

Zwischen Naturgesetz und Wahlfreiheit: Von Darwin zur genetischen Information

KONRAD BACHMANN

Information und Entropie

Information ist erst im zwanzigsten Jahrhundert durch die Arbeiten von Claude Shannon (1948; Shannon & Weaver, 1949) zu einem allgemeinen naturwissenschaftlichen Begriff geworden, und zwar ursprünglich mit der gängigen Bedeutung von (schriftlichen) Nachrichten von Mensch zu Mensch. Es ging darum, wie solche Nachrichten effektiv und fehlerfrei maschinell übermittelt werden können. Shannon sah die Nachricht als lineare Abfolge von verschiedenen Komponenten, die mehr oder weniger beliebig angeordnet sein können (Buchstaben eines Alphabets, Ziffern). Eine bestimmte Anordnung von Komponenten enthält oder ist Information. Information betrifft damit Struktur, die nicht notwendig so durch Naturgesetze festgelegt ist, die aber durch Weitergabe von einem Informationsträger auf andere, durch Reproduktion, länger bestehen bleibt als die einzelnen Informationsträger. Information ist immer an materielle Informationsträger gebunden, seien es Gehirne, Bücher, Inschriften oder elektronische Medien, aber ihre Übertragbarkeit macht sie zur immateriellen Eigenschaft. Information legt beim formbaren Empfänger eine von vielen Möglichkeiten fest, sie trifft Entscheidungen, antwortet auf Fragen. Nachrichten sind Einheiten der Informationsübermittlung, und der *Informationsgehalt* einer Nachricht errechnet sich aus der Anzahl von Zeichen, die bei der Weitergabe dieser Nachricht übertragen werden müssen, und der Wahrscheinlichkeit des Auftretens eines bestimmten Zeichens an einer Stelle der Folge. Shannons binäres System zerlegt die verschie-



Konrad Bachmann, Professor i.R. für Evolutionäre Botanik, Leiter der Abteilung Taxonomie am Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung in Gatersleben, K. Mitglied der Göttinger Akademie seit 1995

denen Wahrscheinlichkeiten formal in Einzelentscheidungen mit Ja-Nein-Antworten mit den Symbolen 1 und 0, jeweils mit einer Wahrscheinlichkeit von $1/2$.

Es zeigte sich sehr bald, dass Shannons Formalismus weit mehr als eine Grundlage für die Technologie der Nachrichtenübermittlung ist. Shannon suchte damals nach einem brauchbaren Namen für das, was wir heute den Informationsgehalt einer Nachricht nennen. Er schwankte zwischen *Information* und *Ungewissheit (uncertainty)*, weil seine Berechnung die Anzahl der möglichen Anordnungen von Zeichen betrifft, aus denen die Nachricht eine bestimmte festlegt (also den Grad der Ungewissheit beim Empfänger, der durch die Nachricht aufgehoben wird). Der Mathematiker John von Neumann empfahl, diese Größe *Entropie* zu nennen: „In the first place, your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage“ (Tribus & McIrvine, 1971). Die Shannonsche Gleichung für den Informationsgehalt stimmt wirklich mit den Gleichungen von Ludwig Boltzmann und Willard Gibbs für die Entropie in der Statistischen Mechanik überein. Die unterschiedlichen Ansätze verbergen sich jeweils in einer Konstante. Nirgendwo wurde die Beziehung zwischen Nachrichtenübermittlung und Statistischer Mechanik (und damit Thermodynamik) derart bedeutungsvoll wie bei der Aufklärung der Informationsweitergabe bei der Fortpflanzung von Organismen, also in der Genetik, und von dort aus bei der Analyse der Entstehung dieser Information, also in der Evolutionsforschung. Darwins Theorie von Variation und Selektion ist letzten Endes eine Theorie von Entropie und Information.

Das Entropiekonzept wurde 1856 von Rudolf Clausius als Lösung für das Problem eingeführt, dass der Erste Hauptsatz der Thermodynamik die Erhaltung der Gesamtenergie bei Energieumsetzungen feststellt, dass es aber unmöglich ist, Wärme vollständig in mechanische Energie umzusetzen. Abhängig von der Temperatur geht ein Teil der Energie, die Entropie, immer als Wärme verloren. Nach dem Zweiten Hauptsatz der Thermodynamik strebt die Entropie in einem geschlossenen System (ohne Austausch von Materie oder Energie) einem Maximum zu. Für den Beobachter („makroskopisch“) ist dieses Maximum ein strukturloses System mit einer einheitlichen Temperatur. Die Statistische Mechanik erklärt das anhand der thermischen Bewegung der Moleküle. Durch ihre Zusammenstöße erreichen die Moleküle im Laufe der Zeit die statistisch wahrscheinlichsten Anordnungen im vorhandenen Raum und eine mittlere Geschwindigkeit, deren Beschreibung also keine Information außer der Anzahl der Moleküle und den statisti-

schen Durchschnittswerten ihrer relevanten Eigenschaften benötigt. Dem Maximum der Entropie des Systems entspricht ein Minimum des Informationsgehaltes. Bei einem *strukturierten* System ist das Ganze mehr als die Summe seiner Teile, und zwar um den zusätzlichen Informationsgehalt. Der Informationsgehalt eines strukturierten Systems aus Molekülen entspricht der Differenz zwischen der maximalen Entropie eines Systems aus denselben Molekülen (errechnet aus allen denkbaren Anordnungen) und der augenblicklichen Entropie aus allen Anordnungen der Moleküle, die als gleiche Struktur wahrgenommen und übertragen werden können. Dem entspricht bei Nachrichten die Anzahl der möglichen Anordnungen von Buchstaben, die dieselbe Nachricht übermitteln.

Bei diesen Überlegungen spielt der Betrachter (der Empfänger) eine wichtige Rolle beim Unterschied zwischen *Struktur* und *Information*. Information betrifft Struktur, soweit sie von einem Empfänger erkannt (also übernommen) und interpretiert wird und die Form, das Verhalten oder das Befinden des Empfängers beeinflusst. Die formale *mikroskopische* oder *reduktionistische* Behandlung von Information als Anordnung von Komponenten verlangt also ein Verständnis dafür, was vom Empfänger als Komponente angesehen wird, also etwa das Auflösungsvermögen seiner Wahrnehmung, und welche Anordnungen davon für den Empfänger relevant sind. Bei der Informationstheorie spielte das anfangs keine Rolle, weil man selbstverständlich von Lesern ausging, die den Buchstaben eines Textes den Inhalt entnehmen. Dabei ist zum Beispiel das Schriftbild konstante und austauschbare Hintergrundinformation. Man kann aber auch einen handgeschriebenen Text darauf untersuchen, was die Handschrift über den oder die Schreiber während des Schreibens aussagt, auch wenn der Text selbst dafür irrelevant ist. Die technische Übermittlung der relevanten Information aus dem Original sieht dann ganz anders aus. Bei der Analyse irgendeiner Art von Informationsübertragung ist das Wahrnehmungsvermögen und das Interesse (die Formbarkeit, die Fragestellung) des Empfängers der entscheidende Faktor. Der Empfänger muss zur Nachricht in Form und Inhalt komplementäre Information besitzen. Viele Missverständnisse bei der Behandlung von Informationsübertragung in der Biologie haben auf der Annahme beruht, dass wir Signale so empfangen und interpretieren, wie sie der natürliche Empfänger wahrnimmt. Information mag also ein naturwissenschaftliches Grundphänomen sein, aber sie unterscheidet sich dadurch von Phänomenen wie Energie oder Masse, dass sie eine subjektive Eigenschaft ist. Nicht nur das. Es ist zwar keineswegs alle Information Kommunikation. Wenn wir aus dem Fenster schauen, um zu sehen, wie das Wetter und wie stark der Verkehr auf der Straße ist, dann ist das für uns

relevante Information, die aber nicht für uns produziert worden ist, anders als bei Wetterberichten oder Verkehrsberichten im Radio. Sehr viel Information entsteht aber wirklich als Teil von Kommunikation und zielt auf die Wahrnehmung und Interpretation durch einen bestimmten Empfänger. Das kann ein anderer Organismus sein, es kann aber auch eine Zelle in einem Organismus oder ein Molekül sein oder ein Apparat, der programmiert ist, bestimmte Signale zu erkennen, zu interpretieren und entsprechend zu funktionieren. Lange bevor es Menschen gab, war Information nicht nur ein selektiv und subjektiv definiertes Phänomen, sie hatte auch einen Sinn und Zweck für den Empfänger und ein Ziel für den Sender. Mit dem Konzept von Information sind Subjektivität, Bedeutung und Zielgerichtetheit zu legitimen naturwissenschaftlichen Themen geworden. Wir müssen uns von dem Gedanken lösen, dass sie Bewusstsein oder Intelligenz voraussetzen.

*Embryonalentwicklung und Evolution:
Die Selbstorganisation zielstrebigter Programme*

Der Zweite Hauptsatz der Thermodynamik betrifft zwar nur geschlossene Systeme ohne Austausch von Material und Energie mit ihrer Umgebung. Er schien aber die allgemeine spontane Tendenz zum Zerfall jeglicher Struktur zu beschreiben, sei es das Zerbröckeln von Steinen, das Zerfallen von Papier oder das unvermeidliche Altern und Sterben von lebenden Organismen, und er schien die spontane Entstehung von komplexen Strukturen auszuschließen. Die Thermodynamik schien keine naturwissenschaftliche Erklärung für das Entstehen von Lebewesen zuzulassen, weder für die Individualentwicklung von der befruchteten Eizelle zum geschlechtsreifen Organismus, noch für die Entstehung der verschiedenen Organismen überhaupt. Die Individualentwicklung ist ein zielgerichteter Vorgang. Sie strebt einem spezifischen Endzustand zu, der sehr genau vorherbestimmt ist, selbst in seinen umweltabhängigen Varianten, seiner *Reaktionsnorm*. Noch im 18. Jahrhundert hat man angenommen, dass die Individualentwicklung nur das Größenwachstum einer winzigen, aber vollständigen Kopie des Organismus ist, und sich darum gestritten, ob diese Kopie im Ei oder in den „Spermaterchen“ vorliegt. Seit Ende des 18. Jahrhunderts wurde es zunehmend deutlich, dass diese *Präformationstheorie* nicht stimmen konnte, sondern dass die Struktur eines Organismus *epigenetisch* entsteht (Das Wort „epigenetisch“ hat in der heutigen Biologie mehrere verschiedene Bedeutungen). Was man seit Aristoteles mit dem Wort *Entelechie* beschrieben hat, also die interne Regulation der Entwicklung eines Systems auf ein vorbestimmtes Endziel hin, wird heute als ein algorithmisches steuerndes Programm mit

Informationsübertragung gesehen. Dass die *Evolution der Organismen* kein derartiger Prozess mit dem Menschen als vorgegebenem Ziel ist, ist zwar selbst bei einem oberflächlichen Vergleich von Individualentwicklung und Evolution deutlich, aber der Gedanke hat sich sehr viel länger gehalten. Noch im 19. Jahrhundert waren die Wörter *Entwicklung* und *Evolution* in der Biologie praktisch austauschbar.

