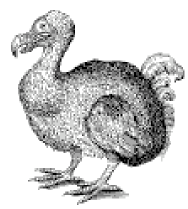
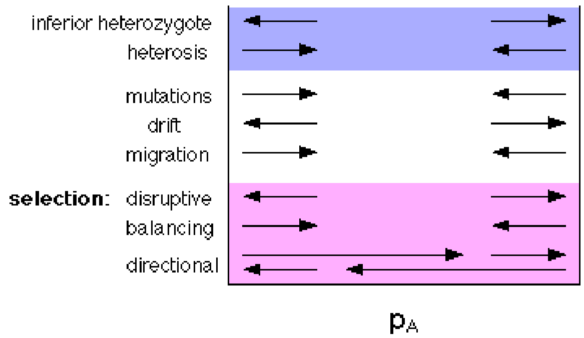
6.891: Computational Evolutionary Biology

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

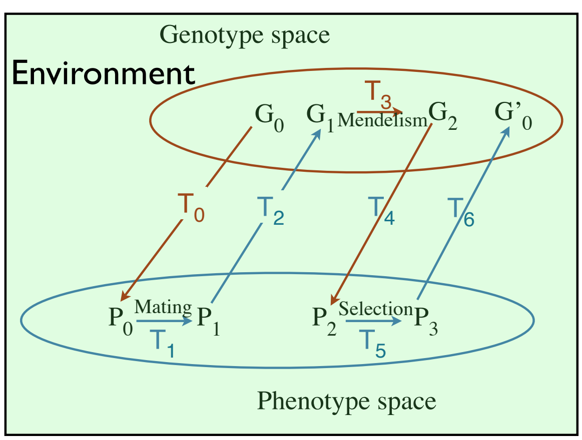


The forces of evolution, II

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



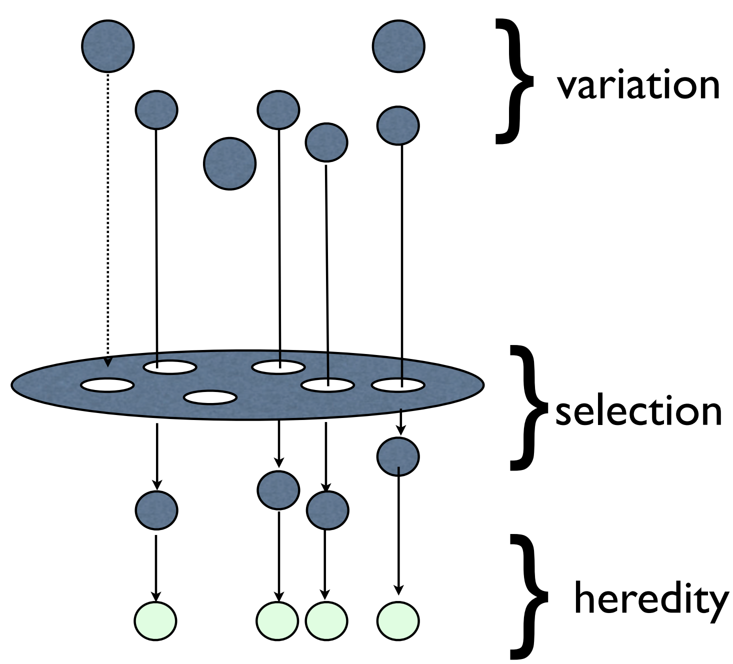
The dynamical system framework



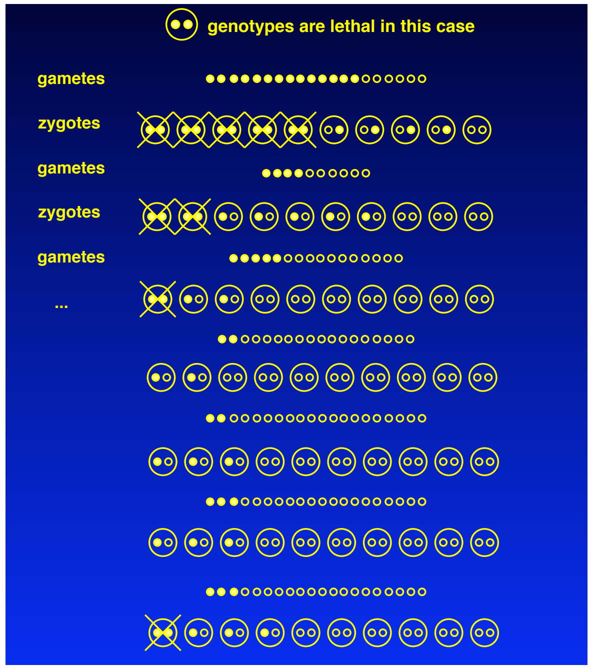
