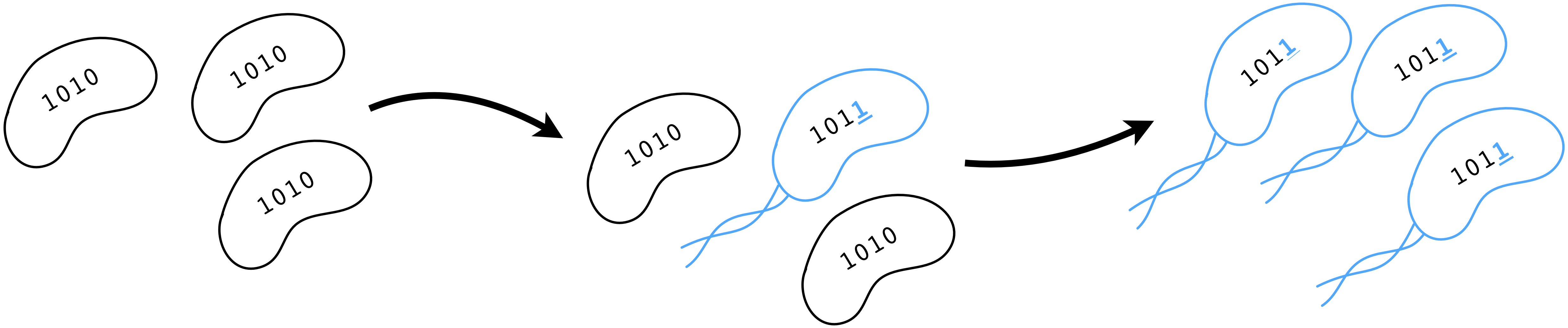
Evolutionary dynamics & ecological diversification in rapidly adapting populations

Benjamin Good

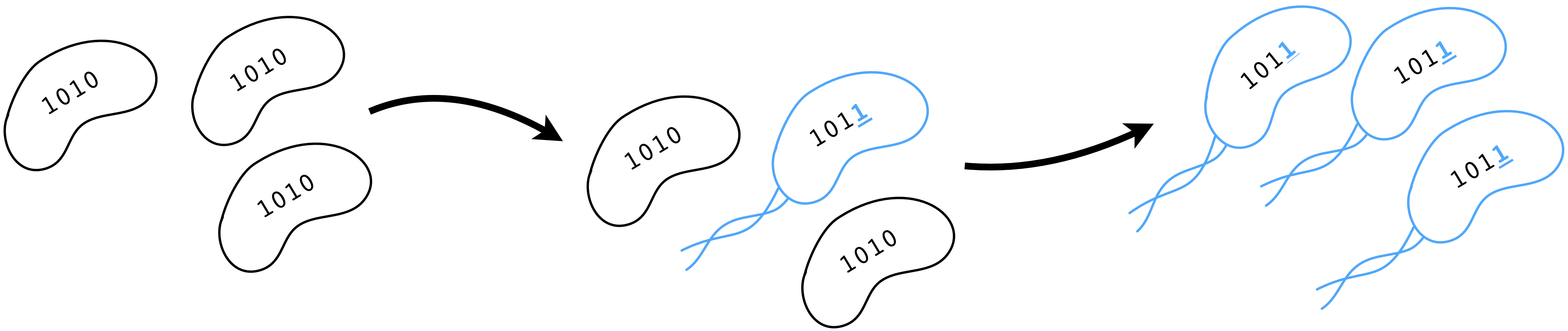
Miller Fellow, UC Berkeley

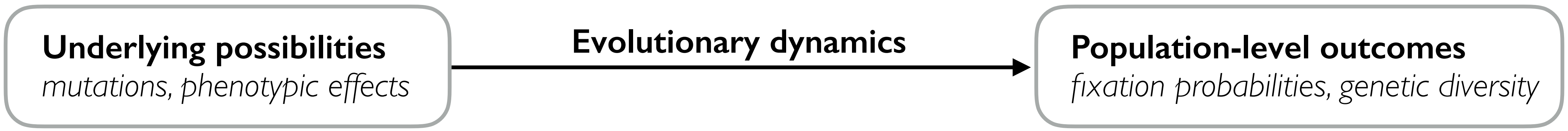
UCSB QBio Seminar, 2/15/18

Evolution as a stochastic dynamical process



Evolution as a stochastic dynamical process



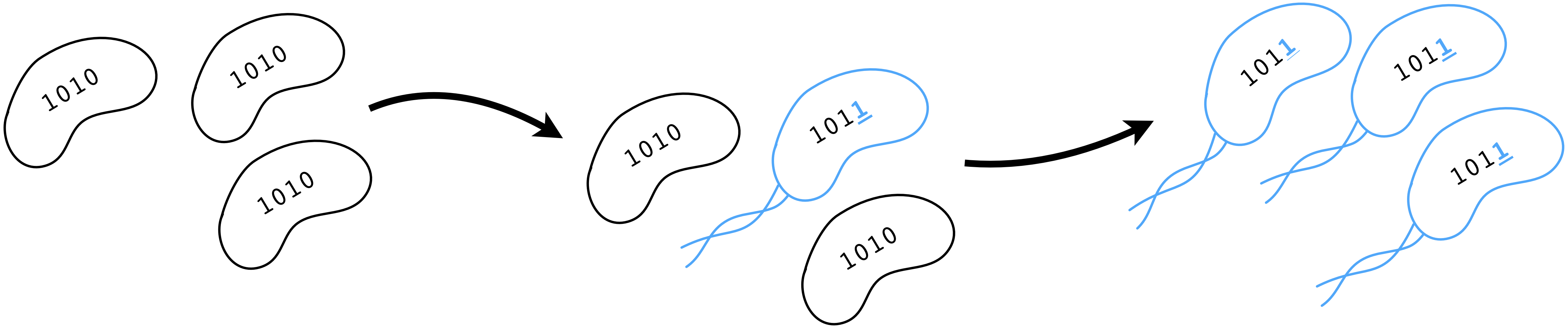


Underlying possibilities
mutations, phenotypic effects

Evolutionary dynamics

Population-level outcomes
fixation probabilities, genetic diversity

Evolution as a stochastic dynamical process

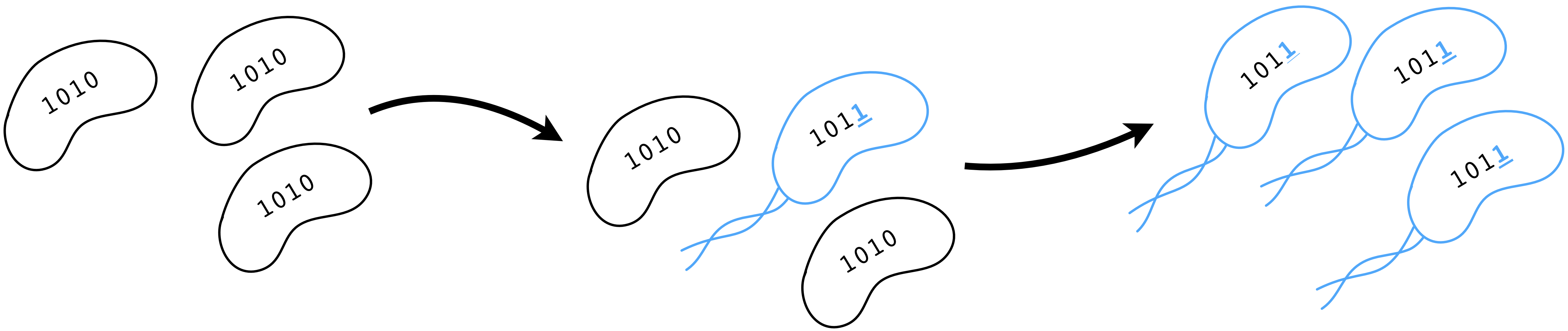


Underlying possibilities
mutations, phenotypic effects

Evolutionary dynamics
(general principles?)

Population-level outcomes
fixation probabilities, genetic diversity

Evolution as a stochastic dynamical process



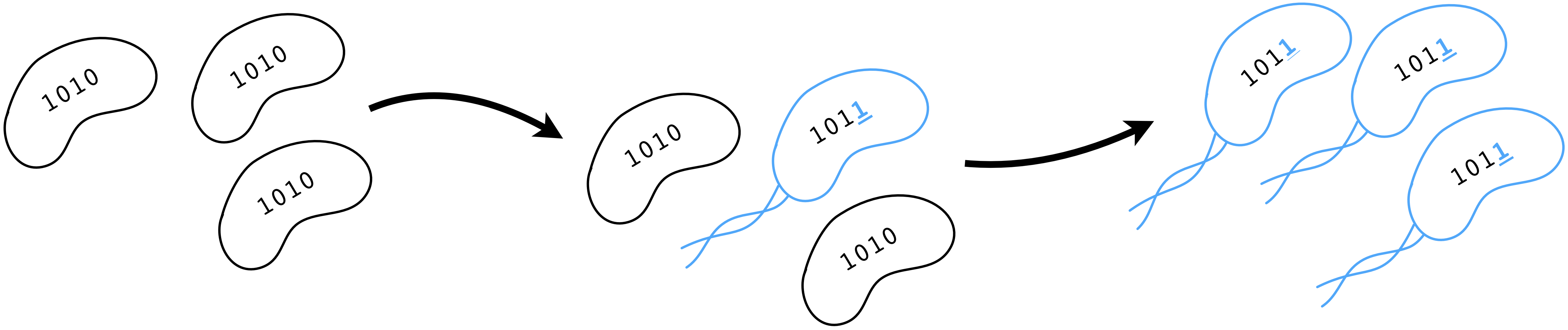
Underlying possibilities
mutations, phenotypic effects

Evolutionary dynamics
(general principles?)

Population-level outcomes
fixation probabilities, genetic diversity

Can we understand this process in a quantitative way, to make predictions ?

Evolution as a stochastic dynamical process



Underlying possibilities
mutations, phenotypic effects

Evolutionary dynamics
(general principles?)

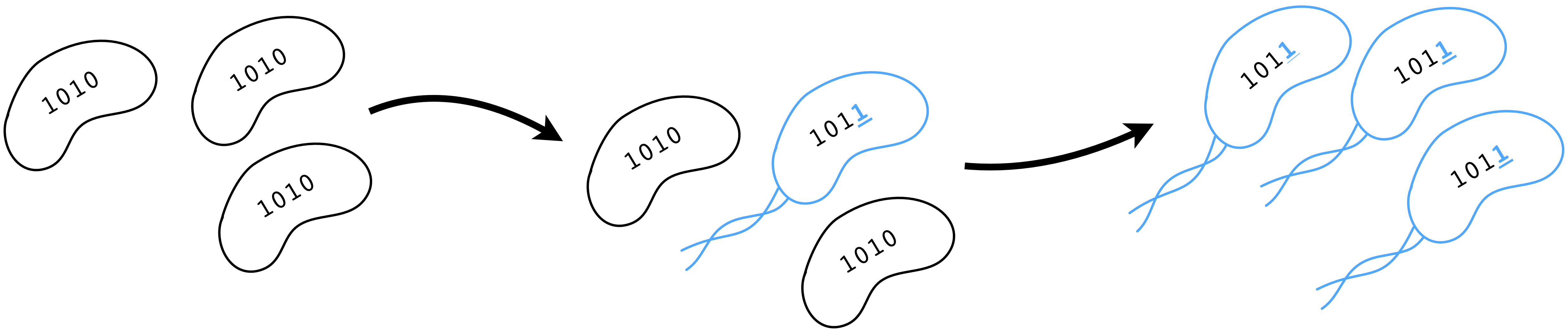
Population-level outcomes
fixation probabilities, genetic diversity

Can we understand this process in a quantitative way, to make ~~predictions~~ ?

short-term predictions

for microbes, cancer cells,
immune repertoires, etc.

Evolution as a stochastic dynamical process



Underlying possibilities
mutations, phenotypic effects
(probe w/ gene editing)

Evolutionary dynamics →
(general principles?)

Population-level outcomes
fixation probabilities, genetic diversity
(observe w/ sequencing)

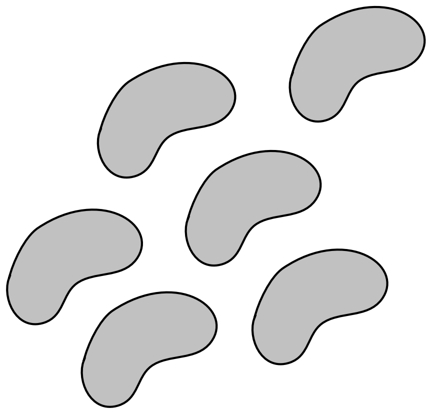
Can we understand this process in a quantitative way, to make ~~predictions~~ ?

short-term predictions

for microbes, cancer cells, immune repertoires, etc.

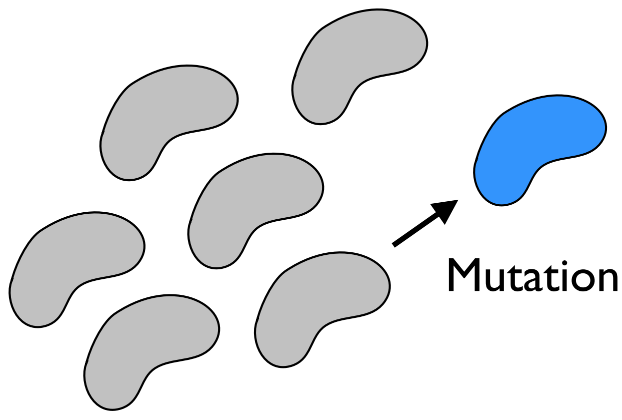
The simplest model of evolution

Population, size N



Generation 0

The simplest model of evolution

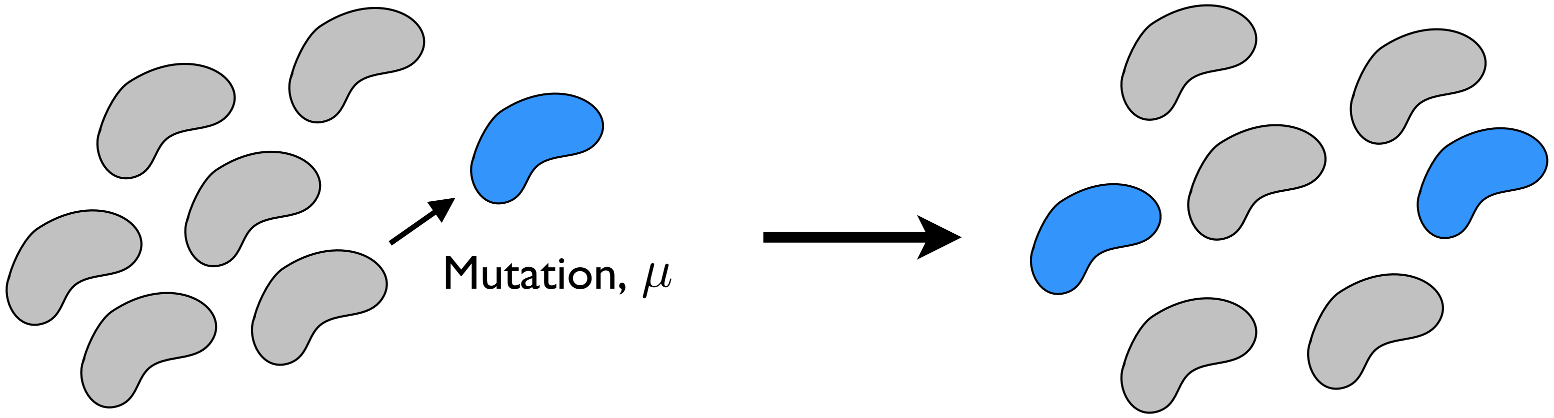


Population, size N

Mutation, μ

Generation 0

The simplest model of evolution



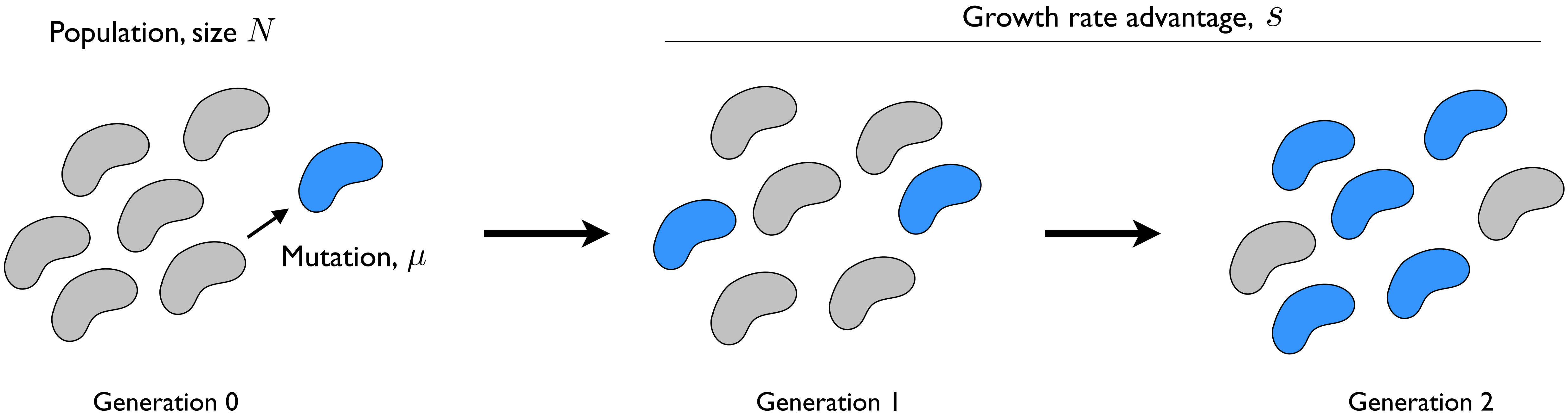
Population, size N

Mutation, μ

Generation 0

Generation 1

The simplest model of evolution



Population, size N

Growth rate advantage, s

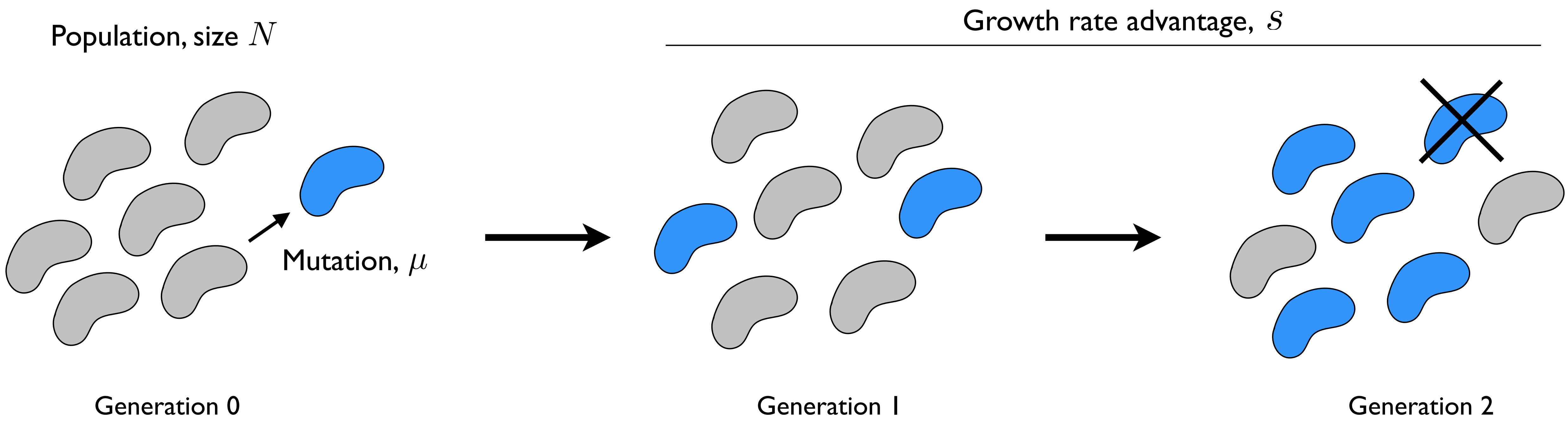
Mutation, μ

Generation 0

Generation 1

Generation 2

The simplest model of evolution



Population, size N

Growth rate advantage, s

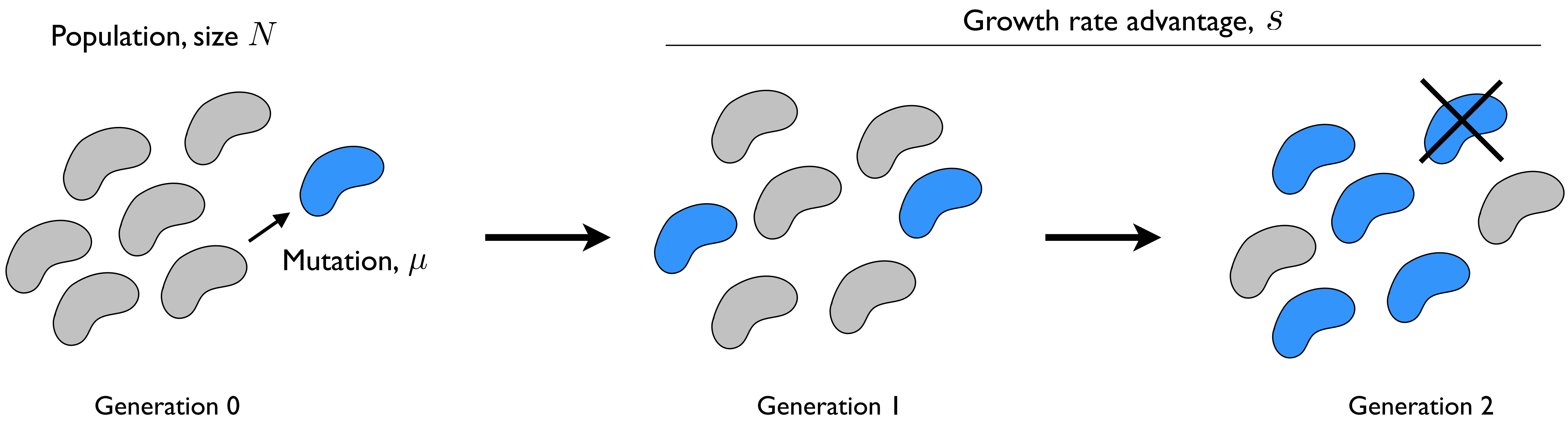
Mutation, μ

Generation 0

Generation 1

Generation 2

The simplest model of evolution



Population, size N

Growth rate advantage, s

Mutation, μ

Generation 0

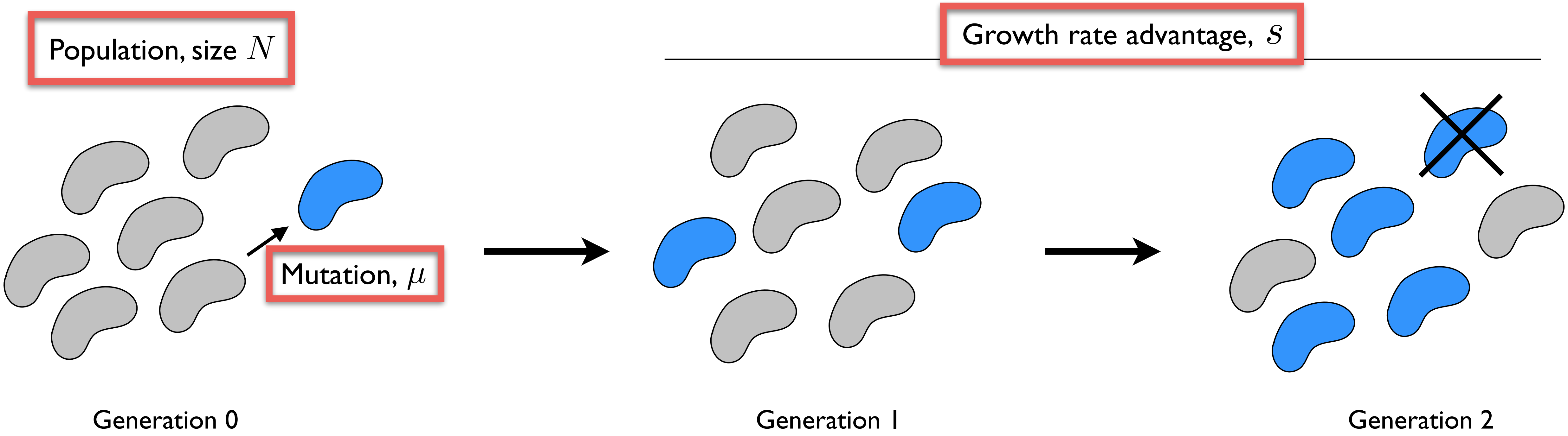
Generation 1

Generation 2

Fraction mutant cells, f, grows as:

$$\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\substack{\text{genetic drift}\\ \textit{(birth-death noise)}}}$$

The simplest model of evolution

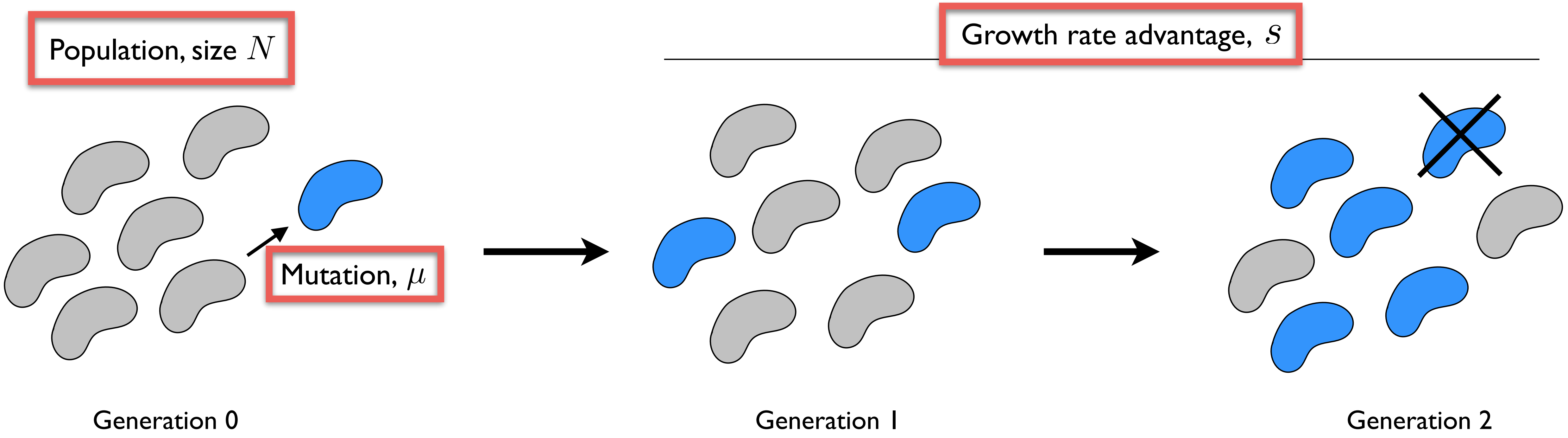


Fraction mutant cells, f, grows as:

$$\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{s f(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\substack{\text{genetic drift}\\ \textit{(birth-death noise)}}}$$

Measure

The simplest model of evolution



Population, size N

Mutation, μ

Growth rate advantage, s

Generation 0

Generation 1

Generation 2

Fraction mutant cells, f, grows as:

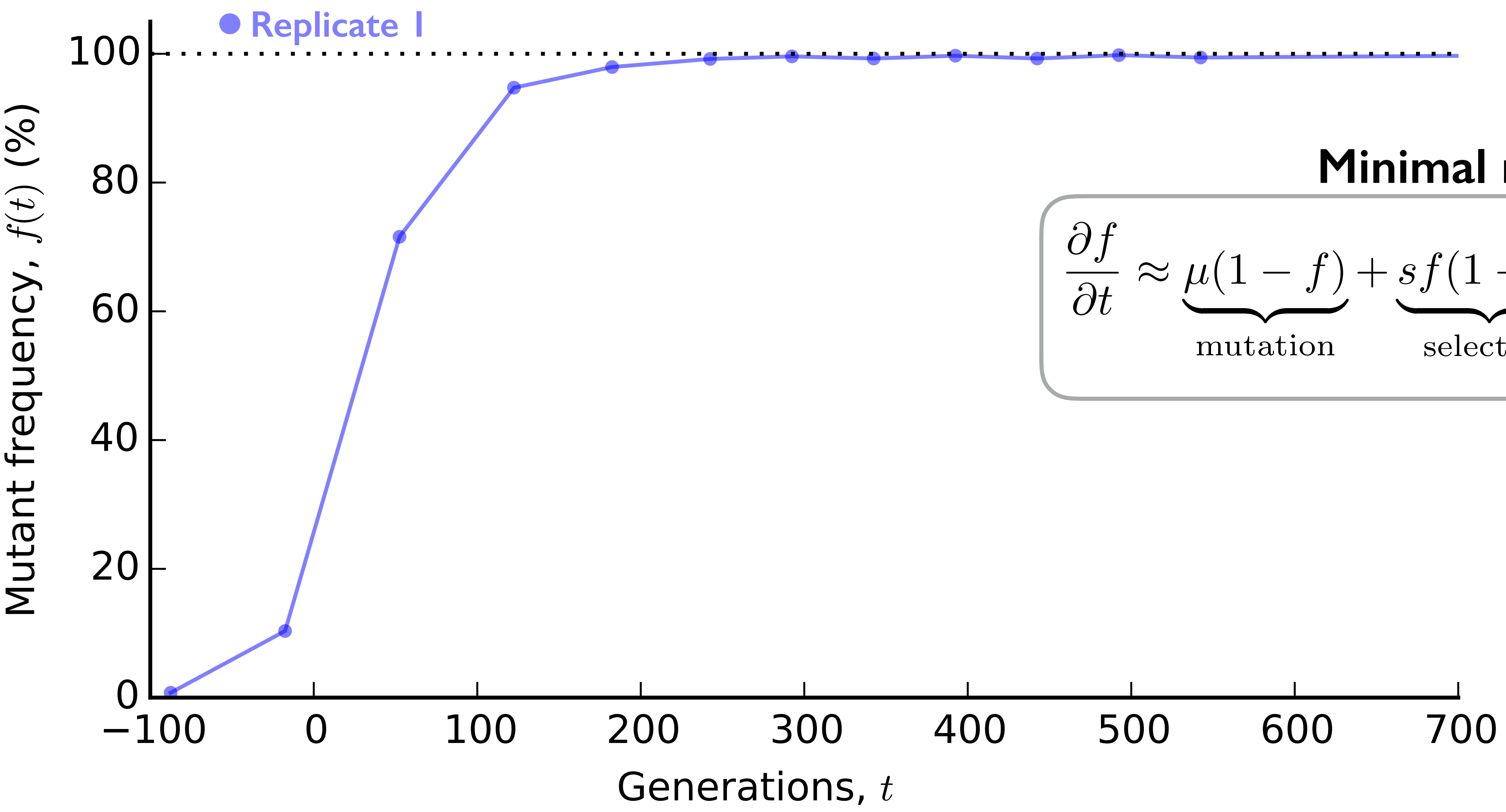
$$\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{s f(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\text{genetic drift}}$$

(birth-death noise)

Measure → ***Predict evolution!***

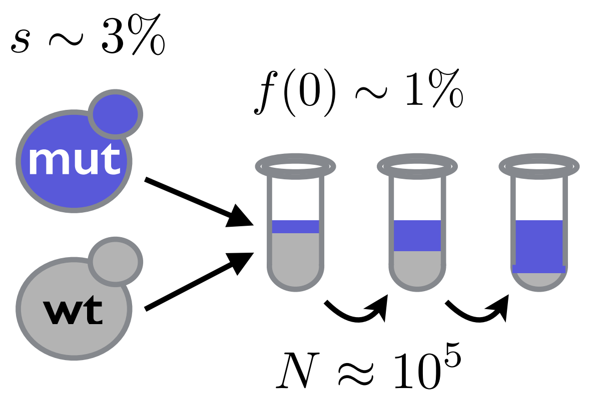
e.g. how long until $f = 100\%$ ("fixation") ?

A laboratory “test” of the simple model



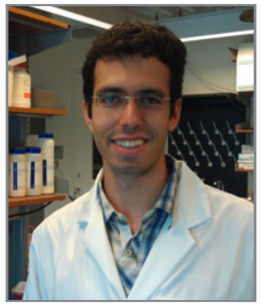
Minimal model

$$\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\text{genetic drift}}$$



+

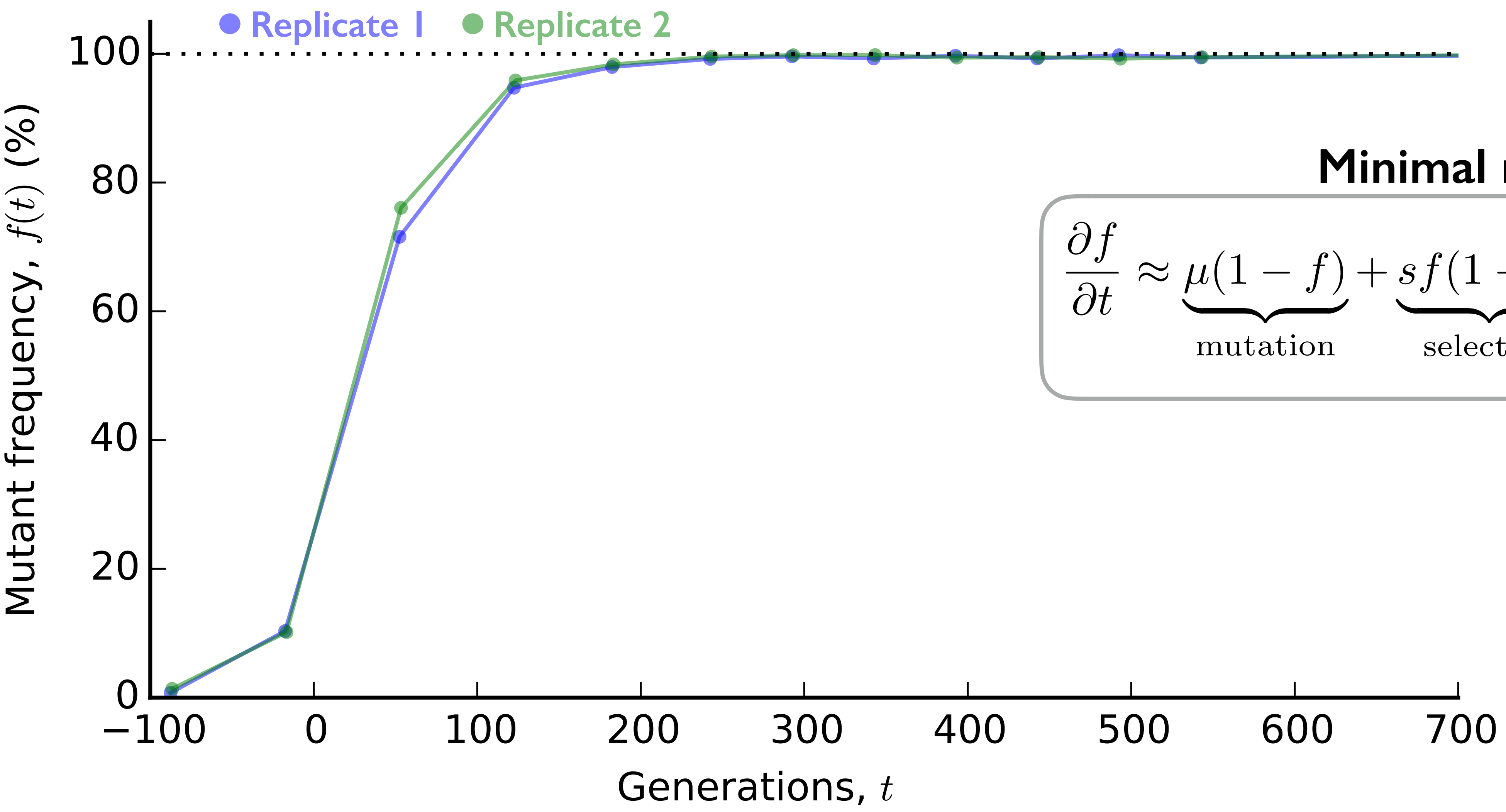
Measure frequency
w/ flow cytometry



Evgeni Frenkel

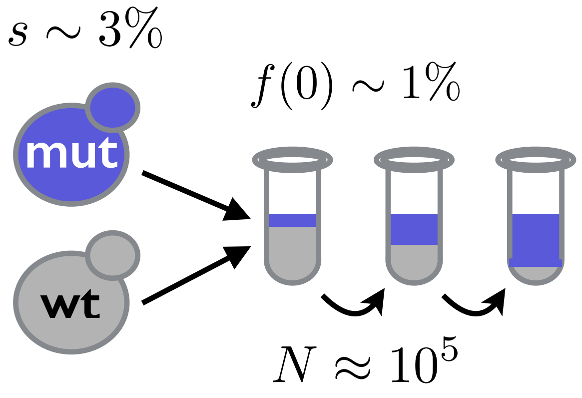
Data from Frenkel, Good, and Desai (2014)

A laboratory “test” of the simple model



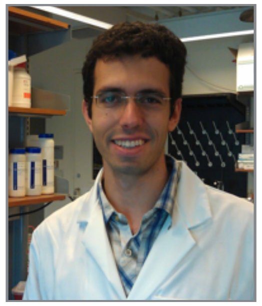
Minimal model

$$\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\eta(t)}_{\text{genetic drift}}$$



+

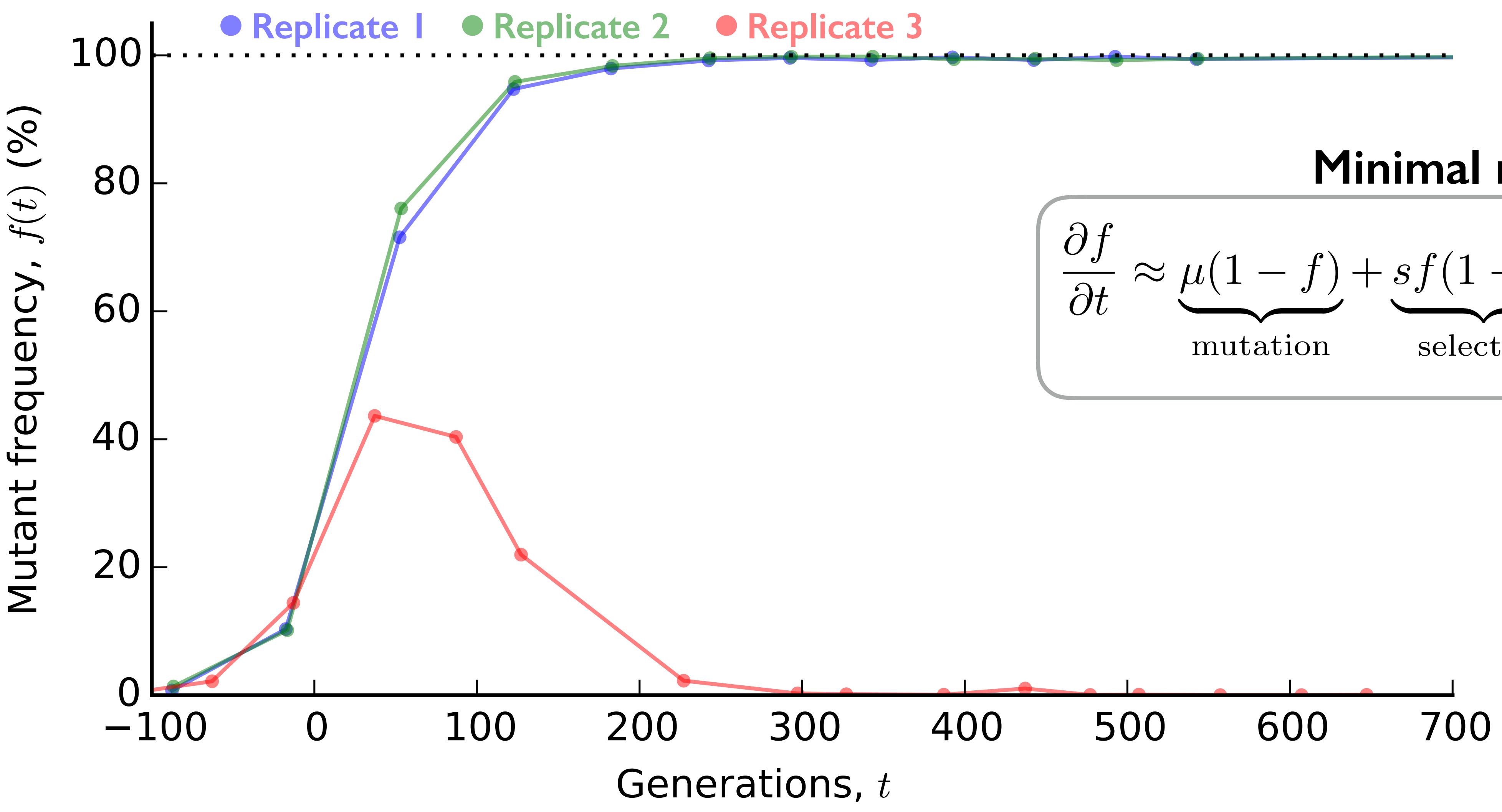
Measure frequency
w/ flow cytometry



Evgeni Frenkel

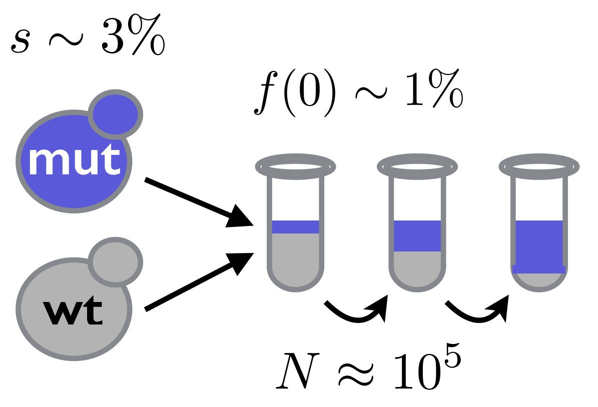
Data from Frenkel, Good, and Desai (2014)

A laboratory “test” of the simple model



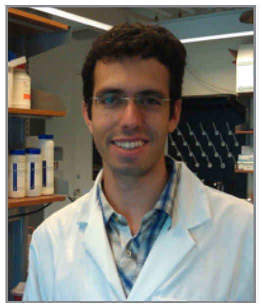
Minimal model

$$\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\text{genetic drift}}$$



+

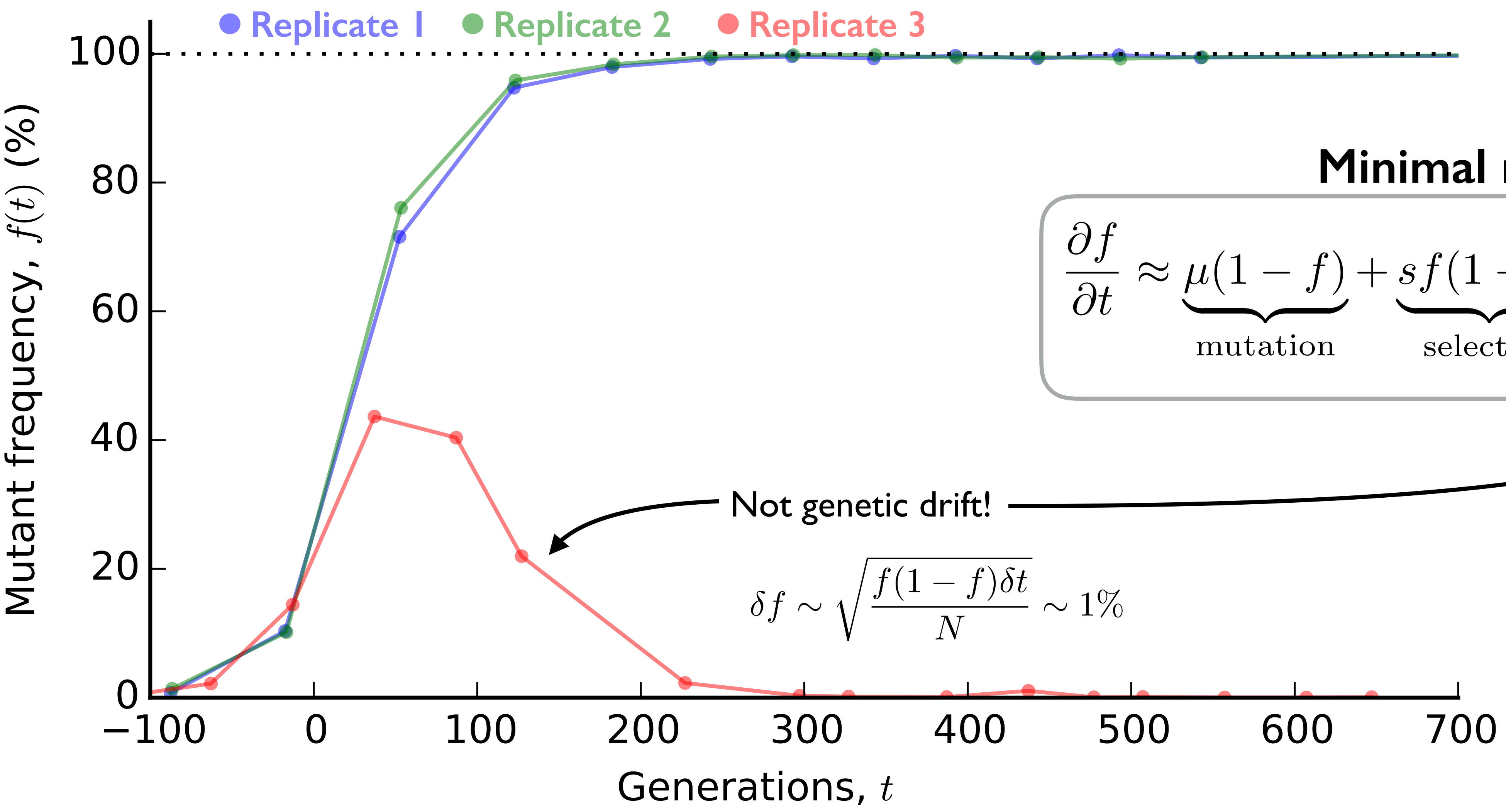
Measure frequency
w/ flow cytometry



Evgeni Frenkel

Data from Frenkel, Good, and Desai (2014)

A laboratory "test" of the simple model

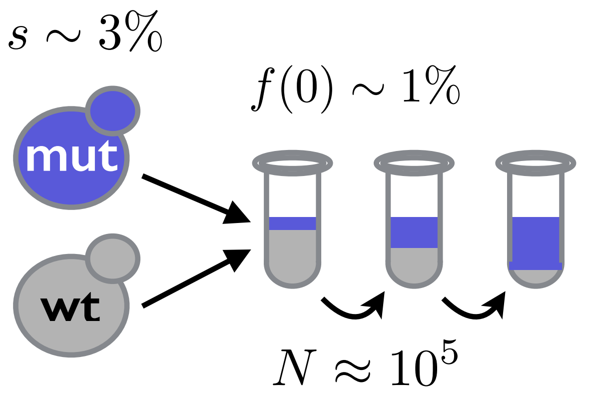


Not genetic drift!

$$\delta f \sim \sqrt{\frac{f(1-f)\delta t}{N}} \sim 1\%$$

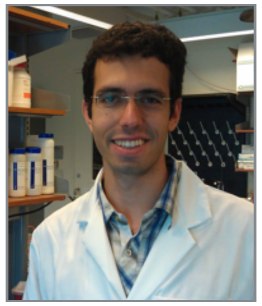
Minimal model

$$\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\eta(t)}_{\text{genetic drift}}$$



+

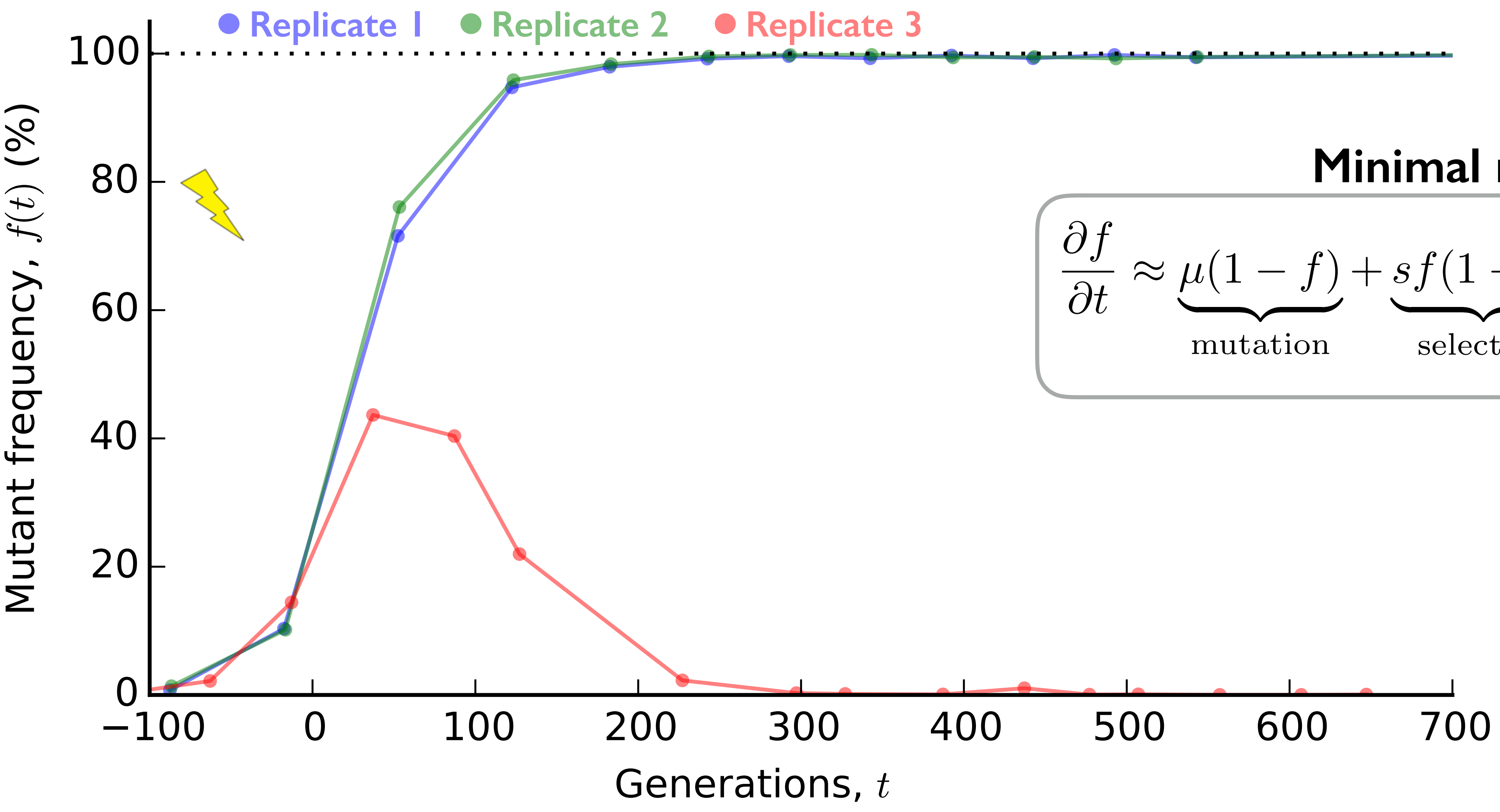
Measure frequency
w/ flow cytometry



Evgeni Frenkel

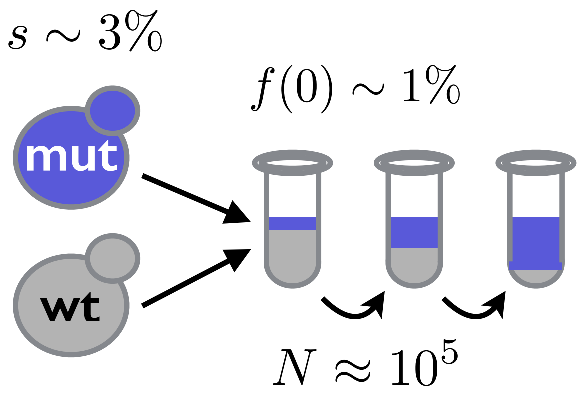
Data from Frenkel, Good, and Desai (2014)

A laboratory “test” of the simple model



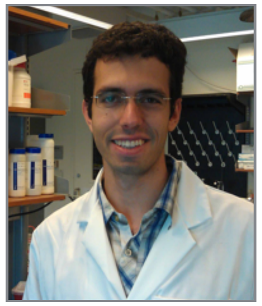
Minimal model

$$\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\text{genetic drift}}$$



+

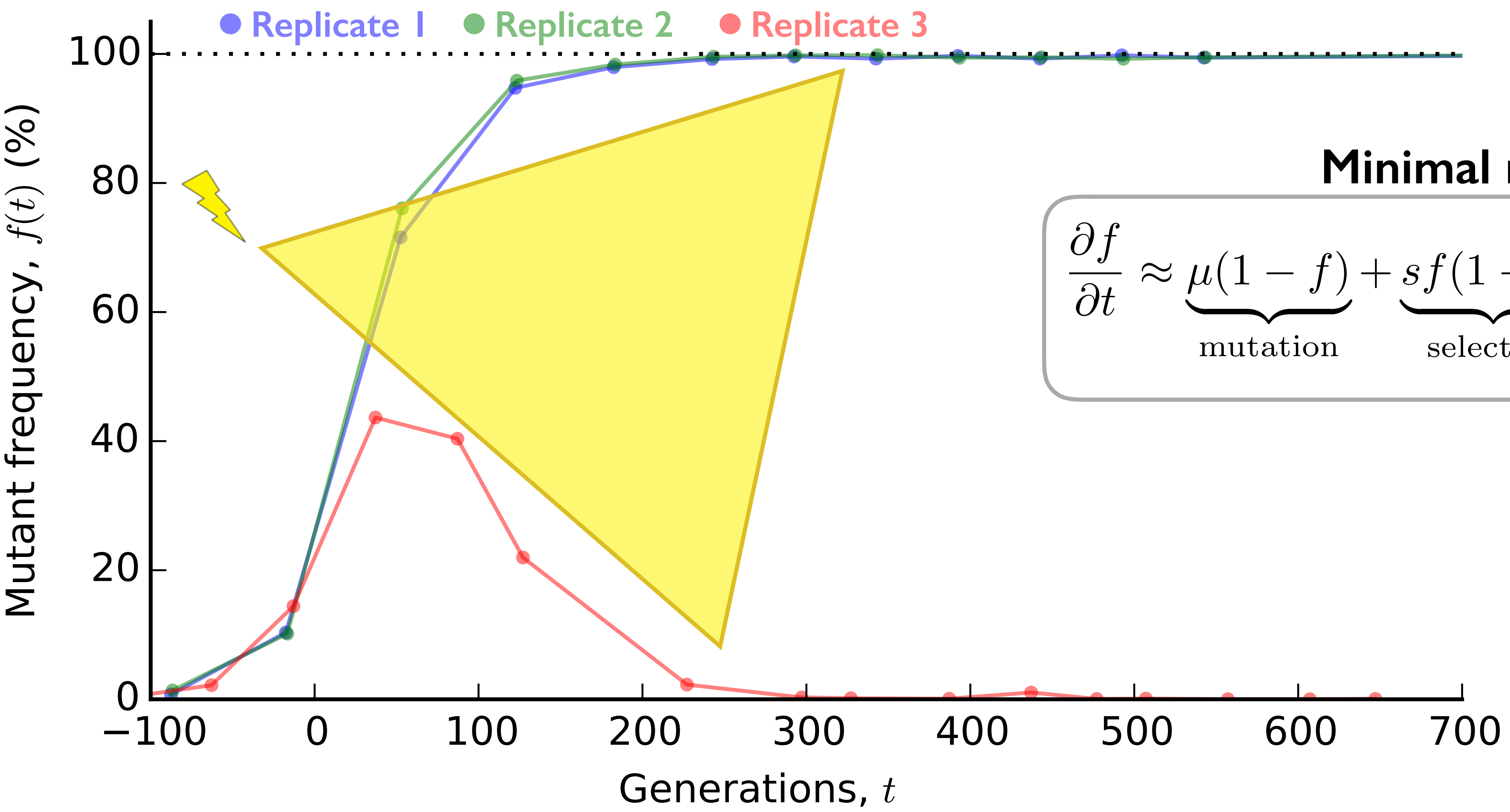
Measure frequency
w/ flow cytometry



Evgeni Frenkel

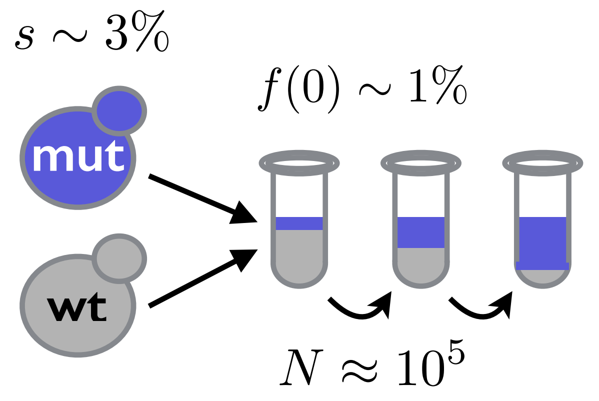
Data from Frenkel, Good, and Desai (2014)

A laboratory "test" of the simple model



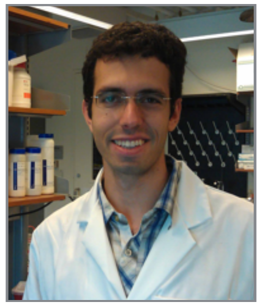
Minimal model

$$\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\text{genetic drift}}$$



+

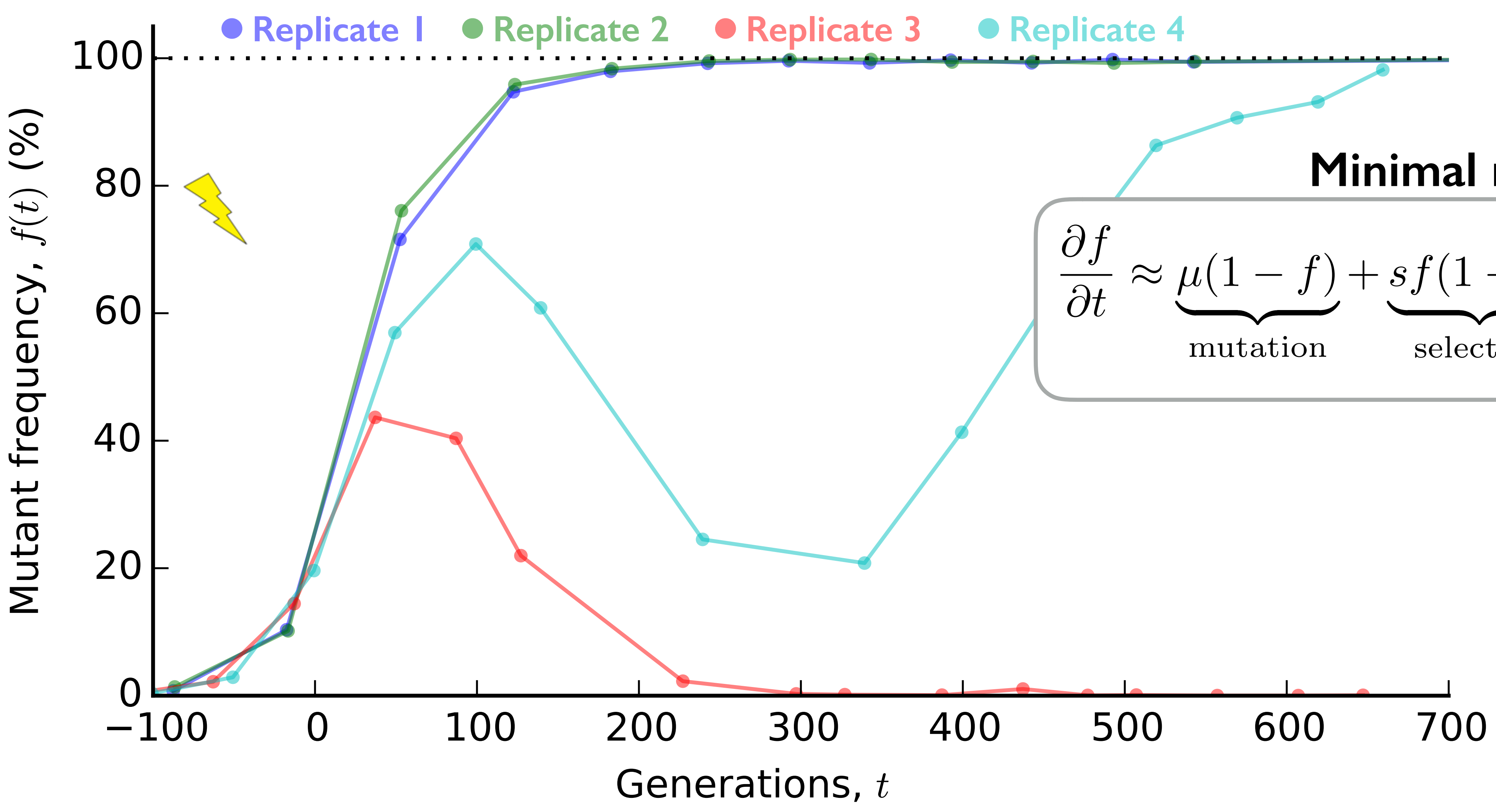
Measure frequency
w/ flow cytometry



Evgeni Frenkel

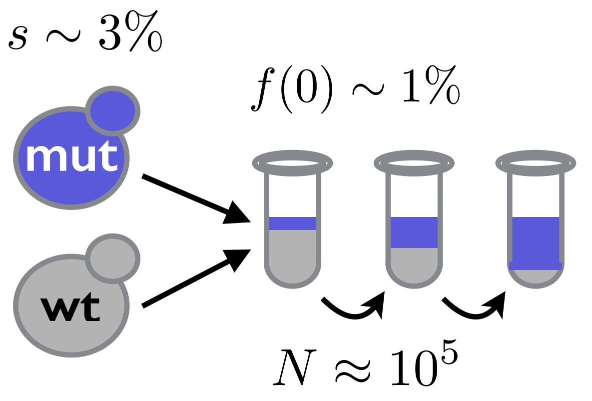
Data from Frenkel, Good, and Desai (2014)

A laboratory “test” of the simple model



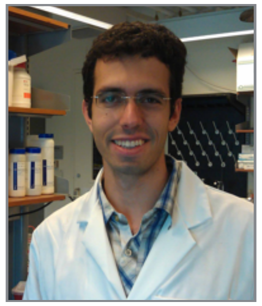
Minimal model

$$\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\text{genetic drift}}$$



+

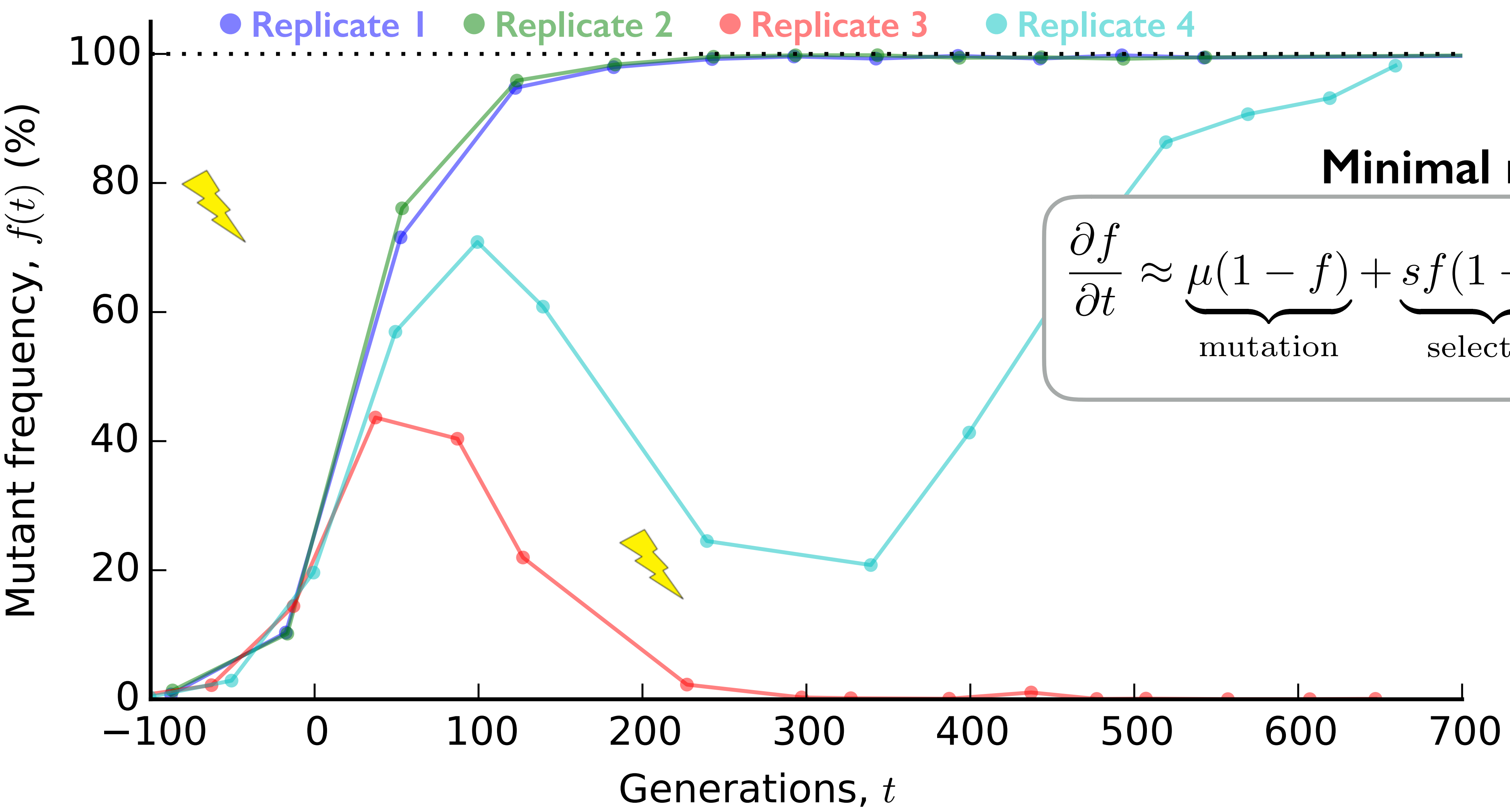
Measure frequency
w/ flow cytometry



Evgeni Frenkel

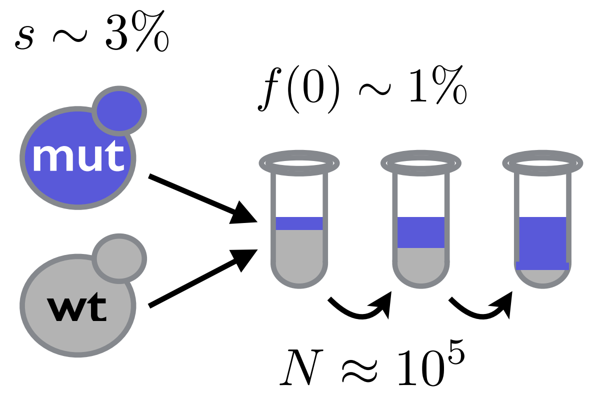
Data from Frenkel, Good, and Desai (2014)

A laboratory "test" of the simple model



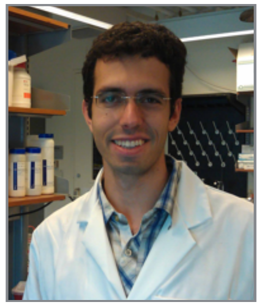
Minimal model

$$\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\text{genetic drift}}$$



+

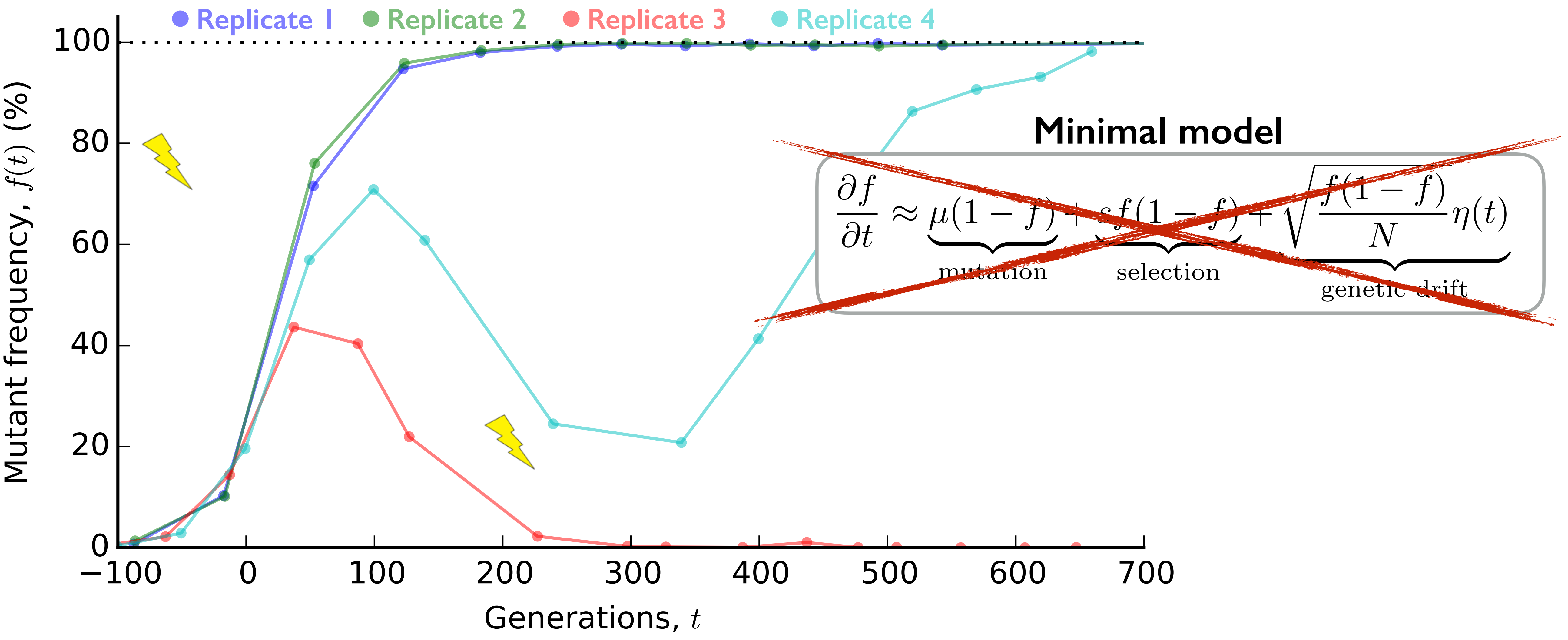
Measure frequency
w/ flow cytometry



Evgeni Frenkel

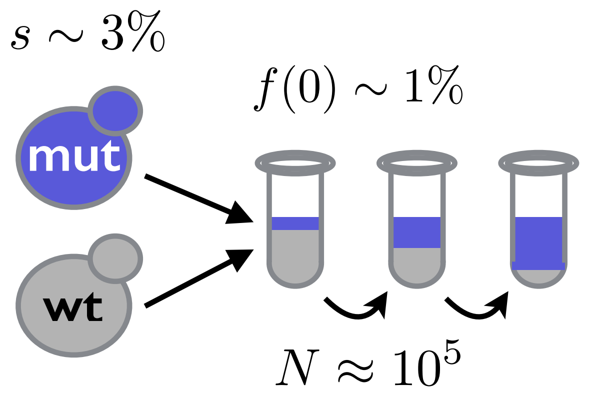
Data from Frenkel, Good, and Desai (2014)

A laboratory "test" of the simple model



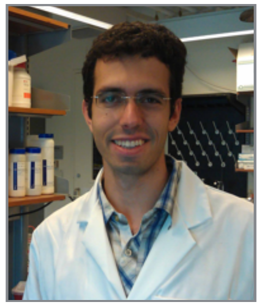
Minimal model

$$\xcancel{\frac{\partial f}{\partial t} \approx \underbrace{\mu(1-f)}_{\text{mutation}} + \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\eta(t)}_{\text{genetic drift}}}$$



+

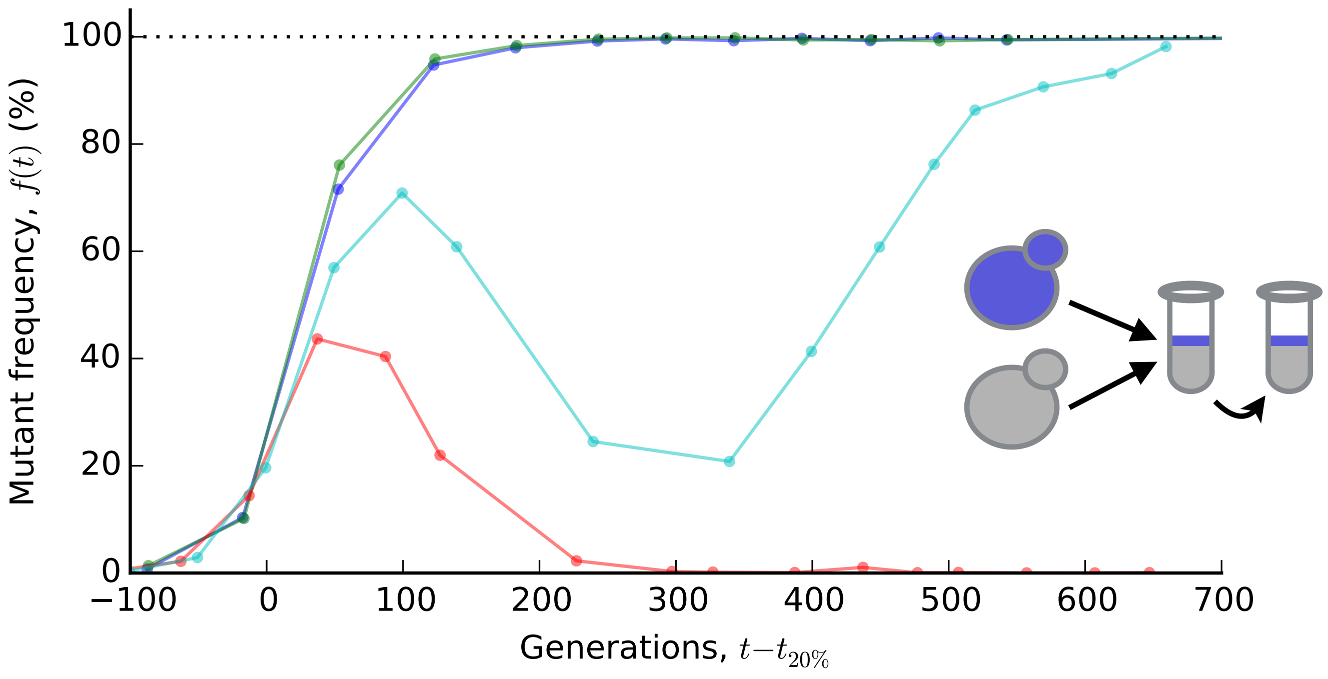
Measure frequency
w/ flow cytometry



Evgeni Frenkel

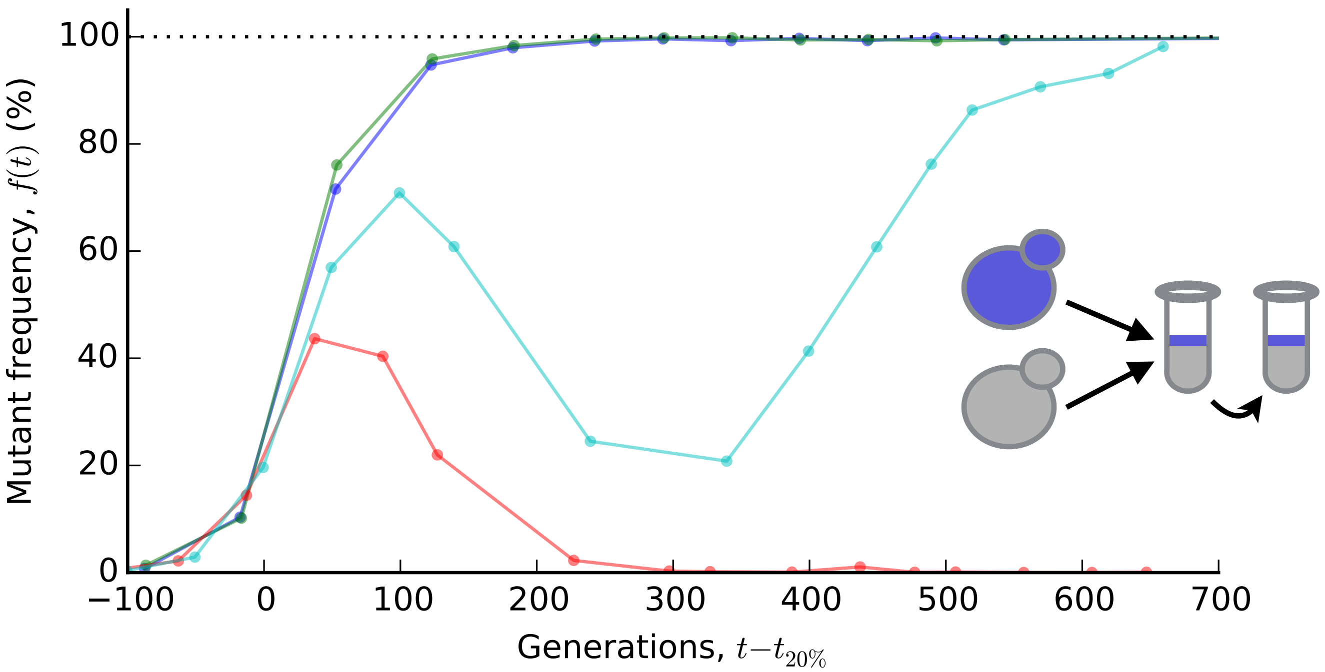
Data from Frenkel, Good, and Desai (2014)

Key observations from laboratory evolution experiments



1. *Large microbial populations can adapt* ***extremely*** *rapidly*

Key observations from laboratory evolution experiments

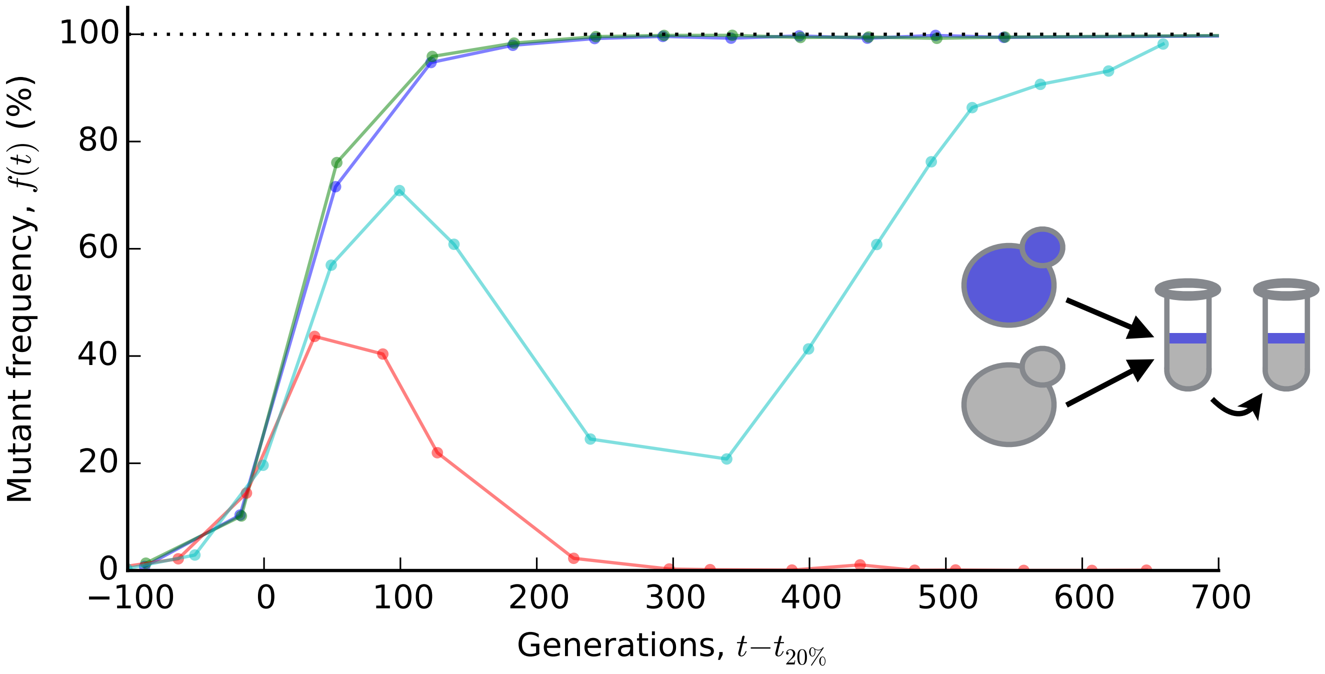


1. *Large microbial populations can adapt* ***extremely*** *rapidly*

↓

2. *Must account for* ***combinations*** *of mutations on same genome*

Key observations from laboratory evolution experiments



1. *Large microbial populations can adapt* ***extremely*** *rapidly*

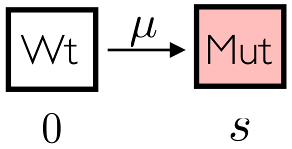
↓

2. *Must account for* ***combinations*** *of mutations on same genome*

"genetic linkage"

Why does genetic linkage make this a hard problem?