Der historische Prozess der Evolution war im 19. Jahrhundert schon vor Darwin die unvermeidliche naturwissenschaftliche Erklärung für die hierarchisch gestaffelte Vielfalt der Organismen und die geologische Abfolge der Fossilien. Was zur Akzeptanz der Theorie fehlte, war ein natürlicher Mechanismus. Die Biologen seiner Zeit haben Darwins Theorie bereitwillig als Beleg dafür akzeptiert, dass ein natürlicher Mechanismus für die Evolution der Organismen gefunden werden kann, und sie haben die historische Verwandtschaft in der Form eines Stammbaums zur Grundlage des Systems der Organismen gemacht. Wirklich verstanden hat die Theorie anfangs kaum einer. Es hat auch kaum einer den Versuch gemacht. In einer Zeit, in der man allgemein glaubte, aus natürlichen Mechanismen Maximen für menschliches Handeln ableiten zu müssen, schien die Darwinsche Theorie mit Schlagworten wie „Überleben des Stärksten“ und „Kampf aller gegen alle“ ein Aufruf zum Umsturz aller traditionellen Werte zu sein. Erklärte Darwinisten haben seine Theorie bedenkenlos mit der gängigen Vorstellung einer zielgerichteten Entwicklung zum Menschen, speziell zu ihrer eigenen Rasse oder Klasse, verbunden und daraus einen natürlichen Auftrag zur Erhaltung dieses Zieles durch Züchtung und Selektion abgeleitet. Es ist hier nicht der Platz, um diese naiven ideologischen Anpassungen, ihre Motivationen und die entsetzlichen Folgen ihrer zynischen Ausnutzung zu untersuchen. Erst das Konzept der *genetischen Information* hat schließlich den Unterschied zwischen Individualentwicklung und Evolution deutlich gemacht. Individualentwicklung ist Fortschritt mit detailliert vorgegebenem spezifischen Ziel, Evolution ist das unvermeidliche Nebenprodukt der Tendenz zur Bewahrung des Erreichten durch Weitergabe von Information. Entwicklung entspricht der materiellen Umsetzung der genetischen Information, die Evolution deren Entstehung und Veränderung. Verbunden sind beide durch den Prozess der Reproduktion, also der Herstellung multipler Kopien durch Informationsübertragung. Genetik, Evolutionsbiologie und Entwicklungsbiologie haben sich lange Zeit überraschend unabhängig voneinander entwickelt. Seit der Mitte des 20. Jahrhunderts hat die Molekularbiologie eine entscheidende Rolle bei der praktischen und der intellektuellen Zusammenführung der verschiedenen Ansätze gespielt.

Organismen als thermodynamisch offene Systeme mit regelnder Information

Ein ganz offensichtlicher Ausgangspunkt zur Abgleichung des Zweiten Hauptsatzes und der Existenz von komplexer dynamischer Struktur ist die Beobachtung, dass lebende Organismen keine thermodynamisch geschlossenen Systeme sind, sondern für ihr Bestehen von einem ununterbrochenen Strom von Materie und Energie abhängen. Erste Information dazu haben schon Antoine Lavoisier und Pierre Simon Laplace 1780 geliefert, als sie die Wärmeproduktion und die Produktion von Kohlendioxyd bei einem Kaninchen mit der einer Kerzenflamme verglichen und zeigten, dass bei beiden das gleiche quantitative Verhältnis vorlag. Auch wir Menschen beziehen unsere Energie (nicht zu verwechseln mit unserer Motivation) offensichtlich aus der „geregelten Verbrennung“ unserer Nahrungsstoffe. Julius Robert Mayer, einer der Entdecker des Satzes von der Erhaltung der Energie, kam zu seiner Entdeckung durch die Beobachtung von Patienten, nicht von Dampfmaschinen. Heutzutage sind wir alle genau über den notwendigen Energiebedarf und über die entsprechenden Brennwerte von Fetten, Kohlehydraten und Eiweißen informiert und nehmen keinen Anstoß am Modell eines Organismus als Verbrennungsmaschine. Die Komplexität, die Stabilität und die Leistungen unserer Körper werden gerade durch ein Verständnis der physikalischen Prinzipien bewundernswert. Dem allen liegt eine Komplexität der Informationsübertragung zugrunde, die alle bisherigen Leistungen menschlicher Technologie weit übertrifft.

Die strukturelle Komplexität von offenen Systemen bleibt erhalten, indem durch Zufuhr von nutzbarer Energie und Abfuhr von Entropie das Erreichen von Gleichgewichtszuständen verhindert wird. Was wie eine konstante Struktur aussieht, ist ein dynamisches Fließgleichgewicht: Unkontrollierter langsamer Zerfall wird durch schnellen geregelten Abbau vorweggenommen und durch Nachschub und Synthese ausgeglichen. Dabei kann die spontane Tendenz zum Erreichen von Gleichgewichten, also den energetisch wahrscheinlichsten Anordnungen der vorhandenen Komponenten, als treibende Kraft genutzt werden. Im Stoffwechsel lebender Organismen betrifft das die Gleichgewichtskonzentrationen von miteinander reagierenden Molekülen. Sie werden verhindert, indem durch Aufnahme oder Synthese ständig eine hohe Konzentration energiereicher Moleküle aufrechterhalten wird, die spontan im Gleichgewicht praktisch völlig verschwinden würden. Der Zerfall dieser Moleküle wird nicht dem langsamen Zufall überlassen, sondern sie werden in einer festgelegten Abfolge von Schritten gezielt dem Gleichgewichtszustand zugeführt. Das ist die „geregelte Verbrennung“ von Nahrungsmolekülen, bei der die Energie nicht sofort im Ganzen als Wärme

abgegeben wird. Bei dem schrittweise kontrollierten Abbau ist es möglich, Abbau- und Synthesereaktionen derart zu koppeln, dass ein Teil der freiwerdenden Energie aus dem Abbauschritt zur Synthese eines energiereichen Produkts genutzt werden kann, das erst später abgebaut wird. Im offenen System kann die Gesamtentropie *verringert* werden, es kann spezifische Struktur entstehen und erhalten bleiben, wenn energiearme Endprodukte und Wärme an die Umgebung abgegeben werden. Dafür braucht der Organismus aber nicht nur eine verlässliche Quelle von einer für ihn nutzbaren Energieform, sondern eine komplexe Apparatur zur informationsgesteuerten Kontrolle der Stoffwechselschritte.

Die entscheidende Rolle beim Informationsfluss in lebenden Organismen spielen Enzyme als Katalysatoren von chemischen Prozessen. Ein Enzym erkennt und bindet die molekularen Komponenten einer bestimmten Reaktion so, dass diese Reaktion bevorzugt und beschleunigt stattfindet. Das entspricht einer bestimmten Auswahl aus allen möglichen Reaktionen, also der Übertragung von Information auf den Stoffwechsel. Serien von Enzymen, von denen eines die Reaktionsprodukte des anderen übernimmt und einen Schritt weiter verarbeitet, leiten den Zellstoffwechsel entlang von Stoffwechselwegen, die einen schrittweise erfolgenden Abbau der Energie und die Übertragung eines erstaunlich großen Anteils der Energie von Abbaureaktionen auf Synthesereaktionen garantieren. Das sind Hunderte von Reaktionen, aber sie bilden zusammen ein raffiniertes Streckennetz für die Weitergabe von Molekülen und Bruchstücken von Molekülen entlang von alternativen Pfaden bis hin zu den energiearmen Molekülen, die als nicht weiter verwertbare Stoffwechselendprodukte an die Umgebung abgegeben werden. Der hierarchisch gestaffelten körperlichen Ähnlichkeit von Organismen entspricht eine hierarchisch gestaffelte Ähnlichkeit ihrer Stoffwechselwege, also ihrer Enzymausstattung. So, wie die vergleichende Morphologie Hinweise auf eine gemeinsame Abstammung aller Organismen gibt, tut das auch die vergleichende Biochemie. Das Interessante dabei ist, dass beide zwar denselben Stammbaum erkennen lassen (wie das die Theorie voraussagt), dass aber entscheidende Phasen der biochemischen Evolution der morphologischen Evolution vorausgegangen sind, so dass sich die Daten aus Morphologie und Biochemie ergänzen. Wieso gerade molekulare Strukturen lebender Organismen so unvorstellbar konservativ erhalten bleiben, wird deutlich, wenn wir die Konstruktionsprinzipien dieser Moleküle anschauen.

*Proteine sind die Empfänger und Vermittler von Information
in lebenden Organismen*

Chemisch sind die Enzyme des Stoffwechsels Eiweiße, Proteine. Im Rahmen der Informationsübertragung in der Zelle sind Proteine die spezifischen molekularen Empfänger und Verarbeiter von Information. Proteine sind große und komplexe Moleküle aus tausenden von Atomen in einer bestimmten dreidimensionalen Anordnung. Ein Eiweißmolekül enthält also sehr viel Information, wenn der Betrachter an der genauen Anordnung der Atome interessiert ist. Im Stoffwechsel sind Eiweiße aber primär Empfänger (Rezeptoren) von Information, und ein Großteil ihrer Struktur ist ein Gerüst für den Teil des Moleküls, mit dem es sein spezifisches Zielmolekül (seinen Liganden, sein Substrat) erkennt. Dazu dient ein Teil der Oberfläche des Proteinmoleküls, der ein negatives Abbild des Ligandenmoleküls ist, so wie das Innere eines Handschuhs für eine Hand. Der Ligand passt genau in diese Bindungsstelle. Er geht keine feste (kovalente) chemische Bindung mit dem Protein ein, sondern wird reversibel durch eine Reihe von molekularen Anziehungskräften, „schwachen Bindungen“, gehalten, die alle zusammen sicherstellen, dass das richtige Molekül gebunden ist. Es ist nicht leicht, ein Protein dabei zu täuschen, aber es gibt keine Kommunikation, die nicht irgendwie manipuliert wird. Die meisten Drogen sind alternative Signale an spezifische Proteine. Die Bindung eines passenden Moleküls wird vom Protein in eine Ja/Nein-Entscheidung umgesetzt, bei der das Protein entweder selbst ausführende Partei ist (wie bei der enzymatischen Katalyse) oder sein Bit an Information an die Zelle übermittelt. Die Reaktion eines Proteins auf die Bindung seines Liganden ist oft das Resultat einer leichten Verformung von bindendem und gebundenem Molekül durch die gegenseitige Anziehung der „schwachen Bindungen“. Bei einem Enzym kann das die spezifische Reaktion des gebundenen Substrats bestimmen. Diese Verformung kann aber auch als Signal von anderen Proteinen erkannt werden, wenn sie Bindungsstellen für Oberflächenstrukturen des Senderproteins haben, die erst bei dessen Verformung entstehen. Das ist der Grundvorgang für den Empfang und die Weitergabe von Information in der Zelle. So können Rezeptorproteine in der Zellmembran mit einer Bindungsstelle an der Außenseite der Zelle Signalmoleküle aus der Umgebung (Hormone, Neurotransmitter) binden und diese Bindung nach innen an Empfängerproteine im Cytoplasma melden. Welche Zelle auf welches Signal reagiert, hängt von ihrer zellspezifischen Ausstattung mit Rezeptorproteinen ab. Komplexere Erkennungsprozesse, einschließlich unserer bewussten Wahrnehmung, beruhen auf einer programmierten Koordinati-

