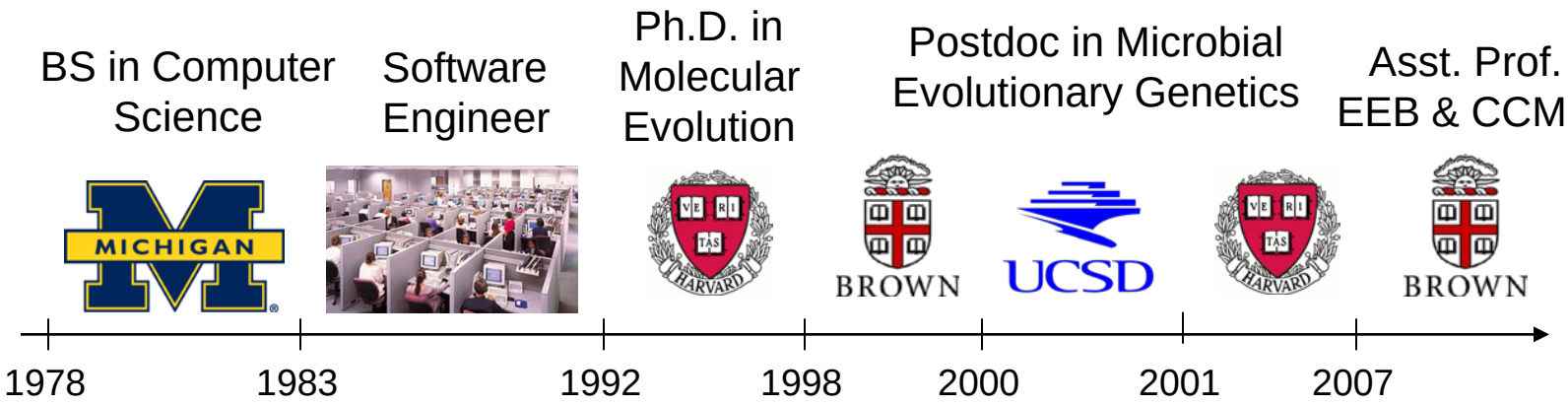
The Simplest Questions About Adaptation That I Know

(and perhaps even some answers!)



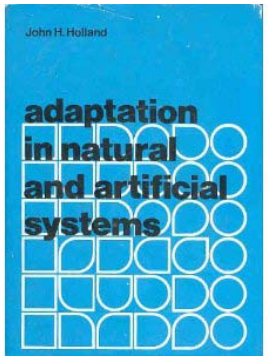
Daniel M. Weinreich

Department of Ecology and Evolutionary Biology, and
Center for Computational Molecular Biology

Brown University

As an Undergraduate I Studied Genetic Algorithms (1983)

John Holland (ca. 1980)



1975

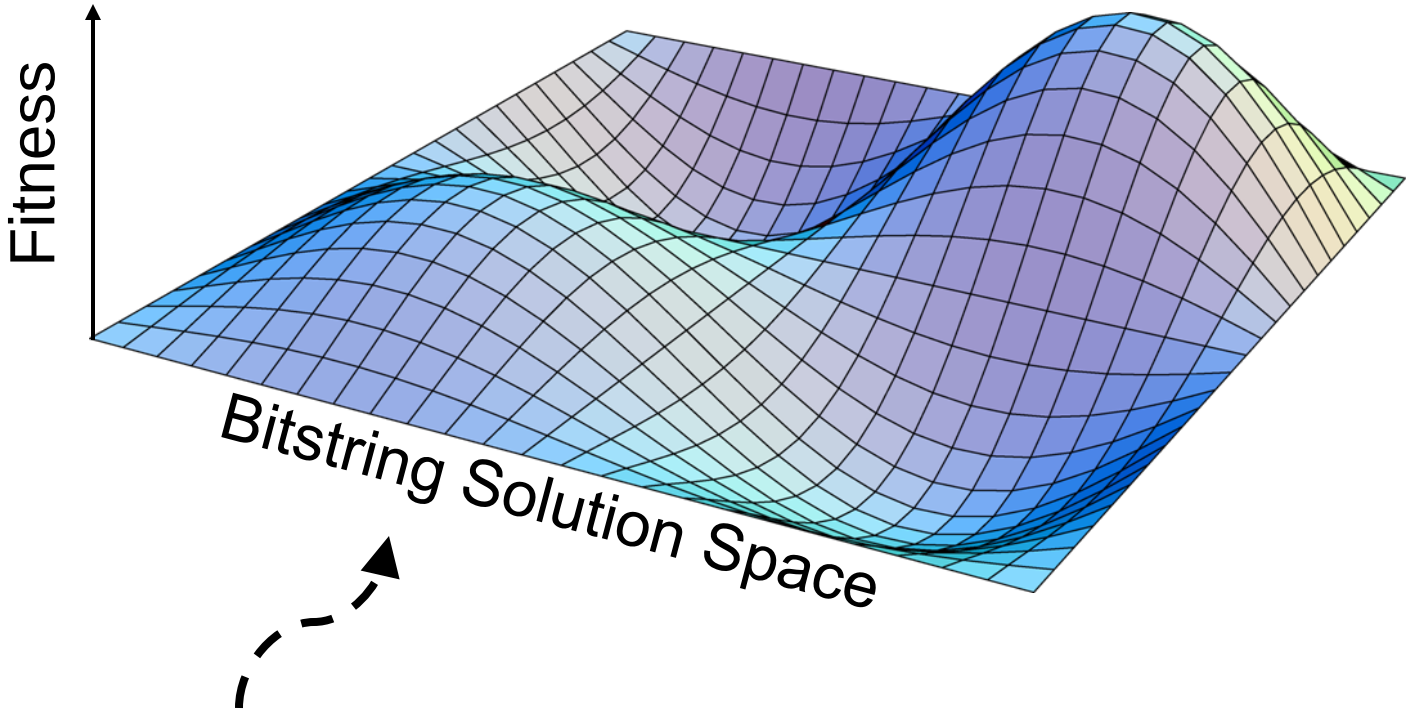
- Requires
  - Bitstring representation of all conceivable solutions
  - Fitness function to evaluate all conceivable solutions
- Initialize population often with ~10^3 random strings.
- Iterate:
  - Selection (e.g. fitness-proportional or tournament or greedy)
  - Reproduction (crossover and/or mutation)
- Stop at some predetermined point.

Fitness Function

Bit Position						
L	…	4	3	2	1	Fitness
0	…	0	0	0	0	3.1
0	…	0	0	0	1	2.7
0	…	0	0	1	0	0.0
0	…	0	0	1	1	1.0
0	…	0	1	0	0	12.3
⋮	…	⋮	⋮	⋮	⋮	⋮
1		1	1	1	0	0.3
1	…	1	1	1	1	1.1

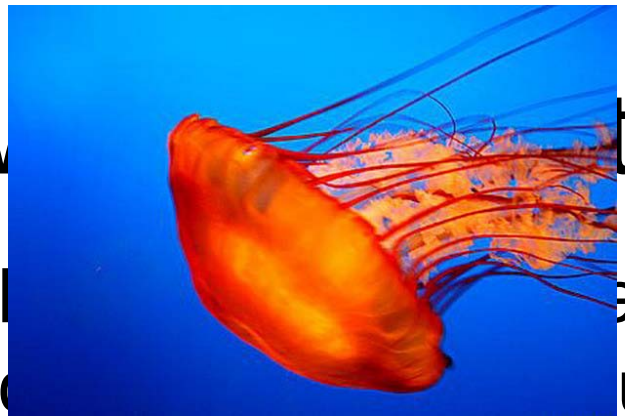
Mapping or lookup table from each solution to a non-negative number.

(Wright's) Fitness Landscape



The *L*-dimensional hypercube (*cf.* Chris Marx' talk last week; Haldane 1932). Readily generalizes to more than two alleles at each locus.

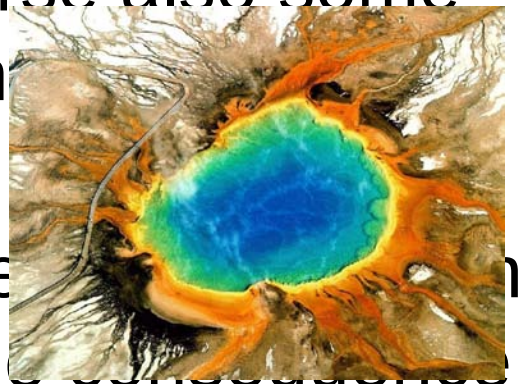
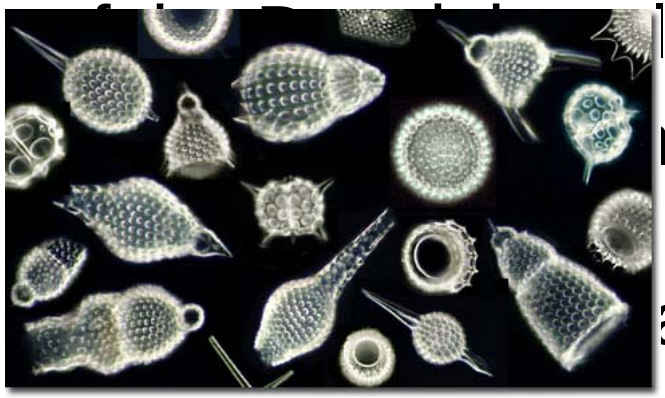
(Wright 1932)


W te

- orit erv
  h s uc
  computational a urse also some
  en n
  on
- nt ge
  regard the staggering biology as the consequence
  lgo
- t c bic
  sc ctl
  am tu .)

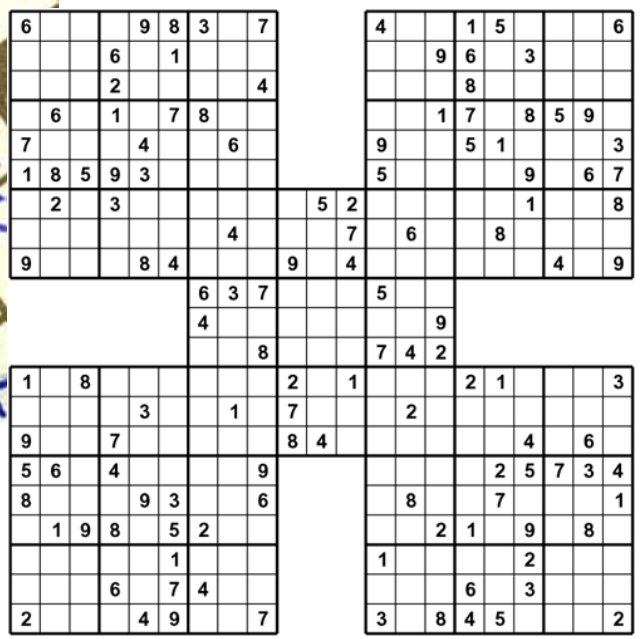


BS in Comp Sci: Fork in the Road

or

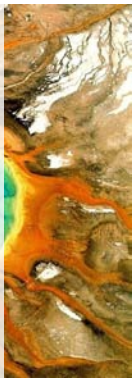
Software Engineering Left Me Hungry…



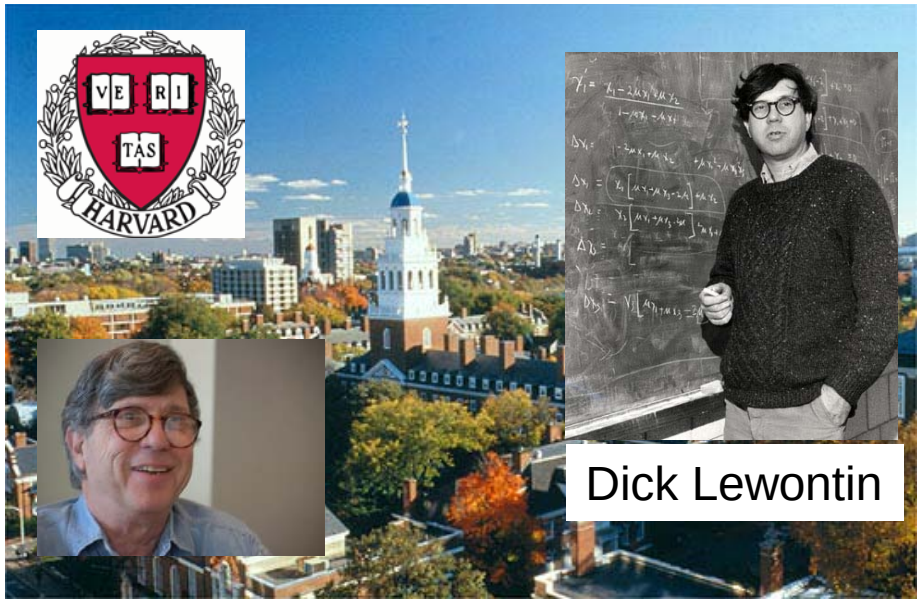
(Not terribly satisfying intellectually.)



Can a computationally-minded approach get intellectual traction in evolutionary genetics?


…Study Evolutionary Biology (1992)



“Find ‘em and Grind ‘em School”

- $\theta = 4N\mu = E(\pi)$

(but where are the dynamics?)

- $\Delta p = p(1 - p)s$

(but where is the epistasis?)

A MOLECULAR APPROACH TO THE STUDY OF GENIC HETEROZYGOSITY IN NATURAL POPULATIONS. II. AMOUNT OF VARIATION AND DEGREE OF HETEROZYGOSITY IN NATURAL POPULATIONS OF *DROSOPHILA PSEUDOOBSCURA*[1]

R. C. LEWONTIN AND J. L. HUBBY

Department of Zoology, University of Chicago, Chicago, Illinois

Genetics **54**: 595–609 August 1966.

Microbial Experimental Evolution

Bruce Levin

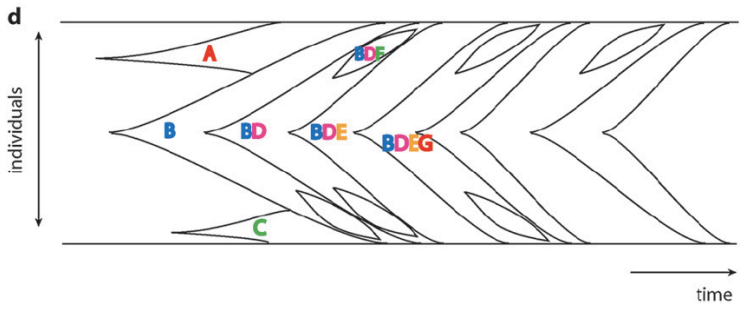
Lin Chao

Christina Burch

'Doing research in population biology without mathematical and/or computer simulation models is like playing tennis without a net or boundary lines'.

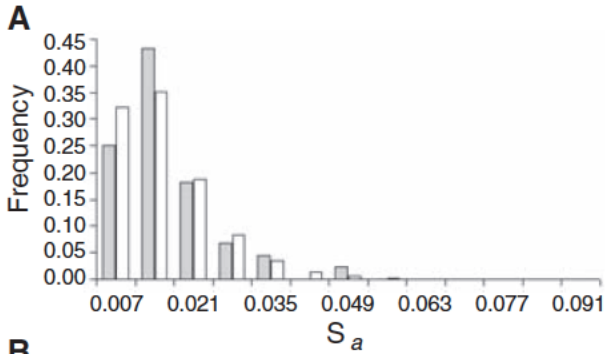
'For us, natural and not-so-natural selection is about dN/dt and not dN/dS.'

Adaptation



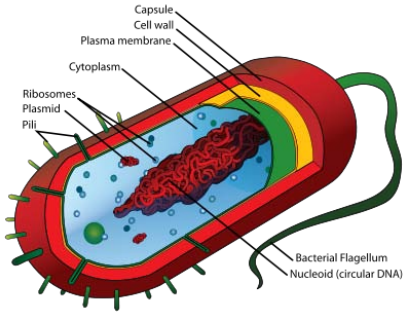
Dynamics of adaptation:

Population structure
Clonal interference
Multiple mutations
Mutation and recombination rates?



Fitness landscape:

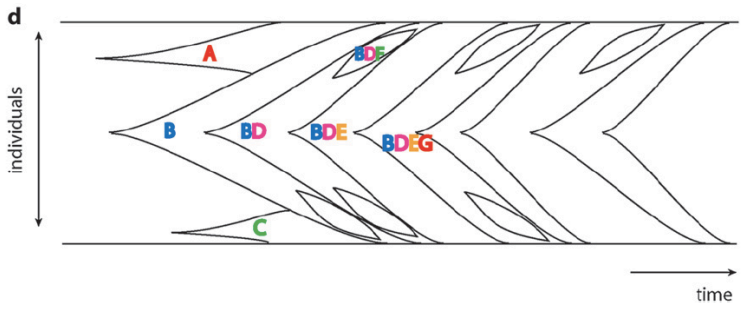
~~Mutation rates~~
Distribution of fitness effects
Epistasis



Environment + cellular architecture:

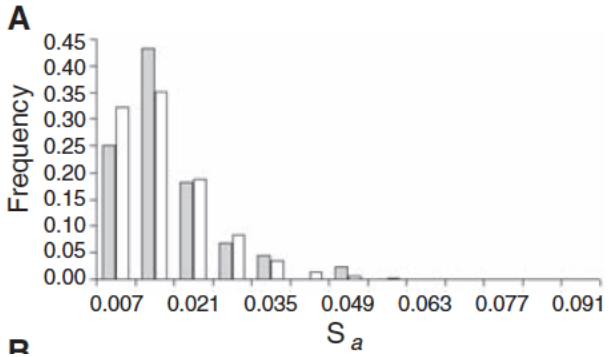
Regulatory networks
Proteins

Adaptation



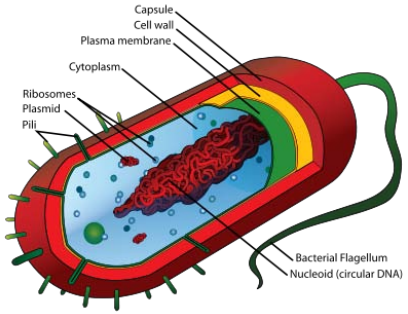
III: Dynamics of adaptation:

Population structure
Clonal interference
Multiple mutations
Mutation and recombination rates?



