DOI:10.1002/biuz.201310498

Forward Genomics – ein neuer Ansatz der vergleichenden Sequenzanalyse:

Vom Phänotyp zum Genotyp

Michael Hiller

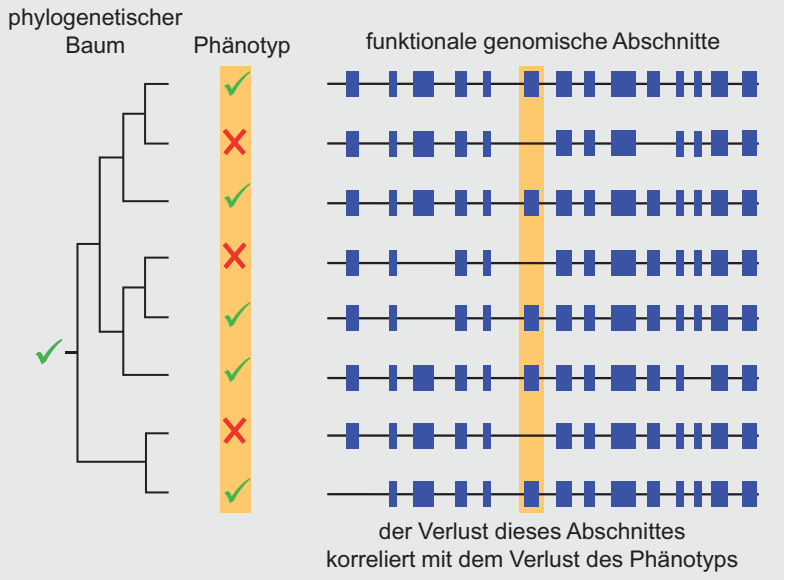
Vergleichende Genomik kann helfen, Änderungen im Genom zu finden, die für phänotypische Unterschiede zwischen Spezies verantwortlich sind.

Im Laufe von vielen Millionen Jahren hat die Evolution zu einer erstaunlichen Vielfalt an Spezies geführt. So können beispielsweise einige Säugetiere wie Fledermäuse fliegen, während andere wie Wale und Delfine in den ozeanischen Lebensraum zurückgekehrt sind. Und auch wir Menschen unterscheiden uns in vielen Merkmalen von unserem nächsten Verwandten, dem Schimpansen. Diese Vielfalt an beobachtbaren Charakteristika (Phänotypen genannt) sehen wir nicht nur innerhalb der Säugetiere, sondern finden sie in allen Tier- und Pflanzengruppen.

Auf molekularer Ebene ist die DNA die Blaupause des Lebens, in der die Phänotypen eines Organismus verschlüsselt sind. Wenn die DNA zweier Individuen gleich ist, sind die Phänotypen dieser Individuen ebenfalls sehr ähnlich, wie man an eineiigen Zwillingen erkennen kann. Daher müssen Unterschiede in der DNA für die vielen phänotypischen Unterschiede zwischen den Arten verantwortlich sein. Mittlerweile ist es technisch möglich, die Gesamtheit der DNA (das Genom) zu entschlüsseln, d. h. (fast) alle A,C,G und T-Buchstaben lesen zu können. Seit ein paar Jahren sind einige Genome von Wirbeltieren, Insekten und anderen Spezies entschlüsselt. Trotzdem wissen wir im Allgemeinen nur sehr wenig darüber, welche Unterschiede im Genom für bestimmte phänotypische Unterschiede verantwortlich sind.

Die vergleichende Genomik kann die Genome verschiedener Arten systematisch vergleichen und darin sowohl nach Ähnlichkeiten als auch nach Unterschieden auf der DNA-Sequenzebene suchen. Warum also ist es dennoch so schwierig, die phänotypischen Änderungen mit genomischen Unterschieden in Zusammenhang zu bringen? Der Hauptgrund dafür ist, dass jedes Paar von Spezies, selbst wenn es so eng miteinander verwandt ist wie Mensch und Schimpanse, viele Millionen von genomischen Veränderungen aufweist und sich gleichzeitig in hunderten oder tausenden Phänotypen unterscheidet.

ABB. 1 | **KORRELATION ZWISCHEN EINEM PHÄNOTYP UND DEM VERLUST EINER FUNKTIONALEN GENOMREGION**



Jede blaue Box stellt eine Region im Genom dar. Wir stellen hier „Änderungen in einer Region" vereinfacht als den Verlust der Region dar. Die meisten Boxen sind in allen Spezies vorhanden oder fehlen nur in wenigen. In diesem Beispiel gibt es nur eine Region, wo das Vorhandensein/Fehlen exakt mit dem Phänotyp korreliert.

