인하대학교 INHA UNIVERSITY

Statis

eLife

Evolutionary Biology

Rafting a waterfall: Artificial selection for collective composition can succeed or fail depending on the initial and target values

Reviewed Preprint
v2 • March 3, 2025
Revised by authors

Can We Select Desired Microbial Collectives? Answers from Statistical Physics

PHYSICAL REVIEW RESEARCH **8**, 013227 (2026)

General theory for group resetting with application to avoidance

Hye Jin Park (Inha University), NSPCS, 2026.07.20

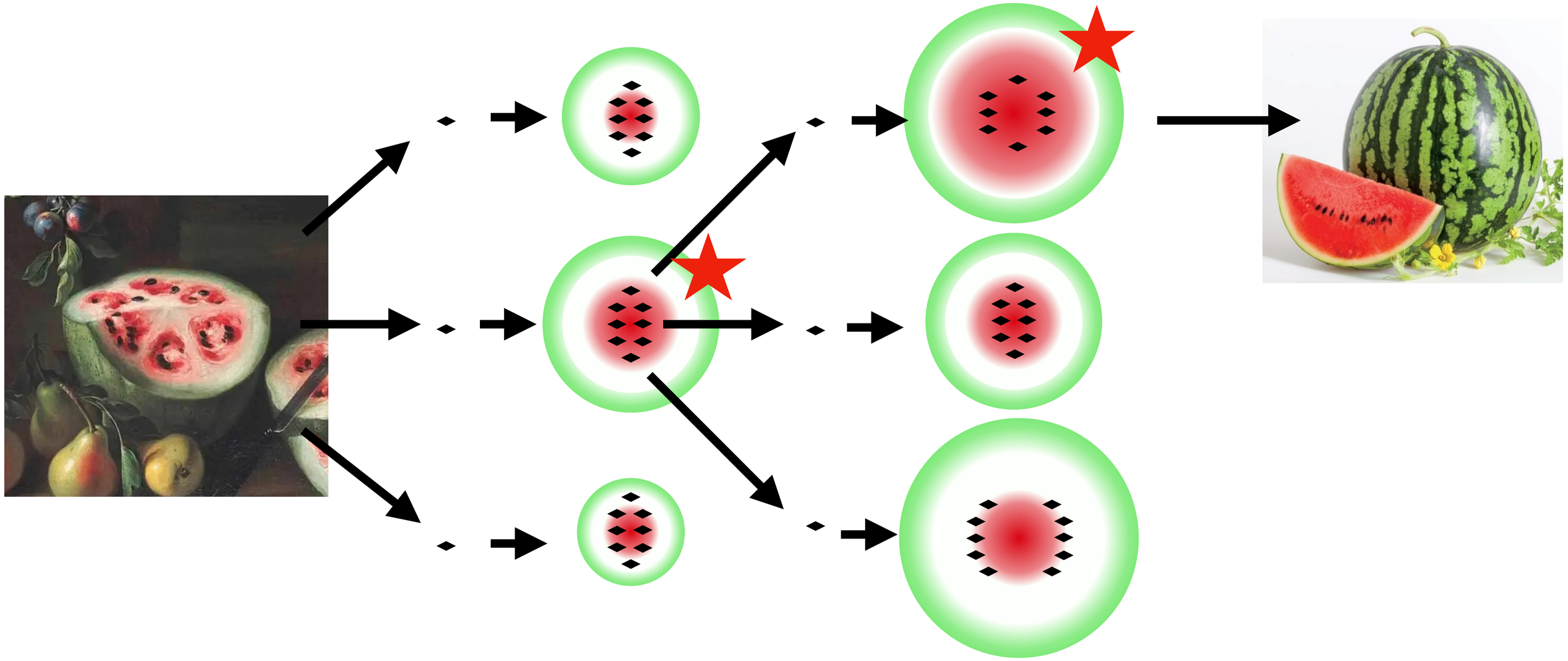
In 17th century painting



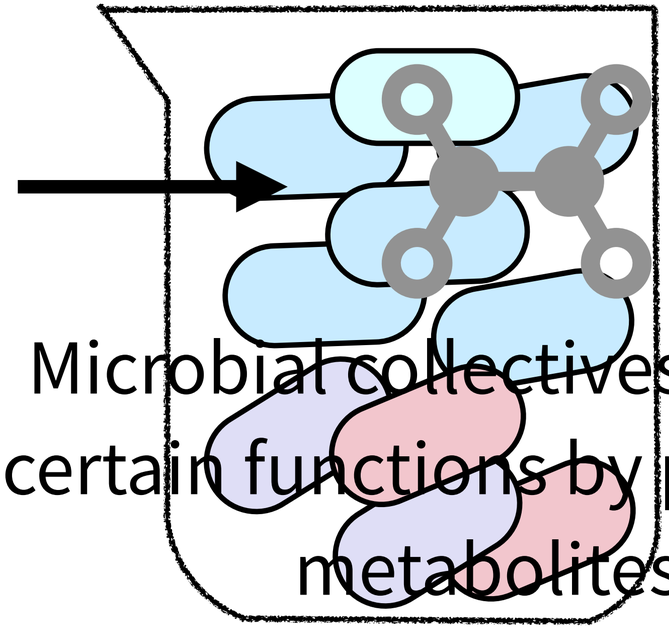
Present watermelon



Artificial selection



Selection on microbial collectives not individuals



Microbial collectives perform certain functions by producing metabolites.

Function

Biofuel production



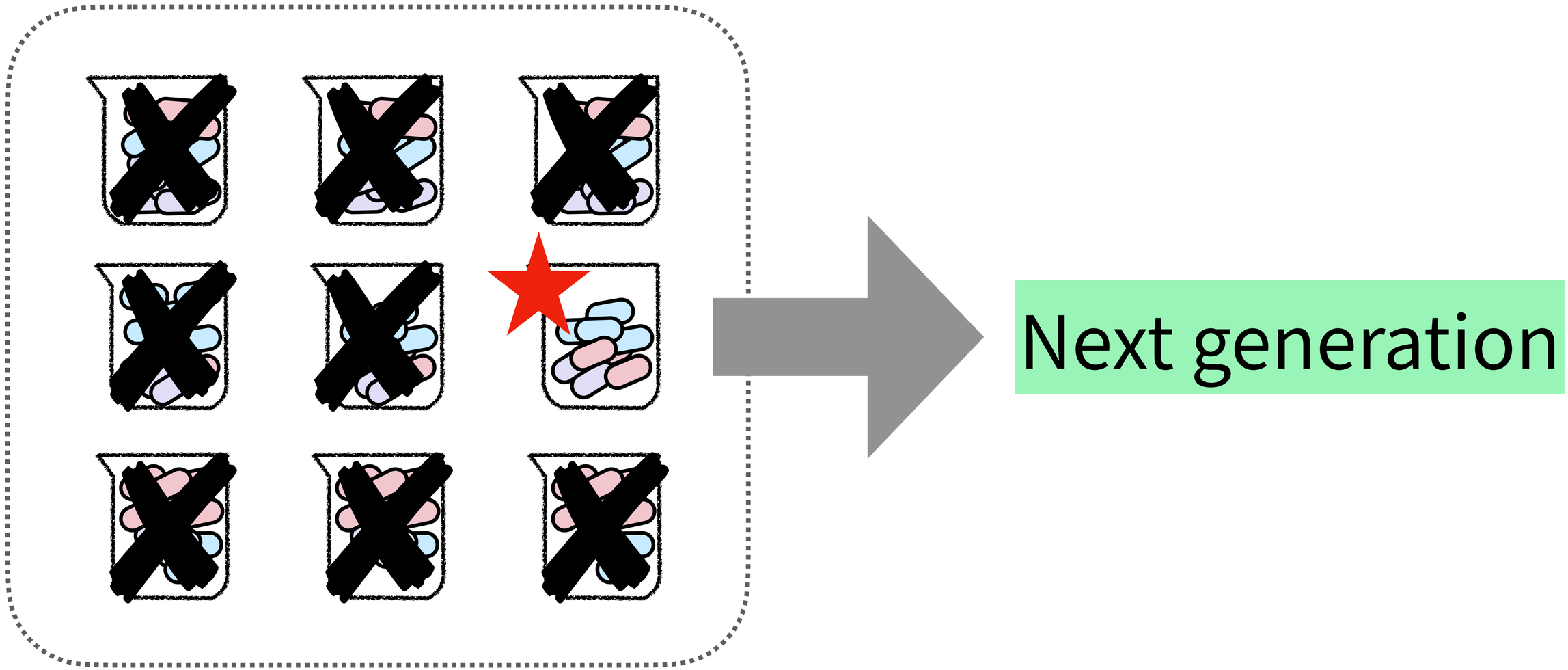
Waste degradation



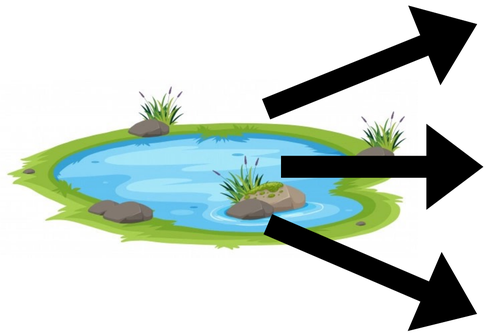
Plant growth

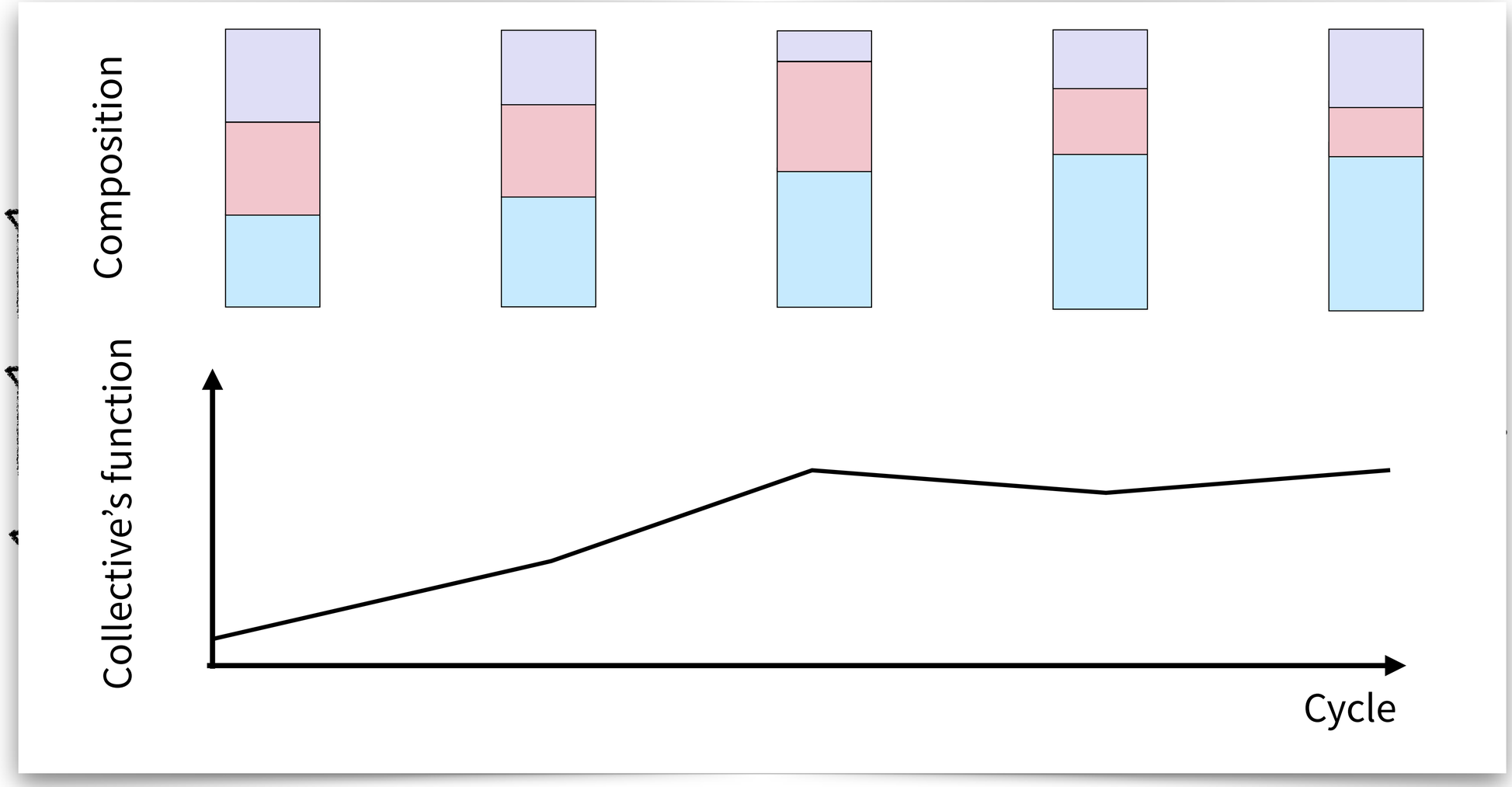


Artificial selection on microbial collectives



Artificial selection on microbial collectives



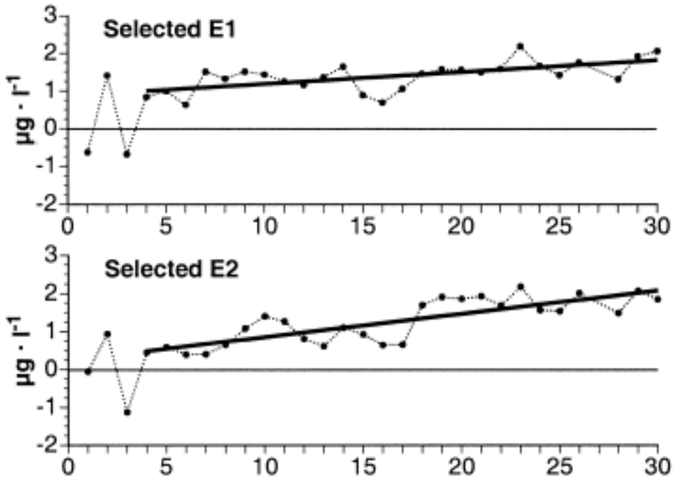


Experiments

Á. Sánchez *et al.*, Annu. Rev. Biophys. **50**, 323-341 (2021)

Function Reducing water pollution

Result



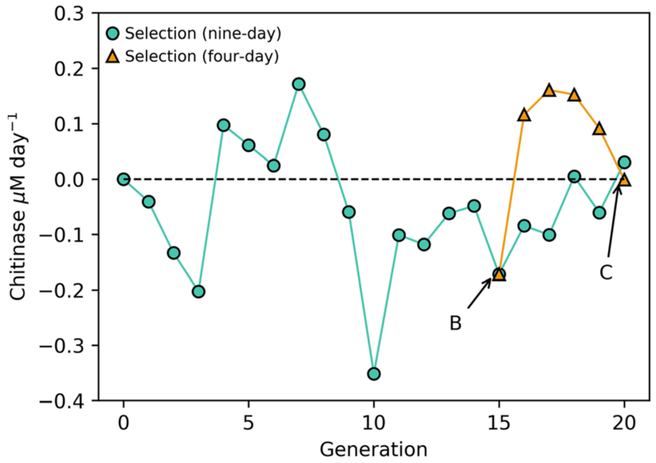
The amount of substance that reduces water pollution increases.

Success of selection!