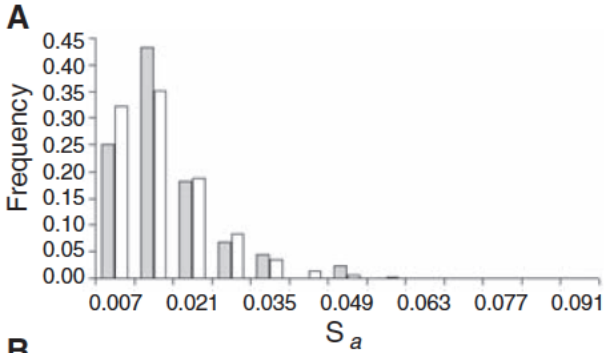
I. Fitness landscape:

~~Mutation rates~~
Distribution of fitness effects
Epistasis



II. Environment + cellular architecture:

Regulatory networks
Proteins



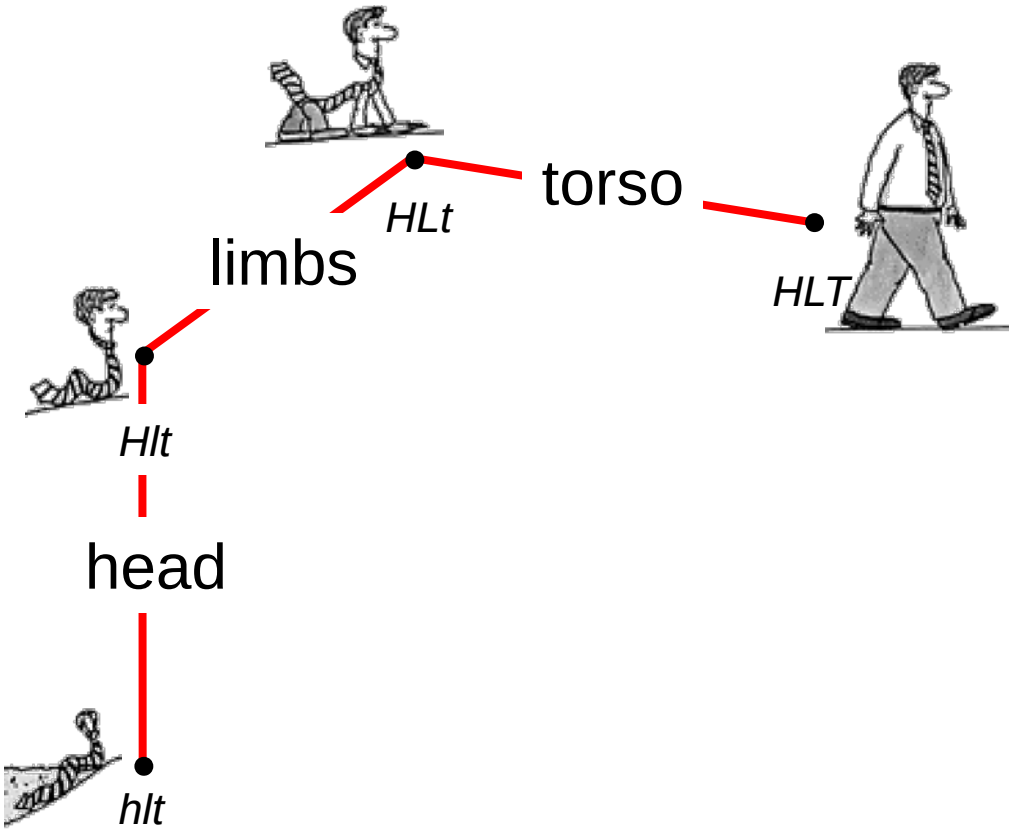
Fitness landscape:
Mutation rates
Distribution of fitness effects
Epistasis

Part I: Fitness Landscapes

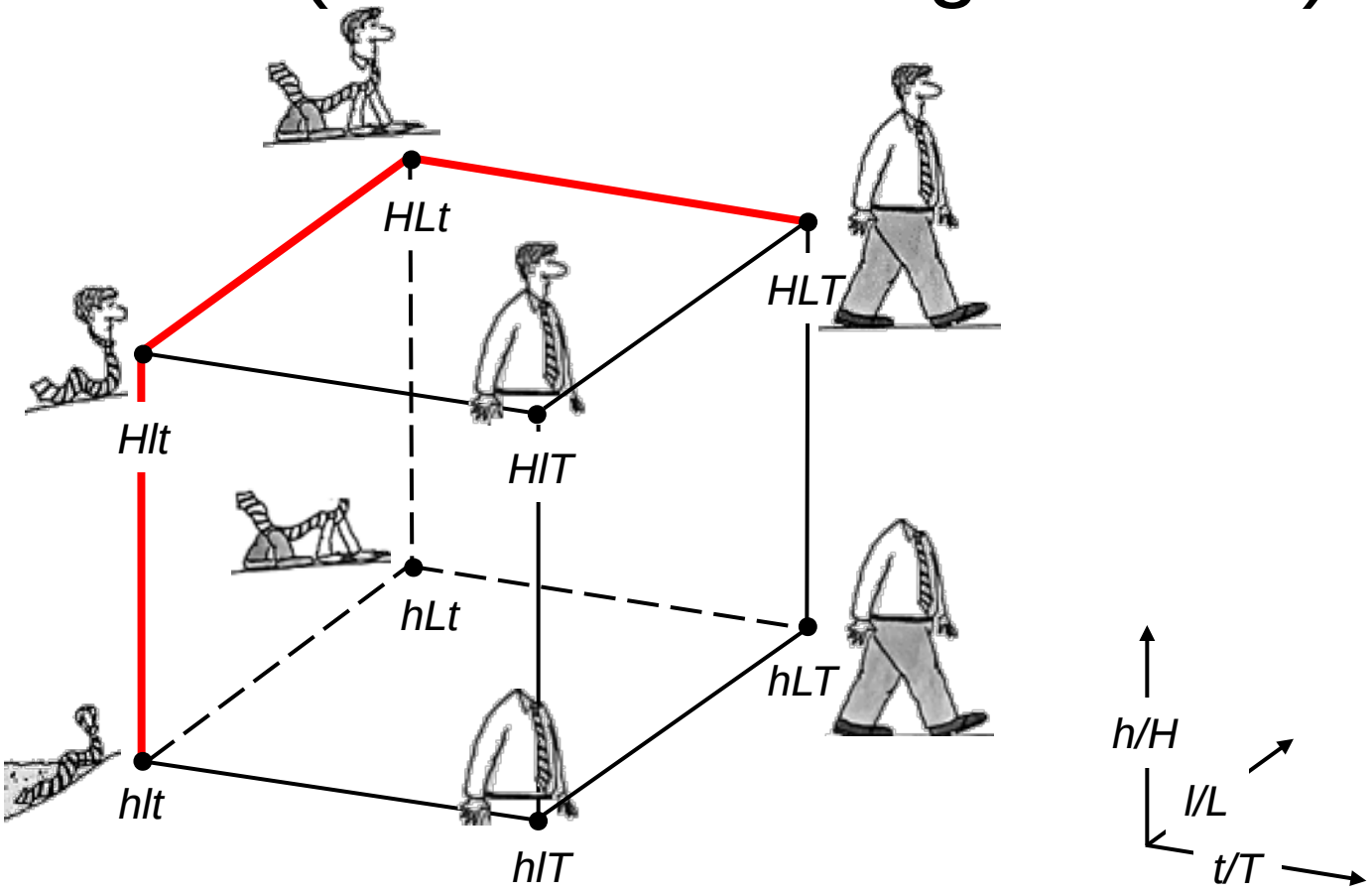
A. How Can We Use Them?
B. What Are They Actually Like?
C. When Do They Fail Us?

A Cartoon of the Problem

Evolution Changes Heritable Phenotypes



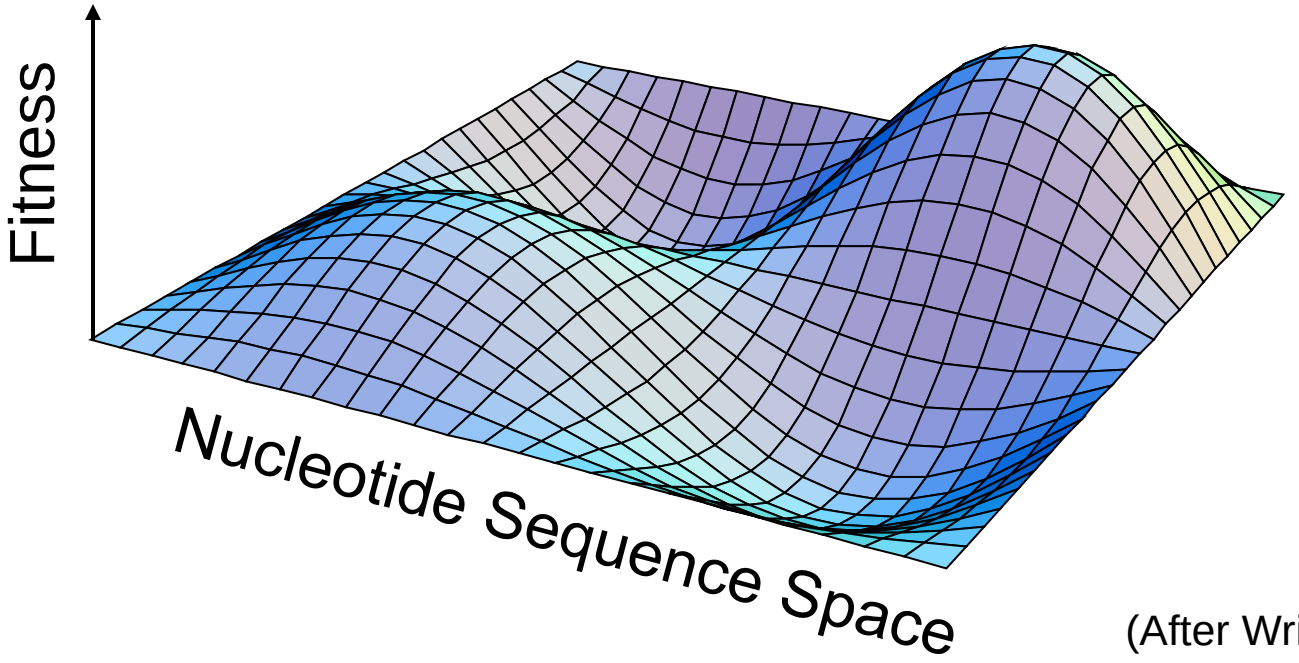
Nucleotide Sequence Space Defines Many Mutationally Equivalent Trajectories (Here Assuming SSWM)



(after Maynard-Smith 1970; see also Wright 1932)

One Question

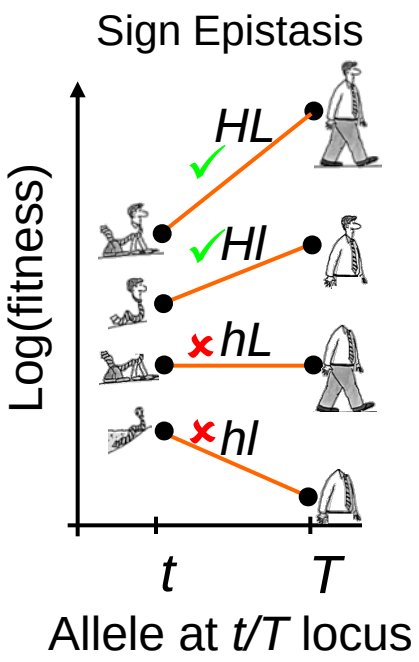
Is natural selection empirically constrained to follow a subset of mutational trajectories to reach high-fitness sequences?

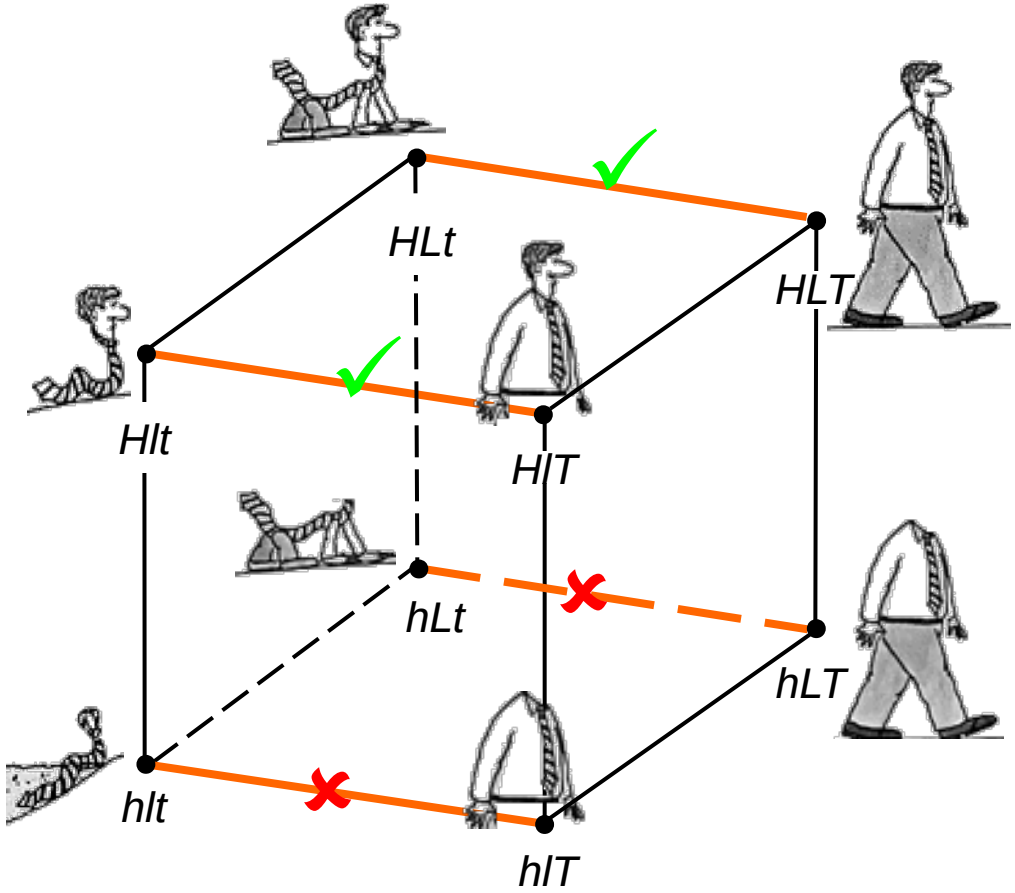


(After Wright 1932)

Theoretical Digression

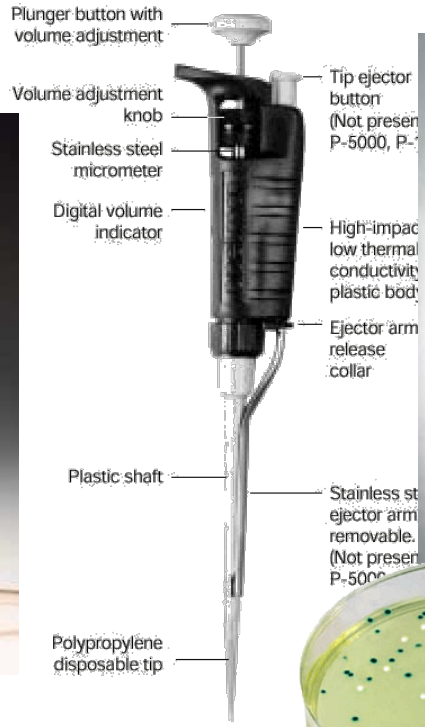
Sign Epistasis Limits Selective Accessibility

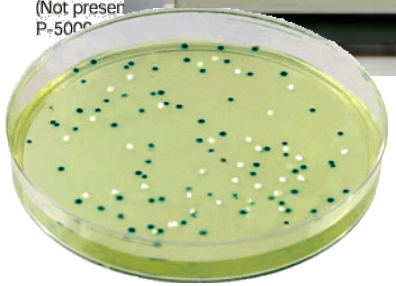




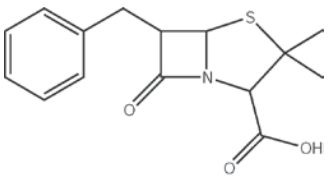
(Weinreich *et al*. 2005)

An Experiment

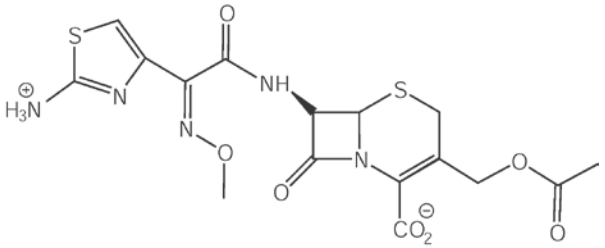




β-lactam Antibiotic Resistance



Penicillin (1946)



Cefotaxime (1992)

- Resistance via hydrolysis of lactam ring by β-lactamase.
- TEM β-lactamases are the major source of plasmid-mediated β-lactam resistance.
- Five mutations in *TEM*^{wt} jointly increase cefotaxime resistance 100,000-fold, and yield an allele called *TEM**.

