

Selektion und Mutation aus informationstheoretischer Sicht – Grenzen der Erhaltbarkeit von genetischer Information

Selection and Mutation from the Informationtheoretical Point of View –
Limits of the Preservation of Genetical Information

Wolfgang Tiefenbrunner

Bundesamt und Forschungszentrum für Landwirtschaft, Spargelfeldstr. 191, A-1220 Wien
Z. Naturforsch. **50c**, 883–894 (1995); eingegangen am 23. Dezember 1994/15. August 1995
Mutation, Selection, Error Threshold, Molecular Evolutionary Rate, Information Theory

There is no doubt that evolution can only operate in accordance with the laws of thermodynamics. To transform this simple perception into a statement that can be quantified it is necessary to examine how the entropy (or the information content) changes by means of the two main factors of evolution: selection and mutation. In the present paper this aim was reached by using the methods of information theory.

That evolution becomes possible the change of information by mutation and selection must not be negativ in the long run. Consequently an upper limit for the information content of the genom must be assumed. If the evolution as a self-optimizing process gains this upper limit it reaches a dynamic equilibrium at which the increase of entropy by mutation balances the decrease of entropy by selection. In particular it was examined under which circumstances of dynamic equilibrium mastergenotypes (genotypes that can preserve their specific sequence over a long time) can exist.

Einleitung

Anfang der 70iger Jahre setzte sich die Erkenntnis durch, daß Mutationen keineswegs seltene – und daher für die Populationsgenetik eher unbedeutende – Ereignisse sind (bis dahin wurde die Bedeutung der Mutationen allein in der Erzeugung neuer genetischer Varianten gesehen). Dann aber zeigten Überlegungen zur präbiotischen Evolution (Eigen, 1971), die in die Quasispeziesstheorie mündeten, daß Mutationen wahrscheinlich auch heute noch wesentlich häufiger sind, als ursprünglich angenommen.

Kernelement dieser Theorie ist die „Fehlerschwelle“, die die Voraussetzungen für die Erhaltbarkeit der spezifischen Information eines Genotyps beschreibt. Die Fehlerschwelle beschreibt auch die Begrenzung der Genomgröße, wenn man davon ausgeht, daß es – damit Evolution möglich ist – in einer Population asexueller Organismen, Viren oder RNA-Moleküle wenigstens einen Genotyp geben muß, der seine spezifische Infor-

mation erhalten kann. Diese Annahme ist (aus historischen Gründen) zwar naheliegend, aber durch die Quasispeziesstheorie nicht beweisbar. Überprüfbar wird sie, wenn man eine andere Möglichkeit findet, Aussagen über die Erhaltbarkeit genetischer Information zu machen.

Unabhängig von dieser Theorie entwickelte sich etwa zur selben Zeit aufgrund empirischer Befunde die Anschauung, daß Mutationen nicht nur relativ häufige Ereignisse sind, sondern daß viele davon – zumindest in relativ großen Populationen – eine sehr hohe Verweildauer haben und Substitutionen oftmals nahezu zufällig erfolgen. Insbesondere die Entdeckung der „molekularen Uhr“ (Zuckerkandl und Pauling, 1962) war es, die die Formulierung der „Neutralen Theorie der molekularen Evolution“ (Kimura, 1968a, b; Kimura, 1987 und andere) initiierte. Zu erklären, warum die molekulare Evolution mit relativ konstanter Rate abläuft, ist denn auch eines ihrer zentralen Anliegen.

Der „Neutralen Theorie“ ist es zwar gelungen, den Zusammenhang zwischen molekularer Evolutionsrate und der neutralen Mutationsrate (der Rate, mit der Mutanten auftreten, die sich in ihrer Fitness vom Wildtyp nicht oder kaum unterscheiden) präzise aufzuzeigen. Sie hat damit aber das

Reprint requests to W. Tiefenbrunner.
Telefax: 0043 28816/5194.

0939–5075/95/1100–0883 \$ 06.00 © 1995 Verlag der Zeitschrift für Naturforschung. All rights reserved.

D



Dieses Werk wurde im Jahr 2013 vom Verlag Zeitschrift für Naturforschung in Zusammenarbeit mit der Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. digitalisiert und unter folgender Lizenz veröffentlicht: Creative Commons Namensnennung-Keine Bearbeitung 3.0 Deutschland Lizenz.

Zum 01.01.2015 ist eine Anpassung der Lizenzbedingungen (Entfall der Creative Commons Lizenzbedingung „Keine Bearbeitung“) beabsichtigt, um eine Nachnutzung auch im Rahmen zukünftiger wissenschaftlicher Nutzungsformen zu ermöglichen.

This work has been digitalized and published in 2013 by Verlag Zeitschrift für Naturforschung in cooperation with the Max Planck Society for the Advancement of Science under a Creative Commons Attribution-NoDerivs 3.0 Germany License.

On 01.01.2015 it is planned to change the License Conditions (the removal of the Creative Commons License condition “no derivative works”). This is to allow reuse in the area of future scientific usage.

eigentliche Problem nur verlagert, weil man nun erklären muß, warum neutrale Mutationen gerade mit der Häufigkeit auftreten, daß daraus resultiert, daß die molekulare Evolution – verglichen mit der phänotypischen – mit bemerkenswert konstanter Rate abläuft. Ihrer Struktur nach ist die „Neutrale Theorie“ aber nicht besonders gut geeignet, generelle Aussagen über den Anteil neutraler Mutationen an der Gesamtheit der Mutationen zu machen.

Ein sehr einfaches, aber fundamentales Modell, das Evolution aus informationstheoretischer Sicht beschreibt, ist nun dazu geeignet, die beiden besprochenen Theorien zu ergänzen bzw. ihre Grundlagen zu überprüfen. Es geht davon aus, daß Evolution als ein Vorgang verstanden werden kann, in dessen Verlauf Information erworben wird und der daher jenen thermodynamischen Restriktionen unterliegt, die für jeden Informationserwerb und jede Informationserhaltung Gültigkeit haben.

In der vorliegenden Arbeit werden die Grundlagen des Modells entwickelt und seine Aussagen zur Erhaltbarkeit genetischer Information präsentiert. In einem weiteren Artikel wird dann die Existenz informationsneutraler Mutationen aufgezeigt und der mögliche Zusammenhang zum Phänomen der „molekularen Uhr“ hergestellt.

Evolution und Informationsfluß im dynamischen Gleichgewicht

Auch die Evolution kann – wenn es um den Informationserwerb geht – den Rahmen, der ihr durch die physikalischen Gesetze vorgegeben ist, nicht sprengen (allerdings sind es nicht die Gesetze der Gleichgewichtsthermodynamik, die diesen Rahmen bilden). Konkret bedeutet dies:

1) Evolution findet fernab vom thermodynamischen Gleichgewicht statt und das bedeutet, daß wir mit Vorgängen zu rechnen haben, die eine Zunahme der Entropie (= eine Abnahme der Information) bewirken, weil jedes System zum thermodynamischen Gleichgewicht bzw. zur Entropiemaximierung strebt.