Wright, R.J., Gibson, M.I. & Christie-Oleza, J.A., *Microbiome* **7,** 85 (2019)

Function Decomposing Chitin

Result



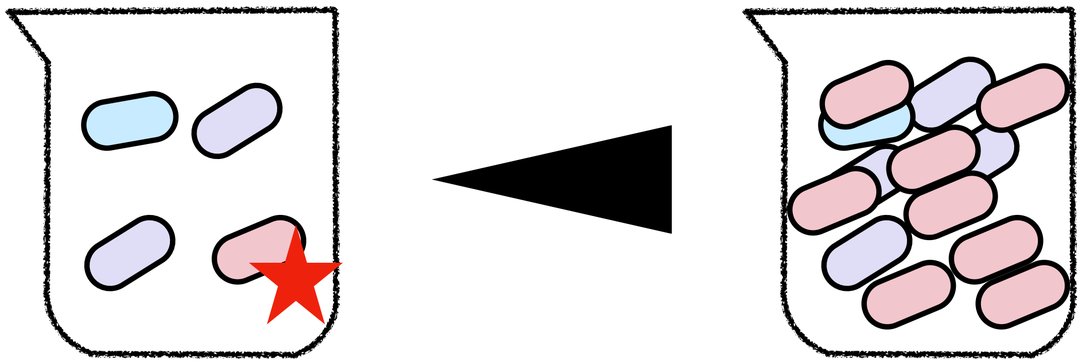
The amount of Chitinase was not increased.

Failure of selection!

Multilevel selection

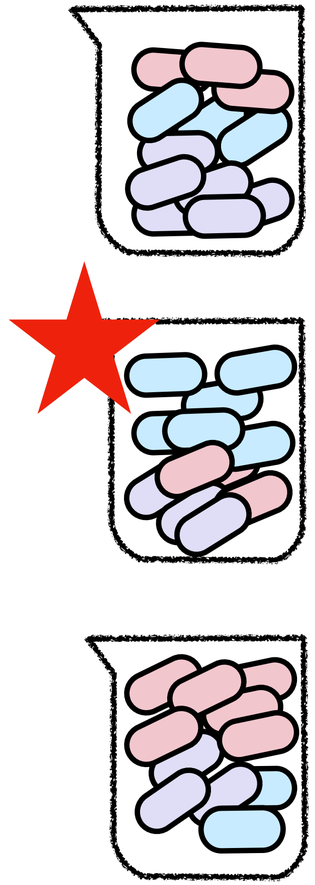
Growth

Individual-level selection during growth stage



Selection

Collective-level selection at selection stage



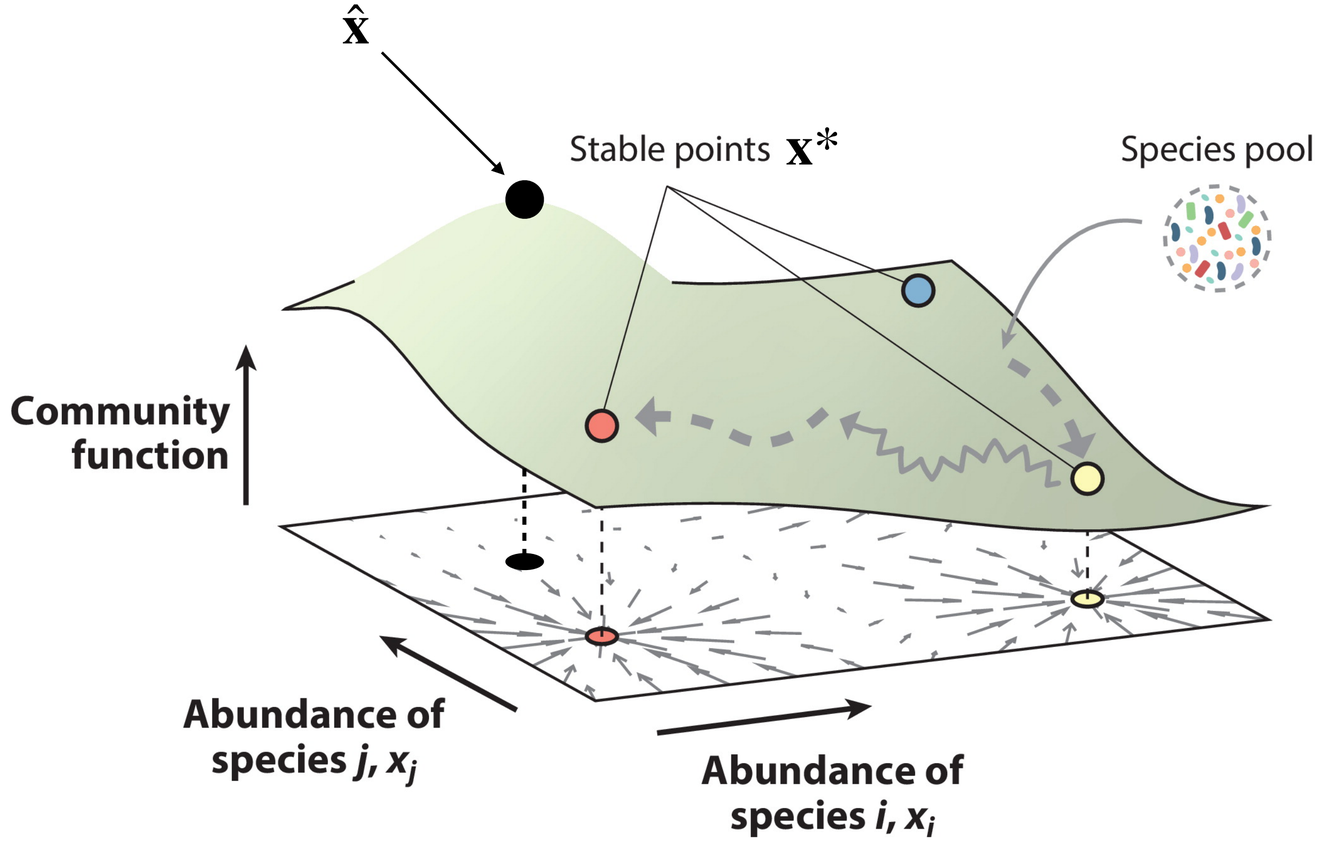
Nature selection VS artificial selection [Sánchez et al., Annu. Rev. Biophys. 50, 323 (2021)]

x_i: abundance of type i

Growth Stable fixed points $\mathbf{x}^*$

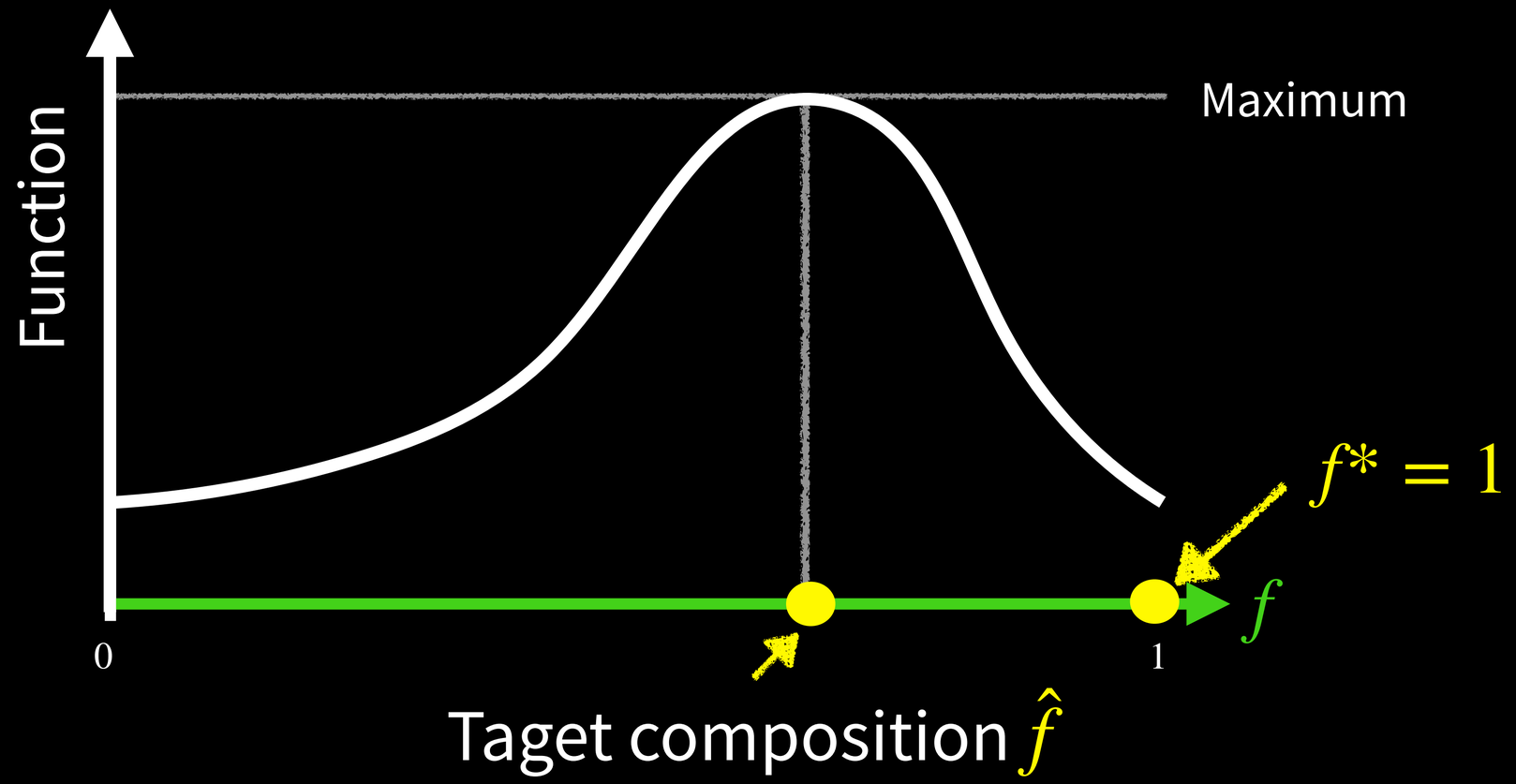
Selection Maximize at $\hat{\mathbf{x}}$

$$\mathbf{x}^* \neq \hat{\mathbf{x}}$$



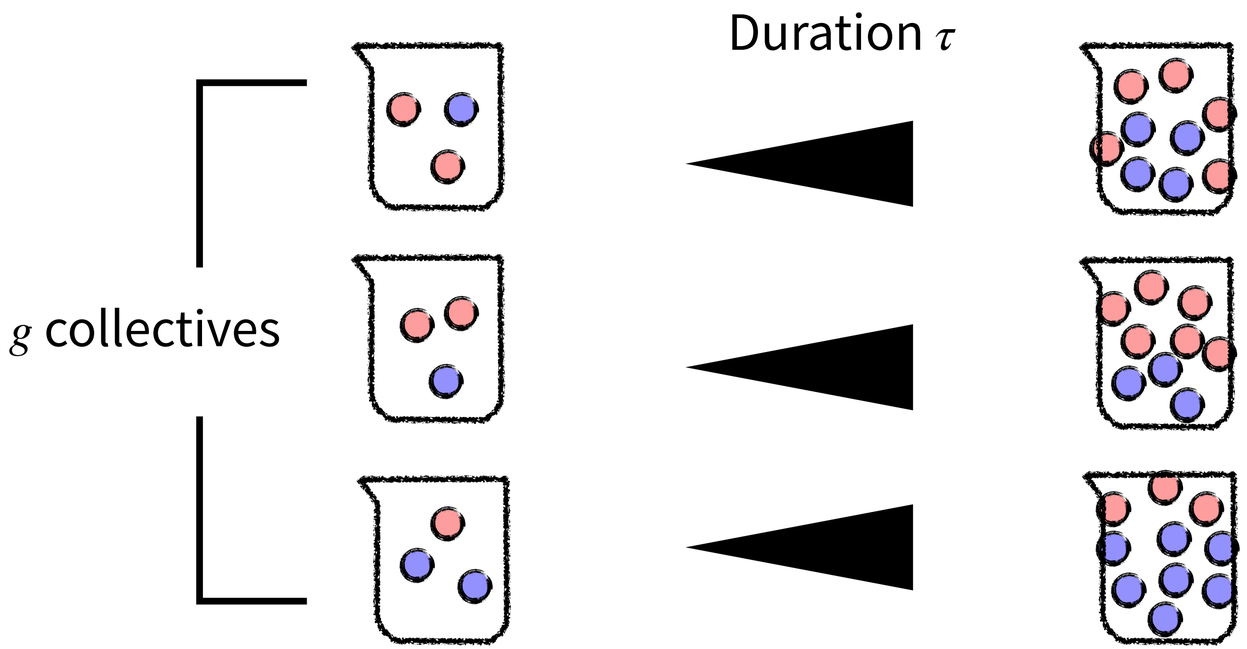
We propose a simple model to analyze the condition where the artificial selection on microbial community is successful.

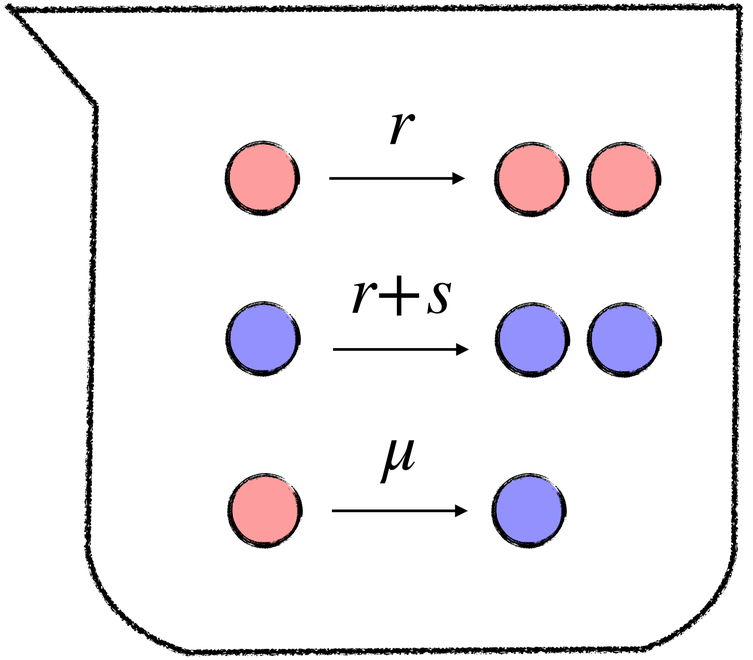
Composition = Frequency $f(w,m) = \dfrac{m}{w+m}$



Model

Growth





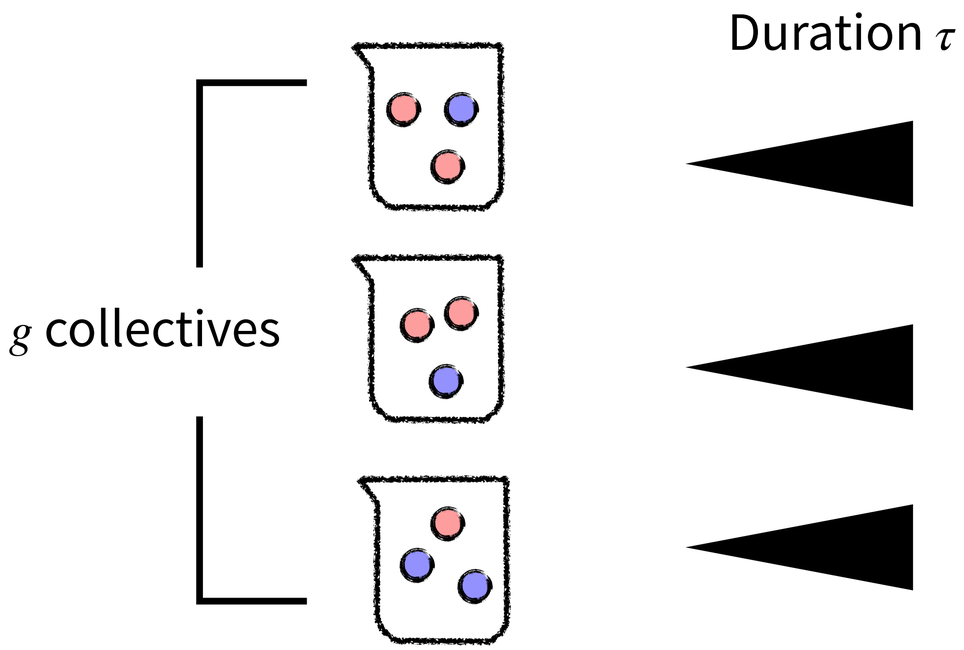
Initial total cell number: N_0
# of wildtype cells: w
# of mutants: m
Initial mutant frequency: f_0

