

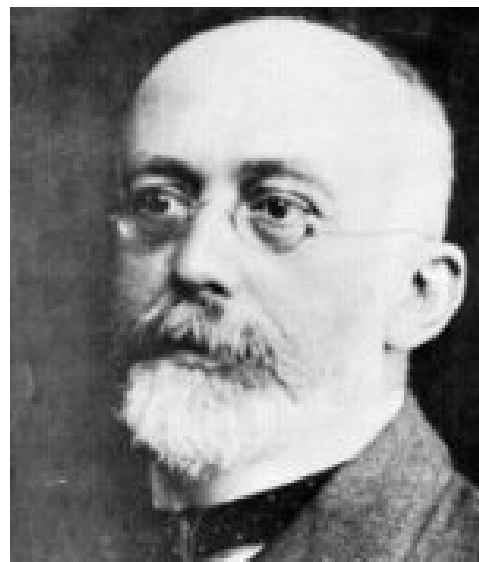
# Populationsgenetik 2: Das Hardy-Weinberg-Gesetz

Peter N. Robinson

Institut für medizinische Genetik  
Charité Universitätsmedizin Berlin

1. Juni 2008

## Hardy & Weinberg





- Godfrey Harold Hardy  
1877–1947
- Englischer Mathematiker

- Wilhelm Weinberg  
(1862-1937)
- Deutscher Arzt

# Hardy-Weinberg-Gleichgewicht

- Gegeben sei eine Population mit einem Locus mit zwei Allelen. Unter den folgenden Annahmen
  - ▶ Die Population ist unendlich groß
  - ▶ Es kommen keine Mutationen vor
  - ▶ Migrationen finden nicht statt
  - ▶ Es liegen keine Unterschiede zwischen den Genotypen in Hinblick auf Fruchtbarkeit oder Überlebensfähigkeit vor

# Hardy-Weinberg-Gleichgewicht

			
		A	a
	A	$p^2$	$pq$
	a	$pq$	$q^2$

- Die Genotyp-Frequenzen sind
  - ▶ AA:  $p^2$
  - ▶ Aa:  $2pq$
  - ▶ aa:  $q^2$

# Hardy-Weinberg-Gleichgewicht

Paarung	Häufigkeit	Nachkommen		
		AA	Aa	aa
$AA \times AA$	$P^2$	$P^2$	-	-
$AA \times Aa$	$2PH$	$PH$	$PH$	-
$AA \times aa$	$2PQ$	-	$2PQ$	-
$Aa \times Aa$	$H^2$	$H^2/4$	$H^2/2$	$H^2/4$
$Aa \times aa$	$2HQ$	-	$HQ$	$HQ$
$aa \times aa$	$Q^2$	-	-	$Q^2$
Gesamt	$(P + H + Q)^2$	$(P + H/2)^2$	$2(P + H/2)(H/2 + Q)$	$(H/2 + Q)^2$
	$= 1$	$= p^2$	$= 2pq$	$= q^2$

- Die Genotyp-Frequenzen verändern sich nicht von Generation  $i$  zu Generation  $i + 1$

## Hardy-Weinberg: Heterozygotenfrequenz in der Population

- Das Hardy-Weinberg-Gesetz kann herangezogen werden, um bei autosomal rezessiven Krankheiten die Heterozygotenfrequenz in der Bevölkerung zu berechnen
- Die Mukoviszidose (auch Zystische Fibrose, CF) ist eine der häufigsten hereditären Krankheiten in Mitteleuropa
- Erbgang: autosomal rezessiv
- Häufigkeit ca. 1:2500

# Hardy-Weinberg: Heterozygotenfrequenz in der Population

- Phänotypisch kann man die Genotypen AA und Aa nicht unterscheiden
- Definitionsgemäß haben Erkrankte den Genotyp aa