Vitamin C-Synthese und wiederholte Evolution

Einige phänotypische Unterschiede haben jedoch eine Eigenschaft, die sehr hilfreich ist: Dieselben phänotypischen Unterschiede kommen in verschiedenen, unabhängig voneinander entwickelten Arten vor, d.h. Evolution hat sich wiederholt. Besonders von Interesse ist dabei der wiederholte Verlust eines bestimmten Phänotyps (Abbildung 1). Ein Beispiel dafür ist die Syn-

these von Vitamin C. Die Mehrheit der Säugetiere kann körpereigenes Vitamin C herstellen. Der Mensch und einige verwandte Primaten sowie das Meerschwein und Fledermäuse haben diese Fähigkeit jedoch verloren und müssen Vitamin C über die Nahrung aufnehmen, um die Krankheit Skorbut zu vermeiden [1]. Die Fähigkeit, das Vitamin zu synthetisieren, war im gemeinsamen Vorfahren aller Säugetiere vorhanden. Daher kann man davon ausgehen, dass im Genom dieses Vorfahren die entsprechende Information für diesen Phänotyp (Vitamin C-Synthese) vorhanden war. Dann erklärt sich der Verlust dieses Phänotyps vermutlich als Verlust dieser genomischen Information, während in den heutigen Vitamin C-synthetisierenden Säugetieren diese Information noch intakt sein sollte. Dies liefert eine Möglichkeit, im Genom vieler Säugetiere die genomischen Regionen und damit die entsprechenden Änderungen zu finden, die für den Unterschied im Vitamin C-Phänotyp verantwortlich sind. Während man unzählige genomische Regionen findet, die Veränderungen zwischen zwei Arten aufweisen, sollte es nur sehr wenige Regionen geben, deren Änderungen in exakt den Spezies vorkommen, die nicht mehr Vitamin C synthetisieren können. Demzufolge führen solche mehrfachen phänotypischen Verluste zu einem spezifischen evolutionären Muster in diesen Genomen, nach dem man suchen kann (Abbildung 1). Das wiederum erhöht die Genauigkeit der Vorhersage eines Zusammenhanges zwischen genomischer Region und phänotypischer Änderung.

