

Dynamik von Genen in Populationen

Thomas Spießer

01.08.2007

Zusammenfassung

Evolution ist die Veränderung der vererbbaaren Merkmale einer Population von Lebewesen von Generation zu Generation.¹ Das Studium solcher genetischen Veränderungen innerhalb von Populationen fällt im allgemeinen in die Domäne der Populationsgenetiker.

Um molekulare Evolution tatsächlich verstehen zu können, muss ein gewisses Basisverständnis allgemeiner Prinzipien der Populationsgenetik vorhanden sein. In dieser Ausarbeitung sollen nur einige dieser Prinzipien vorgestellt werden, mit Hilfe derer man versucht folgende Fragen der Populationsgenetik zu beantworten. Wie zum Beispiel verändert sich die Frequenz eines mutierten Gens unter dem Einfluss evolutionärer Kräfte mit der Zeit, oder wie wird die genetische Variabilität innerhalb von Populationen beibehalten. Ein weiteres Problem ist die Bestimmung der Wahrscheinlichkeit, mit der sich eine neue Mutante komplett gegen seinen Wildtyp durchsetzen kann und ihn komplett ersetzt und in welcher Zeit dies geschieht. Ebenfalls von grossem Interesse für die molekulare Evolution, sind Zufallsereignisse und ihr Einfluss auf die Dynamik von Genen in Populationen.

Um eine Vorstellung davon zu bekommen, wie sich also Genfrequenzen über der Zeit verhalten, muss man vereinfachte mathematische Modelle schaffen, die einem helfen, diese sehr komplexen Prozesse zu verstehen und zu veranschaulichen. In dieser Ausarbeitung beschränke ich mich auf die Modellierung mit Differentialgleichungen, einem sog. deterministischen Modell, abgeleitet aus dem Hardy-Weinberg Gesetz. Dieses Modell soll uns ermöglichen zu verstehen, wie sich die Genfrequenzen unter dem Druck von evolutionstreibenden Kräften, wie der natürlichen Selektion, über der Zeit verhalten. Es muss hier natürlich erwähnt werden, dass diese Art der Modellierung nicht die einzige ist, und es auch noch viele weitere evolutionstreibende Kräfte neben der natürlichen Selektion auftreten, auf die in dieser Ausarbeitung aber nicht weiter eingegangen werden soll.

¹<http://de.wikipedia.org/wiki/Evolution>

1 Einleitung

Man nennt einen Genlocus (Genort) die physikalische Position eines Gens im Genom. Besteht das Genom aus mehreren Chromosomen, so ist die Position innerhalb des Chromosoms gemeint, auf dem sich das Gen befindet. Verschiedene Varianten dieses Gens werden als Allele bezeichnet, die sich alle an derselben Stelle auf dem Chromosom, nämlich dem Genort befinden. In Populationen existieren meist mehrere Allele pro Locus. Der prozentuale Anteil eines Locus, der von einem Allel innerhalb einer Population okkupiert wird, nennt sich Allelfrequenz oder auch Genfrequenz.

Wenn wir in unserem Zusammenhang über Evolution sprechen, dann meinen wir hier den Veränderungsprozess der genetischen Zusammensetzung von Populationen, wobei die grundlegendste Komponente, die Veränderung der Allelfrequenz über der Zeit ist. Betrachtet man also eine Mutation unter einem evolutionärem Gesichtspunkt, so ist sie nur bedeutsam, wenn ihre Frequenz sich mit der Zeit steigert. Erreicht die Frequenz letztendlich 1, so sagt man, hat sich die Mutation in der Population "festgesetzt". Damit die Allelfrequenz einer Mutation überhaupt steigen kann, müssen andere evolutionstreibende Kräfte, neben der Mutation selbst, ins Spiel kommen. Solche Kräfte sind u.a. die natürliche Selektion, zufälliger Gendrift, Rekombination und Migration. Ihr Effekt auf die Veränderung von Allelfrequenzen muss studiert werden, will man Evolution verstehen. Ich beschränke mich hier auf die natürliche Selektion, die als Hauptkraft der Evolution in der klassischen Evolutionslehre gilt. Der zufällige Gendrift gilt als ausgesprochen wichtig, wenn man Evolution auf der molekularen Ebene untersucht und ist das Thema meines Mit-Referenten.

