellierung von Evolutionsvorgängen auf molekularem Niveau dar. Die Computersimulation wurde an Stellen eingesetzt, an welchen man weder durch mathematische Analyse noch durch gezielte Näherungen weitergekommen war (EIGEN und SCHUSTER 1979). Den Abschluss bilden schließlich computergestützte Anwendungen der Spieltheorie, die auf John MAYNARD SMITH zurückgehen und zur Erklärung der Entstehung kooperativen Verhaltens in Gesellschaften verwendet wurden (MAYNARD SMITH 1982, HOFBAUER und SIGMUND 1998, NOWAK 2006). Die Ausbreitung von Verhaltensformen in Raum und Zeit wurde mit Hilfe des Computers untersucht.

5.1 Direkte Simulation von Evolutionsvorgängen am Computer

Zwei Computerprogramme, *Creeper* und *Reaper*, gingen der Entwicklung von *Core War* voraus (SHOCH und HUPP 1982): *Creeper* war einer der ersten Computerwürmer, die durch Computernetze krochen, bestimmte Programme infizierten, terminierten und Kopien von sich selbst anlegten. Das Programm *Reaper* hingegen wanderte durch das Netz auf der Suche nach Kopien

8 Der Begriff *Core* geht auf das heute völlig veraltete ‚random access‘-Speichersystem ‚Magnetic Core Memory‘ zurück. ‚Core‘ lebt auch noch in dem bekannten Begriff ‚Core dump‘ weiter.

von *Creeper*, die es stilllegte und eliminierte. *Creeper* und *Reaper* gehen damit nicht nur *Core War* und den CPU-Simulationen von Evolutionsvorgängen voraus, sondern initiierten auch die unliebsame Entwicklung von Computerviren und -würmern sowie deren Abwehrsoftware (DEWDNEY 1985).

Beim Computerspiel *Core War* (DEWDNEY 1984) gibt es keine spielenden Personen, sondern es sind Computerprogramme, welche im *Memory* der CPU eines Computers situiert sind, sich durch Kopieren vermehren und dadurch miteinander um Speicherplatz konkurrieren. Das Ergebnis der Auseinandersetzung ist Selektion: Das sich am effizientesten vermehrende Computerprogramm bleibt schließlich als einziges über, ganz im Sinne von *Selection of the fittest*. Im Allgemeinen wird *Core War* nicht auf einem wirklichen Computer gespielt, sondern auf einer virtuellen Maschine simuliert. Die Programme für *Core War* werden in der Maschinensprache oder in einer abstrakten Assemblersprache *Redcode* geschrieben. Computerfachleute haben sich eine große Reihe von derartigen Programmen mit verschiedenen Strategien ausgedacht, welche schließlich in eigenen Wettbewerben gegeneinander antraten (DEWDNEY 1987).

Von *Core War* bis zur Implementierung von evolutionsartigen Prozessen in virtuellen Maschinen war es nur mehr ein kleiner Schritt. In die Computerprogramme musste Variation beim Kopiervorgang in Form von Mutation und/oder Rekombination eingebaut werden, wobei das hauptsächliche Problem darin bestand, dass eine hinreichend große Zahl der Programmvarianten ‚lebensfähig‘, d. h. kopierfähig, sein musste, um einen Weiterbestand der Population in den zukünftigen Generationen sichern zu können. Ohne auf die Einzelheiten eingehen zu können, zählen wir hier nur zwei Ansätze zur Simulation von Evolutionsprozessen auf.⁹ Tom RAYS *Tierra* (RAY 1991, MAYNARD SMITH 1992, ADAMI 1995) war im Jahre 1990 eines der ersten Simulationssysteme, dessen Programme also evolutionsbefähigt waren, da sie kopieren und mutieren konnten. Die Programme werden als digitale Organismen betrachtet, welche um zwei Ressourcen, Computerzeit und Speicherplatz, konkurrieren. Im Vergleich zu den bescheidendsten natürlichen evolvierenden Systemen ist die Komplexität der *Tierraschen* Organismen natürlich sehr beschränkt.

Das Interesse der *Artificial Life Community* (ADAMI 1998) an der direkten Simulation von Evolution am Computer war der Anlass zu einer ganzen Reihe von weiteren Entwicklungen, die auf *Core War* und *Tierra* aufbauten. Steen RASMUSSEN und Kollegen entwickelten *Core-world*, um das Entstehen und die Evolution von kooperativen chemischen Reaktionsnetzwerken studieren zu können (RASMUSSEN et al. 1990). Charles OFRIA, Chris ADAMI, Titus BROWN und Claus WILKE entwickelten durch *Tierra* inspiriert eine eigene Software mit dem Namen *Avida* für Studien der Evolution von Populationen digitaler Organismen (WILKE und ADAMI 2002, O'NEILL 2003). Dieses Computerprogramm wird seitdem zum Studium von allgemeinen Evolutionsvorgängen eingesetzt, insbesondere können spezifische Hypothesen oder mathematische Modelle getestet oder verifiziert werden. Als Beispiel sei der Befund erwähnt, dass stationäre Mutantenverteilungen bei kleinen Änderungen von Mutationshäufigkeiten nach Art eines Phasenüberganges drastisch umstrukturiert werden können. Bei hinreichend großen Mutationsraten wird nicht immer der Genotyp mit dem höchsten Fitnessparameter selektiert, sondern kann es auch ein Genotyp mit geringerem Fitnessparameter und höherer Gesamtrepro-

9 Für ein weitergehendes Studium sei eine Monographie von John CASTI (1977) empfohlen; von Christian HEINMANN (2008) stammt ein illustrativer Aufsatz zu ‚Artificial Life‘.

duktionseffizienz als Folge eines stärkeren Mutationsrückflusses aus einer Mutantenumgebung mit höherer Fitness sein. Der nachteilige Fitnessunterschied wird durch den stärkeren Mutationsrückfluss überkompensiert (SCHUSTER und SWETINA 1988). Eine im Jahre 2001 durchgeführte Simulation mit *Avida* konnte diesen Befund ohne die Annahme einer stationären Population bestätigen (WILKE et al. 2001). Der Mikrobiologe Richard LENSKI ergänzte seine Langzeitevolutionsexperimente mit *Escherichia coli*-Bakterien durch Simulationen mit *Avida*, wodurch die *In-silico*-Evolution besonderen Auftrieb erfuhr (LENSKI 2004). Insbesondere bieten die Computereperimente immer die volle Information über die Verteilung aller Genotypen in einer Population, die zurzeit experimentell nicht erhalten werden kann. Dadurch kann die zeitliche Entwicklung nachteiliger und neutraler Mutanten, welche für ein tiefergehendes Verständnis des Evolutionsmechanismus unentbehrlich ist, direkt im Computereperiment verfolgt werden. Ein ungelöstes Problem der Evolutionsforschung betrifft den Ursprung und die evolutionäre Zunahme komplexer Eigenschaften und Funktionen. Ohne bereits eine Antwort geben zu können, war es doch mit Hilfe von Computersimulationen möglich zu klären, dass die Zunahme an Komplexität keinesfalls mit der Zahl der Mutationsschritte korreliert (LENSKI et al. 1999 und 2003; siehe auch Abschnitt 5.2.3).

5.2 Die kinetische Theorie der Evolution

Die Voraussetzungen für das Eintreten von Evolution wurden schon im Abschnitt 3.2 aufgezählt. Wir gehen hier davon aus, dass ein Energie- oder Materialfluss, der das System fern vom Gleichgewichtszustand hält, gegeben ist. Die Suche nach einem möglichst einfachen System, um das Untersuchungsobjekt im Sinne der Physik auf das Wesentliche zu reduzieren, führt unweigerlich auf die Evolution von RNA-Molekülen im Laborexperiment. Evolution *in vitro* ist schon vierzig Jahre alt (JOYCE 2007), und die ursprünglich grundlagenorientierten Arbeiten haben biotechnische Anwendungen in der gezielten Synthese von Biomolekülen gefunden (BRAKMANN und JOHNSSON 2002, KLUSMANN 2006). Die Replikation von RNA-Molekülen im Laborversuch ist ein mit den Methoden der Chemie voll analysierbares evolutionsbefähigtes Experimentalsystem (BIEBRICHER et al. 1983), in welchem die Bedingungen für exponentielles Wachstum präzisiert werden können. Diese bestehen aus einem ausreichend großen Reservoir an aktivierten Bausteinen für die Synthese und einem Überschuss an dem für die Replikation benötigten Enzym α -Replikase. Der experimentelle Aufbau für *In-vitro*-Evolution kann sehr einfach sein und im Überimpfen von kleinen Probenmengen in frisches Replikationsmedium bestehen (*Serial transfer*; SPIEGELMAN 1971) oder höchst komplexe Geräte erfordern, welche eine genaue Konstanz der Reaktionsbedingungen zu erzielen gestatten (WATTS und SCHWARZ 1997). Drei Probleme, bei deren Lösung der Computer eine entscheidende Rolle spielte, werden hier als Illustrationsbeispiele näher betrachtet:

- die Fehlerfortpflanzung bei fortgesetzter Replikation,
- die Gestalt realistischer Fitnesslandschaften und der Ursprung neutraler Varianten und
- der stochastische Verlauf der evolutionären Optimierung.

5.2.1 Fehlerfortpflanzung und Fehlerschranke

Im Einklang mit den experimentellen Daten wird die Replikation eines RNA-Moleküls durch zwei Schritte modelliert (Abb. 4): Die Nukleinsäure (X_i) bindet an die Replikase und initiiert

das Kopieren der Vorlage. Der Kopiervorgang kann eine fehlerlose Kopie (X_i) oder eine Mutante ($X_{j,j \neq i}$) produzieren. Die Evolution einer Population durch Replikation und Mutation wird durch eine Differentialgleichung, die Mutations-Selektionsgleichung beschrieben, welche unmittelbar aus der Selektionsgleichung durch Einführen von Mutationshäufigkeiten folgt (EIGEN 1971):

$$\frac{dx_j}{dt} = \sum_{i=1}^n Q_{ji} f_i x_i - x_j \varphi(t) \quad \text{mit} \quad \varphi(t) = \sum_{i=1}^n f_i x_i, \quad [12]$$

wobei Q_{ji} die Häufigkeit bezeichnet, mit welcher der Genotyp X_j durch eine fehlerhafte Replikation des Genotyps X_i gebildet wird. Korrekte Replikation und Mutation sind parallele chemische Reaktionen, und da einsichtiger Weise jede Kopie entweder fehlerfrei oder fehlerhaft sein muss, gilt die Beziehung: $\sum_{j=1}^n Q_{ji} = 1$. Die anderen Größen sind in ihrer Bedeutung unverändert. Auch an dem Term $\varphi(t)$, welcher den Zuwachs der Population kompensiert, hat sich durch Einführen der Mutationen nichts geändert, und er entspricht nach wie vor der mittleren Fitness der Gesamtpopulation. Die mathematische Analyse der Mutations-Selektionsgleichung ist nicht schwierig (THOMPSON und MCBRIDE 1974, JONES et al. 1976): Alle Mutationshäufigkeiten sind *per definitionem* positive Größen¹⁰, $0 < Q_{ji} < 1$, und alle Fitnessparameter f_i sind nicht-negative Zahlen, und daher sind die Bedingungen für die Gültigkeit des Theorems von PERRON und FROBENIUS (SENETA 1981) erfüllt. Durch – praktisch immer numerisches – Lösen eines Eigenwertproblems (EIGEN et al. 1989) können mit Hilfe des Computers Lösungen $x_j(t)$ berechnet werden. Aus dem n -dimensionalen Eigenwertproblem erhält man n Eigenvektoren $\xi_k(t)$, aus denen die Lösungskurven $x_j(t)$ durch lineare Kombination erhalten werden. Jedem Eigenvektor entspricht ein Eigenwert λ_k , wobei mehrere Eigenvektoren zu ein und demselben Eigenwert gehören können. Ist dies der Fall, spricht man von Entartung des Eigenwertes. Die Gültigkeit des Perron-Frobenius-Theorems hat für die Gestalt der Gesamtheit der Lösungen, welche die Entwicklung der Population beschreibt, entscheidende Bedeutung. Die für die Biologie wichtigsten Ergebnisse sind:

- Der größte Eigenvektor, $\xi_0(t)$, des Eigenwertproblems hat ausschließlich positive Komponenten, wie dies von (relativen) Konzentrationen verlangt werden muss; und
- der größte Eigenwert, λ_0 , der zum größten Eigenvektor $\xi_0(t)$ gehört, ist nicht entartet, wodurch die Eindeutigkeit der Lösung der Differentialgleichung garantiert ist.

Es gilt daher $\lambda_0 > \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots$, und dies bedeutet, dass die Population nach einiger Zeit gegen den größten Eigenvektor konvergiert, denn es gilt für t hinreichend groß: $\exp(\lambda_0 t) \gg \exp(\lambda_1 t) \geq \exp(\lambda_2 t) \geq \exp(\lambda_3 t) \geq \dots$. Der Eigenvektor $\xi_0(t)$ bildet daher die Langzeitleösung der Selektions-Mutationsgleichung und beschreibt die stationäre Verteilung der Mutanten in der Population.

10 Die biologische Bedingung für positive Mutationshäufigkeiten besteht darin, dass es eine – wenn auch noch so kleine – von Null verschiedene Wahrscheinlichkeit gibt, um in einem Replikationsschritt von einer beliebigen Sequenz X_i zu einer anderen Sequenz X_j zu kommen. Die Resultate der mathematische Analyse gelten aber auch für die schwächere Bedingung, dass die Produktion von X_j an der Vorlage X_i über eine endliche Zahl von Zwischenstufen erfolgt: $X_i \rightarrow X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow \dots \rightarrow X_j$.

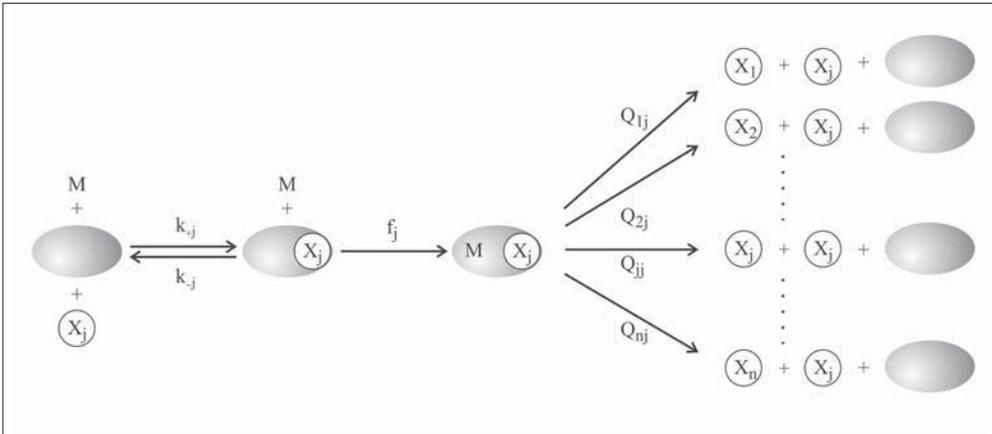


Abb. 4 Korrekte Replikation und Mutation als parallele chemische Reaktionen. Das replizierende Enzym, zum Beispiel Q β -Replikase, bindet zuerst die Nucleinsäure und anschließend die aktivierten Bausteine, die während der Synthese der Kopie eingebaut werden. Die Replikationskinetik wird (formal) in zwei Faktoren aufgeteilt: (i) der die Fitness bestimmende Geschwindigkeitsparameter f_j und (ii) die Mutationshäufigkeit Q_{ij} . Der Prozess führt zu einer korrekten Kopie mit der Frequenz $f_j Q_{ij}$, für erfolgreiche Vererbung muss $Q_{ij} Q_{ij} \forall i \neq j$ erfüllt sein. Außerdem muss jede Kopie entweder korrekt oder fehlerhaft sein und daher gilt: $\sum_{i=1}^n Q_{ij} = 1$. Die aktivierten Monomeren sind im Bild mit M bezeichnet.

