POSITIVE AND NEGATIVE SELECTION
(AND RELATED PROBLEMS)

CLAUDIA BANK

INSTITUTO
GULBENKIAN
DE CIÊNCIA

Evolutionary Dynamics @ IGC:

- How do populations adapt to challenging environments?
  E.g., how does drug resistance evolve?
- Which processes drive speciation & diversification?
- What is the role of interactions in evolution?

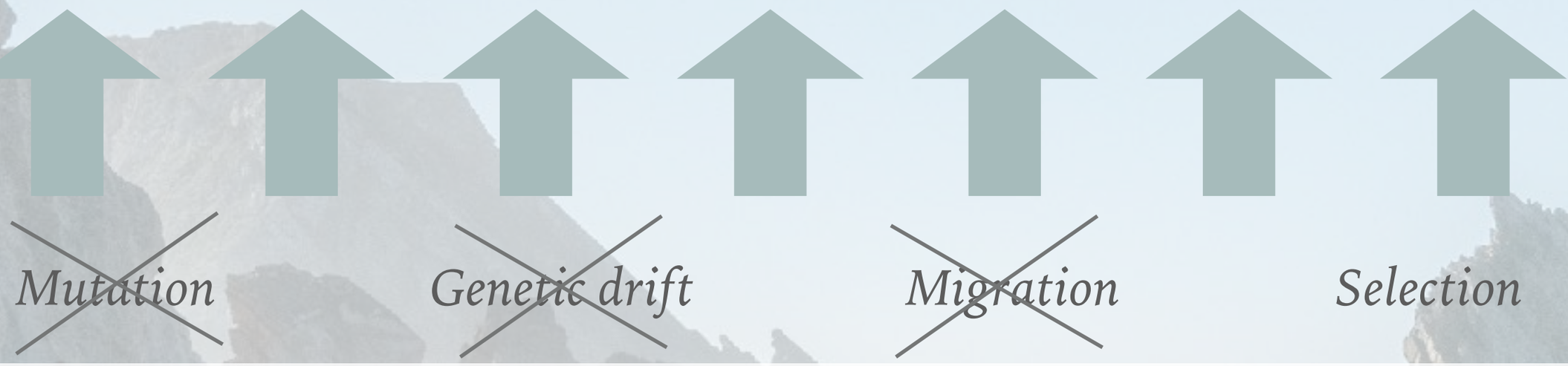
Mutation *Genetic drift* *Migration* *Selection*

What we do

- Study evolutionary processes using simple models
- Evaluate these models using empirical and simulated data
- Use modeling to inform experimental design *a priori*

Evolutionary Dynamics @ IGC:

- How do populations adapt to challenging environments?
  E.g., how does drug resistance evolve?
- Which processes drive speciation & diversification?
- What is the role of interactions in evolution?



What we do

- Study evolutionary processes using simple models
- Evaluate these models using empirical and simulated data
- Use modeling to inform experimental design *a priori*

> It may be said that natural selection is daily and hourly scutinising, throughout the world, every variation, even the slightest; rejecting that which is bad, preserving and adding up all that is good; silently and insensibly working , whenever and wherever opportunity offers, at the improvement of each organic being in relation to its organic and inorganic conditions of life. We see nothing of these slow changes in progress, until the hand of time has marked the long lapse of ages.
>
> *- Darwin, 1859*

> It may be said that natural selection is daily and hourly scutinising, throughout the world, every variation, even the slightest; rejecting that which is bad, preserving and adding up all that is good; silently and insensibly working , whenever and wherever opportunity offers, at the improvement of each organic being in relation to its organic and inorganic conditions of life. We see nothing of these slow changes in progress, until the hand of time has marked the long lapse of ages.
>
> *- Darwin, 1859*

If you were to write a book about evolution, what would you introduce first, and why?

Mutation *Genetic drift* *Migration* *Selection*

NATURAL SELECTION REQUIRES

- Variation
- Inheritance
- Differential reproductive success

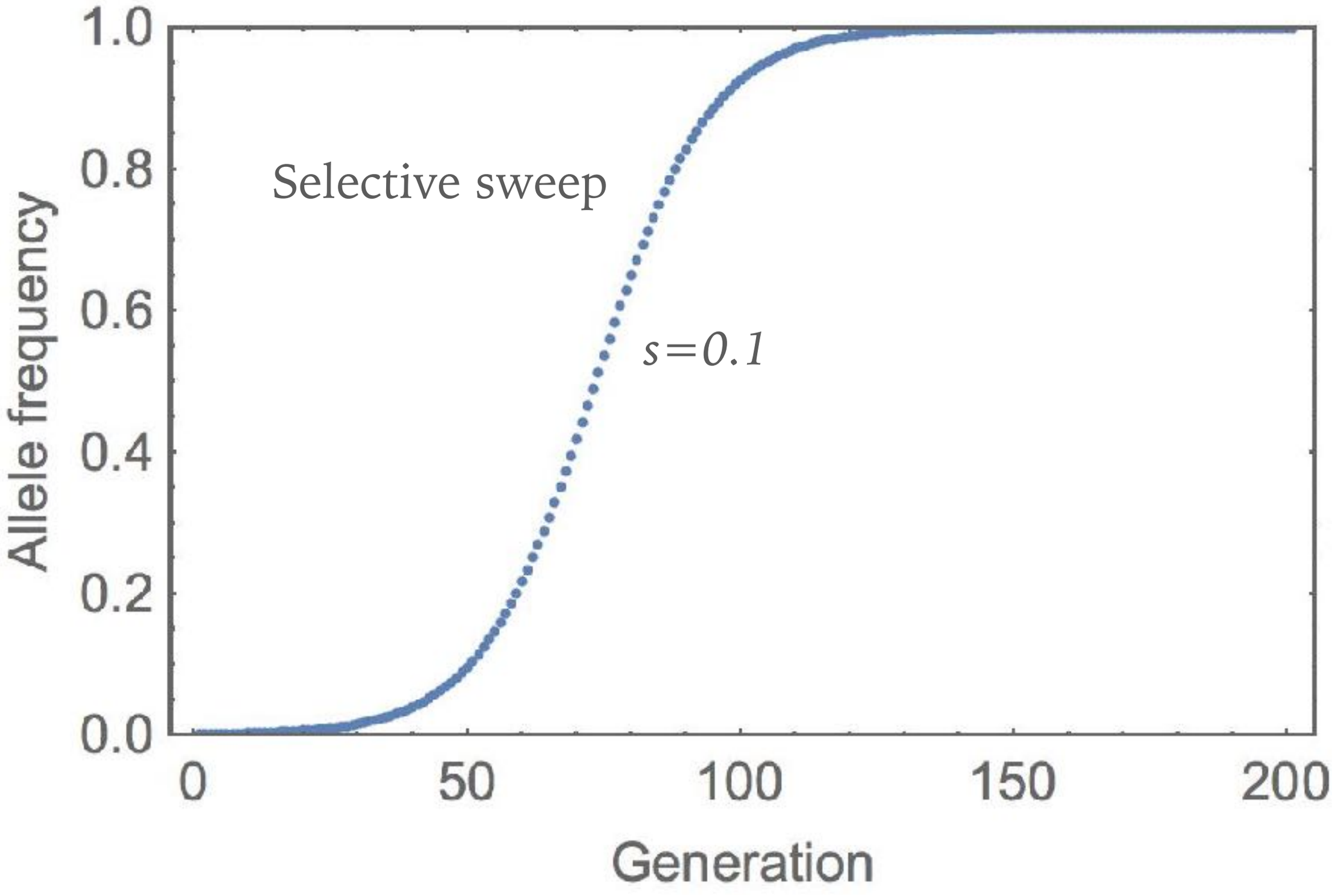
> Although many processes shape evolution, natural selection is special because it creates complex, functioning organisms. All other processes tend to degrade what has been built up by natural selection, simply because these processes act at random with respect to function.

-Barton et al. , Evolution (textbook)

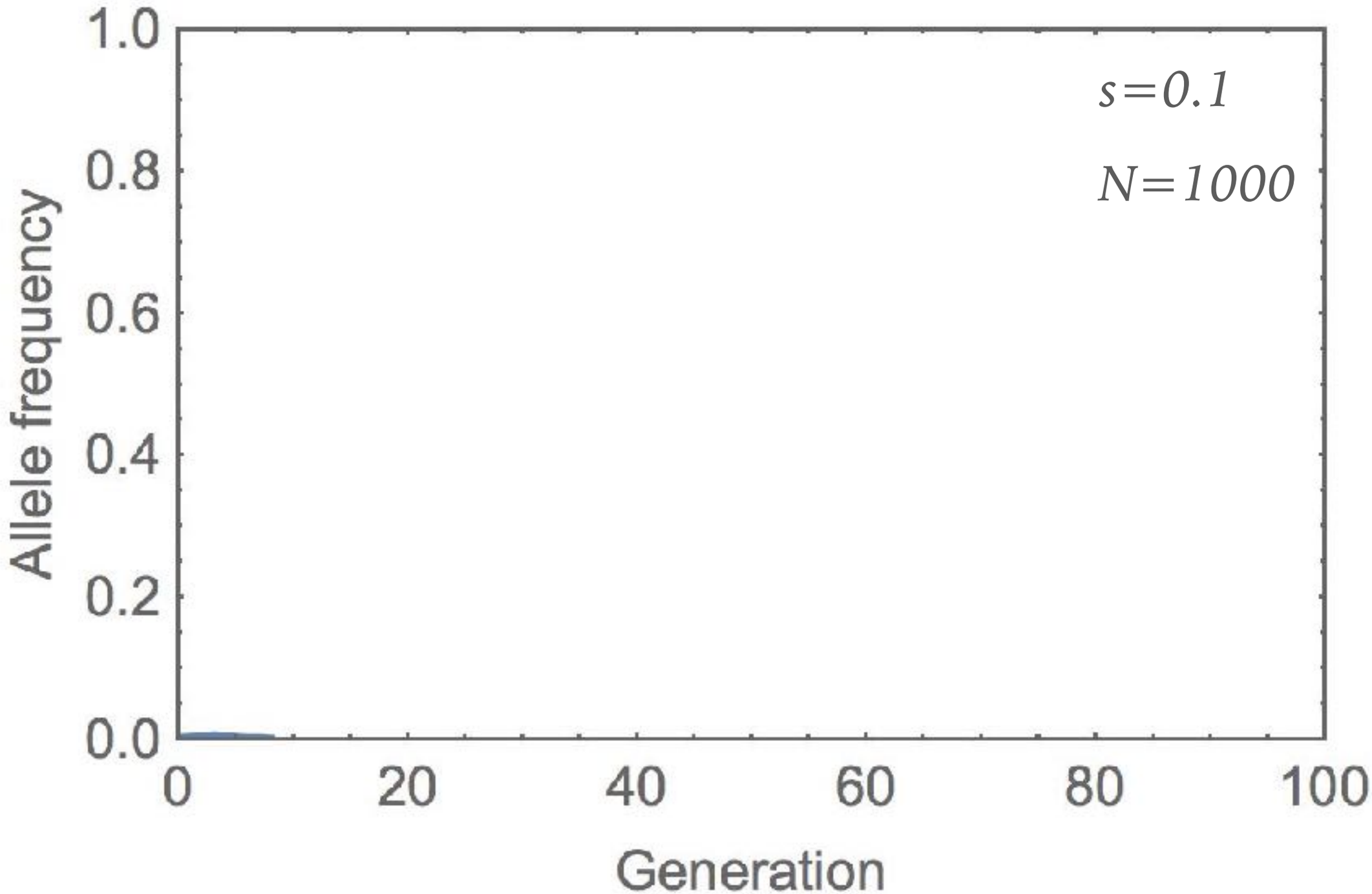
WHAT WE WANT TO KNOW ABOUT SELECTION

- How big/small are adaptive steps?
- What are the proportions of beneficial, neutral, and deleterious mutations?
- How do mutational effects change dependent on the environment?
- How do mutational effects change dependent on the genetic background? (I.e., what is the role of epistasis?)
- What is the role of selection vs. other evolutionary processes in shaping genomes?
- How can we infer the contribution of selection to molecular evolution?

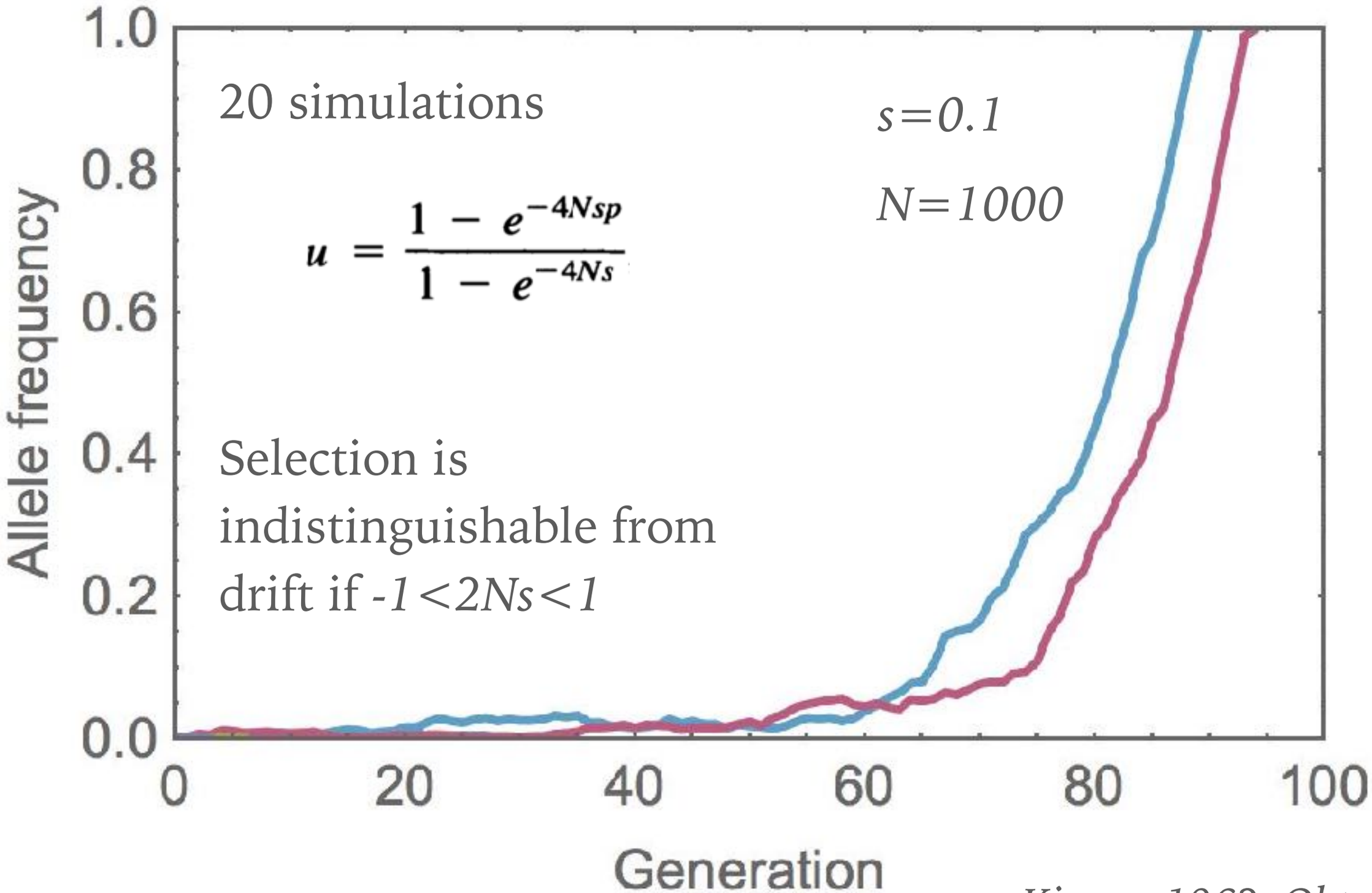
THEORETICALLY, SELECTION IS THE “EASIEST" EVOLUTIONARY FORCE



BUT THEN, GENETIC DRIFT COMES ALONG

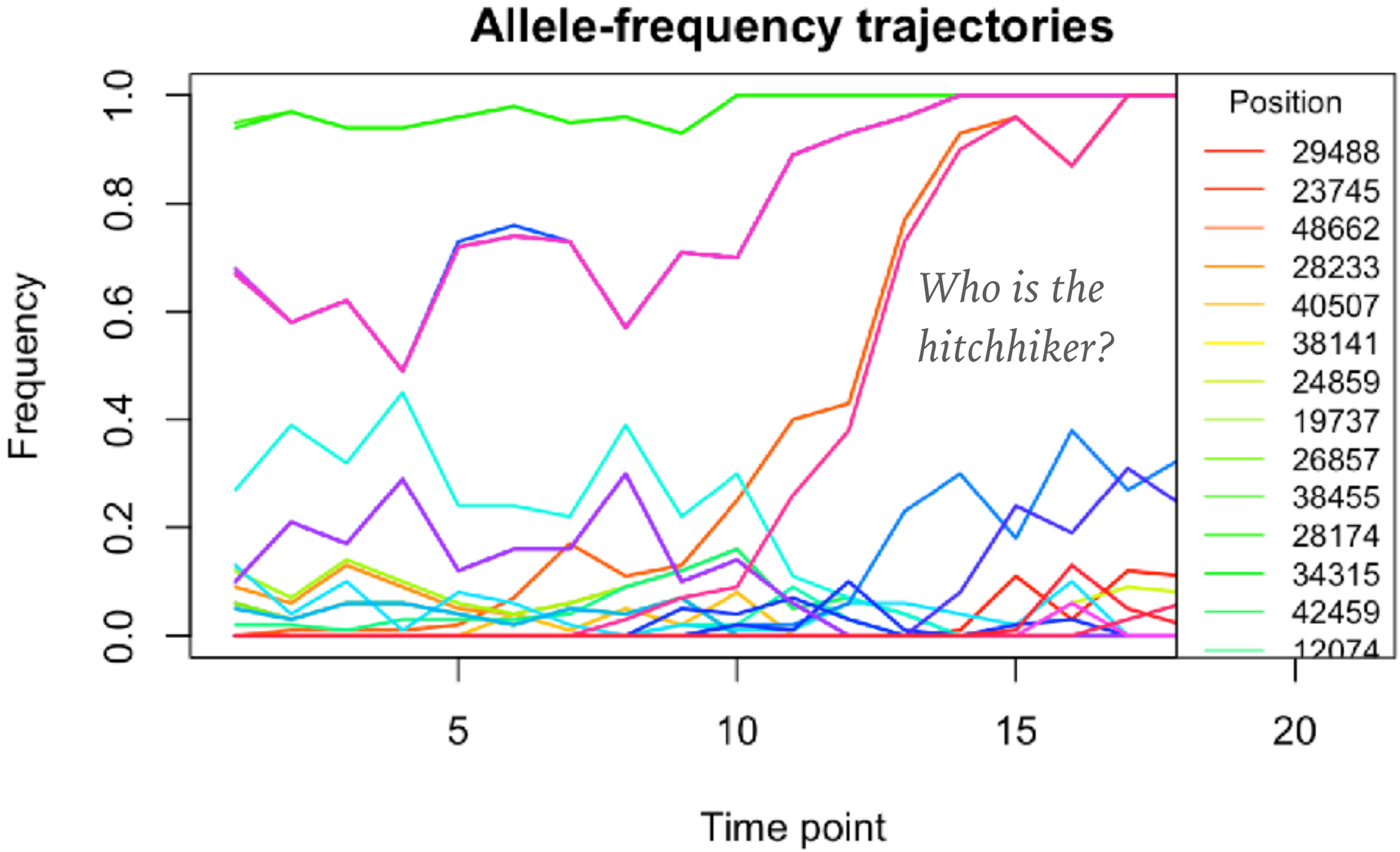


BUT THEN, GENETIC DRIFT COMES ALONG



Kimura 1962, Ohta 1973

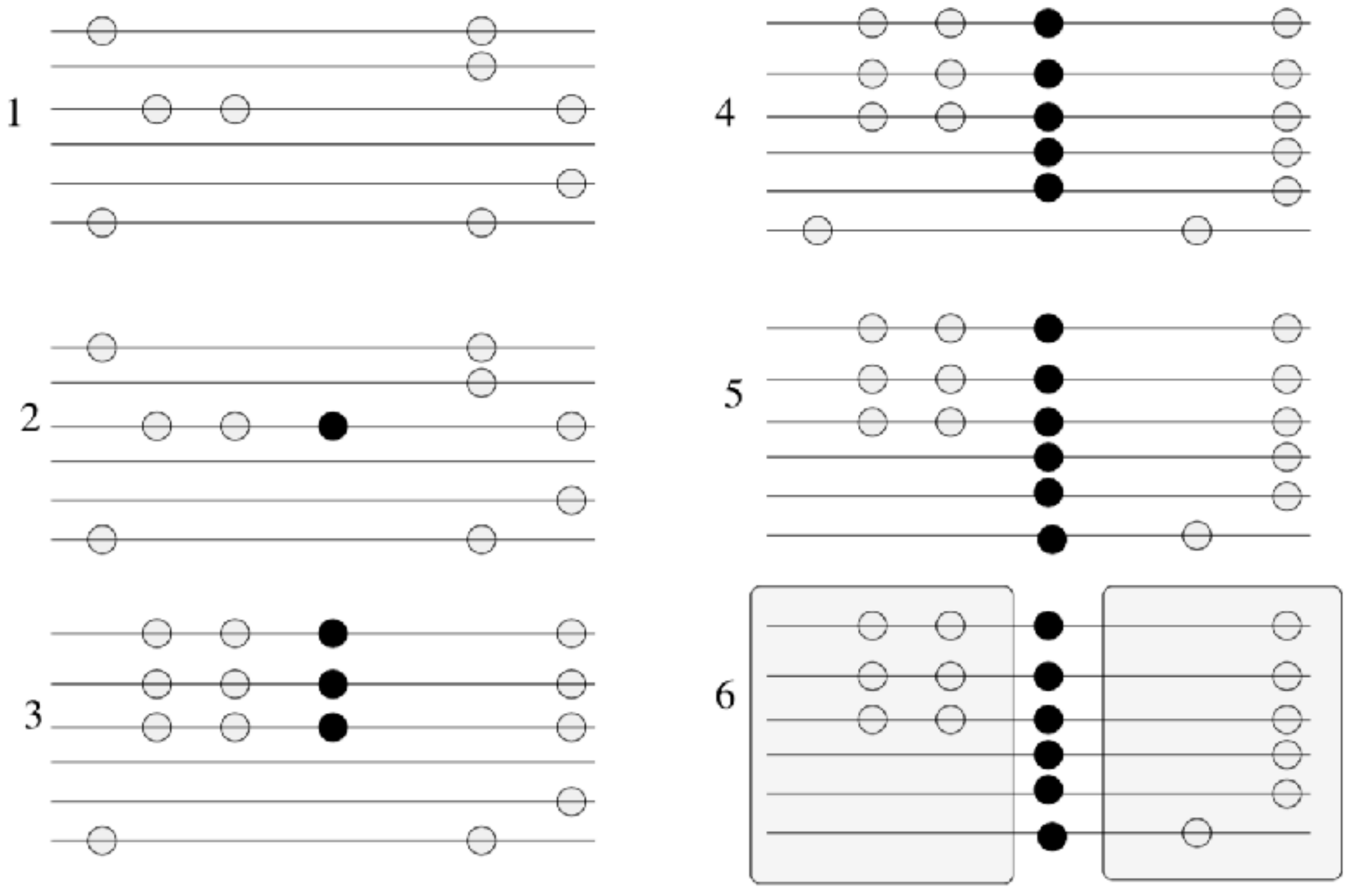
OR LINKAGE


Allele-frequency trajectories
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
Time point
Who is the hitchhiker?
Position
29488
23745
48662
28233
40507
38141
24859
19737
26857
38455
28174
34315
42459
12074

OR POPULATION STRUCTURE, OR EPISTASIS, OR [ADD YOUR FAVORITE HERE]

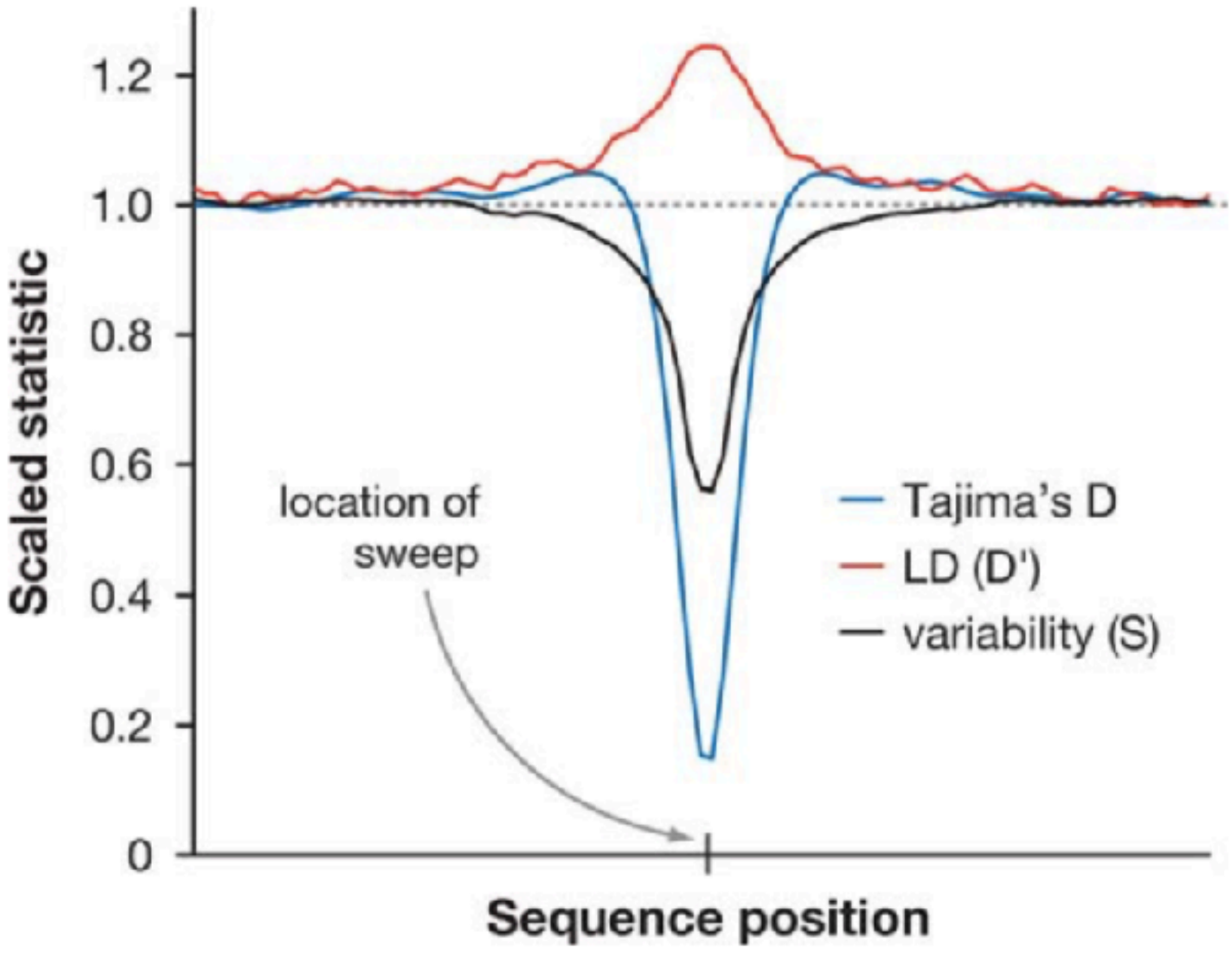
- Keep in mind that selection operates on phenotypic differences among individuals in a population; it does not act on a genotype, much less an allele.