- $\Rightarrow q^2 = \frac{1}{2500}$

- $q = \sqrt{\frac{1}{2500}} = \frac{1}{50}$

AA	Aa
Aa	aa

# Hardy-Weinberg: Heterozygotenfrequenz in der Population

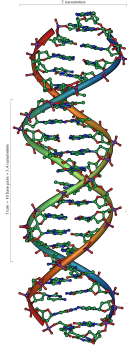
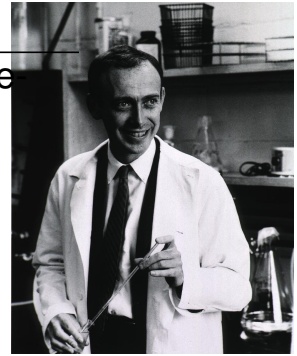
- $p + q = 1 \quad \Rightarrow p = \frac{49}{50}$
- Die Heterozygotenfrequenz beträgt

$$2pq = 2 \cdot \frac{1}{50} \cdot \frac{49}{50} \approx 2 \cdot \frac{1}{50} \cdot 1 = \frac{1}{25}$$

- Etwa 1:25 Personen mitteleuropäischer Herkunft ist heterozygoter Mutationsträger für die CF

# Das Genom von James Watson†

Gen	Chromosom	Krankheit
DPYD	1q22	Dihydropyrimidin-Dehydrogenase Mangel
PDE6B	4p16.3	Retinitis pigmentosa
GNE	9p	Myopathie
ERCC6	10q	Cockayne-Syndrom
MYO7A	11q13.5	Usher-Syndrom 1b
PFKM	12q13.3	Glykogen-Speicherkrankheit
RPGRIP1	14q11	Zapfchen-Stäbchen-Dystrophie
IL12RB1	19p13.1	Mykobakterien-Infektion
NPHS1	19q	Nephrosis
ARSA	22q	Leukodystrophie



Wikipedia commons

† Wheeler et al (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature **452**:872–876

## Fitness

### Fitness

Ein Maß für die Reproduktionswahrscheinlichkeit der verschiedenen Genotypen.

- reduzierte Überlebenswahrscheinlichkeit
- reduzierte Fruchtbarkeit

Die Fitness eines bestimmten Genotyps ist die zu erwartende Anzahl von Nachkommen in der nächsten Generation.

## Absolute Fitness $AA$ , $AB$ und $BB$

Die Fitness ist das Produkt aus der Überlebenswahrscheinlichkeit und (gegeben, dass ein Individuum überlebt) die durchschnittliche Zahl von Gameten, die an die nächste Generation weitergegeben werden.

- $AA$  hat eine Überlebenswahrscheinlichkeit von  $9/10$ . Überlebende geben durchschnittlich 5 Gameten an die nächste Generation
- $AB$  hat eine Überlebenswahrscheinlichkeit von  $8/10$ . Überlebende geben durchschnittlich 4 Gameten an die nächste Generation
- $BB$  hat eine Überlebenswahrscheinlichkeit von  $7/10$ . Überlebende geben durchschnittlich 2 Gameten an die nächste Generation

Die absolute Fitness von  $AA$  ist dann  $9/10 \times 5 = 4,5$ , die von  $AB$  ist  $8/10 \times 4 = 3,2$ , die von  $BB$  ist  $7/10 \times 2 = 1,4$ .

# Fitness

## Relative Fitness $AA$ , $AB$ und $BB$

In der Regel interessieren wir uns mehr für die relative Fitness. Bezogen auf die Fitness von  $AA$  ist die relative Fitness von  $AB$   $3,2/4,5 = 0,71$  und die relative Fitness von  $BB$   $1,4/4,5 = 0,31$ .

# Die Ausbreitung eines günstigen Allels in einer Population

Sei  $A$  das vorteilhafte Allel und  $a$  das nachteilhafte. Sei  $p_n$  die Frequenz von  $A$  in Generation  $n$ . Die Fitness von  $AA$  wird angegeben mit  $1 + s$  und die von  $aa$  mit 1. In Generation  $n$  haben wir:

Genotyp	$AA$	$Aa$	$aa$
Fitness	$1 + s$	$1 + hs$	1
Frequenz der Zygoten	$p_n^2$	$2p_nq_n$	$q_n^2$
Relative Anteil der überlebenden Erwachsenen	$p_n^2(1 + s)$	$2p_nq_n(1 + hs)$	$q_n^2$

Tabelle: Selektion.