Es gibt zwei mathematische Ansätze für die Modellierung von Entwicklungen von Allelfrequenzen in Populationen mit der Zeit, einen deterministischen und einen stochastischen Ansatz. Im deterministischen Modell geht man davon aus, dass die Veränderung in der Allelfrequenz in einer Population von Generation zu Generation in einzigartiger Weise vonstatten geht und deshalb auch eindeutig vorausgesagt werden kann, wenn man die anfänglichen Gegebenheiten kennt. Das stochastische Modell basiert auf der Annahme, dass Veränderungen in der Allelfrequenz in einer Population von einer Generation zu der nächsten nicht genau vorhergesagt werden können, sondern nur Wahrscheinlichkeiten für das Erreichen gewisser Allelfrequenzen bestimmt werden können. Das deterministische Modell ist das mathematisch einfachere aber auch unrealistischere Modell, welches aber oft trotzdem gute Annäherungen erzielt.

Der deterministische Ansatz soll uns im folgenden das Verhalten von Allelfrequenzen in Populationen über der Zeit und unter dem Einfluss natürlicher Selektion, näher bringen.

2 Natürliche Selektion

2.1 Evolutionstheorie nach Charles Darwin

Nach Charles Darwin's Definition, ist natürliche Selektion das Prinzip durch welches jedwede kleinste Variation eines Phänotyps konserviert wird, wenn sie dem Individuum von Nutzen ist.

Darwin beobachtete, dass sich Kinder geringfügig von ihren Eltern unterscheiden und erkannte, dass es eine Überproduktion von Kindern gibt. Den Unterschied im Phänotyp konnte Darwin nicht erklären, da Mendel'sche Regeln, Vererbungslehre und die Existenz von DNS damals noch nicht bekannt waren. Trotzdem erkannte Darwin, dass es durch die Überproduktion von Nachkommen und einer Beschränkung des Lebensraumes zu einer Konkurrenz zwischen Lebewesen kommen muss, und dass sich phänotypische Unterschiede deshalb als Vor- oder auch Nachteil für Individuen herausstellen können. Die individuelle Fähigkeit zu Überleben und Nachkommen zu produzieren, wurde "individual fitness" genannt. Je fitter ein Individuum ist, desto besser ist es an seine Umwelt angepasst und desto mehr besser angepasste Nachkommen kann es produzieren. Dieses Ungleichgewicht in der individuellen fitness bildet die Grundlage für eine Selektion geeigneterer Individuen einer Population, die unter den bestimmten Lebensbedingungen mehr Nachkommen erzeugen können, also die vorteilhaftesten Variationen besitzen. Dieser schrittweise Prozess bildet den natürlichen Auslesevorgang, Selektion, der der Population ein erfolgreicherer Überleben sichern soll.

Darwins Illustration von Schnabelvariationen von Finken auf den Galpagosin-

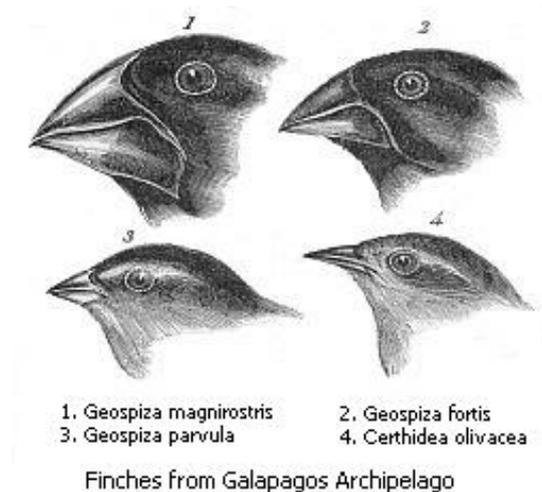


Abbildung 1: Die "Darwin Finken" auf den Galpagosinseln haben diverse Schnabelvariationen, angepasst auf ihre verschiedenen Nahrungsgewohnheiten

seln ist in Abbildung 1 zu sehen. Jeder Schnabel der verschiedenen Spezies ist speziell auf die verschiedenen Nahrungspräferenzen der Vögel angepasst. Die Schnabelform hat sich also wahrscheinlich unter dem Druck natürlicher Selektion entwickelt.