on von vielen solchen Einzeleffekten in vielen Zellen. Auf der dynamischen Bindung von Proteinen mit Proteinen, oft in komplexen Anordnungen, beruhen auch viele Strukturen in der Zelle und alle zellulären Bewegungsvorgänge. Wichtig ist dabei, dass das allgemeine Konstruktionsprinzip von Proteinmolekülen derart flexibel ist, dass damit Bindungsstellen für ein weites Spektrum von Molekülen gebildet werden können. Durch Proteine wird Energie aus der Umgebung in den Energiefluss der Zelle eingespeist, und dabei wird Information aus der Gesamtheit der beteiligten Proteine koordiniert in die dynamische Struktur von Zellen und Organismen umgesetzt. Diese Information liegt in Tausenden von verschiedenen Proteinen, jedes aus Tausenden von Atomen in genauer Anordnung. Man hat einmal gedacht, dass Proteine bei der Synthese von weiteren Kopien ihre Information auf diese Kopien übertragen, aber gerade das ist nicht der Fall, und das betrifft eine der wichtigsten Erkenntnisse zur Evolution. Proteine übertragen Information, für deren Erkennung sie programmiert sind, aber sie übertragen nicht die Information über ihren eigenen Aufbau.

Dass Information vom Interesse des jeweiligen Betrachters abhängt, lässt sich gut an der Rolle der Proteine als Empfänger, Ausführende und Überträger von Information demonstrieren. *Unser* Verständnis ihrer Funktion verlangt Information über die dreidimensionale Anordnung von Tausenden von Atomen. In der Zelle sind Proteine aber *Empfänger* von Information, und die beruht auf der Lage und der chemischen Reaktivität von Punkten der Bindungsstelle, an denen Atome des Proteins mit Atomen des gebundenen Liganden in Berührung kommen, also auf einer Identifikation durch das Abgleichen von einer beschränkten, aber für den Zweck ausreichenden Anzahl von Merkmalen. Schließlich wird diese Identifikation für die Zelle in eine einzelne Ja/Nein-Entscheidung umgesetzt: „Bindungsstelle besetzt“ oder „Bindungsstelle nicht besetzt“. Alle diese Bindungsreaktionen sind reversibel, so dass der Zustand „Bindungsstelle besetzt“ von der Konzentration des Liganden abhängt (wie schnell ein freigeskommenes Ligandenmolekül durch ein anderes ersetzt wird). Das ist wichtig, um die Signalstärke und ihre zeitliche Veränderung verfolgen zu können. Gerade die additiven Wirkungen von „schwachen Bindungen“, die einzeln schon bei normaler Temperatur instabil sind, spielen wichtige Rollen überall in der dynamischen Regulation der Zelle. Neben der Bindung zwischen Protein und Substrat sind solche schwachen Bindungen auch entscheidend für die dreidimensionale Struktur der Proteinmoleküle selbst und damit für den Schlüsselmechanismus der biologischen Informationsübertragung, die Festlegung einer dreidimensionalen, wirksamen Struktur durch eine lineare Folge von Bausteinen.

Makromoleküle: ein generisches Bauprinzip schafft Freiheit für spezifische Information, lineare Information schränkt räumliche Form auf einen Endzustand ein

Proteine sind eine Klasse von *Makromolekülen* der Zelle, lange Ketten aus einem beschränkten Satz von chemischen Grundbausteinen. Bei Proteinen sind das zwanzig verschiedene Aminosäuren. Alle Aminosäuren haben die gleichen zwei reaktiven Gruppen (Amino- und Carboxylgruppe), über die sie in beliebiger Folge zu langen beweglichen Peptidketten verbunden werden können. Die jeweils verschiedenen chemischen Gruppen der aneinanderhängenden Aminosäuren stehen dann abwechselnd rechts und links aus der Kette heraus. Ein Protein besteht aus einigen hundert Aminosäuren in ganz bestimmter linearer Folge. Die chemische Formel für ein bestimmtes Protein kann also durch die Aminosäuresequenz als digitale Nachricht aus einem Alphabet mit zwanzig Zeichen ersetzt werden, wobei Kenntnis der Strukturen der einzelnen Aminosäuren als bekannte, für alle Proteine gleiche Hintergrundinformation vorausgesetzt wird. Mit der linearen Folge sind zwar implizit alle Atome des Proteinmoleküls mit ihren festen (kovalenten) chemischen Bindungen spezifiziert, aber die Kette kann noch erstaunlich frei gedreht und gefaltet werden, und nur eine von astronomisch vielen Verknäuelungen der Kette ist das funktionierende dreidimensionale Proteinmolekül. Stabilisiert wird diese dreidimensionale Struktur durch die Anziehungskräfte zwischen den vielen reaktiven Gruppen der Aminosäuren, die darin einander gegenüber liegen, also durch Hunderte von schwachen Bindungen verschiedener Art. Bei hoher Temperatur kommen die schwachen Bindungen nicht mehr gegen die thermische Bewegung auf, und die Proteinstruktur wird *denaturiert*. Wohl die erstaunlichste Beobachtung in der Geschichte der Proteinchemie war der Befund von Christian Anfinsen (1961), dass ein einzelnes denaturiertes Proteinmolekül unter den richtigen physiologischen Bedingungen spontan wieder seine *native* dreidimensionale funktionelle Struktur annahm. Das ließ sich bald zu der Erkenntnis verallgemeinern, dass bei der Synthese eines bestimmten Proteins in der Zelle nur die Reihenfolge der Aminosäuren spezifiziert wird. Die Kette nimmt dann von selbst die richtige, funktionierende Form an. Das ist keineswegs selbstverständlich. Eine der folgenreichsten Leistungen der Evolution ist die Speicherung der Information über diejenigen Aminosäurefolgen, die bei normaler Temperatur in kürzester Zeit durch die Konkurrenz zwischen schwachen Bindungen und thermischer Bewegung eine energetisch stabilste Faltung finden, die dann auch noch eine spezifische Rolle bei der Informationsübertragung in der Zelle hat. Man kann sich die Aufgabe

vielleicht durch den Vergleich mit einem Klappstuhl verdeutlichen, der sehr viel weniger verbundene Einzelteile hat, bei dem aber die funktionierende Faltung *nicht* die energetisch wahrscheinlichste ist. Man muss zumindest eine Vorstellung vom Ziel, möglichst auch vom Weg dahin haben, also *zusätzliche Information*, um die brauchbare Struktur zu erhalten, und diese Struktur droht immer, bei starken Bewegungen spontan in die energetisch stabilste zu kollabieren. Bei Proteinen in der Zelle wäre so etwas gefährlich. Es gibt ganz wenige derart labile Proteine. Die Creutzfeldt-Jakob-Krankheit und der Rinderwahn scheinen darauf zu beruhen.

Die verschiedenen Strukturbedingungen, die die Aminosäuresequenz eines funktionellen Proteins in einem Organismus gleichzeitig erfüllen muss, machen erst deutlich, was für ein erstaunlicher Prozess die Evolution von Proteinen durch Variation und Selektion ist. Proteine sind übrigens wahrscheinlich eine sekundäre Erfindung der Evolution, die wohl mit selbstreproduzierenden Nukleinsäuren begonnen hat, und zwar mit Ribonukleinsäuren. Auch das sind Makromoleküle, Ketten von Untereinheiten, in diesem Falle von vier verschiedenen Nukleotiden, deren unterschiedliche Bestandteile organische Basen (A, C, G, U) sind. Auch diese Ketten können spontan dreidimensionale Faltungen durch schwache Bindungen zwischen jeweils komplementären Basen (A mit U, G mit C) annehmen, und einige solche Faltungen können katalytisch wirken. Primär kann die Basenfolge von Nukleinsäuren aber über die komplementären Bindungen Information zur Synthese von Kopien von sich selbst liefern. Mit der Erfindung von DNA (Desoxyribonukleinsäure) als Spezialform einer Nukleinsäure für Lagerung und Selbstreproduktion der Information und der Erfindung eines Nukleinsäurecodes für Aminosäuresequenzen von Proteinen war die Trennung zwischen der Erhaltung und Weitergabe der Information (DNA) und ihrer Auseinandersetzung mit der Umwelt (Protein) eingeleitet. Die genetische Information füllte sich mit Instruktionen für Strukturen und Reaktionen, die nur mittelbar dem direkten Ziel ihrer eigenen Weitergabe dienen, indem sie die Reaktionen in der Zelle gegen Schwankungen in den äußeren Bedingungen abschirmen. Die RNA hat dabei eine vermittelnde Rolle sowohl bei der Informationsübertragung (messenger RNA, mRNA) wie beim Übersetzen des Codes (transfer RNA, tRNA) behalten, und RNA hat im Laufe der Evolution eine große Anzahl neuer Funktionen dazugewonnen.

Die gemeinsame Ahnenform aller heute lebenden Organismen vor mehr als dreieinhalb Milliarden Jahren muss bereits ein Einzeller mit DNA und Proteinen gewesen sein, und dieser *Last Universal Common Ancestor* (LUCA) muss auch schon die Information für die meisten Aminosäuresequenzen

enthalten haben, die bis heute der Grundstock für die Funktionen der Zelle sind. Ein Vergleich der Proteine aller lebenden Organismen zeigt, dass die meisten auf ein paar Hundert Urformen rückführbar sind. Viel Proteinevolution seitdem besteht aus kombinatorischer Zusammenstellung dieser Blöcke zu neuen funktionellen Einheiten. Aminosäureaustausche in der Kette als Folgen zufälliger Mutationen finden ununterbrochen in allen Organismen statt, aber die meisten behindern oder zerstören die Funktion des Proteins. Sequenzen ändern sich im Laufe der Evolution, aber derart langsam, je nach der Funktion des Proteins, dass man die Anzahl unterschiedlicher Aminosäuren in den Kopien eines Proteins in verschiedenen Arten oder Gruppen von Organismen näherungsweise als eine *molekulare Uhr* benutzen kann, die zeigt, wann die Arten ihre unabhängige Evolution von einem gemeinsamen Ahnen aus begonnen haben. Molekulare Daten erlauben eine überraschend präzise Stammbaumrekonstruktion einschließlich quantitativer Abschätzungen über die Verlässlichkeit der Schlüsse.

