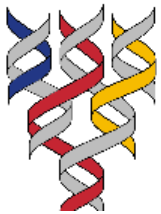


SFB 680
Molecular Basis of
Evolutionary Innovations

Dynamics of adaptation in asexual and sexual populations

Joachim Krug
Institute of Theoretical Physics, University of Cologne

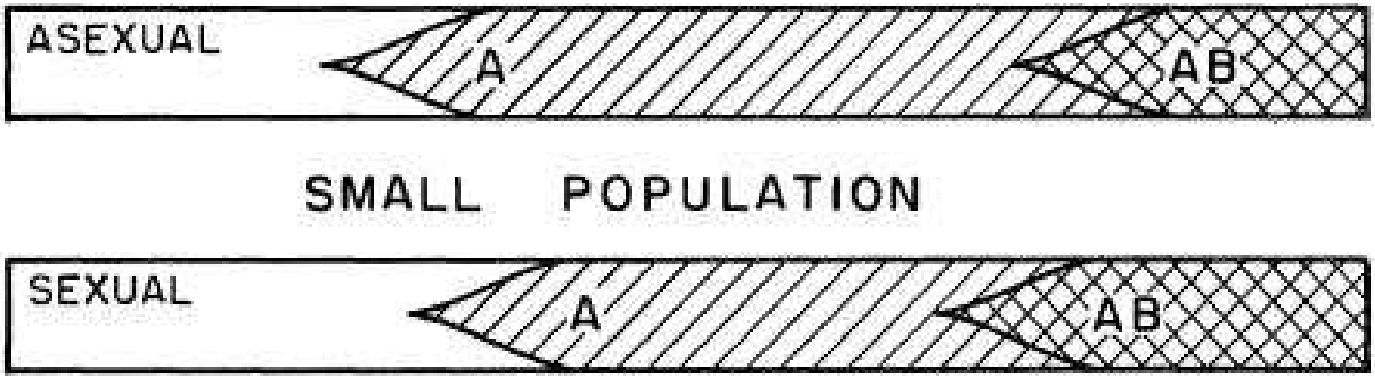
- Clonal interference in asexual populations
- Adaptation in the house-of-cards model
- Effects of recombination in an empirical fitness landscape

Joint work with Su-Chan Park and Arjan de Visser

The Muller-Fisher mechanism for the advantage of sex

J.F. Crow & M. Kimura, Am. Nat. **99**, 439 (1965)

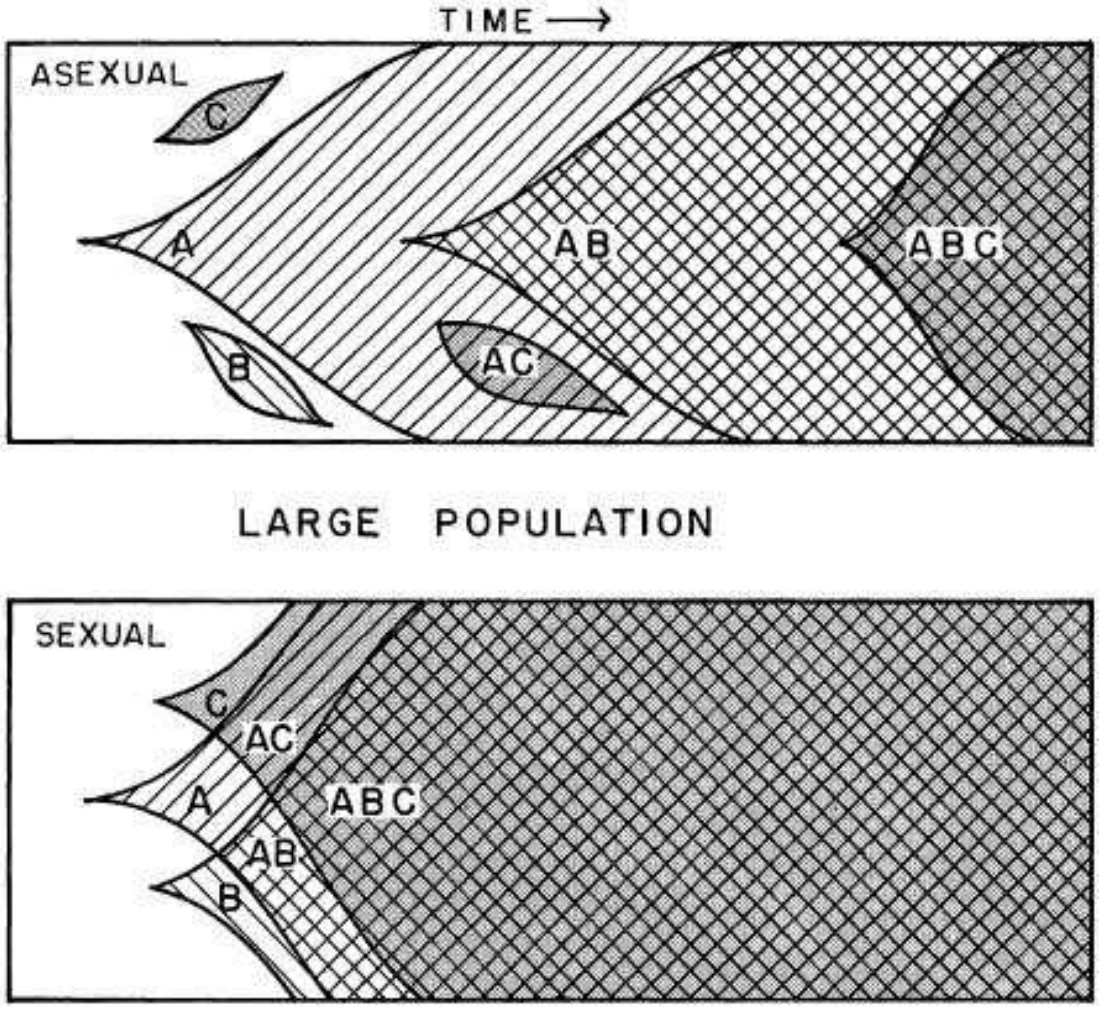
- Dynamics of an adapting population:



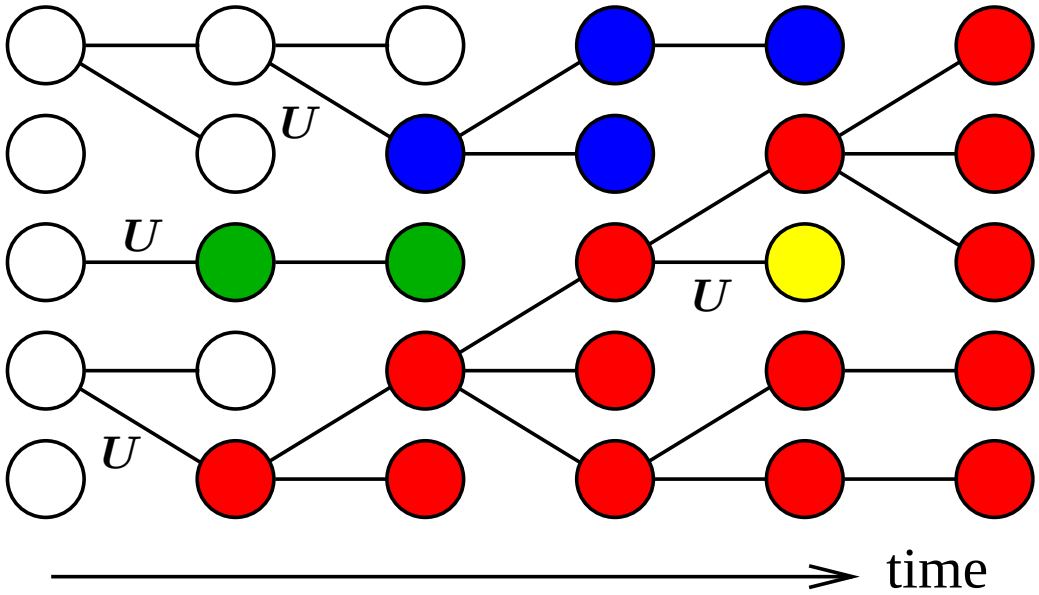
periodic selection

The Muller-Fisher mechanism for the advantage of sex

- **Clonal interference** slows down the adaptation of asexual populations



The Wright-Fisher model



- Constant population size N, discrete non-overlapping generations
- Each individual chooses an ancestor from the preceding generation
- Individual i is chosen with probability $\sim w_i$ Wrightian fitness
- Mutations occur with probability U per individual and generation

Fixation

- When a single mutant of fitness w' is introduced into a monomorphic population of fitness w, the outcome for $t \to \infty$ is either fixation (all w') or loss of the mutation (all w)

- Fixation probability for the Wright-Fisher model (Kimura, 1962)

$$\pi_N(s) \approx \frac{1-e^{-2s}}{1-e^{-2Ns}}, \quad s = \frac{w'}{w} - 1 \qquad \text{selection coefficient}$$

- Under strong selection: ($N|s| \gg 1$) deleterious mutations ($s < 0$) cannot fix, while beneficial mutations ($s > 0$) fix with probability

$$\pi(s) = 1 - e^{-2s} \approx 2s, \quad s \ll 1$$

- Mean time to fixation of a beneficial mutation:

$$t_{\text{fix}} \approx \frac{\ln N}{s}$$

Mutation and fitness model

- Infinite sites approximation:
  Each mutation creates a new genotype, no recurrent mutations