The dynamical system framework

$\mathbf{p}$: some space
T: some mapping, $T\mathbf{p} \rightarrow \mathbf{p}$ ($\mathbf{p}' = T\mathbf{p}$)
sequence $\mathbf{p}$, $T\mathbf{p}$, $T(T\mathbf{p})$,...$T^k(\mathbf{p})$ = <u>orbit</u> of $\mathbf{p}$
$\mathbf{p}' - \mathbf{p} = T\mathbf{p} - \mathbf{p}$

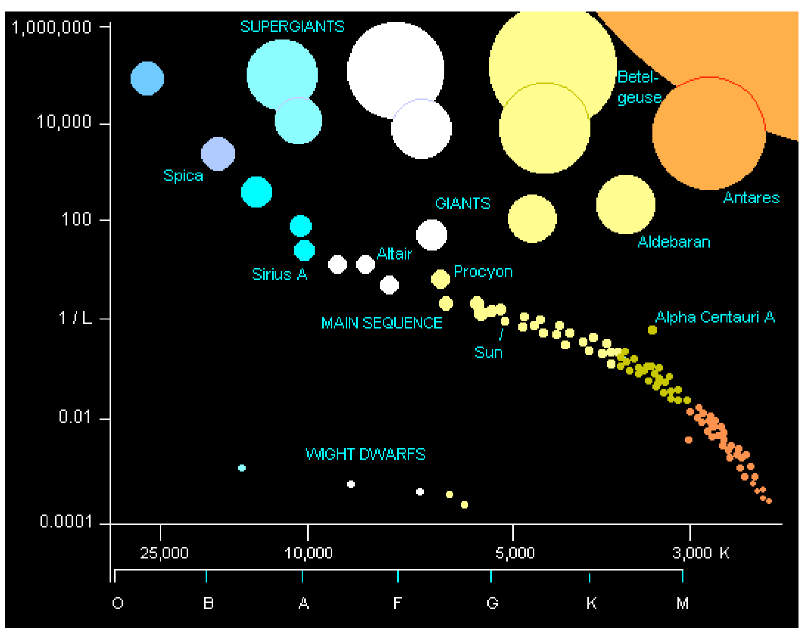
A selectional model of evolution



New reality TV show: "Survivor"