2.2 Erweiterte Evolutionstheorie (Neo-Darwinismus)

Wie bereits erwähnt, wußte Darwin nicht, wie die genetische Variabilität in Populationen zu erklären war. Der Neo-Darwinismus ist eine Erweiterung der Evolutionstheorie und berücksichtigt die modernen Erkenntnisse der Genetik, der Ökologie und Verhaltensbiologie. Sie integriert die Mendel'schen Regeln und erkennt die Mutation als Ursache für die genetische Variabilität in Populationen an.

2.3 Modellierung

Auf der Grundlage, dass wir Individuen einen Fitnesswert zuweisen können und aus der Erkenntnis, dass individuelle Fitness vom Genotyp des Organismus abhängig ist, was Darwin noch nicht wußte, lässt sich das erste einfache Modell formulieren. Hierbei ist noch zu beachten, dass Mutationen sich negativ, neutral oder positiv auf die Fitness der Träger auswirken können. Mutationen, die die Fitness reduzieren, werden in der Regel schnell aus der Population entfernt, weshalb diese Art der Selektion auch *negative oder reinigende Selektion* genannt wird. Ist eine Mutation genauso fit, wie das beste Allel in der Population, ist sie *selektiv neutral* und der Zufall entscheidet, was mit ihr geschieht. Bringt eine Mutation ihrem Träger einen selektiven Vorteil, so wird dies eine *positive Selektion* genannt.

Für unser erstes Modell nehmen wir einen Locus mit zwei Allelen, A_1 und A_2 an, dem jeweils ein negativer, positiver oder neutraler Fitness-Wert w zugeordnet wird. In einem diploiden Organismus wären die Fitness-Werte natürlich abhängig von der Interaktion der beiden Allele des Locus und die entsprechenden Genotypen A_1A_1 , A_1A_2 und A_2A_2 hätten die Fitness-Werte w_{11} , w_{12} und w_{22} . Weiter nehmen wir an, dass die Frequenz des ersten Allels p ist und die Frequenz des zweiten Allels $q = 1 - p$. Unter diesen Voraussetzungen und der Annahme, dass Partner bei der Fortpflanzung zufällig gewählt werden, gilt für die Allelfrequenzen $A_1A_1 = p^2w_{11}$, $A_1A_2 = 2pqw_{12}$ und $A_2A_2 = q^2w_{22}$. Man sagt, dass eine Population, in der das genotypische Verhältniss so bleibt, d.h. sich die Fitness-Werte nicht voneinander unterscheiden $w_{11} = w_{12} = w_{22}$, sich in einem *Hardy-Weinberg Gleichgewicht* befindet.

3 Hardy-Weinberg Gleichgewicht

Das Hardy-Weinberg Gleichgewicht ist ein theoretisches, vereinfachendes Konzept der Genetik, mit dem Allelfrequenzen modelliert oder die Entwicklung bestimmter Allelfrequenzen simuliert werden können. Es sagt, unter einer Reihe von Annahmen, den Übergang von Allelfrequenzen von einer Generation in die nächste voraus. Das Gesetz von Hardy und Weinberg geht dabei von einer idealen Population aus, d.h. eine Population, die unendlich groß ist und in der alle Genotypen im gleichen Verhältniss vorhanden und ebenfalls gleich erfolgreich bei der Produktion von Nachkommen sind. In dieser idealen Population gibt es keine Mutationen und auch keine Migration, d.h. es findet keine Evolution statt. Sind diese Voraussetzungen erfüllt, so verändern sich die Allelfrequenzen durchschnittlich über die Zeit nicht, und die Frequenzen für die nächste Generation sind $A_1A_1 = p^2$, $A_1A_2 = 2pq$ und $A_2A_2 = q^2$.