- Multiplicative model: Fitness of offspring w' related to parental fitness w by

$$w \;\rightarrow\; w' = w(1+s)$$

  with selection coefficient s chosen randomly from a distribution $P(s)$

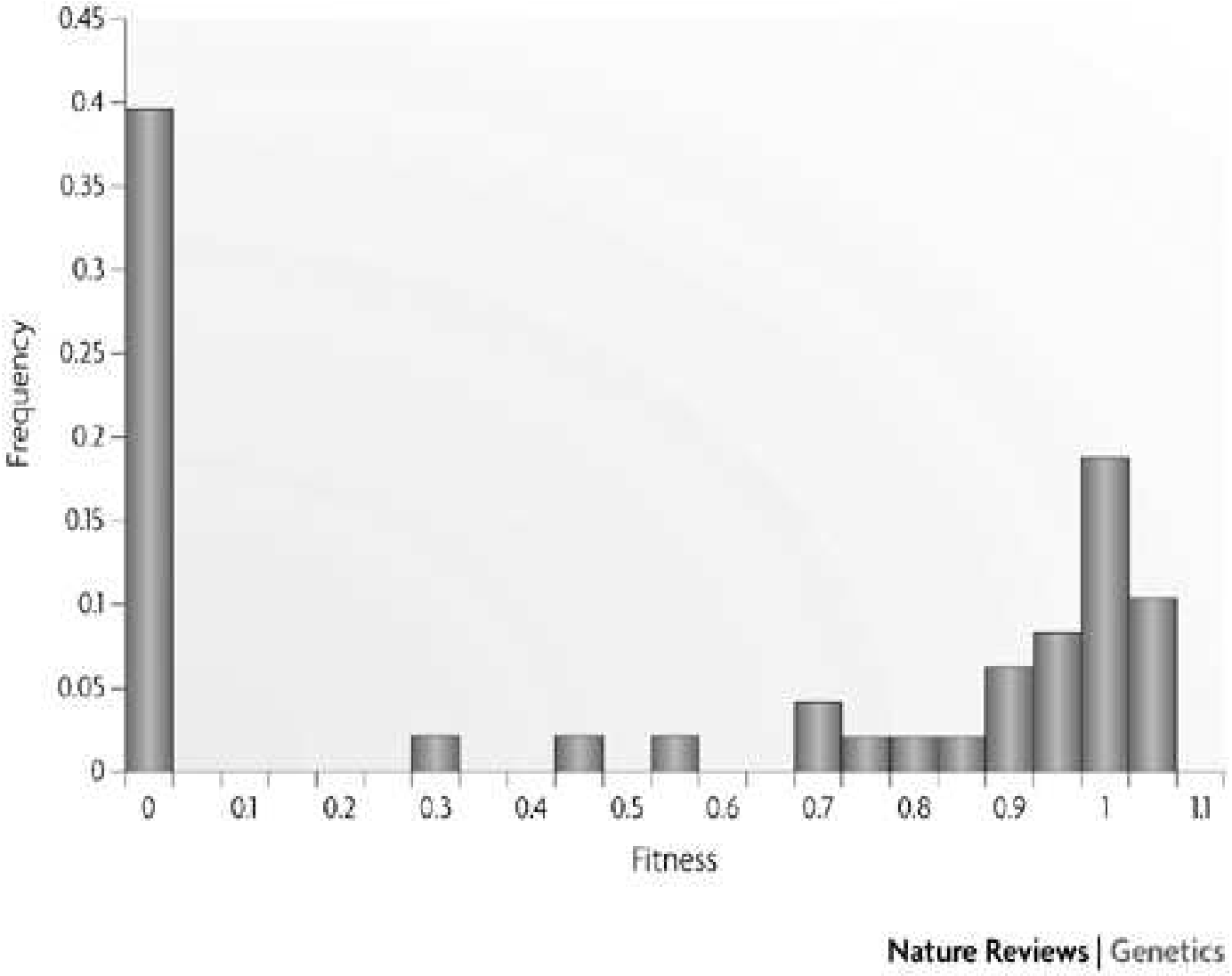
- Extremal statistics arguments suggest that the distribution of selection coefficients for beneficial mutations is exponential:

H.A. Orr, Genetics **163**, 1519 (2003)

$$P_b(s) = s_b^{-1} e^{-s/s_b}, \quad s > 0$$

- Beneficial mutations occur with probability U_b

An empirical fitness distribution (VS virus)

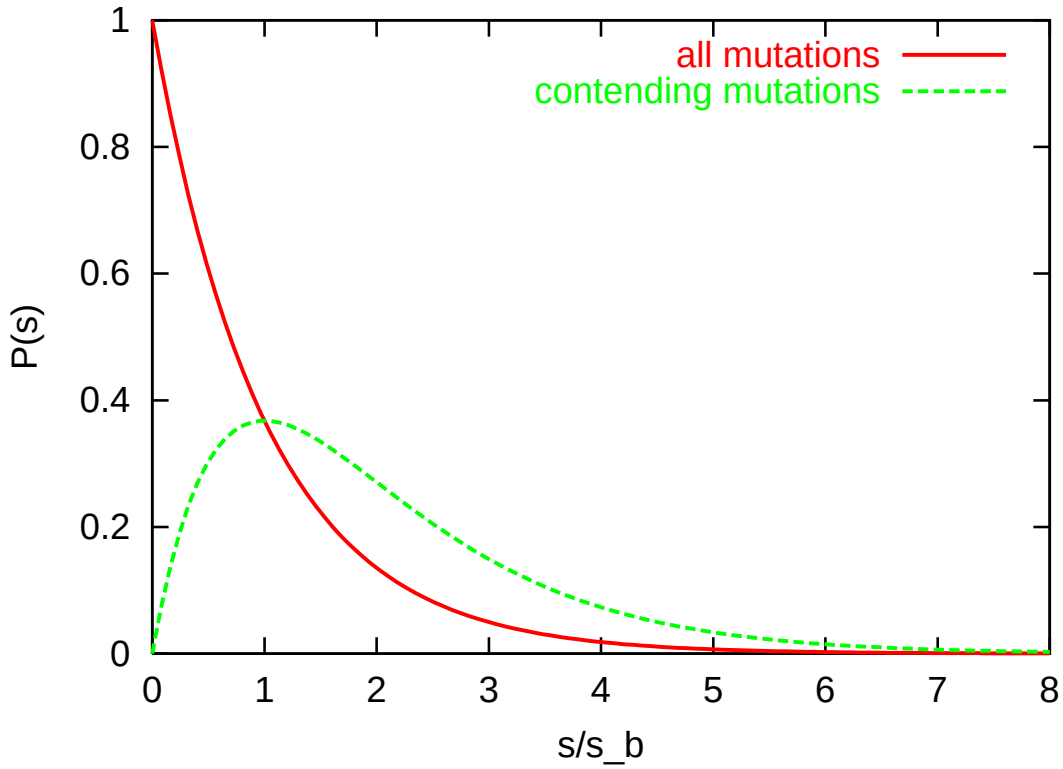


from: A. Eyre-Walker & P.D. Keightley, Nat. Rev. Gen. **8**, 610 (2007)

Contending mutations

- Beneficial mutations are most likely to get lost by genetic drift in the early (stochastic) regime of the fixation process

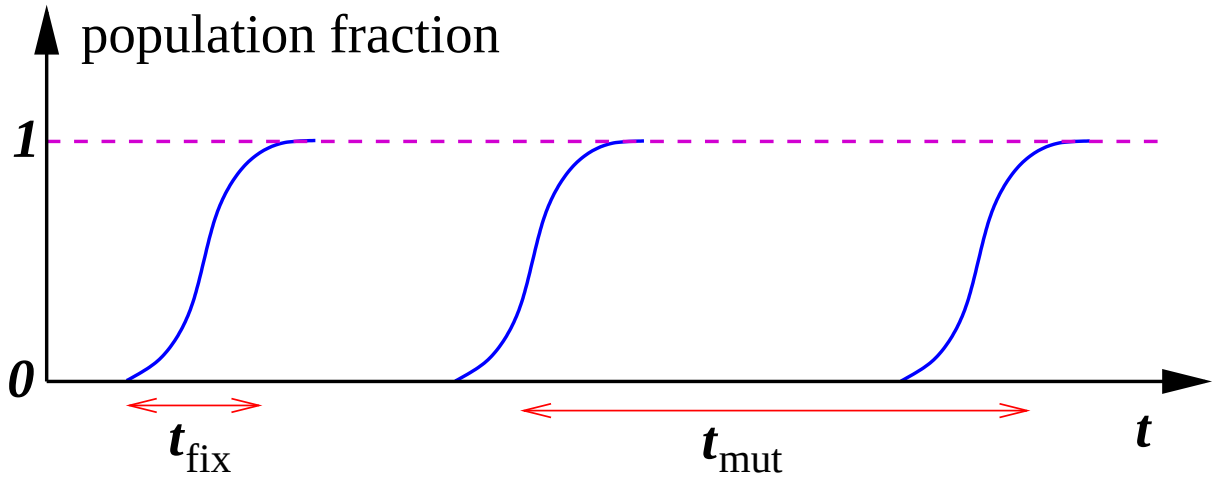
  $\Rightarrow$ mutations become contenders with probability $\pi(s) \approx 2s$



- Probability distribution of contending mutations: $P_c(s) \sim sP_b(s) \sim se^{-s/s_b}$

Periodic selection vs. clonal interference

- Contending mutations arise at rate $2s_bNU_b = 1/t_{\rm mut}$
- Periodic selection requires $t_{\rm fix} \ll t_{\rm mut}$

