BROWN

Simpson's Paradox in Mutation Rate Evolution

C. Scott Wylie, Brown University

Collaborators: Eugene Raynes, Paul Sniegowski, Dan Weinreich

Alternate title:

Why large populations favor mutators and small populations inhibit them

About me…

- PhD in physics (front propagation / population genetics)
- postdoc at Brown (Providence RI, near Boston), trying to be a "real"/experimental biologist
- Want to know "how evolution works", particularly when evolution ≠optimization
- Also work on
  - mutational robustness
  - tradeoffs in enzyme evolution (Part 2, if we get there)
  - protein stability: biophysical basis of fitness landscapes (epistasis)

Outline (Part 1)

1. Intro to mutators / mutation rate evolution

1. Simulations and mathematical analysis of mutation rate evolution
    1. Population size matters, qualitatively: “sign inversion” happens
    2. Sign inversion leads to “Simpson’s paradox”

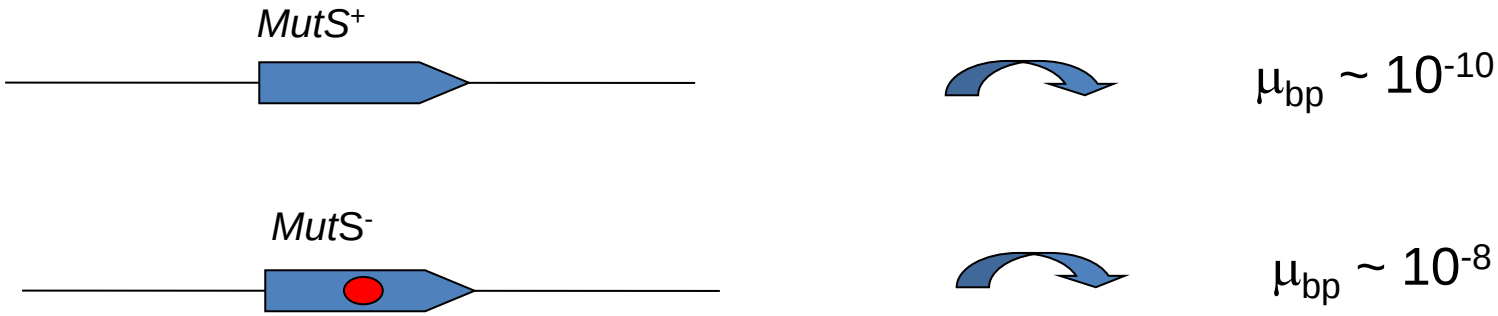
2. Sign inversion and Simpson’s paradox occur in “indirect selection” in general? Mutators are a case study of more general phenomenon?

Jargon that we’ll get to:

- indirect selection
- sign inversion
- Simpson’s paradox

Mutators are variants with high mutation rates

- Proofreading and repair enzymes reduce mutation rate.
- These enzymes, too, can be broken by mutations. Ex:



Mutators are variants with high mutation rates

- Proofreading and repair enzymes reduce mutation rate.
- These enzymes, too, can be broken by mutations. Ex:

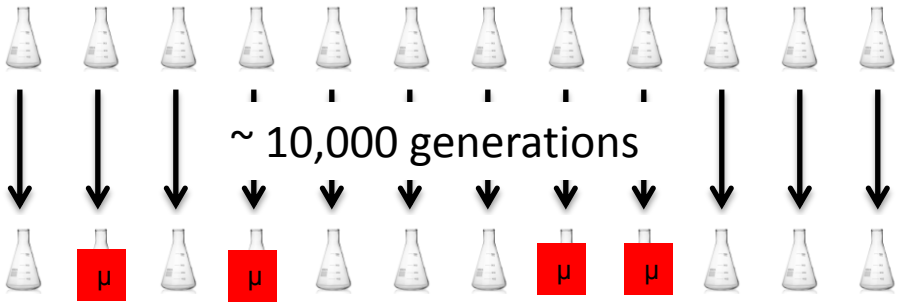
"Defective" genes involved in replication/repair fidelity are called <u>mutator alleles</u>.

Mutators are found in natural and laboratory settings

Mutators play a major role in:

- laboratory microbial populations
- cancer
- antibiotic resistance
- etc. etc.

e.g. Lenski lines: serial E. coli transfer



A puzzle: mutators often take over laboratory microbial populations (e.g. ~1/3 of Lenski's lines). So why aren't wild populations "mutators"?

Bare-bones model of mutator evolution

Simplest simulations that include essential ingredients have 6 parameters:

- M = factor by which mutator's mutation rate is elevated (e.g. 100x)
- U_d = deleterious mutation rate (per genome per genome duplication)
- U_b = beneficial mutation rate (per genome per genome duplication)
- s_d = deleterious selection coefficient (~ % growth rate penalty of new mutation)
- s_b = beneficial selection coefficient (% growth rate advantage of new mutation)
- N = population size (census or effective??)

Typical parameter values for laboratory microbial populations

- $M \sim 100$ for MMR knockouts. Mild mutators certainly possible as well.
- $U_b \sim 10^{-6}$ to 10^{-5} per generation (Perfeito '07, Desai lab, Levy et. al 2015, etc)
- $U_d \sim 10^{-4}$ per generation (e.g. Kibota and Lynch '96)
- $s_b \sim 0.01$ to 0.1 (e.g. Lenski
- $s_d \sim 0.01$ (e.g. Kibota and Lynch '96)

note: shockingly high U_b (U_d/U_b only 10-100). "Real" or idiosyncratic to lab environment?

Dynamics = haploid, asexual Moran model

cartoon illustration:

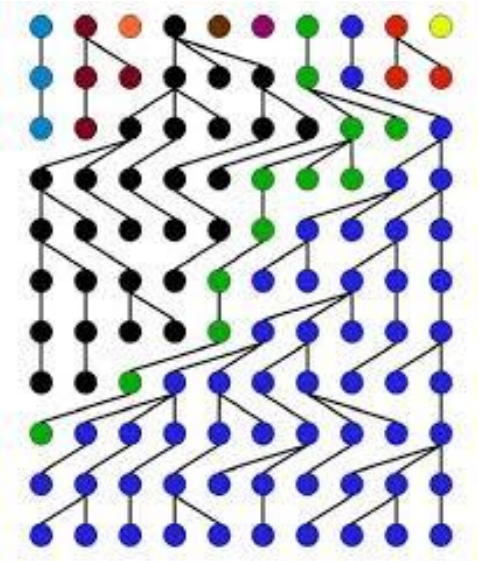


figure credit:
http://culturemath.ens.fr/

children/generation: Poisson w/ mean ~ fitness

Moran's model is standard model of population genetics, incorporating

- drift
- selection
- mutation
- other bells/whistles possible

Other alternatives, e.g. Wright-Fisher, give essentially same results.

forces FOR and AGAINST mutators

- high rate of beneficial muts.

