

Vererbungsmuster und Einführung in die Populationsgenetik

Peter N. Robinson

Institut für medizinische Genetik
Charité Universitätsmedizin Berlin

17. Mai 2009

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

- 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 multiples 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)

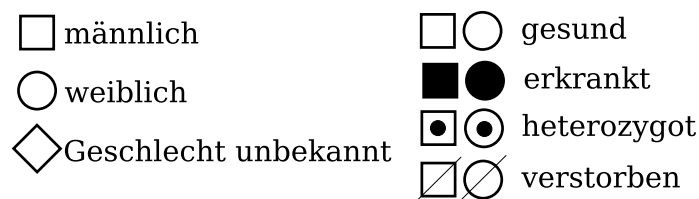
- Phänotypische Merkmale (z.B. Körpergröße, Blutdruck u.v.m.) werden meistens von mehreren Genen und Umweltfaktoren bestimmt
- Bei mendelnden hereditären Erkrankungen ist eine Veränderung in einem Gen sowohl notwendig als auch ausreichend, damit das Krankheitsmerkmal exprimiert wird
- Über 6.000 mendelsche Krankheiten sind heute bekannt[†]



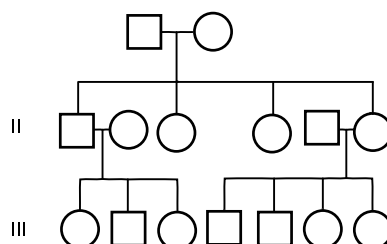
Wikipedia commons

[†]<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>

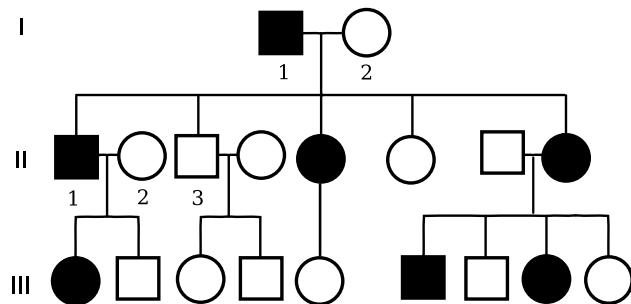
Stammbaumsymbole



□=○ konsanguin (blutsverwandt)