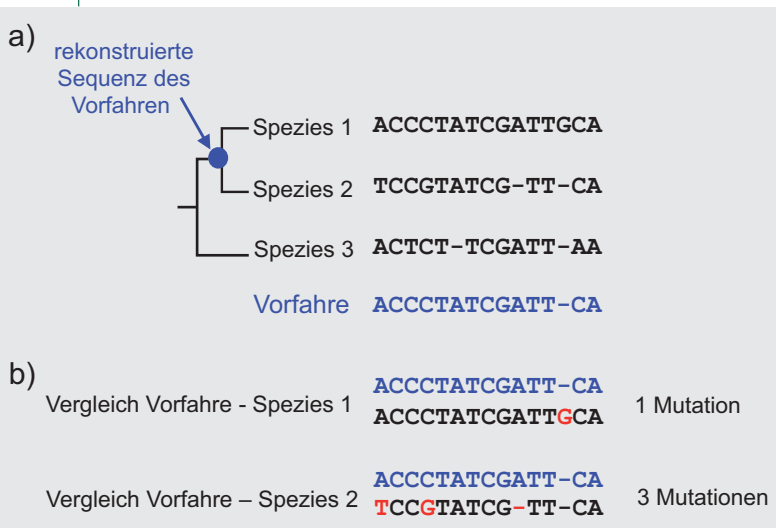
Selektion vs. neutrale Evolution

Das in Abbildung 1 dargestellte Prinzip basiert auf dem Unterschied zwischen neutraler Evolution und Selektion. Wenn ein Phänotyp für Überleben und Fitness einer Art wichtig ist, dann wird auch die genomische Information für diesen Phänotyp wichtig sein und zufällige Mutationen, die diese Information zerstören, werden von der Selektion (oft auch „negative" Selektion genannt) aussortiert. Als Folge evolvieren die Genomabschnitte, die diese Information enthalten, viel langsamer (es häufen sich also weniger Änderungen an) als andere neutrale Abschnitte, die keine wichtige Information beinhalten. Die Regionen, die Informationen für Vitamin C-Synthese enthalten, sind also unter Selektion in allen Spezies, wo dieser Phänotyp wichtig ist. Was passiert aber in den Spezies, die den Phänotyp verloren haben? Wenn eine Art zu einem gewissen evolutionären Zeitpunkt den Phänotyp verliert und das keine negativen Folgen hat (z.B. weil genügend Vitamin C mit der Nahrung aufgenommen wird), dann sind auf einmal alle genomischen Regionen, die ausschließlich Informationen für Vitamin C-Synthese enthalten, nicht mehr wichtig. Folglich evolvieren diese Abschnitte nicht mehr unter Selektion, sondern sie evolvieren neutral. Mutationen, welche die Information zerstören, werden nicht mehr ausselektiert. Daher sammeln sich im Laufe der Zeit nur in den entsprechenden Spezies mehr Änderungen in diesen Abschnitten an. Auf der Ebene der DNA-Sequenz könnte man also Regionen, die ausschließlich die Vitamin C-Information beinhalten, dadurch erkennen, dass sie mehr Mutationen in den nicht synthetisierenden Spezies angesammelt haben.

Rekonstruktion der DNA-Sequenz evolutionärer Vorfahren

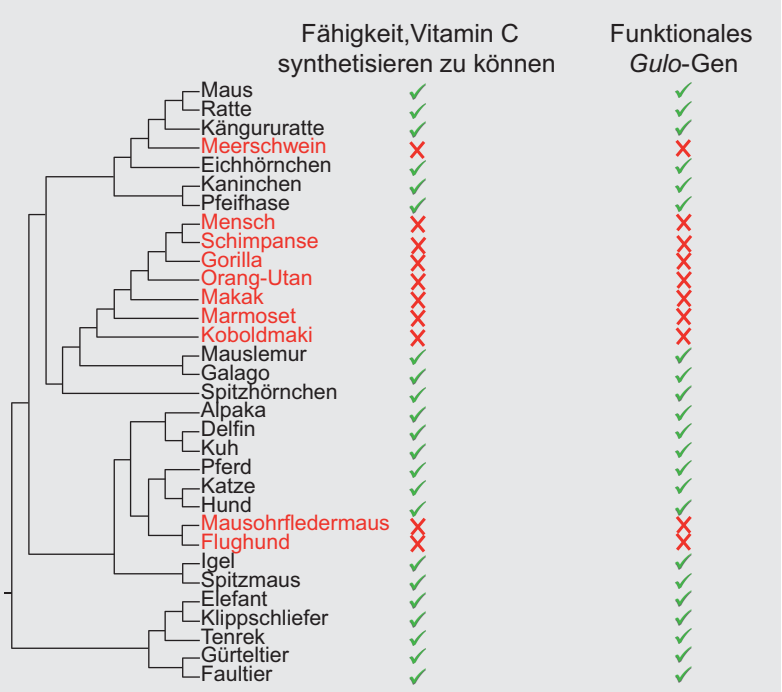
Bevor man aber in den vielen Genomen nach solchen Abschnitten suchen kann, braucht man eine Methode, welche die Anzahl der Mutationen pro Spezies und pro Abschnitt messen kann. Das ist nicht trivial, weil zum Beispiel der paarweise Vergleich einer Region von Mensch zu Schimpanse fast immer weniger Mutationen liefern wird, als der Vergleich von Mensch zu Maus, da der Mensch viel näher mit dem Schimpansen verwandt ist. Wenn man aber für einen Abschnitt die DNA-Sequenz des Vorfahren der Säugetiere kennen würde, könnte man im paarweisen Vergleich auf faire Art und Weise die Anzahl der Mutationen bestimmen. Obwohl man die Sequenz des Vorfahren niemals genau kennen