2) Evolution funktioniert offensichtlich nur, weil sich die Biosphäre stets fernab vom thermodynamischen Gleichgewicht befindet und das bedeutet, daß Prozesse stattfinden müssen, die die Entropiezunahme kompensieren, d. h. ihrerseits

eine Entropieabnahme bewirken. Diese Entropieabnahme kann aber kein willkürliches Ausmaß annehmen, sondern ist durch Energie- und Informationsinput prinzipiell begrenzt. Vorgänge, die eine Entropiezunahme bedingen und solche, die eine Entropieabnahme bewirken, halten sich die Waage und es ergibt sich daher – zeitweise – ein dynamisches, ein „steady state“ – Gleichgewicht.

Jener Vorgang, der im Evolutionsgeschehen verantwortlich für Entropieabnahme – also für Informationszunahme – ist, ist die Selektion. Um die Entropieänderung durch Selektion mit den Methoden der Informationstheorie quantifizieren zu können, muß man die Selektion als Sortierprozeß beschreiben.

Informationsabnahme wird im Evolutionsgeschehen durch Mutationen verursacht (es finden in den Organismen natürlich auch noch andere Vorgänge statt, die eine Entropiezunahme bewirken; für die Evolution entscheidend sind aber nur die Prozesse, die eine Informationsabnahme am genetischen Material der Keimbahn bedingen). Der Informationsverlust durch Mutation muß quantifiziert werden. Dabei ist es erforderlich, auf zwei Ebenen zu arbeiten, denn der phänotypischen Ordnung, an der die Selektion ansetzt, entspricht eine genotypische an der die Mutationen stattfinden.

Selektion und Mutation bedingen also einen Informationsfluß, der in ein dynamisches Gleichgewicht mündet, solange sich der Zu- (Selektion) oder Abfluß (Mutation) nicht ändert. Die Größe, die sich im Gleichgewicht befindet und die einer Beschränkung unterliegt, ist der Informationsgehalt des Genoms. Da das Gleichgewicht dynamisch ist, kann sich die spezifische Information der Genotypen sehr wohl ändern.

Die Möglichkeiten des Informationserwerbs durch Selektion sind ebenfalls begrenzt, wie aus der Theorie des MAXWELL'schen Dämons folgt (Brillouin, 1956). Der Informationsverlust kann allerdings durch die Entwicklung immer besserer Korrekturlesemechanismen, die die Mutationsrate herabsetzen, im Laufe der Evolution verringert werden. Wenn dies geschieht, gelangt das System aus dem bisherigen Gleichgewicht und gerät schließlich in ein neues. Im Folgenden soll Evolution im Gleichgewicht beschrieben werden.

Selektion als Sortierprozeß

Selektion läßt sich als Sortierprozeß beschreiben, allerdings mit der Einschränkung, daß das Sortieren auf bestimmte Weise – nämlich „unscharf“ – geschieht. Das bedeutet, daß der Sortierende sich irren kann und daß die Möglichkeit, daß er sich irrt, nicht vernachlässigbar ist. Warum der Sortierprozeß unscharf sein muß und welche Beschränkungen sich aus dieser Unschärfe ergeben, wird an anderer Stelle gezeigt. Da es dazu dann erforderlich ist, den Kontakt zur Thermodynamik auf einfache Weise herstellen zu können, soll als Ausgangspunkt für die Entwicklung des Modells ein prominenter Sortierer – der Maxwell'sche Dämon – gewählt werden.

Maxwell 1871 beschrieb sein Wesen als: „a being whose faculties are so sharpened that he can follow every molecule in his course, and would be able to do what is at present impossible to us ... Let us suppose that a vessel is divided into two portions, A and B by a division in which there is a small hole, and that a being who can see the individual molecules opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower ones to pass from B to A. He will, thus, without expenditure of work raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics.“

Maxwell's Dämon konnte die Geschwindigkeit der Moleküle beliebig genau messen und reagierte absolut sicher. Er entnimmt daher der Menge der vorhandenen Moleküle eine Teilmenge, nämlich die Menge der hinreichend raschen (oder in die andere Richtung arbeitend der hinreichend langsamen) Moleküle, und jedes der vorhandenen Moleküle gehört mit Sicherheit entweder der Teilmenge an oder nicht.

Die Begriffe „schnell“ und „langsam“ sind aber ihrer Natur nach unscharf. Wir werden dies und die Möglichkeit, daß der Dämon sich irren kann, berücksichtigen: es sei **jedem** vorhandenen Molekül eine gewisse Wahrscheinlichkeit – im allgemeinen ungleich Null und Eins – zugeschrieben, Element der Menge der hinreichend schnellen Moleküle zu sein. Die Menge der hinreichend schnellen Moleküle ist also unscharf („fuzzy“). Weiters nehmen wir an, daß die Wahrscheinlichkeit w zur Menge der hinreichend schnellen Mole-

küle zu gehören, von der Geschwindigkeit des Moleküls abhängt, und zwar so, daß w um so größer ist, je größer die Geschwindigkeit des Moleküls ist. Das ist die Voraussetzung dafür, daß Selektion durch den Dämon stattfinden kann.

Maxwell wollte zeigen, daß man – wenn man jene Eigenschaften hat, die er seinem Wesen zuerkannte – ohne Arbeit leisten zu müssen, einen Temperaturunterschied herstellen kann (und damit in der Lage ist, den zweiten Hauptsatz der Thermodynamik zu verletzen). Daher ließ er seinen Dämon die Moleküle nach deren Geschwindigkeit (eigentlich: kinetische Energie) sortieren. Genausogut könnte der Dämon aber auch nach anderen Eigenschaften sortieren bzw. selektieren, nach Größe, Ausmaß der Lichtabsorption, Polarität oder Sonstigem. Auch muß er nicht unbedingt Moleküle auslesen – genausogut könnten es Phänotypen sein und die Eigenschaft nach der er diese selektiert der Anpassungsgrad. w wäre dann die Wahrscheinlichkeit, zur Menge der hinreichend angepaßten Phänotypen zu gehören, die vom Anpassungsgrad (Fitness), aber auch von anderen Eigenschaften, z. B. von der „Kurzsichtigkeit“ des Dämons abhängt.

Wir wollen nun das Modell verändern, weil es uns schließlich im Endeffekt nicht um Moleküle und deren Geschwindigkeit, sondern um Phänotypen und deren „Angepaßtheit“ gehen wird (der Bequemlichkeit wegen und auch wegen der Modellkonsistenz werden wir vorerst weiterhin von Molekülen und Geschwindigkeiten sprechen).

Zunächst besteht das Ursprungsmodell von Maxwell aus zwei Prozessen, die man getrennt behandeln kann: der Übergang schneller Moleküle von A nach B und dem Transfer langsamer Moleküle von B nach A. Wir betrachten im Folgenden nur mehr ersteren, die Selektion schneller Moleküle. Außerdem wollen wir davon ausgehen, daß der Dämon den Durchgang genau M mal öffnet, also M Moleküle durchläßt, und daß M viel kleiner ist als die Anzahl N der in A enthaltenen Moleküle.

