

Vererbungsmuster und Einführung in die Populationsgenetik

Peter N. Robinson

Institut für medizinische Genetik
Charité Universitätsmedizin Berlin

18. November 2014

Outline

- 1 Einführung
- 2 Vererbungsmuster
- 3 Einführung in die Populationsgenetik
- 4 Genetische Drift

Populationsgenetik

- Fachrichtung in der Biologie, die die Verteilung von Gen(bzw. Allel oder Haplotyp-)Frequenzen in biologischen Populationen studiert.
- Insbesondere wird der Einfluss auf diese Verteilung durch die folgenden vier Faktoren untersucht
 - 1 Selektion (natürliche Auslese)
 - 2 genetische Drift
 - 3 Mutation und Rekombination
 - 4 Migration bzw. Isolation
- Veränderungen innerhalb einer Population
- Variation innerhalb einer Population

Populationsgenetik

- Mathematische Modelle der Populationsgenetik bilden eine Basis für unser Verständnis der Evolution
- Ein Verständnis von genetischer Drift und der "molekularen Uhr" sind Ausgangspunkt für bioinformatische Methoden wie mehrfaches Alignment, Suche nach konservierten DNA-Sequenzen durch Genomvergleiche u.a.

Populationsgenetik

- In diesen und den folgenden zwei Vorlesungen werden wir vier Themen behandeln
 - ▶ Vererbungsmuster
 - ▶ genetische Drift
 - ▶ Hardy-Weinberg-Gleichgewicht
 - ▶ Kopplungsungleichgewicht (LD)

Outline

- 1 Einführung
- 2 Vererbungsmuster**
- 3 Einführung in die Populationsgenetik
- 4 Genetische Drift

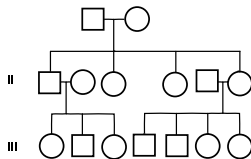
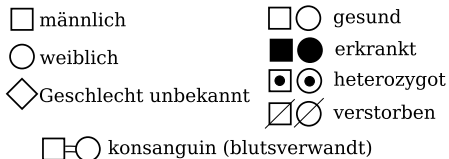
Mendel

- Phänotypische Merkmale (z.B. Körpergröße, Blutdruck u.v.m.) werden meistens von mehreren Genen und Umweltfaktoren bestimmt
- Bei mendelnden hereditären Erkrankungen ist eine Veränderung in einem Gen sowohl notwendig als auch ausreichend, damit das Krankheitsmerkmal exprimiert wird
- Über 7.000 mendelsche Krankheiten sind heute bekannt[†]