Wir können den Effekt eines dominanten, rezessiven oder intermediären Allels mit unterschiedlichen Werten von  $h$  simulieren:

- $h = 1$ : dominant,  $h = 0$ : rezessiv,  $h = 1/2$ : intermediär

# Die Ausbreitung eines günstigen Allels in einer Population

- Die Summe dieser Einträge ist  $Z = p_n^2(1 + s) + 2p_nq_n(1 + hs) + q_n^2$ .
- $Z = [p_n^2 + 2p_nq_n + q_n^2] + p_n^2s + 2hsp_nq_n = 1 + s(p_n^2 + 2hp_nq_n)$
- Da jeder Genotyp jeweils 2 Allele hat, ist die Summe der relativen Anteile der Allele  $2Z$ .

# Die Ausbreitung eines günstigen Allels in einer Population

Die Frequenz des Allels  $A$  in der Generation  $n+1$  ist gleich der Frequenz unter überlebenden Erwachsenen von Generation  $n$ . Da jedes Individuum jeweils 2 Allele hat, und  $AA$ -Individuen zwei  $A$ -Allele beisteuern und  $Aa$  nur eins, haben wir

$$\begin{aligned}p_{n+1} &= \frac{2 \times \text{relativer Anteil } AA + 1 \times \text{relativer Anteil } Aa}{2Z} \\p_{n+1} &= \frac{2 \times p_n^2(1+s) + 1 \times 2p_nq_n(1+hs)}{2[1+s(p_n^2+2hp_nq_n)]} \\p_{n+1} &= \frac{p_n^2(1+s) + p_nq_n(1+hs)}{1+s(p_n^2+2hp_nq_n)}\end{aligned}$$

# Die Ausbreitung eines günstigen Allels in einer Population

Sei  $\delta p_n$  die Veränderung in  $p_n$  über eine Generation, d.h.,

$$\delta p_n = p_{n+1} - p_n = \frac{p_n^2(1+s) + p_nq_n(1+hs)}{1+s(p_n^2+2hp_nq_n)} - p_n \quad (1)$$



# Die Ausbreitung eines günstigen Allels in einer Population

Wir können den Anstieg der Frequenz eines vorteilhaften Allels mit folgendem matlab-Code simulieren (die Funktion setzt Gleichung 1 um):

```
function A = select(p,s,h,ngenerations)
%The following function calculates p_{n+1} given p_n
f=@(p,q,h,s) (p^2*(1+s) + p*q*(1+h*s))/(1+s*(p^2 + 2*h*p*q));

A = [p ];
for i=2:ngenerations
    q=1-p;
    p = f(p,q,h,s);
    A = [ A; p ];
end
```

# Die Ausbreitung eines günstigen Allels in einer Population

```
s=0.1; % Selektionsvorteil
ngenerations = 1200; % Anzahl der Generationen
t= [1:ngenerations]'; % Zeitpunkte fuer Plotten

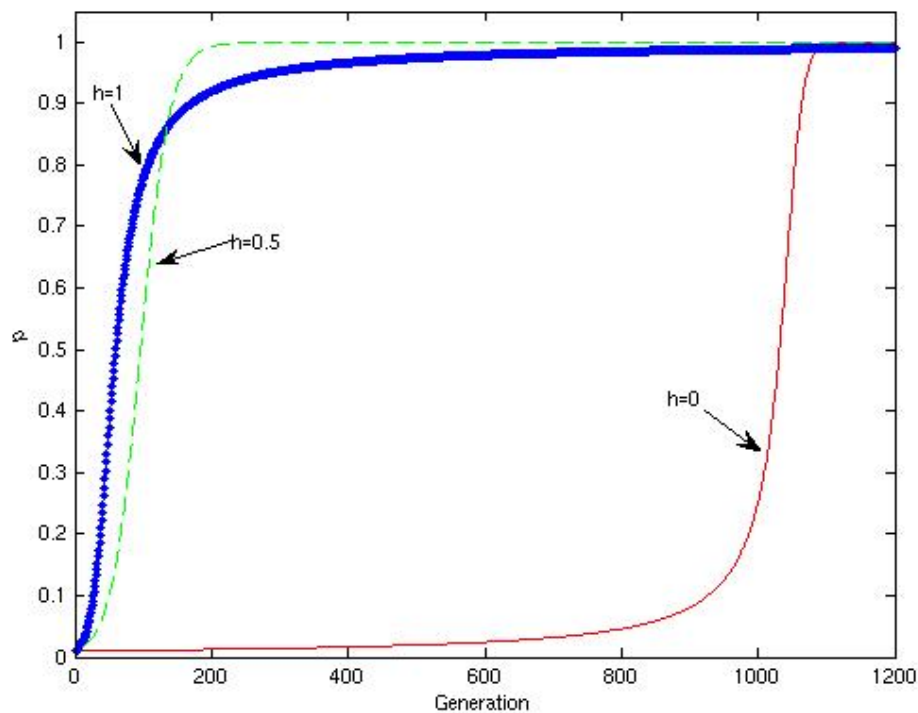
p=0.01; % Anfaengliche Frequenz von A

h=0; % rezessiv
A=select(p,s,h,ngenerations);
h=1; % dominant
B=select(p,s,h,ngenerations);
h=0.5; % intermediaer
C=select(p,s,h,ngenerations);
```