Weiters nehmen wir an, daß wir die Moleküle in k verschiedene Geschwindigkeitsklassen einteilen können. Jeder Klasse können wir nun einerseits eine gewisse relative Häufigkeit x_i zuordnen ($i = 1, \dots, k$), die eben der Anzahl der Moleküle dieser Klasse in Relation zu N ist und andererseits eine bestimmte Wahrscheinlichkeit w_i , das ist die

Wahrscheinlichkeit jedes Moleküls dieser Klasse, zur Menge der hinreichend schnellen Moleküle zu gehören. Die Einteilung in Klassen stellt keine Beschränkung dar, weil auch jedes Molekülindividuum eine Klasse repräsentieren kann.

Wir wollen nun die Frage stellen, wie groß die Wahrscheinlichkeit ist, daß der Dämon, der das **Herannahen eines Moleküls** wahrnimmt, ein Molekül der i -ten Klasse passieren läßt. Diese entspricht offensichtlich dem Produkt der Wahrscheinlichkeiten, daß das herannahende Molekül der i -ten Klasse angehört **und** daß dieses Molekül hinreichend schnell ist. Wenn wir annehmen, daß für eine bestimmte Molekülklasse die Wahrscheinlichkeit, auf das Loch bzw. den Dämon zuzufiegen, nur von deren relativer Häufigkeit abhängt (würde sie von der Eigenschaft, hinreichend schnell zu sein abhängen, könnte der Dämon allein durch fortwährendes beliebiges Öffnen und Schließen bereits einen Temperaturunterschied zwischen denn Kammern erzeugen, und könnte dann **tatsächlich** den 2ten Hauptsatz verletzen), ergibt sich:

$$\begin{aligned} &\text{Wahrscheinlichkeit der Passage} \\ &\text{eines } i\text{-Moleküls} = x_i w_i. \end{aligned} \quad (1)$$

Als nächstes wollen wir die Frage stellen, wie groß die Wahrscheinlichkeit ist, daß **irgendein** Molekül passiert. Diese entspricht natürlich der Summe aller Klassenwahrscheinlichkeiten:

$$\begin{aligned} &\text{Wahrscheinlichkeit der Passage} \\ &\text{irgendeines Moleküls} = \sum_i x_i w_i. \end{aligned} \quad (2)$$

Dies ist auch für ein Molekül die mittlere Wahrscheinlichkeit, hinreichend schnell zu sein.

Weiters wollen wir wissen, wie groß der relative Anteil x_i' der Moleküle der Klasse i unter jenen M Molekülen ist, die mit Hilfe des Dämons von A nach B gelangt sind. Der erwartete relative Anteil der i -Moleküle nach der Passage muß der Wahrscheinlichkeit der Passage eines **i -Moleküls** in Relation zur Wahrscheinlichkeit der Passage **irgendeines** Moleküls entsprechen. Das heißt:

$$x_i' = x_i w_i / \sum_j x_j w_j \quad (i, j = 1, \dots, k). \quad (3)$$

x_i ändert sich während des Entnahmeprozesses nicht wesentlich, weil $M \ll N$ (wir setzen hier allerdings zusätzlich voraus, daß $x_i N$ eine einigermaßen

große Zahl ist). Gleichung (3) ist das sehr einfache Beispiel einer diskreten Selektionsgleichung, wie man sie etwa für sich asexuell fortpflanzende Lebewesen (z. B. parthenogenetische Tiere oder Bakterien) aufstellen könnte (siehe dazu z. B. Ewens, 1979 oder Hofbauer und Sigmund, 1984). Der Dämon betreibt also tatsächlich Selektion! Die Qualität dieses Sortierprozesses hängt nicht von der absoluten Geschwindigkeit der Moleküle ab, dafür aber gewissermaßen vom Unterscheidungsvermögen des Dämons. Dies sei folgendermaßen gezeigt: Gleichung (3) läßt sich in die Form bringen:

$$x_i' / x_i = 1 / \sum_j (x_j w_j / w_i), \quad (4)$$

d.h. die Relationen w_j / w_i entscheiden über die Änderung der relativen Häufigkeiten im Selektionsprozess – wenn wir also den Dämon einmal bei sehr hohen Wahrscheinlichkeiten der Moleküle hinreichend schnell zu sein selektieren lassen, das zweite Mal, wenn alle Moleküle exakt um das Tausendfache (oder einen beliebigen anderen konstanten Betrag) verminderte w_i -Werte aufweisen, erhalten wir das gleiche Ergebnis (im ersten Fall weisen die Moleküle auch eine höhere Geschwindigkeit auf – wenn auch nicht notwendigerweise eine tausendmal höhere).

Die Relation $w_i / \sum_j x_j w_j$ bringt die Wahrscheinlichkeit eines i -Moleküls hinreichend schnell zu sein in Zusammenhang mit der mittleren Wahrscheinlichkeit zu hinreichender Schnelligkeit. Sie gibt an, wie sich ein i -Molekül in dieser Eigenschaft im Mittel von irgendeinem Molekül unterscheidet. Wir definieren dies als Unterscheidbarkeit u_i .

$$u_i = w_i / \sum_j x_j w_j. \quad (5)$$

Selektion kann sowohl als Interaktion zwischen Individuen verstanden werden – und wird dann (wie heute meist üblich) als Konkurrenzphänomen dargestellt – als auch als Interaktion zwischen Individuum und Umwelt, was eine Beschreibung als Sortierprozess zur Folge hat, wie das hier der Fall war. Beide Sichtweisen führen letztlich zum gleichen Ergebnis, aber sie veranlassen unterschiedliche weiterführende Fragen.

In der Informationstheorie gibt es Standardverfahren zur Behandlung der Frage, wie sich durch Einwirken eines Sortierprozesses die Entropie än-

dert (Gerthsen und Vogel, 1993; Brillouin, 1956). Diese Verfahren werden wir nun anwenden.

Entropieänderung durch Selektion

Als Endprodukt des Selektionsprozesses erhält man eine Komposition – verschieden häufige Molekülsorten, die sich hinsichtlich des Merkmals, nach dem die Selektion erfolgte, unterscheiden. Gleichung (3) gibt die relativen Häufigkeiten an, mit denen die einzelnen Molekülklassen schließlich angetroffen werden. Wir können nun unseren ziemlich gut informierten Dämon mit seiner hervorragenden Fähigkeit, die einzelnen Molekülsorten zu unterscheiden, durch einen anderen ersetzen, von dem wir uns vorstellen, daß er so kurz-sichtig ist, daß er zwar erkennen kann, daß sich ein Molekül dem Loch nähert, aber überhaupt nicht in der Lage ist, diese nach irgendeiner Eigenschaft – z. B. der Geschwindigkeit – zu sortieren. Bei der Annäherung eines Moleküls öffnet oder schließt er das Loch nach gutdünken – rein zufällig also. Er soll aber immerhin noch so gut sehen, daß er in der Lage ist wahrzunehmen, ob ein Molekül passiert hat und nachdem er genau M Moleküle passieren ließ, seine Tätigkeit zu beenden.