Before:



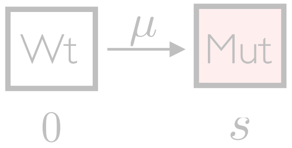
(2 states)

Frequency of ***mutant***:

$$\frac{\partial f}{\partial t} = sf(1-f) + \ldots$$

Why does genetic linkage make this a hard problem?

Before:

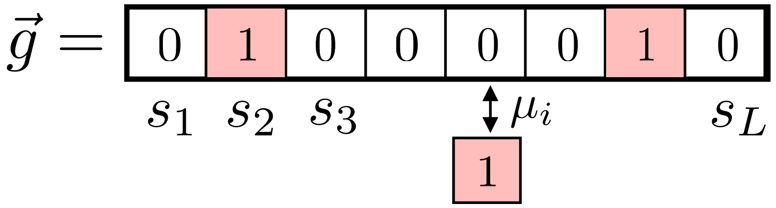


(2 states)

Frequency of ***mutant***:

$$\frac{\partial f}{\partial t} = sf(1-f) + \ldots$$

Linkage: *model full* genome (* 2^L *states)*



Frequencies of ***genotypes***:

$$\frac{\partial f(\vec{g})}{\partial t} = [X(\vec{g}) - \sum_{\vec{g}'} X(\vec{g}')f(\vec{g}')]f(\vec{g}) + \ldots$$

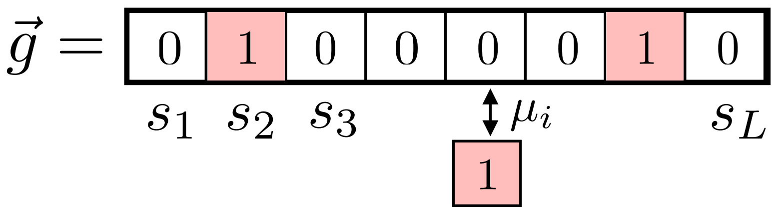
Why does genetic linkage make this a hard problem?

Before: $\underset{0}{\boxed{\text{Wt}}} \xrightarrow{\mu} \underset{s}{\boxed{\text{Mut}}}$ *(2 states)*

Frequency of ***mutant:***

$$\frac{\partial f}{\partial t} = sf(1-f) + \ldots$$

Linkage: *model full* genome (* 2^L *states)*



Frequencies of ***genotypes:***

$$\frac{\partial f(\vec{g})}{\partial t} = [X(\vec{g}) - \sum_{\vec{g}'} X(\vec{g}')f(\vec{g}')]f(\vec{g}) + \ldots$$

Fitness of ***genotype:***

$$X(\vec{g}) \approx \sum_{i=1}^{L} s_i g_i + \text{epistasis}$$

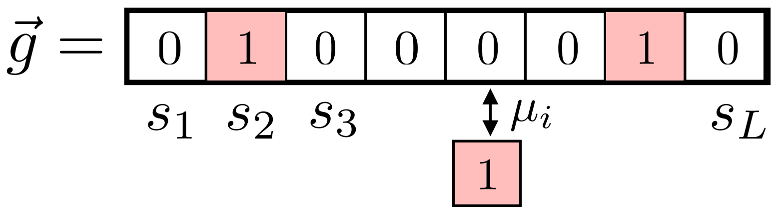
Why does genetic linkage make this a hard problem?

Before: $\underset{0}{\boxed{\text{Wt}}} \xrightarrow{\mu} \underset{s}{\boxed{\text{Mut}}}$ *(2 states)*

Frequency of ***mutant***:

$$\frac{\partial f}{\partial t} = sf(1-f) + \ldots$$

Linkage: *model full* genome (2^L states)*



Frequencies of ***genotypes***:

$$\frac{\partial f(\vec{g})}{\partial t} = [X(\vec{g}) - \sum_{\vec{g}'} X(\vec{g}')f(\vec{g}')]f(\vec{g}) + \ldots$$

Fitness of ***genotype***: $X(\vec{g}) \approx \sum_{i=1}^{L} s_i g_i + \text{epistasis}$

Why does genetic linkage make this a hard problem?

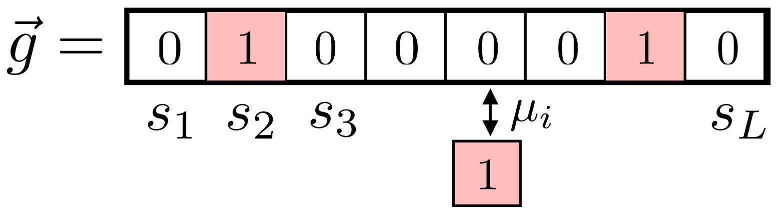
Before: Wt $\xrightarrow{\mu}$ Mut *(2 states)*

$0 \qquad s$

Frequency of ***mutant***:

$$\frac{\partial f}{\partial t} = sf(1-f) + \ldots$$

Linkage: *model full* genome (2^L states)*

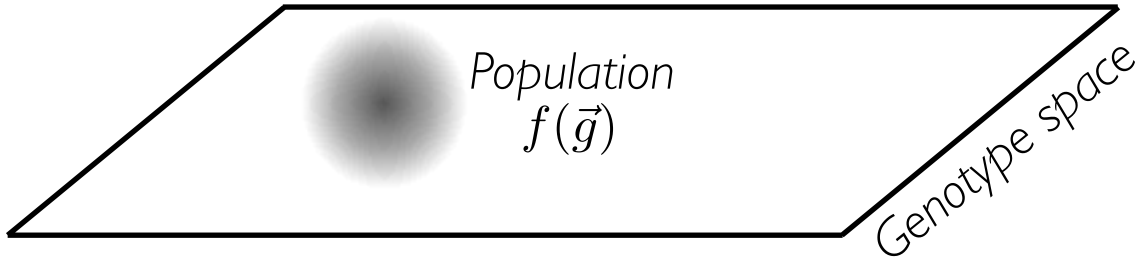


Frequencies of ***genotypes***:

$$\frac{\partial f(\vec{g})}{\partial t} = [X(\vec{g}) - \sum_{\vec{g}'} X(\vec{g}')f(\vec{g}')]f(\vec{g}) + \ldots$$

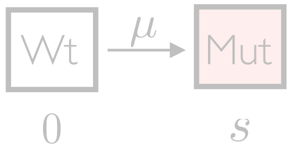
Fitness of ***genotype***: $X(\vec{g}) \approx \sum_{i=1}^{L} s_i g_i + \text{epistasis}$

1. ***Population size (N) is large, but # genotypes vastly larger!***



→ ***always "out-of-equilibrium"***

Why does genetic linkage make this a hard problem?

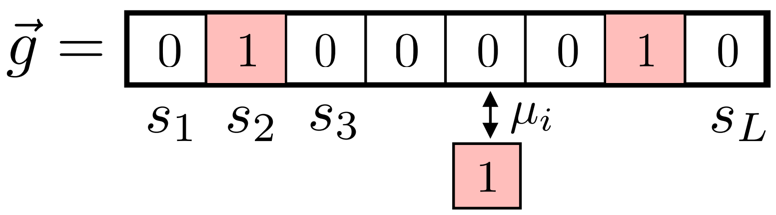
Before: 

(2 states)

Frequency of ***mutant***:

$$\frac{\partial f}{\partial t} = sf(1-f) + \dots$$

Linkage: *model full* genome (2^L states)*

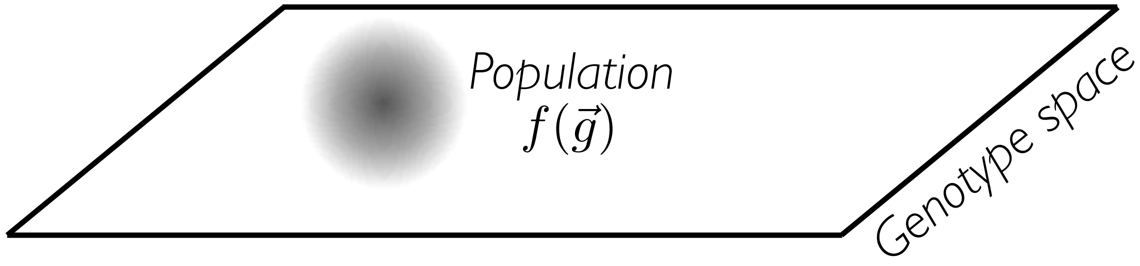


Frequencies of ***genotypes***:

$$\frac{\partial f(\vec{g})}{\partial t} = [X(\vec{g}) - \sum_{\vec{g}'} X(\vec{g}')f(\vec{g}')]f(\vec{g}) + \dots$$

Fitness of ***genotype***: $X(\vec{g}) \approx \sum_{i=1}^{L} s_i g_i + \text{epistasis}$

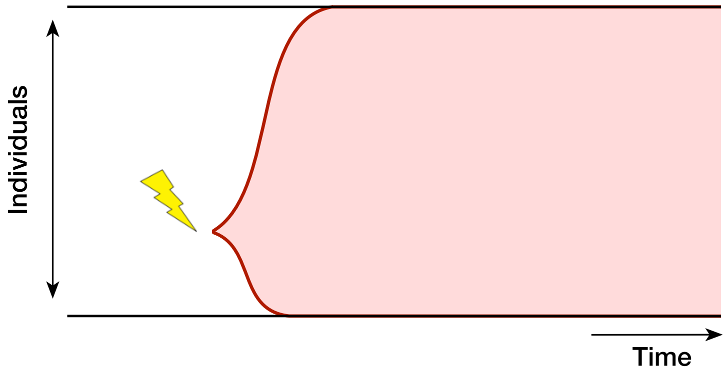
1. Population size (N) is large,
but # genotypes vastly larger!



→ ***always "out-of-equilibrium"***

---

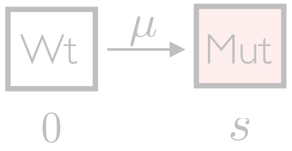
2. Individual mutations still take over!



→ ***noise matters @ pop'n level***

Why does genetic linkage make this a hard problem?

Before:

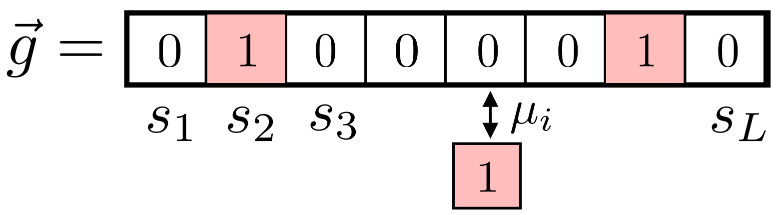


(2 states)

Frequency of ***mutant***:

$$\frac{\partial f}{\partial t} = sf(1-f) + \ldots$$

Linkage: *model full* genome (*2^L *states)*



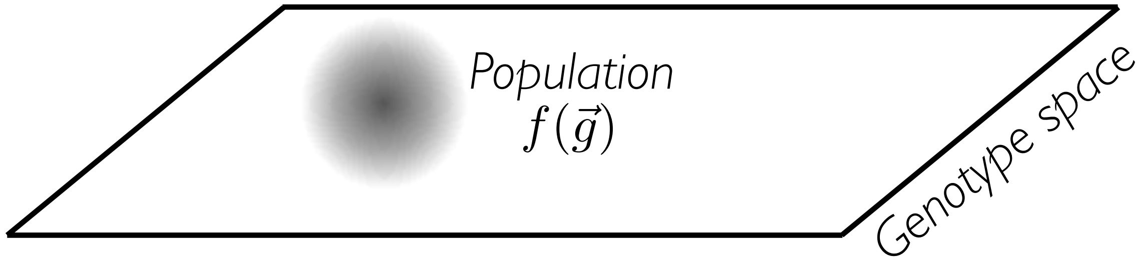
Frequencies of ***genotypes***:

$$\frac{\partial f(\vec{g})}{\partial t} = [X(\vec{g}) - \sum_{\vec{g}'} X(\vec{g}')f(\vec{g}')]f(\vec{g}) + \ldots$$

Fitness of ***genotype***:

$$X(\vec{g}) \approx \sum_{i=1}^{L} s_i g_i + \text{epistasis}$$

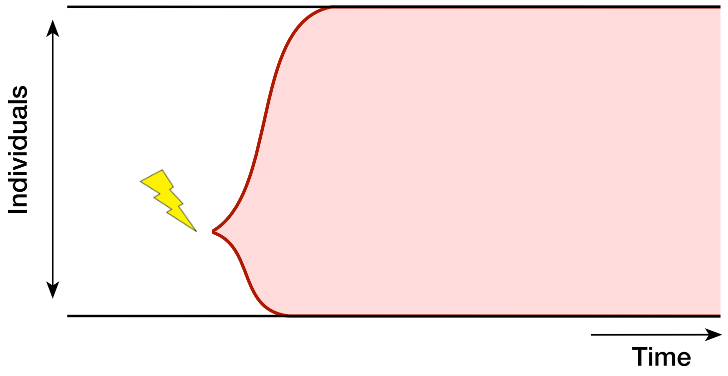
1. Population size (N) is large, but # genotypes vastly larger!



→ ***always "out-of-equilibrium"***

---

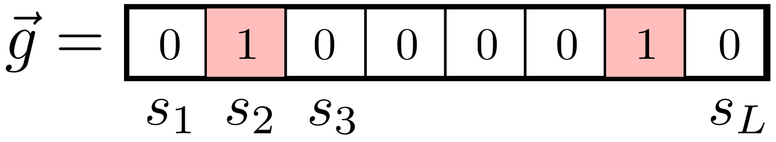
2. Individual mutations still take over!



→ ***noise matters @ pop'n level***

---

3. Explosion of parameters!



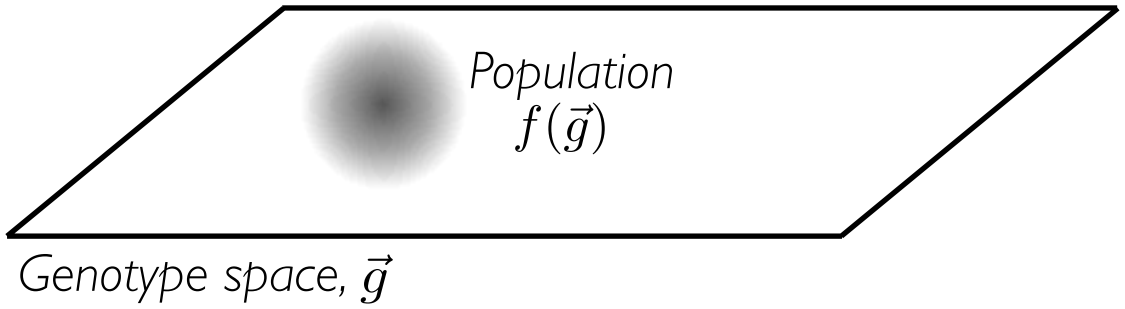
→ ***which combinations matter? how might we measure them?***

How do we think about genetic linkage theoretically?

Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

How do we think about genetic linkage theoretically?

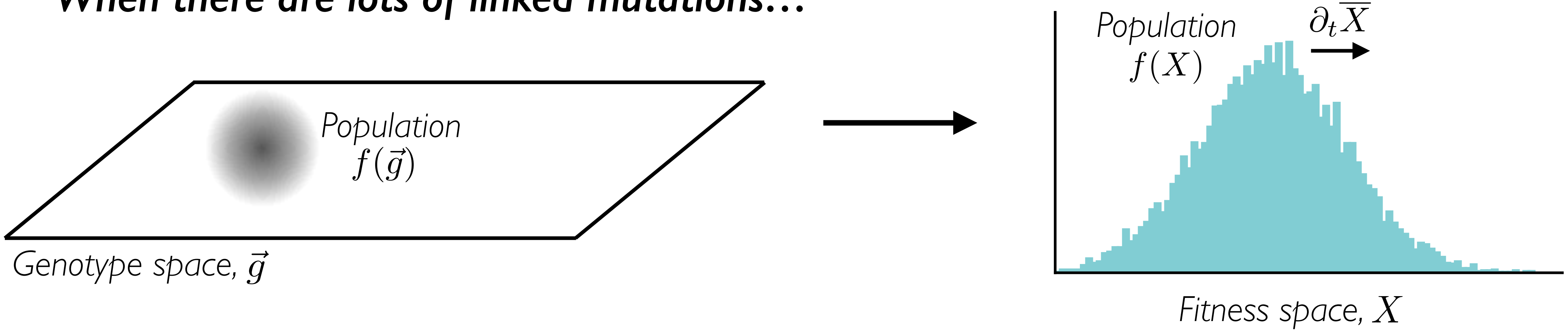
When there are lots of linked mutations…



Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

How do we think about genetic linkage theoretically?

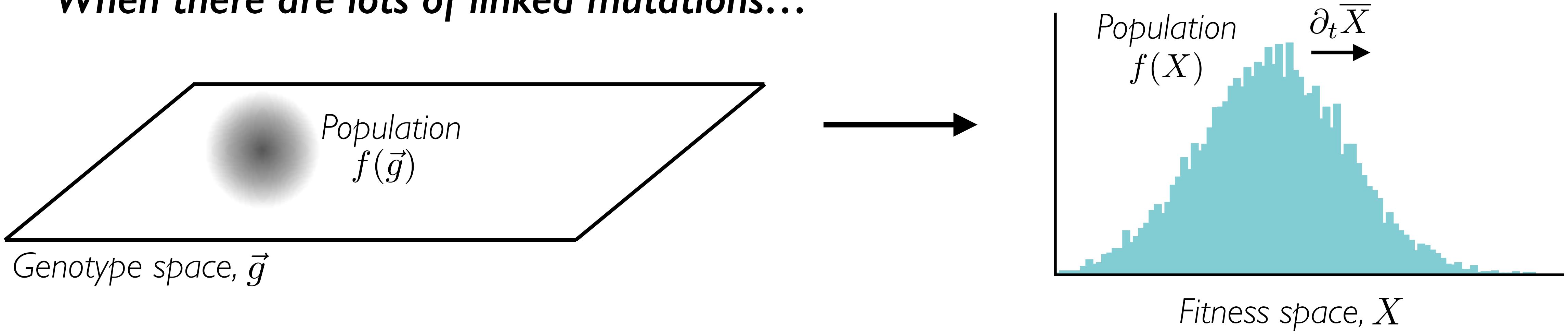
When there are lots of linked mutations…



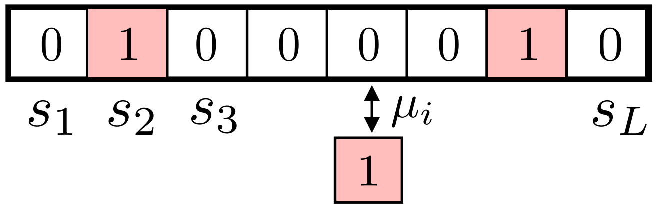
Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

How do we think about genetic linkage theoretically?

When there are lots of linked mutations…



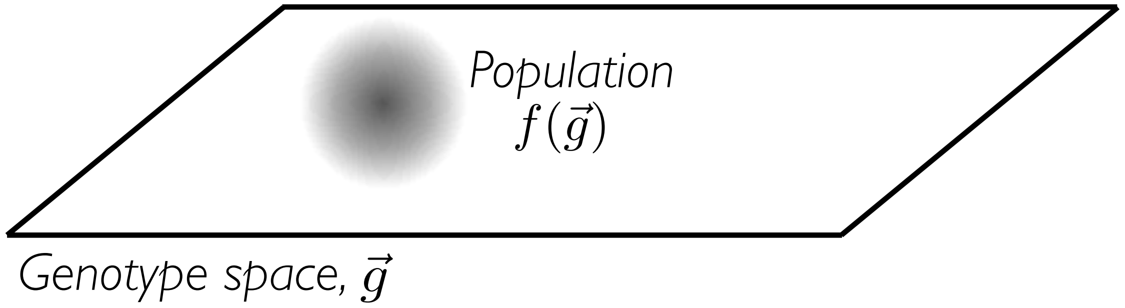
Before: $\vec{g} =$

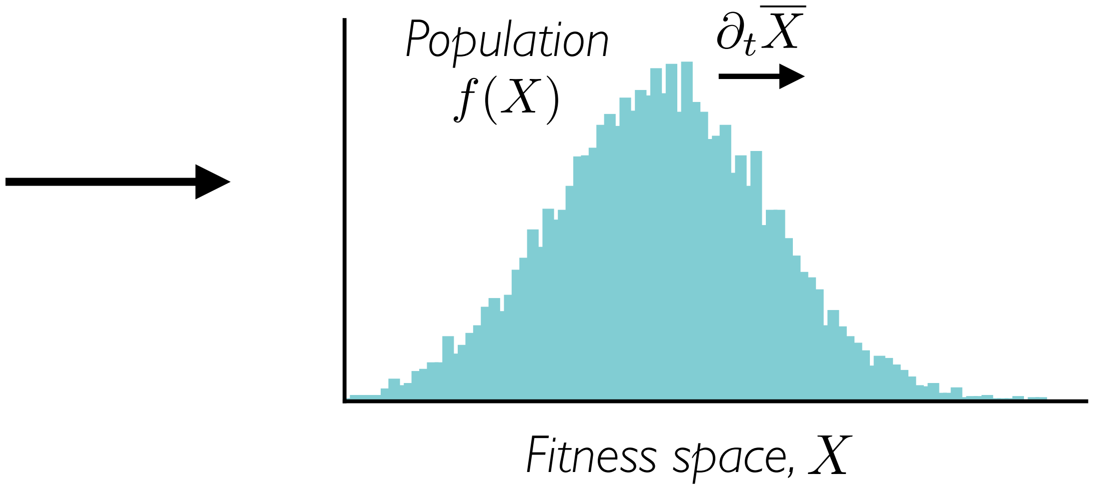


Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

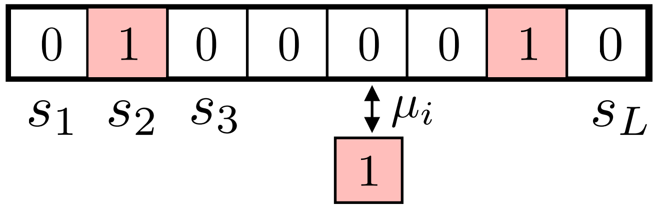
How do we think about genetic linkage theoretically?

When there are lots of linked mutations…



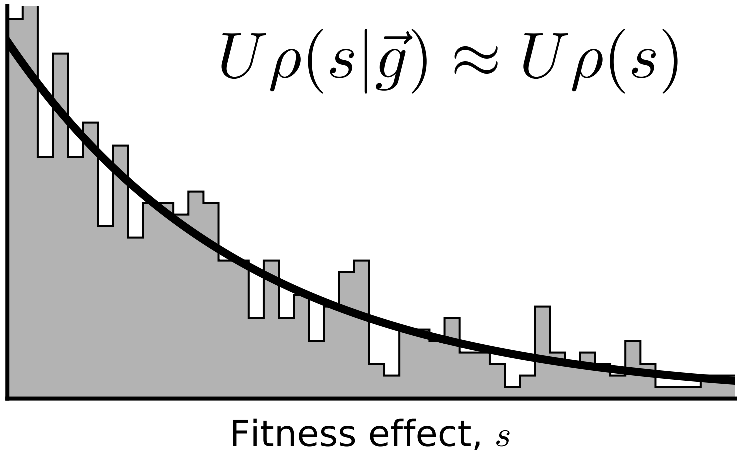


Before: $\vec{g} =$



Now: $X \rightarrow X + s$

Statistical distribution of fitness effects (DFE),

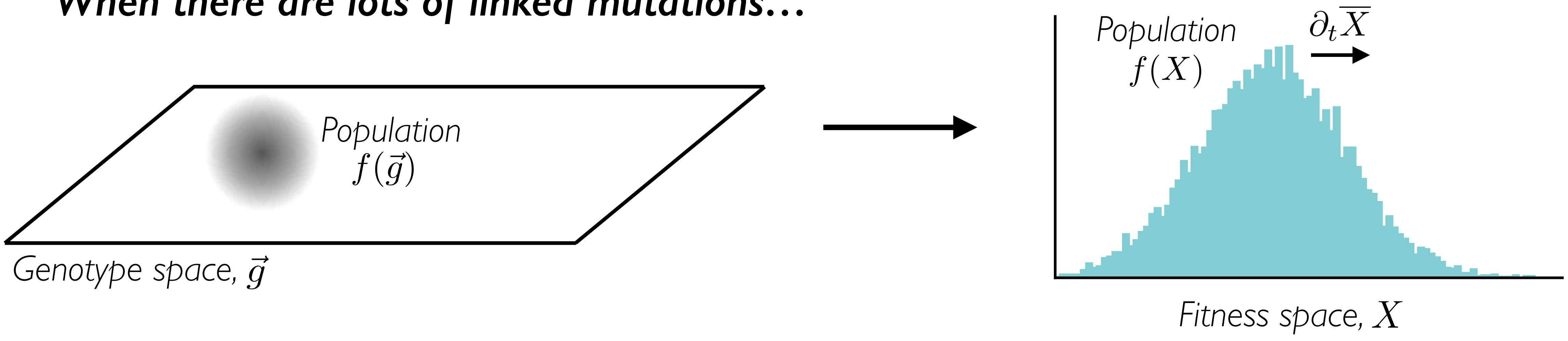


$$U\rho(s|\vec{g}) \approx U\rho(s)$$

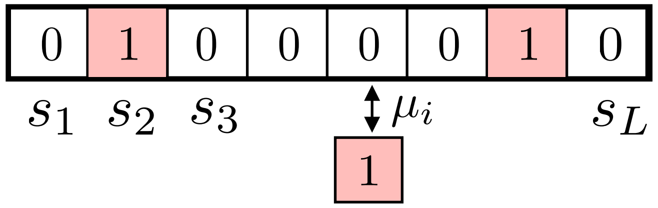
Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

How do we think about genetic linkage theoretically?

When there are lots of linked mutations…

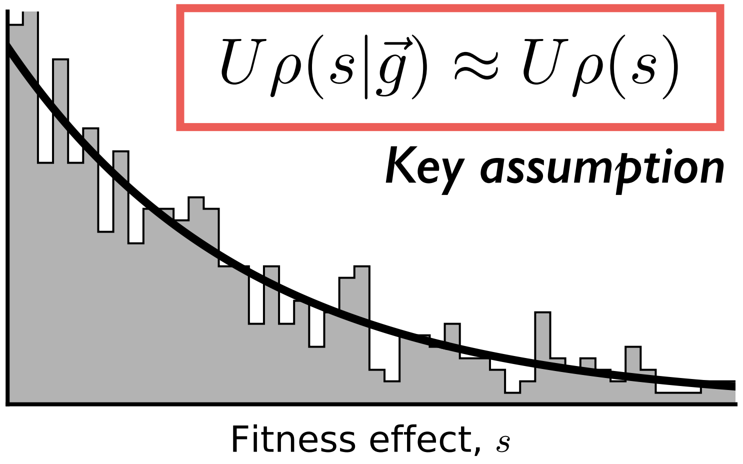


Before: $\vec{g} =$



Now: $X \to X + s$

Statistical distribution of fitness effects (DFE),



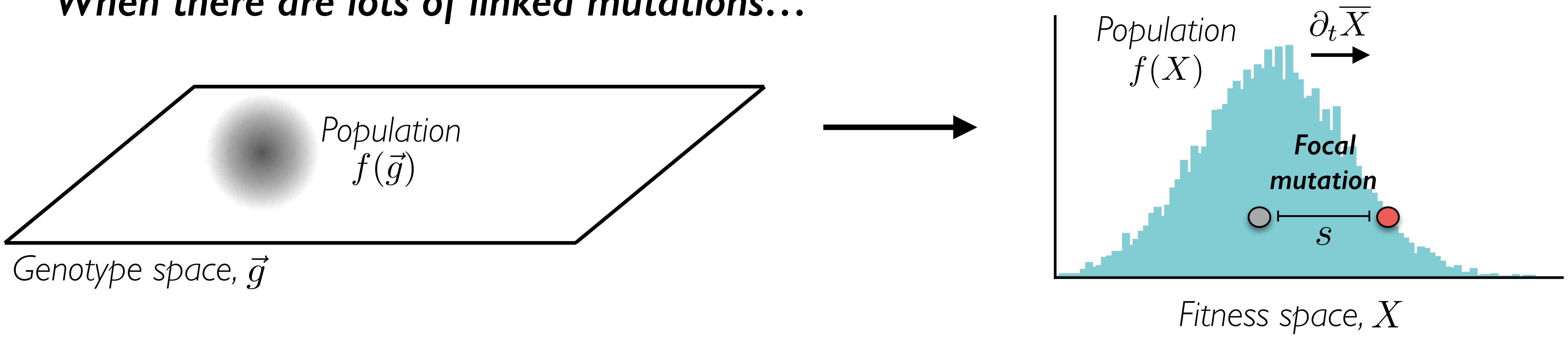
$$\boxed{U\rho(s|\vec{g}) \approx U\rho(s)}$$

Key assumption

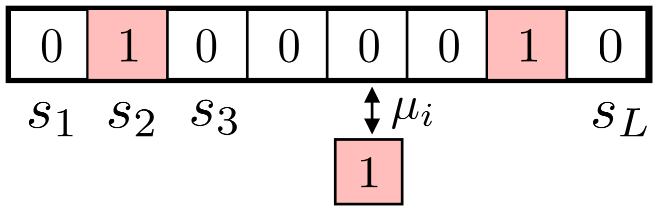
Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

How do we think about genetic linkage theoretically?

When there are lots of linked mutations…

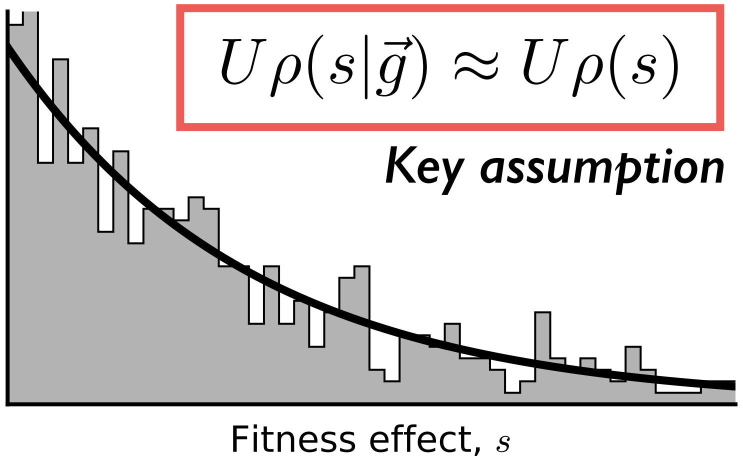


Before: $\vec{g} =$



Now: $X \to X + s$

Statistical distribution of fitness effects (DFE),



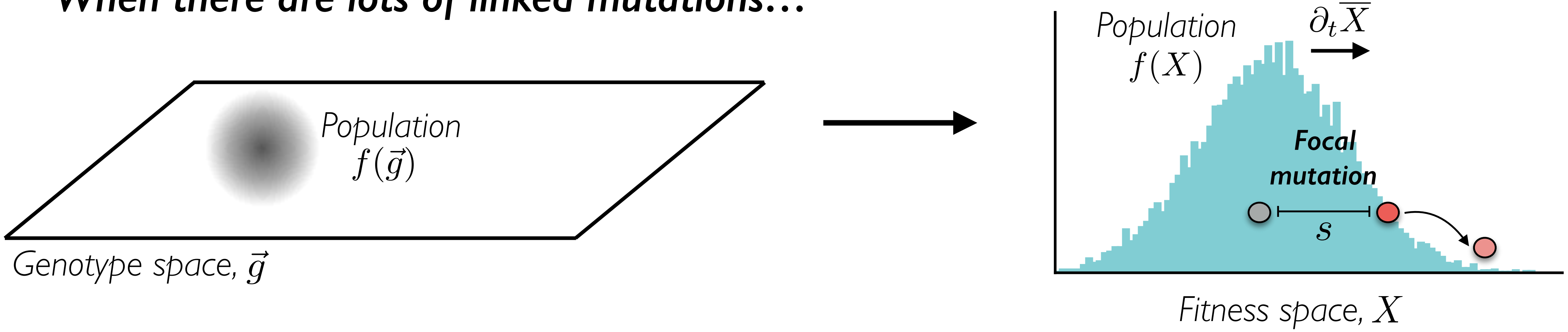
$$\boxed{U\rho(s|\vec{g}) \approx U\rho(s)}$$

Key assumption

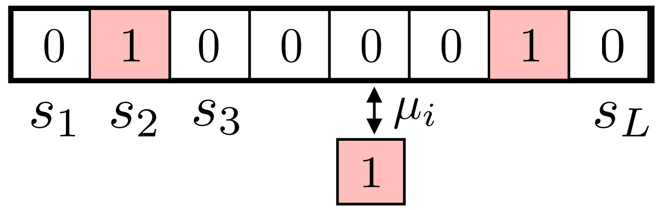
Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

How do we think about genetic linkage theoretically?

When there are lots of linked mutations…

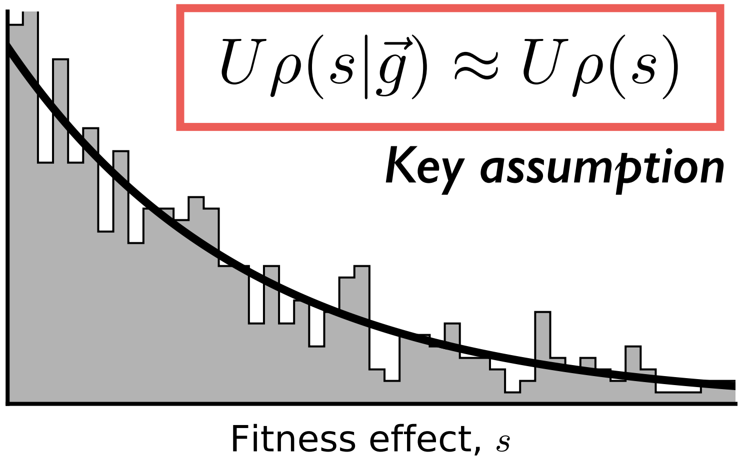


Before: $\vec{g} =$



Now: $X \rightarrow X + s$

Statistical distribution of fitness effects (DFE),

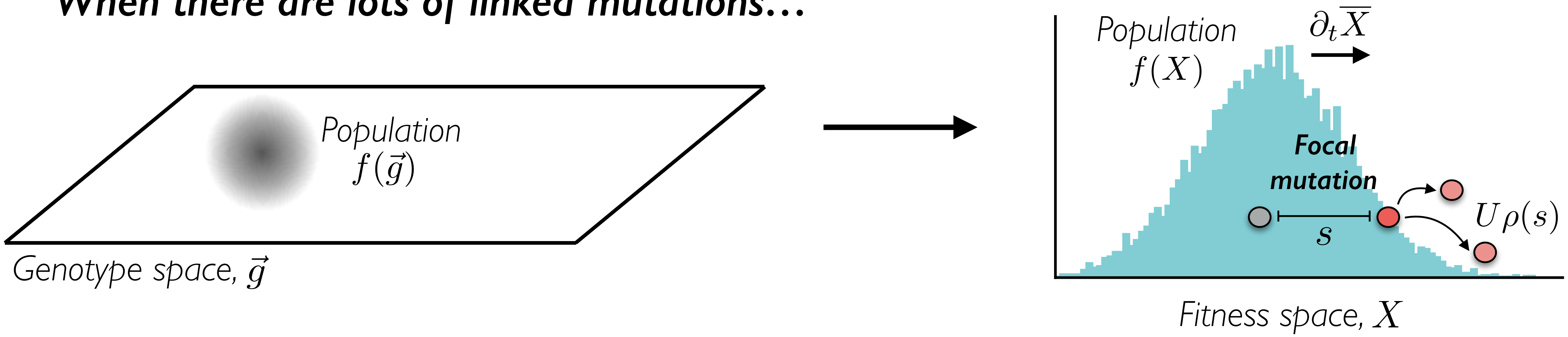

$$U\rho(s|\vec{g}) \approx U\rho(s)$$

Key assumption

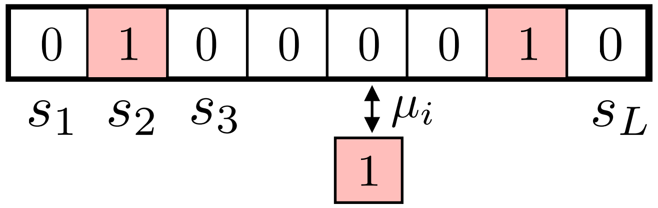
Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

How do we think about genetic linkage theoretically?

When there are lots of linked mutations…



Before: $\vec{g} =$

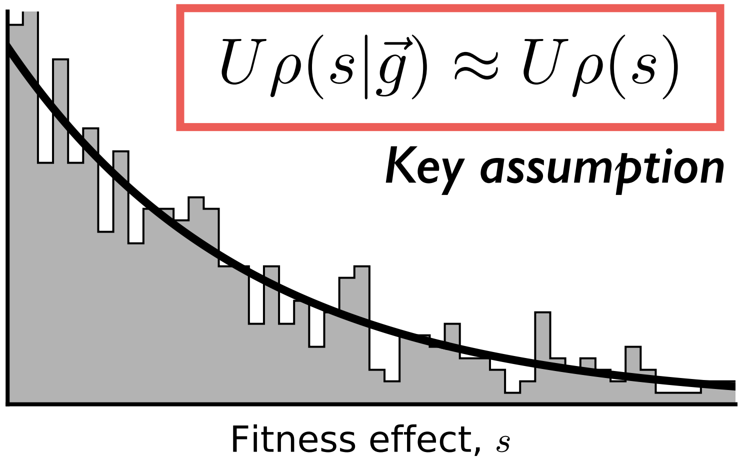


Now: $X \to X + s$

Statistical distribution of fitness effects (DFE),

$$\boxed{U\rho(s|\vec{g}) \approx U\rho(s)}$$

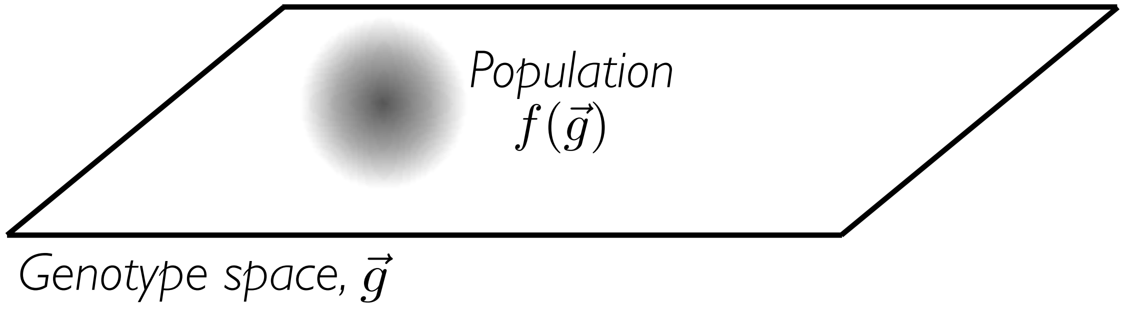
Key assumption

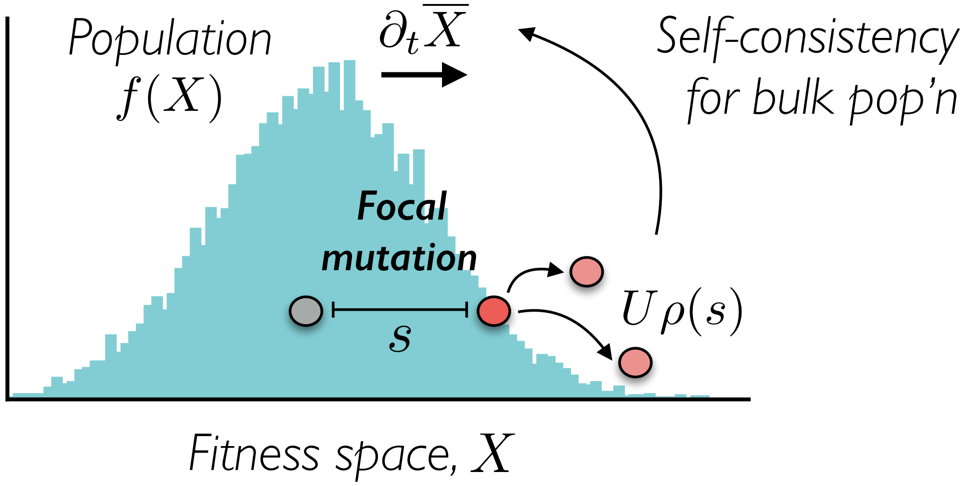


Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

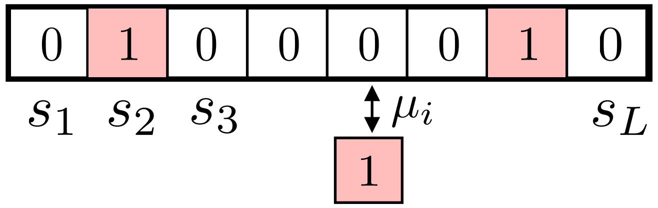
How do we think about genetic linkage theoretically?

When there are lots of linked mutations…



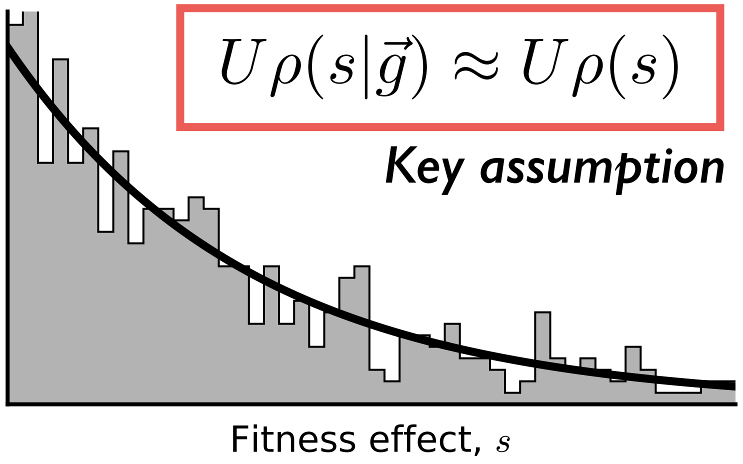


Before: $\vec{g} =$



Now: $X \to X + s$

Statistical distribution of fitness effects (DFE),



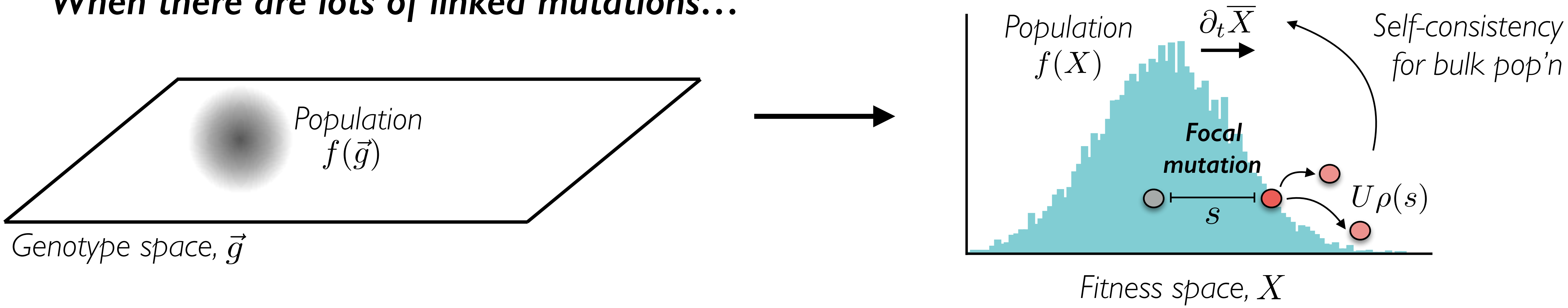
$$\boxed{U\rho(s|\vec{g}) \approx U\rho(s)}$$

Key assumption

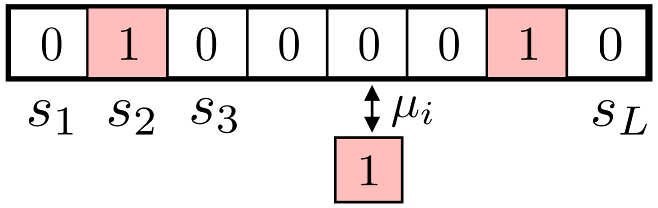
Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

How do we think about genetic linkage theoretically?

When there are lots of linked mutations…

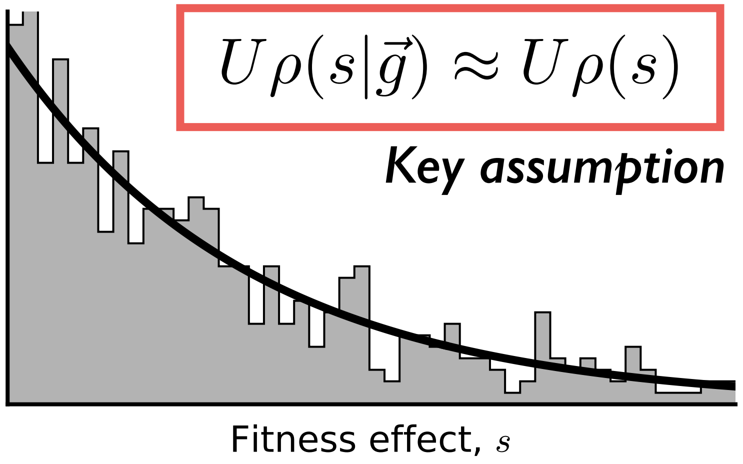


Before: $\vec{g} =$



Now: $X \to X + s$

Statistical distribution of fitness effects (DFE),



$$\boxed{U\rho(s|\vec{g}) \approx U\rho(s)}$$

Key assumption

Predictions for

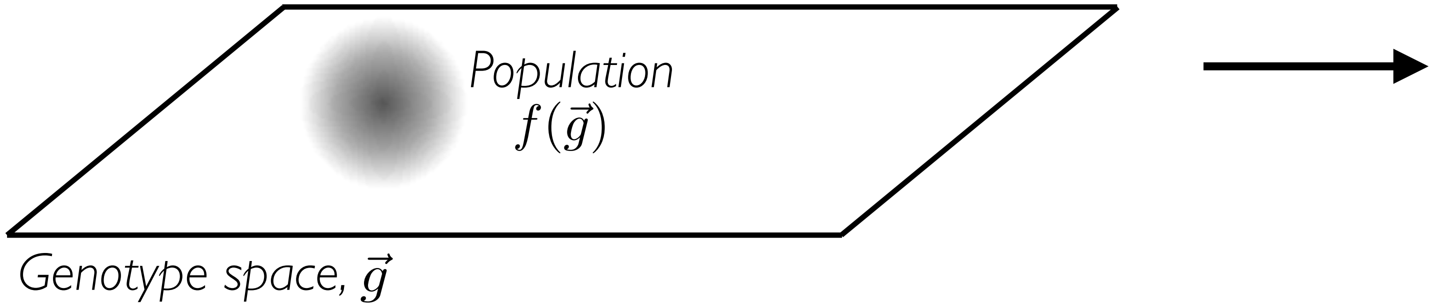
1. *Common* observables in DNA sequence data (e.g., substitution rates, genetic diversity)*

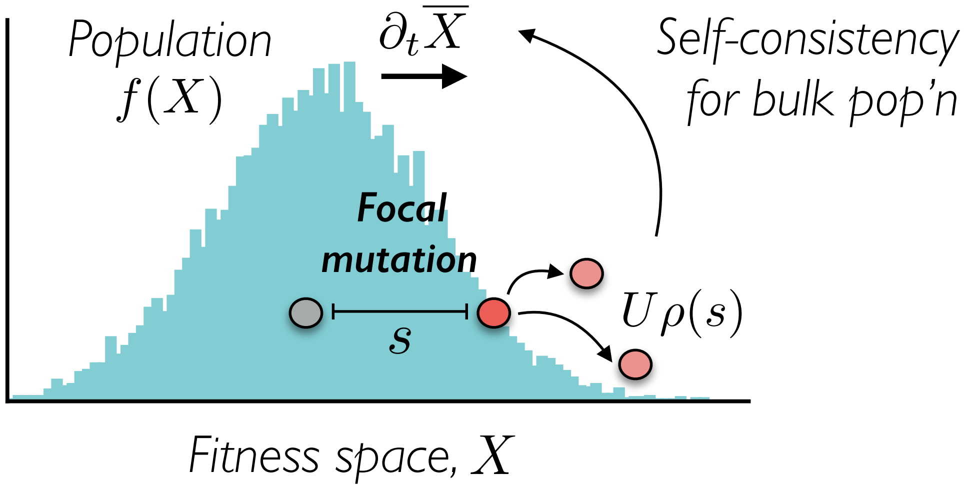
2. *Dependence on underlying parameters,* $U\rho(s)$

Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

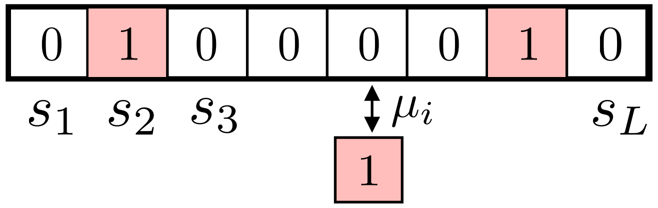
How do we think about genetic linkage theoretically?

When there are lots of linked mutations…





Before: $\vec{g} =$

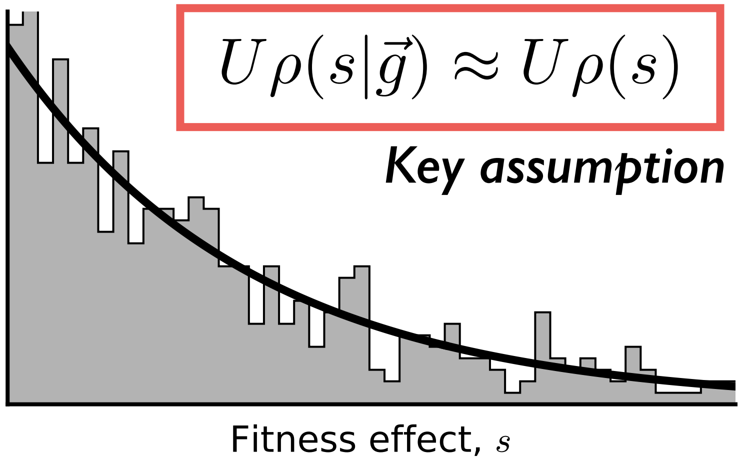


Now: $X \to X + s$

Statistical distribution of fitness effects (DFE),

$$\boxed{U\rho(s|\vec{g}) \approx U\rho(s)}$$

Key assumption



Predictions for

1. *Common* observables in DNA sequence data (e.g., substitution rates, genetic diversity)*

2. *Dependence on underlying parameters,*

$$U\rho(s) \approx \underset{(U_b^*,\, s_b^*)}{\text{strongly beneficial}} + \underset{(U_d^*,\, -s_d^*)}{\text{strongly deleterious}} + \underset{(U_n^*,\, s_n^*)}{\text{nearly neutral}}$$

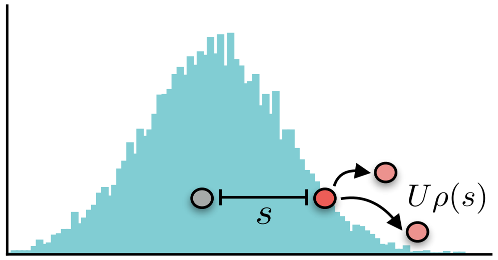
Good et al (2012); Good and Desai (2013; 2014; 2016)
see also Neher et al (2010); Hallatschek (2011); Fisher (2013)

Key insights from theory

Key insights from theory

1. In rapidly evolving populations…

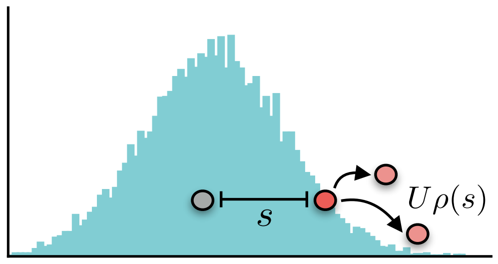
fates of mutations are strongly influenced by ***dynamical processes***, in addition to their ***inherent biological effects***.



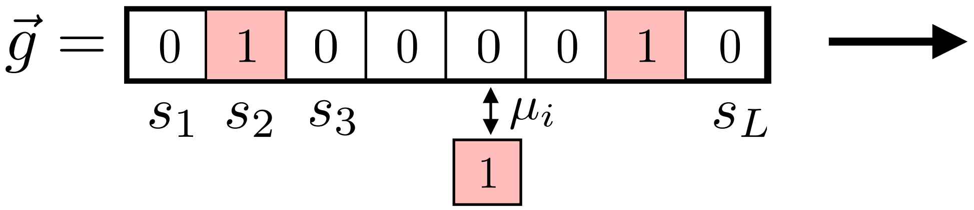
Key insights from theory

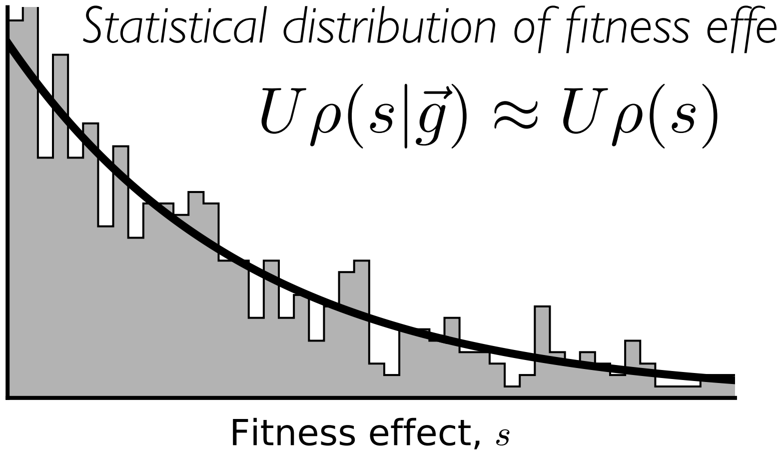
1. In rapidly evolving populations…

fates of mutations are strongly influenced by ***dynamical processes***, in addition to their ***inherent biological effects***.



2. Think statistically!





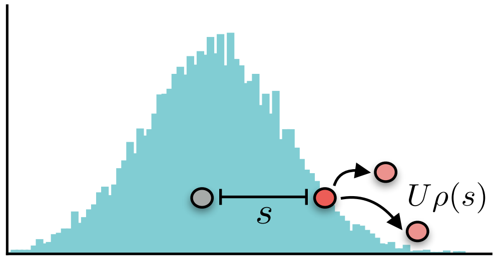
Statistical distribution of fitness effects (DFE),

$$U\rho(s|\vec{g}) \approx U\rho(s)$$

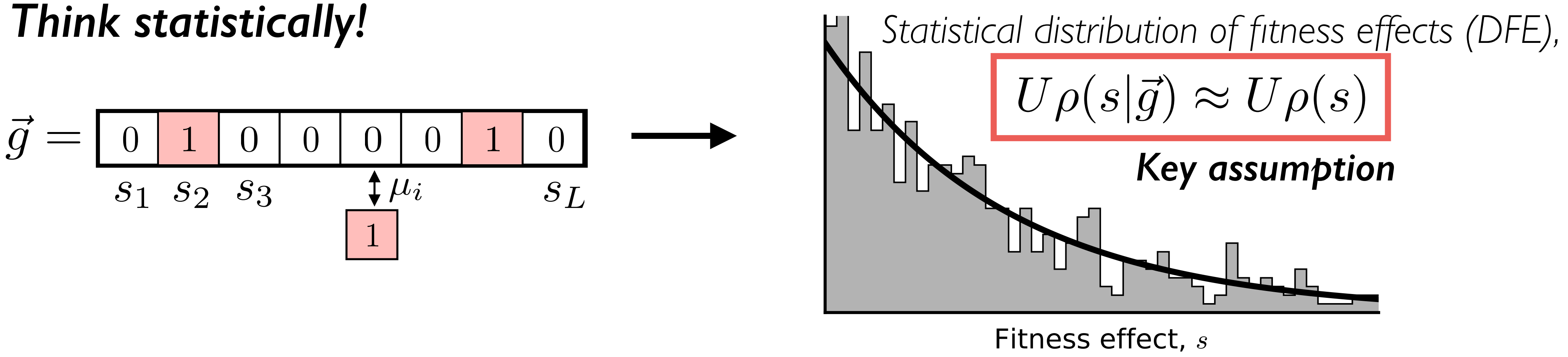
Key insights from theory

1. In rapidly evolving populations…

fates of mutations are strongly influenced by ***dynamical processes***, in addition to their ***inherent biological effects***.



2. Think statistically!



Statistical distribution of fitness effects (DFE),

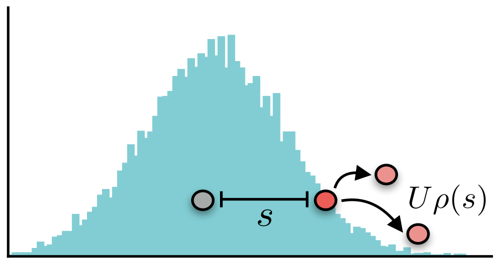
$$\boxed{U\rho(s|\vec{g}) \approx U\rho(s)}$$

Key assumption

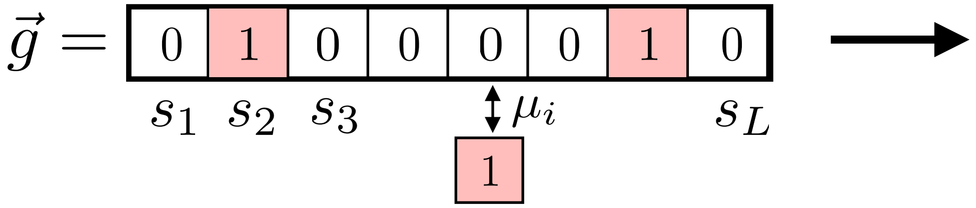
Key insights from theory

1. In rapidly evolving populations…

fates of mutations are strongly influenced by ***dynamical processes***, in addition to their ***inherent biological effects***.



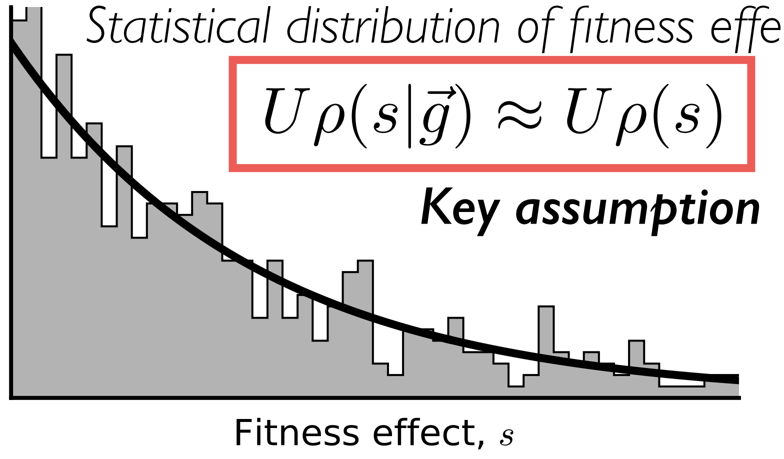
2. Think statistically!



Statistical distribution of fitness effects (DFE),

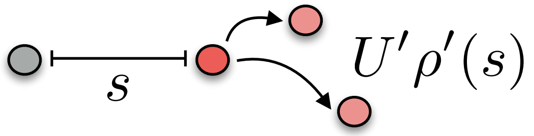
$$\boxed{U\rho(s|\vec{g}) \approx U\rho(s)}$$

Key assumption



Future direction:

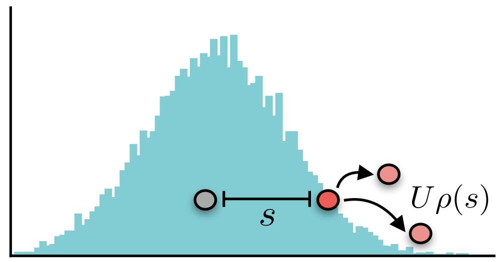
"Macroscopic" epistasis ⟶ *mutations that change rates & fitness effects of other mutations*



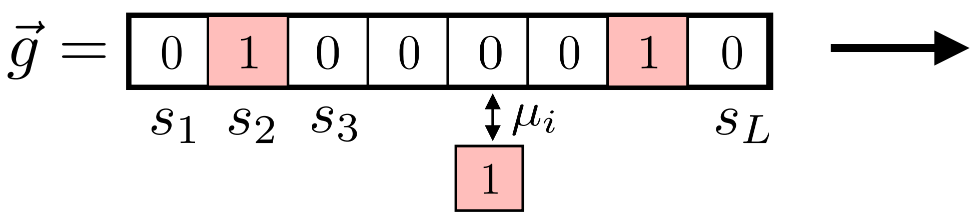
Key insights from theory

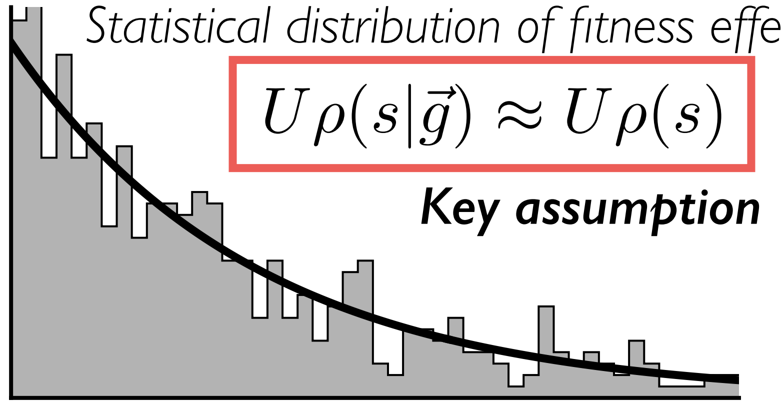
1. In rapidly evolving populations…

fates of mutations are strongly influenced by ***dynamical processes***, in addition to their ***inherent biological effects***.



2. Think statistically!





Statistical distribution of fitness effects (DFE),

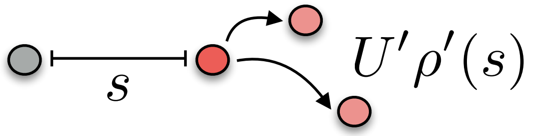
$$\boxed{U\rho(s|\vec{g}) \approx U\rho(s)}$$

Key assumption

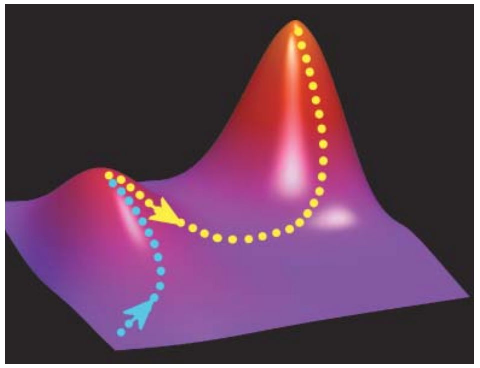
Fitness effect, s

Future direction:

"Macroscopic" epistasis → *mutations that change rates & fitness effects of other mutations*



Questions: *when is invasion* $\neq$ *optimal for population?*
How far can populations "see" in fitness landscape?