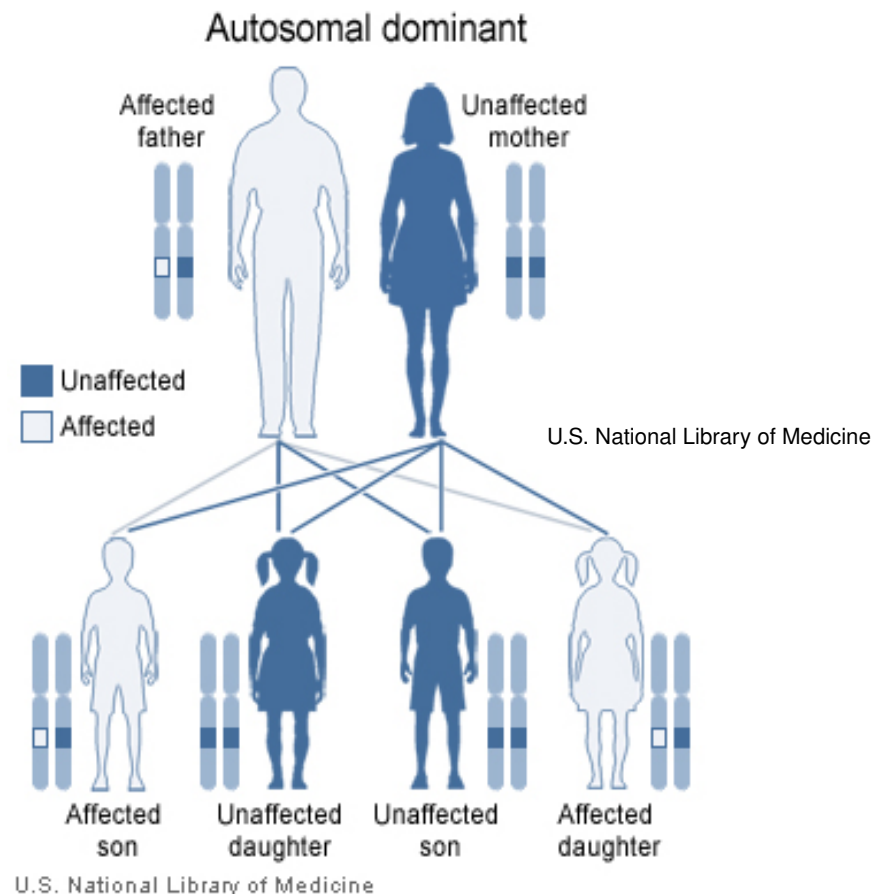


Autosomal dominant

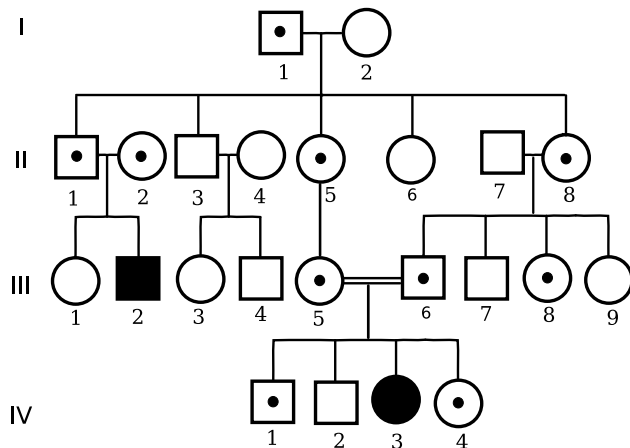


- Vertikale Transmission (Elternteil zu Kind)
- Vererbung unabhängig vom Geschlecht
- Vater-zu-Sohn-Vererbung möglich
- Merkmalsausprägung im heterozygoten Zustand
- Wiederholungswahrscheinlichkeit für Nachkommen einer betroffenen Person: 50%

Autosomal dominant

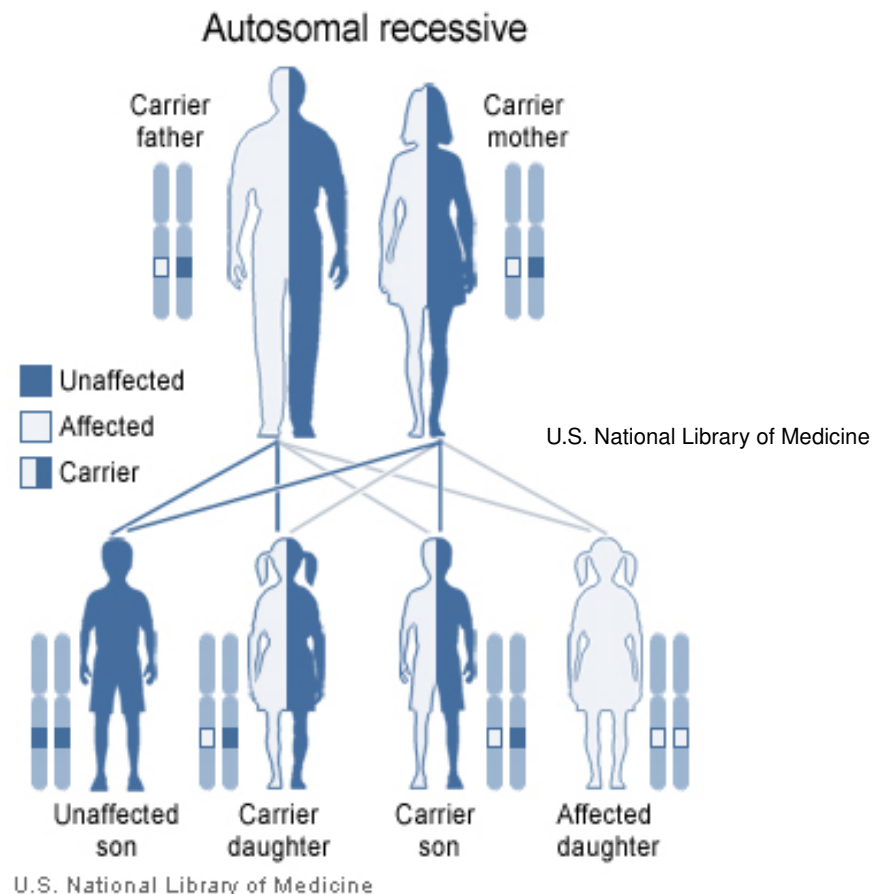


Autosomal rezessiv

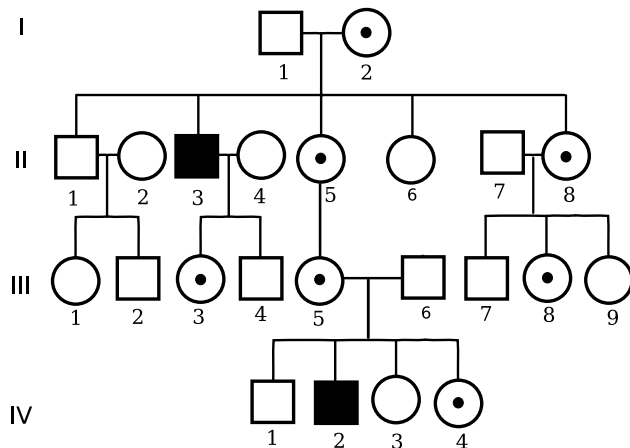


- Betroffene Personen in der Regel Kinder gesunder, heterozygoter Eltern
- Vertikale Vererbung wird nicht beobachtet (Ausnahme: Pseudodominanz)
- Vererbung unabhängig vom Geschlecht
- Risiko für Geschwister einer betroffenen Person, bzw. Nachkommen heterozygoter Eltern: 25%
- Häufiger in konsanguinen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

