

Evolution gesehen durch die Brillen der Physiker und der Biologen

Peter Schuster, Institut für Theoretische Chemie der Universität Wien

Die Entwicklung der Naturwissenschaften wurde Jahrhunderte lang durch die Physik und ihr Verhältnis zur Mathematik bestimmt. In gleicher Weise kann man zu Recht behaupten, dass die heutige Mathematik ohne Physik nicht das wäre, was sie ist. Völlig anders ist die gegenseitige Beziehung zwischen der erst etwa zweihundert Jahre alten Biologie zur Mathematik: Die konventionelle Theorie der Evolution wurde weitestgehend ohne eine einzige mathematische Formel entwickelt, obwohl sie auch schon zu Zeiten Darwins unschwer quantitativ formulierbar gewesen wäre. Physiker lieben allgemeine Konzepte, die in eine exakte Form gegossen werden können, wogegen konventionelle Biologen Details und Sammlungen hübscher Ausnahmen bevorzugen. Mit der Molekularbiologie nahmen Chemie, Physik und in ihrem Tross auch die Mathematik Einzug in die Biologie und auch in die Evolutionsforschung. Die heutigen Forschungs- und Analysetechniken stehen ohne massive Unterstützung durch Informatik und Mathematik auf verlorenem Posten. Im Referat wird eine mathematisierbare und mathematisierte Evolutionstheorie vorgestellt, die in ihrer Anwendung auf einfache Systeme präzise überprüfbare quantitative Aussagen machen kann. Insbesondere gelang durch die Konzeption von zellfreien, zur Evolution befähigten Experimentalsystemen der Brückenschlag zwischen chemischer Kinetik und Biologie. Wie durch ein paar Beispiele belegt wird, können die in den letzten Jahren entdeckten neuen Regulationsmechanismen der Zellentwicklung und Vererbung nur durch quantitative Modelle auf molekularer Basis verstanden werden. Ohne quantitative Konzepte und ausgefeilte computertechnische Methoden ist die heutige Datenflut in der Biologie nicht beherrschbar. Zur Bewältigung der anstehenden Probleme werden sowohl ‚Big-Data‘-Verfahren zur Erfassung, Speicherung und Abruf der ungeheuren Datenmengen als auch eine neue umfassende Theoretische Biologie zur sinnvollen Einordnung der Ergebnisse benötigt.

Der vorliegende Beitrag soll eine Art Plädoyer darstellen für eine neue theoretische Biologie, die aus den Fehlern der Vergangenheit lernt und sich den Grundfragen der gegenwärtigen Biologie stellt. Eine solche umfassende Disziplin muss auf drei Pfeilern errichtet werden: Experimente, welche bisher nie dagewesene Datenmengen erheben, rigorose Mathematik zur Definition und Analyse neuer Begriffe basierend auf Physik und Chemie als ordnende Grundlage und Computerunterstützung zur Modellierung und Datenverarbeitung. Nur so wird es möglich werden, die wichtigen Daten von den belanglosen zu trennen.

Prolog aus dem 17. und 18. Jahrhundert

Die Anfänge der neuzeitlichen Naturwissenschaft werden üblicherweise ins 17. Jahrhundert gelegt und Isaac Newtons Gravitationsgesetz wird zumeist als erster großer Meilenstein der heutigen Physik angesehen. Keine andere Entdeckung illustriert so prägnant die Rolle der Mathematik in der Physik zum Vorteil beider Disziplinen: Messbar sind Wegstücke, die in bestimmten Zeitabschnitten durchlaufen werden aber zur mathematischen Beschreibung der Bahnen bewegter Körper dienen glatte Kurven. Es musste ein Übergang von der diskreten Welt in das Kontinuum gefunden werden und dies gelang unabhängig voneinander Leibniz und Newton durch die Erfindung des Differentialquotienten:

$$\lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt} .$$

Die Physik hatte eine für die Mechanik unentbehrliche Methode zur Beschreibung der Bewegung auf Bahnen gewonnen, der Mathematik wurde von der Physik die neue Disziplin der Differentialrechnung geschenkt. Zu erwähnen ist auch Leonhard Euler, der unter anderem die heute gebräuchliche Symbolik in die Mechanik einführte.