ABB. 2 | BESTIMMUNG DER ANZAHL DER MUTATIONEN IN EINEM GENOMABSCHNITT



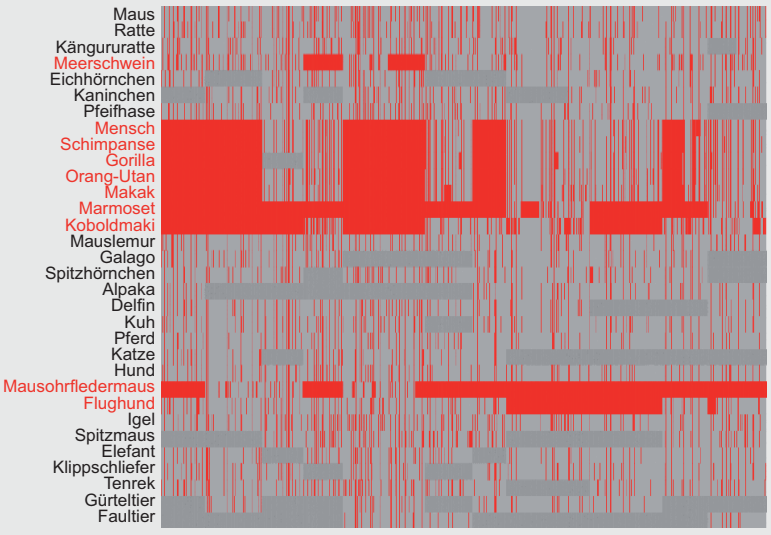
a) Für das dargestellte Sequenzalignment einer Region im Genom kann man die wahrscheinlichste Sequenz des Vorfahren zweier Spezies berechnen, wobei man Spezies 3 verwendet, um zwischen verschiedenen Möglichkeiten zu unterscheiden. Zum Beispiel kann für die erste Spalte der Vorfahre entweder ein A oder ein T haben. Wenn man nun sieht, dass Spezies 3 ein A in der ersten Spalte hat, ist das wahrscheinlichste Szenario, dass der Vorfahre ebenfalls ein A hatte und dieses A in Spezies 2 zu einem T mutiert ist. Das Fehlen von Basen aufgrund von Insertionen oder Löschungen in einigen Spezies ist als Strich „–" dargestellt.
b) Für diesen Genomabschnitt vergleichen wir die Sequenz des Vorfahren jeweils paarweise mit den Sequenzen von Spezies 1 und 2 und zählen die Anzahl der Mutationen. Dadurch können wir pro Abschnitt und pro Spezies die Anzahl der Mutationen bestimmen.

ABB. 3 | VITAMIN-C-SYNTHESE UND *GULO*-GEN



Die Fähigkeit, Vitamin C synthetisieren zu können, ist mehrfach im Laufe der Evolution verloren gegangen und korreliert perfekt mit dem Verlust des *Gulo*-Gens.

ABB. 4 | SEQUENZALIGNMENT DES *GULO*-GENS



Jeder rote Strich oder rote Bereich stellt eine Mutation oder Löschung eines Bereichs dar, die in der entsprechenden Spezies aufgetreten ist. Hellgraue Bereiche sind identisch zum Vorfahren. Dort sind keine Mutationen aufgetreten. Dunkelgraue Bereiche sind Genomabschnitte, wo die DNA-Sequenz noch nicht vollständig sequenziert wurde. Man erkennt, dass *Gulo* zwar in allen Spezies Mutationen aufweist, aber dass die Anzahl der Mutationen in den nicht Vitamin C synthetisierenden Spezies (rot gekennzeichnet) deutlich höher ist.

wird, kann man diese doch relativ gut schätzen, wenn man die Sequenzen vieler Nachkommen kennt. Die Sequenz des Säugetiervorfahren zum Beispiel kann man mit 98%iger Genauigkeit berechnen, durch so genannte „ancestral reconstruction" [2]. Dabei schätzt man für jede Position in einem Abschnitt die wahrscheinlichste DNA-Base des Vorfahrens unter Berücksichtigung eines gegebenen Wahrscheinlichkeitsmodells der Sequenzevolution (Abbildung 2). Allerdings sind die heutigen Genomsequenzen nicht absolut genau. So weisen einige Regionen erhöhte Sequenzierfehlerraten auf und andere Regionen sind einfach noch nicht sequenziert. Vorsicht ist hier geboten, denn diese Fehlerquellen sehen auf den ersten Blick genau wie Mutationen beziehungsweise Löschungen von DNA-Abschnitten aus. Wenn man also genomweit für viele Regionen die Zahl der Mutationen messen will, muss man Kenntnis über solche Artefakte haben, um für die entsprechende Spezies keine Mutationszahl zu berechnen – kein Wert ist hier besser als ein falscher Wert.