Wildtype

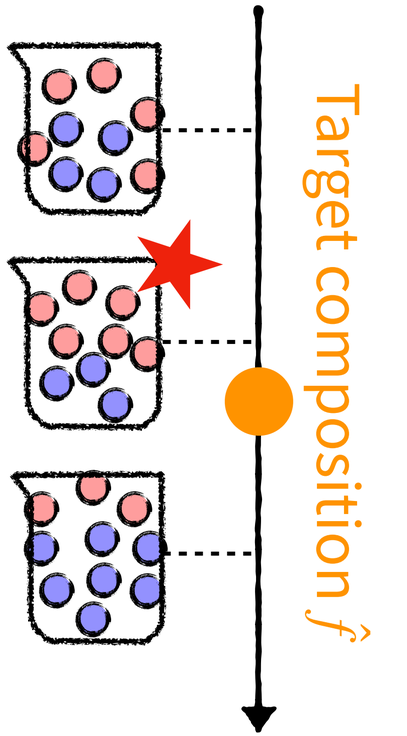
Mutant type

Model

Growth



Selection

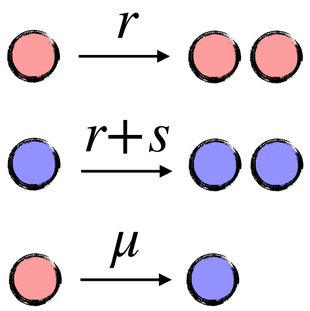


Initial total cell number: N_0

# of wildtype cells: w

# of mutants: m

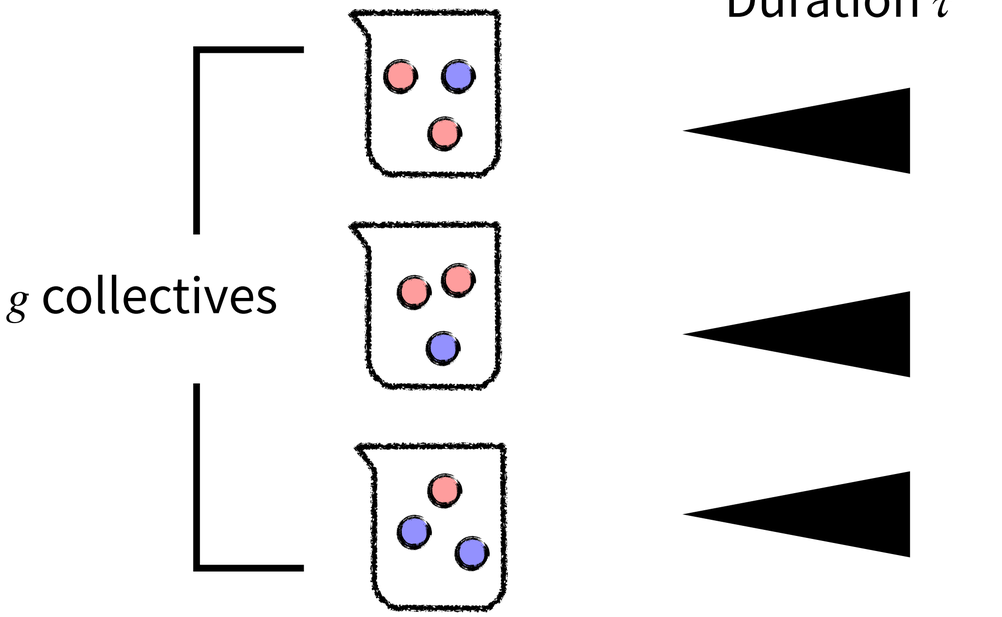
Initial mutant frequency: f_0



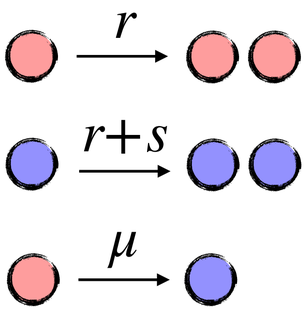
$$f = \frac{m}{w+m}$$

Model

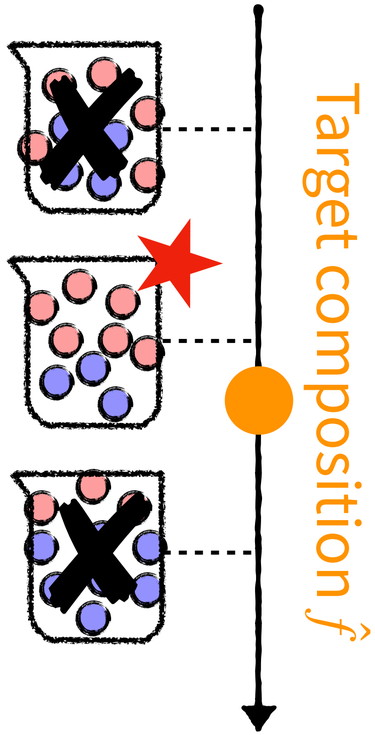
Growth



Initial cell number: N_0
# of wildtype cells: w
# of mutants: m
Initial mutant frequency: f_0

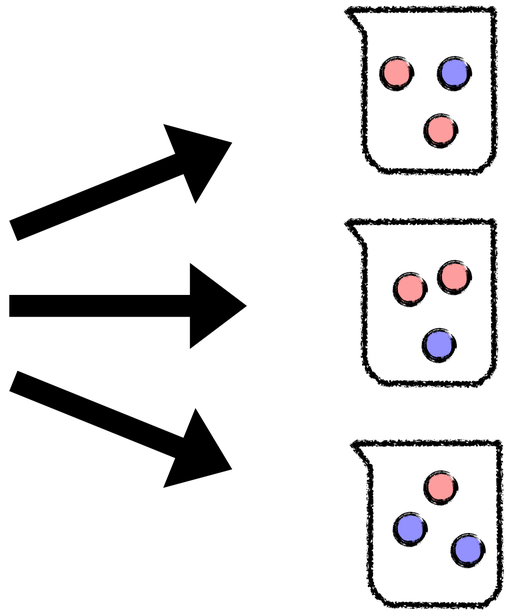


Selection



$$f = \frac{m}{w+m}$$

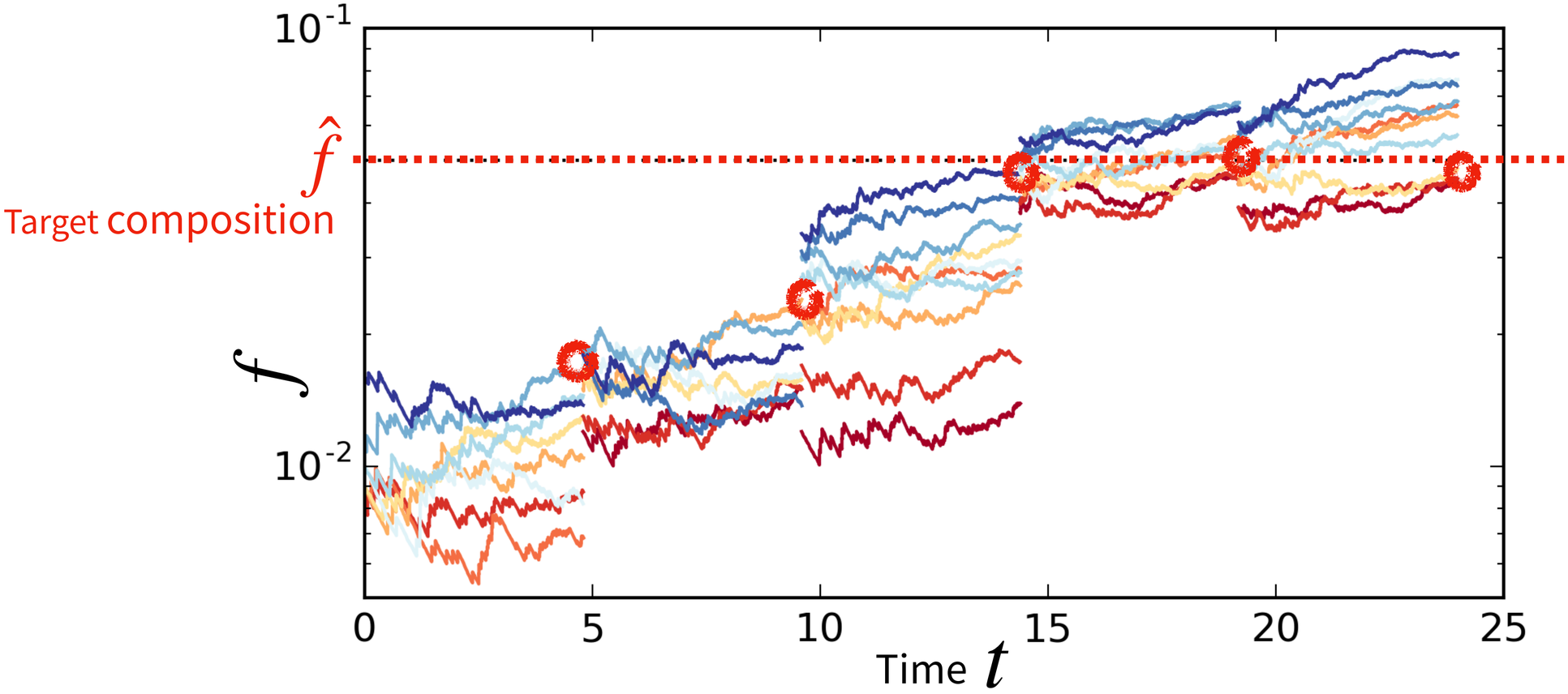
Reproduction



Total cell numbers
in each collective: N_0

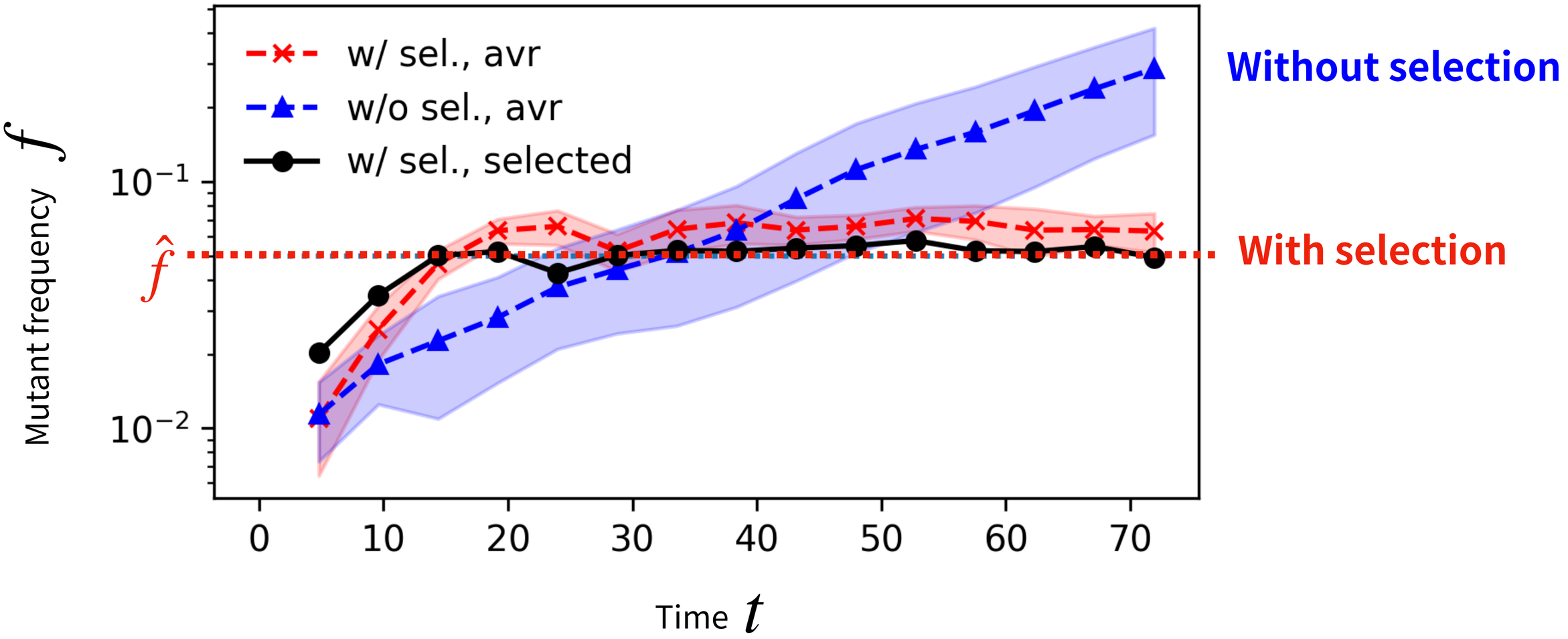
Simulation result

$g = 10,\ N_o = 1000,\ \bar{f}_o = 0.01,\ r = 0.5,\ s = 0.05,\ \mu = 10^{-4}, \tau = \log(g+1)/r \approx 4.8,\ \hat{f} = 0.05$