Wir fragen nun nach der Wahrscheinlichkeit P , mit der dieses Wesen durch seine Zufallstätigkeit jene Molekülkomposition erhält, die der gut informierte, viel weniger kurzsichtige Dämon durch gezielte Arbeit zustande gebracht hätte. Per definitionem besteht zwischen dieser Wahrscheinlichkeit P und der Entropie S folgender Zusammenhang:

$$S = C \ln P. \quad (6)$$

In der Physik hat der konstante Faktor C eine physikalische Dimension (Energie/Grad) und entspricht der BOLTZMANNschen Konstante. Im außerphysikalischen Bereich wählt man der Einfachheit halber oftmals $C = 1$ (was auch wir tun werden). In der Informationstheorie wird P mit dem Informationsgehalt I in Zusammenhang gebracht:

$$I = -\text{ld } P = -\ln P / \ln 2, \quad (7)$$

wobei ld der Logarithmus dualis (der Logarithmus zur Basis 2) ist. Abgesehen von einer Konstante entspricht daher der Informationsgehalt der „Negentropie“, der negativen Entropie.

Die Wahrscheinlichkeit einer bestimmten Komposition aus insgesamt M Molekülen, bestehend aus k unterschiedlichen „Sorten“ bzw. Klassen, wobei die Anzahl der Moleküle der i -ten Klasse durch $m_i' = M x_i'$ gegeben ist, wird durch:

$$\frac{M!}{\prod_i m_i'!} \quad (8)$$

mögliche Folgen (die durch die Reihenfolge, mit der die einzelnen Moleküle passieren, gegeben ist) bestimmt. Jede dieser Folgen tritt mit einer Wahrscheinlichkeit

$$\prod_i x_i^{m_i'} \quad (9)$$

auf. Dies deshalb, weil bei zufälligem Öffnen und Schließen des Lochs die Wahrscheinlichkeit der Passage eines i -Moleküls nur von der relativen Häufigkeit x_i abhängt. Weil $M \ll N$ ist und wir annehmen, daß alle m_i große Zahlen darstellen, ändert sich durch die Passage eines i -Moleküls die relative Häufigkeit x_i im Behälterteil A nicht. Da es sich bei der Passage unterschiedlicher Moleküle um unabhängige Ereignisse handelt, erhält man die Wahrscheinlichkeit, die Realisierung einer bestimmten Folge festzustellen, indem man die Wahrscheinlichkeiten der Einzelereignisse – also der einzelnen Passagen – multipliziert.

Die Wahrscheinlichkeit P , zufällig jene Komposition zu erhalten, die der „wohlinformierte“ Dämon durch Selektion erreicht, ist damit:

$$P = \frac{M!}{\prod_i m_i'!} \prod_i x_i^{m_i'} \quad (10)$$

mit der Nebenbedingung:

$$\sum_i m_i' = M. \quad (11)$$

Mit Hilfe der Stirlingschen Formel

$$M! \cong \frac{M^M}{e^M} \sqrt{2\pi M} \quad (12)$$

läßt sich Gl. (10) wesentlich vereinfachen, wenn wir voraussetzen, daß die einzelnen Komponenten m_i' einigermaßen große Zahlen repräsentieren. Insbesondere durch logarithmische Transformation:

$$\ln M! \approx M \ln M - M + 1/2 \ln(2\pi M). \quad (13)$$

Unter dem Logarithmus spielen die kleinen Ungenauigkeiten der Stirlingschen Formel eine sehr

geringe Rolle. Man kann sich daher für größere M schon mit den beiden ersten Gliedern begnügen:

$$\ln M! \approx M \ln M - M. \quad (14)$$

Angewendet auf Gl. (10) ergibt Gl. (14) (und Gl. (6)):

$$S = \ln P = M \ln M - M - \sum (m_i' \ln m_i') + \sum m_i' + \sum (m_i' \ln x_i). \quad (15)$$

Da $\sum m_i' = M$, heben sich das 2te und das 4te Glied weg:

$$S = M \ln M - \sum (m_i' \ln m_i') + \sum (m_i' \ln x_i). \quad (16)$$

Wir ersetzen nun nach $m_i' = m_i u_i$ (siehe Gln. (3) und (5)) und nach $x_i = m_i/M$:

$$S = M \ln M - \sum (m_i u_i \ln m_i u_i) + \sum (m_i u_i \ln m_i/M) \quad (17)$$

$$S = M \ln M + \sum \left(m_i u_i (\ln m_i - \ln M - \ln m_i - \ln u_i) \right) \quad (18)$$

$$S = M \ln M - \sum \left(m_i u_i (\ln M + \ln u_i) \right) \quad (19)$$

$$S = M \ln M - \ln M \sum m_i u_i - \sum (m_i u_i \ln u_i). \quad (20)$$

Wegen $M = \sum m_i u_i$ erhält man schließlich:

$$S = - \sum (m_i u_i \ln u_i) = -M \sum (x_i u_i \ln u_i). \quad (21)$$

In Worten: die Information der Komposition ergibt sich indem man den Logarithmus der Unterscheidbarkeitswerte aller Individuen, die **nach** erfolgter Selektion noch übrig sind, addiert. Wichtig ist, daß eine Klasse von Eigenschaftsträgern, deren $u_i = 0$ ist, nach erfolgter Selektion nicht mehr vorkommt, weil bei ihnen auch $w_i = 0$ ist (d. h., sie werden mit Sicherheit nicht passieren); daher kann auch $\ln 0$ in Gl. (21) nicht auftreten.

Wir wollen nun das Modell weiterentwickeln. Selektion – wie sie hier beschrieben wird – bedeutet, daß man nach einem bestimmten Kriterium aus einem „Behälter“, der sehr viele Moleküle – oder sonstige „Eigenschaftsträger“ – enthält, vergleichsweise wenige entnimmt. Als Modell für die Evolution ist das natürlich nicht ausreichend, weil die Selektion in Wirklichkeit ständig wirkt, die Moleküle im Behälter A also selbst eine Komposition darstellen, die das Ergebnis eines Selektionsprozesses ist

und auch die aus A selektierten Moleküle wieder Grundlage neuerlicher Selektion werden sollen. Da wir angenommen haben, daß $M \ll N$ (es wird jedesmal nur eine kleine Anzahl Moleküle aus den ursprünglich vorhandenen ausgelesen), ist dieser Prozeß nicht beliebig oft wiederholbar, weil sich die Anzahl der Moleküle mit jedem Selektionsschritt vermindert.

Daher muß an jeden Selektionsschritt eine Phase der Vermehrung angeschlossen werden, die erreichen soll, daß wieder die ursprüngliche Molekül- oder Phänotypenanzahl N vorliegt.

Von dieser Vermehrungsphase wollen wir fordern:

- daß sich die relativen Häufigkeiten x_i' , wie sie nach erfolgter Selektion vorliegen, durch die Vermehrung nicht ändern;
- daß die Vermehrung fehlerfrei geschieht, also keine neuen Komponenten auftreten (dies ist genaugenommen bereits durch die erste Forderung erfüllt).

Nach erfolgter Selektion und Vermehrung haben wir wieder N Moleküle. Die Entropie der vorliegenden Komposition ist gegeben durch:

$$S = -N \sum (x_i u_i \ln u_i), \quad (22)$$

was daran liegt, daß sich die mittlere Entropie pro Individuum ($-\sum (x_i u_i \ln u_i)$) durch die Vermehrung nicht ändert, sondern lediglich die Gesamtzahl der beteiligten Individuen (bzw. dadurch, daß man die Gesamtentropie durch Addition der Einzelentropien erhält).

