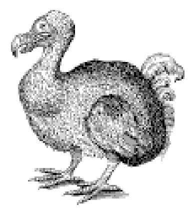
6.891: Computational Evolutionary Biology

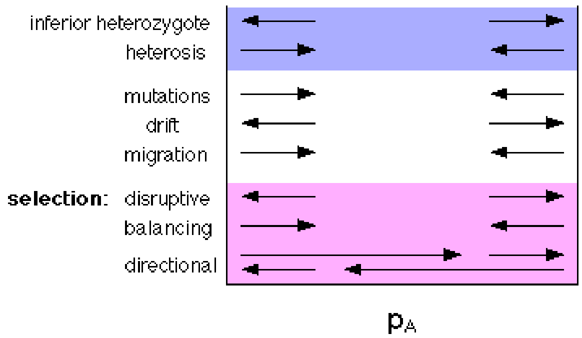
R.C. Berwick & a cast of thousands
Today: the forces of evolution



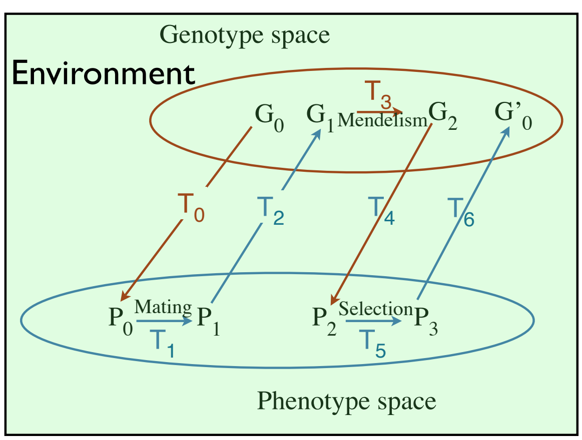
The forces of evolution, II

The deterministic model:
- *F=ma* for gene dynamics: review
- The algebra of natural selection: the lab

Why biology is not like physics: what goes off the rails - frequency dependent fitness
Does selection maximize fitness?
Does sex make you fitter?
The multivariate case: sickle cell anemia example
Change or die: the case for mutation



The dynamical system framework



The algebra of selection - J.B.S. Haldane, 1924
1 gene in 2 different forms (alleles)

genotype	AA	Aa	aa
frequency	p^2	$2pq$	q^2
relative fitness	w_{11}	w_{12}	w_{22}
after selection			

← survivors

$$\bar{w} \equiv \text{ mean fitness } \equiv p^2 w_{11} + 2p(1-p)w_{12} + (1-p)^2 w_{22}$$

$$\bar{w}_1 \equiv \text{ mean fitness of A } \equiv p^2 w_{11} + p(1-p)w_{12}$$

Algebra II

$$\bar{w} \equiv \text{ mean fitness } \equiv p^2 w_{11} + 2p(1-p)w_{12} + (1-p)^2 w_{22}$$

$$\bar{w}_1 \equiv \text{ mean fitness of A } \equiv p^2 w_{11} + p(1-p)w_{12}$$

fitness ratios (scaled):

$$\frac{p^2 w_{11}}{\bar{w}} : \frac{2pqw_{12}}{\bar{w}} : \frac{q^2 w_{22}}{\bar{w}}$$

Newton's *F=ma* for evolutionary systems
Basic dynamical system map: compute p' from p

$$p' = \frac{A \text{ survivors}}{\text{all survivors}} = \frac{p^2 w_{11} + \frac{1}{2} \times 2p(1-p)w_{12}}{p^2 w_{11} + 2p(1-p)w_{12} + (1-p)^2 w_{11}}$$

$$p' = \frac{p(pw_{11} + p(1-p)w_{12})}{p^2 w_{11} + 2p(1-p)w_{12} + (1-p)^2 w_{11}} = p\frac{\overline{w}_1}{\overline{w}}$$