Transformational Evolution - the main sequence



	Mortal	Immortal
Variational		
Transformational		

Fisher’s proof of mud slides

$x =$ 1st parent's deviation from mean

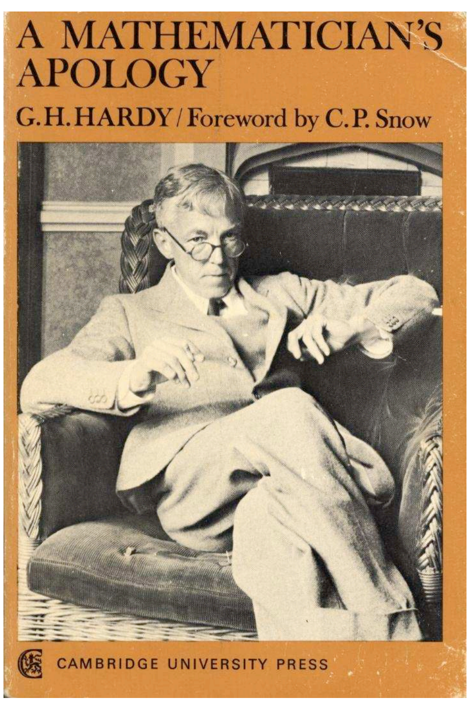
$y =$ 2nd parent's deviation from mean

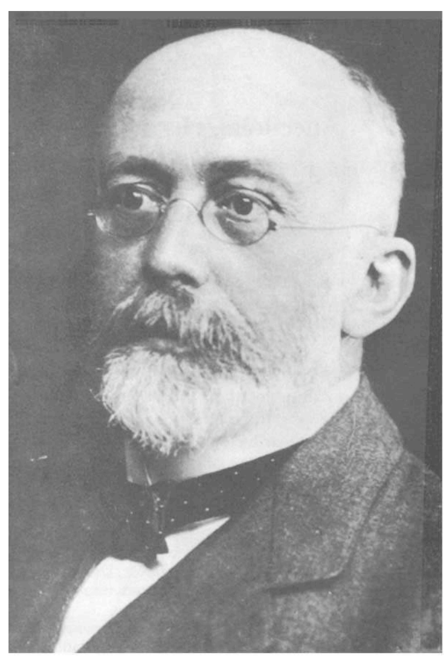
variance $= E(x^2)$

$$\mathrm{var}(\frac{1}{2}(x+y)) = E\left(\left[\frac{1}{2}(x^2+y^2)\right]^2\right) = E\left[\frac{1}{4}(x^2+2xy+y^2)\right] =$$

$$E\left[\frac{1}{4}(2x^2)\right] = \frac{1}{2}E(x^2)$$

Gregor Mendel saves Darwin?





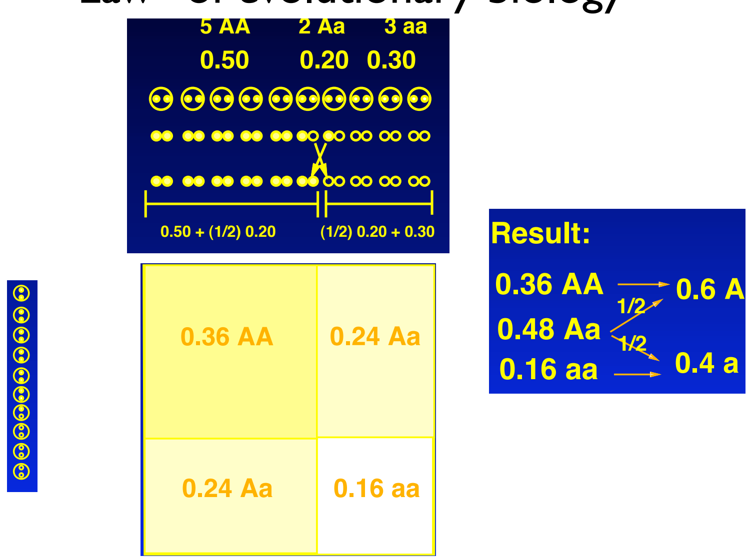
Nature Reviews | Genetics

Terminology

A,a = different forms of the same gene
AKA "alleles"
p,q = frequencies of alleles
two chromosomes in each eurkaryotic cell - diploid - so possible <u>genotypes</u> are:

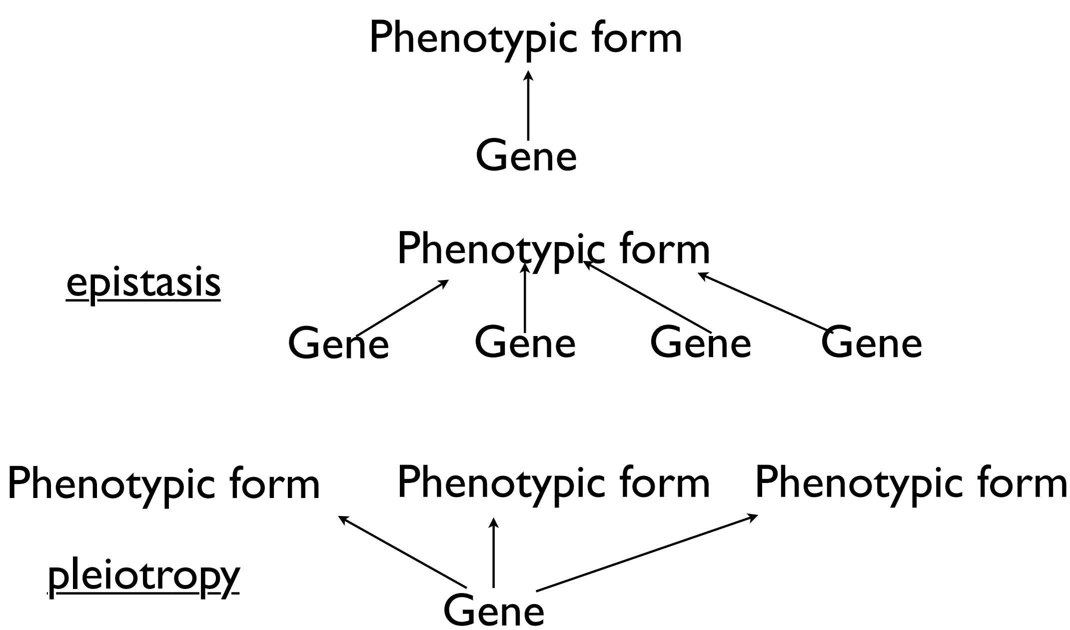
AA = homozygote *A*
Aa = heterozygote *A,a*
aa = homozygote *a*