Um die formale mathematische Analyse des Mutations-Selektionsproblems mit Leben zu erfüllen, stellen wir uns Anfangspopulationen mit beliebigen Verteilungen der Genotypen vor. Unter konstanten äußeren Bedingungen entwickeln sich diese Population durch Reproduktion, Mutation und Selektion, und nach hinreichend langer Zeit sind alle Populationen gleich geworden und verändern sich von außen gesehen nicht mehr. Man spricht von einem stationären Zustand, in dem sich die Veränderungen der relativen Konzentrationen durch korrekte Replikation, Mutation und Selektion gerade kompensieren, sodass von außen keine Änderung beobachtbar ist. In einer asexuellen Spezies bildet diese stationäre Verteilung von Mutanten, die durch den Eigenvektor $\xi_0(t)$ bestimmte stationäre Population, das genetische Reservoir und wurde in Analogie zur Spezies bei sexueller Vermehrung *Quasispezies* genannt (EIGEN und SCHUSTER 1977). Die Ausbildung von Quasispezies wurde nicht nur bei *In-vitro*-Evolutionsexperimenten mit RNA-Molekülen beobachtet, sie spielt auch eine vorrangige Rolle bei der Beschreibung und Analyse der Evolution von Viruspopulationen (DOMINGO 2006, DOMINGO et al. 2008). Durch das Auftreten von Mutanten in der Quasispezies kann die Population im stationären Zustand nicht mehr die maximale mittlere Fitness erreichen. Diese wäre nur in einer homogenen Population erfüllt, welche ausschließlich aus der Sequenz maximaler Fitness besteht, und es gilt daher $\lim_{t \rightarrow \infty} \phi(t) < f_m$: Aus dem Darwinschen Optimierungsprinzip wird eine, wenn auch überaus wirksame Optimierungsheuristik.

Die Häufigkeit von Mutationen hat einen entscheidenden Einfluss auf die Struktur der Quasispezies. Bei verschwindenden Mutationsraten wird gemäß der Selektionsgleichung die Variante mit der höchsten Fitness selektiert, und die Quasispezies ist eine homogene Population, bestehend aus diesem Genotyp, der als *Mastersequenz* charakterisiert wird. Im anderen Extremfall nehmen wir an, dass korrekte Replikation und Mutationen gleiche Wahrscheinlichkeiten aufweisen. Diese *Zufallsreplikation* führt dazu, dass im stationären Zustand alle Sequenzen gleich häufig sind: Die Quasispezies besteht aus der Gleichverteilung. In der Realität

finden wir stets Situationen vor, welche zwischen den beiden Extremen liegen. Um diese mathematisch mit geschlossenen Ausdrücken analysieren zu können, sind wir auf Näherungen angewiesen. Eine solche Näherung besteht in der Vernachlässigung der Erzeugung der Mastersequenz als fehlerhafte Kopien der Mutanten oder *Neglect of mutational backflow* (EIGEN 1971). Für diesen Ansatz, den wir hier als nullte Näherung bezeichnen, nimmt die Mutations-Selektionsgleichung die unten gezeigte Form an und entspricht formal der Selektionsgleichung. Die stationäre Konzentration der Mastersequenz in der nullten Näherung, $\bar{x}_m^{(0)}$, ist einfach zu berechnen:

$$\frac{dx_m^{(0)}}{dt} = (Q_{mm}f_m - \varphi)x_m^{(0)} \quad \text{mit} \quad \bar{x}_m^{(0)} = \frac{Q_{mm}\sigma_m - 1}{\sigma_m - 1}. \quad [13]$$

Hierin bedeuten Q_{mm} die Häufigkeit, mit welcher die Mastersequenz korrekt kopiert wird, und σ_m die *Superiorität* der Mastersequenz, welche als Quotient zwischen dem Fitnessparameter der Mastersequenz und der mittleren Fitness der restlichen Population berechnet wird:

$$\sigma_m = \frac{f_m}{f_{-m}} \quad \text{mit} \quad \bar{f}_{-m} = \frac{\sum_{i=1, i \neq m}^n f_i x_i}{\sum_{i=1, i \neq m}^n x_i} = \frac{\sum_{i=1, i \neq m}^n f_i x_i}{1 - x_m}. \quad [14]$$

Um die Abhängigkeit der stationären Konzentration der Mastersequenz als eine Funktion der Replikationsgenauigkeit einfach berechnen zu können, werden *Punktmutationen* betrachtet, welche einzelne Nukleotide austauschen. Außerdem müssen wir noch die vereinfachende Annahme einer *Uniform error rate* treffen, welche davon ausgeht, dass die spontane Mutationshäufigkeit¹¹ weder von der Natur des zu mutierenden Nukleotids noch von der Position entlang der Sequenz abhängt.¹² Bezeichnen wir diese einheitliche Häufigkeit einer Mutation pro Nukleotid und Replikation mit p , dann erhalten wir für die Mutationsterme zwischen allen Sequenzen:

$$Q_{ji}(p) = (1 - p)^{\lambda - d_H(X_j, X_i)} p^{d_H(X_j, X_i)}, \quad Q_{mm}(p) = (1 - p)^\lambda. \quad [15]$$

Die Länge der Polynukleotidsequenz ist hier mit λ bezeichnet und $d_H(X_i, X_j)$ stellt den Hammingabstand¹³ zwischen den beiden Sequenzen X_i und X_j dar. Einsetzen für Q_{mm} ergibt die Abhängigkeit der stationären Konzentration von der Mutationsrate, $\bar{x}_m^{(0)}(p)$. Mit zunehmender Mutationsrate wird der Anteil der Mastersequenz an der Quasispezies immer kleiner, bis schließlich die stationäre Konzentration der Mastersequenz an der Stelle $p = p_{cr}$ verschwindet. Diesen kritischen Wert der Mutationsrate bezeichnen wir als Fehlerschranke:

-
- 11 Die spontane Mutation steht im Kontrast zu induzierten Mutationen, welche durch Einwirkung von Strahlung oder Chemikalien erzeugt werden.
- 12 Experimentelle Daten zeigen, dass diese Annahme nur eine Näherung darstellt, denn es gibt sogenannte ‚Hot spots‘ für Mutationen. Dies sind Positionen, an welchen die Mutationshäufigkeit deutlich höher ist als an den restlichen Nukleotiden. Dessen ungeachtet ist der Ansatz einer *Uniform error rate* durchaus brauchbar, wenn wir annehmen, dass die verwendete Mutationshäufigkeit einen Mittelwert darstellt.
- 13 Der Hammingabstand gibt die Zahl der Positionen an, in welchen sich zwei Sequenzen unterscheiden (HAMMING 1950, 1986). Wesentlich für die Berechnung des Hammingabstandes ist das Alignment, die Art der Gegenüberstellung der beiden Sequenzen. Wir gehen hier davon aus, dass die Sequenzen gleich lang sind und die beiden Enden einander gegenüber gestellt werden.

$$p_{\bar{x}_m^{(0)}=0} = p_{cr} = 1 - \sigma_m^{-\frac{1}{\lambda}} \approx \frac{\ln \sigma_m}{\lambda} \quad \text{oder} \quad \lambda_{\max} \approx \frac{\ln \sigma_m}{p}. \quad [16]$$

Das Verschwinden der Mastersequenz ist natürlich ein Artefakt der nullten Näherung, denn die Gültigkeit des Perron-Frobenius-Theorems garantiert, dass keine Konzentration null werden kann (Abb. 5). Zwar kann diese Näherung durch störungstheoretische Ansätze verbessert werden (SWETINA und SCHUSTER 1982), aber die Verbesserungen sind mühsam zu berechnen und führen nicht sehr weit.

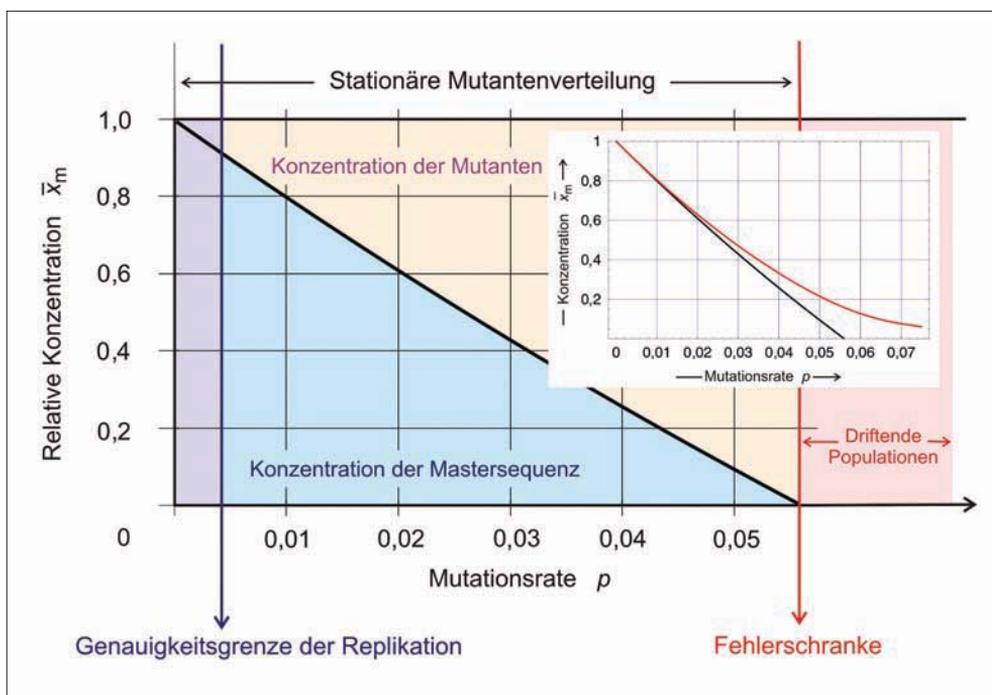


Abb. 5 Quasispezies und Fehlerschranke. Gezeigt ist die stationäre Konzentration der Mastersequenz als Funktion der Mutationsrate p . Für verschwindende Mutationsrate wird die Mastersequenz selektiert, $\lim_{p \rightarrow 0} \bar{x}_m = 1$, mit steigender Mutationsrate nimmt der Anteil der Mutationen in der Quasispezies ständig zu, und der Anteil der Mastersequenz wird entsprechend immer geringer. In nullter Näherung verschwindet die Konzentration der Mastersequenz bei $p_{cr} = 0,056$ (für $\lambda = 6$ und $\sigma = 1,4131$). Im Insert ist die Konzentration in nullter Näherung $\bar{x}_m^{(0)}(p)$ mit der exakten Lösung $\bar{x}_m(p)$ verglichen, welche gegen den Wert $\lim_{p \rightarrow \tilde{p}} \bar{x}_m = 1/64 = 0,015625$ konvergiert. Der Bereich einer stationären Quasispezies in realen Systemen wird bei hohen Mutationsraten durch die Fehlerschranke $p = p_{cr}$ begrenzt, bei höheren Mutationsraten, $p > p_{cr}$, bricht infolge der Akkumulation von Fehlern die Vererbung einer stabilen Polynukleotidsequenz zusammen, und Populationen driften nach Art eines Irrflugs durch den Raum aller Sequenzen. Eine zweite Grenze der Replikation, die Genauigkeitsgrenze, ist durch die Fehlerrate des Replikationsprozesses gegeben. Natürliche Mutationsraten sind in Tabelle 1 zusammengestellt.

Hier endet der einfache mathematische Zugang zum Fehlerfortpflanzungsproblem, und es beginnt die Entdeckungsreise mit Hilfe des Computers. Die Berechnung einer Population mit allen möglichen Sequenzen stößt auch bei relativ kurzen Kettenlängen auf unüberwindliche

Grenzen: Schon bei Kettenlänge $10=\lambda$ hat man für natürliche Oligonukleotide ein Eigenwertproblem $10^6 \times 10^6$ zu lösen, für binäre Sequenzen bleibt immerhin noch die Größe 1000×1000 . Der letztere Fall ist mit den heutigen Computern und Algorithmen eine Routineangelegenheit, für den ersteren Fall muss man schon die Struktur der Matrix ausnützen, um mit den Ressourcen, Computerzeit und Speicherplatz das Auslangen zu finden. Für besonders einfache Verteilungen der Fitnessparameter, beispielsweise für die *Single-Peak*-Landschaft (Abb. 8), in welcher es nur zwei Fitnessparameter gibt, einen für die Mastersequenz und einen zweiten für alle anderen Sequenzen, nehmen alle Sequenzen, welche einen bestimmten Hammingabstand d_H von der Mastersequenz, $X_m \equiv X_0$ haben, die gleiche stationäre Konzentration an und können daher zu einer Fehlerklasse Γ_k zusammengefasst werden:

$$\Gamma_k: \{X_i \in \Gamma_k \text{ iff } d_H(X_k, X_0) = k\} \quad \bar{y}_k(p) = \sum_{i=1, X_i \in \Gamma_k} \bar{x}_i(p). \quad [17]$$

In Abbildung 6 ist als Beispiel die Abhängigkeit der Quasispezies auf einer *Single-Peak*-Landschaft für eine Kettenlänge $100 = \lambda$ gezeigt. In der Nähe der kritischen Mutationsrate p_{cr} beobachten wir drei unterschiedliche Phänomene (SCHUSTER 2010a):

- eine Abnahme der stationären Konzentration der Mastersequenz $\bar{y}_0^{(0)}(p) = \bar{x}_0^{(0)}$ auf einen sehr kleinen Wert;
- eine völlige Umstrukturierung der Quasispezies in einem sehr kleinen Intervall Δp ; und
- einen Übergang in die Gleichverteilung, obwohl der Wert der kritischen Mutationsrate noch weit vom Punkt der Zufallsreplikation entfernt ist: $p_{cr} \ll \tilde{p} = \kappa^{-1}$, worin κ die Zahl der Buchstaben im Nukleotidalphabet bedeutet – für natürliche Nukleinsäuren gilt $\kappa = 4$.