SELECTION LEAVES TRACES IN GENOMES



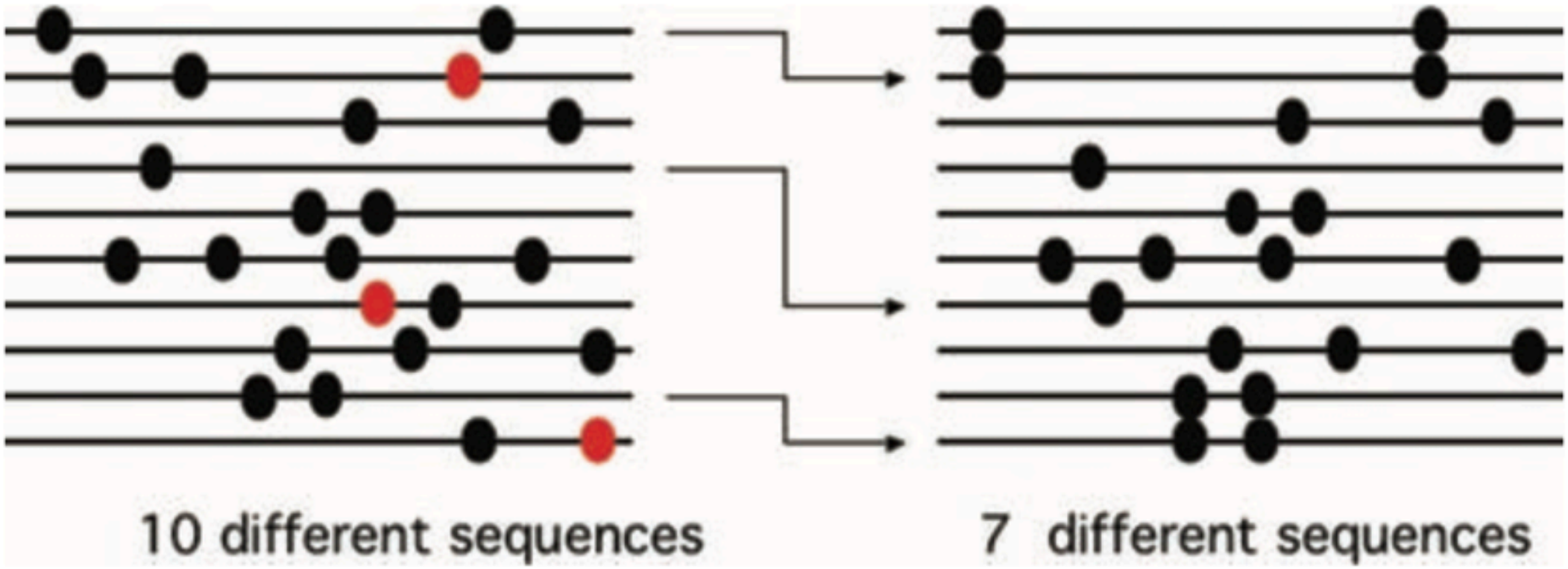
Pavlidis and Alachiotis 2017

TRACES OF A SELECTIVE SWEEP



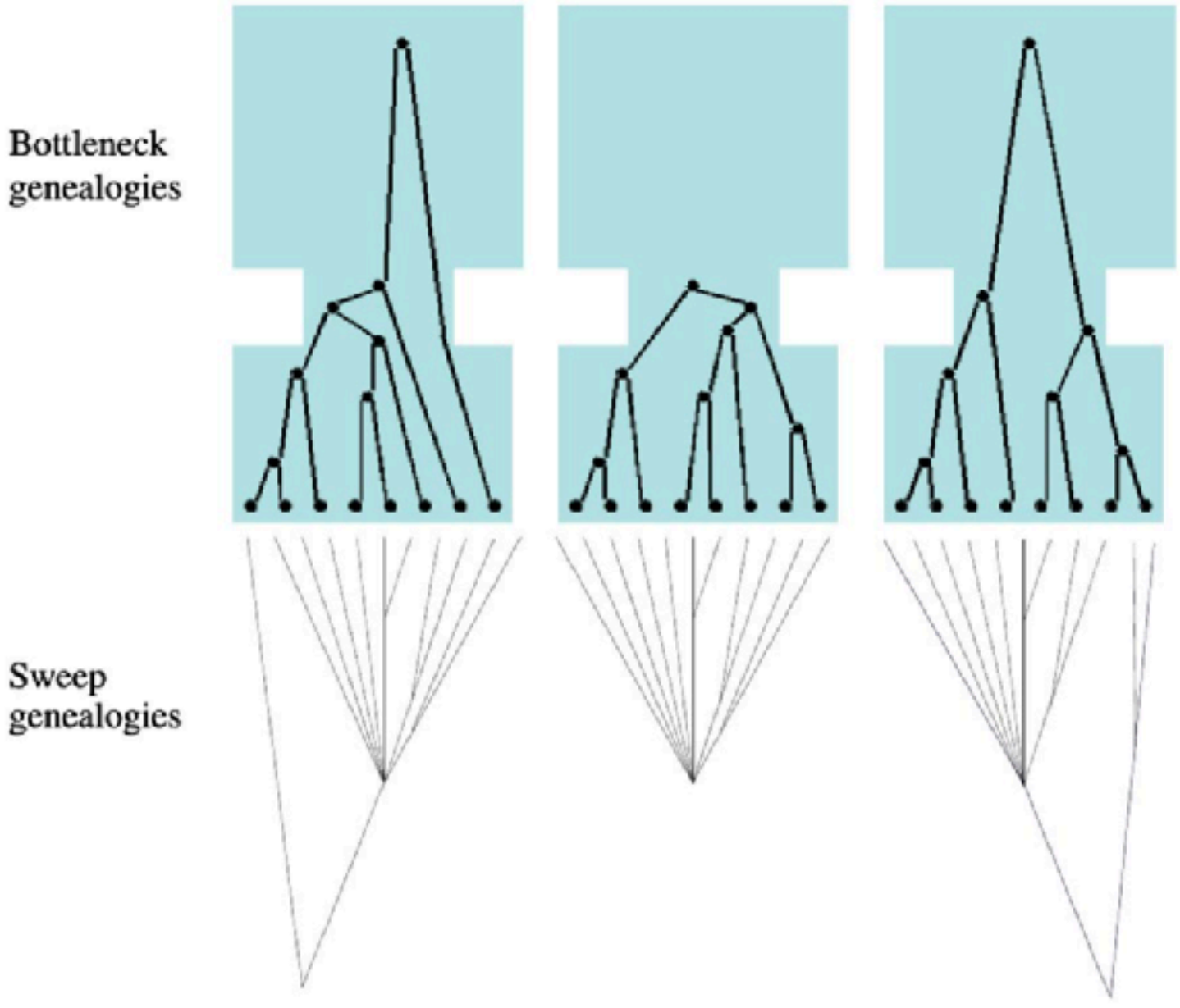
Nielsen 2005

TRACES OF BACKGROUND SELECTION



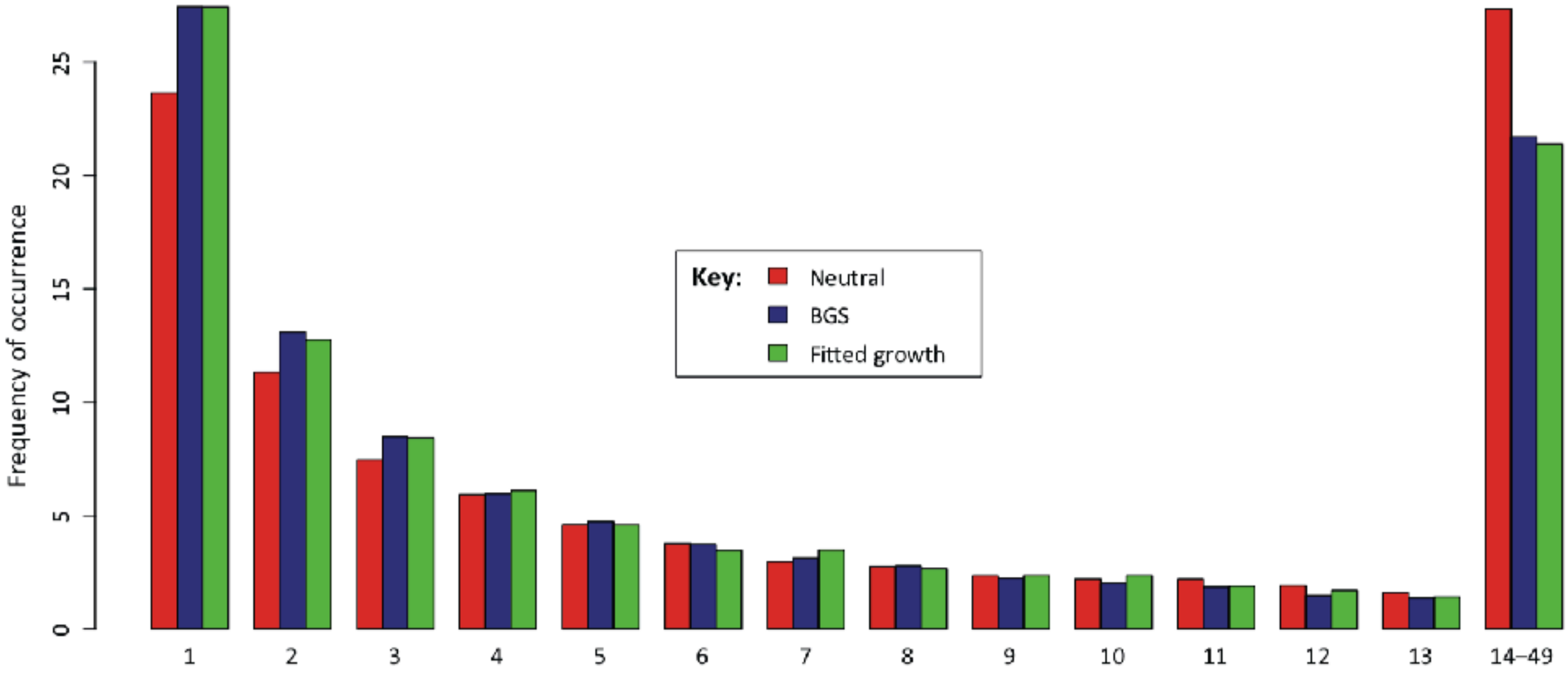
Charlesworth 2013

BUT IT IS DIFFICULT TO DISTINGUISH SELECTION FROM DEMOGRAPHY



Pavlidis and Alachiotis 2017

BUT IT IS DIFFICULT TO DISTINGUISH SELECTION FROM DEMOGRAPHY



Bank et al. 2014

SELECTION VS DEMOGRAPHY MEMORY

Bottleneck

Frequency-dependent selection

Selective sweep

Population growth

Population structure

Background selection

Selective sweep

Hitchhiking

PROGRESS IN POPULATION-GENETIC SELECTION INFERENCE

- genome-wide data and additional information
- Haplotype data and statistics
- Many (orthogonal) inference methods can be used in parallel (SFS-based, haplotype based, comparative)
- Two-step approach: infer demography from putatively neutral regions, then use the inferred demographic model for selection scan (e.g., Pavlidis et al. 2013) - joint inference in the future?xs
- Use simulations to validate results
- Use info from experimental evolution
- Obtain time-serial data for increased statistical power

Bank et al. 2014

WHAT WE WANT TO KNOW ABOUT SELECTION

- How big/small are adaptive steps?
- What are the proportions of beneficial, neutral, and deleterious mutations?
- How do mutational effects change dependent on the genetic background? (I.e., what is the role of epistasis?)
- What is the role of selection vs. other evolutionary processes in shaping genomes?
- How can we infer the contribution of selection to molecular evolution?

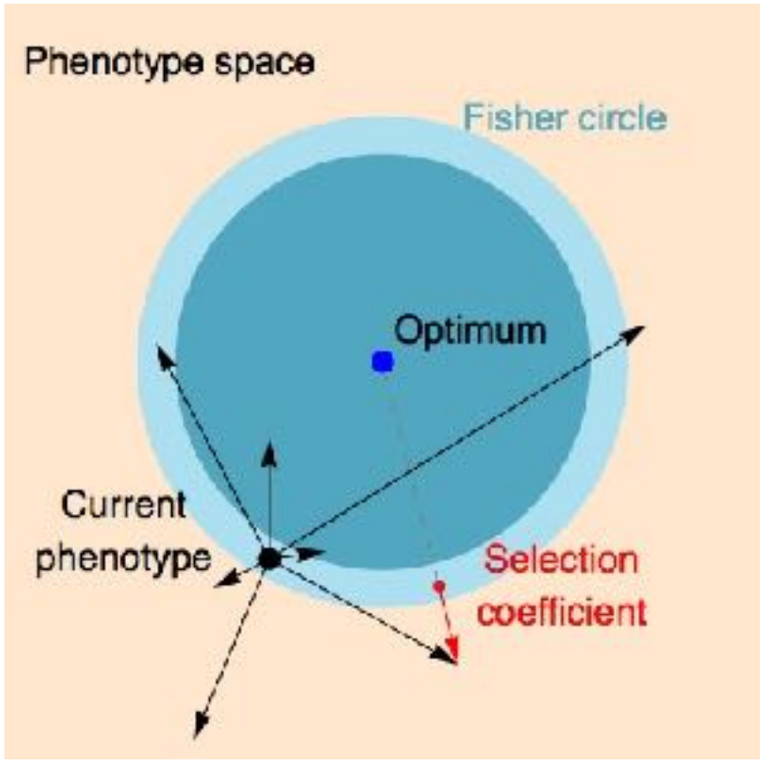
What do we expect adaptation to be like THEORETICALLY?

ADAPTATION VS ADAPTATIONS

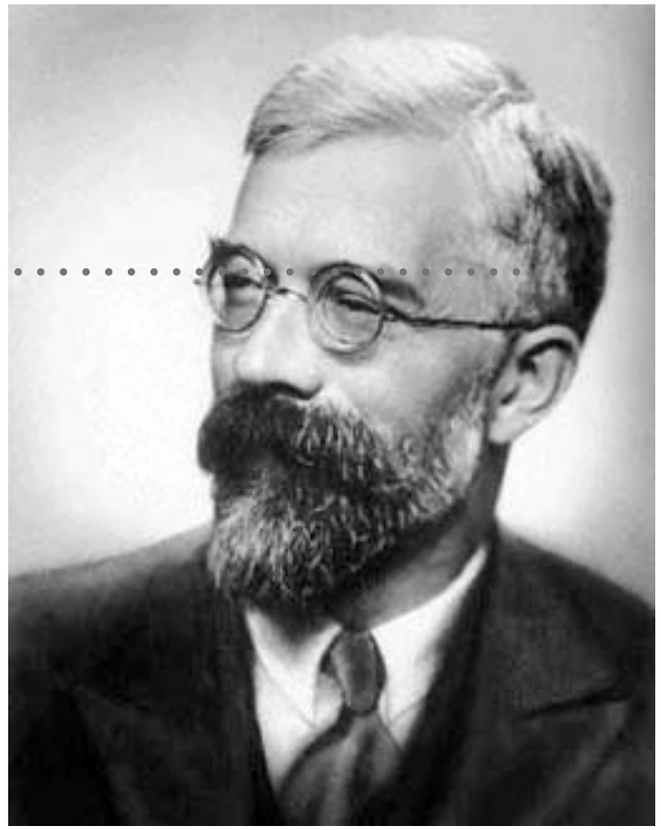
- Adaptation: the process of increasing (mean) fitness of a population in an environment
- An adaptation: a trait that increases its carrier's fitness in a specific environment, and that has spread bc of of the direct action of natural selection for its function

TWO MODELS OF ADAPTATION

FISHER'S GEOMETRIC MODEL



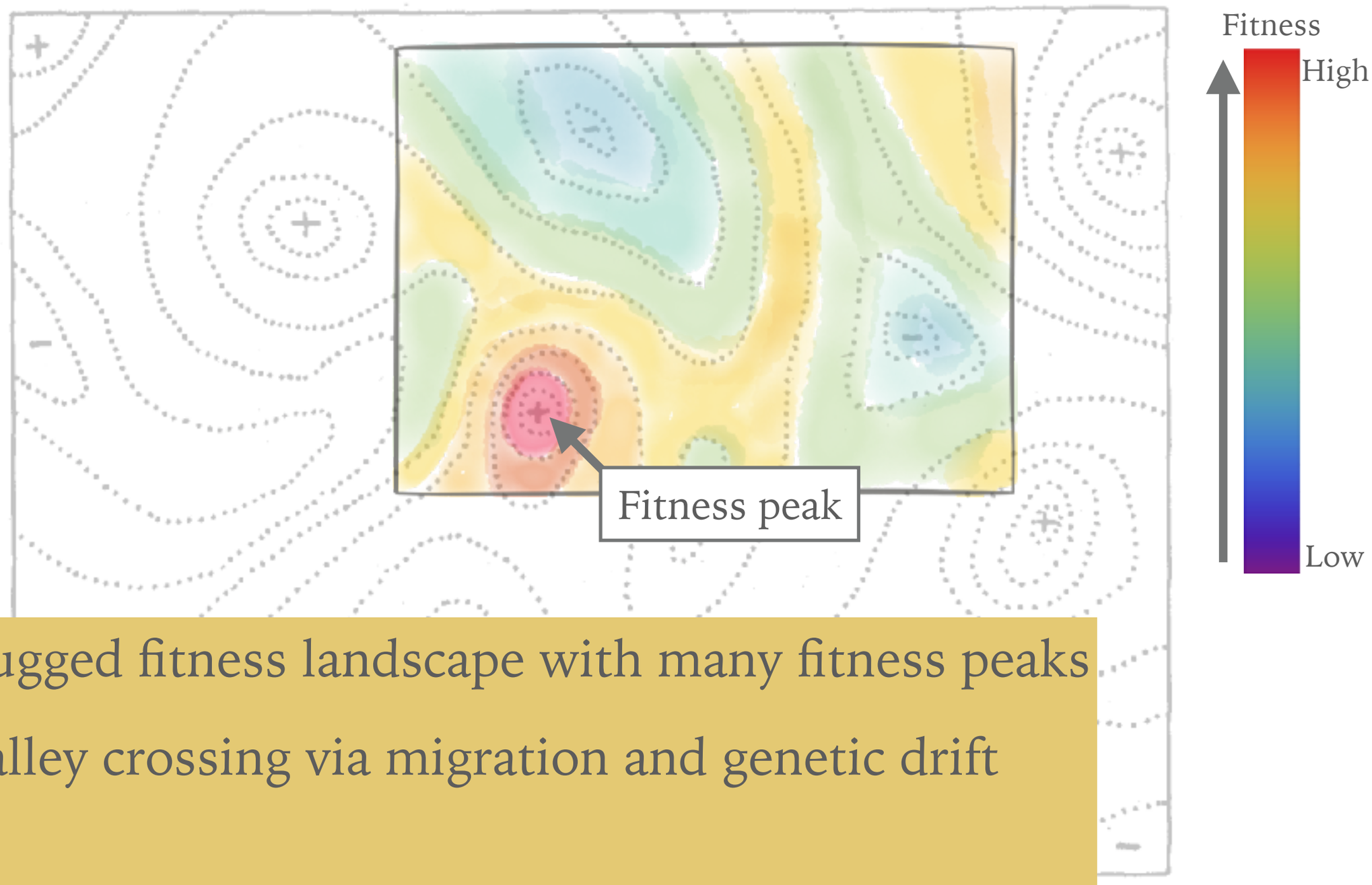
Fisher, 1930



- More challenging environment
  => more beneficial mutations

- Philosophy: Large populations, a single fitness optimum

WRIGHT'S SHIFTING BALANCE



- Rugged fitness landscape with many fitness peaks
- Valley crossing via migration and genetic drift
- Philosophy: Small, structured populations

Wright, 1932

Which team are you on, Team Fisher or Team Wright, and why?

WHAT WE WANT TO KNOW ABOUT SELECTION

- How big/small are adaptive steps?
- What are the proportions of beneficial, neutral, and deleterious mutations?
- How do mutational effects change dependent on the environment?
- What is the shape of the distribution of fitness effects (DFE)?

ESTIMATES OF MEAN BENEFICIAL EFFECT SIZE FROM POLYMORPHISM DATA

- s=0.002 (Li and Stephan 2006; Jensen et al. 2008)
- s=0.01 (MacPherson et al. 2008)
- s=0.00001 (Andolfatto 2007)

- For known phenotype: s=0.102 (Linnen et al. 2009)

AN EXPERIMENTAL APPROACH TO THE DFE: DEEP MUTATIONAL SCANNING

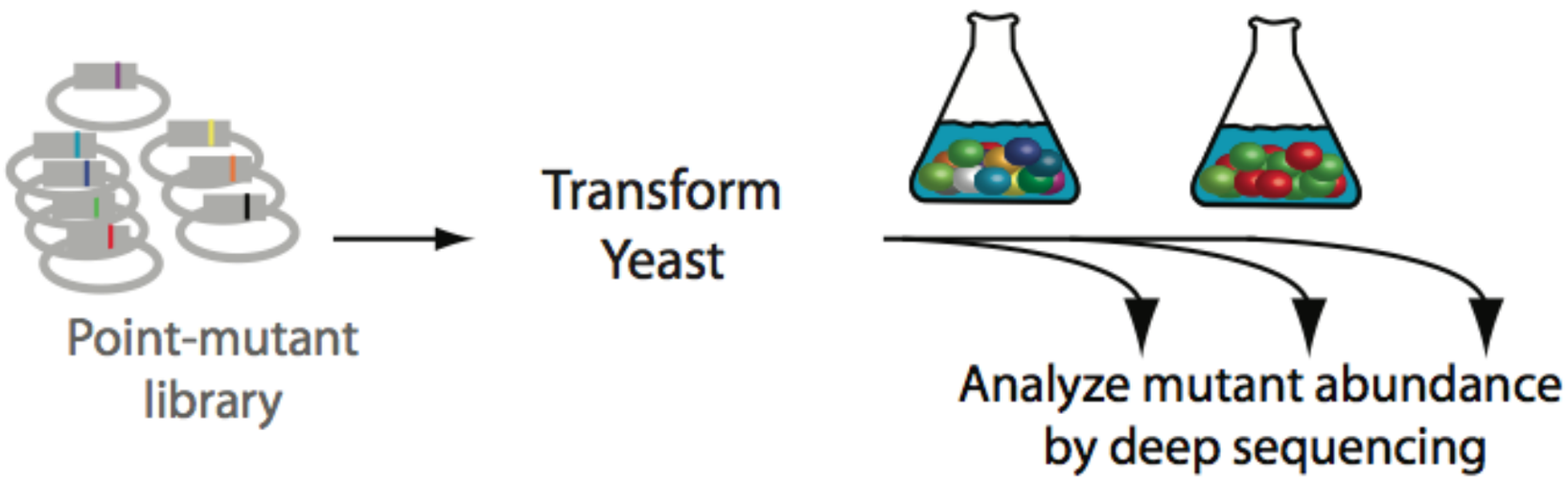
- Systematic high-throughout sampling of hundreds of chosen mutations (including those that are strongly deleterious)

Deep mutational scanning results in a (almost "evolution-free") snapshot of the DFE

- Genetic background is precisely controlled (minimized potential for secondary mutations)

Ryan Hietpas

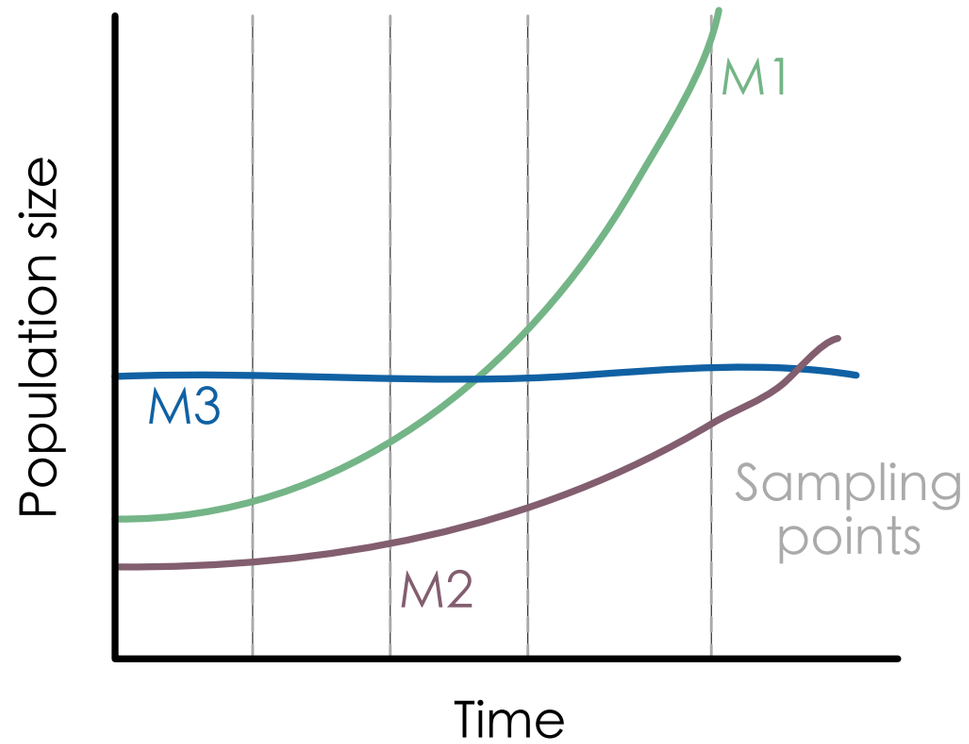
Jeff Jensen



Hietpas, Jensen & Bolon, PNAS, 2011

DEEP MUTATIONAL SCANNING FROM A MODELER'S POINT OF VIEW

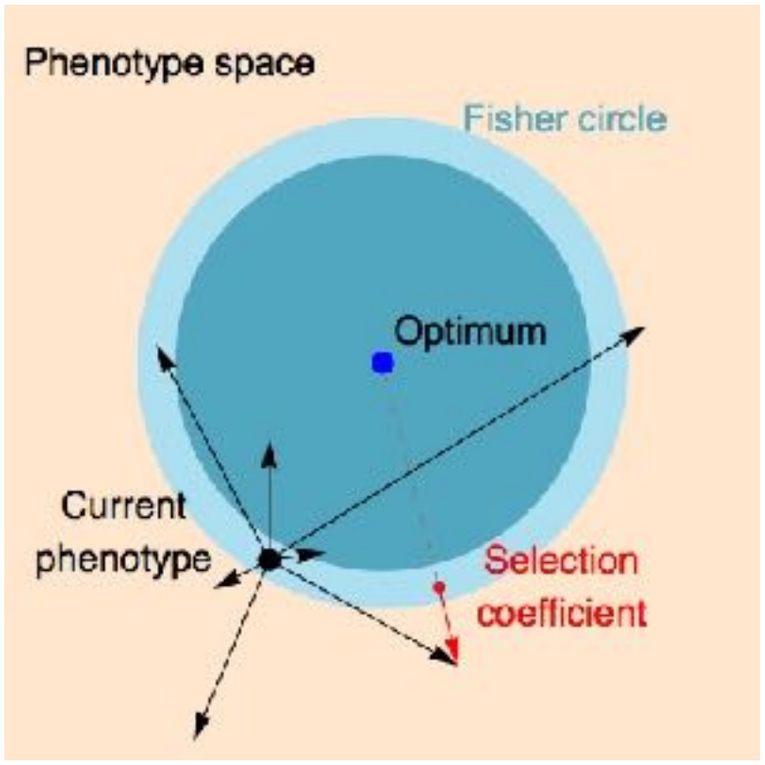
- Exponential growth of hundreds of mutants, each with its own growth rate/selection coefficient
- Sequencing corresponds to multinomial sampling of mutants independently at each sampling time



- <1% fitness differences detectable

For the “Fisherians”: the shape of the DFE across environments

FISHER'S GEOMETRIC MODEL



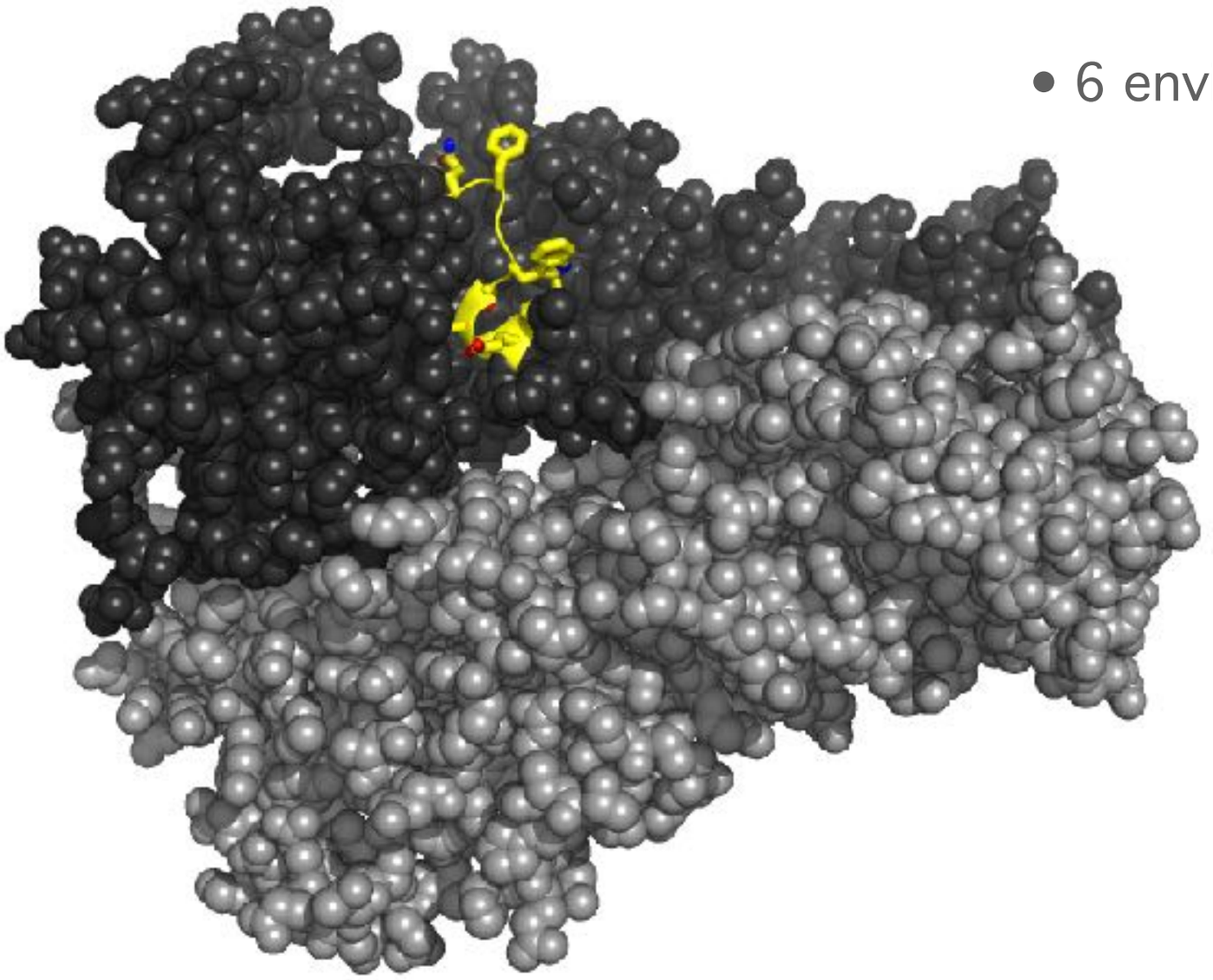
Fisher, 1930

Hypotheses:

- Relocation of the optimum or the current phenotype in a new environment can increase the distance to the optimum and hence the potential for beneficials.
- The distribution of beneficial mutations is bounded or exponential.