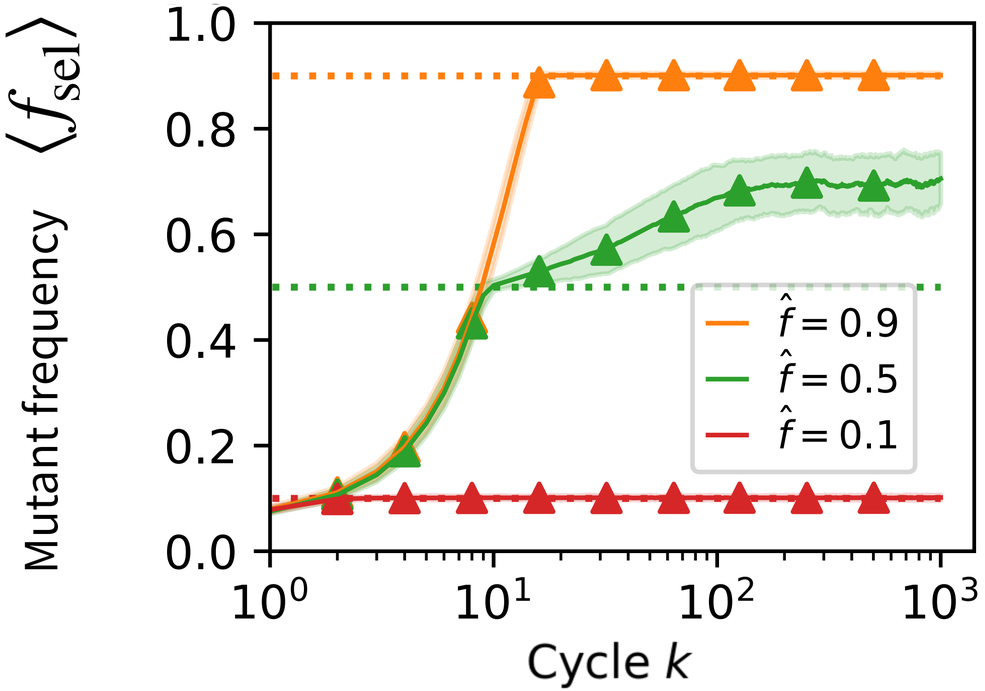
Effects of selection

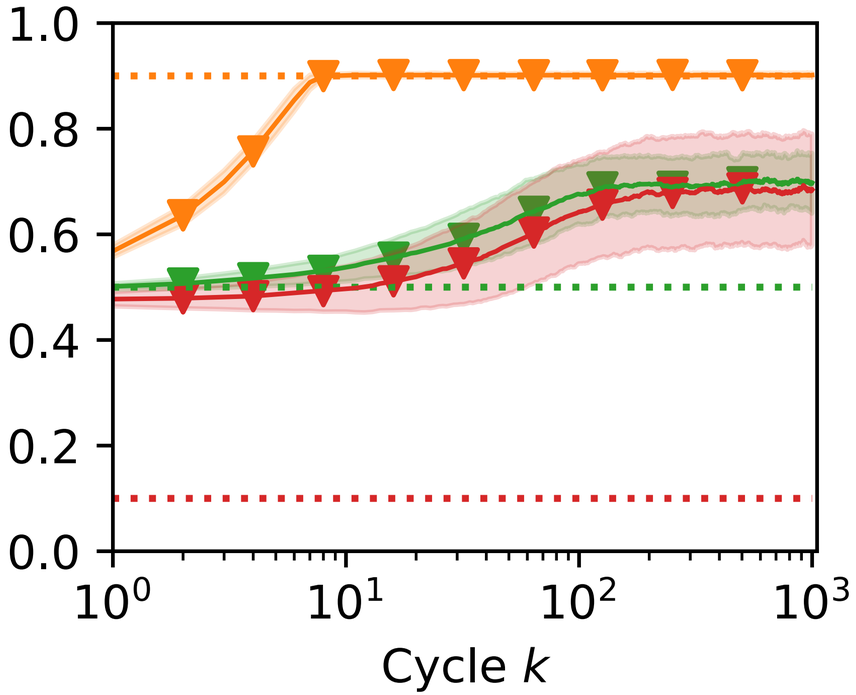
$g = 10,\ N_o = 1000,\ \bar{f}_o = 0.01,\ r = 0.5,\ s = 0.05,\ \mu = 10^{-4}, \tau = \log(g+1)/r \approx 4.8,\ \hat{f} = 0.05$



Various target compositions $\hat{f}$ and initial compositions f_0

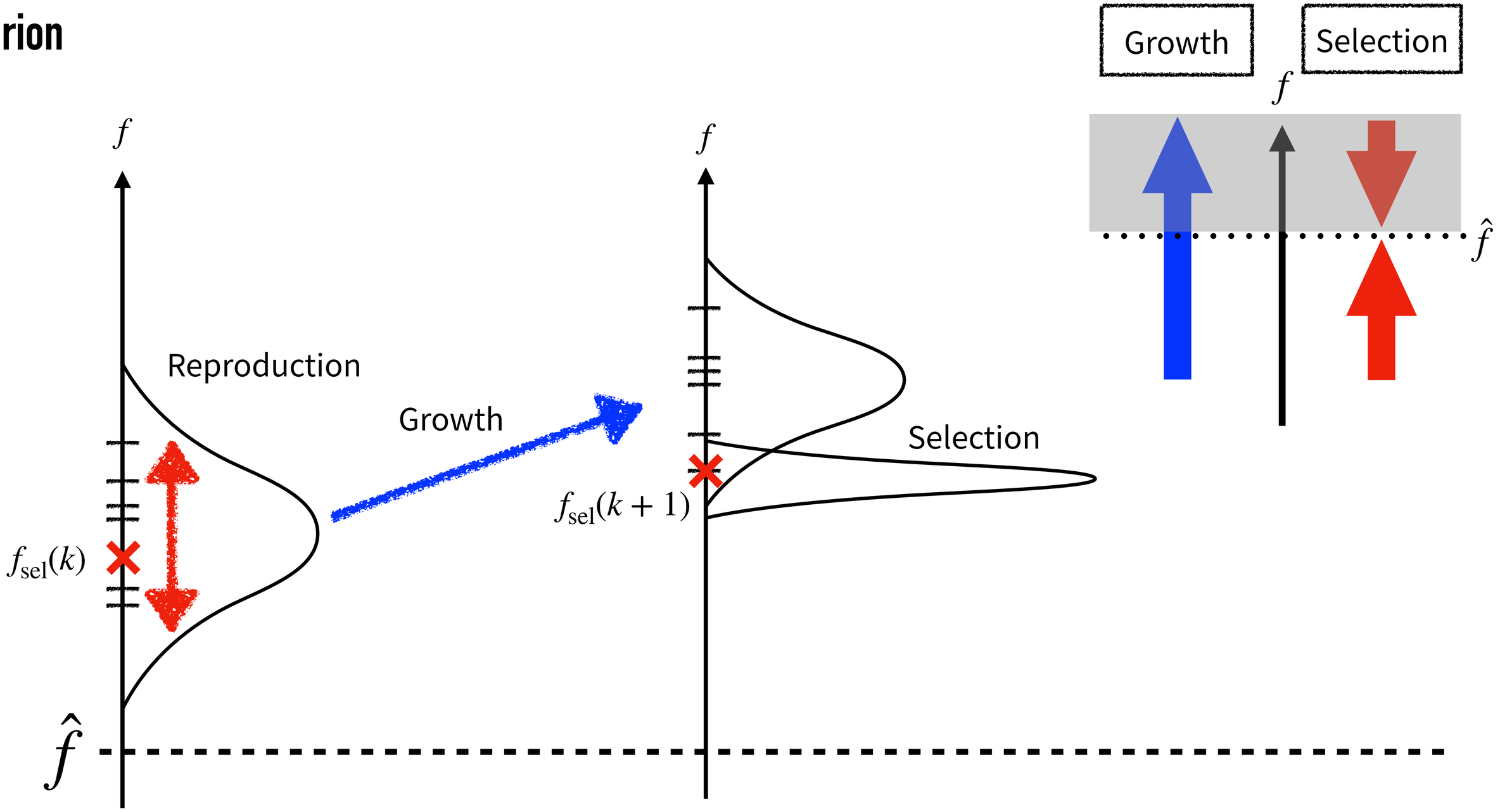
Not all target frequencies can be achieved.



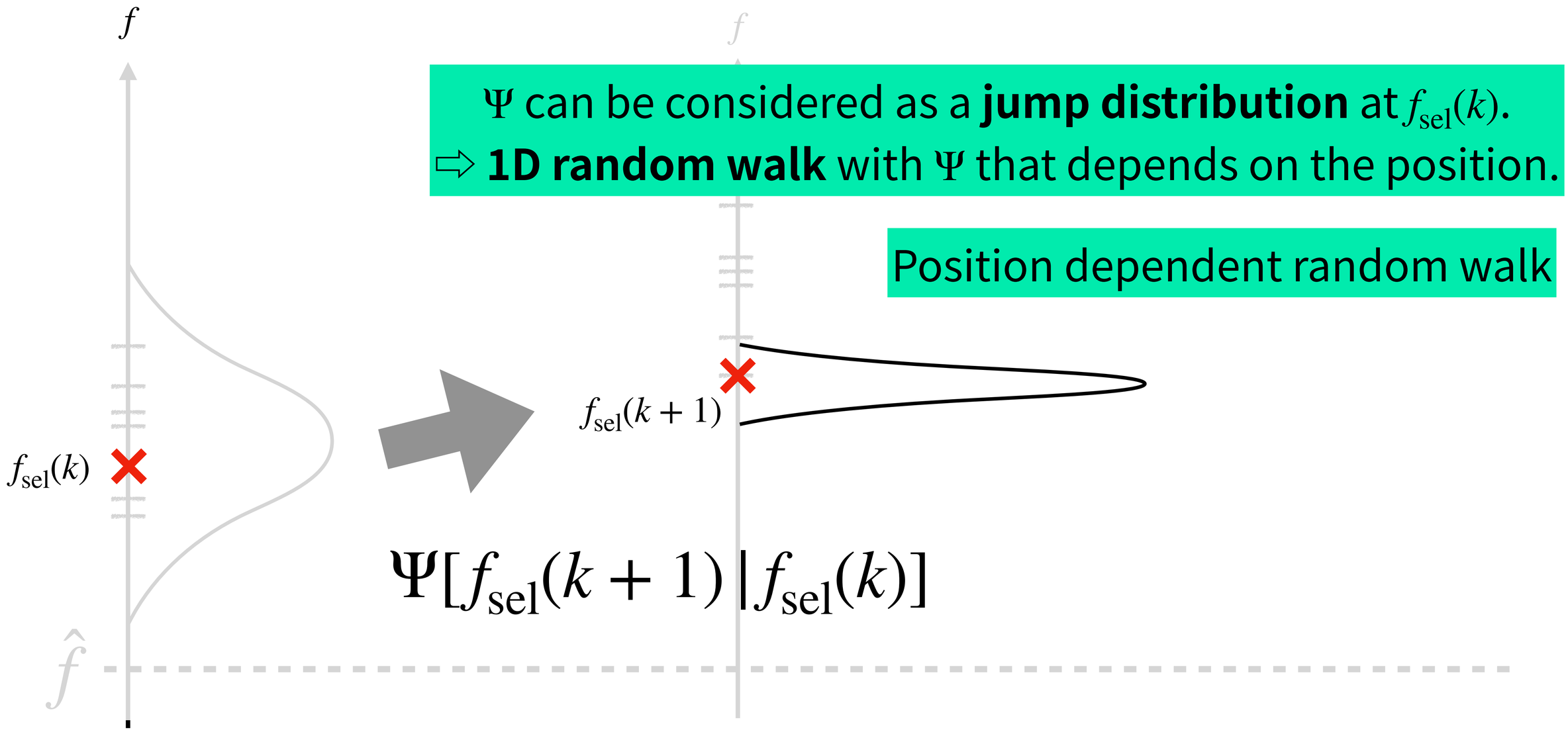


$g = 10,\ N_o = 1000,\ r = 0.5,\ s = 0.03,\ \mu = 10^{-4}, \tau = \log(g+1)/r \approx 4.8$

Criterion


Growth
Selection
f
$\hat{f}$
f
Reproduction
Growth
f
Selection
$f_{\mathrm{sel}}(k)$
$f_{\mathrm{sel}}(k+1)$
$\hat{f}$

Criterion



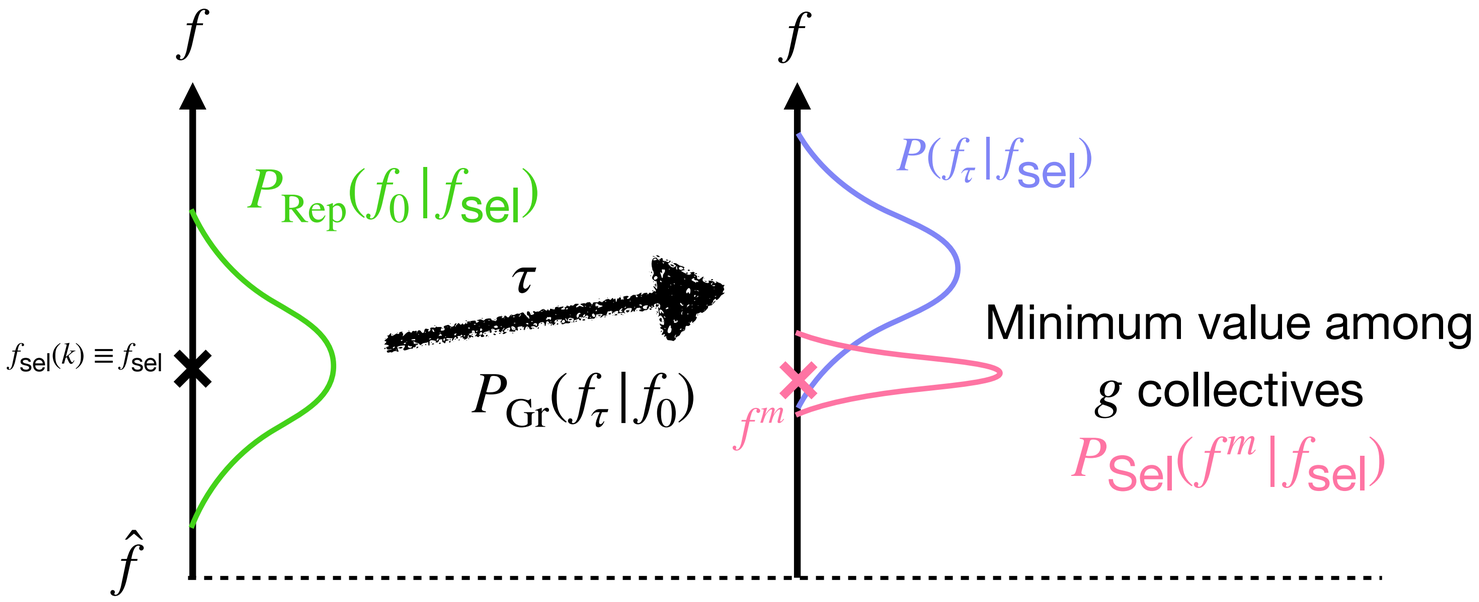
Ψ can be considered as a **jump distribution** at $f_{\mathrm{sel}}(k)$.
⇨ **1D random walk** with Ψ that depends on the position.