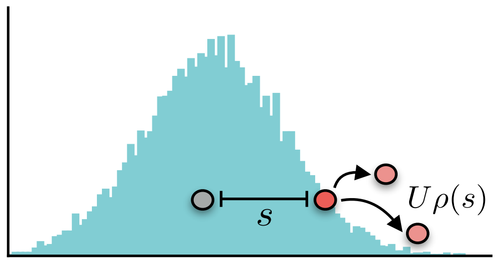


Poelwijk, Kiviet* et al (2007)*

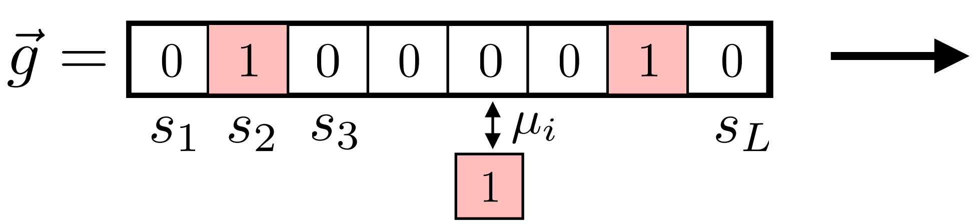
Key insights from theory

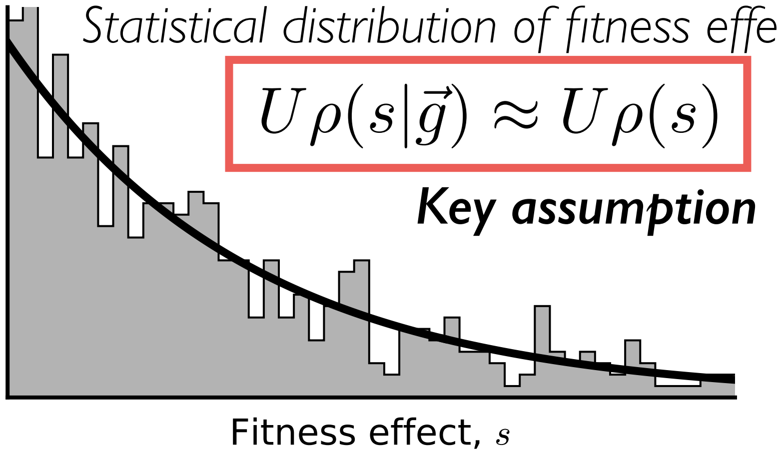
1. In rapidly evolving populations…

fates of mutations are strongly influenced by ***dynamical processes***, in addition to their ***inherent biological effects***.



2. Think statistically!





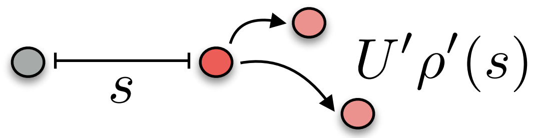
Statistical distribution of fitness effects (DFE),

$$\boxed{U\rho(s|\vec{g}) \approx U\rho(s)}$$

Key assumption

Future direction:

"Macroscopic" epistasis → *mutations that change rates & fitness effects of other mutations*

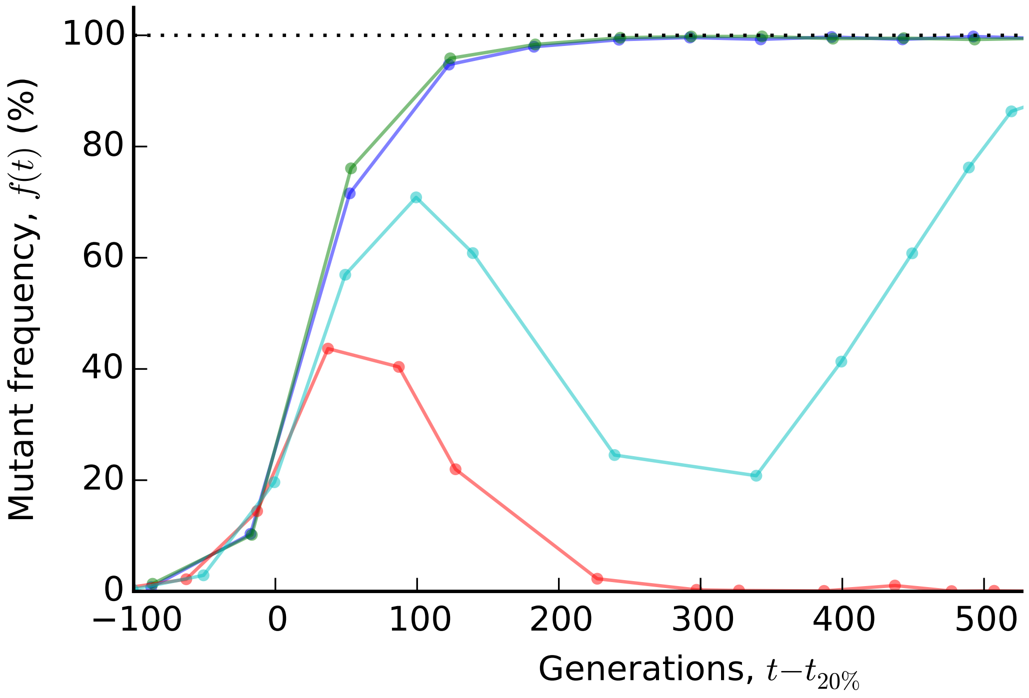


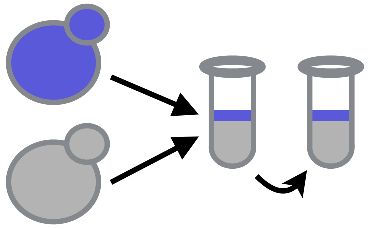
vs traditional "microscopic" expansion

$$X(\vec{g}) \approx \sum_i s_i g_i + \sum_{ij} \epsilon_{ij} g_i g_j + \dots$$

Questions: *when is invasion* $\neq$ *optimal for population?*
How far can populations "see" in fitness landscape?

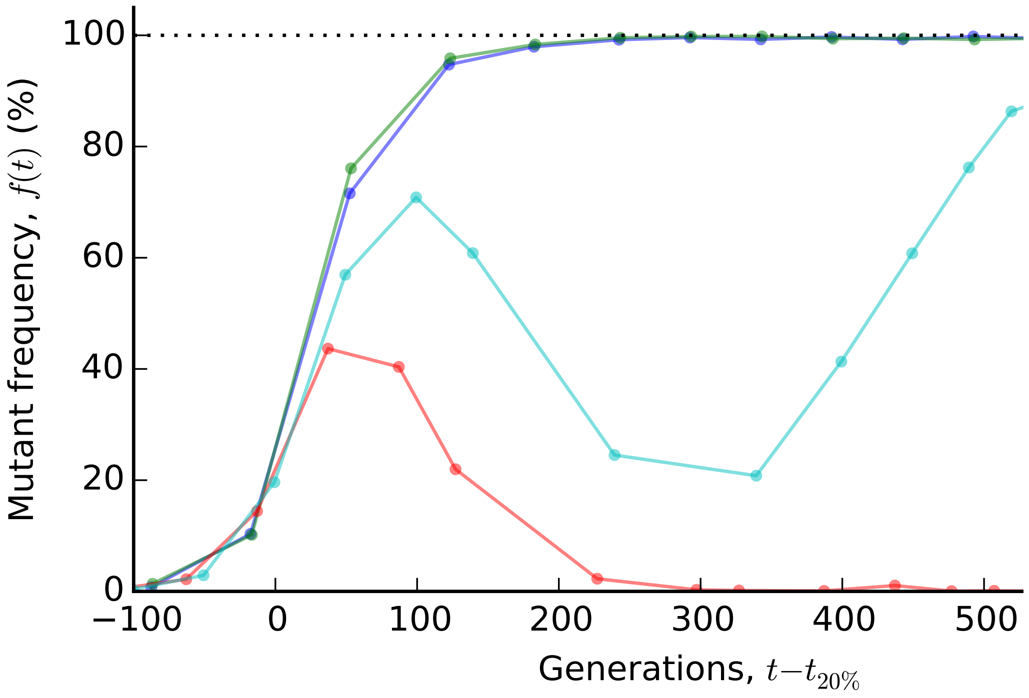
But how relevant are these models, really?

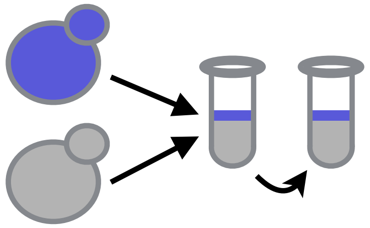




Short-term adaptation
to ***new*** environment…

But how relevant are these models, really?

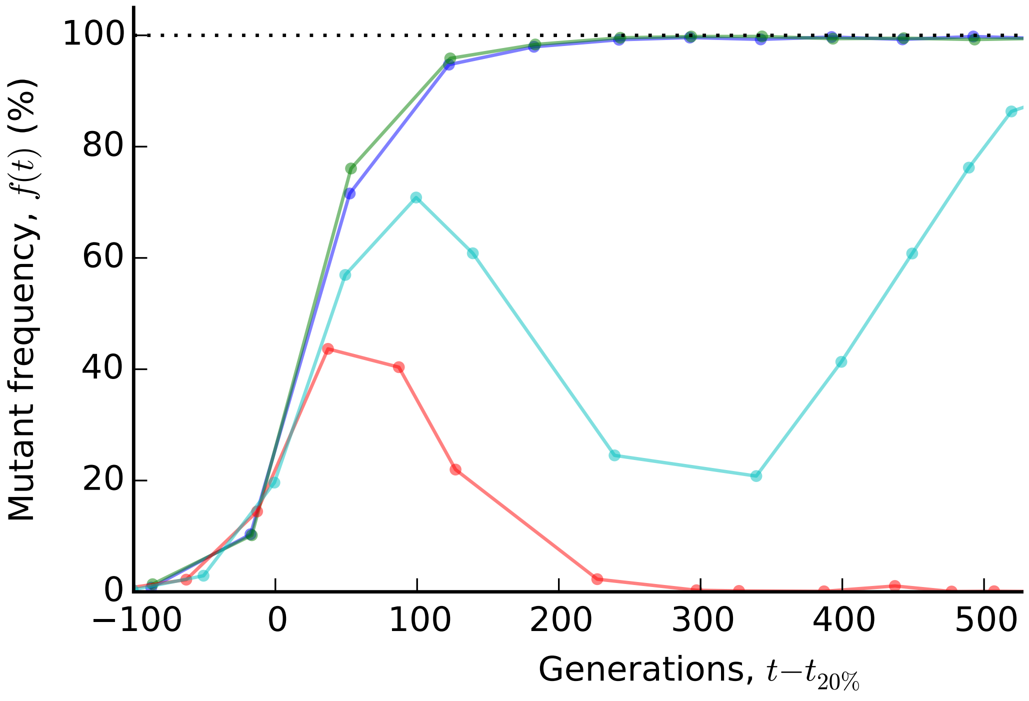


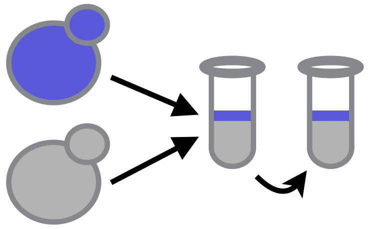


Short-term adaptation
to ***new*** environment…

Does ***long-term*** adaptation look different?

But how relevant are these models, really?



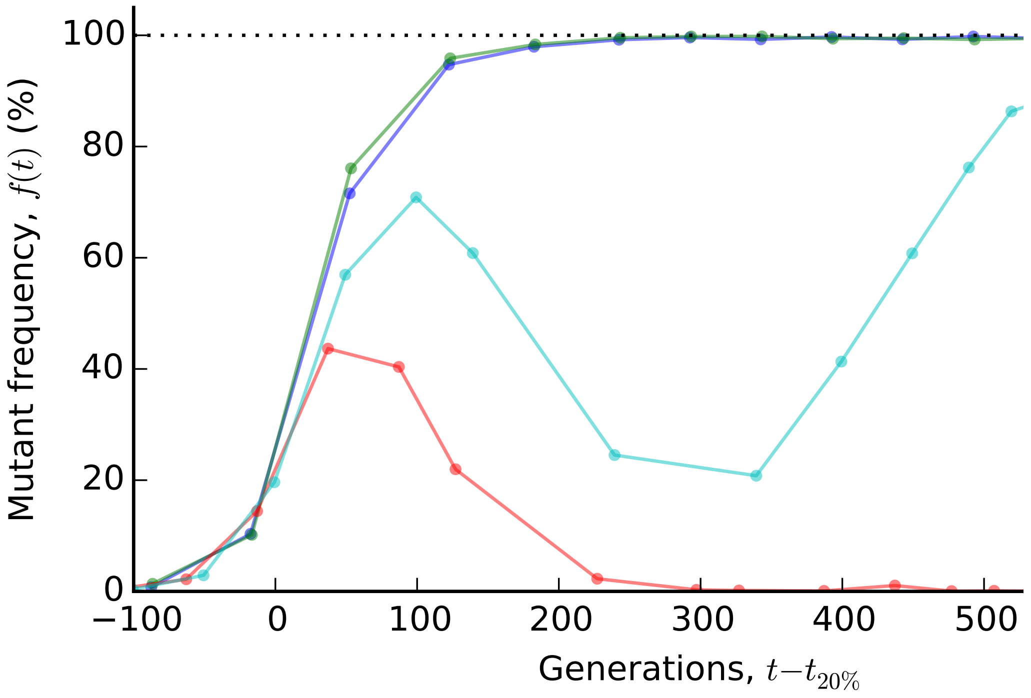


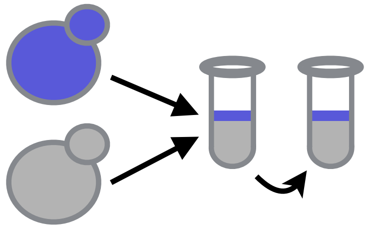
Short-term adaptation to ***new*** environment…

Does ***long-term*** adaptation look different?

Questions: Do dynamics eventually become "simple" again?

But how relevant are these models, really?





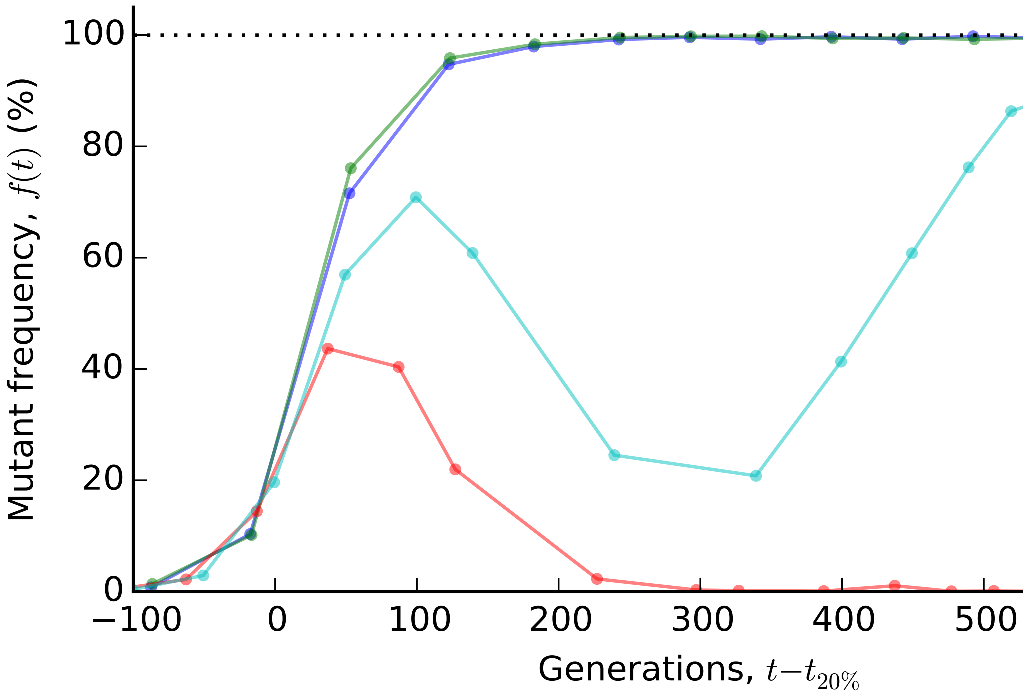
Short-term adaptation
to ***new*** environment…

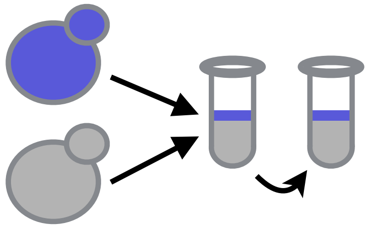
Does ***long-term*** adaptation look different?

Questions: Do dynamics eventually become “simple” again?

or do new complications arise that require revisions to the model?

But how relevant are these models, really?



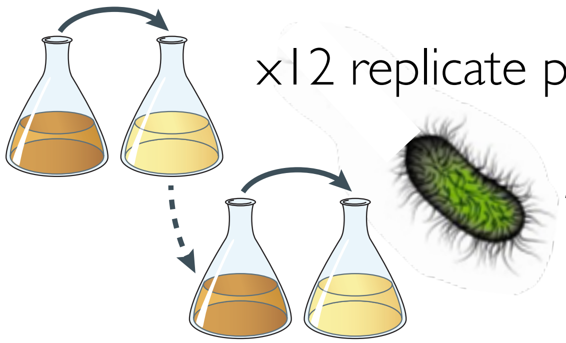


Short-term adaptation to ***new*** environment…

Does ***long-term*** adaptation look different?

A long-term experiment, "the LTEE"

Richard Lenski

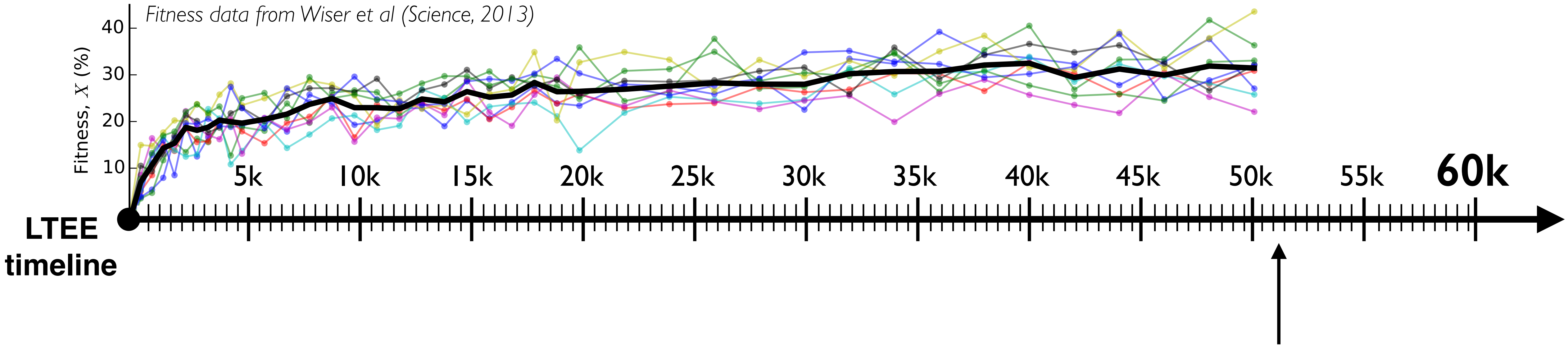


×12 replicate pop'ns

E. coli

× 30 years…
(& counting!)

Long-term adaptation to a constant environment

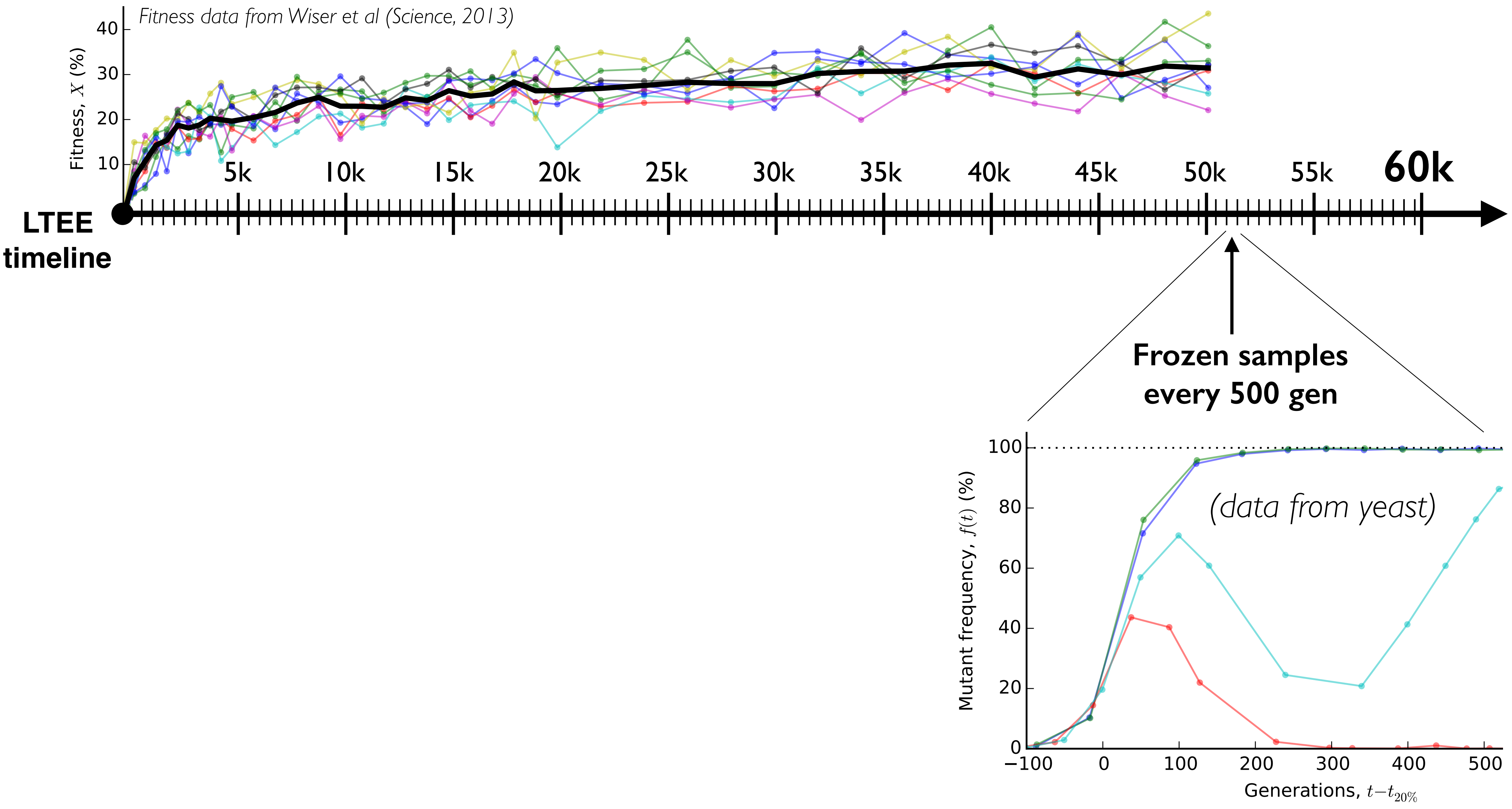


Fitness data from Wiser et al (Science, 2013)

LTEE timeline

Frozen samples every 500 gen

Long-term adaptation to a constant environment



Fitness data from Wiser et al (Science, 2013)

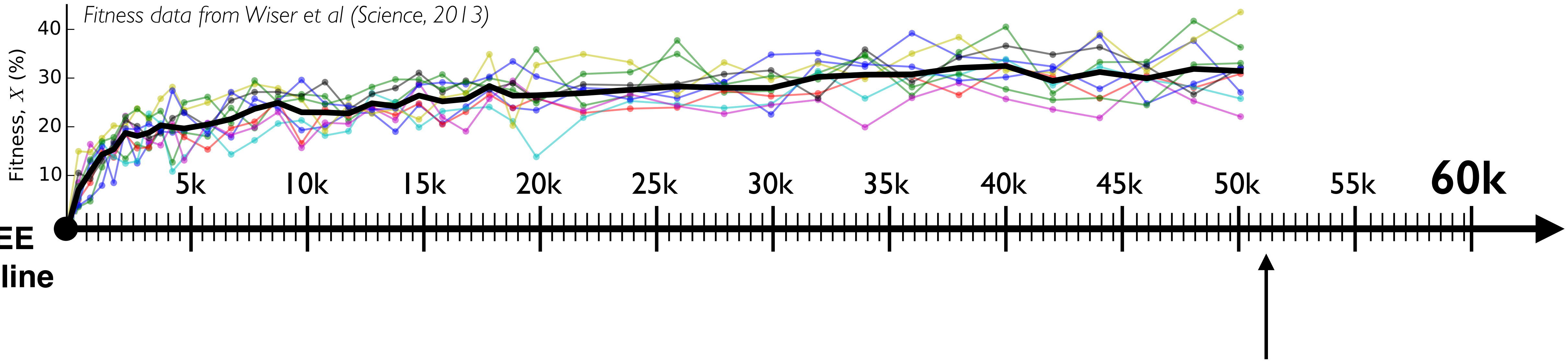
LTEE timeline

Frozen samples every 500 gen

(data from yeast)

Problem: Samples too far apart?

Long-term adaptation to a constant environment



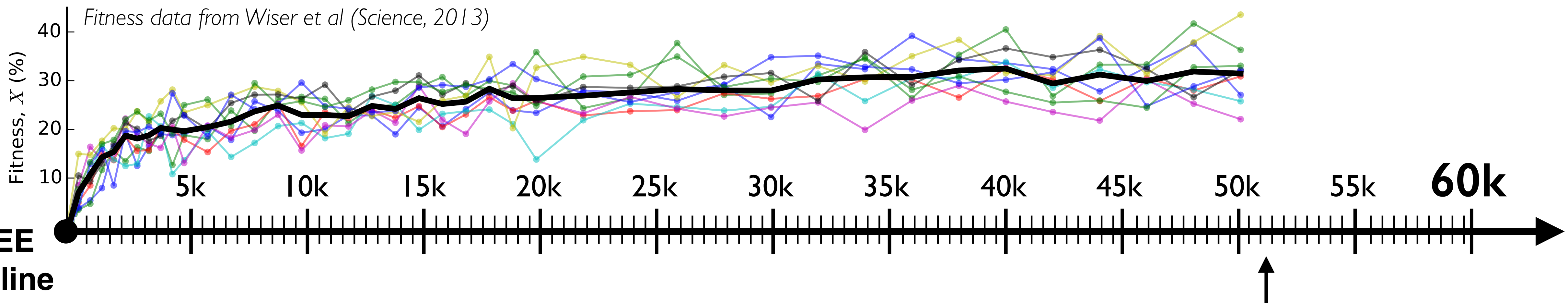
LTEE timeline

Frozen samples every 500 gen

Solution: Fit to ***model*** suggests that we can see dynamics!

[***"knobs":*** $U\rho(s) \approx (U_b^*, s_b^*) + (U_n^*, 0)$]

Long-term adaptation to a constant environment


Fitness data from Wiser et al (Science, 2013)

LTEE timeline

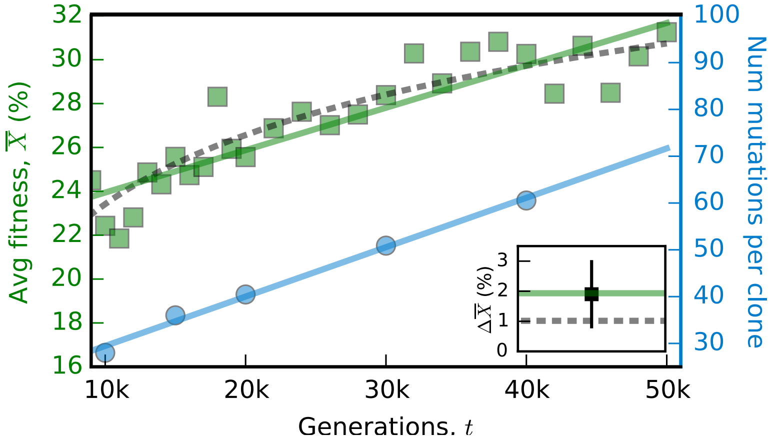
Frozen samples every 500 gen

Solution: Fit to ***model*** suggests that we can see dynamics!

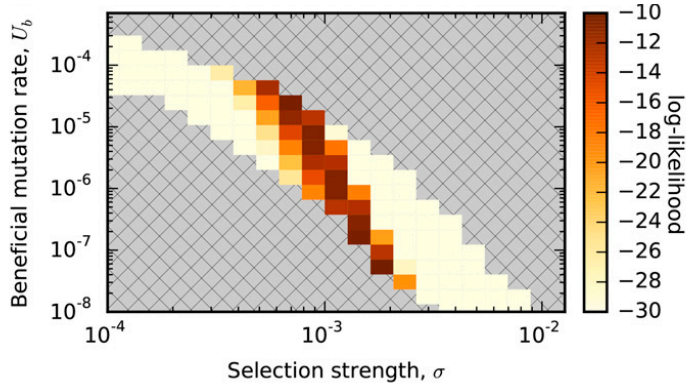
[***"knobs":*** $U\rho(s) \approx (U_b^*, s_b^*) + (U_n^*, 0)$]

---

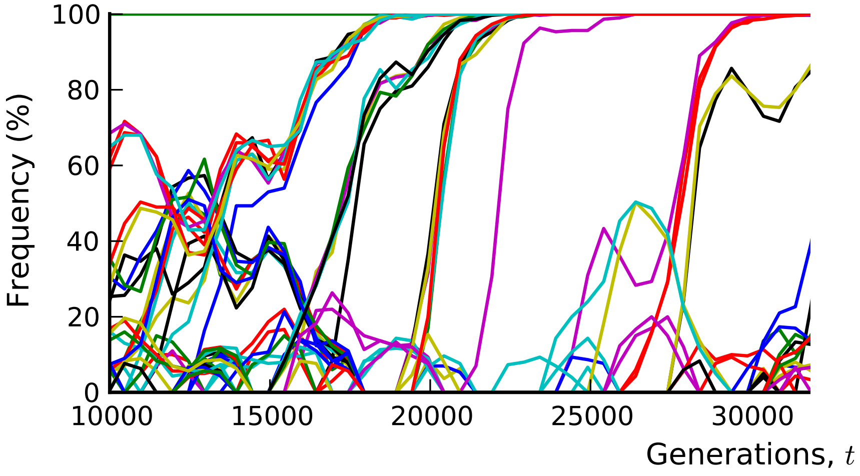
1. Previous data



2. Fit* to model

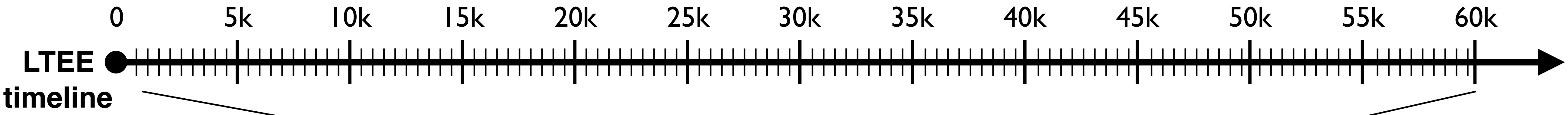


3. Simulated sequence data



Good and Desai (Genetics, 2015)

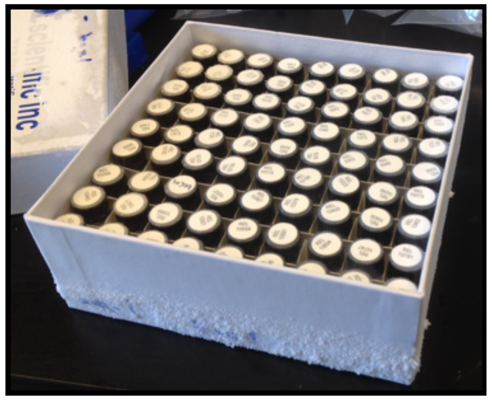
Reconstructing a “molecular fossil record”



w/ Michael McDonald

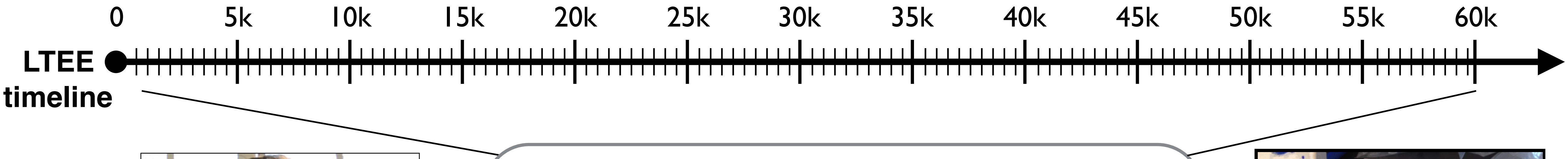
Sequencing strategy

Frozen mixed-population samples ~ 500 gen
x 12 populations
= ~1500 timepoints @ 10-100x coverage



“Ancient” DNA from Ara-1

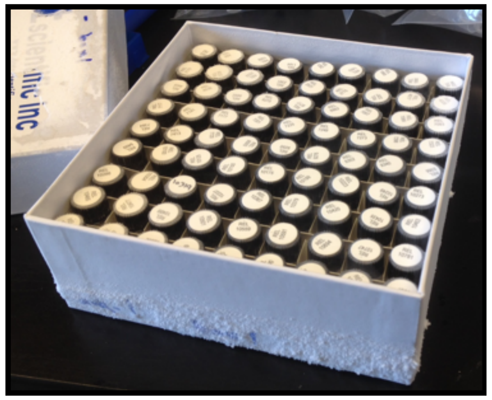
Reconstructing a "molecular fossil record"



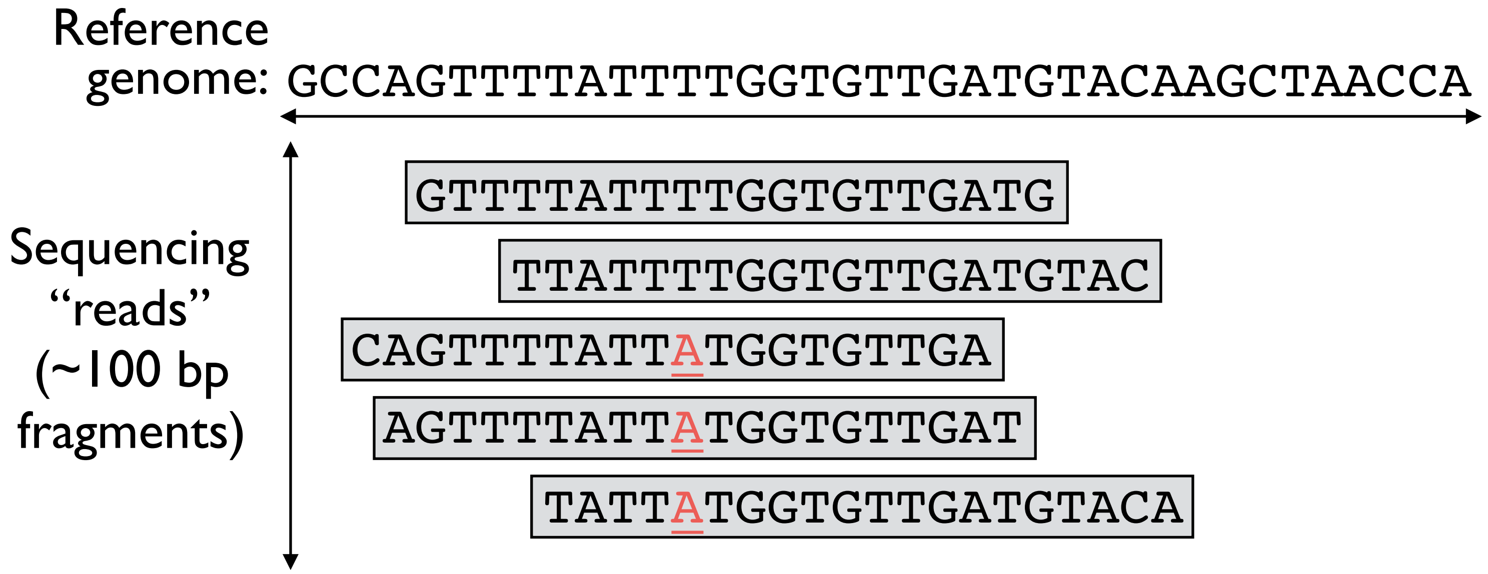
w/ Michael McDonald

Sequencing strategy

Frozen mixed-population samples ~ 500 gen
x 12 populations
= ~1500 timepoints @ 10-100x coverage



"Ancient" DNA from Ara-1

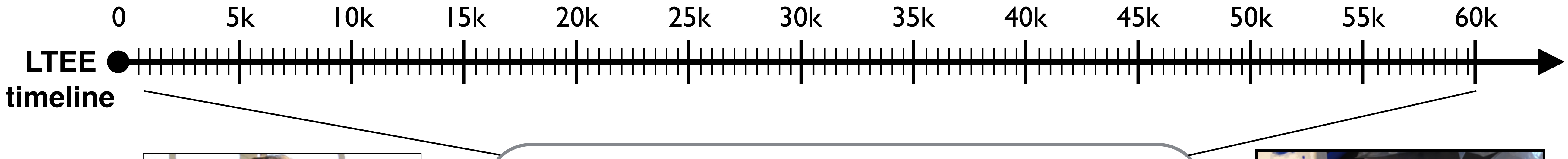


Reference genome: GCCAGTTTTATTTTGGTGTTGATGTACAAGCTAACCA

Sequencing "reads" (~100 bp fragments)

GTTTTATTTTGGTGTTGATG
TTATTTTGGTGTTGATGTAC
CAGTTTTATTATGGTGTTGA
AGTTTTATTATGGTGTTGAT
TATTATGGTGTTGATGTACA

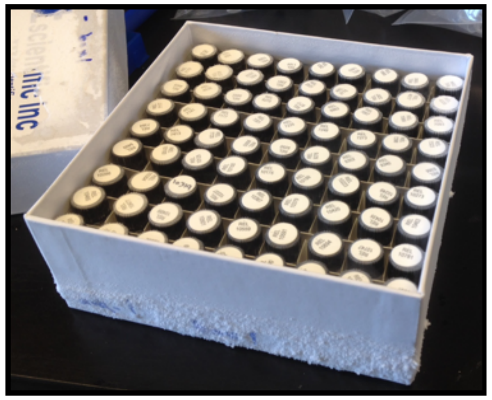
Reconstructing a "molecular fossil record"



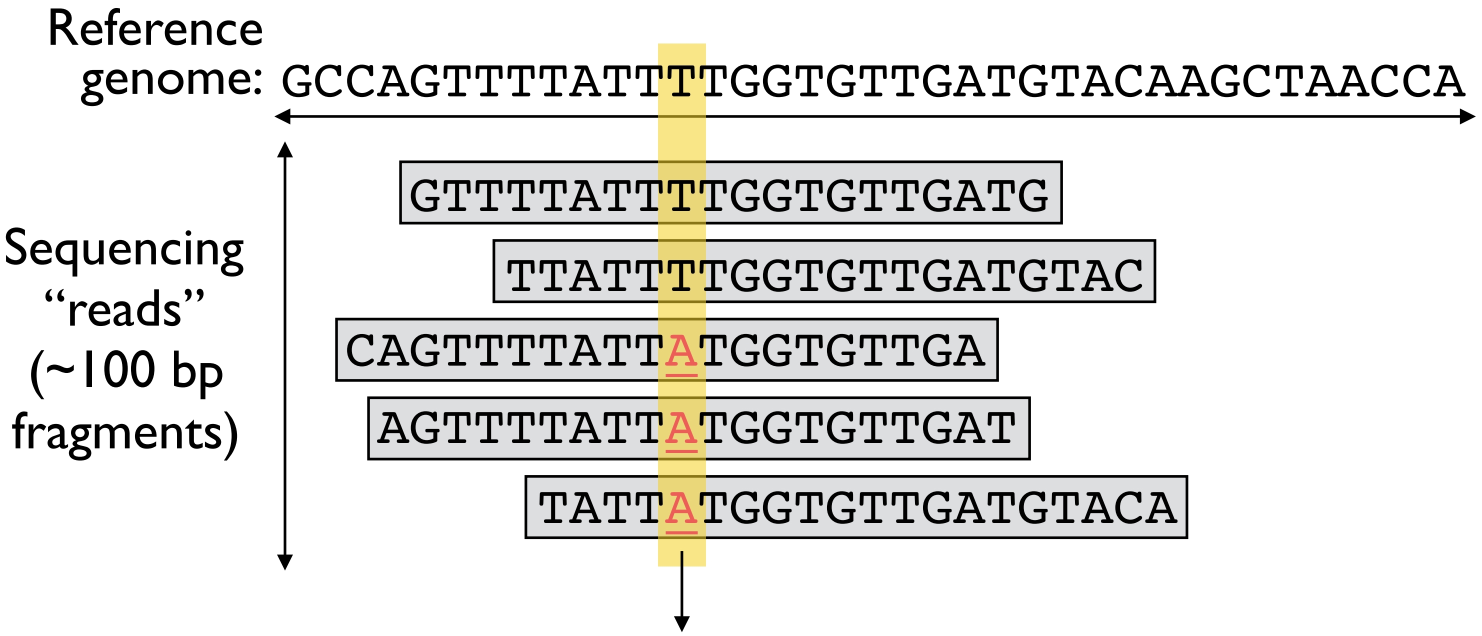
w/ Michael McDonald

Sequencing strategy

Frozen mixed-population samples ~ 500 gen
x 12 populations
= ~1500 timepoints @ 10-100x coverage



"Ancient" DNA from Ara-1



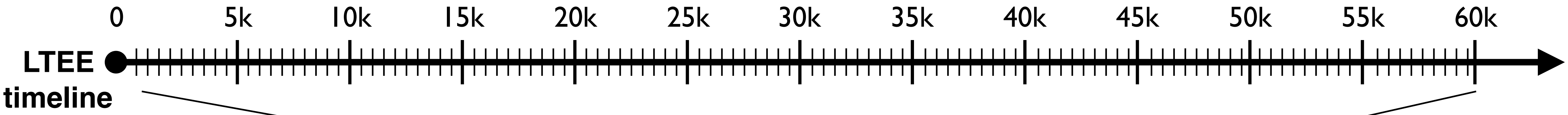
Reference genome: `GCCAGTTTTATTTTGGTGTTGATGTACAAGCTAACCA`

Sequencing "reads" (~100 bp fragments)

`GTTTTATTTTGGTGTTGATG`
`TTATTTTGGTGTTGATGTAC`
`CAGTTTTATTATGGTGTTGA`
`AGTTTTATTATGGTGTTGAT`
`TATTATGGTGTTGATGTACA`

"T→**A**" mutation @ 60% frequency
in this population in this timepjoint

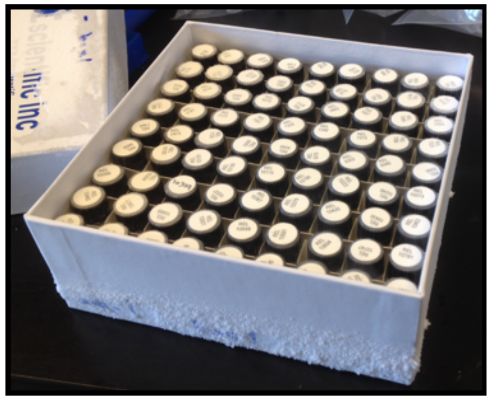
Reconstructing a "molecular fossil record"



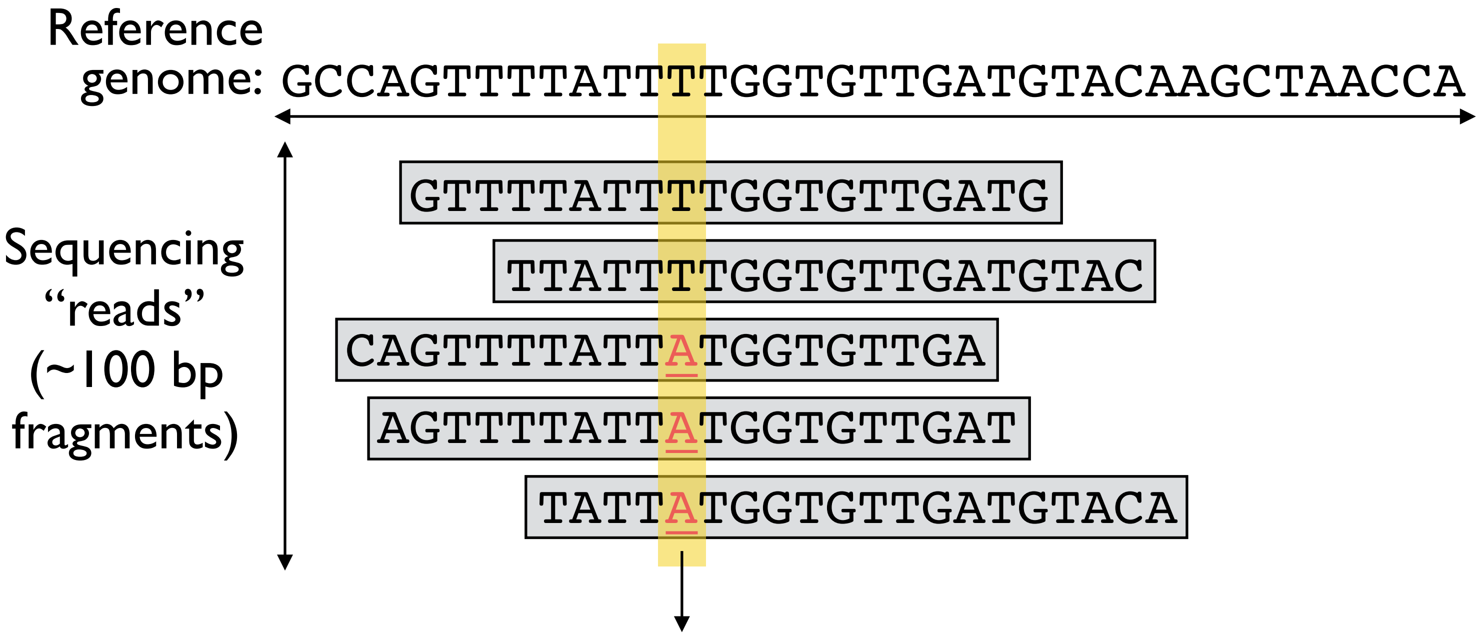
w/ Michael McDonald

Sequencing strategy

Frozen mixed-population samples ~ 500 gen
× 12 populations
= ~1500 timepoints @ 10-100x coverage



"Ancient" DNA from Ara-1

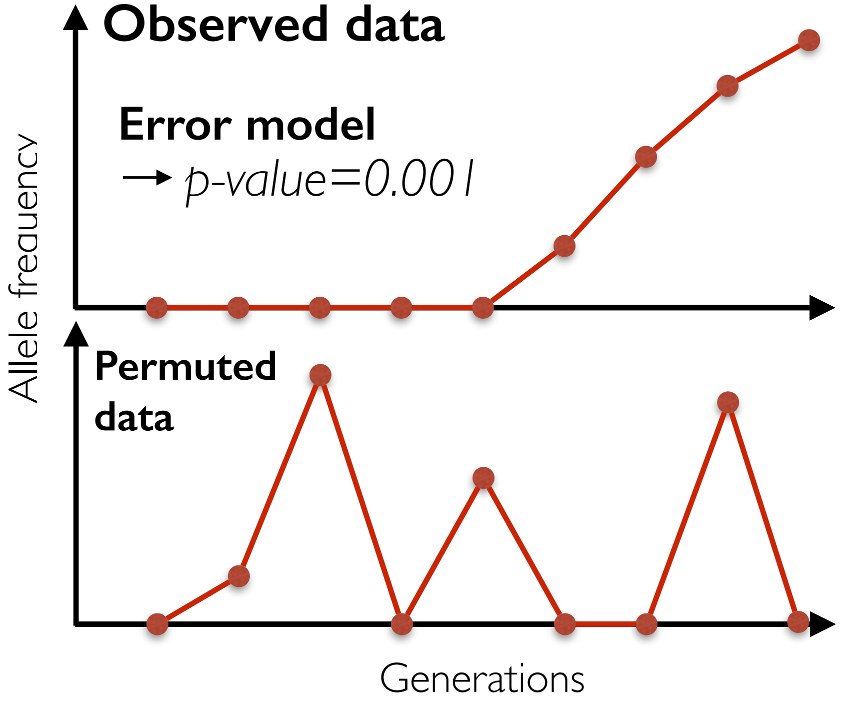


Reference genome: GCCAGTTTTATTTTGGTGTTGATGTACAAGCTAACCA

Sequencing "reads" (~100 bp fragments)

```
GTTTTATTTTGGTGTTGATG
TTATTTTGGTGTTGATGTAC
CAGTTTTATTATGGTGTTGA
AGTTTTATTATGGTGTTGAT
TATTATGGTGTTGATGTACA
```

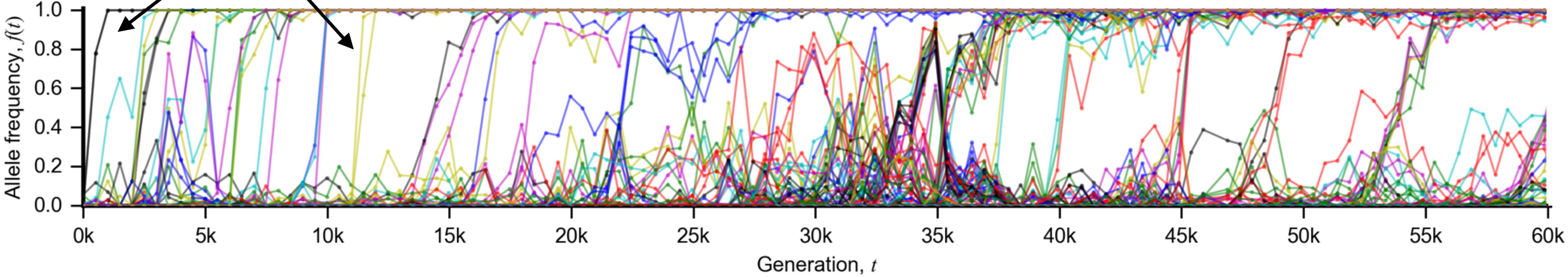
"T→**A**" mutation @ **60%** frequency
in this population in this timepjoint



Dynamics of molecular evolution over 60,000 generations

frequency trajectories
of different mutations

One population ("Ara+2") through time

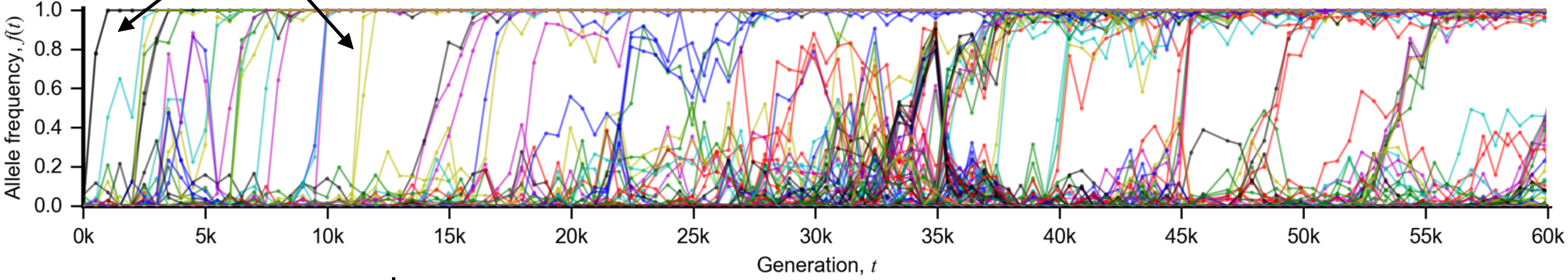


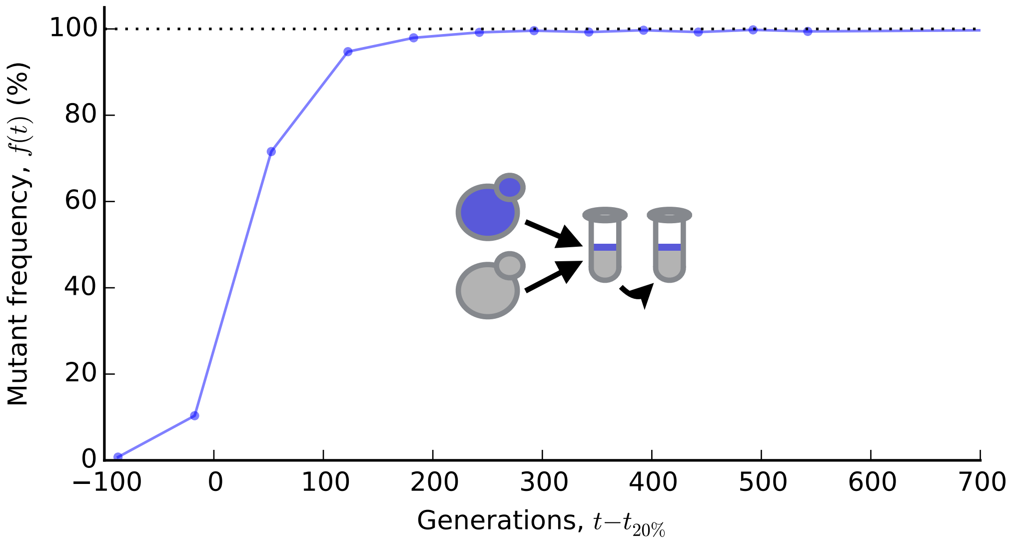
Good, McDonald* et al (Nature, 2017)*

Dynamics of molecular evolution over 60,000 generations

frequency trajectories
of different mutations

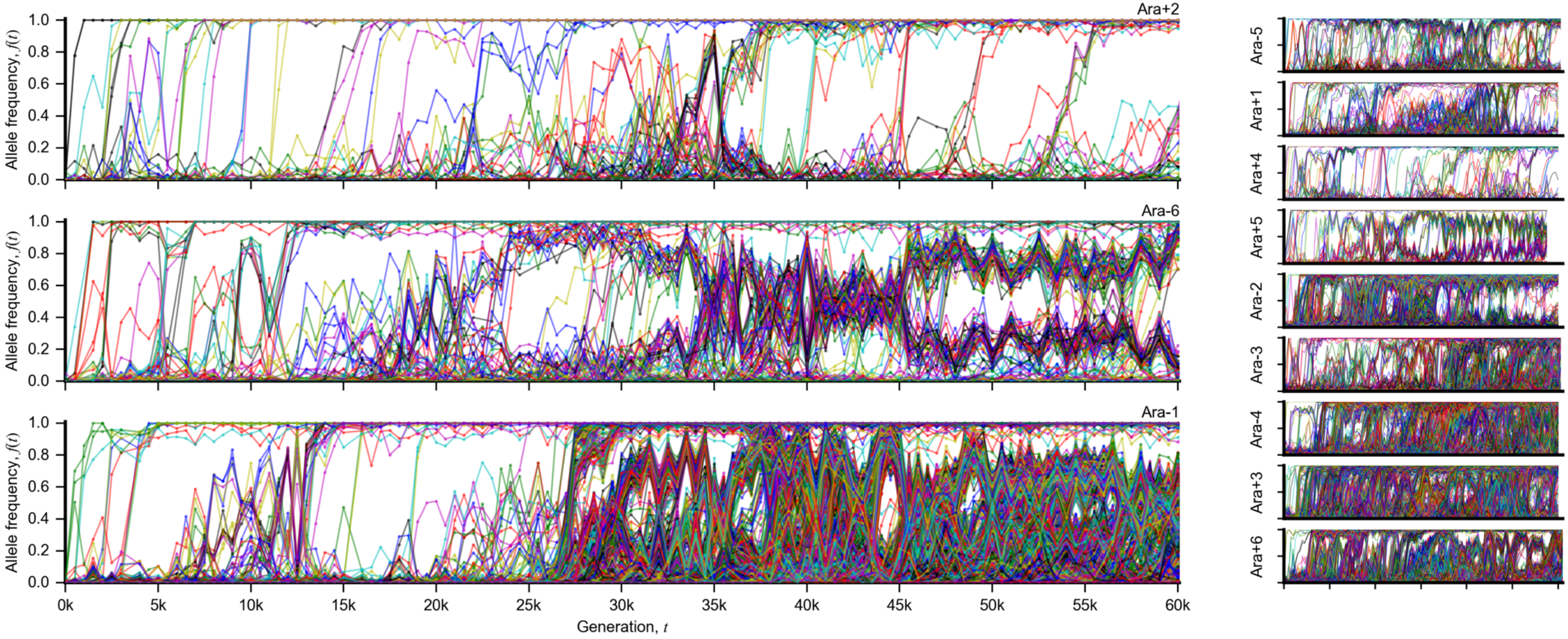
One population (“Ara+2”) through time





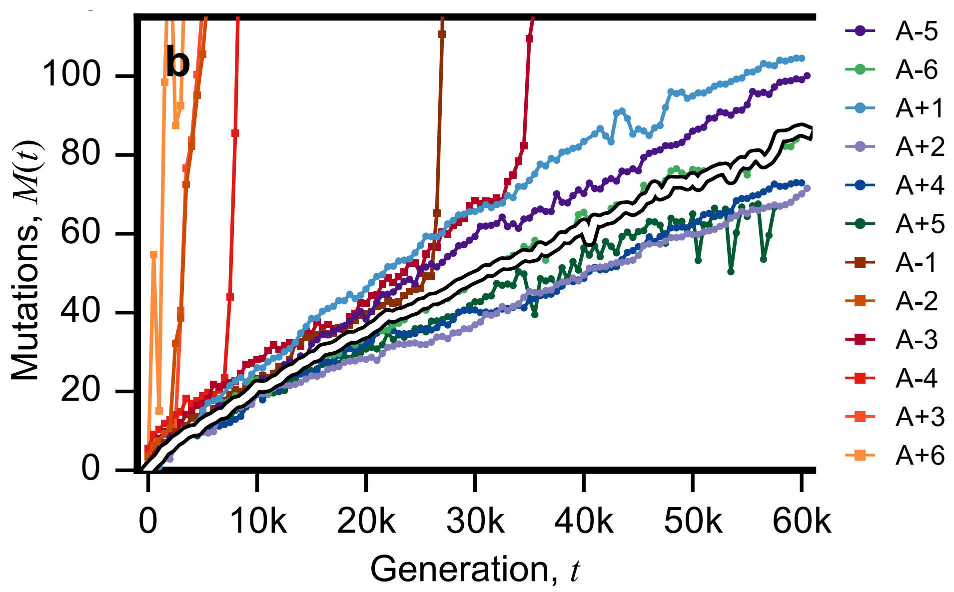
Good, McDonald* et al (Nature, 2017)*

Dynamics of molecular evolution over 60,000 generations in 12 replicate populations



Good, McDonald* et al (Nature, 2017)*

Dynamics of molecular evolution over 60,000 generations

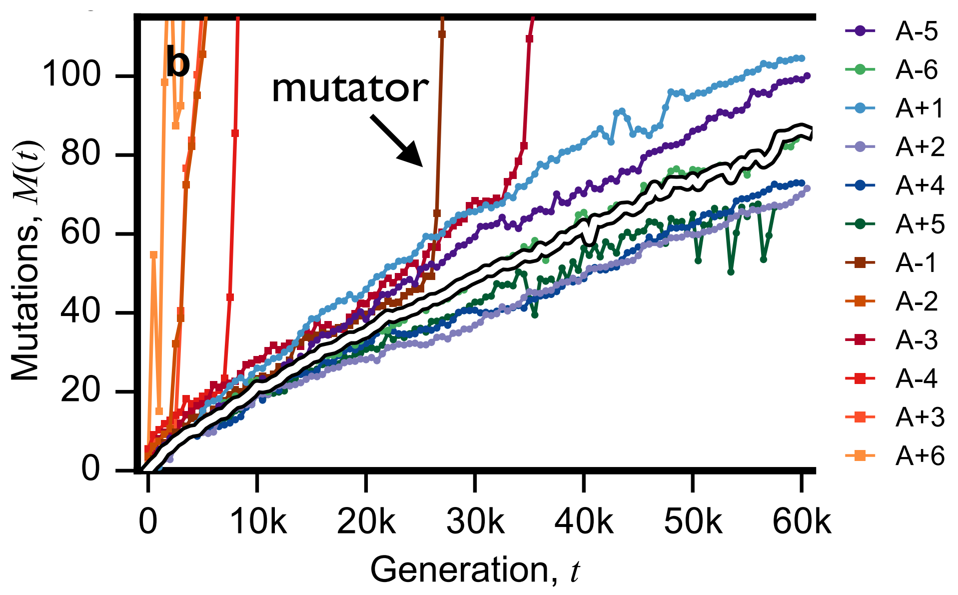


Avg # mutations / cell

$$M(t) \equiv \sum f_i(t)$$

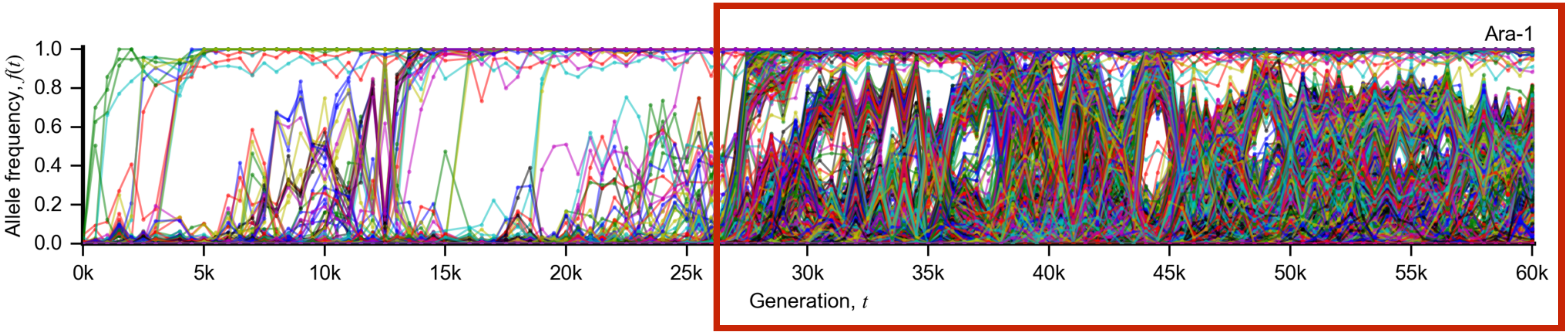
Good, McDonald* et al (Nature, 2017)*

Dynamics of molecular evolution over 60,000 generations



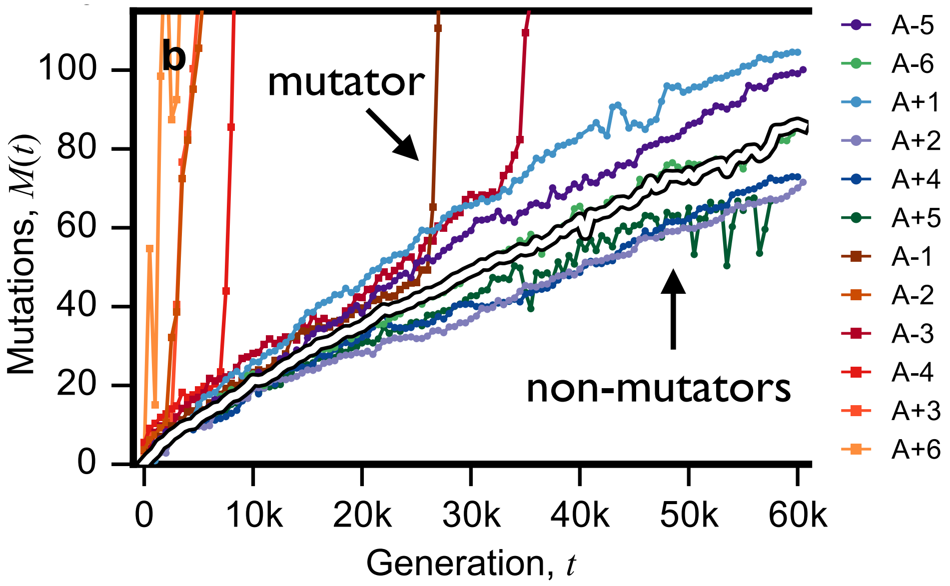
Avg # mutations / cell

$$M(t) \equiv \sum f_i(t)$$



Good, McDonald* et al (Nature, 2017)*

Dynamics of molecular evolution over 60,000 generations

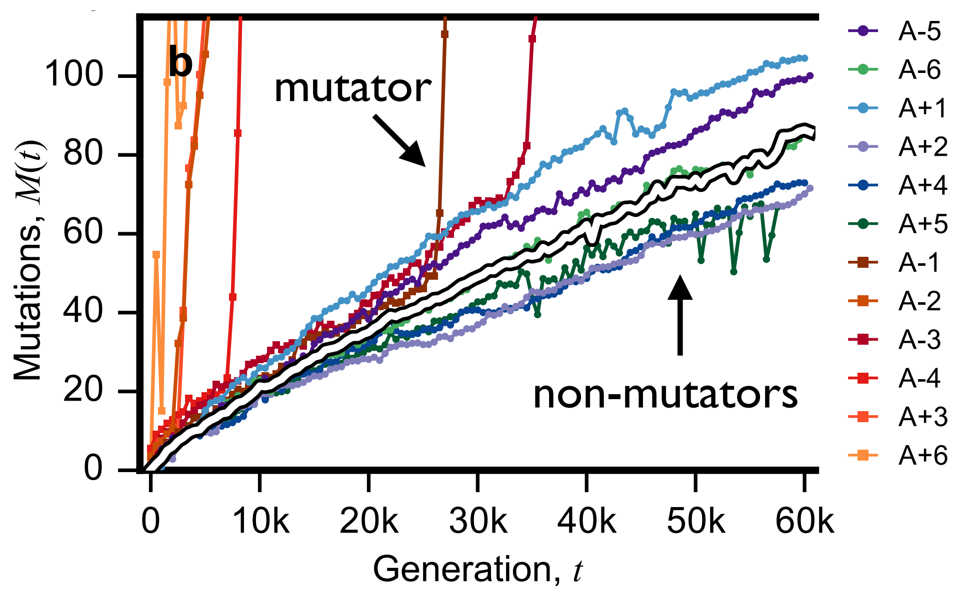


Avg # mutations / cell

$$M(t) \equiv \sum f_i(t)$$

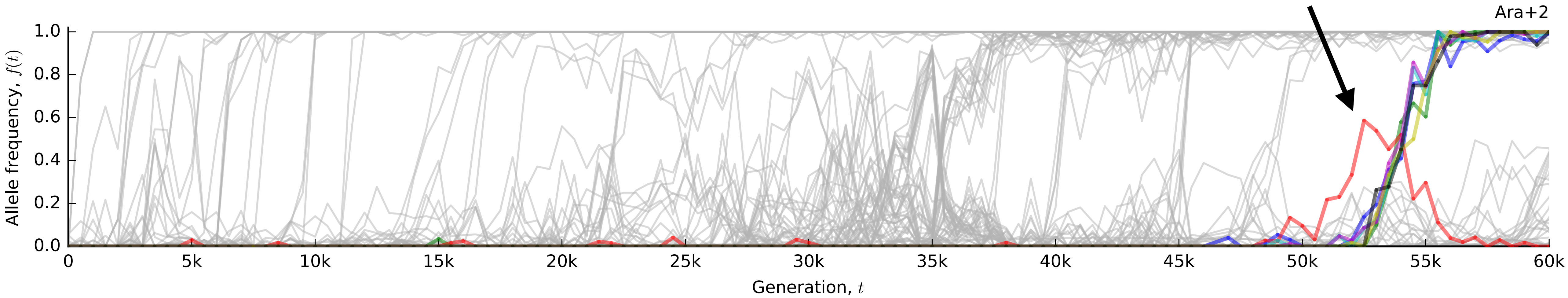
Good, McDonald* et al (Nature, 2017)*

Dynamics of molecular evolution over 60,000 generations



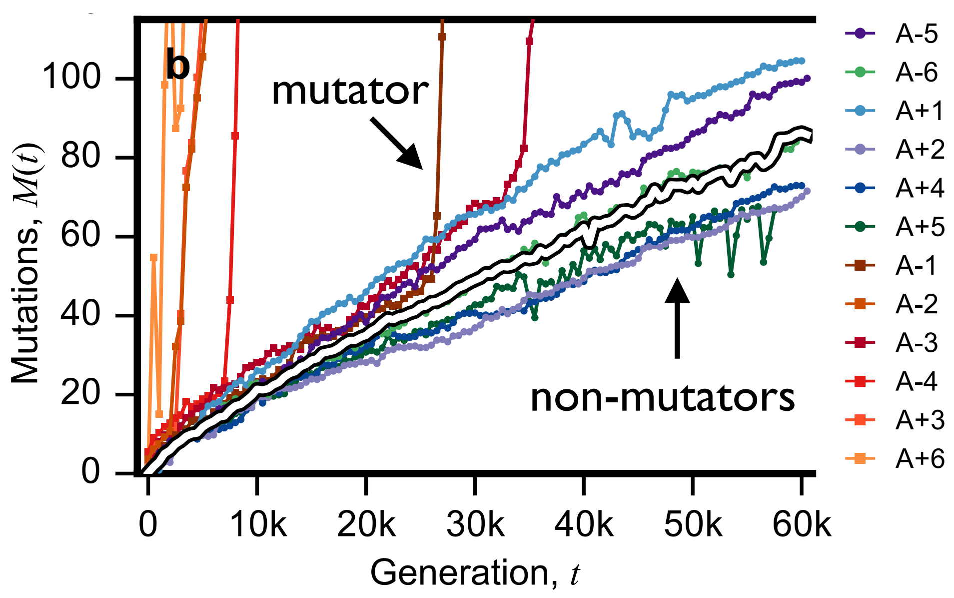
Avg # mutations / cell

$$M(t) \equiv \sum f_i(t)$$



Good, McDonald* et al (Nature, 2017)*

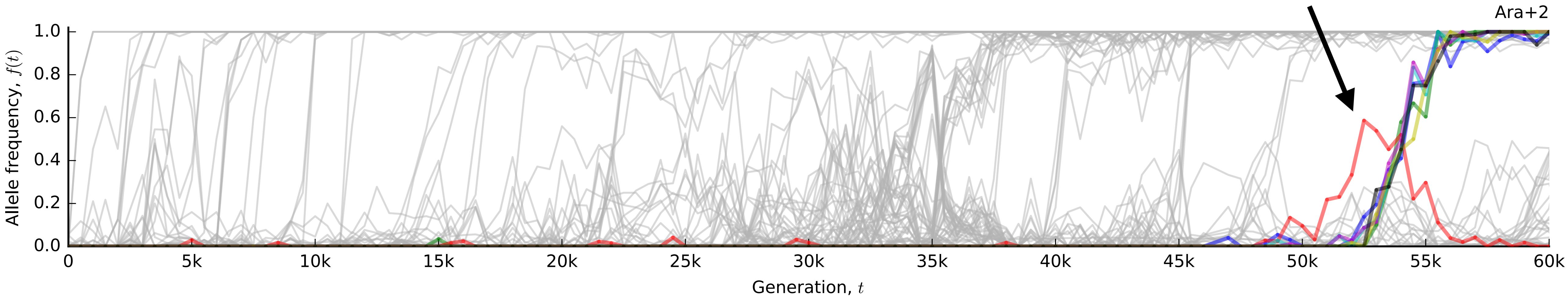
Dynamics of molecular evolution over 60,000 generations



Avg # mutations / cell

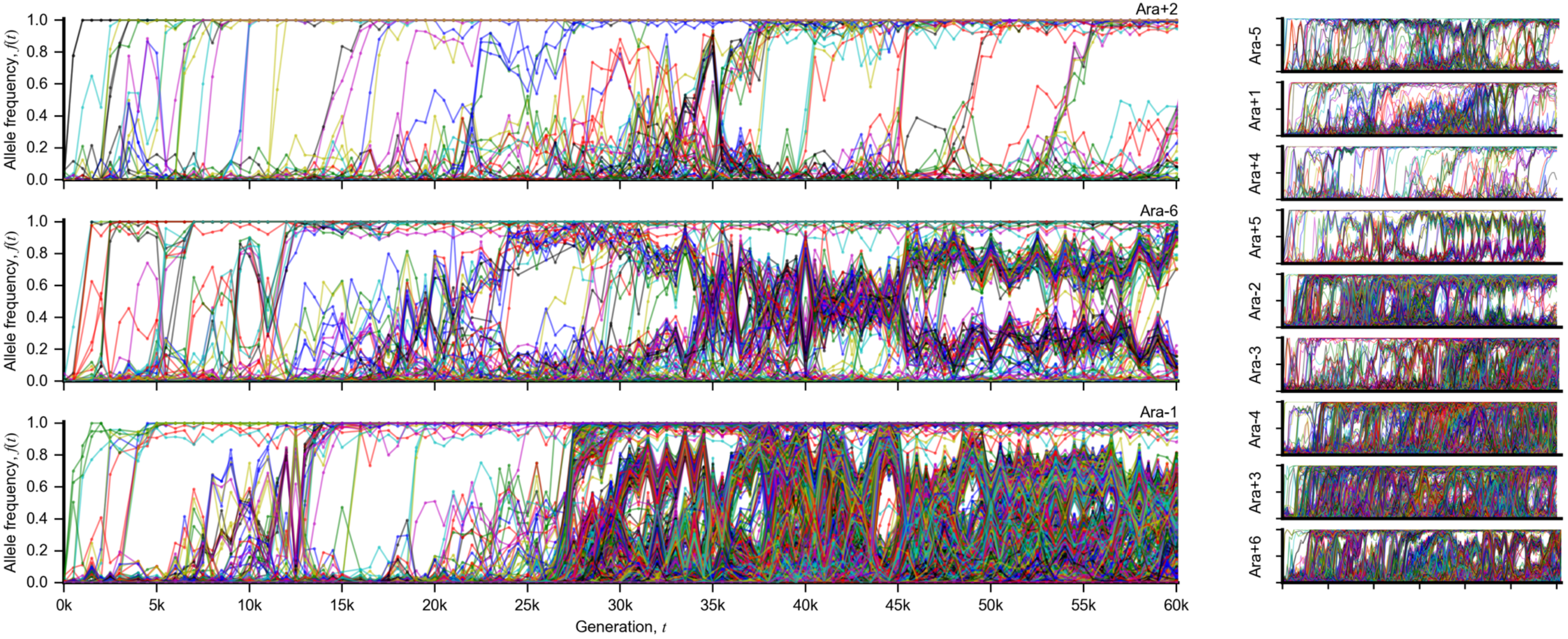
$$M(t) \equiv \sum f_i(t)$$

Adaptation ***still*** proceeds at rapid pace; linkage important!



