

Univ.-Prof. Dr. Joachim Hermisson

Moderne Evolutionsforschung in Wien: Entwicklungen und Herausforderungen

150 Jahre nach Charles Darwins "Origin of Species" führen neue DNA-Sequenzierungstechnologien und moderne mathematische Methoden zu erstaunlichen Erkenntnissen über die Bedeutung und Reichweite natürlicher Selektion. In der Evolutionsforschung zeichnet sich der Anfang einer neuen integrativen Phase ab, in der Konzepte aus Ökologie, Genetik, Entwicklungsbiologie, Bioinformatik und Mathematik auf der Ebene molekularer Sequenzdaten verknüpft werden. Wien ist mit zahlreichen Forschungsgruppen ein Zentrum dieser Entwicklung.

Zwei fundamentale Erkenntnisse kennzeichnen Darwins Evolutionstheorie: zum einen die Evolution als die Veränderung biologischer Arten über die Zeit und ihre Abstammung von gemeinsamen Vorfahren, zum anderen die Selektion als die Triebfeder von Adaptation, der Anpassung von Arten an ihre Umwelt. Während die Tatsache von Evolution binnen weniger Jahre in der ganzen wissenschaftlichen Welt akzeptiert wurde, hatte es die Selektion sehr viel schwerer. Erst mit der Verknüpfung von Darwins Theorie und Mendels Vererbungslehre in der Populationsgenetik wurden die Probleme auf der theoretischen Ebene gelöst. Auch an eindeutigen empirischen Belegen für natürliche Selektion und Adaptation gibt es heute keinen Mangel. Dennoch sind selbst fundamentale Fragen immer noch offen und bis vor kurzem für die Forschung kaum zugänglich: Wie häufig sind Adaptationsereignisse in der Geschichte einer biologischen Art? Wie groß sind die einzelnen adaptiven Schritte? Wie stark ist die Selektion? Welche Gene oder DNA-Abschnitte und welche phänotypischen Merkmale sind betroffen?

Jahrzehntelange Feldforschungen

Die klassische Adaptationsforschung setzt am Phänotyp an. Für gewöhnlich ist es nicht schwierig, für phänotypische Merkmale plausible adaptive Szenarien zu entwerfen. Ein eindeutiger Nachweis, der über solches "adaptive story telling" hinausgeht, ist jedoch sehr viel schwieriger und - wenn überhaupt - oft nur mit enormem Aufwand möglich. Detaillierte Erkenntnisse - wie die Aufschlüsselung der Evolution der Schnabelform bei den Darwinfinken auf den Galapagosinseln - sind das Resultat jahrzehntelanger Feldforschung. Der phänotypische Ansatz erfordert zudem die Konzentration auf einzelne Merkmale, die mutmaßlich unter sehr starker Selektion stehen. Ein Gesamtbild aller Adaptionsereignisse innerhalb einer Art kann so nicht gewonnen werden.

Neutrale Mutationen

Seit der einfachen Verfügbarkeit von Genomdaten wird deshalb versucht, Adaptation auch auf der genotypischen Ebene nachzuweisen. Durch die vergleichende Genomanalyse (zwischen Arten) gelingt aber nur in Einzelfällen der klare Nachweis, dass Unterschiede in den DNA- bzw. Aminosäuresequenzen tatsächlich die Folge positiver Selektion sind. In der Tat gibt es eine plausible Alternative: Nach der vom japanischen Genetiker Motoo Kimura (1924-1994) vorgeschlagenen Neutraltheorie der Evolution sollte die weit überwiegende Anzahl der DNA-Unterschiede (sogenannte molekulare Divergenzen) selektiv neutral sein, also weder vorteilhaft noch nachteilig. Neutrale Mutationen sind wahrscheinlich sehr viel häufiger als positiv selektierte. Durch genetische Drift (also Zufallsschwankungen in ihrer Frequenz) können sich aber auch solche Mutationen in einer Population ausbreiten und letztendlich zu Unterschieden zwischen verschiedenen Arten führen.