THE SHAPE OF THE DFE IN CHALLENGING ENVIRONMENTS

The data set



- 9 aa region from Hsp90 (aa positions 582-590) in Saccharomyces cerevisiae

- 6 environments:

30°C	30°C+0.5M NaCl
36°C	36°C+0.5M NaCl
25°C	25°C+0.5M NaCl

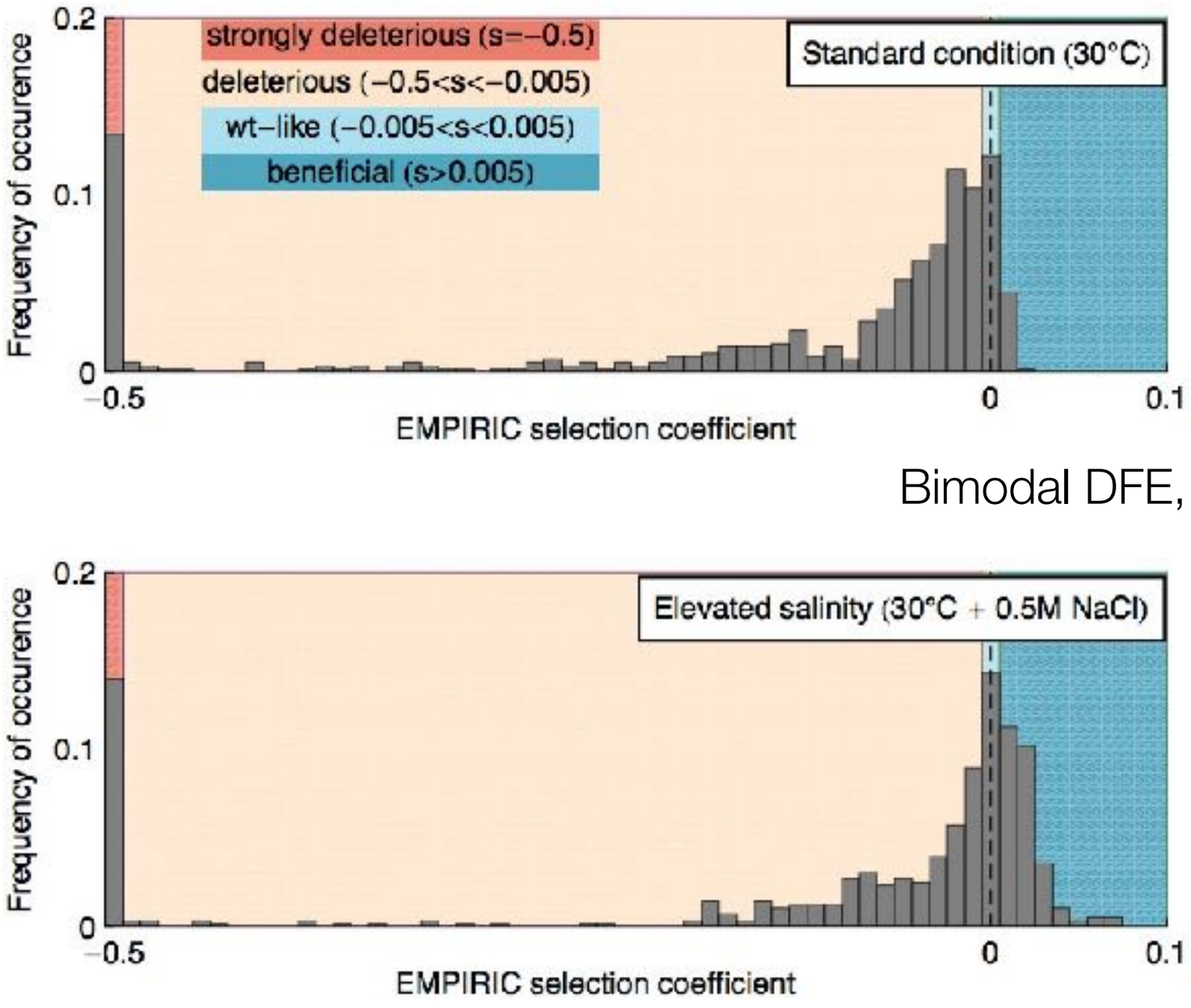
- Fitness data for every possible codon at each aa position (i.e. the same 560 mutations per environment)

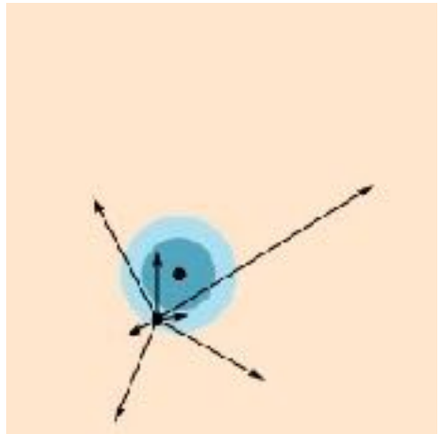
Relative growth of wt:

1	0,45
0,83	0,33
0,63	0,3

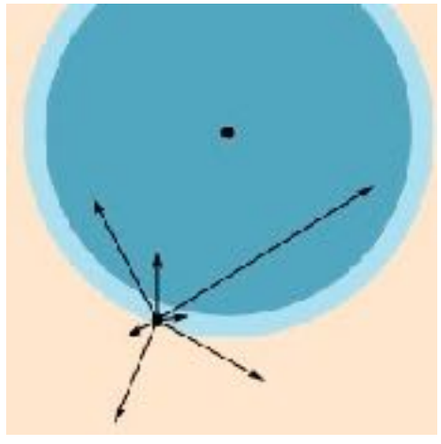
Data obtained by Ryan Hietpas @ UMassMed

The shape of the full DFE





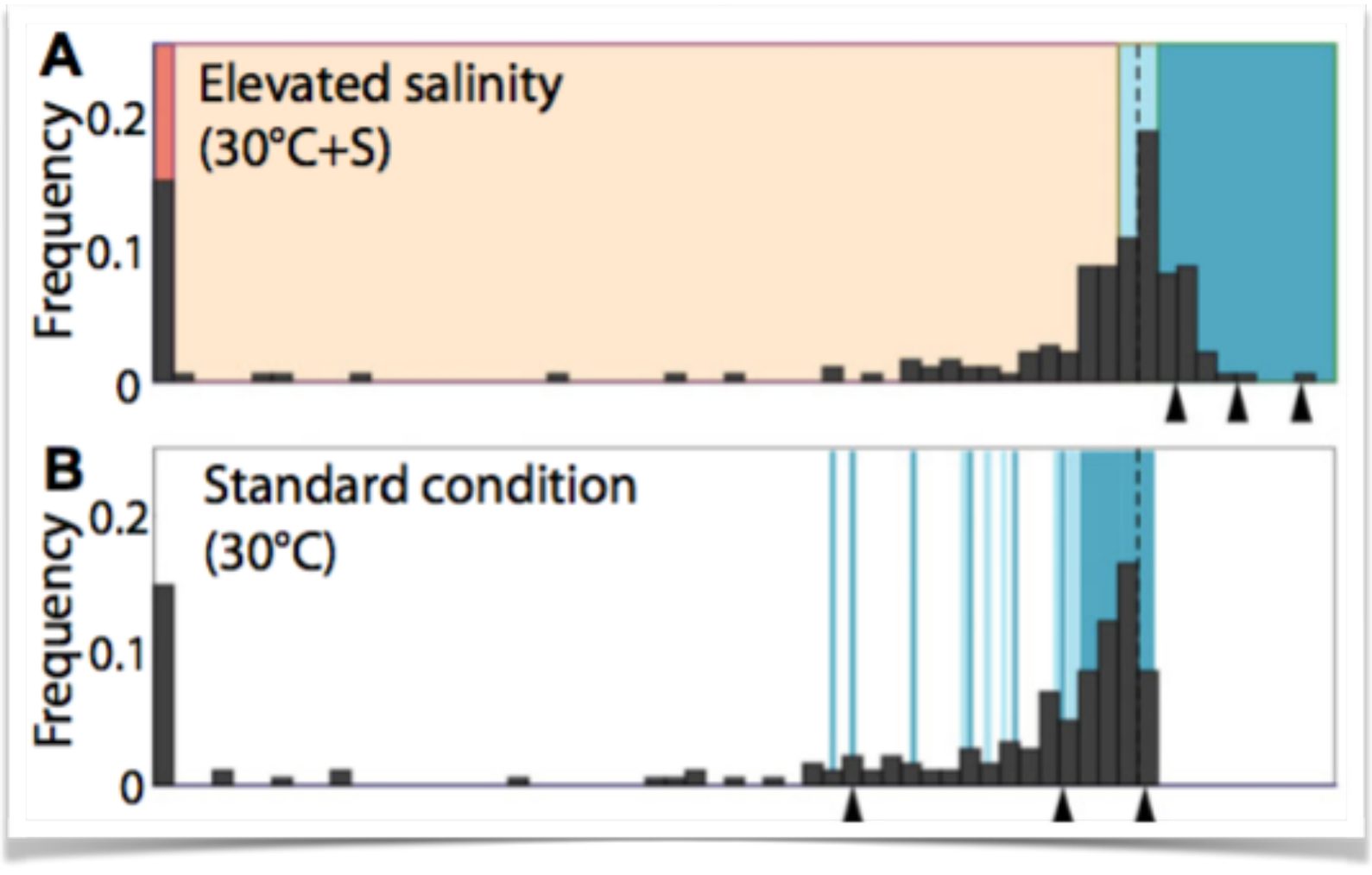
Bimodal DFE, few beneficials - close to optimum



Increased number of beneficials, increased variance - far from optimum

Hietpas, Bank et al. 2013

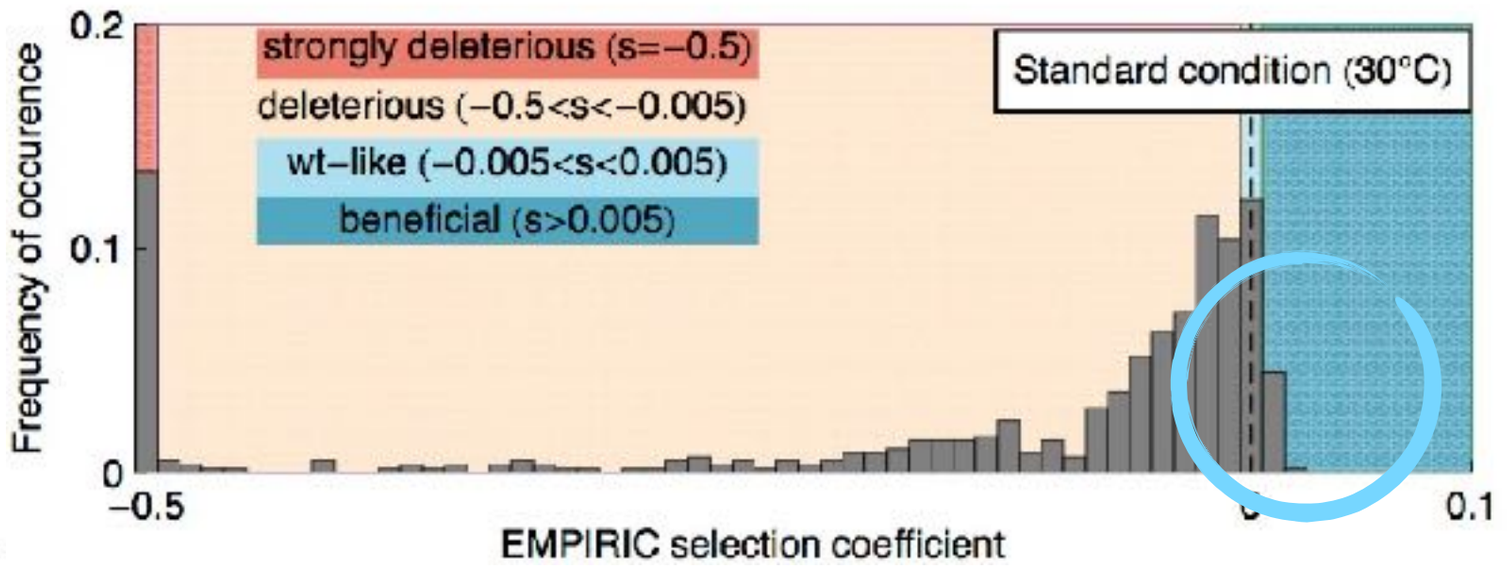
COSTS OF ADAPTATION

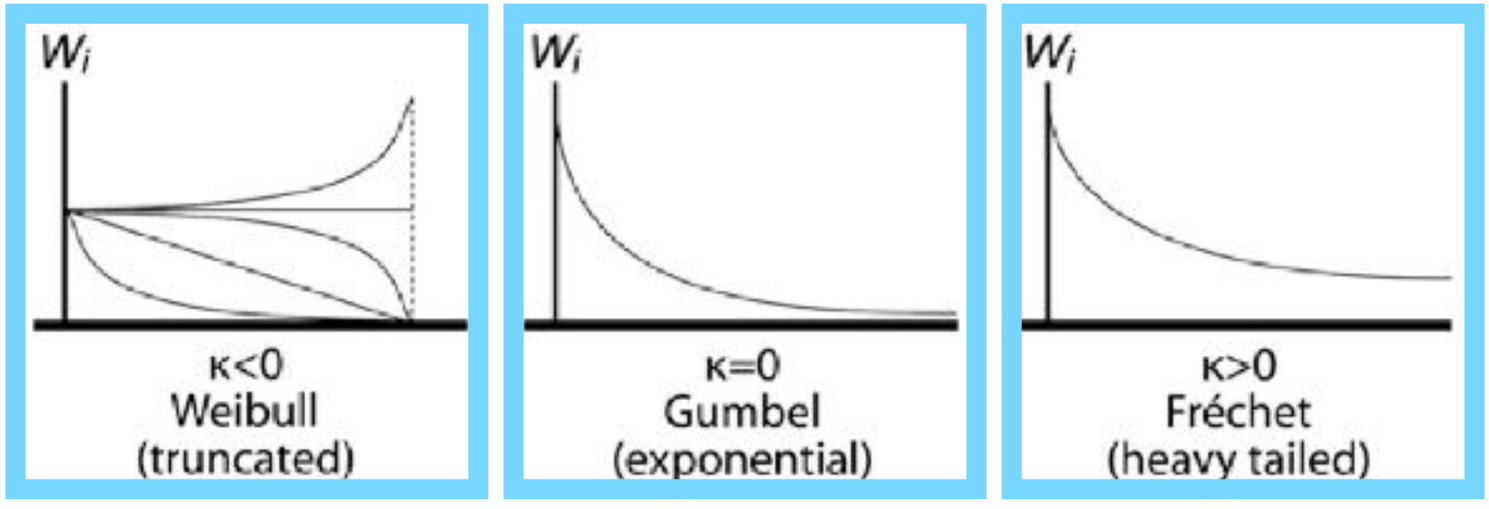


THE SHAPE OF THE BENEFICIAL TAIL OF THE DFE

HOW PREDICTABLE IS ADAPTIVE EVOLUTION?

- Fit Generalized Pareto distribution to beneficial tail
- Kappa parameter determines tail shape



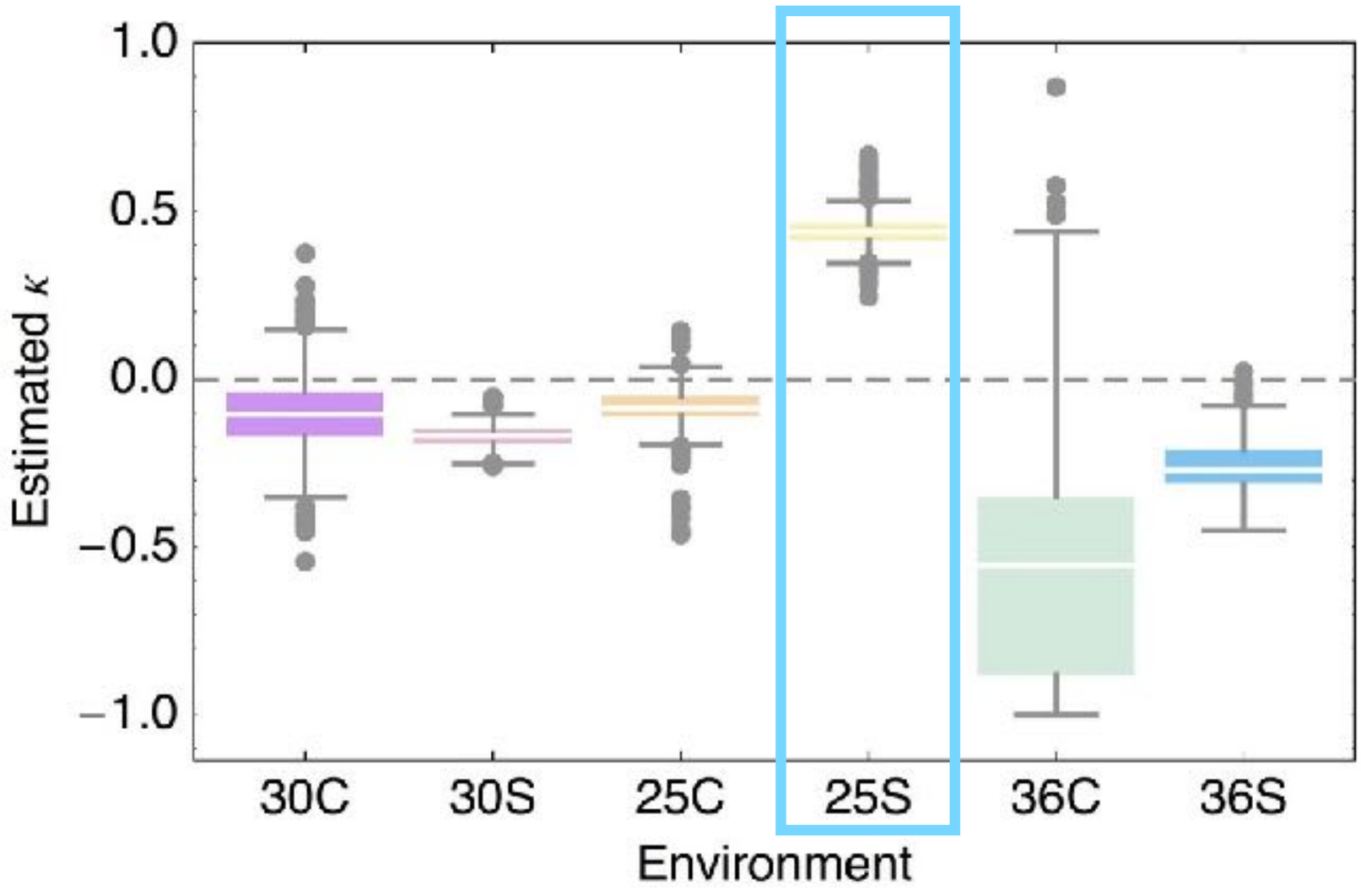


- Unbounded DFE, highly unpredictable mutational effects
- Not captured by FGM

From Beisel et al., Genetics, 2007

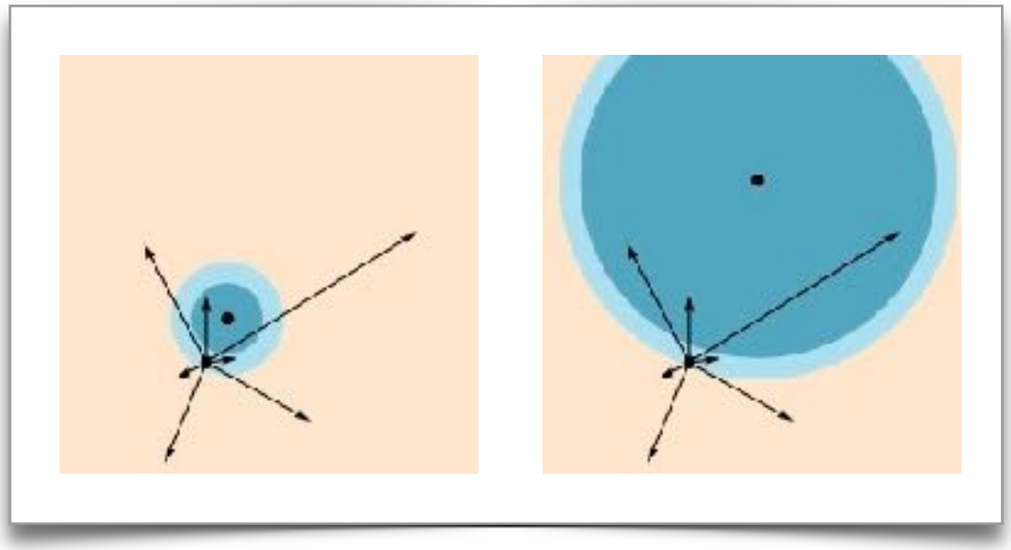
TAIL SHAPE PARAMETER IN CHALLENGING ENVIRONMENTS

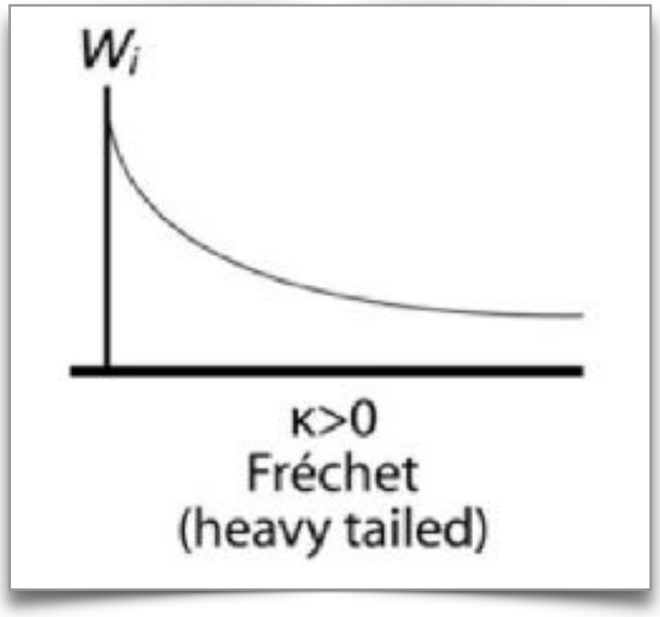
S. cerevisiae EMPIRIC data from Hsp90



SUMMARY – TEAM FISHER

- In response to a novel environmental challenge the number and size of beneficial mutations increases, and costs of adaptation are observed - in agreement with predictions from Fisher's geometric model when the optimum is displaced.
- Following severe environmental challenges, the step size of adaptive mutations might be highly unpredictable.

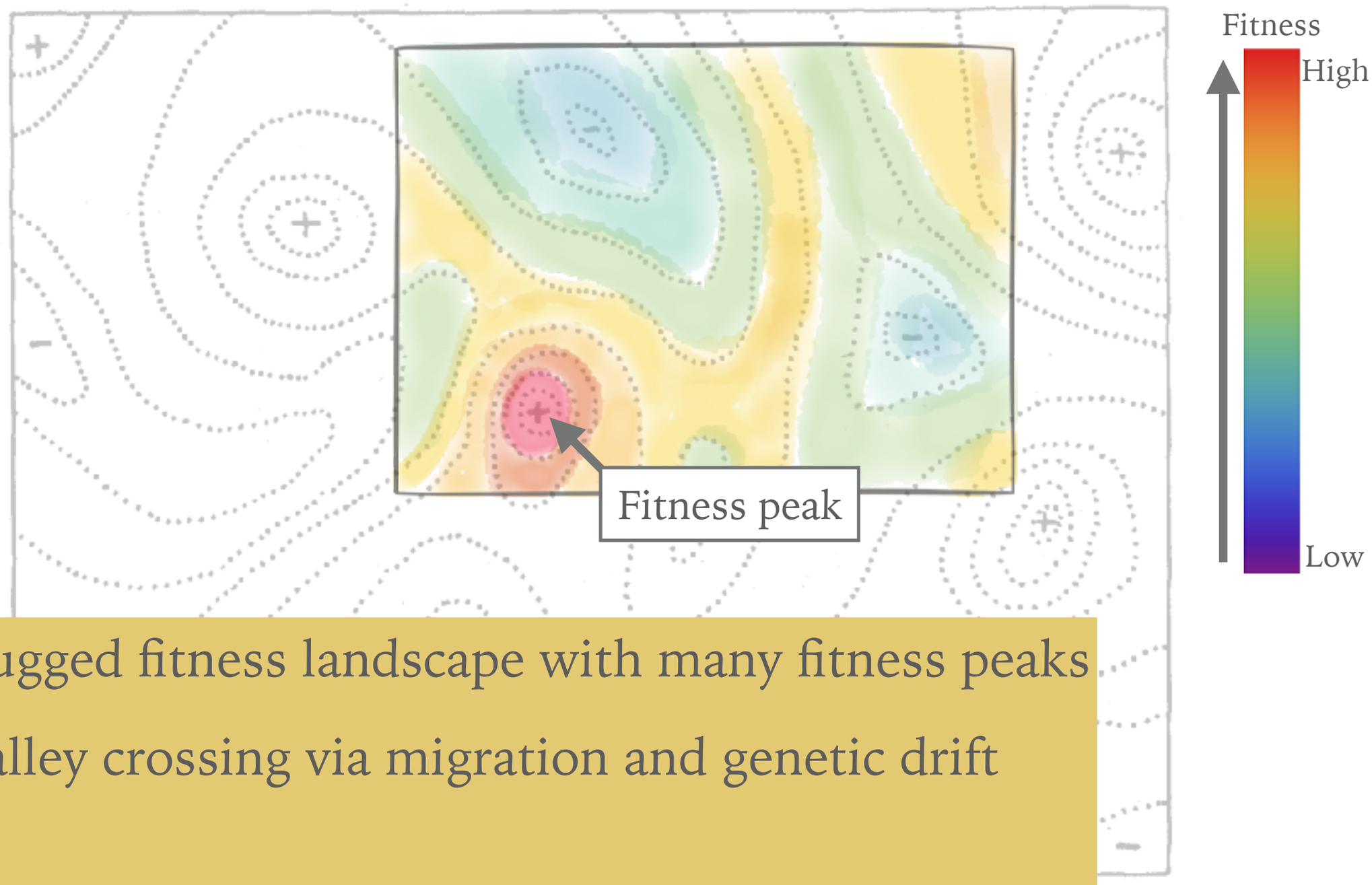




But what about epistasis?

(Spoiler: this is the part for the “Wrightians”)

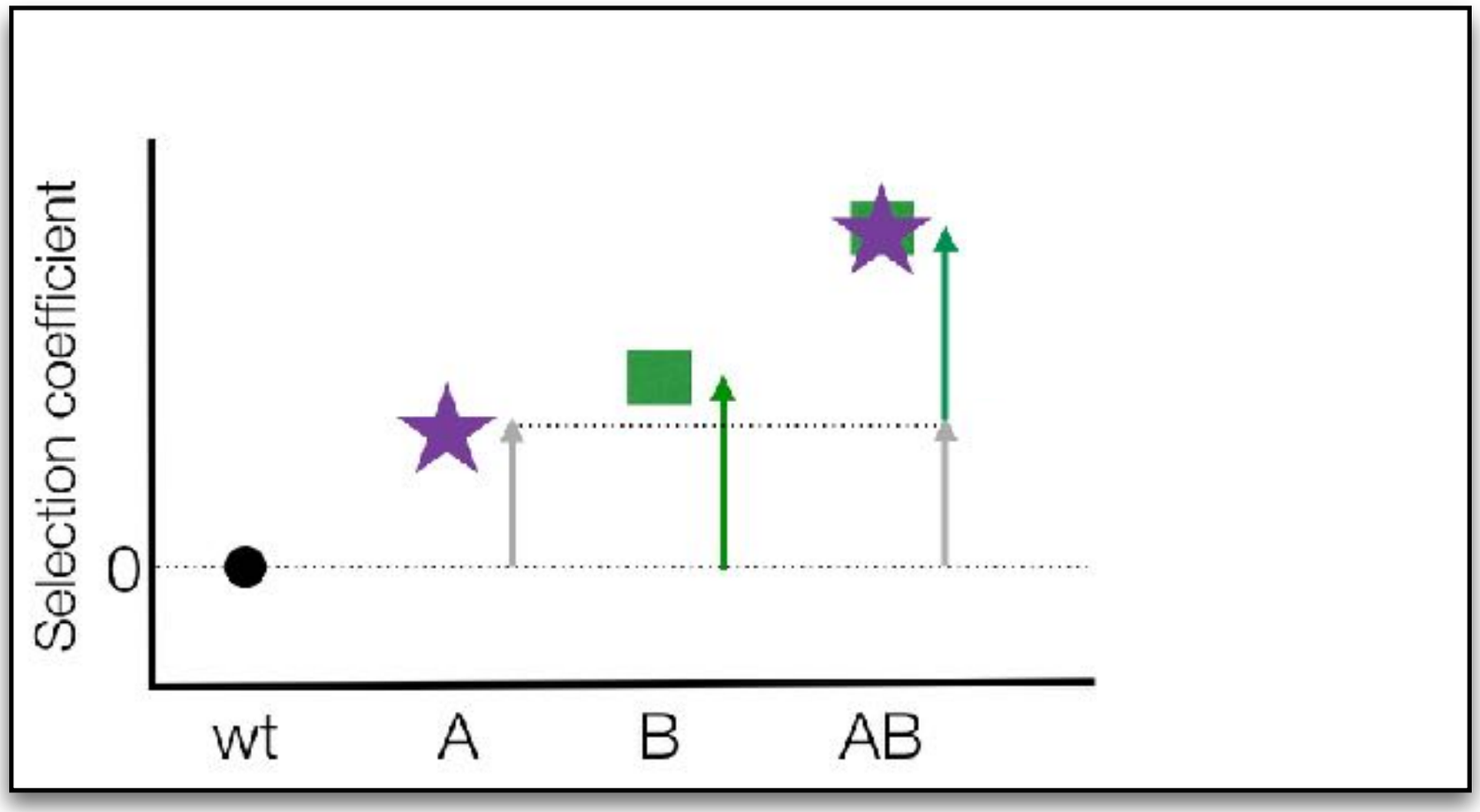
WRIGHT'S SHIFTING BALANCE



- Rugged fitness landscape with many fitness peaks
- Valley crossing via migration and genetic drift
- Philosophy: Small, structured populations

Wright, 1932

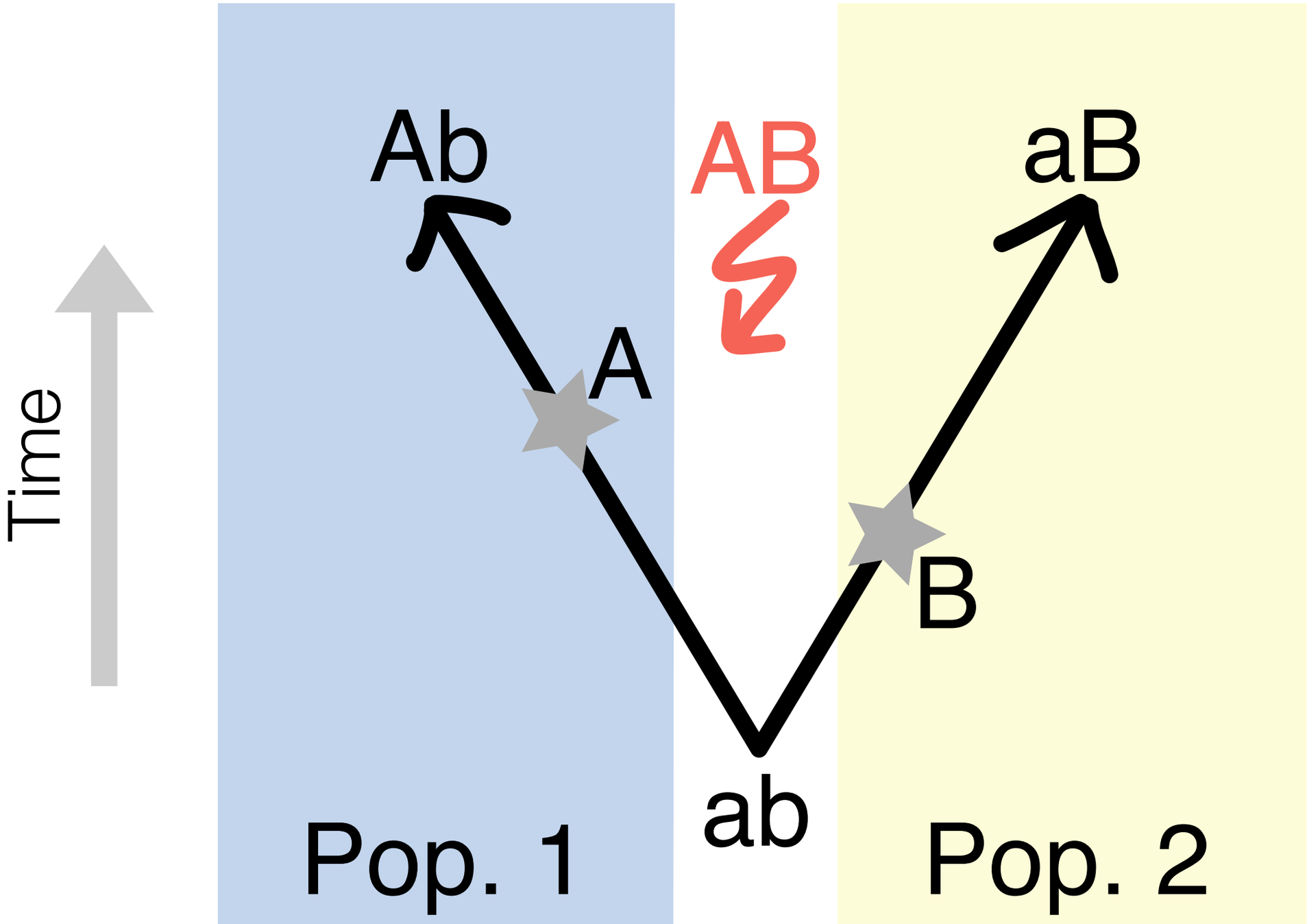
WHAT IS EPISTASIS?