Position dependent random walk

$$\Psi[f_{\mathrm{sel}}(k+1)\,|\,f_{\mathrm{sel}}(k)]$$

Conditional probability $\Psi[f_{\mathrm{sel}}(k+1)\,|\,f_{\mathrm{sel}}(k)]$

$$P_{\mathrm{Rep}}(f_0\,|\,f_{\mathrm{sel}}) \approx \frac{N_0!}{(N_0 f_{\mathrm{sel}})!(N_0 - N_0 f_{\mathrm{sel}})!} f_{\mathrm{sel}}^{N_0 f_{\mathrm{sel}}} (1 - f_{\mathrm{sel}})^{N_0 - N_0 f_{\mathrm{sel}}}$$



$$P(f_\tau\,|\,f_{\mathrm{sel}}) = \int df_0 \;\; P_{\mathrm{Gr}}(f_\tau\,|\,f_0) P_{\mathrm{Rep}}(f_0\,|\,f_{\mathrm{sel}})$$

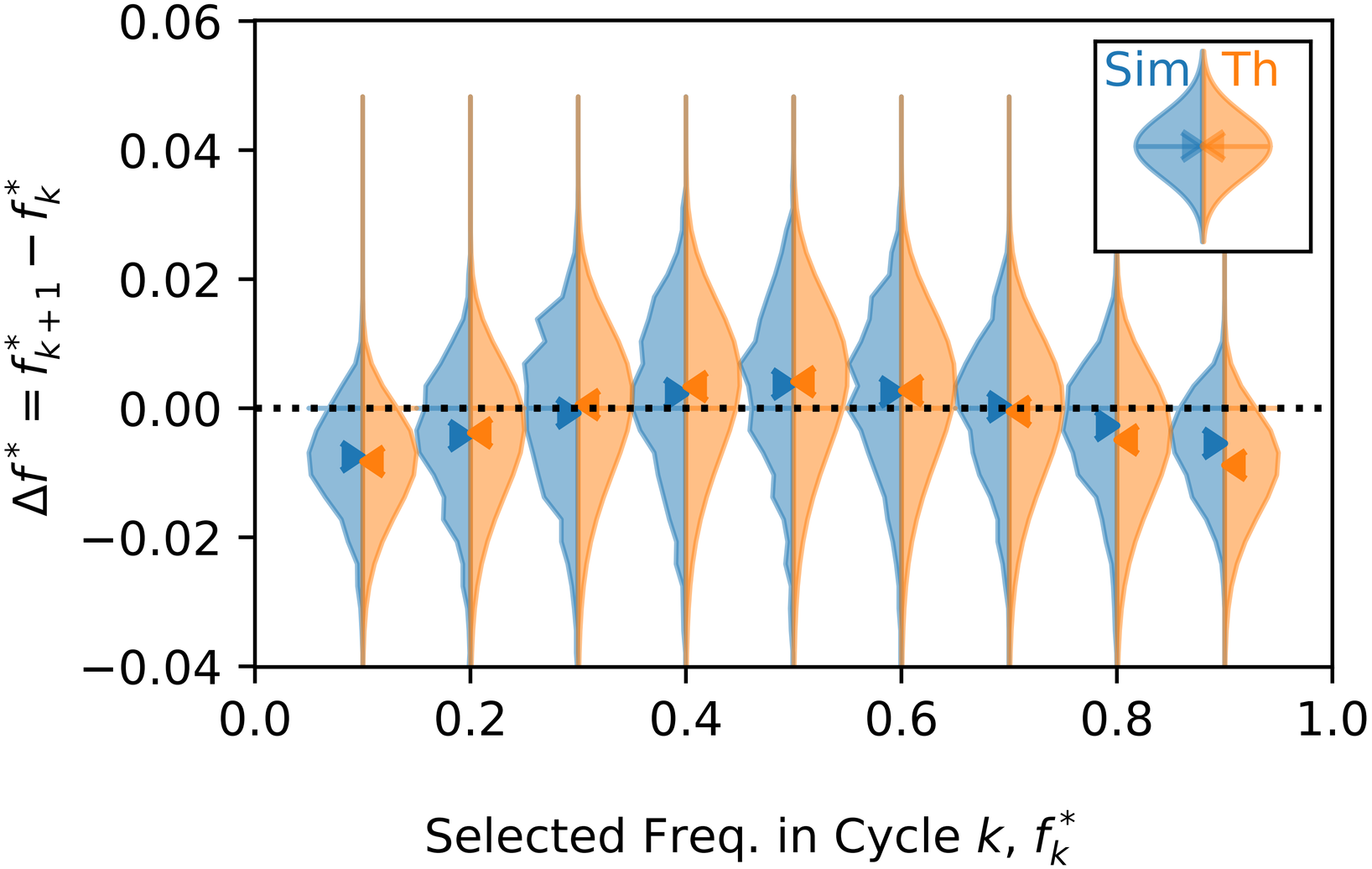
$$f^m = \min\{f^{(1)}, f^{(2)}, \cdots, f^{(g)}\}$$

$$P_{\mathrm{Sel}}(f^m\,|\,f_{\mathrm{sel}}) = g(1 - F(f^m))^{g-1} P(f^m\,|\,f_{\mathrm{sel}})$$

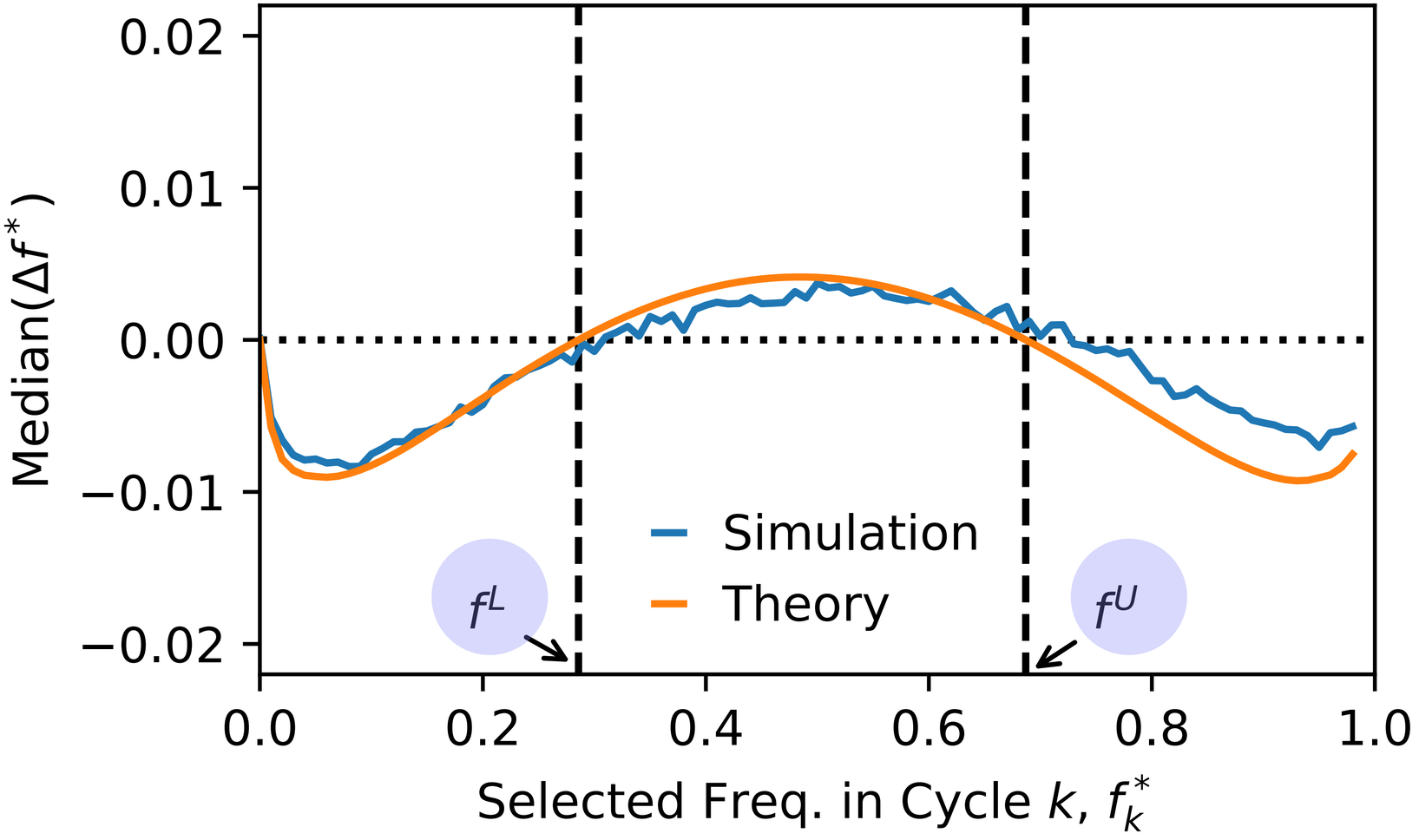
$$F(f^m) = \mathrm{Prob}[f^m < f] = \int_{f^m}^{\infty} P(f\,|\,f_{\mathrm{sel}}) df$$

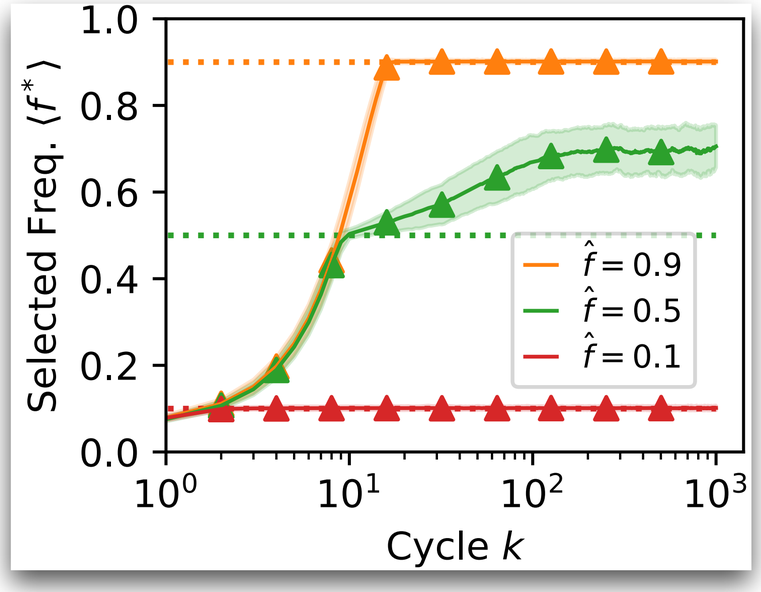
$$= \Psi[f_{\mathrm{sel}}(k+1)\,|\,f_{\mathrm{sel}}(k)]$$

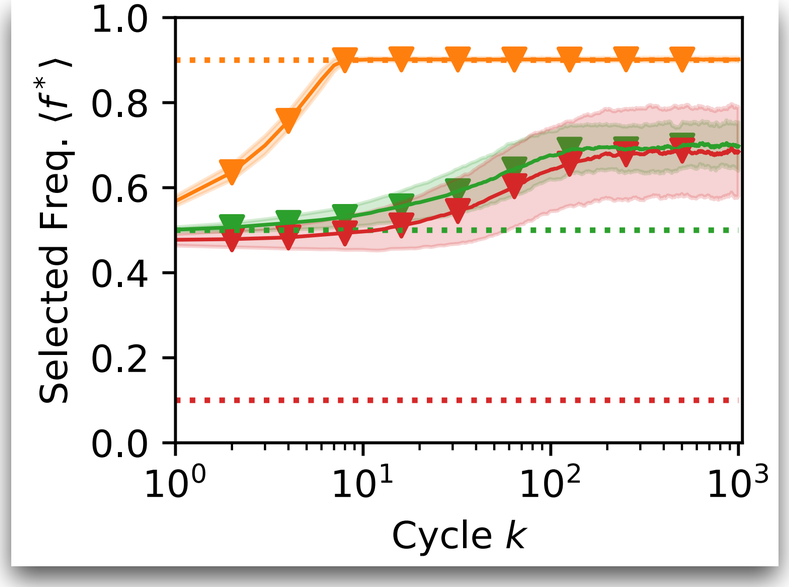
Conditional probability $\Psi[f_{\mathrm{sel}}(k+1)\,|\,f_{\mathrm{sel}}(k)]$