Das Vitamin C-synthetisierende Enzym

Für Vitamin C-Synthese sucht unser Ansatz also nach Regionen, die mehr Mutationen in den entsprechenden Primaten, dem Meerschwein und Fledermäusen aufweisen. Dieser Ansatz ist ähnlich zur klassischen Genetik, die für einen gegebenen phänotypischen Unterschied durch Kreuzung und Kartierung von Mutationen das entsprechende Gen findet (genannt *Forward Genetics*). Im Prinzip macht unsere Methode dasselbe, durchsucht aber ganze Genome nach Abschnitten mit diesem speziellen Muster, weshalb wir sie als *Forward Genomics* bezeichnen [3]. Angewendet auf den Vitamin C-Phänotyp und die Genome von 27 Säugetieren, finden wir eine einzige Stelle im Genom (Abbildung 3). Diese Stelle enthält das Gen *Gulo* (*gulonolactone (L-) oxidase*), dessen bereits bekannte Funktion ist, ein für die Vitamin C-Synthese verantwortliches Schlüsselenzym zu codieren. Passend dazu fanden wir, dass dieses Gen in exakt den nicht Vitamin C-synthetisierenden Spezies inaktiviert ist (Abbildung 4), obwohl Bruchstücke des Gens noch vorhanden sind. Dies ist ein klares Indiz für neutrale Evolution und zeigt, dass die genomische Information für diesen Phänotyp nicht mehr wichtig und nicht mehr vollständig vorhanden ist. Obwohl man die Funktion von *Gulo* schon biochemisch aufgeklärt hat und dieses Gen von vornherein ein guter Kandidat war, um den wiederholten Verlust von Vitamin C-Synthese in verschiedenen Säugetieren zu erklären, konnte unsere Forward Genomics-Methode dieses Gen sehr spezifisch nur mit Hilfe der vergleichenden Analyse mehrerer Genome finden.

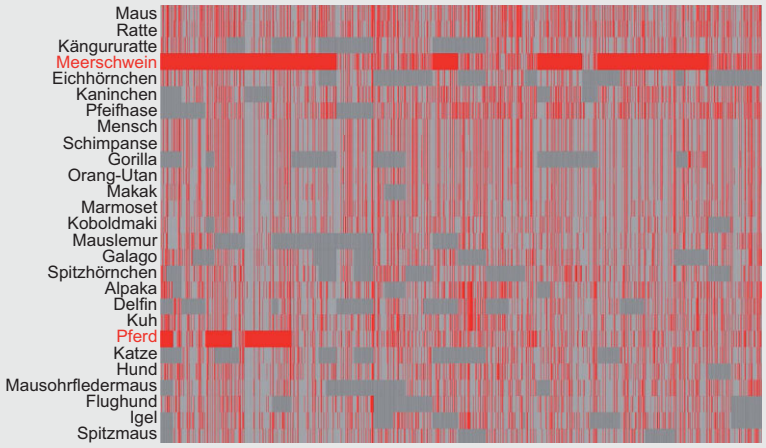
Ein Krankheitsgen, das in natürlichen Spezies fehlt

Können wir *Forward Genomics* auch auf andere phänotypische Unterschiede anwenden? Eine weitere Ver-

änderung, die wiederholt im Laufe der Evolution auftrat, ist das Vorhandensein von Phospholipiden in der Gallenflüssigkeit von Säugetieren. Während viele Spezies eine messbare Konzentration von Phospholipiden in der Galle aufweisen, ist diese Konzentration bei Meerschweinen und Pferden (zwei unabhängige Abstammungslinien in der Evolution) extrem niedrig. Wenn wir – äquivalent zu Vitamin C – mit der *Forward Genomics*-Methode nach Regionen suchen, die mehr Mutationen in diesen beiden Spezies aufweisen, dann finden wir eine kleine Anzahl solcher Regionen. Das Durchsehen der bekannten Funktionen der Gene in diesen Abschnitten zeigte, dass nicht jeder der gefundenen Abschnitte für diesen Phänotyp relevant ist. Trotzdem liefert eines der wenigen Gene, das *Abcb4* Gen, vermutlich die Erklärung für diesen phänotypischen Unterschied [3]. *Abcb4* codiert für ein Protein, das spezifisch in der Leber hergestellt wird und dort Phospholipide in den Gallensaft transportiert. Ähnlich zu *Gulo* ist *Abcb4* auch im Meerschwein und im Pferd inaktiviert, denn größere Teile der Information in diesem Gen fehlen (Abbildung 5).