Von Darwin zur DNA

Ich habe oben schon angedeutet, dass Evolutionsbiologie und Genetik in ihrer intellektuellen Entwicklung voneinander unabhängiger waren, als man das heute annehmen sollte. In der Genetik ging es um den Mechanismus der Weitergabe von Information von Generation zu Generation, in der Evolutionsbiologie um die Veränderung (und erst später um den Ursprung) dieser Information. Der empirische Nachweis von erblichen Unterschieden zwischen den Individuen einer Art war eine der beiden Komponenten der Darwinschen Theorie. Darwin entkräftete damit endgültig die Vorstellung von einer Art als unveränderlicher platonischer Idee (griechisch εἶδος, lateinisch *species*), die in den materiellen Individuen mehr oder weniger präzise abgebildet ist. Noch Linné hatte diese Vorstellung mit seinem System in der Biologie verankert. Nach dieser Vorstellung sollte man hinter den zufälligen Unterschieden zwischen den Individuen einer Art die allen gemeinsamen wesentlichen (*essentiellen*) Merkmale der Art finden können. Diese Auffassung heißt deshalb Essentialismus. Darwin zeigte, dass sich das Bild einer Art aus mehr oder weniger unabhängigen erblichen Merkmalen zusammensetzt, die bei verschiedenen Individuen derselben Art in alternativen oder multiplen Merkmalszuständen vorliegen können. Selektion bei der Vererbung kann die Häufigkeit und Zusammenstellung der Merkmalszustände ändern, bis eine neue Art entsteht. Eine evolutionäre Verwandtschaft aller Organismen verlangte die prinzipielle Veränderlichkeit *aller* Merkmale.

Damit veränderte Darwin die Art von einer gegebenen Grundeinheit der Organismenvielfalt in das vorübergehende, mehr oder weniger verlässlich abgrenzbare dynamische Resultat einer normalisierenden Selektion. Eine Grundtendenz zur ungerichteten Variation in alle Richtungen wird dabei in eine deutlich, aber nicht nach einem einheitlichen Schema lokal strukturierte Vielfalt verändert. Ohne es zu wissen, führte Darwin damit parallel zu der gleichzeitigen Entwicklung der Thermodynamik auf einer höheren Organisationsebene das Prinzip der spontanen Tendenz zum Zerfall von Struktur ein. Die andere Komponente seiner Theorie, das Prinzip der Selektion, zeigte, wie unter diesen Umständen komplexe Strukturen entstehen und erhalten bleiben können. Was fehlte, war die Einsicht in die fundamentale Bedeutung der (identischen) Replikation und Vermehrung von Einheiten durch Informationsweitergabe. Die Tatsache konnte als selbstverständlich vorausgesetzt werden, über den Mechanismus wusste man nichts. Darwin hatte die erbliche Variabilität von Merkmalen empirisch abgesichert, aber er wusste nichts über den Ursprung der Merkmalsunterschiede und nichts über den Mechanismus von deren Weitergabe. Er konnte noch nicht einmal ausschließen, dass die Merkmalsunterschiede bei der Züchtung von Pflanzen und Tieren nicht etwa erst durch die künstliche Haltung und Selektion entstanden. Sachliche Kritik an seiner Theorie konnte da angreifen. Darwin war sich bewusst, dass seine Theorie letztlich von der Aufklärung des Vererbungsmechanismus abhing. Er hat versucht, aus allen relevanten Beobachtungen eine Theorie der Vererbung abzuleiten (Darwin, 1868). Diese Vererbungstheorie ist historisch interessant, aber falsch. Sie zeigt übrigens, dass Darwin (selbstverständlich) trotz seiner Betonung der erblichen Variation von identischer Replikation von Merkmalen (in verschiedenen Varianten) als Grundvorgang der Vererbung ausging.

Zur selben Zeit hatte Gregor Mendel schon aus gezielten sorgfältigen Experimenten eine Theorie der Vererbung abgeleitet, die in einem Punkt ihrer Zeit weit voraus war. Mendel machte einen deutlichen Unterschied zwischen sichtbaren Merkmalen und hypothetischen Trägern genetischer Information, für die Wilhelm Johannsen 1909 den Namen *Gene* einführte. Alternative Merkmalszustände wie rote oder weiße Blütenfarbe wurden nun als Resultat von alternativen Formen von Genen, den *Allelen* eines Gens „für Blütenfarbe“, interpretiert. Zwei (gleiche oder verschiedene) davon liegen in jeder Körperzelle, je eines davon gelangt in einer Reduktionsteilung in eine Keimzelle. Überraschende Effekte bei der Vererbung von Merkmalen, vor allem das Überspringen von Generationen, konnten nun durch die Interaktion der beiden verschiedenen Allele des entsprechenden Gens in einem Individuum erklärt werden. Gleichzeitig mit der Neuentdeckung

der Mendelschen Theorie wurde auch der spontane Ursprung von erblichen Varianten, *Mutation*, nachgewiesen. Die ersten Fälle betrafen aber nicht den typischen Fall der Mutation eines Allels in ein neues Allel, und das hat für lang anhaltende Missverständnisse gesorgt.

Dass den hypothetischen Genen wirkliche materielle Strukturen entsprechen mussten, war sicher, aber solange die chemische Struktur von Genen nicht bekannt war, musste man ihr Verhalten indirekt aus ihren Effekten auf Merkmale erschließen. Dabei wurde deutlich, dass es keine Eins-zu-Eins-Entsprechung zwischen Genen und (beliebigen) Merkmalen geben konnte. Verschiedene Interaktionen von mehreren Genen mussten zur Erklärung der Vererbung eines Merkmals postuliert werden, und ein Gen konnte gleichzeitig Effekte auf verschiedene Merkmale haben. Zudem wurde die genetische Komponente bei der Ausbildung eines Merkmals durch die unvermeidlichen Umwelteffekte auf praktisch alle Merkmale überdeckt. Man musste erst einmal mit statistischen Methoden das genetische Signal von den Umwelteffekten bei der Variation eines Merkmals in einer Stichprobe von Organismen trennen, um dann, auch mit statistischen Methoden, Aussagen über die vorhandene genetische Variation zu treffen. Die statistische Genetik von Merkmalen wurde ein wertvolles Instrument zur Analyse und Planung von Projekten in der Tier- und der Pflanzenzucht. Dabei waren natürlich nur Gene betroffen, die in verschiedenen Allelen vorlagen. Neumutationen spielten bei diesen kurzzeitigen Analysen kaum eine Rolle. Der Bestand an variablen Genen und die vorhandenen Allele hängen von der untersuchten Stichprobe, der jeweiligen Population, ab. Je mehr Gene in verschiedenen Allelen vorliegen, desto mehr kann durch geschickte Selektion erreicht werden. Das ließ sich von der Züchtungsforschung auf die Evolution durch natürliche Selektion extrapolieren. Die statistische Populationsgenetik wurde zur Grundlage der Synthetischen Evolutionstheorie aus Mendelscher Genetik und Darwinischer Selektionstheorie. Es ist verständlich, dass sie sich auf augenblickliche Kurzeitevolution in natürlichen Populationen bis hin zur Artbildung konzentrierte. Sewall Wright fasste das in den vierziger Jahren in der Definition zusammen: *Evolution ist die Änderung der Häufigkeiten von Allelen in Populationen*. Das bedeutete nicht, dass man annahm, alle Organismen „von der Amöbe zum Menschen“ hätten dieselben Gene und unterschieden sich nur in den allelen Formen dieser Gene. Es betraf die augenblicklichen Evolutionsvorgänge, die jederzeit überall nebeneinander in den verschiedenen Populationen von Organismen stattfinden. Die Beziehung zwischen dieser *Mikroevolution* in Populationen und der *Makroevolution* von ganz verschiedenen Organismen typen war umstritten. Die Synthetische Theorie nahm

an, dass Mikroevolution über hinreichend lange Zeit in Makroevolution übergeht. Wir wissen jetzt, dass das so ist, aber der Prozess ist nicht so stetig und statistisch feinkörnig, wie man sich das damals vorstellte. Gelegentlich passiert einmal etwas, was im Augenblick nicht auffällt, was sich aber im Nachhinein als eine qualitative Weichenstellung für die Evolution der Nachkommen dieser Population erweist. Für die dramatischeren dieser Episoden hat John Maynard Smith den Ausdruck *Major Evolutionary Transitions* geprägt. Wo diese evolutionären Neuerfindungen neue Lebensweisen erschlossen haben und nicht bis zum Aussterben mit den bisherigen Formen konkurrierten, haben die Vorher- und die Nachher-Versionen in unterschiedlichen Nachkommenschaften überlebt und können in heute lebenden Organismen verglichen werden. Das erlaubt uns jetzt detaillierte experimentelle Analysen von vielen entscheidenden Neuerungen im Laufe der Evolution.

Die Evolution der genetischen Information umfasst mehr als den Austausch von Allelen von Genen. Aber ohne direkten Zugang zu Genen und Allelen war es beinahe unmöglich, Genaueres darüber herauszufinden. Zur Lösung dieser Aufgabe trugen verschiedene Forschungsrichtungen bei. Aus der Erkenntnis, dass sich die Chromosomen im Zellkern nach den Mendelschen Regeln verhalten, entstand die Cytogenetik, die von einer linearen Anordnung der Gene auf den Chromosomen ausging. Chromosomen bestehen aus Proteinen und DNA. Es war bekannt, dass DNA aus vier Nucleotiden als Untereinheiten besteht, die durch Phosphodiesterbindungen verknüpft sind. Aber während die entscheidende Rolle von Proteinen bei praktisch allen Prozessen in der Zelle immer deutlicher wurde, war wenig oder nichts über die genaue Struktur und Funktion von DNA bekannt. Der Arzt Archibald Garrod hatte schon 1902, also kurz nach der Wiederentdeckung der Mendelschen Theorie, festgestellt, dass die Stoffwechselkrankheit Alkaptonurie nach dem Mendelschen Schema vererbt wird, und er hatte damit eine Verbindung zwischen einem Gen und einem Stoffwechselweg entdeckt. In den vierziger Jahren untersuchten George Beadle und E. L. Tatum die Beziehung zwischen Genmutationen und Stoffwechseleffekten bei einem Pilz, der sich besonders für solche Experimente eignete, und fanden, dass Mutationen in einzelnen Genen den Ausfall von einzelnen Enzymen zur Folge hatten (Beadle & Tatum, 1941). Ein Gen schien also die Information für ein Enzym, ein Protein, zu enthalten. Kurz danach gelang zwei Forschergruppen mit verschiedenen Ansätzen der verblüffende Nachweis, dass nicht Protein, sondern DNA die genetische Information enthält und überträgt (Avery & al., 1944; Hershey & Chase, 1952). Als dann 1953 James Watson und Francis Crick die molekulare Struktur der DNA aufklär-

wen, konnten sie darin sowohl die Grundlage der Informationscodierung wie die Grundlage der Informationsübertragung erkennen.