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

Dynamics: Compute Δ p and also w vs. p gives the 'jet fuel' formula for gene frequency change under selection

$$p' - p = \frac{p\overline{w}_1}{\overline{w}} - \frac{p\overline{w}}{\overline{w}} \;\Rightarrow\; \Delta p = \frac{p(\overline{w}_1 - \overline{w})}{\overline{w}} \text{ with 2 alleles,}$$

$$\overline{w} = p\overline{w}_1 + (1-p)\overline{w}_2, \text{ so substituting:}$$

$$\Delta p = \frac{p(1-p)(\overline{w}_1 - \overline{w}_2)}{\overline{w}} \text{ and now substitute for } (\overline{w}_1 - \overline{w}_2) = \frac{d\overline{w}}{dp} =$$

$$\Delta p = \frac{p(1-p)}{\overline{w}}\frac{d\overline{w}}{dp} = \frac{p(1-p)}{}\frac{d\ln(\overline{w})}{dp}$$

"The company you keep"

Understanding the basic $\triangle p$ formula

Fitness of organisms in which A finds itself

$$\triangle p = p\frac{\boxed{\bar{w}_1} - \bar{w}}{\boxed{}}$$

Divided by fitness of all organisms

F=ma

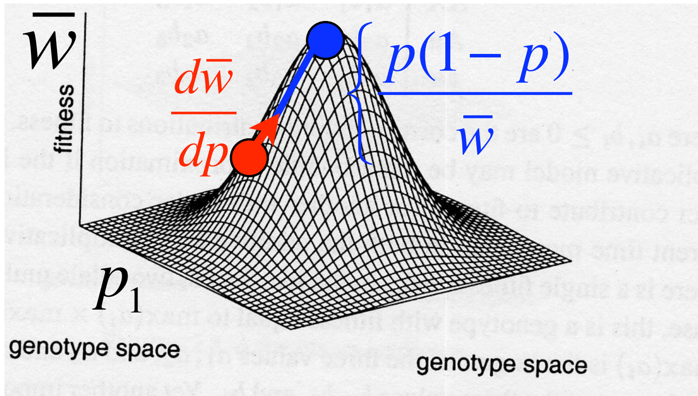
The jet fuel formula for 'evolutionary change'

$$\Delta p = \boxed{\frac{p(1-p)}{\overline{w}}}\boxed{\frac{d\overline{w}}{dp}}$$

<u>Amount</u> of change in *p* (variance component)

<u>Direction</u> of change (slope of $\overline{w}$ wrt *p*, + or –)

Natural selection works by local gradient ascent



What happens to variation under selection?

Suppose (relative) fitnesses are:

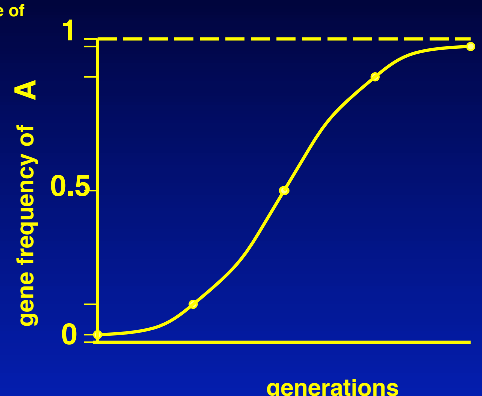
AA	Aa	aa
$(1+s)^2$	1+s	1

x (1+s) x (1+s)

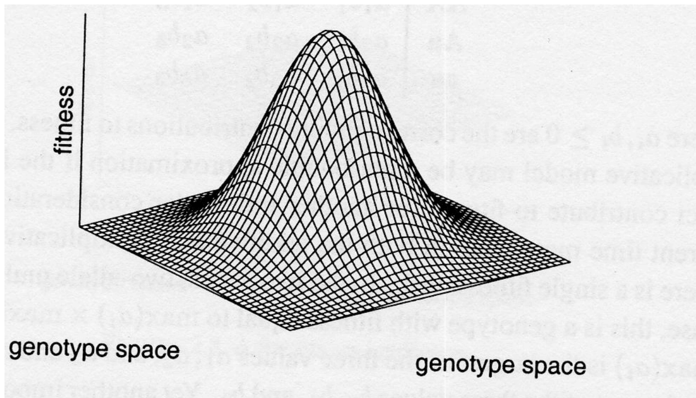
So in this example each change of a to A multiplies the fitness by (1+s), so that it increases it by a fraction s.