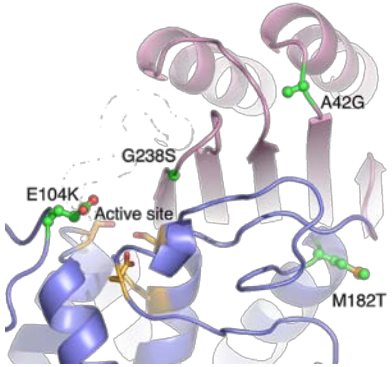
The Question:

• What is the topography of the fitness function lying between *TEM*^{wt} to *TEM**?

– I used reverse genetics to construct each of the 2^5 = 32 alleles defined by all combinations of these five mutations, and assayed the cefotaxime resistance of each.

– On the premise that natural selection acts to increase cefotaxime resistance, this defines the fitness function between these two alleles.

The Data



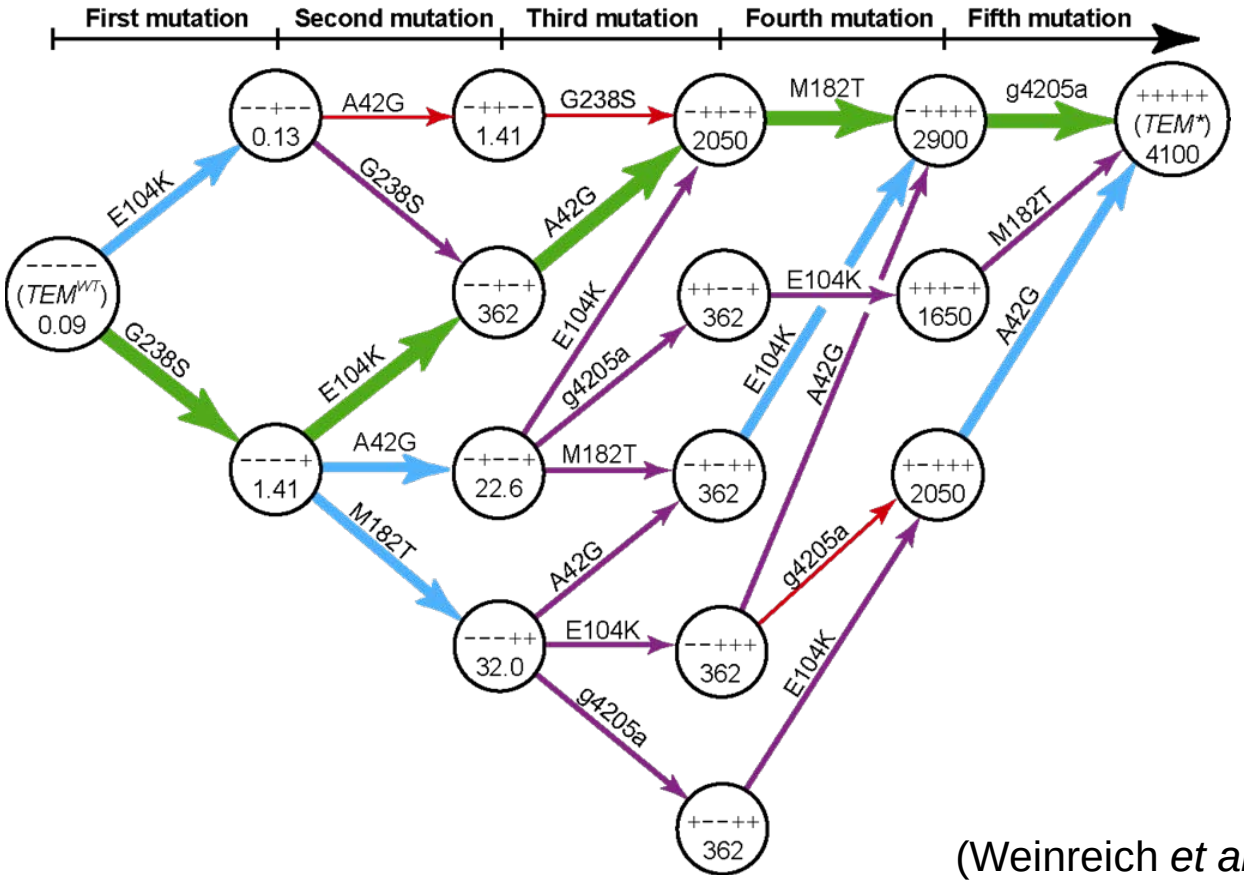
Mutational State						Resistance (μg/ml)		
g4205a	A42G	E104K	M182T	G238S	Allele	Rep 1	Rep 2	Rep 3
–	–	–	–	–	*TEM*^{wt}	0.0884	0.0884	0.0884
–	–	–	–	+		1.41	1.41	1.41
–	–	–	+	–		0.0711	0.0884	0.0711
–	–	–	+	+		32.0	32.0	32.0
–	–	+	–	–		0.130	0.130	0.130
–	–	+	–	+		362.	362.	362.
⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮
+	+	+	+	–		1.41	1.41	2.0
+	+	+	+	+	*TEM**	4096.	4096.	4096.

Four of the five mutations in *TEM** exhibit sign epistasis

Mutation	Number of TEM alleles on which mean mutational effect on resistance is			Mean proportional increase
	Positive	Negative	Negligible	
g4205a	8	2	6	1.4
A42G	12	0	4	5.9
E104K	15	1	0	9.7
M182T	8	3	5	2.8
G238S	16	0	0	1×10^3

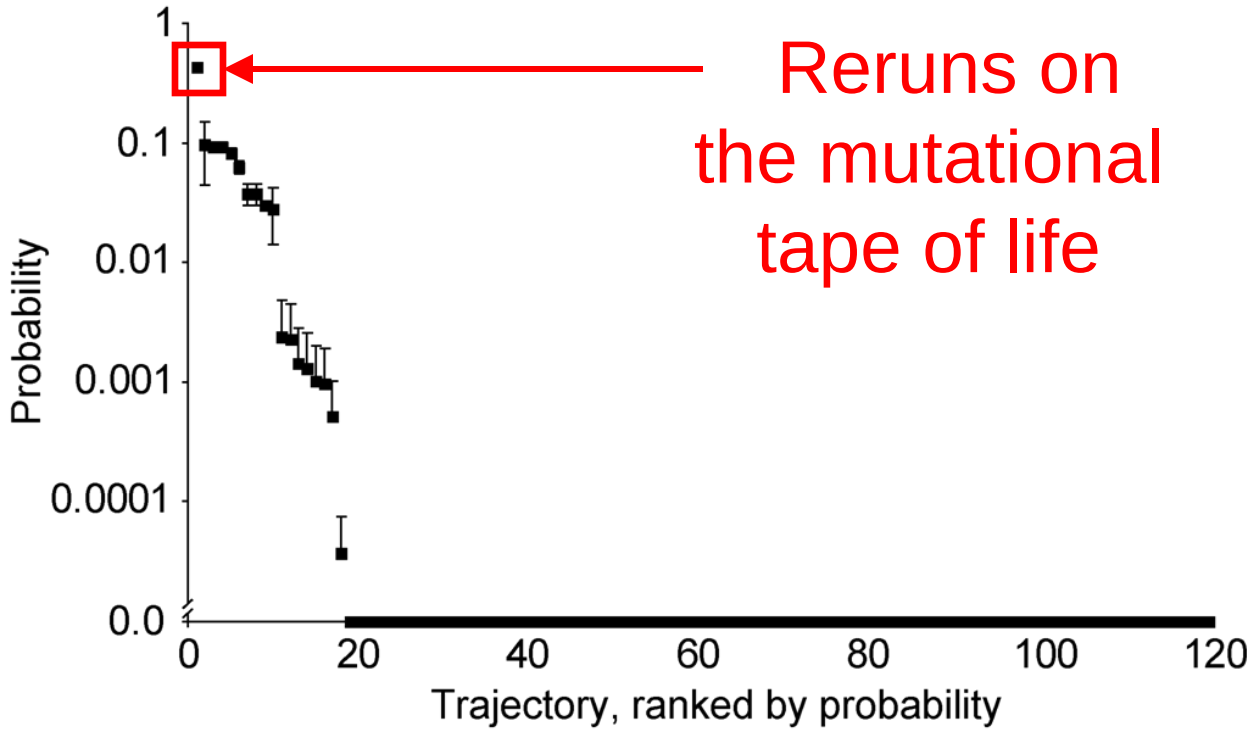
(Weinreich *et al*. 2006)

Only 18 of 120 trajectories are selectively accessible



(Weinreich *et al*. 2006)

Sharp Bias in Probabilities of Realization among Accessible Trajectories



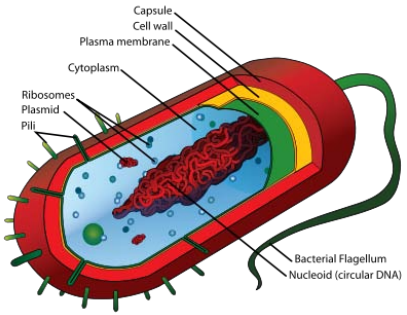
(Weinreich *et al*. 2006)

What About In Other Systems?

- Within a gene: Considerable sign epistasis
  - Isopropyl Malate Dehydrogenase (Lunzer *et al*. 2005)
  - Dihydrofolate Reductase (Lozovsky *et al*. 2009)
  - Ancestral Hormone Receptors (Thornton lab)
  - (How about a structural gene? E.g. β-tubulin?)
- Between genes: Less sign epistasis
  - Methylotrophy (Marx lab)
  - *E. coli* in minimal media (Cooper, Lenski labs)
  - *E. coli* multidrug resistance (Gordo lab)
  - (How about two genes whose products interact? E.g. DHFR and DHPS?)

Fitness Landscape Limitations

- Violations of SSWM (*cf.* Rouzine, Desai, Neher among others). Cannot represent a population by just a single point.
- More sites. (Table grows exponentially and predictions about trajectory realizations are conditioned on a known endpoint.)
- Varying environment including frequency dependence. (We can encode a discrete environment in the fitness function but what's its "mutational" model? *N.B.* the reversing environment of Gore.)



Environment + cellular architecture
Regulatory networks
Proteins

Part II: Environmental + Cellular Architecture

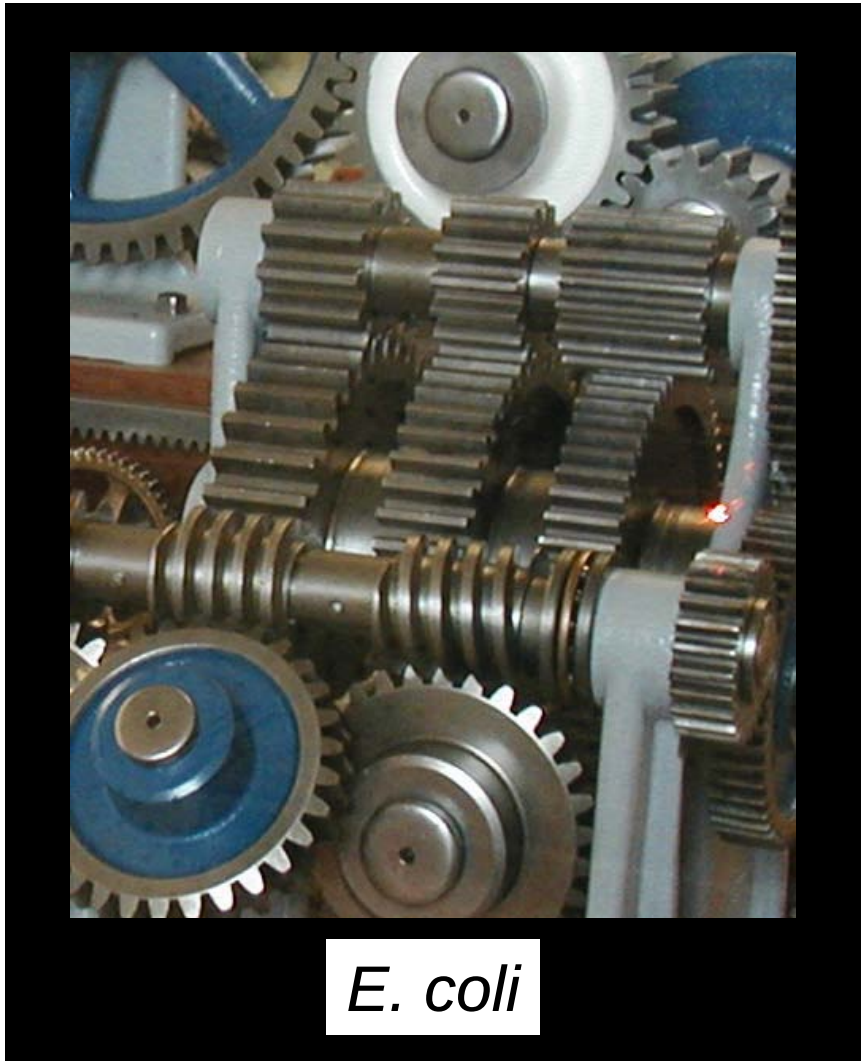
A. Protein Biology of β-Lactamase
B. ΦX174 Life History Evolution
C. Theory
    1. Fisher's Geometric Model
    2. Metabolic Control Analysis

} We're funded here!