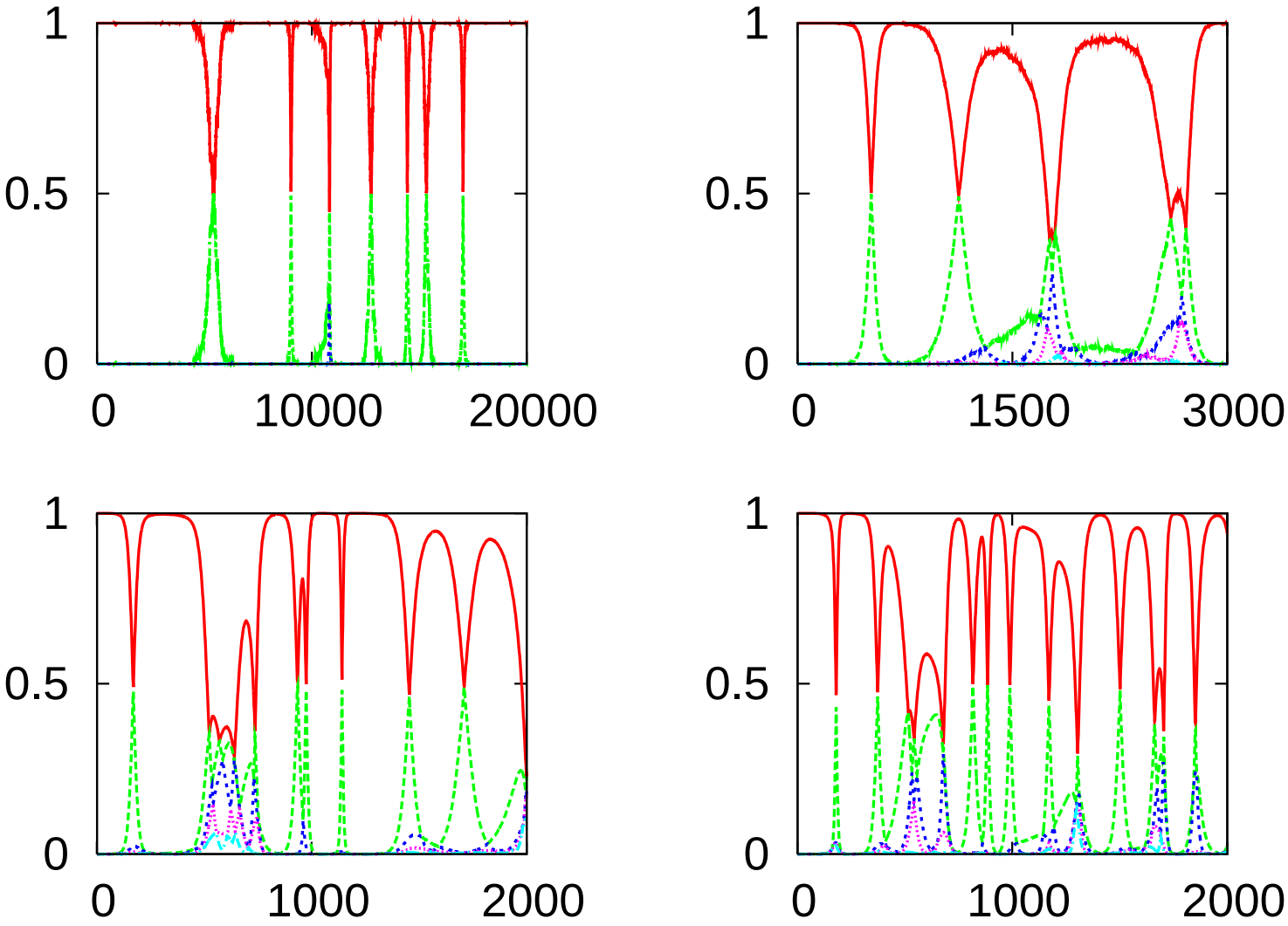


- In the periodic selection regime every contending mutation is fixed

  $\Rightarrow$ rate of adaptation $R = 2s_b/t_{\rm mut} = 4s_b^2NU_b$

- Beneficial mutations interfere when $t_{\rm fix} \gg t_{\rm mut}$ or $2NU_b \ln N \gg 1$
- Clonal interference is inevitable for large N [provided U_b is constant!]

Wright-Fisher dynamics for $U_b = 10^{-6}, s_b = 0.02$



$N = 10^4, 10^5, 10^6, 10^7$

The Gerrish-Lenski theory of clonal interference

P.J. Gerrish, R.E. Lenski, Genetica **102/103**, 127 (1998)

- **Key idea:** A contending mutation s survives clonal competition if no superior mutation $s' > s$ arises during the time to fixation of s.

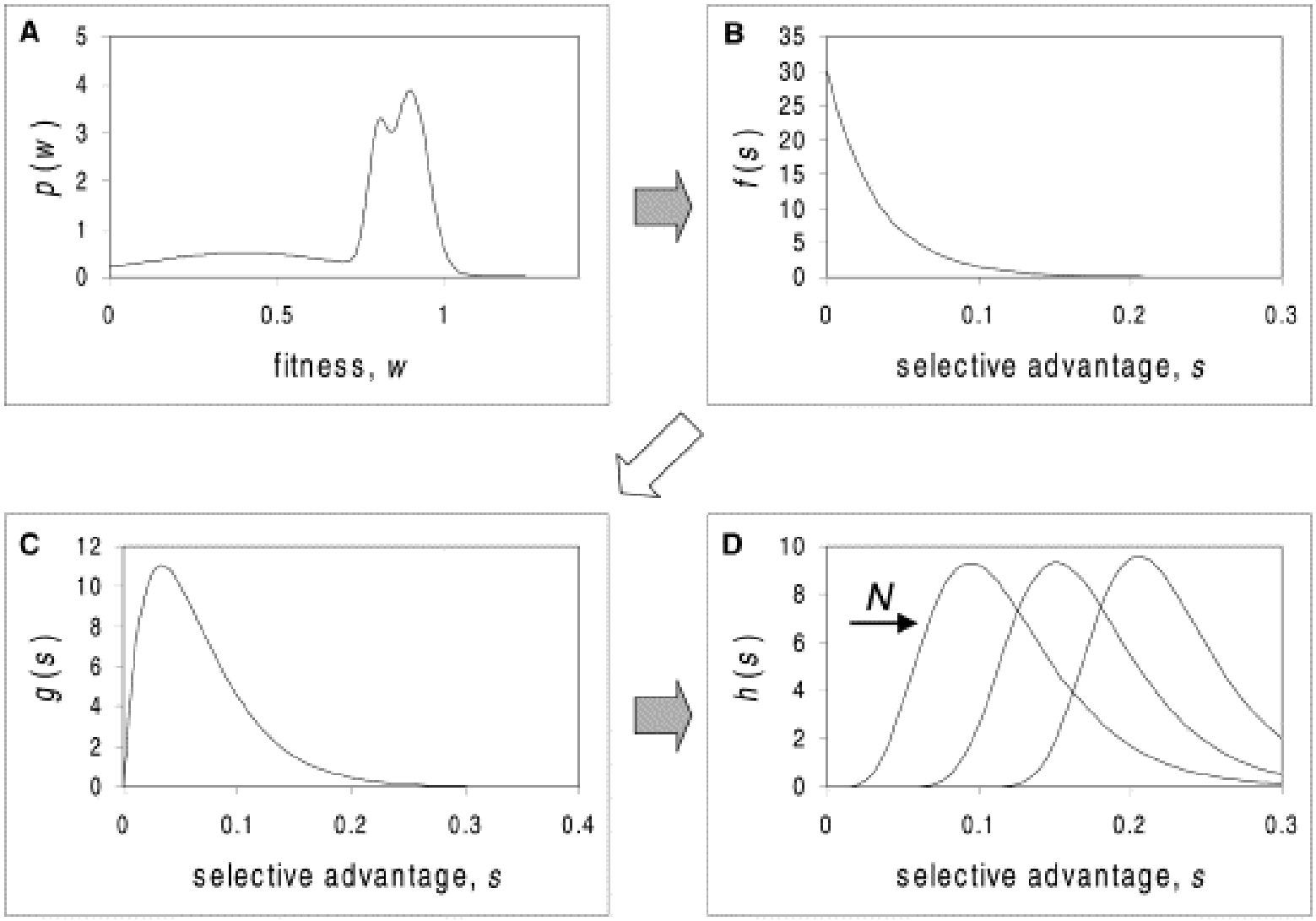
- The survival probability is $\exp[-\lambda(s)]$ with

$$\lambda(s) = NU_b t_{\text{fix}} \int_s^\infty ds'\, P_c(s') = \frac{N \ln NU_b}{s} \int_s^\infty ds'\, \pi(s') s_b^{-1} \exp[-s'/s_b]$$

- GL theory does not (explicitly) account for the complex interaction of different clones. In particular, the possibility of beneficial mutations arising within a growing clone (multiple mutations) is ignored.