Hardy-Weinberg as the simplest evolutionary dynamical system: "Newton's First Law" of evolutionary biology

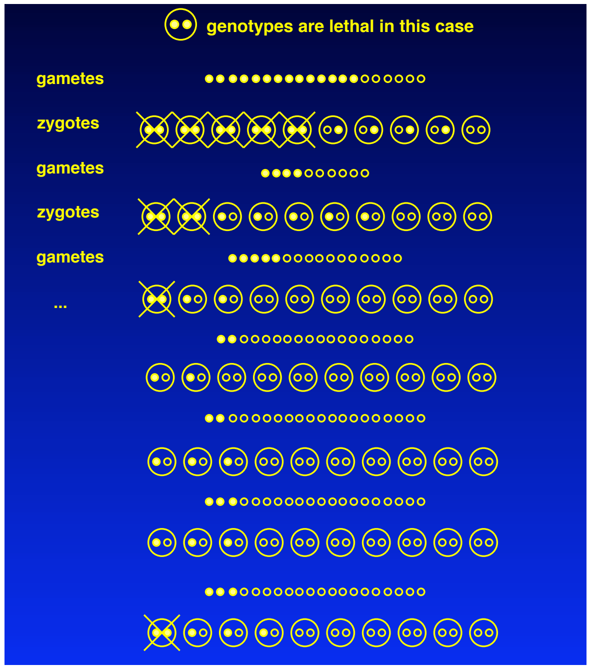


How realistic are the H-W assumptions?
What biological model are we proposing?

Model myth vs. reality



New reality TV show: "Survivor"



Survivor by the ratings numbers

Genotypes:	AA	Aa	aa
relative fitnesses:	1	1	0.7 (assume these are viabilities)

Initial gene frequency of A = 0.2

Initial genotype frequencies (from Hardy–Weinberg)

(newborns)	0.04	0.32	0.64
	x 1	x 1	x 0.7

Survivors (these are relative viabilities)

0.04 + 0.32 + 0.448 = Total: 0.808

genotype frequencies among the survivors: (divide by the total)

0.0495 0.396 0.554

gene frequency

A: 0.0495 + 0.5 x 0.396 = 0.2475

a: 0.554 + 0.5 x 0.396 = 0.7525

genotype frequencies: (among newborns)

0.0613 0.3725 0.5663

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):

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

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

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

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

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

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

F=ma

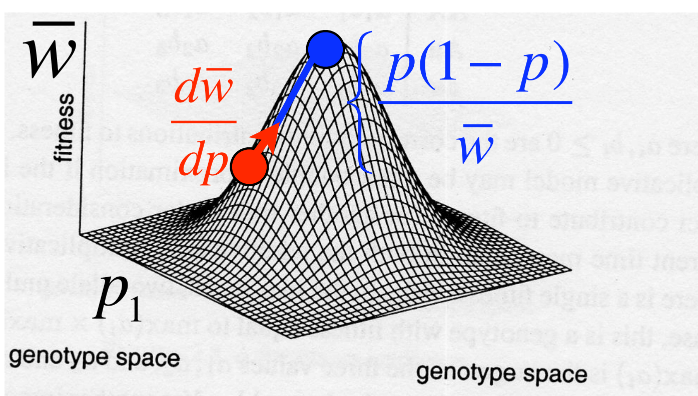
The jet fuel formula for 'evolutionary change'

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

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

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

The shape of things now



Mean fitness always increases...

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

But...this is <u>not</u> the same things as maximizing fitness...