Thus Far We Have Ignored…

β-lactamase Allele ⟶

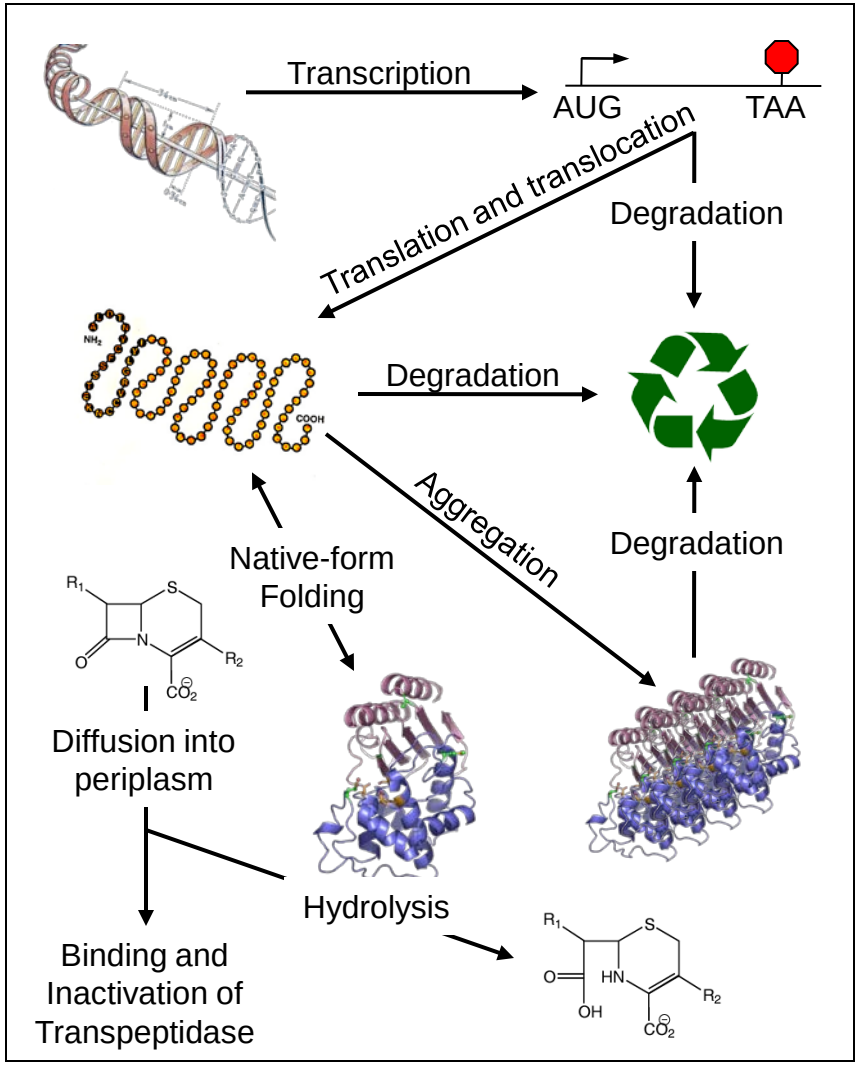
Resistance ⟵



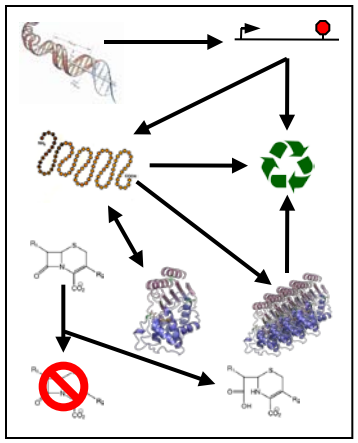
E. coli

…the underlying biology

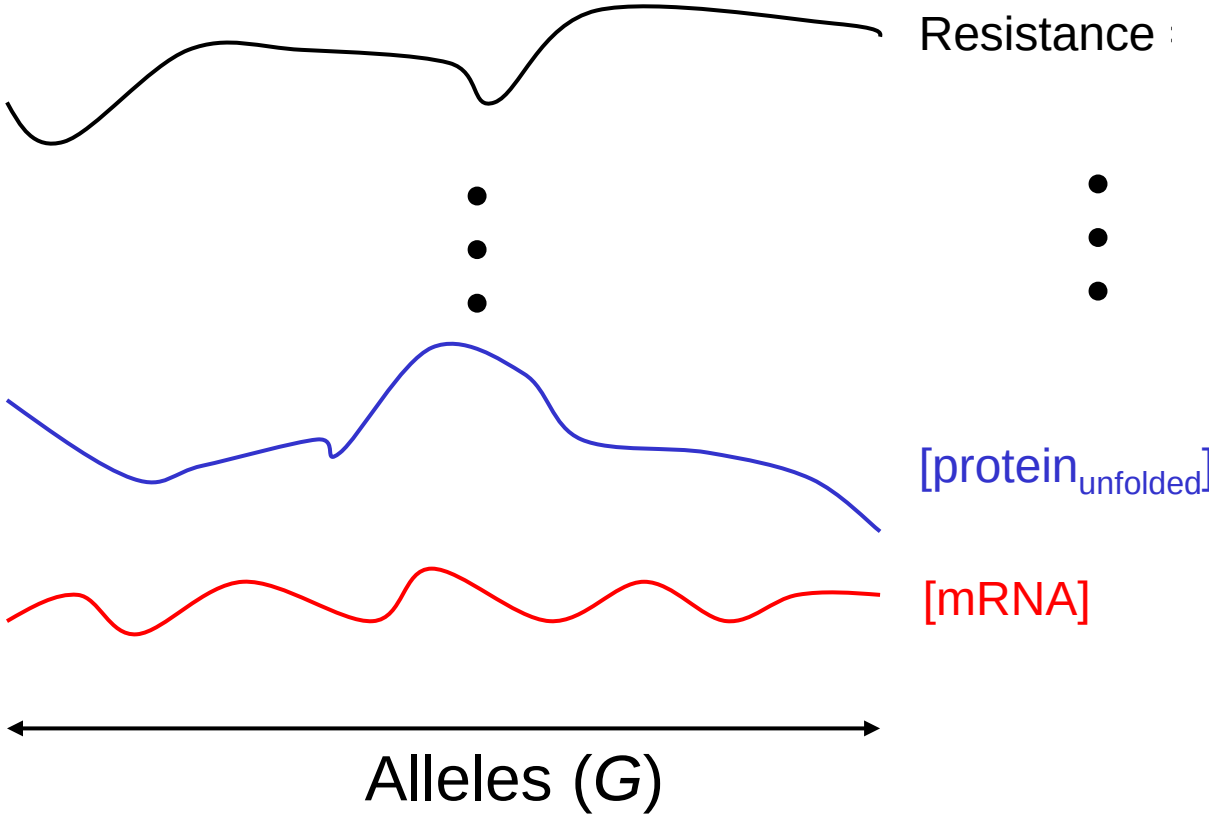
β-lactamase Allele →



Resistance ←

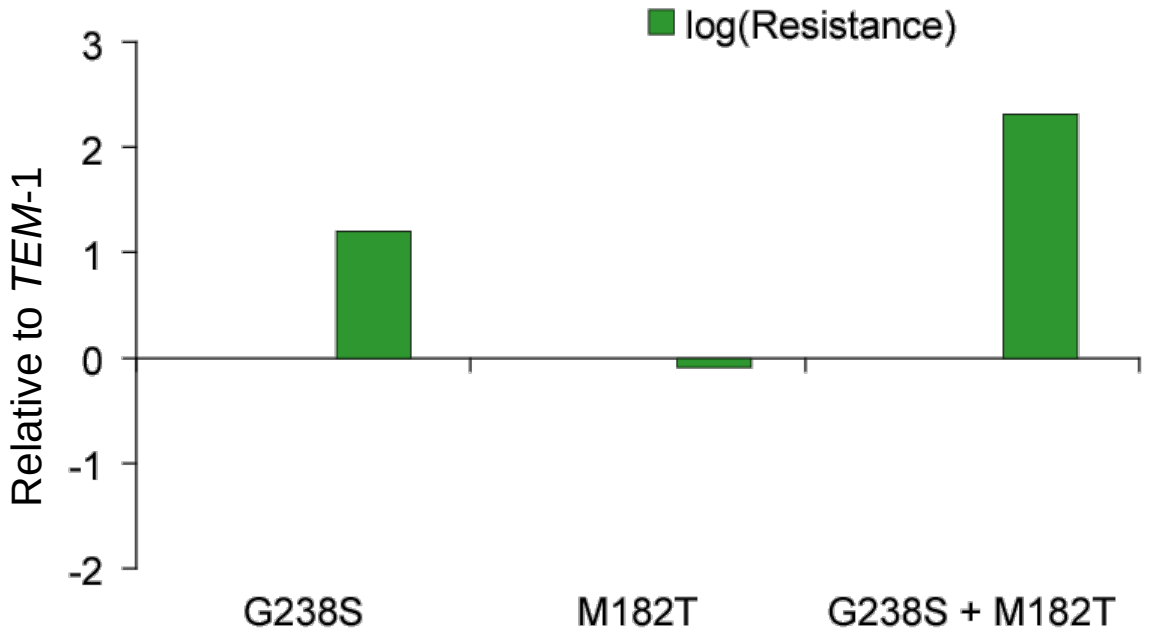


The Big Picture: Modeling Drug Resistance Evolution in Terms of Proximal Mechanisms



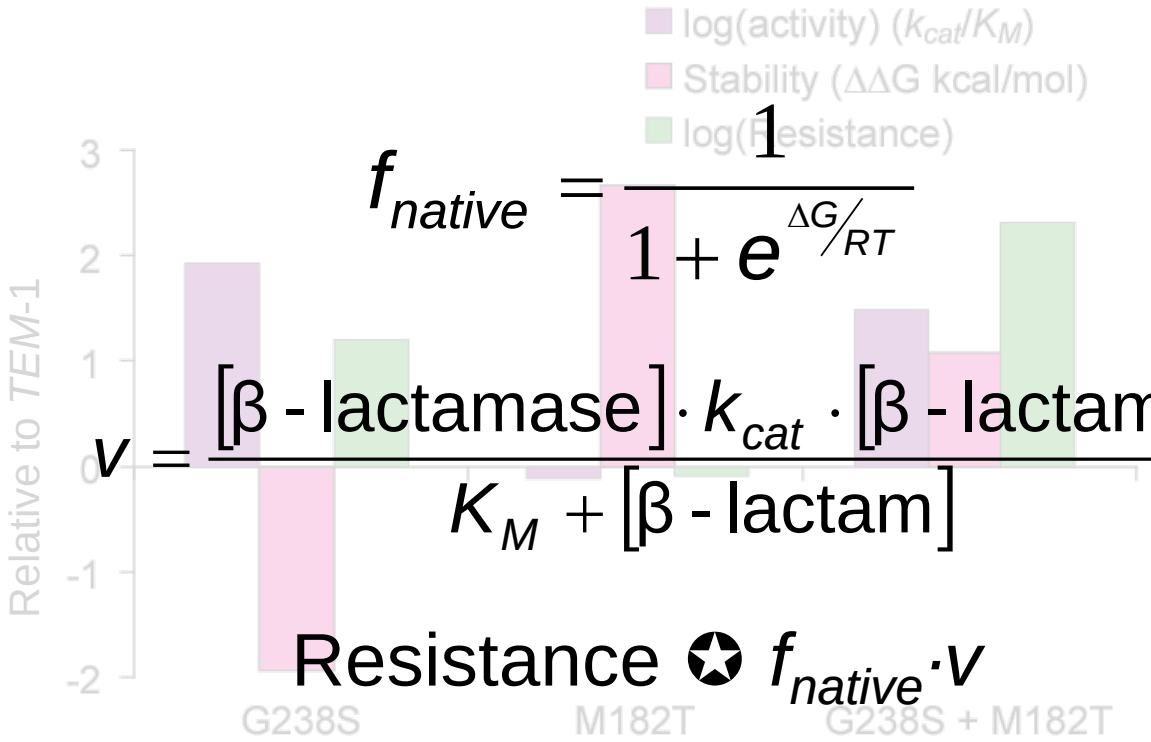
Nothing in Biology Makes Sense Except in the Light of Evolution – Th. Dobzhansky, 1973

Nothing in Evolution Makes Sense Except in the Light of Biology - A. Dean, 2001



(Weinreich *et al*. 2006)

Nothing in Evolution Makes Sense Except in the Light of Biology - A. Dean, 2001



$$f_{native} = \frac{1}{1 + e^{\Delta G / RT}}$$

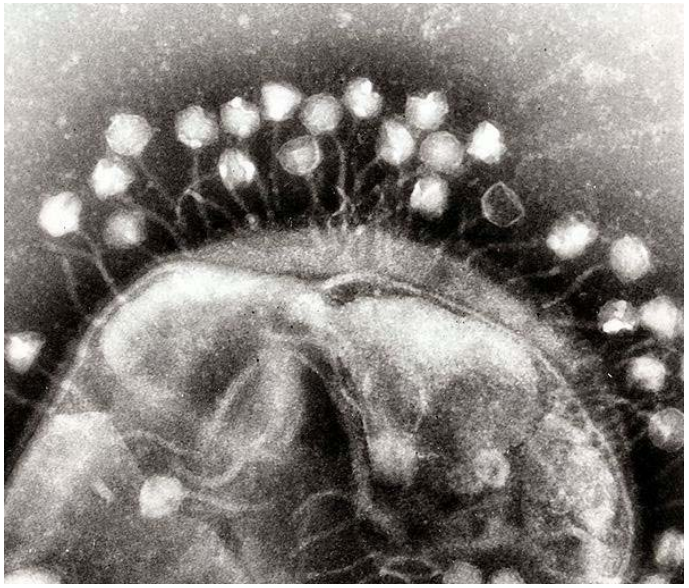
$$v = \frac{[\beta\text{-lactamase}] \cdot k_{cat} \cdot [\beta\text{-lactam}]}{K_M + [\beta\text{-lactam}]}$$

Resistance ✪ $f_{native} \cdot v$

(Wang *et al*. 2002; Weinreich *et al*. 2006)

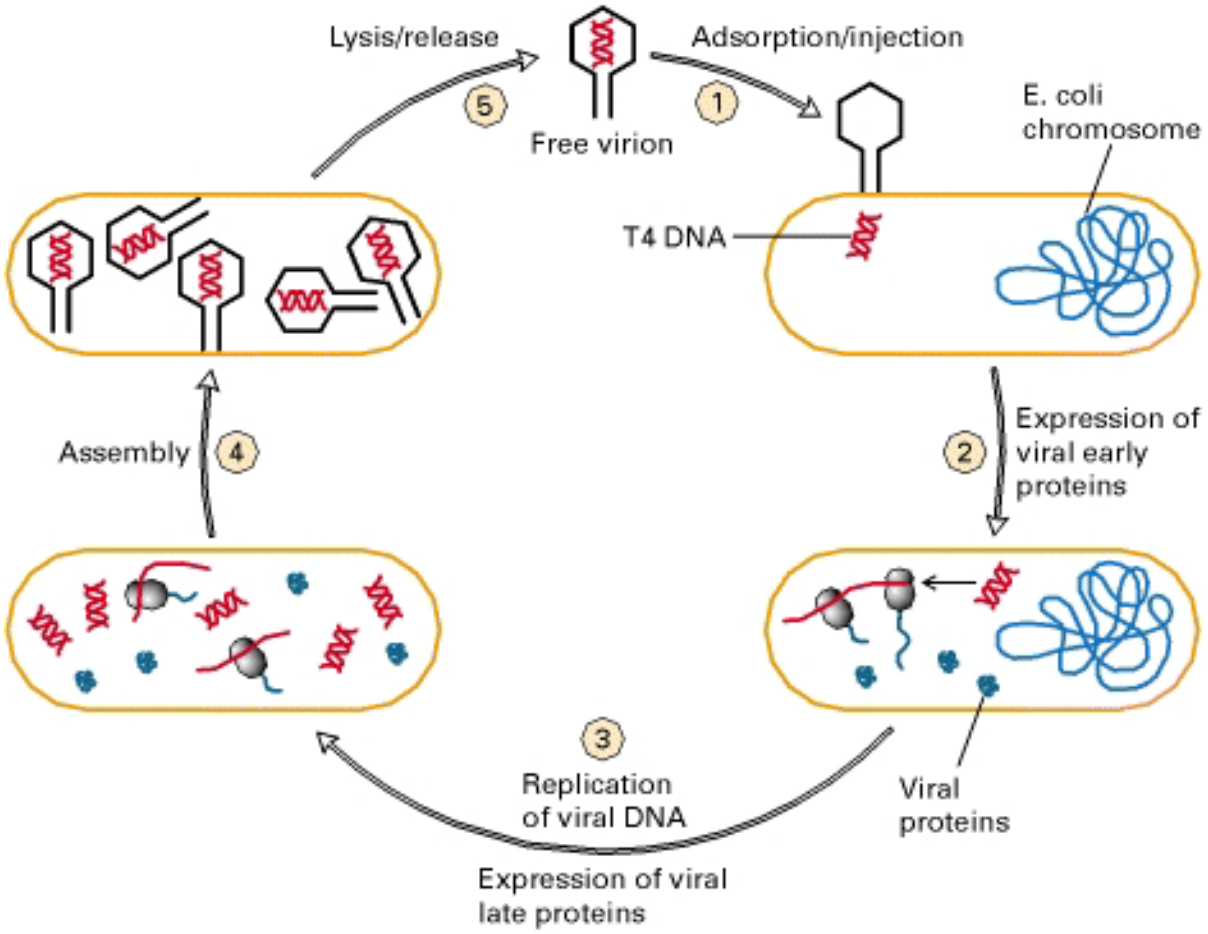
Life History Evolution in Phage

- Phage from φᾰγεῖν, to eat.
- (Bacterio)phage are a genome in a membrane but are metabolically inert.
- Discovered by Twort (1915) and d'Hérelle (1917).
- Fundamental discoveries in molecular genetics made with phage in the '40's.
- First genome sequenced (1997).
- Key top-down regulator of marine microbes.



100 nm

Lytic Phage Life Cycle



Decomposing Phage Growth Rate

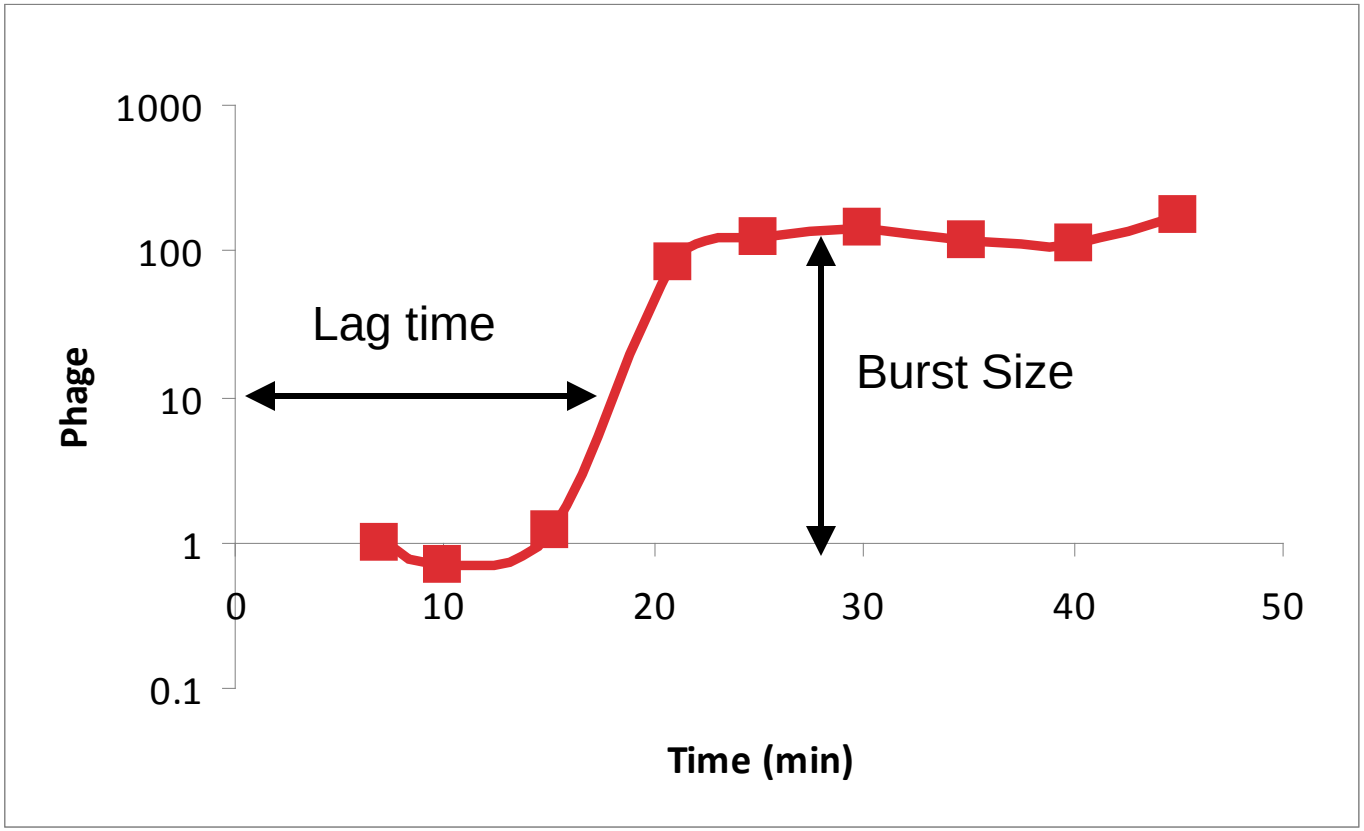
- Bull *et al.* (2006) suggest that growth rate (w) is:

$$w = -x + kC(Be^{-L(d+w)} - 1)$$

- Adsorption rate (k)
  - Rate at which the phage attach to the bacteria
- Burst size (B)
  - Number of phage that emerge from a lysed cell (for one infecting phage)
- Lag time (L)
  - Amount of time between adsorption and lysis
- Phage death rate (x)