"Junk-DNA"

Ende des zwanzigsten Jahrhunderts war die Auseinandersetzung zwischen SelektionistInnen und NeutralistInnen ein prägendes Thema der Evolutionsforschung. Insbesondere auf der Ebene der DNA wurde die Bedeutung neutraler Evolution durch genetische Drift betont. Da die weitaus größten Teile des Genoms in vielen Arten nicht-kodierend sind (also im Gegensatz zu den Genen nicht in ein Protein übersetzt werden), galten sie im Wesentlichen als funktionslose "Junk DNA" und daher evolutionär neutral. Aber auch in den kodierenden Bereichen der Gene wurde Selektion überwiegend als Kraft wahrgenommen, die schädliche Mutationen aus der Population entfernt (negative Selektion). Noch bis vor wenigen Jahren galt deshalb positive Selektion als sehr seltenes Ereignis. Erkenntnisse der jüngsten Zeit führen jedoch zu einer grundlegenden Neubewertung dieses Bildes.

Neue Methoden

Den Anstoß für neue Entwicklungen gaben - wie oft - technologische Fortschritte. Eine neue Generation von Sequenzierungsmaschinen, die seit drei bis vier Jahren am Markt eingeführt wird, beschleunigt die Gewinnung von Rohdaten um mehrere Größenordnungen. Seit den Tagen des "human genome projects" sind die Sequenzierungskosten um einen Faktor 1.000 bis 10.000 gefallen (eine genaue Angabe ist wegen der schnell veränderlichen Preise nicht möglich). Dementsprechend ist Anfang 2008 ein neues internationales Projekt zur Sequenzierung von mehr als 1.000 menschlichen Genomen gestartet worden - in einer Zeit von nur drei Jahren.

Entscheidend ist, dass mit diesen Hochleitungsverfahren ein neuer Datentyp zugänglich wird: genomweite Polymorphismusdaten. Das Hauptziel der klassischen Sequenzierung - wie etwa im "human genome project" - war es gewesen, Referenzsequenzen zu generieren, die die Art (ob Fruchtfliege, Maus oder Mensch) im Baum des Lebens repräsentiert. Mit den neuen "high throughput"-Verfahren ist es nun erstmals möglich, nicht nur einzelne Sequenzen zu gewinnen, sondern für jede Art die Genome von hunderten oder tausenden Individuen einzeln zu sequenzieren. Während Referenzgenome über die Unterschiede zwischen den Arten Auskunft geben, geht es bei der Sequenzierung von Stichproben um die Varianz innerhalb einer Art. Diese Unterschiede zwischen den DNA Sequenzen von Individuen einer Art bezeichnet man als Polymorphismen. In einer Stichprobe bilden die Polymorphismen entlang des Genoms charakteristische Muster.

Von der Phylogenetik zur Populationsgenetik

Der Schwerpunkt der Forschung verschiebt sich damit von der Phylogenetik, die sich mit der Rekonstruktion des Baums des Lebens beschäftigt, zur Populationsgenetik (oder -genomik), die beschreibt, wie Polymorphismen in einer Population durch Neumutationen entstehen und von den evolutionären Kräften (Selektion, Rekombination, Drift) beeinflusst werden. Schon seit rund 30 Jahren weiß man, dass das Polymorphismus-Muster nicht nur den genetischen Zustand der gegenwärtigen Population abbildet, sondern auch Informationen über die jüngere evolutionäre Geschichte speichert. Dies gilt insbesondere für Adaptationsereignisse in Folge positiver natürlicher Selektion. Im Zentrum steht dabei die Erkenntnis, dass die rasche Ausbreitung einer positiv selektierten Genvariante das Polymorphismus-Muster in benachbarten DNA-Abschnitten, die nicht selbst unter Selektion stehen ("linked neutral DNA"), beeinflusst und in charakteristischer Weise prägt. Solche "selective sweeps" bleiben je nach Populationsgröße mehrere 1000 bis 100.000 Generationen sichtbar. Man kann deshalb hoffen, die charakteristischen Signale in genomweiten Polymorphismusdaten zu finden und so ein umfassendes Bild über die jüngere adaptive Geschichte einer Population zu erhalten.