Autosomal rezessiv

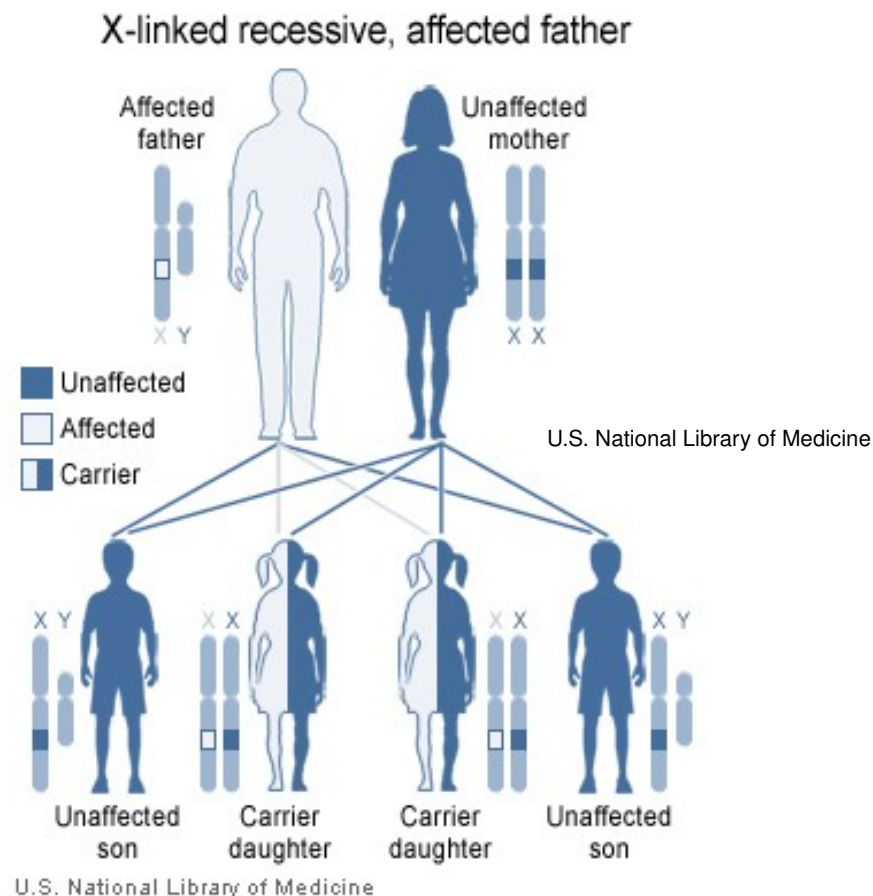


X-chromosomal rezessiv

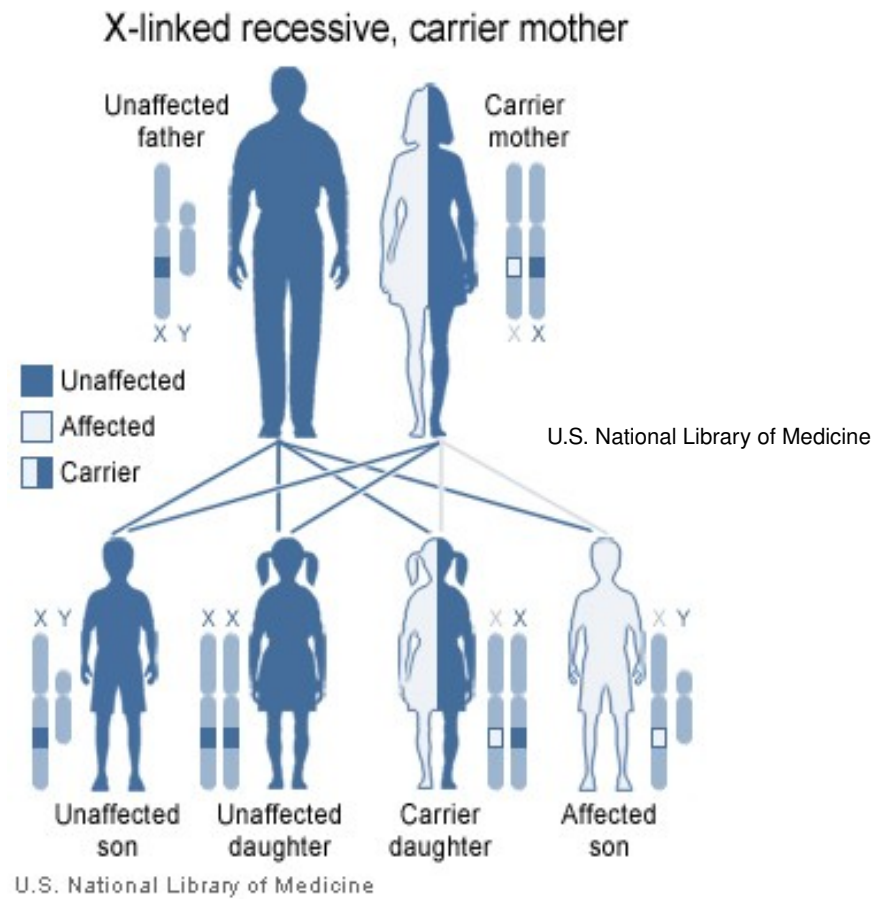


- Vater-zu-Sohn-Vererbung unmöglich
- Alle Töchter betroffener Väter erben das mutierte Gen
- Unbetroffene Männer sind nie Überträger
- Risiko für Söhne weiblicher Überträgerinnen: 50%
- Risiko für Töchter weiblicher Überträgerinnen: 50% selber Überträgerin zu sein