Interessanterweise gibt es Mutationen im Mensch, die ebenfalls dieses Gen inaktivieren und diese Individuen haben wie Meerschwein und Pferd extrem niedrige Phospolipidkonzentrationen im Gallensaft. Aber im Unterschied zu diesen beiden Spezies leiden die betroffenen Menschen an einer Krankheit, die mit erhöhtem Risiko für Gallensteine und Zerstörung der Gallenkanälchen verbunden ist [4]. Oftmals ist eine Lebertransplantation die einzige Möglichkeit der Heilung. Es stellt sich also die Frage, warum zwei natürlich vorkommende Spezies nicht an diesen schädlichen Folgen leiden, obwohl sie dasselbe Gen schon vor mehreren Millionen Jahren verloren haben. Vermutlich gab es andere phänotypische und genetische Änderungen in diesen beiden Spezies, die die schädlichen Folgen des *Abcb4*-Verlustes kompensieren. Wenn man herausfinden könnte, was diese kompensierenden Änderungen sind, könnte man vielleicht neue Behandlungsstrategien für die betroffenen Patienten entwickeln. Diese Idee unterscheidet sich von der traditionellen Krankheitsforschung, wo man typischerweise ein Gen in einem Modellorganismus (oftmals der Maus) gezielt inaktiviert und die Aspekte der Krankheit in der Mutante analysiert. Im Gegensatz dazu gibt es in diesem Fall schon natürlich vorkommende Spezies, bei denen das Gen inaktiviert ist. Hier könnte man untersuchen, weshalb diese Spezies keine Krankheitssymptome zeigen oder warum diese sich nicht schädlich auswirken. Während die Galle vom Pferd nicht gut untersucht ist, sind möglicherweise ein paar relevante Aspekte beim Meerschwein schon bekannt. Die Gallensalze, welche für die Fettverdauung wichtig sind und die die Gallenkanälchen zerstören, falls Phospholipide in der Galle fehlen, sind im Meerschwein weniger hydrophob. Die Gallensalze haben daher vermutlich ein weniger schädliches Potenzial als in anderen Spezies wie zum Beispiel im Menschen [5].

ABB. 5 | SEQUENZALIGNMENT DES *ABCB4*-GENS

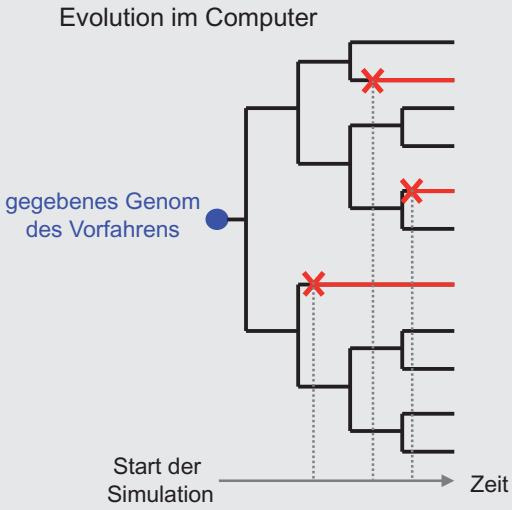


Analog zu *Gulo* in Abb. 4, weist *Abcb4* Mutationen in allen Spezies auf, aber im Meerschwein und Pferd, wo die Phospholipid-Konzentration in der Galle extrem niedrig ist, sind mehr Mutationen aufgetreten und größere Bereiche des Gens fehlen.

Evolution im Computer

Simulationsstudien sind nützlich um zu untersuchen, wie gut *Forward Genomics* auf andere unabhängig voneinander verlorene Phänotypen anwendbar ist. Im Computer kann man simulieren, wie sich ein Genom im Laufe der Evolution ändert, wobei man realistische Wahrscheinlichkeitsmodelle für Mutationen verwendet, neutrale Evolution und Selektion berücksichtigt sowie gut verstandene Prinzipien der Sequenzevolution von Genen und anderen funktionellen Genombereichen einschließt. Natürlich vereinfachen diese Simulationen eine in Wirklichkeit viel komplexere Evolution, aber sie sind nützlich, um *Forward Genomics* zu bewerten und zu optimieren, da man viele verschiedene Szenarien simulieren und störende Faktoren genau kontrollieren kann. Wir starten diese Simulation mit einem Teil des heutigen menschlichen Genoms, das wir als Modell für das Genom des Säugetiervorfahrens nehmen. Wir kennen das Genom des Vorfahrens also in dieser Simulation. Dann evolvieren wir das Genom des Vorfahrens im Computer, wobei wir die Phylogenie der Säugetiere nachstellen. Den unabhängigen Verlust eines Phänotyps können wir simulieren, indem wir zufällig ein Gen heraussuchen und dieses neutral in den Spezies evolvieren lassen, bei denen der Phänotyp fehlt. In allen anderen Spezies evolviert dieses Gen unter Selektion (Abbildung 6). Wenn alle Genome der heutigen Spezies „evolviert sind", können wir wie oben erklärt pro Genomabschnitt und Spezies die Anzahl der Mutationen bestimmen und dann nach Abschnitten suchen,

ABB. 6 | **SCHEMATISCHE DARSTELLUNG DER KÜNSTLICHEN EVOLUTION IM COMPUTER**



Wir starten mit einem bekannten Genom und simulieren die Evolution dieses Genoms in einem phylogenetischen Baum. An bestimmten Zeitpunkten simulieren wir den Verlust eines Phänotyps in voneinander unabhängigen Spezies (rote Kreuze) in dem wir ein Gen neutral evolvieren lassen (rote Kanten). Am Ende der Simulation suchen wir in den evolvierten Genomen nach den charakteristischen Mutationsmustern, die nach dem unabhängigen Verlust des Phänotyps zu erwarten sind. Natürlich kann niemand vorhersagen, wie sich ein Genom wirklich in Zukunft ändern wird.