Die Bekanntheit von Shannons Informationstheorie machte es leicht, die genetische Information in der DNA in einer unregelmäßigen, aber festgelegten Abfolge der vier verschiedenen Nukleotide, hier A, G, T und C, zu vermuten. Besonders aufregend war aber die Entdeckung, dass DNA eine Doppelhelix aus *zwei* umeinander gewundenen Nukleotidketten ist, die gegenläufig nebeneinander liegen, wobei jeweils einem A in einer Kette ein T in der komplementären und einem C ein G gegenüberliegt. Diese Paarungen beruhen auf spezifischen „schwachen Bindungen“ und schaffen einen konstanten Abstand zwischen den beiden Ketten. Zusammen mit schwachen Bindungen zwischen aufeinander folgenden Basenpaaren sorgen sie für eine ausreichend stabile Paarung der Ketten im normalen Temperaturbereich. Diese Bindungen zwischen den komplementären Ketten verhindern auch weitgehend die Bildung dreidimensionaler Faltungen innerhalb der Einzelketten. Sie stabilisieren und sie verstecken die Sequenzinformation bei der Lagerung. Für die Informationsübertragung müssen die beiden Ketten voneinander getrennt werden. Eine Nukleotidkette kann dann als Matrize für die Anlagerung von passenden Nukleotiden bei der Synthese der komplementären Kette dienen. Bei der Replikation der DNA geschieht das gleichzeitig an beiden Einzelketten, die jeweils zu Doppelmolekülen ergänzt werden (Meselson & Stahl, 1958). Auch wenn für diesen Replikationsprozess eine ganze Reihe verschiedener Proteine als Apparatur und Werkzeuge notwendig sind, besteht die Informationsübertragung wirklich nur in der spezifischen Basenpaarung zwischen komplementären Nukleotiden. Die neusynthetisierten Doppelketten werden von entsprechenden Proteinen auf ihre Struktur hin kontrolliert. Das nennt man zwar „Korrekturlesen“, aber der Inhalt der Information wird nicht geprüft. Mutationen, die sich nicht auf die Helixstruktur auswirken, schaffen neue Allele. Wegen dieser rein formalen Kontrolle können zum Beispiel auch informationstragende Sequenzen in die bestehenden DNA-Ketten eingespleißt werden (von Viren oder von Biochemikern), und diese zusätzlichen Sequenzen werden sauber mitkopiert und von der Zelle als genetische Information interpretiert. Das weitere Schicksal aller Änderungen der DNA-Sequenz des Genoms beruht ausschließlich auf ihrem Effekt auf das Überleben und die Reproduktion der Zelle oder des Organismus, also auf Darwinscher Selektion. Nirgendwo ist das Verhältnis zwischen richtungsloser struktureller Variation bei der Vermehrung von Informationsträgern und der Entstehung oder dem Verlust von sinnvoller Information derart gut zu verfolgen wie bei der Evolution der DNA von Genomen. Jetzt, da wir die vollständige

Basensequenz ganzer Genome kennen, sehen wir erst, wie sehr das Bild vom Genom als Perlenkette von Genen für Proteine noch unsere Vorstellung von der Natur als perfektem Uhrwerk aus vorgegebenen Elementen unter der Kontrolle von schematisch einfachen Gesetzen widerspiegelt. Was wir jetzt sehen, ist eine Dynamik mit Informationsübertragung, Variation und Selektion, die im Freiraum aller möglichen Anordnungen hin und wieder wirksame komplexe Strukturen schafft und dann oft und lange genug erkennbar reproduziert, so dass sie uns wie invariante Grundeinheiten der Natur vorkommen.

Das C-Wert-Paradox und die Evolution von Information

Sobald bekannt wurde, dass DNA der molekulare Träger der genetischen Information ist, begannen Biochemiker, die Gesamtmenge der DNA in den Genomen verschiedener Arten zu bestimmen. Sie erwarteten, dass die Menge der genetischen Information proportional zur Komplexität der jeweiligen Art sei, dass der DNA-Gehalt des Genoms also ein Maß für die Komplexität eines Organismus sein sollte. Alfred Mirsky und Hans Ris veröffentlichten 1951 die ersten vergleichenden Daten zu den Genomgrößen einer Reihe verschiedener Tierarten. Die DNA-Menge im haploiden Genom ist der *C-Wert*, entweder als Anzahl von Basenpaaren oder in Picogramm (10^{-12} g). Dieser C-Wert war, wie erwartet, für alle Individuen einer Art innerhalb der damaligen Messgenauigkeit der gleiche. Bei einigen Milliarden Basenpaaren in einem Genom konnten ein paar Millionen mehr oder weniger allerdings nicht erkannt werden. Das kam später (Greilhuber, 2005; Check, 2005). Das aufregende Ergebnis damals war, dass die C-Werte verschiedener Arten, also die Mengen an genetischem Material, bei Pflanzen und Tieren nichts mit der geschätzten Komplexität der Art zu tun haben. Je mehr Arten untersucht wurden, desto deutlicher wurde dieses Ergebnis. Gerade bei Wirbeltieren gab es da Überraschungen. Einige Salamander und Lungenfische haben 50-mal soviel DNA im Zellkern wie wir Menschen, und bei den Säugetieren hat das Erdferkel noch einmal zwei Drittel soviel wie wir. Die C-Werte von 19 Arten derselben Gattung von Kröten (*Bufo*) formen eine log-normale Verteilung zwischen 8 und 16 pg. Die C-Werte bei Pflanzen variieren über einen ähnlichen Bereich wie die der Tiere (Bennett et al., 1982). Dieses unvorhergesehene Resultat ist unter dem Namen *C-Wert-Paradox* (Thomas, 1971) in die Geschichte eingegangen.

Wir wissen heute recht genau, was zur Variation der C-Werte beiträgt (Gregory, 2005). Nirgendwo sonst haben wir einen besseren Einblick in

die laufende Produktion genetischer Komplexität. Das beginnt mit den verschiedenen Arten von Mutationen. Sie reichen von Austausch, Verlusten oder Einschüben einzelner Nukleotide in der Sequenz („Punktmutationen“) über lokale Verdoppelungen oder Verluste von Sequenzen verschiedener Länge bis hin zur Verdoppelung des gesamten Genoms, wenn an entscheidender Stelle einer Chromosomenverdoppelung keine Zellteilung folgt. Dazu kommen Austausche und Einwanderungen von Sequenzen fremder Genome über Parasiten, vor allem von einigen Arten Viren, die ihre Genome in das Genom der Wirtszelle einbauen können, und durch Hybridisierung zwischen Arten. Die verschiedenen Arten von Mutation sind bei verschiedenen Gruppen von Organismen verschieden häufig, und ihre Bedeutung hängt obendrein von der jeweiligen Populationsgröße ab. Man hat zum Beispiel die Häufigkeit von Sequenzmutationen beim Bakterium *E. coli* unter optimalen (nicht-mutagenen) Umständen auf einen Austausch aller 54 Milliarden kopierter Nukleotide geschätzt (Drake, 1999). Die vier-einhalb Millionen Basenpaare im Genom werden dann also in den meisten Fällen ohne einen einzigen Fehler kopiert. Trotz dieser anscheinend vernachlässigswerten Fehlerrate kann ein einziges Bakterium in ein paar Litern Nährlösung über Nacht so viele Nachkommen produzieren, dass man davon ausgehen kann, dass darin irgendwo jedes beliebige Basenpaar mutiert ist. Dann kommt Selektion ins Spiel. Die Vermehrungsrate nimmt zwar ab, wenn die Umstände sich verschlechtern, aber die Vermehrung darf nie aufhören, und letztlich müssen Teilungsrate und Absterberate ins Gleichgewicht kommen, wenn sich die Population im Rahmen der vorhandenen Ressourcen halten will. Das führt unweigerlich zu einer (unbewussten, automatischen) Konkurrenz zwischen den Zellen, welche vor der nächsten Teilung absterben, und denen, welche sich noch einmal teilen. Dabei haben die Zellen mit neuen Mutationen in der Regel die schlechteren Chancen. Selektion bewahrt vor allem das bewährte Erreichte vor statistischem Zerfall. Positive Evolution ist ein gelegentlicher Nebeneffekt, auch wenn sie auf lange Zeit Erstaunliches erreicht. Die präzise Replikation der DNA unter normalen Umständen ist eines der Ergebnisse davon. Die Population wird nur dann nicht aussterben, wenn laufend ein Überschuss an funktionierenden Zellen produziert wird, unter denen durch Konkurrenz um die begrenzten Ressourcen die effektivsten auch die häufigsten bleiben. Unter normalen Umständen ist eine minimale Mutationsrate die beste Strategie für das Überleben einer Population.

Man sollte meinen, dass dann die Evolution irgendwann bei einem optimalen Genom festläuft, das durch keine Mutation verbessert werden kann und seine Mutationsrate gerade durch eine entsprechend effektive Vermeh-

rungsrate ausgleicht. Die lokal und zeitlich verschiedenen Umweltansprüche verhindern das, aber alle *typischen* oder *regelmäßigen* Strukturen und Prozesse in der Biologie sind evolutionäre Erfindungen, die zwar auch immer Variationen unterworfen sind, aber zumindest lokale Optima darstellen, bei denen Variation keine grundsätzlich neuen Trends in Gang setzen kann. Die Zellstruktur von Bakterien und Archäen ist ein gutes Beispiel. Bakterien und Archäen sind zwei uralte Gruppen von einzelligen Organismen, bei denen das Genom in der Regel ein einziges, ringförmig geschlossenes DNA-Molekül ist, das wirklich einigermaßen einer Perlenkette von Genen gleicht. Weil bei ihnen die DNA innerhalb der Zelle nicht in einem separaten Zellkern eingeschlossen ist, werden Bakterien und Archäen trotz ihrer unterschiedlichen Zellchemie oft als *Prokaryoten* dem dritten Typ von heute lebenden Zellen, den *Eukaryoten* (Amöben, Pilze und Tiere, Pflanzen und einige andere Gruppen), gegenübergestellt. Prokaryoten sind seit Milliarden von Jahren erfolgreich. Sie kommen überall vor. Wir brauchen nur daran zu denken, dass jeder von uns bis zu einhundert Billionen (*zehn hoch vierzehn*) Prokaryotenzellen von Hunderten verschiedenen Arten beherbergt (Gill et al., 2006), weit mehr, als unser Körper eigene Zellen hat. Aber die Bauprinzipien von Prokaryotenzellen setzen ihrer Evolution enge Grenzen. Ihre Genome enthalten meist einige Millionen Nukleotidpaare. Die größten, mit etwa 13 Millionen, sind etwa so groß wie die kleinsten Eukaryotengenome. Mindestens zwei Gruppen von Bakterien zeigen unabhängig voneinander Ansätze zur Vielzelligkeit, aber in Milliarden Jahren sind keine wirklich vielzelligen Prokaryoten entstanden. Der Erfolg der Prokaryoten beruht darauf, dass sie für jede erreichbare Konstellation von Umweltbedingungen auf oder unter der Erde spezielle physiologische Adaptationen gefunden haben, darunter für Bedingungen, in denen jeder Eukaryot absterben würde. Grob vereinfacht ist die Evolution von Prokaryoten durch Zunahme an physiologischer Diversität charakterisiert, die der Eukaryoten durch Zunahme an organischer Komplexität. Einzellige *Eukaryoten* haben immer wieder unabhängig voneinander echte Vielzelligkeit entdeckt und sind zu Ahnengruppen von großen vielzelligen Pflanzen, von Tieren, Pilzen, Rotalgen oder Braunalgen geworden.