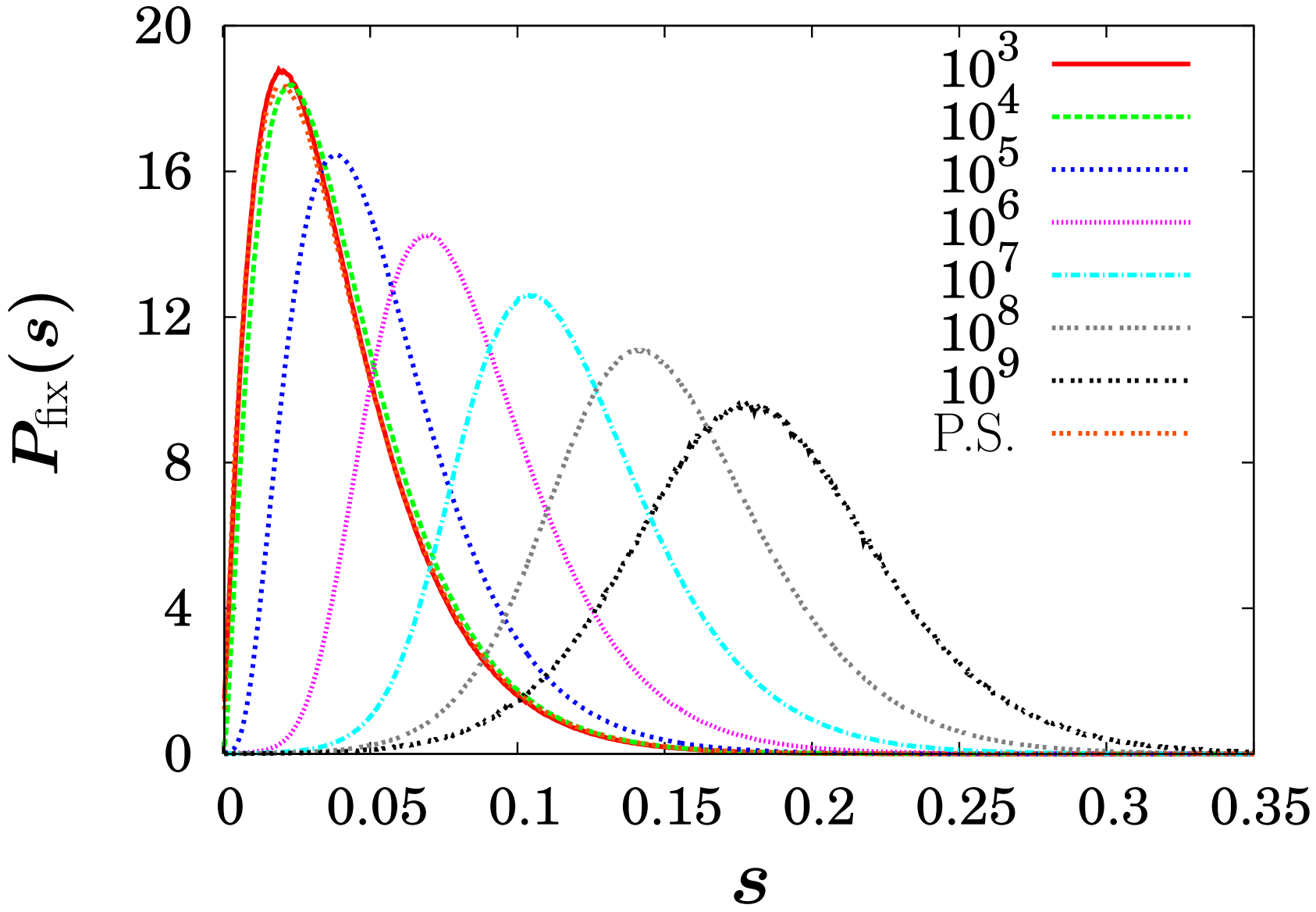
- **Qualitative predictions:**
  Clonal interference reduces the rate of substition $E[k]$ but increases the mean selection coefficient of fixed mutations $E[s]$.

Summary so far:



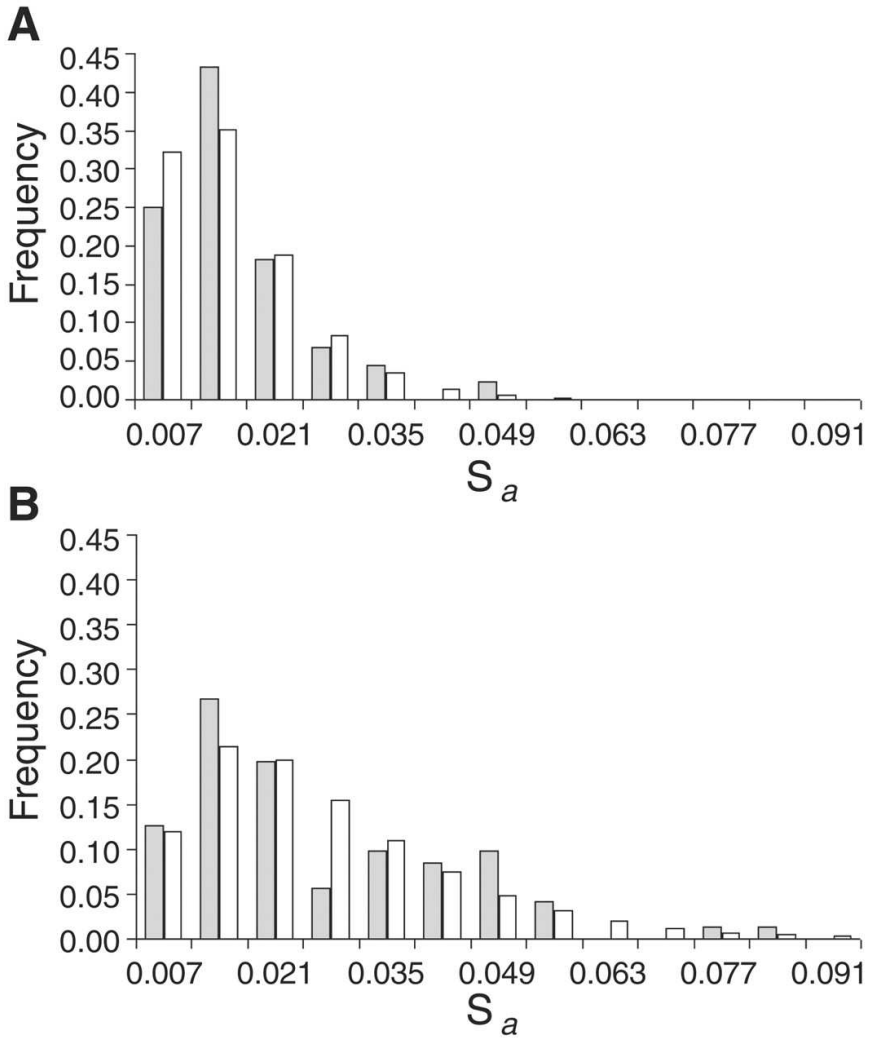
D.E. Rozen, J.A.G.M. de Visser, P.J. Gerrish, Curr. Biol. **12**, 1040 (2002)

Distribution of fixed mutations: Simulation



Distribution of mutational effects: Experiments (*E. coli*)

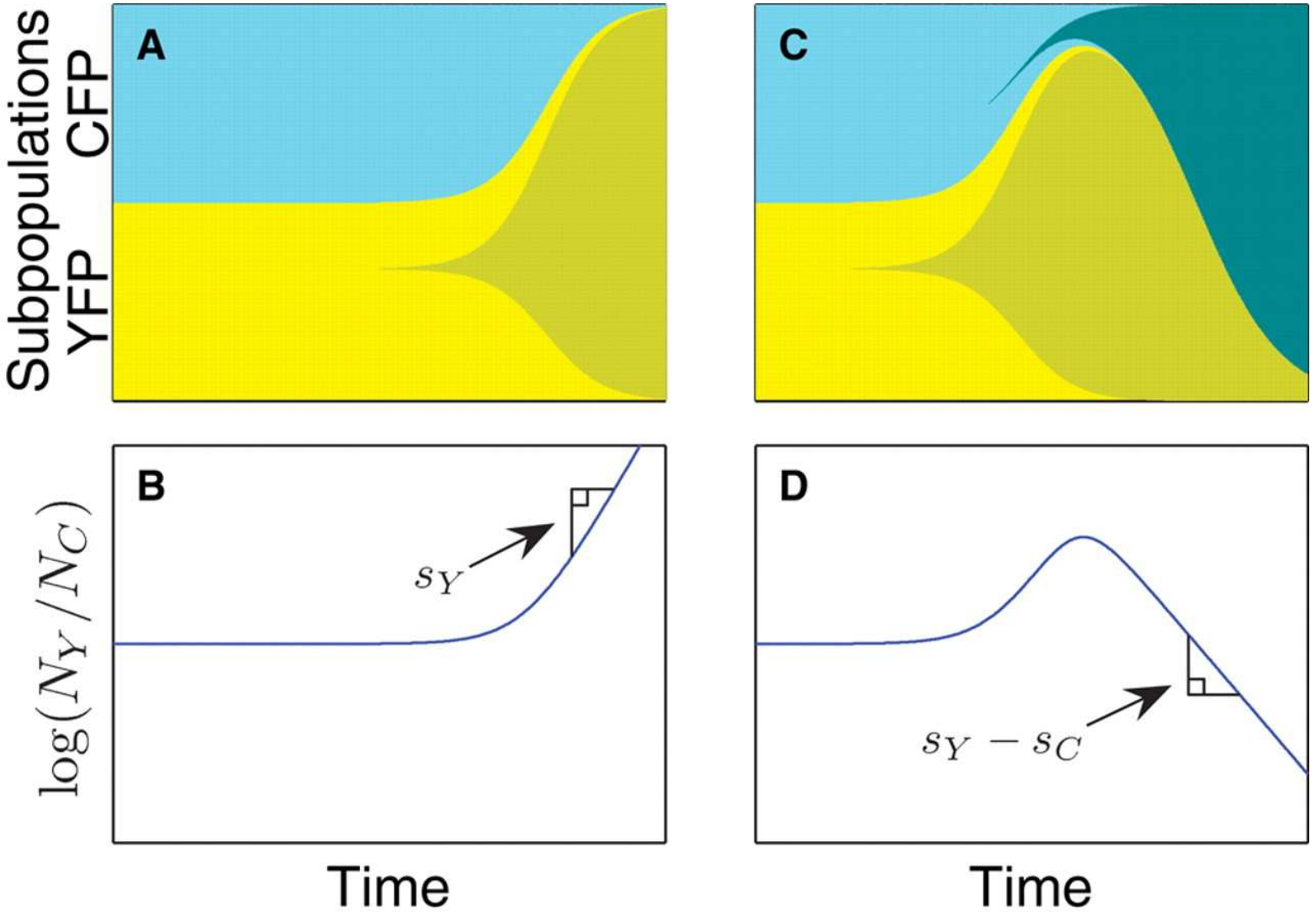
L. Perfeito et al., Science **317**, 813 (2007)