Aber wann genau ist ein Polymorphismus-Muster charakteristisch genug, um auf Selektion schließen zu können? Die theoretische Beschreibung von Polymorphismus-Mustern erfolgt im Rahmen mathematischer Modelle, die auch die Basis für statistische Testverfahren darstellen. Tatsächlich wäre ohne die Entwicklung neuer anspruchsvoller mathematischer Verfahren eine Interpretation der biologischen Sequenzdaten nicht möglich gewesen. Polymorphismus-Daten werden heute in der Regel im Rahmen der Koaleszenz-Theorie ("coalescent theory") beschrieben. Das ist ein maßgeschneiderter Formalismus aus dem Gebiet der stochastischen Prozesse, der Evolution rückwärts in der Zeit betrachtet und so den Effekt evolutionärer Kräfte auf die Genealogie (d.h. den gemeinsamen Stammbaum) von Stichproben aus einer Population beschreibt. Mit der Verfügbarkeit genomweiter Daten wurden diese Verfahren entsprechend erweitert und Tests zur Identifikation von Selektionsmustern entwickelt.

Unerwartete Resultate

Was konnte man von einem solchen Zugang erwarten? Zweifel waren in jedem Fall angebracht. So können mit vielen Methoden nur evolutionär sehr junge Ereignisse erfasst werden. Außerdem ist das Polymorphismus-Muster auch unter neutraler Evolution sehr variabel. Neutralität als Nullhypothese ist entsprechend schwer zu widerlegen, Tests auf positive Selektion sollten nur die Spitze des Eisbergs erkennen. Wenn Adaptation außerdem so selten wäre, wie weithin angenommen, würde man kaum etwas finden. Vor diesem Hintergrund müssen die Resultate der ersten genomweiten Adaptationsstudien überraschen.

Starke Beschleunigung des Adaptationsprozesses

Für zahlreiche untersuchte Arten, für die bereits Daten zur Verfügung stehen, wie Mensch, Maus oder Fruchtfliege (*Drosophila*), erhält man Selektionssignale in unerwarteter - ja fast unglaublicher - Menge. Für eine evolutionär sehr kurze Zeit werden mit unterschiedlichen Testverfahren Hunderte oder Tausende von Adaptationsereignissen vorhergesagt. Insbesondere für den Menschen sind die veröffentlichten Werte extrem. Mehrere tausend Adaptationen in nur rund 80.000 Jahren, oder mehr als sieben Prozent aller Gene in nur 3.000 Generationen: Das ist schwer zu begreifen und nur mit einer starken Beschleunigung des Adaptationsprozesses in jüngster Zeit erklärbar. Neue Selektionsdrücke als Folge neuer Krankheiten (wie Malaria) und der Umstellung der Lebensweise (Ackerbau) werden als mögliche Gründe genannt.

Aber auch über etwas längere Zeiträume kommt man zu erstaunlichen Ergebnissen. Über den Vergleich innerartlicher und zwischenartlicher Variation (Polymorphismus und Divergenz) lässt sich der Anteil der Aminosäure-Unterschiede zwischen Arten abschätzen, der adaptiv ist, also Resultat positiver Selektion. Für nahe verwandte Drosophila-Arten ergibt sich so ein Anteil von rund 50 Prozent - bei Berücksichtigung schwacher Selektion sogar bis zu 95 Prozent - ein klarer Gegensatz zu den Vorhersagen der Neutraltheorie der Evolution.

Gibt es überhaupt "Junk-DNA"?