WHY SHOULD WE CARE?

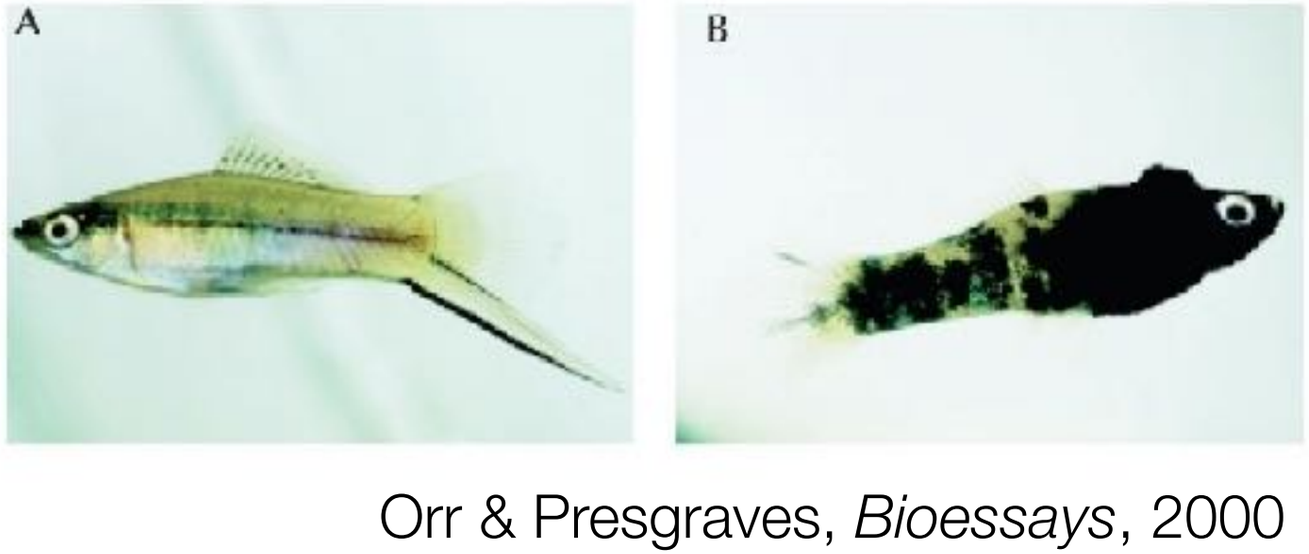
- epistasis creates non-random associations between loci (LD)
- Ruggedness of fitness landscape is a determinant of predictability/repeatability of evolution
- accumulation of epistatic alleles is basis of the most widely accepted model for allopatric speciation

THE DOBZHANSKY-MULLER MODEL

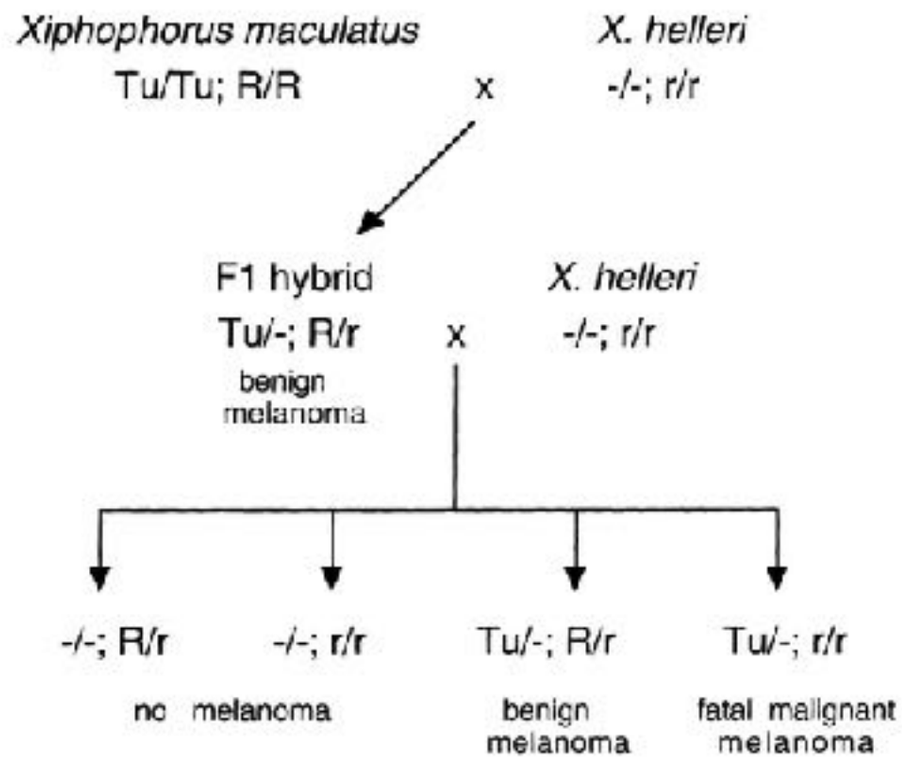


WHAT IS THE EVIDENCE?

DMIS



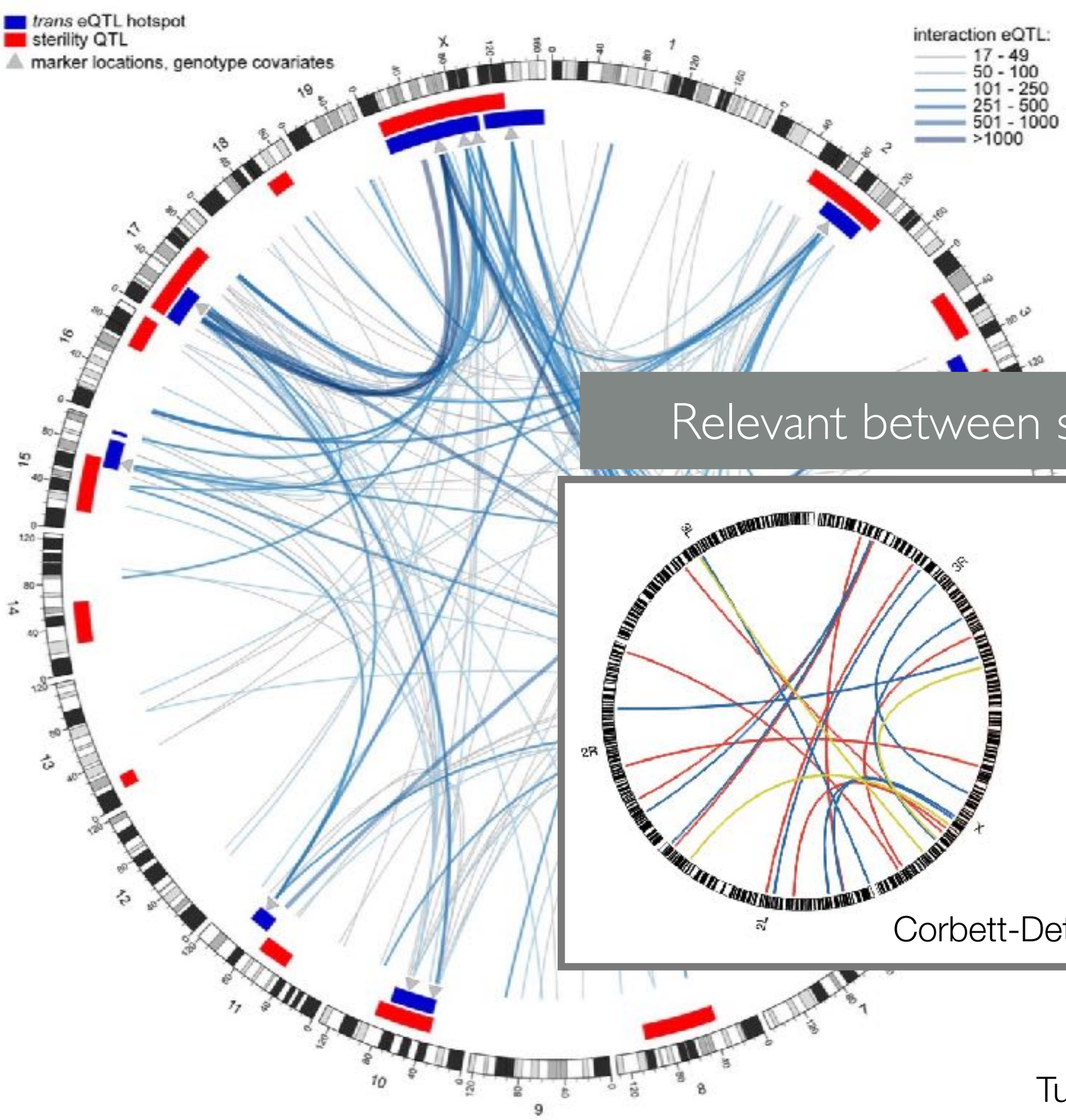
Orr & Presgraves, *Bioessays*, 2000



- sexual selection on tumor gene
- interaction with promoter of repressor gene
- ongoing gene flow

Scarpino et al., MBE, 2013

DMIS



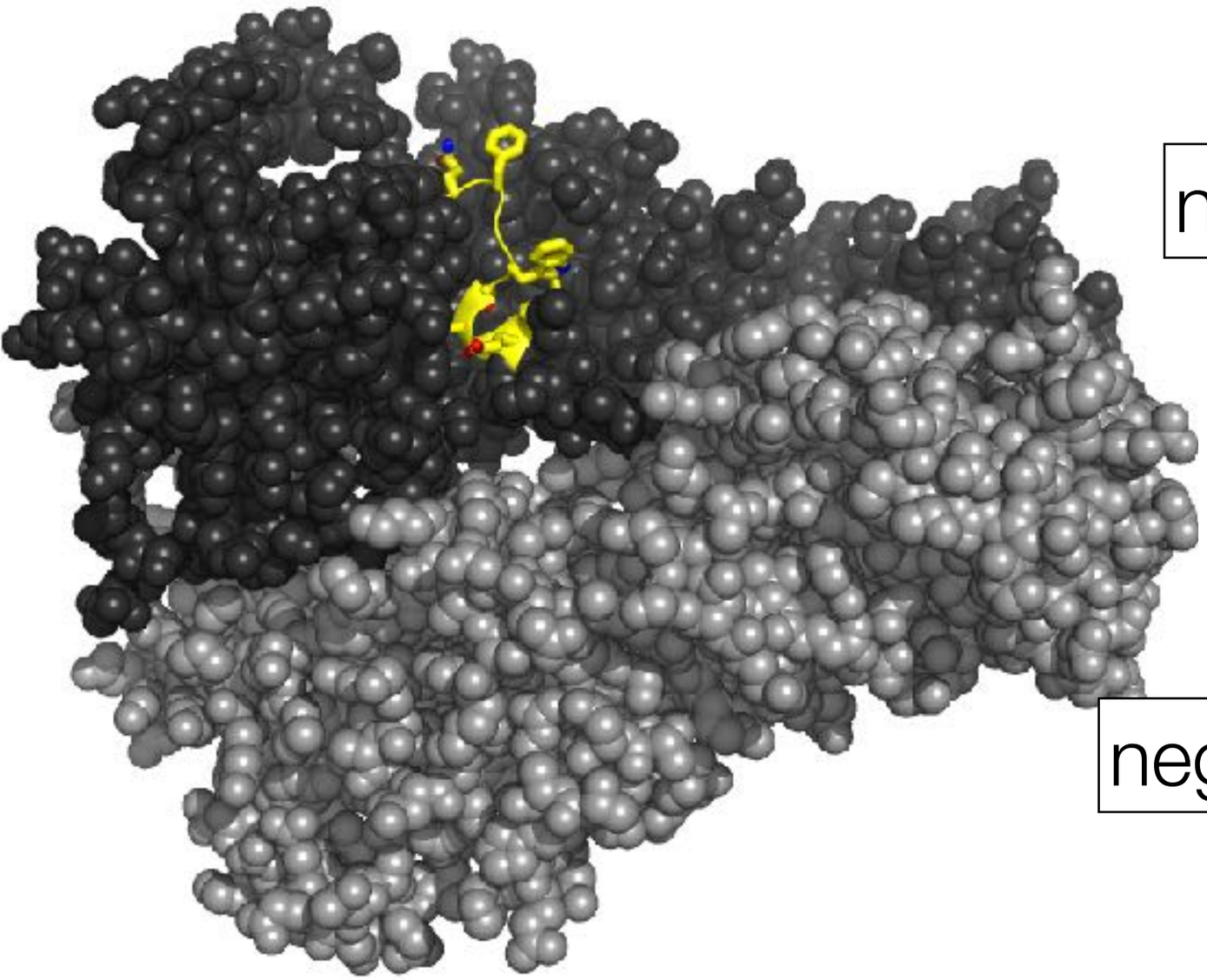
Relevant between species only?

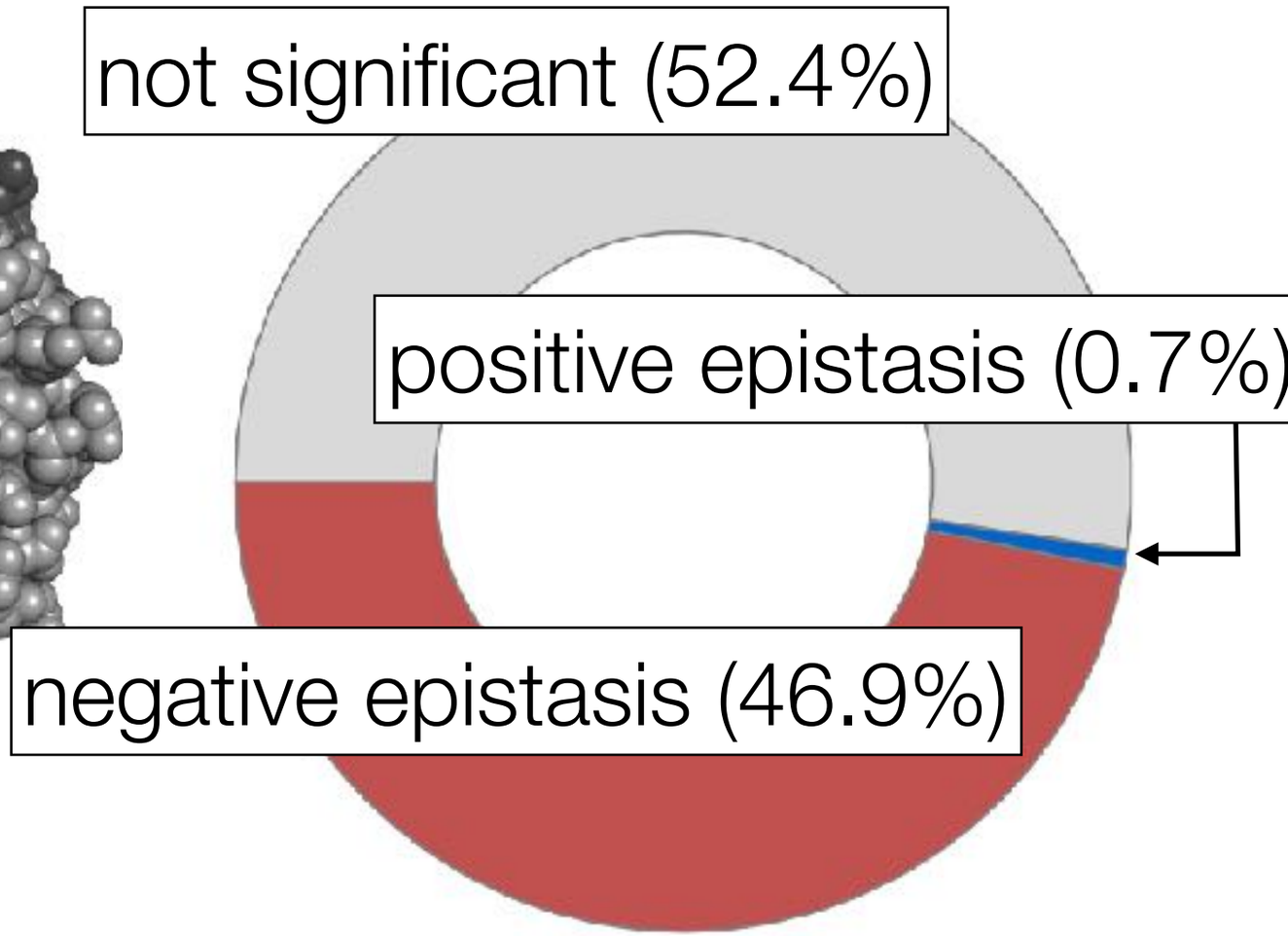
Maybe not!

Corbett-Detig et al., *Nature*, 2013

Turner et al., *PlosGen*, 2014

PAIRWISE EPISTASIS WITHIN A PROTEIN



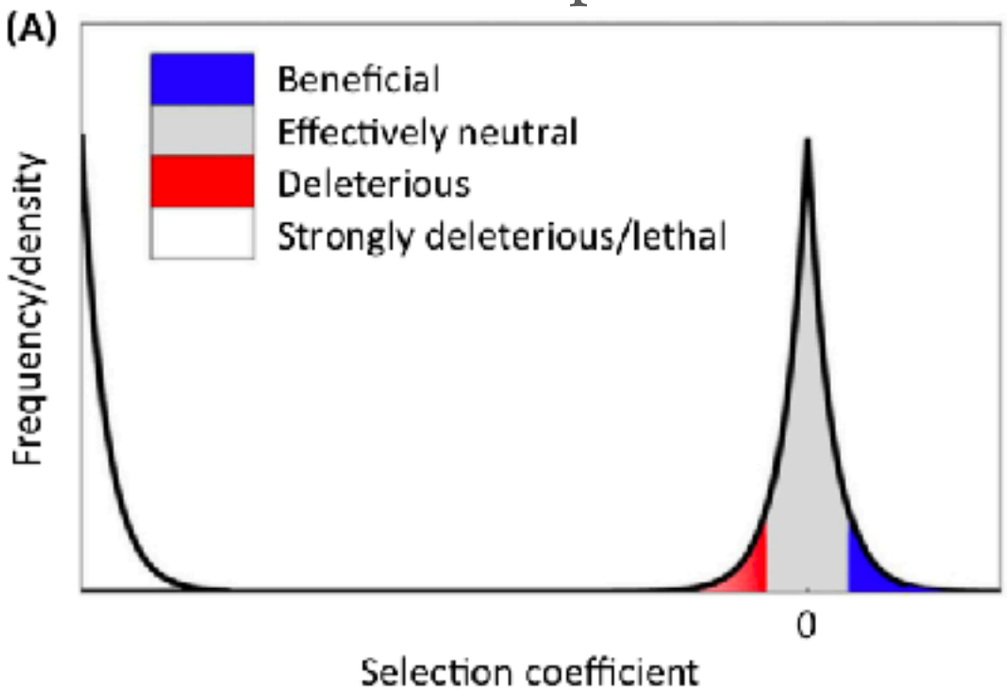


Bank et al. 2015

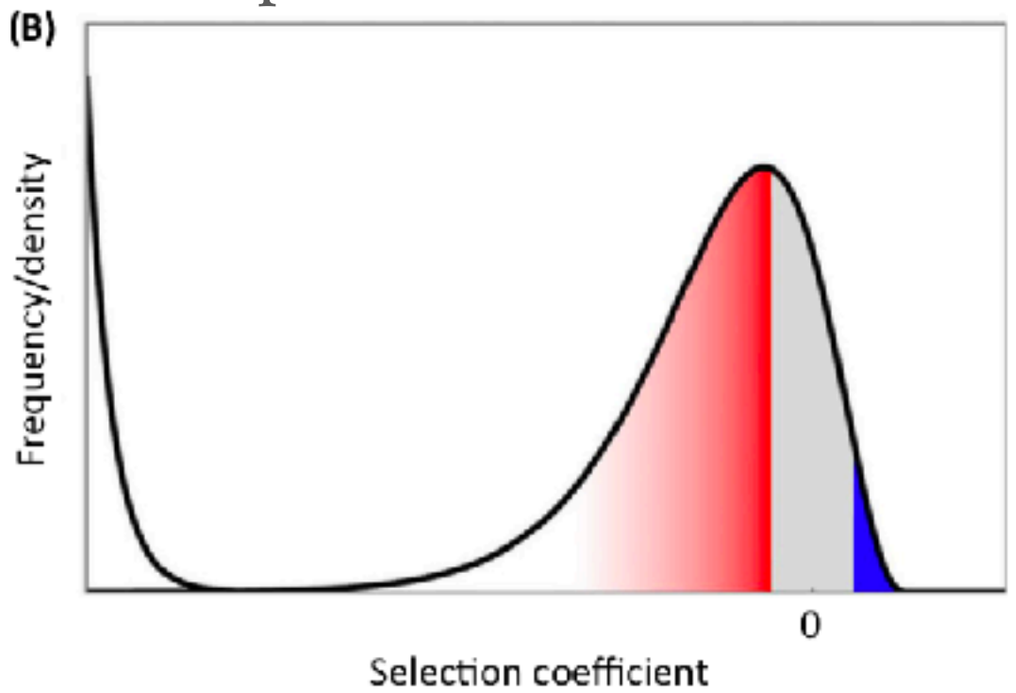
SYNOPSIS OF PART 1

- Selection is both simple and difficult, but certainly important
- DFE looks different than what has been assumed in most studies
- Epistasis seems to be common

"Old" assumption

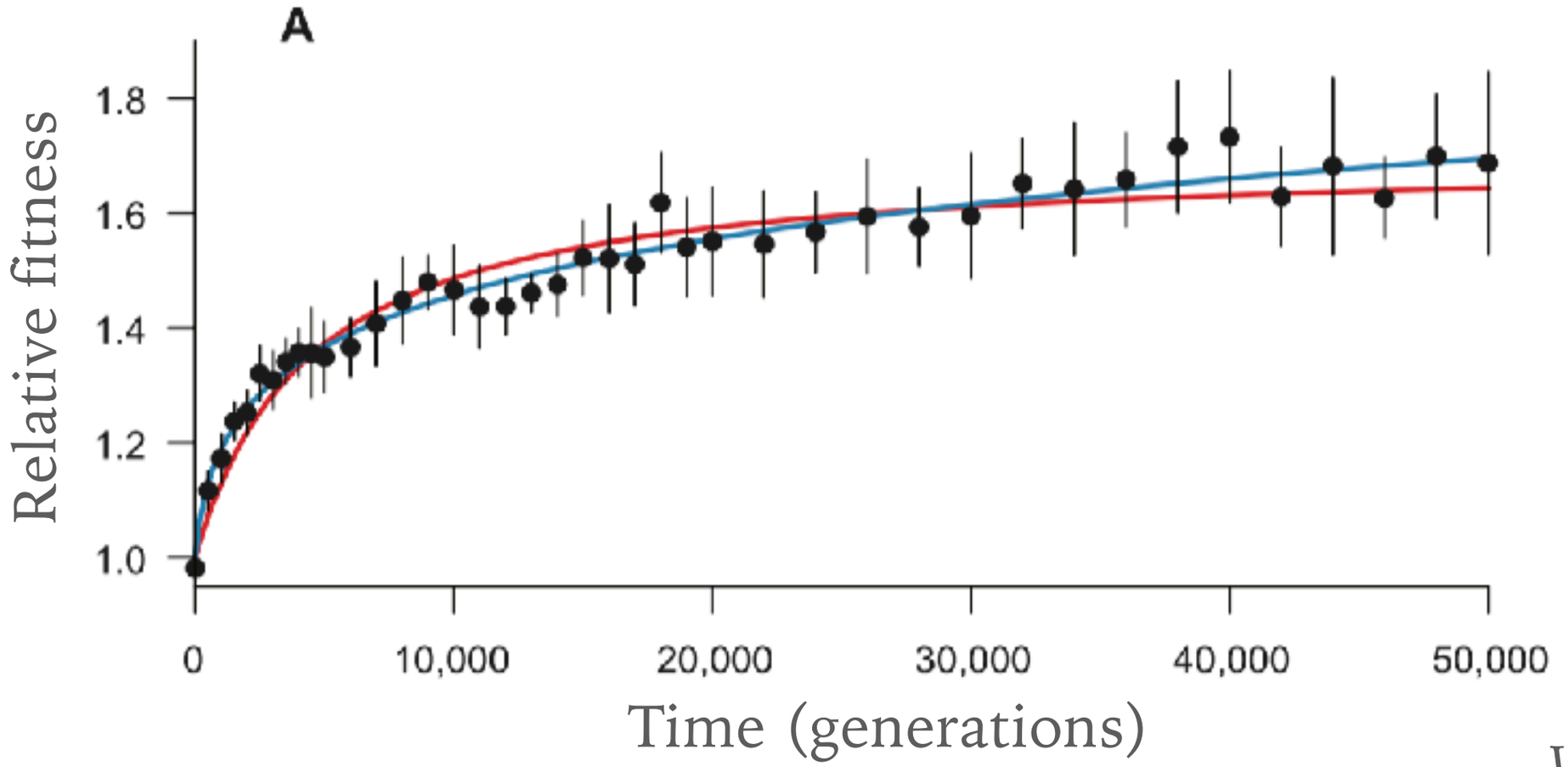


Experimental evidence



FISHER OR WRIGHT?

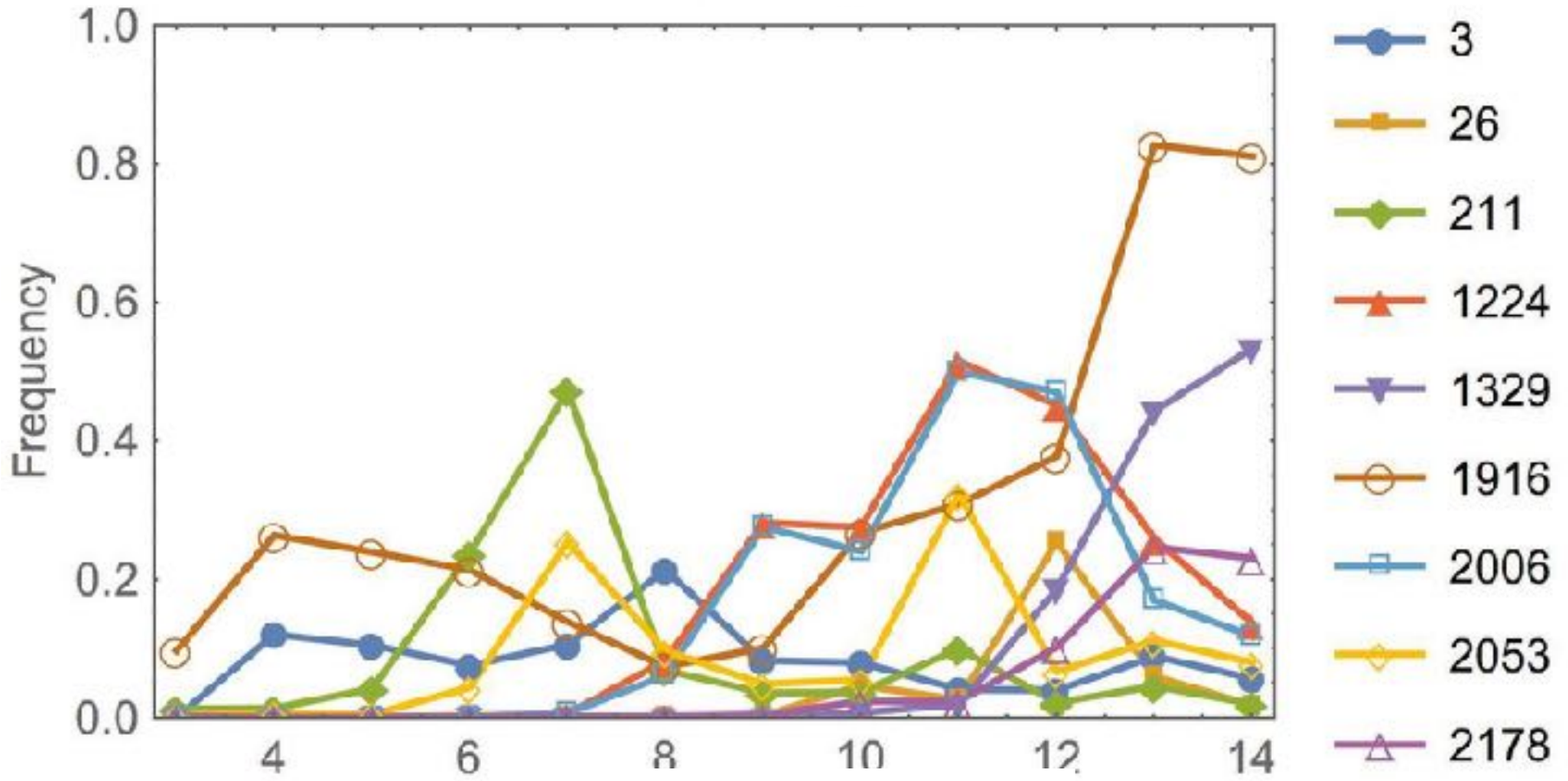
- From an ecological point of view, frequent bottlenecks seem likely.
- But adaptation is also miraculous in constant environments with high population sizes - how is that possible?