A: $N = 2 \times 10^4$ **B:** $N = 10^7$

Measuring selection coefficients in evolution experiments

M. Hegreness et al., Science **311**, 161 (2006)



GL-theory: Quantitative predictions for large N

C.O. Wilke, Genetics **167**, 2045 (2004); S.C. Park & JK, PNAS **104**, 18135 (2007)

- Rate of substitution: $\gamma \approx 0.577215\ldots$ Euler's constant

$$E[k] \approx \frac{s_b}{\ln N}[\ln(U_b N \ln N) + \gamma - 1] \to s_b$$

- Mean selection coefficient of fixed mutations:

$$E[s] \approx s_b[\ln(U_b N \ln N) + \gamma]$$

- Rate of adaptation: $E[w]$: mean population fitness

$$R = \lim_{t\to\infty} \frac{\ln E[w]}{t} \approx E[s]E[k] \to s_b^2 \ln(U_b N \ln N)$$

- Logarithmic dependence on the mutation supply NU_b

Extremal statistics estimates

- Probability to find a selection coefficient larger than S :

$$\mathrm{Prob}[s > S] = \int_s^\infty P_b(s)\, ds = e^{-s/s_b}$$

- The largest selection coefficient $s_{\max}$ in t generations is determined by

$$\mathrm{Prob}[s > s_{\max}] = \frac{1}{NU_b t} \;\Leftrightarrow\; s_{\max} = s_b \ln(NU_b t)$$

- Self-consistency requires that $t = t_{\mathrm{fix}}(s_{\max}) = \ln N / s_{\max}$

$$\Rightarrow\; s_{\max} = s_b \ln(NU_b \ln N / s_{\max}) \;\Rightarrow\; s_{\max} = E[s] \approx s_b \ln(NU_b \ln N)$$

- Rate of substitution:

$$E[k] = \frac{1}{t_{\mathrm{fix}}(s_{\max})} = \frac{s_{\max}}{\ln N} \approx \frac{s_b}{\ln N} \ln(NU_b \ln N)$$

Other mutation distributions

- Extremal statistics for $P_b(s) \sim \exp[-(s/s_b)^\beta]$ yields

$$s_{\max} \sim s_b (\ln N)^{1/\beta}, \quad E[k] \sim s_b (\ln N)^{1/\beta - 1}, \quad R \sim s_b^2 (\ln N)^{2/\beta - 1}$$

- Compare to behavior for mutations of single strength $P_b(s) = \delta(s - s_0)$:

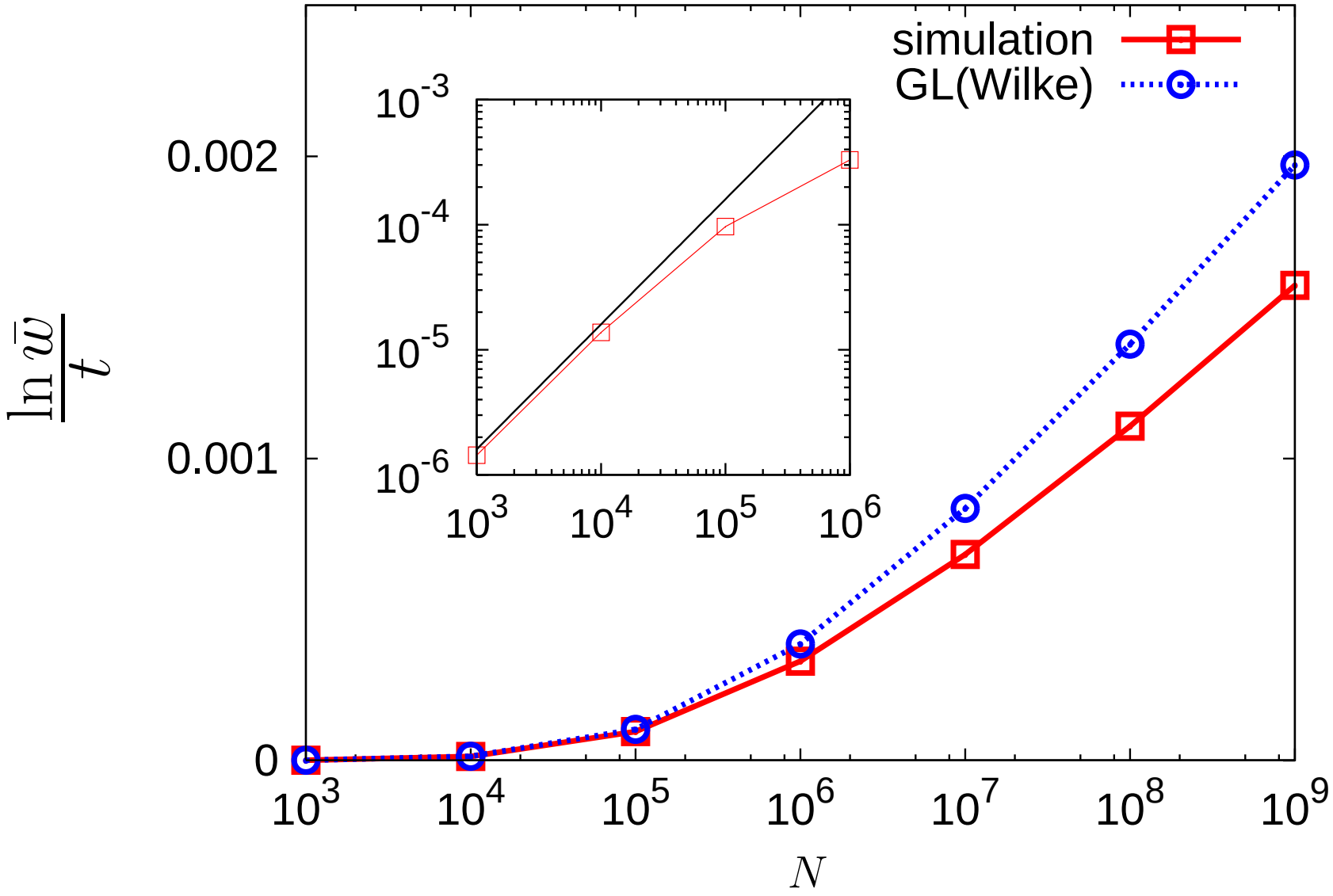
$$R \approx \frac{2 s_0^2 \ln N}{\ln^2 (U_b/s_0)} \sim s_0^2 \ln N$$

M.M. Desai & D.S. Fisher, Genetics **176**, 1759 (2007)

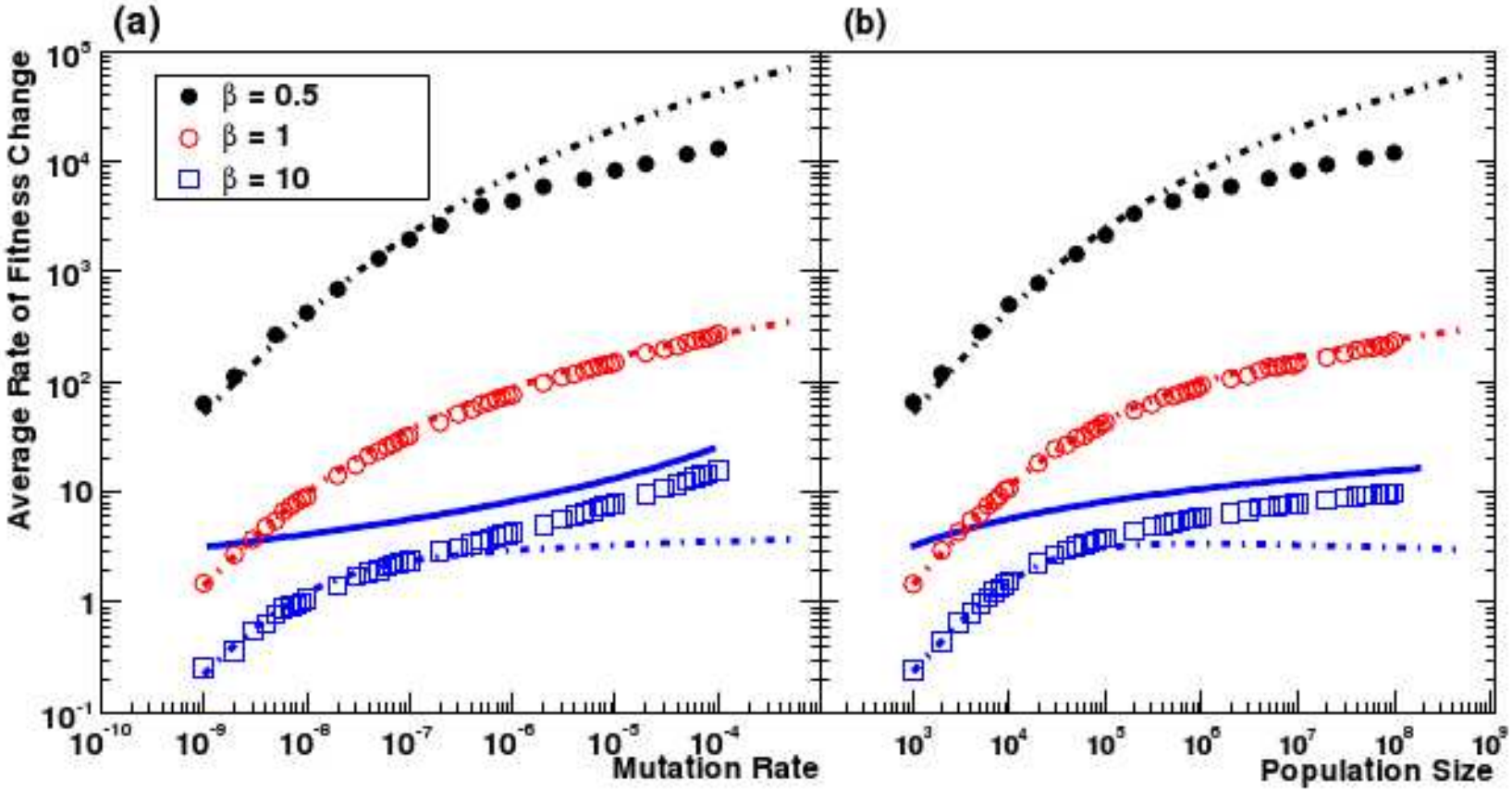
- Adaptation driven by
  (i) single mutations of large effect for $\beta < 1$
  (ii) multiple mutations of average effect for $\beta > 1$