Good, McDonald* et al (Nature, 2017)*

…but something is weird

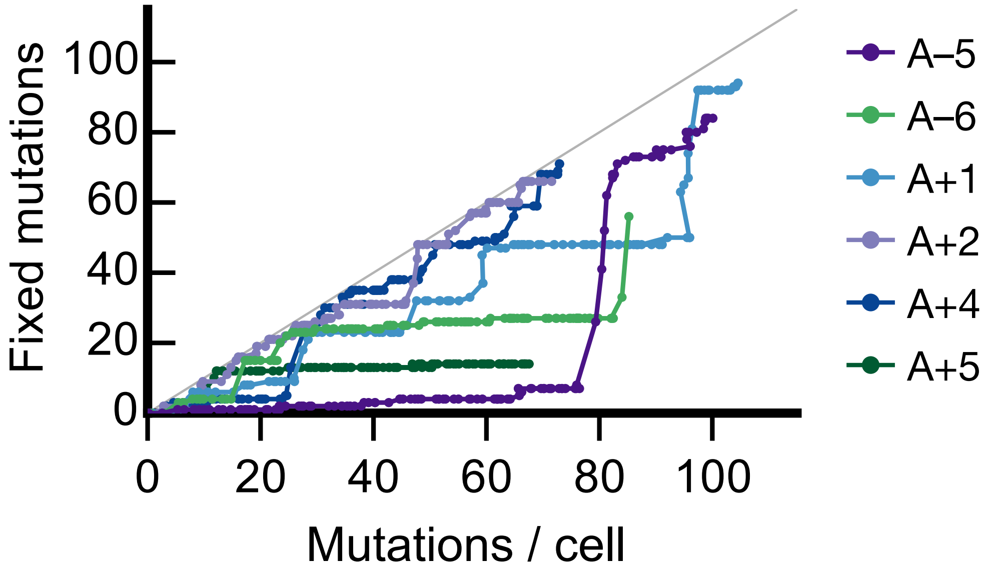


Good, McDonald* et al (Nature, 2017)*

Rapid evolution, but deficit of fixed mutations

Model predicts:

“fixed” mutations ~ avg # mutations / cell



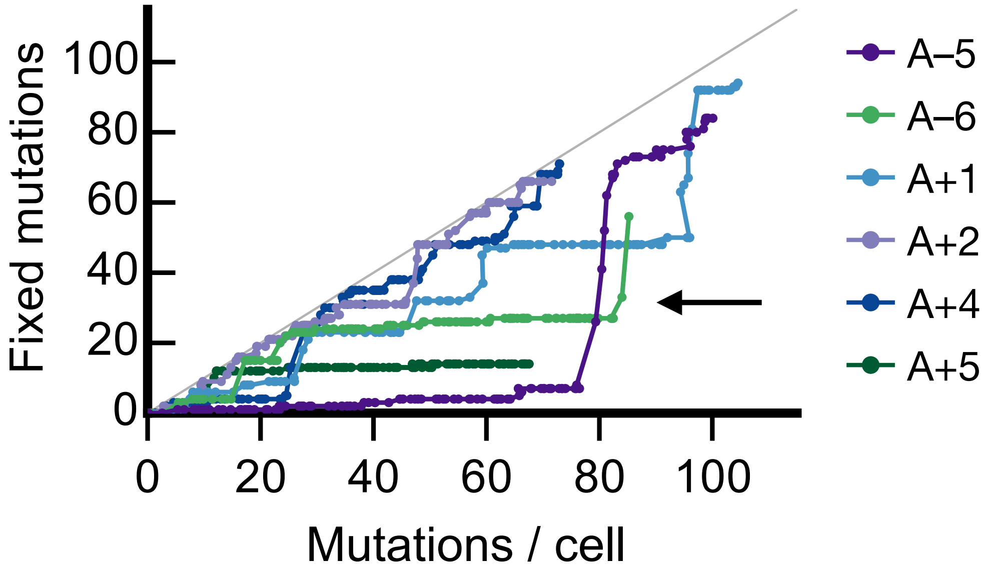
Good, McDonald* et al (Nature, 2017)*

Rapid evolution, but deficit of fixed mutations

Model predicts:

"fixed" mutations ~ avg # mutations / cell

Too few fixed muts!



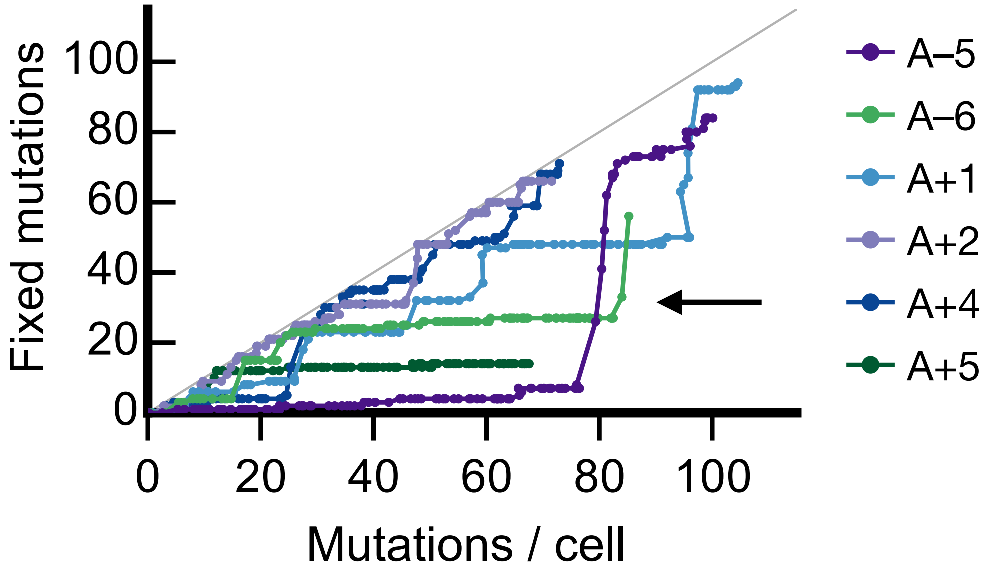
Good, McDonald* et al (Nature, 2017)*

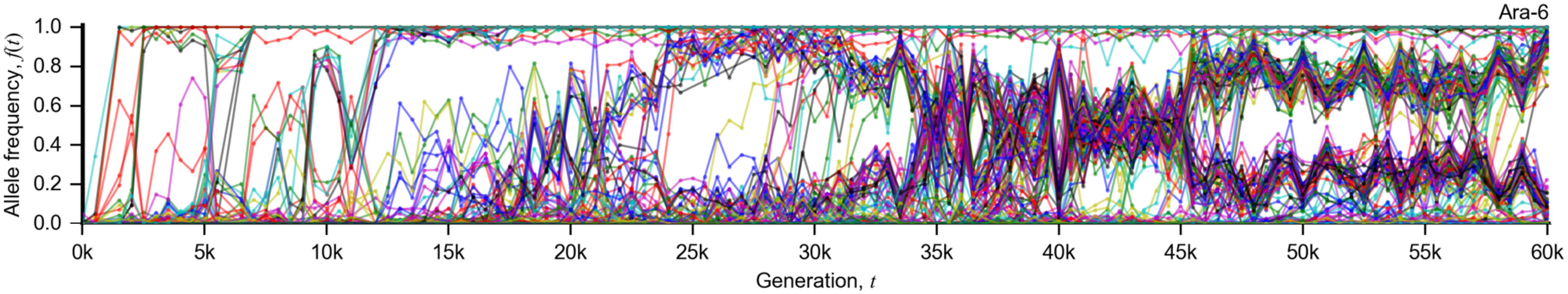
Rapid evolution, but deficit of fixed mutations

Model predicts:

"fixed" mutations $\sim$ avg # mutations / cell

Too few fixed muts!

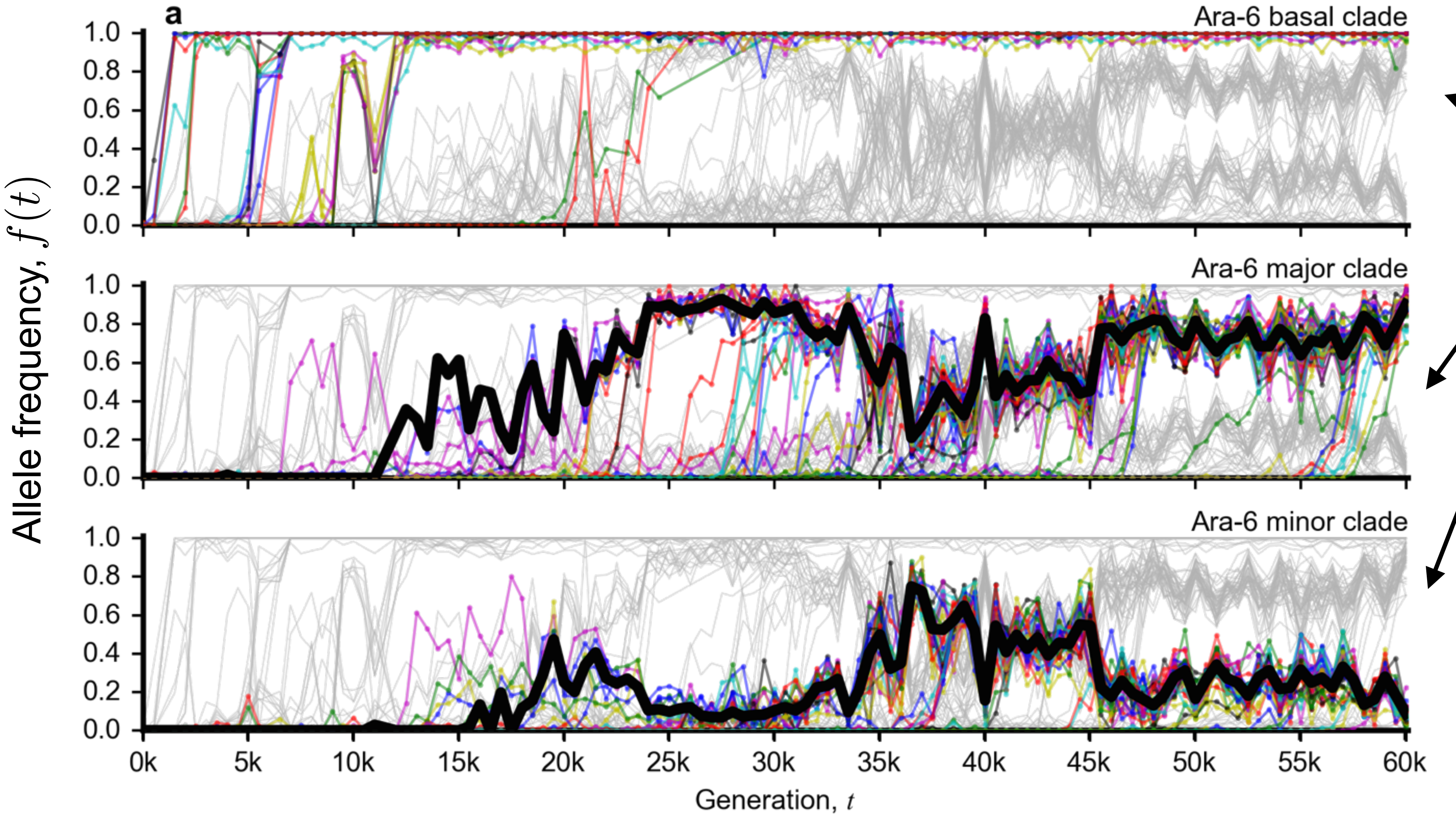




Mutations "pile up" in 2 clades that coexist for $\gtrsim 10,000$ ***gens***

Good, McDonald* et al (Nature, 2017)*

Inferring clade dynamics with hidden Markov models

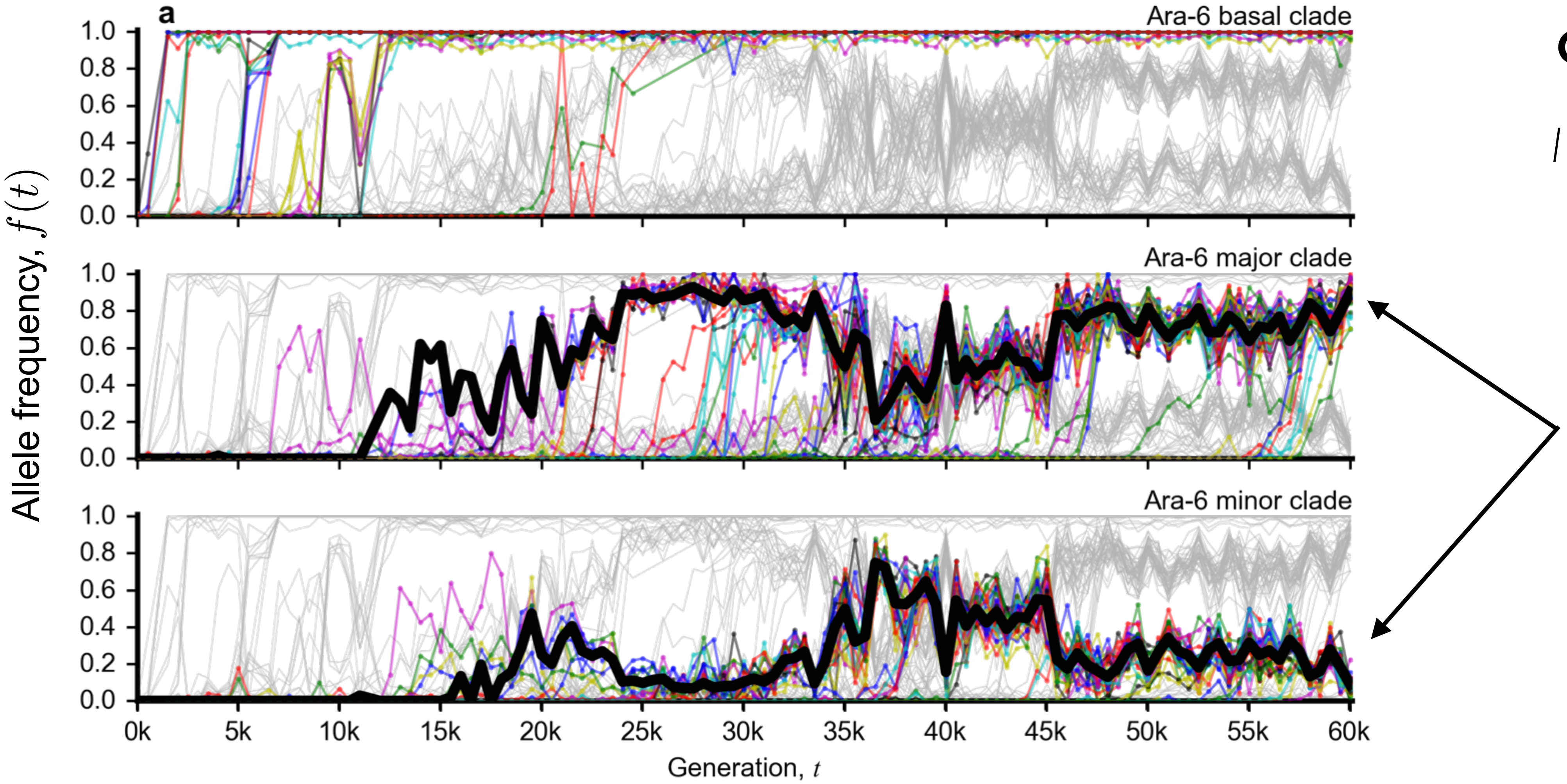


CladeHMM

1. Classifies mutations into clades

Good, McDonald* et al (Nature, 2017)*

Inferring clade dynamics with hidden Markov models



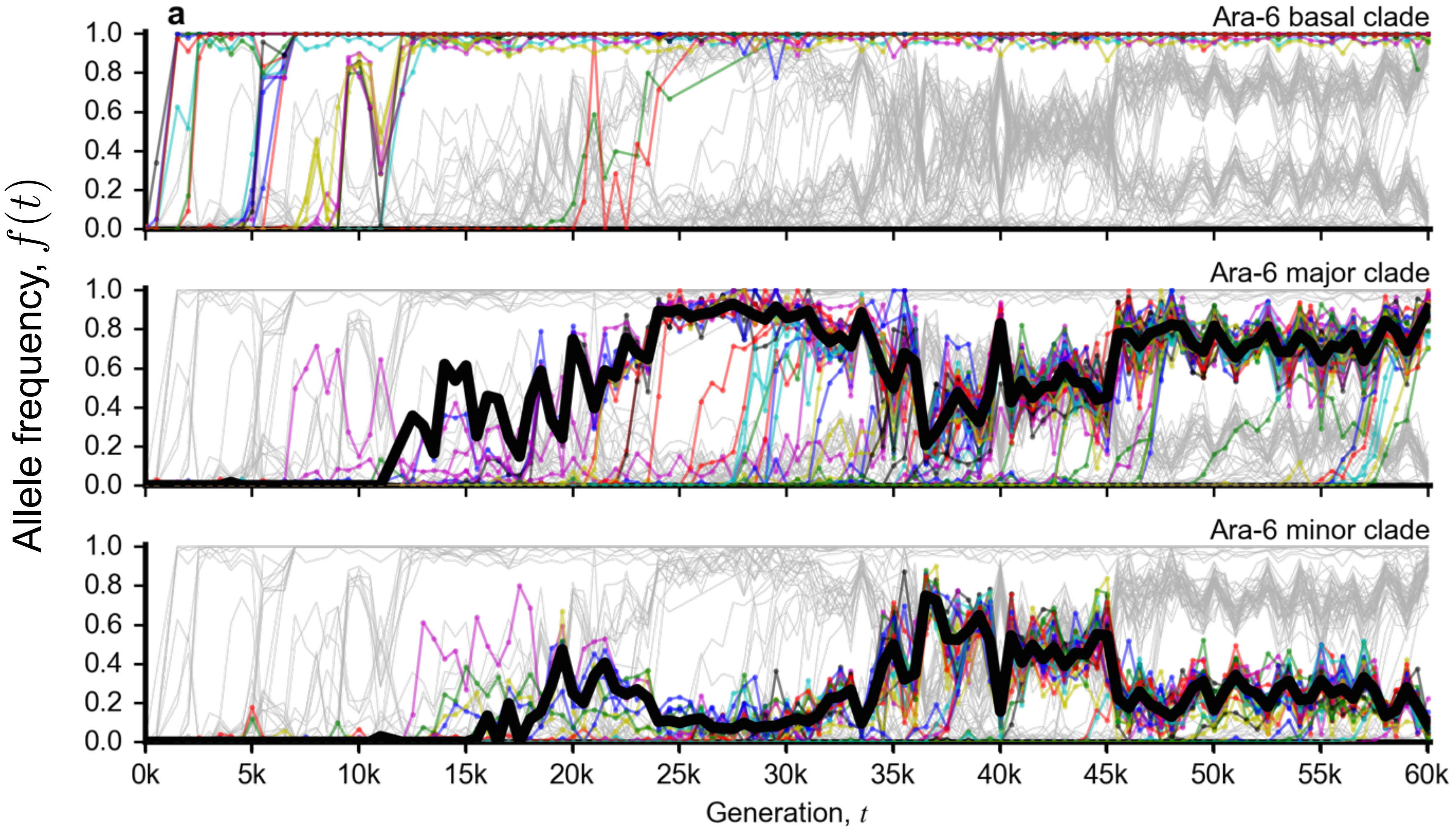
CladeHMM

1. Classifies mutations into clades

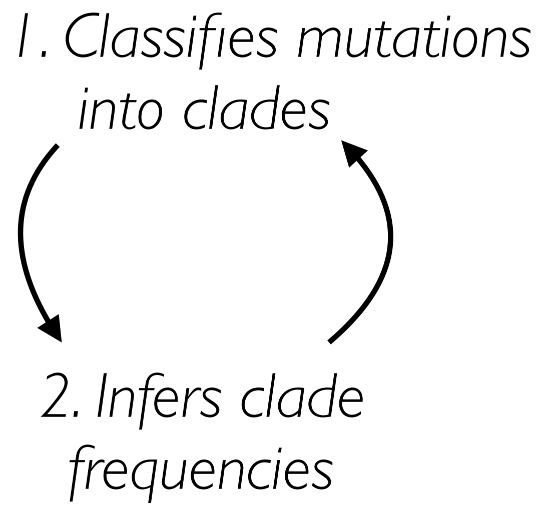
2. Infers clade frequencies

Good, McDonald* et al (Nature, 2017)*

Inferring clade dynamics with hidden Markov models



CladeHMM

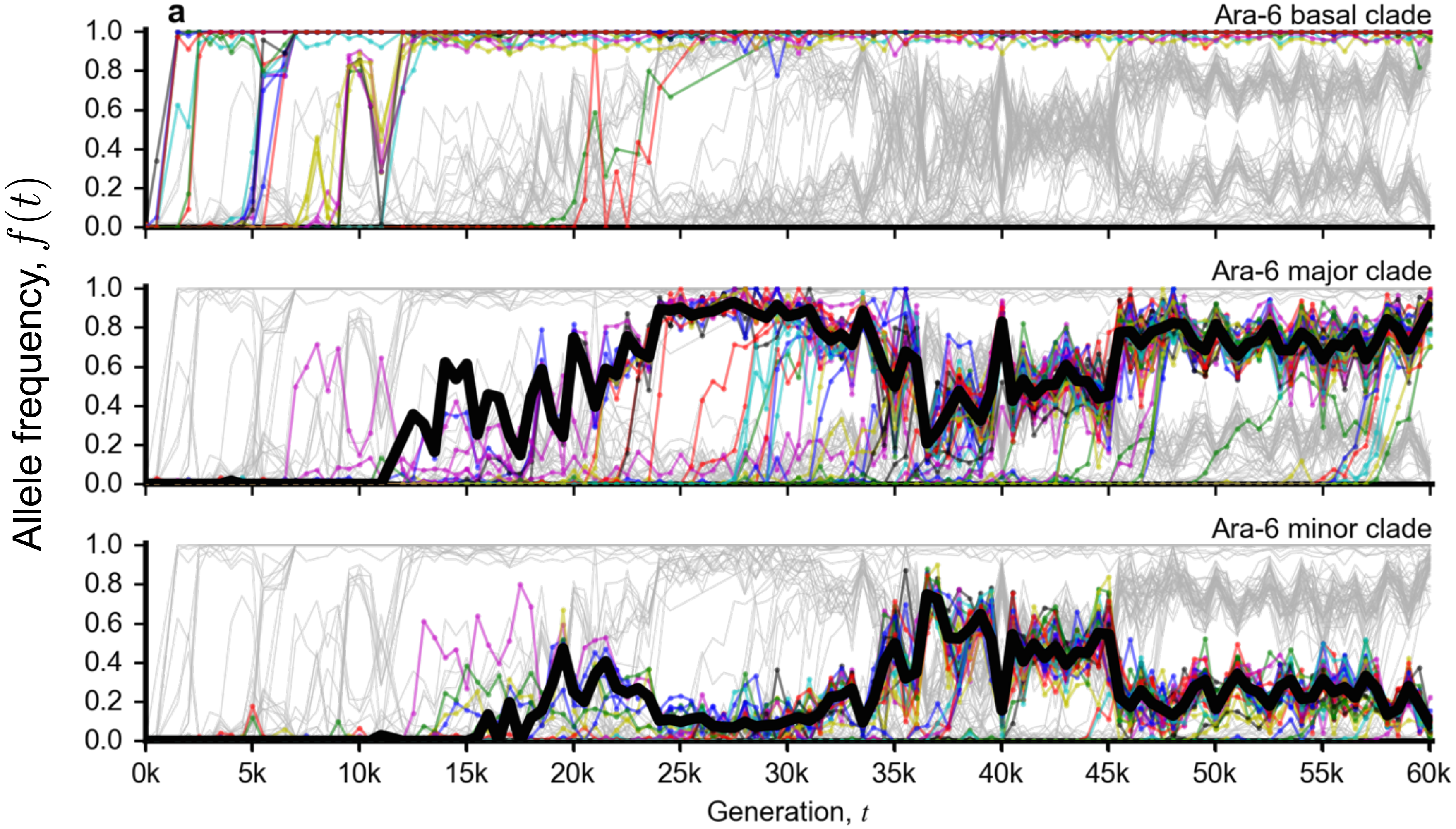


1. Classifies mutations into clades

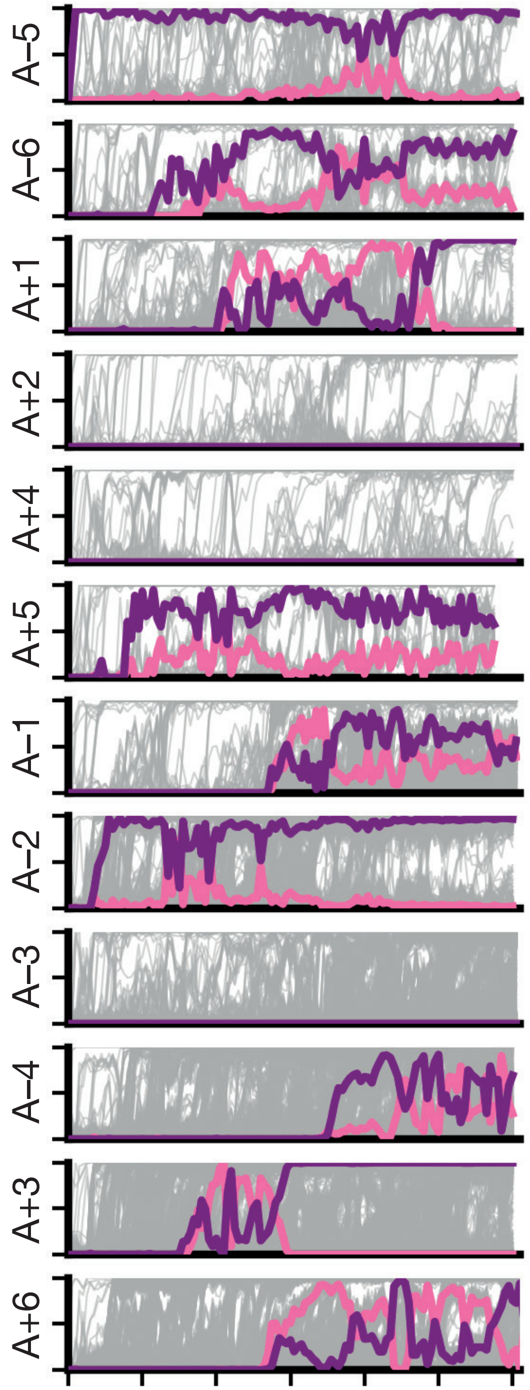
2. Infers clade frequencies

Good, McDonald* et al (Nature, 2017)*

Inferring clade dynamics with hidden Markov models

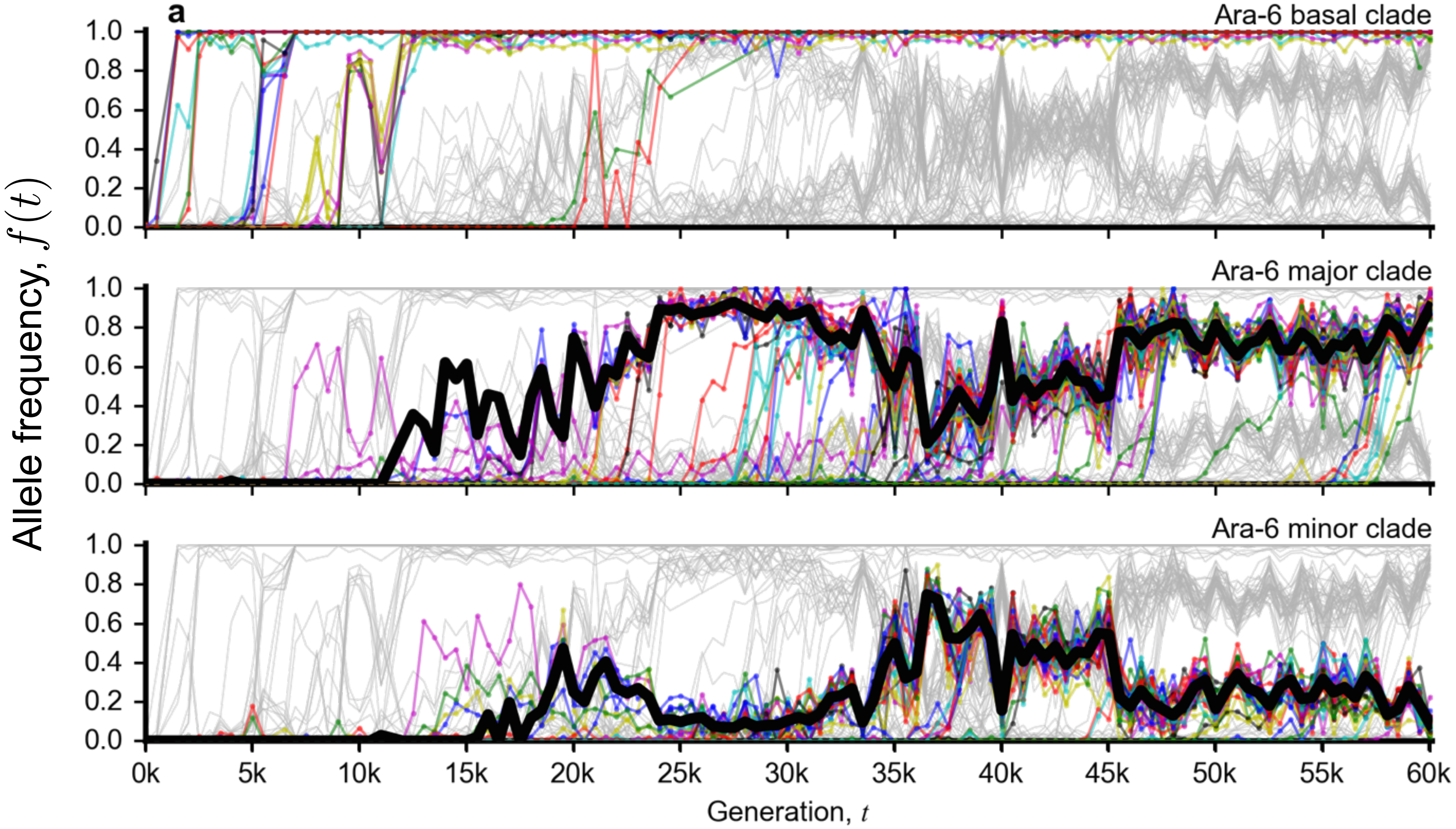


All 12 pop'ns

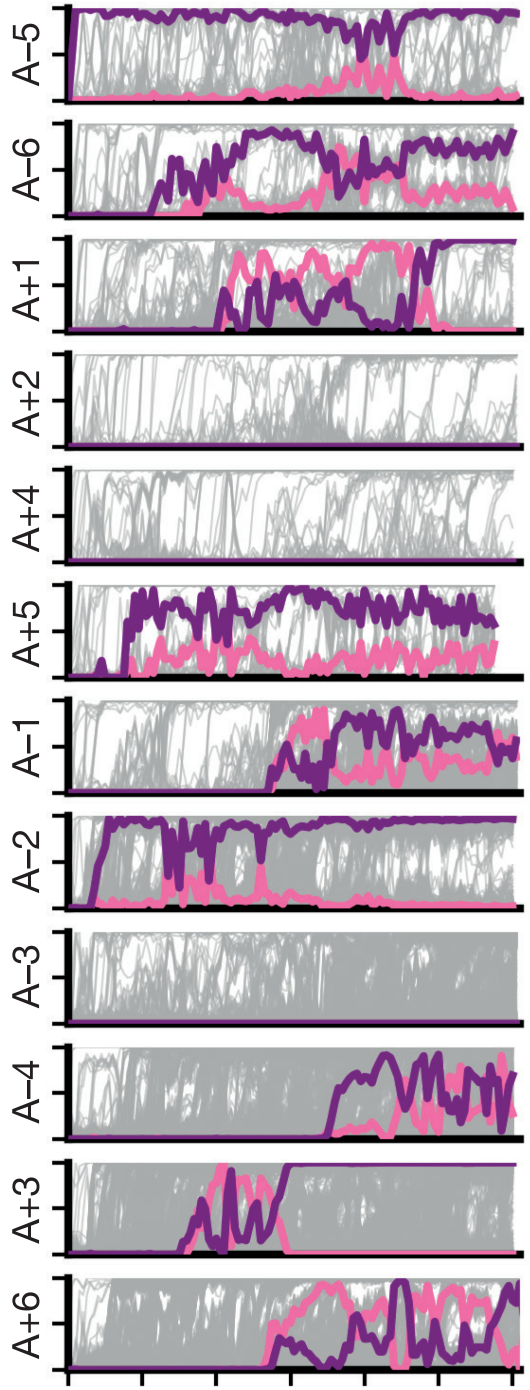


Good, McDonald* et al (Nature, 2017)*

Inferring clade dynamics with hidden Markov models



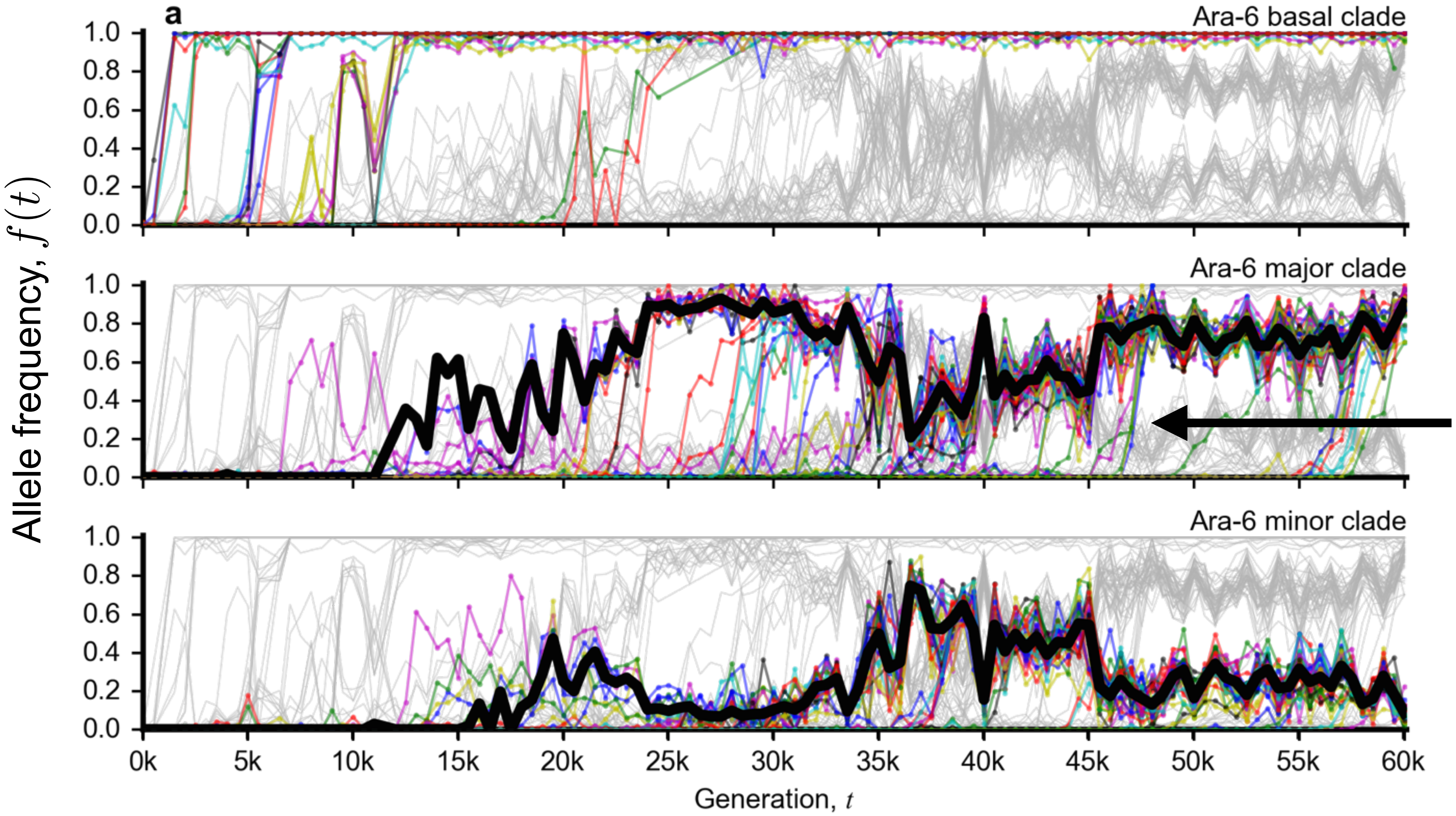
All 12 pop'ns



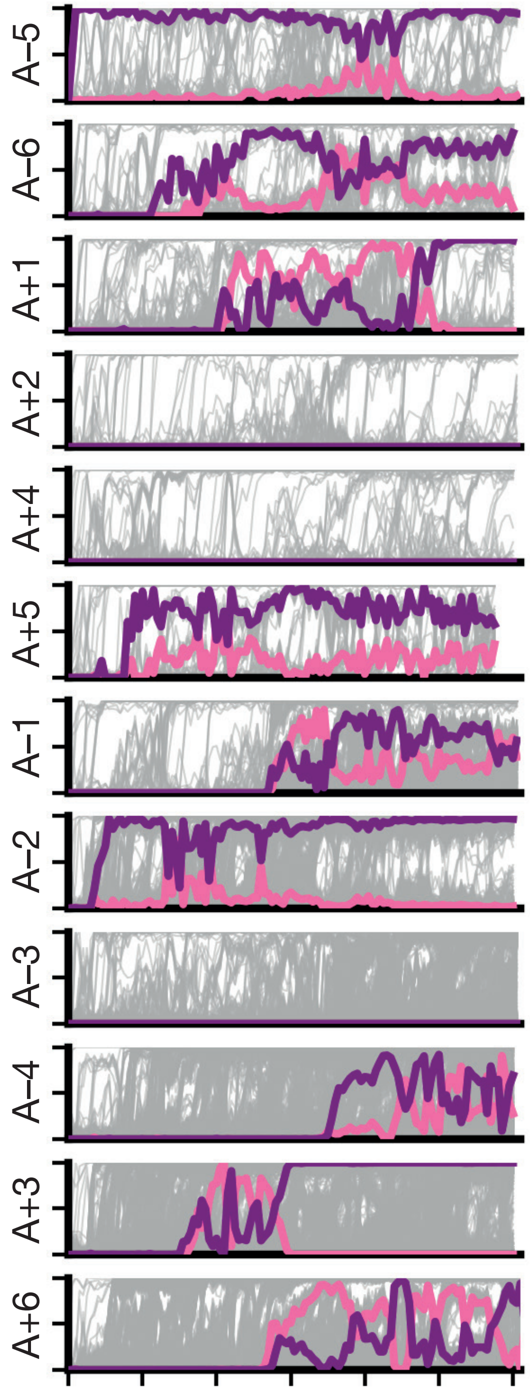
1. Long-lived clades are ***common*** in the LTEE.

Good, McDonald* et al (Nature, 2017)*

Inferring clade dynamics with hidden Markov models



All 12 pop'ns

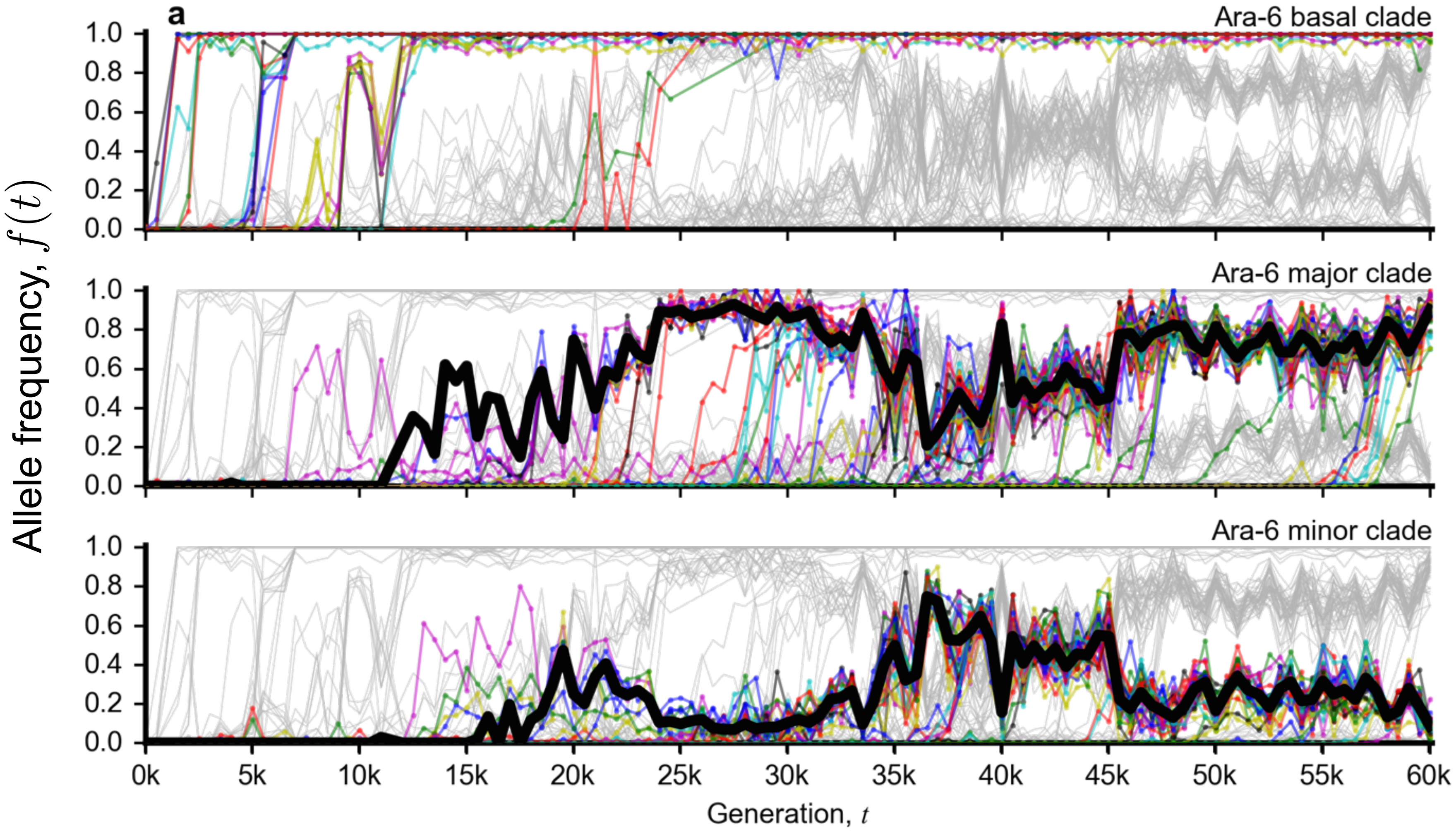


2. Mutations continue to fix ***within*** clades.

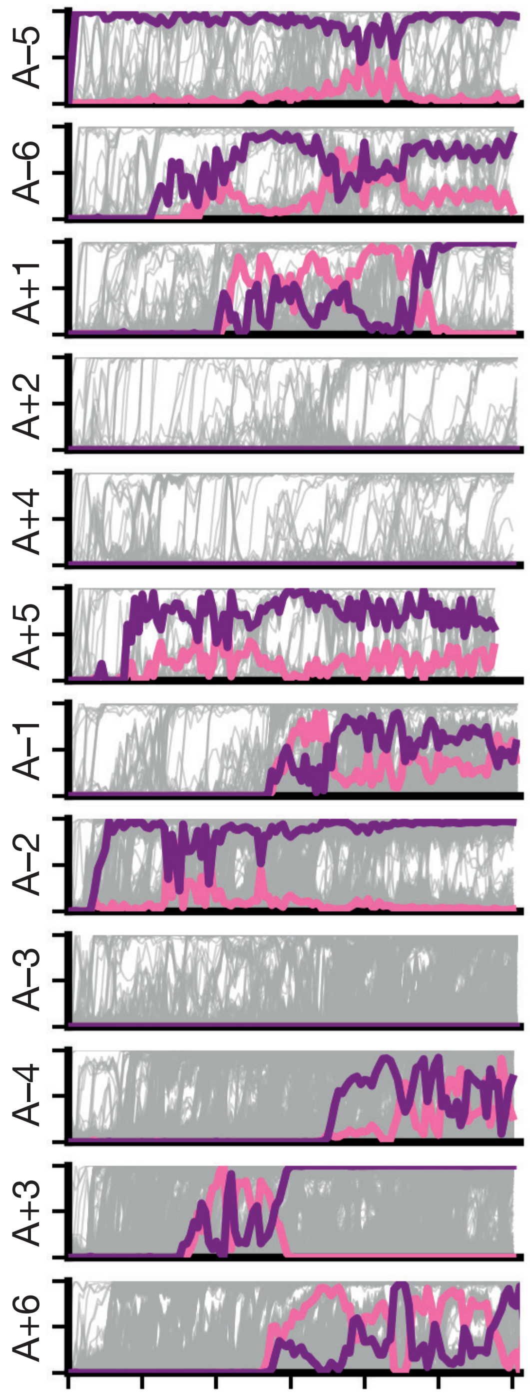
$$\tau_{\text{within}} \ll \tau_{\text{clade}}$$

Good, McDonald* et al (Nature, 2017)*

Inferring clade dynamics with hidden Markov models



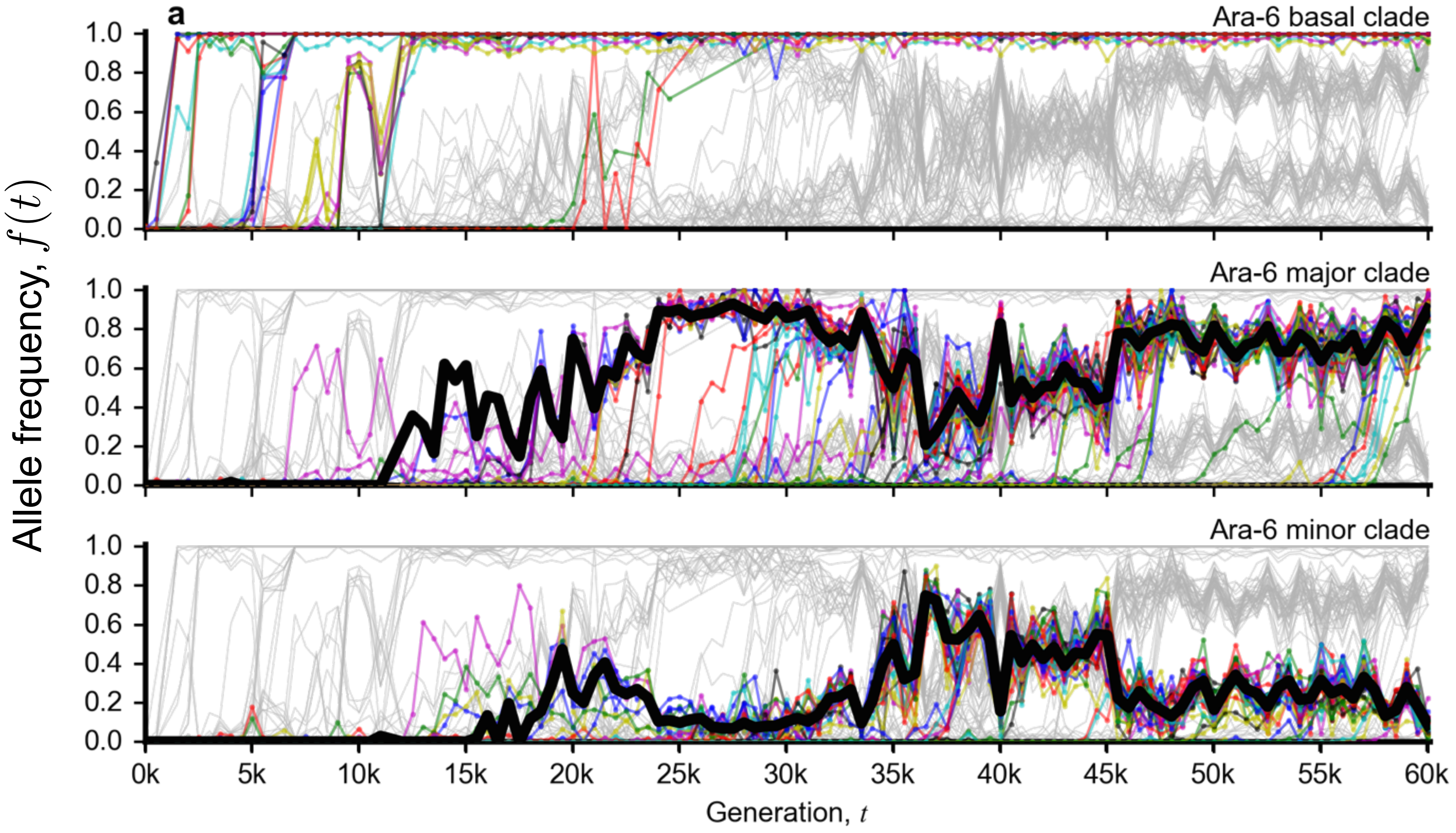
All 12 pop'ns



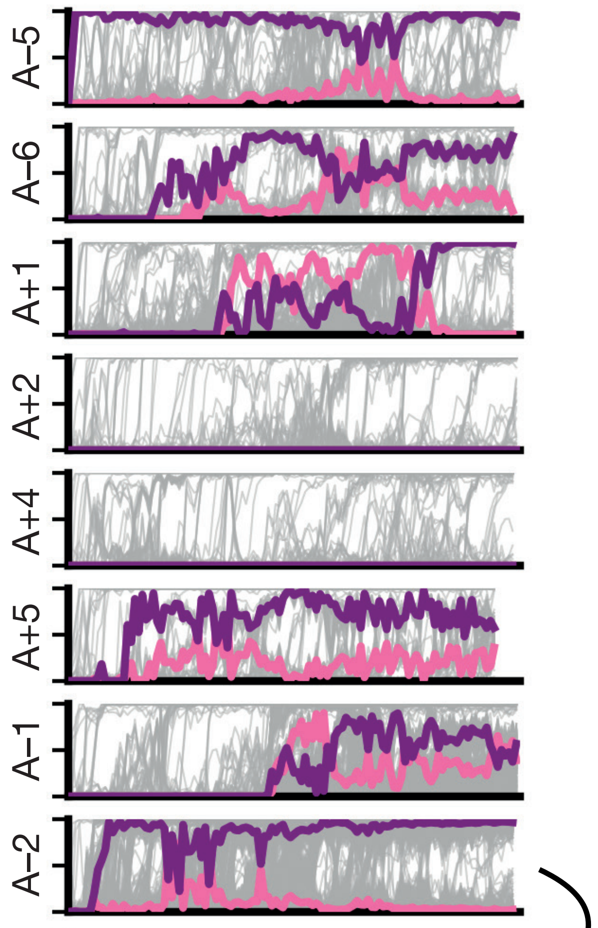
3. Separation of timescales ***inconsistent*** with null model of adaptation with linkage

Good, McDonald* et al (Nature, 2017)*

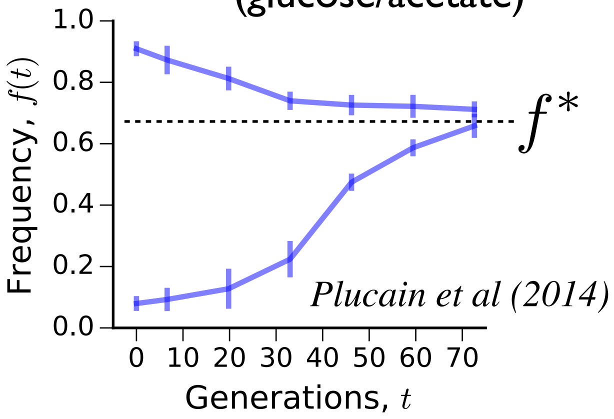
Inferring clade dynamics with hidden Markov models



All 12 pop'ns



Stable coexistence
(glucose/acetate)

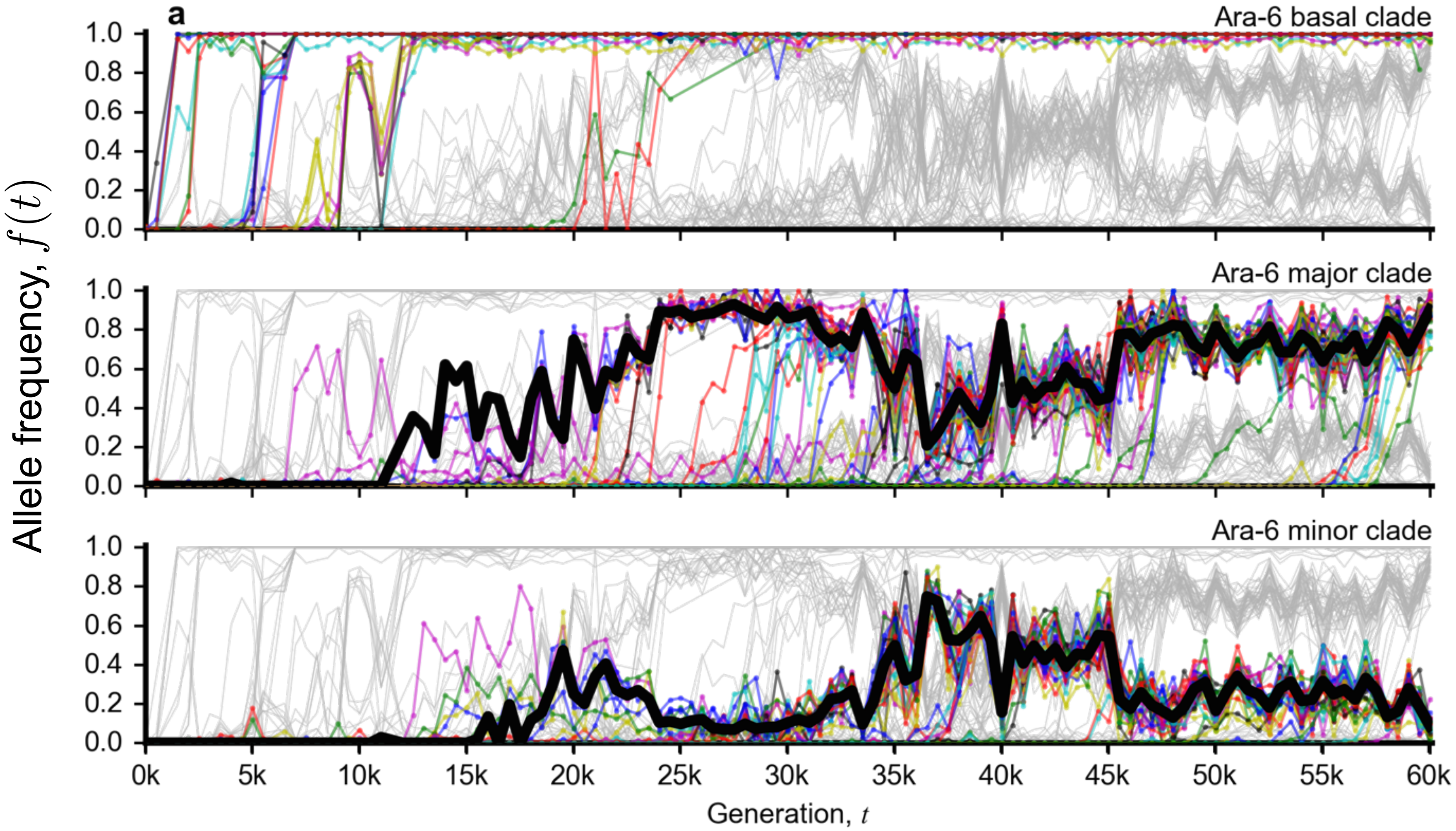


Plucain et al (2014)

3. Separation of timescales ***inconsistent*** with null model of adaptation with linkage

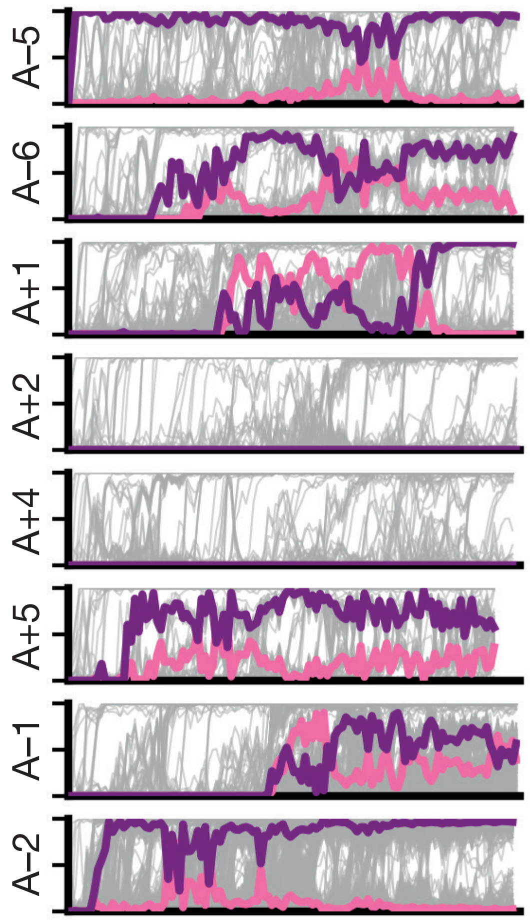
Good, McDonald* et al (Nature, 2017)*

Inferring clade dynamics with hidden Markov models

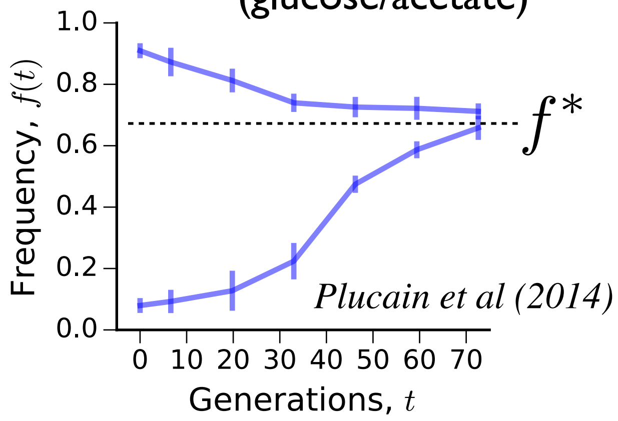


4. But ***quasi-stable*** on evolutionary timescales

All 12 pop'ns



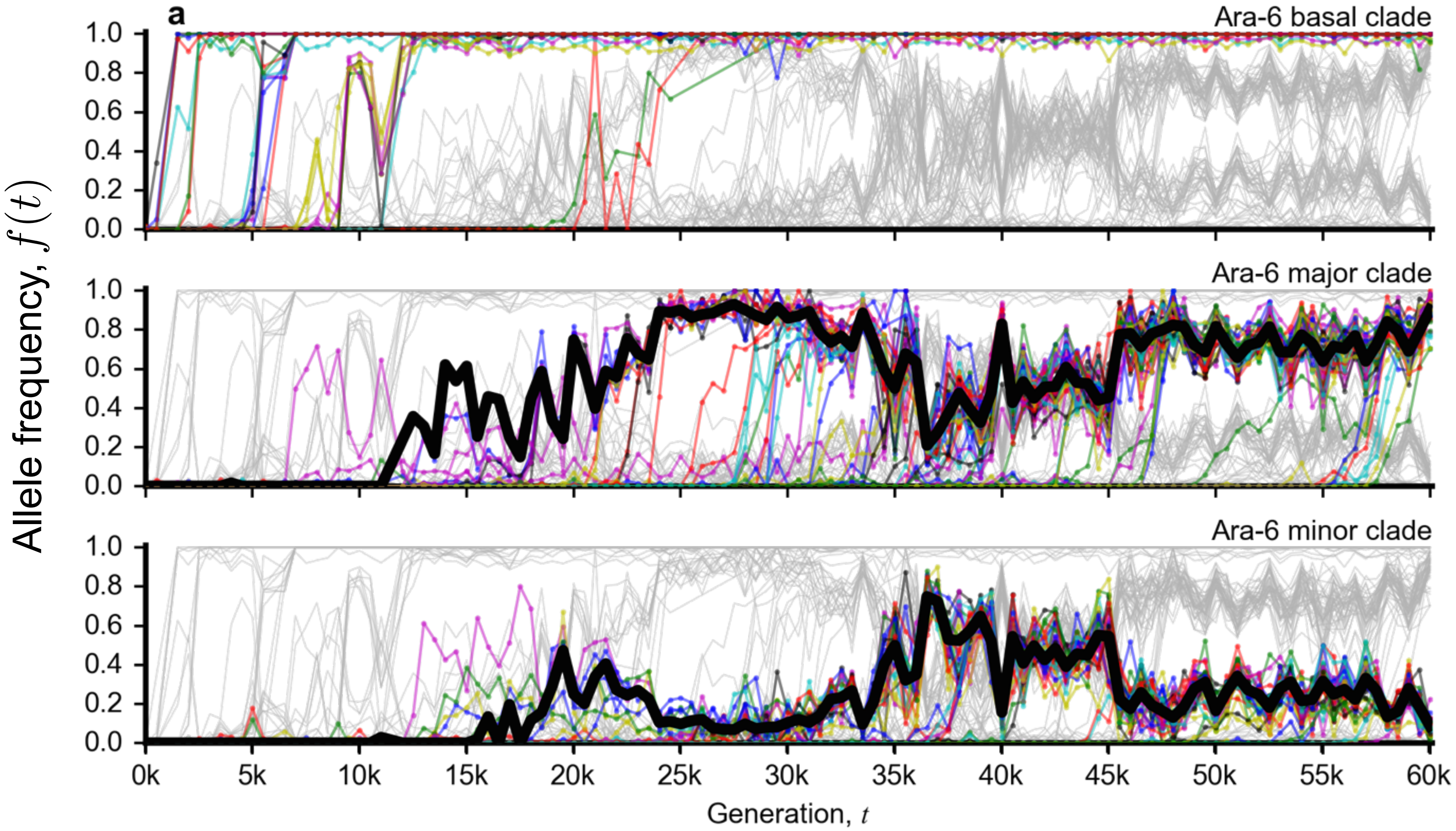
Stable coexistence
(glucose/acetate)



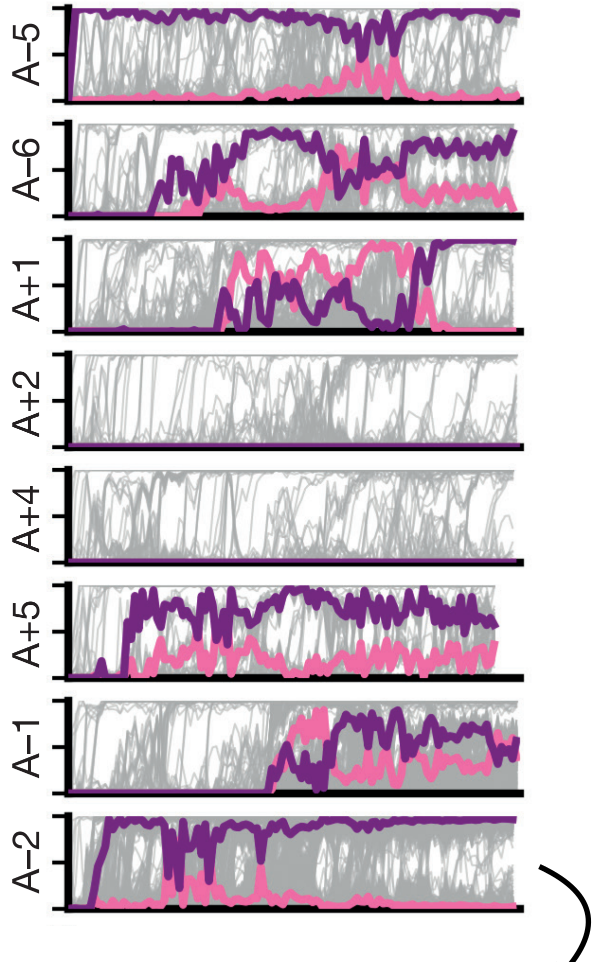
Plucain et al (2014)

Good, McDonald* et al (Nature, 2017)*

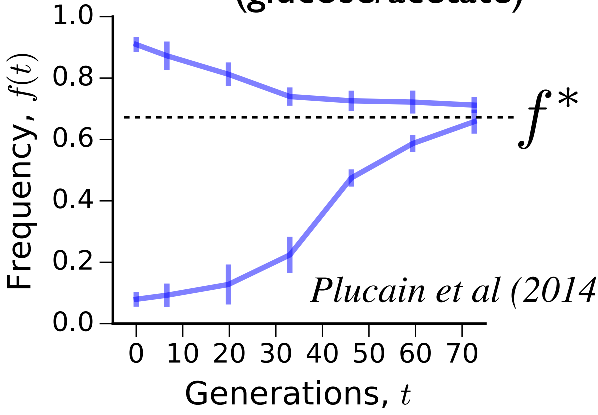
Inferring clade dynamics with hidden Markov models



All 12 pop'ns



Stable coexistence
(glucose/acetate)

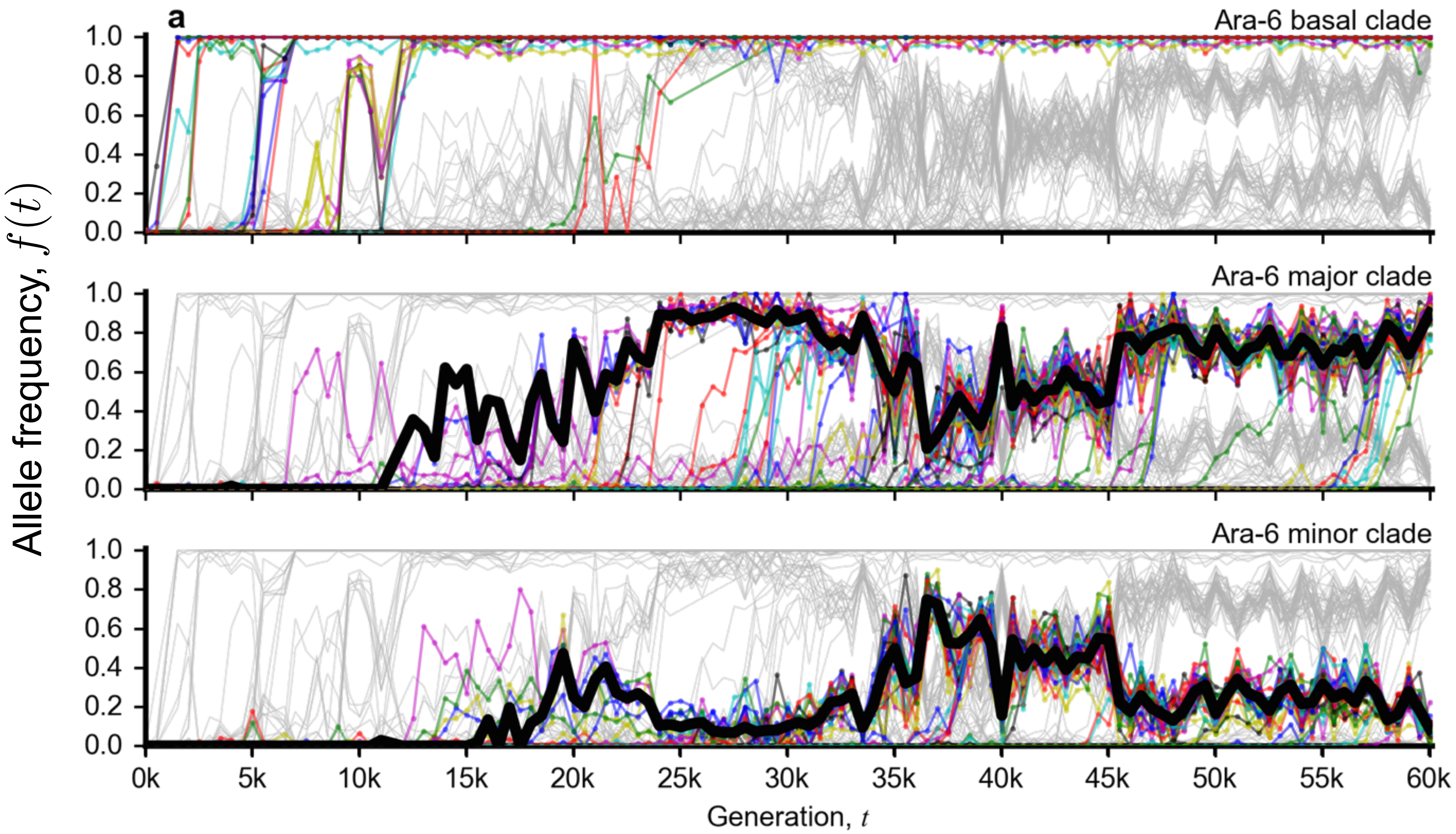


Plucain et al (2014)

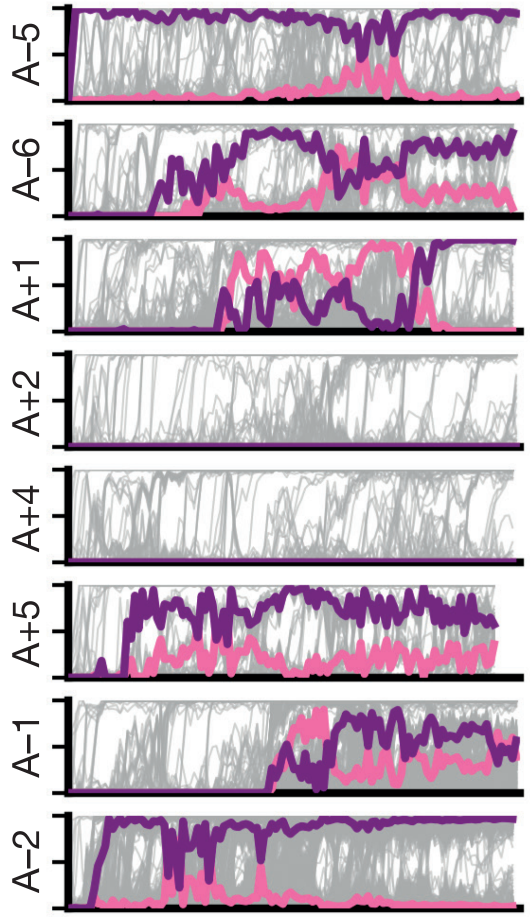
5. High variability in clade freqs across replicates (both at endpoint and over time)

Good, McDonald* et al (Nature, 2017)*

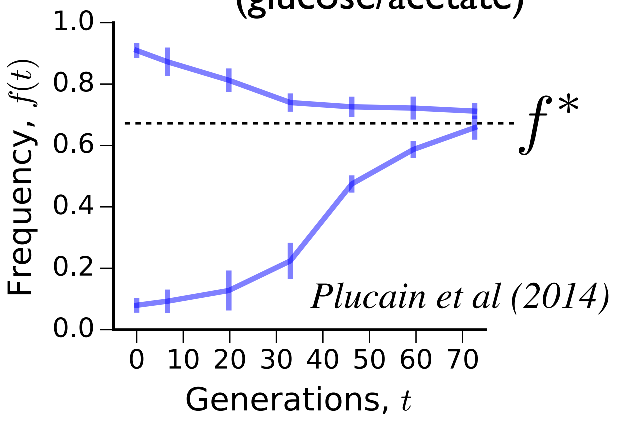
Inferring clade dynamics with hidden Markov models



All 12 pop'ns



Stable coexistence
(glucose/acetate)