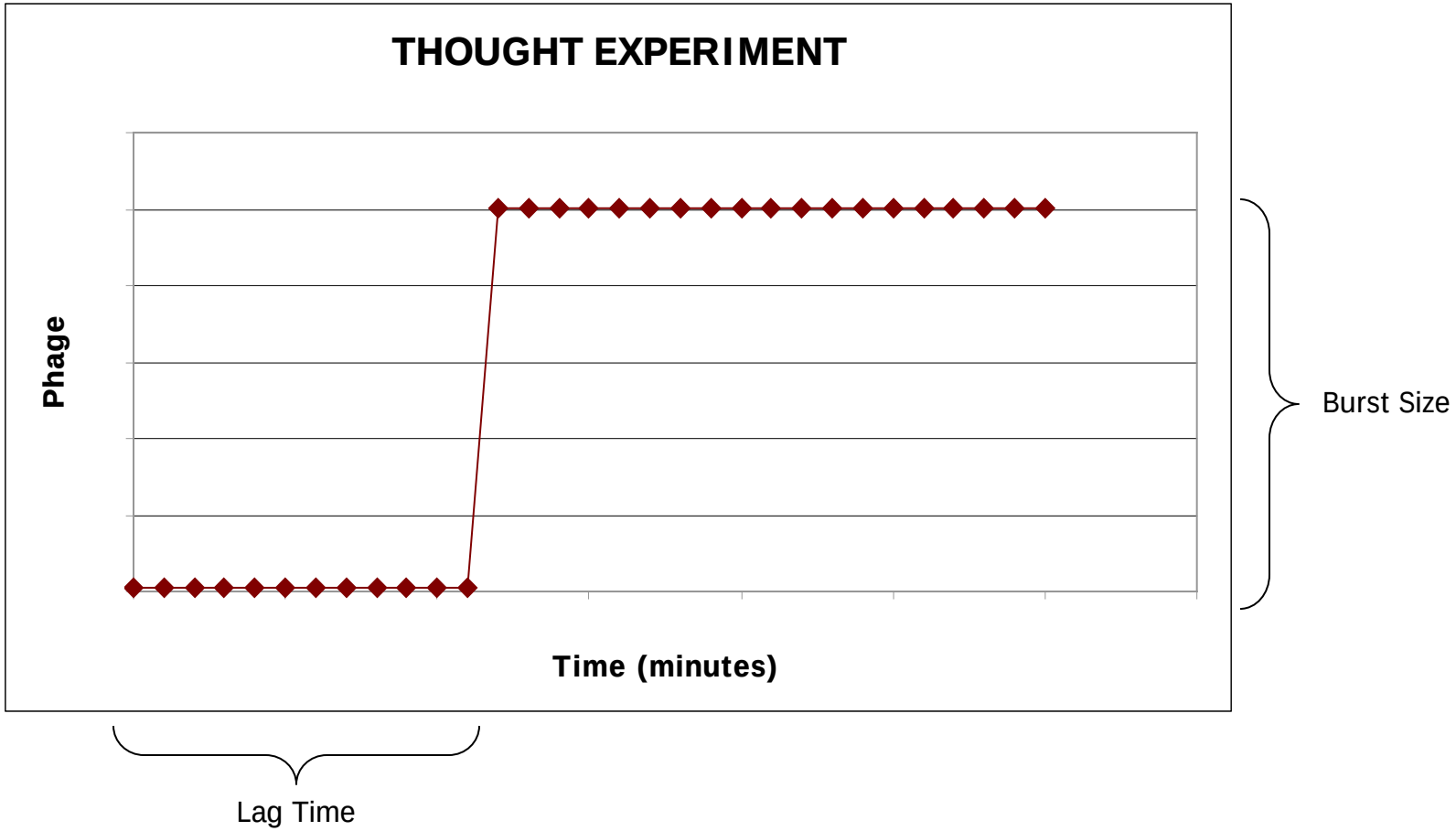
- Host Constants
  - C is the host cell density
  - d is the host cell death rate

Bulk Culture Assay



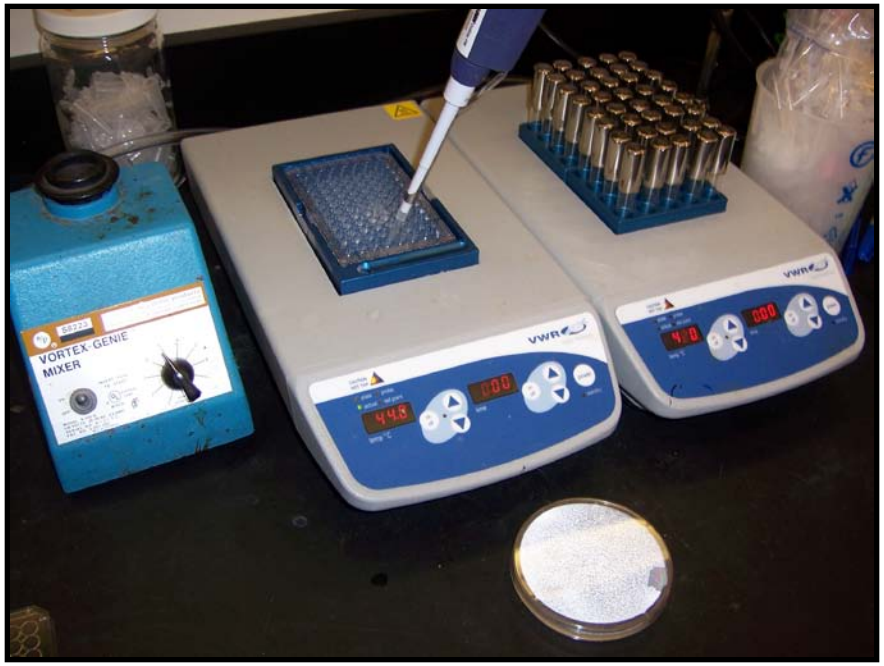
(Weinreich and Knies, unpub)

Can We Characterize Burst Size and Lag Time for an Individual Phage?

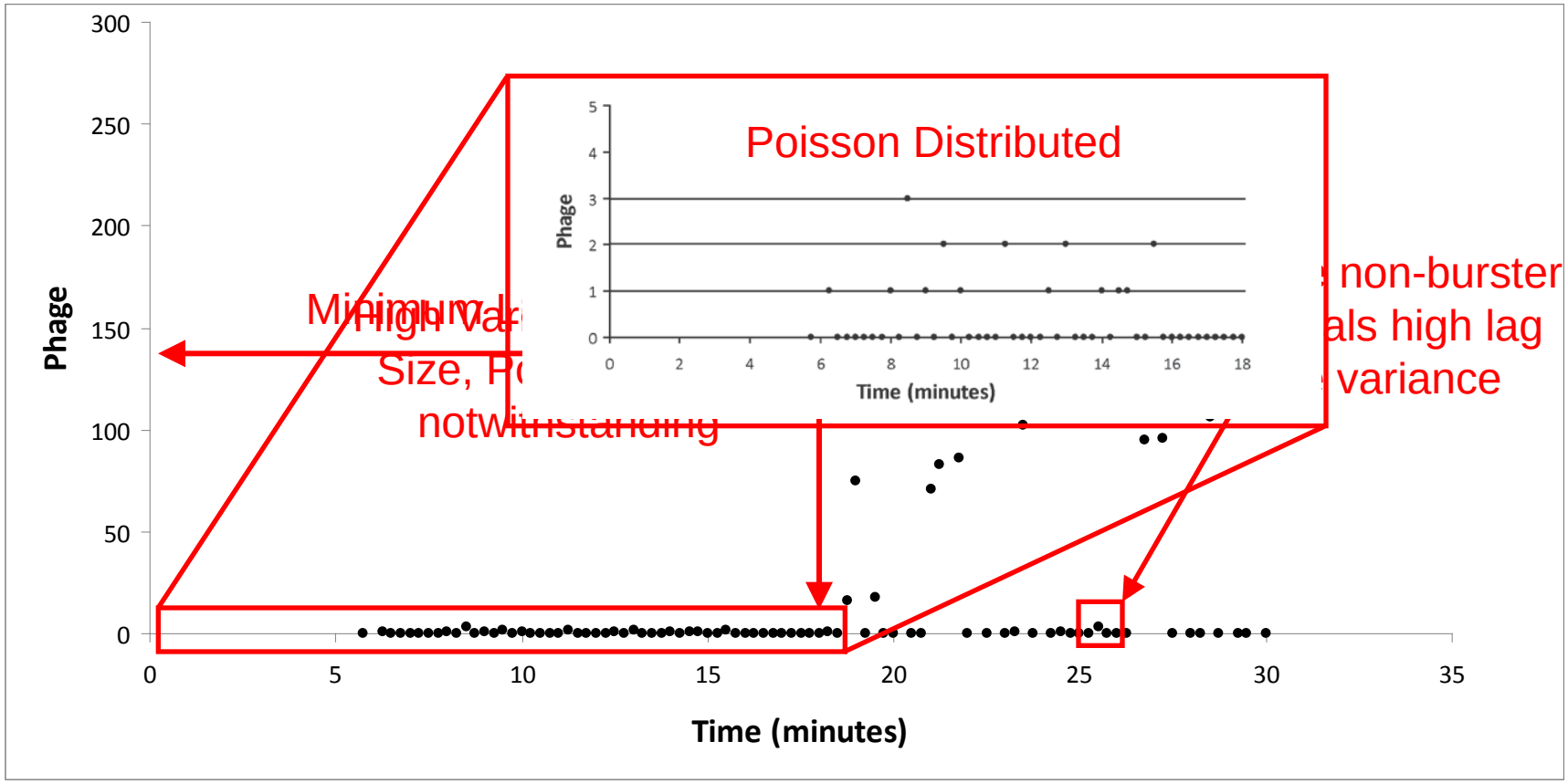


Method

- In each of 60 wells we add 100µl host cells and on average ~½ phage particle.
- Every 30 seconds we titrate the total number of phage in the next well.
- Note that this is destructive sampling.

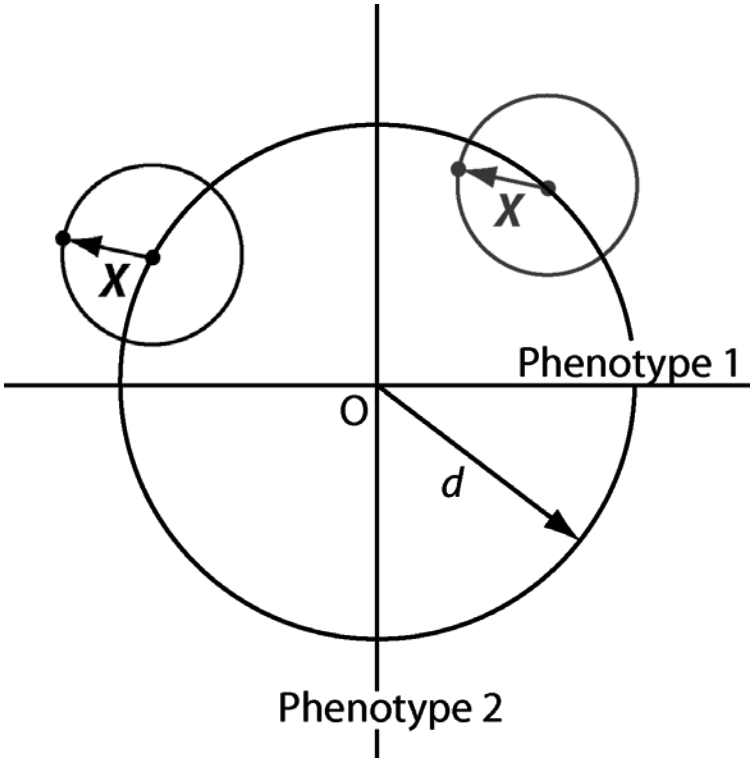


Sample Data



Fascinating problem: How to simultaneously estimate two distributions from temporal data?

RA Fisher's Geometric Model of Adaptation



All traits are under stabilizing selection and all mutations are pleiotropic.

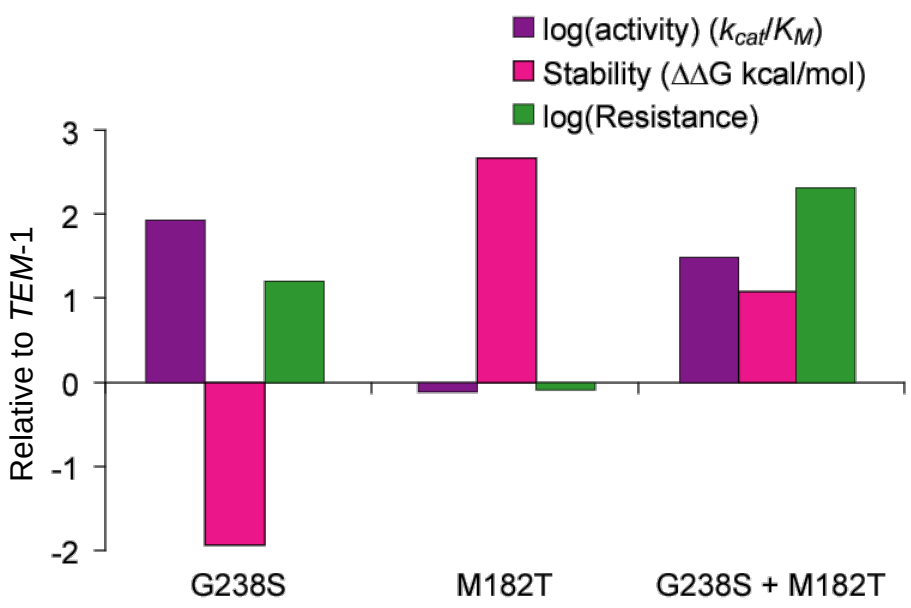
MISSENSE MEANDERINGS IN SEQUENCE SPACE: A BIOPHYSICAL VIEW OF PROTEIN EVOLUTION

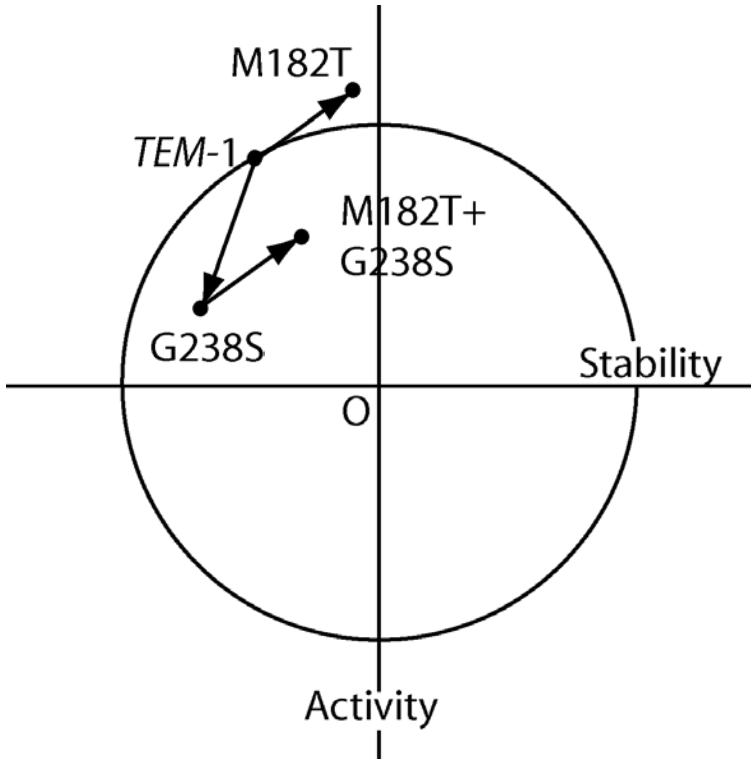
Mark A. DePristo, Daniel M. Weinreich and Daniel L. Hartl

678 | SEPTEMBER 2005 | VOLUME 6 www.nature.com/reviews/genetics

1. To be successful, a protein must perform a function (e.g. bind or catalyze), but it must also
   - Successfully fold into its native form
   - Avoid aggregation
   - Avoid premature degradation
2. Intermediate trait values are often optimal (called ‘stabilizing selection’).
3. Most mutations influence more than one trait (they act ‘pleiotropically’).