Neben der Häufigkeit der Adaptationsereignisse überrascht auch ihre Verteilung auf dem Genom. Es zeigt sich, dass sehr viel mehr Regionen auf dem Genom unter Selektion stehen als noch vor wenigen Jahren für möglich gehalten. Adaptation ist keineswegs auf die Protein kodierenden Abschnitte beschränkt. Ein wesentlicher Anteil betrifft Abschnitte zwischen den Genen, die eine wichtige Rolle für die Genregulation spielen (d.h., sie bestimmen die Menge an Protein, die von einem Gen produziert wird). Für Drosophila wird geschätzt, dass bis zur Hälfte der gesamten nicht-kodierenden DNA unter Selektion steht. Und je nach betrachtetem Abschnitt werden 20 bis 60 Prozent aller nicht kodierenden Sequenzunterschiede zwischen Arten als adaptiv klassifiziert. Da in den meisten Organismen der bei weitem größte Teil des Genoms nicht-kodierend ist (bei Drosophila 80 Prozent, beim Menschen 95 Prozent), ist Adaptationen auf diesen Abschnitten möglicherweise sogar häufiger als auf den Proteinen. Adaptation wäre damit auf dem Genom so weit verbreitet, dass sich bereits die Frage stellt, ob es überhaupt wesentliche Abschnitte der DNA gibt, die keine Funktion haben. Mit dem Begriff "Junk-DNA" ist man jedenfalls sehr vorsichtig geworden.

Es besteht keine Frage, dass diese Resultate das Potenzial haben, unser Verständnis von Adaptation nachhaltig zu verändern. Einstweilen müssen sie sich aber erst im Lichte nachfolgender Forschung bewähren. Während die erste Runde genomweiter Analysen eindrucksvoll das Potenzial des neuen Ansatzes demonstriert, bleiben viele Fragen vorerst offen. Warum, zum Beispiel, scheinen einige Modellorganismen wie die Schaumkresse (*Arabidopsis*) ein völlig anderes Bild zu zeigen? Und vor allem: Wie sicher ist die Interpretation der Daten? Die Entwicklung der Methoden befindet sich noch in einem frühen Stadium, Fehler im Datenmaterial und alternative Deutungsmöglichkeiten müssen in Zukunft ausgeschlossen werden.

Weiterer Forschungsbedarf

Auf der empirischen Seite ist die Qualität von "high throughput"-Sequenzdaten für viele Analysemethoden noch nicht ausreichend. Dies betrifft zum einen die relativ hohe Fehlerrate bei der Bestimmung der Basenfolge (den "base calls"). Zum anderen können mit vielen "high throughput"-Verfahren nur jeweils sehr kurze zusammenhängende DNA-Abschnitte gelesen werden, was die Analyse von statistischen Korrelationen zwischen verschiedenen Abschnitten (linkage disequilibrium) schwierig macht. Es ist aber abzusehen, dass die damit verbundenen Probleme bald gelöst sein werden. So verbessert sich die Qualität der Daten mit jeder neuen Sequenzergeneration. Vor allem aber werden die Sequenzierungskosten auch in Zukunft weiter fallen. Magnus Nordborg, neuberufener Direktor am Gregor-Mendel-Institut, sagt voraus, dass künftige BiologiestudentInnen im Rahmen eines Praktikums ihr eigenes Genom sequenzieren werden. Geringere Kosten verbessern effektiv auch die Qualität schlicht dadurch, dass Fehler über eine vielfache Resequenzierung desselben Genoms korrigiert werden können. Die Hauptaufgabe der empirischen Evolutionsforschung wird sich langfristig auf die funktionale Analyse verschieben, also der Identifikation der phänotypischen

Adaptationen, denen die potenziell vorteilhaften Genvarianten zugrunde liegen. Dieser Schritt bleibt in jedem Fall eine große Herausforderung.