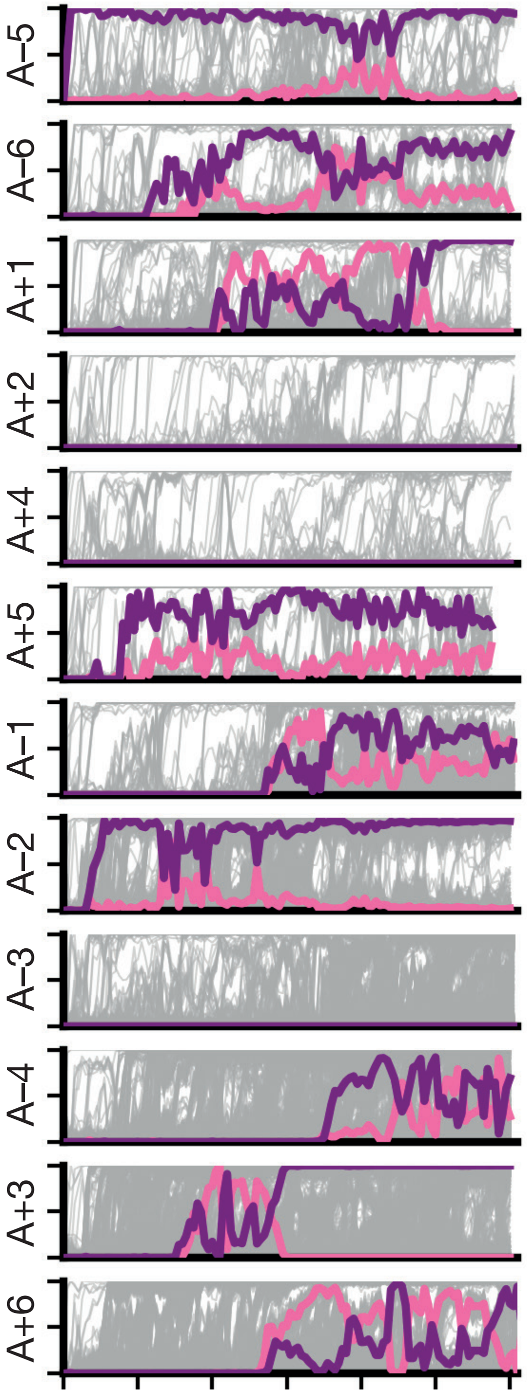


Plucain et al (2014)

Stochastic evolutionary forces play important role in shaping structure of community

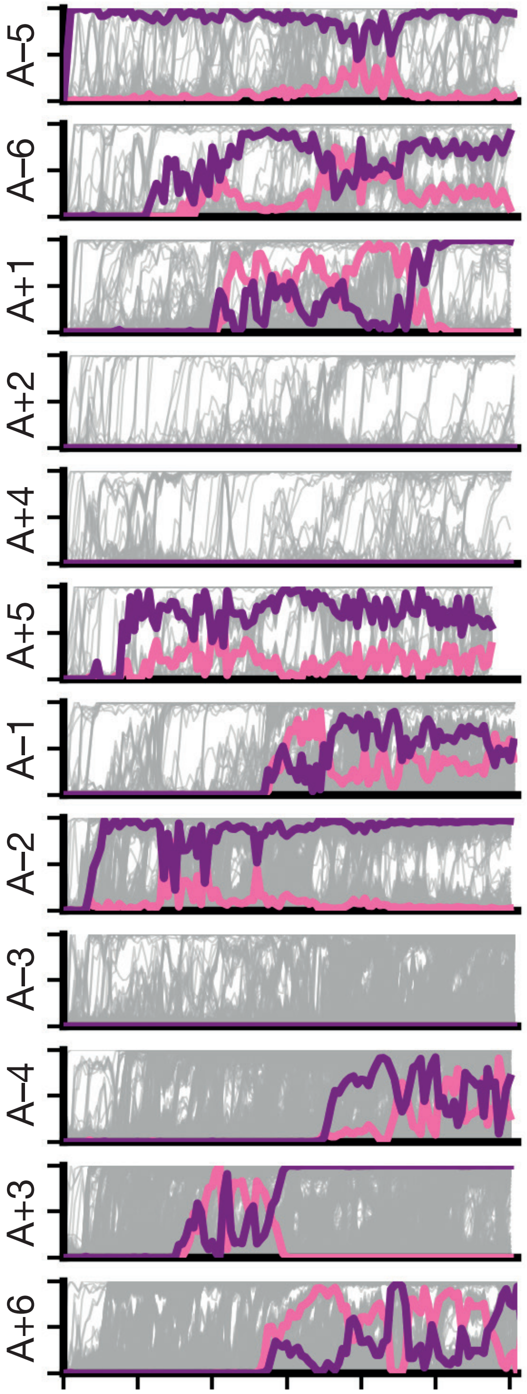
Good, McDonald* et al (Nature, 2017)*

Coupled ecological and evolutionary processes



+
other laboratory systems…

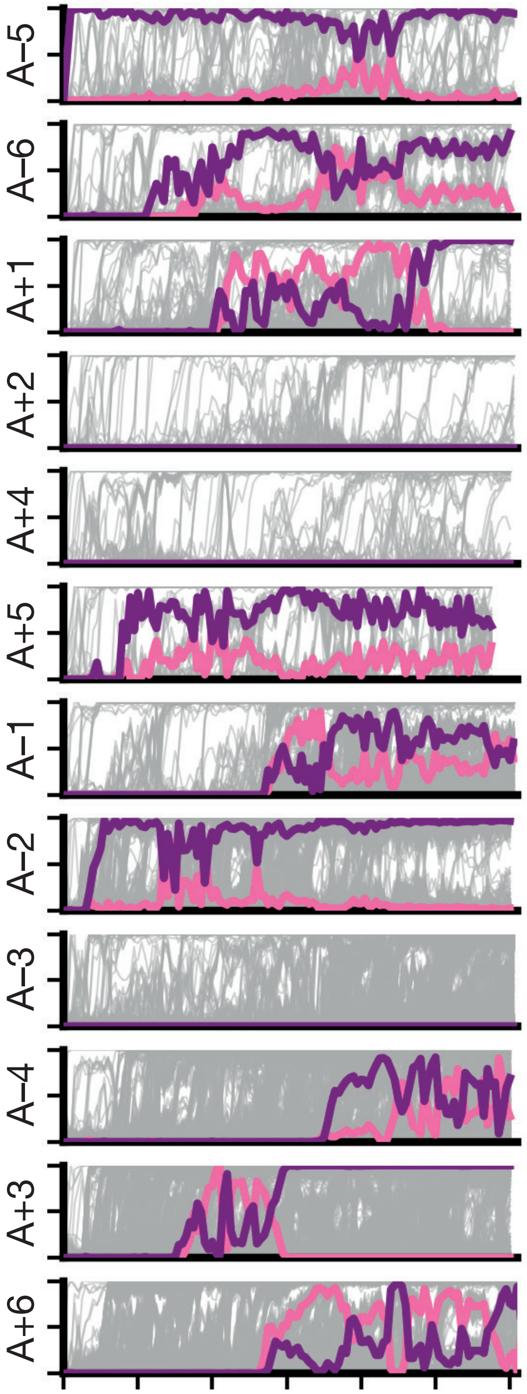
Coupled ecological and evolutionary processes



\+
other laboratory
systems…

Many open questions:

Coupled ecological and evolutionary processes

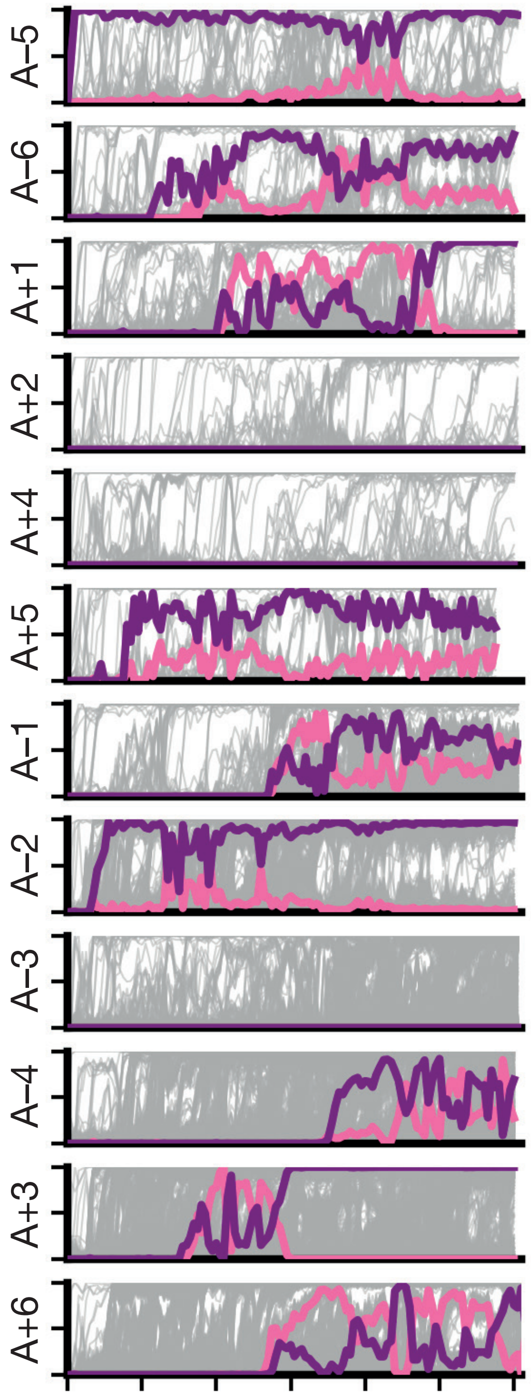


+
other laboratory systems…

Many open questions:

1. What explains common ***emergence*** , but varied ***outcomes***?

Coupled ecological and evolutionary processes



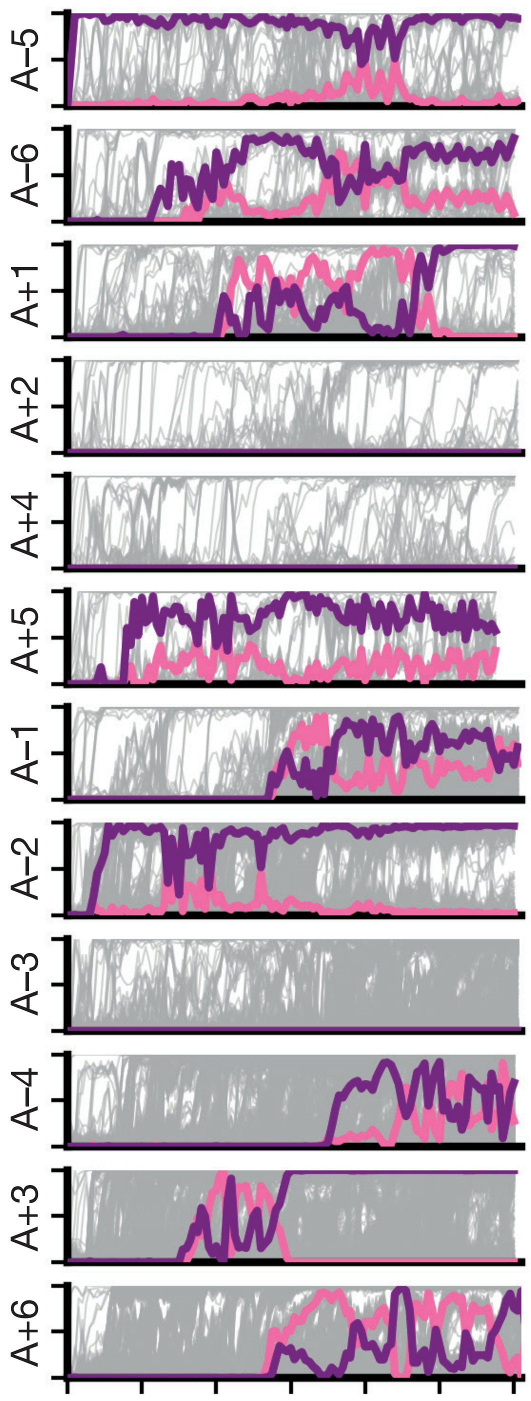
+
other laboratory systems…

Many open questions:

1. What explains common ***emergence*** , but varied ***outcomes***?

   Niche-specific adaptation?
   More general coupling between fitness and stable freq?

Coupled ecological and evolutionary processes



+
other laboratory systems…

Many open questions:

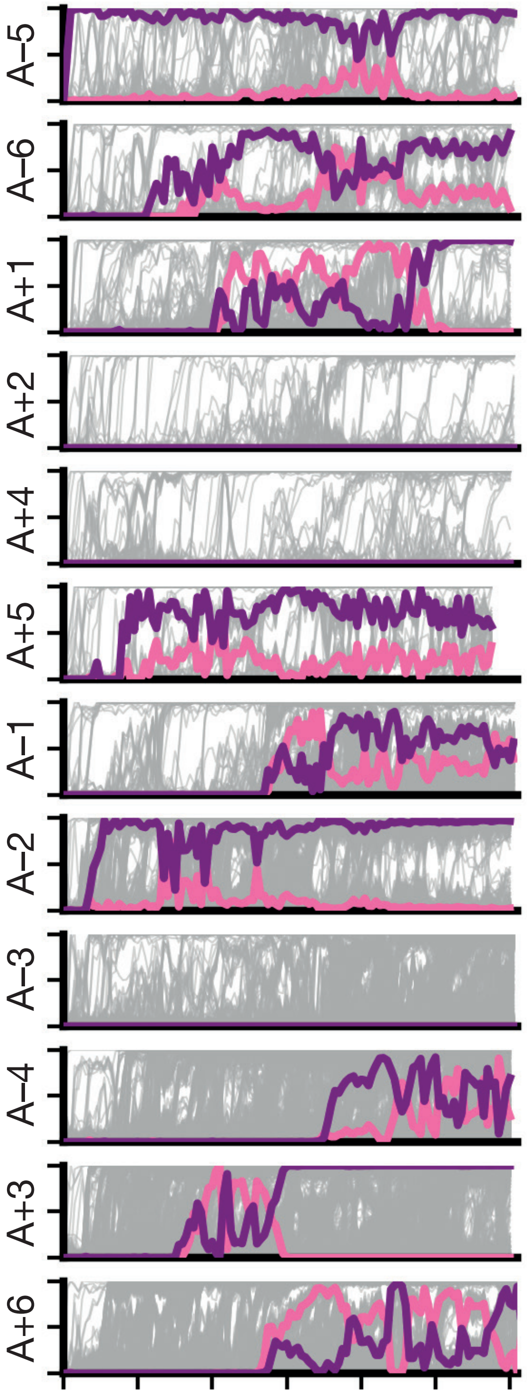
1. What explains common ***emergence*** , but varied ***outcomes***?

Niche-specific adaptation?
More general coupling between fitness and stable freq?

---

2. Do ecological interactions ***accelerate*** evolution?

Coupled ecological and evolutionary processes



+
other laboratory systems…

Many open questions:

1. What explains common ***emergence*** , but varied ***outcomes***?

Niche-specific adaptation?
More general coupling between fitness and stable freq?

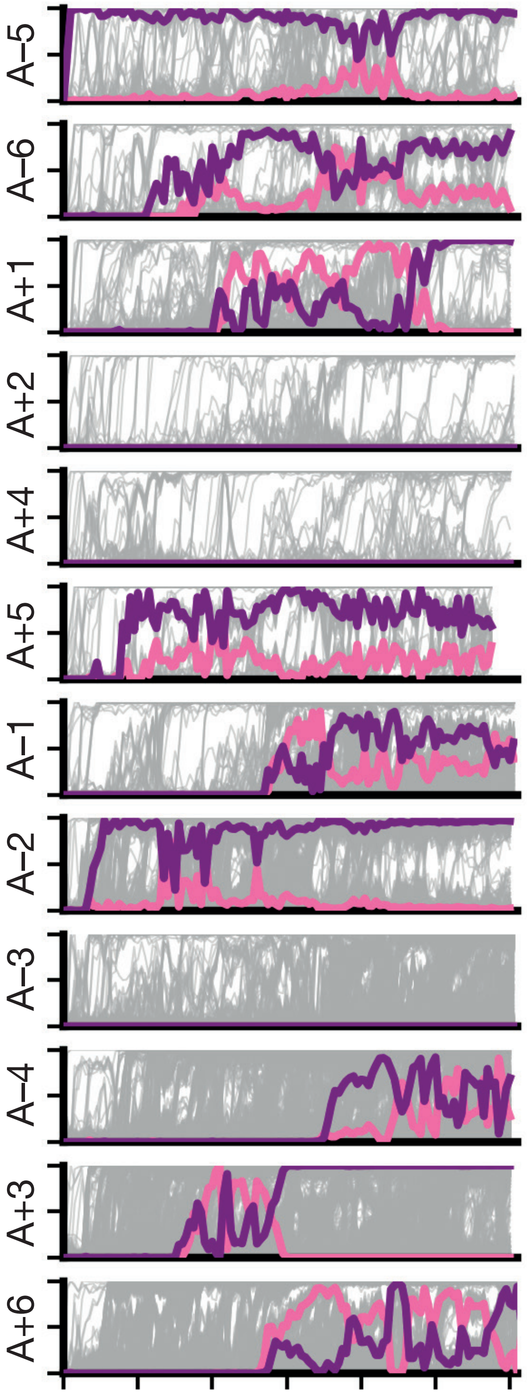
---

2. Do ecological interactions ***accelerate*** evolution?

---

3. What can we hope to ***infer*** about underlying eco/evo processes from this kind of sequence data?

Coupled ecological and evolutionary processes



+
other laboratory systems…

Many open questions:

1. What explains common ***emergence*** , but varied ***outcomes***?

Niche-specific adaptation?
More general coupling between fitness and stable freq?

---

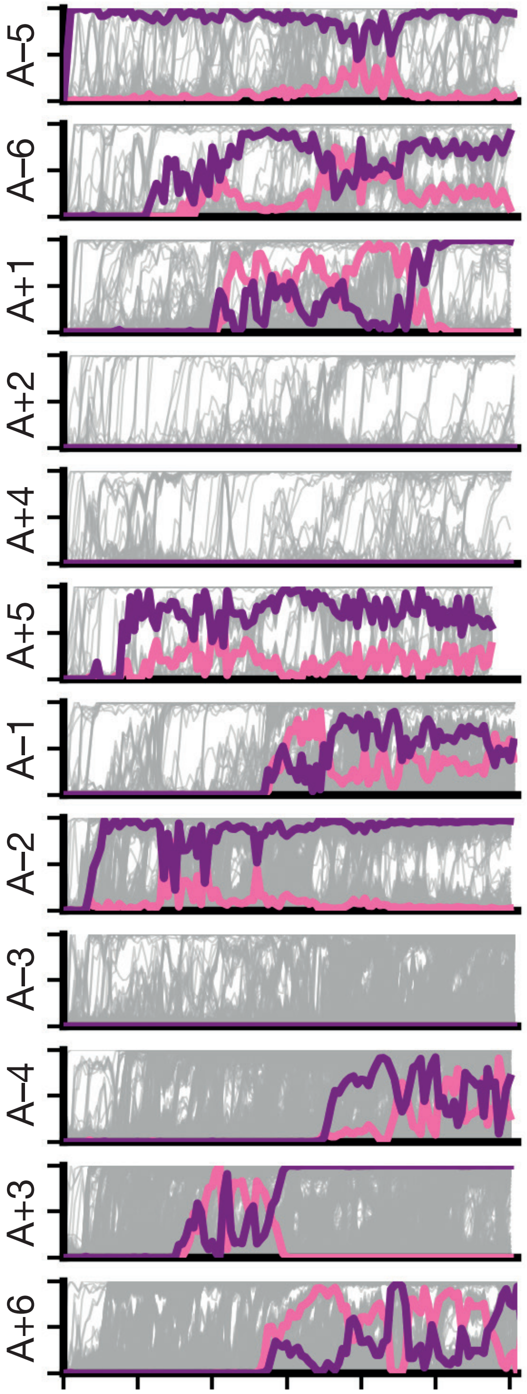
2. Do ecological interactions ***accelerate*** evolution?

---

3. What can we hope to ***infer*** about underlying eco/evo processes from this kind of sequence data?

Problem: most population genetic models neglect ecology.

Coupled ecological and evolutionary processes



+

other laboratory systems…

Many open questions:

1. What explains common ***emergence*** , but varied ***outcomes***?

Niche-specific adaptation?
More general coupling between fitness and stable freq?

---

2. Do ecological interactions ***accelerate*** evolution?

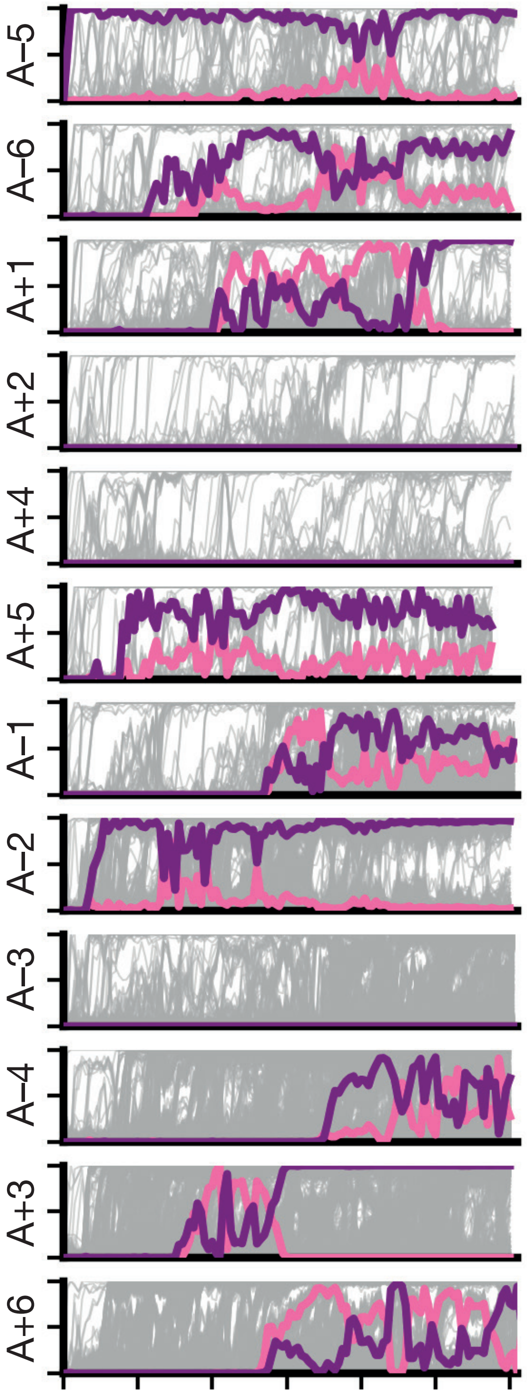
---

3. What can we hope to ***infer*** about underlying eco/evo processes from this kind of sequence data?

Problem: most population genetic models neglect ecology.

Easy to ***reject*** "no ecology", but not sure what to ***expect*** instead.

Coupled ecological and evolutionary processes



+
other laboratory systems…

Many open questions:

1. What explains common ***emergence*** , but varied ***outcomes***?

Niche-specific adaptation?
More general coupling between fitness and stable freq?

---

2. Do ecological interactions ***accelerate*** evolution?

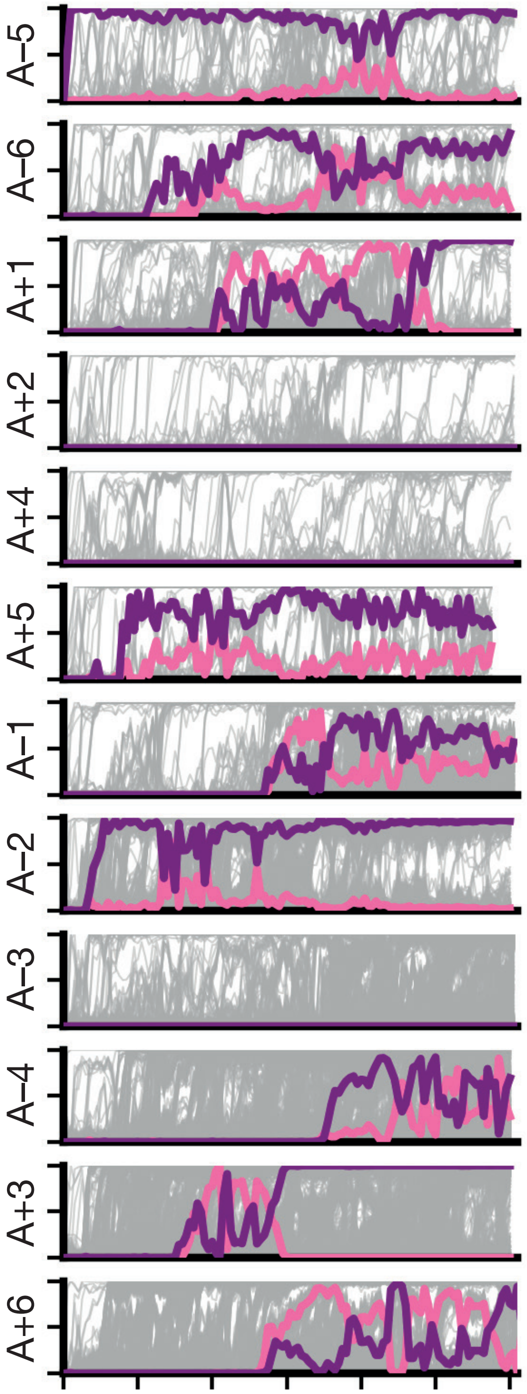
---

3. What can we hope to ***infer*** about underlying eco/evo processes from this kind of sequence data?

Problem: most population genetic models neglect ecology.

Easy to ***reject*** "no ecology", but not sure what to ***expect*** instead.
e.g. are these trajectories surprising? or not?

Coupled ecological and evolutionary processes



+
other laboratory systems…

Many open questions:

1. What explains common ***emergence*** , but varied ***outcomes***?

Niche-specific adaptation?
More general coupling between fitness and stable freq?

---

2. Do ecological interactions ***accelerate*** evolution?

---

3. What can we hope to ***infer*** about underlying eco/evo processes from this kind of sequence data?

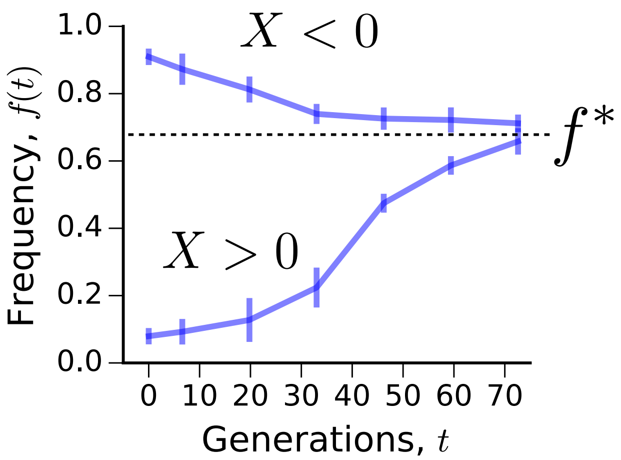
Problem: most population genetic models neglect ecology.

Easy to ***reject*** "no ecology", but not sure what to ***expect*** instead.
e.g. are these trajectories surprising? or not?

Need quantitative theory

How to model feedback between ecology and evolution?

1. The general case

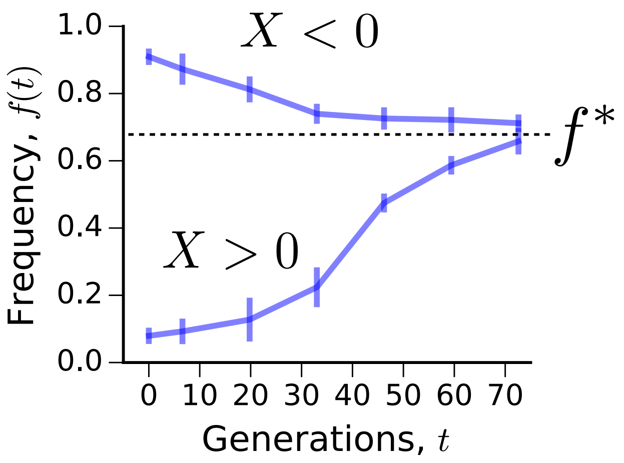


Frequency-dependent selection *(another type of interaction)*

$$X(\vec{g}) \to X[\vec{g}, \{f(\vec{g}')\}]$$

How to model feedback between ecology and evolution?

1. The general case



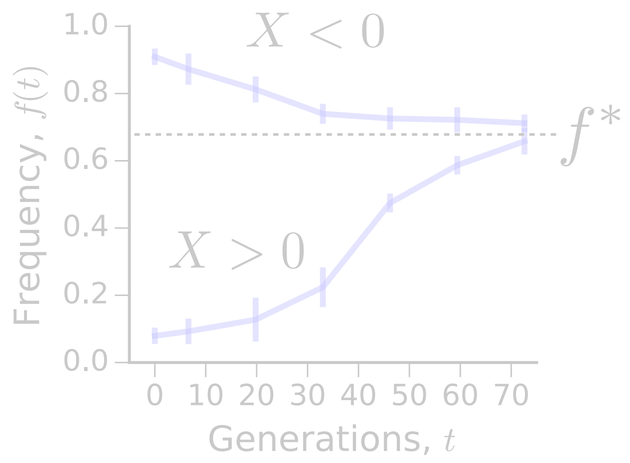
Frequency-dependent selection *(another type of interaction)*

$$X(\vec{g}) \to X[\vec{g}, \{f(\vec{g}')\}] \approx X(\vec{g}) + \sum_{\vec{g}'} \underset{\uparrow}{c(\vec{g}, \vec{g}')} f(\vec{g}') + \ldots$$

hard to model!
(many params)

How to model feedback between ecology and evolution?

I. The general case



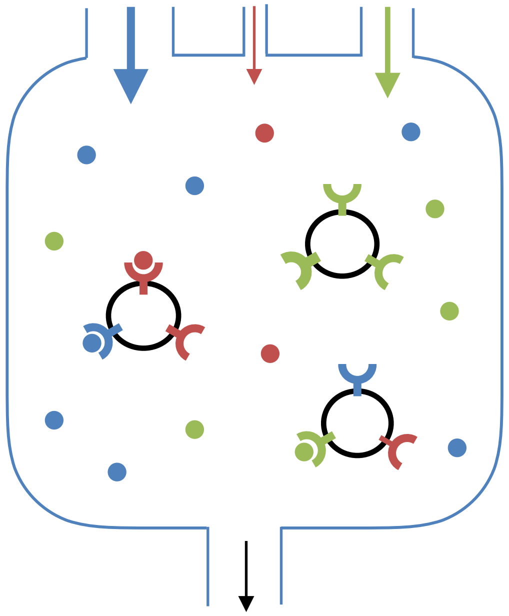
Frequency-dependent selection *(another type of interaction)*

$$X(\vec{g}) \rightarrow X[\vec{g}, \{f(\vec{g}')\}] \approx X(\vec{g}) + \sum_{\vec{g}'} c(\vec{g}, \vec{g}')f(\vec{g}') + \ldots$$

hard to model!
(many params)

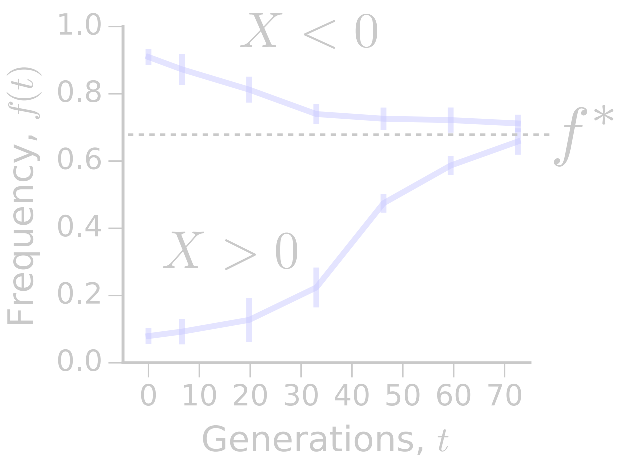
II. An empirically-motivated toy model

a Competition for $\mathcal{R} > 1$ resources



How to model feedback between ecology and evolution?

I. The general case



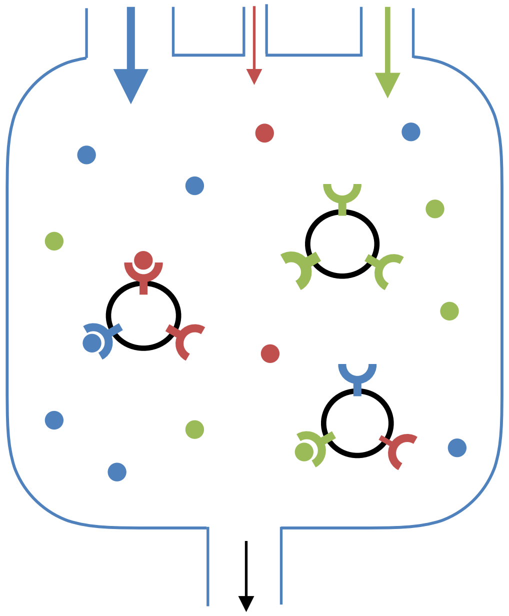
Frequency-dependent selection *(another type of interaction)*

$$X(\vec{g}) \to X[\vec{g}, \{f(\vec{g}')\}] \approx X(\vec{g}) + \sum_{\vec{g}'} c(\vec{g},\vec{g}')f(\vec{g}') + \dots$$

hard to model!
(many params)

II. An empirically-motivated toy model

a Competition for $\mathcal{R} > 1$ resources

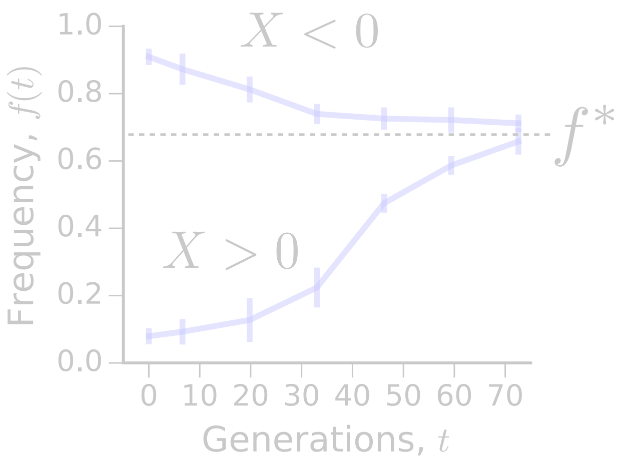


"Environmental supply vector"

β_i = % biomass supplied by resource i

How to model feedback between ecology and evolution?

I. The general case



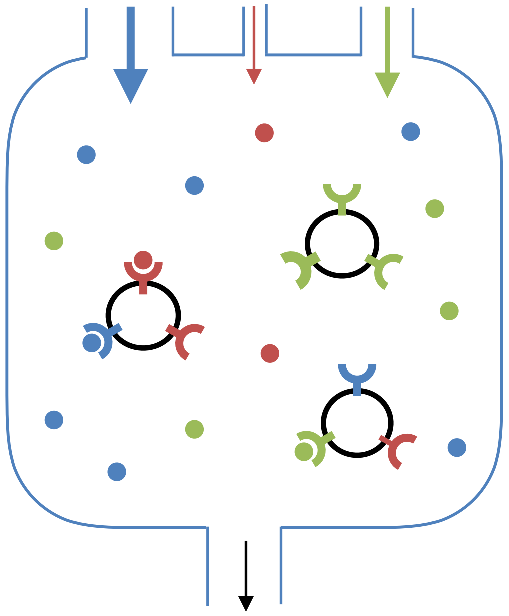
Frequency-dependent selection *(another type of interaction)*

$$X(\vec{g}) \to X[\vec{g}, \{f(\vec{g}')\}] \approx X(\vec{g}) + \sum_{\vec{g}'} c(\vec{g}, \vec{g}') f(\vec{g}') + \ldots$$

hard to model!
(many params)

II. An empirically-motivated toy model

a Competition for $\mathcal{R} > 1$ resources



“Environmental supply vector”

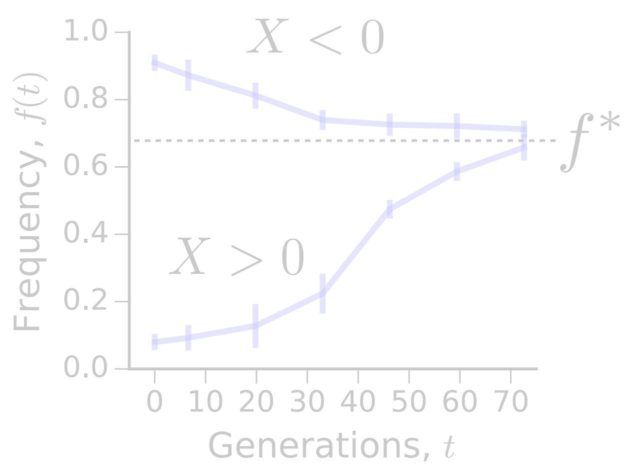
β_i = % biomass supplied by resource i

“Resource strategy”

$\alpha_{\mu,i}$ = % energy strain μ spends to import resource i

How to model feedback between ecology and evolution?

I. The general case



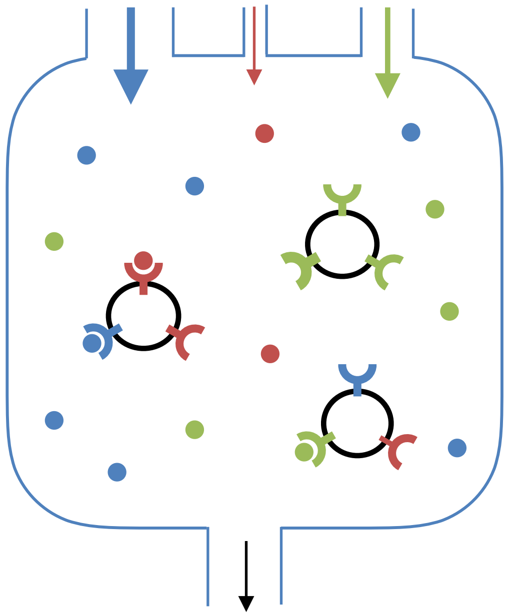
Frequency-dependent selection *(another type of interaction)*

$$X(\vec{g}) \to X[\vec{g}, \{f(\vec{g}')\}] \approx X(\vec{g}) + \sum_{\vec{g}'} c(\vec{g},\vec{g}')f(\vec{g}') + \dots$$

hard to model!
(many params)

II. An empirically-motivated toy model

a Competition for $\mathcal{R} > 1$ resources



“Environmental supply vector”

β_i = % biomass supplied by resource i

“Resource strategy”

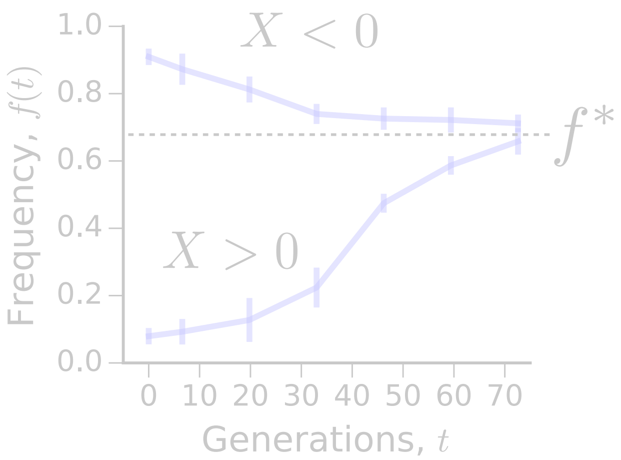
$\alpha_{\mu,i}$ = % energy strain μ spends to import resource i

“General fitness”

X_μ = max growth rate of strain μ

How to model feedback between ecology and evolution?

I. The general case



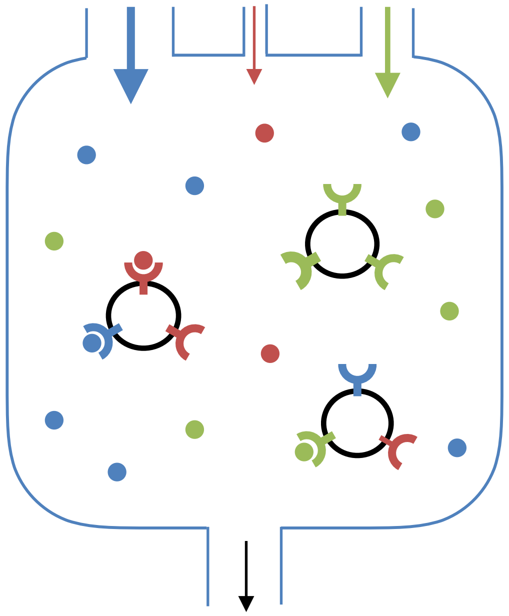
Frequency-dependent selection *(another type of interaction)*

$$X(\vec{g}) \to X[\vec{g}, \{f(\vec{g}')\}] \approx X(\vec{g}) + \sum_{\vec{g}'} c(\vec{g}, \vec{g}') f(\vec{g}') + \dots$$

hard to model!
(many params)

II. An empirically-motivated toy model

a Competition for $\mathcal{R} > 1$ resources



Frequencies of strains

$$\frac{\partial f_\mu}{\partial t} \approx \sum_i \alpha_{\mu,i} \left[X_\mu - \overline{X}_i\right] f_\mu + \text{genetic drift}$$

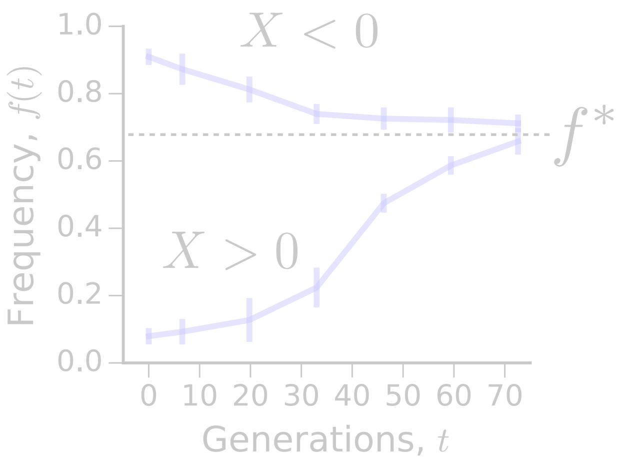
"Resource-specific mean fitness"

$$\overline{X}_i \equiv \log\left(\sum_\mu \frac{\alpha_{\mu,i}}{\beta_i} e^{X_\mu} f_\mu\right)$$

"Environmental supply vector"	*"Resource strategy"*	*"General fitness"*
β_i = % biomass supplied by resource i	$\alpha_{\mu,i}$ = % energy strain μ spends to import resource i	X_μ = max growth rate of strain μ

How to model feedback between ecology and evolution?

I. The general case



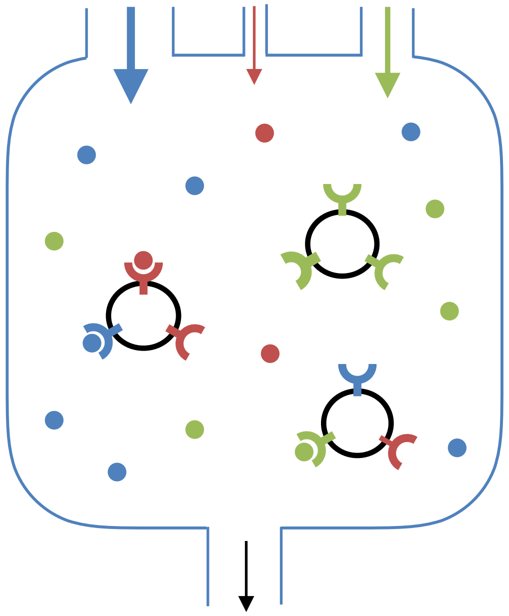
Frequency-dependent selection *(another type of interaction)*

$$X(\vec{g}) \to X[\vec{g}, \{f(\vec{g}')\}] \approx X(\vec{g}) + \sum_{\vec{g}'} c(\vec{g}, \vec{g}') f(\vec{g}') + \dots$$

hard to model!
(many params)

II. An empirically-motivated toy model

a Competition for $\mathcal{R} > 1$ resources



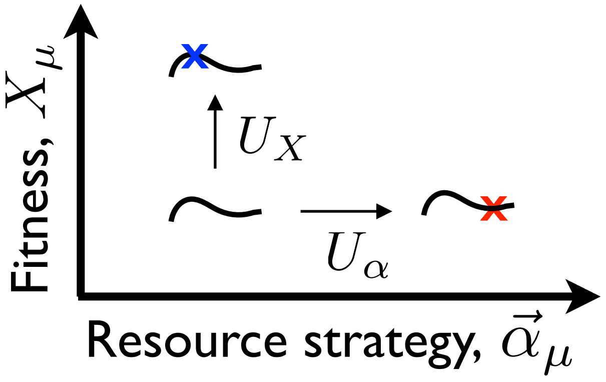
Frequencies of strains

$$\frac{\partial f_\mu}{\partial t} \approx \sum_i \alpha_{\mu,i} \left[X_\mu - \overline{X}_i\right] f_\mu + \text{genetic drift}$$

"Resource-specific mean fitness"

$$\overline{X}_i \equiv \log\left(\sum_\mu \frac{\alpha_{\mu,i}}{\beta_i} e^{X_\mu} f_\mu\right)$$

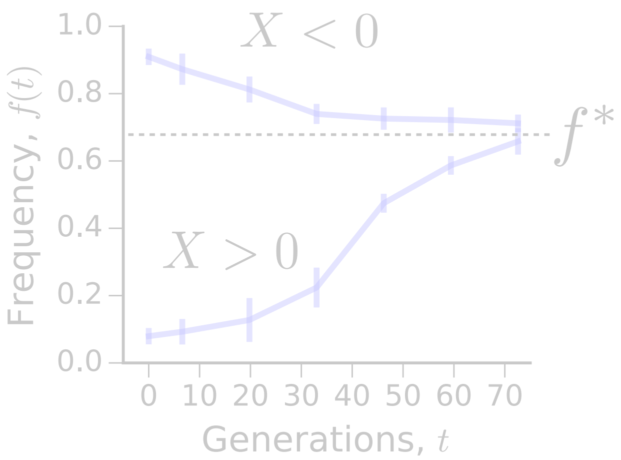
b Heritable mutations



"Environmental supply vector"	*"Resource strategy"*	*"General fitness"*
β_i = % biomass supplied by resource i	$\alpha_{\mu,i}$ = % energy strain μ spends to import resource i	X_μ = max growth rate of strain μ

How to model feedback between ecology and evolution?

I. The general case



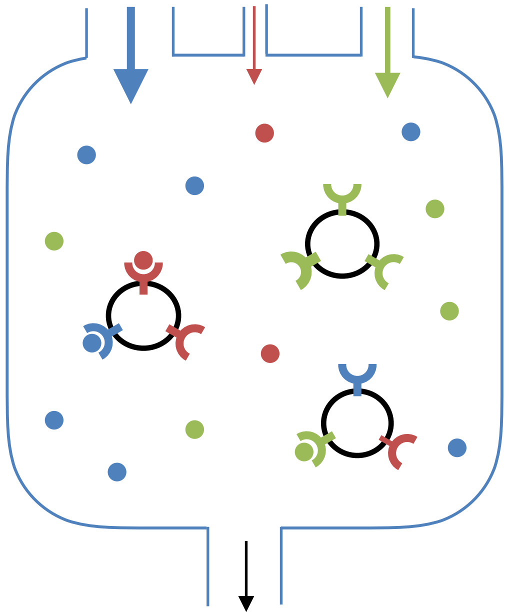
Frequency-dependent selection *(another type of interaction)*

$$X(\vec{g}) \to X[\vec{g}, \{f(\vec{g}')\}] \approx X(\vec{g}) + \sum_{\vec{g}'} \underbrace{c(\vec{g}, \vec{g}')}_{\uparrow} f(\vec{g}') + \dots$$

hard to model!
(many params)

II. An empirically-motivated toy model

a Competition for $\mathcal{R} > 1$ resources



Frequencies of strains

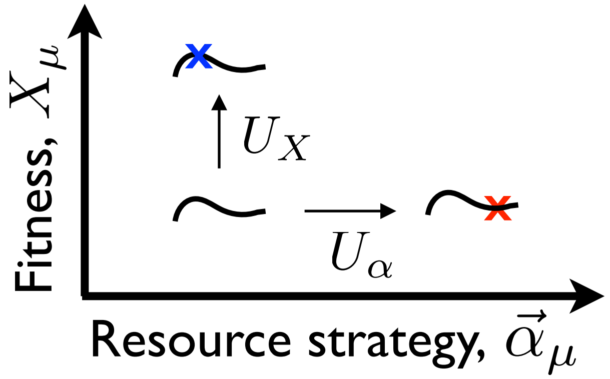
$$\frac{\partial f_\mu}{\partial t} \approx \sum_i \alpha_{\mu,i} \left[X_\mu - \overline{X}_i \right] f_\mu + \text{genetic drift}$$

"Resource-specific mean fitness"

$$\overline{X}_i \equiv \log \left(\sum_\mu \frac{\alpha_{\mu,i}}{\beta_i} e^{X_\mu} f_\mu \right)$$

Experiments: lots of "non-metabolic" adaptation going on at same time!

b Heritable mutations



"Environmental supply vector"

β_i = % biomass supplied by resource i

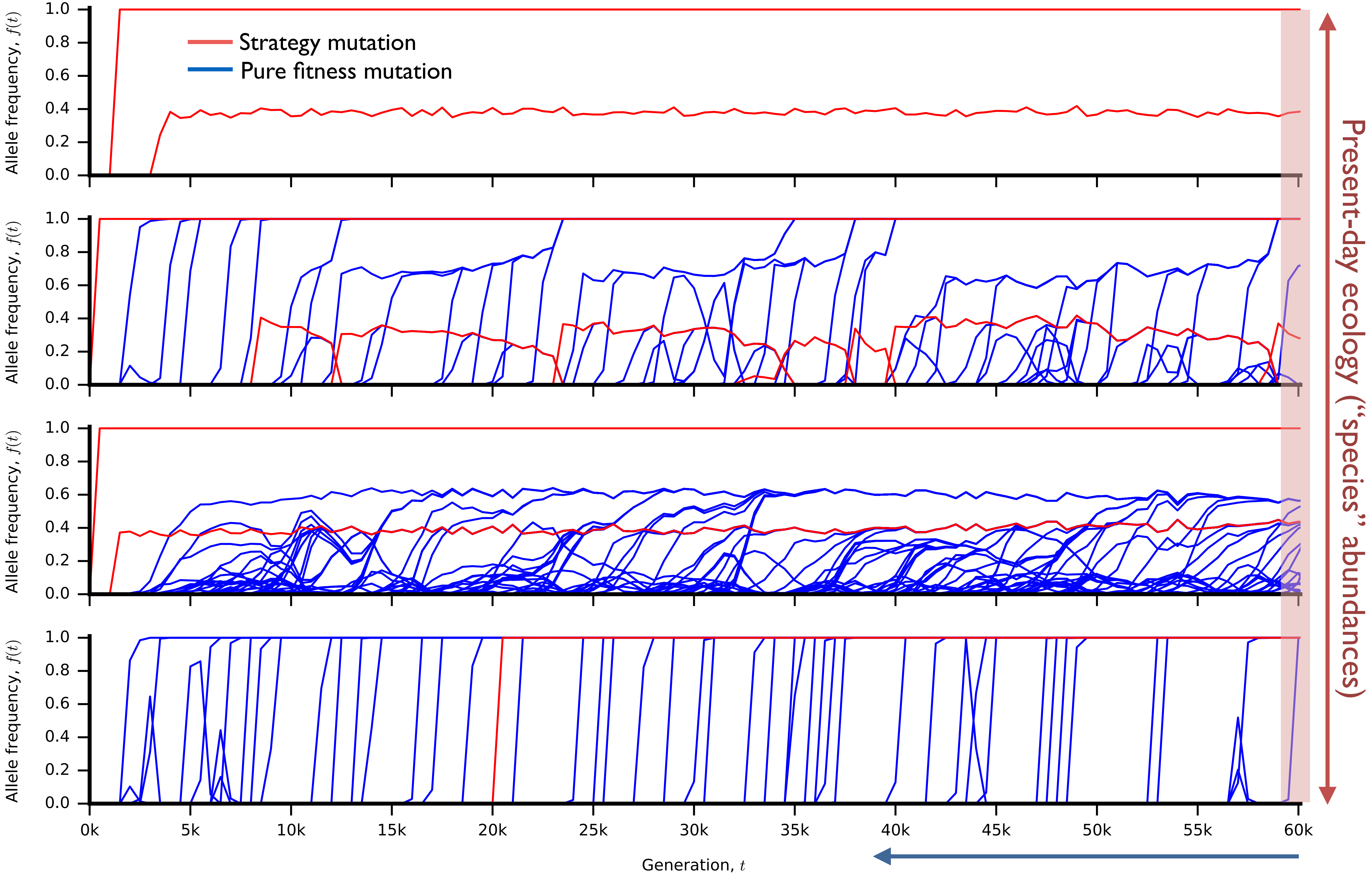
"Resource strategy"

$\alpha_{\mu,i}$ = % energy strain μ spends to import resource i

"General fitness"

X_μ = max growth rate of strain μ

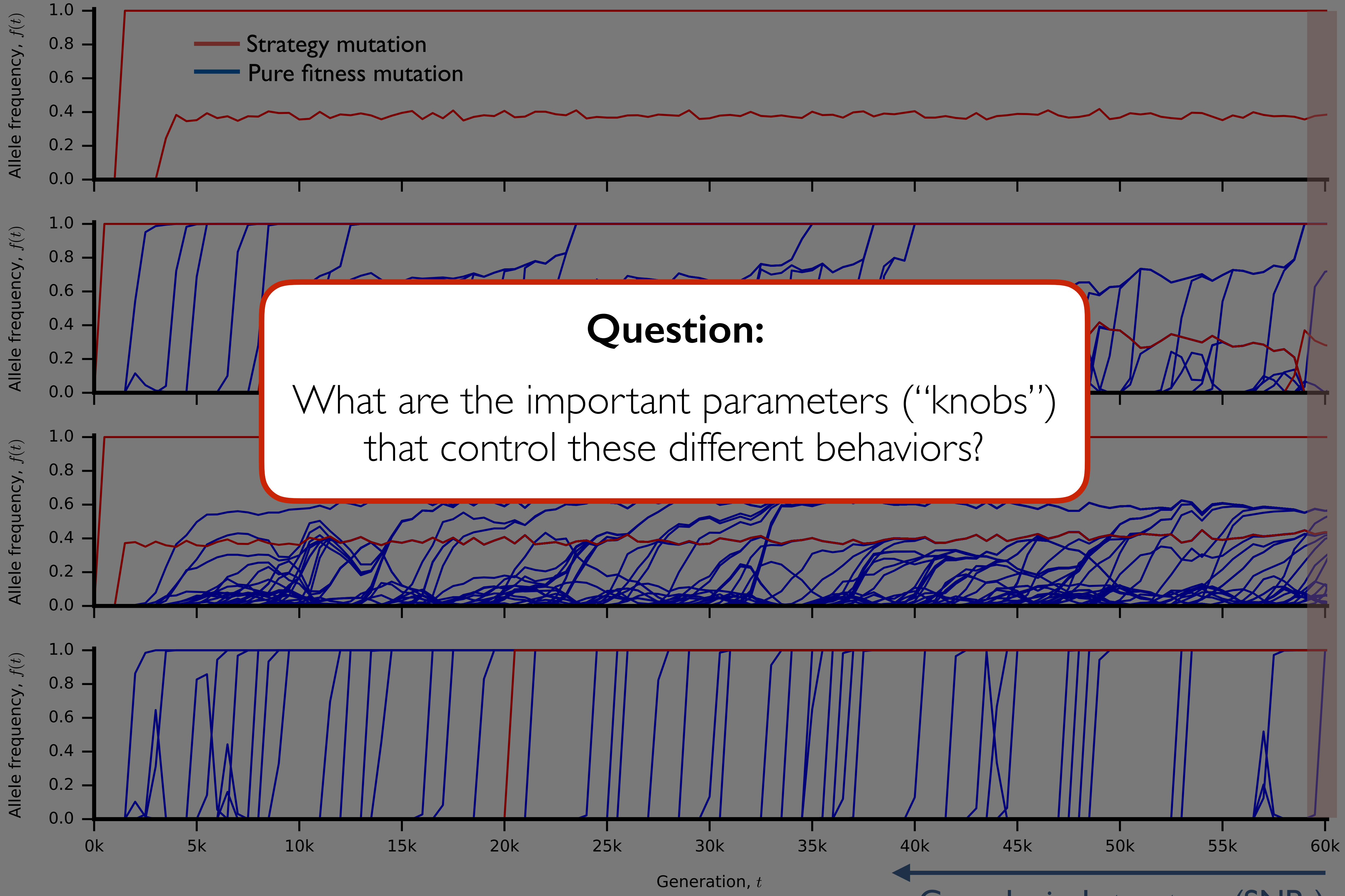
Simulations w/ same ecological opportunities, different supply of "fitness" mutations



Present-day ecology ("species" abundances)

Genealogical structure (SNPs)

Simulations w/ same ecological opportunities, different supply of "fitness" mutations



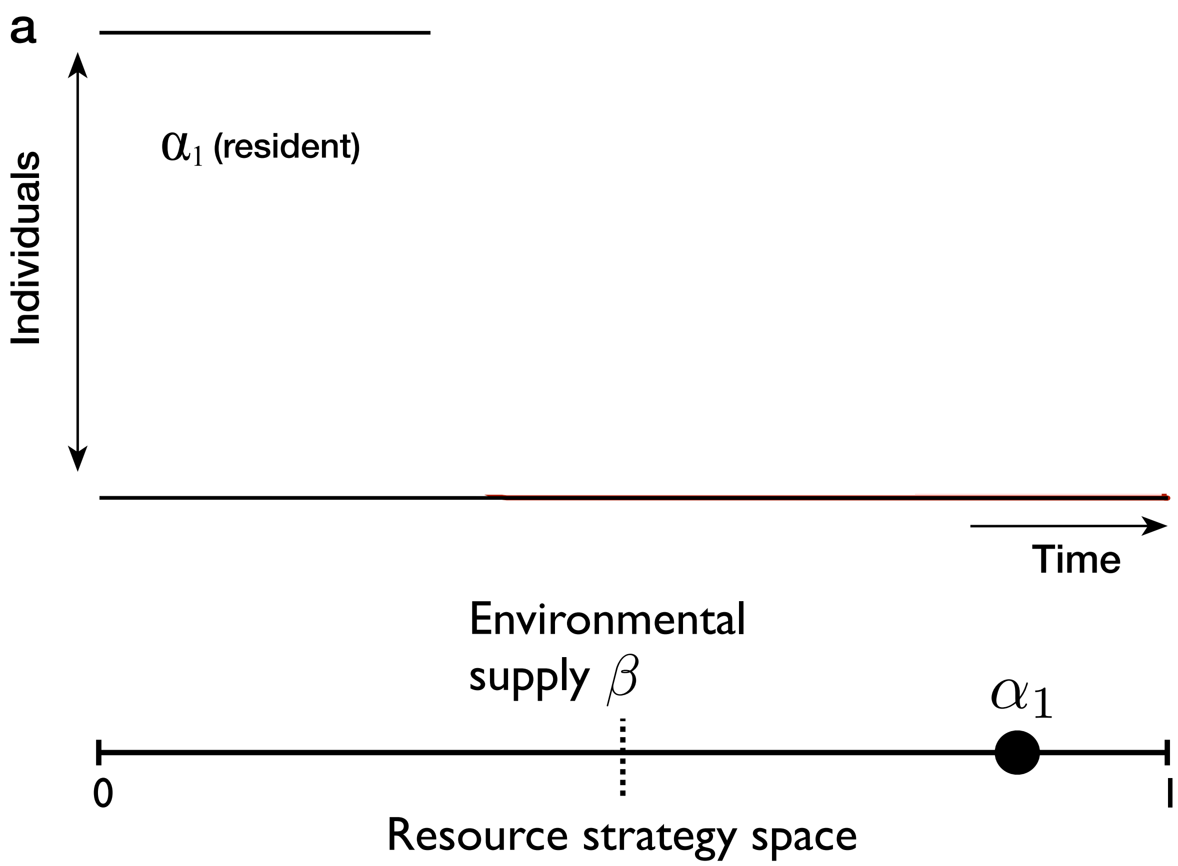
Question:

What are the important parameters ("knobs") that control these different behaviors?

Present-day ecology ("species" abundances)

Genealogical structure (SNPs)

“Neutral” case: selection for *ecosystem* to match environment



“Environmental supply vector”

β = % biomass supplied by resource 1

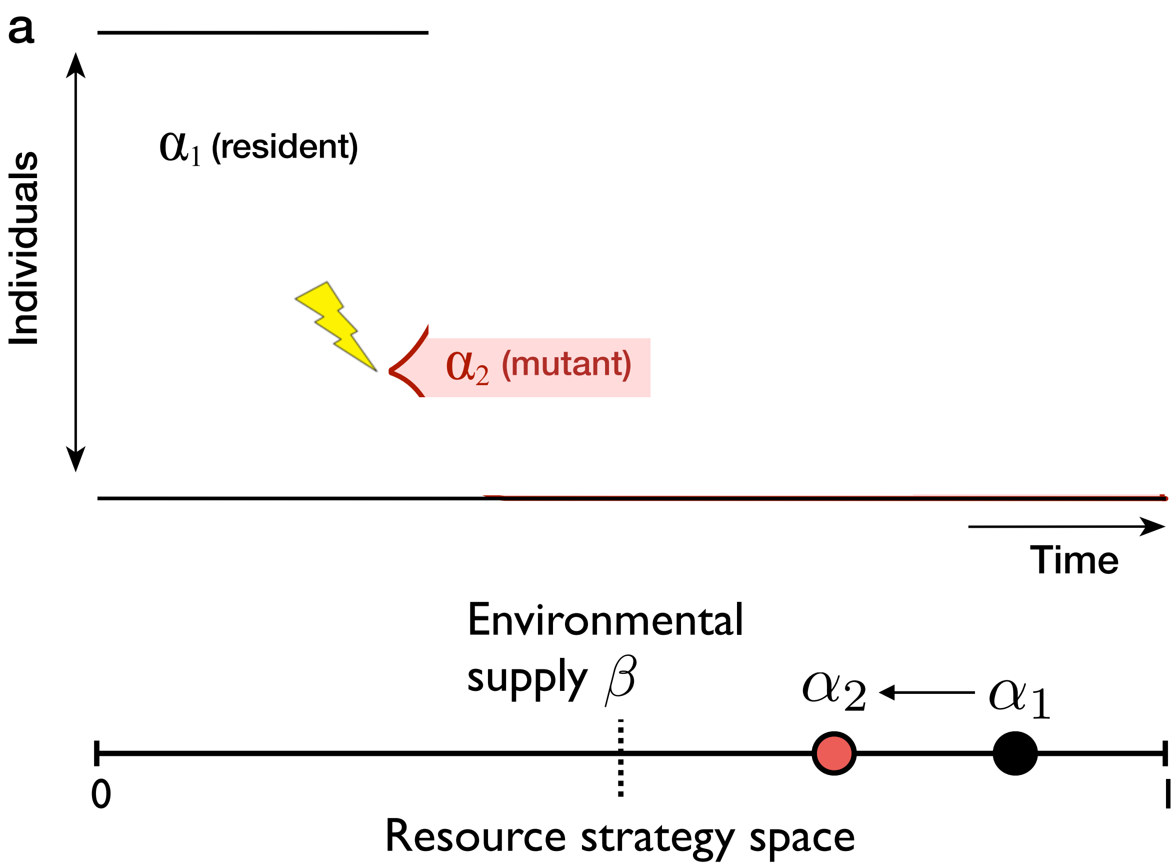
“Resource strategy”

α_μ = % energy strain μ spends to import resource 1

“General fitness”

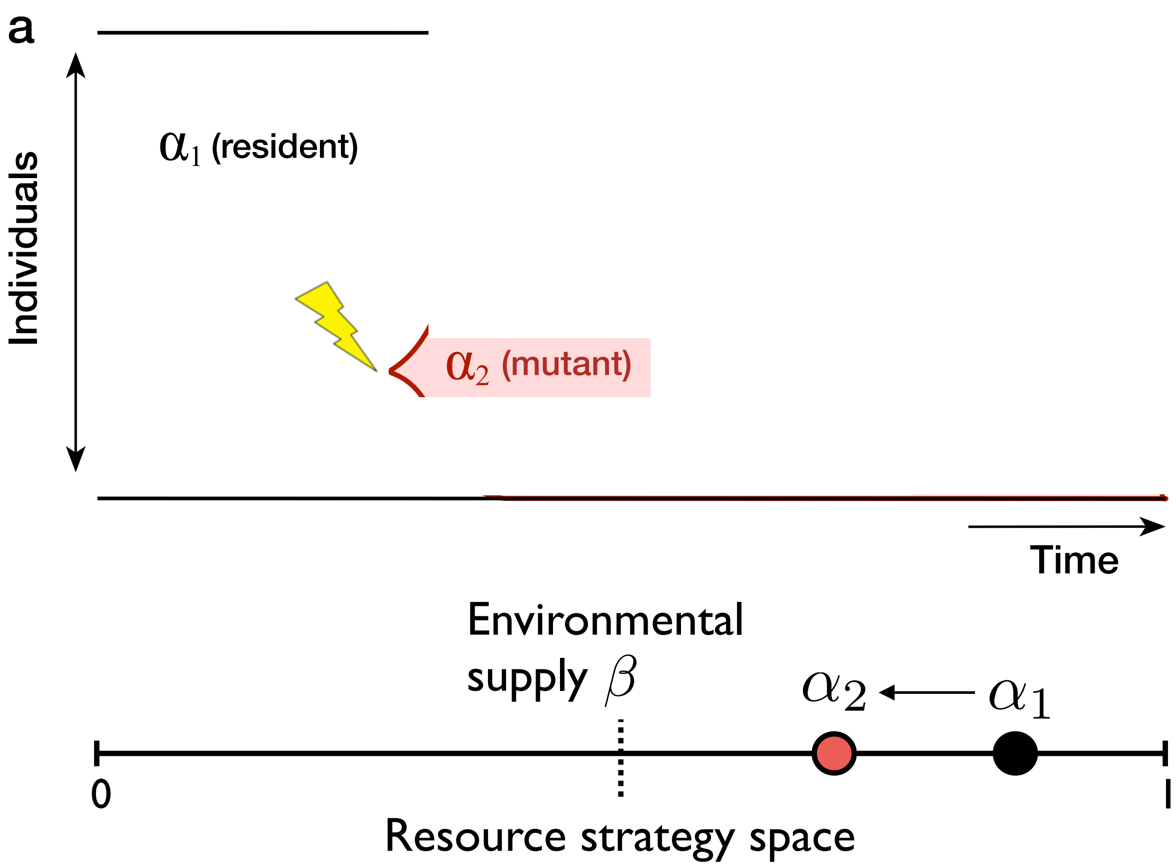
X_μ = max growth rate of strain μ

“Neutral” case: selection for *ecosystem* to match environment



“Environmental supply vector”	“Resource strategy”	“General fitness”
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

“Neutral” case: selection for *ecosystem* to match environment



“Ecological” selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_2) \propto \Delta\alpha(\beta - \alpha_1)$$

“Environmental supply vector”

β = % biomass supplied by resource 1

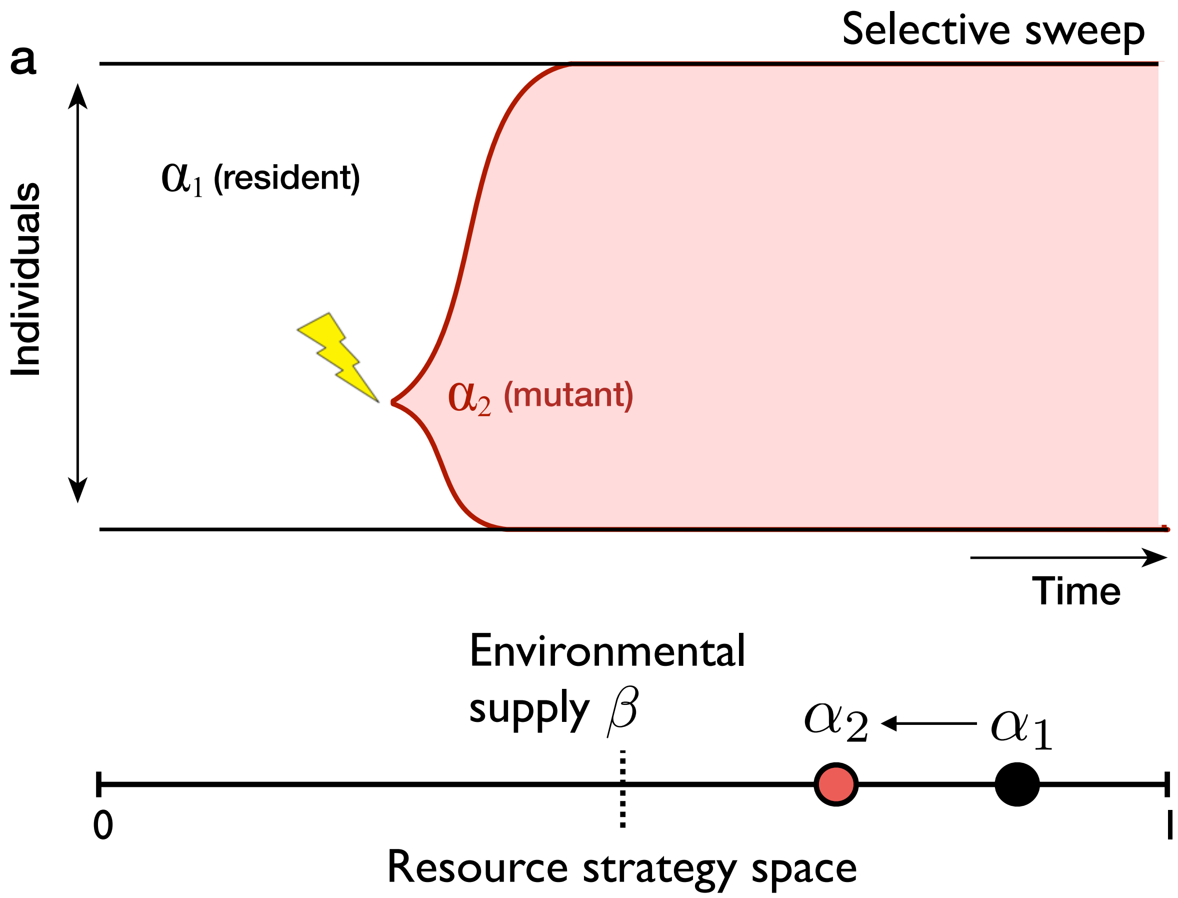
“Resource strategy”

α_μ = % energy strain μ spends to import resource 1

“General fitness”

X_μ = max growth rate of strain μ

“Neutral” case: selection for *ecosystem* to match environment

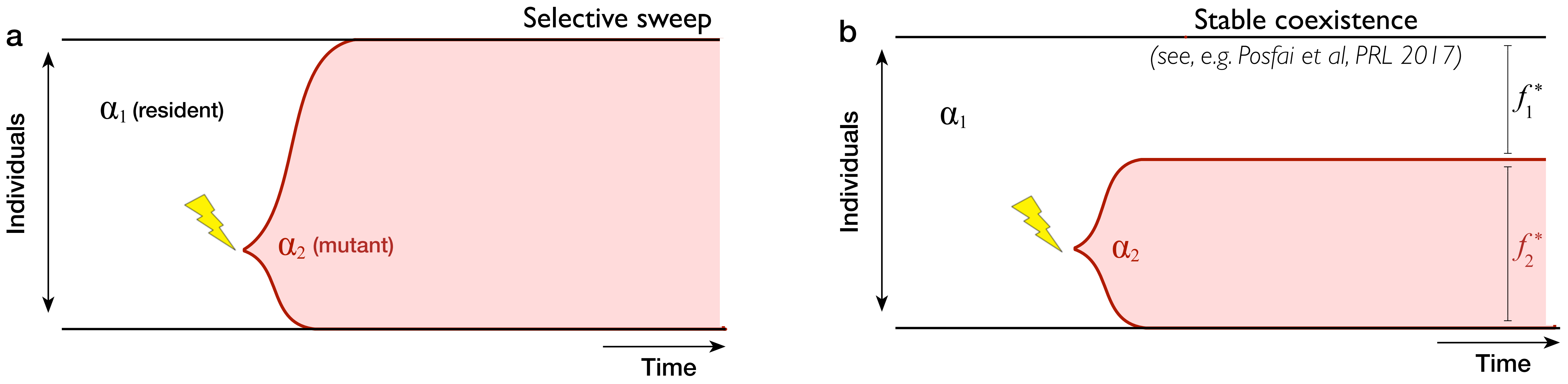


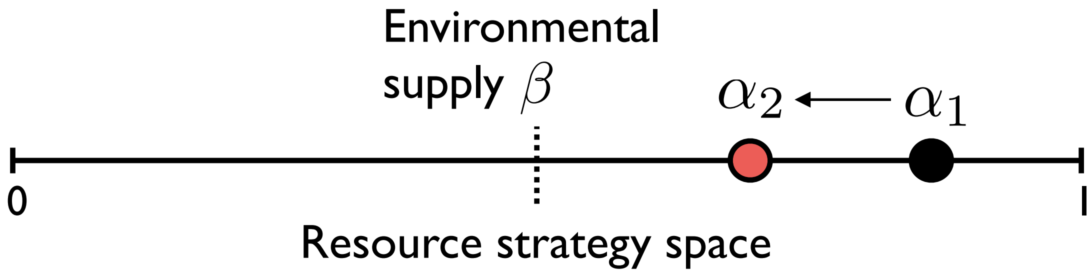
“Ecological” selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_2) \propto \Delta\alpha(\beta - \alpha_1)$$