Upper and Lower Boundaries



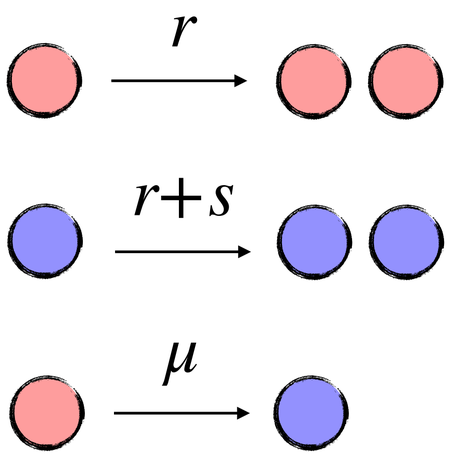




Parameters

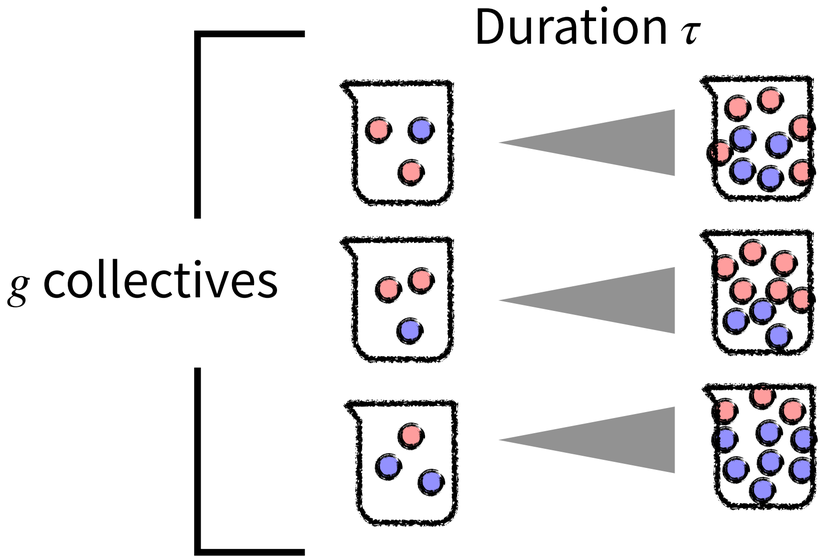
Microbial collectives

$$r, s, \mu, \hat{f}$$



Experimental set up

$$g, \tau, N_0, f_0$$



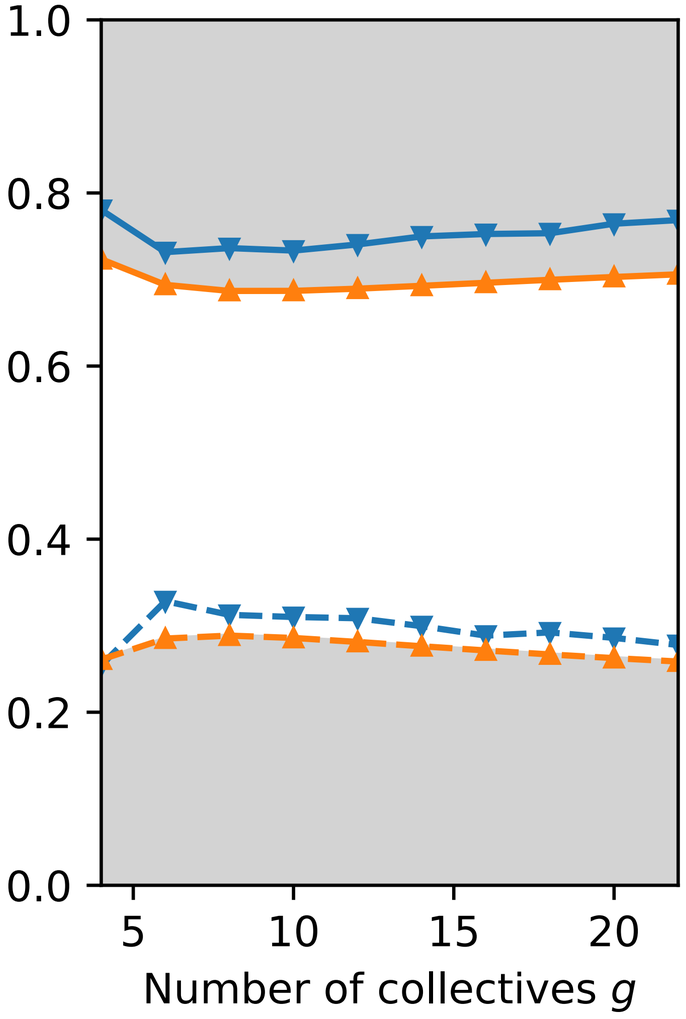
Newborns' total cell number: N_0

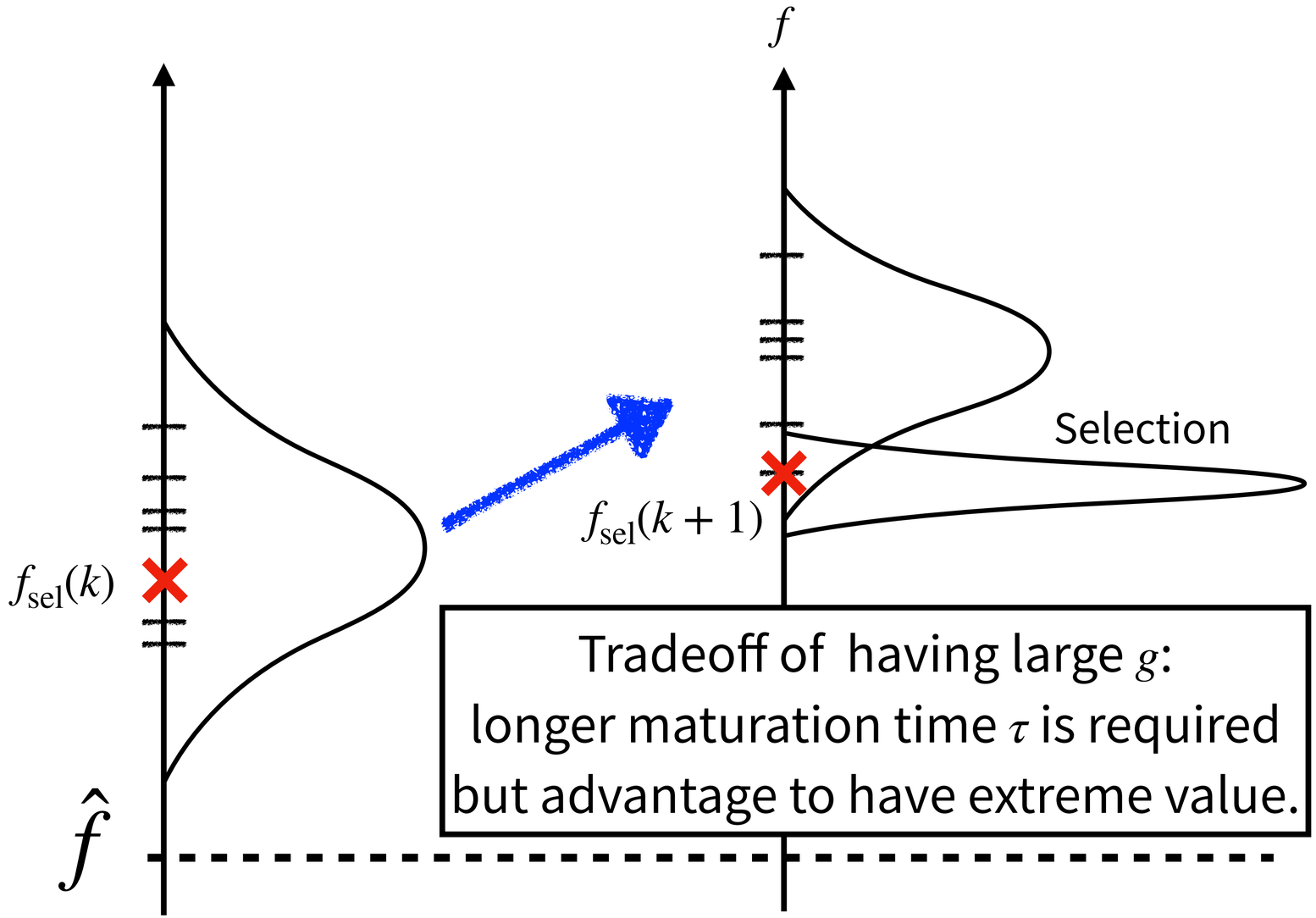
Initial frequency: f_0

$g = 10,\ \tau = \log(g+1)/r,\ N_o = 1000,\ r = 0.5,\ s = 0.03,\ \ \mu = 10^{-4},$

The number of collectives g

An optimal number of collectives for successful selection exists.

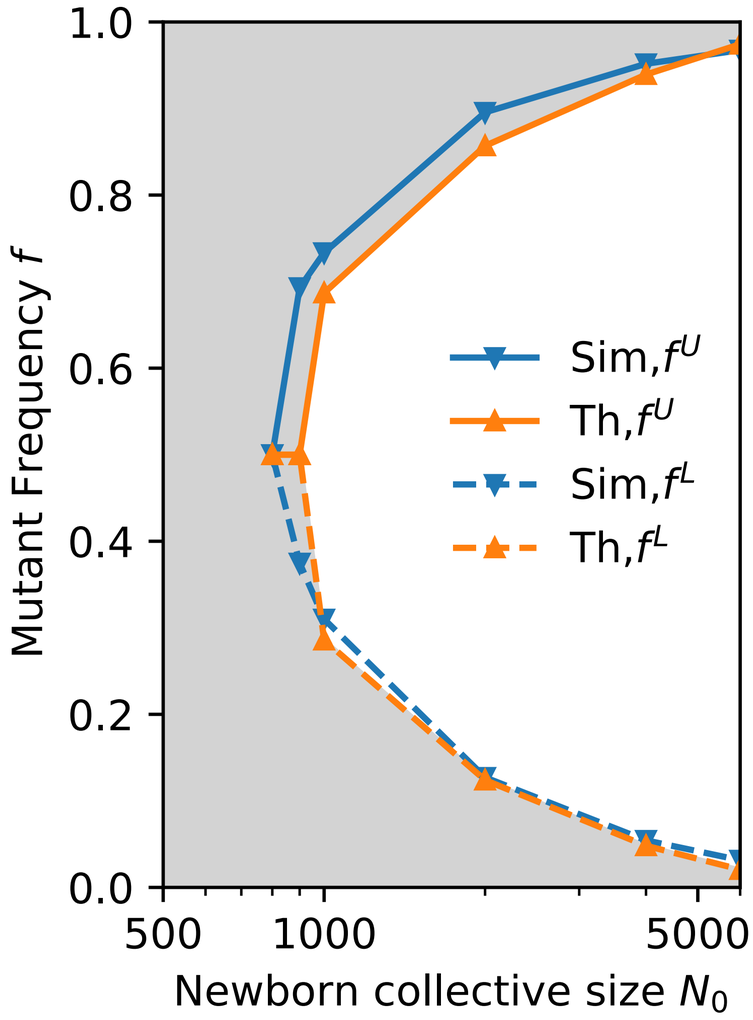


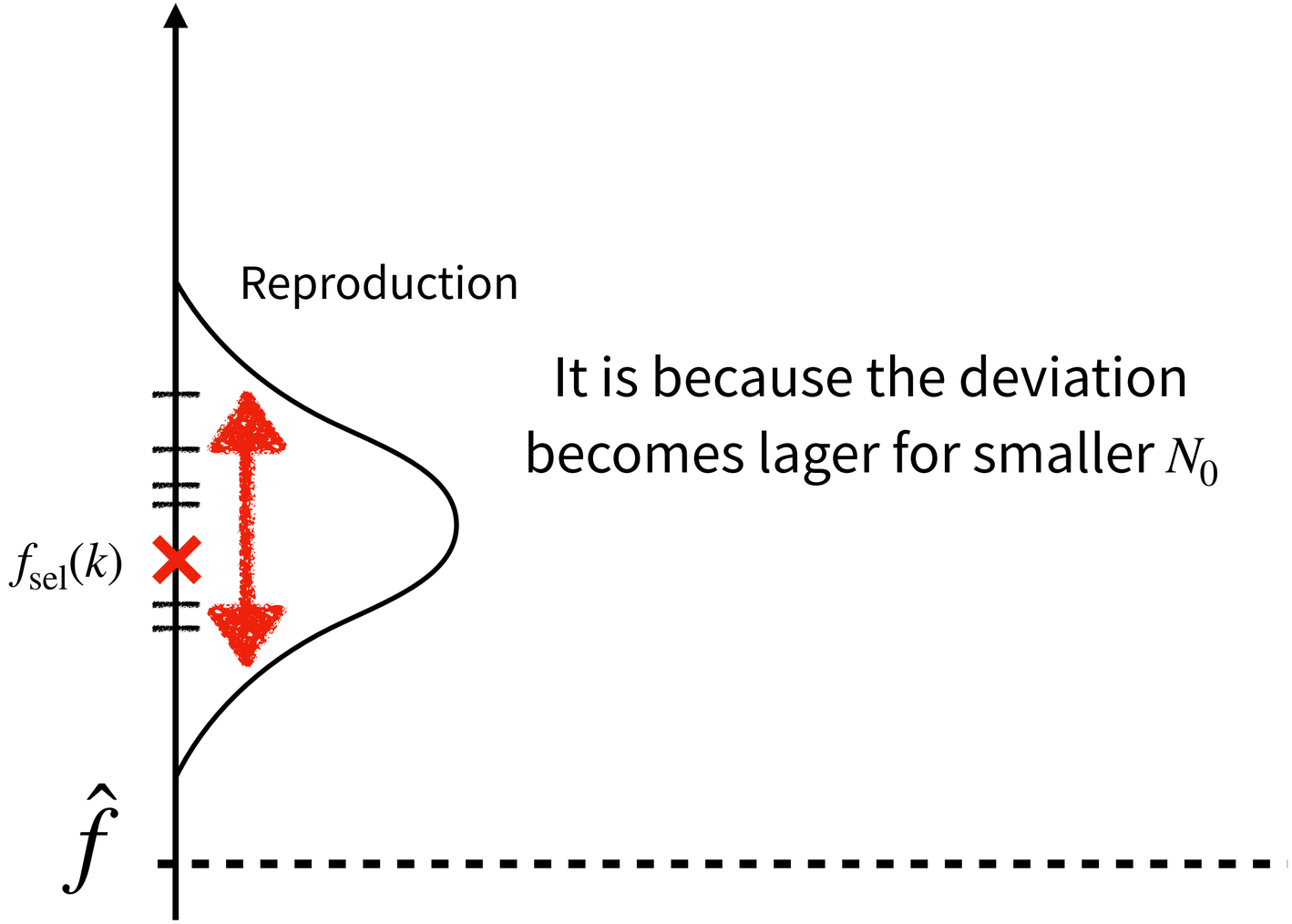


$g = 10,\ \tau = \log(g+1)/r,\ N_o = 1000,\ r = 0.5,\ s = 0.03,\ \ \mu = 10^{-4},$

Newborn cell number N_0

Smaller N_0 is advantageous to select the target composition.

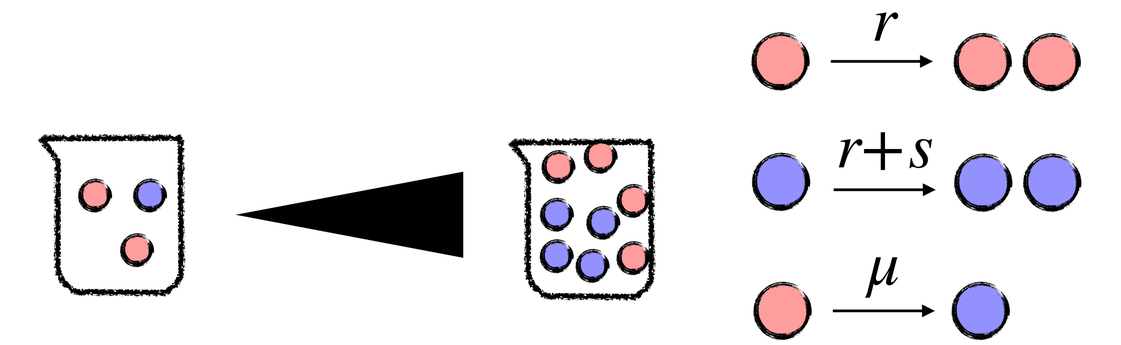


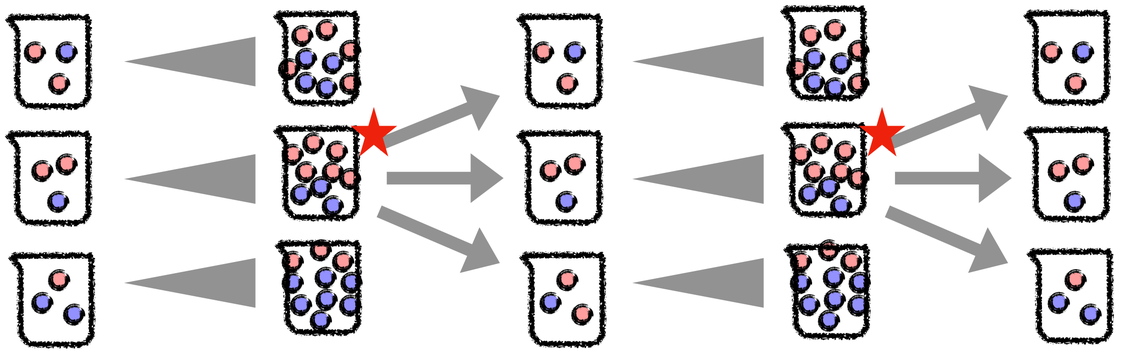


It is because the deviation becomes lager for smaller N_0

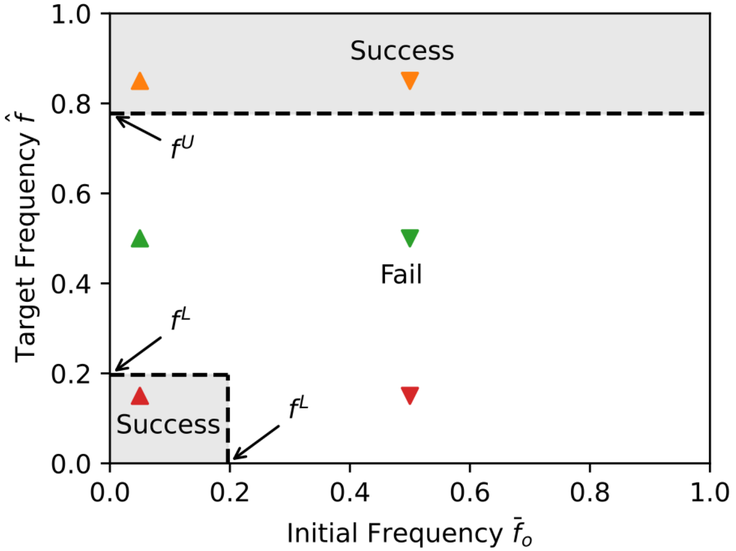
Summary

1. Proposing a simple model to understand the condition of successful artificial selection

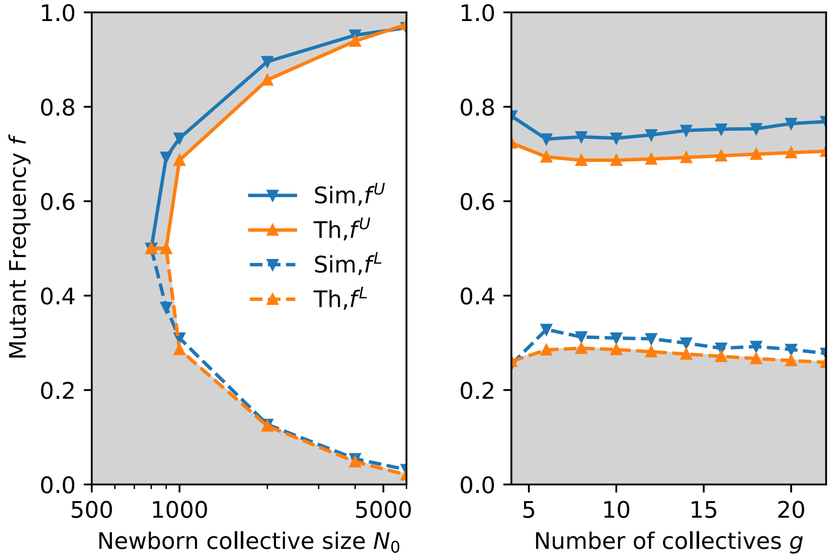




2. Analyzing conditions for successful selection



3. Finding the optimal experimental setup



Artificial selection on ‘microbial collectives’