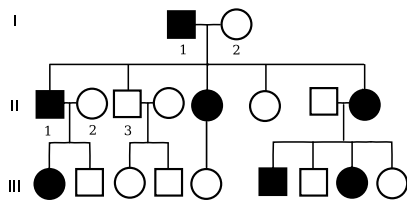


[†]<http://omim.org>

Stammbaumsymbole

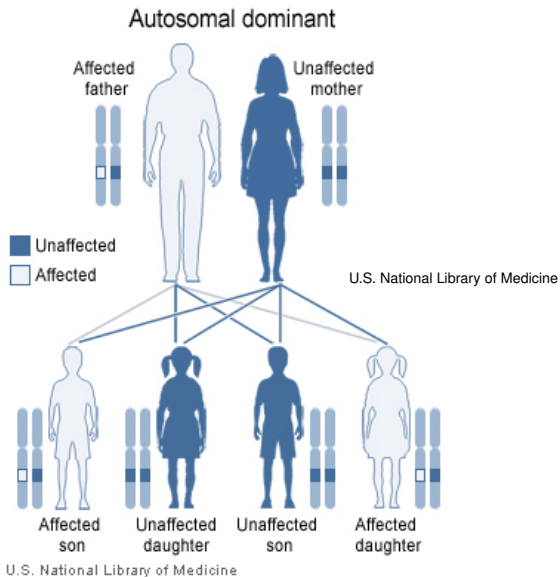


Autosomal dominant

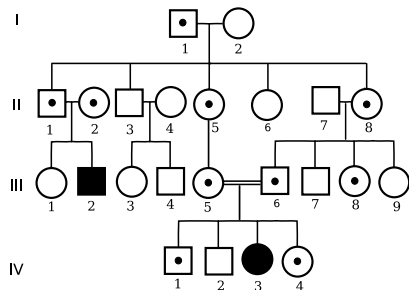


- Vertikale Transmission (Elternteil zu Kind)
- Vererbung unabhängig vom Geschlecht
- Vater-zu-Sohn-Vererbung möglich
- Merkmalsausprägung im heterozygoten Zustand
- Wiederholungswahrscheinlichkeit für Nachkommen einer betroffenen Person: 50%

Autosomal dominant

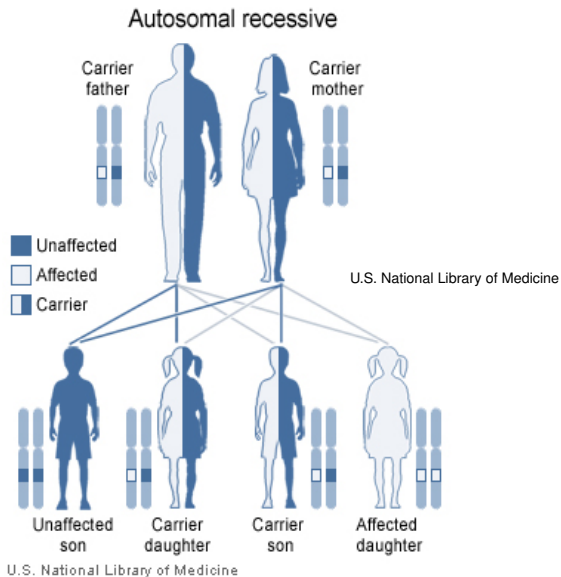


Autosomal rezessiv

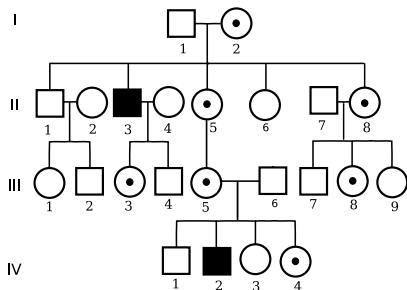


- Betroffene Personen in der Regel Kinder gesunder, heterozygoter Eltern
- Vertikale Vererbung wird nicht beobachtet (Ausnahme: Pseudodominanz)
- Vererbung unabhängig vom Geschlecht
- Risiko für Geschwister einer betroffenen Person, bzw. Nachkommen heterozygoter Eltern: 25%
- Häufiger in konsanguinen Ehen