Im Folgenden wird für uns allerdings nicht die Entropie nach erfolgter Selektion und Vermehrung, sondern die Entropieänderung durch Selektion und Vermehrung, entscheidend sein. Um diese errechnen zu können, müssen wir die Entropie des Systems vor erfolgter Selektion kennen. Wir verwenden abermals unseren sortierungsunfähigen, kurzsichtigen Dämon, den wir wieder M Moleküle entnehmen lassen, gehen diesmal aber davon aus, daß die Komposition in der die entnommenen Moleküle vorliegen, der wahrscheinlichsten entspricht. Nach der Vermehrungsphase erhalten wir einen Zustand, der dem Ausgangszustand praktisch entspricht und kennen somit die Ausgangsentropie (diese Schlußfolgerung ist nur dann zulässig, wenn die Voraussetzung, daß alle m_i einigermaßen große Zahlen repräsentieren, erfüllt ist).

Die wahrscheinlichste Komposition liegt vor, wenn $x_i' = x_i$ bzw. $m_i' = m_i$. In Analogie zu Gl. (16) erhalten wir:

$$S_{\max} = M \ln M - \sum (m_i \ln m_i) + \sum (m_i \ln x_i) \quad (23)$$

$$S_{\max} = M \ln M - \sum \left(m_i \ln (x_i M) \right) + \sum (m_i \ln x_i) \quad (24)$$

$$S_{\max} = M \ln M - \sum \left(m_i (\ln x_i + \ln M) \right) + \sum (m_i \ln x_i) \quad (25)$$

$$S_{\max} = M \ln M - \ln M \sum m_i. \quad (26)$$

Wegen $\sum m_i = M$ erhalten wir schließlich:

$$S_{\max} = 0. \quad (27)$$

Genaugenommen ist dieses Ergebnis selbstverständlich. Wir haben mit Gl. (22) die Informationsmenge abgeschätzt, die dem System durch die Tätigkeit des Dämons zugeführt wird. Deswegen kommen wir auch zur Schlußfolgerung, daß ohne Tätigkeit des Dämons die Systementropie Null ist. Daher können wir nunmehr feststellen, daß die durch Selektion und Vermehrung erfolgende Entropieänderung durch folgende Gleichung gegeben ist:

$$\Delta S_{\text{sel}} = -N \sum (x_i u_i \ln u_i). \quad (28)$$

Es gilt $u_i > 1$ für ein Individuum, wenn dessen Wahrscheinlichkeit, hinreichend angepaßt zu sein, größer ist, als die mittlere Wahrscheinlichkeit zu hinreichender Angepaßtheit in der Ausgangspopulation. Der Logarithmus von u_i ist dann positiv, bei $u_i < 1$ ist er negativ. Wenn also durch Selektion die Anzahl jener Individuen gegenüber der Ausgangssituation zunimmt, deren w_i größer ist, als das im Mittel für ein Individuum der Ausgangspopulation der Fall war, nimmt damit die Entropie der Gesamtpopulation ab, bzw. deren Informationsgehalt zu. Dies entspricht genau unserer intuitiven Erwartung.

Die durch die Selektion verursachte mittlere Entropieänderung pro Individuum (wir werden sie in der Folge auch als Selektionsleistung bezeichnen) Δs_{sel} ist gegeben durch:

$$\Delta s_{\text{sel}} = - \sum (x_i u_i \ln u_i). \quad (29)$$

Mutationen aus informationstheoretischer Sicht

Wir interessieren uns hier für Mutationen ausschließlich unter dem Gesichtspunkt des durch sie verursachten Informationsverlustes. Daher ist es möglich, eine Reihe vereinfachender Annahmen zu formulieren, ohne das deswegen die Aussagekraft des resultierenden Modells allzusehr leidet.

Weiters müssen wir für die Bereiche, wo kontroverse Ansichten bestehen und eine endgültige Klärung noch nicht in Sicht ist, vorläufige Arbeitshypothesen akzeptieren. Als Arbeitshypothesen formulieren wir: 1) Die DNA unterschiedlicher Lebewesen kann mit unterschiedlicher Rate mutieren; 2) das Genom mutiert in allen Bereichen annähernd gleich schnell.

Als Vereinfachungen wollen wir akzeptieren: a) Um abschätzen zu können, wie sich der Informationsgehalt durch Mutationen ändert, genügt es, Punktmutationen zu berücksichtigen. Diese Annahme scheint wegen der gegenüber anderen Mutationsereignissen wesentlich größeren Häufigkeit von Punktmutationen zulässig. b) Die Mutationswahrscheinlichkeit sei von der Position und der Art des gerade vorliegenden Zeichens unabhängig.

Die genetische Information liegt in Form einer Zeichenkette vor, der ein Zeichensatz mit vier Symbolen zugeordnet ist. Unter den beschriebenen Voraussetzungen ist die Anzahl der Positionen (der Nukleotide), die pro Zeiteinheit im Gesamtgenom mutieren binomialverteilt (bzw., da n groß und p sehr klein ist poissonverteilt), wobei der Mittelwert v der Binomialverteilung (bzw. der Poissonverteilung) durch

$$v = np \quad (30)$$

gegeben ist, wenn n die Länge der Zeichenkette (die Genomlänge des haploiden Genoms) und p die Mutationswahrscheinlichkeit pro Position (Nukleotid) in der angenommenen Zeiteinheit ist. v ist die mittlere Anzahl von Mutationen pro haploidem Genom und Zeiteinheit und wird als Mutationsrate pro Gamet (oder haploidem Genom) bezeichnet.

Neben der Mutationsrate v ist auch noch die Zerfallsrate z , die den Informationszerfall (der Information des Genotyps, aber nicht der des Phänotyps) durch Mutationen beschreibt, wichtig. Wir gehen davon aus, daß eine bestimmte Zeichenkette in sehr vielen identischen Kopien (Genome mit identischer Sequenz) vorliegt. Mutationsbedingt wird die relative Häufigkeit von Genomen mit ursprünglicher Sequenz abnehmen. Wir fragen nach dieser relativen Häufigkeit.

Die Wahrscheinlichkeit, daß eine bestimmte Position innerhalb der angenommenen Zeiteinheit

nicht mutiert, ist $1-p$. Unter den oben beschriebenen Voraussetzungen ist dann die Wahrscheinlichkeit, daß das ganze Genom nicht mutiert $(1-p)^n$. Es gilt daher:

$$z = (1-p)^n, \quad (31)$$

wenn wir Rückmutationen als vernachlässigbar betrachten.

Wenn n groß und p sehr klein ist, gilt näherungsweise:

$$z \approx e^{-np} \approx e^{-v}. \quad (32)$$

Entropieänderung durch Mutationen

Jede zufällige Änderung eines einzigartigen Textes – wie etwa der genetischen Information – stellt einen Informationsverlust dar (man berücksichtige dabei, daß die Informationstheorie keineswegs a priori unterscheidet, ob ein Text etwas „bedeutet“ oder nicht – sie hat nicht die semantische, sondern die absolute Information zum Inhalt). Nach Mutation aller Textpositionen ist jeder beliebige Text einer bestimmten Länge möglich und die Information ist dann null.