Es bedarf vermutlich keines zweiten Beispiels zur Illustration der Beziehung zwischen Physik und Mathematik aber dennoch sei exemplarisch das deterministische Chaos zur Illustration des Fortbestandes dieses innigen Verhältnisses genannt. Henri Poincaré hatte schon zu Ende des 19. Jahrhunderts das Prinzip chaotischer Bahnen in einer Arbeit für einen von König Oskar II von Schweden und Norwegen ausgelobten Preis dargelegt. Deterministisches Chaos ist charakterisiert durch unregelmäßige Bahnen, die ungewöhnlich sensibel auf kleinste Änderungen der Anfangsbedingungen reagieren. Im Jahre 1963 veröffentlichte der Meteorologe Edward Lorenz ein einfaches dynamisches System, welches deterministisches Chaos zeigt, und begründete damit zusammen mit einigen anderen die Chaostheorie, einen neuen Zweig der Mathematik dynamischer Systeme.

Die im 18. Jahrhundert vorherrschende Einstellung der Physiker und Philosophen zur Mathematik wurde überaus präzise von Immanuel Kant in seiner 1786 verfassten Schrift über Metaphysische Anfangsgründe der Naturwissenschaft zum Ausdruck gebracht:

„ ... Ich behaupte nur, dass in jeder besonderen Naturlehre nur so viel eigentliche Wissenschaft angetroffen werden könne, als darin Mathematik anzutreffen ist.“

Ob man heute dieser Ansicht zustimmen kann oder nicht, sei vorderhand in den Raum gestellt. Am Schluss dieses Exposés werden wir noch darauf zurückkommen.

Evolutionstheorie und Mathematik

Der Name Biologie als Naturwissenschaft des Lebendigen ist jung im Vergleich zur Physik, die sich im 16. Jahrhundert als eigenständige Wissenschaft von der Philosophie abnabelte. Er taucht um die Wende von 18. zum 19. Jahrhundert an mehreren Stellen gleichzeitig auf und wurde von Jean-Baptiste Lamarck und Gottfried Reinhold Treviranus geprägt. Ganz allgemein sah man zwischen der unbelebten Natur von Physik und Chemie einerseits und der Physik andererseits eine unüberblickbare Kluft:

„ ... es ist für Menschen ungereimt, ... zu hoffen, dass noch etwa dereinst ein Newton aufstehen könne, der auch nur die Erzeugung eines Grashalms nach Naturgesetzen, die keine Absicht geordnet hat, begreiflich machen werde... “.

Mit anderen Worten ausgedrückt heißt dies: Die Gesetze der Physik können nicht zu einem Verstehen der verschiedenen Formen des Lebens beitragen.

Erst mit den bahnbrechenden Untersuchungen von Charles Darwin und Russell Edgar Wallace über Selektion und Entstehung der Arten bekam die Biologie ihr eigenständiges Gesicht in Form der Evolutionstheorie, die in der Folge auch als das Alleinstellungsmerkmal

der Biologie verstanden wurde. Niemand hat dies so prägnant formuliert wie der Genetiker Theodosius Dozhansky in seinem Aufsatz übertitelt:

„Nothing in biology makes sense except in the light of evolution.“

Charles Darwin wurde zwar von Ernst Haeckel enthusiastisch als Newton des Grashalms im Kantschen Sinne gefeiert aber wie Evelyn Fox Keller richtig ausführt, hat er keine Brücke zwischen der belebten und der unbelebten Natur geschlagen. Dies blieb der Molekularbiologie des 20. Jahrhunderts vorbehalten.