Wiser et al. 2013

SELECTION INFERENCE FROM TIME-SERIAL SNP DATA

CAN WE ESTIMATE SELECTION COEFFICIENTS FROM TRAJECTORIES?

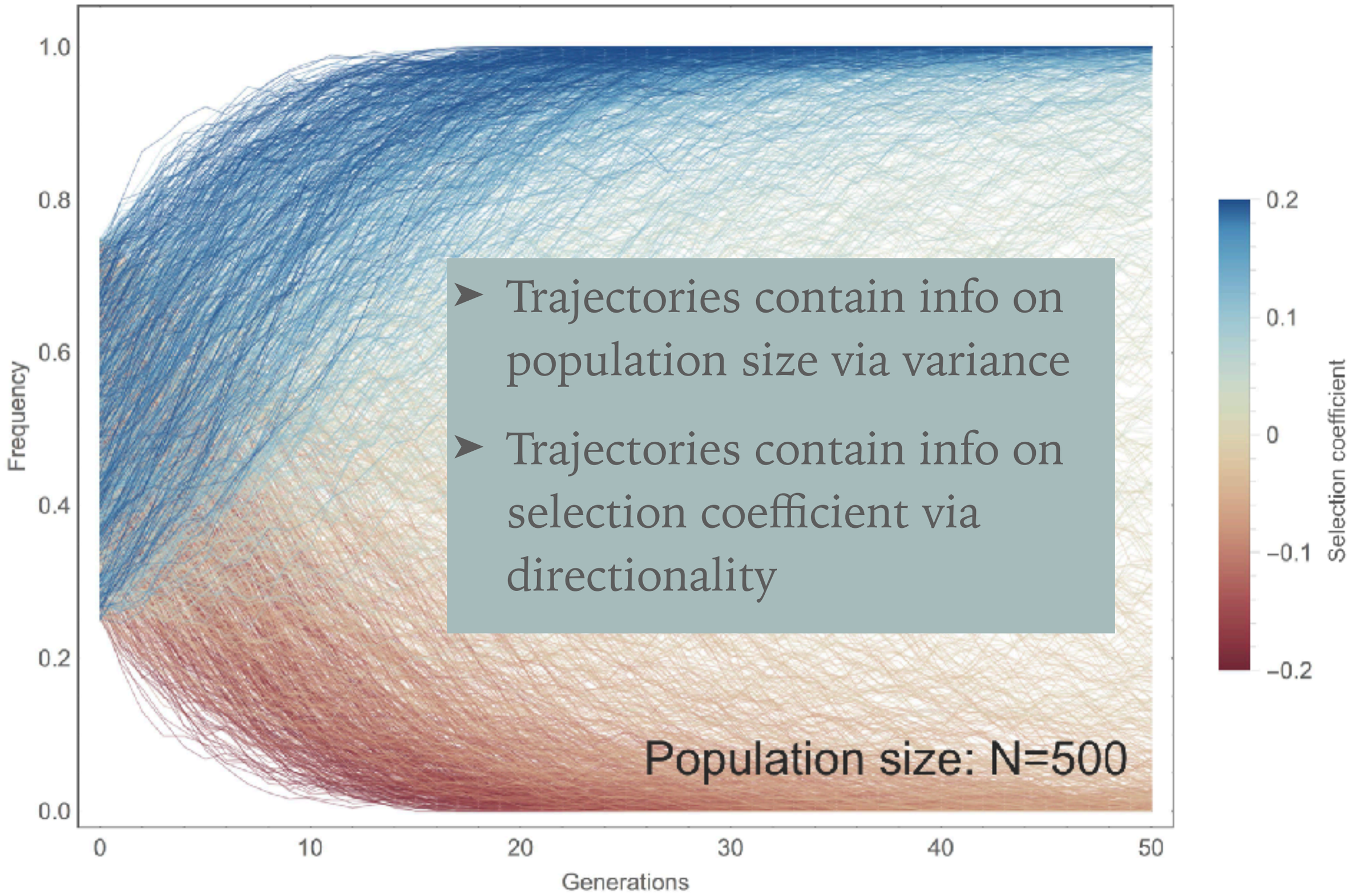


Output: allele-frequency trajectories of mutants along the genome

WFABC*

**Foll, Poh et al. 2014, Foll, Shim et al. 2014*

THE WRIGHT–FISHER MODEL



- Trajectories contain info on population size via variance
- Trajectories contain info on selection coefficient via directionality

SUMMARY STATISTICS FOR WFABC

$$Fs = \frac{(x-y)^2}{z(1-z)} \text{ and } Fs' = \frac{1}{t}\frac{Fs[1-1/(2\tilde{n})]-2/\tilde{n}}{(1+Fs/4)\left[1-1/\left(n_y\right)\right]}$$

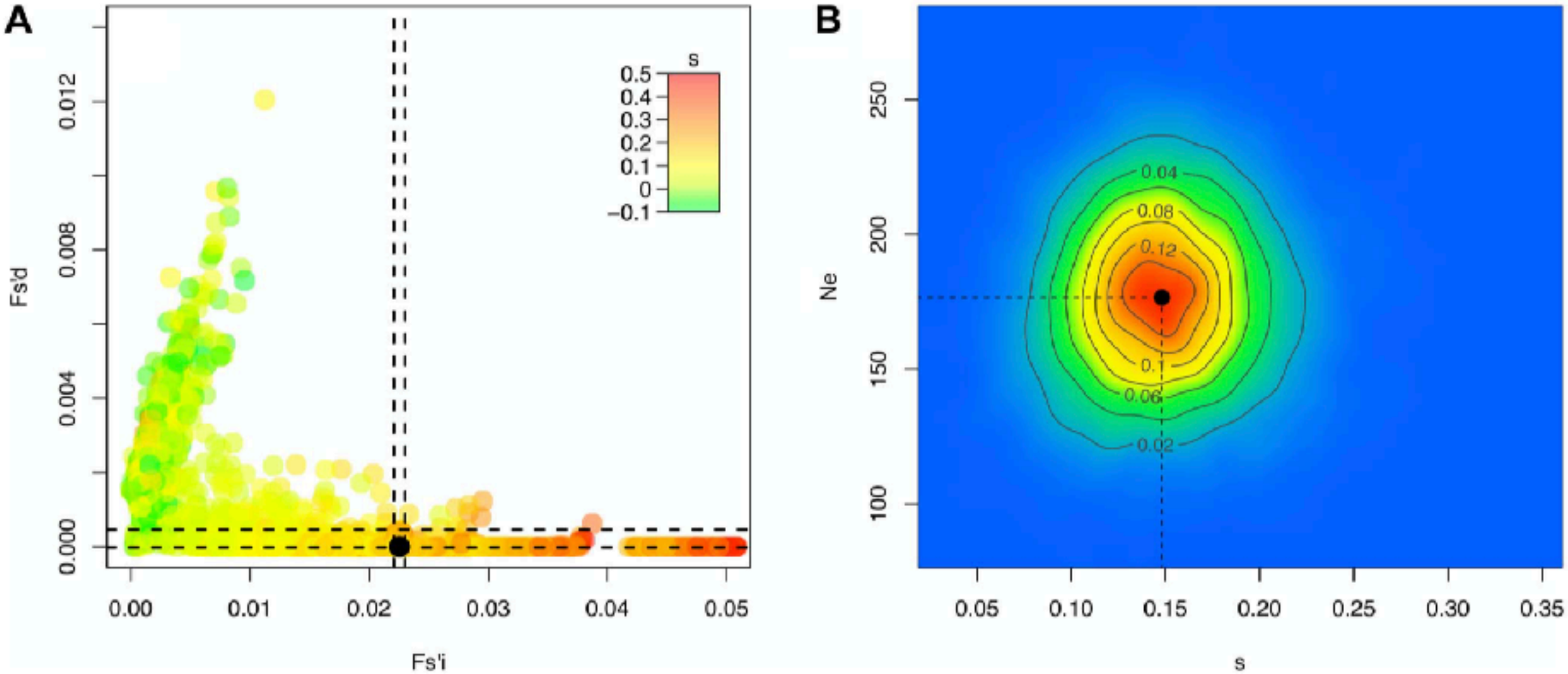
- x, y: minor allele frequencies at two consecutive time points
- $z=(x\text{-}y)/2$
- $\tilde{n}$: harmonic mean of sample sizes n_x and n_y
- t generations between sampling points
- *Fs'* is averaged over sites and times

$$N_e = 1/Fs' \text{ or } N_e = 1/(2Fs')$$

Jorde & Ryman 2007

SUMMARY STATISTICS FOR WFABC

- $U(X_i) = (Fsd_i, Fsi_i)$: for each single trajectory split *Fs'* into *Fsd'* and *Fsi'* prime to determine the directional components in the trajectory.

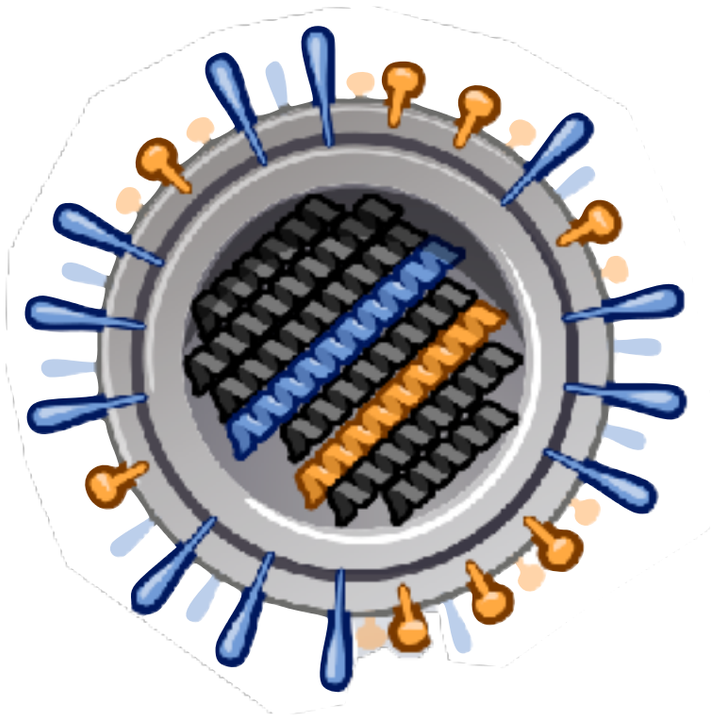


Foll, Poh et al. 2014

WFABC – A SOFTWARE TO INFER EFFECTIVE POPULATION SIZE AND SELECTION FROM TIME-SERIAL DATA

- Input: allele-frequency trajectories (min. 3 time points)
- wfabc_1: Infer effective population size from whole data set
- wfabc_2: Infer selection coefficient from individual trajectories
- ABC method. Output: posterior probabilities

Developed by Matthieu Foll; Foll, Poh et al. 2014, Foll, Shim et al. 2014



INFLUENZA

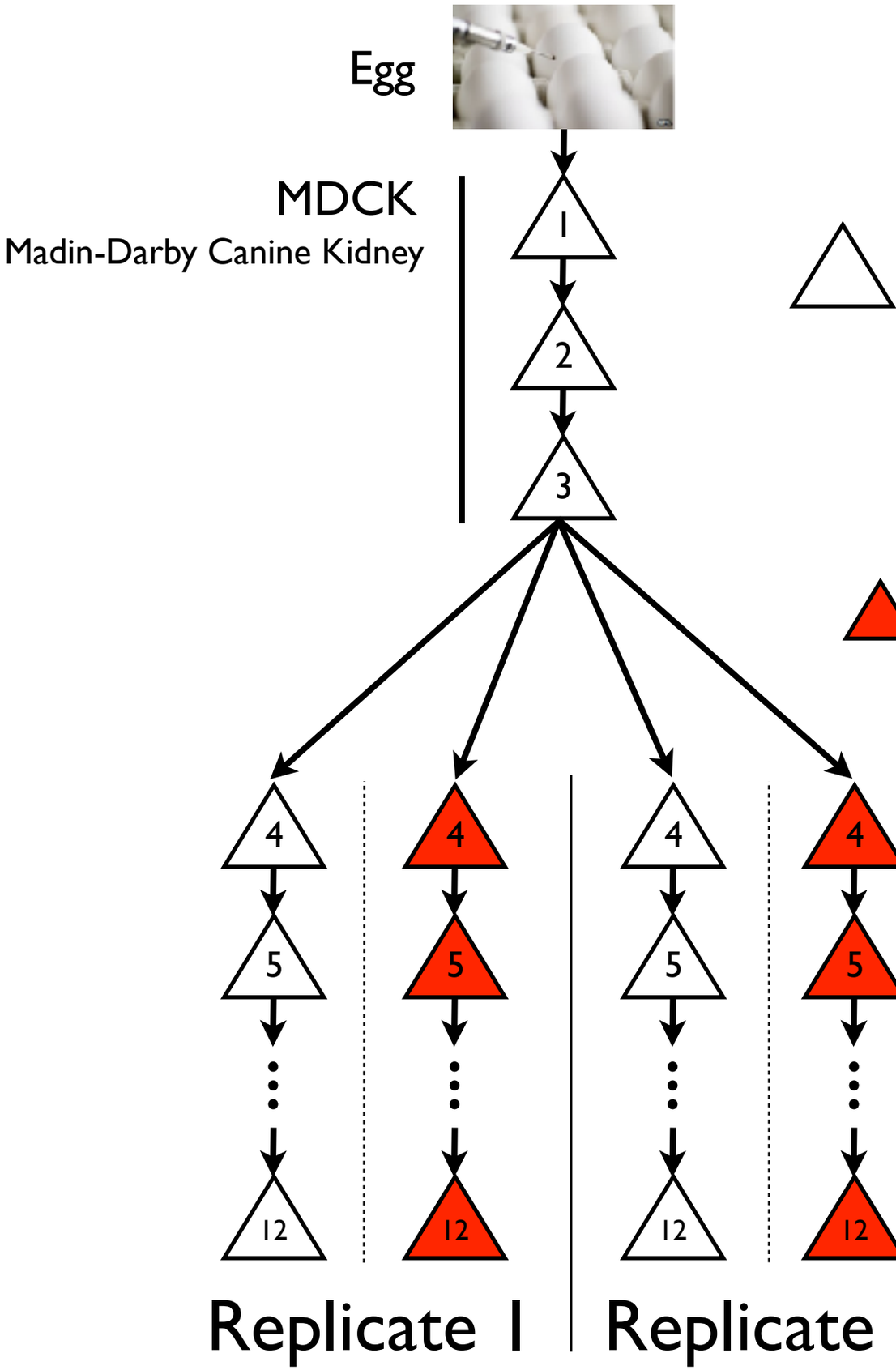
- responsible for 150,000-200,000 deaths each year
- high motivation to develop effective vaccines and treatments

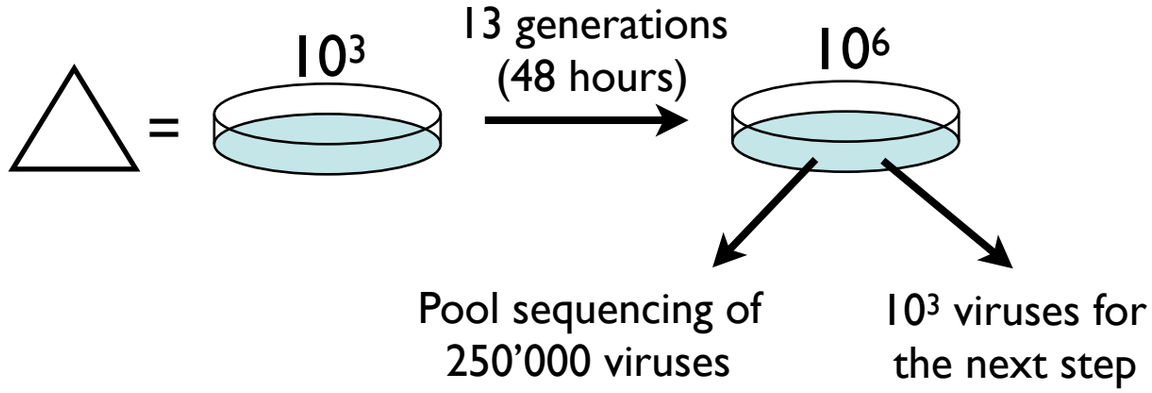
OSELTAMIVIR (TAMIFLU)

- competitive inhibitor of neuraminidase: prevents viral particles from being released by infected cells
- Resistance by single mutation in NA spread rapidly in natural populations

FAVIPIRAVIR

- increases mutation rate by interfering with viral RNA-dependent RNA polymerase - lethal mutagenesis
- Approved in Japan, in trials in USA





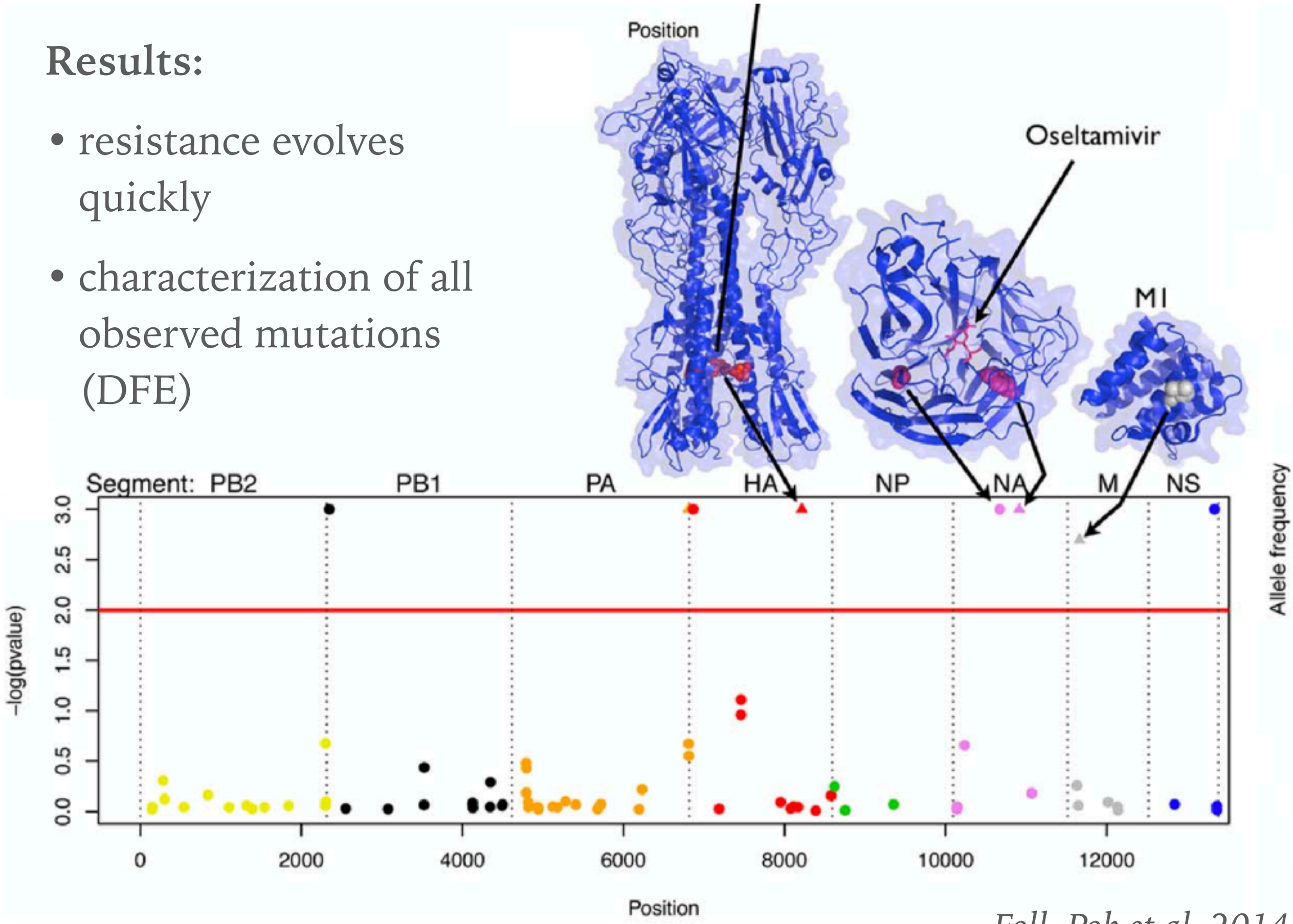
= antiviral drug (oseltamivir=Tamiflu)

- No drug:
  'standard' environment
- Drug:
  challenging environment

Foll, Poh et al. 2014

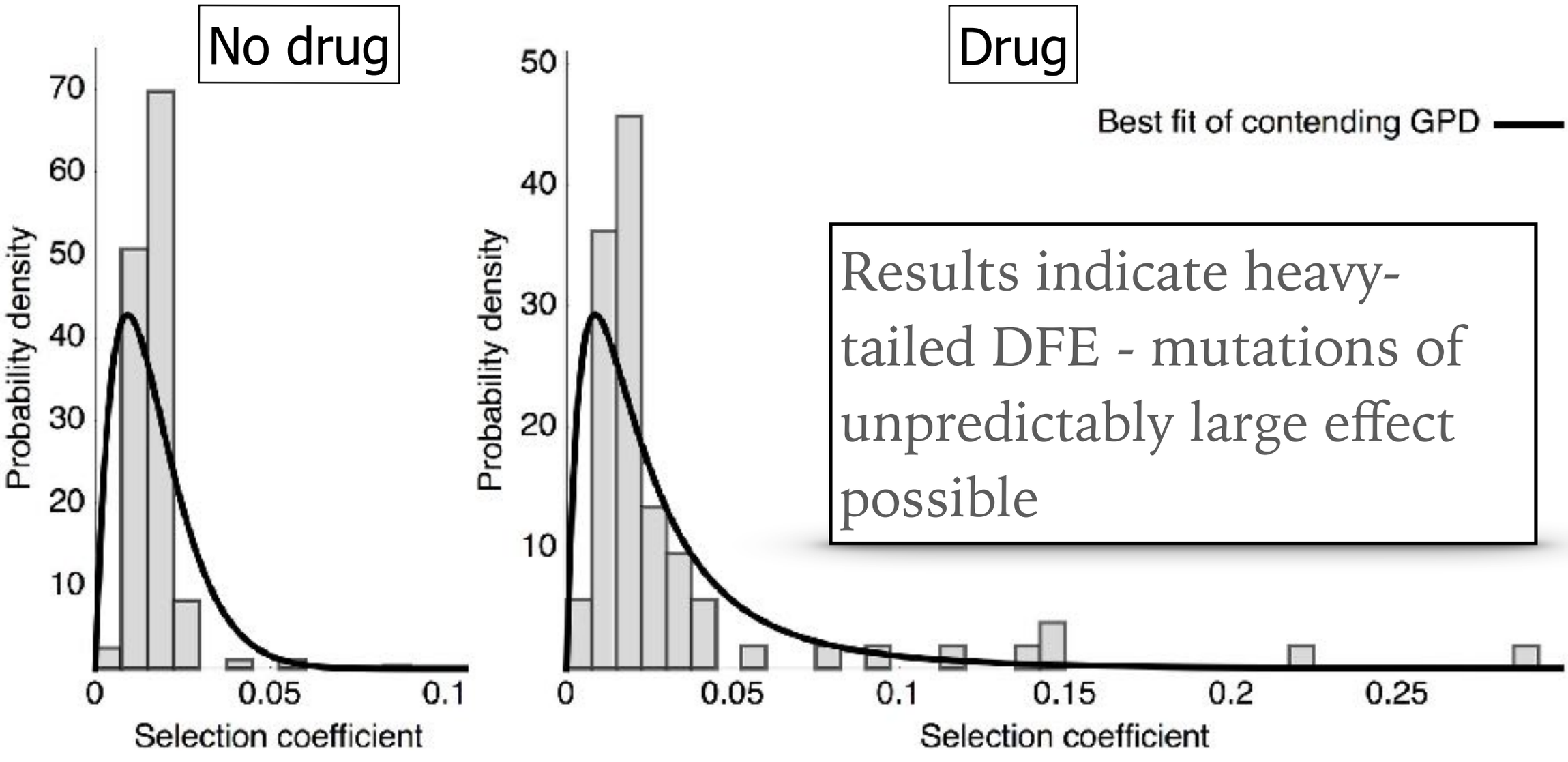
Results:

- resistance evolves quickly
- characterization of all observed mutations (DFE)



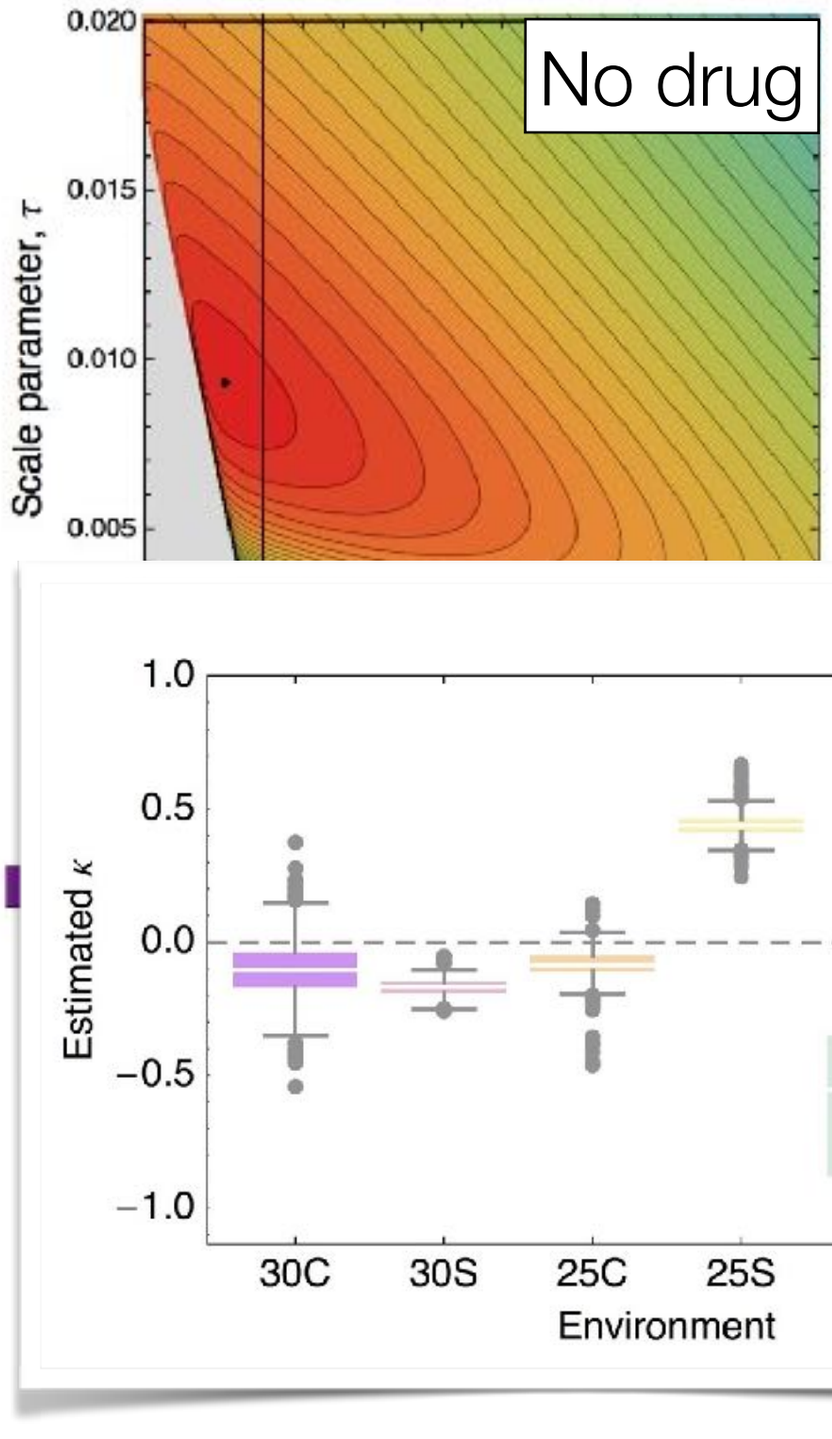
Foll, Poh et al. 2014

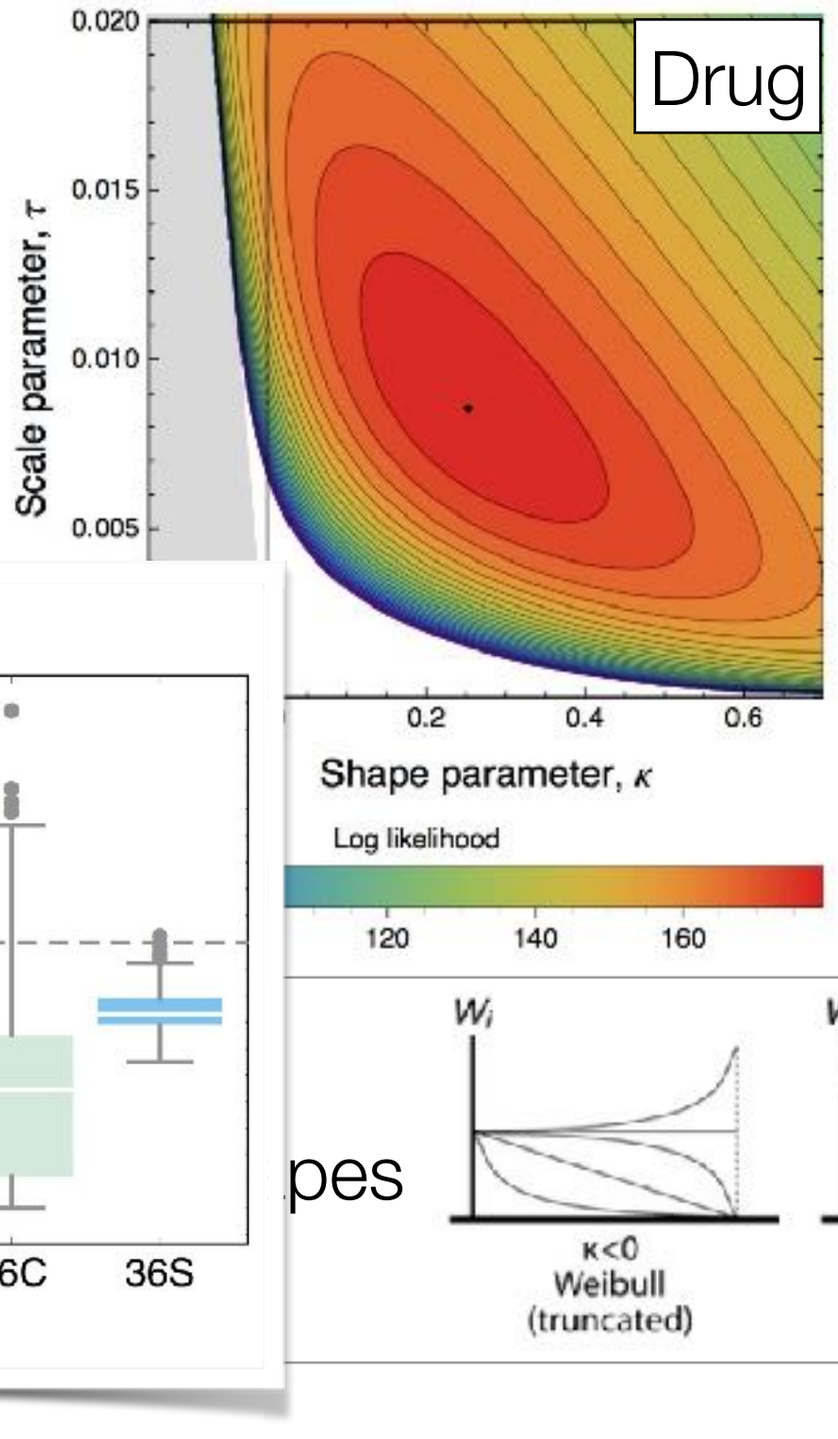
SHAPE OF THE DFE (OF MUTATIONS THAT REACH >2% FREQUENCY)