Prinzipiell geht durch eine Mutation diejenige Information verloren, die nötig wäre, den ursprünglichen Zustand wiederherzustellen. Wir müssen also die Frage stellen, wieviel Information ein Nukleotid (eine Position) überhaupt speichert, wieviel Information im ursprünglichen Zustand also vorhanden war.

Während wir in Zusammenhang mit der Selektion die Entropie einer Komposition ermittelt haben, geht es jetzt um den Informationsgehalt einer Sequenz (eines spezifischen Polynukleotids). Dazu müssen wir zunächst die Wahrscheinlichkeit, mit der eine bestimmte Zeichenkette durch einen Zufallsprozess zustandekommt (etwa durch zufällige Entnahme von Symbolen aus einem Sack, wobei wir annehmen wollen, daß sich die relative Häufigkeit der unterschiedlichen Zeichen im Sack durch die Entnahme nicht ändert), feststellen.

Sei h_i die relative Häufigkeit des i -ten Symbols ($i = 1, \dots, k$). Wenn wir annehmen, daß auch in der Zeichenkette der Länge n die relative Häufigkeit der unterschiedlichen Symbole h_i ist (was am wahrscheinlichsten ist), dann gilt:

Wahrscheinlichkeit für das zufällige Zustandekommen der Zeichenkette =

$$\prod_i h_i^{nh_i}. \quad (33)$$

Nach Gl. (6) (mit $C = 1$) ist daher die erwartete Entropie:

$$S = n \sum_i h_i \ln h_i. \quad (34)$$

Die mittlere Entropie pro Zeichen beträgt daher $\sum h_i \ln h_i$. Am größten ist der mittlere Informationsgehalt (hier berechnet nach $I = -\sum h_i \ln h_i$, vergl. mit Gl. (6) und (7)), wenn alle vier Symbole (Basen) der Polynukleotidkette mit gleicher Häufigkeit vorkommen. Er beträgt dann 1.39 (wenn wir zur Berechnung des mittleren Informationsgehalts den Logarithmus dualis verwenden, erhalten wir die Information in bit- Einheiten. Der Informationsgehalt pro Position beträgt dann 2 bit). Weil aber diese Bedingung nicht erfüllt ist und aus anderen, gleich zu besprechenden Gründen, ist der Informationsgehalt niedriger. Jedenfalls erlaubt diese Abschätzung die Aussage, daß der Informationsverlust durch eine Punktmutation Δs_{mut} im Mittel nicht größer ist, als 2 bit (Δs_{mut} entspricht dem mittleren Informationsgehalt pro Zeichen).

Jede Form der Redundanz reduziert den mittleren Informationsgehalt und damit auch Δs_{mut} . Eine offensichtliche Form der Redundanz ist die Inhaltswiederholung. Eine weniger offensichtliche ist die Redundanz des genetischen Codes, die die Information des Phänotyps betrifft: Eine Änderung an der DNA bedeutet nicht unbedingt eine Veränderung des transkribierten Proteins und somit oftmals keinen phänotypischen Informationsverlust (die Degeneration des Codes könnte standardmäßig in einer Theorie der Entropieänderung durch Mutation berücksichtigt werden; hier ist dies nicht geschehen, da es auch informationstragende DNA-Abschnitte gibt, die für kein Protein kodieren).

In diesem Zusammenhang muß man erkennen, daß zwar Mutationen am Genom stattfinden, dieses aber die Information für den Organismus speichert. Der Genotyp liefert die spezifische Information für den Phänotyp, der gewissermaßen die „Dreh-scheibe“ sowohl für den Informationsgewinn durch Selektion, als auch für den Informationsverlust durch Mutation darstellt. Letztlich interessiert uns der Informationsverlust der den Phänotyp betreffenden Information. Dieser ist aber oftmals schwer zu quantifizieren, da es vom Protein, seiner Funktion und seiner Umgebung abhängt, ob eine genetische

sche Änderung einen Informationsverlust für den Organismus zur Folge hat.

Geringe ΔS_{mut} -Werte weisen im allgemeinen auch „neue“ Gene – d. h. Gene, die erst vor relativ kurzer Zeit durch Duplikation entstanden sind und eine neue Funktion erworben haben – auf. Ein Analogiebeispiel soll das verständlich machen: jedem Programmierer geschieht es gelegentlich, daß er erkennt, daß ein Programm, das er vor längerer Zeit geschrieben hat, sich nun auch zur Lösung einer anderen Aufgabe einsetzen läßt. Ein speziell für diesen Zweck geschriebenes Programm würde aber im allgemeinen weniger Speicherplatz erfordern. Bezogen auf die semantische Information liegt also eine relativ geringe Informationsdichte vor. Sehr „alte“ Gene, die in ihrer Funktion schon sehr lange tätig sind (wie etwa die Histogene) haben jedenfalls in der Regel einen relativ hohen mittleren Informationsgehalt. Wahrscheinlich muß auch ein hoher, mittlerer Informationsgehalt pro Nukleotid erst durch die Evolution erworben werden.

Insgesamt scheint es daher sehr schwierig zu sein, den mittleren Informationsverlust durch Punktmutation bestimmen zu wollen. Dennoch ist es möglich ΔS_{mut} empirisch zu bestimmen. ΔS_{mut} , die Informationsdichte, bestimmt nämlich die Häufigkeit informationsneutraler Mutationen, die wiederum die (meßbare) molekulare Evolutionsrate beeinflusst (die nähere Analyse dieses Zusammenhangs ist nicht unkompliziert und unterbleibt hier, weil sie zur Behandlung des Themas des vorliegenden Artikels nicht erforderlich ist).

Wir wollen nun die mutationsbedingte Entropieänderung der Gesamtpopulation ΔS_{mut} ermitteln. Die Entropieänderung pro Individuum (und Zeiteinheit) hängt von der mittleren Entropieänderung pro Mutation und der Anzahl der Mutationen am Genom ab.

$$\begin{aligned} \text{Entropieänderung pro Individuum} = \\ v\Delta S_{\text{mut}} = np\Delta S_{\text{mut}} \end{aligned} \quad (35)$$

Daraus folgt:

$$\Delta S_{\text{mut}} = Nnp\Delta S_{\text{mut}} \quad (36)$$

Das evolutionsdynamische Gleichgewicht und die Fehlerschwelle

Evolution setzt den Erhalt von Ordnung bzw. Information voraus. Information geht ständig verloren und daraus folgt die Notwendigkeit, daß In-

formation erworben wird. Über längere Zeiträume muß der Informationserwerb den Verlust zumindest ausgleichen. Die Bilanz der Entropieänderung muß daher negativ sein oder zumindest null, d. h. die Entropie im Laufe der Zeit kleiner werden oder gleich bleiben. Dies bedeutet, daß

$$\Delta S_{\text{mut}} + \Delta S_{\text{sel}} \leq 0 \quad (37)$$

sein muß. Weiters folgt (Gl. (28) u. Gl.(36)):

$$Nnp\Delta S_{\text{mut}} - N\sum x_i u_i \ln u_i \leq 0 \quad (38)$$

$$n\Delta S_{\text{mut}} \leq \frac{\sum x_i u_i \ln u_i}{p} \quad (39)$$

ΔS_{mut} , die mittlere Entropieänderung pro Mutation entspricht quantitativ dem mittleren Informationsgehalt pro Nukleotid. $n\Delta S_{\text{mut}}$ entspricht daher dem Informationsgehalt des (haploiden) Genoms und Ungleichung 39 stellt also eine Begrenzung für diesen Informationsgehalt dar, die offenbar nur von der Selektionsleistung und der Mutationsrate abhängt.