Autosomal rezessiv

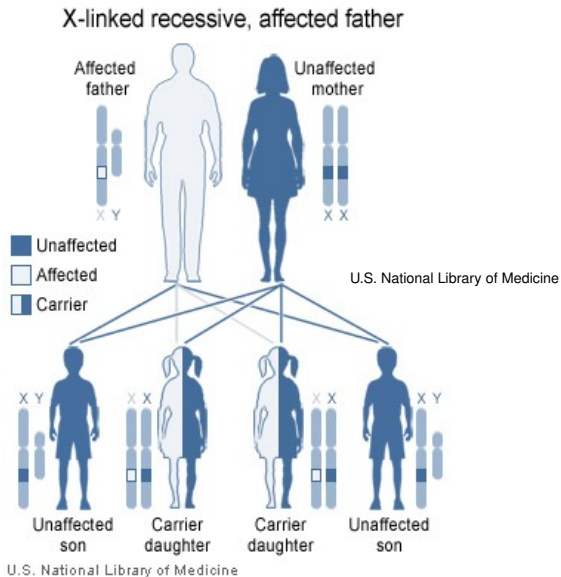


X-chromosomal rezessiv

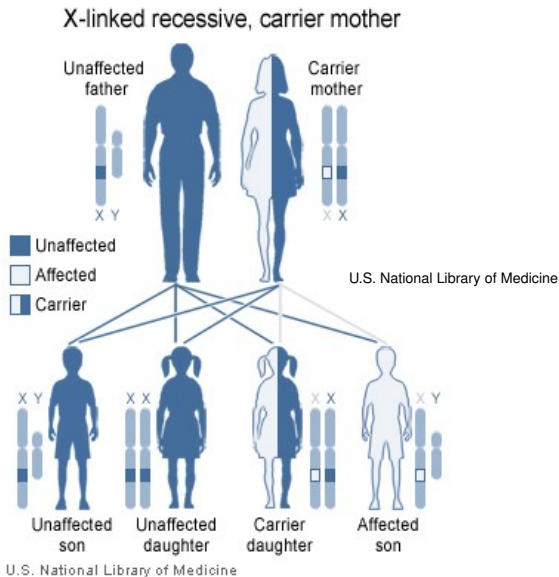


- Vater-zu-Sohn-Vererbung unmöglich
- Alle Töchter betroffener Väter erben das mutierte Gen
- Unbetroffene Männer sind nie Überträger
- Risiko für Söhne weiblicher Überträgerinnen: 50%
- Risiko für Töchter weiblicher Überträgerinnen: 50% selber Überträgerin zu sein

X-chromosomal rezessiv: Vater betroffen

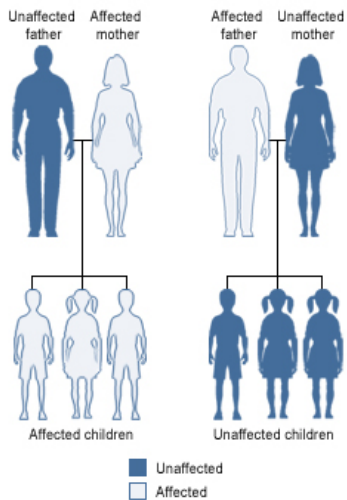


X-chromosomal rezessiv: Mutter Konduktorin



Mitochondrial

Mitochondrial



U.S. National Library of Medicine

U.S. National Library of Medicine

Weitere Erbgänge

- X-chromosomal dominant
- Y-chromosomal
- Imprinting (Vorlesung 11)
- Antizipation (Vorlesung 11)

Outline

- 1 Einführung
- 2 Vererbungsmuster
- 3 Einführung in die Populationsgenetik**
- 4 Genetische Drift

Populationsgenetik (1)

Im folgenden sollen zwei Modelle vorgestellt werden:

- Selektion (natürliche Auslese)
- genetische Drift

Beides kann die genetische Zusammensetzung einer Population verändern.

Mutation ohne Selektion

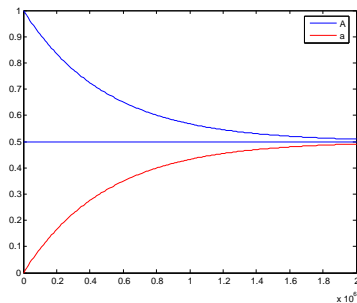
- Betrachten wir ein Mutationen zwischen zwei Allelen A und a mit Mutationsrate u und v :



- p_i (p_{i+1}): Häufigkeit von Allel A in Generation i ($i+1$)
- Die Definition der Mutationsraten impliziert

$$p_{i+1} = p_i(1 - u) + (1 - p_i)v$$

Mutation ohne Selektion



- Zwei Allele A und a , Mutationsrate $A \rightarrow a$ und $a \rightarrow A$ pro Generation 10^{-6}
- Häufigkeiten der Allele A und a gleichen sich nach 2 Millionen Generation an
- Fazit: Mutation ist kritisch für die Evolution, verändert jedoch ohne andere Kräfte die Allelfrequenz **recht langsam**