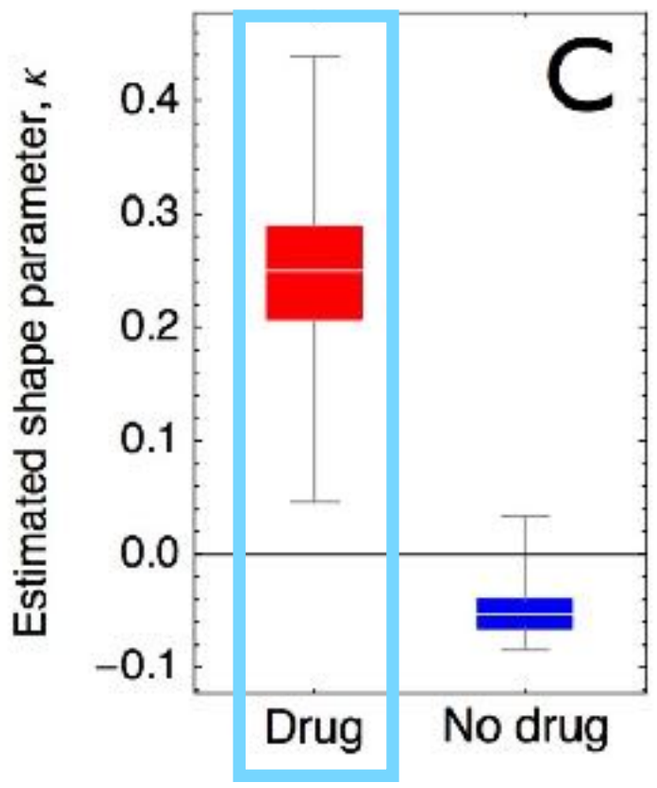


Foll, Poh et al. 2014

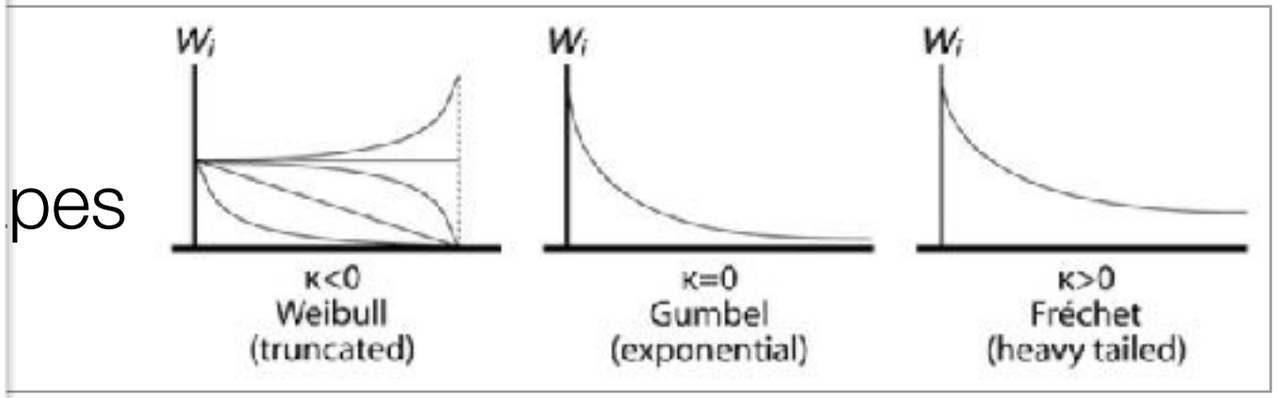
Tail shape parameter in challenging environments







pes



Foll, Poh et al. 2014

EXPERIMENTAL EVOLUTION OF INFLUENZA VIRUS

- Useful approach to study (resistance) evolution, both from a medical and an evolutionary point of view
- Artificial setup allows us to monitor various aspects of the dynamics - bottleneck size, absolute growth rate, genome-wide allele frequencies, cell culture quality - a great testing ground for population-genetic methods

FINALLY SOMETHING ABOUT NEGATIVE SELECTION

WHAT MAKES MUTATION-RATE ENHANCERS EXCITING

MEDICALLY

- could be used against a range of different viruses
- resistance is assumed to be difficult to achieve

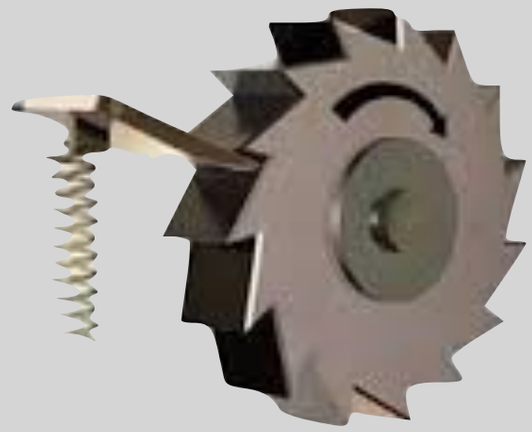
EVOLUTIONARILY

- existing body of theory on the potential effects of high mutation rates
- proposed mechanisms of extinction versus rapid adaptation - potential for validation?

MUTATIONAL MELTDOWN/LETHAL MUTAGENESIS

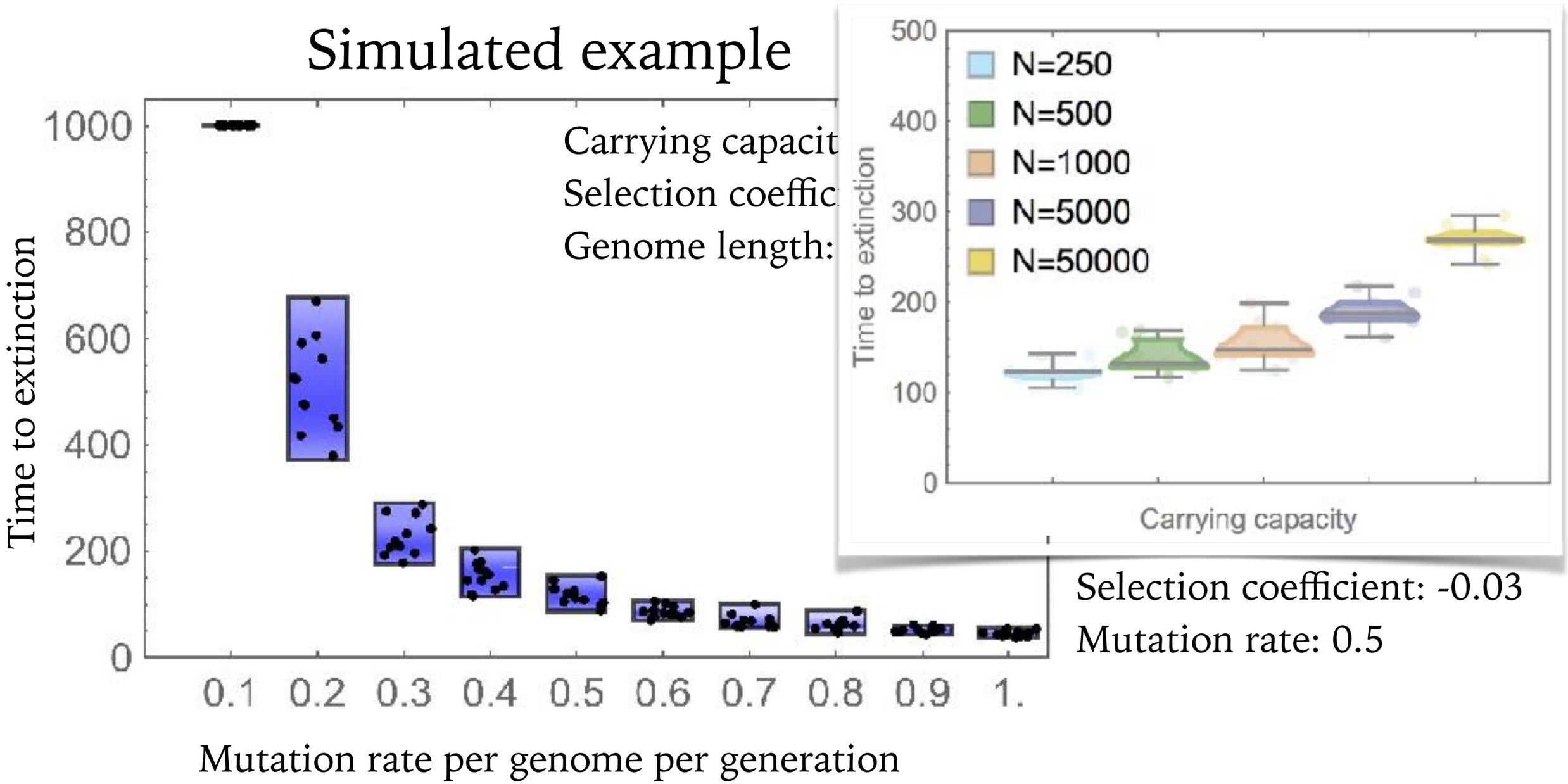
- a population goes extinct because it accumulates too many deleterious mutations (such that the absolute growth rate becomes <1) - this can be caused by mutation pressure or random genetic drift (or both)

Muller's ratchet: the step-wise loss of the fittest genotype due to accumulation of deleterious mutations in asexual populations



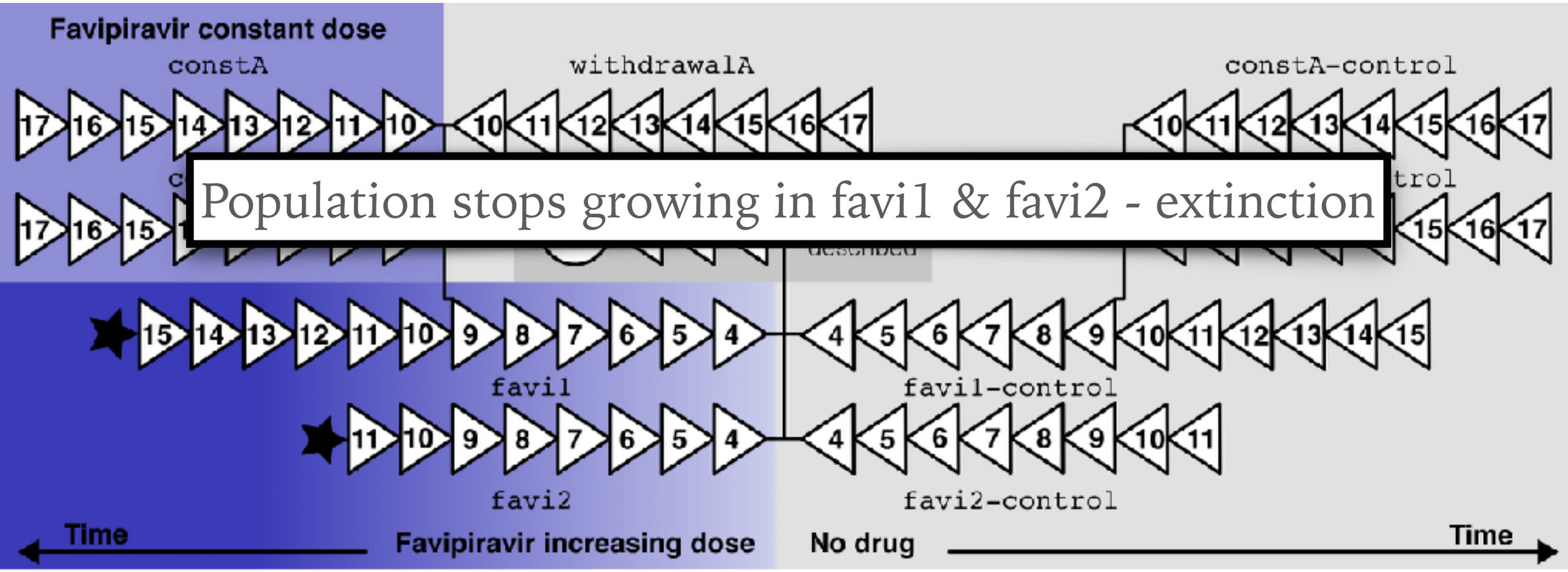
E.g., Lynch et al., 1990, Evolution

MUTAGENIC DRUGS AGAINST RNA VIRUS INFECTIONS



- Mutagenic drug favipiravir approved for use against influenza in Japan and discussed as promising candidate drug against various RNA viruses.

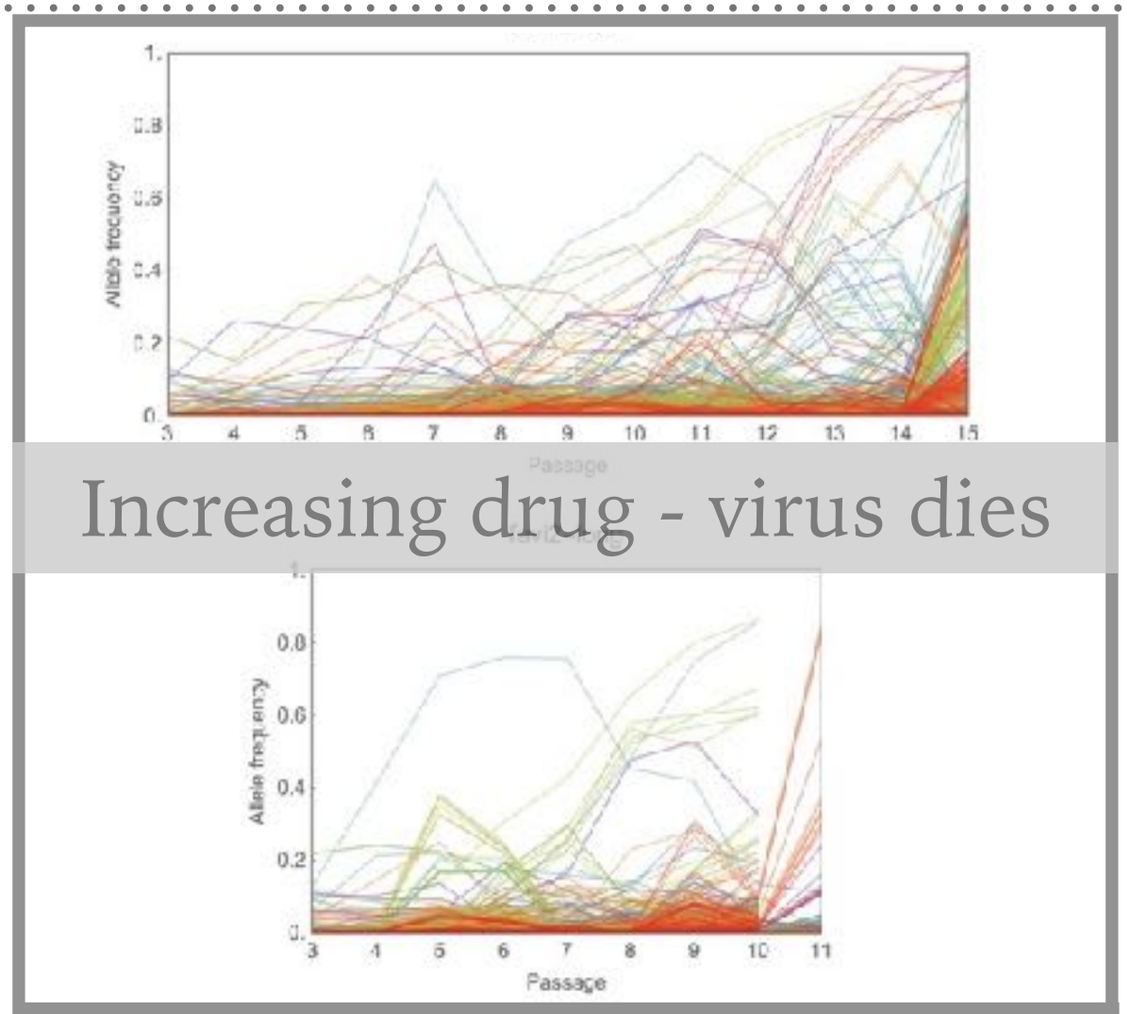
EXPERIMENTAL APPROACH



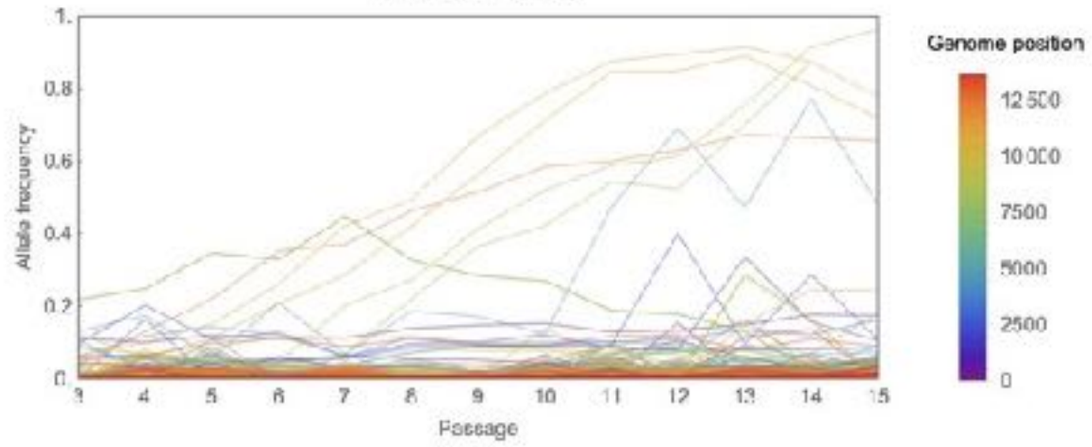
Population stops growing in favi1 & favi2 - extinction

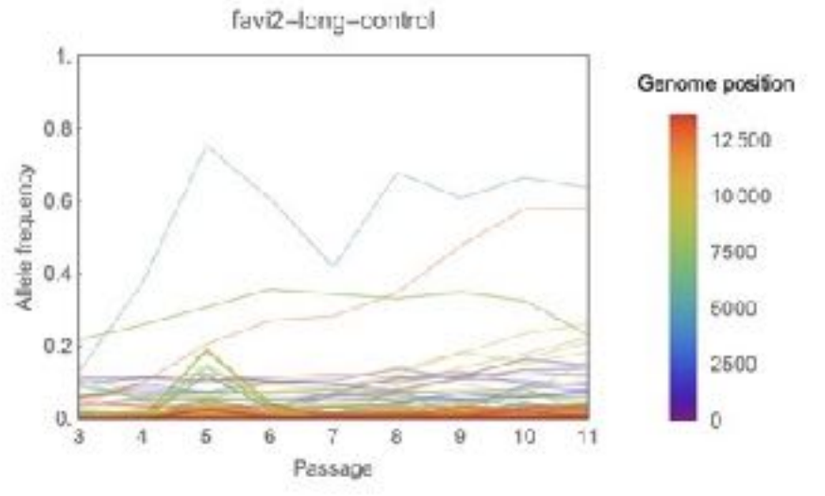
- Describe evolutionary dynamics under different drug treatments
- study potential for resistance mutations