The time for gene frequency change, in generations, turns out to be:

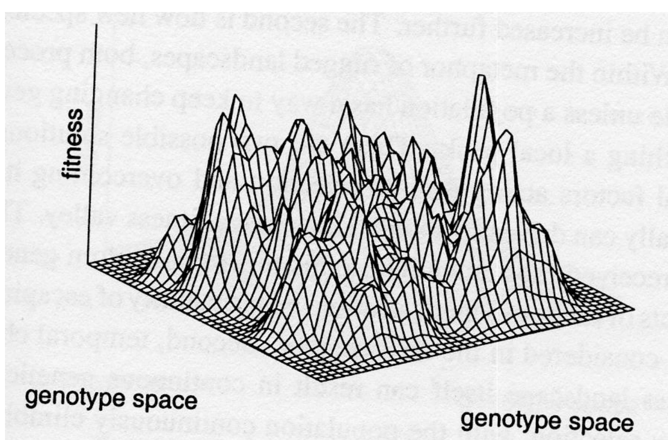
s	change of gene frequencies: 0.01 – 0.1	0.1 – 0.5	0.5 – 0.9	0.9 – 0.99
1	3.46	3.17	3.17	3.46
0.1	25.16	23.05	23.05	25.16
0.01	240.99	220.82	220.82	240.99
0.001	2399.09	2198.02	2198.02	2399.09



The shape of things now



The shape of things now



Mean fitness always increases...

$$\Delta\bar{w} = 2p(1-p)\{w_{11}p + w_{12}(1-2p) - w_{22}(1-p)\}^2$$
$$\times \{w_{11}p^2 + (w_{12} + \frac{1}{2}w_{11} + \frac{1}{2}w_{22})p(1-p) + w_{22}(1-p)^2\}\bar{w}^{-2}$$

But...this is <u>not</u> always the same thing as globally maximizing fitness...

Solving the fundamental recurrence equation

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

w_{11}	w_{12}	w_{22}	
$1+s$	$1+sh$	1	e.g.,
1	1	1	
$(1+s)$	$(1+s)$	1	"dominance"
1	$(1+s)$	$(1+s)$	"recessive"
$1+s$	$1+sh$	1	"heterozygote over/under dominant"

NB: only a few special cases have explicit solutions!

Haploid case (no Hardy-Weinberg sexual mixing)

Fitness ratios 1+*s* : 1 (for fitness A:a)
s= selection coefficient

$$\frac{p_A^{(t)}}{p_a^{(t)}} = (1+s)^t \frac{p_A^{(0)}}{p_a^{(0)}},$$

Rates of change in gene frequencies

$$\frac{p_A^{(t)}}{p_a^{(t)}} = (1+s)^t \frac{p_A^{(0)}}{p_a^{(0)}},$$

$$t = \left[\ln\left(\frac{p_A^{(t)}}{p_a^{(t)}}\right) - \ln\left(\frac{p_A^{(0)}}{p_a^{(0)}}\right)\right] / \ln(1+s).$$

s=0.01; A=0.01

$$\begin{aligned} t &= \left[\ln\left(\tfrac{0.99}{0.01}\right) - \ln\left(\tfrac{0.01}{0.99}\right)\right] / \ln(1.01) \\ &= [\ln 99 - \ln(1/99)] / \ln(1.01) \\ &= 923.6115 \text{ generations} \end{aligned}$$