- excessive deleterious mutations
- beneficial mutations from non-mutators

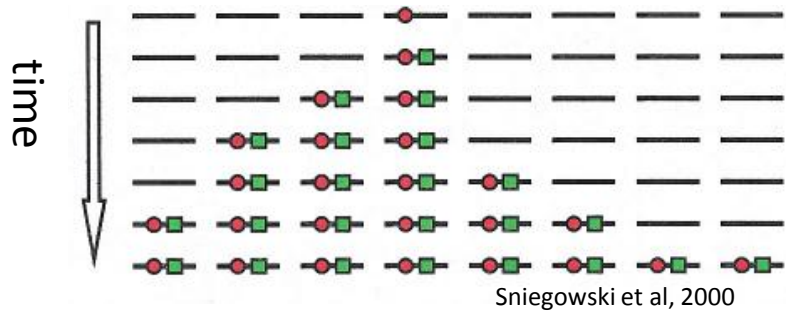
What does fitness mean for mutators?

Simplistic notion: evolution maximizes fitness. And fitness = number offspring left per generation. So, we can ask: "are mutators fitter than neutral expectation"?

But the simplistic approach is unproductive. If we assume that mutators have no direct selective effect, then

initially: mutators are neutral
later, after mutations occur:

- mutators usually linked to deleterious mutations and thus disfavored.
- occasionally hitchhike with beneficial mutations and thus favored.



Sniegowski et al, 2000

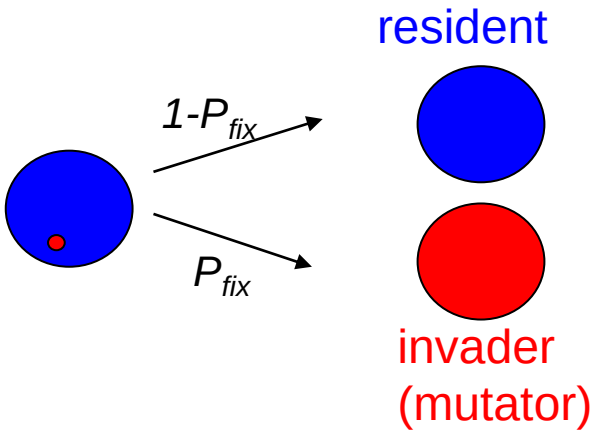
red: mutator allele
green: beneficial mutation (driver)

mutators (can be) indirectly selected via linkage with directly selected sites.

Message: "fitness" is ambiguous/stochastic because selection depends on random events (mutations) happening at random times.

Fixation probability measures long term fitness

A better, more lineage-centric approach is via fixation probability:



Ultimately, two possible fates of any lineage:

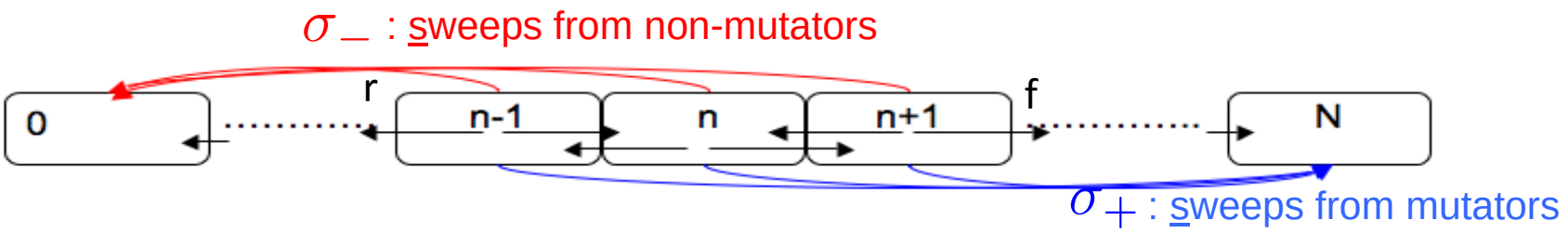
- extinction
- fixation (achieve 100% frequency)

$$\text{long term fitness} = N \times P_{\text{fix}} + 0 \times (1 - P_{\text{fix}}) = N \times P_{\text{fix}}$$

"ordinary fitness": number of descendants left by a single individual during 1st generation
"long-term fitness": number of descendants left eventually.

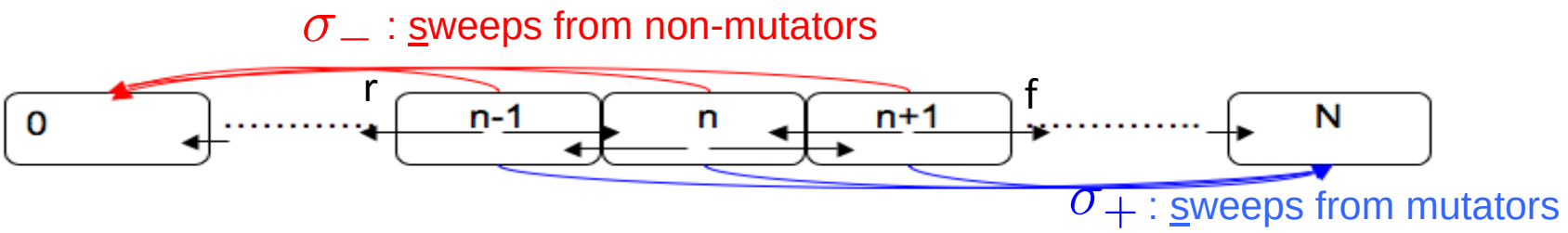
- fitness = 1 means selectively neutral (lineage neither grows nor shrinks)
- Ordinarily, the sign of (fitness – 1) is the same in both short and long term: short-term predicts long-term, at least qualitatively.
- When analyzing mutators, we must think long-term. Goal: calculate P_{fix} for mutators.