Jedem aufmerksamen Leser fällt sofort auf, dass Jahrhundertwerke über biologische Theorien vollkommen frei von mathematischen Überlegungen sind. Zwei hochberühmte Beispiele sind Charles Darwins „Origin of Species“ im 19. Jahrhundert und Ernst Mayrs „The Growth of Biological Thought“, das vor etwa 30 Jahren erschienen ist. Auch das 1942 bei Cambridge University Press erschienene von den meisten Biologen als Anfang einer theoretischen Biologie apostrophierte Buch „On Growth and Form“ von D’Arcy Wentworth Thompson kommt weitestgehend ohne Mathematik aus. Woran liegt die überaus schwache Nutzung formal mathematischen Denkens durch die Biologen? Sind biologische Inhalte nicht formalisierbar? Dagegen spricht die ungeheure Abstraktion von Darwin und Wallace, die zu einem allgemeinen Prinzip aus einer Fülle von Beobachtungen führte. Wie wir in der Folge zeigen werden, sind biologische Konzepte sehr wohl formalisierbar, wenn man die nötigen Fakten kennt.

Das Eintreten von Evolution auf der Basis von natürlicher Selektion nach Darwins Prinzip ist an drei notwendige und hinreichende Voraussetzungen gebunden:

1. Vermehrung mit Vererbung,
2. Variation und
3. Selektion.

Vermehrung von unabhängig agierenden Individuen führt zu exponentiellem Wachstum. Variation als Rekombination des Erbmaterials ist bei sexueller Reproduktion unvermeidlich und Mutationen als Kopierfehler des genetischen Materials treten zwangsläufig auf. Beides war Darwin unbekannt. Selektion ist eine Konsequenz exponentiellen Wachstums bei endlichen Ressourcen. Kann das Darwinsche Selektionsprinzip mit der zur Zeit Darwins zur Verfügung stehenden Mathematik formalisiert werden? Die Antwort ist ein klares: Ja! Wir stellen diese Behauptung anhand ganz einfacher Beispiele unter Beweis.

Der mittelalterliche Mathematiker Leonardo da Pisa, der auch unter dem Namen Fibonacci bekannt ist, hat Vermehrung durch ein Hasenmodell illustriert: Hasen bekommen Nachwuchs zu diskreten Zeitpunkten, etwa jedes Frühjahr, Hasen bringen Junghasen auf die Welt, die erwachsen werden müssen und daher beim nächsten Zeitpunkt noch keine Jungen haben, und Fibonacci’s Hasen leben ewig. Die Zahl der Hasenpaare wächst nach dem Gesetz der sogenannten Fibonaccifolge: $n_{k+1} = n_k + n_{k-1}$. oder 1,1,2,3,5,8,13,21,34,55,89,... . Diese Folge kann sehr gut mit Hilfe einer geometrischen Reihe angenähert werden:

$$n_k = n_{k-1} (1 + r) = n_0 (1 + r)^k \quad \text{mit} \quad n_0 = 1/\sqrt{5} \quad \text{und} \quad r = (\sqrt{5} - 1)/2. \quad (1)$$

Der Nationalökonom Thomas Robert Malthus hat 1798 in seinem berühmt gewordenen Buch über „Principle of Populations“ das Wachstum der menschlichen Population mit einer geometrischen Reihe dargestellt. Gleichung (1) geht davon aus, dass Populationen diskret

sind und dass sich die Individuen einer Generation nicht mit jenen einer anderen Generation paaren. Beide Voraussetzungen werden durch die Aufstellung eines kontinuierlichen Modells unnötig. Von der diskreten Gleichung (1) gelang man zum kontinuierlichen Darstellung durch die einfache Annahme, dass die Generationszeiten immer kürzer gemacht werden. Wir identifizieren die Generationszahl k mit der Zeit t und verkürzen nun den Zeitschritt $\Delta t = \vartheta$ für die Zählung der Individuen sukzessive: $\vartheta = 1/m$ mit $m = 1,2,3,\dots$. Einsichtiger Weise ist die Zahl der Nachkommen in einem kürzeren Zeitabschnitt geringer aber im Mittel proportional zur verstrichenen Zeit, und wir müssen in Gleichung (1) r durch r/m ersetzen. Nun können wir leicht den Übergang zur kontinuierlichen Darstellung durchführen:

$$\lim_{m \rightarrow \infty} \left(1 + \frac{r}{m}\right)^m = e^r \quad \text{und} \quad n(t) = n(0) e^{rt} \quad (2)$$