“Environmental supply vector”	*“Resource strategy”*	*“General fitness”*
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

“Neutral” case: selection for *ecosystem* to match environment





“Ecological” selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_2) \propto \Delta\alpha(\beta - \alpha_1)$$

“Environmental supply vector”

β = % biomass supplied by resource 1

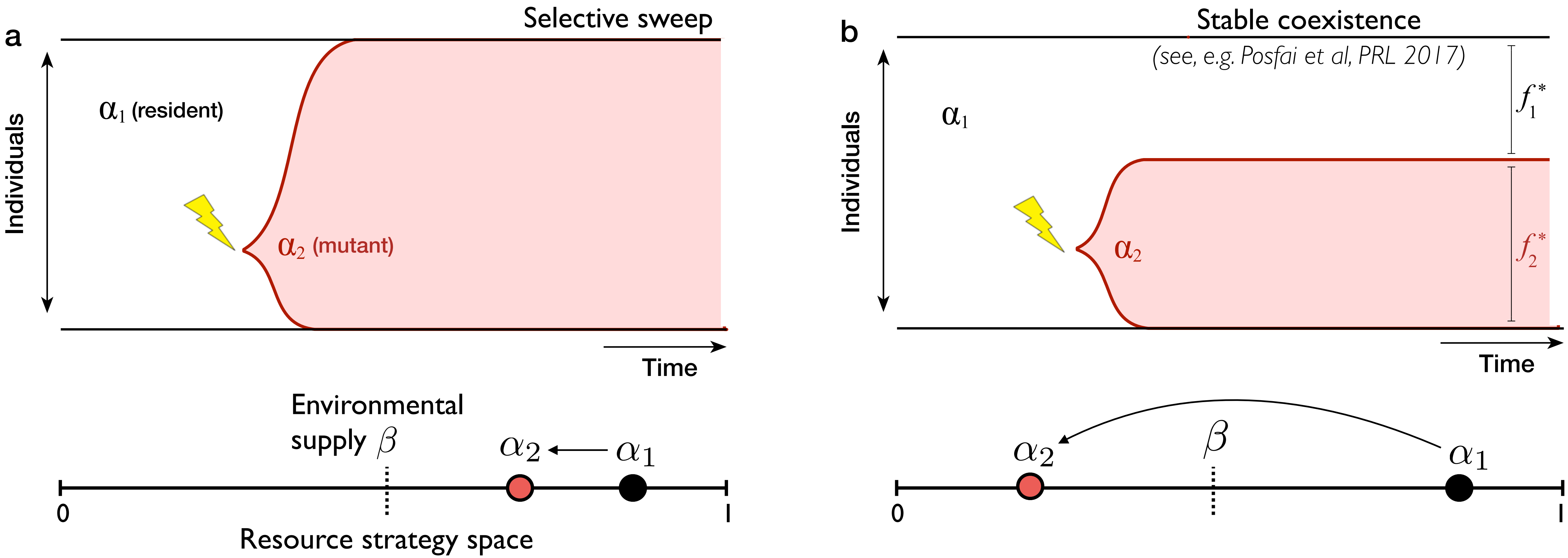
“Resource strategy”

α_μ = % energy strain μ spends to import resource 1

“General fitness”

X_μ = max growth rate of strain μ

“Neutral” case: selection for *ecosystem* to match environment



“Ecological” selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_2) \propto \Delta\alpha(\beta - \alpha_1)$$

“Environmental supply vector”

β = % biomass supplied by resource 1

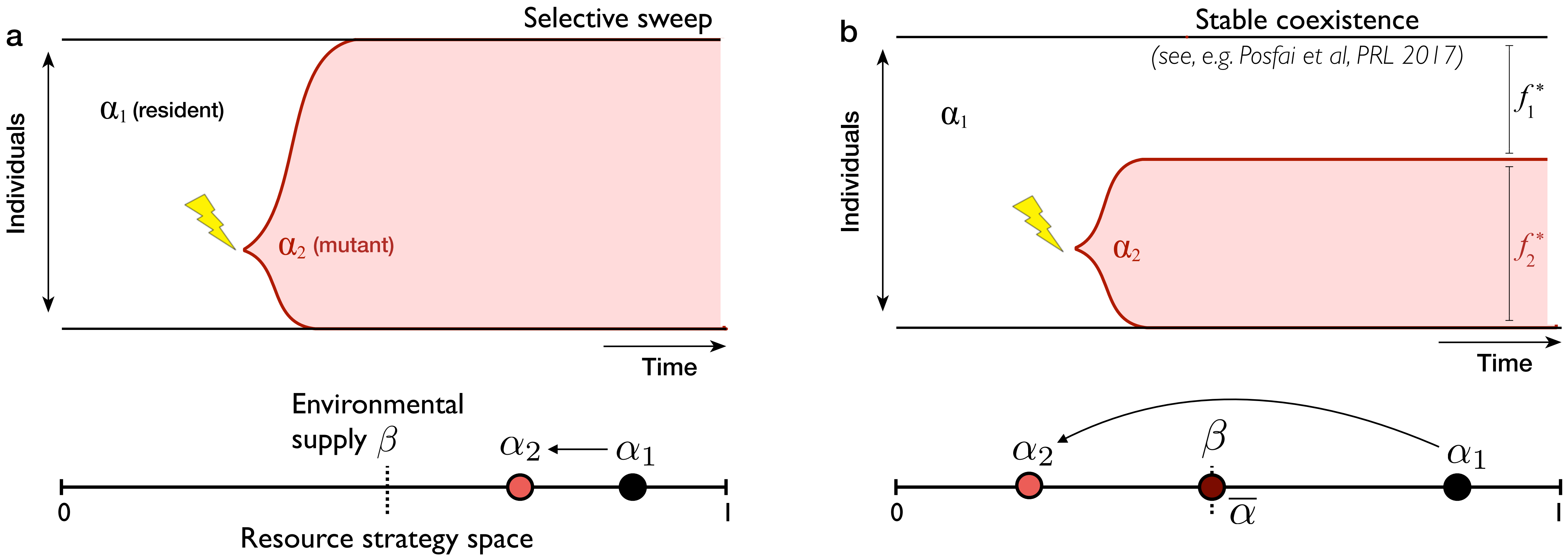
“Resource strategy”

α_μ = % energy strain μ spends to import resource 1

“General fitness”

X_μ = max growth rate of strain μ

“Neutral” case: selection for *ecosystem* to match environment



“Ecological” selection pressure:

$$s_{\mathrm{eff}}(\alpha_1 \to \alpha_2) \propto \Delta\alpha(\beta - \alpha_1)$$

“Environmental supply vector”

β = % biomass supplied by resource 1

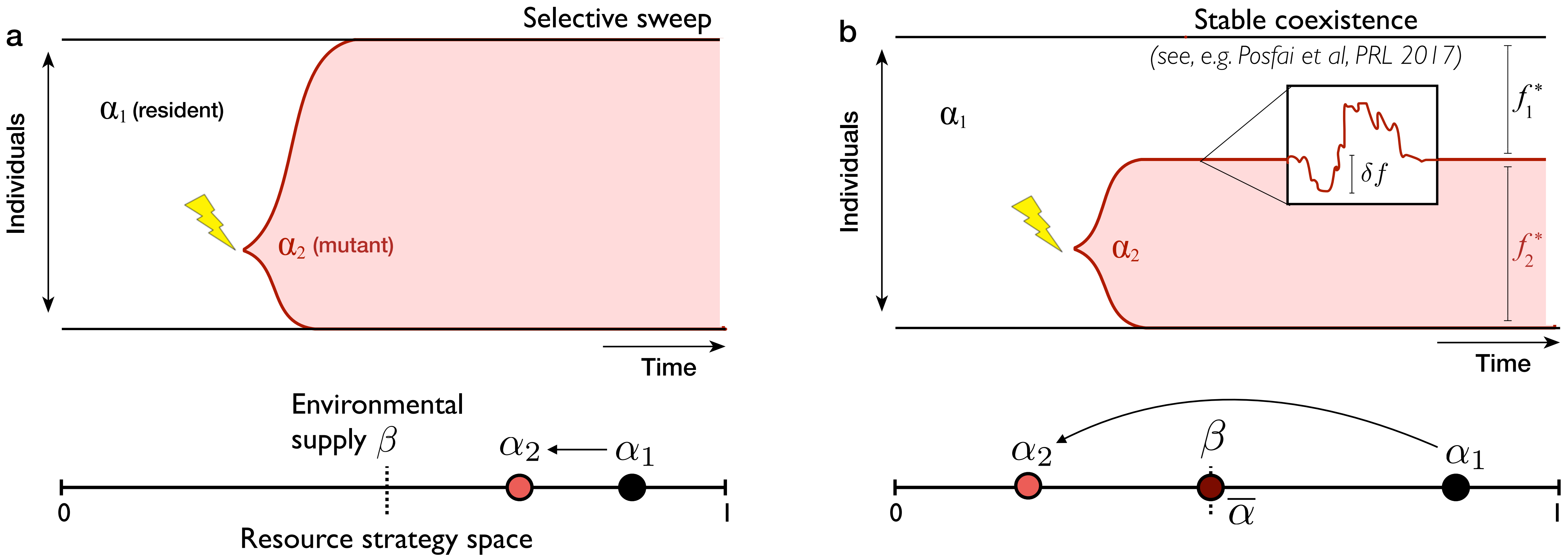
“Resource strategy”

α_μ = % energy strain μ spends to import resource 1

“General fitness”

X_μ = max growth rate of strain μ

"Neutral" case: selection for *ecosystem* to match environment



"Ecological" selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_2) \propto \Delta\alpha(\beta - \alpha_1)$$

"Environmental supply vector"

β = % biomass supplied by resource 1

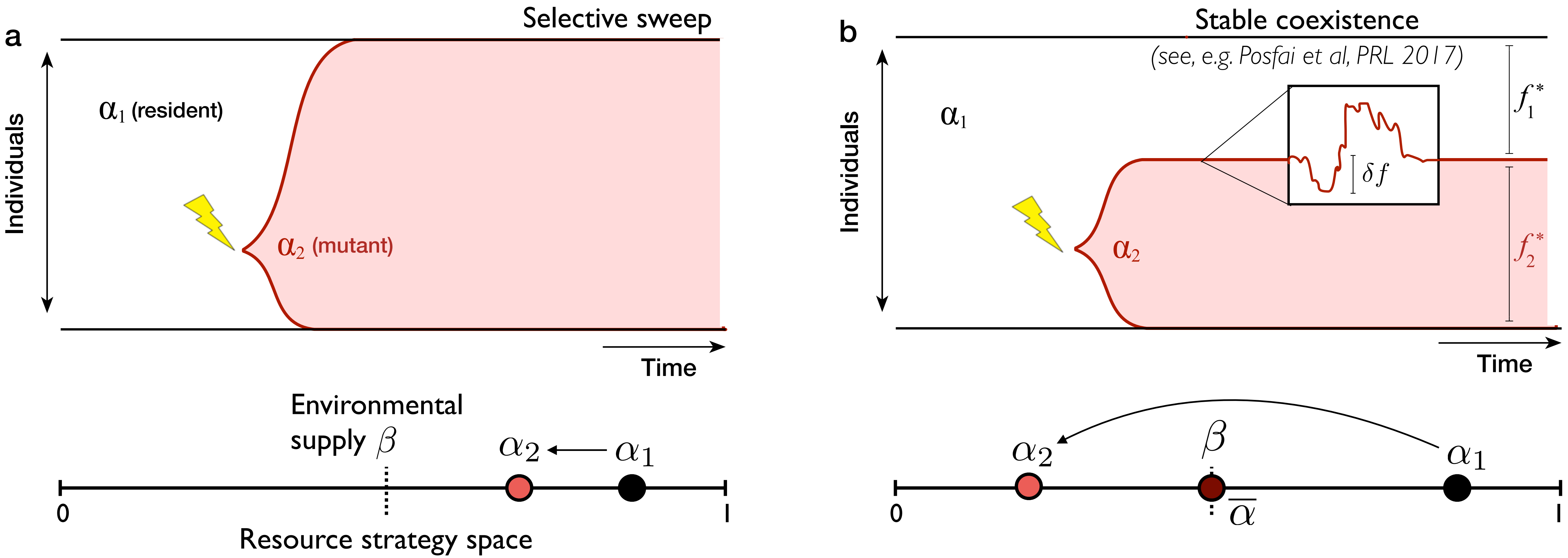
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

“Neutral” case: selection for *ecosystem* to match environment



“Ecological” selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_2) \propto \Delta\alpha(\beta - \alpha_1)$$

Marginal stability against add’l muts

$$s_{\text{eff}}(\alpha_i \to \alpha_3) \sim \mathcal{O}(1/N)$$

“Environmental supply vector”

β = % biomass supplied by resource 1

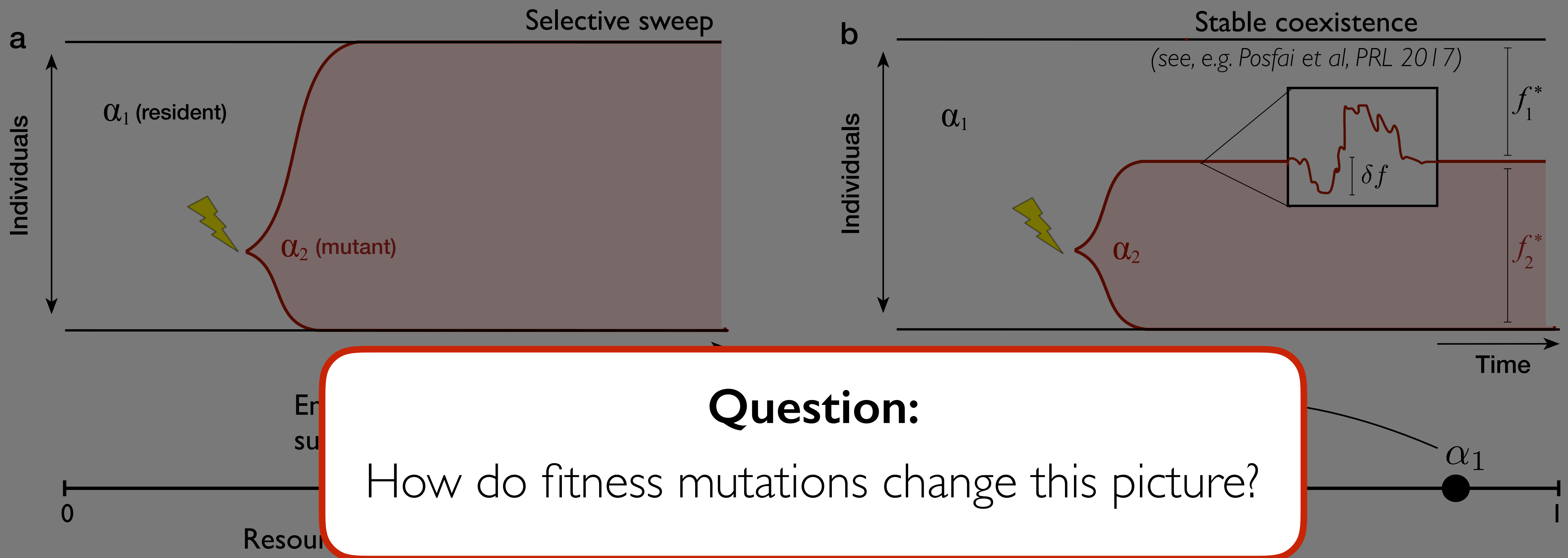
“Resource strategy”

α_μ = % energy strain μ spends to import resource 1

“General fitness”

X_μ = max growth rate of strain μ

“Neutral” case: selection for *ecosystem* to match environment



Question:

How do fitness mutations change this picture?

“Ecological” selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_2) \propto \Delta\alpha(\beta - \alpha_1)$$

Marginal stability against add’l muts

$$s_{\text{eff}}(\alpha_i \to \alpha_3) \sim \mathcal{O}(1/N)$$

“Environmental supply vector”

β = % biomass supplied by resource 1

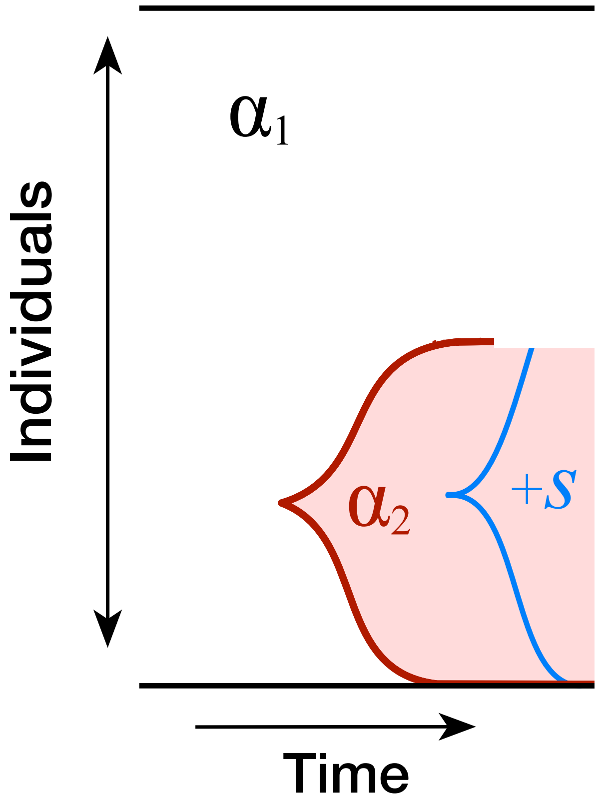
“Resource strategy”

α_μ = % energy strain μ spends to import resource 1

“General fitness”

X_μ = max growth rate of strain μ

Diversification-selection balance



“Environmental supply vector”

β = % biomass supplied by resource 1

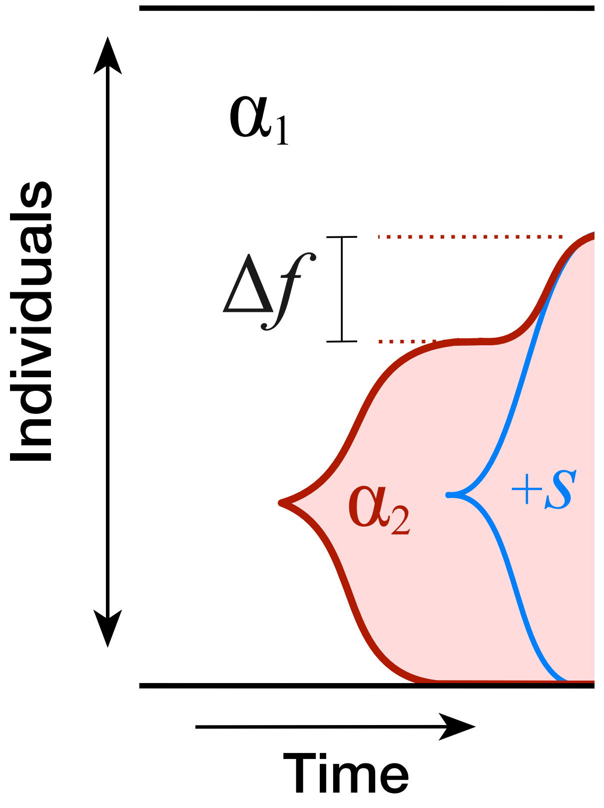
“Resource strategy”

α_μ = % energy strain μ spends to import resource 1

“General fitness”

X_μ = max growth rate of strain μ

Diversification-selection balance



"Environmental supply vector"

β = % biomass supplied by resource 1

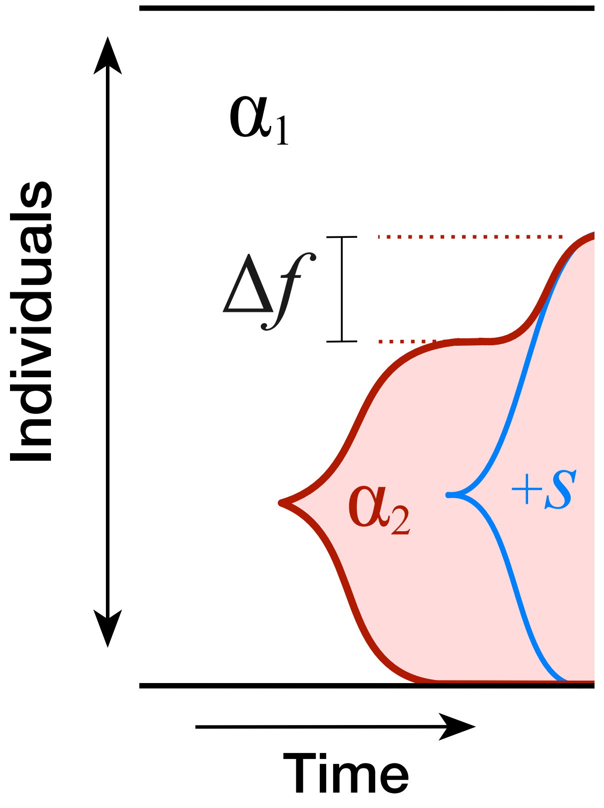
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

Diversification-selection balance



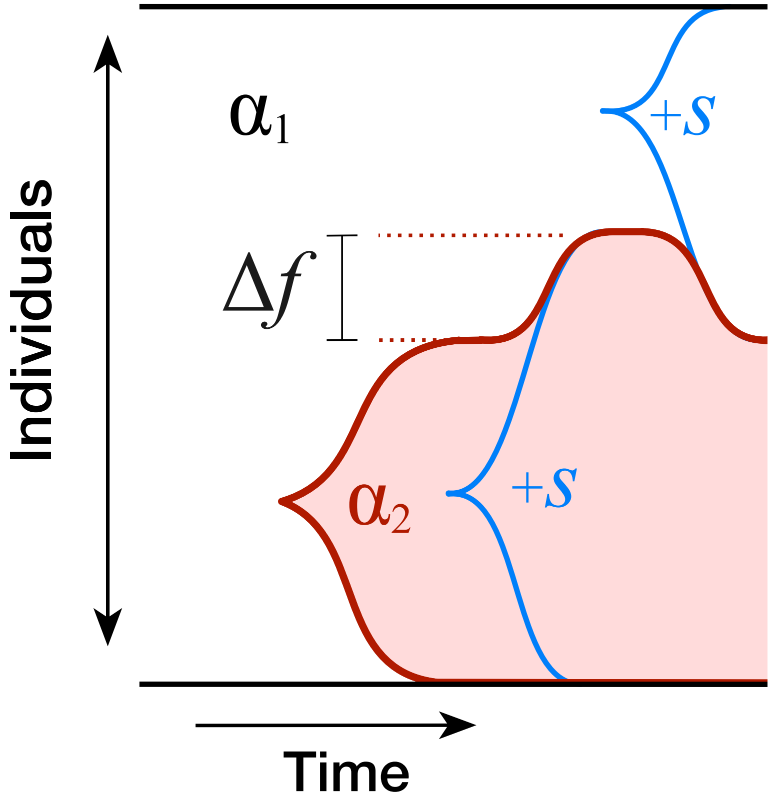
1. Fitness differences perturb equilibria:

$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta\alpha^2}{\beta(1-\beta)}$$

"Evolutionary robustness"

"Environmental supply vector"	"Resource strategy"	"General fitness"
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

Diversification-selection balance



1. Fitness differences perturb equilibria:

$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta\alpha^2}{\beta(1-\beta)}$$

"Evolutionary robustness"

"Environmental supply vector"

β = % biomass supplied by resource 1

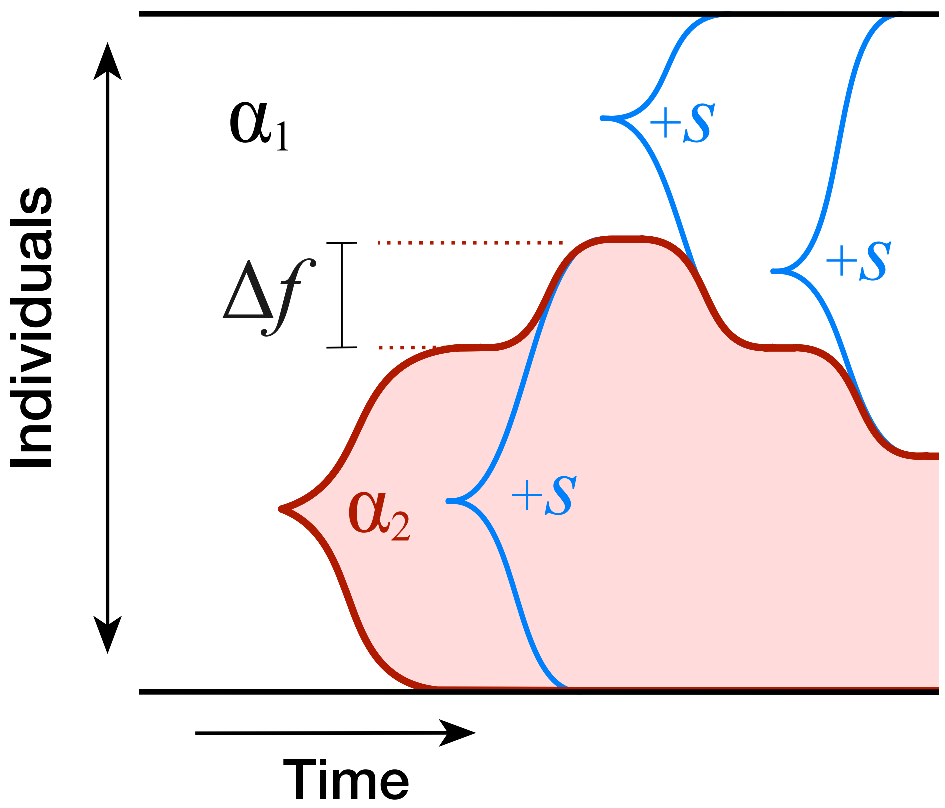
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

Diversification-selection balance



1. Fitness differences perturb equilibria:

$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta\alpha^2}{\beta(1-\beta)}$$

"Evolutionary robustness"

"Environmental supply vector"

β = % biomass supplied by resource 1

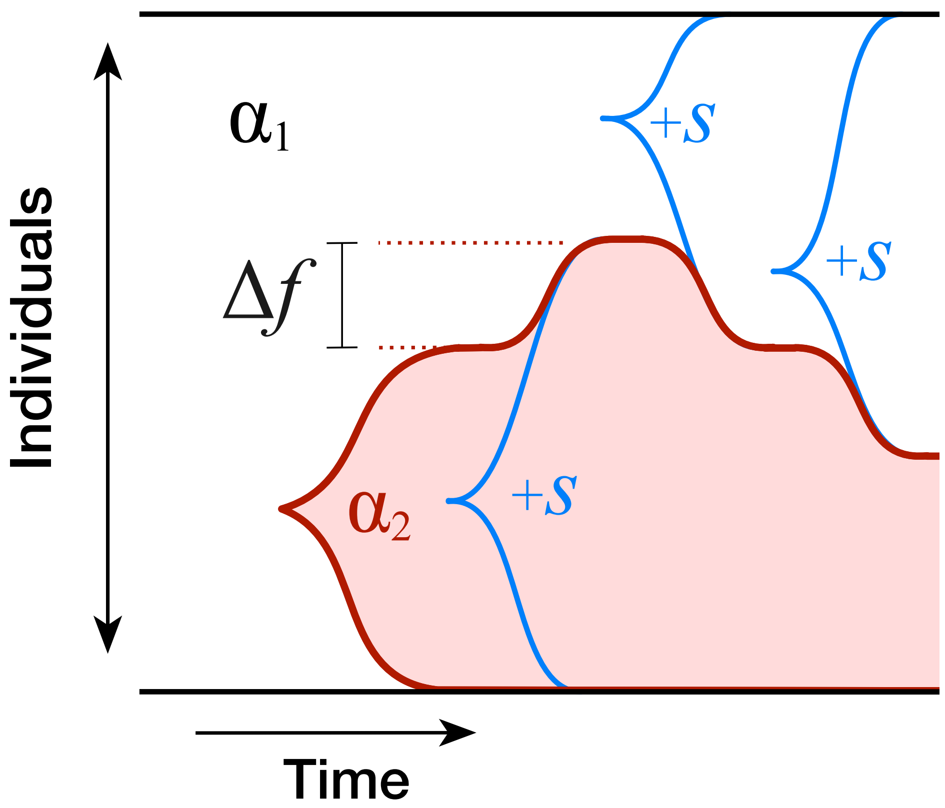
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

Diversification-selection balance



1. Fitness differences perturb equilibria:

$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta\alpha^2}{\beta(1-\beta)}$$ *"Evolutionary robustness"*

2. Fitness "race" between clades *(weak mutation limit)*

$$\frac{\partial \Delta X}{\partial t} = N U_X s^2 \left[2 f^*(\Delta X) - 1\right] + \sqrt{N U_X s^3}\,\eta(t)$$

"Environmental supply vector"

β = % biomass supplied by resource 1

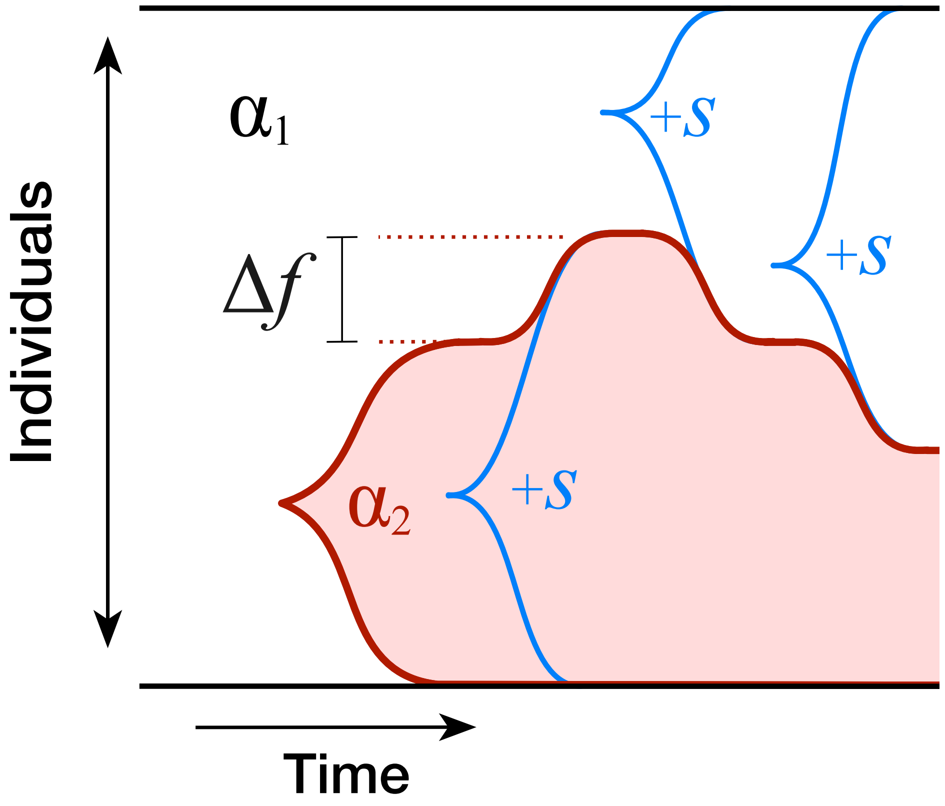
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

Diversification-selection balance



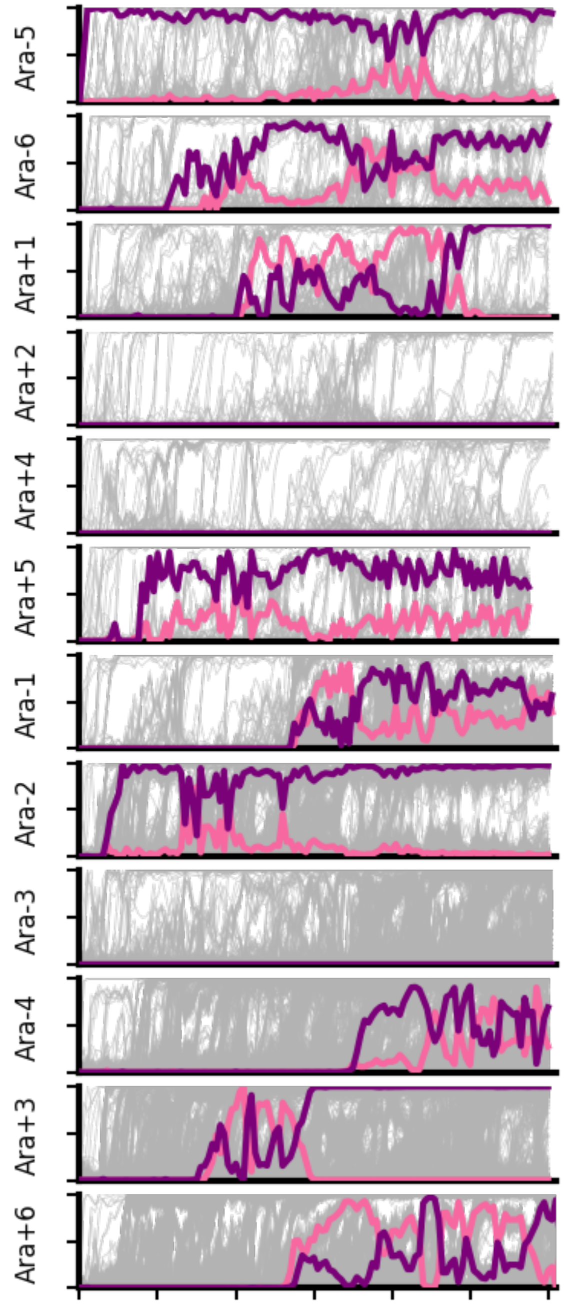
1. Fitness differences perturb equilibria:

$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta\alpha^2}{\beta(1-\beta)}$$ *"Evolutionary robustness"*

2. Fitness "race" between clades *(weak mutation limit)*

$$\frac{\partial \Delta X}{\partial t} = N U_X s^2 \left[2 f^*(\Delta X) - 1\right] + \sqrt{N U_X s^3}\,\eta(t)$$

e.g., Lenski lines



"Environmental supply vector"

β = % biomass supplied by resource 1

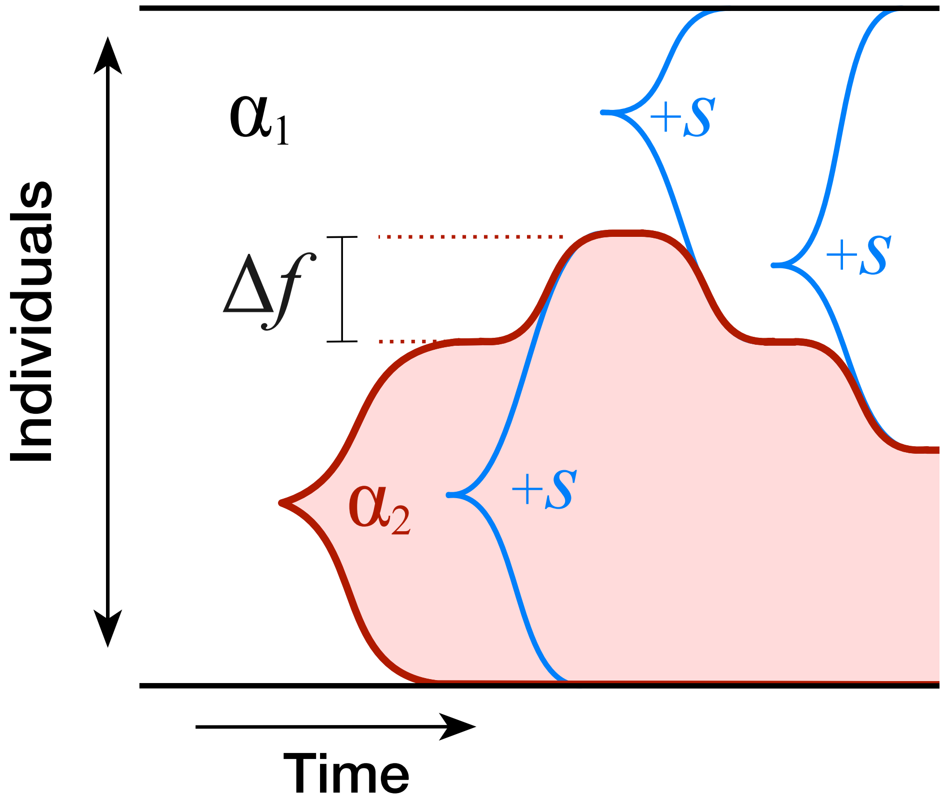
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

Diversification-selection balance



1. Fitness differences perturb equilibria:

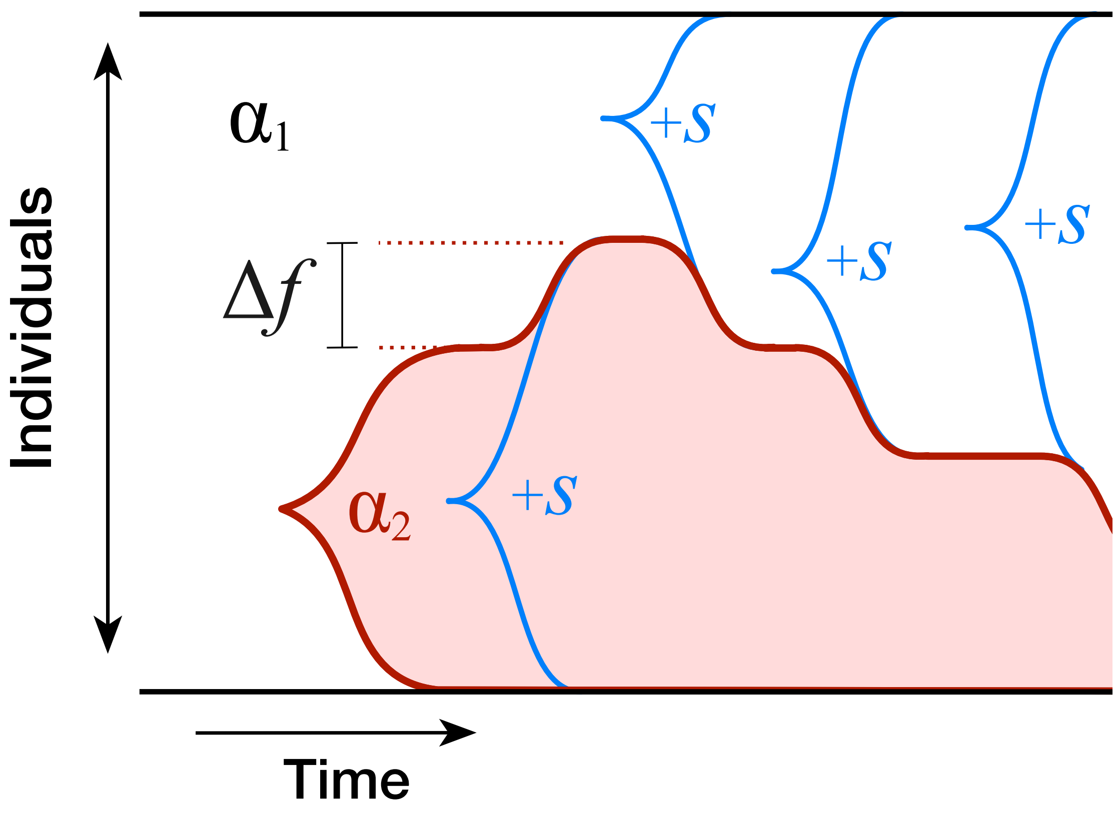
$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta\alpha^2}{\beta(1-\beta)}$$ *"Evolutionary robustness"*

2. Fitness "race" between clades *(weak mutation limit)*

$$\frac{\partial \Delta X}{\partial t} = N U_X s^2 \left[2 f^*(\Delta X) - 1\right] + \sqrt{N U_X s^3}\,\eta(t)$$

"Environmental supply vector"	"Resource strategy"	"General fitness"
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

Diversification-selection balance



1. Fitness differences perturb equilibria:

$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta\alpha^2}{\beta(1-\beta)}$$

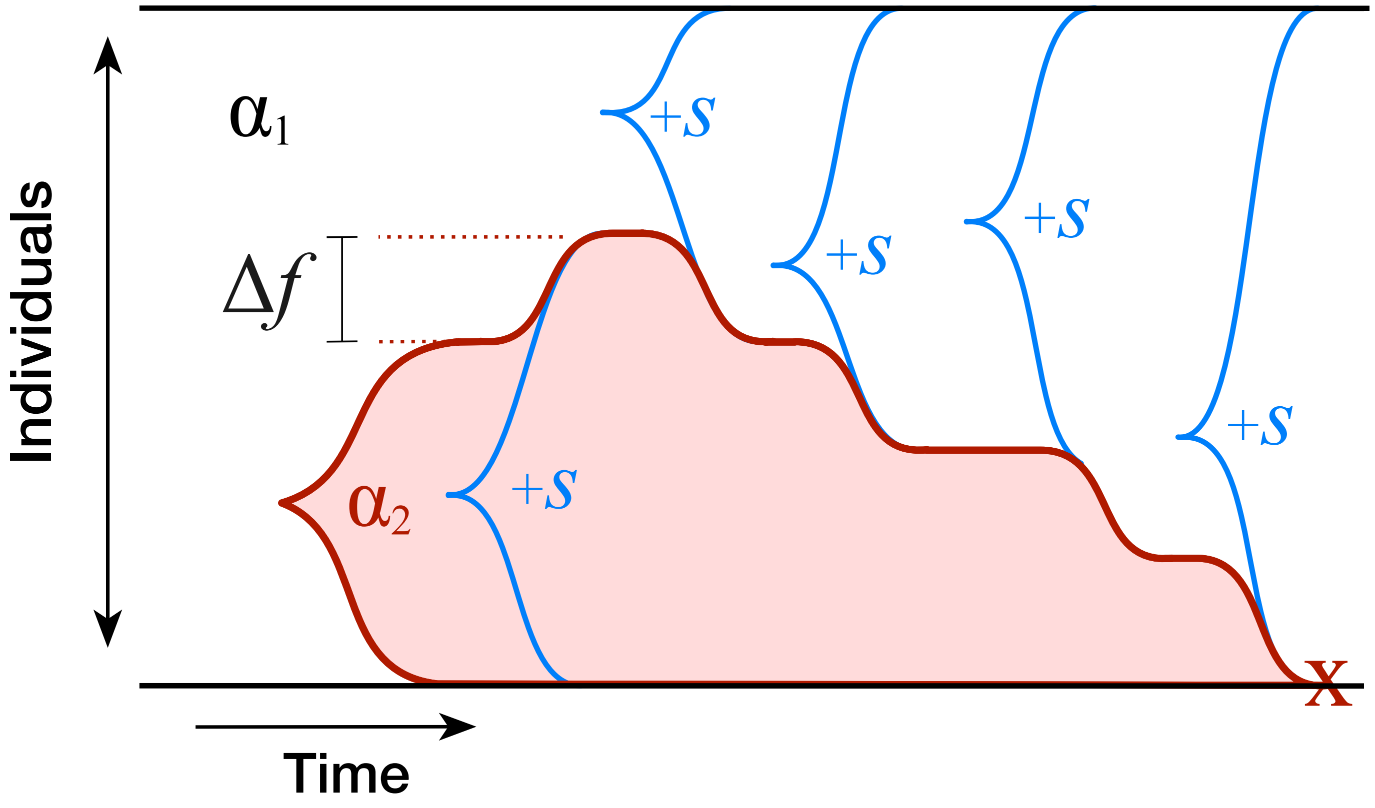
"Evolutionary robustness"

2. Fitness "race" between clades *(weak mutation limit)*

$$\frac{\partial \Delta X}{\partial t} = N U_X s^2 \left[2f^*(\Delta X) - 1\right] + \sqrt{N U_X s^3}\,\eta(t)$$

"Environmental supply vector"	"Resource strategy"	"General fitness"
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

Diversification-selection balance



1. Fitness differences perturb equilibria:

$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta\alpha^2}{\beta(1-\beta)}$$

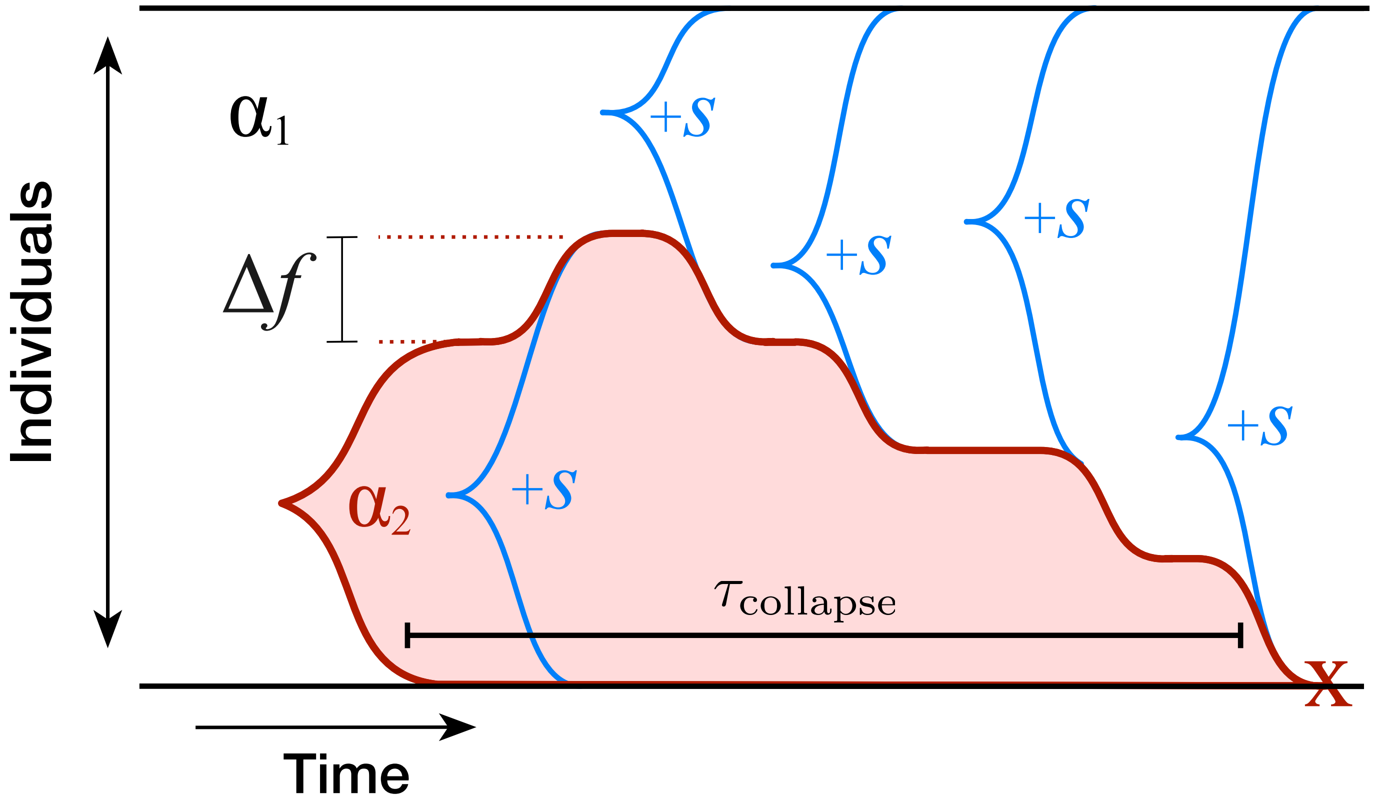
"Evolutionary robustness"

2. Fitness "race" between clades *(weak mutation limit)*

↳
$$\frac{\partial \Delta X}{\partial t} = N U_X s^2 \left[2f^*(\Delta X) - 1\right] + \sqrt{N U_X s^3}\,\eta(t)$$

"Environmental supply vector"	"Resource strategy"	"General fitness"
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

Diversification-selection balance



1. Fitness differences perturb equilibria:

$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta\alpha^2}{\beta(1-\beta)}$$

"Evolutionary robustness"

2. Fitness "race" between clades *(weak mutation limit)*

$$\frac{\partial \Delta X}{\partial t} = N U_X s^2 \left[2 f^*(\Delta X) - 1\right] + \sqrt{N U_X s^3}\,\eta(t)$$

"Environmental supply vector"

β = % biomass supplied by resource 1

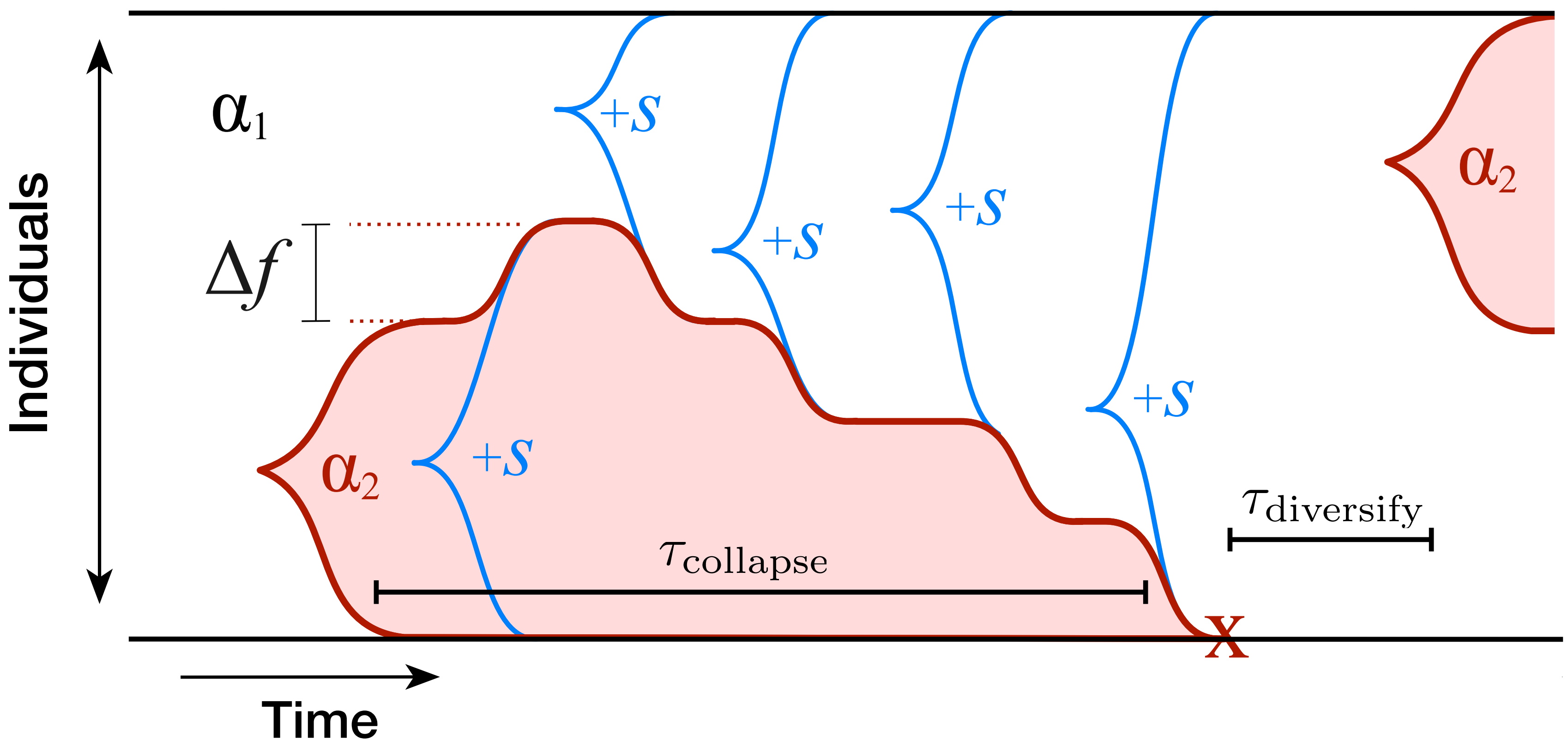
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

Diversification-selection balance



1. Fitness differences perturb equilibria:

$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta\alpha^2}{\beta(1-\beta)}$$

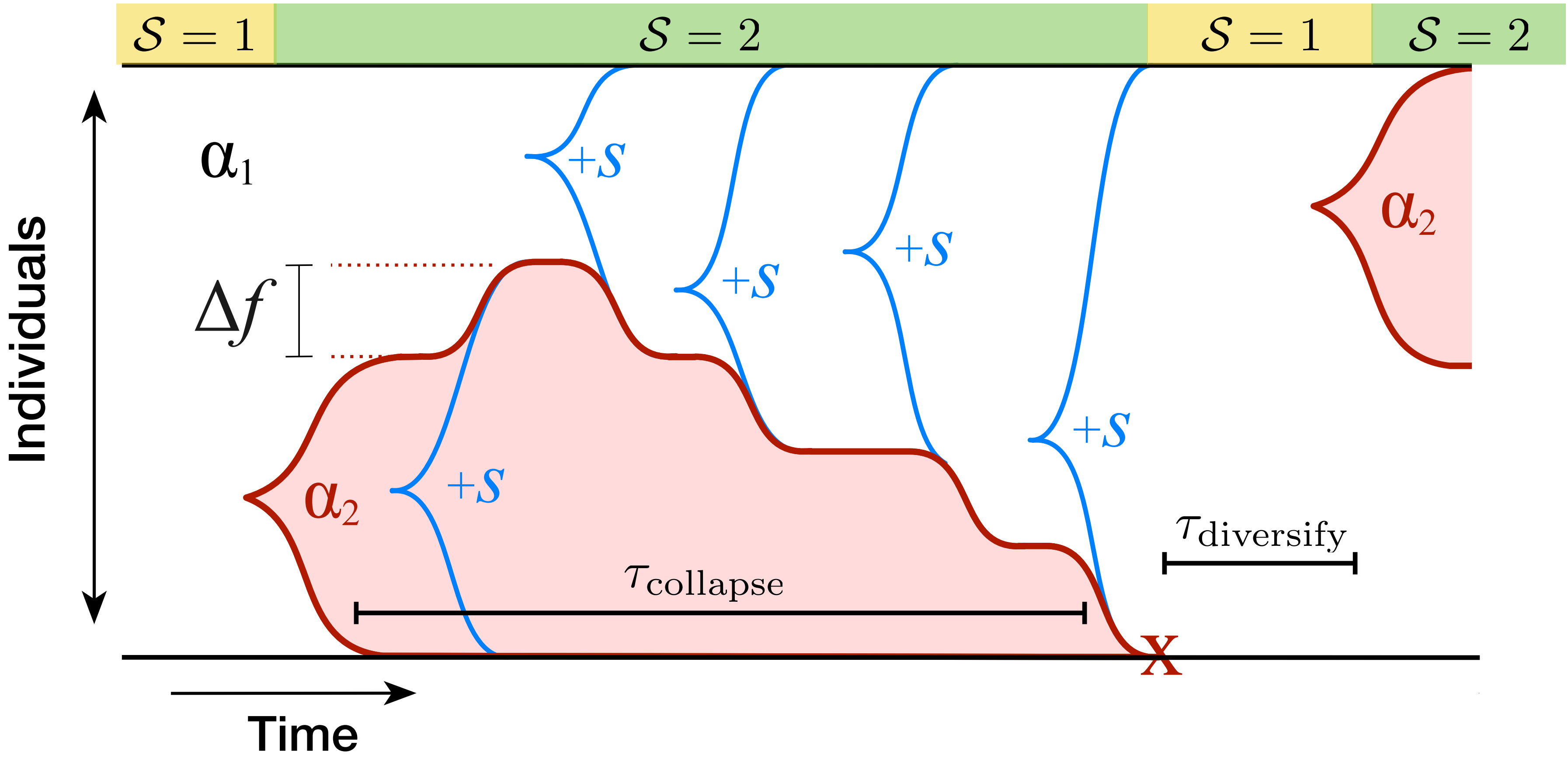
"Evolutionary robustness"

2. Fitness "race" between clades *(weak mutation limit)*

$$\frac{\partial \Delta X}{\partial t} = N U_X s^2 \left[2 f^*(\Delta X) - 1\right] + \sqrt{N U_X s^3}\,\eta(t)$$

"Environmental supply vector"	*"Resource strategy"*	*"General fitness"*
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

Diversification-selection balance



1. Fitness differences perturb equilibria:

$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta\alpha^2}{\beta(1-\beta)}$$

"Evolutionary robustness"

2. Fitness "race" between clades *(weak mutation limit)*

↳
$$\frac{\partial \Delta X}{\partial t} = N U_X s^2 \left[2 f^*(\Delta X) - 1\right] + \sqrt{N U_X s^3}\,\eta(t)$$

Diversification-selection balance

$$\frac{\Pr[\mathcal{S} = 2]}{\Pr[\mathcal{S} = 1]} \sim \frac{\tau_{\text{collapse}}}{\tau_{\text{diversify}}}$$

"Environmental supply vector"

β = % biomass supplied by resource 1

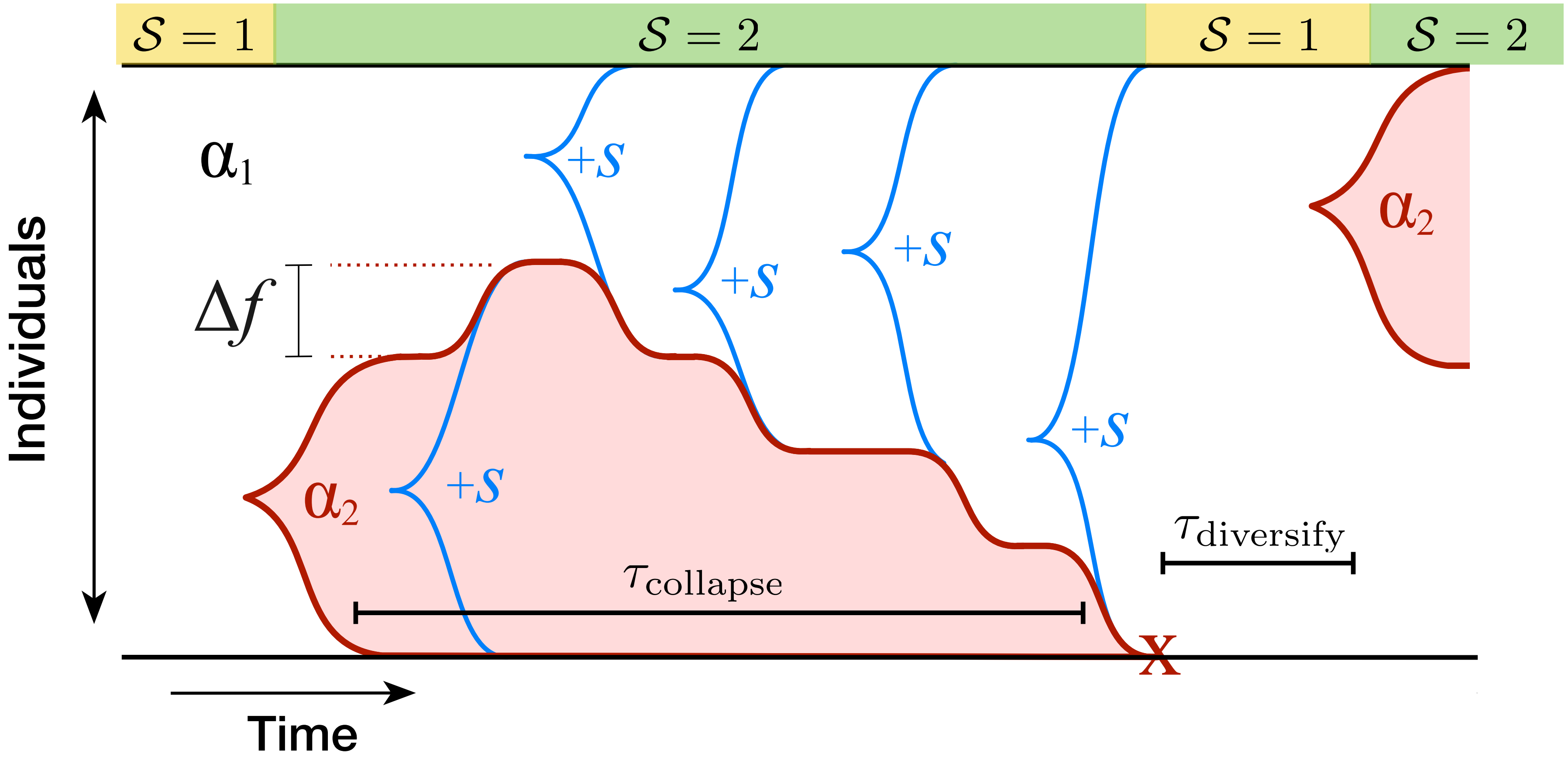
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

Diversification-selection balance



1. Fitness differences perturb equilibria:

$$\Delta f \approx \frac{\Delta X}{X^*}; \quad X^* \approx \frac{\Delta \alpha^2}{\beta(1-\beta)}$$

"Evolutionary robustness"

2. Fitness "race" between clades *(weak mutation limit)*

↳ $$\frac{\partial \Delta X}{\partial t} = N U_X s^2 \left[2 f^*(\Delta X) - 1\right] + \sqrt{N U_X s^3}\,\eta(t)$$

Diversification-selection balance

$$\frac{\Pr[\mathcal{S}=2]}{\Pr[\mathcal{S}=1]} \sim \frac{\tau_{\text{collapse}}}{\tau_{\text{diversify}}}$$

E.g., $$\sim \frac{U_\alpha}{U_X}\left(\frac{X^*}{s}\right)^2 \log\left(\frac{X^*}{s}\right)$$

"Environmental supply vector"

β = % biomass supplied by resource 1

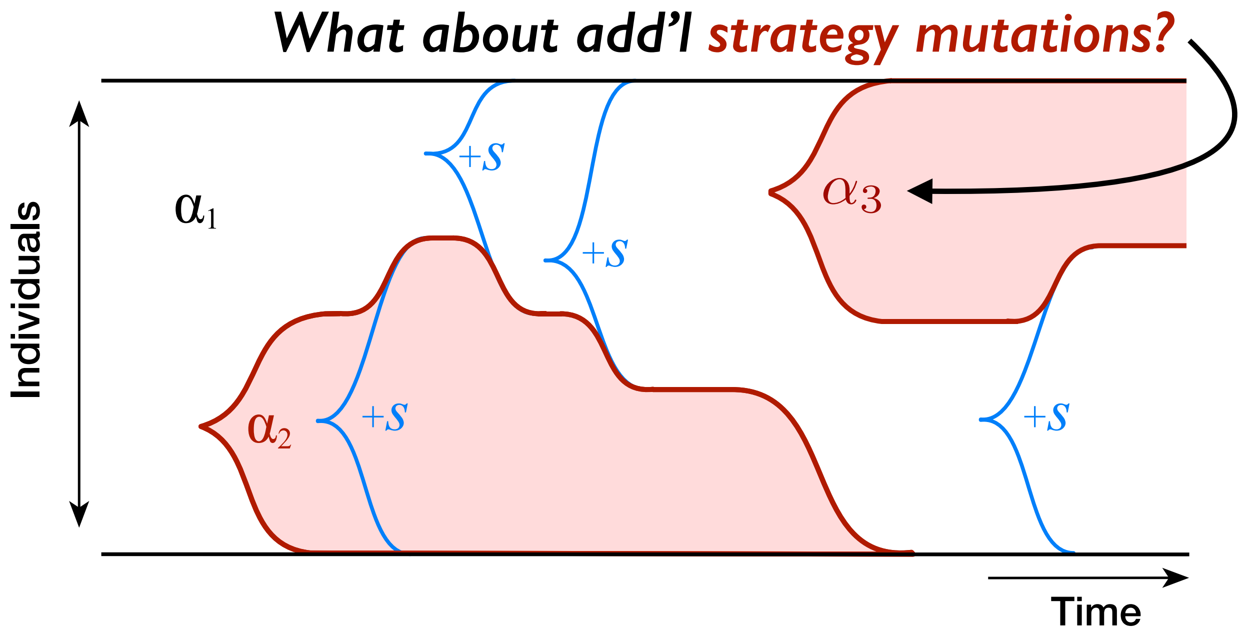
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

Emergent selection for further ecological tuning



"Environmental supply vector"

β = % biomass supplied by resource 1

"Resource strategy"

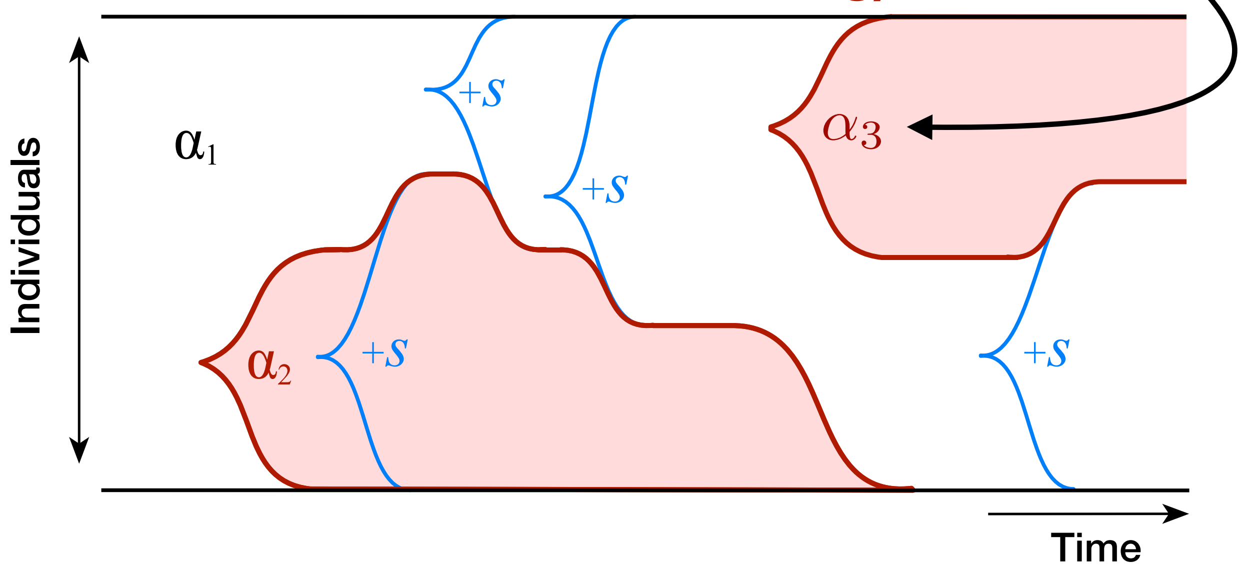
α_μ = % energy strain μ spends to import resource 1

"General fitness"

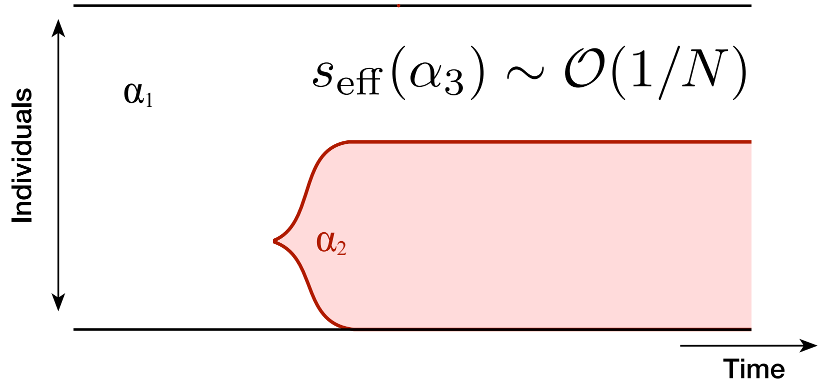
X_μ = max growth rate of strain μ

Emergent selection for further ecological tuning

What about add'l strategy mutations?



Recall neutral case: *(Posfai et al 2017)*



$s_{\text{eff}}(\alpha_3) \sim \mathcal{O}(1/N)$

"Environmental supply vector"

β = % biomass supplied by resource 1

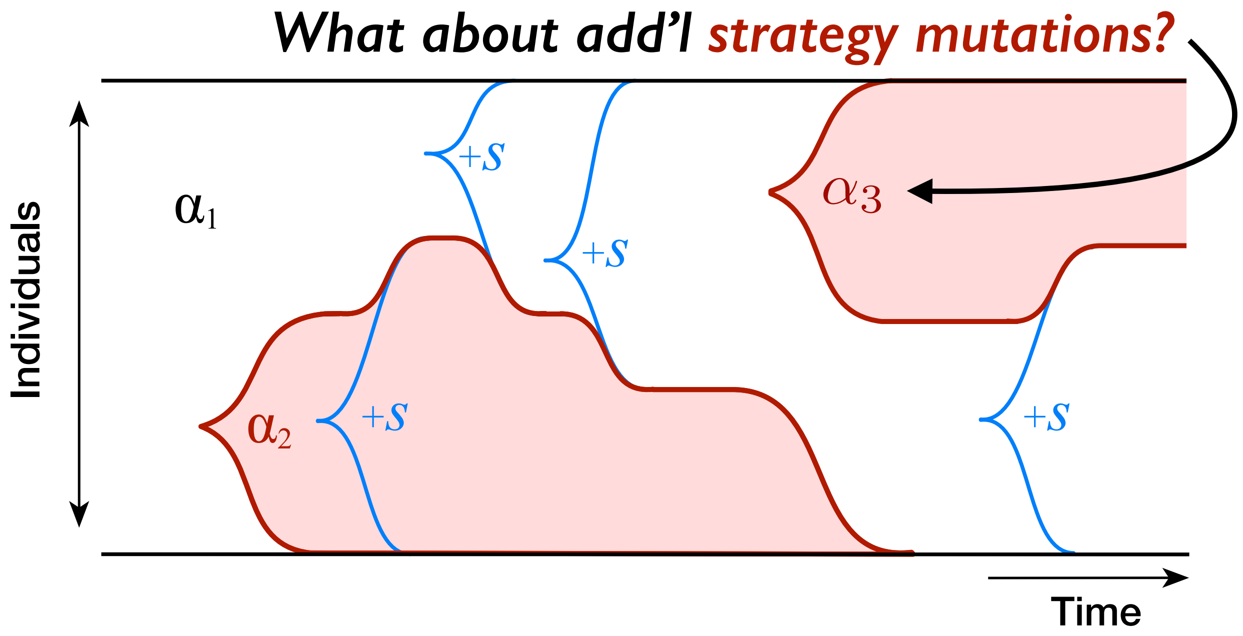
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

Emergent selection for further ecological tuning



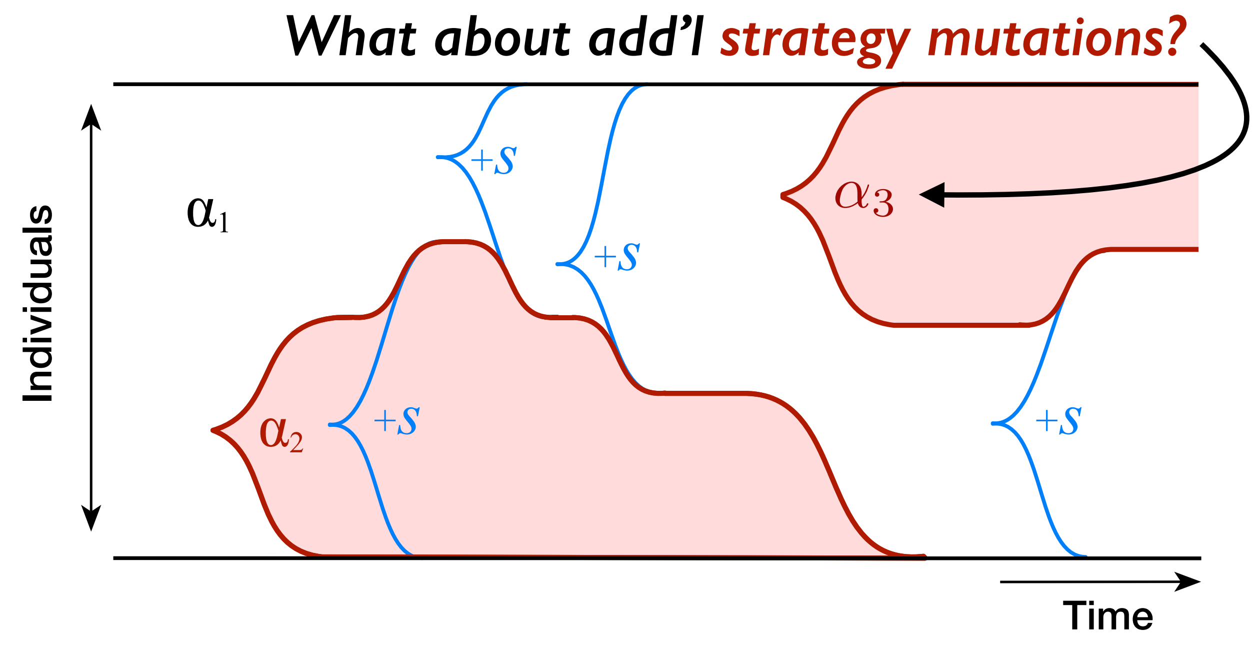
Induced ecological selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_3) \approx \frac{\alpha_3 - \alpha_1}{\alpha_2 - \alpha_1} \cdot (X_1 - X_2)$$

"*Environmental supply vector*"	"*Resource strategy*"	"*General fitness*"
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

Emergent selection for further ecological tuning

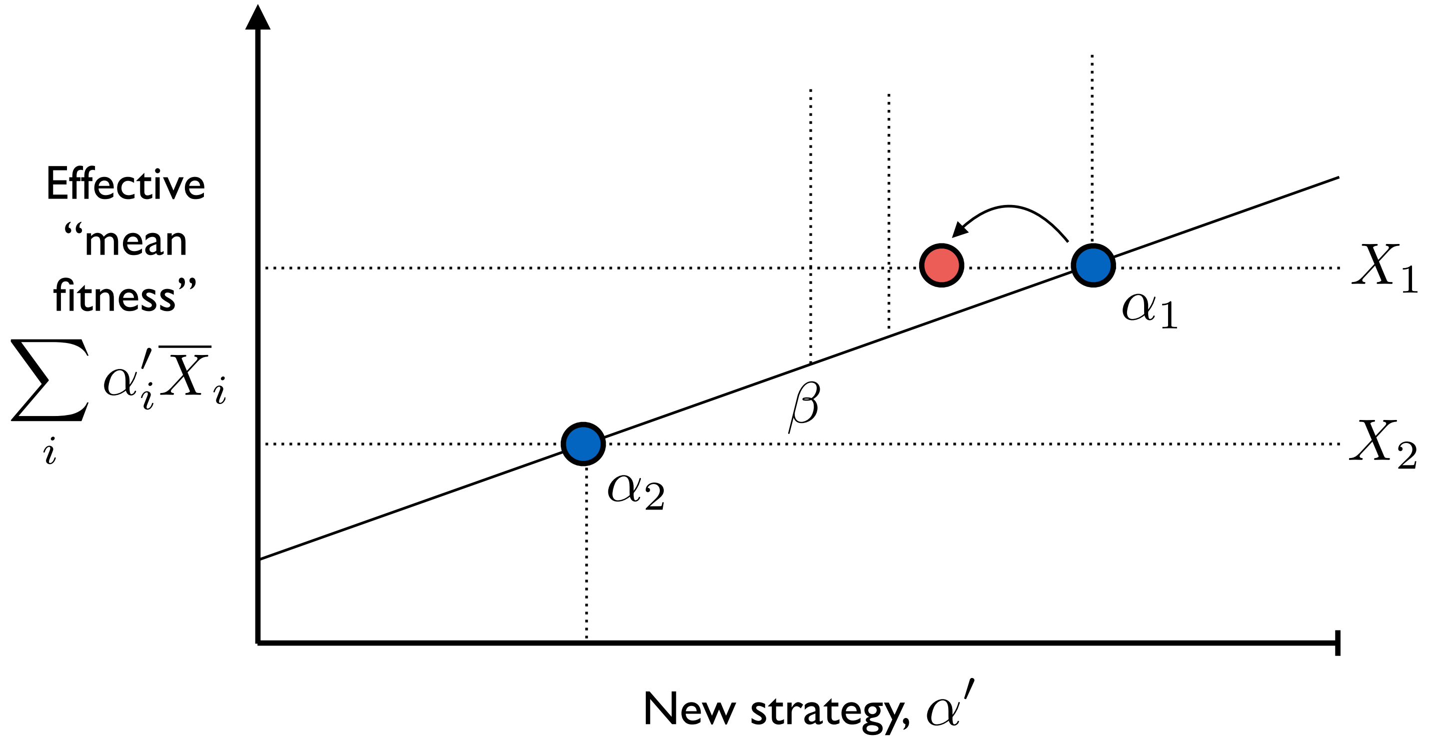
What about add'l strategy mutations?