Special cases of fitness - geometric ratios

$$\begin{array}{ccc} AA & Aa & aa \\ (1+s)^2 & 1+s & 1 \end{array}$$

$$\begin{aligned} \bar{w}_A &= (1+s)^2 p + (1+s)(1-p) \\ &= (1+s)[p(1+s)+1-p] \\ &= (1+s)[1+sp], \\ \bar{w}_a &= 1+sp, \end{aligned}$$

Geometric fitness ratios

$$\begin{aligned} \bar{w} &= p^2(1+s)^2 + 2p(1-p)(1+s) + (1-p)^2 \\ &= [p(1+s)+1-p]^2 \\ &= (1+sp)^2. \end{aligned}$$

$$\begin{aligned} p' &= \frac{p(1+s)(1+sp)}{(1+sp)^2} \\ &= \frac{p(1+s)}{1+sp}, \\ \frac{p'}{1-p'} &= \frac{p}{1-p}\frac{(1+s)(1+sp)}{1+sp} \\ &= \boxed{} \end{aligned}$$

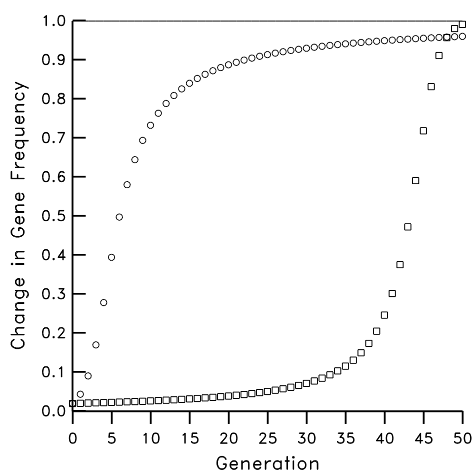
Asexual case (logistic growth)

$$\frac{p_A^{(t)}}{p_a^{(t)}} = (1+s)^t \frac{p_A^{(0)}}{p_a^{(0)}},$$

Additive fitness ratios

AA	Aa	aa
$1 + 2s$	$1 + s$	1

No analytic solutions for p in terms of t!



The course of gene frequency change over 50 generations when fitnesses AA, Aa, and aa are 2.3 : 2.3 : 1 (circles) and 2.3 : 1 : 1 (squares). Initial frequency A is 0.02.

Additive fitness ratios

$$1 + 2s : 1 + s : 1$$

$$\bar{w} \; = \; 1 + 2sp,$$

$$p' \; = \; \frac{p(1 + s + sp)}{1 + 2sp}.$$

$$\Delta p \; = \; \frac{sp(1 - p)}{1 + 2sp}.$$

Times required to change through various gene frequency ranges when $s = 0.01$.

		Favored Allele		
From	To	Dominant	Multiplicative	Recessive
0.001	0.01	232.07	231.32	90,231.2
0.01	0.1	249.89	240.99	9,239.79
0.1	0.5	308.61	220.82	1,019.72
0.5	0.9	1,019.72	220.82	308.61
0.9	0.99	9,239.79	240.89	249.89
0.99	0.999	90,231.2	231.32	232.07