Interessanterweise stellt sich heraus, dass die mit der nullten Näherung berechnete Fehlerschranke fast genau mit dem aus den numerischen Rechnungen auf einer *Single-Peak*-Landschaft abgeleiteten Wert übereinstimmt. Eine exakte Lösung der Quasispezies auf einer *Single-Peak*-Fitnesslandschaft kann mit Hilfe von Methoden aus der statistischen Mechanik von Polymeren erhalten werden (GALLUCCIO 1997). Bei Fehlerraten jenseits der Fehlerschwelle, $\tilde{p} > p > p_{cr}$, entspricht die stationäre Mutantenverteilung der Gleichverteilung. Für Moleküle der Kettenlänge $\lambda = 100$ hätte dies zur Konsequenz, dass alle $4^{100} \simeq 1,6 \times 10^{60}$ Sequenzen in gleicher Menge vorhanden sein müssten. Bei dem Minimum von einem Molekül pro Sequenz entspräche dies einer Populationsgröße von $1,6 \times 10^{60}$, die gegenüber den größten Proben mit etwa 10^{15} RNA-Molekülen überastronomisch ist. Jenseits der Fehlerschwelle kann daher die Quasispezies nicht realisiert werden, man beobachtet an ihrer Stelle Populationen, welche im Raum der Sequenzen driften (siehe auch Abschnitt 5.2.3). Für Genetik und Evolution hat das Fehlen einer stationären Lösung eine unmittelbare Konsequenz: Es gibt keine längerfristige Vererbung von Nukleinsäuresequenzen. Optimierung von Eigenschaften ist höchstens auf der Ebene der Phänotypen möglich, wenn viele Genotypen die gleiche Fitness aufweisen und daher gegenüber Selektion neutral sind (Abschnitt 5.2.3).

Quasispezies und Fehlerschranke haben eine weit über die *In-vitro*-Evolution hinausgehende Bedeutung. Insbesondere haben die Konzepte in die Virologie Eingang gefunden und werden für die Entwicklung neuer antiviraler Wirkstoffe angewendet (DOMINGO 2005, 2006,

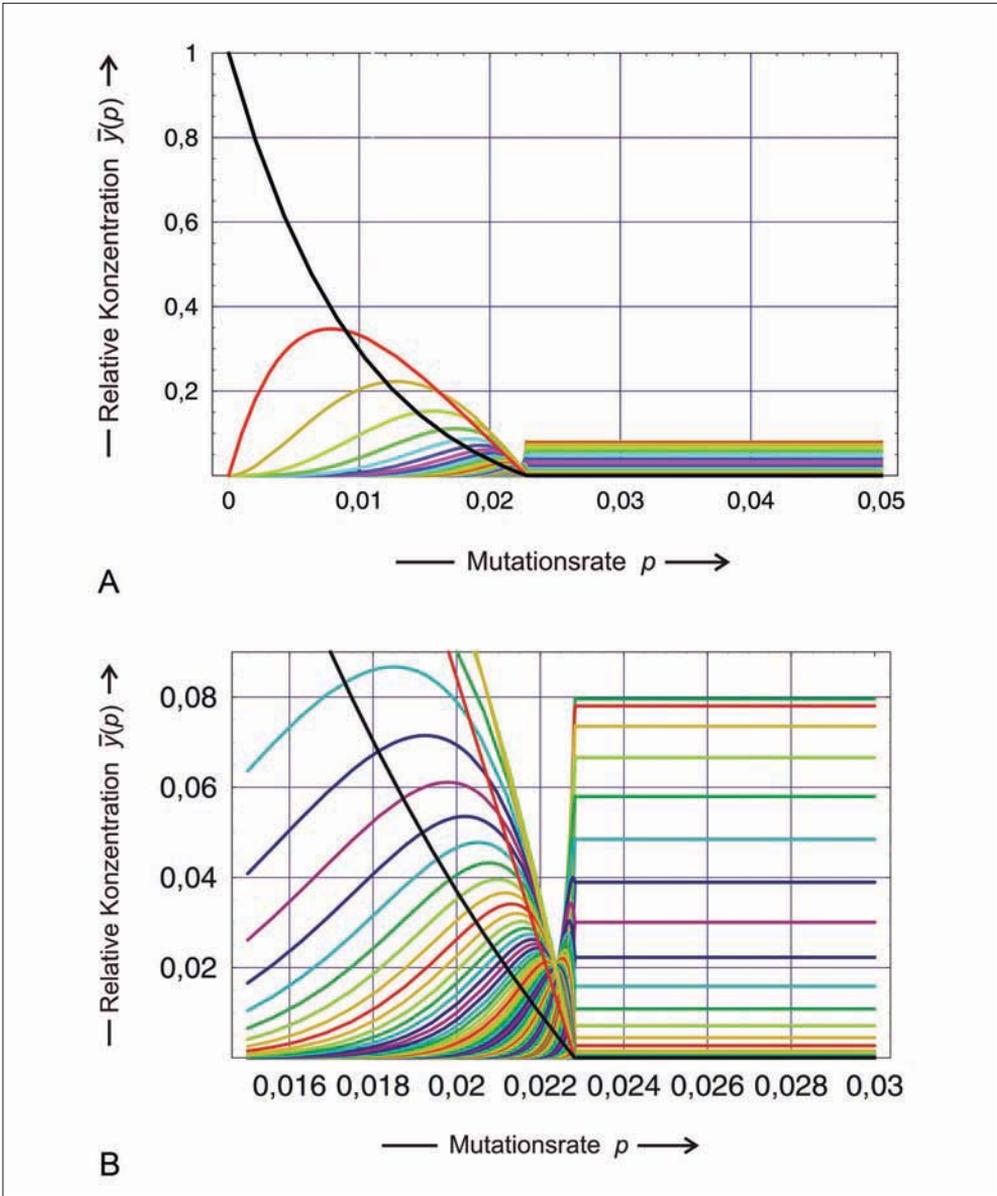


Abb. 6 Fehlerschranke auf einer *Single-Peak*-Landschaft. Die Fehlerschranke ist als Auftragung der (relativen) Konzentration gesamtter Fehlerklassen $\bar{y}_k(p) = \sum_{i=1, d_H(x_i, x_0)=k}^n \bar{x}_i(p)$ dargestellt. Die nullte Fehlerklasse ist durch die Mastersequenz gegeben, $X_0 \equiv X_m$, und die Zugehörigkeit zu einer Fehlerklasse k ist durch eine vorgegebene Hammingdistanz von der Mastersequenz, $d_H(X_i, X_0) = k$, bestimmt. Kurven für die einzelnen Fehlerklassen sind durch unterschiedliche Farben der Kurven illustriert: $\bar{y}_0(p) = \bar{x}_0(p)$ schwarz, $\bar{y}_1(p)$ rot, $\bar{y}_2(p)$ gelb, etc. Die untere Abbildung ist eine Vergrößerung des oberen Bildes. Sie lässt die Schärfe des Überganges von der strukturierten Quasispezies zur Gleichverteilung und die Abnahme der Konzentration der Mastersequenz besonders gut erkennen.

Parameterwahl: $\lambda = 100, f_m = 10, f_{k,k \neq m} = 1$, und daher gilt $\sigma_m = 10, p_{cr} = 0,02276$ und $\bar{y}_0(\tilde{p}) = \bar{x}_0(\tilde{p}) = 1/2^{100} = 7,9 \times 10^{-31}$.

Tab. 1 Genomlängen und Mutationsraten für verschiedene Klassen von Organismen (DRAKE 1993, DRAKE et al. 1998, DRAKE und HOLLAND 1999). Mit kB wird die Einheit eine Kilobase = 1000 Nukleotide abgekürzt. Die Mutationsrate μ ist der Mittelwert aus den Produkten Genomlänge \times Mutationsrate p .

Klasse	Genomlänge λ (kB)	Mutationsrate p	Mutationsrate μ
RNA-Viren	4,2 – 13,6	$7,3 \times 10^{-5} - 1,5 \times 10^{-3}$	0,67
Retroelemente ¹⁴	5,9 – 9,2	$3,3 \times 10^{-6} - 2,4 \times 10^{-5}$	0,1
DNA-Organismen	$6,4 - 42 \times 10^3$	$7,2 \times 10^{-11} - 7,2 \times 10^{-7}$	0,0034
Höhere Eukaryoten	$1,8 \times 10^7 - 8,0 \times 10^7$ ($8,0 \times 10^7 - 3,2 \times 10^9$)	$5,0 \times 10^{-11} - 2,3 \times 10^{-10}$	0,0033

DOMINGO et al. 2008)¹⁵. Hier sei noch auf einen empirischen Befund hingewiesen, der zeigt, dass das über die Fehlerschranke hergeleitete Produkt aus Genomlänge und spontaner Mutationsrate, $\mu = p\lambda$, konstant und für eine bestimmte Klasse von Organismen charakteristisch ist (Tab. 1). Die Werte variieren von etwa einer Mutation pro Replikation bei RNA-Viren über eine Mutation pro zehn Replikationen bei Retroelementen bis zu einer Mutation in 300 Replikationen bei Bakterien und höheren Organismen.

5.2.2 Realistische Fitnesslandschaften

Das Konzept einer Fitnesslandschaft geht zurück auf Sewall WRIGHT, einen der drei Begründer der Populationsgenetik. Er illustrierte u. a. den evolutionären Optimierungsprozess mit der einfachen Metapher eines Wanderers in einer bergigen Landschaft, der nicht bergab gehen kann (WRIGHT 1932, 1988): Werte für die Fitness einzelner Genotypen sind aufgetragen über einem Raum aller Genotypen, wodurch eine Art Landschaft entsteht. Der Abstand zwischen zwei Genotypen entspricht ihrer *genetischen Distanz*, und dies bedeutet, dass verwandte Genotypen einer Art im Raum der Genotypen einen Cluster bilden. In Sinne der Darwinschen Theorie besetzen Arten durch die Selektion bedingt lokale Optima auf der Evolutionslandschaft. WRIGHTS Metapher wurde heftig kritisiert, insbesondere von FISHER, der ins Treffen führte, dass zwei Annahmen unzutreffend wären:

- die Unabhängigkeit der Fitnesswerte von der Population und
- die zeitliche Invarianz der Fitnesslandschaften.

Neben FISHERS Kritik war es die empirische Unzugänglichkeit von Fitnesswerten, die das Konzept Fitnesslandschaft bis in die 1980er Jahre zum reinen Illustrationsbeispiel degradierte. Erst mit dem zunehmenden Wissen von Molekularbiologie und Biophysik änderte sich die

14 Retroelemente sind Retroviren und Retrotransposons, DNA-Organismen sind DNA-Viren und Bakterien sowie *Saccharomyces cerevisiae* (Hefe) und *Neurospora crassa* (Schlauchpilz). Bei den höheren Eukaryoten bezieht sich der Wert von μ auf eine Zellteilung und die *effektive Genomgröße*, welche jene Teile des Genoms ausklammert, in denen fast nur neutrale Mutanten vorkommen (die tatsächliche Genomgröße ist in Klammern angegeben).

15 Eine vor drei Jahren begonnene Diskussion, ob der Effekt der Erhöhung der Mutationsrate durch antivirale Pharmaka durch Überschreiten der Fehlerschwelle oder durch letale Mutagenese – ‚lethal mutagenesis‘ (BULL et al. 2005, 2007) – beschrieben wird, ist im Grunde genommen überflüssig, wenn man die kinetische Mutations-Selektionsgleichung korrekter Weise auf eine abnehmende, d. h. zeitlich variierende Populationsgröße anwendet.

Situation: Fitnesslandschaften werden als ein Spezialfall von Hyperflächen¹⁶ verstanden, welche in Physik und Chemie standardmäßig zur Trennung von hinreichend langsamen Dynamiken von den darunter liegenden schnelleren Prozessen verwendet werden. Als Beispiel seien hier nur die Energiehyperflächen der Quantenchemie im Rahmen der Born-Oppenheimer-Näherung genannt (CRAMER 2002), auf denen die Bewegungen der Atomkerne getrennt von den Elektronenbewegungen ablaufen. Vorrangige Bedeutung gewann das Konzept der Energieflächen bei den Ansätzen, die Strukturen der Biopolymeren mit den Methoden der Physik zu verstehen (ONUCHIC et al. 1997). Wir können uns hier auf die wesentlichsten Gesichtspunkte beschränken, da kürzlich ein umfangreicher Übersichtsartikel erschienen ist (SCHUSTER 2009).

Die Grundlage für das Konzept der Hyperflächen in der Biologie ist der Ansatz, die Funktionen von Biopolymeren durch zwei aufeinander folgende Abbildungen zu beschreiben:

$$S_j = \Psi(X_i) \quad f_k = \Phi(S_j)$$

Sequenz → Struktur → Funktion

Die Funktion eines Moleküls oder eines komplexeren Gebildes, beispielsweise die Fitness, ist eine Eigenschaft der molekularen Struktur S , und diese ist ihrerseits durch die Sequenz X bestimmt. Mathematisch lässt sich dies durch zwei Abbildungen oder mathematische Funktionen, $S = \Psi(X)$ und $f = \Phi(S)$, ausdrücken. Als Struktur versteht man hier im Allgemeinen die thermodynamisch stabile Struktur mit minimaler freier Energie. Im Falle von Biopolymeren, Proteinen oder Nucleinsäuren, sind diese Funktionen zum Teil schon bekannt. Ein theoretisch und experimentell besonders gut untersuchtes Beispiel bilden die Sekundärstrukturen von Ribonucleinsäuren (SCHUSTER et al. 1994, REIDYS et al. 1997, REIDYS und STADLER 2002, SCHUSTER 2006), deren Abbildungen als Modelle für kompliziertere Fälle dienen können (Abb. 7).

Im Fall von Biopolymeren, Nucleinsäuren oder Proteinen, gibt es viele Hinweise, dass ein relativ hoher Anteil der Mutationen keinen oder keinen merkbaren Einfluss auf die Eigenschaften des Biomoleküls aufweist. Als Beispiel seien die synonymen Codons¹⁷ bei der Translation erwähnt. Dies führt in der Folge zu Neutralität gegenüber Selektion (KIMURA 1983, OHTA und GILLESPIE 1996, OHTA 2002), und diese äußert sich in *genetischer Drift*: In endlichen Populationen kommt es durch die Stochastik der Vermehrung auch in Abwesenheit von Fitnessdifferenzen zur Anreicherung eines Genotyps, wobei die Auswahl der überlebenden Variante zufällig ist, und man spricht dann von zufälliger *Fixierung*, d. h. Anreicherung bis zu (praktisch) 100 %, eines Genotyps in der Population. Ein einmal fixierter Genotyp wird im Mittel nach einer Zeit $\tau = \mu^{-1}$ durch einen anderen (zufällig ausgewählten) Genotyp ersetzt, wobei μ die auf das gesamte Genom bezogene Mutationsrate darstellt. Interessanterweise hängt diese Zeit für die Ablöse eines Genotyps nicht von der Populationsgröße N ab. Das gegenwärtige Bild hinsichtlich der Wirksamkeit von Mutationen geht davon aus, dass in einer Spezies, die sich in einem stationären Ökosystem befindet, etwa die Hälfte der Mutationen nachteilig sind und durch Selektion eliminiert werden, etwa ebenso viele Mutationen sind neutral, und nur wenige Mutationen sind vorteilhaft und bestimmen die weitere Evolution.

16 Eine Hyperfläche ist das Analogon zu einer Fläche im dreidimensionalen Raum in anderen, praktisch immer höheren Dimensionen.

17 Bei der Übersetzung von Nucleotidsequenzen der DNA in die Aminosäuresequenzen der Proteine stehen 64 Codons den 20 Aminosäureresten gegenüber. Zwangsläufig gibt es daher Redundanz bei der Kodierung. Codons, welche für dieselbe Aminosäure kodieren, werden als synonym bezeichnet.