How to calculate eventual P_{fix} for 1 mutator



n = # mutators
f = probability (n → n+1)
r = probability (n → n-1)
P_n=fixation prob. of n mutators

How to calculate eventual P_{fix} for 1 mutator

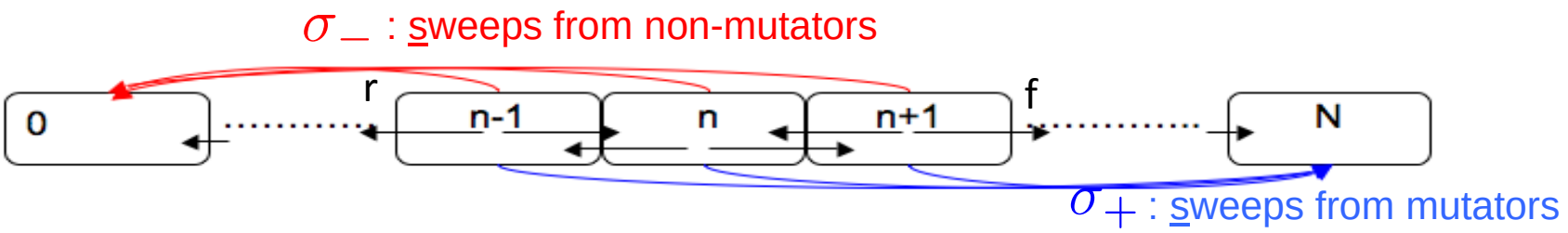


fix now

$$P_1 = \overbrace{\sigma_+}+$$

n = # mutators
f = probability (n → n+1)
r = probability (n → n-1)
P_n=fixation prob. of n mutators

How to calculate eventual P_{fix} for 1 mutator



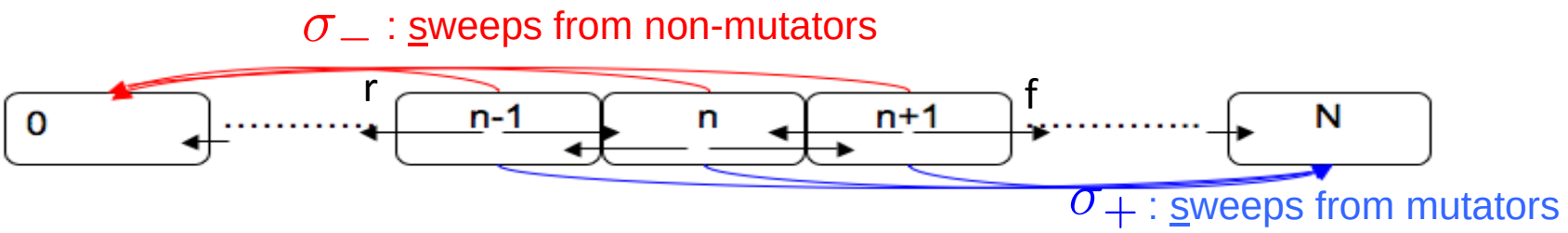
fix now

increase by 1, then fix

$$P_1 = \sigma_+ + f P_2 +$$

n = # mutators
f = probability (n → n+1)
r = probability (n → n-1)
P_n=fixation prob. of n mutators

How to calculate eventual P_{fix} for 1 mutator



fix now

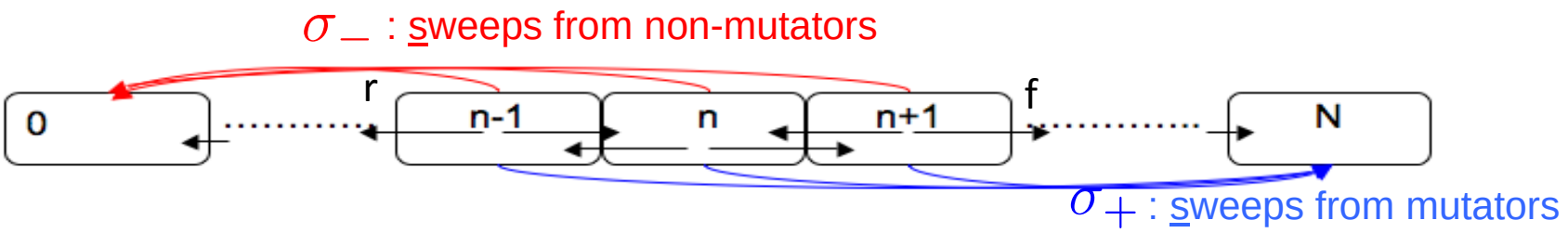
increase by 1, then fix

nothing happens; fix later

$$P_1 = \sigma_+ + f P_2 + (1 - (f + r + \sigma_+ + \sigma_-))P_1$$

n = # mutators
f = probability (n → n+1)
r = probability (n → n-1)
P_n=fixation prob. of n mutators

How to calculate eventual P_{fix} for 1 mutator



fix now

increase by 1, then fix

nothing happens; fix later

$$P_1 = \sigma_+ + f P_2 + (1 - (f + r + \sigma_+ + \sigma_-))P_1$$

This is the "backward recursion equation" for P_i. We have to evaluate f, r, $\sigma_\pm$ for this eq. to be of any use…

n = # mutators
f = probability (n → n+1)
r = probability (n → n-1)
P_n=fixation prob. of n mutators

f, r, and $\sigma_{\pm}$ in terms of experimental parameters (N, M, U_b, U_d, s_b, s_d)

key assumption #1: deleterious mutations always go extinct (sooner or later: doesn't matter when). If that's true, then we just have to keep track of # error-free mutators:

$$r = e^{-MU_d} \approx 1 - MU_d$$
$$f = e^{-U_d} \approx 1 - U_d$$

notes/observations:

- r and f don't depend on s_d (b/c we don't care about the *kinetics* of extinction/fixation).
- When $M > 1$, $r > f$: selection against mutators b/c of deleterious mutations

key assumption #2: beneficial mutations occur "one-at-a-time" (neglect clonal interference)

- This is a *terrible* approximation for large laboratory populations (b/c $NU_b >> 1$).
- But I'm focusing on small (or recently bottlenecked) populations, where assumption is OK.
- So, probability that beneficial (driver) fixes is $\approx s_b$. And "rate of sweeps", ($\sigma_{\pm}$) given by

$$\sigma_+ \approx MU_b s_b$$
$$\sigma_- \approx (N-1)U_b s_b$$

M: mutators' elevated mutation rate

N-1: non-mutators severely outnumbered initially

Recursion equation is easily solved in most interesting case: mutators on verge of favored/disfavored (i.e. neutral)

fix now

increase by 1, then fix

nothing happens; fix later

$$P_1 = \sigma_+ + f P_2 + (1 - (f + r + \sigma_+ + \sigma_-))P_1$$

neutrality condition: $P_1 = 1/N$, $P_2 = 2/N$. After some algebra,

$$f - r + N\sigma_+ = \sigma_-$$

Trading $f, r, \sigma_\pm$ for experimental parameters, we arrive at this:

$$\boxed{N U_b s_b - U_d > 0}$$: conditions favoring mutators

Factors favoring mutators:

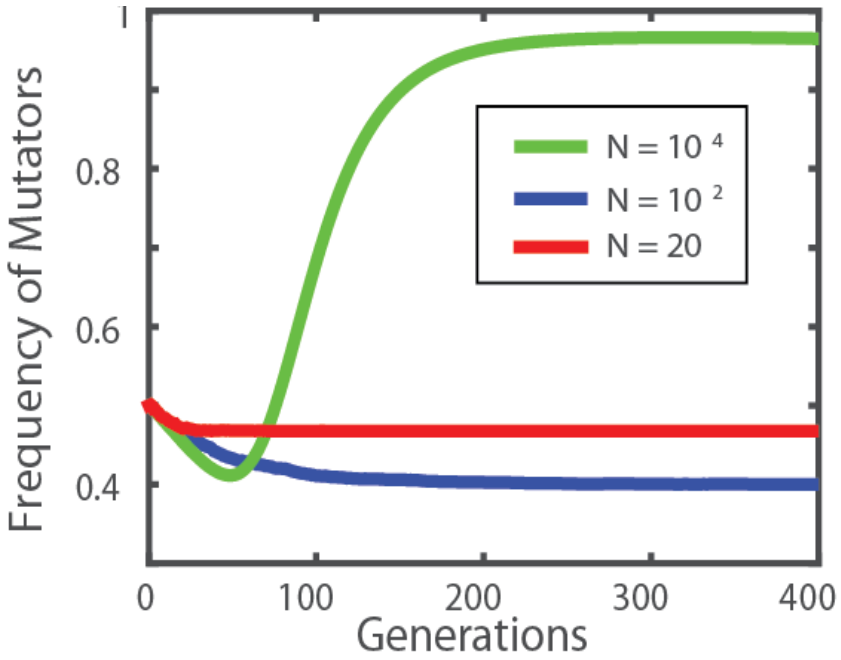
- Large U_b
- Small U_d
- Large s_b
- Large N? Can flip sign of inequality just by changing N?? "sign inversion"

Population size dictates the direction of selection on mutation rate

Analytic predictions re: "sign inversion" are borne out in explicit simulations, which

- relax "assumption #1" (that deleterious mutations must go extinct)
- relax "assumption #2" (which precluded clonal interference)

Average dynamics over many replicates



upward selection pressure on U
downward selectionpressure on U
"no" selection pressure on U

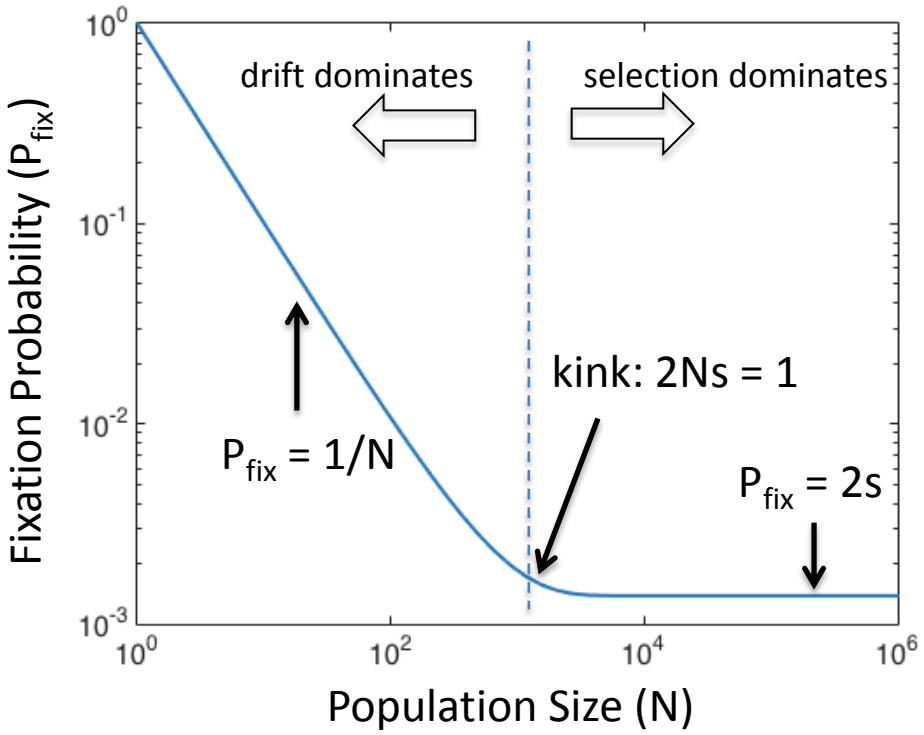
This is WEIRD!

- deeply population-level effect
- evolutionary outcome can't be predicted from the properties of individuals.
- Is this "evolutionary cell biology??"

rest of talk: dig deeply into why sign inversion happens and what are consequences

Sign inversion occurs because mutators are nearly impervious to drift

solid line is Kimura's formula: $$P_{\text{fix}} = \frac{1 - e^{-s}}{1 - e^{-Ns}}$$