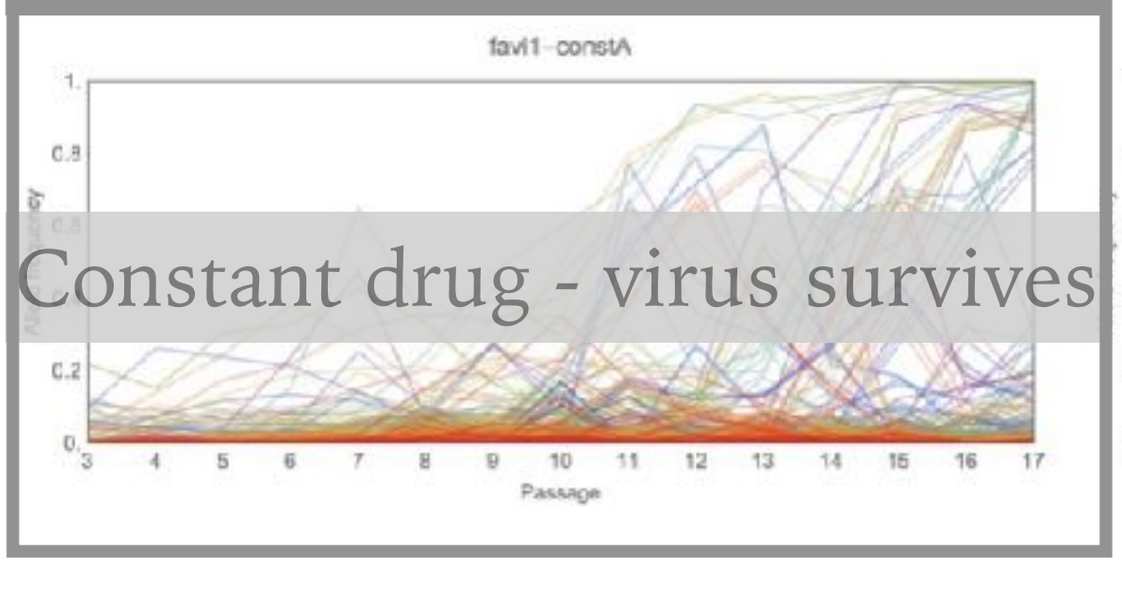
INFLUENZA A LABORATORY EVOLUTION UNDER MUTAGENIC DRUG TREATMENT



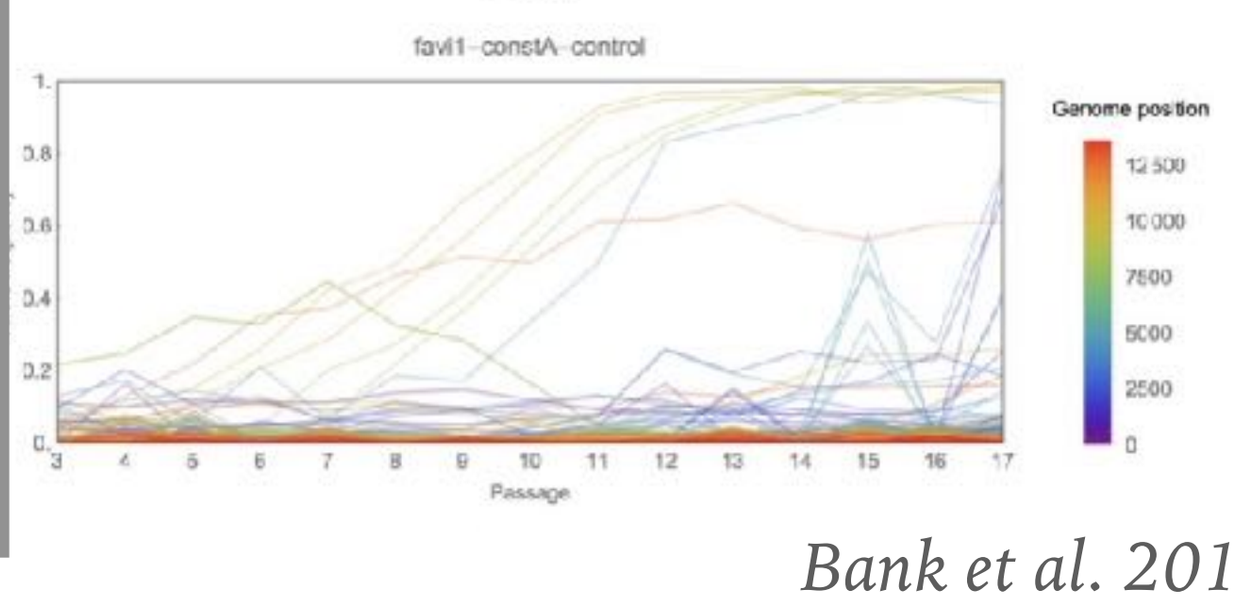
Increasing drug - virus dies



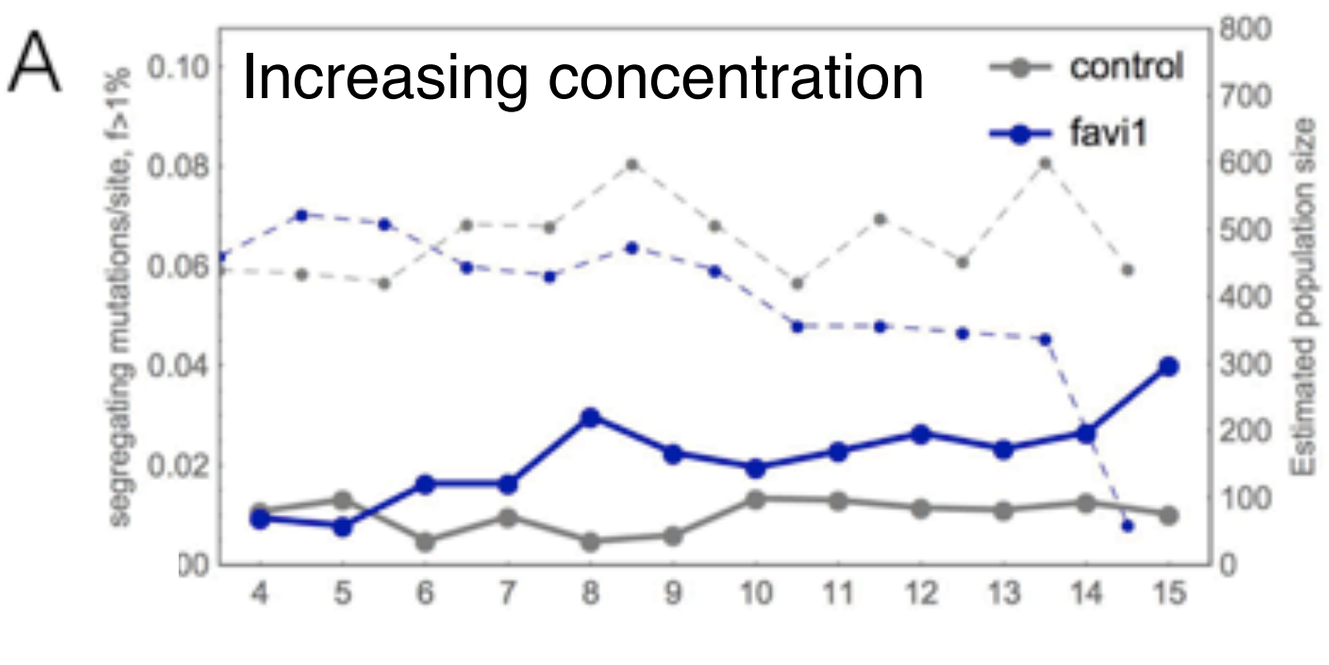


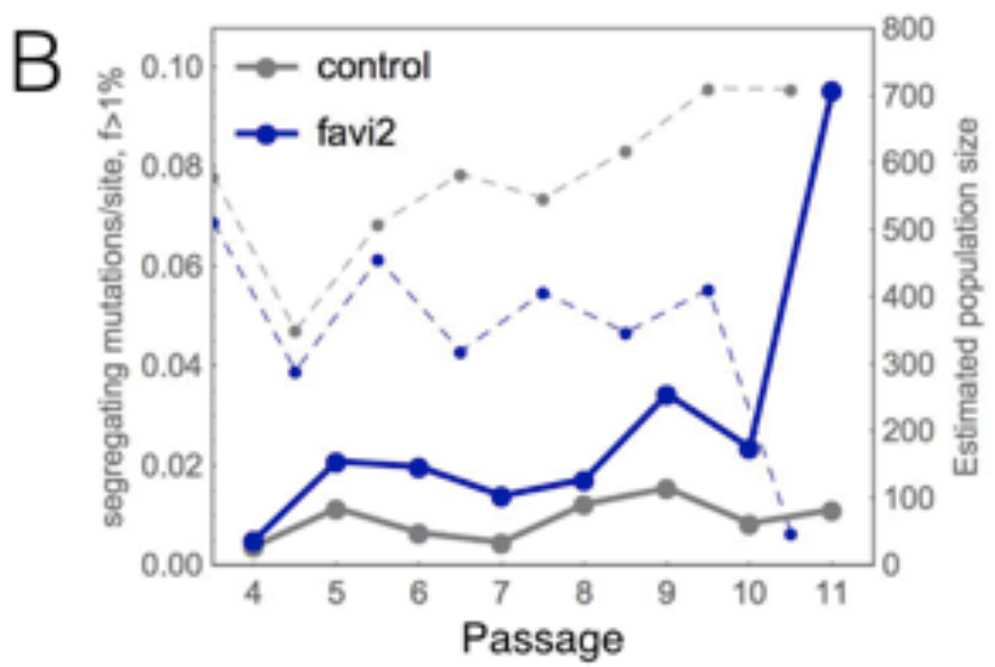


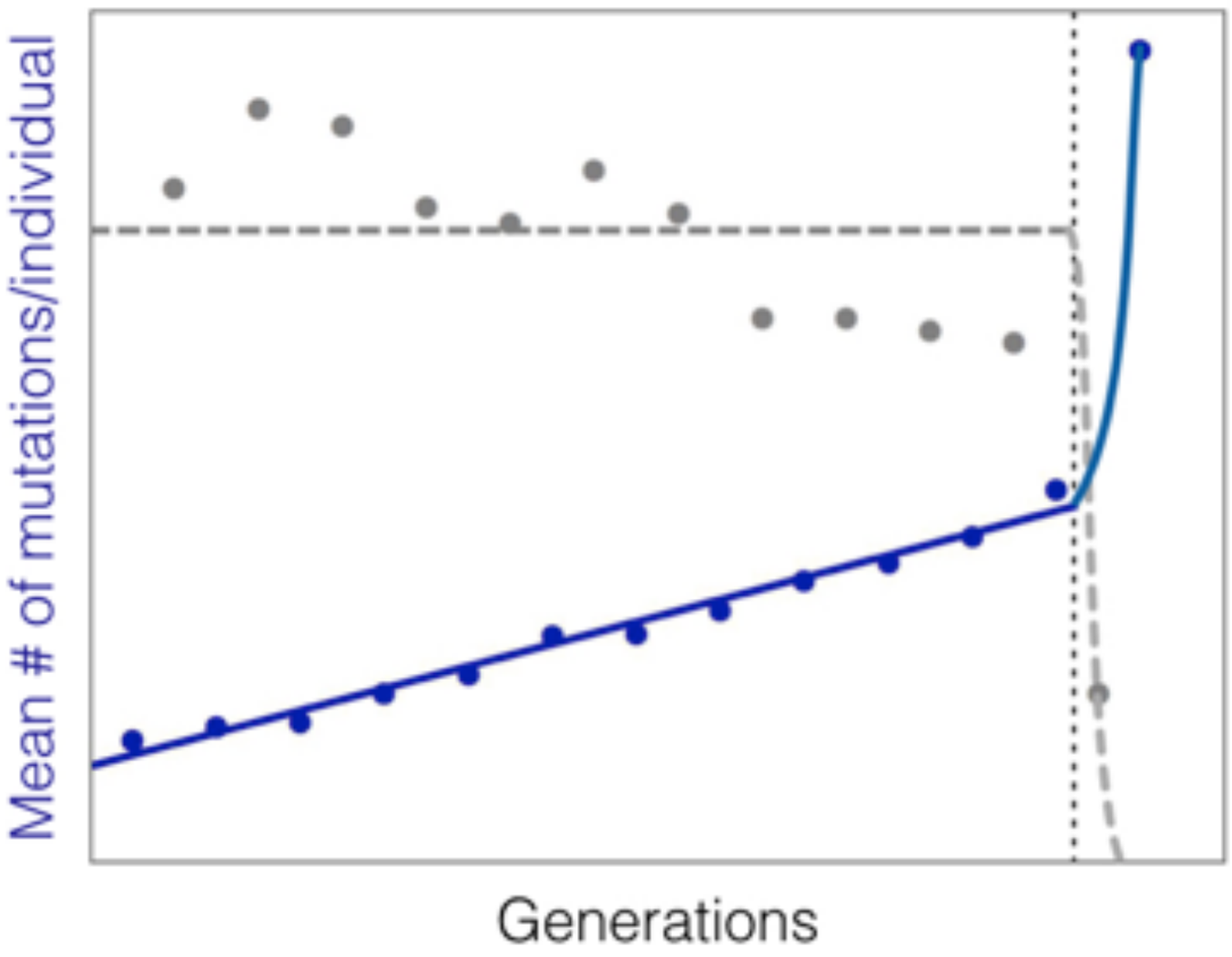
Constant drug - virus survives



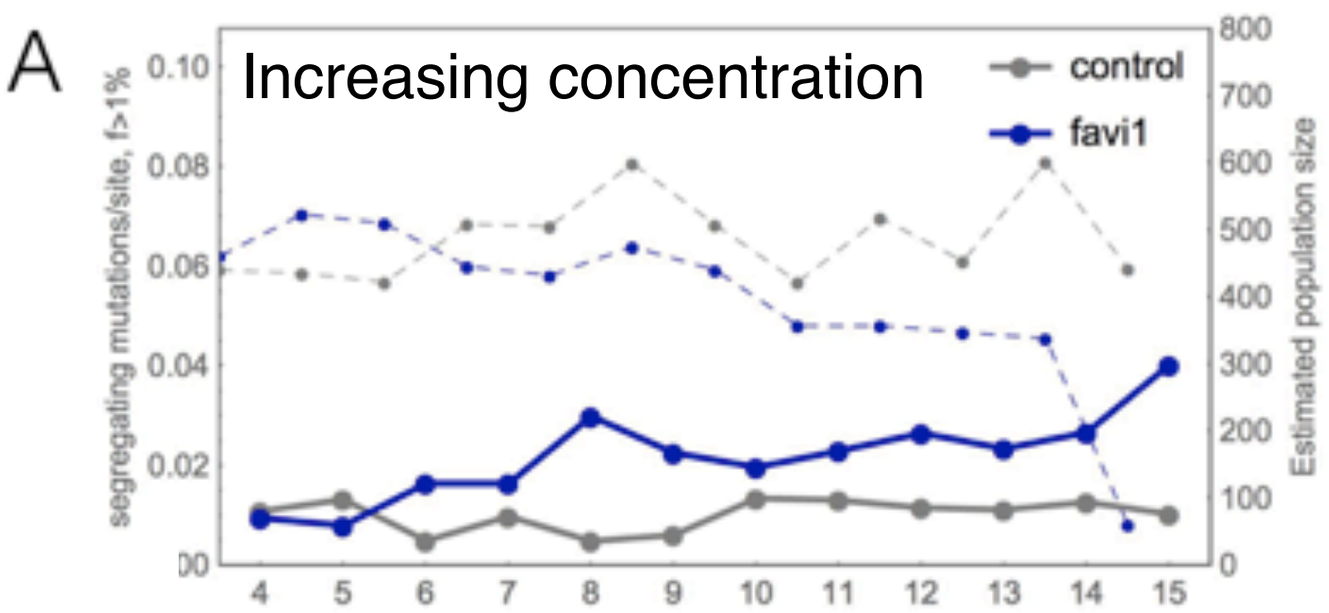
Bank et al. 2016

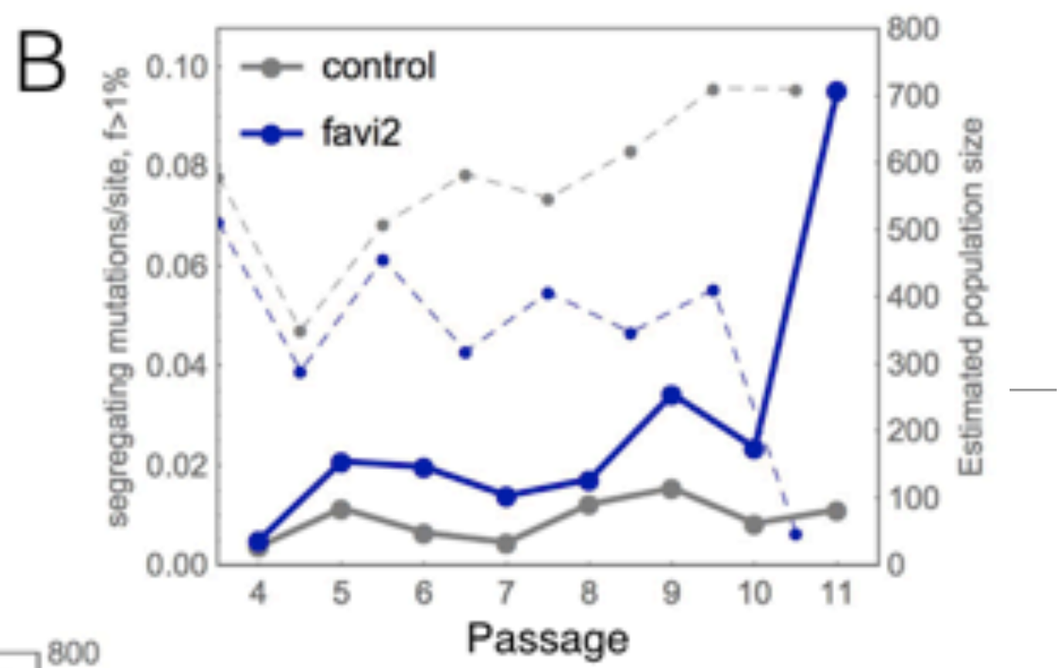


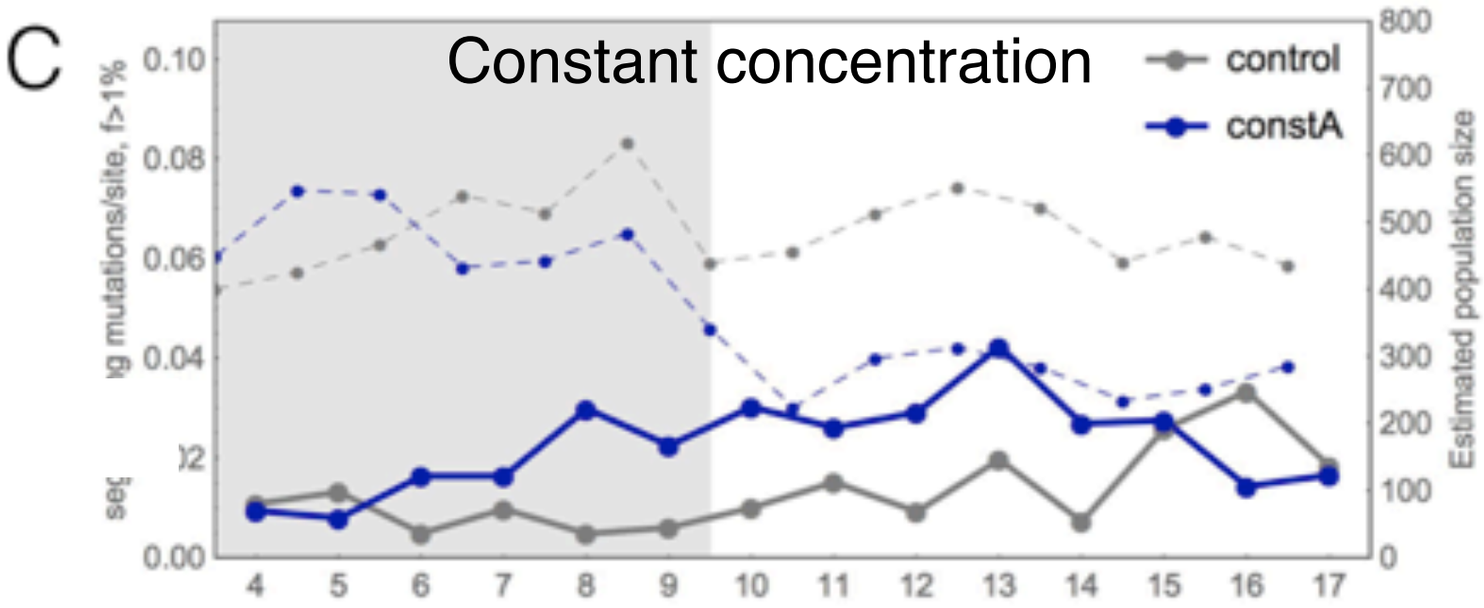


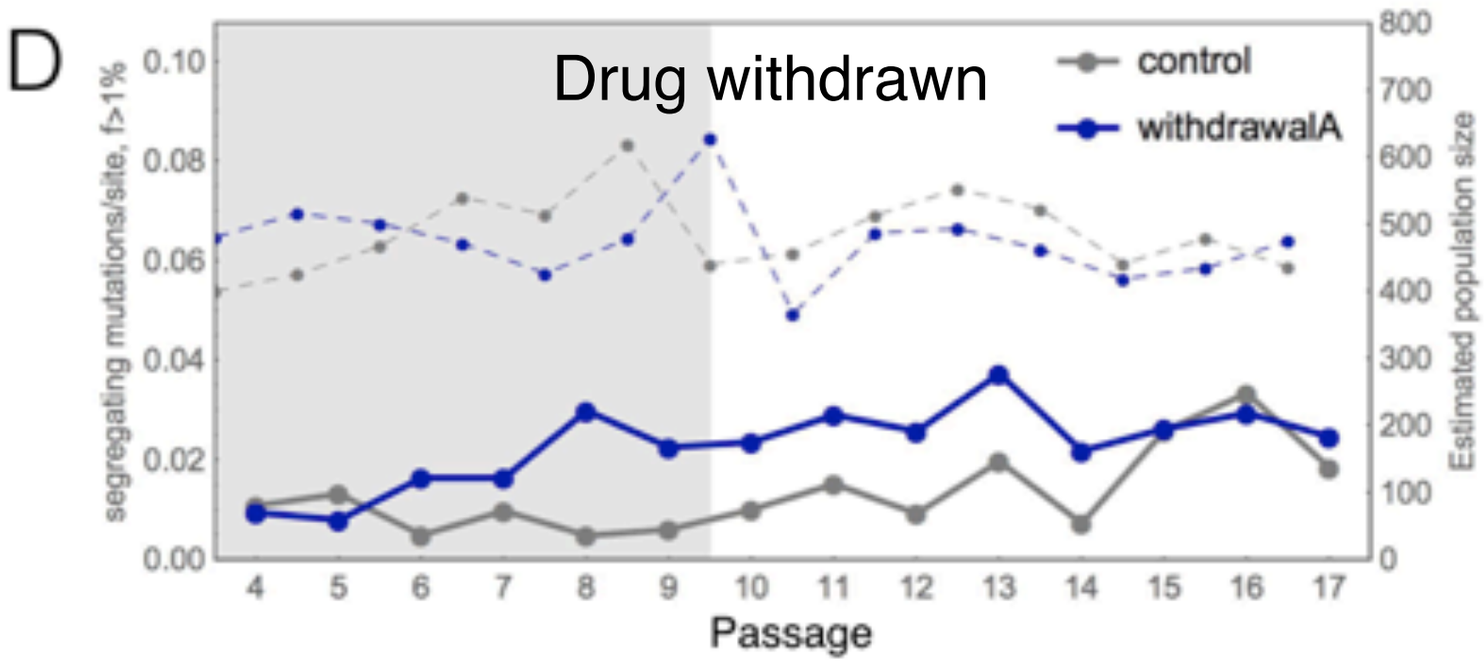


- Clear increase in number of segregating mutations
- Decrease in effective population size, esp. right before extinction
- Dynamics follow prediction from Lynch et al., 1993 for mutational meltdown







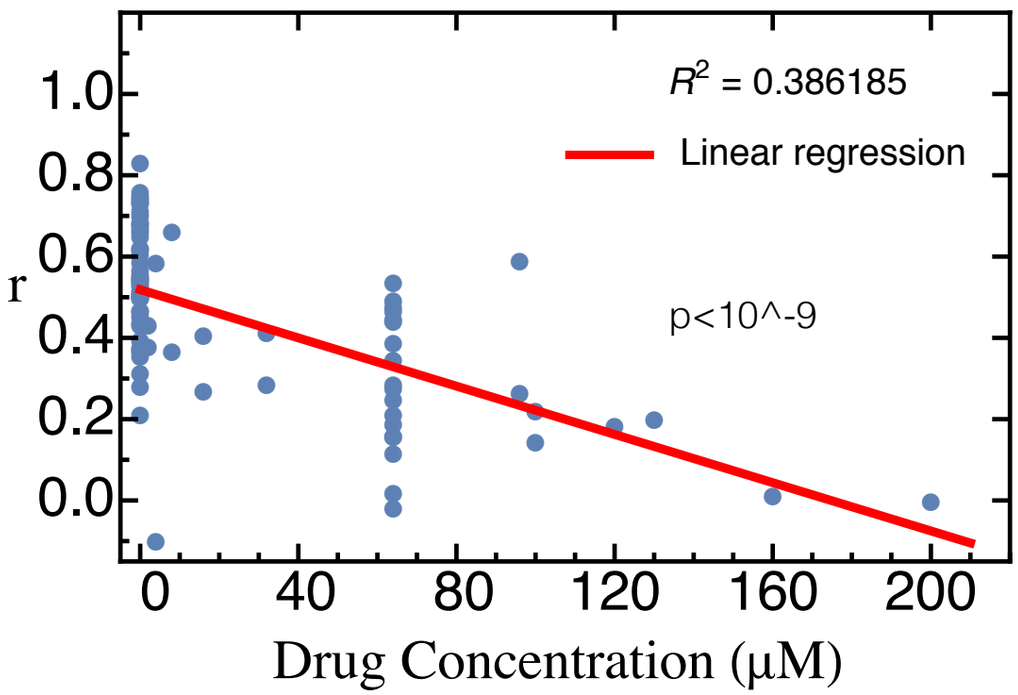


- Immediate recovery of effective population size but not # of segregating mutations in withdrawal
- survival of population at constant dose, but at low effective pop. size

Absolute growth rates

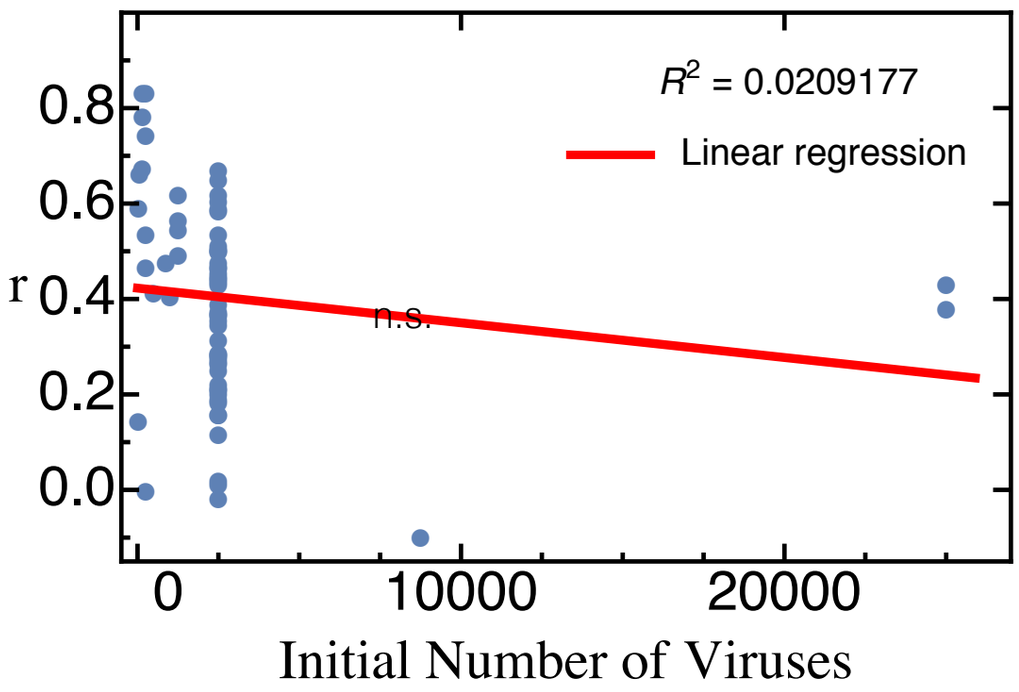
Growth rate decreases w drug concentration

A



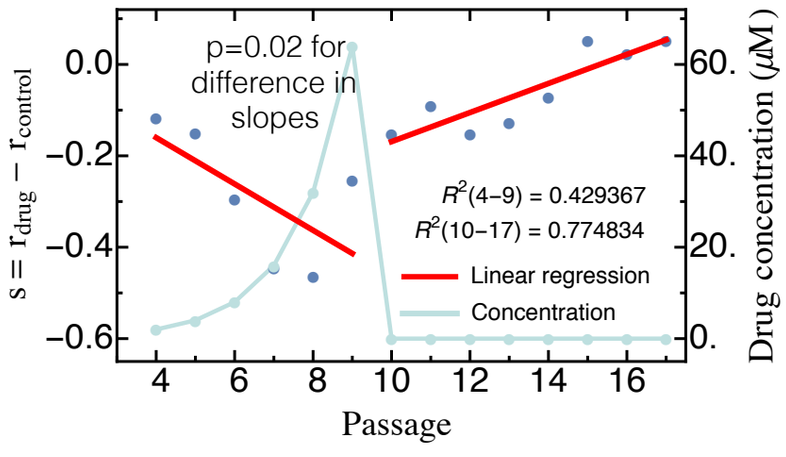
No correlation between growth rate and MOI

B

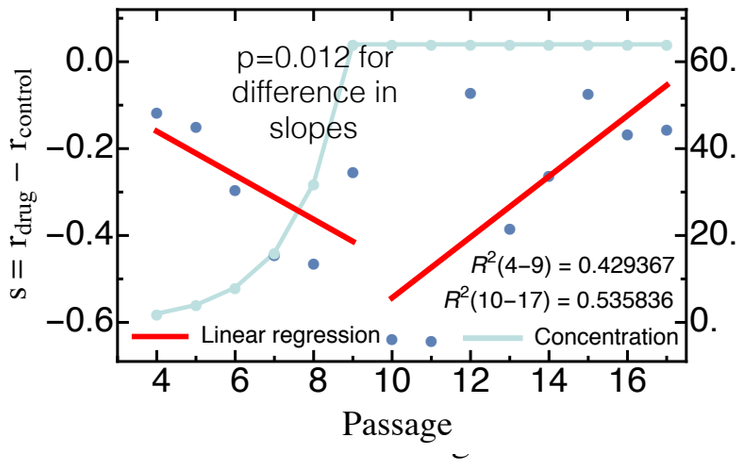


- obtained via MOI and virus output

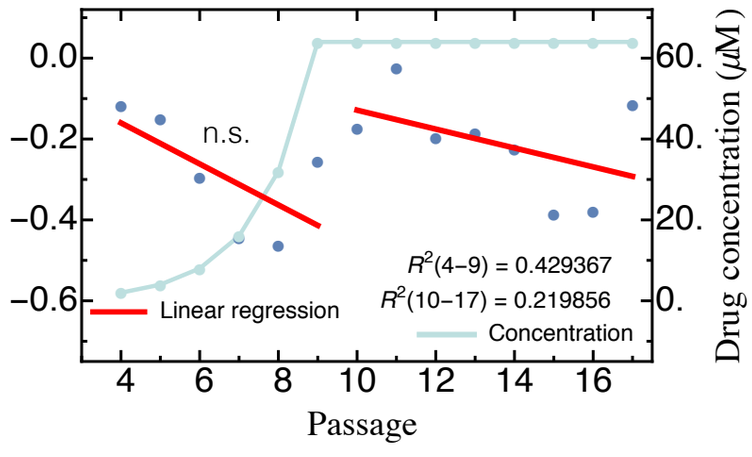
Absolute growth rates



WithdrawalA



ConstA

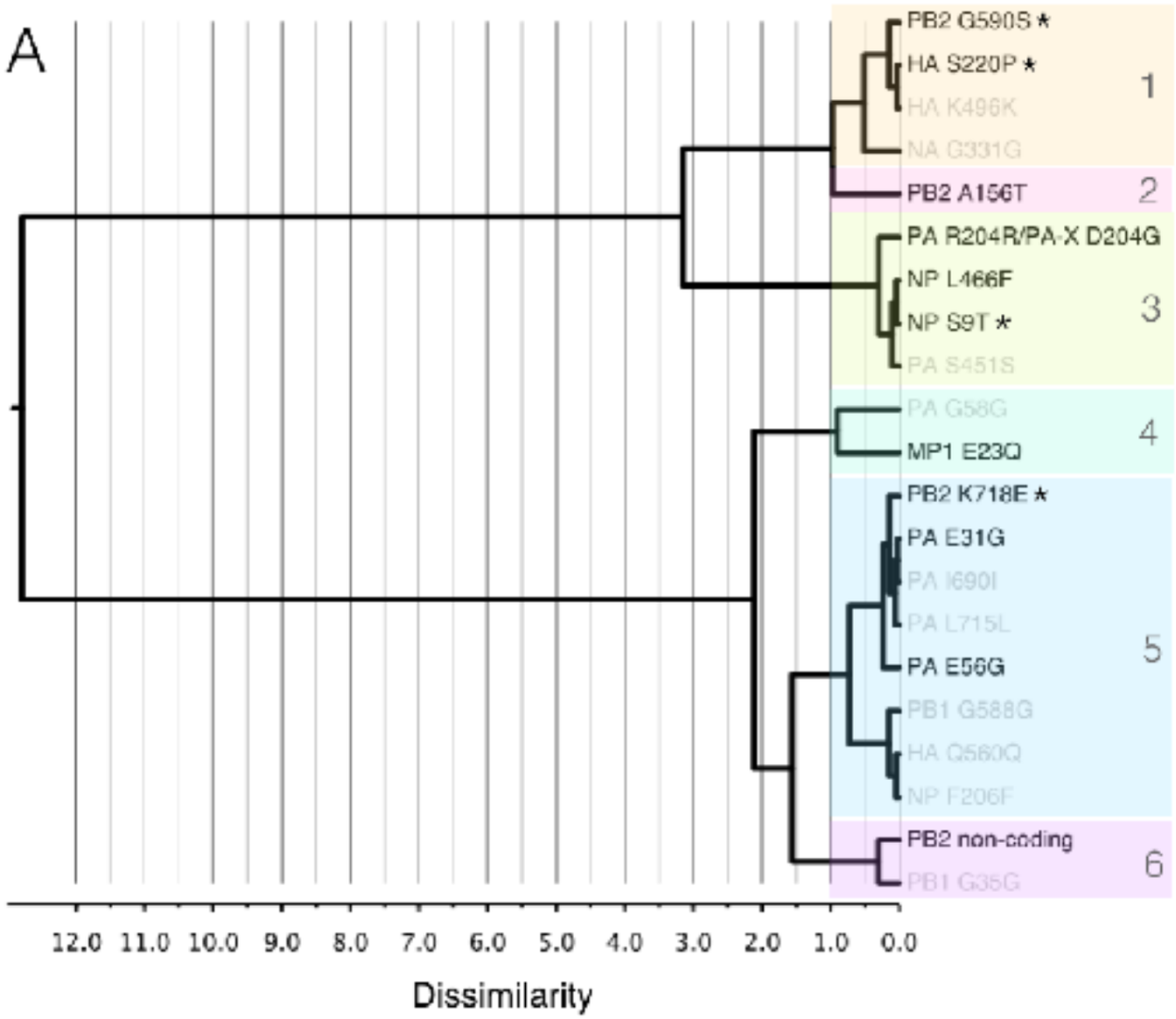


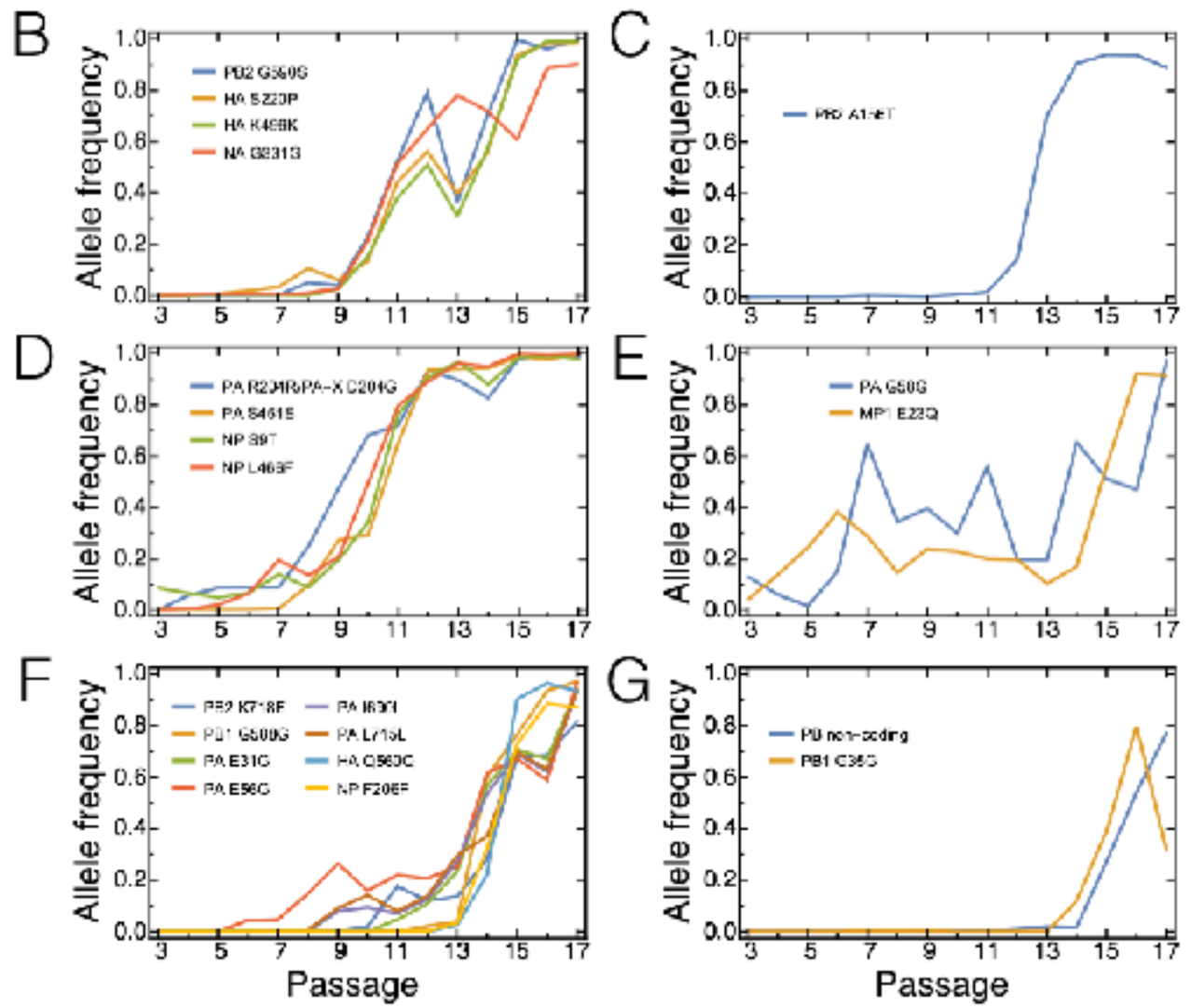
ConstB

- recovery of growth rate in withdrawal
- evidence for evolving recovery on constA - resistance?
- indeed greatest number of (and most compelling evidence for) adaptations in constA!