Konkurrenz

- Erst mal wird ein einfaches Modell des bakteriellen Wachstums erstellt, womit man die Konkurrenz simulieren kann
- logistisches vs. logarithmisches Wachstum

Bakterielles Wachstum

- *E. coli*: Zellteilung (2 Tochterzellen) alle 20 Minuten in logarithmischer Phase
- Zellzahl verdoppelt sich demnach alle 20 Minuten



Wikipedia commons



Einfaches Modell

- Betrachte eine Population mit x Individuen.
- Sei $r\Delta t$ die Wahrscheinlichkeit, dass ein einzelnes Bakterium in der Zeit Δt sich teilt.
- Die Veränderung in der Anzahl der Bakterien (Δx) ist dann

$$\Delta x = x \cdot r\Delta t$$

- Für $\Delta t \rightarrow 0$ erhält man die Differentialgleichung:

$$\frac{dx}{dt} = rx$$

Einfaches Modell (2)

Für $f(x) = Ce^{rx}$ gilt

$$f'(x) = rCe^{rx} = rf(x)$$

Für $x = 0$ gilt $f(0) = Ce^0 = C \cdot 1$. Wir schreiben $x_0 = f(0)$ und erhalten:

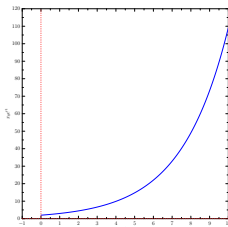
$$x = x_0 e^{rt}$$

Einfaches Modell (3)

Die Lösung dieser Differentialgleichung ist daher

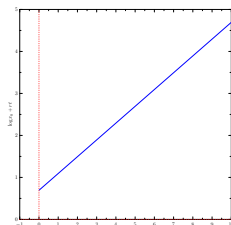
$$x = x_0 e^{rt} \quad (1)$$

wobei $x_0 = f(0)$, d.h., die Anzahl der Individuen zu Zeit $t = 0$.



Indem wir beide Seiten logarithmieren, erhalten wir

$$\log x = \log x_0 + rt \quad (2)$$

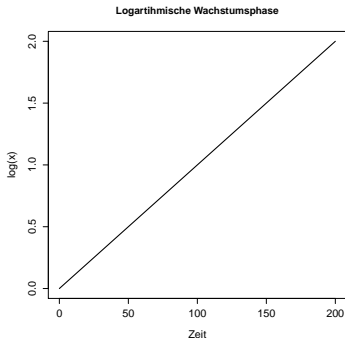


R-Skript: Logarithmisches Wachstum

```
x0 <- 1.0
r <- 0.01
t <- 0:200 ## 0,1,2,...,200
lnx <- log(x0) + r*t
plot(t,lnx, type='l',lwd=2,xlab="Zeit",ylab="log(x)",
      main="Logarithmische Wachstumsphase",cex.lab=1.25,
      cex.axis=1.25)
```

R: Logarithmisches Wachstum

- Die Befehle können jeweils einzeln auf der R-Kommandozeile eingegeben werden oder als Skript in einer Datei (z.B. `logphase.R`) gespeichert werden und durch Eingabe von `>source("logphase.R")` ausgeführt werden.
- Der `plot`-Befehl erzeugt dann die folgende Abbildung:



Logistisches Wachstum

- Logistisches Wachstum = *Wachstumsprozess mit Selbstbegrenzung*
- Nach einer logarithmischen Wachstumsphase sind die Ressourcen verbraucht.
- Intuitiv kann man sagen, dass sich die Population einer Größe, nähert, die längerfristig stabil bleiben kann

$$\frac{dx}{dt} = rx(1 - x/K) \quad (3)$$

- Die Konstante K gibt die *Kapazität* an.