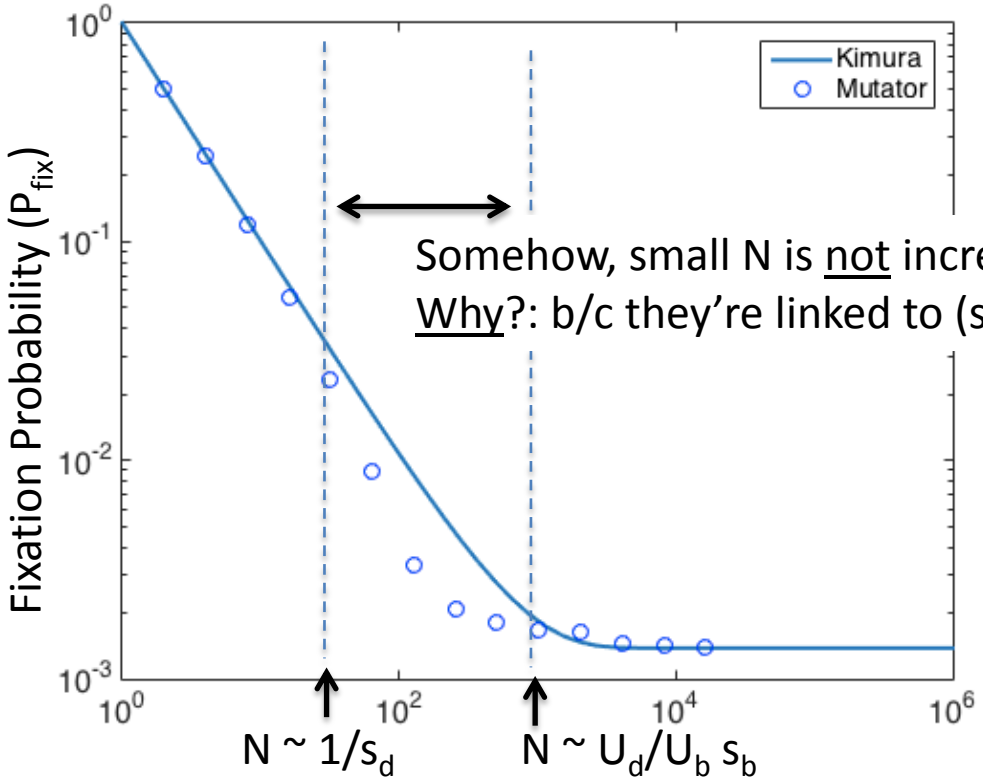


Seems weird at first, but true: Small *N* increases P_{fix} (b/c it's easier to hit the absorbing state at n=N).

In this sense, "drift" increases chances of fixation

Sign inversion occurs because mutators are nearly impervious to drift

solid line is fit to Kimura's formula: $P_{\mathrm{fix}} = \dfrac{1 - e^{-s}}{1 - e^{-Ns}}$ circles = mutator simulations



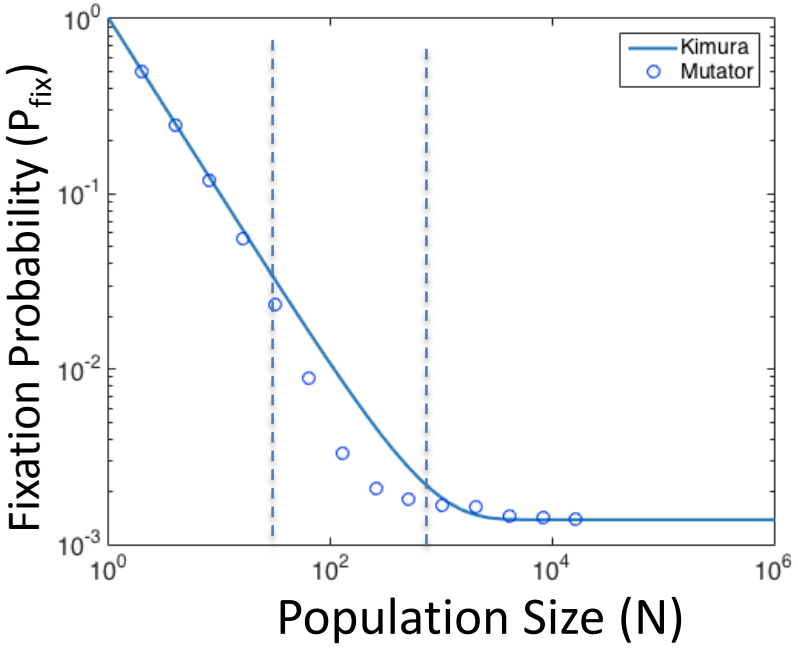
Somehow, small N is <u>not</u> increasing P_{fix} for mutators

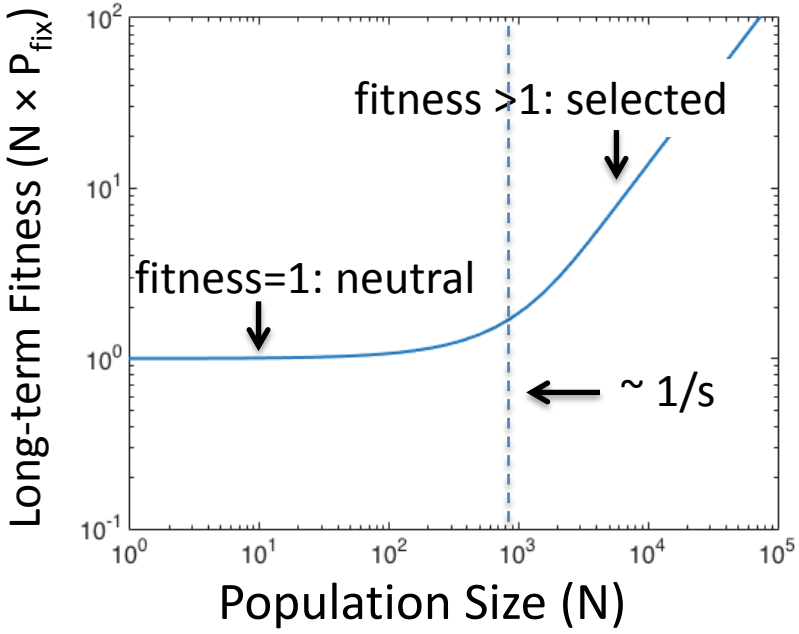
<u>Why</u>?: b/c they're linked to (strong) driver/anchor mutations

<u>key point</u>: 2 critical N values for mutators, but only 1 for ordinary mutations.

Sign inversion occurs because mutators are nearly impervious to drift

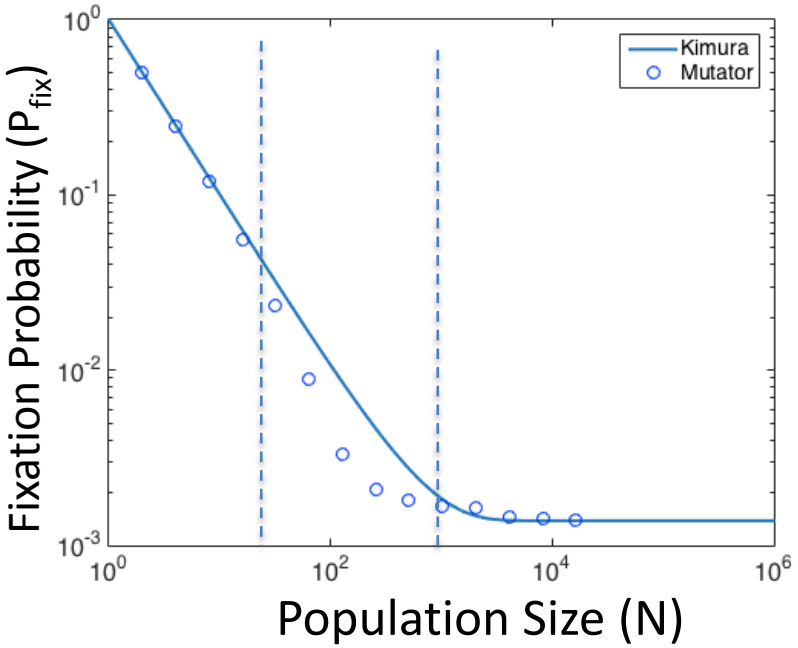
solid line is fit to Kimura's formula: $$P_{\text{fix}} = \frac{1 - e^{-s}}{1 - e^{-Ns}}$$