Wie gesagt, ist das Hardy-Weinberg Gesetz ein sehr vereinfachendes, theoretisches Konzept, denn weder existiert eine unendlich große Population, noch sind Populationen frei von evolutionstreibenden Kräften. Nichtsdestotrotz, bietet uns das Gesetz die Möglichkeit Frequenzverläufe zu modellieren und somit das Verhalten von Allelfrequenzen in Populationen über der Zeit in Theorie zu studieren.

4 Selektionstypen

Die eigentliche Modellierung der Veränderung der Allelfrequenzen unter Selektionsdruck erfolgt nun, indem wir den verschiedenen Genotypen verschiedene Fitness-Werte zuteilen. Ausgehend von dem oben beschriebenen Modell, sind die Allelfrequenzen, wie bereits erwähnt, für die nächste Generation p^2w_{11} für A_1A_1 , $2pqw_{12}$ für A_1A_2 und q^2w_{22} für A_2A_2 . Die Frequenz des Allels A_2 wird also in der nächsten Generation:

$$q' = \frac{pqw_{12} + q^2w_{22}}{p^2w_{11} + 2pqw_{12} + q^2w_{22}} \quad (1)$$

Das Maß der Veränderung der Allelfrequenz von A_2 pro Generation erstreckt sich also auf $\Delta q = q' - q$ und es gilt:

$$\Delta q = \frac{pq[p(w_{12} - w_{11}) + q(w_{22} - w_{12})]}{p^2w_{11} + 2pqw_{12} + q^2w_{22}} \quad (2)$$

In den folgenden zwei Beispielen wird nun davon ausgegangen, dass A_1 das originale Allel in der Population ist und durch einen Mutation in Generation 0 ein neues Allel A_2 in der Population auftaucht, welches sich in seinem Fitness-Wert von A_1 unterscheidet.

In unserem ersten Beispiel haben die beiden Homozygoten verschiedene Fitness-Werte und der Fitness-Wert der Heterozygote ist genau der Mittelwert dieser beiden. Man nennt dies einen *Kodominanten Selektionsmodus*.

Nehmen wir an, dass der Fitness-Wert für $A_1A_1 = 1$ ist und der Wert für $A_2A_2 = 1 + 2s$ ist, dann ist der Wert für $A_1A_2 = 1 + s$ und nach Gleichung 2 ist das

Maß der Veränderung der Allelfrequenz von A_2 pro Generation:

$$\Delta q = \frac{spq}{1 + 2spq + 2sq^2} \quad (3)$$

Mit Hilfe dieser Gleichung kann die Allelfrequenz von A_2 in jeder Generation bestimmt werden. In Abbildung 2 kann man sehen, wie sich die Allelfrequenz von A_2 bei einem Wert von $s = 0,01$ über einer Zeit von 1200 Generationen verhält. Es wird offensichtlich, dass sich das neue Allel A_2 nach etwa 1000 Generationen

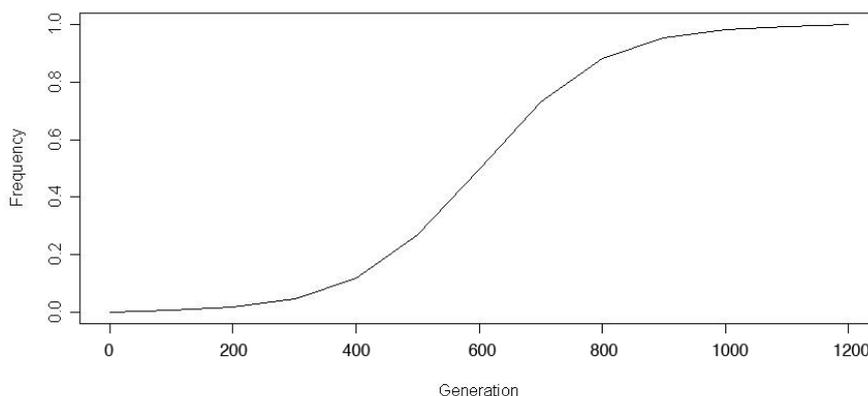


Abbildung 2: Frequenz eines kodominanten, bevorteilten Alleles mit $s = 0,01$ nach Auftauchen durch Mutation in Generation 0

durchgesetzt und sich in der Population "festgesetzt" hat. Dies ist auch nicht verwunderlich, denn A_2 hat durch die Zuweisung des höheren Fitness-Wertes einen Selektionsvorteil. Diese Art der Selektion nennt sich auch *gerichtete Selektion*.