Logistisches Wachstum (2)

Ist $x \ll K$ am Anfang des Wachstums, dann gilt $\frac{dx}{dt} \approx rx$, was im Einklang mit der Beobachtung steht, dass sich Populationen am Anfang i.d.R. exponentiell vermehren.

$$\frac{dx}{dt} = rx(1 - x/K) \approx rx(1 - 0) \text{ für } x \ll K \quad (4)$$

Logistisches Wachstum (3)

- Ist $x < K$, dann ist $\frac{dx}{dt} > 0$, d.h., die Population wächst.

$$\frac{dx}{dt} = rx(1 - x/K) > 0 \text{ für } x < K$$

- Ist $x > K$, d.h., die aktuelle Population liegt oberhalb der Kapazität, dann ist $\frac{dx}{dt} < 0$ und die Population verringert sich.

$$\frac{dx}{dt} = rx(1 - x/K) < 0 \text{ für } x > K$$

Logistisches Wachstum (4)

- Die Lösung der logistischen Gleichung[†] lautet:

$$x = \frac{Ce^{rt}}{1 + Ce^{rt}/K} \quad (5)$$

Die Konstante C lässt sich anhand des Wertes x_0 und der Kapazität K berechnen als

$$C = \frac{Kx_0}{K - x_0}$$

[†]Im Skript zur Vorlesungsreihe wird die Herleitung erklärt.

Logistisches Wachstum: R

```
x0 <- 1
r <- 0.01
K <- 1.5
t <- 0:750 ## 0,1,2,...,750
C <- (K*x0)/(K-x0)

## Die Funktion gilt fuer t als Skalar oder Vektor
logistic <- function(t) {
  ( C * exp(r*t)) / (1 + C*exp(r*t)/K)
}

x <- logistic(t)
## alternativ: x <- sapply(t, logistic)

plot(t,x,type='l',xlab="Zeit",ylab="Wachstum",
      main="Logistisches Wachstum",cex.lab=1.25,
      cex.axis=1.25)
```

Logistisches Wachstum: R (2)

- Die folgenden Zeilen definieren eine Funktion in R:

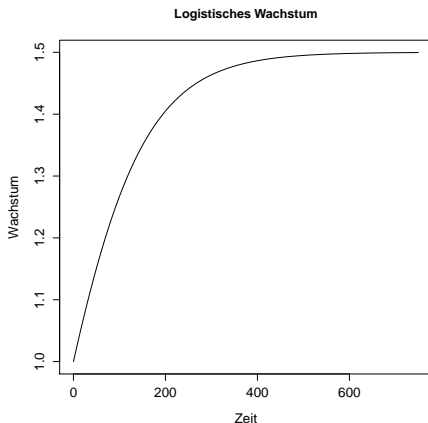
```
logistic <- function(t) {  
  ( C * exp(r*t)) / (1 + C*exp(r*t)/K)  
}
```

- `logistic` berechnet die Formel für die Lösung zur logistischen Gleichung

$$x = \frac{Ce^{rt}}{1 + Ce^{rt}/K} \quad (6)$$

- `apply(t, logistic)` : ruft die Funktion `logistic` mit jedem Element von `t` auf und gibt einen Vektor mit den jeweiligen Ergebnissen zurück

Logistisches Wachstum



- $x_0 = 1$

$$x = \frac{Ce^{rt}}{1 + Ce^{rt}/K} \quad (7)$$

Logistisches Wachstum

- für $t = 0$ haben wir[†]

$$\begin{aligned}x &= \frac{Ce^{rt}}{1 + Ce^{rt}/K} \\ &= \frac{C \cdot 1}{1 + C \cdot 1/K} \\ &\approx C\end{aligned}$$

[†] $C \ll K$, im Beispiel haben wir $C = 0,1$ und $K = 1,5$

Logistisches Wachstum

- für $t \gg 0$ haben wir

$$\begin{aligned}x &= \frac{Ce^{rt}}{1 + Ce^{rt}/K} \\ &\approx \frac{Ce^{rt}}{Ce^{rt}/K} \\ &= K\end{aligned}$$