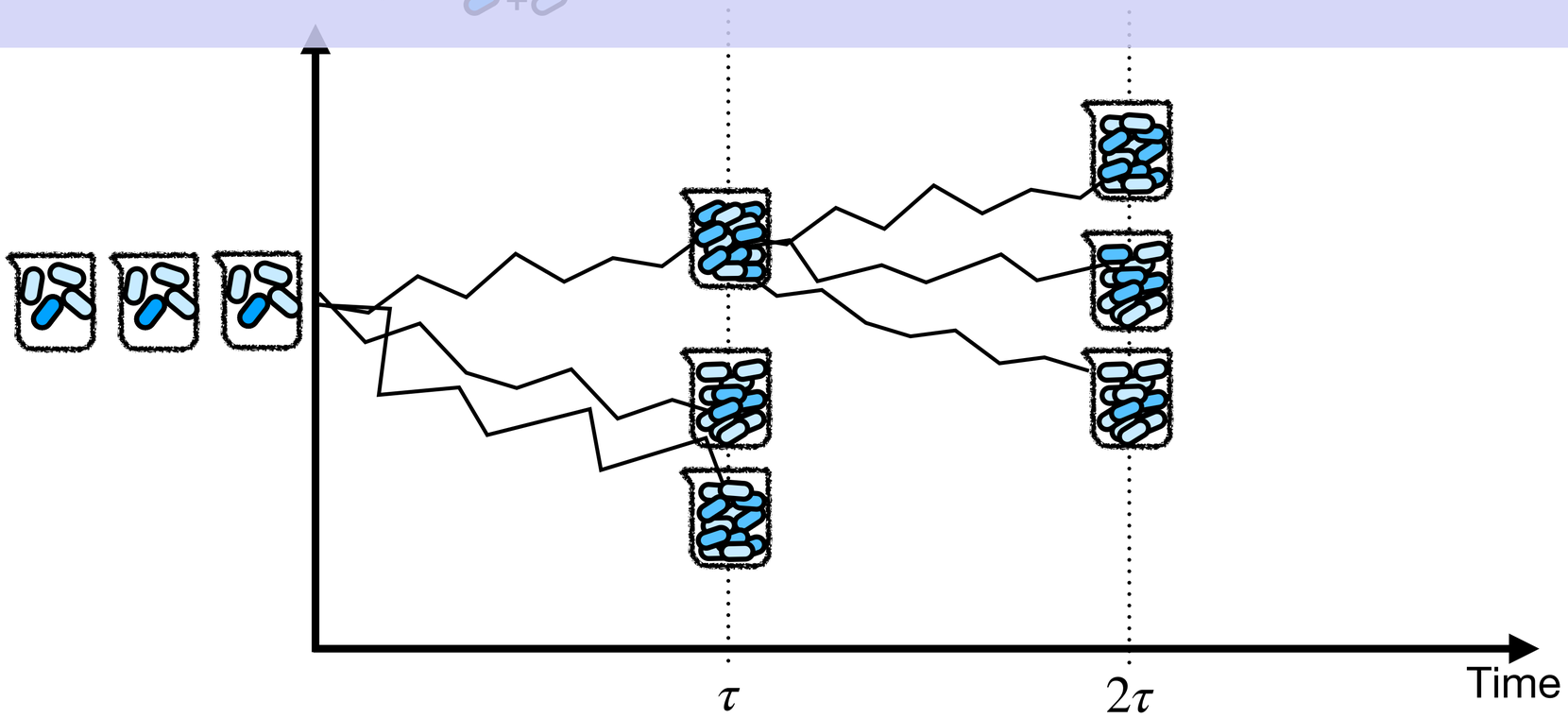
wild-type microbe , grows faster

produces a public good

Frequency x

Frequency x

Group resetting problem!



Group search and resetting in nature

Resetting position is determined by the ***collective information*** of agents!

Group hunting of wolves



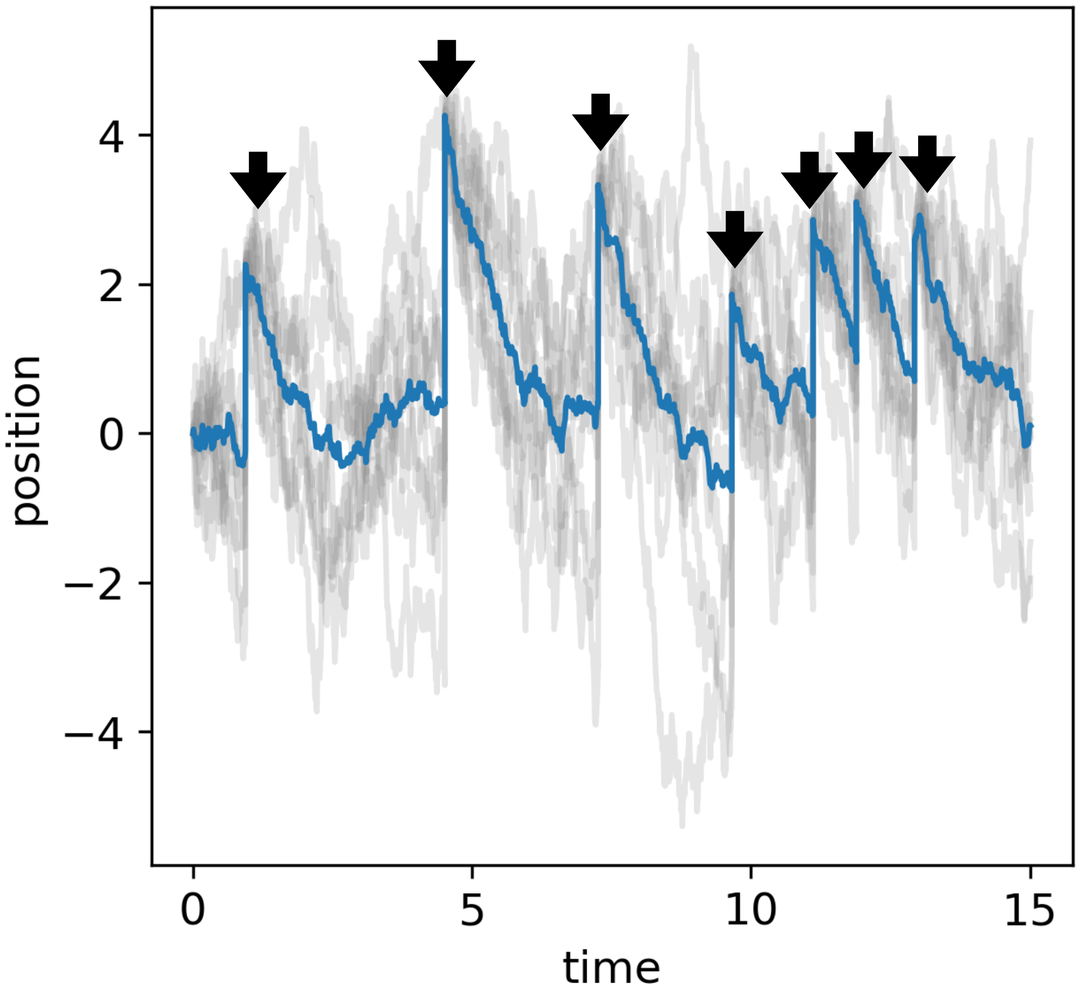
Search prey individually, and reset position to prey when it is discovered

Group foraging of ants



Search food individually, and reset position (or path) to the found food.

Group resetting process



Random walks

Multiple random walkers independently walk

Resetting

Reset all positions to the present optimum

Fokker-Planck equation for the mean position

Fokker-Planck equation

Resetting position distribution $R(\zeta, t)$

$$\frac{\partial}{\partial t}P(\zeta,t) = \underbrace{-\frac{\partial}{\partial \zeta}\left[A(\zeta)P(\zeta,t)\right]}_{\text{drift}} + \underbrace{\frac{\partial^2}{\partial \zeta^2}\left[\frac{D}{n}P(\zeta,t)\right]}_{\text{diffusion}} \underbrace{- rP(\zeta,t) + rR(\zeta,t)}_{\text{resetting}}$$

$\zeta = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$: **mean position** of individual walkers

$P(\zeta, t)$: Probability of finding the mean position at ζ at time t

Fokker-Planck equation for the mean position

Resetting position distribution $R(\zeta, t)$

$$R(\zeta,t) = \underbrace{\left(1 - \int_0^t \phi(\tau)d\tau\right) K_n(\zeta\,|\,\zeta_0;t)}_{\text{The first resetting occurs at } t} + \underbrace{\int_0^t d\tau\phi(\tau)\int_{-\infty}^{\infty} d\zeta' K_n(\zeta\,|\,\zeta';\tau)R(\zeta',t-\tau)}_{\text{The last resetting happened at } t-\tau \text{ to } \zeta'}$$

$\phi(\tau)$: **Waiting-time distribution** between consecutive resetting events.

$K_n(\zeta\,|\,\zeta';\tau)$: **Distribution of reset position** $\zeta = X$ after time τ since the last reset to $\zeta' = X'$.

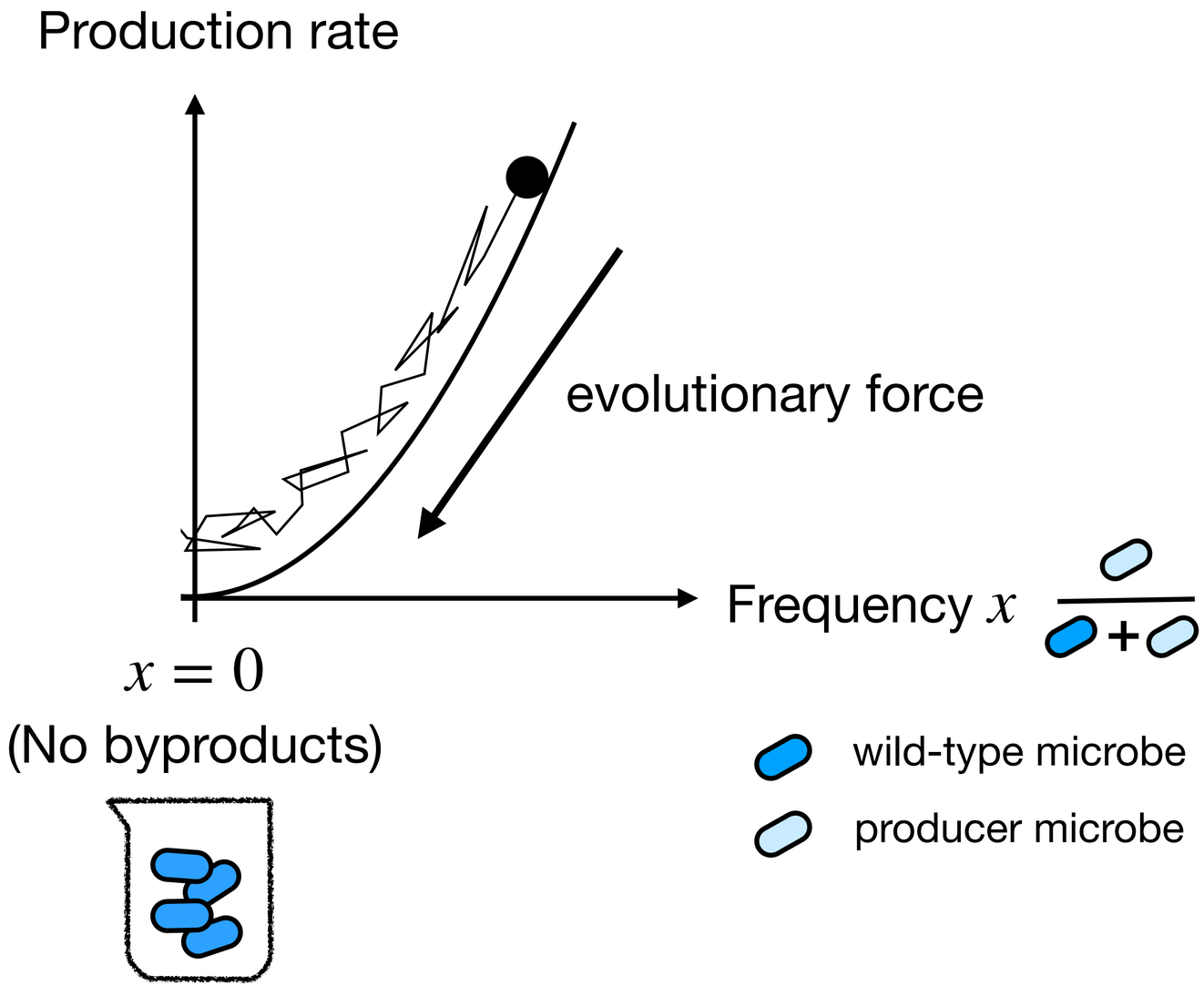
Fokker-Planck equation for the mean position

$$\frac{\partial}{\partial t}P(\zeta,t) = -\frac{\partial}{\partial \zeta}\left[A(\zeta)P(\zeta,t)\right] + \frac{\partial^2}{\partial \zeta^2}\left[\frac{D}{n}P(\zeta,t)\right] - rP(\zeta,t) + rR(\zeta,t)$$

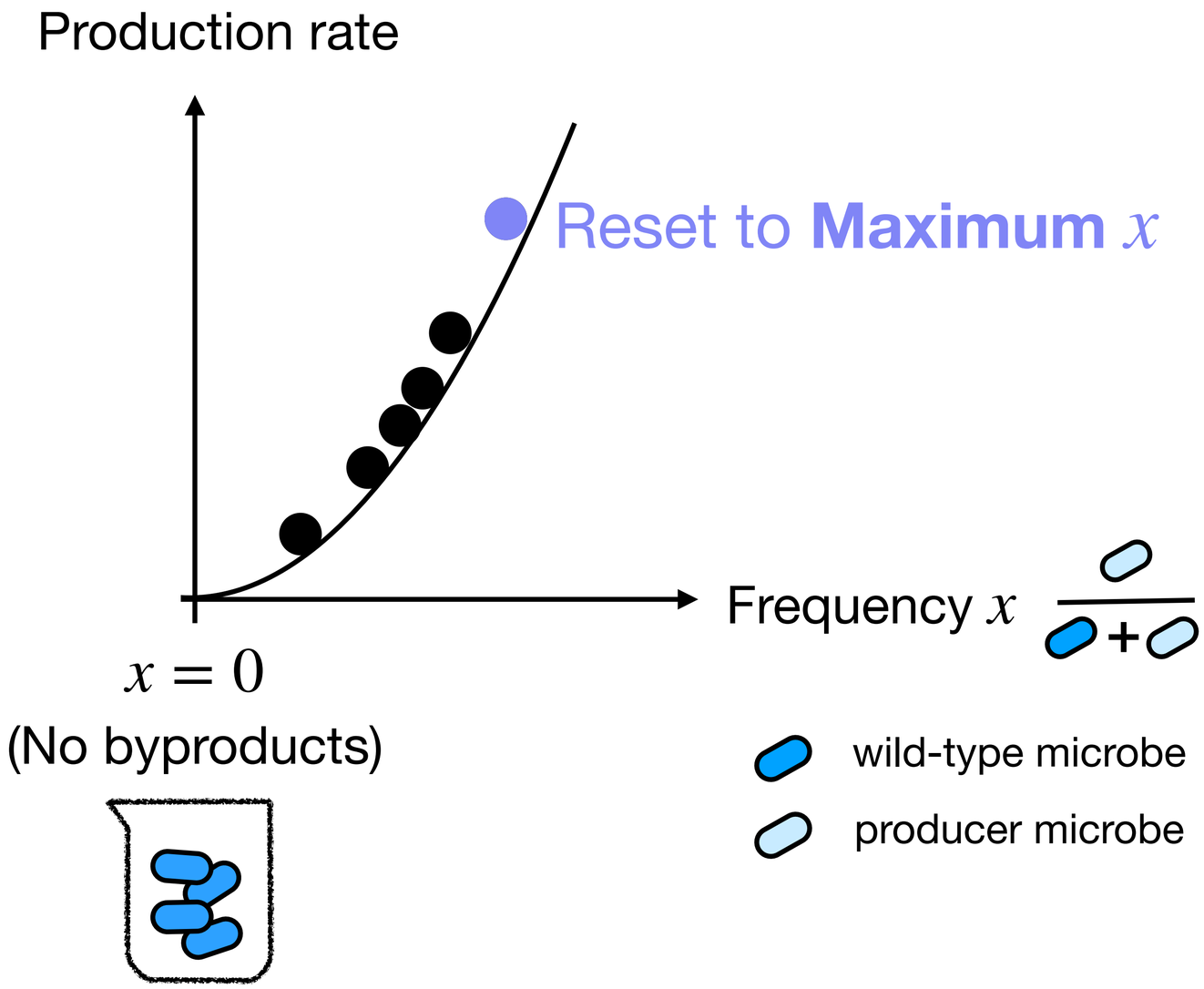
$$R(\zeta,t) = \left(1 - \int_0^t \phi(\tau)d\tau\right) K_n(\zeta\,|\,\zeta_0;t) + \int_0^t d\tau\phi(\tau)\int_{-\infty}^{\infty} d\zeta' K_n(\zeta\,|\,\zeta';\tau)R(\zeta',t-\tau)$$

For fixed reseting point ζ_0, $K_n(\zeta\,|\,\zeta';\tau) = \delta(\zeta-\zeta_0)$, and thus $R(\zeta,t) = \delta(\zeta-\zeta_0)$, reproducing the well-known results.