*Vielzelligkeit, Sex und unerkannt mitgeschleppte Kopierfehler
als Quelle neuer Information*

Die Struktur der eukaryotischen Zellen mit ihren linearen Chromosomen und ihren Membransystemen war eine Voraussetzung für die Evolution der

komplexen genetischen Regulationsmechanismen, wie sie Vielzelligkeit verlangt. Der entscheidende Fortschritt war aber die Erfindung der Sexualität, also der Verschmelzung von zwei Einzellern mit je einem Genom (haploid) zu einer Zelle mit zwei Genomen (diploid), und der komplementären Reduktionsteilung, vor der die jeweils homologen Chromosomen der beiden Genome sich präzise der Länge nach nebeneinander legen und reziprok Stücke austauschen. Entstanden ist dieser seltsame Mechanismus sicher als eine Antwort auf mutagene Stresssituationen. Bei Prokaryoten gibt es eine ganze Reihe vergleichbarer Prozesse, aber eben nicht diesen. Vollständiger reziproker Stückaustausch zwischen Genomen kann verschiedene Mutationen in zwei gepaarten Chromosomen auf einem Chromosom zusammenbringen und die entsprechenden fehlerfreien Stellen auf dem anderen. Dieser Mechanismus kann aber auch unabhängige nützliche Mutationen aus zwei verschiedenen Abstammungslinien zusammenbringen und mit dieser Rekombination von unabhängigen Erfindungen die Evolution enorm beschleunigen. Das ist einer der großen Vorteile von Sex. Er wird dadurch möglich, dass in diploiden Zellen ein funktionierendes Allel in einem Genom ein funktionsloses Allel desselben Gens im anderen Genom oft voll kompensiert und das mutierte Allel nur dann schädlich wirkt, wenn es von den Keimzellen beider Eltern übertragen wird. Das ist der Grund für die Vermeidung von Inzucht und ein Mechanismus, der es erlaubt, neue Allele lange in einer Population mitzuschleppen und in verschiedenen Kombinationen auf ihre Wirkung zu testen. Sex ist früh in der Evolution einzelliger Eukaryoten entstanden und hat sich als eine *Major Evolutionary Transition* erwiesen, die neue Möglichkeiten der Evolution eröffnet hat, deren Resultate dann aber auch eine Rückkehr zu einer ausschließlich asexuellen Vermehrung durch Teilung beinahe unmöglich gemacht haben. Sexuelle Vermehrung ist uns derart selbstverständlich, dass man erst in den siebziger Jahren ernstlich untersucht hat, warum sich asexuelle Vermehrung mit Weitergabe ganzer Genome nicht durchgesetzt hat. Die Tendenz dazu bleibt bestehen, und viele Organismen wechseln zwischen sexueller und asexueller Vermehrung.

Echte Vielzelligkeit beruht darauf, dass die klonalen, genetisch identischen Nachkommen einer Zelle verbunden bleiben und durch Integration und Arbeitsteilung einen komplexen Organismus bilden, in dem die einzelnen Zellen schließlich nur noch Bauelemente sind. Aus komplexen Systemen (Zellen) werden Komponenten eines Systems mit höherer Komplexität (vielzellige Organismen). Das verlangt zusätzliche Information. Zu den Genen für die Grundfunktionen einer Zelle kommt ein genetisches Programm für die fortschreitende funktionelle und morphologische Diffe-

renzierung der Zellen im Verlauf ihrer Teilungen, bei der zellinterne Prozesse und Kommunikation zwischen den Zellen zusammenspielen, und es verlangt eine organismusweite Kommunikation zwischen Zellen für die integrierte Funktion des gesamten Organismus. Dieses zusätzliche Programm kann nicht aus der Änderung von Allelhäufigkeiten in vorhandenen Genen entstehen, es benötigt neue Gene und neue Regulationsmechanismen, also größere Genome. Die wahrscheinlichste Quelle für mehr DNA sind zufällige Duplikationen im Genom. Die konnten schon ohne fortgeschrittene Sequenzierungsmethoden nachgewiesen werden. Direkte Wiederholungen von einzelnen Nukleotiden, Gruppen von Nukleotiden und längeren Stücken der DNA schienen überall in Genomen vorzuliegen, ohne dass ihnen eine Funktion zugewiesen werden konnte. Ob solche Wiederholungen (neuartige) funktionelle Information enthalten und woher sie die bekommen haben, war anfangs unbekannt. Zwei Modelle aus der Zeit um 1970 haben sich bei der Antwort als wegweisend erwiesen.

Roy Britten und Eric Davidson (1969) nahmen an, dass die wiederholten Sequenzen, die anscheinend nicht in RNA und sicher nicht in Proteine umgeschrieben wurden, regulierende Funktion haben sollten. Vor Gene geschaltet, sollten sie den entsprechenden Enzymen zeigen, wann und wo das Produkt eines Gens benötigt wird. Auch wenn gerade die meisten der Sequenzen, von denen Britten und Davidson damals ausgingen, nicht dazu dienen, ist deren Modell eines Regulationscodes in der DNA inzwischen zu einem der aktivsten Gebiete der Molekulargenetik geworden. Dabei geht es um einen Code, der auf der Bindung zwischen Proteinen und (doppelsträngiger) DNA beruht. Dass die DNA über Proteine, die sie codiert und die an DNA binden, sich selbst reguliert, war schon vor der Entdeckung des Tripletcodes für die Proteinsynthese bekannt. Die Entdeckung von Francois Jacob und Jacques Monod (1961), wie das Darmbakterium *E. coli* bei Glucosemangel seinen Zuckerstoffwechsel umstellt, ist das Lehrbuchbeispiel. Vielzellige Eukaryoten haben diesen Mechanismus der Verbindung von Umweltinformation und innerer Regulation der Zelle zu Netzwerken von Kommunikation zwischen Zellen ausgebaut. Und das ist nur einer der intrazellulären Regulationscodes von Vielzellern. Im Gegensatz zum klassischen „genetischen Code“ ist dies übrigens kein universaler Code, sondern seine Evolution in verschiedenen Gruppen ist ein wichtiger Faktor der Makroevolution.

Komplementär dazu ging es Susumu Ohno um den Ursprung neuer Gene. In seinem Buch *Evolution by Gene Duplication* (1970) zeigte er, dass im Genom gelegentlich ganz ähnliche Gene wiederholt vorkommen, manchmal direkt hintereinander *in tandem*. Die zufällig komplette Verdoppelung

eines Gens würde redundante Information produzieren. Mutationen in einer der Kopien könnten die Funktion dieser Kopie modifizieren und im Laufe der Evolution daraus eine Variante unter eigener Regulation machen (Hittinger & Carroll, 2007). Wir wissen jetzt, dass auf diese Weise ganze Familien von Genen für ähnliche Proteine innerhalb eines Genoms entstanden sind. Das Paradebeispiel waren damals die Gene für Hämoglobin. Ein funktionierendes Hämoglobin besteht aus vier Proteinketten, zwei identischen Produkten eines α -Hämoglobingens und zwei von einem β -Hämoglobingens. In dieser Zusammenstellung können die Moleküle sich so beeinflussen, dass sie bei einer ganz bestimmten Sauerstoffkonzentration von Bindung zu Abgabe übergehen. So wird Aufnahme von Sauerstoff in den Lungen und Abgabe in den Geweben programmiert. Im menschlichen Genom liegen aber vier funktionelle Gene vom α -Typ *in tandem* auf Chromosom 16 und fünf vom β -Typ *in tandem* auf Chromosom 11. Bei der Entwicklung vom Embryo über den Fötus zum Erwachsenen werden Hämoglobine von fortschreitend anderen dieser Gene synthetisiert. Damit wird Aufnahme und Abgabe von Sauerstoff im Embryo, im Fötus und im Erwachsenen den jeweiligen Umgebungen angepasst. Bei Fröschen hilft ein ähnlicher Mechanismus beim Übergang vom Wasserleben der Kaulquappe zum Landleben der Frösche.

Das war ein Beispiel für die subtilen Möglichkeiten zur Evolution über die zufällige Duplikation von Genen. Ohno wies aber auch darauf hin, dass eine zufällige Genverdoppelung keineswegs automatisch neue Funktionen schaffen muss. Es ist viel wahrscheinlicher, dass eine Mutation die Funktion einer redundanten Kopie zerstört, ohne dass das einen Effekt auf den Organismus hat. So eine funktionslose Kopie würde im Laufe der Zeit beliebige Mutationen ansammeln und unweigerlich im Verlauf von etwa vier Millionen Jahren regelmäßiger Weitergabe zu Buchstabensalat zerfallen. Und wirklich liegen zwischen den vier menschlichen α -Genen drei funktionslose Kopien und zwischen den fünf β -Genen zwei solche *Pseudogene*. Dies sind noch erkennbare, also nur die jüngsten abortiven Verdoppelungen in der langen Geschichte der Hämoglobine. Ohno (1972) nannte solche Sequenzen, die der Selektion entgehen und die langsam zur Unkenntlichkeit mutieren, *Junk DNA*, molekularen Müll. Sein Argument war stichhaltig und mit Beispielen belegt, aber der Gedanke, dass eventuell ein großer Teil des Genoms, unseres Genoms, Müll sein sollte, war damals vielen Biologen (ich gehörte dazu) unangenehm. Darwinsche Evolution beginnt zwar mit zufälliger Variation, und der größte Teil dieser Variation ist sicher nutzlos oder schädlich, aber wir nahmen an, dass rigorose Selektion nicht nur Schädliches, sondern auch Überflüssiges entferne und dass alle Details eines

Organismus jederzeit optimal zur Funktion des Organismus beitragen. Das ist bei allem Staunen über die funktionelle Komplexität von Organismen ein Vorurteil. Für die genetische Information gilt es erst recht nicht.