Wann erreicht x die halbe Kapazität?

$$\begin{aligned}\frac{K}{2} &= \frac{Ce^{rt}}{1 + Ce^{rt}/K} = \frac{KCe^{rt}}{K + Ce^{rt}} \\ K + Ce^{rt} &= 2Ce^{rt} \\ K &= Ce^{rt} \\ rt &= \log K - \log C \\ t &= \frac{\log K - \log C}{r}\end{aligned}$$

in unserem Beispiel haben wir

$$t = \frac{\log 1.5 - \log 0.1074}{0.01} = 263.91 \text{Sek.}$$

Konkurrenz

- Unter welchen Umständen wird ein Organismus einen anderen von einer gegebenen Nische verdrängen?
- Veränderte Fassung von der logistischen Gleichung, wobei x und y jeweils die Anzahl von Bakterien vom Typ X und Y darstellen sollen

$$\begin{aligned}\frac{dx}{dt} &= r_1 x (1 - (x + y)/K_1) \\ \frac{dy}{dt} &= r_2 y (1 - (x + y)/K_2)\end{aligned}$$

- Falls $K_1 \neq K_2$, wird der eine Organismus den anderen komplett ersetzen.

Konkurrenz (2)

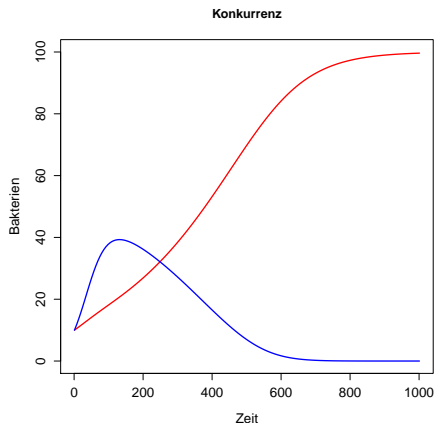


Abbildung: Konkurrenz zwischen zwei Bakterienstämme. $r_1 = 0.1$, $r_2 = 0.4$, $K_1 = 100$, $K_2 = 60$. Sowohl x und y beginnen mit jeweils 10 Individuen. y hat eine höhere Wachstumsrate als x aber eine geringere Kapazität.

Outline

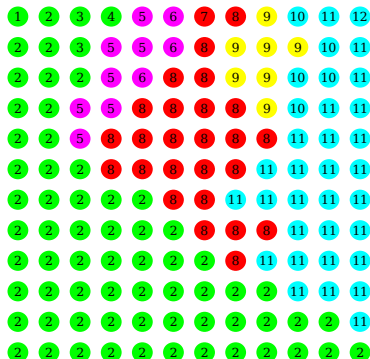
- 1 Einführung
- 2 Vererbungsmuster
- 3 Einführung in die Populationsgenetik
- 4 Genetische Drift**

Genetische Drift

- Die oben beschriebenen Modelle sind deterministisch
- Insbesondere bei kleinen Populationen kommt es zu zufälligen Veränderungen der Verteilungen der Allele.
- Solche Fluktuationen werden als genetische Drift (englisch, *drift*=sich treiben lassen).

Genetische Drift (2)

- Beispiel: 12 Individuen mit fünf unterschiedlichen Allelen
- In jeder Generation gehen manche Allele verloren (per Zufall)
- Nach 12 Generationen sind alle Individuen Nachkommen des ursprünglichen Individuums 2



Genetische Drift und natürliche Auslese

- Die Folge von genetischer Drift ist, dass es ohne das Wirken anderer Mechanismen (insbesondere ohne das Wirken der “natürlichen Auslese“) zu Schwankungen bei den Allel- und Genotyphäufigkeiten kommt.
- Sowohl Selektion wie auch Drift beeinflussen Allelhäufigkeiten in einer Population
- Im Allgemeinen ist der Einfluss der genetischen Drift umso bedeutsamer, je kleiner die Population
- Relevanz für die Medizin: z.B. hohe Inzidenz bestimmter hereditären Erkrankungen in manchen Populationsisolaten