Bedenkt man, daß die Evolution ein sich selbst optimierender Prozess ist und daß das Genom jene Information trägt, die dem Organismus die Anpassung an und das Überleben in seiner Umwelt ermöglicht, dann ist es naheliegend anzunehmen, daß nach Möglichkeit durch die Evolution die Gesamtinformation des Genoms maximiert wird. D. h. daß wir häufig, wenn nicht sogar im Regelfall beobachten werden, daß

$$n\Delta S_{\text{mut}} = \frac{\sum x_i u_i \ln u_i}{p} \quad (40)$$

Es gibt in der Literatur eine Ungleichung, die ähnliche – wenn auch nicht identische – Aussagen wie Ugl. (39) macht, aber das Produkt eines völlig anderen Denkansatzes ist. Es lohnt sich, einen Vergleich mit diesem als „Fehlerschwelle“ bekannt gewordenen Konzept anzustellen.

Eigen hat die Fehlerschwelle erstmals 1971 präsentiert. Spätere Arbeiten von Eigen *et al.* (1981), Schuster und Sigmund (1980), Schuster und Sweitina (1982), Hofbauer und Sigmund (1984) und Maynard Smith (1989) beschäftigen sich ebenfalls mit diesem Thema.

Der Grundgedanke ist, daß es in der Population (oder „Quasipopulation“, denn Eigen entdeckte die Fehlerschwelle in Zusammenhang mit seinen Arbeiten über präbiotische Evolution und die

Evolution von relativ kurzen RNA-Molekülen) wenigstens einen Genotyp geben sollte, der seine spezifischen Eigenschaften über Generationen unverändert erhalten kann. Damit er dazu in der Lage ist, muß jedes Individuum, das Träger dieses Genotyps ist, im Durchschnitt wenigstens mehr als einen Nachkommen haben, der Träger desselben Genotyps ist, der also nicht mutiert ist. Für diesen Genotyp – der ursprünglich als Mastergenotyp oder Stammsequenz bezeichnet wurde – muß also gelten, daß das Produkt aus Vermehrungsrate r und Zerfallsrate z (siehe Gl. (31) und Gl. (32) – z gibt den relativen Anteil der nicht mutierten Nachkommen an) größer als eins ist:

$$rz > 1, \quad (41)$$

woraus folgt:

$$re^{-np} > 1 \quad (42)$$

$$e^{-np} > 1/r \quad (43)$$

$$-np > -\ln r \quad (44)$$

$$n < (\ln r)/p. \quad (45)$$

Ugl. (39) und Ugl. (45) unterscheiden sich in einigen wesentlichen Punkten. Zunächst einmal ist Ugl. (39) eine Grenze für die Erhaltbarkeit der Gesamtinformation; bei ihrer Herleitung wurde die Gesamtpopulation berücksichtigt. Ugl. (45) stellt eine Grenze für die Erhaltbarkeit einer spezifischen Sequenz dar; bei ihrer Herleitung wurde nur ein einziger Genotyp berücksichtigt, der von der Gesamtpopulation dadurch „entkoppelt“ wurde, daß die Anzahl der Nachkommen r als konstant angesehen wurde. Tatsächlich ist die mittlere Anzahl der Nachkommen pro Individuum – wie im Abschnitt über Selektion durch den MAXWELLSchen Dämon behandelt – eine Funktion der Wahrscheinlichkeit jedes Individuums der Population, zur Menge der hinreichend angepaßten Phänotypen zu gehören. Ist eine Population im Gleichgewicht, kann man natürlich die Anzahl der Nachkommen eines Individuums mit bestimmtem Genotyp als einigermaßen konstant betrachten. Wir wollen daher r durch $u_1 = x_1'/x_1$ ersetzen, wobei wir festlegen, daß u_1 die größte in der Population vorkommende mittlere Nachkommenanzahl sein soll.

Wir fragen nun nach den Bedingungen, die erfüllt sein müssen, damit in einer Population, die im evolutionsdynamischen Gleichgewicht ist, we-

nigstens **ein** Genotyp, der in der Lage ist seine spezifische Sequenz zu erhalten (ein Mastergenotyp), vorkommen **kann**. Dazu bringen wir Gl. (40) in die Form:

$$n = \frac{\sum x_i u_i \ln u_i}{\Delta s_{\text{mut}} p} \quad (46)$$

und ersetzen n in Gl. (45) (mit u_1 statt r):

$$\frac{\sum x_i u_i \ln u_i}{\Delta s_{\text{mut}} p} < \frac{\ln u_1}{p}, \quad (47)$$

woraus folgt:

$$\Delta s_{\text{mut}} > \frac{\sum x_i u_i \ln u_i}{\ln u_1}. \quad (48)$$

Der rechtsseitige Ausdruck in Ugl. (48) ist sicher kleiner oder gleich eins, weil die mittlere Anzahl der Nachkommen pro (die Selektion) überlebendem Individuum nicht größer sein kann, als die größte in der Population vorkommende mittlere Nachkommenanzahl. Allerdings kann man annehmen, daß in realen Populationen der Unterschied nicht sehr groß ist und der kritische Wert für Δs_{mut} nahe bei eins liegt.

Der mittlere Informationsgehalt pro Nukleotid kann maximal 1.39 betragen, wenn die vier unterschiedlichen Nukleotide mit gleicher Häufigkeit vorkommen. Selbst wenn das der Fall wäre, käme er allein durch die Redundanz des genetischen Codes in den Bereich von 1. Im allgemeinen wird der mittlere Informationsgehalt aber noch weit tiefer liegen, deutlich unter 1, weil es im Genom relativ viele Abschnitte geringer Informationsdichte gibt (es ist jedoch zu beachten, daß es auch nichtcodierende Abschnitte hoher Informationsdichte gibt). Dies dürfte auch für die präbiotische Evolution richtig sein, weil gerade für sie sinngemäß gilt, was im vorigen Abschnitt über die geringen Δs_{mut} -Werte neuer Gene gesagt wurde.

Daß Δs_{mut} im allgemeinen deutlich kleiner als eins sein dürfte (so daß häufig Ugl. (48) nicht erfüllt sein wird) hat zur Konsequenz, daß man vermutlich oft in einer Population keinen Mastergenotyp antreffen wird (es ist dann praktisch jeder in der Population auftretende Genotyp einzigartig) – aus dessen Erhaltbarkeit kann man daher auch nicht auf Eigenschaften der Gesamtpopulation schließen.