Wesentlicher Forschungsbedarf besteht aber auch auf der Seite der mathematischen Modellbildung. Die gegenwärtigen Modelle der molekularen Populationsgenetik beruhen durchwegs auf stark vereinfachenden Annahmen über die genetischen und ökologischen Voraussetzungen von natürlicher Selektion. Betrachtet wird ein Szenario von Adaptation auf einem isolierten Gen-Locus, ausgehend von einer einzelnen vorteilhaften Neumutation. Die Population ist idealerweise nicht strukturiert und von konstanter Größe. Fast durchgehend wird zudem ein räumlich und zeitlich konstanter Selektionsdruck vorausgesetzt. All diese Annahmen stehen in Widerspruch zu Erkenntnissen über Adaptationsprozesse in der Natur: Natürliche Populationen sind variabel und oft stark strukturiert. Selektion ist die Folge räumlich und zeitlich veränderlicher Umweltbedingungen. Sie wirkt nicht direkt auf einzelne Gen-Loci, sondern auf phänotypische Merkmale, die von vielen interagierenden Genen bestimmt werden. Ihr Ursprung ist oft nicht eine einzelne Neumutation, sondern Gen-Varianten, die in der lokalen Ursprungspopulation bereits vorhanden waren oder durch Genfluss aus anderen Populationen importiert werden.

Es ist bekannt, dass all diese Faktoren einen erheblichen Einfluss auf DNA-Polymorphismus-Muster haben. Für die Abschätzung von Häufigkeit und Stärke von Adaptationsereignissen und die Identifikation "adaptiver Gene" ist ihre Einbeziehung deshalb von zentraler Bedeutung. Die Aufgabe für zukünftige Forschung ist deshalb die Anpassung der Modelle an realistische biologische Szenarien. Diese ist vor allem deshalb anspruchsvoll, da sie nur in Kombination detaillierter Kenntnisse aus Ökologie, Genetik und Molekularbiologie mit fortgeschrittenen mathematischen und bioinformatischen Methoden möglich ist. Für die Evolutionsforschung bedeutet dies nach Jahren der Spezialisierung in verschiedene Teilbereiche die Notwendigkeit zu Integration und Kooperation. Kompetenzen aus verschiedenen Labors und Arbeitsgruppen müssen verknüpft werden.

Evolutionsforschung in Wien

Für Forschungsstandorte, die im Bereich Evolution den wissenschaftlichen Fortschritt an vorderster Front gestalten wollen, bedeutet dies eine erhebliche Anforderung. Insbesondere ist eine kritische Masse an Forschungsgruppen erforderlich, um die verschiedenen Kompetenzbereiche abzudecken - von molekular bis ökologisch und von empirischer Forschung zur mathematischen Methodenentwicklung. Weltweit haben nur wenige Standorte hierfür die notwendigen Voraussetzungen. Aufbauend auf einer langen Tradition und in jüngster Zeit durch gezielte Neuberufungen verstärkt, ist Wien heute auch im internationalen Vergleich in einer hervorragenden Ausgangsposition. Mehr als 20 Gruppen aus den Bereichen Biologie, Biochemie, Bioinformatik, Statistik und Mathematik haben sich der Evolution als Arbeitsgebiet verschrieben.

Während die Selektion als Motor des Adaptationsprozesses in der Biologie heute nicht mehr in Frage steht, sind für ein Gesamtbild des Prozesses selbst elementare Fragen immer noch offen. Nach 150 Jahren besteht nun zum ersten Mal die Chance, viele dieser Fragen auf der Basis von molekularen Sequenzdaten der Forschung zugänglich zu machen. Die Antworten werden unser künftiges Verständnis von Evolution prägen. Und die Wiener Forschung wird hierzu ihren Beitrag leisten können.

Univ.-Prof. Dr. Joachim Hermisson ist Stiftungsprofessor für Mathematik und Biowissenschaften an den Max F. Perutz Laboratories und der Fakultät für Mathematik.