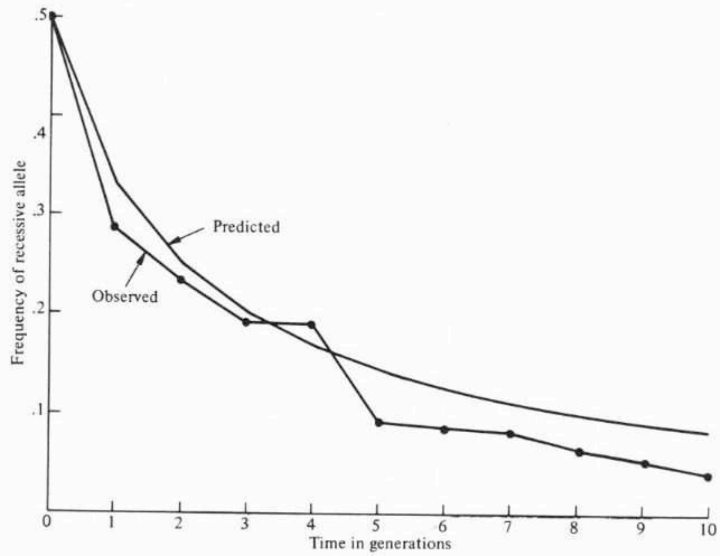
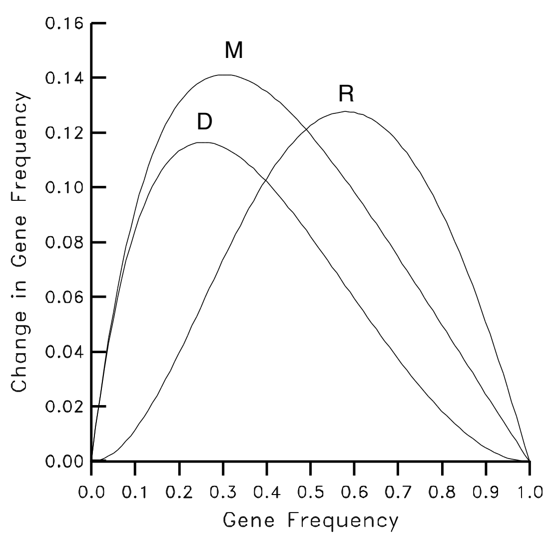


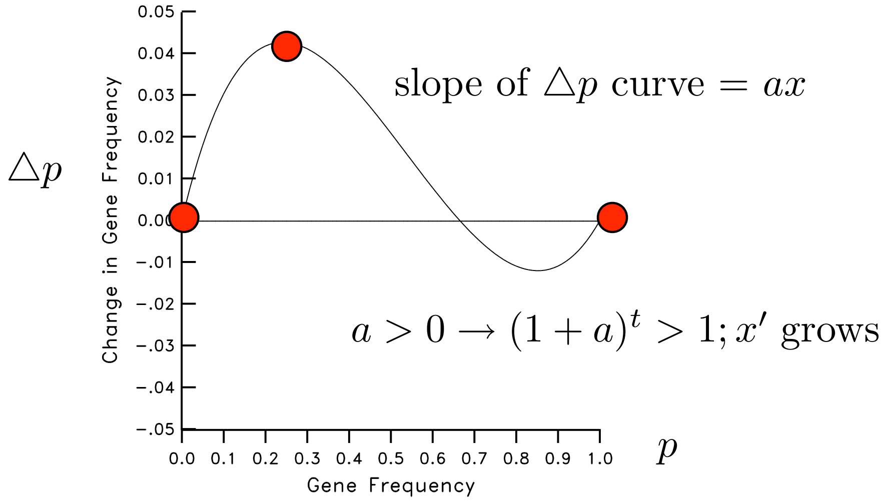
FIGURE 3.4. *Experiment illustrating selection against a recessive lethal gene. The frequency of the recessive allele is on the vertical axis, time in generations is on the horizontal axis.* [Data from B. Wallace (1963), The elimination of an autosomal lethal from an experimental population of *Drosophila melanogaster*, *Amer. Natur.* **97**: 65–66.]



Change of the gene frequency plotted against gene frequency of A for cases in which the favored allele is dominant (D), multiplicative (M) and recessive (R). Fitnesses of AA : Aa : aa genotypes were respectively 2.3 : 2.3 : 1, 5.29 : 2.3 : 1, and 2.3 : 1 : 1.

Stability Considerations

$$
\begin{aligned}
x' &= p + \Delta p - p_e &&= p_e + x + \Delta p - p_e \\
&= x + \Delta p &&\simeq x + ax &&= x(1+a)
\end{aligned}
$$



The change in gene frequency (Δp) plotted against the gene frequency in a case of overdominance where fitnesses of $AA : Aa : aa$ are 0.85 : 1 : 0.7.