Induced ecological selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_3) \approx \frac{\alpha_3 - \alpha_1}{\alpha_2 - \alpha_1} \cdot (X_1 - X_2)$$

Graphical picture



"Environmental supply vector"

β = % biomass supplied by resource 1

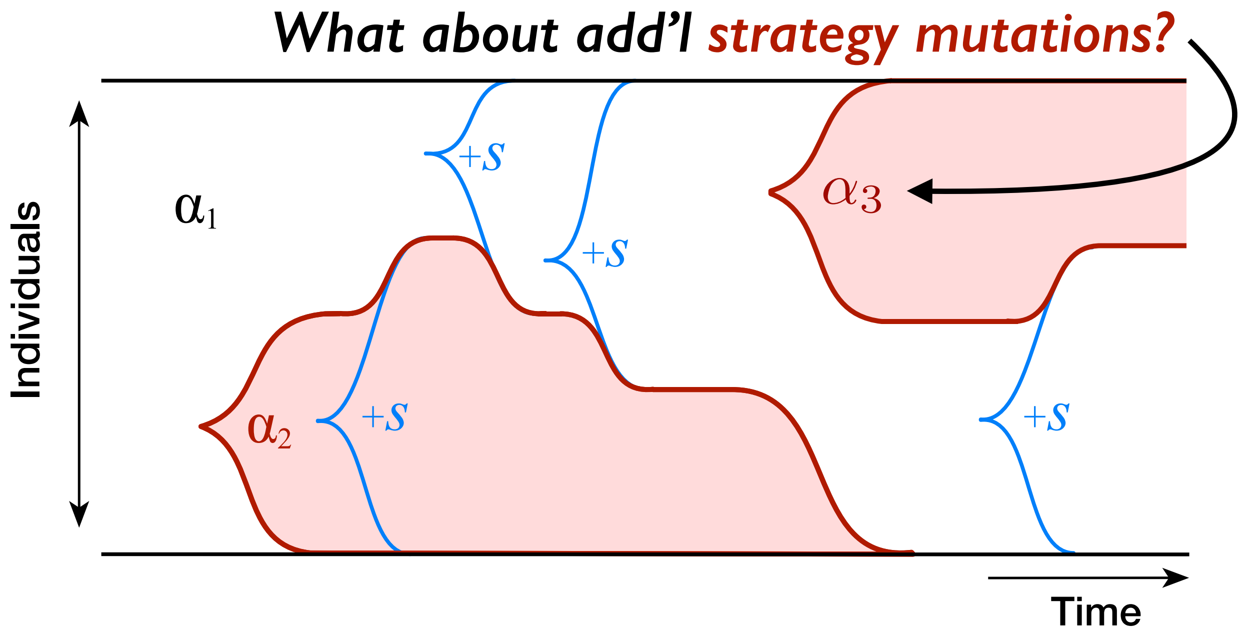
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

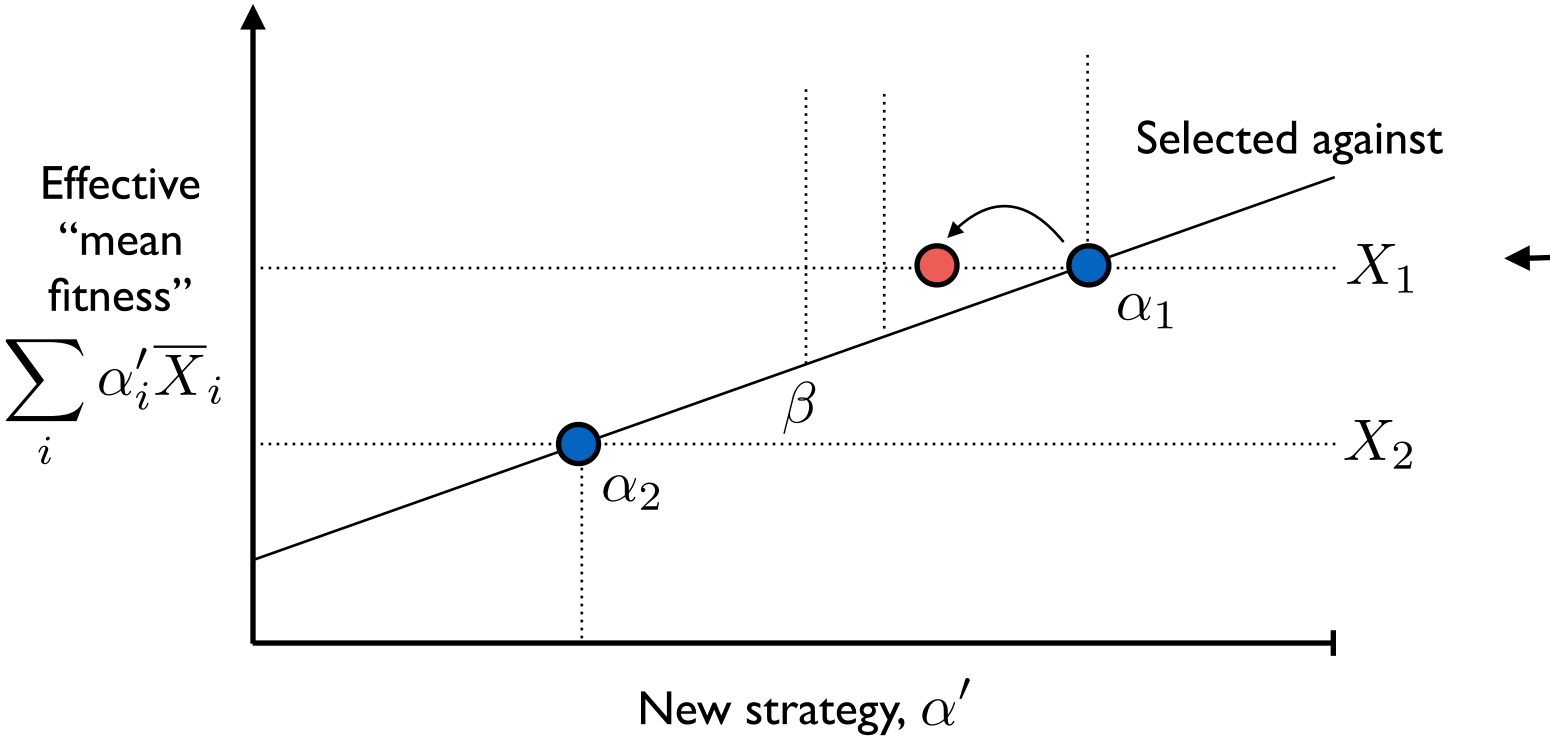
Emergent selection for further ecological tuning



Induced ecological selection pressure:

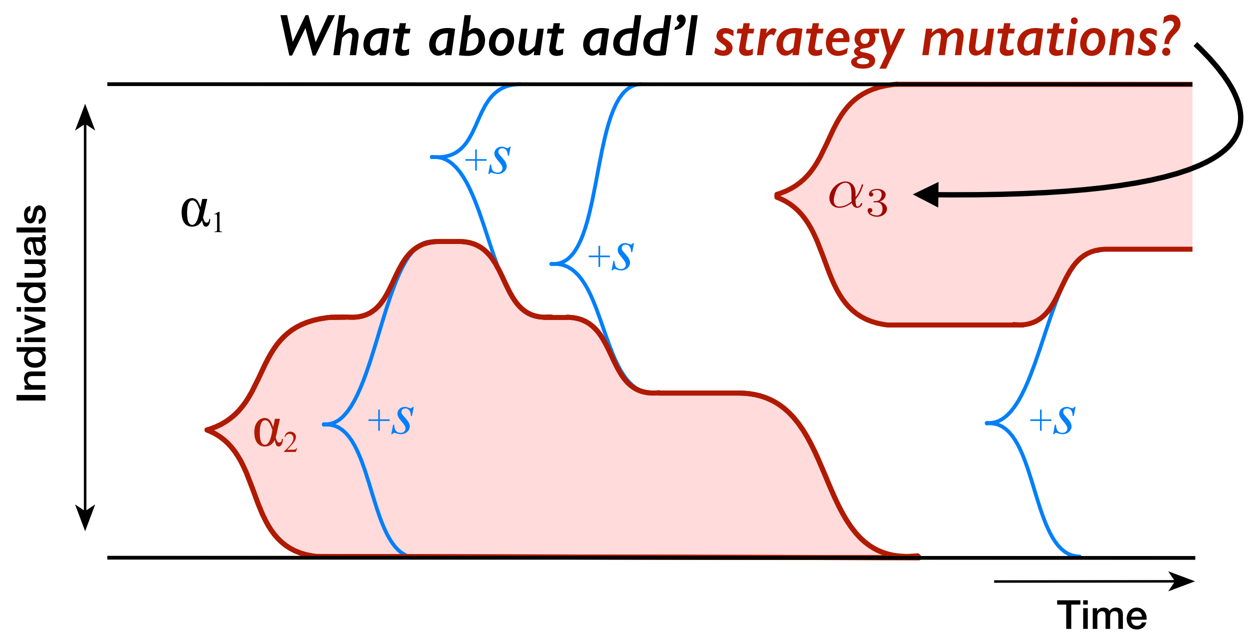
$$s_{\text{eff}}(\alpha_1 \to \alpha_3) \approx \frac{\alpha_3 - \alpha_1}{\alpha_2 - \alpha_1} \cdot (X_1 - X_2)$$

Graphical picture



"Environmental supply vector"	*"Resource strategy"*	*"General fitness"*
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

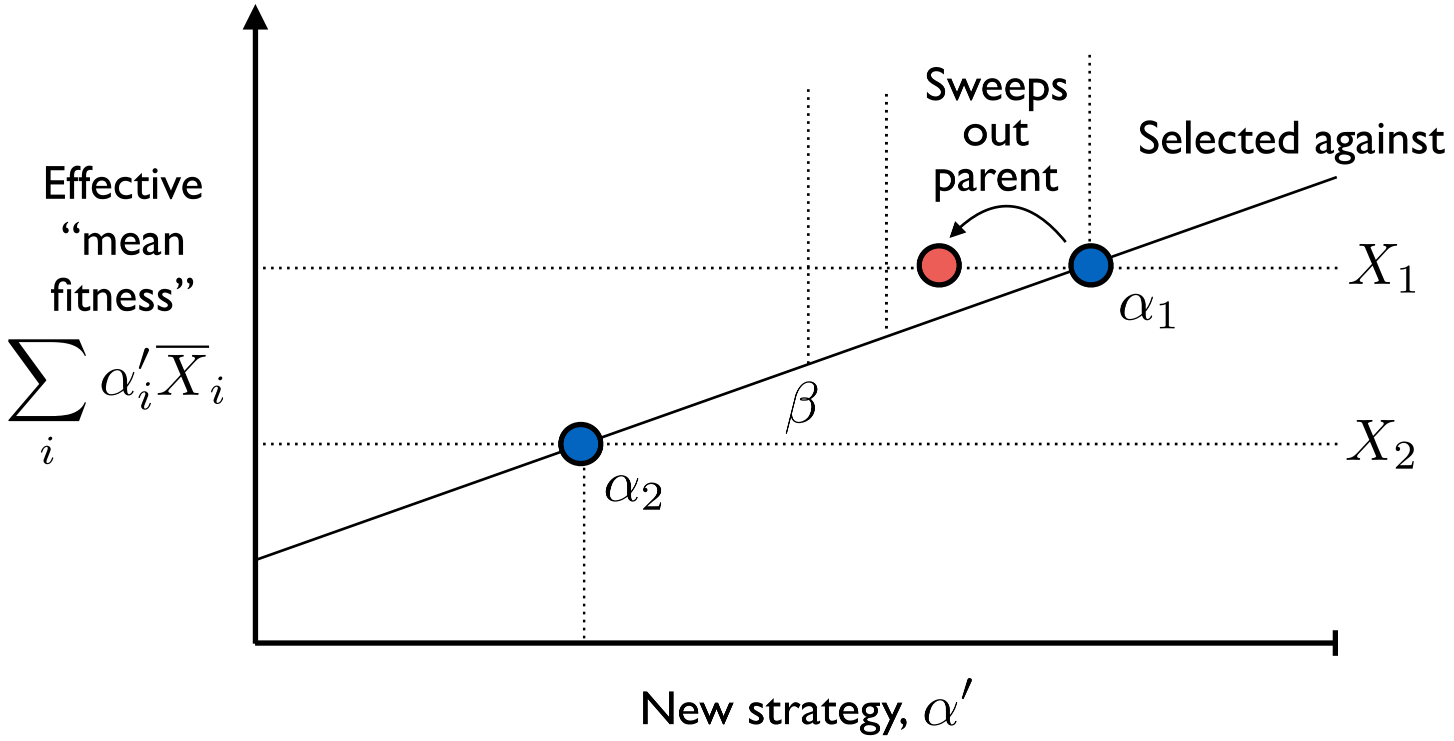
Emergent selection for further ecological tuning



Induced ecological selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_3) \approx \frac{\alpha_3 - \alpha_1}{\alpha_2 - \alpha_1} \cdot (X_1 - X_2)$$

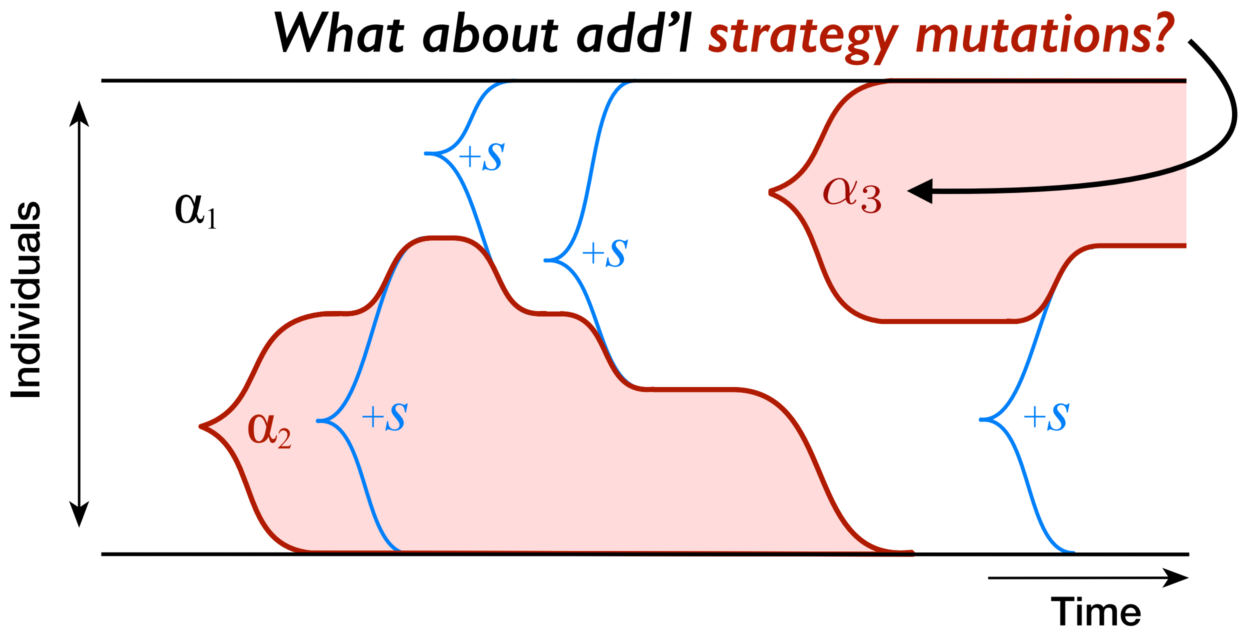
Graphical picture



"Environmental supply vector"	*"Resource strategy"*	*"General fitness"*
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

Emergent selection for further ecological tuning

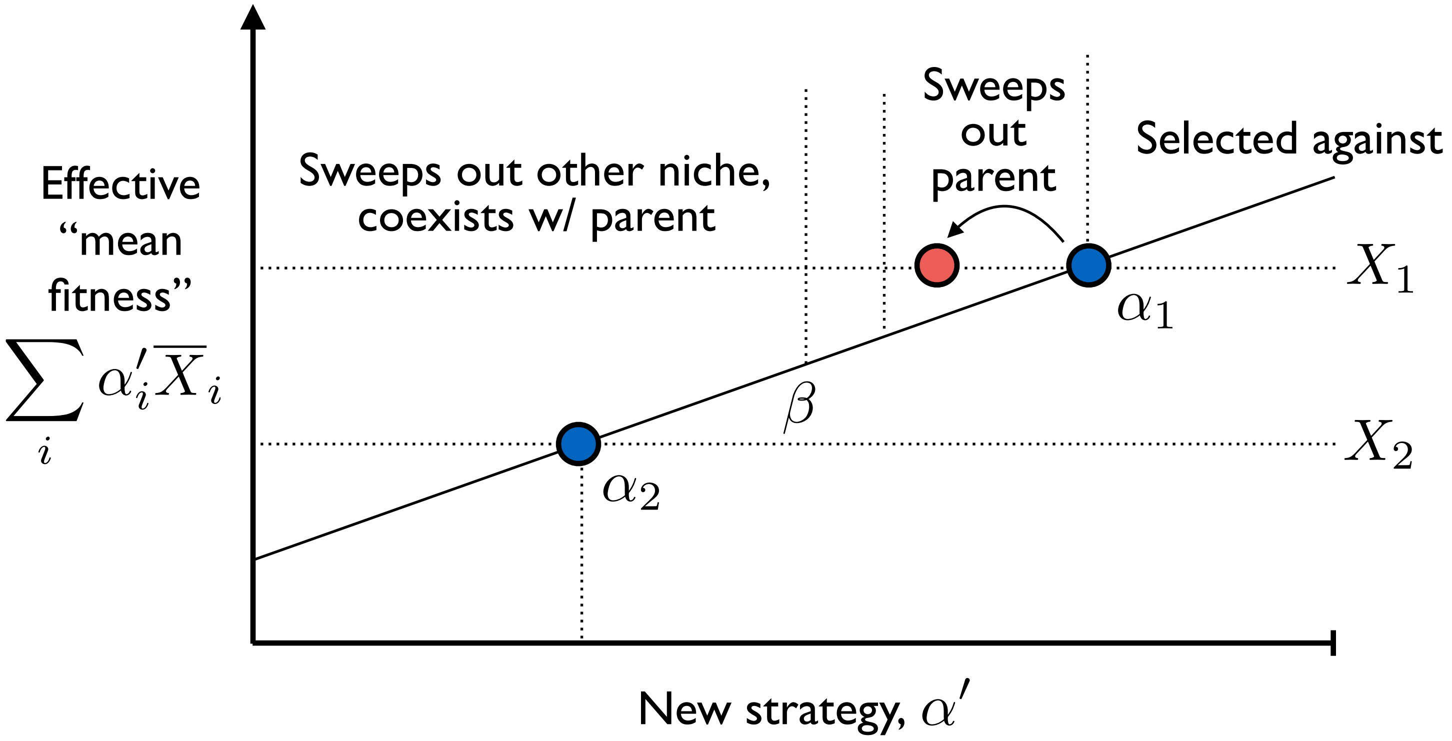
What about add'l strategy mutations?



Induced ecological selection pressure:

$$s_{\mathrm{eff}}(\alpha_1 \to \alpha_3) \approx \frac{\alpha_3 - \alpha_1}{\alpha_2 - \alpha_1} \cdot (X_1 - X_2)$$

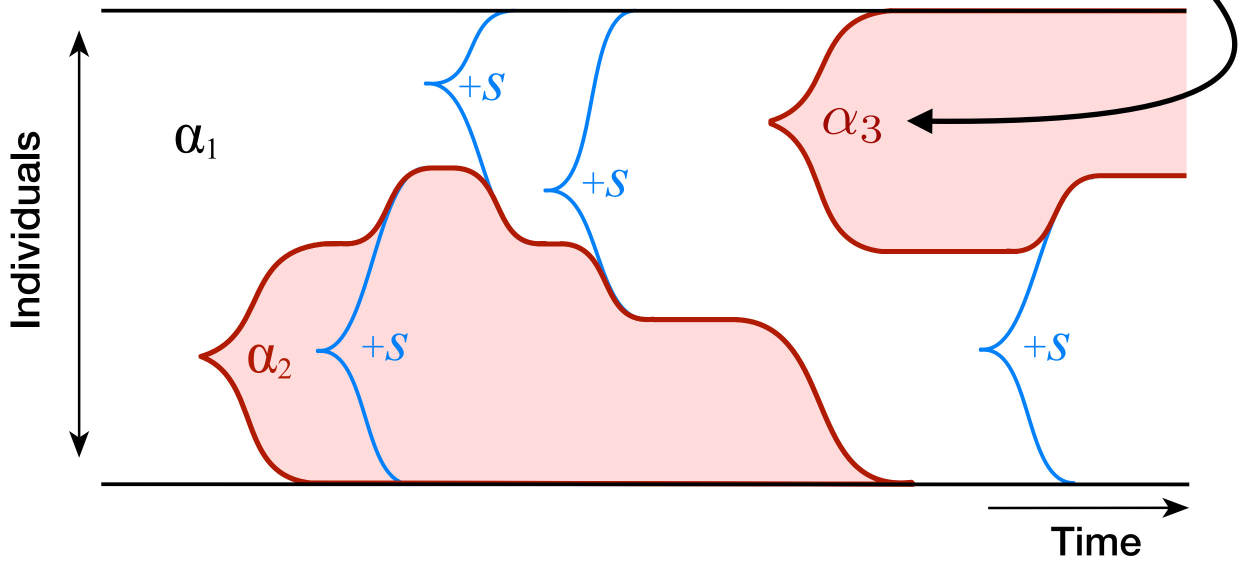
Graphical picture



"Environmental supply vector"	*"Resource strategy"*	*"General fitness"*
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

Emergent selection for further ecological tuning

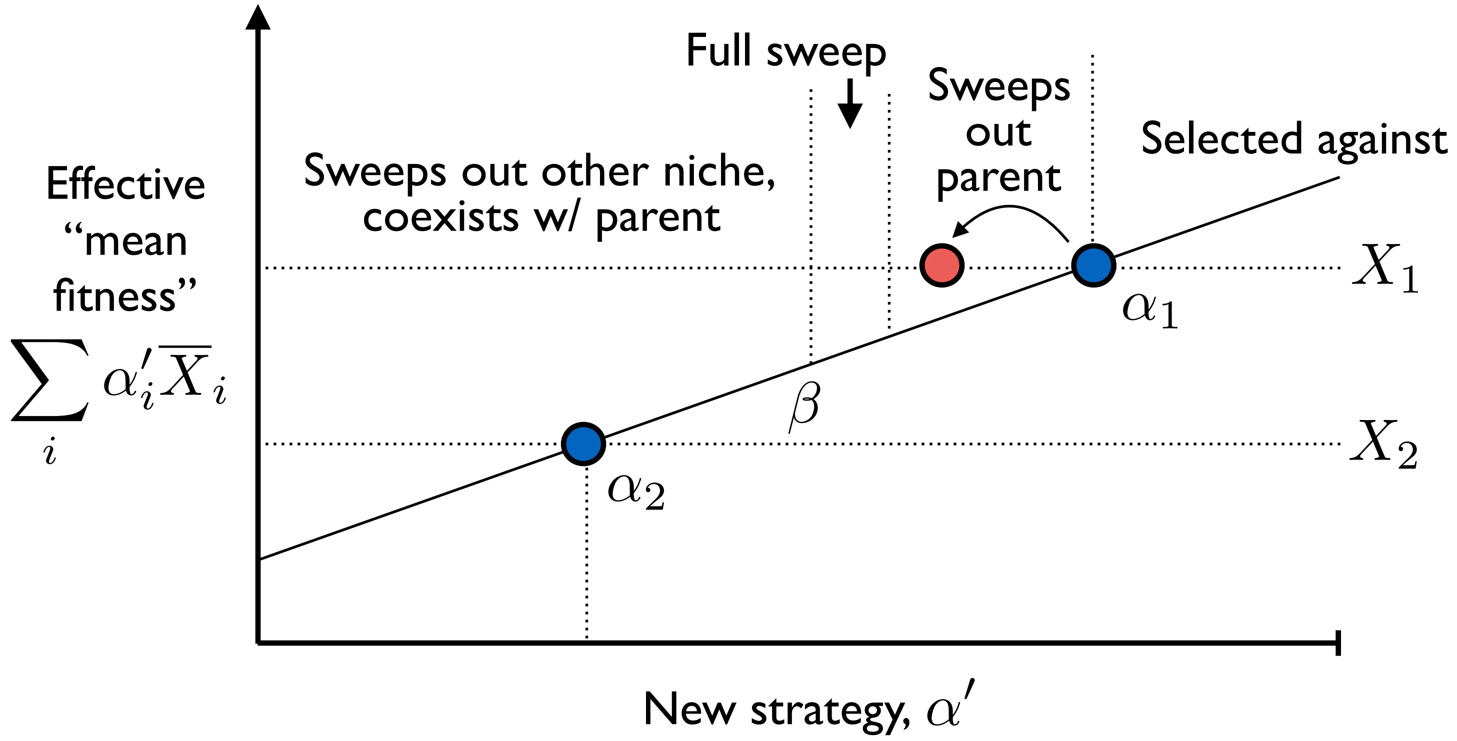
What about add'l strategy mutations?



Induced ecological selection pressure:

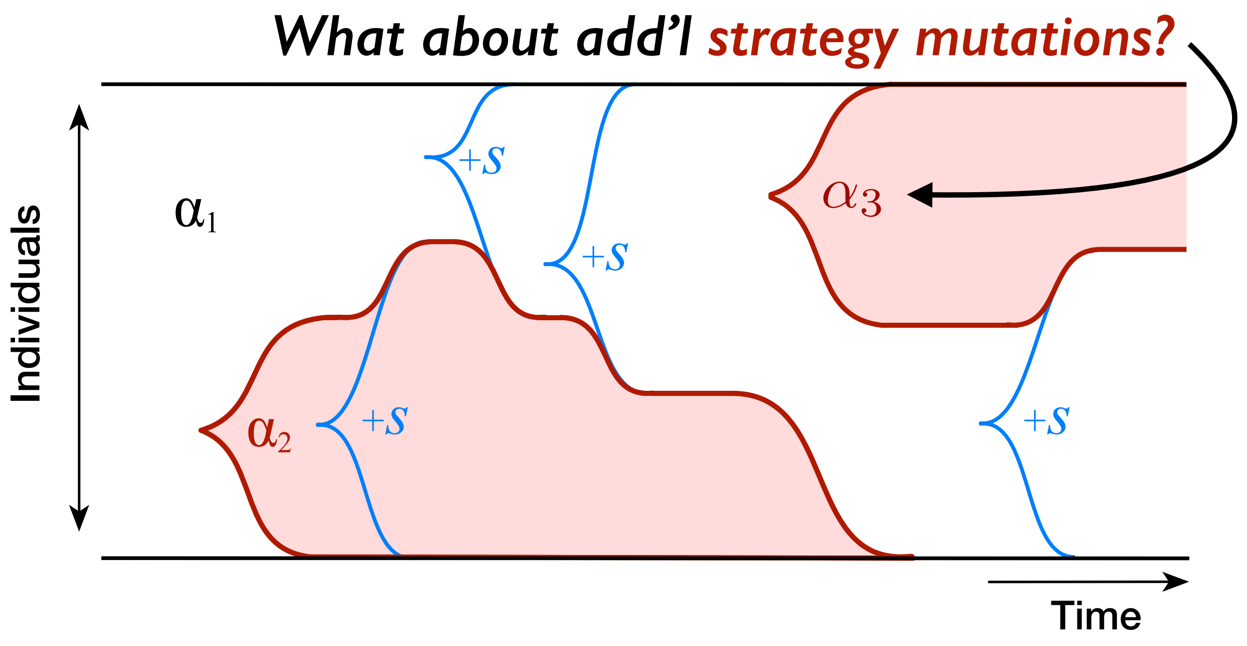
$$s_{\text{eff}}(\alpha_1 \to \alpha_3) \approx \frac{\alpha_3 - \alpha_1}{\alpha_2 - \alpha_1} \cdot (X_1 - X_2)$$

Graphical picture



"Environmental supply vector"	*"Resource strategy"*	*"General fitness"*
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

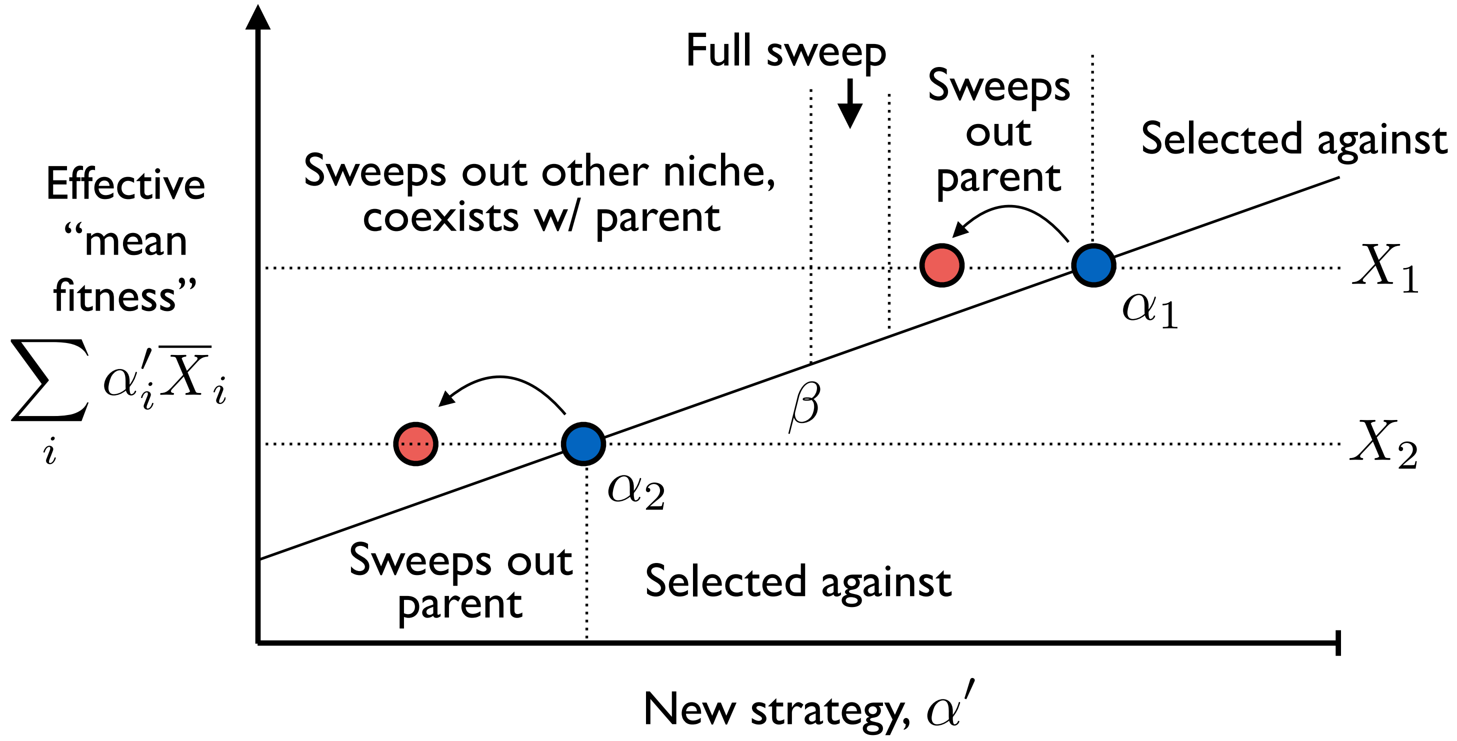
Emergent selection for further ecological tuning



Induced ecological selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_3) \approx \frac{\alpha_3 - \alpha_1}{\alpha_2 - \alpha_1} \cdot (X_1 - X_2)$$

Graphical picture



"Environmental supply vector"

β = % biomass supplied by resource 1

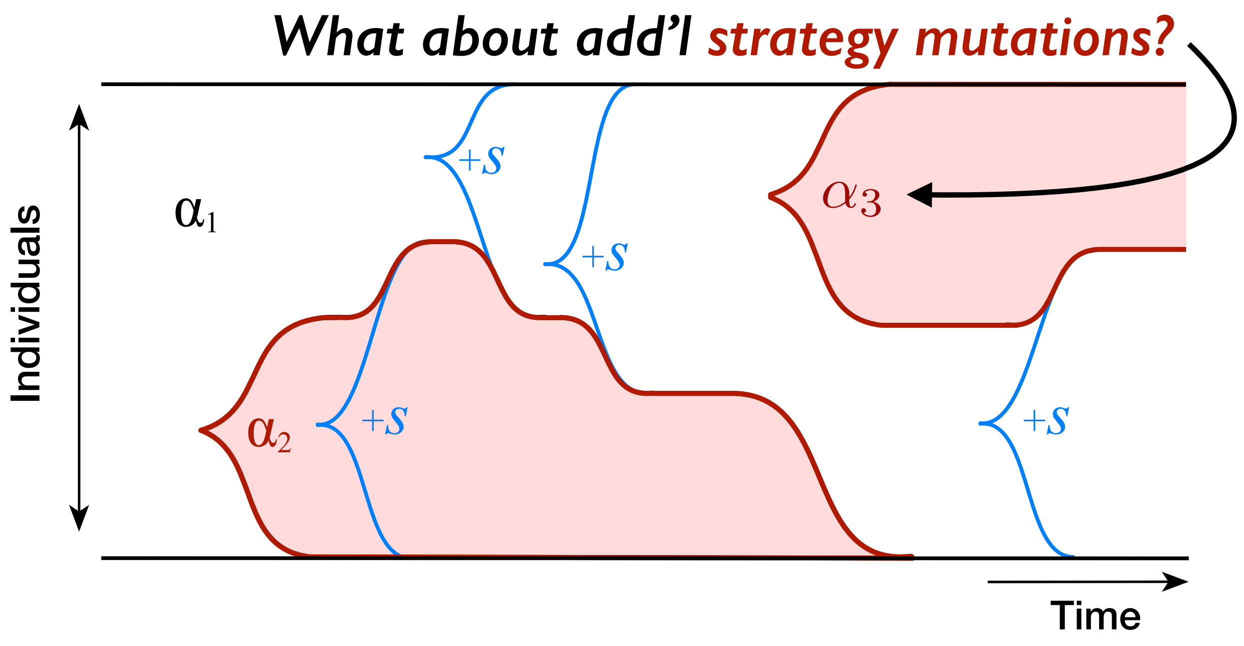
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

"General fitness"

X_μ = max growth rate of strain μ

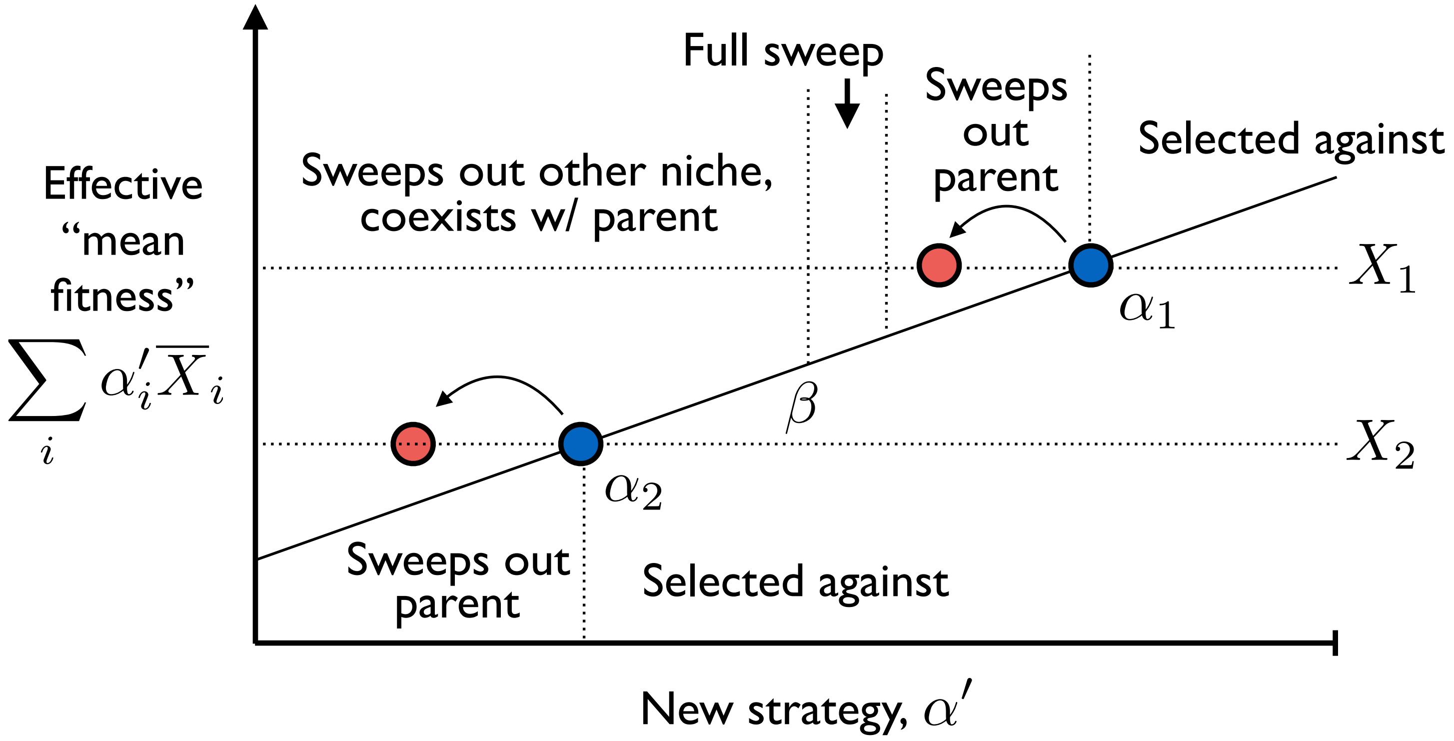
Emergent selection for further ecological tuning



Induced ecological selection pressure:

$$s_{\mathrm{eff}}(\alpha_1 \to \alpha_3) \approx \frac{\alpha_3 - \alpha_1}{\alpha_2 - \alpha_1} \cdot (X_1 - X_2)$$

Graphical picture

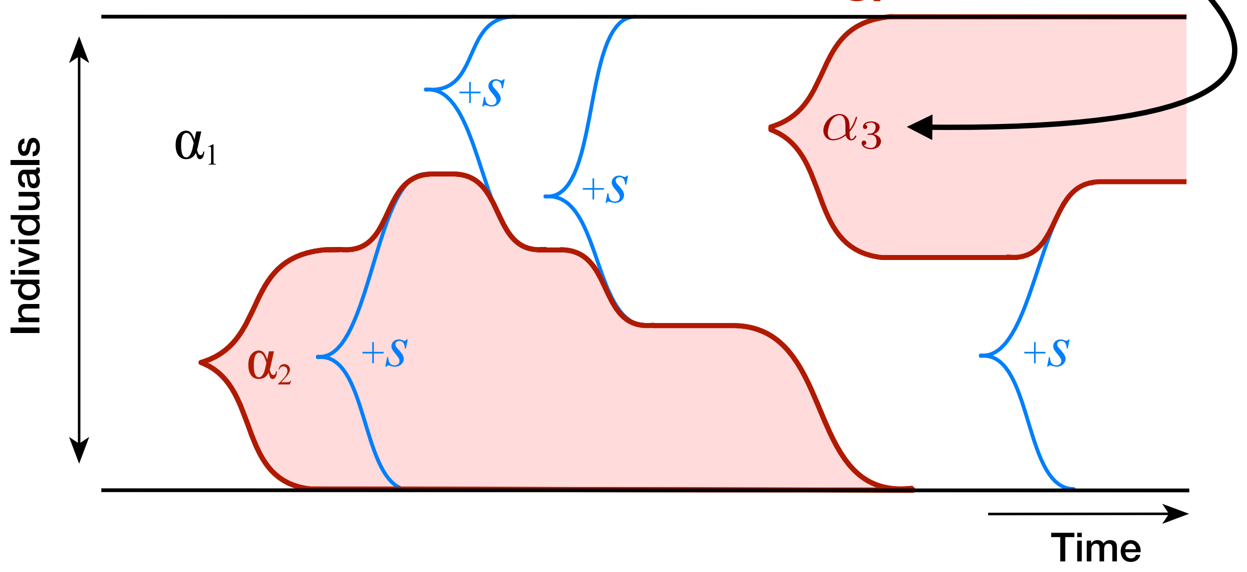


Selection for fit "generalists" and less-fit "specialists"

"Environmental supply vector"	"Resource strategy"	"General fitness"
β = % biomass supplied by resource 1	α_μ = % energy strain μ spends to import resource 1	X_μ = max growth rate of strain μ

Emergent selection for further ecological tuning

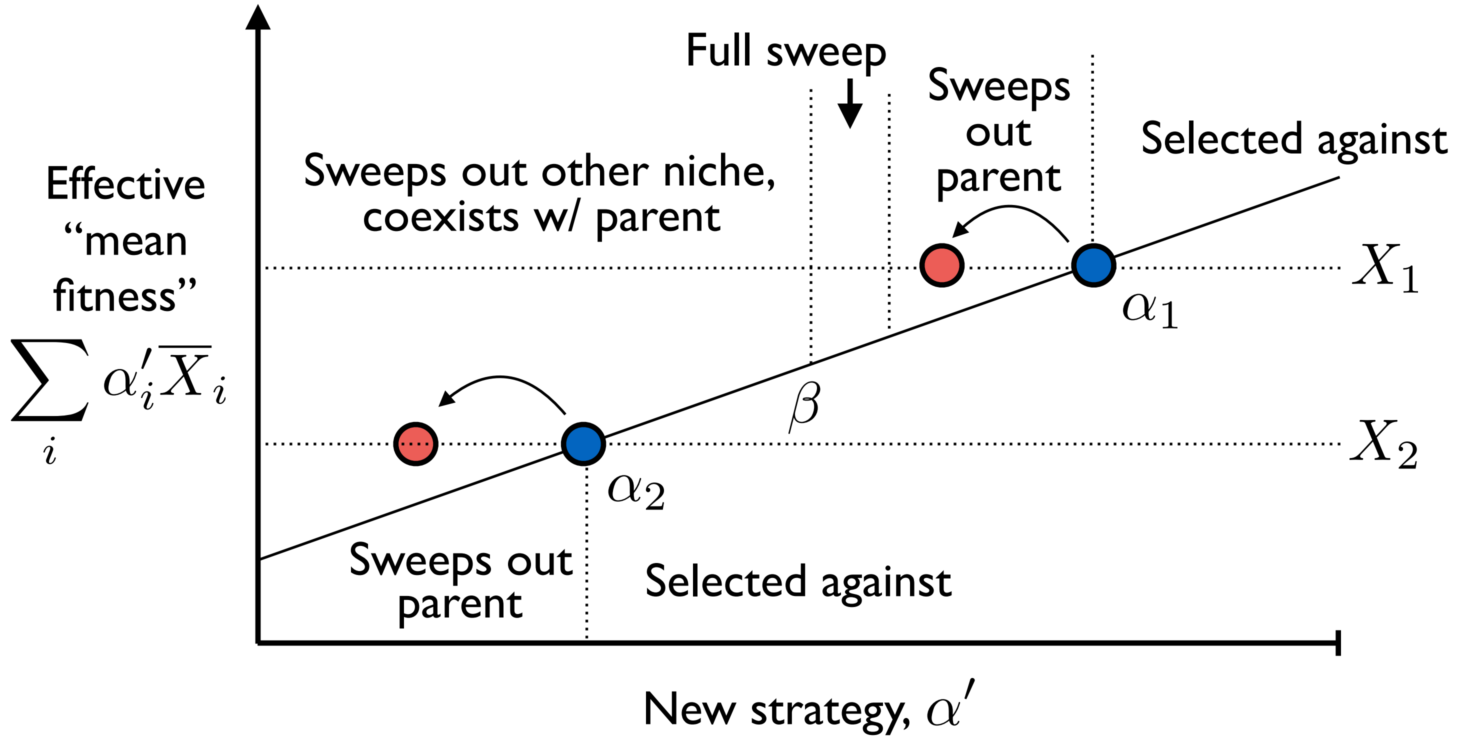
What about add'l strategy mutations?



Induced ecological selection pressure:

$$s_{\text{eff}}(\alpha_1 \to \alpha_3) \approx \frac{\alpha_3 - \alpha_1}{\alpha_2 - \alpha_1} \cdot (X_1 - X_2)$$

Graphical picture



Selection for fit "generalists" and less-fit "specialists"

Test in laboratory experiments?

"Environmental supply vector"

β = % biomass supplied by resource 1

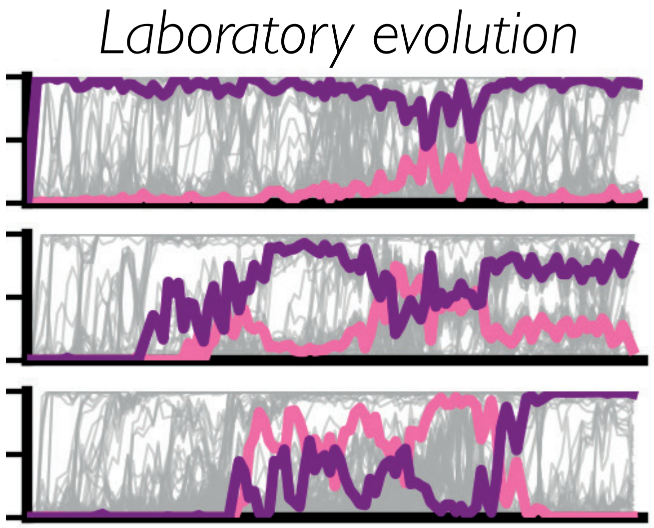
"Resource strategy"

α_μ = % energy strain μ spends to import resource 1

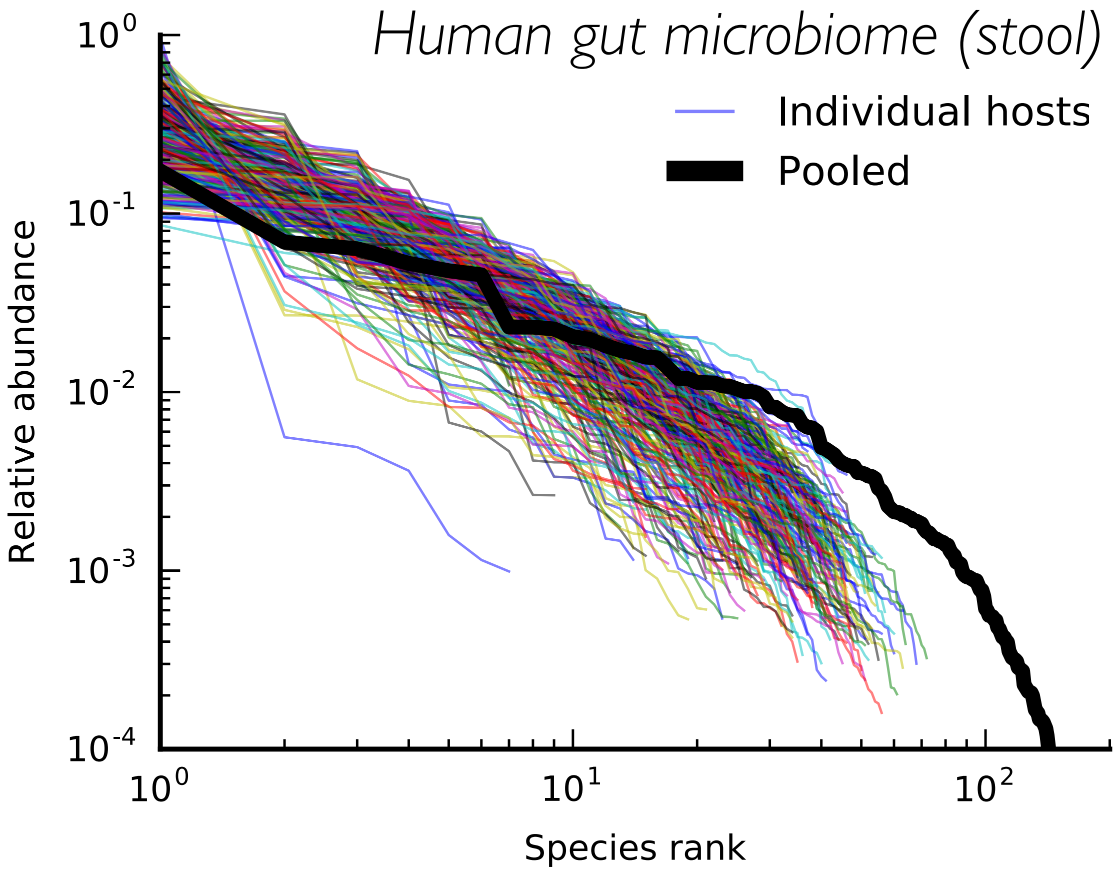
"General fitness"

X_μ = max growth rate of strain μ

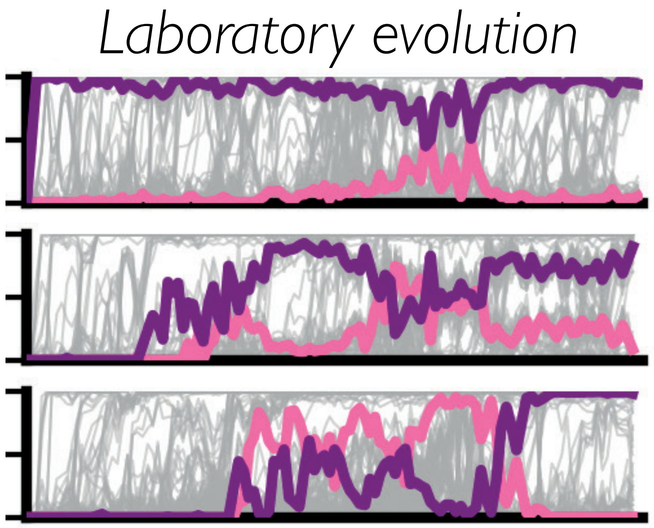
Future directions: evolution in highly diverse ecosystems



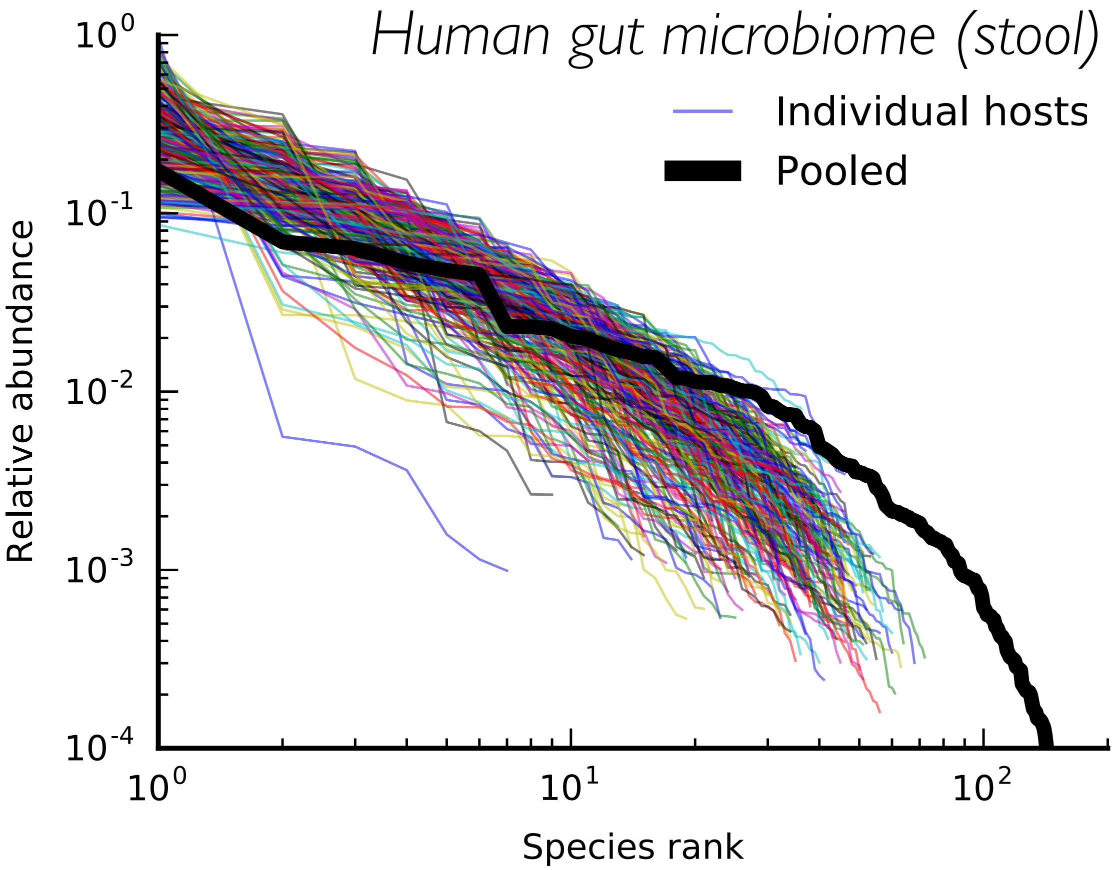
VS →



Future directions: evolution in highly diverse ecosystems

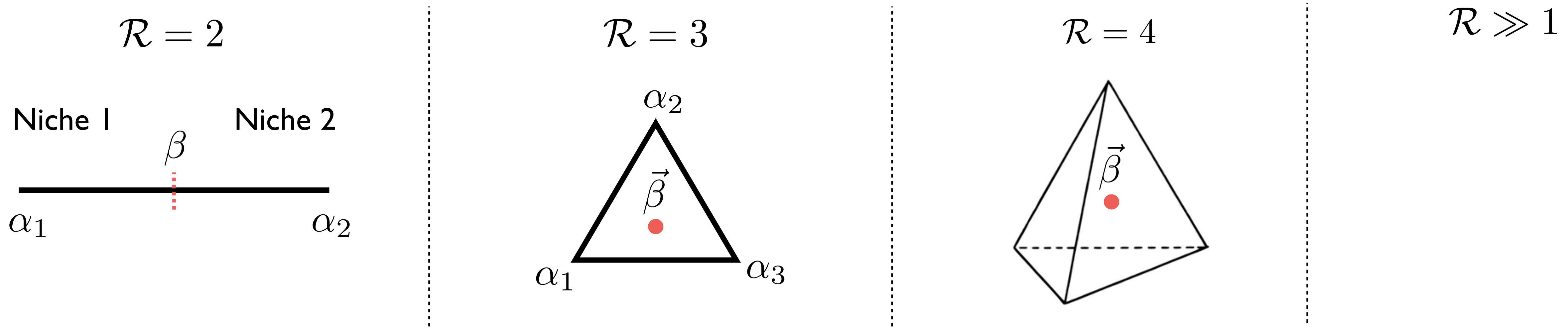


vs



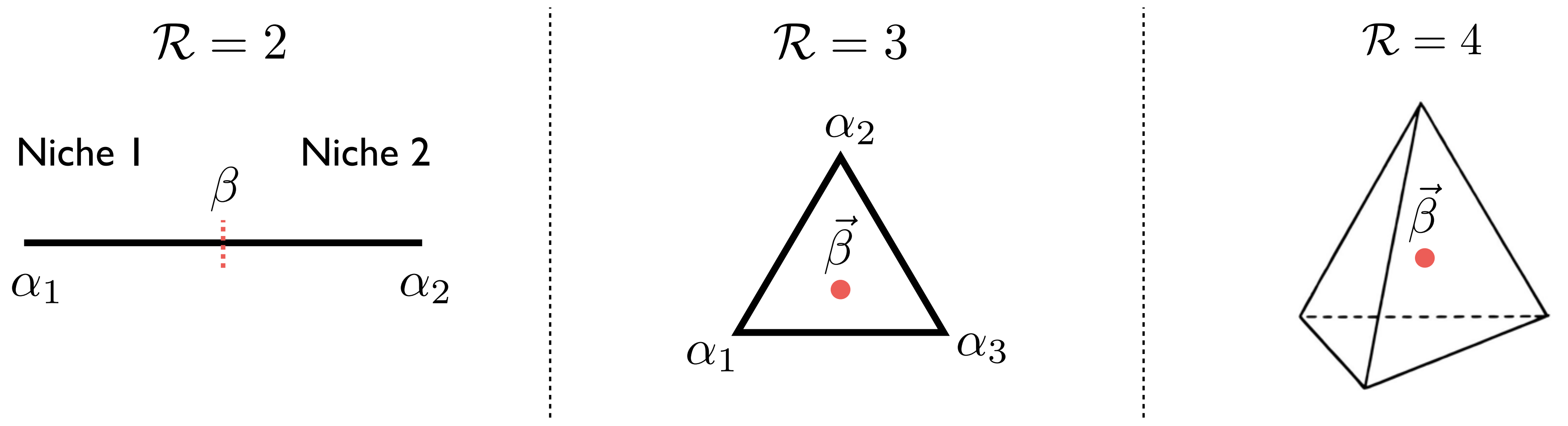
Questions: *How does evolution alter community structure?*
How does the community alter evolutionary dynamics?
Can we make connections to sequence data?

Future direction 1: highly diverse consumer-resource models



increasing # of coexisting strains ⟶

Future direction 1: highly diverse consumer-resource models



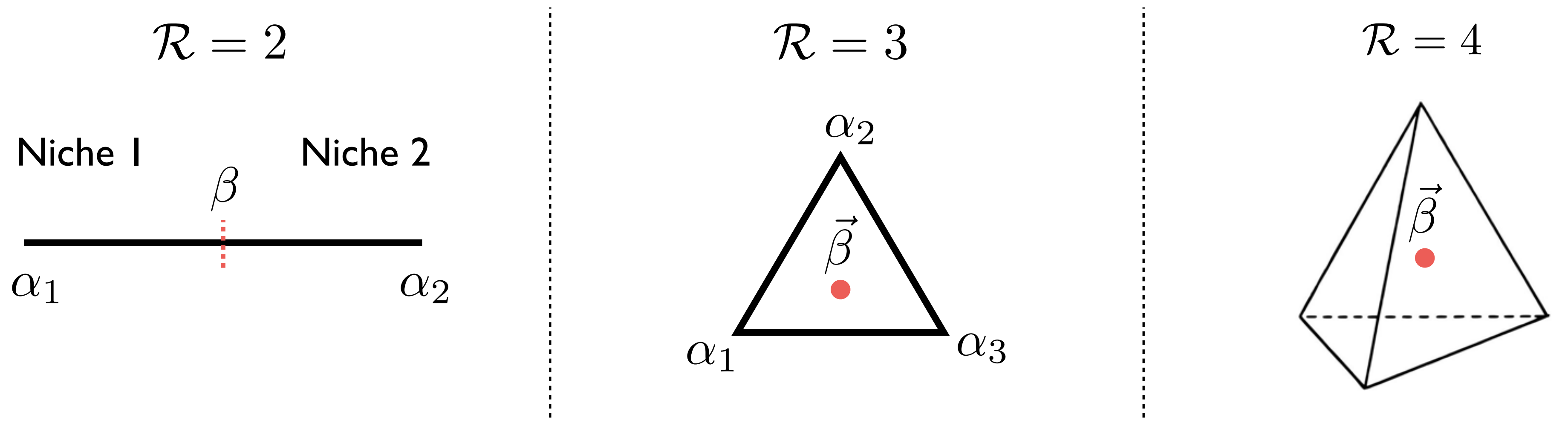
$\mathcal{R} \gg 1$

New behavior?

or just
harder math?

increasing # of coexisting strains ⟶

Future direction 1: highly diverse consumer-resource models



$\mathcal{R} \gg 1$

New behavior?

or just
harder math?

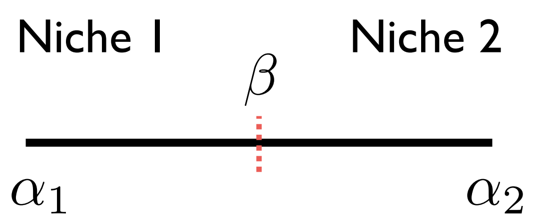
increasing # of coexisting strains ⟶

Preliminary results:

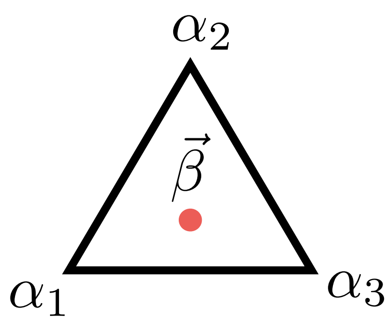
In ***"saturated"*** ecosystems ($\mathcal{S} = \mathcal{R}$)

Future direction 1: highly diverse consumer-resource models

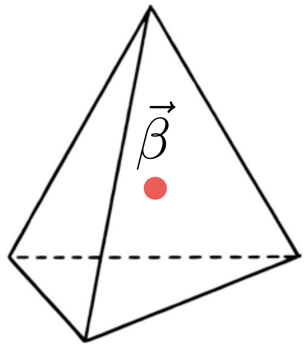
$\mathcal{R} = 2$



$\mathcal{R} = 3$



$\mathcal{R} = 4$



$\mathcal{R} \gg 1$

New behavior?

or just
harder math?

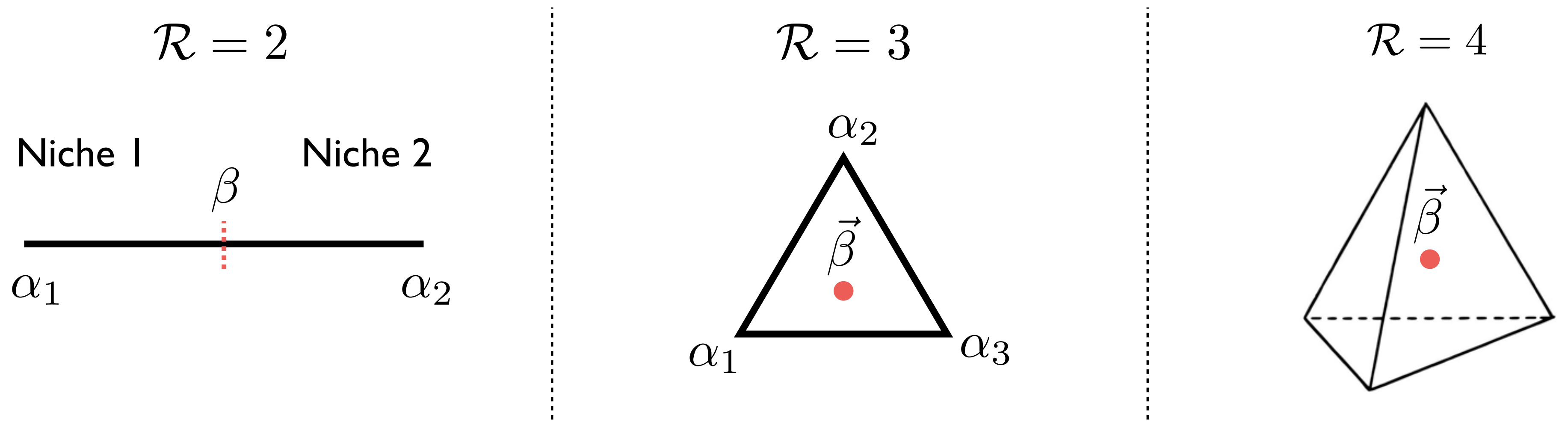
increasing # of coexisting strains →

Preliminary results:

In ***"saturated"*** ecosystems ($\mathcal{S} = \mathcal{R}$)

↳ matrix generalizations
of previous results $(\alpha_{\mu,i})^{-1}$

Future direction 1: highly diverse consumer-resource models



$\mathcal{R} \gg 1$

New behavior?

or just
harder math?

increasing # of coexisting strains ⟶

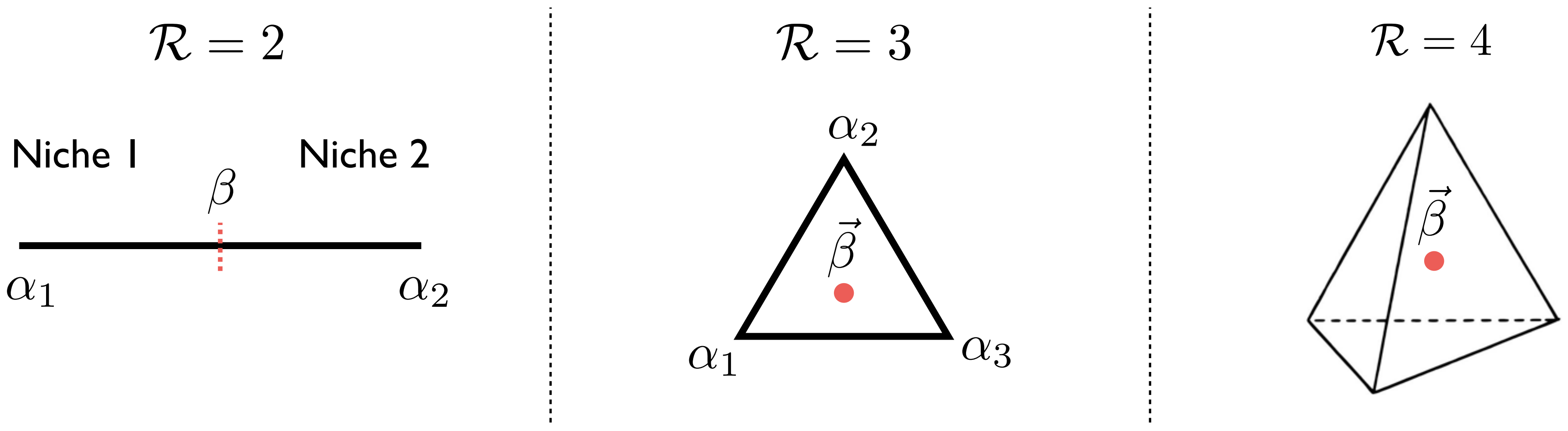
Preliminary results:

In ***"saturated"*** ecosystems ($\mathcal{S} = \mathcal{R}$)

↳ matrix generalizations of previous results $(\alpha_{\mu,i})^{-1}$

E.g. emergent selection for mutations from high- to low-fitness resources

Future direction 1: highly diverse consumer-resource models



$\mathcal{R} \gg 1$

New behavior?

or just
harder math?

increasing # of coexisting strains ⟶

Preliminary results:

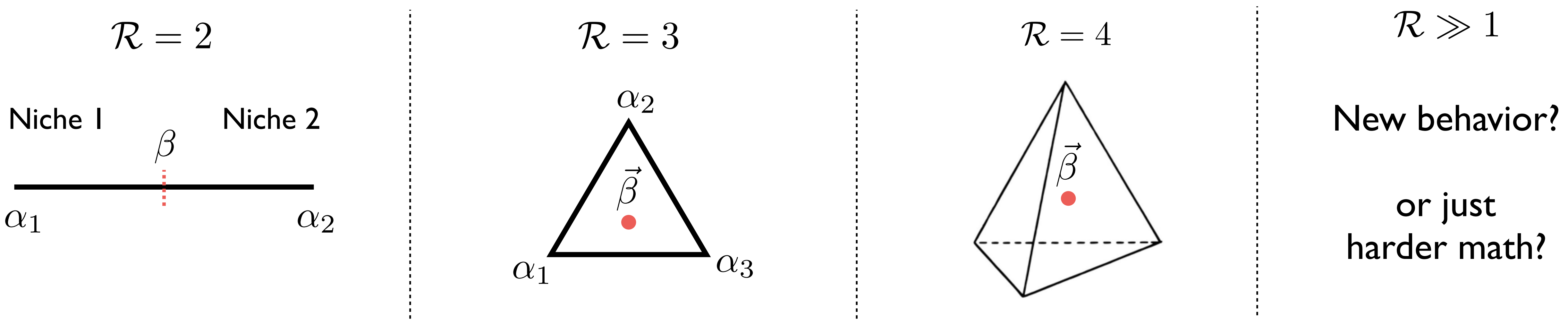
In ***"saturated"*** ecosystems ($\mathcal{S} = \mathcal{R}$)

↳ matrix generalizations of previous results $(\alpha_{\mu,i})^{-1}$

E.g. emergent selection for mutations from high- to low-fitness resources

But coexistence ≠ saturation in general!