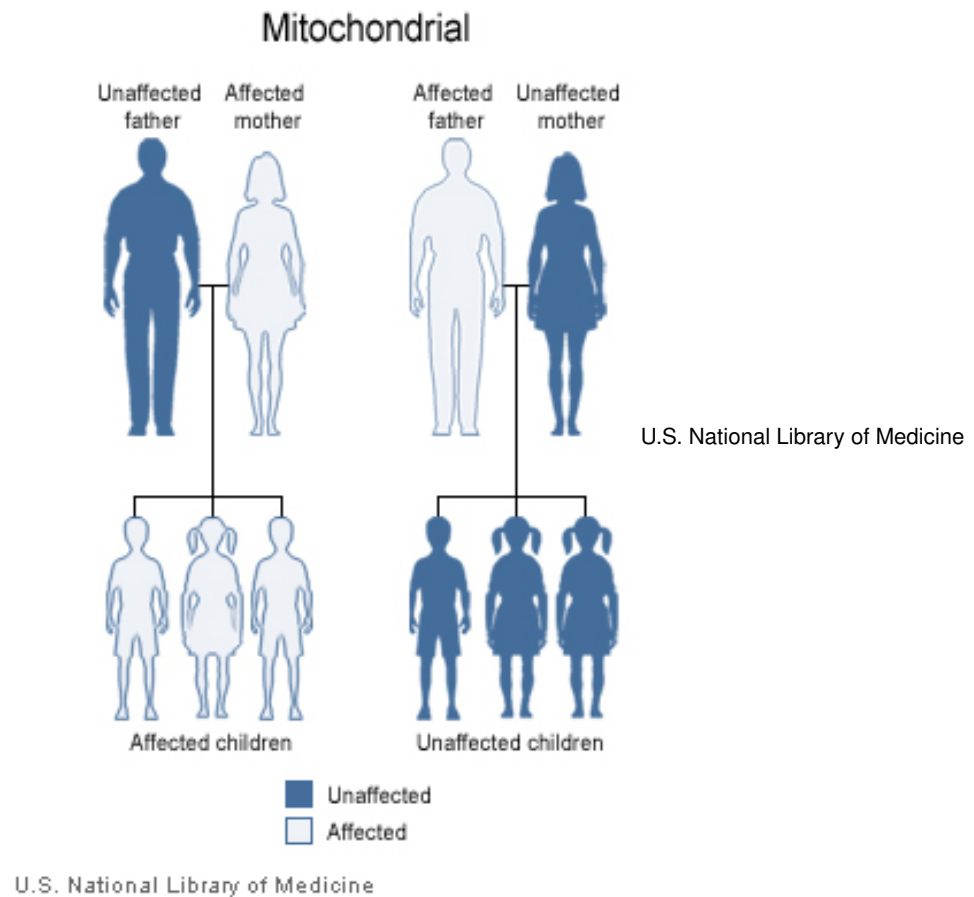
X-chromosomal rezessiv



X-chromosomal rezessiv



Mitochondrial



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

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

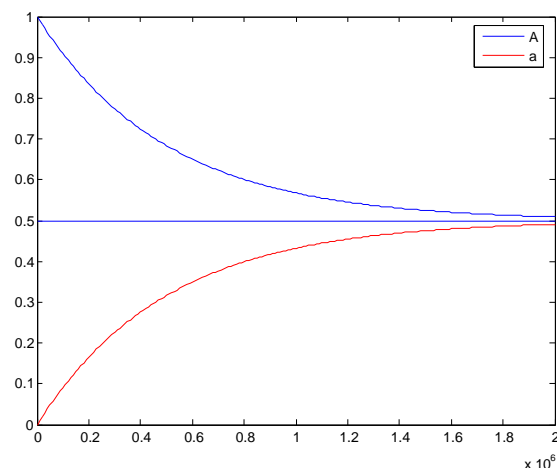
- Betrachten wir ein Mutationen zwischen zwei Allelen A und a mit Mutationsrate u und v :