Das Konzept der Fehlerschwelle (Gl. (45)) läßt sich prinzipiell nur auf asexuelle Organismen anwenden. Bei sich sexuell vermehrenden Organismen ist ein bestimmter Genotyp nicht mehr oder nur noch in Ausnahmefällen erhaltbar. Im Koppungsgleichgewicht ist die relative Häufigkeit eines jeden haploiden Genotyps in der Population durch das Produkt der relativen Häufigkeiten seiner Allele gegeben und hängt nicht unmittelbar von der Häufigkeit des Genotyps in der letzten Generation und von dessen Vermehrungsrate ab. Würde bei sexuellen Lebewesen ein Zusammenhang zwischen der Erhaltbarkeit eines Genotyps und der Genomgröße existieren, müßte die Genomgröße mit der Allelhäufigkeit schwanken!

Im Folgenden soll untersucht werden, wie sich Diploidie und sexuelle Vermehrung (Meiosis) auf die Erhaltbarkeitsbedingung (Gl. (39)) auswirken.

Diploidie und Rekombination

Sexuelle Vermehrung setzt die zeitweise Vereinigung von väterlichem und mütterlichem Chromosomensatz in einer Zelle voraus. Bei vielen Einzellern stellt diese diploide Zelle, die Zygote, ein Überdauerungsstadium dar. Bei sehr vielen Vielzellern besteht hingegen der Organismus aus diploiden Zellen, die daher länger der Selektion ausgesetzt sind, als die haploide Zellgeneration (die Gameten).

Der Übergang zur Diploidie ändert an der Entropieänderung durch Selektion insofern nichts, als Gl. (28) ihre Gültigkeit beibehält. Die Selektion verläuft gleich, unabhängig davon ob die Phänotypen haploid oder diploid sind und daher kommt man bei der Beschreibung dieses Vorgangs auch zum selben Resultat.

Anders ist es (vermeintlich) mit der Entropieänderung durch Mutation: in Gl. (36) ändert sich bei Vorhandensein von zwei Chromosomensätzen pro Zelle zumindest die Anzahl der Nukleotide.

Für die Anzahl der Nukleotide pro diploider Zelle schreiben wir n' .

Sicher beeinflußt die Diploidie unmittelbar weder die Populationsgröße N noch die Mutationsrate pro Nukleotid p . Wir schreiben daher für die Gesamtentropieänderung $\Delta S_{\text{mut}}'$ durch Mutation bei Diplonten:

$$\Delta S_{\text{mut}}' = N n' p \Delta s_{\text{mut}}', \quad (49)$$

wobei offensichtlich gilt: $n' = 2n$.

$\Delta s_{\text{mut}}'$, die mittlere Entropieänderung pro Nukleotid durch Mutation bei Diplonten hängt vom Informationsgehalt des diploiden Genoms ab. Väterlicher und mütterlicher Chromosomensatz stimmen an etwa 99,9% der Nukleotidpositionen überein und der Gesamtinformationsgehalt des diploiden Genoms ist daher nur sehr geringfügig größer, als der des haploiden Genoms. Daher ist der mittlere Informationsgehalt pro Nukleotid des diploiden Genoms nur etwa halb so groß, wie der des haploiden, d. h. $\Delta s_{\text{mut}}' \cong \frac{1}{2} \Delta s_{\text{mut}}$. Und daraus folgt für die Gesamtentropieänderung (Gl. (49)): $\Delta S_{\text{mut}}' \cong \Delta S_{\text{mut}}$. Die Erhaltbarkeitsbedingung Gl. (39) behält also für Diplonten ihre Gültigkeit.

Auch durch die Rekombination ändert sich der Informationsgehalt nicht: die beiden Rekombinanten besitzen offensichtlich die gesamte genetische Information des Elternpaares. Empirisch läßt sich ebenfalls zeigen, daß die Meiose die Erhaltbarkeit von genetischer Information nicht beeinflußt: aus dem Übergang von der sexuellen Vermehrung zur parthenogenetischen (wie er selbst bei Landwirbeltieren, z. B. manchen Eidechsen und Rennechsen, vorkommt) resultiert keine oder keine drastische Änderung der Genomgröße.

Danksagung

Meiner Frau, Mag. Astrid Tiefenbrunner, möchte ich für die Anfertigung des englischsprachigen Textteils herzlich danken. Herrn Dr. Walter Zislavsky danke ich für die kritische Durchsicht des Originalmanuskripts.

- Alberts B., Bray D., Lewis J., Raff M., Keith R. und Watson J. D. (1986), *Molekularbiologie der Zelle*. VCH-Verlagsgesellschaft, 1310 pp.
- Brillouin L. (1956), *Science and Information Theory* (2nd edition, 1962). Academic Press, London, 351 pp.
- Eigen M. (1971), Selforganisation of matter and the evolution of biological macromolecules. *Die Naturwissenschaften* **58**, 465–523.
- Eigen M., Gardiner W., Schuster P. und Winkler-Oswatitsch R. (1981), Ursprung der genetischen Information. *Spektrum der Wissenschaft* **6**, 36–56.
- Eigen M. (1987), *Stufen zum Leben*. Piper, München, Zürich, 311 pp.
- Ewens W. J. (1979), *Mathematical Population Genetics*. Springer Verlag, Berlin, Heidelberg, New York, 325 pp.
- Falconer D. S. (1984), *Einführung in die quantitative Genetik*. Verlag Eugen Ulmer, Stuttgart, 472 pp.
- Gerthsen C. und Vogel H. (1993), *Physik*. 17. Auflage. Springer Verlag, Berlin, Heidelberg, New York, 943 pp.
- Hofbauer J. und Sigmund K. (1984), *Evolutionstheorie und dynamische Systeme: mathematische Aspekte der Evolution*. Verlag Paul Paray, Berlin, Hamburg, 213 pp.
- Kimura M. (1968), Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res., Camb.* **11**, 247–269.
- Kimura M. (1968), Evolutionary rate at the molecular level, *Nature* **217**, 624–626.
- Kimura M. (1987), *Die Neutralitätstheorie der molekularen Evolution*. Parey, Berlin, Hamburg, 303 pp.
- Klein J., Satta Y., Takahata N., O'Huigin C. (1993), Trans-specific MHC polymorphism and the origin of species in primates. *J. Med. Primatol.* **22**, 57–64.
- Maynard Smith J. (1989), *Evolutionary Genetics*. Oxford University Press, Oxford, 325 pp. (Deutsche Fassung: *Evolutionsgenetik*, Thieme 1992).
- Sarich V. M. und Wilson A. C. (1967), Rates of the albumin evolution in primates. *Proc. Natl. Acad. Sci. USA* **58**, 142–148.
- Sarich V. M. und Wilson A. C. (1967), Immunological time scale for hominid evolution, *Science* **158**, 1200–1204.
- Schuster P. und Swetina J. (1982), Selfreplication with errors: a model for polynucleotide replication. *Biophys. Chem.* **16**, 329–345.
- Schuster P. und Sigmund K. (1980), Self-organisation of biological macromolecules and evolutionary stable strategies. In: *Dynamics of Synergetic Systems*. Springer Verlag, Berlin, Heidelberg, New York, pp. 156–169.
- Zuckerkindl E. und Pauling L. (1962), Molecular disease, evolution, and genic diversity. In: *Horizons in Biochemistry* (M. Kasha und B. Pullman, Herausg.). Academic Press, New York, 604 pp.