CAN VIRUSES ADAPT TO MUTAGENIC DRUG TREATMENTS?

- "Adaptation" in this context means survival/persistence of a pathogen or other health threat despite exposure to drug, immune system, novel environments, etc.
- **By which mechanisms can viruses escape from mutagenic drug treatment? Can we detect the signatures of such adaptation? What are the dangers of mutagenic drugs?**
- An example of **evolutionary rescue**: an adaptation spreads in a population that is otherwise doomed to extinction due to a change in the environment





- similarity between trajectories indicates hitchhiking/ joint selection
- WFABC candidates in constA provide focal set, which then can be refined
- clusters indicate potential “adaptation story"

Data set	# candidates	Extinction observed?	Increased # mutations?	Indication of recovery?	Reduced Ne?
favi1	5	yes	yes	no	yes
favi2	3	yes	yes	no	yes
constA	18	no	yes	yes	yes
constB	6	no	yes	unclear	yes
withdrawalA	1	no	yes	yes	no

Mechanism is working

Drug challenges populations

Data set	# candidates	Extinction observed?	Increased # mutations?	Indication of recovery?	Reduced Ne?
favi1	5	yes	yes	no	yes
favi2	3	yes	yes	no	yes
constA	18	no	yes	yes	yes
constB	6	no	yes	unclear	yes
withdrawalA	1	no	yes	yes	no

Mutational meltdown

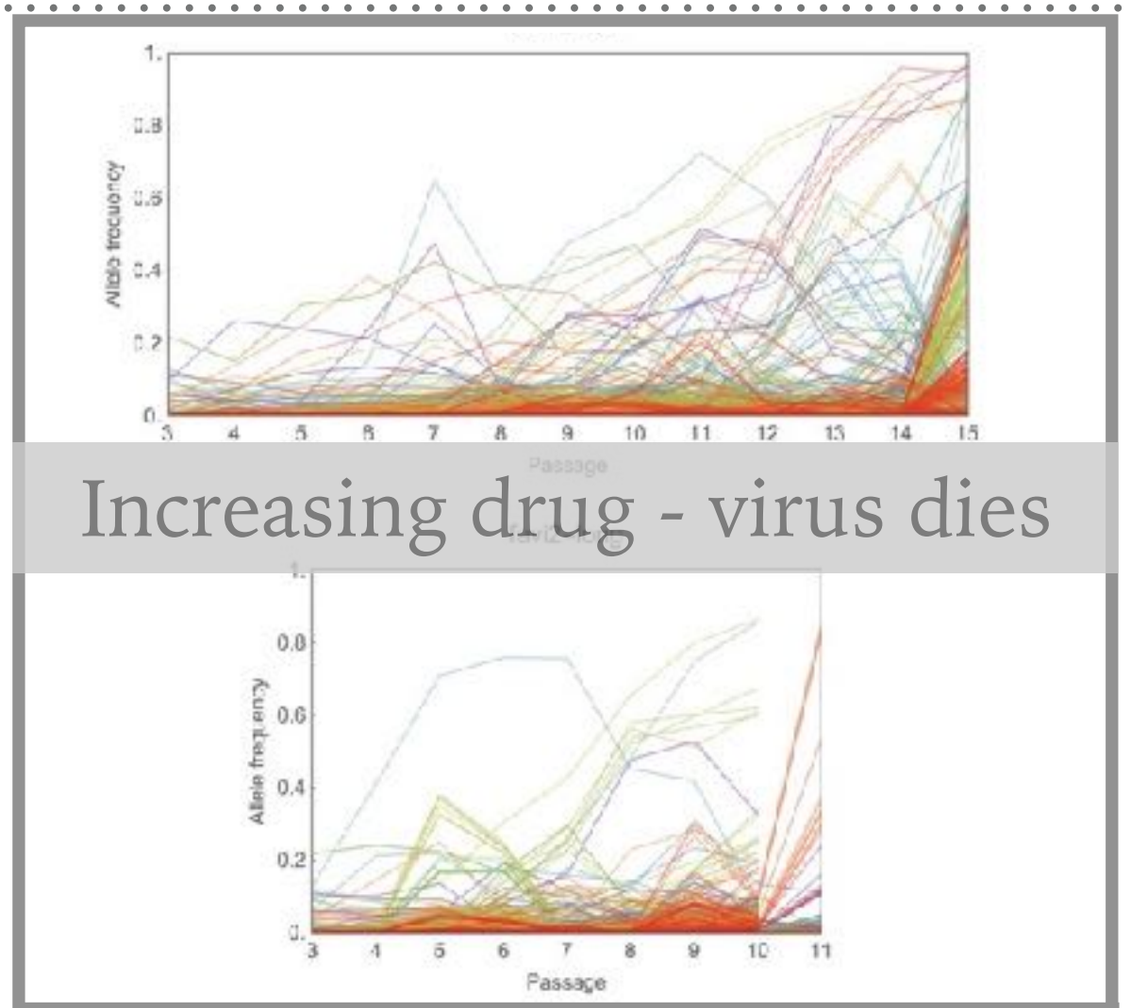
Data set	# candidates	Extinction observed?	Increased # mutations?	Indication of recovery?	Reduced Ne?
favi1	5	yes	yes	no	yes
favi2	3	yes	yes	no	yes
constA	18	no	yes	yes	yes
constB	6	no	yes	unclear	yes
withdrawalA	1	no	yes	yes	no

Resistance evolution? Evolutionary rescue?

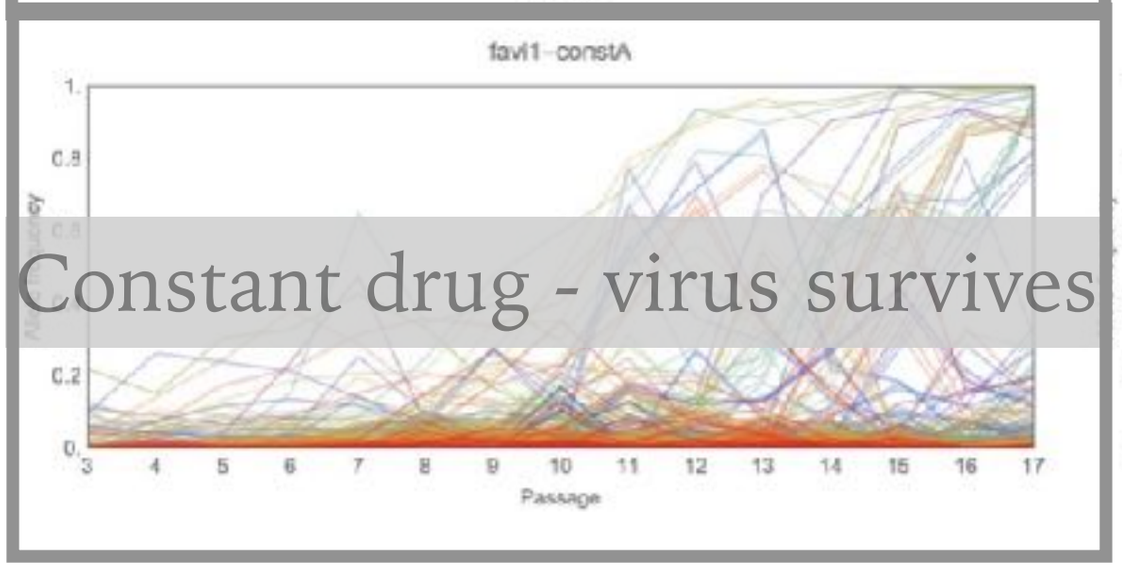
SUMMARY/CONCLUSION OF THE STUDY

- We observe mutational meltdown in action - i.e., the drug is effective in favi1 & favi2.
- We see potential for resistance evolution (à la evolutionary rescue?) under constant doses of favipiravir - i.e., drug doses have to be sufficiently high for success, otherwise the increase in mutation rate may even allow for a speedup of adaptation
- Novel time-serial approaches enable the identification of candidates, which can be tested functionally in the future.

WHERE'S THE CATCH?

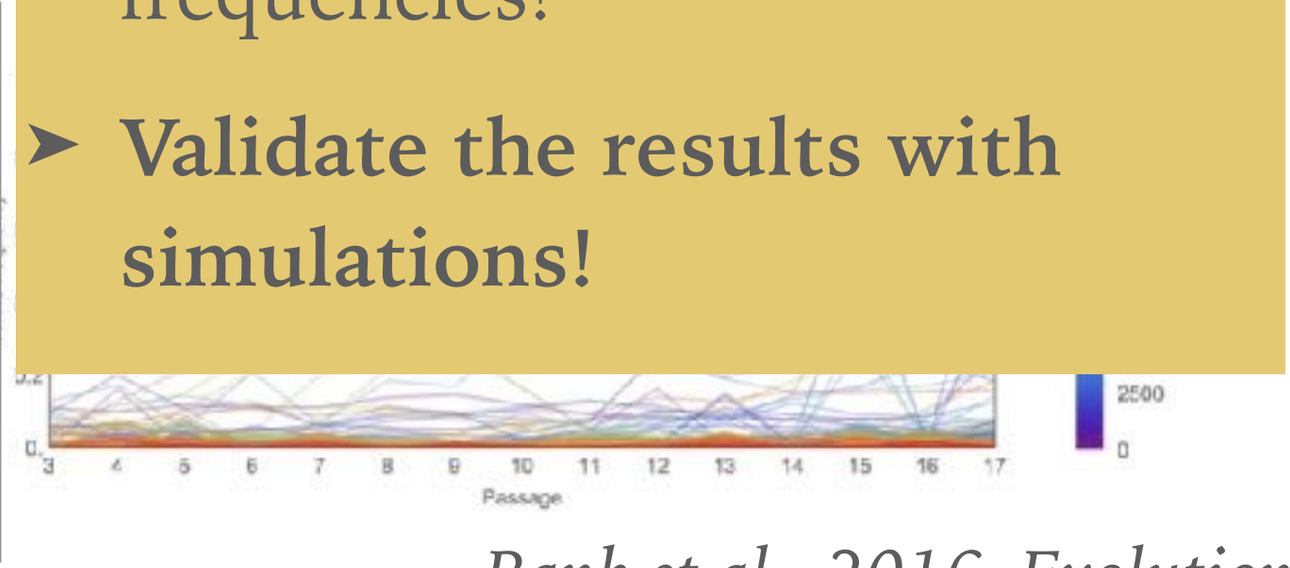


Increasing drug - virus dies



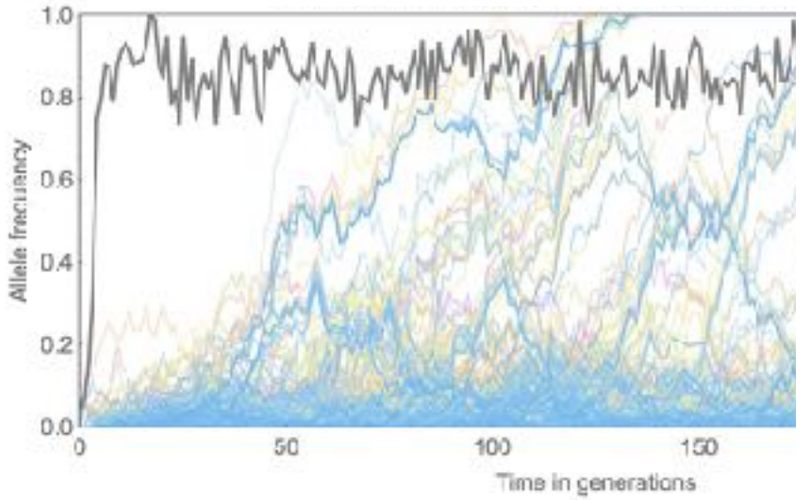
Constant drug - virus survives

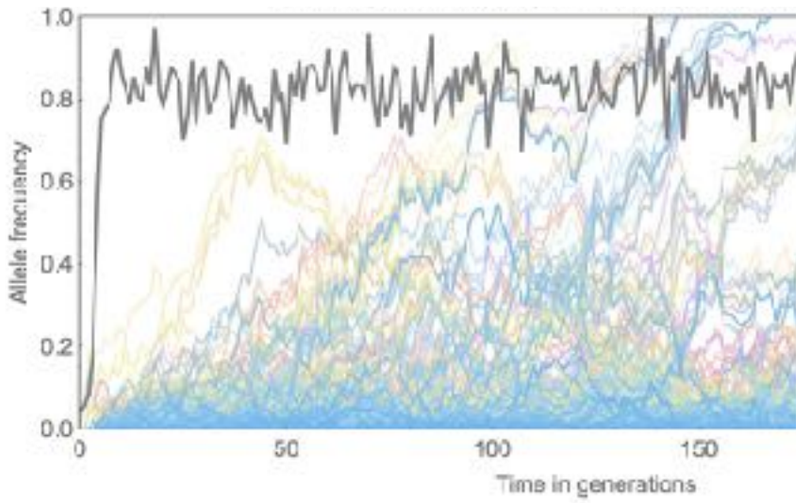
- How can the virus adapt to the drug? What is the signature of different adaptation mechanisms?
- How good are our methods for detection of candidate loci?
- How informative are allele frequencies?
- **Validate the results with simulations!**

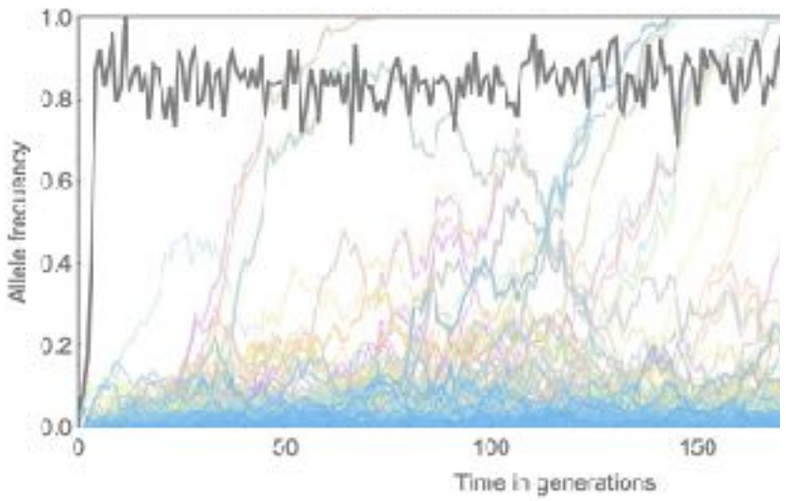


Bank et al., 2016, Evolution

SIMULATE EVOLUTION OF A CLONAL POPULATION WITH HIGH MUTATION RATES







POTENTIAL MECHANISMS OF RESCUE FROM INCREASED MUTATION RATES

- "**traditional" beneficial mutations** that increase growth rate: only a temporary fix because they will not stop the ratchet
- a mutation rate modifier that reduces the mutation rate below the critical level: **evolution of drug resistance**
- a modifier of the fitness distribution, i.e. a mutation that changes mutational effects genome-wide: **evolution of drug tolerance**

Important to note: both weaker and stronger effects of (deleterious) mutations can slow down the ratchet (Gordo & Charlesworth 2000)

Tolerance could be the most dangerous mechanism of adaptation to mutagenic drugs because it allows the virus to propagate at high mutation rates, which may allow rare/unseen/complex beneficial mutations to invade subsequently.

TODAY'S QUESTIONS

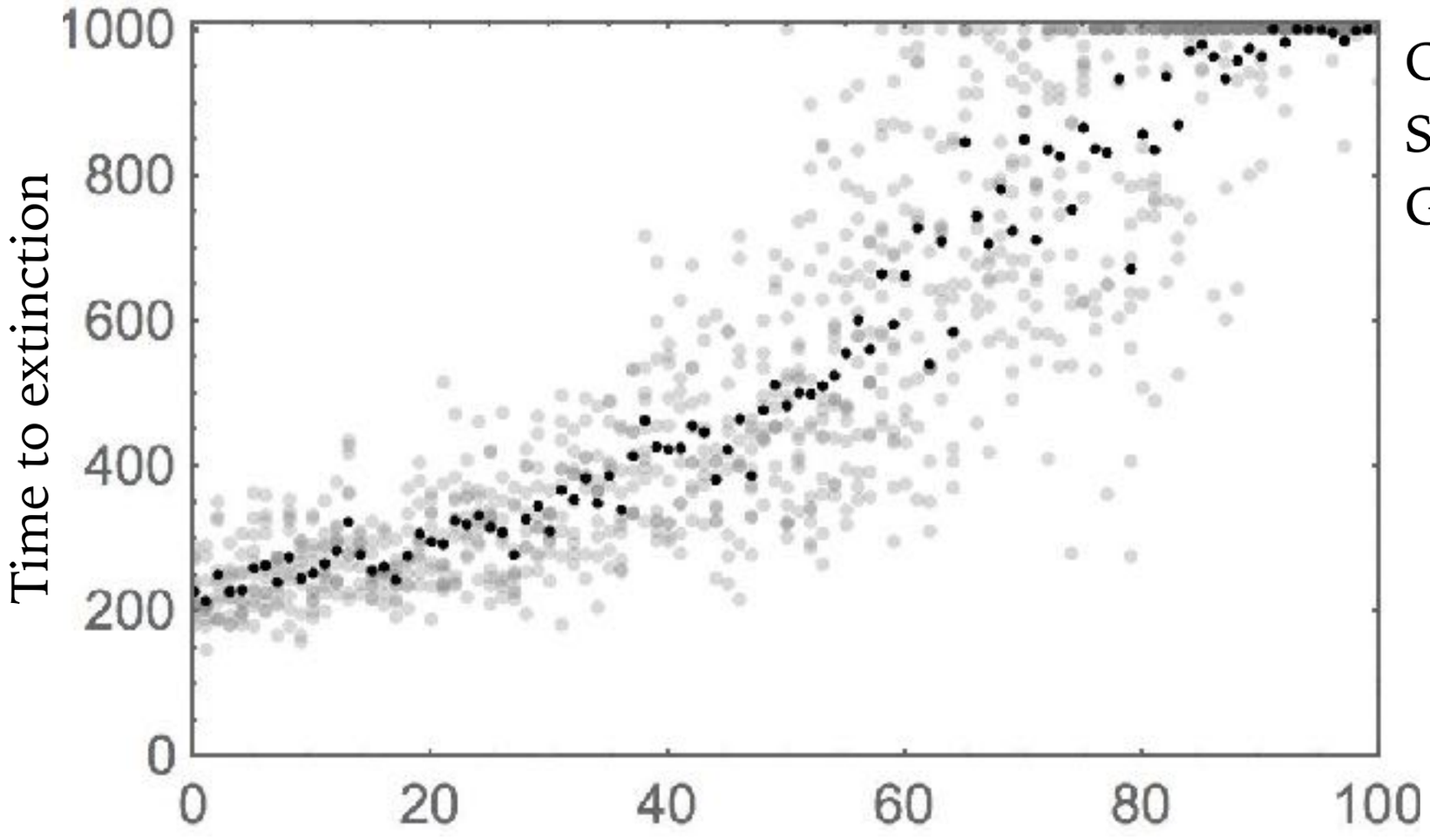
- How does the availability of “traditional" beneficials prolong extinction times?
- When does a mutation rate modifier invade?
- In which conditions does a modifier of the distribution of fitness effects (DFE) invade?

SIMULATION DETAILS

- Genome with L di-allelic loci [1000]
- Carrying capacity C of the clonal population [250], initial population size C_0 [invasion size: 10]
- Initial absolute growth rate R [2]
- Arbitrary distribution of fitness effects [-0.05; multiplicative]
- Mutation rate μ per genome per generation [0.3]
- Record haplotypes in each generation, stop if no extinction has occurred after 1000 generations (transmission/immune reaction)
- l loci with “adaptive" mutations; either beneficial, mutation rate modifier, or DFE modifier

For now: focus on extinction time & “rescue" probability - Later: compare trajectories

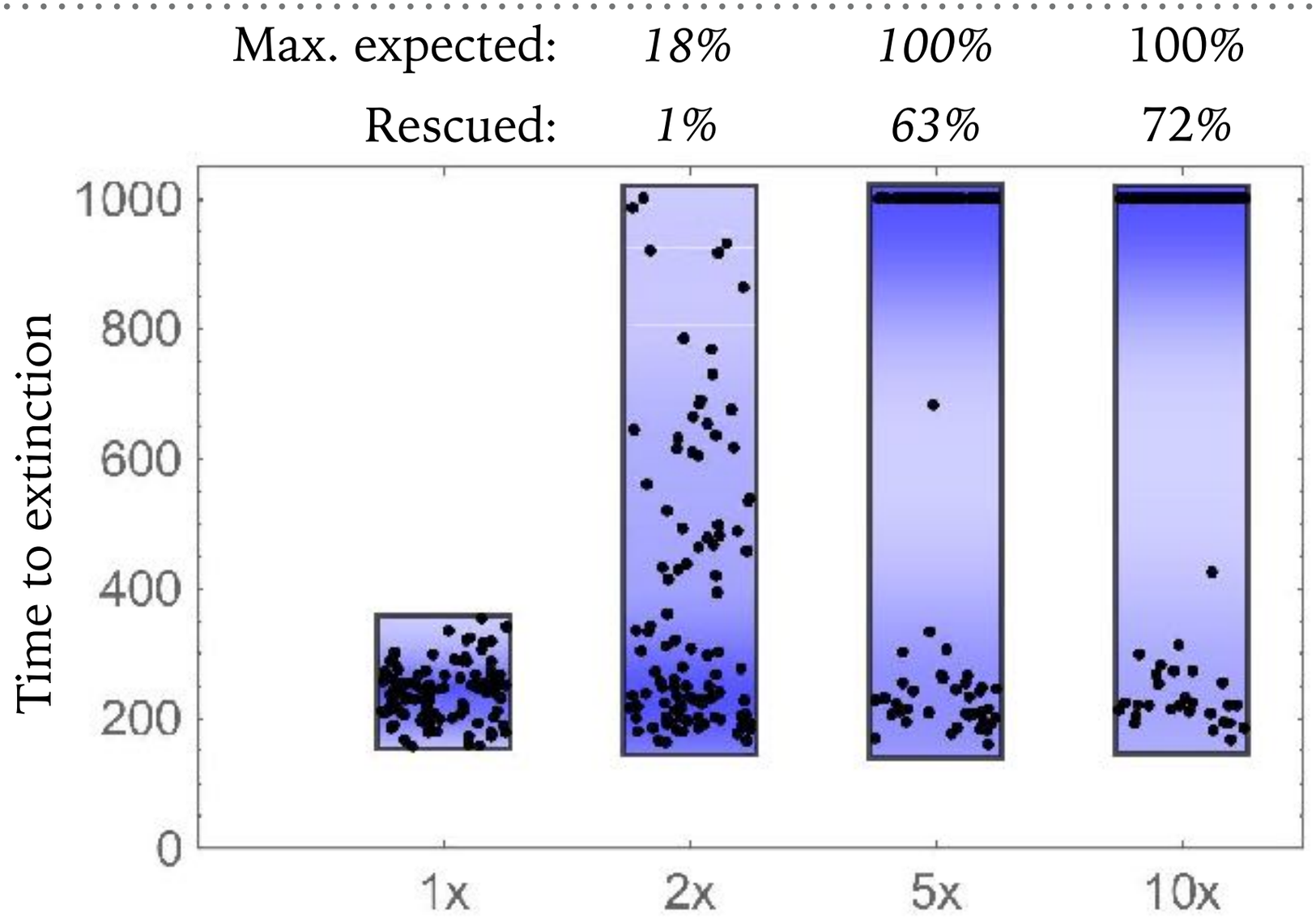
EXTINCTION TIMES WITH BENEFICIAL MUTATIONS



Carrying capacity: 250
Selection coefficient: +/-0.05
Genome length: 1000

- Many beneficials necessary to allow for significantly prolonged time to extinction.
- Clonal interference impedes efficient spread of multiple beneficials and increases variance in extinction times.

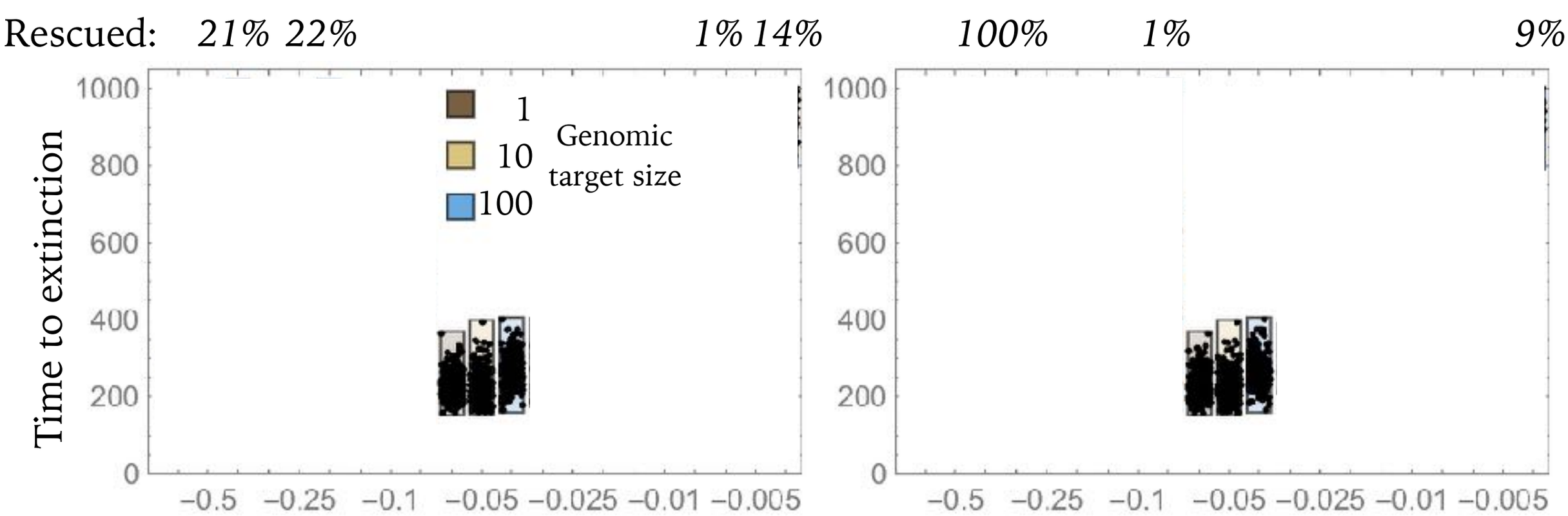
INVASION OF A MUTATION RATE MODIFIER



Carrying capacity: 250
Selection coefficient: +/-0.05
Genome length: 1000

- Mutation rate modifier of sufficient strength readily invades and rescues the population with high probability.

INVASION OF A DFE MODIFIER



- Both types of modifiers can invade; “chaperone” modifier invades easily but rarely rescues; “negative" modifier only invades under specific conditions but then rescues reliably.

CONCLUSIONS

- Extinction process is rather deterministic over a large range of the parameter space.
- Many available beneficials are needed to prolong the extinction time (e.g., to successful transmission of the virus).
- If available, mutation rate modifiers readily invade and make the population resistant to mutagenic treatment.
- DFE modifiers in both directions can invade and make the virus tolerant to high mutation rates. This is possibly the most dangerous adaptation mechanism, because it could modify virus evolution also in absence of the drug.

ACKNOWLEDGEMENTS

TEAM


Alex Blanckaert


Inês Fragata


Hermina Ghenu


Marco Louro


Mark Schmitz

COLLABORATORS

Maria João Amorim, IGC
Thomas Bataillon, Aarhus University
Daniel Bolon, Pam Cote, Ryan Hietpas,
UMass Medical School
Roger Butlin, University of Sheffield
Mónica Bettencourt-Dias, IGC
Isabel Gordo, IGC
Jeffrey Jensen, Arizona State University
Rees Kassen, University of Ottawa
Jonna Kulmuni, Helsinki University
David Liberles, Temple University
Sebastian Matuszewski, EPFL
Vitor Sousa, University of Lisbon
Alex Wong, Carleton University

FUNDING

FUNDAÇÃO CALOUSTE GULBENKIAN

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

jpiamr
JOINT PROGRAMMING INITIATIVE ON ANTIMICROBIAL RESISTANCE