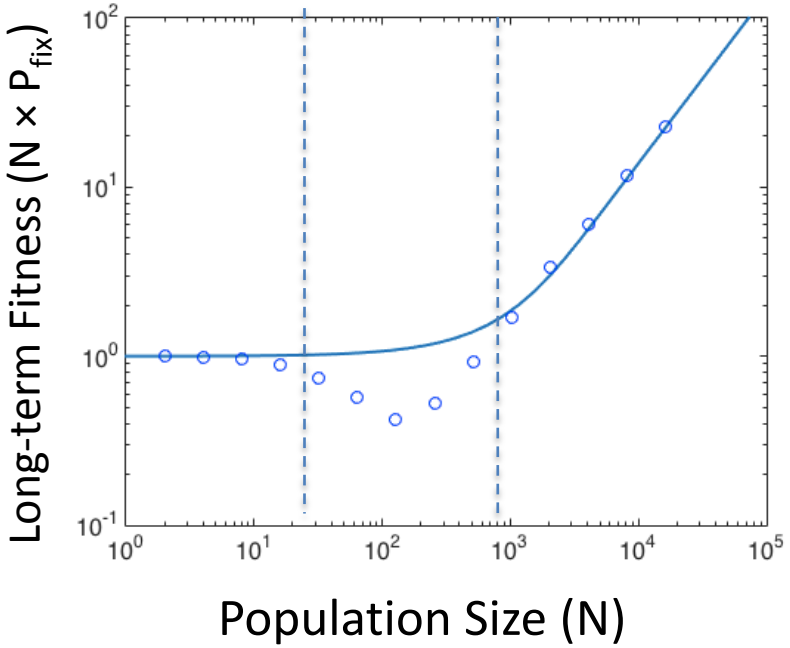




sign inversion <u>impossible</u> in textbook case (Kimura's formula): $NP_{fix} >= 1$

Sign inversion occurs because mutators are nearly impervious to drift



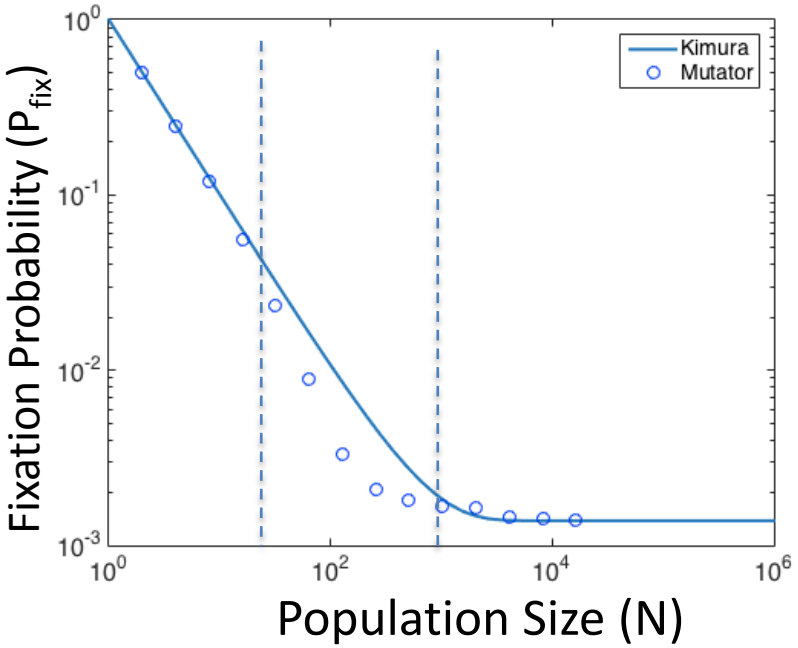


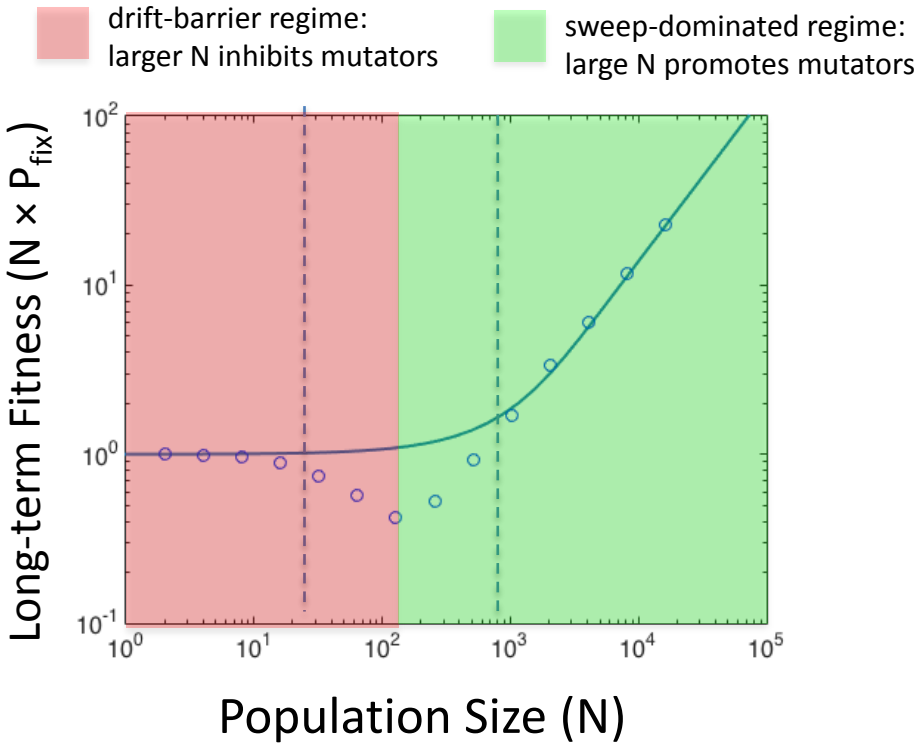
"Dip" on right-hand-side is direct consequence of "lag" of left-hand-side.