Die Ansammlung von Informationsmüll im Genom könnte zwar das Fehlen einer Korrelation zwischen Genomgröße und organismischer Komplexität erklären, aber solche Mengen von überflüssiger DNA in jeder Zelle sollten auch ohne codierende Funktion Effekte auf Überleben und Fortpflanzung haben und damit eine Gegenselektion hervorrufen. Vor der Erkenntnis, dass Konkurrenz zwischen Organismen letztlich Konkurrenz zwischen alternativer Information ist, wurde damals Effizienz der Organismen im Umgang mit Energie als mögliche Grundwährung bei der Konkurrenz zwischen Organismen angesehen. Da fiel es auf, dass dieselben Nukleotide, aus denen Nukleinsäuren bestehen, auch die zentrale Rolle bei der Speicherung und Übertragung von Energie im Stoffwechsel spielen. Das illustriert, wie weit die Evolution einmal Vorhandenes für alle möglichen Anwendungen verwendet, aber es machte die Ansammlung und Weitergabe von molekularem Müll zu einer unerklärlichen Energieverschwendung. Und doch gab es ohne jeden Zweifel überflüssige DNA. Wenn man Sichtweisen in der Naturwissenschaft danach beurteilt, welche von ihnen die Beobachtungen einfacher erklärt und voraussagt, wird man hier wieder die Asymmetrie zwischen der genetischen Information und dem Organismus erkennen. Trotz ihrer notwendigen gegenseitigen Abhängigkeit ist es letztlich die Information, die sich den Organismus als Reproduktionsapparat, zum Überleben und zur Abschirmung von der Umwelt schafft, nicht der Organismus, der die Information zu seiner Vermehrung einsetzt. Information bleibt über die endliche Lebensdauer von Informationsträgern hinaus bestehen. Organismen mit sexueller Fortpflanzung können „sich“ grundsätzlich nicht vermehren. Der Informationsfluss in der Zelle verläuft von Nukleinsäuren zu Proteinen, aber nicht umgekehrt. Nur ein kleiner Teil der DNA im Genom eines komplexen Vielzellers codiert für Proteine. Wir wissen noch nicht, wieviel der übrigen DNA im Genom Teil des „Entwicklungsprogramms“ für den Organismus ist, aber Genome enthalten auf jeden Fall sehr viel DNA, die wegfällen kann, ohne dass es einen merklichen Effekt auf den Organismus hat (Nóbrega et al., 2004).

Das Genom als Gesellschaft von Genen. Parasitische DNA

Darwinsche Selektion beruht auf der Konkurrenz alternativer Formen der gleichen Information um deren Weitergabe (Allele von Genen oder, allge-

mein, „alternative Antworten auf dieselbe Frage“). Das ist Konkurrenz zwischen replizierenden Informationsträgern um Bausteine und Energie zur Synthese von Kopien. Aber neben dieser Darwinschen Konkurrenz zwischen replikationsfähigen Einheiten gibt es die Konkurrenz von Information, die sich nicht selbst replizieren kann, um Replikationsmechanismen. Gene können sich nicht allein replizieren. Sie sind auf die Zusammenarbeit im Genom bei der Programmierung eines Apparates für ihre Replikation angewiesen. Zum Bild vom Genom als Abfolge einer bestimmten Anzahl Gene an festen Plätzen, deren Allele um diesen Platz konkurrieren, kommt die Rangelei von neuer Information um einen dauerhaften Platz in einem Genom. Für Duplikationen bestehender Information hat Ohno das erläutert. Ein neuer Platz im Genom bleibt erhalten, wenn dort eine Funktion lokalisiert ist, die zum Überleben und zur Replikation des gesamten Genoms beiträgt. Kein Wunder, dass Gene mit zentralen Rollen bei der DNA-Replikation oder der Proteinsynthese seit Milliarden Jahren fester und erkennbarer Bestandteil aller Genome sind. Spezialfunktionen haben es nicht so leicht, und funktionslose DNA kann nicht nur unkontrolliert degenerieren, sondern auch ohne Effekt verlorengehen.

Was überraschte, war die Erkenntnis, dass ein großer, wenn nicht gar der größte Teil der DNA eines Vielzellergenoms aus parasitischer DNA besteht, die ohne Gegenleistung vom Replikationsmechanismus der Zelle profitiert. Parasitische DNA verschiedener Art gibt es auch in Prokaryoten, genauso wie regulatorische DNA, Genduplikationen und Pseudogene, aber bei Prokaryoten reduziert Selektion die Genomgröße rigoros auf ein Minimum, und Sequenzen, die nicht für Überleben und Fortpflanzung notwendig sind, haben dort keine lange Verweildauer. Die Toleranz von eukaryotischen Zellen für große Mengen DNA im Genom, auf der die Evolution komplexer genetischer Regulationssysteme beruht, macht diese Zellen auch relativ tolerant für parasitische DNA, besonders in ihrer inaktivierten, funktionslosen Form als molekularer Müll. Anstatt sich einen Platz im Genom durch einen Beitrag zur Funktion des Organismus zu sichern, entgeht solche DNA dadurch der Selektion, dass sie sich an verschiedenen Stellen in mehrfachen Kopien ins Genom einbaut und sich oder eine redundante Kopie von sich von einer Stelle im Genom in eine andere umsetzt. Daher der Name *Transposons* oder *Springende Gene*. Im menschlichen Genom liegen mehr als drei Millionen Kopien verschiedener Arten von Transposons, beinahe alles die Überbleibsel von alten Episoden explosionsartiger Verbreitung im Genom, die von der Zelle zwar nicht gezielt ausgeschnitten, aber in ihrer Funktion blockiert werden.

Genome wachsen, bis die physiologischen Effekte adaptiver Evolution im Wege stehen

Die große Menge von nicht notwendiger und eventuell schädlicher DNA erklärt sich dadurch, dass eine beliebige Verdoppelung von DNA im Genom meist weniger nachteilig ist als ein beliebiger Verlust von DNA, vor allem in kompakten Genomen. Genome haben daher die Tendenz, langsam anzuwachsen. Auch wenn eukaryotische Zellen im Gegensatz zu Prokaryoten erstaunlich tolerant für zusätzliche DNA sind und wenn die Lieferung von energiereichen Nukleotiden für die Replikation des jeweiligen Genoms eine vorrangige Aufgabe des Zellstoffwechsels ist, macht sich die Menge der DNA irgendwann doch in der Struktur und der Funktion des Organismus bemerkbar. Am einfachsten ist das zu sehen, wenn man (durch Unterdrückung einer Zellteilung) einen Organismus mit drei oder vier Genomen pro Zelle statt der normalen zwei erhält (Triploide oder Tetraploide statt Diploide), ohne dass man dabei die Information in diesen Genomen verändert. In den meisten Fällen wird das Zellvolumen etwa proportional größer, die Zellen teilen sich etwas langsamer und haben pro Gewichtseinheit eine geringere Stoffwechselaktivität. Bei krautigen Pflanzen ändert sich mit der Zellgröße die Größe des Organismus. Praktisch alle großen und großblütigen gärtnerischen Zuchtvarianten sind künstliche *polyploide* Derivate von Wildpflanzen. Das (unabhängig entstandene) Entwicklungsprogramm von Tieren strebt in der Regel eine vorgegebene Endgröße an. Die meisten tetraploiden Tiere haben die halbe Zahl doppelt so großer Zellen. Das kann nachteilige Effekte haben, wenn zum Beispiel doppelt so große rote Blutkörperchen durch Kapillaren fließen, deren dickere Wandzellen das Lumen einengen. Die Effekte einer Reduzierung der Zahl der Nervenzellen im Gehirn sind bisher kaum untersucht worden. Ein plötzlicher Übergang von normaler Diploidie zu Tetraploidie kann aber auch andere überraschende Effekte haben. Ein berühmtes Beispiel aus der Natur sind zwei nordamerikanische Laubfroscharten, die auch ein Experte nicht am Aussehen unterscheiden kann, die diploide *Hyla chrysoscelis* und die tetraploide *Hyla versicolor*. Die Frequenz der Rufe der Männchen ist bei der tetraploiden Art halb so schnell wie bei der diploiden Art (McLister et al., 1994), und die jeweiligen Weibchen scheinen bevorzugt auf die Rufe der passenden Männchen zu reagieren.

Auch die kontinuierliche Variation von Genomgrößen zwischen verwandten Arten ist mehr oder weniger genau mit Zellgrößen und anderen Parametern korreliert. Zum Beispiel sind Pflanzen ab einer bestimmten Genomgröße obligat mehrjährig, und Wirbeltiere, die fliegen können (Fleder-

mäuse, flugfähige Vögel, gerade auch Kolibris), haben kleinere Genome als flugunfähige Gruppen. Vor allem Selektion auf einen Lebensstil mit überdurchschnittlich hoher Stoffwechselaktivität oder einer besonders schnellen Generationenfolge bevorzugt Genome, in denen möglichst viel entbehrliche DNA verlorengegangen ist. Nicht zufällig sind die Lieblingsobjekte von Genetikern Organismen mit mehreren Generationen pro Jahr und kleinen Genomen. Dazu gehörten die Taufliege *Drosophila melanogaster* und der Kreuzblütler *Arabidopsis thaliana*. Diese kleinen und gut charakterisierten Genome sind auch unter den ersten voll sequenzierten gewesen. Dabei hat sich unter anderem gezeigt, dass das Genom von *A. thaliana* im Laufe der Evolution der zweikeimblättrigen Pflanzen ein verdreifachtes Ahnengenom ererbt und dieses bei seiner weiteren Evolution noch zweimal verdoppelt hat, so dass das jetzige winzige Genom aus den Resten von zwölf Kopien verschiedenen Alters des Ahnengenoms der höheren Pflanzen zusammengestückt ist (Tang et al., 2008). Übrigens zeigt auch das menschliche Genom noch deutliche Spuren von zwei vollen Verdoppelungen in der frühen Evolution der Wirbeltiere, eine vor und eine nach der Abspaltung der Vorfahren der Neunaugen und der Schleimfische, deren Genom nur auf zwei Kopien des Urgenoms der Chordatiere zurückgeht. Es ist erstaunlich, wie viele Details der Entstehungsgeschichte eines Genoms sich aus dem Textvergleich der Basenfolgen mehr oder weniger nahe verwandter Arten rekonstruieren lassen. Im Grunde genommen ist der Evolutionsprozess sehr viel konservativer, als wir einmal angenommen haben. Beinahe alle Neuerungen sind Umbau, Kombination und Wiederverwendung von Vorhandenem. Von all den unendlich vielen Mutationen aller Art, die jeden Tag überall neu entstehen, halten sich auf Dauer nur die allerwenigsten. Die Fähigkeit diploider Eukaryoten mit sexueller Fortpflanzung, eine Menge von Allelen und eine Menge von nichtkodierender DNA weiter zu reproduzieren, auch wenn sie keinen direkten Vorteil bringen, und ihre Fähigkeit, verschiedene Neuerungen aus verschiedenen Abstammungslinien innerhalb einer Art oder von eng verwandten Arten zusammenzubringen und neu zu kombinieren, ist die Grundlage einer genetischen Komplexität, deren Umfang wir immer noch nicht voll erfasst haben. Der automatische Opportunismus der Selektion, die Nützliches erhält, woher es auch kommen mag, zeigt sich auch darin, dass immer mehr funktionale Neuerungen entdeckt werden, die aus Teilen parasitischer DNA oder aus dem Müll funktionsloser Duplikationen entstanden sind (Biemont & Vieira, 2006; Xiao et al., 2008).