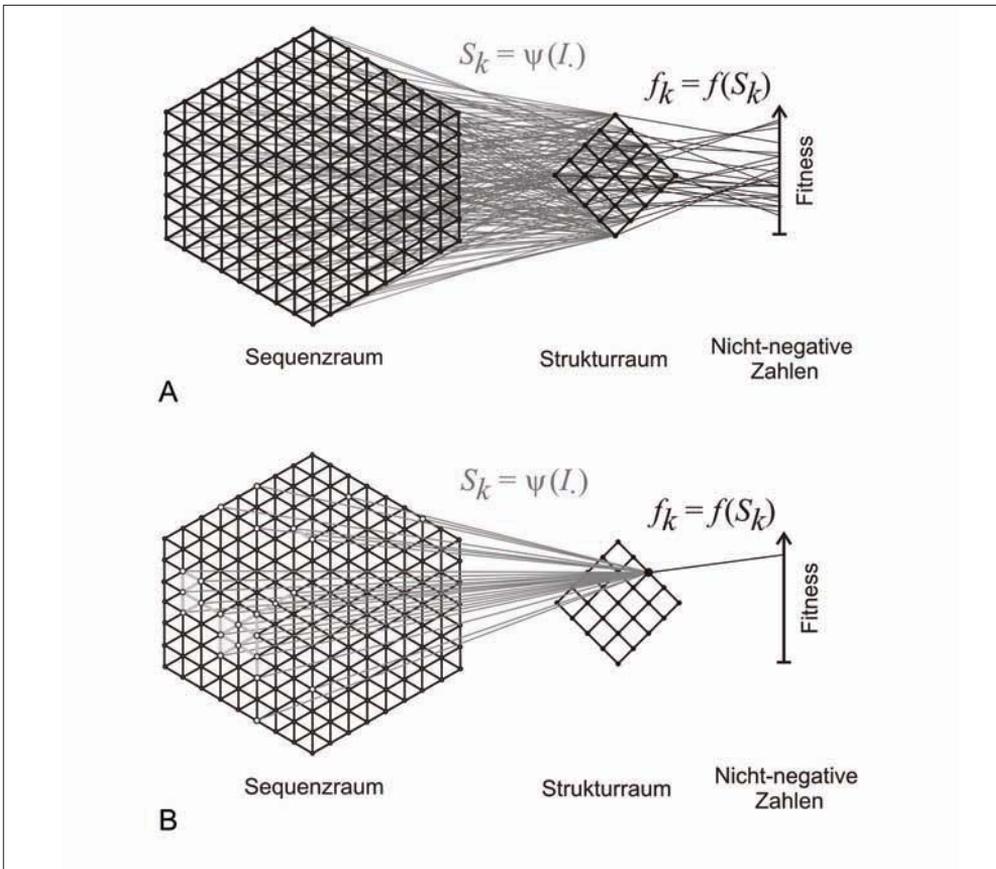


Abb. 7 Sequenzstruktur-Abbildung von RNA-Molekülen und neutrale Netzwerke. Die beiden Bilder zeigen schematisch die beiden konsekutiven Abbildungen aus dem Raum der (RNA-)Sequenzen in den Raum der (RNA-)Strukturen und von dort in die nicht-negativen reellen Zahlen am Beispiel von Fitnesswerten. Es gibt viel mehr Sequenzen als Strukturen, und daher ist die Sequenz-Strukturabbildung nicht eindeutig umkehrbar. Das untere Bild stellt das Urbild einer Struktur, S_k , im Raum der Sequenzen dar (B): Es besteht aus mehreren Sequenzen (hier 25, die durch schwarz umrandete weiße Punkte dargestellt sind), welche gemeinsam das neutrale Netzwerk der Struktur S_k bilden (nächste Nachbarn sind durch hellgraue Linien verbunden). Die Abbildung der Strukturen in die reellen Zahlen kann ebenso Neutralität aufweisen in dem Sinne, dass mehrere Strukturen dieselbe Fitness besitzen. Sequenz- und Strukturraum sind hier zum Zweck der Darstellung zweidimensional gezeichnet. In der Realität sind beide Räume hochdimensional.

Neutralität geht weit über *synonyme Codons* hinaus, und ein Grund dafür findet sich schon in der Sequenz-Strukturbeziehung von Biopolymeren. Im Allgemeinen gibt es sehr viel mehr Sequenzen als biologisch relevante Strukturen,¹⁸ und dies hat zur Konsequenz, dass mehrere

18 Jede Sequenz faltet in eine eindeutig bestimmte Struktur, wenn man alle Details hinreichend genau auflöst. Dessen ungeachtet gibt es viele sehr kleine Unterschiede in den atomar aufgelösten Strukturen, welche zu keinen messbaren Unterschieden in den Funktionen der Moleküle führen. Ein einprägsames Illustrationsbeispiel ist der Austausch eines Basenpaares in einer doppelhelikalen Region einer RNA durch ein anderes: Der Ersatz ist auf der Ebene der Atomkoordinaten merkbar, er spielt aber (fast immer) keine Rolle für die weniger detaillierte biologisch relevante Struktur des Moleküls.

Sequenzen auf eine Struktur abgebildet werden. Im besonderen Fall der RNA-Sekundärstrukturen kann die Sequenz-Strukturabbildung untersucht werden (Abb. 7; SCHUSTER et al. 1994, SCHUSTER 2006). Die Menge der Sequenzen, welche ein und dieselbe S(ekundär)struktur ausbilden, wird das neutrale Netzwerk der Struktur genannt, es stellt das Urbild einer Struktur im Raum der Sequenzen dar und entspricht formal einer (nicht eindeutigen) Umkehrung der Abbildung Ψ :

$$G_k = \Psi(S_k) \div \{X_i | \Psi(X_i) = S_k\}. \quad [18]$$

Neutrale Netzwerke sind jene Unterräume der Räume der Genotypen, in welchen die neutrale Evolution stattfindet (Abschnitt 5.2.3).

Ein weiteres Charakteristikum realistischer Fitnesslandschaften ist ihre *Zerklüftetheit* – *Ruggedness*: Im Sinne des Hammingabstandes oder einer Mutationsdistanz nahe beisammen liegende Sequenzen können sehr unterschiedliche oder ähnliche Fitness aufweisen und auch neutral sein. Verändert beispielsweise der Austausch einer einzigen Aminosäure das aktive Zentrum eines katalytisch wirksamen Enzyms, dann kann die Mutation letal sein, führt aber die Mutation zu einem synonymen Codon, dann wird sich kein bemerkbarer Effekt einstellen. Von gezielten Veränderungen an Proteinsequenzen ist bekannt, dass aufeinander folgende Mutationen nicht additiv sind. Es gibt verschiedentlich *kompensatorische Mutationen*, das sind Mutationen, welche den negativen Effekt einer vorangegangenen Mutation wettmachen, ohne den ursprünglichen Genotyp wieder herzustellen. Besonders einfach sind kompensatorische Mutationen an den Nukleotidpaaren in einer Doppelhelix zu verstehen: Mutieren wir beispielweise im Watson-Crick-Paar $\mathbf{A}=\mathbf{U}$ das \mathbf{U} , $\mathbf{U} \rightarrow \mathbf{C}$, für $\mathbf{A} \cdot \cdot \mathbf{C}$ ist keine Paarung in der Doppelhelix möglich, und die Helix wird destabilisiert. Eine weitere Mutation $\mathbf{A} \rightarrow \mathbf{G}$ kompensiert den, die Doppelhelix destabilisierenden, Effekt, da ein anderes Watson-Crick-Paar, $\mathbf{G}=\mathbf{C}$, gebildet wird. Ein reales Beispiel aus der Virologie betrifft die RNA des Hepatitisdelta-Virus (HDV; KUMAR et al. 1993). Kompensationseffekte werden bei Proteinstrukturen (WILSON et al. 1992), bei Proteinfaltung (NISHIMURA et al. 2005) und bei Wechselwirkungen zwischen Proteinen (JUCOVIC und HARTLEY 1996, MARTIN et al. 1999) beobachtet und analysiert. Kompensatorische Mutationen sind ein Grund für zerklüftete Landschaften, da die Fitness in der Folge der Mutationen hinunter und hinaufgeht. Eine weitere Ursache für *Ruggedness* resultiert aus der Nicht-Lokalität der Wechselwirkungen, welche für die Ausbildung der Strukturen von Biomolekülen maßgeblich sind. Eine einzige Punktmutation kann daher zu einer radikalen Änderung der molekularen Struktur Anlass geben. Nahezu alle häufig verwendeten einfachen Modelle für Fitnesslandschaften sind glatt und daher kritisch zu betrachten. Dies gilt insbesondere für die in der Populationsgenetik häufig verwendeten additiven und multiplikativen Fitnesslandschaften (Abb. 8), bei denen angenommen wird, dass jede Mutation die vom selektierten Genotyp wegführt eine (additive) konstante Verringerung der Fitness oder eine Multiplikation mit einem Faktor kleiner eins bewirkt. Auch *Single-Peak*-Fitnesslandschaften sind weit von der Realität entfernt. Die gewaltigen Steigerungen in der Computerleistung und in den Speichermöglichkeiten machen es heute möglich, in gewissem Ausmaß auch realistische, zerklüftete Landschaften zu untersuchen und mit empirischen Landschaften zu vergleichen (SCHUSTER 2010b). Ausgangspunkt für ein weitgehend realistisches Modell ist eine *Single-Peak*-Landschaft, welcher zufällig gewählte Fitnessbeiträge mit einer vorzuziehenden Bandbreite d überlagert werden (Abb. 8).

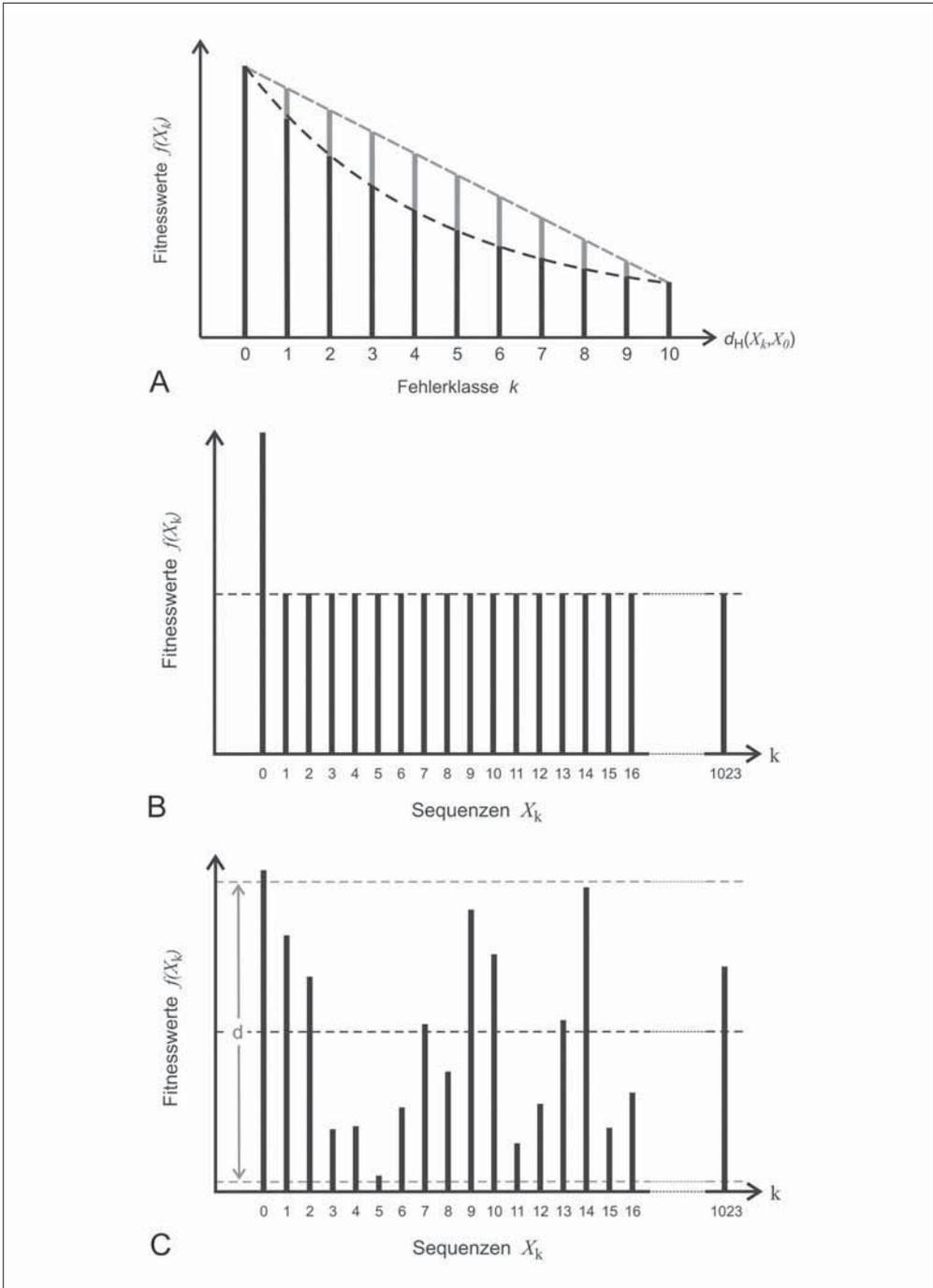


Abb. 8 Modelle für Fitnesslandschaften. (A) Das oberste Bild zeigt zwei Landschaften, welche in der Populationsgenetik häufig verwendet werden: (i) die additive Landschaft, in welcher jede Mutation, die von der Mastersequenz wegführt zu einer Abnahme des Fitnesswertes um einen konstanten Betrag führt, mit $f_k = f_0 - k \cdot a$ mit $k = d_H(X_0, X_i)$,

Eine Schlüsselfrage betrifft die Fehlerfortpflanzung auf den verschiedenen modellhaften und realitätsnahen Fitnesslandschaften. Interessanterweise sind es von den oben genannten Beispielen nur die zwei glatten Landschaftsmodelle, die additive und die multiplikative Landschaft, welche keine Fehlerschranken-artigen Übergänge aufweisen. Neben der *Single-Peak*-Landschaft zeigen alle hinreichend zerklüfteten Modelle Fehlerschranken, welche bei zunehmender Streuung der Fitnesswerte erhalten bleiben. Die Position der Fehlerschranke bewegt sich mit zunehmender Bandbreite d zu kleineren Mutationsraten. Die Rolle von neutralen Genotypen in der Mutations-Selektionsgleichung ist von besonderer Bedeutung, da sie einen zu KIMURAS neutraler Theorie alternativen Zugang zu den Mutantenverteilungen in Langzeitpopulationen vermittelt. Vor mehr als zwanzig Jahren konnte der Fall zweier neutraler Sequenzen X_i und X_j mit $f_i = f_j$ im Grenzfall verschwindender Mutationsraten exakt gelöst werden (SCHUSTER und SWETINA 1988). Das Ergebnis hängt vom Hammingabstand der beteiligten Genotypen ab und man erhält:

- $\lim_{p \rightarrow 0} \bar{x}_i(p) = \lim_{p \rightarrow 0} \bar{x}_j(p) = 0,5$ für nächste Nachbarn ($d_H(X_i, X_j) = 1$),
- $\lim_{p \rightarrow 0} \bar{x}_i(p) = \alpha$, $\lim_{p \rightarrow 0} \bar{x}_j(p) = 1 - \alpha$ für übernächste Nachbarn ($d_H(X_i, X_j) = 2$),
- $\lim_{p \rightarrow 0} \bar{x}_i(p) = 1$, $\lim_{p \rightarrow 0} \bar{x}_j(p) = 0$ oder $\lim_{p \rightarrow 0} \bar{x}_i(p) = 0$, $\lim_{p \rightarrow 0} \bar{x}_j(p) = 1$ für alle übrigen Fälle ($d_H(X_i, X_j) \geq 3$), wie es der neutralen Theorie entspricht.