We hypothesize that the "dip" occurs in other instances of indirect selection, e.g.:

- beneficial mutations in a changing environment
- modifiers of recombination rate

Sign inversion occurs because mutators are nearly impervious to drift





"Dip" on right-hand-side is direct consequence of "lag" of left-hand-side.

We hypothesize that the "dip" occurs in other instances of indirect selection, e.g.:

- beneficial mutations in a changing environment
- modifiers of recombination rate

Further connection with drift-barrier theory

fix now

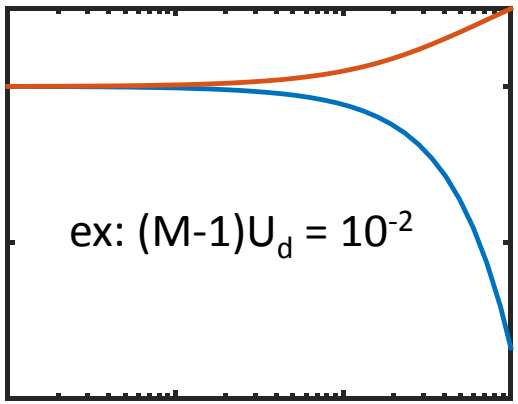
increase by 1, then fix

nothing happens; fix later

$$P_1 = \sigma_+ + f P_2 + (1 - (f + r + \sigma_+ + \sigma_-)) P_1$$

Drift barrier theory (typicall?) neglects beneficial mutations (U_b=0). In terms of the equation written above, this means that $\sigma_{\pm}$ =0 (no sweeps). This leads to:

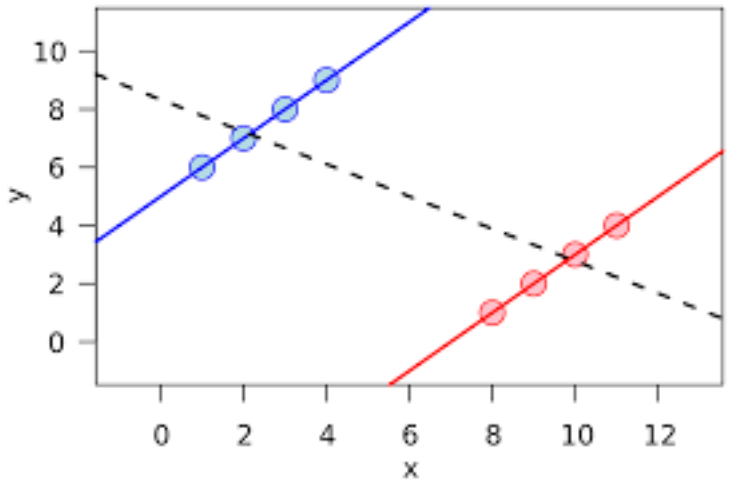
$$N P_{\text{fix}} \approx \frac{1 - e^{(M-1)U_d}}{1 - e^{N(M-1)U_d}}$$



This recovers the drift barrier prediction:
red: anti-mutators favored in large populations
blue: mutators favored in small populations

But only if U_b=0!

Simpson's paradox (in general)

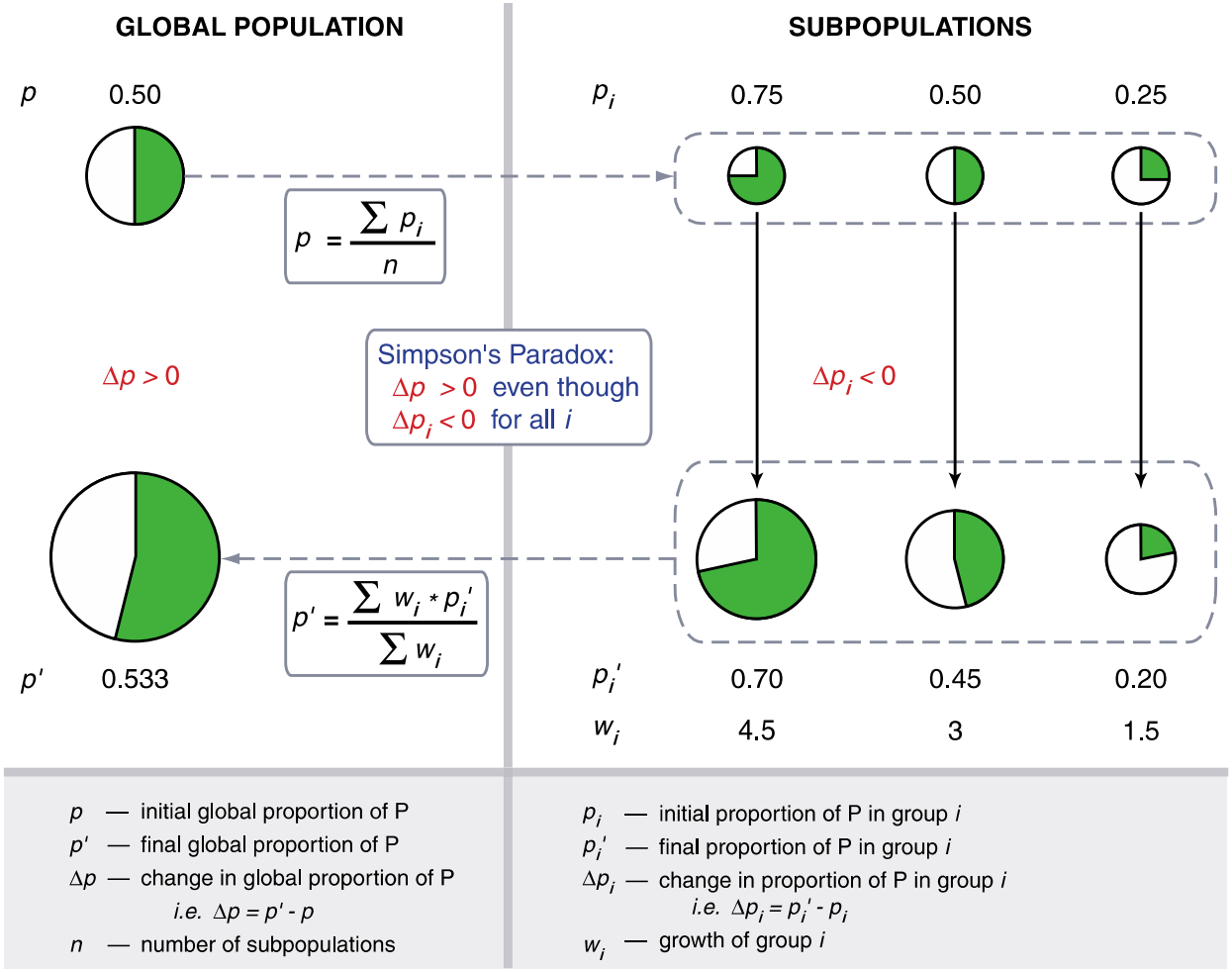


Roughly:

The direction of a trend changes when "things" are pooled.