# Die Ausbreitung eines günstigen Allels in einer Population

```
clf;  
  
plot(t,A,'r-');  
hold on;  
plot(t,B,'b. ');  
plot(t,C,'g—');  
axis([0 ngenerations 0 1.05]);
```

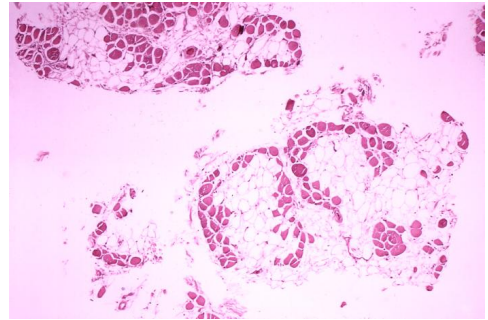
## Die Ausbreitung eines günstigen Allels



- Selektion. Die Geschwindigkeit, womit sich ein vorteilhaftes Allel in der Population ausbreitet, hängt davon ab, ob sich das Allel dominant ( $h = 1$ ), rezessiv ( $h = 0$ ) oder intermediär (z.B.  $h = 0.5$ ) auswirkt.

# Nehmen genetische bedingte Krankheiten zu oder ab?

- Beispiel Duchenne-Muskeldystrophie
- Mutationen im Dystrophingen
- Zunehmende Muskelschwäche
- Zwischen dem 7. und 12. Lebensjahr rollstuhlpflichtig
- Erhebliche Einschränkung der Lebenserwartung
- Betroffene Patienten sind i.d.R. nicht fortpflanzungsfähig



Muskelbiopsie bei Probanden mit Duchenne-Muskeldystrophie. Die Muskelfasern (rot) sind durch Fettzellen ersetzt. Wikipedia commons.

## Neumutationen

- Neumutationsrate für ausgewählte menschliche Gene/genetisch bedingte Krankheiten
  - ▶ Achondroplasie:  $0,5 - 1 \times 10^{-5}$
  - ▶ Retinoblastom:  $5 \times 10^{-6}$
  - ▶ Neurofibromatose Typ 1:  $0,5 - 1 \times 10^{-4}$
  - ▶ Duchenne-Muskeldystrophie:  $5 \times 10^{-5}$
  - ▶ ...
- Gleichgewicht zwischen Reproduktionsnachteil und Neumutationsrate

- Neumutationsrate für ausgewählte menschliche Gene/genetisch bedingte Krankheiten
  - ▶ Achondroplasie:  $0,5 - 1 \times 10^{-5}$
  - ▶ Retinoblastom:  $5 \times 10^{-6}$
  - ▶ Neurofibromatose Typ 1:  $0,5 - 1 \times 10^{-4}$
  - ▶ Duchenne-Muskeldystrophie:  $5 \times 10^{-5}$
  - ▶ ...
- Gleichgewicht zwischen Reproduktionsnachteil und Neumutationsrate