Stability Analysis

$$
\begin{aligned}
x' &= p + \Delta p - p_e &&= p_e + x + \Delta p - p_e \\
&= x + \Delta p &&\simeq x + ax &&= x(1+a)
\end{aligned}
$$

Locally stable if:

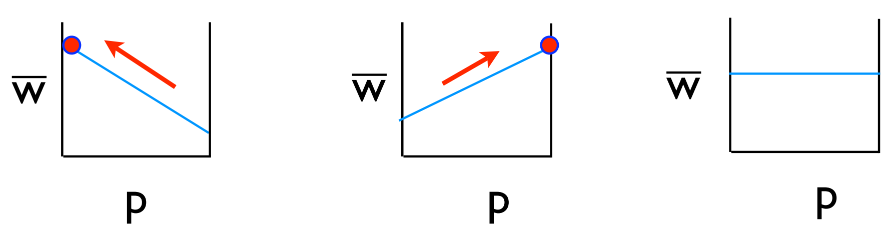
$$-2 \;<\; \left[\frac{d(\Delta p)}{dp}\right]_{p\,=\,p_e} \;<\; 0,$$

Dynamical system analysis of 'adaptive topography'
or mean fitness vs. p

$$\bar{w} = p^2[(w_{11} - w_{12}) + (w_{22} - w_{12})] - 2p(w_{11} - w_{12}) + w_{22}$$

$$w_{12} = \frac{w_{11} + w_{22}}{2}$$

One locus, 2 allele case: 7 graphs, p vs. $\bar{w}$



'Degenerate' case: quadratic mean fitness, with
$w_{12}= (w_{11}+w_{22})/2$

Dynamical system analysis of 'adaptive topography' or mean fitness vs. p - nondegenerate case

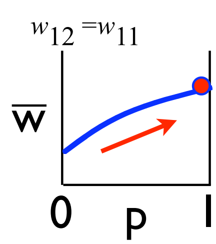
$$\bar{w} = p^2[(w_{11} - w_{12}) + (w_{22} - w_{12})] - 2p(w_{11} - w_{12}) + w_{22}$$

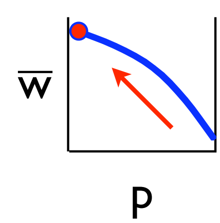
$w_{12} \neq \dfrac{w_{11} + w_{22}}{2}$ so in this case, formula for $\bar{w}$ is a parabola.

There are 4 further subcases, depending on the ordering of the w_{ij}

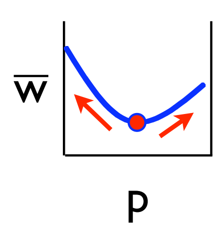
$$\hat{p} = \frac{w_{22} - w_{12}}{(w_{11} - w_{12}) + (w_{22} - w_{12})}$$

The four nonlinear cases - selection at one locus, 2 alleles - adaptive topography

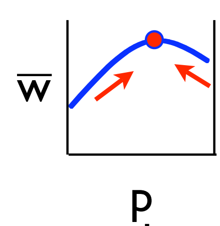




$$\therefore \hat{p} = \frac{w_{22} - w_{12}}{(w_{11} - w_{12}) + (w_{22} - w_{12})} = \frac{w_{22} - w_{12}}{w_{22} - w_{12}} = 1$$