Numerische Rechnungen auf realistischen Fitnesslandschaften mit Neutralität haben gezeigt, dass die genannten Resultate nicht nur im Grenzfall verschwindender Mutationsraten gelten, sondern fast exakt für den gesamten Bereich von $p = 0$ bis zur Fehlerschranke $p = p_{cr}$ gültig sind. Die beiden ersten äußern sich nachweisbar in Populationen insbesondere in Viruspopulationen dadurch, dass es keine eindeutige *Konsensussequenz*¹⁹ gibt. Gibt es mehr als zwei neutrale nächste Nachbarn, dann bildet sich ein stationärer Cluster von mengenmäßig stark gekoppelten Sequenzen aus. Die Konzentrationsverhältnisse werden durch den größten Eigenvektor der Adjazenzmatrix bestimmt. Für die Virologie haben derartige Cluster Bedeutung, da sie ein genetisches Reservoir für die Entwicklung pathogener Eigenschaften im Wirtorganismus darstellen.

5.2.3 Stochastische Simulation der evolutionären Optimierung

Eine ganze Reihe von stochastischen Phänomenen hat ganz entscheidenden Einfluss auf das evolutionäre Geschehen. Dementsprechend gibt es auch eine beachtliche Zahl von mathema-

19 Unter einer Konsensussequenz versteht man die gemittelte Sequenz einer Population. Bei einer Quasispezies ist die Konsensussequenz im allgemeinen (nahezu) identisch mit der Masterssequenz.

und (ii) die multiplikative Landschaft, in welcher die Fitness mit der Entfernung von der Mastersequenz durch konsequente Multiplikation mit einem Faktor kleiner 1 gebildet wird, mit $f_k = f_0 \cdot \beta^k$ mit $k = d_H(X_0, X_i)$. Für beide Fälle wurde im Bild $\lambda = 10$ gewählt. (B) Das mittlere Bild zeigt die *Single-Peak*-Landschaft für $\lambda = 10$ und binäre Sequenzen (und damit gilt $n = 1024$). Es gibt nur zwei Fitnesswerte, f_0 und f_k für $k \neq 0$. (C) Das unterste Bild zeigt schließlich eine Fitnesslandschaft, welche den aus der Molekularbiologie bekannten Fitnessverteilungen nahe kommt. Sie wurde aus der *Single-Peak*-Landschaft erzeugt durch Überlagerung mit zufälligen Inkrementen, die innerhalb einer Bandbreite d um f_k herum zentriert sind. Die natürlichen Werte sind zwar nicht zufallsverteilt, aber sie korrelieren nicht mit dem Hammingabstand von der Mastersequenz, sodass eine Überlagerung mit Beiträgen aus gleichverteilten Zufallszahlen die tatsächlichen Verhältnisse viel besser wiedergibt als jeder systematische Zusammenhang mit der Hammingdistanz.

tischen Ansätzen für die stochastische Analyse der Populationsdynamik (BLYTHE und MCKANE 2007). Hier beschränken wir uns auf Computersimulationen, welche von dem kinetischen Ansatz der Evolution in einem chemischen Reaktionsnetzwerk ausgehen und aus diesem ein System von Markov-Prozessen modellieren. Diesen stochastischen Prozessen entsprechen *chemische Mastergleichungen*, welche den Ausgangspunkt für die numerische Berechnung bilden (GARDINER 2009). Lösungen dieser Mastergleichungen können durch ein von Daniel GILLESPIE (2007) vorgeschlagenes und implementiertes Verfahren angenähert werden. Dabei wird ein Ensemble von einzelnen Computerläufen – üblicherweise *Trajektorien* genannt – Lauf für Lauf berechnet und statistisch ausgewertet. Für die stochastische Analyse wird das Mutations-Selektionssystem in einen Flussreaktor (CSTR = *Continuously Stirred Tank Reactor*) eingebettet, in welchem das verbrauchte Material durch Zufluss einer Vorratslösung ergänzt und die durch Replikation erzeugte Überzahl an RNA-Molekülen durch Abfluss ausgeglichen wird (FONTANA und SCHUSTER 1987, 1998, FONTANA et al. 1989).

Die Anwendung des Gillespieschen Algorithmus auf die Mutations-Selektionsgleichung führt zu einigen wesentlichen Ergänzungen des aus den Differentialgleichungen erhaltenen Bildes. Die hier präsentierten Daten stammen von Ansätzen zur Optimierung von RNA-Strukturen mit einer Zielvorgabe (FONTANA und SCHUSTER 1998), die Ergebnisse waren bei Läufen ohne vorgegebene Zielstrukturen (FONTANA et al. 1989) ganz ähnlich:

- Der stochastische Prozess benötigt eine minimale Populationsgröße, um für hinreichend lange Zeiten stabil zu bleiben und Evolution zu ermöglichen. In dem in Abbildung 9 beschriebenen Computerexperiment liegt die kritische Molekülzahl bei ungefähr $N = 20$: Für $N < 18$ stirbt die Population fast mit Wahrscheinlichkeit 1 aus, wogegen für $N > 22$ fast immer das Ziel erreicht wird.
- Die einzelnen Computerläufe streuen sehr stark, und zwar sind sie sowohl hinsichtlich der zwischenzeitlich durchlaufenen Strukturen als auch hinsichtlich der Dauer der Zielsuche und der Zahl der Replikationen (Tab. 2) stark divergent: Keine zwei Läufe weisen Zwischenstufen mit denselben Sequenzen und Strukturen auf.
- Die Optimierung verläuft nicht kontinuierlich, sondern in Stufen. Schnelle Phasen mit rascher Annäherung an die Zielstruktur wechseln mit zum Teil lange andauernden, quasi-stationären Perioden ab (Abb. 9).

Tab. 2 Statistik der Computerläufe zur Optimierung der Sekundärstruktur von RNA-Molekülen. Angegeben sind die Ergebnisse der Trajektorienstatistik von 20 bis 1000 einzelnen Läufen. Die angegebene Zeit ist die Echtzeit des Gillespie-Algorithmus in der gleichen Zeiteinheit wie die verwendeten Reaktionsgeschwindigkeitsparameter. Mittelwert und Standardabweichung wurden unter der Annahme einer an die Daten angepassten Log-Normalverteilung berechnet.

Populationsgröße N	Zeit bis zur Erreichung des Zieles		Zahl der Replikationen [10^7]			
	Mittelwert	Standardabweichung	Mittelwert	Standardabweichung		
1000	900	+ 1380	- 542	1,2	+ 3,1	- 0,9
2000	530	+ 880	- 330	1,4	+ 3,6	- 1,0
3000	400	+ 670	- 250	1,6	+ 4,4	- 1,2
10000	190	+ 230	- 100	2,3	+5,6	- 1,6
30000	110	+ 97	- 52	3,6	+6,8	- 2,3
100000	62	+ 50	- 28			

Die ausgeprägt starke Streuung zwischen den einzelnen Trajektorien findet eine zwangslose Erklärung durch die Analyse der Prozesse, welche auf den quasistationären Optimierungsplateaus ablaufen. Zwei verschiedene Szenarien wurden registriert:

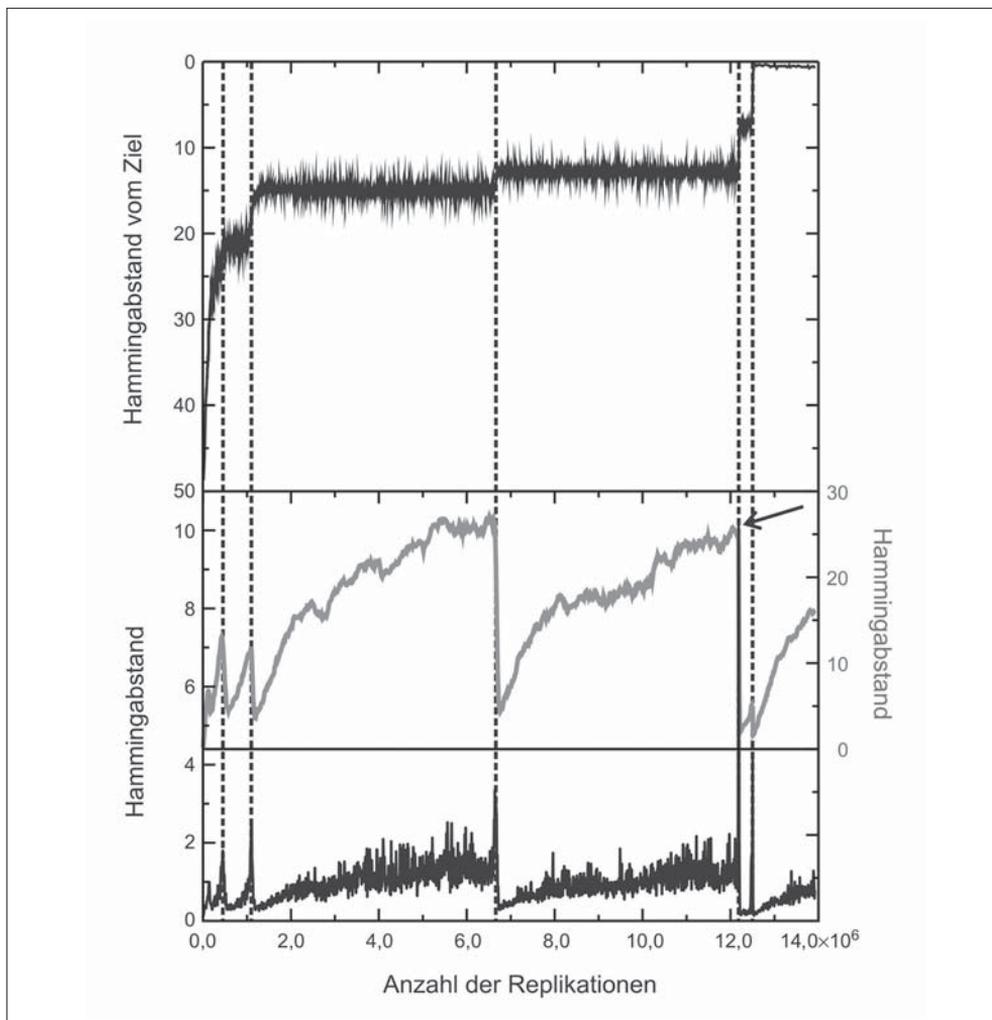


Abb. 9 Eine Trajektorie der evolutionären Optimierung von RNA-Strukturen. Die oberste Abbildung zeigt den mittleren Strukturabstand einer Population von $N = 1000$ RNA-Molekülen von der vorgegebenen Zielstruktur als Funktion der Anzahl durchgeführter Replikationen (schwarz). Dieser Abstand wird als Hammingdistanz zwischen den als ‚Strings‘ von Zeichen (Symbolketten) kodierten Strukturen erhalten und nimmt mit zunehmender Fitness im Sinne einer geringeren Distanz von der Zielstruktur ab. Die Abnahme ist nicht kontinuierlich, sondern erfolgt in stark ausgeprägten Stufen. Die Kurve in der Mitte (grau) zeigt die Breite der Population im Sequenzraum, die Kurve in der unteren Abbildung stellt die Geschwindigkeit dar, mit welcher der Schwerpunkt der Population durch den Sequenzraum driftet. Am Beginn einer jeden quasistationären Epoche der Evolution ist die Population schmal und breitet sich entlang des Fitnessplateaus laufend aus, bis sie am Ende des Plateaus kollapsartig schrumpft und wieder eine ähnliche Breite wie am Beginn der Epoche einnimmt. Die Erklärung des beobachteten Verhaltens ist einfach: Die Population vollzieht einen Diffusionsprozess auf einem neutralen Netzwerk, bis sie schließlich in einen Bereich kommt, von wo aus der Optimierungsprozess weiter fortgesetzt werden kann. Nach einer raschen Optimierungsphase, in welcher die Population ihrem Ziel näher kommt, ist ein Punkt erreicht, von dem aus keine weiteren Verbesserungen durch einzelne Mutationsschritte möglich sind, und die Diffusion beginnt von Neuem. Die Bewegung des Populationsschwerpunktes (unteres Bild) unterstützt diese Interpretation voll: Während der Ausbreitung am neutralen Netz ist nur eine sehr langsame und zufällige Bewegung des Schwerpunktes bemerkbar, aber am Ende des Plateaus tritt in bemerkenswerter Synchronisation mit dem Kollaps der Population eine sprunghafte Verschiebung des Schwerpunktes ein: Durch Se-

- Die in der Population dominierende Struktur²⁰ ist konstant, nur die Sequenzen variieren, und es findet neutrale Evolution im Sinne KIMURAS statt (HUYNEN et al. 1996). Der Prozess entspricht einem Irrflug auf dem neutralen Netzwerk der Struktur. Die Zahl der auf einem Plateau durchlaufenen neutralen Mutanten ist in sehr guter Näherung proportional zur Zahl der stattgefundenen Replikationen.
- Der Prozess auf einem Plateau involviert mehrere nahe verwandte Strukturen mit identischer Fitness und entspricht einem Irrflug im Raum dieser neutralen Strukturen. Beide Prozesse, Irrflug auf dem neutralen Netzwerk oder Irrflug im Raum der Strukturen, werden solange fortgesetzt, bis sich in der unmittelbaren Umgebung eine Sequenz höherer Fitness findet, ein Genotyp, mit welchem der Optimierungsprozess in Richtung auf das Ziel fortgesetzt werden kann. Diese Interpretation des Evolutionsvorganges wird stark unterstützt durch das Verhalten der Population während der quasistationären Perioden (Abb. 9): Die Ausdehnung der Population im Sequenzraum ist am Beginn eines Plateaus gering und nimmt bis zum Plateauende laufend zu, und die Population kollabiert dann in ganz kurzer Zeit. Der Populationsschwerpunkt bleibt während der quasistationären Epoche konstant und springt am Plateauende in eine neue Position. Mit anderen Worten heißt dies: Auf dem Plateau dehnt sich die Population in einer Art *Suchprozess* nach allen Richtungen aus, und, nachdem ein Punkt im Sequenzraum gefunden wurde, an welchem eine erfolgreiche Fortsetzung der evolutionären Optimierung möglich ist, schieft die gesamte Population dorthin und setzt die Optimierung fort. Dementsprechend fallen am Plateauende drei Ereignisse zusammen: eine kollapsartige Verringerung der Breite der Population, eine sprungartige Verschiebung des Populationsschwerpunkts und der Beginn einer neuen adaptiven Phase.