## Reproduktionsnachteil und Neumutationsrate

- Beispiel einer autosomal dominanten Erkrankung
- $p$ : Häufigkeit des mutierten, dominanten Allels
- $q \approx 1$  die Häufigkeit des normalen Allels.
- Erkrankte (Heterozygoten):  $2pq \approx 2p$
- Reproduktionsnachteil um Faktor  $s \Rightarrow$  Anteil von Nachkommen von  $Aa$ -Individuen innerhalb der Population  $2ps$
- Da nur jedes zweite Nachkommen eines  $Aa$ -Individuums das  $A$ -Allel ererbt, reduziert sich der Anteil des Allels  $A$  um  $\frac{1}{2} \times 2ps$
- Ein Gleichgewicht besteht, falls diese Abnahme durch eine Neumutationsrate  $\mu$  kompensiert wird, d.h.

$$\mu = ps \quad \Rightarrow \quad \hat{p} = \frac{\mu}{s}$$

# Autosomal rezessive Erkrankungen

Betrachten wir nun eine autosomal rezessive Erkrankung, bei der homozygote Mutationsträger schwer erkrankt sind und nicht reproduzieren (Tabelle 2).

	AA	Aa	aa	Allelfrequenz von a
Vor Selektion	$p^2$	$2pq$	$q^2$	$q$
Nach Selektion	$\frac{p^2}{p^2 + 2pq}$	$\frac{2pq}{p^2 + 2pq}$	0	?

**Tabelle:** Selektion bei einer autosomal rezessiven letalen Erkrankung.

## Häufigkeit von $q$ vor und nach Selektion

- AA- und aa-Individuen tragen keine Allele zur nächsten Generation bei
- Aa-Individuen vererben ein a-Allel an durchschnittlich jeden zweiten Nachkommen.
- Daher ist die Häufigkeit vom a-Allel in der nächsten Generation  $1/2 \times p(\text{Aa})$ .

$$\begin{aligned}
 q_{n+1} &= \frac{1}{2} \times p(\text{Aa}) \\
 &= \frac{1}{2} \times \frac{2p_n q_n}{p_n^2 + 2p_n q_n} \\
 &= \frac{q_n}{p_n + 2q_n} && \blacklozenge (p_n + q_n = 1) \\
 &= \frac{q_n}{1 + q_n}
 \end{aligned}$$

# Selektion bei autosomal rezessiven Erkrankungen

- Wir können nun fragen, welche Auswirkung die Selektion bei autosomal rezessiven Erkrankungen auf die Häufigkeit von mutierten Allelen in der Population hat.
- Nehmen wir als Beispiel die seltene autosomal rezessive Erkrankung Galaktosämie (Prävalenz in der Bevölkerung 1:40.000 Menschen, d.h.,  $q = 0,005$ ).

## Selektion bei autosomal rezessiven Erkrankungen

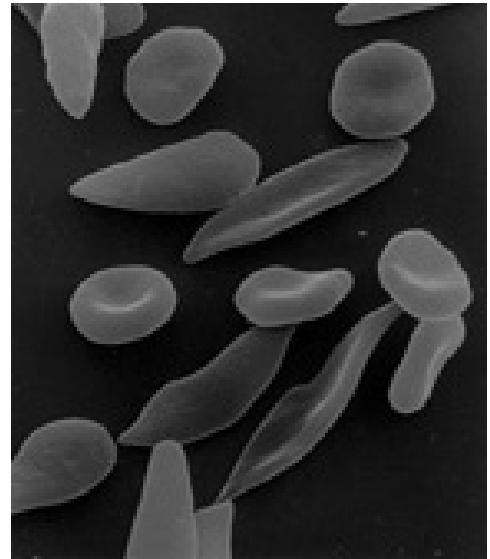
Wie viele Generationen vergehen, bis sich die Häufigkeit  $q$  halbiert hat?

```
f=@(q) q/(1+q);  
q=0.005;  
i=0;  
while q>0.0025  
    q=f(q);  
    i=i+1;  
end  
display([ num2str(i) ' Generationen '])
```