Im zweiten Beispiel hat die Heterozygote den höchsten Fitness-Wert. Man nennt dies einen *Überdominanten Selektionsmodus*.

Wir nehmen wieder an, dass der Fitness-Wert für $A_1A_1 = 1$ ist. Der Wert für A_2A_2 soll gleich $1 + s_2$ sein und der Wert für $A_1A_2 = 1 + s_1$. Hierbei soll $s_1 > 0$ und $s_1 > s_2$ sein. Nach Gleichung 2 ist das Maß der Veränderung der Allelfrequenz von A_2 pro Generation:

$$\Delta q = \frac{-pq(2s_1q - s_2q - s_1)}{1 + 2pqs_1 + q^2s_2} \quad (4)$$

In Abbildung 3 kann man nun die Veränderung in der Allelfrequenz beobachten, die man bei *überdominanter Selektion* erhält. Man nennt diese Art der Selektion auch *balancierende* oder auch *stabilisierende Selektion*, da sich nach einiger Zeit ein stabiles Gleichgewicht zwischen den beiden Allelen einstellt und beide Allele in der Population nebeneinander existieren. In diesem Beispiel sind die Eingangsfrequenzen von oben nach unten: 0,99, 0,75, 0,25 und 0,0. Die Fitness-Werte sind besonders groß gewählt, $s_1 = 0,25$ und $s_2 = 0,125$, weswegen die

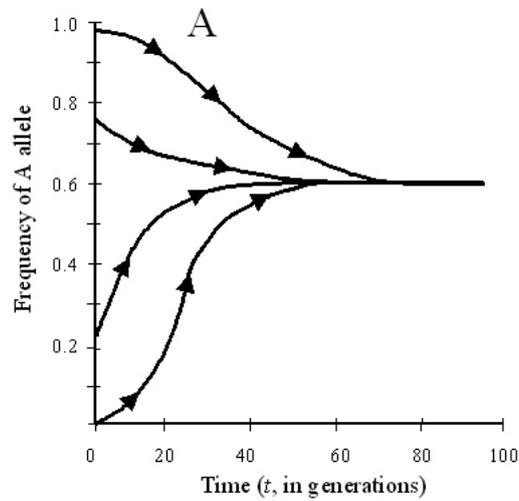


Abbildung 3: Veränderung der Allelfrequenz unter überdominantem Selektionsdruck

Veränderung in der Allelfrequenz besonders schnell vonstatten geht und sich das stabile Gleichgewicht bei $q = 0,667$ bereits nach ca. 60 Generationen einstellt.

Unser Ziel war es das Verhalten von Allelfrequenzen in Populationen über der Zeit zu untersuchen und zu verstehen. Wir haben festgestellt, dass stark vereinfachende, mathematische Modelle existieren, die es uns ermöglichen die Dynamik von Genen in Populationen zu modellieren und haben hier den Versuch unternommen, durch die Anwendung des deterministischen Ansatzes das Verhalten von Allelfrequenzen in Populationen über der Zeit und unter dem Einfluß von natürlicher Selektion zu simulieren, um es somit besser verstehen zu können. Mit Hilfe des Hardy-Weinberg Gesetzes und all seinen benötigten Einschränkungen, nur zwei Allelen und nur einer Mutation, ist es uns gelungen trotz aller Ferne zur Realität, ein Gefühl für das Verhalten von Allelfrequenzen in Populationen zu bekommen und vielleicht sind wir somit dem allgemeinen Verständniss von Evolution einen Schritt näher gekommen.

Literature

- Molecular Evolution. Wen-Hsiung Li, 1997, Chapter 2: Dynamics of genes in populations, 31-52, Sinauer Associates
- A review on the 'Nearly Neutral Theory of Evolution': Ohta (1992) The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* 23:263-86
- www.wikipedia.com