Application for avoiding problem

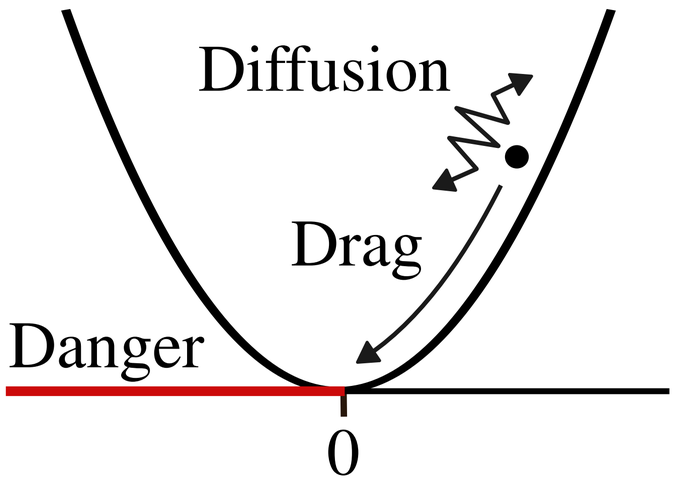


Application for avoiding problem



Application for avoiding problem

Ornstein-**U**hlenbeck process

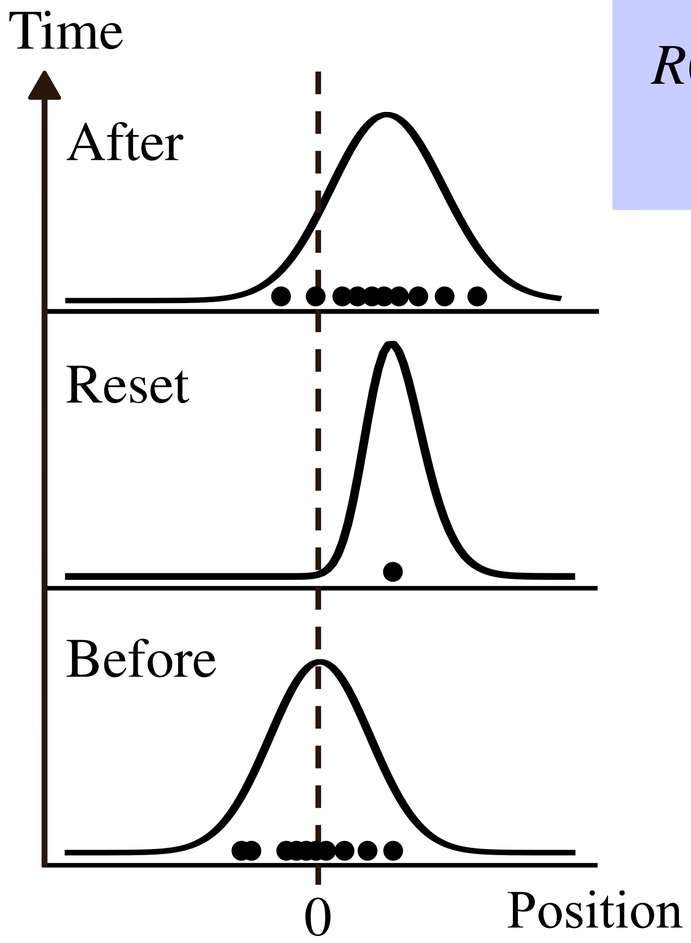


n Brownian particles diffuse in potential $V(x) = \frac{1}{2}kx^2$

$$\frac{\partial}{\partial t}P(\zeta,t) = \frac{\partial}{\partial \zeta}\left[k\zeta P(x,t)\right] + \frac{D}{n}\frac{\partial^2}{\partial \zeta^2}\left[P(\zeta,t)\right] - rP(\zeta,t) + rR(\zeta,t)$$

Application for avoiding problem

Reset to **Maximum** x



$$\frac{\partial}{\partial t}P(\zeta,t)=\frac{\partial}{\partial \zeta}\left[k\zeta P(x,t)\right]+\frac{D}{n}\frac{\partial^2}{\partial \zeta^2}\left[P(\zeta,t)\right]-rP(\zeta,t)+rR(\zeta,t)$$

$$R(\zeta,t)=\left(1-\int_0^t \phi(\tau)d\tau\right)K_n(\zeta\,|\,\zeta_0;t)+\int_0^t d\tau\phi(\tau)\int_{-\infty}^{\infty}d\zeta' K_n(\zeta\,|\,\zeta';\tau)R(\zeta',t-\tau)$$

Poisson process with resetting rate r, $\phi(t)=re^{-rt}$

By Fisher–Tippett–Gnedenko theorem

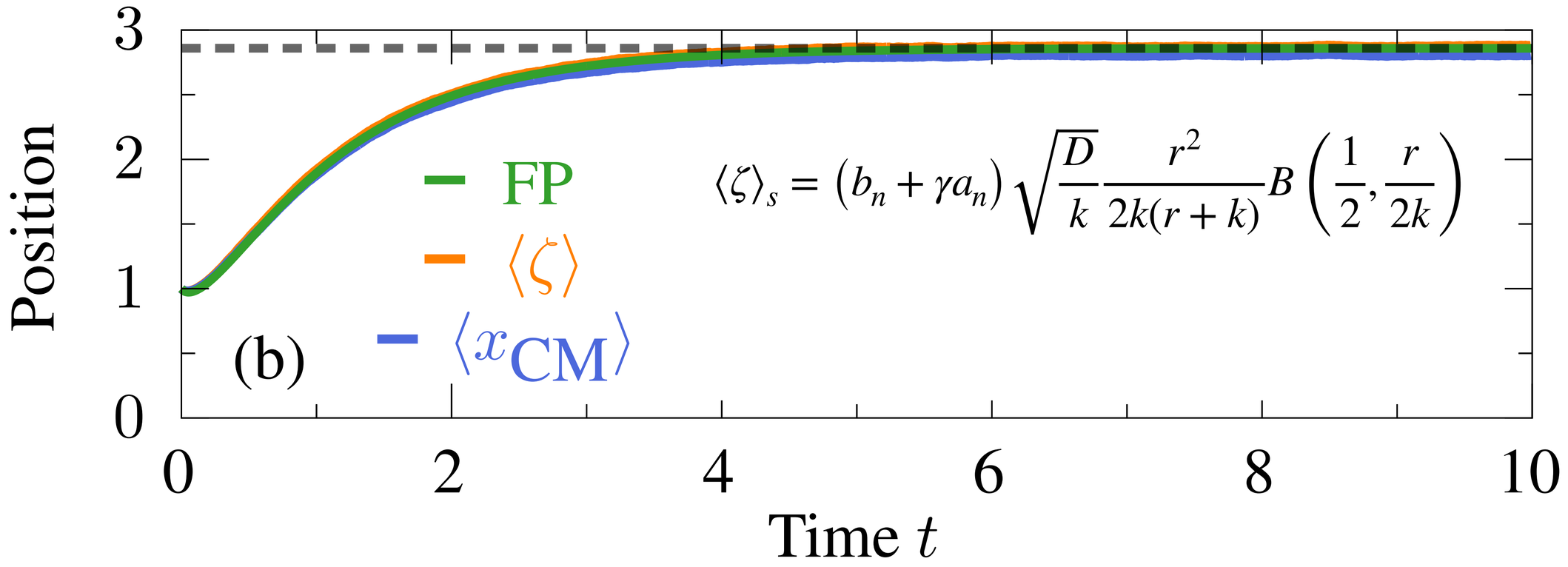
$$K_n(\zeta\,|\,\zeta';\tau)\sim \text{Gumbel}(\mu+b_n\sigma,a_n\sigma)$$

$$b_n=\frac{1}{a_n}=F^{-1}(1-\frac{1}{n})$$

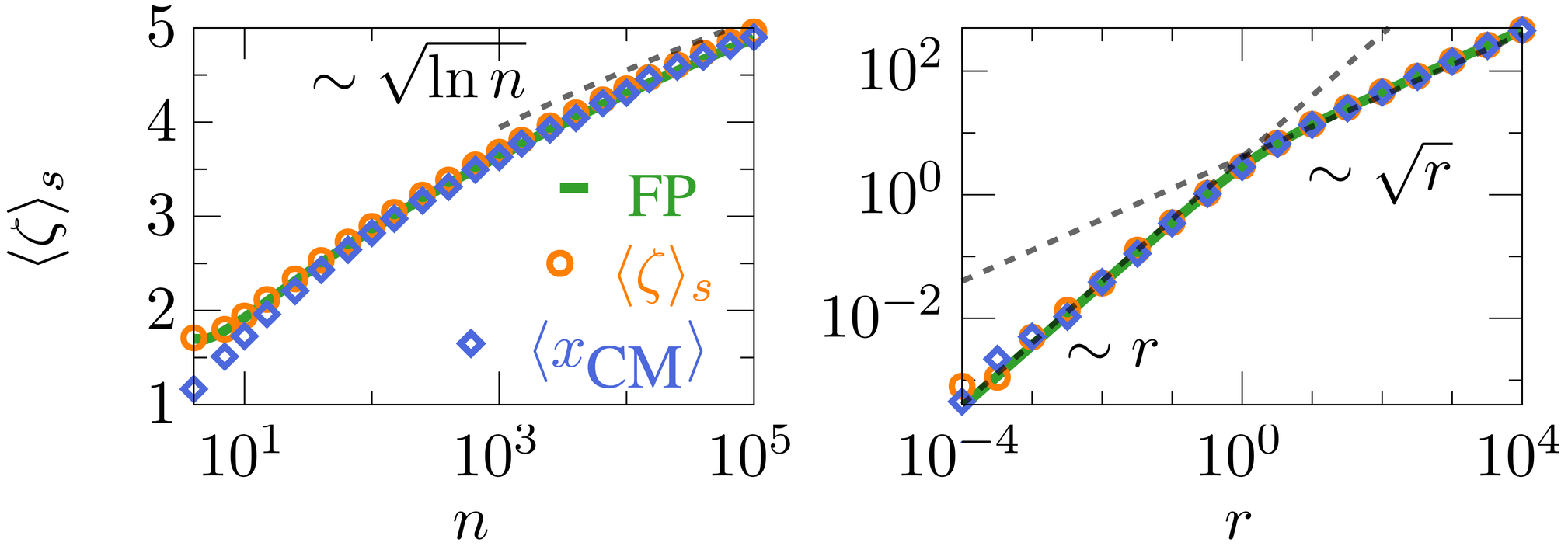
$F^{-1}(y)$: inverse CDF of $\mathcal{N}(0,1)$

Comparison with simulations

$(n, r, k, D) = (100,1,1,2)$



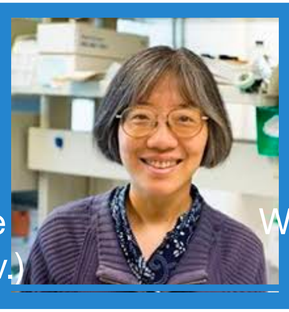
$\langle \zeta \rangle$ at the stationary state



A larger number of walkers and frequent resetting increase the mean position in the stationary state. increasing r is more effective than increasing n.

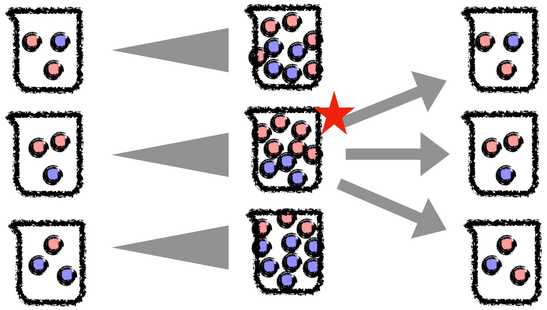


Juhee Lee
(Umeå Univ.)

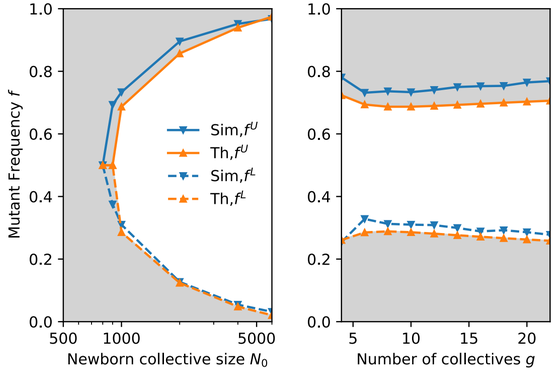


Wenying Shou
(UCL)

1. We propose a **simple model** to understand the **conditions for successful artificial selection**.



2. We analyze these conditions, and identify the **optimal experimental setup**.



Umeå
Univ.



Juhee Lee

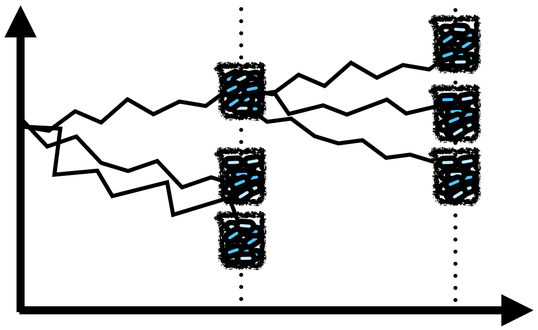


Seong-Gyu Yang



Ludvig Lizana

1. We present a **general framework for group resetting**, focusing on the walkers' mean-position dynamics.



2. We find **strategies that keep the system away from unwanted states**.