- The "standard case" $\beta = 1$ is marginal

Rate of adaptation: Simulations



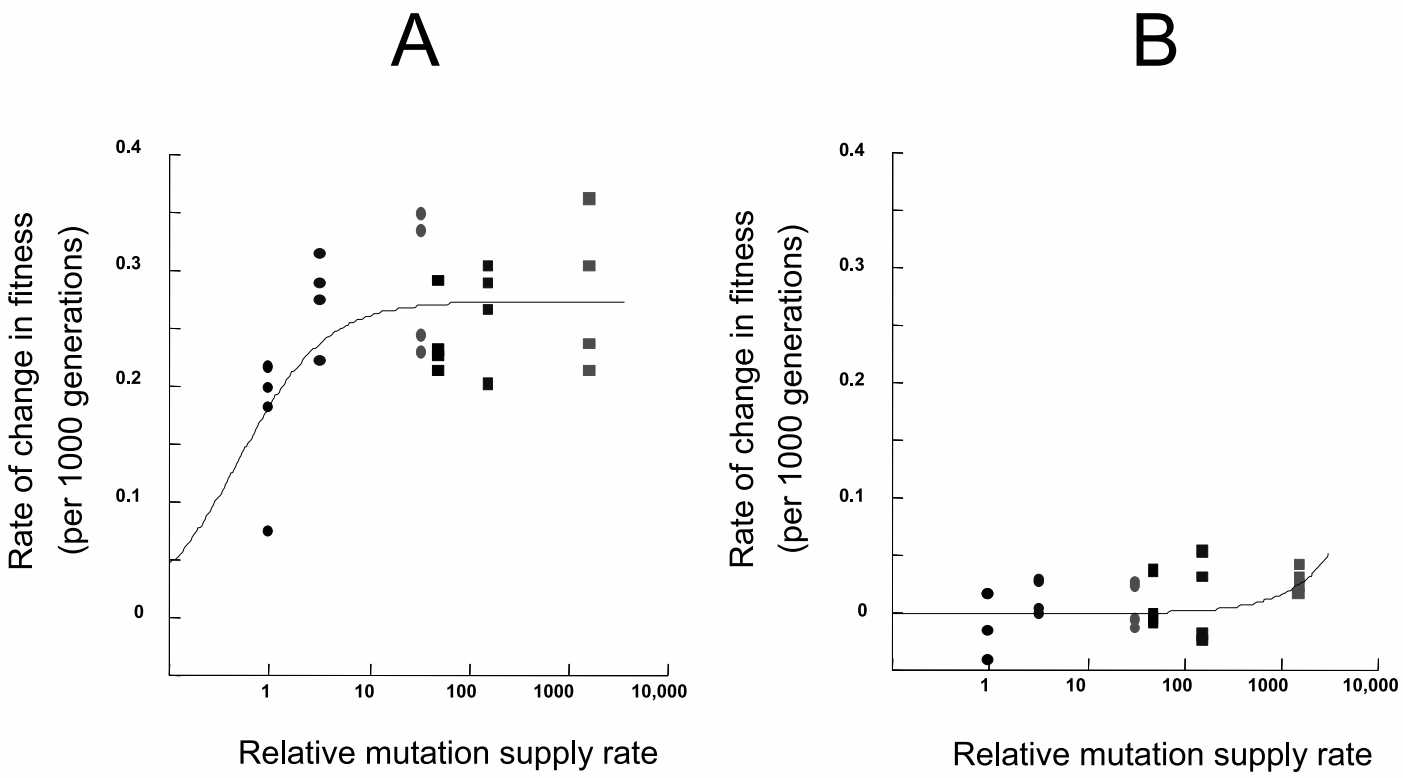
Simulations for other mutation distributions



C.A. Fogle, J.L. Nagle, M.M. Desai, arXiv:0804.1116v1

Rate of adaptation: Experiments

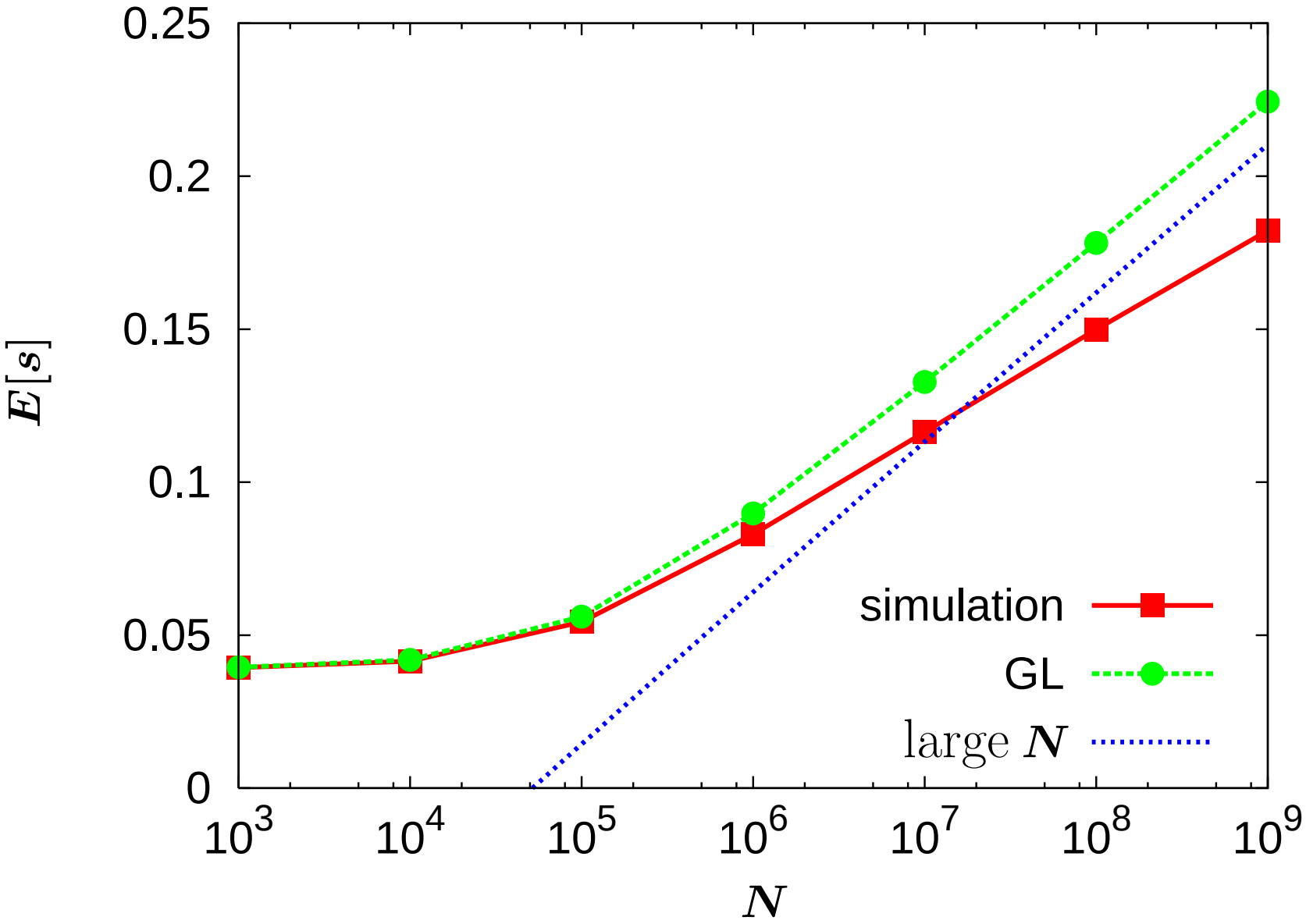
J.A.G.M. de Visser et al., Science **283**, 404 (1999)