- Antwort: 200, oder ca. 6.000 Jahre

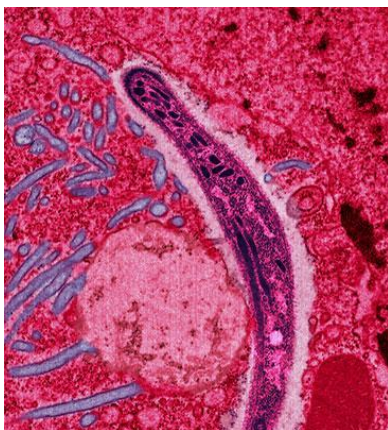
# Sichelzellanämie

- Mutation im  $\beta$ -Globingen
- Autosomal rezessiv, d.h. Erkrankte sind in Bezug auf die Mutation homozygot
- Die roten Blutzellen verformen sich bei Sauerstoffarmut zu sichelförmigen Gebilden
- Anämie
- rezidivierende Krisen/Durchblutungsstörungen mit starken Schmerzen
- Die Lebenserwartung ist vermindert



Sichelförmige rote Blutkörperchen. Wikipedia commons.

## Heterozygotenvorteil



Sichelförmige rote Blutkörperchen. Wikipedia commons.



*Anopheles gambiae*. Überträger des

Malariaparasiten *P. falciparum*. Wikipedia commons.

Rote Blutkörperchen, die in Bezug auf eine Sichelzellmutation heterozygot sind, verformen sich nach Infektion mit *P. falciparum* und werden umso schneller von der Blutbahn entfernt. Dies könnte zum Heterozygotenvorteil beitragen.

# Heterozygotenvorteil

Heterozygote Mutationsträger sind gegen Malaria resistent, weshalb die  $\beta$ -Globin-Mutation in Malariagebieten sehr verbreitet ist.

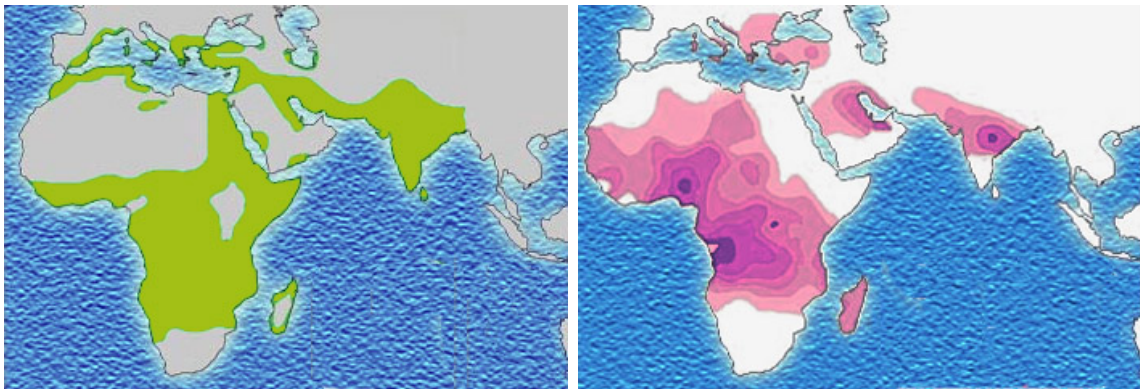


Abbildung: Links: Verbreitung der Malaria falciparum, rechts Verbreitung des Sichelzellallels. Bildquelle: Wikipedia Commons.

# Heterozygotenvorteil

- $s$  der Selektionsnachteil von Personen ohne das Sichelzellallel ( $AA$ )
- $t$  der Selektionsnachteil von homozygoten Mutationsträgern ( $aa$ ).

	$AA$	$Aa$	$aa$	Gesamt
Vor Selektion	$p^2$	$2pq$	$q^2$	1
Fitness	$1 - s$	1	$1 - t$	
Nach Selektion	$p^2(1 - s)$	$2pq$	$q^2(1 - t)$	$1 - sp^2 - tq^2$

Tabelle: Selektion zu Gunsten von Heterozygoten.