Genetische Drift (3)

- Betrachten wir folgendes **deterministisches** Modell
 - ▶ Eine sich asexuell fortpflanzende Population mit N Individuen
 - ▶ Zwei Allele an einem Locus A und a .
 - ▶ p : Anteil von A in Population; q : Anteil von a
 - ▶ $pN + qN = N$.
 - ▶ p' und q' : Anteile von a und A in der nächsten Generation.
- Falls jedes Individuum in der Population genau einen Nachkommen hat, dann gilt $p = p'$ und $q = q'$.

Genetische Drift (4)

- Jedes der N Nachkommen in jeder Generation wird zufällig einem der N Eltern zugewiesen.
- Dann gilt eine **Wahrscheinlichkeit** von p für jedes Kind, vom Typ a zu sein bzw. q vom Typ A zu sein
- Die Wahrscheinlichkeit, dass es dann x Individuen vom Typ A in der nächsten Generation folgt der Binomialverteilung

$$P(x) = \binom{N}{x} p^x q^{N-x} = \frac{N!}{(N-x)!x!} p^x q^{N-x} \quad (8)$$

- wir können denselben Effekt durch den relativ einfachen R-Skript erreichen.

Genetische Drift - R

```
N <- 100  ## Populationsgroesse
p <- 0.5  ## Anteil v. Allel A in der Population
ngenerations <- 200

population <- sample(0:1 ,N, replace=T)

A <- vector(mode="numeric" ,N)
```

- `sample(0:1,N,replace=T)` erzeugt einen $N \times 1$ Vektor mit zufällig gewählten Elementen aus $\{0, 1\}$
- `A <- vector(mode="numeric",N)` erzeugt einen leeren, $N \times 1$ Vektor, dessen i te Element die Anzahl von A Individuen in Generation i speichern wird

Genetische Drift - R

```
for (i in 1:n generations) {  
  A[i] ← sum(population)  
  idx ← ceiling(runif(N,0,N))  
  population ← population[idx]  
}
```

- Der Vektor `population` hat N Elemente (eins für jedes Individuum in der Population). `population[i] == 1`, falls Individuum i ein A-Allel bzw. `0`, falls Individuum i ein a-Allel hat.

Genetische Drift - R

```
idx ← ceiling(runif(N,0,N))  
population ← population[idx]
```

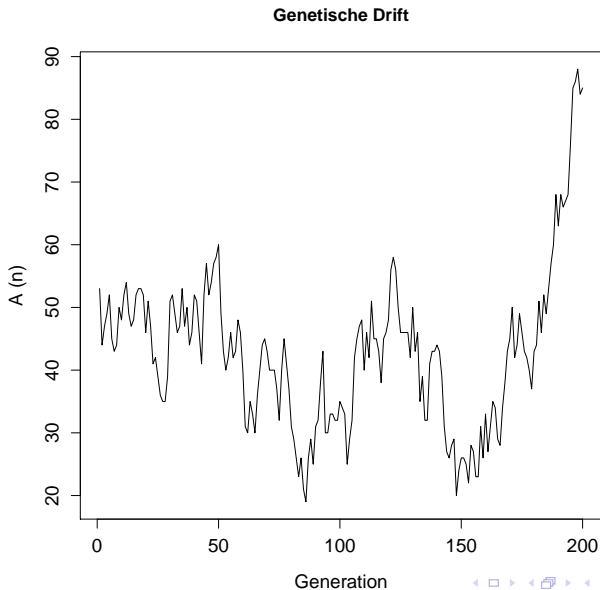
- `runif(N, 0, N)` erzeugt N Zahlen zw. 0 und N nach der Uniformverteilung
- `ceiling` (“Dach”) verwandelt sie jeweils in die nächsthöhere Ganzzahl
- Wir verwenden diese Zahlen nun als Index für die nächste Generation
- Die Wahrscheinlichkeit, dass ein Index eines Elements des Wertes “1” gezogen wird, hängt von dem relativen Anteil von A ab. Dies simuliert die genetische Drift
- Der Befehl `population ← population[idx]` “vertauscht” die Indices