non-adapted

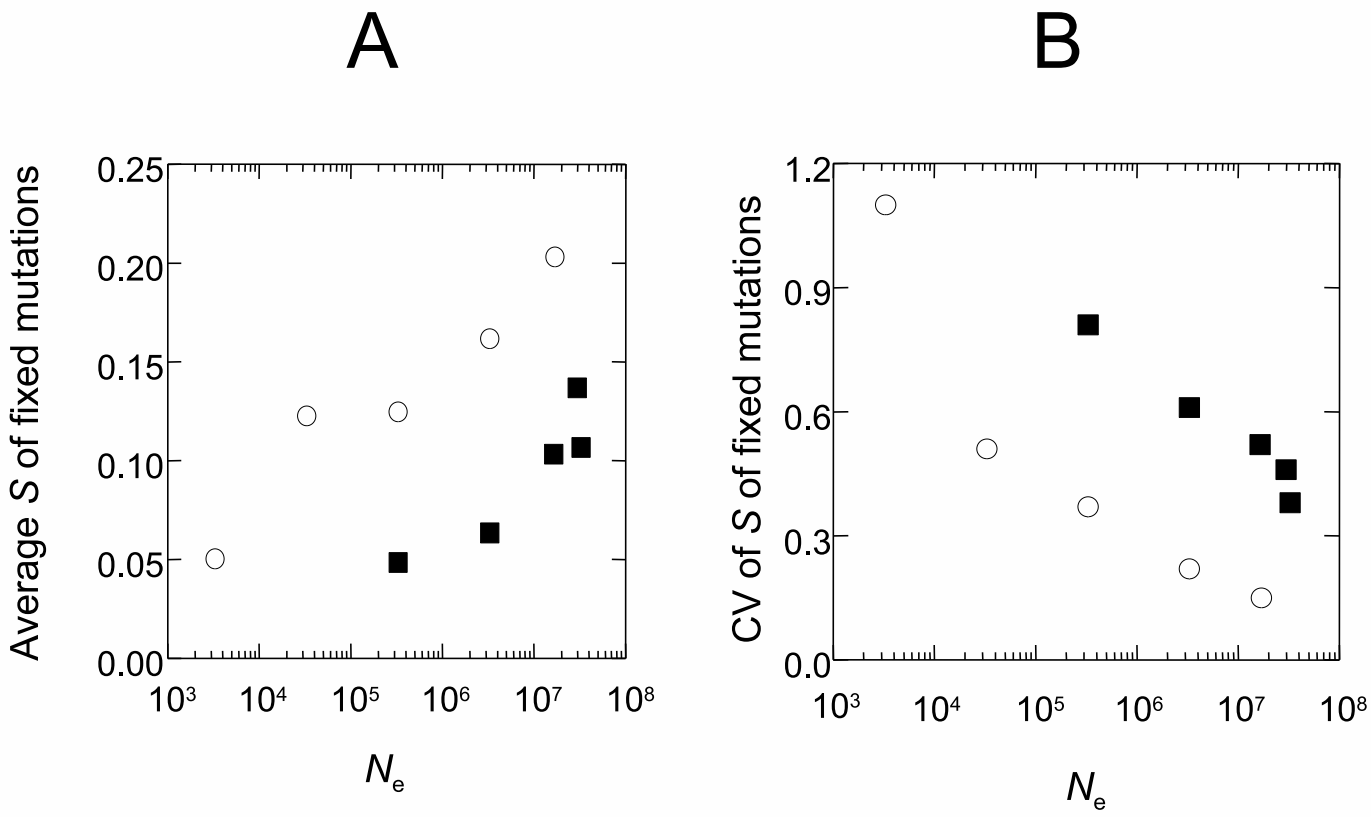
adapted

Selection coefficient of fixed mutations: Simulations

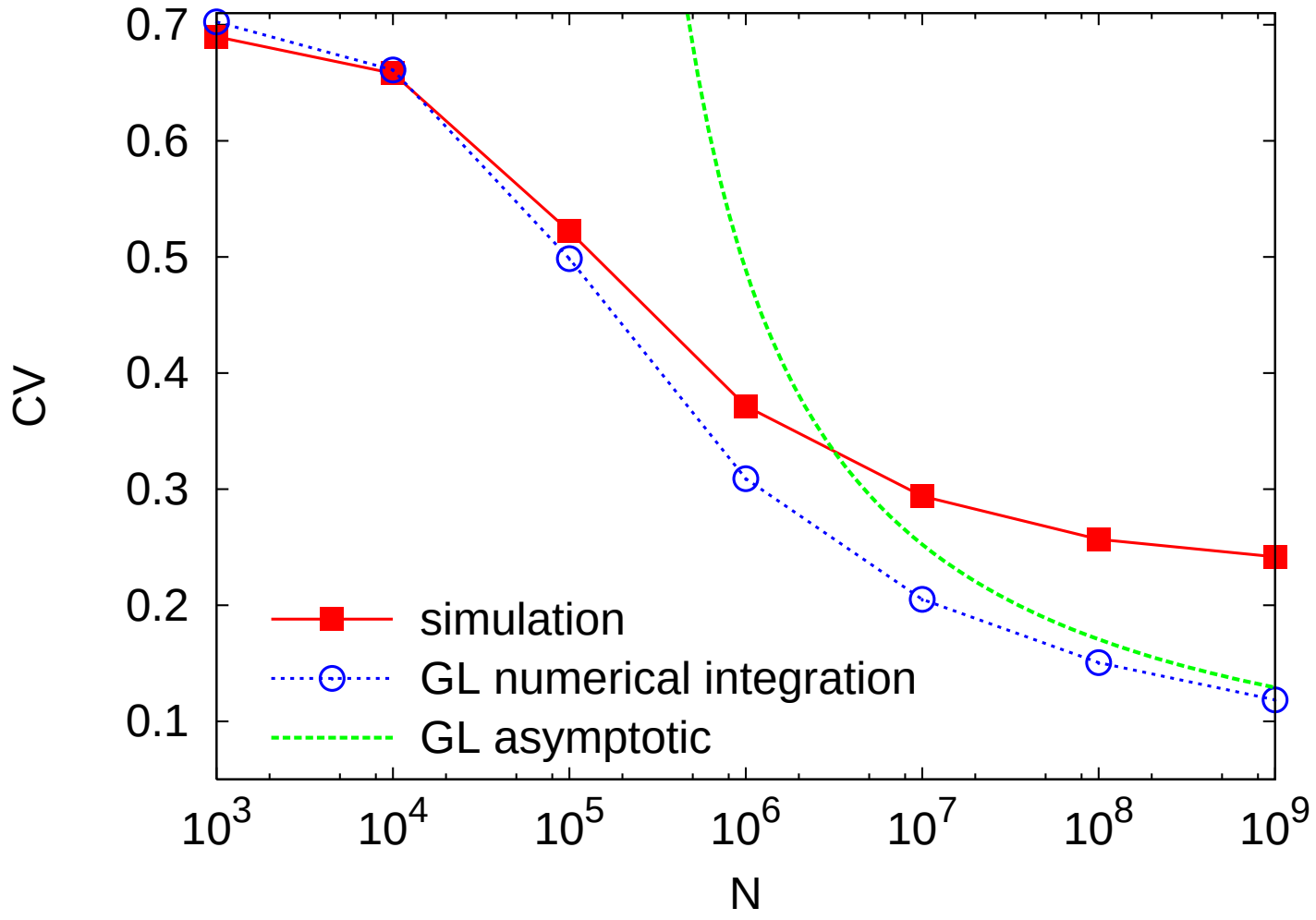


Selection coefficient of fixed mutations: Experiments

J.A.G.M. de Visser & D.E. Rozen, J. Evol. Biol. **18**, 779 (2005)

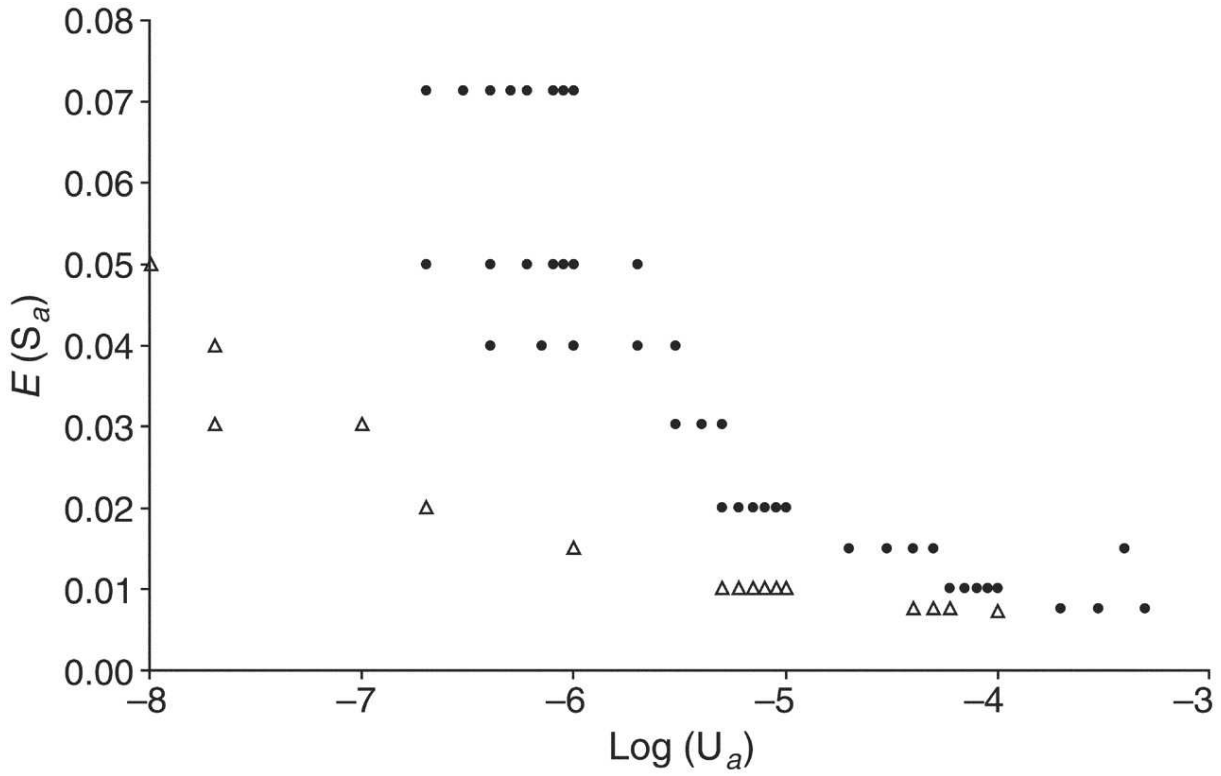


Coefficient of variation from simulations



Reconstruction of model parameters from experiments?