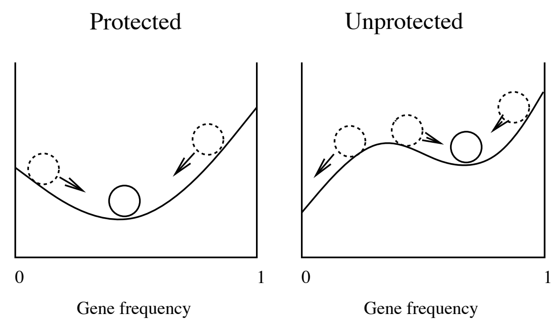


underdominance

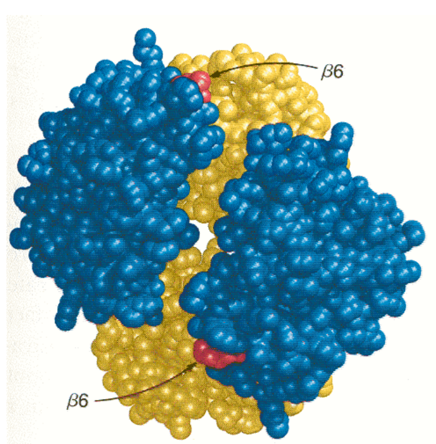


overdominance

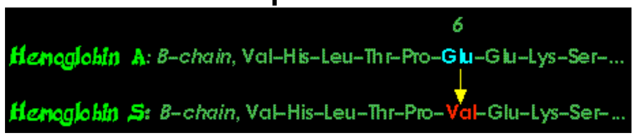
Balanced and unbalanced polymorphism



Variation: within species - Human hemoglobin



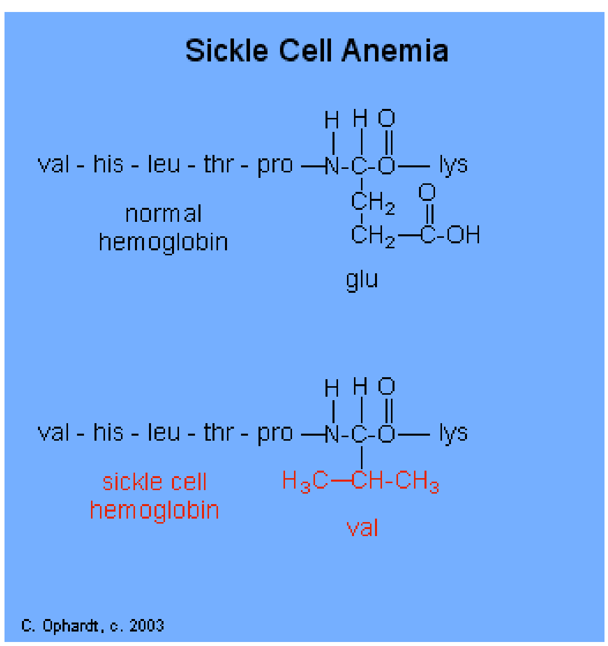
Human variation at genetic code level (genotype) to variation in protein to variation in...



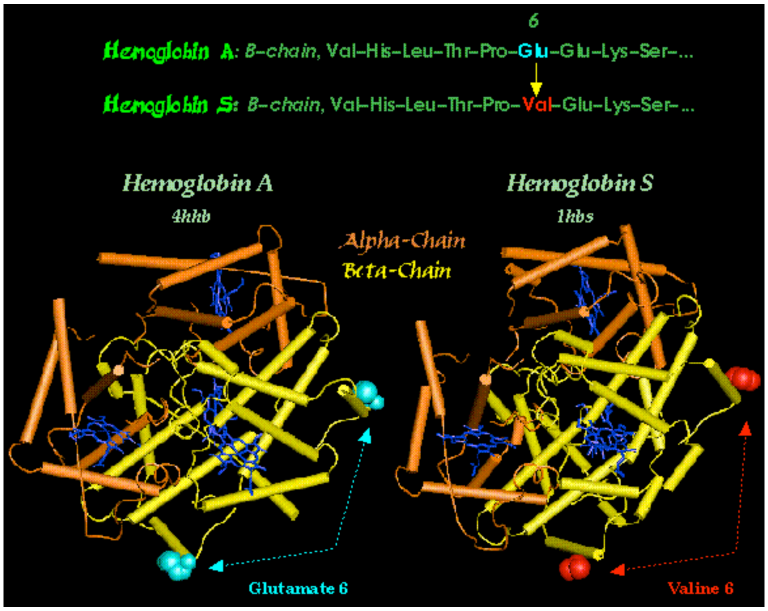
ATG GTG CAC CTG ACT CCT **GA** GAG AAG TCT GCC GTT ACT
ATG GTG CAC CTG ACT CCT **GT** GAG AAG TCT GCC GTT ACT