Ob Fibonacci-Folge oder geometrische Reihe beide können durch eine Exponentialfunktion angenähert beziehungsweise ersetzt werden: Vermehrung von Organismen im biologischen Sinn führt letztlich immer auf exponentielles Wachstum

Die Chemiker haben in Form der Reaktionskinetik ein schlagkräftiges Instrumentarium entwickelt, um kontinuierliche zeitliche Entwicklungen von Populationen aller Art zu berechnen. Wir werden hier die chemische Reaktionskinetik auf Fragen der biologischen Evolution anwenden. Vermehrung nach der chemischen Reaktionsgleichung



führt zu exponentiellem Wachstum in der Zahl der Individuen $x = [X]$:^a Mit A werden die Ressourcen bezeichnet, aus welchen X aufgebaut wird. Durch die Schreibweise (A) deuten wir an, dass dieses Reservoir so groß ist, dass es durch die Synthese von X nicht nennenswert geschmälert wird. Dann ergibt sich:

$$\frac{1}{m-1} \frac{dx}{dt} = kx \Rightarrow x(t) = x(0) e^{(m-1)kt} = x(0) e^{ft} \quad (4)$$

Die linke Gleichung ist eine reaktionskinetische Differentialgleichung, welche die momentane Änderung in der Zahl der Individuen in der Population, der Populationsgröße, zum Ausdruck bringt, m ist die Gesamtzahl der Individuen nach einem Vermehrungsschritt. In der chemischen Kinetik wird die Zahl der in einem Schritt produzierten Nachkommen, hier $m-1$, durch den stöchiometrischen Faktor $1/(m-1)$ berücksichtigt, und k ist ein Reaktionsparameter, der in der Biologie multipliziert mit der Zahl der Nachkommen, $k(m-1) = f$, als „Fitness“ der Spezies bezeichnet wird. Der Vergleich der beiden Gleichungen (2) und (4) zeigt, dass die Zahl der Nachkommen die Natur der Wachstumsfunktion – exponentielles Wachstum – nicht beeinflusst.

Werden für die Vermehrung zwei oder mehr Exemplare der zu vermehrenden Spezies gebraucht,



^a Mit den eckigen Klammern wird die zahlenmäßige Menge einer Variablen ausgedrückt. Es kann sich dabei ebenso um Individuen in einer Population wie um die Konzentration einer Substanz handeln.

so spricht man von Autokatalyse höherer Ordnung. Wie man leicht berechnen kann, folgt die Autokatalyse höherer Ordnung einem anderen Wachstumsgesetz:

$$\frac{dx}{dt} = k x^2 \Rightarrow x(t) = x(0) \frac{1}{1 - x(0) k t} \Rightarrow t_{\text{cr}} = \frac{1}{x(0) k} \quad (6)$$

Da die entstehende Wachstumsgleichung eine Hyperbel darstellt spricht man von hyperbolischem Wachstum. Ein Charakteristikum des hyperbolischen Wachstums besteht darin, dass unendliche Ressourcen vorausgesetzt die Wachstumskurve bei einem kritischen Zeitpunkt t_{cr} unendlich werden kann. Einsichtiger weise gibt es keine unendlichen Ressourcen. In Abbildung 1 sind Wachstumskurven, die verschiedenen Gesetzen folgen miteinander verglichen.

Die Besonderheit exponentiellen Wachstums ist mehrfach: (i) Das exponentielle Wachstum steht an der Grenze zwischen den Wachstumskurven, die zu endlichen Zeiten den Wert unendlich erreichen und jenen, welche erst für unendliche Zeiten unendlich werden, (ii) die interne Dynamik in unbeschränkt wachsenden Populationen ist identisch mit jener in beschränkt wachsenden oder in stationären Populationen und (iii) nur exponentielles Wachstum führt zwangsweise auf Selektion, denn bei subexponentielles Wachstum kann Koexistenz eintreten und in hyperbolisch wachsenden Populationen kann eine Variante, die sich einmal durchgesetzt hat, nicht mehr durch eine fittere Spezies verdrängt werden.