*Information ist Einschränkung von Freiheit:
Konkurrenz und Kooperation von Informationsträgern*

Darwinsche Selektion beruht auf der Konkurrenz der Träger *verschiedener* erblicher Eigenschaften um die gleichen Ressourcen. Solange Informationsträger dieselbe Information übermitteln, sind sie für diese Information beliebig austauschbar. Welche von ihnen die Information weitergeben und welche vorher absterben, ist dann ohne Konkurrenz und ohne Selektion dem Zufall überlassen. Träger derselben Information können zusammenarbeiten und dabei zugunsten von anderen Individuen, altruistisch, auf Fortpflanzung und Leben verzichten, wenn dieses Opfer die Persistenz der gemeinsamen Information fördert. Das ist das Geheimnis komplexer Vielzelligkeit: sie ist der Zusammenschluss von Zellen, die durch wiederholte Teilung, also klonale, asexuelle Fortpflanzung entstanden sind und Kopien desselben Genoms enthalten. Das Entwicklungsprogramm läuft in einer Art Metagenom über die Kommunikation zwischen Kopien dieses Genoms ab. Mit ganz wenigen, genau kontrollierten Ausnahmen behalten aber alle Zellen die volle genetische Information. Wenn Mutationen die Kopie des Genoms in einer der Zellen verändern, kommt es leicht zum Ausbruch dieser Zelle und ihrer Nachkommen aus der bedingungslosen Unterordnung der Körperzellen unter die klonale Gemeinschaft, also zu Krebserkrankungen.

Kein anderer Zusammenschluss von Trägern gemeinsamer Information ist derart bedingungslos, weil in keinem anderen Fall die Information aller Teilnehmer völlig identisch ist. Dann motiviert gemeinsame identische Information zur Zusammenarbeit, und gleichzeitig motivieren unterschiedliche Versionen zur Konkurrenz. Diese so einfache und so grundlegende Erkenntnis ist erst durch Veröffentlichungen von William Hamilton (1964) in ihrer vollen Bedeutung erkannt worden. Gerade sexuelle Fortpflanzung sorgt ja dafür, dass die unterschiedliche Information der Eltern bei der Keimzellenbildung durch den obligatorischen reziproken Austausch von Stücken zwischen den beiden Kopien jedes Chromosoms neu zusammengestellt, *rekombiniert* wird. Im Laufe des 20. Jahrhunderts ist schrittweise deutlich geworden, dass auch bei uns Menschen kein Genom dem anderen gleicht. Jeder ist genetisch einmalig und damit ein biologischer Konkurrent von jedem anderen. Auf der Ebene von Organismen ist das ein Kampf aller gegen alle. Die einmalige individuelle Allelzusammenstellung schließt aber gleichzeitig aus, dass es dabei um die Persistenz des eigenen Genoms geht. Sie endet ja mit dem Tod des Organismus. Ein Organismus mit sexueller Fortpflanzung konkurriert also mit anderen um die Weitergabe der Allele,

die er zufällig bei der Befruchtung mitbekommen hat. Von denen sind bei zwei Individuen viele identisch und viele verschieden. Die verschiedenen motivieren zur Konkurrenz, die gemeinsamen zur Kooperation. Jedes Gen teilt durch seine Allele die Population anders in konkurrierende Parteien auf. Der Grad von spontaner Kooperation und Konkurrenz im Verhalten zwischen Individuen einer Art lässt sich oft erstaunlich genau aus dem relativen Anteil identischer und verschiedener Allele voraussagen, wenigstens insoweit die Individuen diese Anteile in anderen erkennen oder abschätzen können. Als Schreckbild vom Organismus als Marionette an den Fäden seiner Gene hat diese Interpretation eine heftige Debatte um den Determinismus unseres Verhaltens und unseren freien Willen ausgelöst. Dabei ist meist übersehen worden, dass verschiedene Gene einander widersprechende Strategien verfolgen, dass aber alle diese Gene nur in ihrer jeweiligen Zusammenarbeit einen Organismus lenken können. Verhaltensweisen werden von einem genetischen Programm gesteuert, bei dem die verschiedenen Interessen der Komponenten nach historisch gewachsenen Regeln gegeneinander verrechnet werden. Die (hormonelle, neuronale) steuernde Information, die schließlich dabei in einem Organismus entsteht, und erst recht die Information, die dabei zwischen Organismen ausgetauscht wird, sind aus einer sehr komplexen Interaktion der Gene abgeleitete Produkte. Der Organismus mag von seinen Genen gesteuert werden, aber die gegensätzlichen Interessen der verschiedenen Gene in jeder Situation wirken gerade bei komplexen Organismen bis in das Verhalten durch. Widerstreitende Motivationen sind eine alltägliche Erfahrung im menschlichen Verhalten. Sie sind die Grundlage unserer Entscheidungsfreiheit und der damit verbundenen Pflicht, Entscheidungen zu treffen. Vielleicht beruht unsere einmalige Fähigkeit, die Widersprüchlichkeit unserer verschiedenen Motivationen zu erkennen, überhaupt zu verstehen, dass es dabei um die Konkurrenz von Information geht, und gelegentlich bewusst und rational einzugreifen, darauf, dass unser unbewusst programmiertes Entscheidungssystem gerade wegen seiner Fähigkeit, viele Faktoren zu berücksichtigen, immer wieder in multiplen Möglichkeiten festläuft. Das Problem wird dann an das langsame und umständliche bewusste und rationale System verwiesen. Nicht, dass das eine völlig unabhängige und unbeeinflussbare Instanz ist. Ethische Prinzipien sollen dann als erprobte Gewichtungsfaktoren Entscheidungen im Interesse der Gemeinschaft beeinflussen. Das Ziel dabei ist, die chaotischen Konsequenzen von selbstregulierenden Zufallsentscheidungen durch weitere (vernünftige, ethische) Information einzuschränken oder zu verhindern, wo multiple Entscheidungen möglich und etwa gleich wahrscheinlich sind, aber verschiedene Konsequenzen haben.

Die Charakterisierung der Information als ordnendes Prinzip dort, wo Naturgesetze Wahlfreiheit zulassen, wirft neue Fragen auf, aber sie hat sich als bahnbrechendes heuristisches Prinzip zum Verständnis der Evolution von organischer Komplexität erwiesen.

Literatur

- Anfinsen, C. B., E. Haber, M. Sela & F. H. White Jr.. 1961. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the USA* 47: 1309–1314.
- Avery, O. T., C. M. McCleod & M. McCarty. 1944. Studies on the chemical nature of the substance inducing transformation in pneumococcal types. *Journal of Experimental Medicine* 79: 137–158.
- Beadle, G. W. & E. L. Tatum. 1941. Genetic control of biochemical reactions in *Neurospora*. *Proceedings of the National Academy of Sciences of the USA* 27: 499–506.
- Bennett, M. D., J. B. Smith & J. S. Heslop-Harrison 1982. Nuclear DNA amounts in angiosperms. *Proceedings of the Royal Society of London. B* 216: 179–199.
- Biémont, C. & C. Vieira. 2006. Junk DNA as an evolutionary force. *Nature* 443: 521–524.
- Britten, R. J. & E. H. Davidson 1969. Gene regulation for higher cells. A theory. *Science* 165: 349–358.
- Check, E. 2005. Patchwork people. (News Feature). *Nature* 437:1084–1086.
- Darwin, C. R. 1868. *The Variation of Animals and Plants under Domestication*. 2 vols. London: John Murray.
- Drake, J. W. 1999. The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. *Annals of The New York Academy of Sciences* 18: 100–107.
- Gill, S. R., M. Pop, R. T. DeBoy, P. B. Eckburg, P. J. Turnbaugh, B. S. Samuel, J. I. Gordon, D. A. Relman, C. M. Fraser-Liggett & K. E. Nelson. 2006. Metagenomic analysis of the human gut biome. *Science* 312: 1355–1359.
- Gregory, T. R. (ed.) 2005. *The Evolution of the Genome*. Amsterdam: Elsevier, Academic Press.
- Greilhuber, J. 2005. Intraspecific variation in genome size in angiosperms: identifying its existence. *Annals of Botany* 95: 91–98.
- Hamilton, W. D. 1964. The genetical theory of social behavior. *Journal of Theoretical Biology* 7: 16–32.
- Hershey, A. D. & D. M. Chase. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* 36: 39–56.
- Hittinger, C. T. & S. B. Carroll. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449: 677–681.
- Jacob, F. & J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3: 318–356.
- McLister, D., E. D. Stevens & J.P. Bogart. 1994. Comparative contractile dynamics of calling and locomotor muscles in three hylid frogs. *Journal of Heredity* 85: 322–325.
- Meselson, M. & F. W. Stahl. 1958. The replication of DNA in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the USA* 44: 671–682.

- Mirsky, A. E. & H. Ris. 1951. The DNA content of animal cells and its evolutionary significance. *Journal of General Physiology* 34: 451–462.
- Nóbrega, M. A., Y. Zhu, I. Plajzer-Frick, V. Afzal & E. M. Rubin. 2004. Megabase deletions of gene deserts result in viable mice. *Nature* 431: 988–993.
- Ohno, S. 1970. *Evolution by Gene Duplication*. Heidelberg: Springer Verlag.
- Ohno, S. 1972. So much junk DNA in our genomes. *Brookhaven Symposium in Biology* 23: 366–370.
- Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27, 379–423 and 623–656.
- Shannon, C. E. & W. Weaver 1949. *The Mathematical Theory of Communication*. Chicago IL: University of Illinois Press.
- Tang, H., J. E. Bowers, X. Wang, R. Ming, M. Alam & A. H. Paterson. 2008. Synteny and collinearity in plant genomes. *Science* 320: 486–488.
- Thomas, C. A., 1971. The genetic organisation of chromosomes. *Annual Review of Genetics* 5: 237–256.
- Tribus, M. & E. C. McIrvine. 1971. Energy and information. *Scientific American* 224 (September): 179–188.
- Watson, J. D. and F. H. C. Crick. 1953. Molecular structure of nucleic acids. A structure for desoxyribose nucleic acid. *Nature* 171: 737–738.
- Xiao, H., N. Jiang, E. Schaffner, E. J. Stockinger & E. van der Knaap. 2008. A retrotransposon-mediated gene duplication underlies morphological variation in tomato fruit. *Science* 319: 1527–1530.