- 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

- Erstmal 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 $f(0) = C \cdot 1$. Wir schreiben $x_0 = f(0)$ und erhalten:

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

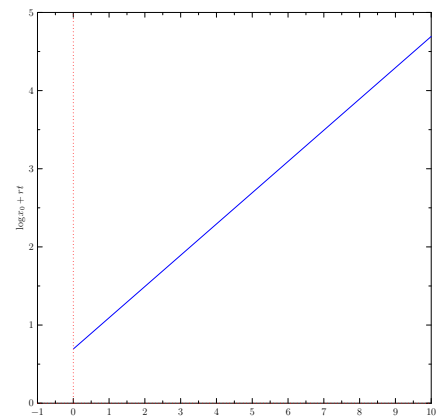
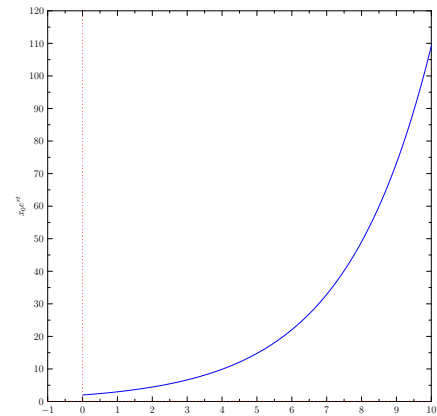
Einfaches Modell (3)

Die Lösung dieser Differentialgleichung ist daher

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

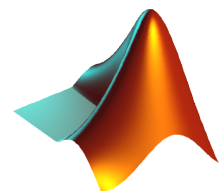
wobei $x_0 = x(0)$, d.h., die Anzahl der Individuen zu Zeit $t = 0$. Indem wir beide Seiten logarithmisieren, erhalten wir

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

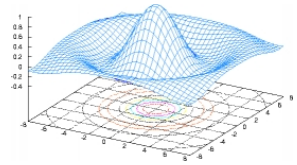
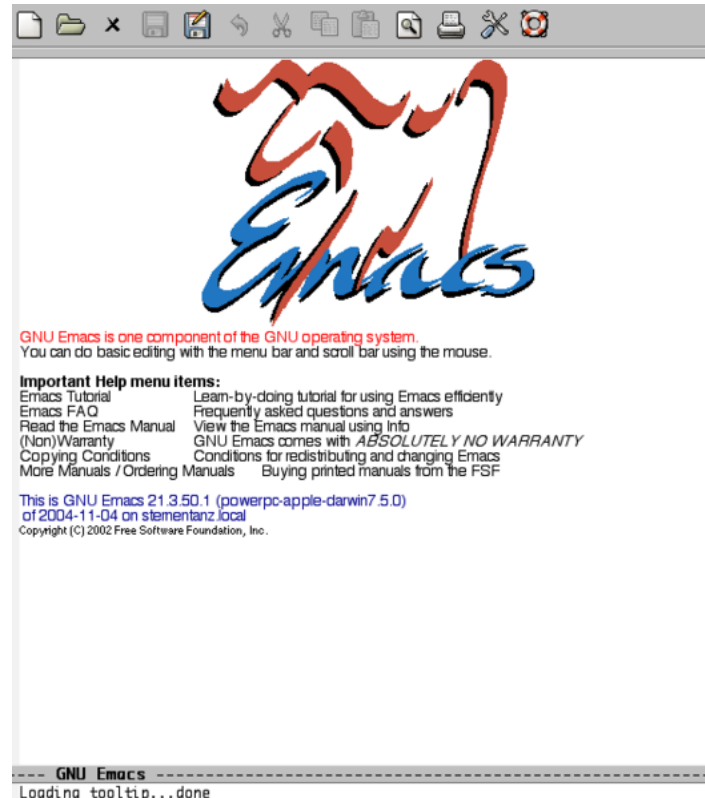


matlab

```
load predictions
predictions = sign(predictions)
%-- 4/17/08 9:42 PM --%
%-- SA
%-- 4/19/08 11:06 AM --%
%-- svmclassify
%-- 4/19/08 3:33 PM --%
%-- %w load('digit237m.mat')
%-- X = X.digit237m;
%-- Y = X.digit237m;
%-- %w load('digit237m.mat')
%-- X = X.digit237m;
%-- Y = X.digit237m;
%-- svmtrain('train', X, Y);
%-- %w load('digit237m.mat')
%-- X1 = X.digit237m;
%-- Y1 = Y.digit237m;
%-- svmtest('test', X1, Y1);
```



- matlab: **MATrix LABoratory**, aber auch eine einfache zu erlernende Umgebung zur Lösung mathematischer Probleme und zur grafischen Darstellung der Ergebnisse