Alle höheren Organismen vermehren sich sexuell entweder obligat wie die Säugetiere oder fast immer wie viele andere Spezies, beispielsweise Reptilien, und sexuelle Vermehrung führt auch zu exponentiellem Wachstum. Hier sollten Protestrufe aus Ihrer Mitte kommen! Brauche man doch auch zwei Individuen für die sexuelle Vermehrung, und haben wir nicht gerade eben gehört, dass der Prozess $2X \rightarrow 3X$ zu hyperbolischem Wachstum führt? Das ist richtig aber die Einheit der Vermehrung ist weder das weibliche noch das männliche Individuum sondern das durch Vereinigung der beiden Individuen entstehende Paar wie sich ganz einfach durch die Anwendung von Mathematik zeigen lässt. Man kann verschieden komplizierte Modelle für die sexuelle Vermehrung aufstellen und durchrechnen, sie alle führen auf exponentielles Wachstum. Ein besonders einfaches Modell der sexuellen Vermehrung umfasst die folgenden zwei Schritte:



Ein Weibchen X vereinigt sich mit einem Männchen Y zu einem Paar, welches Nachkommen zeugen kann. Paarbildung und Paarauflösung werden durch die Reaktionsgleichung (7a) beschrieben und die Jungen werden in einem zweiten Reaktionsschritt (7b) produziert. In der Natur sind die beiden stöchiometrischen Faktoren, $n, m = 0, 1, 2, \dots$, von Wurf zu Wurf gewissen Schwankungen unterworfen. Der Einfachheit halber nehmen wir hier $n = m = 1$ an. Da zwei Individuen für die Zeugung von Nachkommen notwendig sind, scheint es naheliegend zu vermuten, dass wir es mit einer modifizierten Version des Systems in Gleichung (5) zu tun haben. Um dies zu klären, betrachten das Gleichung (7a,b) entsprechende Differentialgleichungssystem mit den Variablen $[X] = x$, $[Y] = y$ und $[X \cdot Y] = z$:

$$\begin{aligned}\frac{dx}{dt} &= -k_1 x y + l_1 z + k_2 z = \frac{dy}{dt} \\ \frac{dz}{dt} &= k_1 x y - l_1 z\end{aligned}\tag{8}$$

Die beiden Reaktionsparameter k_1 und l_1 beschreiben die Häufigkeit der Paarbildung beziehungsweise der Paarauflösung, und k_2 ist der Reaktionsparameter für die Produktion der Nachkommenschaft. Maßgeblich für das Wachstum sind die Gesamtzahl der weiblichen und männlichen Individuen, $x_0 = x + z$ beziehungsweise $y_0 = y + z$. Die in Abbildung 2 gezeigte numerische Integration der Differentialgleichung (8) lässt zwei Phasen der Vermehrung erkennen: (i) die Bildung von Paaren X-Y und (ii) die Vermehrung der Paare. Die letztere erfüllt exponentielles Wachstum nahezu perfekt. Ein vereinfachender technischer Trick sei noch erwähnt: Nimmt man an, dass die Paarbildung und -auflösung hinreichend rasch erfolgt, dann kann man einen quasistationären Zustand annehmen, der durch den Ansatz $dz/dt = 0$ beschrieben wird. Eine Variable kann dadurch eingespart werden und wir erhalten:

$$\frac{dx_0}{dt} = \frac{k_2}{2K} \cdot \left(1 + K(2x_0 + \vartheta) - \sqrt{(1 + K(2x_0 + \vartheta))^2 - 4K^2 x_0 (x_0 + \vartheta)} \right) \approx k_2 x_0\tag{9}$$

Darin bedeuten $K = k_1 / l_1$ die Gleichgewichtskonstante und $\vartheta = y_0(t) - x_0(t) = y(t) - x(t)$ die Differenz zwischen männlichen und weiblichen Individuen. Für große Populationen, $x_0 \gg \vartheta$ und $y_0 \gg \vartheta$, gilt in der Tat das exponentielle Wachstumsgesetz (4). Damit ist auf der Basis der mathematischen Analyse zu zeigen, dass im Rahmen einer in der Praxis erfüllten Näherung das exponentielle Wachstumsgesetz für sexuelle Vermehrung zwangsweise gilt.