Es lohnt zu erwähnen, dass die Langzeitevolution einer Population von *Escherichia coli*-Bakterien (LENSKI et al. 1991) durch das Irrflugmodell auf neutralen Netzwerken interpretiert werden kann. Stufen auf der phänotypischen Ebene des Evolutionsprozesses wurden schon in einer frühen Phase des Jahrzehnte dauernden Experiments beobachtet (ELENA et al. 1996), wogegen die genotypischen Mutationen mit kontinuierlich verlaufen (PAPADOPOULOS et al. 1999). Nach 31 500 Generationen oder zwanzig Jahren wurde eine folgenschwere Innovation in einem (einzigen) der zwölf verschiedenen Anfangskolonien beobachtet (BLOUNT et al. 2008): Diese Kolonie entwickelte eine Art Kanal für die Aufnahme von Zitrat aus dem Nährmedium – Zitrat wird als Puffersubstanz verwendet und kann normalerweise von *E. coli* nicht aufgenommen werden – und erschloss dadurch eine neue Nahrungsquelle, was zu viel stärkerem Wachstum führte. Die Mutation in Richtung Zitrataufnahme wurde reproduziert. Frühere Isolate aus eben dieser Kolonie entwickelten früher oder später auch den vorteilhaften Zitratkanal, aber keine der anderen elf Kolonien war dazu in der Lage. Dieses Experiment demons-

20 Dominierend bezieht sich hier auf das Auftreten der Struktur in einer Folge von Strukturen, welche von der Anfangsstruktur zur Zielstruktur führen und Relay-Serie genannt wurde. (Für Einzelheiten siehe FONTANA und SCHUSTER 1998.)

lektion bewegt sich die Population in ganz wenigen Generationen dorthin, wo eine Sequenz höherer Fitness gefunden wurde. (Die Spitze der Verschiebung des Schwerpunktes beim vorletzten Sprung am rechten Rand des Bildes ist zur besseren Erkennbarkeit mit einem Pfeil gekennzeichnet.) Bei dem vorliegenden Computerexperiment wurde eine Mutationsrate von $p = 0,001$ gewählt. (Weitere Details finden sich in SCHUSTER 2006.)

triert unmissverständlich die Rolle von Kontingenz in der Evolution. Die Innovation kommt nicht durch eine unwahrscheinliche Mutation zustande, sondern durch eine Mutation mit üblicher Wahrscheinlichkeit von einer speziellen Gegend im Sequenzraum, in welche die eine Kolonie zum Unterschied von allen anderen elf gewandert ist. Das Ergebnis passt sehr gut zu den Ergebnissen der Computersimulationen auf Landschaften mit ausgedehnten neutralen Netzwerken. Wegen der großen Zahl unabhängiger Richtungen im Sequenzraum führt praktisch jede neue Computersimulation in eine andere Richtung, und es ist nahezu auszuschließen, dass zwei oder mehrere Simulationen dieselben Zwischenstrukturen durchlaufen. Geschichte ist ein essentielles Merkmal der Biologie, was von Theodosius DOBZHANSKY (DOBZHANSKY et al. 1977) im geflügelten Satz, „Nothing makes sense in biology except in the light of evolution“, ausgedrückt wurde.

5.3 Spieltheorie und Kooperation

Das Studium von Tiergesellschaften rückte in den Brennpunkt der Interessen der Biologen, als die empirisch gestützte Annahme von Beziehungen zwischen Verhalten und genetischer Prädisposition sinnvolle Modelle für die Evolution von Sozietäten ermöglichte. Ein Meilenstein der mathematischen Biologie war die Entwicklung eines spieltheoretischen Modells für die Konfliktaustragung durch John MAYNARD SMITH (MAYNARD SMITH und PRICE 1973, MAYNARD SMITH 1974). Der Grundgedanke für das Modell ist, wie gesagt, das Vorhandensein einer genetischen Komponente für Verhalten, welche eine Präferenz ihrer Träger für das Spielen einer bestimmten Verhaltensstrategie zur Folge hat. Eine erfolgreiche Strategie bewirkt dann eine größere Fitness der Träger im Vergleich zu den Trägern weniger erfolgreicherer Strategien, und sie wird sich mittels des Darwinschen Mechanismus in den zukünftigen Generationen ausbreiten. John MAYNARD SMITH prägte den Begriff der evolutionär stabilen Strategie (ESS), welche eine Verfeinerung des Nash-Gleichgewichts in der Ökonomie darstellt (NOWAK 2006): Wenn eine evolutionär stabile Strategie von einer gesamten Population gespielt wird, dann kann sie von keiner anderen Strategie unterwandert und abgelöst werden. Die Basis der spieltheoretischen Analyse von Auseinandersetzungen in den Gesellschaften bildet die Auszahlungsmatrix, $A = \{a_{ij}; i, j = 1, 2, \dots, n\}$. Ihre Elemente geben an, wie groß Gewinn oder Verlust der beiden Partner nach einem Zusammentreffen sein wird, in welchem sie die Strategien V_j beziehungsweise V_i spielten (Abb. 10).

Gemischte Strategien werden durch Linearkombinationen beschrieben:

$$V_p = \sum_{i=1}^n p_i V_i \quad V_q = \sum_{j=1}^n q_j V_j \quad \text{und daher} \quad a_{pq} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} q_j p_i. \quad [19]$$

Da die Matrix A nicht symmetrisch zu sein braucht, ist $a_{pq} \neq a_{qp}$. Die Bedingung für eine evolutionär stabile Strategie lässt sich nun ganz einfach formulieren. Die gemischte Strategie V_p ist eine ESS, wenn eine der beiden Bedingungen

$$a_{pp} > a_{qp} \quad \text{oder} \quad a_{pp} = a_{qp} \quad \text{und} \quad a_{pq} > a_{qq} \quad \text{für alle } V_q \quad [20]$$

erfüllt ist (MAYNARD SMITH 1974). In Worten ausgedrückt: Entweder erhält Strategie V_p gegen V_p gespielt eine höhere Auszahlung als eine beliebige Strategie V_q gegen V_p gespielt oder, wenn es Strategien V_q gibt, die gegen V_p ebenso gut abschneiden wie V_p , dann muss V_p gegen V_q gespielt mehr Auszahlung erhalten als V_q gegen V_q .

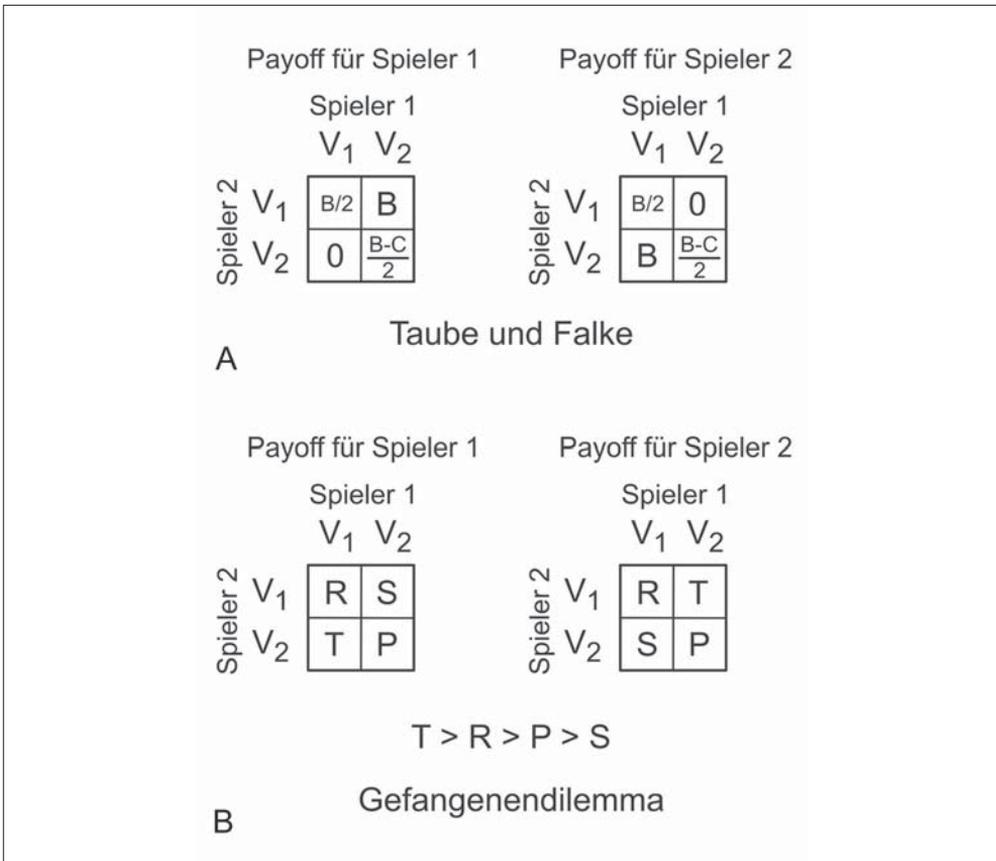


Abb. 10 Die Auszahlungsmatrix bei Spielen in der Verhaltensbiologie. Durch die Auszahlungsmatrix oder ‚Payoff‘-Matrix sind Gewinn und Verlust der beiden Spieler bei einer Auseinandersetzung festgelegt. Im oberen Spiel ‚Taube‘ (Strategie V_1) und ‚Falke‘ (Strategie V_2) streiten die beiden Partner um eine Beute im Werte B, das Risiko einer Verwundung bei einem Kampf wird mit $-C$ angesetzt. Treffen zwei ‚Tauben‘ aufeinander, dann hat jede dieselbe Chance auf den Gewinn (oder sie teilen die Beute). Trifft eine ‚Taube‘ auf einen ‚Falken‘, dann erhält letzterer die Beute, und die ‚Taube‘ geht leer aus. Zwei ‚Falken‘ beginnen zu kämpfen, und es besteht das Risiko einer Verwundung. Die ‚Payoff‘-Matrix für Spieler 2 ist die transponierte ‚Payoff‘-Matrix des Spielers 1, da außer den verschiedenen Strategien keine Unterschiede zwischen den Spielern angenommen werden. Im unteren Spiel wird das ‚Gefangenendilemma‘ modelliert (V_1 : Kooperation, V_2 : Defektspiel): R misst die Kosten der Untersuchungshaft für beide Partner, wenn sie kooperieren und nicht gestehen, T ist der ‚Payoff‘ für den Kronzeugen, der gesteht, während sein Partner die volle Strafe S absitzen muss, und P ist schließlich das gegenüber S günstigere Ergebnis, wenn beide gleichzeitig gestehen. Da $T > R$ und $P > S$ erfüllt sind, ist nicht zu kooperieren durch die höhere individuelle Auszahlung von beiden vorzuziehen, obwohl Kooperation für beide zusammen das global beste Ergebnis bietet.

Die Dynamik der spieltheoretischen Ansätze wurde auch als Differentialgleichungen formuliert, und damit kann die Evolution der Verteilung von Verhaltensstrategien in Populationen durch dynamische Spiele modelliert werden (HOFBAUER et al. 1979, SCHUSTER et al. 1981a–c, SCHUSTER und SIGMUND 1983, TAYLOR und JONKER 1978). Mit V_1, V_2, \dots, V_n bezeichnen wir die verschiedenen Strategien, und durch die Variablen x_1, x_2, \dots, x_n werden die Häufigkeiten beschrieben, mit welchen die jeweiligen Strategien in der Population gespielt werden. Für die zeitliche Entwicklung ergibt sich

$$\frac{dx_j}{dt} = x_j \left(\sum_{i=1}^n a_{ji} x_i - \varphi(t) \right) \text{ mit } \varphi(t) = \sum_{i=1}^n \sum_{k=1}^n a_{ki} x_i x_k \text{ und } \sum_{i=1}^n x_i = 1, \quad [21]$$

worin mit a_{ji} , wie oben angegeben, die Auszahlung an den Spieler bedeutet, welcher Strategie V_j spielt, während sein Gegner Strategie V_i gewählt hat. Diese Gleichung ist formal identisch mit der Fisherschen Selektionsgleichung. Ein fundamentaler Unterschied besteht jedoch hinsichtlich der Eigenschaften der Matrix A : Bei FISHER war die Matrix *per definitionem* symmetrisch, und dies hat alle Formen komplexer Dynamik ausgeschlossen. Die Auszahlungsmatrix A unterliegt jedoch keiner solchen Beschränkung, und die Lösungskurven der spieldynamischen Gleichung zeigen alle bekannten Phänomene nicht-linearer Differentialgleichungen. Ein anderer wichtiger Gesichtspunkt betrifft stabile Lösungen des spieltheoretischen Ansatzes und der Differentialgleichung. Es kann gezeigt werden, dass jede ESS einem asymptotisch stabilen Zustand der Differentialgleichung entspricht. Die Umkehr gilt allerdings nicht: Es gibt asymptotisch stabile Zustände, welchen keine ESS entspricht (HOFBAUER et al. 1979), und deshalb hat der spieldynamische Ansatz einen klaren Vorteil bei der Analyse stabiler Lösungen.

Die spieltheoretischen oder spieldynamischen Modelle behandeln im Wesentlichen zwei Fragestellungen:

- das Auftreten von Kampfritualen bei Auseinandersetzungen mit Gegnern derselben Art, die darauf abzielen, die Stärken der Konfliktpartner abzuschätzen und schwere Verletzungen zu vermeiden; und
- das Auftreten von altruistischem und kooperativem Verhalten in Gesellschaften.

Die Überlegungen zu Altruismus wurden durch das Prinzip der Verwandtenselektion – *Inclusive fitness* – von William HAMILTON (1964) auf eine überprüfbare quantitative Basis gestellt. Altruistische Unterstützung von Verwandten auf Kosten der eigenen Fitness lohnt sich evolutionär genetisch gesehen, wenn der Verwandtschaftsgrad in Bezug zu den Kosten groß genug ist. Die Gleichung HAMILTONS,

$$r b > c, \quad [22]$$

mit dem Verwandtschaftsgrad r , der die Wahrscheinlichkeit, ein Allel gemeinsam mit dem unterstützten Verwandten zu besitzen angibt;²¹ b ist der Nutzen des Verwandten, und c schließlich sind die Kosten für den Altruisten. Die Evolution von Kooperation und der Ursprung des Altruismus wurden zu Hauptthemen der biologischen Verhaltensforschung (AXELROD und HAMILTON 1981). Ausgehend von diesen ersten Ansätzen entstand eine neue Disziplin, die Soziobiologie (WILSON 2000).

Von besonderem Interesse für das Entstehen von Kooperation in Gesellschaften erwies sich ein als Gefangenendilemma bekanntes spieltheoretisches Problem: Zwei Gefangene haben gemeinsam ein Verbrechen begangen und sind in Untersuchungshaft. Die Anklage stützt sich ausschließlich auf Indizien und versucht, einen der beiden zu einem Geständnis zu überreden, wobei ihm eine Amnestie als Kronzeuge in Aussicht gestellt wird und der zweite Gefangene ein hohes Strafausmaß zu erwarten hat. Gestehen beide, haben sie eine mittlere Strafe zu ge-

21 Der Verwandtschaftsgrad ist einfach zu berechnen. Er ist 1 für eineiige Zwillinge, 0,5 zwischen Eltern und Kindern, 0,25 zwischen normalen Geschwistern, etc. Besondere Verwandtschaftsgrade treten bei Ameisen und Bienen auf, bei denen nur die Weibchen diploid sind, wogegen die Männchen nur einen Chromosomensatz tragen.

wärtigen. Wenn die beiden aber kooperieren und nicht gestehen, dann werden sie nach einiger Zeit wegen der Aussichtslosigkeit des Gerichtsverfahrens aus der Untersuchungshaft entlassen (Abb. 10). Das Dilemma besteht nun darin, dass Kooperieren die beste Lösung für die beiden Gefangenen darstellt, aber mit dem Risiko behaftet ist, dass der Gegner nicht kooperiert und der Verlust besonders hoch wird. Daher ist es die sichere Lösung für jeden der beiden Spieler, nicht zu kooperieren, obwohl dabei ein schlechteres Ergebnis in Kauf genommen werden muss. Das Gefangenendilemma wird zu einem interessanten wissenschaftlichen Problem, wenn es eine endliche, aber unbekannte Zahl von Auseinandersetzungen gibt. Für dieses als iteriertes Gefangenendilemma bekannte Problem wurde eine große Zahl von verschiedenen Strategien erdacht (SIGMUND 1995, NOWAK 2006). Es seien nur zwei einfache, aber recht erfolgreiche Strategien erwähnt:

- ‚Wie Du mir, so ich Dir‘ – ‚Tit for tat‘ – besteht darin, sich im nächsten Schritt genau so zu verhalten, wie der Partner im vorigen Spiel; und
- ‚Bei Gewinn beharren, bei Verlust wechseln‘ – ‚Win-stay, lose-shift‘ – wiederholt die letzte Strategie, wenn sie erfolgreich war, und wechselt bei Verlust.