ABB. 7 | **VIELE PHÄNOTYPEN HABEN ÄNDERUNGEN IN UNABHÄNGIGEN EVOLUTIONÄREN LINIEN**

a)

b)

	Zahl der Phänotypen	Prozent der Phänotypen mit Änderungen in unabhängigen Abstammungslinien
Primaten	166	37%
Fledermäuse	207	45%
Anglerfische	88	45%
Summe	461	42%

a) Wenn man die Phylogenie der Spezies kennt, lässt sich auch bestimmen, welche Phänotypen die gleichen Änderungen in voneinander unabhängigen Arten haben (hervorgehoben in orange).
b) Prozent dieser unabhängig-geänderten Phänotypen für drei Datensätze mit Phänotypen von Primaten, Fledermäusen und Anglerfischen (aus [3]).

die mehr Mutationen in allen Spezies haben, bei denen der Phänotyp fehlt.

Da wir das zufällig herausgesuchte Gen kennen, aber nur das typische Mutationsmuster verwenden, um relevante Abschnitte zu finden, können wir genau bewerten, wie viele richtige und falsche Abschnitte gefunden wurden. Die Simulation des Vitamin C- und Phospholipid-Phänotyps zeigte, dass wir auch in diesen simulierten Genomen recht gut die korrekten Gene für beide Phänotypen finden können. Das charakteristische Mutationsmuster, das im Laufe der Evolution entsteht, wenn ein Abschnitt in vielen Spezies unter Selektion und in anderen Spezies neutral evolviert, ist also nicht Zufall, sondern zu erwarten. Weiterhin konnten wir zeigen, dass *Forward Genomics* für eine große Anzahl von verschiedenen simulierten Szenarien unabhängiger Verluste von Phänotypen einige der korrekten Abschnitte mit großer Genauigkeit finden konnte. Das lässt hoffen, dass angewendet auf „echte" Genome für viele andere Phänotypen Abschnitte gefunden werden können, die für diese Phänotypunterschiede verantwortlich sind.

Das Auftreten derselben phänotypischen Änderungen in unabhängigen Spezies ist eine wichtige Voraussetzung für die Anwendung von *Forward Genomics*. Damit stellt sich natürlich die Frage, wieviele solcher Phänotypen (abgesehen von der Vitamin C-Synthese und Phospholipiden in der Galle) es gibt? Um die Phylogenie von Spezies zu bestimmen, werden heute standardmäßig molekulare Sequenzdaten verwendet. In der Vergangenheit hingegen verwendete man phänotypische Daten. In der Tat gibt es viele so genannte Phänotypmatrizen, die systematisch für viele Spezies (Zeilen in der Matrix) oft hunderte von phänotypischen Merkmalen (Spalten) speichern. Wenn man die Phylogenie dieser Spezies gut kennt, kann man jeden Phänotypen auf dem phylogenetischen Baum kennzeichnen und zählen, wie viele unabhängige Äste im Baum dieselben Änderungen aufweisen (Abbildung 7).

Wir haben genau dies für drei solcher Phänotypmatrizen getan und fanden heraus, dass ungefähr 40% der Phänotypen dieselben Änderungen in mindestens zwei unabhängigen Abstammungslinien aufweisen. Das bedeutet, dass unsere Methode auf etliche weitere Phänotypen angewendet werden kann, wenn die Genome dieser Spezies sequenziert sind. Dies wird vermutlich noch ein paar Jahre dauern, aber die Sequenziertechnologie entwickelt sich mit rasender Geschwindigkeit. So haben sich zwei große internationale Projekte das Ziel gesetzt, das Genom von Tausenden von Wirbeltieren und Insekten in den nächsten Jahren zu sequenzieren [6, 7]. Das liefert eine einmalige Möglichkeit, mit Hilfe der vergleichenden Genomik die traditionellen Studien von Phänotypen verschiedener Spezies mit der Molekularbiologie zusammenzubringen. *Forward Genomics* wird vermutlich nicht für jeden Phänotyp ent-