Beschränkte Ressourcen wurden von Pierre-François Verhulst bereits 1828 mit Hilfe eines noch heute durchgehend verwendeten Modells eingeführt, welches als logistische oder Verhulst-Gleichung bekannt ist. Verhulst geht davon aus, dass jedes Ökosystem nur eine begrenzte Zahl von Individuen ernähren kann. Dieser Beschränkung wird dadurch Rechnung getragen, dass ein zusätzlicher Term in Gleichung (4), $-f x^2/C$, eingeführt wird.:

$$\frac{dx}{dt} = f x \left(1 - \frac{x}{C} \right) \Rightarrow x(t) = \frac{x(0) \cdot C}{x(0) + (C - x(0)) \cdot e^{-ft}}\tag{10}$$

Die Wirkung dieses zusätzlichen Terms auf dx/dt ist leicht illustriert: Bei kleinen x -Werten, $x \ll C$, ist der negative Term gegenüber $f x$ vernachlässigbar klein und exponentielles Wachstum wird beobachtet (Abbildung 3), kommen die Werte von x in die Nähe der Kapazität C , $x \approx C$, dann geht das Wachstum dx/dt gegen null. Die resultierende Kurve wird als logistisches Wachstum charakterisiert und gilt als Prototyp exponentiellen Wachstums bei begrenzten Ressourcen.

Die neue Biologie: Epigenetik und 'Big-Data'

In den späten 1990iger Jahren wurden Entdeckungen an RNA-Molekülen gemacht, welche zeigten, dass das damalige molekularbiologische Bild der Genregulation wohl für Bakterien einigermaßen zutreffend ist, dass es aber bei höheren Vielzellerorganismen zumindest ganz entscheidend erweitert werden muss. Die Aktivierung eines Gens durch Faktoren für die Transkription ist ungleich komplizierter, es gibt andere Mechanismen als die Synthese von Repressorproteinen um Gene still zu legen, und die Genregulation erfolgt auch auf der Stufe der transkribierten RNA. Dasselbe Stück DNA kann in verschiedenen Geweben dadurch in verschiedene Proteine übersetzt werden, dass die RNA vor der Übersetzung verschieden prozessiert wird. Mechanismen für epigenetische Vererbung wurden entdeckt und aufgeklärt, die lange Zeit für unmöglich gehalten wurden. Allmählich beginnt sich ein komplexes Bild der Regulation des Metabolismus im Vielzellerorganismus abzuzeichnen, welches auch ein völlig neues Licht auf die Entwicklungsbiologie wirft. Gleichzeitig wurden ganz neue Techniken erfunden und perfektioniert, welche die rasche Erhebung von früher unvorstellbar großen Datenmengen ermöglichen: 'Big-Data' wurde zum Schlagwort für die neue Situation in der Biologie und in der Medizin erhofft man sich die Einführung neuer Behandlungsmethoden, die beispielsweise eine 'personalized medicine' ermöglichen werden könnten: Jedem Patienten seine maßgeschneiderte Therapie!

Auf manchen Teilgebieten der Biologie ist eine auch nur Überblicks mäßige Erfassung der erhobenen Daten durch den Experimentalwissenschaftler nicht mehr möglich. Aufbereitung und Analyse der Daten erfolgen ausschließlich durch Computereinsatz und der Biologe ist abhängig von der Verlässlichkeit der von ihm verwendeten Software. Allerdings wird durch die große und stark wachsende Zahl von Molekulargenetikern, die zur Zeit forschen, die Gefahr unentdeckter gravierender 'Bugs' immer kleiner. Dennoch, der Erfolg der neuen Biologie wird nicht zuletzt davon abhängen, dass eine neue umfassende theoretische Biologie entwickelt wird, in deren methodischem Repertoire Mathematik, Informatik und Computerwissenschaften unabdingbare Bestandteile darstellen.

Fragment muss noch vervollständigt werden.