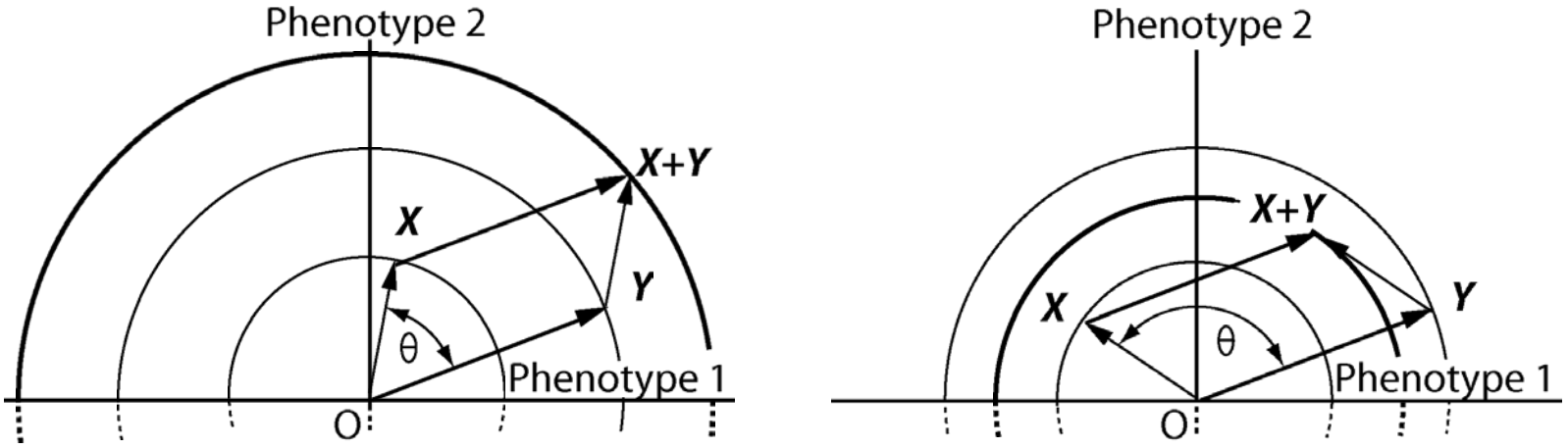
β-lactamase and Fisher





Weinreich (2010)

What Can We Learn from Epistasis between Two Deleterious Mutations

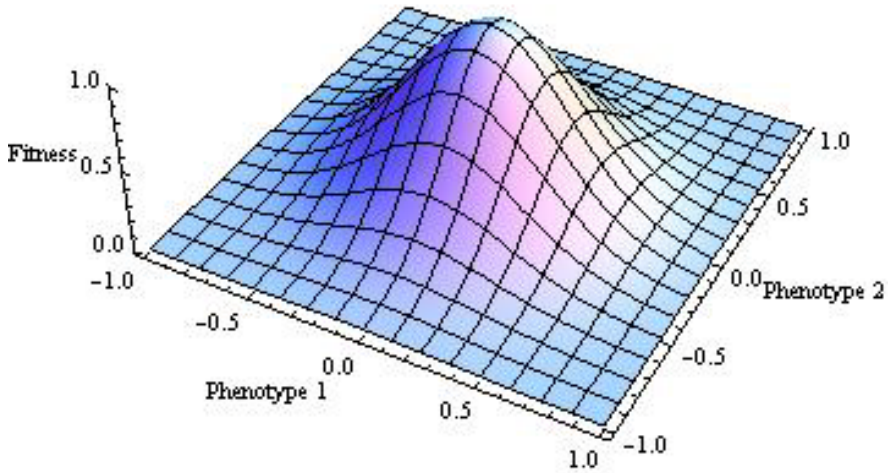


Comparing fitness effects of deleterious mutations singly with their effect in pairs appears to yield insight into the phenotypic angle θ between them.

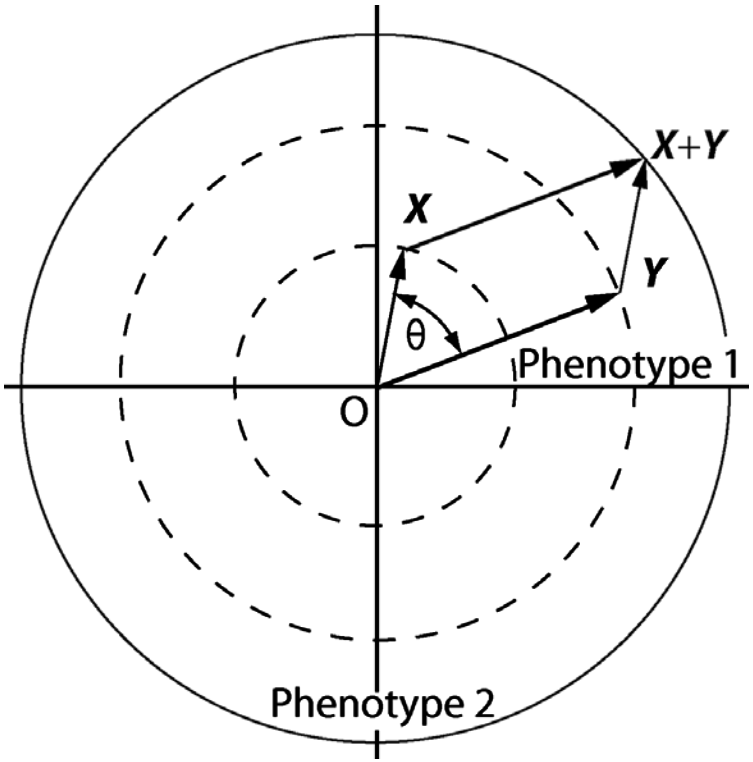
Make Two Assumptions…

Gaussian Fitness:

$W_z = \mathrm{Exp}[-\frac{1}{2}s|\boldsymbol{Z}|^2]$



Mutations are additive in phenotype space:



…to Cash Out the Algebra

$$W_{X+Y} = \mathrm{Exp}\{\ln[W_X] + \ln[W_Y] + 2\cdot(\ln[W_X]\cdot\ln[W_Y])^{1/2}\cdot\cos\theta\}$$

$$\theta = \arccos\left\{\frac{\ln\left[W_{X+Y} / W_X \cdot W_Y\right]}{2\sqrt{\ln W_X \cdot \ln W_Y}}\right\}$$

Given two mutations ***X*** and ***Y***:

1. We can use reverse genetics to put ***X***, ***Y*** and ***X*** + ***Y*** on the wild-type background.
2. Given fitness values for all three genotypes (***X, Y*** and ***X*** + ***Y***), we can compute θ between ***X*** and ***Y***.

How Might This Work?

	Site 1	Site 2	Site 3	Site 4	Site 5
Site 1	0	$\theta_{1,2}$	$\theta_{1,3}$	$\theta_{1,4}$	$\theta_{1,5}$
Site 2	$\theta_{1,2}$	0	$\theta_{2,3}$	$\theta_{2,4}$	$\theta_{2,5}$
Site 3	$\theta_{1,3}$	$\theta_{2,3}$	0	$\theta_{3,4}$	$\theta_{3,5}$
Site 4	$\theta_{1,4}$	$\theta_{2,4}$	$\theta_{3,4}$	0	$\theta_{4,5}$
Site 5	$\theta_{1,5}$	$\theta_{2,5}$	$\theta_{3,5}$	$\theta_{4,5}$	0

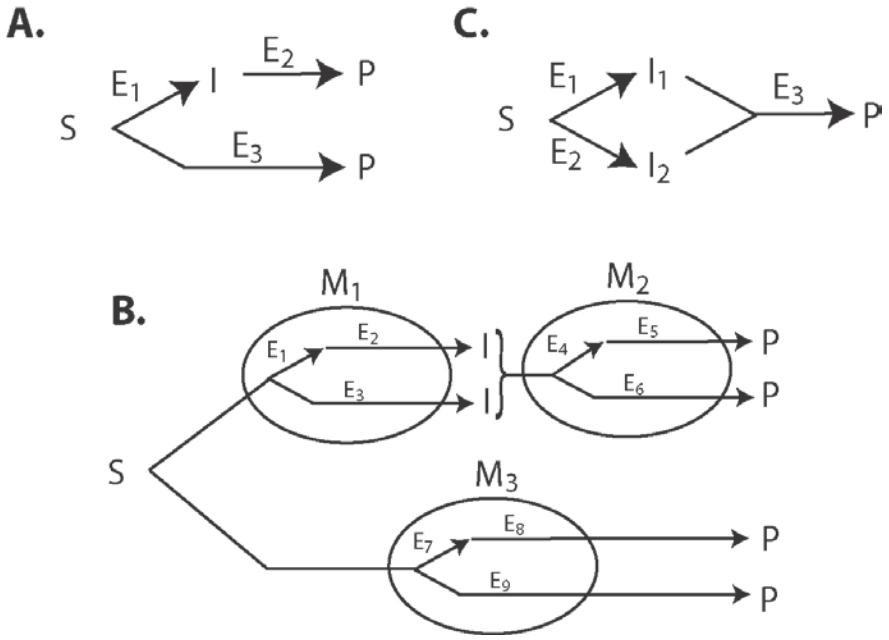
How Might This Work?

	Site 1	Site 2	Site 3	Site 4	Site 5
Site 1	0	$\theta_{1,2}$	$\theta_{1,3}$	$\theta_{1,4}$	$\theta_{1,5}$
Site 2	$\theta_{1,2}$	0	$\theta_{2,3}$	$\theta_{2,4}$	$\theta_{2,5}$
Site 3	$\theta_{1,3}$	$\theta_{2,3}$	0	$\theta_{3,4}$	$\theta_{3,5}$
Site 4	$\theta_{1,4}$	$\theta_{2,4}$	$\theta_{3,4}$	0	$\theta_{4,5}$
Site 5	$\theta_{1,5}$	$\theta_{2,5}$	$\theta_{3,5}$	$\theta_{4,5}$	0

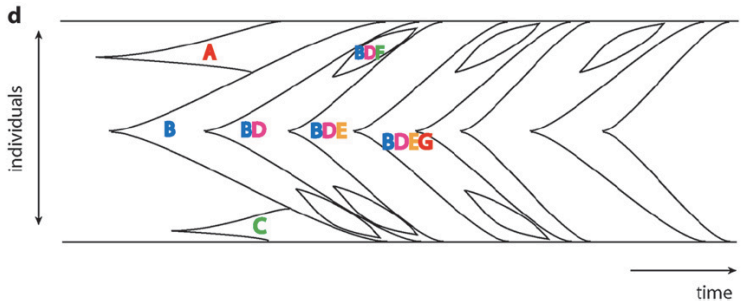
Inferences

- Extreme θ (≈ 0° or ≈ 180°) ⇕ Mutations ***X*** and ***Y*** affect the same phenotypes. If many mutations all have small pairwise values of θ, then perhaps they don't all need to be characterized for phenotypic effects.
- θ ≈ 90° ⇕ Mutations ***X*** and ***Y*** affect distinct phenotypes. If many mutations all have pairwise values of θ ≈ 90°, then there must be many phenotypes. (Formally, the vectors corresponding to these many mutations together form a basis of phenotype space whose dimensionality is given by the number of positive eigenvalues of the previous matrix.)

Metabolic Control Analysis



MCA predicts that deleterious mutations in pairs of genes will have patterns of epistasis reflecting network topology. Can we use data on epistasis between deleterious mutations in pairs of genes to make inferences about network topology? (Szathmáry 1993, Segre *et al*. 2005)



Dynamics of adaptation:
Population structure
Clonal interference
Multiple mutations

Part III: Dynamics of Adaptation

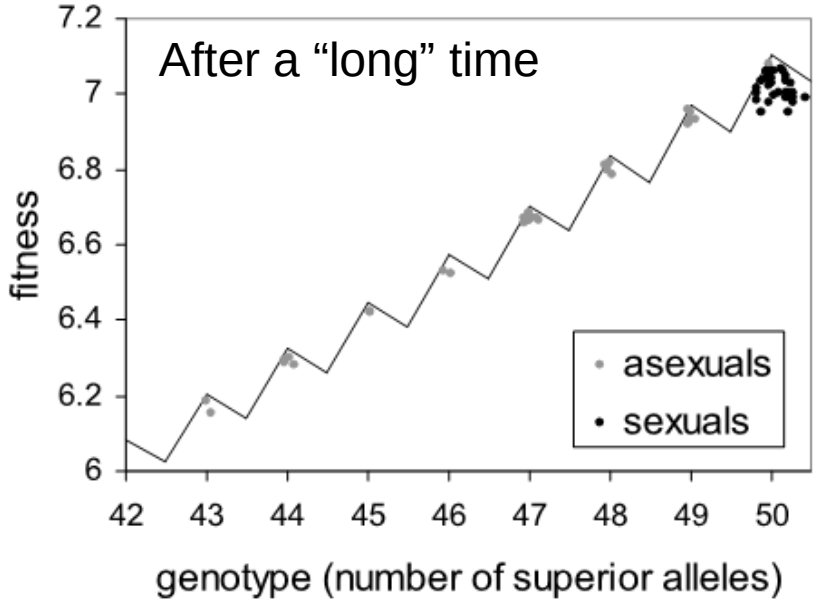
Recombination

1. Landscapes Solved in Linear Time with Sex
2. Crossing Fitness Valleys with Recombination: Whither the Landscape?

Genome Structure and the Benefit of Sex

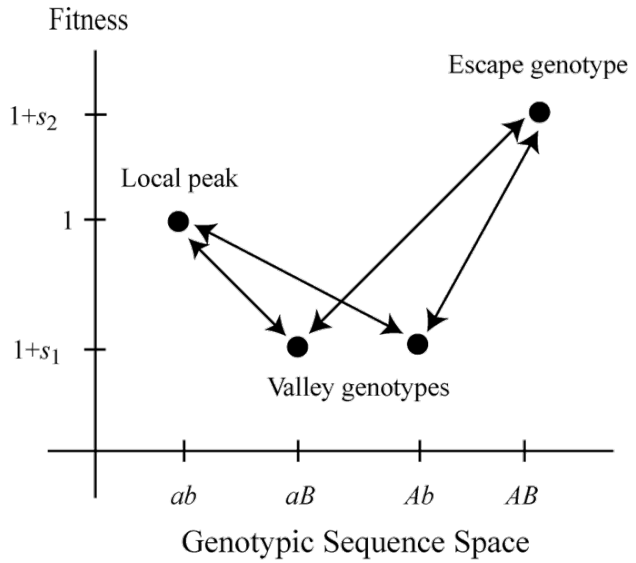
Locus 1	Locus 2	Locus 3	Locus 4	…	Locus *g*

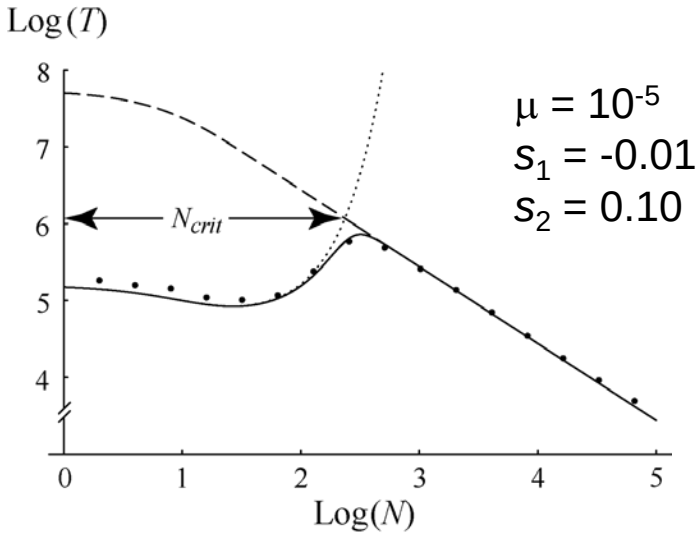
$W_G = \Pi w_i$ (thus, no epistasis *among* loci)
w_i is rugged (thus, sharp sign epistasis within loci)



(Watson, Weinreich and Wakeley 2010)

Time Between Adjacent Peaks





$\mu = 10^{-5}$
$s_1 = -0.01$
$s_2 = 0.10$

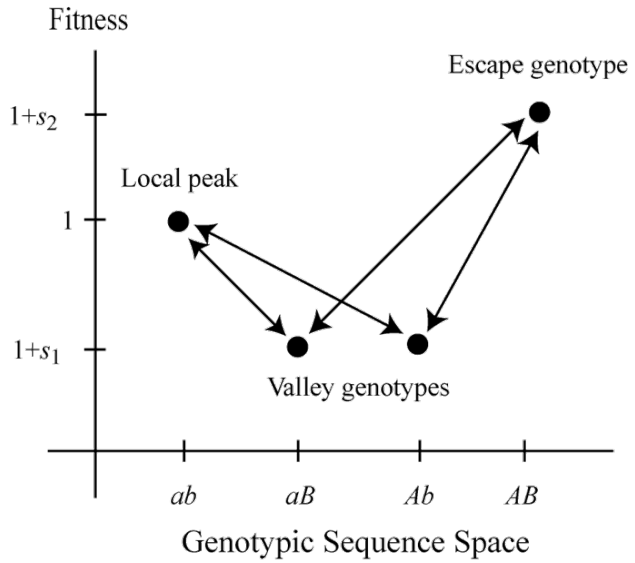
Sequential Fixation: $T_{seq} = \frac{1}{2N\mu \cdot \pi(N,-s_1)} + \frac{1}{N\mu \cdot \pi(N,s_1+s_2)}$