sprechende Abschnitte im Genom finden und vermutlich auch nicht alle relevanten Abschnitte finden können; dafür ist Evolution einfach zu komplex. Trotzdem zeigen unsere Simulationsversuche, dass beim systematischen Anwenden für zahlreiche Phänotypen zumindest einige der relevanten Genomabschnitte gefunden werden können. Daher werden die vergleichende Genomik und Methoden wie *Forward Genomics* bei der Erforschung hilfreich sein, wie die phänotypische Vielfalt, die wir überall in der Natur beobachten, in der DNA verschlüsselt ist.

Literatur

[1] C. L. Linster, E. van Schaftingen, Vitamin C. Biosynthesis, recycling and degradation in mammals, FEBS J. 2007, 274, 1–22.

[2] M. Blanchette, E. D. Green, W. Miller, D. Haussler, Reconstructing large regions of an ancestral mammalian genome in silico, Genome Res. 2004, 14, 2412–2423.

[3] M. Hiller, B. T. Schaar, V. B. Indjeian, D. M. Kingsley, L. R. Hagey et al., A „Forward Genomics" Approach Links Genotype to Phenotype using Independent Phenotypic Losses among Related Species, Cell reports 2012, doi: 10.1016/j.celrep.2012.08.032

[4] A. Davit-Spraul, E. Gonzales, C. Baussan, E. Jacquemin, The spectrum of liver diseases related to ABCB4 gene mutations: pathophysiology and clinical aspects, Semin Liver Dis 2010, 30, 134–146.

[5] A. F. Hofmann, L. R. Hagey, M. D. Krasowski, Bile salts of vertebrates: structural variation and possible evolutionary significance, J. Lipid Res. 2010, 51, 226–246.

[6] D. Haussler, S. O'Brien, O. Ryder, F. Barker, M. Clamp et al., Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species, J. Hered. 2009, 100, 659–674.

[7] G. E. Robinson, K. J. Hackett, M. Purcell-Miramontes, S. J. Brown, J. D. Evans et al., Creating a buzz about insect genomes, Science 2011, 331, 1386.

Zusammenfassung

Trotz der Sequenzierung der Genome vieler Spezies wissen wir nur sehr wenig darüber, welche Änderungen in der DNA für die vielen phänotypischen Unterschiede zwischen Spezies verantwortlich sind. Der Hauptgrund ist, dass es sowohl viele genomische als auch phänotypische Unterschiede gibt. Eine neue Methode der vergleichenden Genomik versucht diese Frage für Phänotypen zu beantworten, welche in unabhängigen Abstammungslinien verlorengegangen sind. Dazu sucht man nach genomischen Regionen, die in den Spezies, wo der Phänotyp fehlt, eine größere Anzahl an Mutationen haben. Die Sequenzierung von tausenden neuen Genomen in naher Zukunft wird es möglich machen, mit Hilfe der vergleichenden Genomik systematisch nach den DNA-Unterschieden zu suchen, welche mit phänotypischen Unterschieden assoziiert sind.

Summary

***Forward Genomics* – a comparative genomics approach to link phenotype to genotype**

Despite availability of several sequenced genomes, we know very little about the specific changes in the DNA that underlie phenotypic differences between species. The main reason is that species differ by both numerous genomic and phenotypic changes. A new comparative genomics method addresses this question by for phenotypes with independent evolutionary losses by searching for genomic regions that exhibit an elevated number of mutations in exactly these phenotype-loss species. The near future sequencing of thousands of novel genomes will make it possible to use comparative genomics to systematically search for such DNA changes that are associated with phenotypic differences.

Schlagworte

Vergleichende Genomanalyse, DNA Sequenz, Phänotyp, Evolution, Genetik, Krankheitsgen, Vitamin C

Der Autor

Michael Hiller, geb. 1977, studierte medizinische Informatik in Leipzig, promovierte in Bioinformatik in Jena und Freiburg, arbeitete für drei Jahre an der Stanford University in Kalifornien und ist seit 2011 Gruppenleiter in Dresden am Max Planck Institut für Physik Komplexer Systeme und am Max Planck Institut für molekulare Zellbiologie und Genetik.

Korrespondenz:
Michael Hiller
Max Planck Institute of Molecular Cell Biology and Genetics
Pfotenhauerstr. 108
D-01307 Dresden
E-Mail: hiller@mpi-cbg.de oder hiller@pks.mpg.de