- Open-source/Gnu matlab-ähnliche Sprache
- <http://www.gnu.org/software/octave/>

matlab/Octave-Skripten

- Einführung in den Vorlesungen, bei Fragen mail an peter.robinson@charite.de
- Skripten für die Vorlesungen können von meiner AG-Homepage heruntergeladen werden: <http://compbio.charite.de>
- Die Skripten können/sollen verwendet werden, um die Algorithmen besser zu verstehen bzw. den Einfluss unterschiedlicher Parameterwerte zu prüfen.

matlab/Octave-Skript: Logarithmisches Wachstum

```
x0=1.0;
r=0.01;
t=linspace(0,200,10); % 10 linear verteilte Punkte zw. 0 und 200
lnx = log(x0)+r*t;    % Vektoroperation, alles auf einmal!
plot(t,lnx,'ro-');    % plot x,y in rot ('r'), als Kreise ('o')
                        % verbunden mit Linie ('-')
axis([0 300 0 3]);    % Achsen einstellen [xmin xmax,ymin ymax]
xlabel('Zeit');        % Beschriftung der X-Achse
ylabel('Wachstum')
title('Logarithmische Wachstumsphase');
```

matlab/Octave-Skript: Logarithmisches Wachstum

- Die Kommandos können eins nach dem anderen in matlab eingegeben werden oder als Skript in einer Datei (z.B. `logphase.m`) gespeichert werden und durch eingabe von `>logphase` ausgeführt werden.
- Das `plot`-Kommando erzeugt dann die folgende Abbildung:

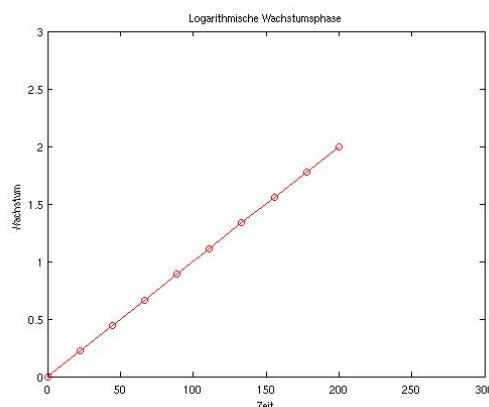


Abbildung: Logarithmische Wachstumsphase

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

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

- 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: matlab

```
x0=0.1;
r=0.01;
K=1.5;

t=linspace(0,750,100); % 100 linear verteilte Punkte
C=(K*x0)/(K-x0);
f=@(t) C.*exp(r.*t) ./ (1 + C.*exp(r.*t)./K);
x=f(t); % Vektoroperation, alles auf einmal!
plot(t,x,'ro-'); % plot x,y in rot ('r'), als Kreise ('o')
               % verbunden mit Linie ('-')
axis([0 750 0 2]); % Achsen einstellen [xmin xmax,ymin ymax]
xlabel('Zeit') % Beschriftung der X-Achse
ylabel('Wachstum') % Beschriftung der Y-Achse
title('Logistisches Wachstum');
```

Logistisches Wachstum: matlab (2)

- Der Code ist ähnlich wie im ersten Skript
- Die folgende Zeile definiert eine anonyme Funktion

```
f=@(t) C.*exp(r.*t) ./ (1 + C.*exp(r.*t)./K);
```

- f berechnet die Formel für die Lösung zur logistischen Gleichung

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

- Die Schreibweise $r.*t$ bedeutet eine elementweise Multiplikation, $./$ eine elementweise Division und ist erforderlich, um die Berechnungen für alle Elemente des Vektors t auf einmal durchzuführen.