MVHLTPEEKSAVT (E is the single letter abbreviation for glutamic acid)
MVHLTP**V**EKSAVT (V is the single letter abbreviation for valine)

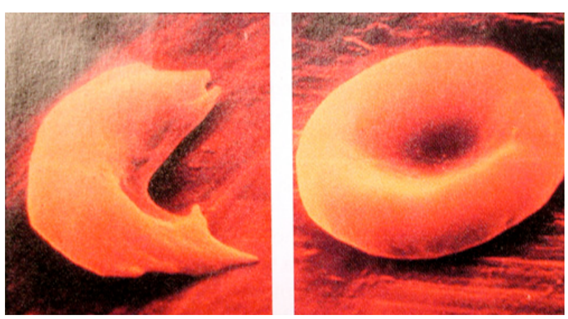
Glutamic acid is a hydrophilic amino acid. Valine is a hydrophobic amino acid.



Variation: different “allelomorphs” or “alleles” (Bateson, 1908)



Variation in “phenotype” = ‘form that shows’



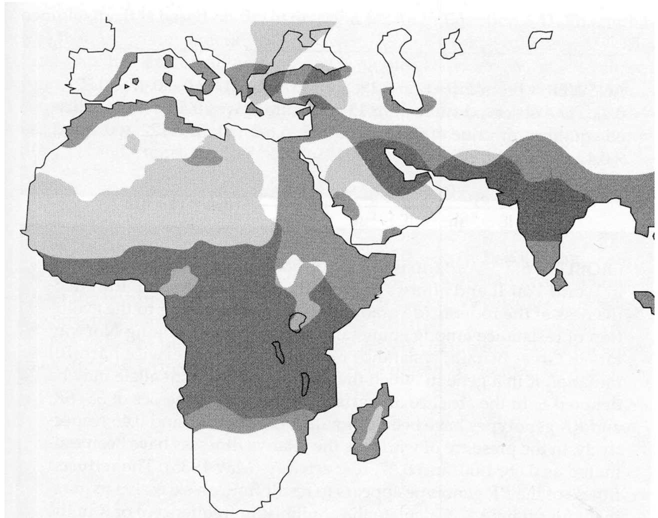


Figure 6.5 The medium gray areas show the incidence of falciparum malaria in Africa, the Middle East, and southern Europe in the 1920s before mosquito control programs were implemented. The light gray areas are regions with a high incidence of sickle-cell anemia. The extensive overlap in the distributions (darkest shade) was an early indication that there might be some causal connection. (After Cavalli-Sforza 1974.)

Analysis of Hb-a, HB-s, and HB-c data (from Cavalli-Sforza, 1977)

	AA	SS	CC	AS	AC	SC
Observed	25374	67	108	5482	1737	130
Expected	25616	307	75	4967	1769	165
Obs/Exp	0.99	0.22	1.45	1.10	0.98	0.79
Relative fitness	0.89	0.20	1.31	1	0.89	0.70

Suppose just *A*, *S* alleles

$$\hat{p}_S = \frac{w_{22} - w_{12}}{(w_{11} - w_{12}) + (w_{22} - w_{12})} = \frac{0.2 - 1.0}{0.89 - 2.0 + 0.2} = 0.1209$$

p_A=0.8791

$\bar{w} = 0.90$

Suppose just a <u>few</u> *C* alleles introduced

$$w_C = p_A w_{AC} + p_S w_{SC} + p_C w_{CC},\ \text{when } p_c \approx 0,$$

$$w_C = p_A w_{AC} + p_S w_{SC} = 0.8670$$

C cannot invade when rare, even though this yields global fitness!