Es gab in der Vergangenheit richtige Turniere zwischen verschiedenen Strategien im iterierten Gefangenendilemma (AXELROD 2006). Der Einfluss der räumlichen Ausbreitung von erfolgreichen Strategien wurde ebenfalls ausführlich studiert (NOWAK und MAY 1992, NOWAK et al. 1994). Der Einfachheit halber benutzten einige der dafür herangezogenen Computerrechnungen diskrete Zeitschritte und gleichzeitige Aktualisierung aller Zellen des zugrundeliegenden zellulären Automaten – ‚Synchronous updating‘. Eine sorgfältige Untersuchung (HUBERMAN und GLANCE 1993) hat gezeigt, dass drastische Unterschiede zwischen kontinuierlicher und diskreter Zeit oder synchroner und asynchroner Aktualisierung auftreten können. Um methodische Artefakte zu vermeiden, ist es daher angezeigt, diskrete Zeitschritte nur zu verwenden, wenn diese durch das biologische Umfeld vorgegeben werden, beispielsweise bei Modellen, welche Jahreszeiten kennen. Praktisch alle neueren Simulationen sind frei von diesen Artefakten und haben das wichtigste Resultat voll bestätigt: Durch die räumliche Dimension können Strategien, welche sonst aussterben würden, erhalten bleiben, und überlegene Strategien können verschwinden. Der Grund dafür besteht in erster Linie darin, dass die Ausbreitung von Mustern von den lokalen Verhältnissen abhängt, welche in der gesamten räumlichen Domäne sehr unterschiedlich verteilt sein können.

6. Perspektiven

Die Evolutionsforschung der Zukunft wird sich einigen Problemen widmen müssen, welche bis dato weitestgehend ausgeklammert waren. Die vordringliche Aufgabe ist eine neue Synthese in der Biologie, welche das zurzeit explodierende molekulare Wissen mit den empirischen Befunden der makroskopischen Biologie vereint. Am besten passt zu dieser Aufgabe ein Satz von John MAYNARD SMITH, den er in Bezug auf die leidige Holismus-Reduktionismus-Debatte geprägt hat (MAYNARD SMITH 1986): „What should be the attitude of a biologist working on whole organisms to molecular biology? It is, I think, foolish to argue that we (the macroscopic biologists) are discovering things that disprove molecular biology. It would be more sensible to say to molecular biologists that there are phenomena that they will one day have to interpret in their terms.“

Es scheint, dass dieser Tag nicht mehr allzu lange auf sich warten lassen wird. In der Systembiologie wird versucht, ganze Zellen oder sogar Organismen auf molekularer Ebene zu modellieren. Es wird dringend notwendig sein, eine echte Synthese der beiden Disziplinen durchzuführen, um eine fruchtbringende Ergänzung zu erreichen. Verwirrende molekulare Details und schwer zu kontrollierende Datenmengen stehen beispielsweise den zwar mathematisch präzise definierten, aber mechanistisch unklaren Effekten der Populationsgenetik gegenüber. Es wird der Ruf immer lauter nach einer einheitlichen, umfassenden theoretischen Biologie (BRENNER 1999), die naturgemäß um Evolutionstheorie und Entwicklungsbiologie herum zentriert sein müsste, ganz in Sinne einer Theorie für *Evo-devo* (*Evolution and development*). Beispiele äquivalenter oder fast äquivalenter Beschreibungen gibt es zu Genüge: Spieltheorie und Differentialgleichungen für dynamische Spiele wurden erwähnt (Anschnitt 5.3), die Populationsgenetik asexueller reproduzierender Organismen und die molekularkinetische Theorie der Evolution bilden einen weiteren Fall (WILKE 2005).

Es braucht nicht betont zu werden, dass die Biologie, insbesondere die gegenwärtige molekulare Biologie eine Wissenschaft des Komplexen schlechthin ist. Eine oft gestellte Frage betrifft daher den Ursprung von Komplexität in der biologischen Evolution. Aus der Sicht der kinetischen Evolutionstheorie kann diese Frage zumindest teilweise beantwortet werden. Die Differentialgleichungen, die stochastischen Prozesse und insbesondere ihre Lösungen mögen kompliziert sein, komplex sind sie nicht. Hier steckt die Komplexität in den Beziehungen zwischen Genotypen und Phänotypen, welche bei höheren Organismen auch die gesamte Entwicklung von der befruchteten Zelle bis zum erwachsenen Phänotyp mit einschließen. Bei der Evolution der überaus unübersichtlichen genetischen Regulationsnetzwerke kommt es zu Duplikationen von ganzen Genomen, wobei dann in der Folge der größte Teil der duplizierten Gene wieder verloren geht und die verbliebenen und neuen Gene in einer mehr oder minder zufälligen Reihenfolge hinterlässt (KELLIS et al. 2004). Man muss sich auch in Erinnerung rufen, dass der Löwenanteil des menschlichen Genomes in RNA-Moleküle transkribiert wird, von deren Funktionen wir zurzeit nur eine ganz vage Vorstellung haben (*Encode Project Consortium* 2007). Jedenfalls ist noch für viele Überraschungen in der zukünftigen Biologie gesorgt, Mathematik und Computersimulation werden ein ungeheuer weites Betätigungsfeld finden.

Anmerkung nach Redaktionsschluss

Nach Fertigstellung des Manuskriptes ist im Wissenschaftsjournal *Nature* eine augenscheinlich grundlegende Arbeit zu dem auf der Seite 204 angesprochenen Thema der „Verwandtenselektion – Inclusive Fitness“ erschienen (NOWAK et al. 2010), die in der soziobiologischen *Community* einiges Aufsehen erregt und eine Menge Widerspruch geerntet hat (OKASHA 2010). Die Autoren zeigen mit Hilfe einer sorgfältigen mathematischen Analyse, dass der Begriff der *Verwandtenselektion* entbehrlich erscheint und die beobachteten Fälle von Altruismus durch eine „richtig angewandte“ konventionelle Theorie der natürlichen Selektion einfacher und besser erklärt werden können.

Literatur

- ADAMI, C.: Learning and complexity in genetic auto-adaptive systems. *Physica D* 80, 154–170 (1995)
- ADAMI, C.: Introduction to Artificial Life. New York: Springer 1998
- ALBERTS, B., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K., and WALTER, P.: Molecular Biology of the Cell. New York: Garland Science, Taylor & Francis Group 2008
- APPEL, K., and HAKEN, W.: Every map is four colourable. Part I: Discharging. *Illinois Journal of Mathematics* 21, 429–290, Part II: Reducibility, *ibid.* 21, 491–567 (1977)
- AXELROD, R.: The Evolution of Cooperation. New York: Basic Books, Perseus Books Group 2006
- AXELROD, R., and HAMILTON, W. D.: The evolution of cooperation. *Science* 211, 1390–1396 (1981)
- BIEBRICHER, C. K., EIGEN, M., and GARDINER, W. C. Jr.: Kinetics of RNA replication. *Biochemistry* 22, 2544–2559 (1983)
- BLOUNT, Z. D., BORLAND, C. Z., and LENSKI, R. E.: Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 105/33, 7899–7906 (2008)
- BLYTHE, R. A., and Mc KANE, A. J.: Stochastic models of evolution in genetics, ecology and linguistics. *J. Statistical Mechanics: Theory and Experiment* 2007:P07018 (2007)
- BRAKMANN, S., and JOHNSON, K.: Directed Molecular Evolution of Proteins or How to Improve Enzymes for Biocatalysis. Weinheim (Bergstraße): Wiley-VCh DE 2002
- BRENNER, S.: Theoretical biology in the third millennium. *Phil. Trans. Roy. Soc. London B* 354, 1963–1965 (1999)
- BULL, J. J., ANCEL MYERS, L., and LACHMANN, M.: Quasispecies made simple. *PLoS Comp. Biol.* 1, 450–460 (2005)
- BULL, J. J., SANJUÁN, R., and WILKE, C. O.: Theory of lethal mutagenesis for viruses. *J. Virology* 81, 2930–2939 (2007)
- CARNOT, S.: Réflexions sur la puissance du feu et sur les machines propres à développer cette puissance. Paris: Bacheliers 1824; und englische Übersetzung: THURSTON, R. H. (Ed.): Reflections on the Motive Power of Heat and on Machines Fitted to Develop this Power. New York: John Wiley & Sons 1890
- CASTI, J. L.: Would-be-worlds. New York: John Wiley & Sons 1997
- CRAMER, C. J.: Essentials of Computational Chemistry. Theories and Models. Chichester, UK: John Wiley & Sons 2002
- DARWIN, C.: On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. London: John Murray 1859
- DEWDNEY, A. K.: In the game called ‘Core War’ hostile programs engage in a battle of bits. *Scientific American* 250/5, 15–19 (1984)
- DEWDNEY, A. K.: A ‘Core War’ bestiary of viruses, worms and other threats to computer memories. *Scientific American* 252/3, 14–19 (1985)
- DEWDNEY, A. K.: A program called ‘MICE’ nibbles its way into victory at the first ‘Core War’ tournament. *Scientific American* 250/1, 8–11 (1987)
- DIRAC, P. A. M.: Quantum mechanics of many-electron systems. *Proc. Roy. Soc. London A* 123, 714–733 (1929)
- DOBZHANSKY, T., AYALA, F. J., STEBBINS, G. L., and VALENTINE, J. W.: Evolution. San Francisco: W. H. Freeman and Company 1977
- DOMINGO, E. (Ed.): Virus entry into error catastrophe as a new antiviral strategy. *Virus Research* 107/2, 115–228 (2005)
- DOMINGO, E. (Ed.): Quasispecies: Concepts and Implications for Virology. Berlin: Springer 2006
- DOMINGO, E., PARRISH, C. R., and HOLLAND, J. J. (Eds.): Origin and Evolution of Viruses. 2nd ed. Amsterdam, NL: Elsevier 2008
- DRAKE, J. W.: Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. USA* 90, 4171–4175 (1993)
- DRAKE, J. W., CHARLESWORTH, R., CHARLESWORTH, D., and CROW, J. F.: Rates of spontaneous mutation. *Genetics* 148, 1667–1686 (1998)
- DRAKE, J. W., and HOLLAND, J. J.: Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. USA* 96, 13910–13913 (1999)
- DUNLAP, R. A.: The Golden Ratio and Fibonacci Numbers. Singapore: World Scientific 1997
- EIGEN, M.: Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–523 (1971)
- EIGEN, M., MCCASKILL, J., and SCHUSTER, P.: The molecular quasispecies. *Adv. Chem. Phys.* 75, 149–263 (1989)
- EIGEN, M., and SCHUSTER, P.: The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften* 64, 541–565 (1977)

- EIGEN, M., and SCHUSTER, P.: The hypercycle. A principle of natural self-organization. Part B: The abstract hypercycle. *Naturwissenschaften* 65, 7–41 (1978a)
- EIGEN, M., and SCHUSTER, P.: The hypercycle. A principle of natural self-organization. Part C: The realistic hypercycle. *Naturwissenschaften* 65, 341–369 (1978b)
- EIGEN, M., and SCHUSTER, P.: *The Hypercycle. A Principle of Natural Self-Organization*. Berlin: Springer 1979
- ELENA, S. F., COOPER, V. S., and LENSKI, R. E.: Punctuated evolution caused by selection of rare beneficial mutations. *Science* 272, 1802–1804 (1996)
- ELENA, S. F., and LENSKI, R. E.: Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nature Rev. Genet.* 4, 457–469 (2003)
- Encode Project Consortium*: Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816 (2007)
- EULER, L.: *Introductio in Analysin Infinitorum*. Vol. I und II. 1748. Ins Deutsche übertragen von H. MASER. Einleitung in die Analysis des Unendlichen. Erster und Zweiter Teil. Berlin: Julius Springer 1885
- FISHER, R. A.: *The Genetical Theory of Natural Selection*. Oxford, UK: Clarendon Press 1930
- FISHER, R. A.: Average excess and average effect of a gene substitution. *Annals of Eugenics* 11, 53–63 (1941)
- FONTANA, W., SCHNABL, W., and SCHUSTER, P.: Physical aspects of evolutionary optimization and adaptation. *Phys. Rev. A* 40, 3301–3321 (1989)
- FONTANA, W., and SCHUSTER, P.: A computer model of evolutionary optimization. *Biophys. Chem.* 26, 123–147 (1987)
- FONTANA, W., and SCHUSTER, P.: Continuity in evolution. On the nature of transitions. *Science* 280, 1451–1455 (1998)
- GÁCS, P.: Reliable cellular automata with self-organization. *J. Statist. Phys.* 103, 45–267 (2001)
- GALLUCCIO, S.: Exact solution of the quasispecies model in a sharply peaked fitness landscape. *Phys. Rev. E* 56, 4526–4539 (1997)
- GARDINER, C. W.: *Stochastic Methods. A Handbook for the Natural and Social Sciences*. Berlin: Springer 2009
- GARDNER, M.: Mathematical Games. The fantastic combinations of John Conway's new solitaire game "life". *Scientific American* 223/10, 120–123 (1967)
- GALILEI, G.: *Il Saggiatore*. Edition Nationale. Bd. 6. Florenz 1896
- GIERER, A., and MEINHARDT, H.: A theory of biological pattern formation. *Kybernetik* 12, 30–39 (1972)
- GILLESPIE, D. T.: Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* 58, 35–55 (2007)
- GONTHIER, G.: Formal proof – The four-color theorem. *Notices Amer. Math. Soc.* 55, 1382–1393 (2008)
- GORDON, R., and BELOUSSOV, L.: From observations to paradigms. The importance of theories and models. An interview with Hans Meinhardt. *Int. J. Dev. Biol.* 50, 103–111 (2006)
- HAMILTON, W. D.: The genetic evolution of social behavior. *J. Theoret. Biol.* 7, 1–16 (1964)
- HAMMING, R. W.: Error detecting and error correcting codes. *Bell. Syst. Tech. J.* 29, 147–160 (1950)
- HAMMING, R. W.: *Coding and Information Theory*. 2nd ed. Englewood Cliffs, N J.: Prentice Hall 1989
- HEINEMANN, C.: Artificial Life Environment. Simulation evolutionärer Prozesse am Computer. *Informatik Spektrum* 31/1, 55–61 (2008)
- HODGKIN, A. L., and HUXLEY, A. F.: A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544 (1952)
- HOFBAUER, J., SCHUSTER, P., and SIGMUND, K.: A note on evolutionary stable strategies and game dynamics. *J. Theoret. Biol.* 81, 609–612 (1979)
- HOFBAUER, J., and SIGMUND, K.: *Evolutionary Games and Replicator Dynamics*. Cambridge, UK: Cambridge University Press 1998
- HOLLEY, R. W., APGAR, J., EVERETT, G. A., MADISON, J. T., MARQUISEE, M., MERRILL, S. H., PENSWICK, J. R., and ZAMIR, A.: Structure of a ribonucleic acid. *Science* 147, 1462–1465 (1965)
- HOPFIELD, J. J.: Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* 79, 2554–2558 (1982)
- HUBERMAN, B. A., and GLANCE, N. S.: Evolutionary games and computer simulations. *Proc. Natl. Acad. Sci. USA* 90, 7716–7718 (1993)
- HUYNEN, M. A., STADLER, P. F., and FONTANA, W.: Smoothness within ruggedness. The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA* 93, 397–401 (1996)
- ILACHINSKI, A.: *Cellular Automata. A Discrete Universe*. New Jersey: World Scientific 2001
- JONES, B. L., ENNS, R. H., and RANGNEKAR, S. S.: On the theory of selection of coupled macromolecular systems. *Bull. Math. Biol.* 38, 15–28 (1976)
- JOYCE, G. F.: Forty years of *in vitro* evolution. *Angew. Chem. Int. Ed.* 46, 6420–6436 (2007)