Logistisches Wachstum

- für $t = 0$ haben wir^b

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

^b $C \ll K$, im Beispiel haben wir $C = 0,1$ und $K = 1,5$

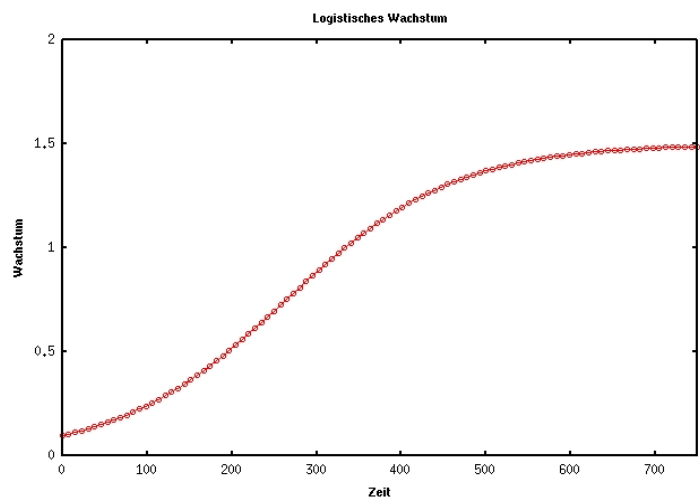


Abbildung: Logarithmische Wachstumsphase

Logistisches Wachstum

- für $t \gg 0$ haben wir

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

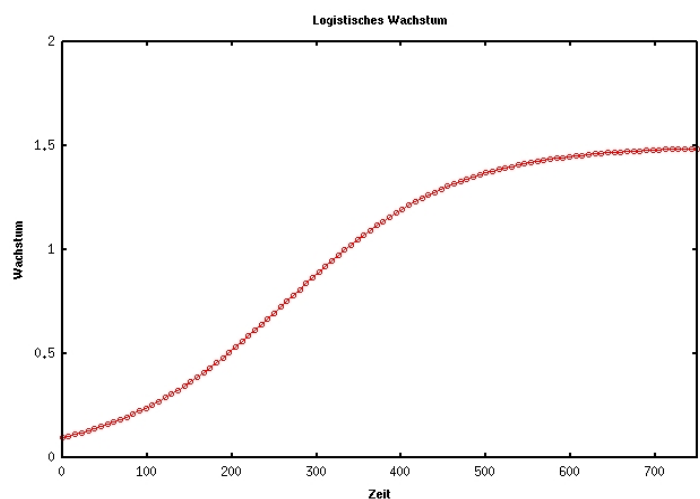


Abbildung: Logarithmische Wachstumsphase

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)

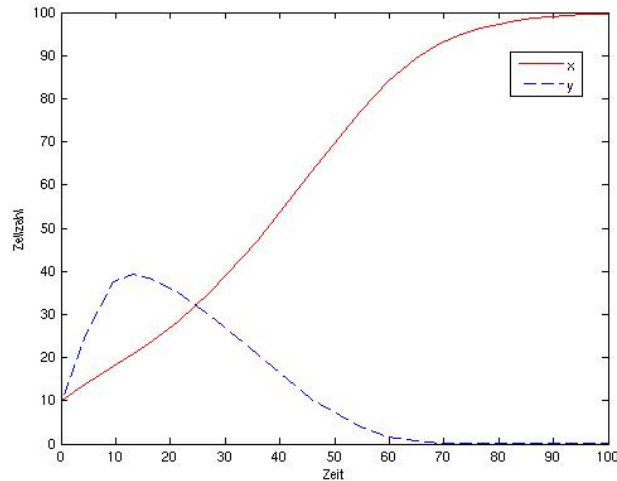


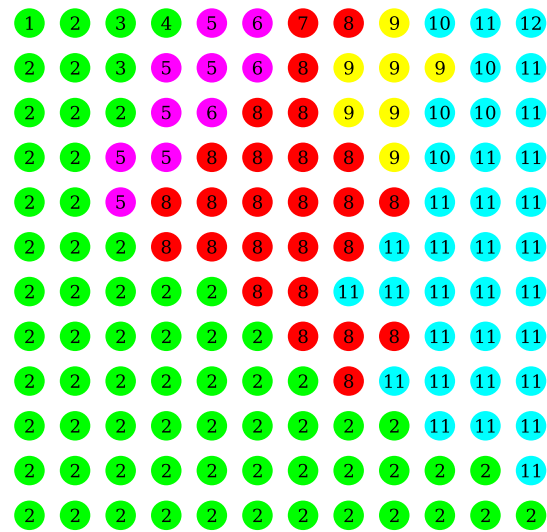
Abbildung: Konkurrenz zwischen zwei Bakterienstämme. $r_1 = 0.1$, $r_2 = 0.4$, $K_1 = 100$, $K_2 = 60$. Sowohl x und y beginnen mit jeweils 10 Individuen. y hat eine höhere Wachstumsrate als x aber eine geringere Kapazität.

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, das 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

$$P(x) = \binom{N}{x} p^x q^{N-x} = \frac{N!}{(N-x)!x!} p^x q^{N-x} \quad (7)$$

- wir können denselben Effekt durch den relativ einfachen matlab/octave-Skript erreichen.

Genetische Drift - matlab

```
N=100; % Populationsgroesse
p=0.5; % Anteil Typ A in Population
ngenerations = 200;

pop=rand(N,1)>0.5;

A=zeros(ngenerations,1);
```

- `pop=rand(N,1)>0.5;` erzeugt einen $N \times 1$ Vektor mit zufällig gewählten Elementen aus $\{0, 1\}$
- `A=zeros(ngenerations,1);` erzeugt einen $ngenerations \times 1$ Vektor, dessen i te Element die Anzahl von A Individuen in Generation i speichern wird

Genetische Drift - matlab

```
for i=1:ngenerations
    A(i) = sum(pop);
    ind=ceil(N*rand(N,1));
    pop=pop(ind);
end
```

- Der Vektor `pop` hat N Elemente (eins für jedes Individuum in der Bevölkerung). `pop(i)` ist 1, falls Individuum i ein A-Allel bzw. 0, falls Individuum i ein a-Allel hat.
- Das Kommando `sum(pop)` berechnet die Summe aller Elemente von `pop`, was in diesem Fall dasselbe ist wie die Anzahl der A-Allele ist.