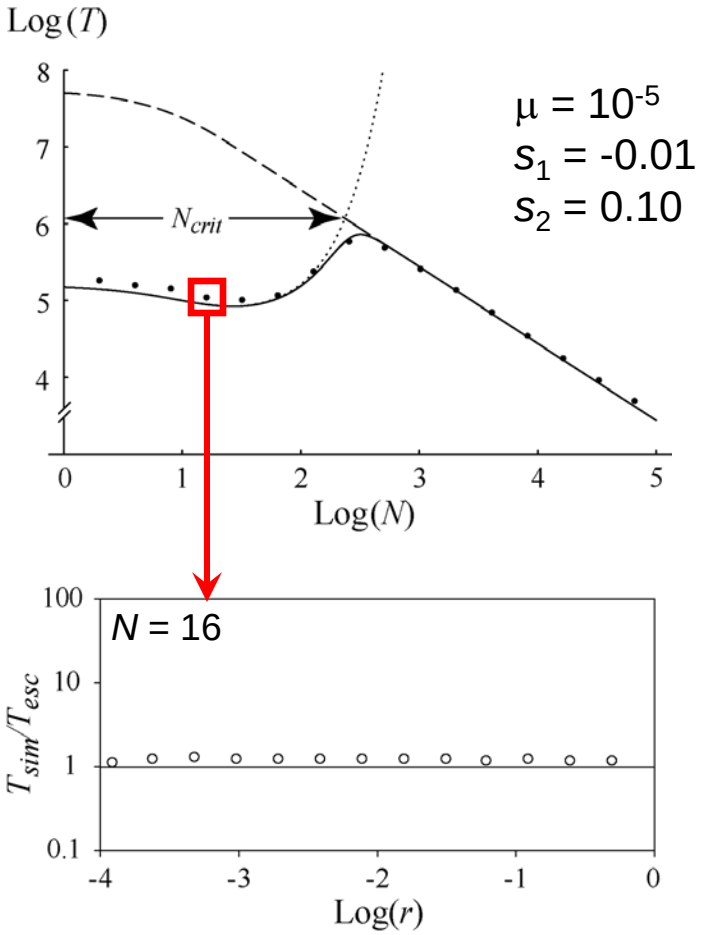
Simultaneous Fixation: $T_{sim} = \frac{s_1}{2N\mu^2 \cdot \pi(N,s_2)}$

Net: $T_{esc} = \frac{1}{\frac{1}{T_{seq}} + \frac{1}{T_{sim}}}$

(Carter and Wagner 2002, Iwasa et al 2004, Weinreich and Chao 2005)

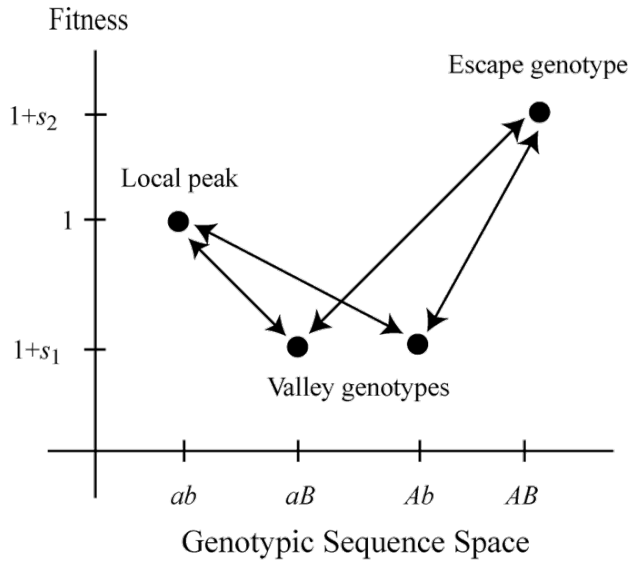
What about Recombination?

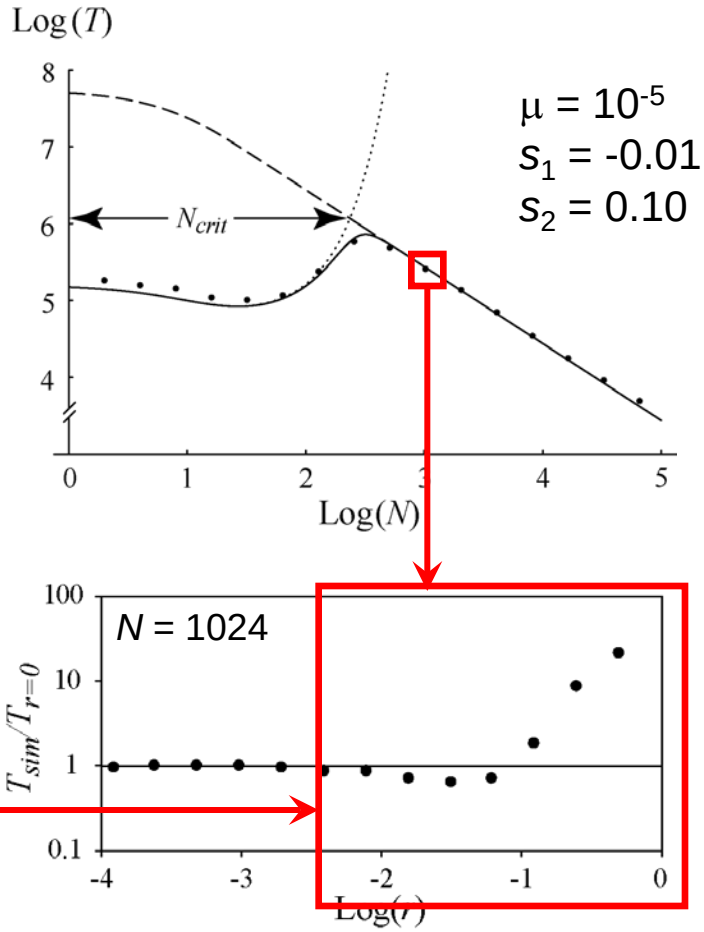




(Weinreich and Chao 2005)

What about Recombination?





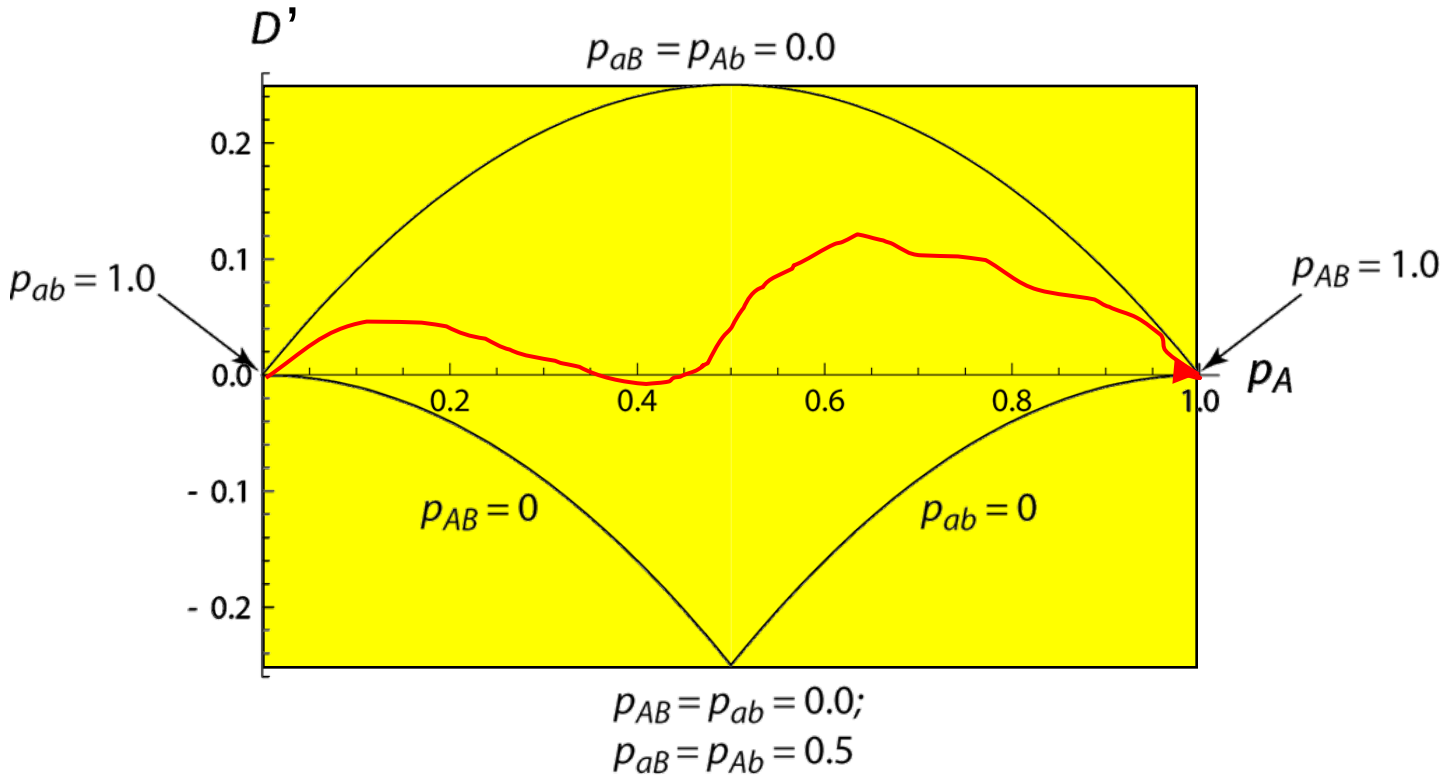
Can we visualize these trajectories?

(Weinreich and Chao 2005)

Need to Track Frequencies

- Genotype frequencies? Somewhat ugly because each difference equation depends on all four state variables.
- We've employed allele frequencies and linkage disequilibrium because it respects the genetics of mutation and recombination, which act atomically on loci and breakpoints between loci.
  - $\Delta_r p_A = \Delta_r p_b = 0$; $\Delta_r D = -r \cdot D$
  - $\Delta_\mu p_A = \mu(1 - 2p_A)$; $\Delta_r D = -4D(1 - \mu)\mu$

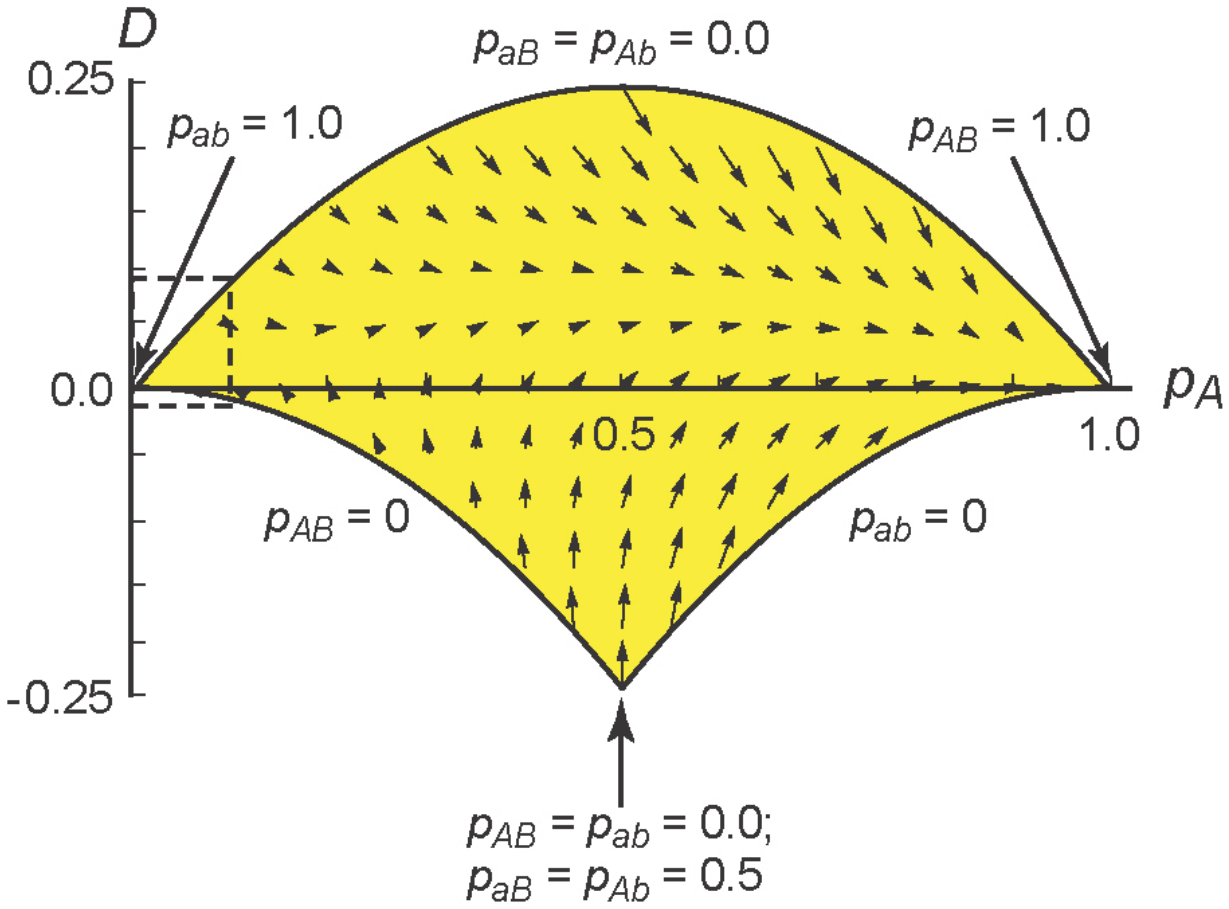
(See also Lewontin's *D'*)



In general, how do recombination, mutation and selection determine what trajectory an evolving population will follow?