Future direction 1: highly diverse consumer-resource models



New behavior?

or just
harder math?

increasing # of coexisting strains ⟶

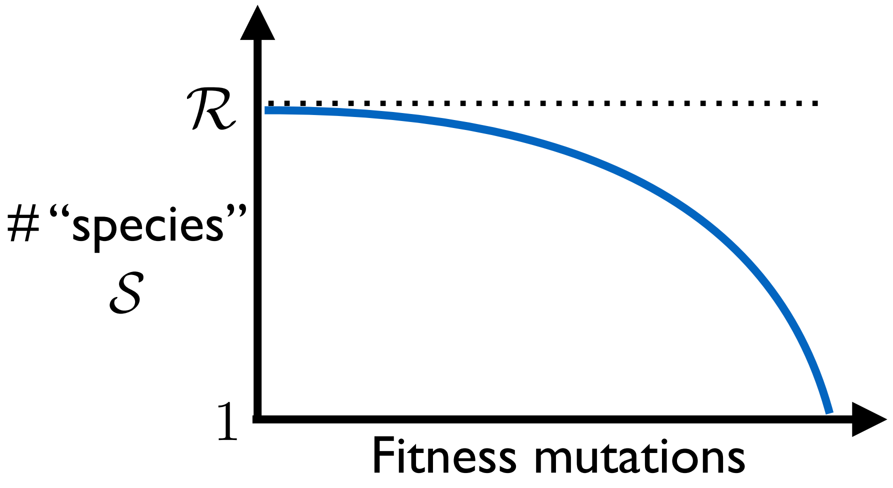
Preliminary results:

In ***"saturated"*** ecosystems ($\mathcal{S} = \mathcal{R}$)
↳ matrix generalizations of previous results $(\alpha_{\mu,i})^{-1}$

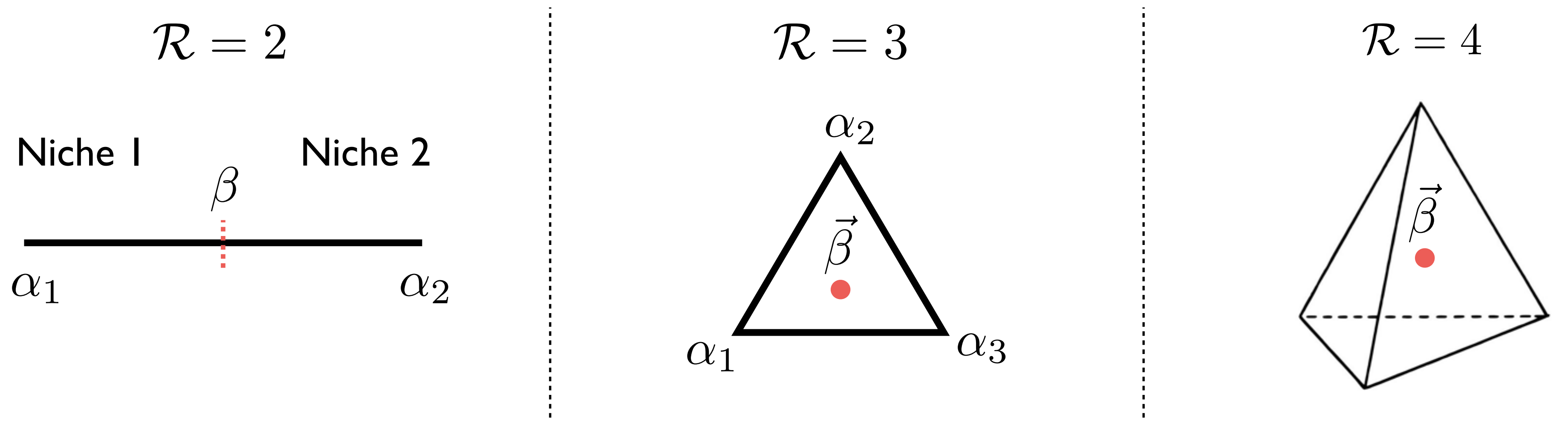
E.g. emergent selection for mutations from high- to low-fitness resources

But coexistence ≠ saturation in general!

Diversification-selection balance



Future direction 1: highly diverse consumer-resource models



$\mathcal{R} \gg 1$

New behavior?

or just
harder math?

increasing # of coexisting strains ⟶

Preliminary results:

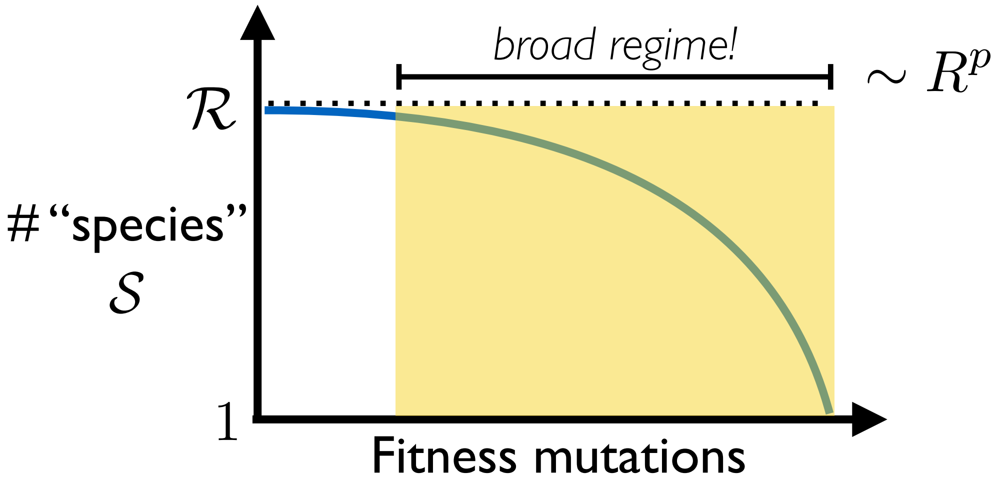
In ***“saturated”*** ecosystems ($\mathcal{S}=\mathcal{R}$)

↳ matrix generalizations of previous results $(\alpha_{\mu,i})^{-1}$

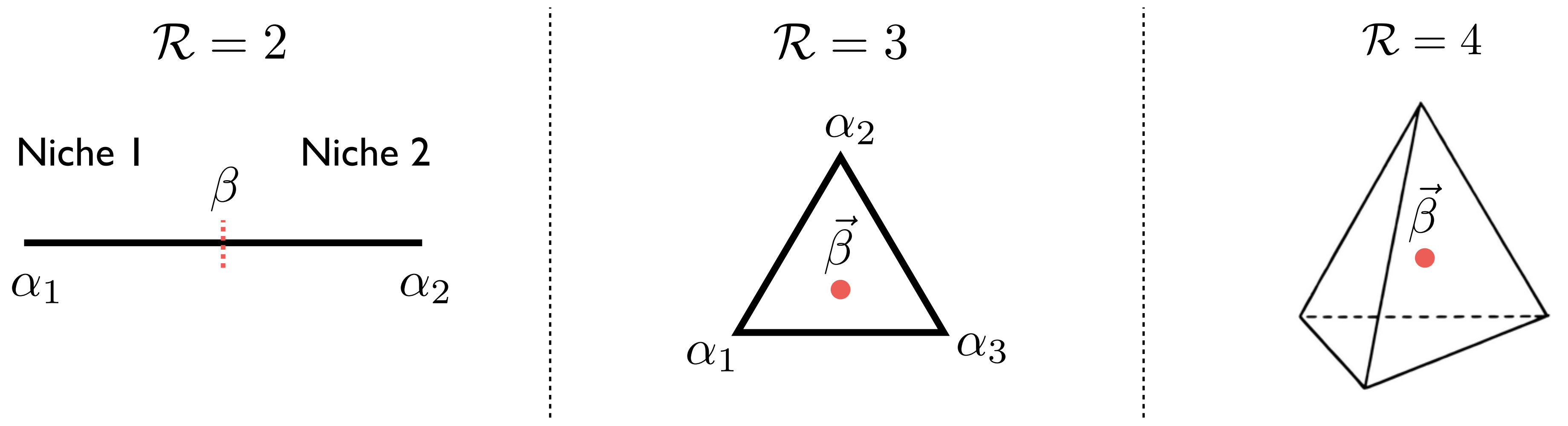
E.g. emergent selection for mutations from high- to low-fitness resources

But coexistence ≠ saturation in general!

Diversification-selection balance



Future direction 1: highly diverse consumer-resource models



$\mathcal{R} \gg 1$

New behavior?

or just
harder math?

increasing # of coexisting strains ⟶

Preliminary results:

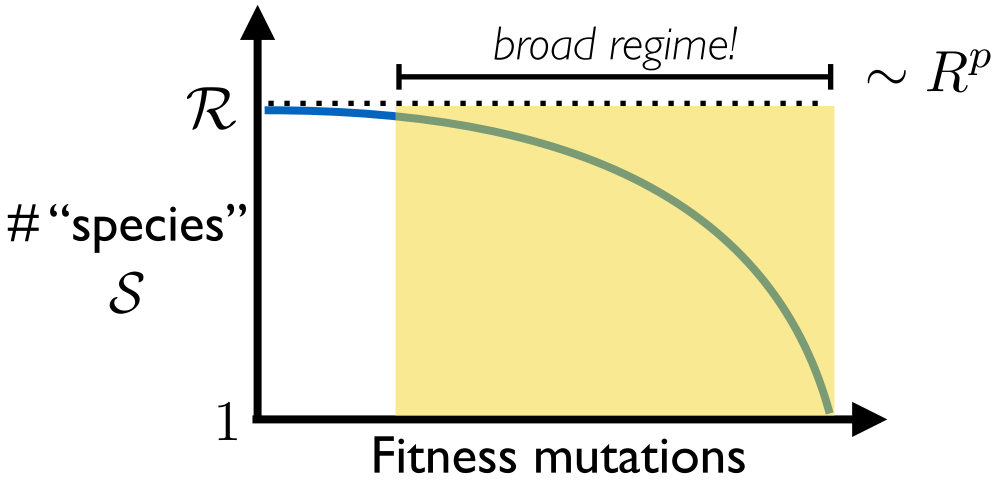
In *“saturated”* ecosystems ($\mathcal{S} = \mathcal{R}$)

↳ matrix generalizations of previous results $(\alpha_{\mu,i})^{-1}$

E.g. emergent selection for mutations from high- to low-fitness resources

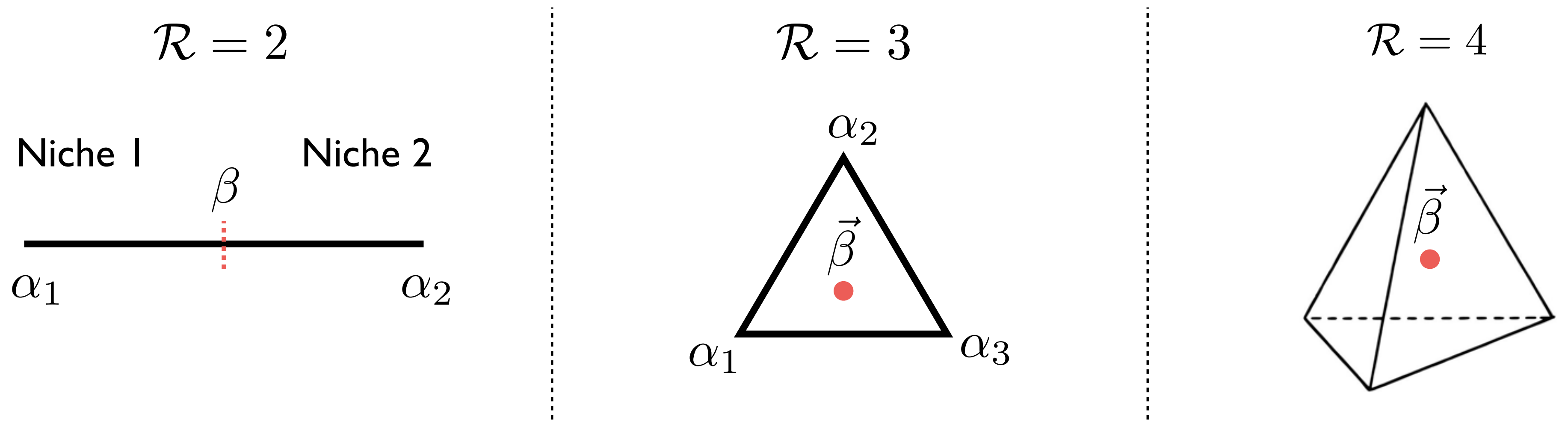
But coexistence ≠ saturation in general!

Diversification-selection balance



Typical structure of steady-state ecosystem?

Future direction 1: highly diverse consumer-resource models



$\mathcal{R} \gg 1$

New behavior?

or just
harder math?

increasing # of coexisting strains ⟶

Preliminary results:

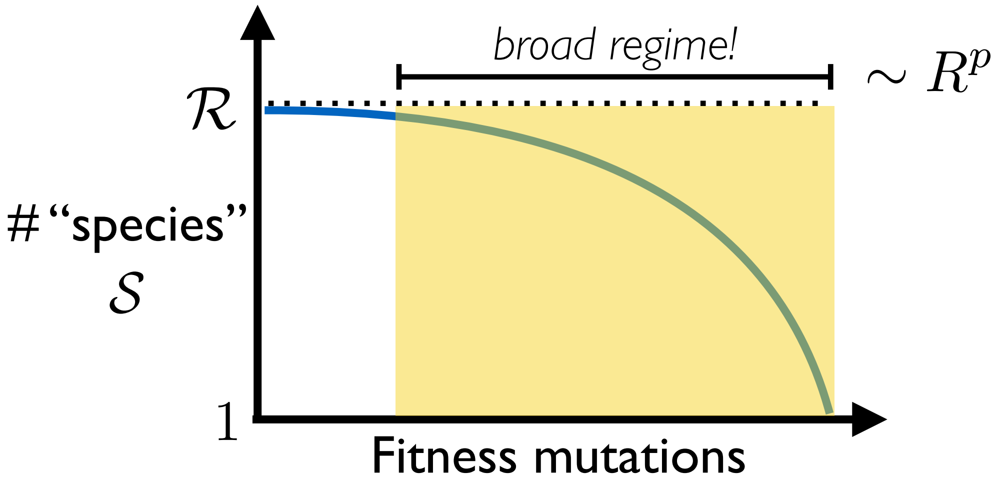
In ***"saturated"*** ecosystems ($\mathcal{S} = \mathcal{R}$)

↳ matrix generalizations
of previous results $(\alpha_{\mu,i})^{-1}$

E.g. emergent selection for mutations from high- to low-fitness resources

But coexistence ≠ saturation in general!

Diversification-selection balance



Typical structure of steady-state ecosystem?

Statistical structure of genealogies?

Lots of ***potential*** models… how do we choose?

↓

empirical data from ***natural microbial communities***

Lots of ***potential*** models… how do we choose?

↓

empirical data from ***natural microbial communities***

(ideally, inspire new laboratory model systems)

Lots of ***potential*** models… how do we choose?

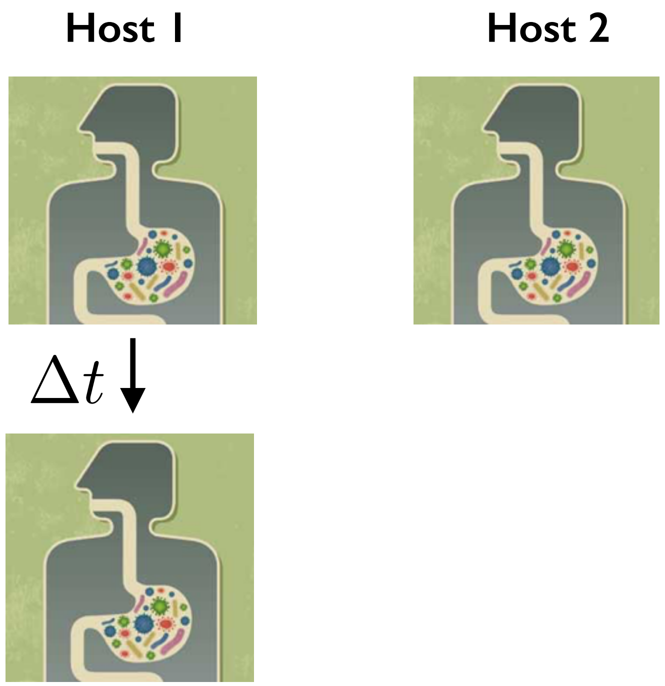
↓

empirical data from ***natural microbial communities***

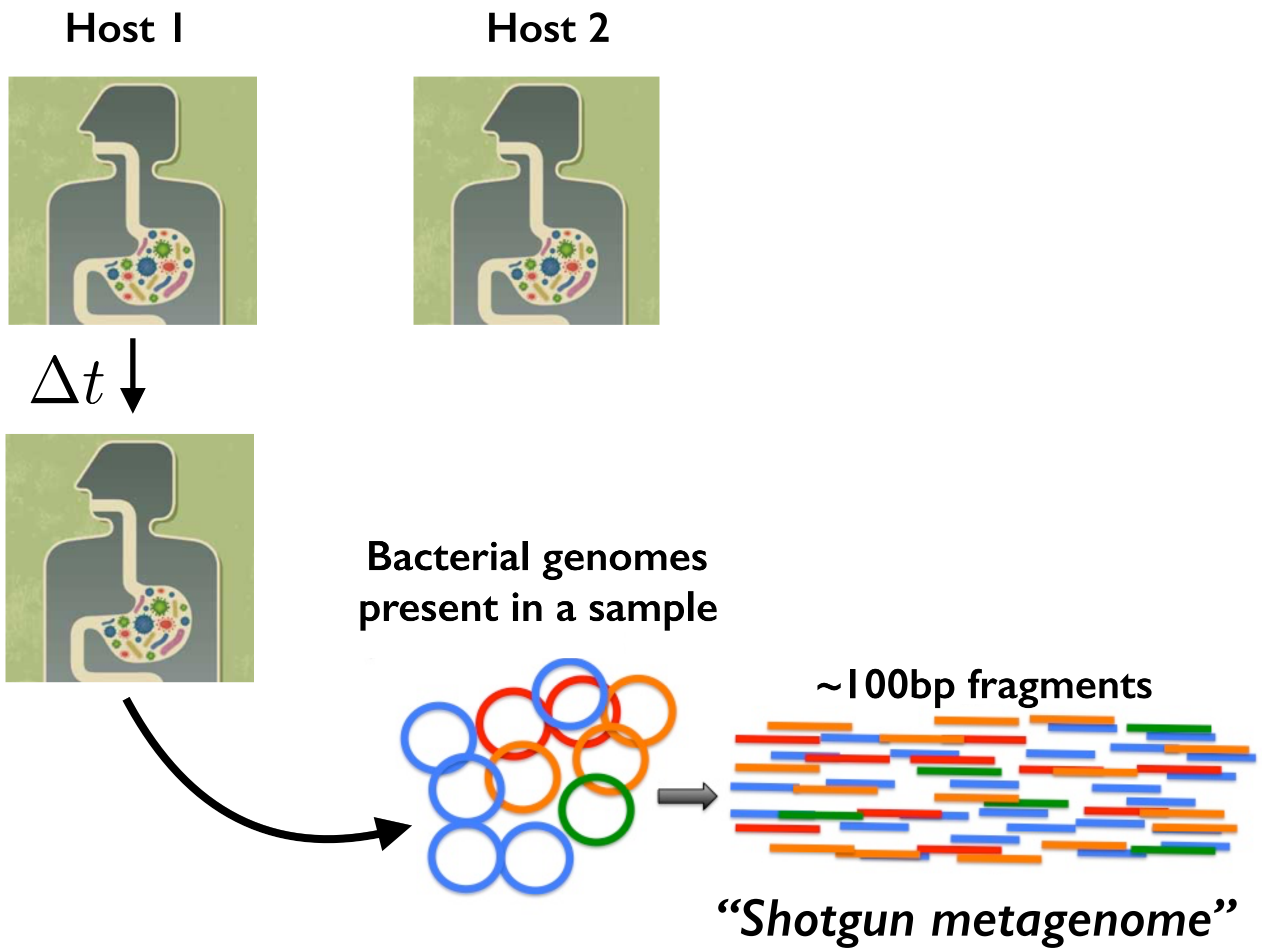
(ideally, inspire new laboratory model systems)

But few observations of evolution so far

Future direction 2: evolution in the gut microbiome



Future direction 2: evolution in the gut microbiome



Host 1

Host 2

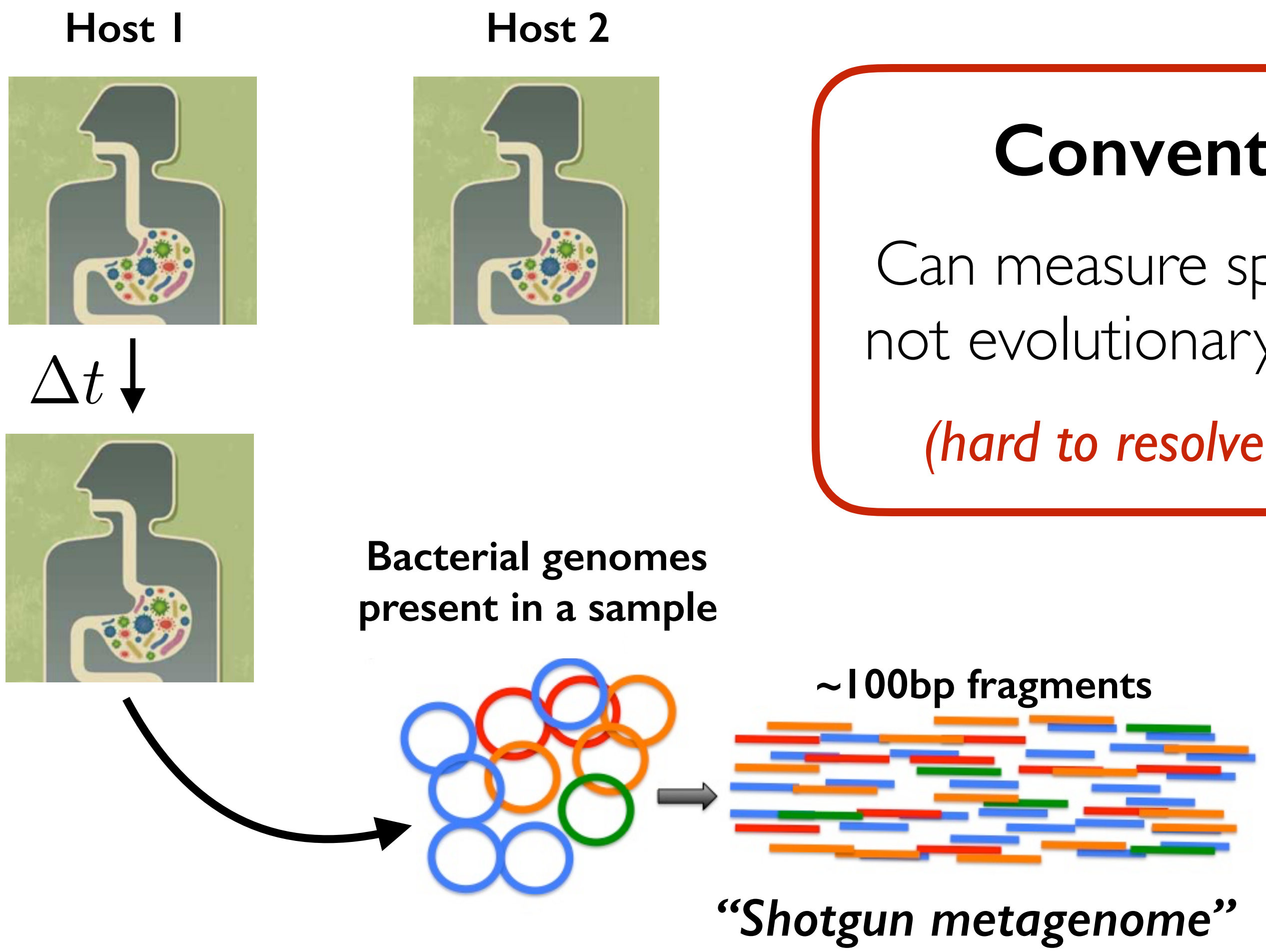
Δt

Bacterial genomes present in a sample

~100bp fragments

"Shotgun metagenome"

Future direction 2: evolution in the gut microbiome



Host 1

Host 2

Δt

Bacterial genomes present in a sample

~100bp fragments

"Shotgun metagenome"

Conventional* wisdom:

Can measure species-level dynamics, but not evolutionary dynamics within species.

(hard to resolve "strains" from short reads)

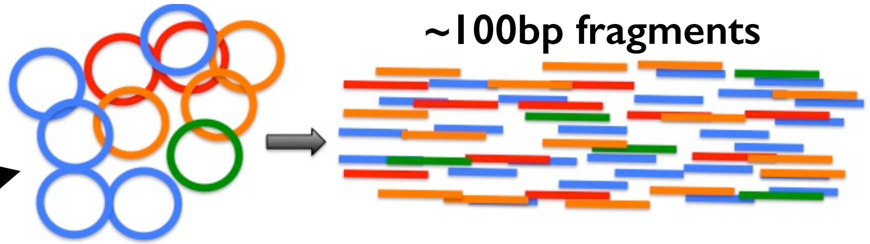
Future direction 2: evolution in the gut microbiome

Host 1

Host 2

Δt ↓

Bacterial genomes present in a sample



~100bp fragments

"Shotgun metagenome"

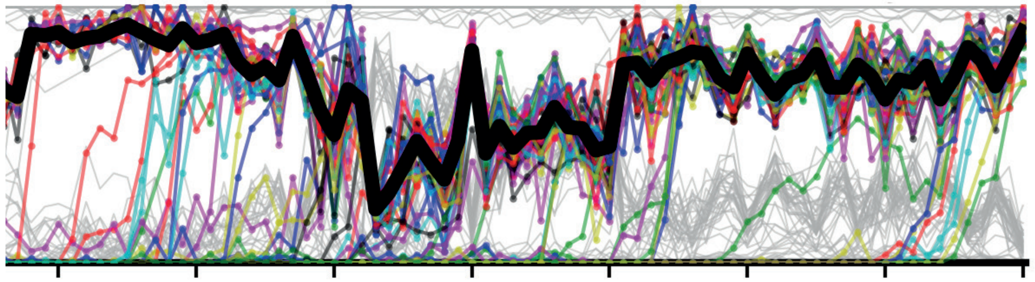
Conventional* wisdom:

Can measure species-level dynamics, but not evolutionary dynamics within species.

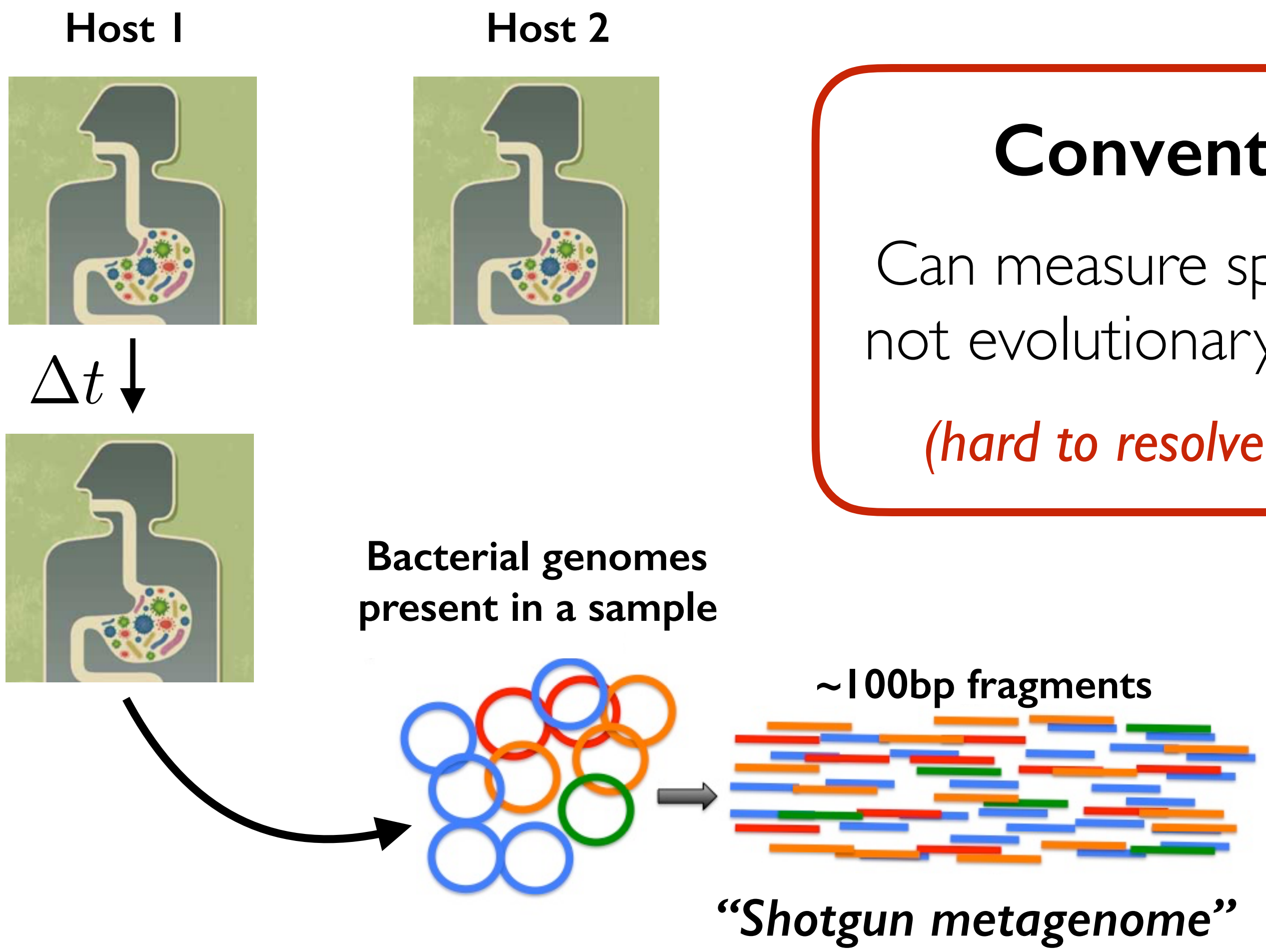
(hard to resolve "strains" from short reads)

?

Evolution in LTEE "community"



Future direction 2: evolution in the gut microbiome



Conventional* wisdom:

Can measure species-level dynamics, but not evolutionary dynamics within species.

(hard to resolve "strains" from short reads)

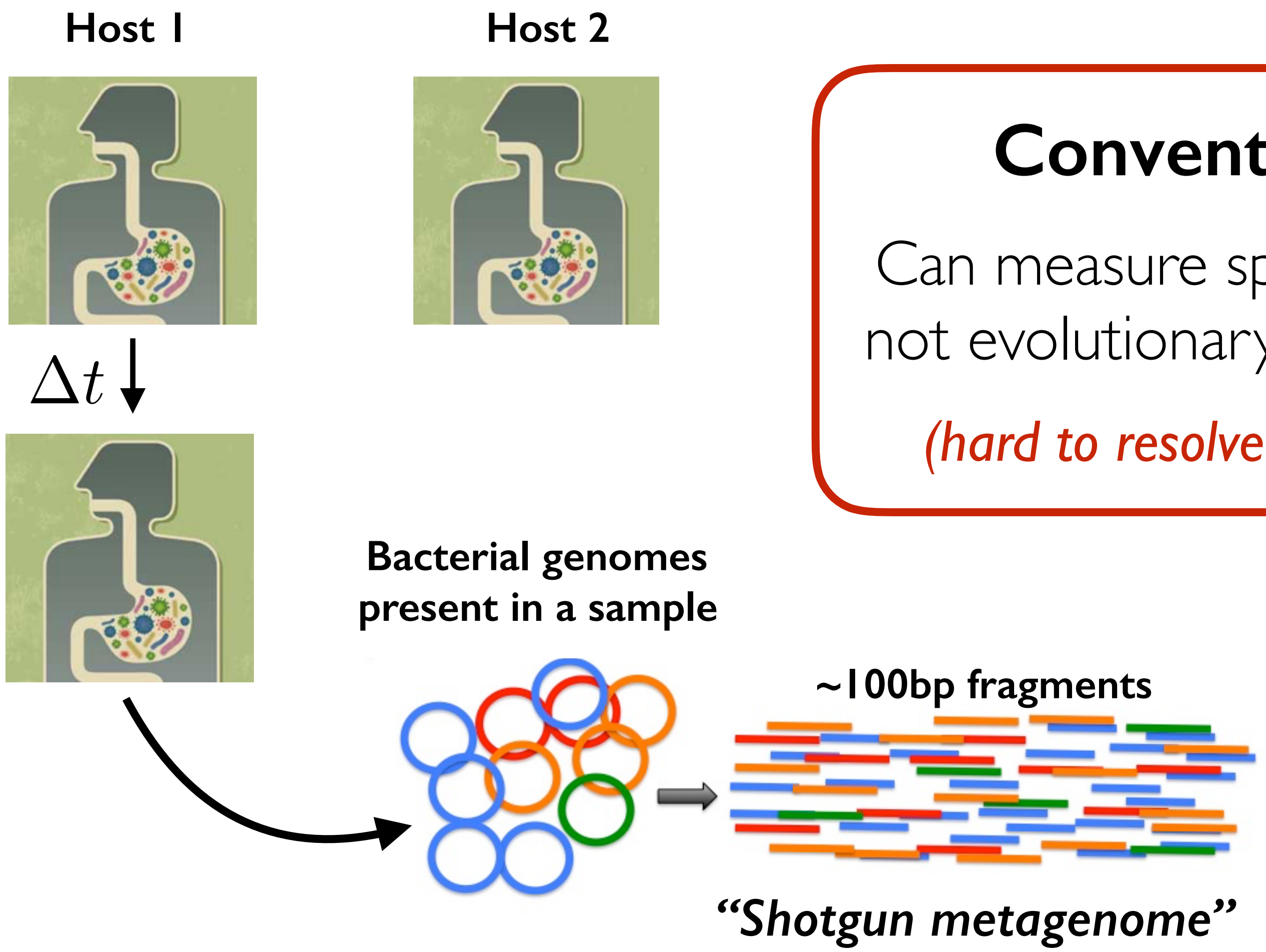
Preliminary work:

With careful, model-based approach, one __can__ extract evolutionary info from complex gut metagenomes

w/ **Nandita Garud**
(UCSF)

Proof of principle, see: Garud, Good*, et al (biorXiv, 2017)*

Future direction 2: evolution in the gut microbiome



Conventional* wisdom:

Can measure species-level dynamics, but not evolutionary dynamics within species.

(hard to resolve "strains" from short reads)

Preliminary work:

*With careful, model-based approach, one **can** extract evolutionary info from complex gut metagenomes*

Methods for measuring evolutionary changes within and across hosts in ~30 gut species

w/ **Nandita Garud**
(UCSF)

Proof of principle, see:
Garud, Good*, et al*
(biorXiv, 2017)

Preliminary findings and future questions

1. **Across-host genealogical signatures challenge existing pop gen models**

   What models can produce these patterns?

   Global signatures of adaptation? or stasis?

Preliminary findings and future questions

---

1. Across-host genealogical signatures challenge existing pop gen models

What models can produce these patterns?

Global signatures of adaptation? or stasis?

---

2. Gut bacteria can evolve within hosts on human-relevant timescales

Is linkage as important as in experimental evolution?

Correlations w/ changes in species abundance? w/ evolution in co-colonizing species?

Preliminary findings and future questions

---

1. Across-host genealogical signatures challenge existing pop gen models

What models can produce these patterns?

Global signatures of adaptation? or stasis?

---

2. Gut bacteria can evolve within hosts on human-relevant timescales

Is linkage as important as in experimental evolution?

Correlations w/ changes in species abundance? w/ evolution in co-colonizing species?

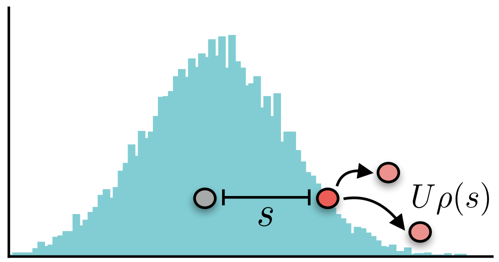
---

3. Bacterial recombination plays greater role

Will require new theoretical models of selection + linkage + recombination / HGT

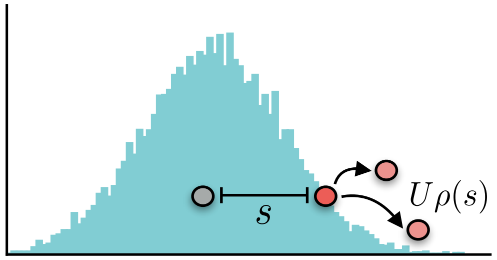
Summary

In rapidly evolving populations, fates of mutations are strongly influenced by ***dynamical processes***, in addition to their ***inherent biological effects***.

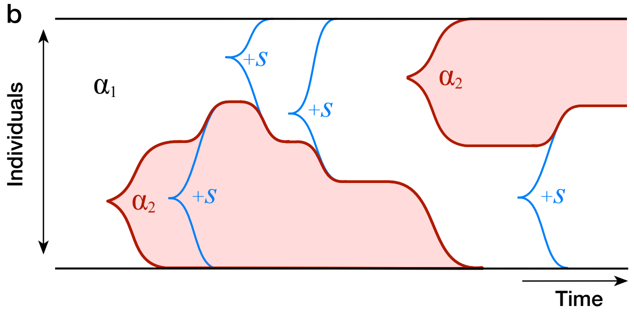


Summary

In rapidly evolving populations, fates of mutations are strongly influenced by ***dynamical processes***, in addition to their ***inherent biological effects***.

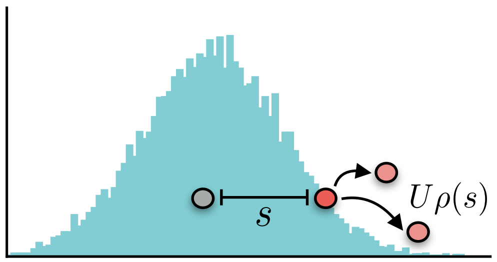


In simple ecological settings, ***rapid evolution*** can also play a important role in determining the ***structure of the community***, and vice versa.

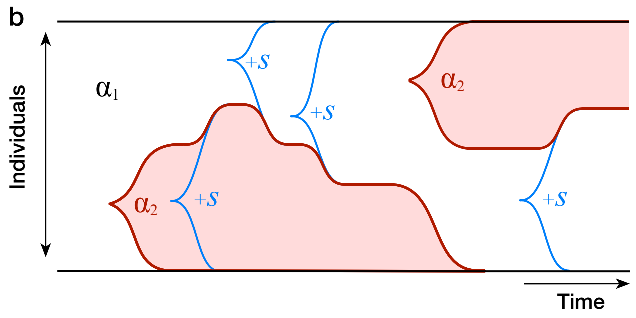


Summary

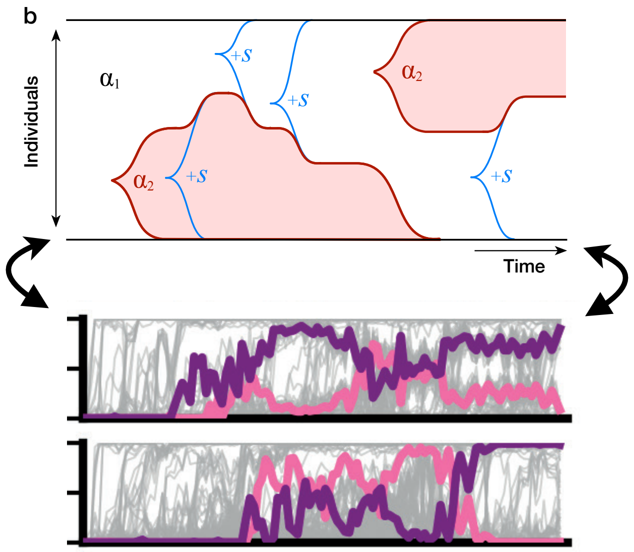
In rapidly evolving populations, fates of mutations are strongly influenced by ***dynamical processes***, in addition to their ***inherent biological effects***.



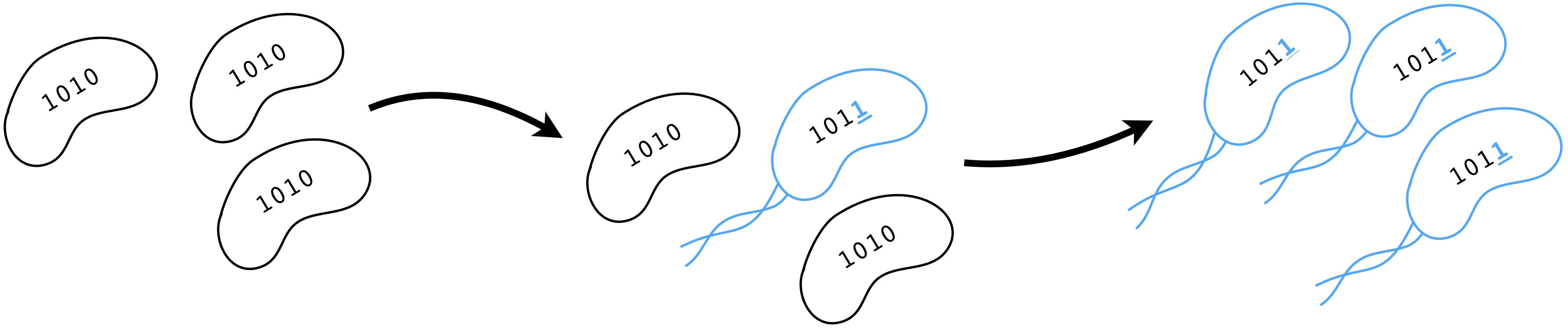
In simple ecological settings, ***rapid evolution*** can also play a important role in determining the ***structure of the community***, and vice versa.



Integrated approach of ***empirically-guided theory*** and ***theoretically-motivated data analysis*** and ***experimental design*** is a promising path toward understanding these processes.



Thanks!



LTEE Sequencing

Mike McDonald (Monash)
Jeff Barrick (UT)
Rich Lenski (MSU).
Michael Desai (Harvard)

Good*, McDonald* et al
(*Nature*, 2017)

Eco-evolutionary theory

Stephen Martis,
Oskar Hallatschek
(UC Berkeley)

Good et al, (in prep)

Evolution in the microbiome

Nandita Garud, Katie Pollard (UCSF)
M. Roodgar, M. Snyder (Stanford)

Garud*, Good*, et al
Evolutionary dynamics of bacteria in the gut microbiome within and across hosts
(*biorXiv* 210955)

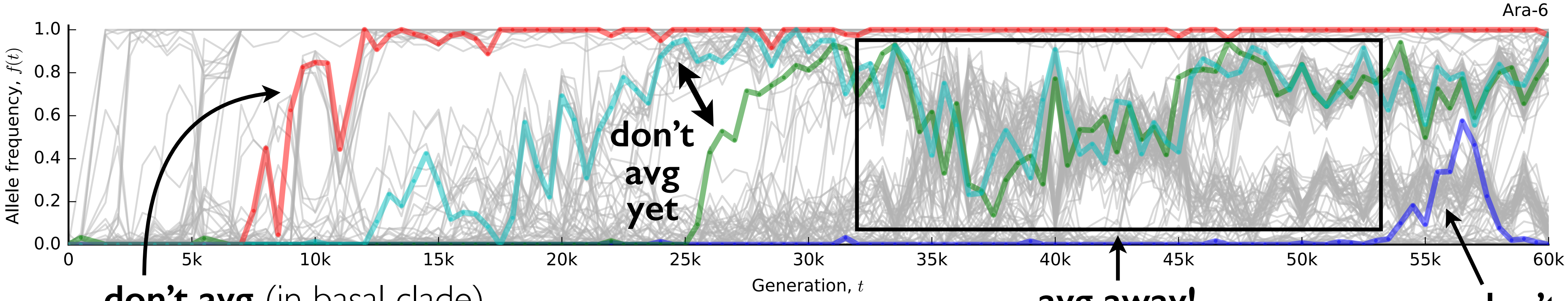
Other collaborators: Ivana Cvijovic (Harvard), Matt Melissa (Harvard), Genya Frenkel (Whitehead), Dan Rice (Chicago), Elizabeth Jerison (Stanford), Daniel Balick (HMS), Richard Neher (Biozentrum), Aleksandra Walczak (ENS), Igor Rouzine (Gladstone)

Miller Institute for Basic Research in Science, University of California, Berkeley

Inferring clade dynamics with hidden Markov models

Basic idea: individual mutations noisy ⟶ avg together to measure clade freq

Problem: which mutations to average, which times to average?



don't avg yet

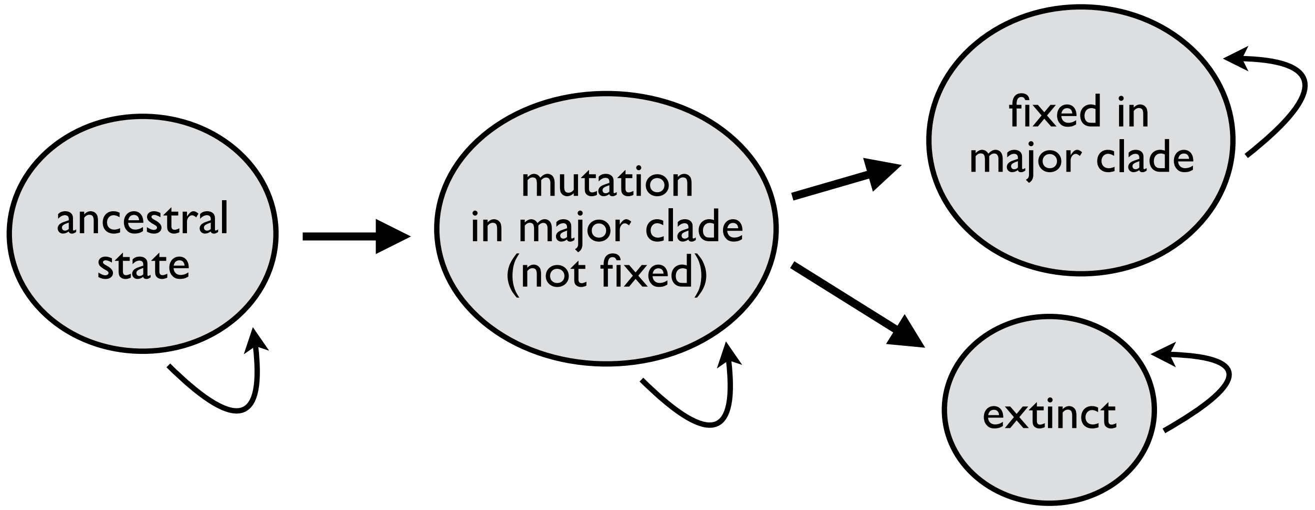
don't avg (in basal clade)

avg away!

don't avg (goes extinct)

Solution: infer with custom hidden Markov model (***cladeHMM***)
⟶ like clustering, but phylogenetically-aware

E.g.



ancestral state

mutation in major clade (not fixed)

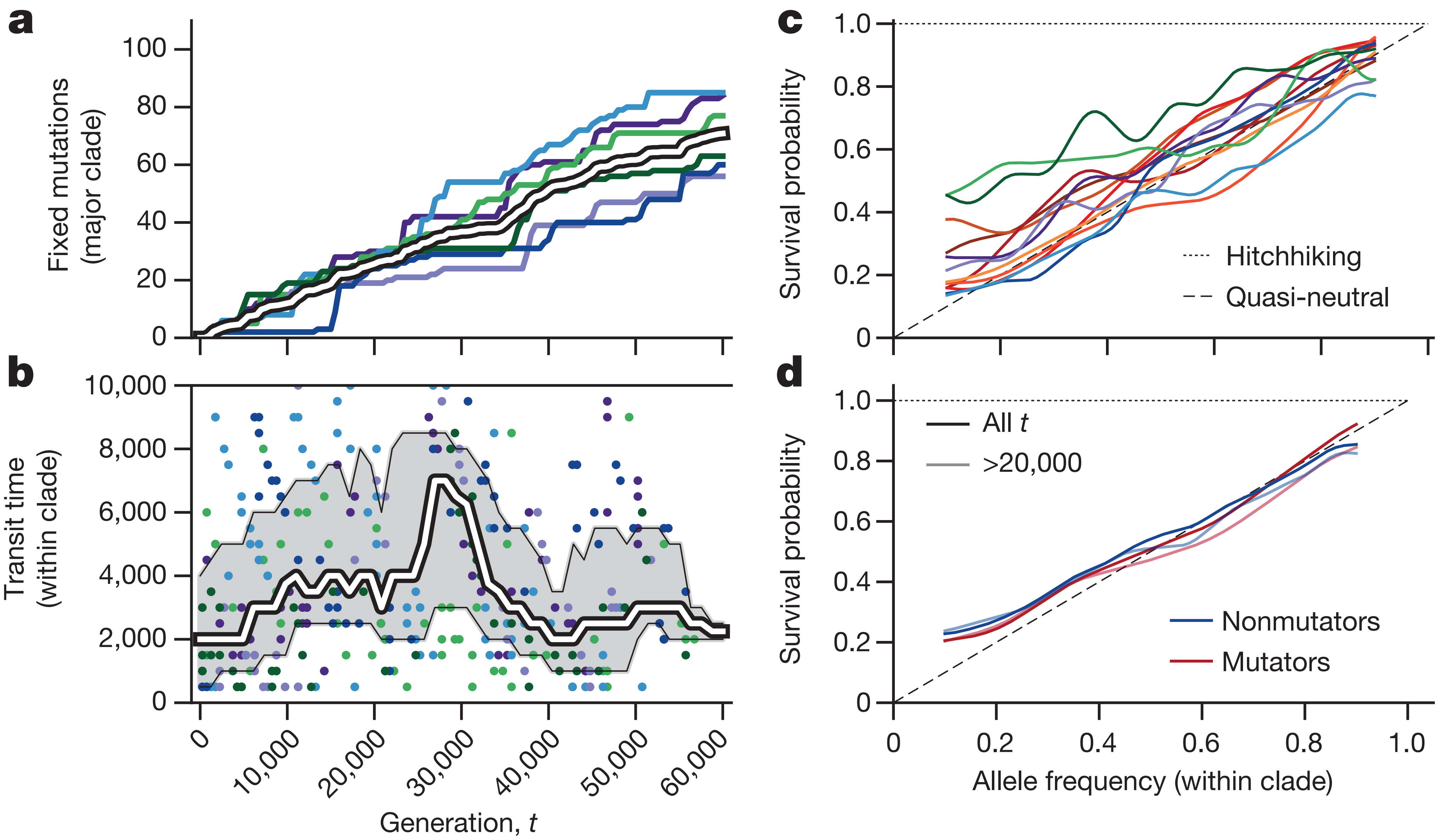
fixed in major clade

extinct

\+ etc.

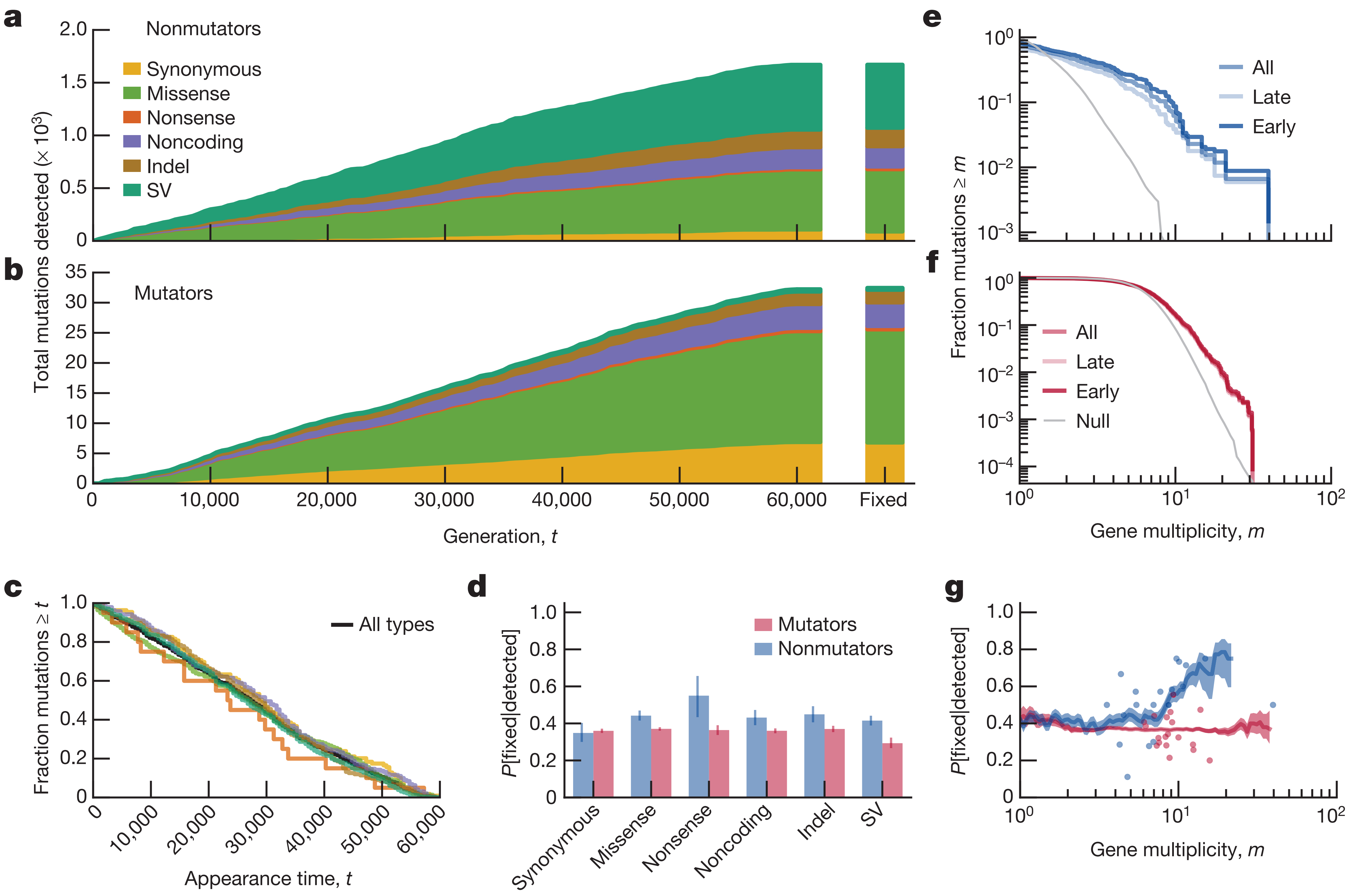
Good, McDonald* et al (Nature, 2017)*

Rapid adaption and clonal interference within clades



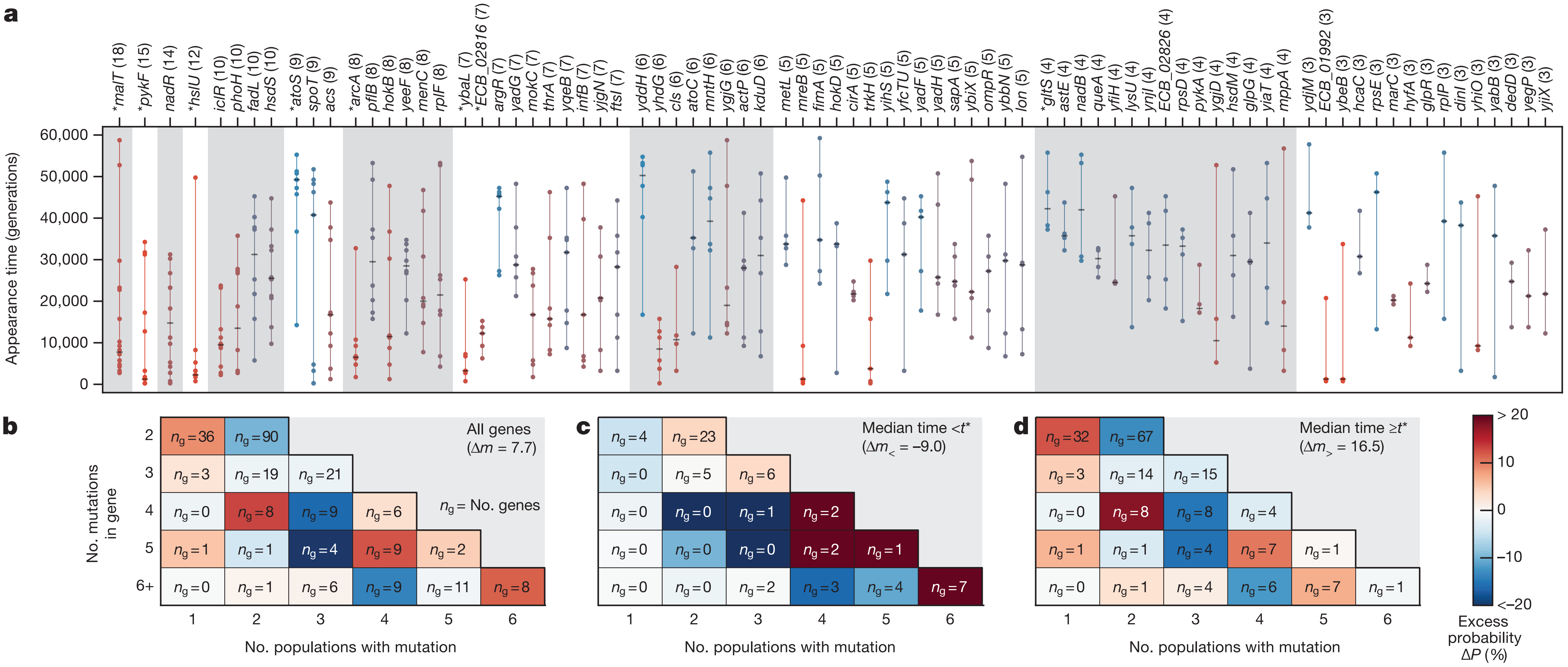
Good, McDonald* et al (Nature, 2017)*

Parallelism at the genetic level



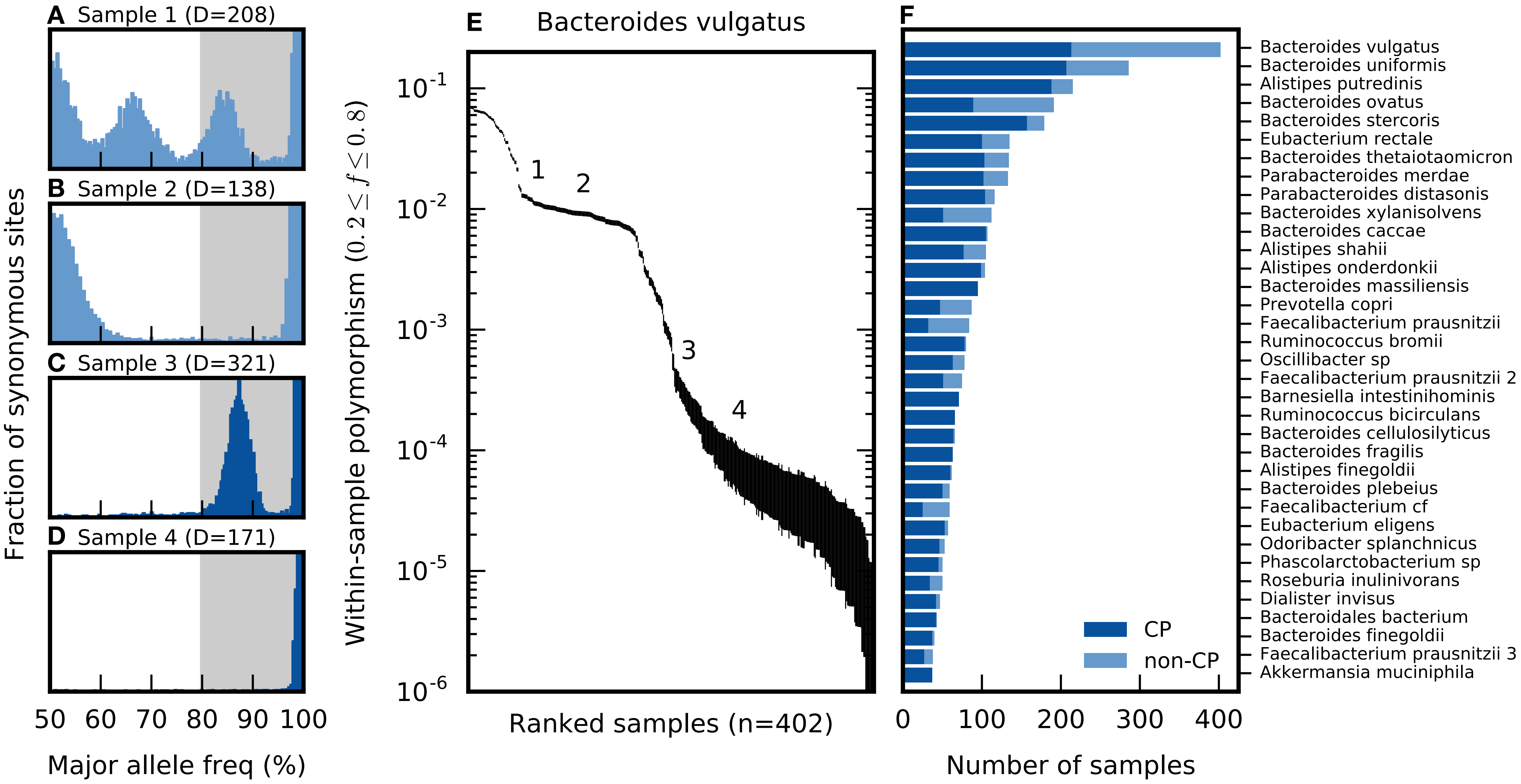
Good, McDonald* et al (Nature, 2017)*

Patterns of epistasis and contingency at the gene level



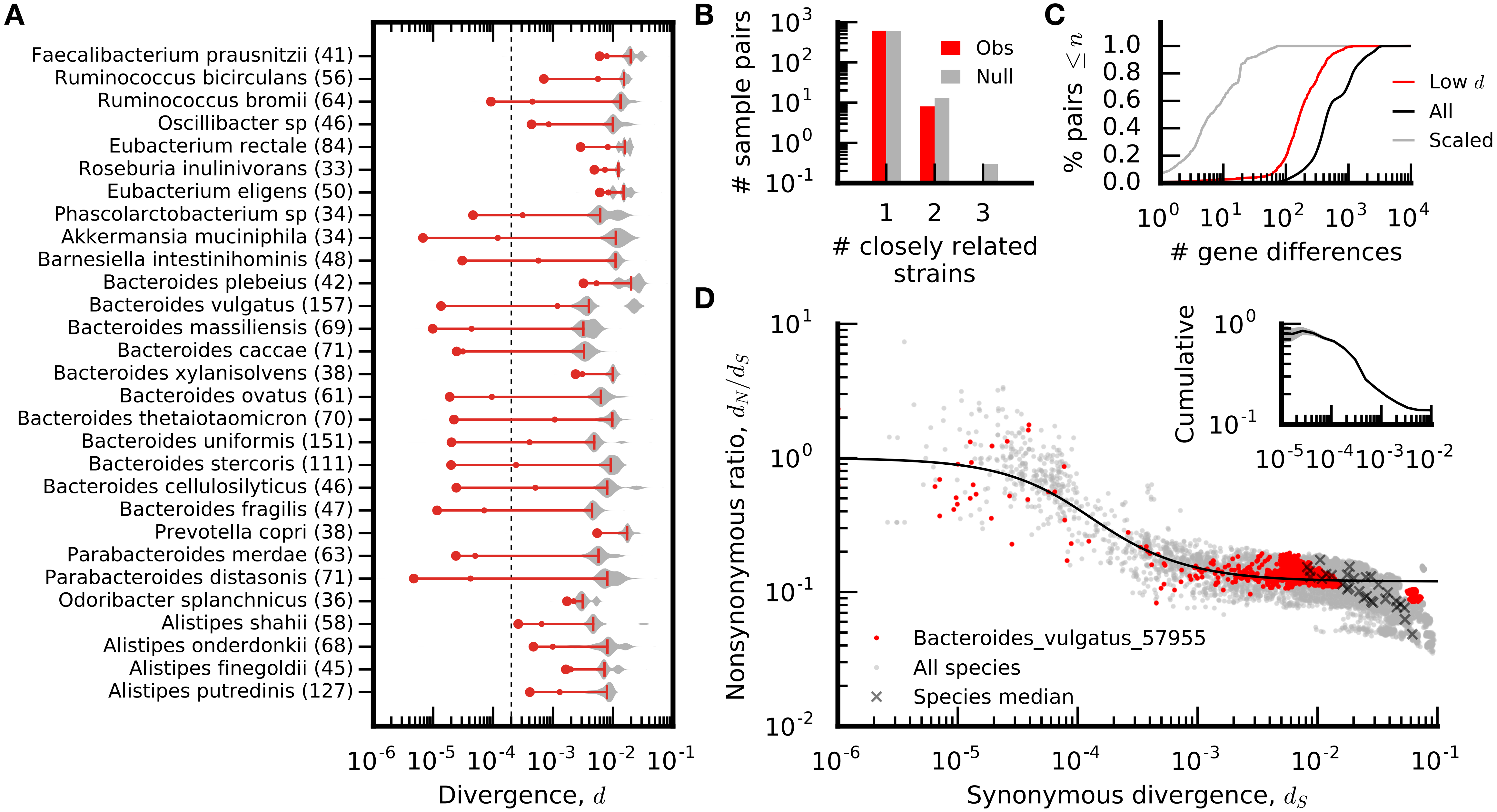
Good, McDonald* et al (Nature, 2017)*

Patterns of genetic diversity within hosts



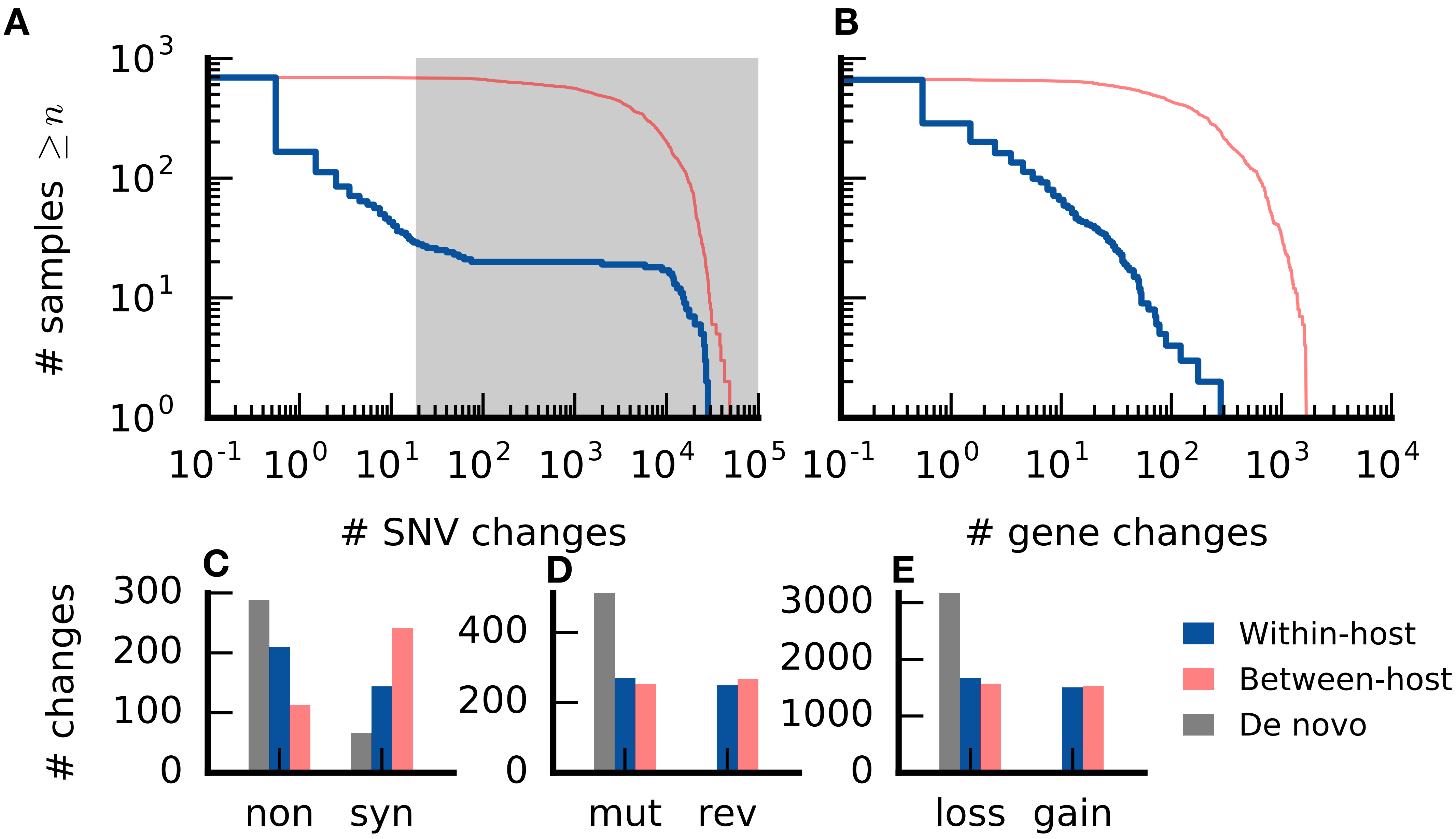
Garud, Good*, et al (biorxiv 2017)*

Patterns of genetic divergence across hosts



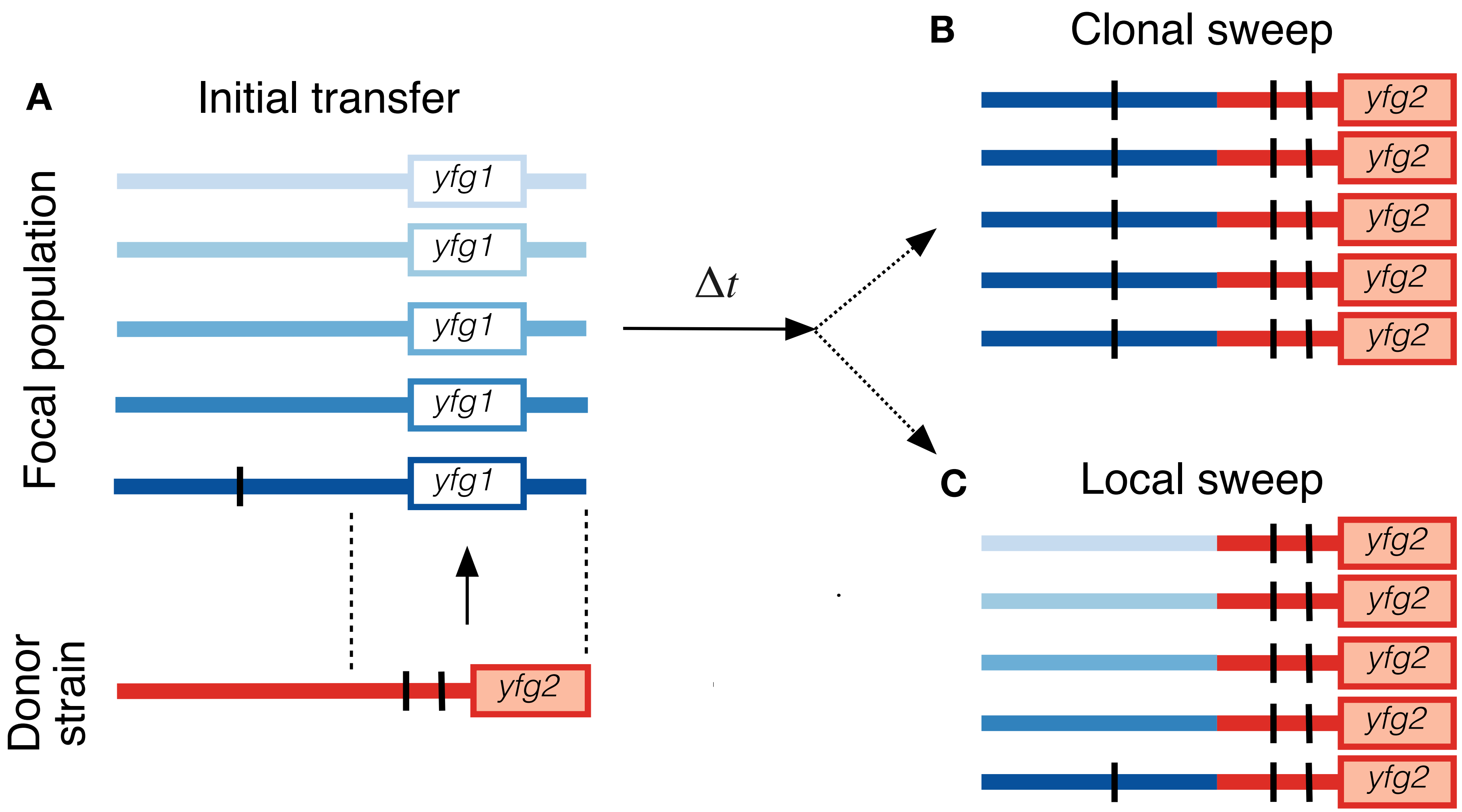
Garud, Good*, et al (biorxiv 2017)*

Short-term succession within hosts



Garud, Good*, et al (biorxiv 2017)*

Putative model of within-host evolution



Garud, Good*, et al (biorxiv 2017)*