Genetische Drift - R

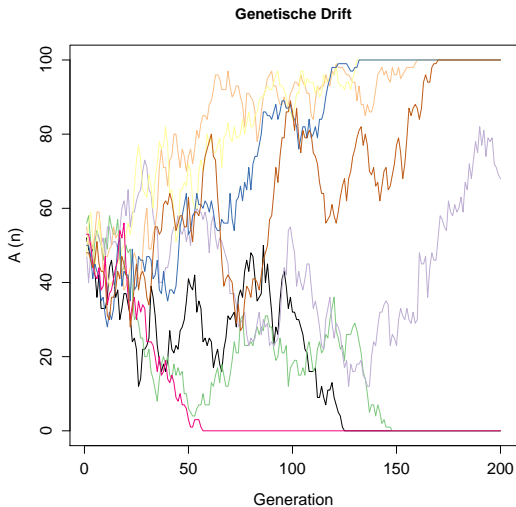
```
gen <- 1:n generations  
  
plot(gen,A,type='l',xlab="Generation",ylab="A (n)",  
      main="Genetische Drift",cex.axis=1.25,cex.lab=1.25)
```

- Nun plotten wir das Ergebnis

Genetische Drift - R



Genetische Drift - R



● Was fällt Ihnen auf?

Genetische Drift - R

- Hausaufgabe: Schreiben Sie eine Funktion, um die oben dargestellte Simulation durchzuführen, führen sie mehrmals durch und plotten Sie das Ergebnis im selben Fenster (Hinweis: Die Befehle, `plot`, `lines`, `pdf` und `dev.off()` könnten sich als nützlich erweisen).

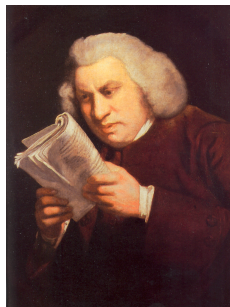
Genetische Drift: Zusammenfassung

- 1 In endlichen Populationen schwanken die Häufigkeit der Allele durch reinen Zufall, ohne das Wirken der natürlichen Auslese.
- 2 Je kleiner die Population, desto ausgeprägter sind die Schwankungen
- 3 In einer endlichen Population kommt es mit einer Wahrscheinlichkeit von 100% zur Fixierung eines Allels (Dies gilt nur für idealisierte Populationen, bei denen z.B. keine neuen Mutationen auftreten, wie in unserer Simulation).

Aufgabe: Prüfe den Einfluss Einfluss des Parameters N (Populationsgröße) auf die genetische Drift mit R (für Übung #2)

The End of the Lecture as We Know It

- Kontakt:
peter.robinson@charite.de
- Für zu Hause
 - ▶ R installieren
 - ▶ Kapitel Populationsgenetik im Skript lesen
 - ▶ R-Skripten in dieser Vorlesung ausführen, Bedeutung der Parameter prüfen (Hausaufgabe)



Lectures were once useful; but now, when all can read, and books are so numerous, lectures are unnecessary. If your attention fails, and you miss a part of a lecture, it is lost; you cannot go back as you do upon a book... People have nowadays got a strange opinion that everything should be taught by lectures. Now, I cannot see that lectures can do as much good as reading the books from which the lectures are taken. I know nothing that can be best taught by lectures, except where experiments are to be shown. You may teach chymistry by lectures. You might teach making shoes by lectures!

Samuel Johnson, quoted in Boswell's Life of Johnson (1791).