- JUCOVIC, M., and HARTLEY, R.: Protein-protein interaction: A genetic selection for compensating mutations at the barnase-barstar interface. *Proc. Natl. Acad. Sci. USA* 93, 2343–2347 (1996)
- KANDEL, E. R., SCHWARTZ, J. H., and JESSELL, T. M.: Principles of Neural Science. 4th ed. New York: McGraw-Hill 2000
- KELLIS, M., BIRREN, B. W., and LANDER, E. S.: Proof and evolutionary analysis of ancient gene duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428, 617–624 (2004)
- KENDREW, J. C., DICKERSON, R. E., STRANDBERG, B. E., HART, R. G., and DAVIES, D. R.: Structure of myoglobin. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* 185, 422–427 (1960)
- KIM, S. H., SUDDATH, F. L., QUIGLEY, G. J., MCPHERSON, A., SUSSMANN, J. L., WANG, A. H. J., SEEMAN, N. C., and RICH, A.: Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* 185, 435–439 (1974)
- KIMURA, M.: The Neutral Theory of Evolution. Cambridge, UK: Cambridge University Press 1983
- KLEINERT, A.: Der messende Luchs. Zwei verbreitete Fehler in der Galilei-Literatur. *NTM – Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin* 17, 199–206 (2009)
- KLUSSMANN, S.: The Aptamer Handbook. Functional Oligonucleotides and Their Applications. Weinheim: Wiley-VCH, (Bergstraße) 2006
- KUMAR, P. K. R., SUH, Y.-A., TAIRA, K., and NISHIKAWA, S.: Point and compensating mutations to evaluate essential stem structures of genomic HDV ribozyme. *FASEB J.* 7, 124–129 (1993)
- LENSKI, R. E.: The future of evolutionary biology. *Ludus vitalis* XII 21, 67–89 (2004)
- LENSKI, R. E., OFRIA, C., COLLIER, T. C., and ADAMI, C.: Genome complexity, robustness, and genetic interactions in digital organisms. *Nature* 400, 661–664 (1999)
- LENSKI, R. E., OFRIA, C., PENNOCK, R. T., and ADAMI, C.: The evolutionary origin of complex features. *Nature* 423, 139–144 (2003)
- LENSKI, R. E., ROSE, M. R., SIMPSON, S. C., and TADLER, S. C.: Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *The American Naturalist* 138, 1315–1341 (1991)
- LOTKA, A. J.: Elements of Physical Biology. Philadelphia, PA.: Williams & Wilkins 1925
- MALTHUS, T. R.: An Essay on the Principle of Population. Oxford World's Classics Reprint. 1798
- MARTIN, C., HARTLEY, R., and MAUGUEN, Y.: X-ray structural analysis of compensating mutations at the barnase-barstar interface. *FEBS Lett.* 452, 128–132 (1999)
- MARTÍ-RENO, M. A., STUART, A. C., FISER, A., SÁNCHEZ, R., MELO, F., and ŠALI, A.: Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291–325 (2000)
- MAXAM, A., and GILBERT, W.: A new method of sequencing DNA. *Proc. Natl. Acad. Sci. USA* 74, 560–564 (1977)
- MAYNARD SMITH, J.: The theory of games and the evolution of animal conflicts. *J. Theoret. Biol.* 47, 209–221 (1974)
- MAYNARD SMITH, J.: Evolution and the Theory of Games. Cambridge, UK: Cambridge University Press 1982
- MAYNARD SMITH, J.: The Problems of Biology. Oxford, UK: Oxford University Press 1986
- MAYNARD SMITH, J.: Byte-sized evolution. *Nature* 355, 772–773 (1992)
- MAYNARD SMITH, J., and PRICE, G. R.: Logic of animal conflict. *Nature* 246, 15–18 (1973)
- MAYR, E.: The Growth of Biological Thought. Diversity, Evolution, and Inheritance. Cambridge, MA: The Belknap Press of Harvard University Press 1982
- MEINHARDT, H.: Models of Biological Pattern Formation. London: Academic Press 1982
- MENDEL, G.: Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereins in Brünn* 4, 3–47 (1866)
- MENDEL, G.: Über einige aus künstlicher Befruchtung gewonnenen Hieracium-Bastarde. *Verhandlungen des naturforschenden Vereins in Brünn* 8, 26–31 (1870)
- MILLS, D. R., PETERSON, R. L., and SPIEGELMAN, S.: An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. USA* 58, 217–224 (1967)
- MOUNT, D. W.: Bioinformatics. Sequence and Genome Analysis. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press 2004
- NEEDLEMAN, S. B., and WUNSCH, C. D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48, 443–453 (1970)
- NEET, K. E., and LEE, J. C.: Biophysical characterization of proteins in the post-genomic era of proteomics. *Mol. Cell. Proteomics* 1, 415–420 (2002)
- NEWTON, I.: Philosophiae naturalis principia mathematica. Regiae Societas. 1686 Englische Übersetzung von A. MOTTE. *Mathematical Principles of Natural Philosophy*. New York: Daniel Adee 1846
- NISHIMURA, C., LIETZOW, M. A., DYSON, H. J., and WRIGHT, P. E.: Sequence determinants of a protein folding pathway. *J. Mol. Biol.* 351, 383–392 (2005)
- NOWAK, M. A.: Evolutionary Dynamics. Exploring the Equations of Life. Cambridge, MA: The Belknap Press of Harvard University Press 2006

- NOWAK, M. A., BONHOEFFER, S., and MAY, R. M.: Spatial games and the maintenance of cooperation. *Proc. Natl. Acad. Sci. USA* *91*, 4877–4881 (1994)
- NOWAK, M. A., and MAY, R. M.: Evolutionary games and spatial chaos. *Nature* *359*, 826–829 (1992)
- NOWAK, M. A., TARNITA, C. E., and WILSON, E. O.: The evolution of eusociality. *Nature* *466*, 1057–1062 (2010)
- OHTA, T., and GILLESPIE, J. H.: Development of neutral and nearly neutral theories. *Theoretical Population Biology* *49*, 128–142 (1996)
- OHTA, T.: Near-neutrality in evolution of genes and gene regulation. *Proc. Natl. Acad. Sci. USA* *99*, 16134–16137 (2002)
- OKASHA, S.: Fisher’s fundamental theorem of natural selection – A philosophical analysis. *British Journal for the Philosophy of Science* *59*, 319–351 (2008)
- OKASHA, S.: Altruism researchers must cooperate. *Nature* *467*, 653–655 (2010)
- O’NEILL, B.: Digital evolution. *PloS Biology* *1*, 11–14 (2003)
- ONUCHIC, J. N., LUTHEY-SCHULTEN, Z., and WOLYNES, P. G.: Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* *48*, 545–600 (1997)
- PAPADOPOULOS, D., SCHNEIDER, J., MEIER-EISS, J., ARBER, W., LENSKI, R. E., and BLOT, M.: Genomic evolution during a 10,000-generation experiment with bacteria. *Proc. Natl. Acad. Sci. USA* *96*, 3807–3812 (1999)
- PERUTZ, M. F., ROSSMANN, M. G., CULLIS, A. F., MUIRHEAD, H., WILL, G., and NORTH, A. C. T.: Structure of haemoglobin. A three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. *Nature* *185*, 416–422 (1960)
- PHIZCKY, E., BASTIAENS, P. I. H., ZHU, H., SNYDER, M., and FIELDS, S.: Protein analyses on a proteomic scale. *Nature* *422*, 208–215 (2003)
- RAMACHANDRAN, K. I., DEEPA, G., and NAMBOORI, K.: *Computational Chemistry and Molecular Modeling: Principles and Applications*. Berlin: Springer 2008
- RASMUSSEN, S., KNUDSEN, C., FELDBERG, R., and HINDSHOLM, M.: The coreworld: Emergence and evolution of cooperative structures in a computational chemistry. *Physica D* *42*, 111–134 (1990)
- RAY, T. S.: An approach to the synthesis of life. In: LANGTON, C. G., TAYLOR, C., FARMER, J. D., and RASMUSSEN, S. (Eds.): *Artificial Life II*; pp. 371–408. Redwood City, CA: Addison-Wesley 1991
- REIDYS, C. M., STADLER, P. F., and SCHUSTER, P.: Generic properties of combinatorial landscapes: Neutral networks of RNA secondary structures. *Bull. Math. Biol.* *59*, 339–397 (1997)
- REIDYS, C. M., and STADLER, P. F.: Combinatorial landscapes. *SIAM Review* *44*, 3–54 (2002)
- ROBERTUS, J. D., LADNER, J. E., FINCH, J. T., RHODES, D., BROWN, R. S., CLARK, B. F. C., and KLUG, A.: Structure of yeast phenylalanyl tRNA at 3 Å resolution. *Nature* *250*, 546–551 (1974)
- SANGER, F., NICKLEN, S., and COULSON, A.: DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* *74*, 5463–5467 (1977)
- SHOCH, J. J., and HUPP, J. A.: The ‘worm’ programs – Early experience with a distributed computation. *Communications of the ACM* *25*, 172–180 (1982)
- SCHUSTER, P.: Prediction of RNA secondary structures: From theory to models and real molecules. *Reports on Progress in Physics* *69*, 1419–1477 (2006)
- SCHUSTER, P.: Genotypes and phenotypes in the evolution of molecules. *European Review* *17*, 281–319 (2009)
- SCHUSTER, P.: Mathematical modeling of evolution. Solved and open problems. *Theory Biosci.* (in press, 2010a)
- SCHUSTER, P.: Quasispecies and error threshold on realistic rugged landscapes. Preprint 2010b
- SCHUSTER, P., FONTANA, W., STADLER, P. F., and HOFACKER, I. L.: From sequences to shapes and back – A case study in RNA secondary structures. *Proc. Roy. Soc. London B* *255*, 279–284 (1994)
- SCHUSTER, P., and SIGMUND, K.: Replicator dynamics. *J. Theor. Biol.* *100*, 533–538 (1983)
- SCHUSTER, P., SIGMUND, K., HOFBAUER, J., and WOLFF, R.: Selfregulation of behaviour in animal societies. I. Symmetric contests. *Biol. Cybernetics* *40*, 1–8 (1981a)
- SCHUSTER, P., SIGMUND, K., HOFBAUER, J., WOLFF, R., GOTTLIEB, R., and MERZ, P.: Selfregulation of behaviour in animal societies. II. Games between two populations without selfinteraction. *Biol. Cybernetics* *40*, 9–15 (1981b)
- SCHUSTER, P., SIGMUND, K., HOFBAUER, J., WOLFF, R., GOTTLIEB, R., and MERZ, P.: Selfregulation of behaviour in animal societies. III. Games between two populations with selfinteraction. *Biol. Cybernetics* *40*, 17–25 (1981c)
- SCHUSTER, P., and SWETINA, J.: Stationary mutant distributions and evolutionary optimization. *Bull. Math. Biol.* *50*, 636–660 (1988)
- SENETA, E.: *Non-negative matrices and Markov Chains*. 2nd edit. New York: Springer 1981
- SIGMUND, K.: *Games of Life. Explorations in Ecology, Evolution and Behavior*. New York: Penguin 1995
- SMITH, T. F., and WATERMAN, M. S.: Identification of common subsequences. *J. Mol. Biol.* *147*, 195–197 und Comparison of biosequences. *Adv. Appl. Math.* *2*, 482–489 (1981)

- SPIEGELMAN, S.: An approach to the experimental analysis of precellular evolution. *Quart. Rev. Biophys.* 4, 213–253 (1971)
- SWETINA, J., and SCHUSTER, P.: Self-replication with errors. A model for polynucleotide replication. *Biophys. Chem.* 16, 329–354 (1982)
- TAYLOR, P. D., and JONKER, L. B.: Evolutionary stable strategies and game dynamics. *Math. Biosciences* 40, 145–156 (1978)
- THOMPSON, C. J., and MCBRIDE, J. L.: On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules. *Math. Biosci.* 21, 127–142 (1974)
- TURING, A. M.: The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. London B* 237, 37–72 (1952)
- VERHULST, P. F.: Notice sur la loi que la population poursuit dans son accroissement. *Correspondance mathématique et physique* 10, 113–121, und 1845. *Recherches mathématiques sur la d'accroissement da la population.* *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles* 18 (Art. 1), 1–45 (1838)
- VOLTERRA, V.: Variazioni et fluttuazioni del numero d'individui in specie animali conviventi. *Mem. R. Acad. Naz. dei Lincei Ser.VI*, 2, 31–113 (1926)
- WATSON, J. D., and CRICK, F. H. C.: A structure for deoxyribonucleic acid. *Nature* 171, 737–738 (1953)
- WATTS, A., and SCHWARZ, G. (Eds.): *Evolutionary biotechnology – From theory to experiment.* *Biophys. Chem.* 66, 67–284 (1997)
- WEIMAR, J. R.: Cellular automata for reaction-diffusion systems. *Parallel Computing* 23, 1699–1715 (1997)
- WILKE, C.O.: Quasispecies theory in the context of population genetics. *BMC Evolutionary Biology* 5, 44 (2005)
- WILKE, C. O., and ADAMI, C.: The biology of digital organisms. *Trends Ecol. Evol.* 17, 528–532 (2002)
- WILKE, C. O., WANG, L. J., OFRIA, C., LENSKI, R. E., and ADAMI, C.: Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* 412, 331–333 (2001)
- WILSON, E. O.: *Sociobiology. The New Synthesis.* 25th anniversary ed. Cambridge, MA: The Belknap Press of Harvard University Press 2000
- WILSON, K. P., MALCOLM, B. A., and MATHEWS, B. W.: Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme. *J. Biol. Chem.* 267, 10842–10849 (1992)
- WRIGHT, S.: The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: JONES, D. F. (Ed.): *Proceedings of the Sixth International Congress on Genetics*; pp. 356–366, Vol. 1. Ithaca, NY: Brooklyn Botanic Garden 1932
- WRIGHT, S.: Surfaces of selective value revisited. *The American Naturalist* 131, 115–131 (1988)
- YARMUSH, M. L., and JAYARMAN, A.: Advances in proteomic technologies. *Annu. Rev. Biomed. Eng.* 4, 349–373 (2002)

Prof. Dr. Peter SCHUSTER
Universität Wien
Institut für Theoretische Chemie
Währingerstraße 17
1090 Wien
Österreich
Tel.: +43 1 427752743
Fax: +43 1 427752793
E-Mail: pks@tbi.univie.ac.at