## Heterozygotenvorteil

- AA-Individuen tragen keine a-Allele zur nächsten Generation bei.
- Aa-Individuen vererben im Durchschnitt das a-Allel an jeden zweiten Nachkommen, d.h.,  $2pq \times 1/2$
- aa-Individuen tragen nach Selektion ein a-Allel an jeden Nachkommen ( $q^2(1 - t) \times 1$ ).

$$\begin{aligned}q_{n+1} &= \frac{1/2 \times 2pq + q^2(1 - t)}{(1 - sp^2 - tq^2)} \\&= \frac{pq + q^2 - tq^2}{1 - sp^2 - tq^2} \\&= \frac{q[p + q] - tq^2}{1 - sp^2 - tq^2} && \blacklozenge (p + q = 1) \\&= \frac{q - tq^2}{1 - sp^2 - tq^2}\end{aligned}$$

## Heterozygotenvorteil

- Wir können die Veränderung von einer Generation zur nächsten berechnen als<sup>†</sup>

$$\begin{aligned}\delta q_n &= q_{n+1} - q_n \\&= \frac{p_n q_n [sp_n - tq_n]}{1 - sp_n^2 - tq_n^2}\end{aligned}$$

<sup>†</sup>s. Skript, Kapitel 2, für Einzelheiten.

# Heterozygotenvorteil

Wir können fragen, ob  $q_n$  steigt oder sinkt von Generation zu Generation, d.h., welches Vorzeichen  $\delta q_n$  hat. Bei der Sichelzellanämie in einem bestimmten Malariagebiet hat man folgende Selektionsnachteile geschätzt:

- $s$  (Selektion gegen Personen ohne das Sichelzellallel  $AA$ ): 0,2039
- $t$  (Selektion gegen homozygoten Mutationsträgern,  $aa$ ): 0,8302

## Heterozygotenvorteil: matlab

Wir plotten  $\delta q_n$  gegen  $q$  für alle möglichen Werte von  $q$ .

```
s=0.2039;
t=0.8302;

f = @(p,q,s,t) p*q*(s*p-t*q)/(1-s*p^2-t*q^2);

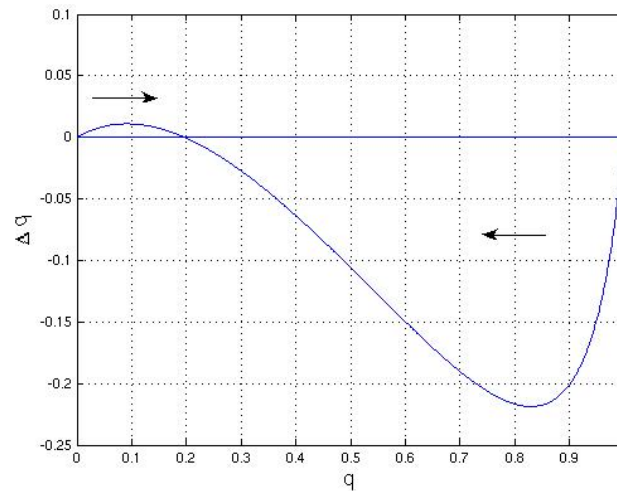
deltaq = [ ];

for q=0:0.001:1
    p=1-q;
    deltaq = [ deltaq ; f(p,q,s,t) ];
end
```

- Die anonyme Funktion  $f$  berechnet

$$f = \delta q_n = \frac{p_n q_n [s p_n - t q_n]}{1 - s p_n^2 - t q_n^2}$$

# Heterozygotenvorteil: matlab



**Abbildung:** Veränderung der Allelfrequenz  $\delta q$  für unterschiedliche Werte von  $q$  bei der Sichelzellanämie. Für  $q < 0,2$  steigt  $q$  von Generation zu Generation. Für  $q > 0,2$  reduziert sich dieser Wert. Eine Heterozygotenfrequenz von bis 31,7% wurde in einigen Populationen in Malariagebieten gefunden, was in etwa  $2pq = 2 \times 0,8 \times 0,2$  entspricht.

## Am Schluss

- Diese Vorlesungsdias stehen unter der GNU-Lizenz für freie Dokumentation (<http://www.gnu.org/licenses/fdl.txt>).
- Kontakt: [peter.robinson@charite.de](mailto:peter.robinson@charite.de)