Genetische Drift - matlab

```
for i=1:ngenerations
    A(i) = sum(pop);
    ind=ceil(N*rand(N,1));
    pop=pop(ind);
end
```

- $N * \text{rand}(N,1)$ multipliziert die Zufallszahlen $\in (0, 1)$ mal N , was Zahlen $\in (0, N)$ erzeugt. Das `ceil`-Kommando erhöht jede reelle Zahl auf die nächstgrößere Ganzzahl. Zum Beispiel

$$\boxed{\text{rand}(3,1)} = \begin{bmatrix} 0.039425 \\ 0.762976 \\ 0.432788 \end{bmatrix} \Rightarrow \boxed{3 * \text{rand}(3,1)} = \begin{bmatrix} 0.11827 \\ 2.28893 \\ 1.29837 \end{bmatrix}$$
$$\Rightarrow \boxed{\text{ceil}(3 * \text{rand}(3,1))} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

Genetische Drift - matlab

```
for i=1:ngenerations
    A(i) = sum(pop);
    ind=ceil(N*rand(N,1));
    pop=pop(ind);
end
```

- Mit `pop=pop(ind)` vertauschen wir die Indices.

Genetische Drift - matlab

```
clf; % clear current figure.  
gen = [1:ngenerations]';  
plot(gen,A,'ro-');  
  
axis([ 0 ngenerations 0 100]); % Achsen einstellen  
xlabel('Generation'); % Beschriftung der X-Achse  
ylabel('Anteil der Individuen mit Allel A');  
title('Genetische Drift');
```

- Nun plotten wir das Ergebnis

Genetische Drift - matlab

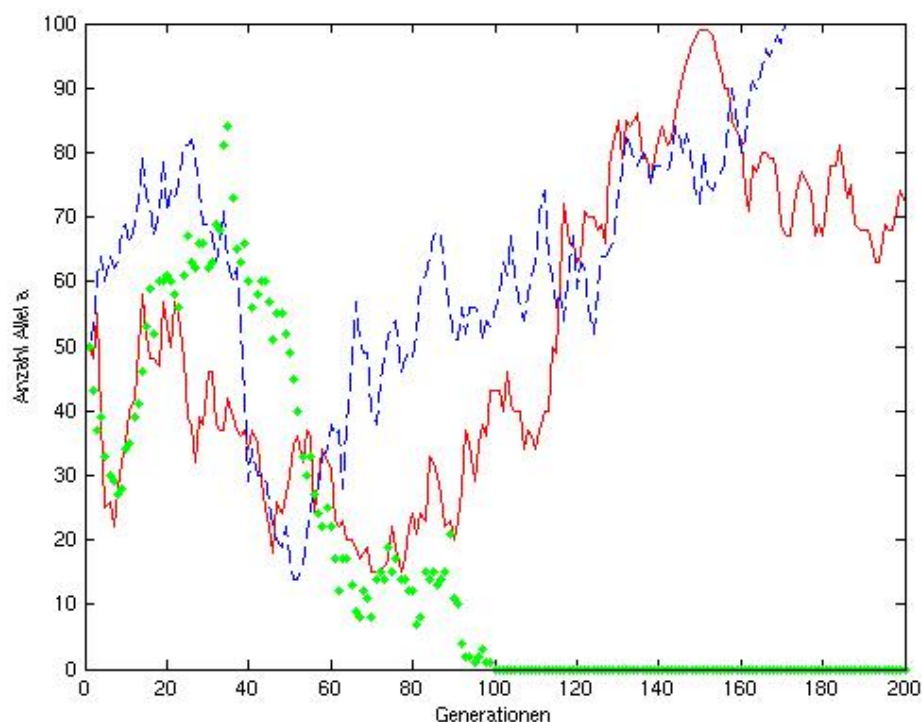
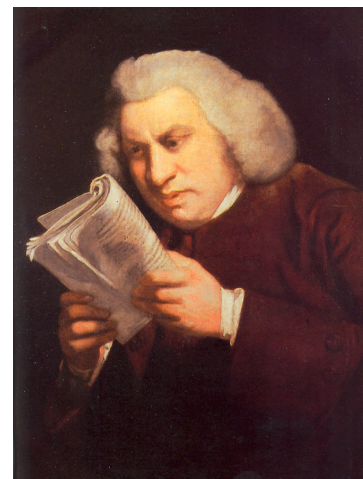


Abbildung: Genetische Drift. Drei Simulationen sind dargestellt.

- 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).

The End of the Lecture as We Know It

- Kontakt:
peter.robinson@charite.de
- Für zu Hause
 - ▶ Octave installieren
 - ▶ Kapitel Populationsgenetik im Skript lesen
 - ▶ octave-Skripten in dieser Vorlesung ausführen, Bedeutung der Parameter prüfen (Hausaufgabe für Übung 1)



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).