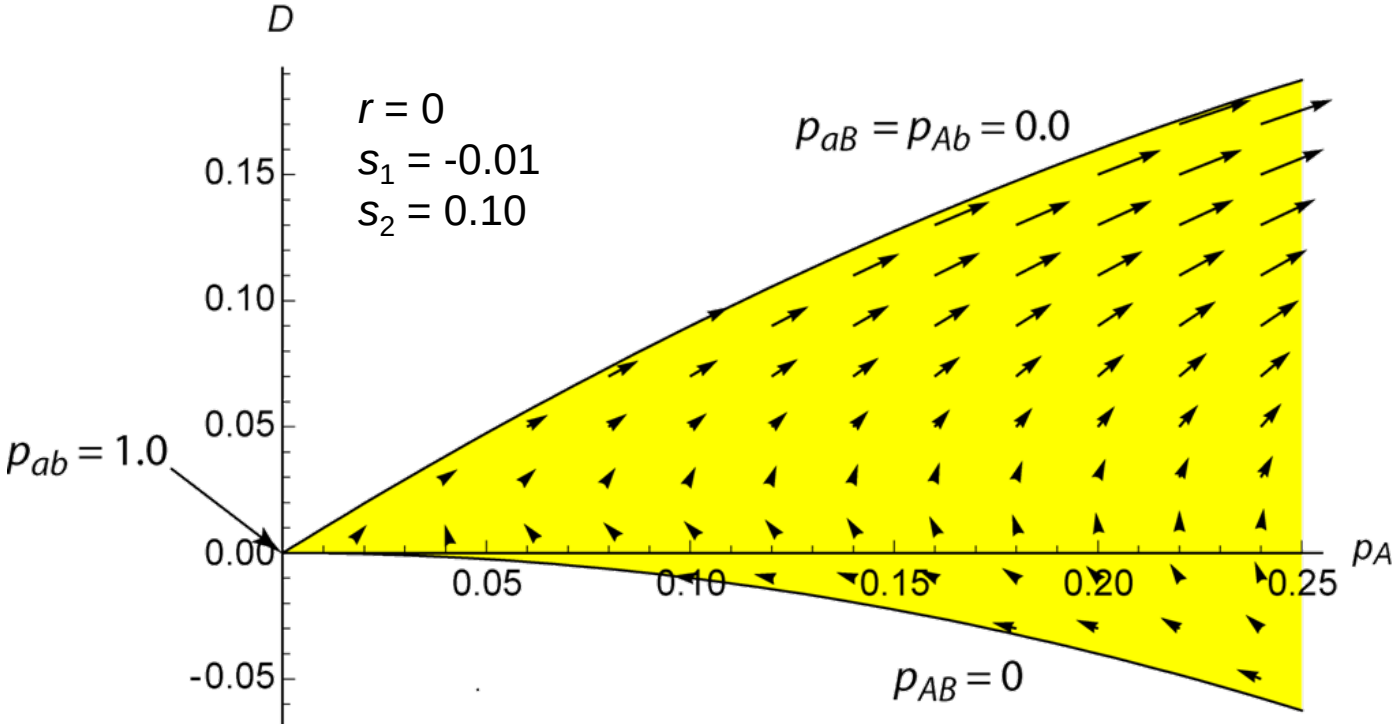
Lewontin 1988

Recombination + Two Fitness Maxima

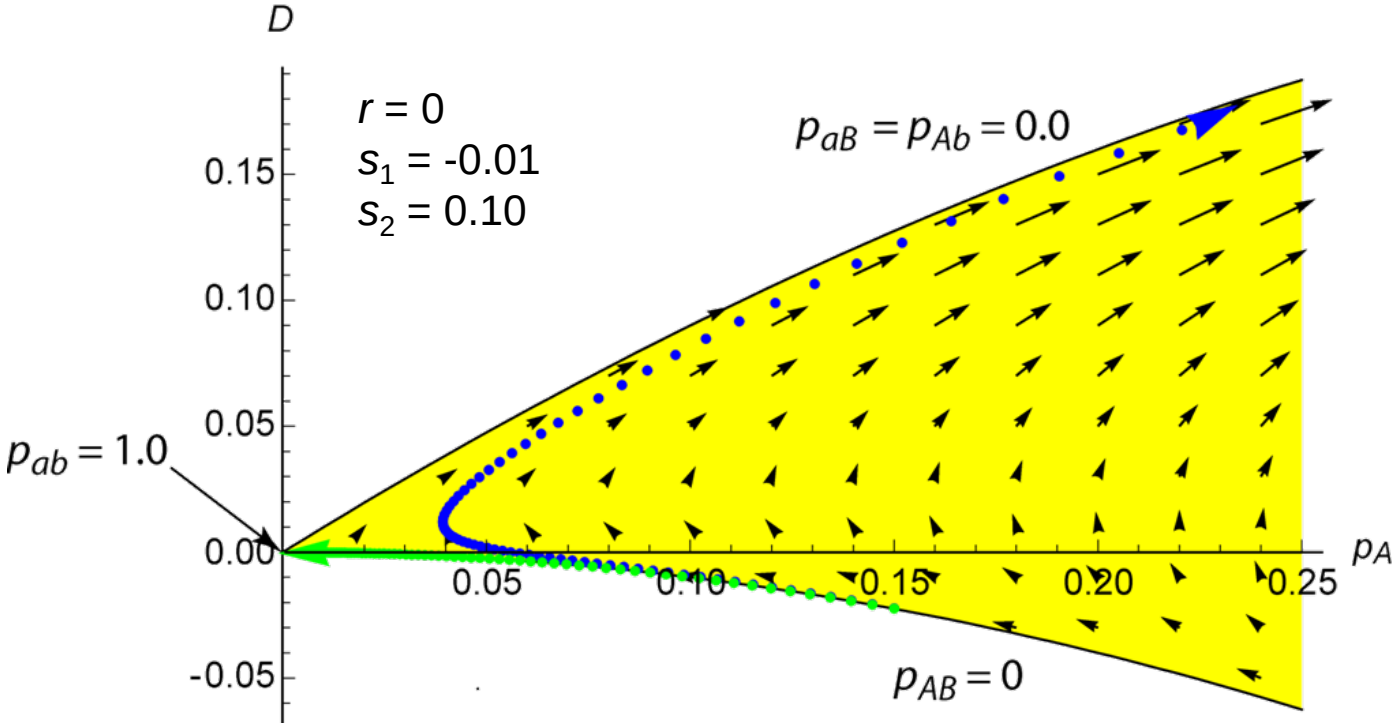


$r = 0$
$s_1 = -0.01$
$s_2 = 0.10$

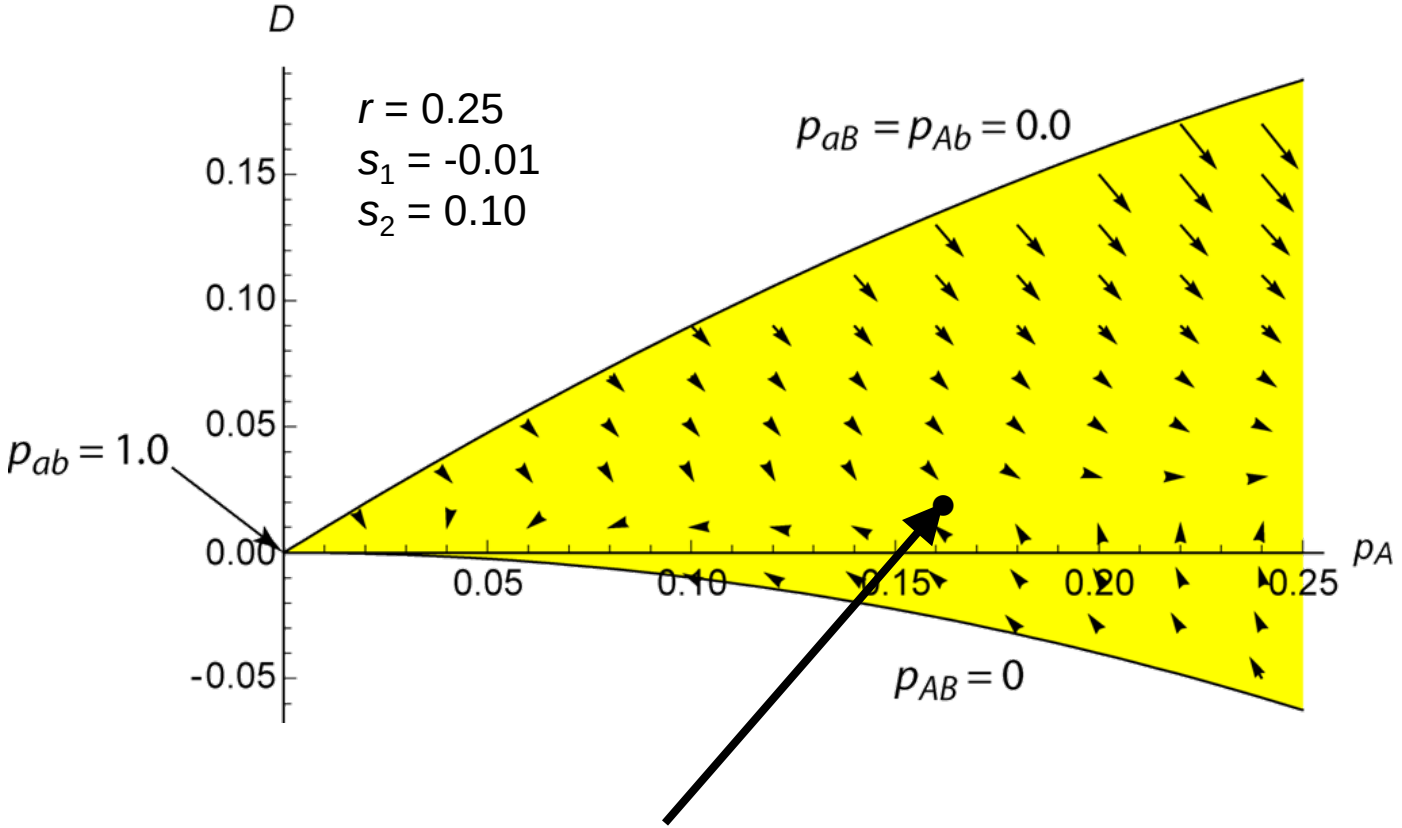
Recombination + Two Fitness Maxima



Recombination + Two Fitness Maxima

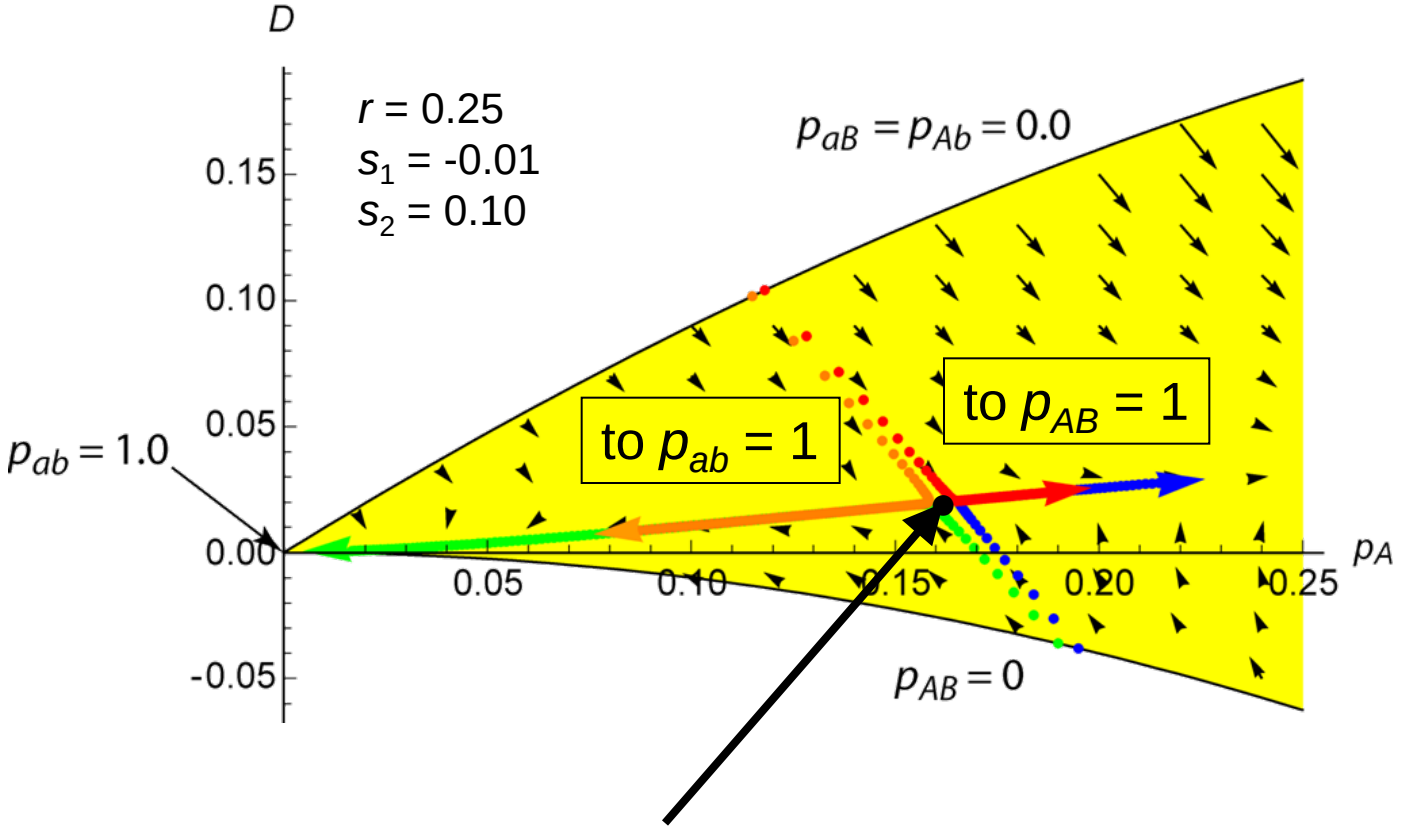


Recombination + Two Fitness Maxima



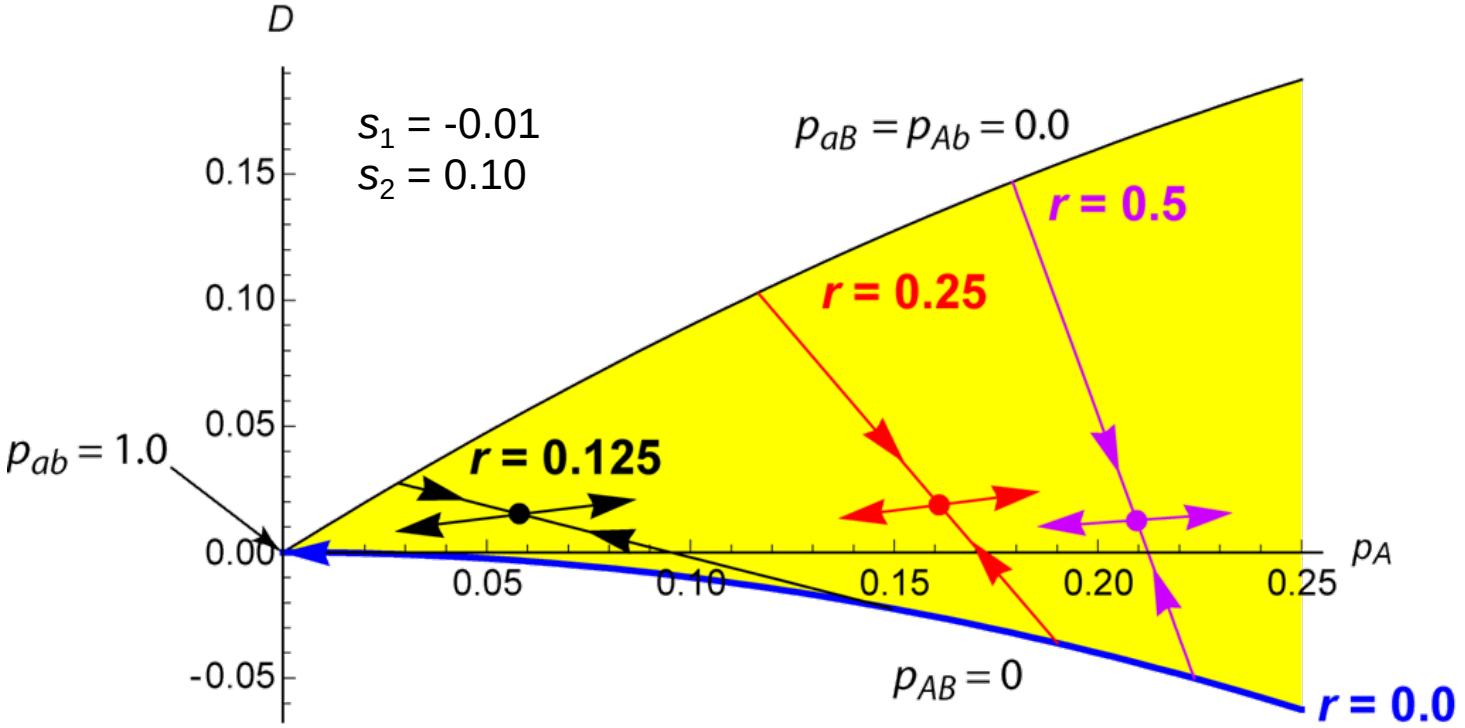
Saddle fixed-point emerges when $r > s_2/(1 + s_2)$; see also Crow and Kimura (1965).

Recombination + Two Fitness Maxima



Saddle fixed-point emerges when $r > s_2/(1 + s_2)$; see also Crow and Kimura (1965).

Recombination + Two Fitness Maxima

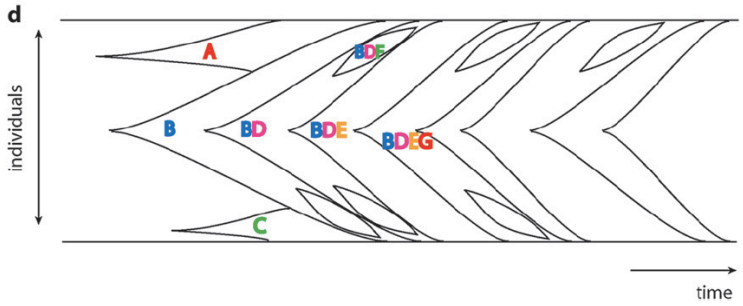


Can analytically locate the fixed points and approximate the corresponding boundaries between basins of attraction as a function of selection and recombination.

Whither the Fitness Landscape?

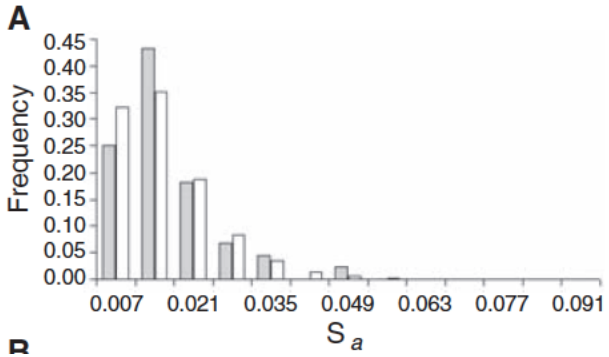
- Although the fitness landscape isn't continuous, subject to SSWM assumptions it is predictive. Why? Because it's a potential function, and the local gradient defines the direction an evolving population is likely to move.
- Is there a potential function over $p_A \times D$ - space? No. In point of fact our vector field $(\Delta p_A, \Delta D)$ corresponds to no potential function. (Formally, $\partial(\Delta p_A)/\partial D \neq \partial(\Delta D)/\partial p_A$.)
- Interestingly, violating SSWM appears to also render the fitness landscape less predictively useful because the fate of any lineage now depends on the fitnesses of whoever else is cosegregating.
- Speculate: Predictive population landscapes do not exist.

Adaptation



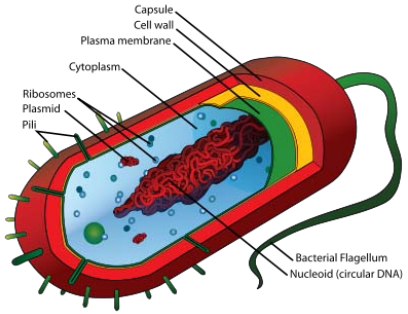
Dynamics of adaptation:

Population structure
Clonal interference
Multiple mutations



Fitness landscape:

Mutation rates
Distribution of fitness effects
Epistasis



Environment + cellular architecture:

Regulatory networks
Proteins

Acknowledgements

- Dan Hartl
- Richard Watson
- Nigel Delaney
- Mark DePristo
- Kyle Brown
- Jen Knies
- Brendan Hickey
- Jeffrey Yuan
- Meghan Hollibaugh Baker

Brown University
Division of Biology and Medicine,
Center for Computational Molecular Biology