See the excellent wikipedia on this!

Previous (experimental) example of Simpson's paradox in microbial evolution:



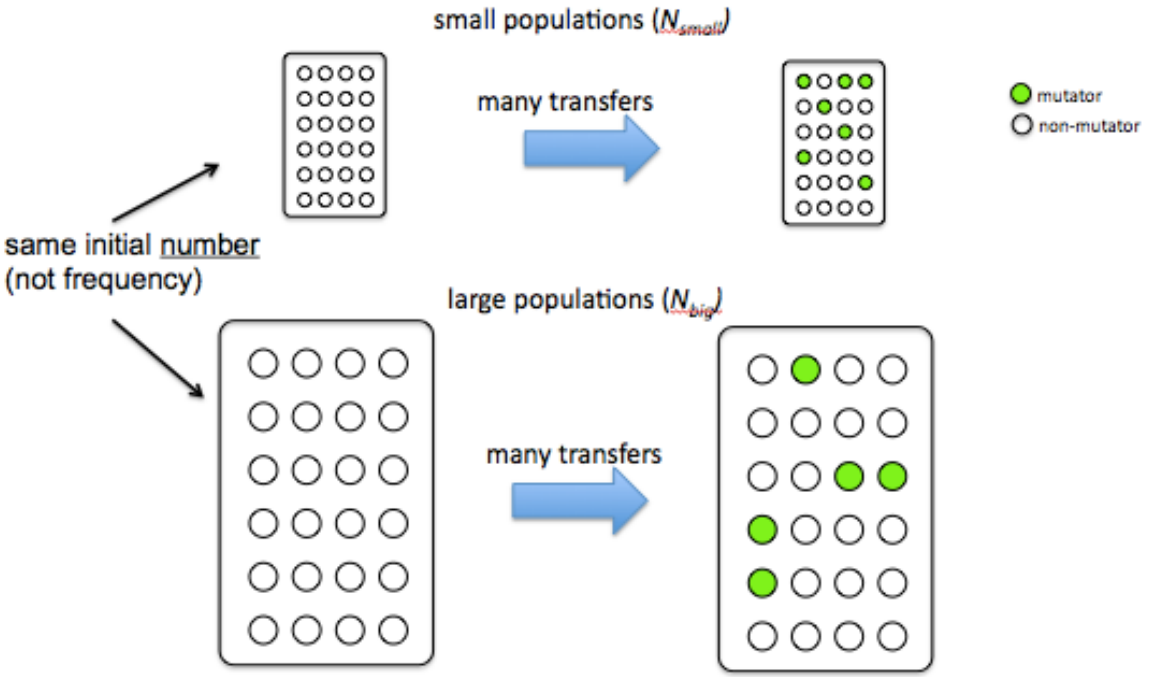
Chuang et. al 2009

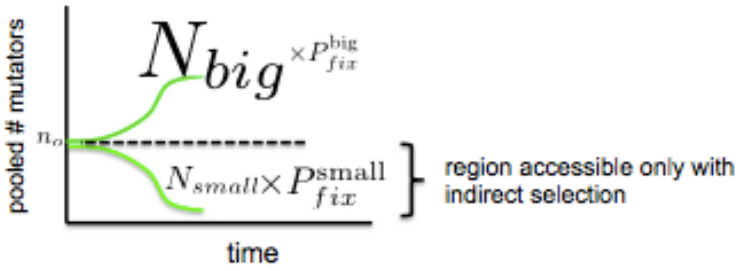
green: "producers"

white: "non-producers"

producers (≈ cooperators) lose <u>every</u> battle but win the war!!

Simpson's paradox in mutation rate evolution





$P_{fix}^{\text{small}} > P_{fix}^{\text{big}}$: small populations favor mutators on well-by-well basis.

$N_{\text{small}} \cdot P_{fix}^{\text{small}} < N_{\text{big}} \cdot P_{fix}^{\text{big}}$: small populations inhibit mutators when pooled.

Simpson's paradox

Future directions

1. Test theoretical predictions using fluorescently labeled yeast. Control population size with periodic bottlenecks (this works in simulations).

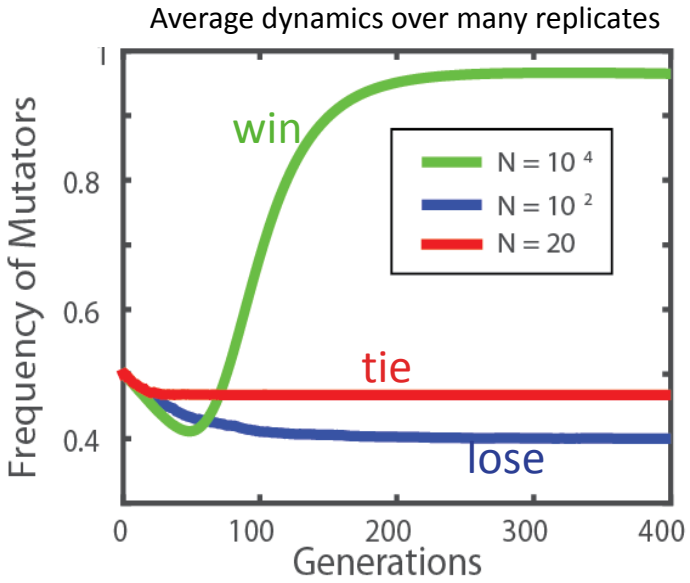
Eugene Raynes

Paul Sniegowski

1. One man's replicate wells are another man's spatially structured meta-population. Is spatial structure sufficient to induce sign-inversion and Simpson's paradox in :
    1. simulations
    2. laboratory yeast experiments

Summary

1. Mutators “lose” in small populations and “win” in large ones: “sign inversion”
2. We understand this effect analytically.
3. Sign inversion leads to Simpson’s paradox
4. The mechanistic origin of Simpson’s paradox may apply generically to other instances of indirect selection.



Thank you for your attention and, **crucially**, your wise criticisms!