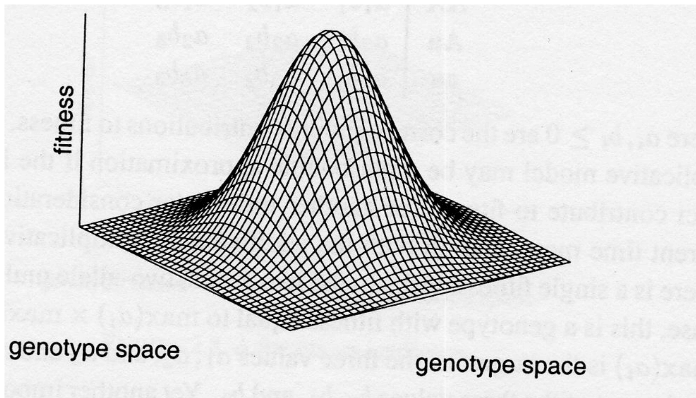
$dp = kp(1-p)dt$ or

$\int \frac{1}{p(1-p)}dp = dt$ so

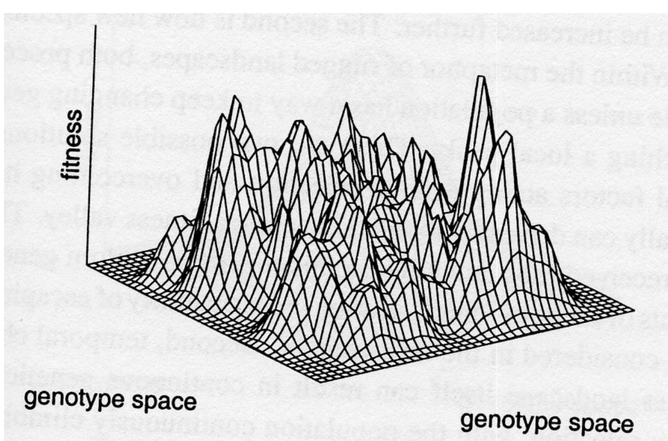
$$t(p_1, p_2) = \int_{p_1}^{p_2} \frac{1}{p(1-p)}dp$$

p	0.001-0.01	0.01-0.1	0.1-0.5	0.5-0.9	0.9-0.99	0.99-0.999
w11>w12> w22	462	480	439	439	480	462
dominance, w11=w12> w22	232	250	309	1.020	9.240	90,231

The shape of things now



The shape of things now



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"

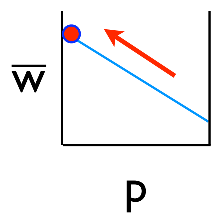
NB: only a few special cases have explicit solutions!

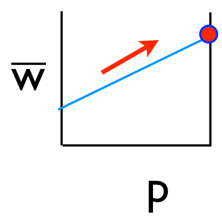
Dynamical system analysis of 'adaptive topography' or mean fitness vs. *p*

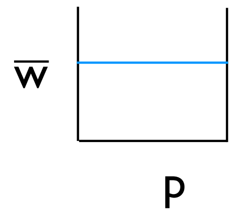
$$\bar{w} = p^2[(w_{11} - w_{12}) + (w_{22} - w_{12})] - 2p(w_{11} - w_{12}) + w_{22}$$

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

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







'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

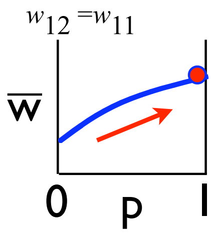
$$\bar{w} = p^2[(w_{11} - w_{12}) + (w_{22} - w_{12})] - 2p(w_{11} - w_{12}) + w_{22}$$

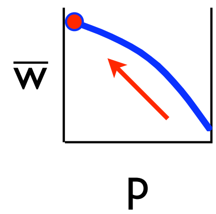
$w_{12} \neq \dfrac{w_{11} + w_{22}}{2}$ so in this case, formula for $\bar{w}$ is a parabola.

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

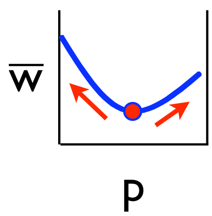
$$\hat{p} = \frac{w_{22} - w_{12}}{(w_{11} - w_{12}) + (w_{22} - w_{12})}$$

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

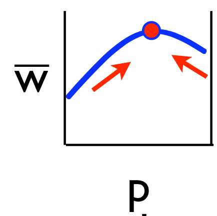




$$\therefore \hat{p} = \frac{w_{22} - w_{12}}{(w_{11} - w_{12}) + (w_{22} - w_{12})} = \frac{w_{22} - w_{12}}{w_{22} - w_{12}} = 1$$



underdominance



overdominance