L. Perfeito et al., Science **317**, 813 (2007)

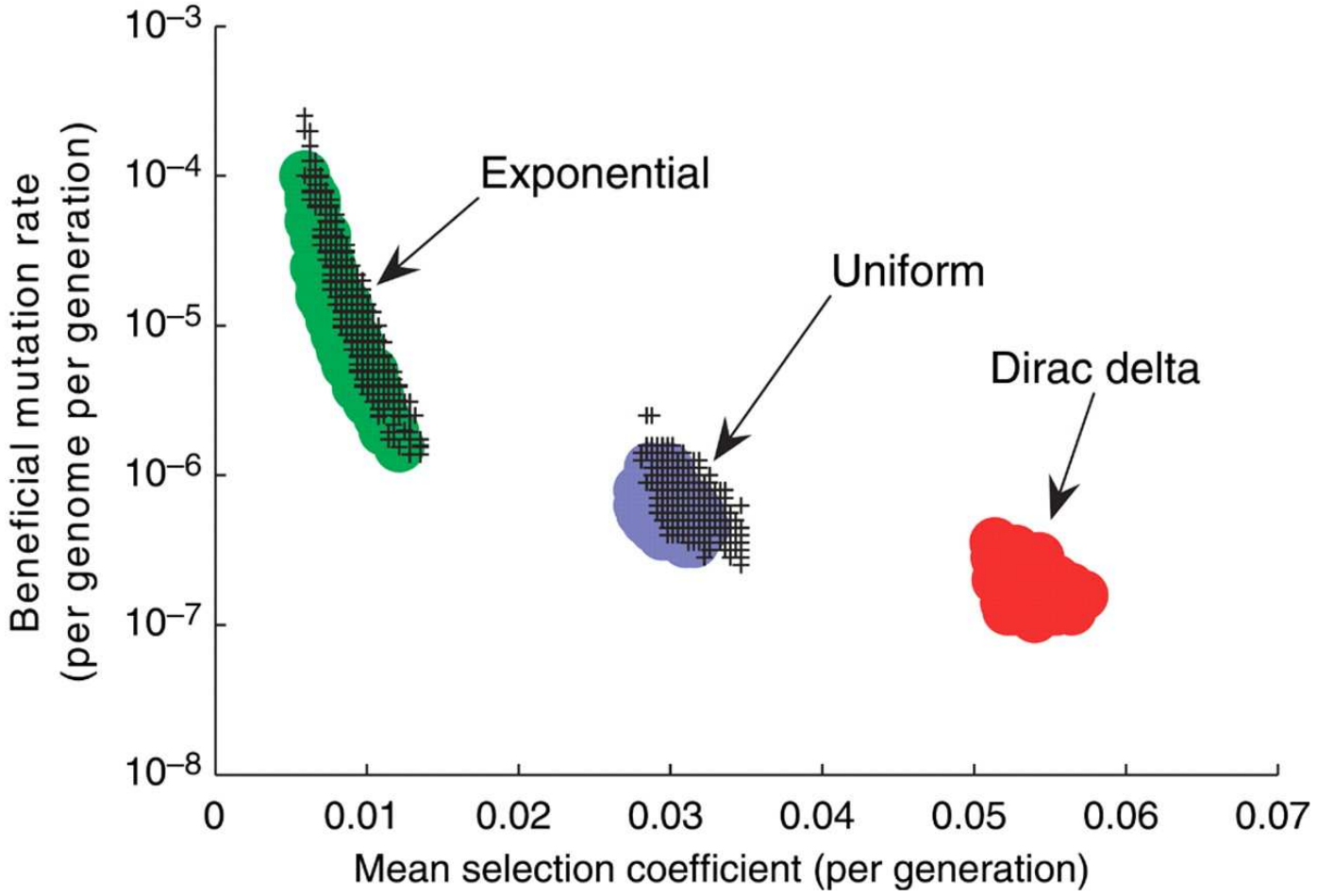


● $N = 2 \times 10^4$ $\quad\quad$ △ $N = 10^7$

- Figure shows simulated parameters consistent with experiments

Reconstruction of model parameters from experiments?

M. Hegreness et al., Science **311**, 161 (2006)



- Estimate of U_b and s_b depends on the choice of $P_b(s)$!

The role of multiple mutations

Structure of the substitution process

- In the presence of multiple mutations, the process of origination of fixed mutations must be distinguished from the process of fixation:

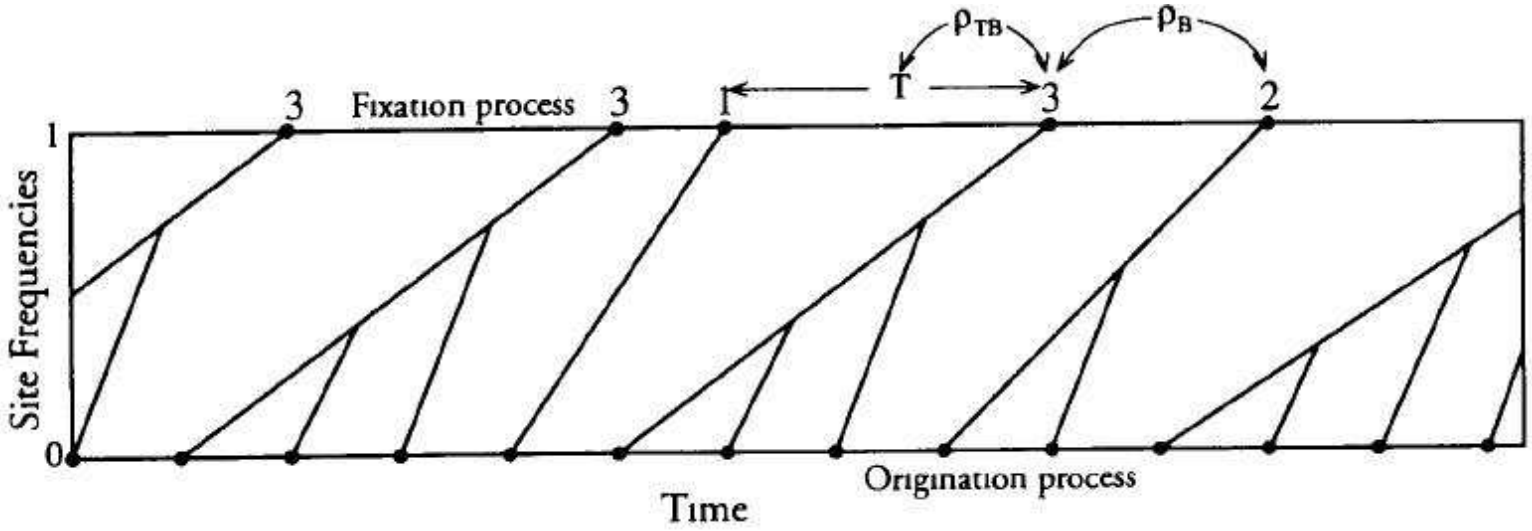
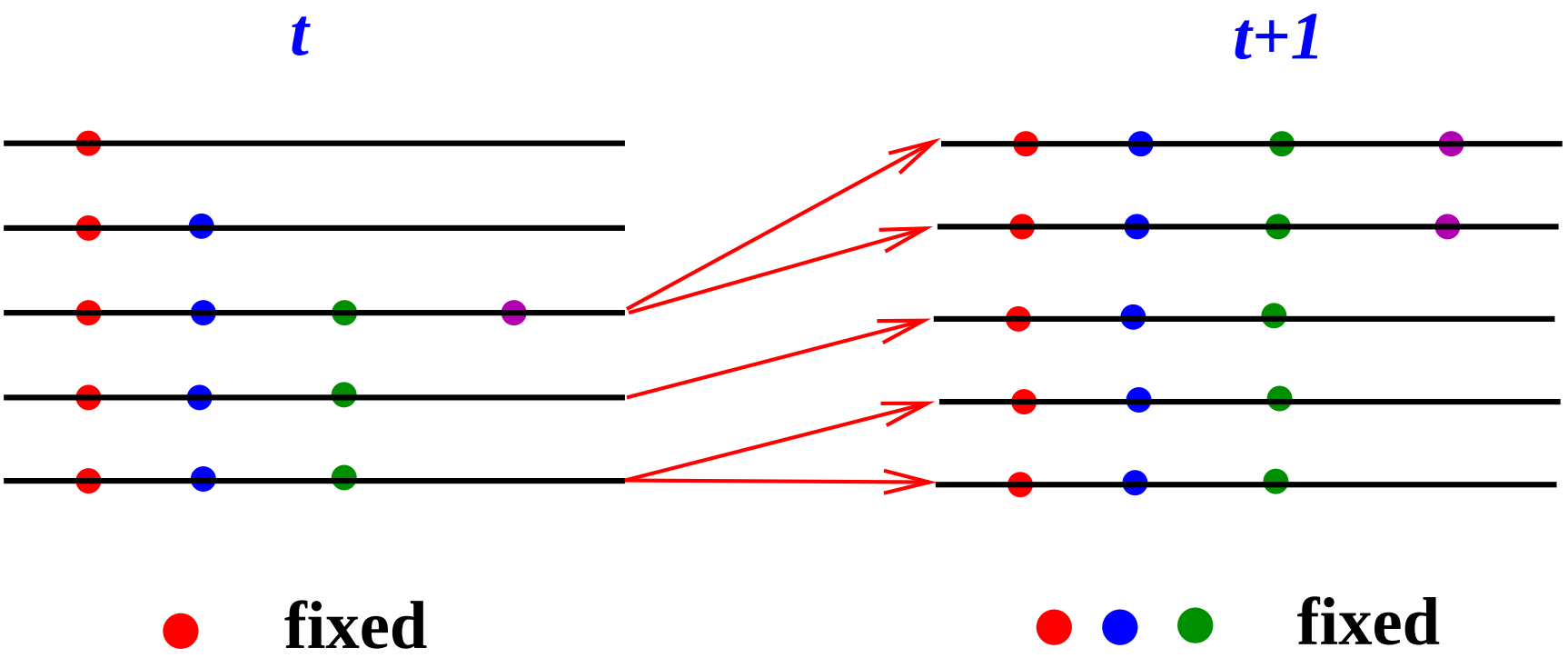


FIG. 1. A diagram of the trajectories of mutations that ultimately fix in the population.

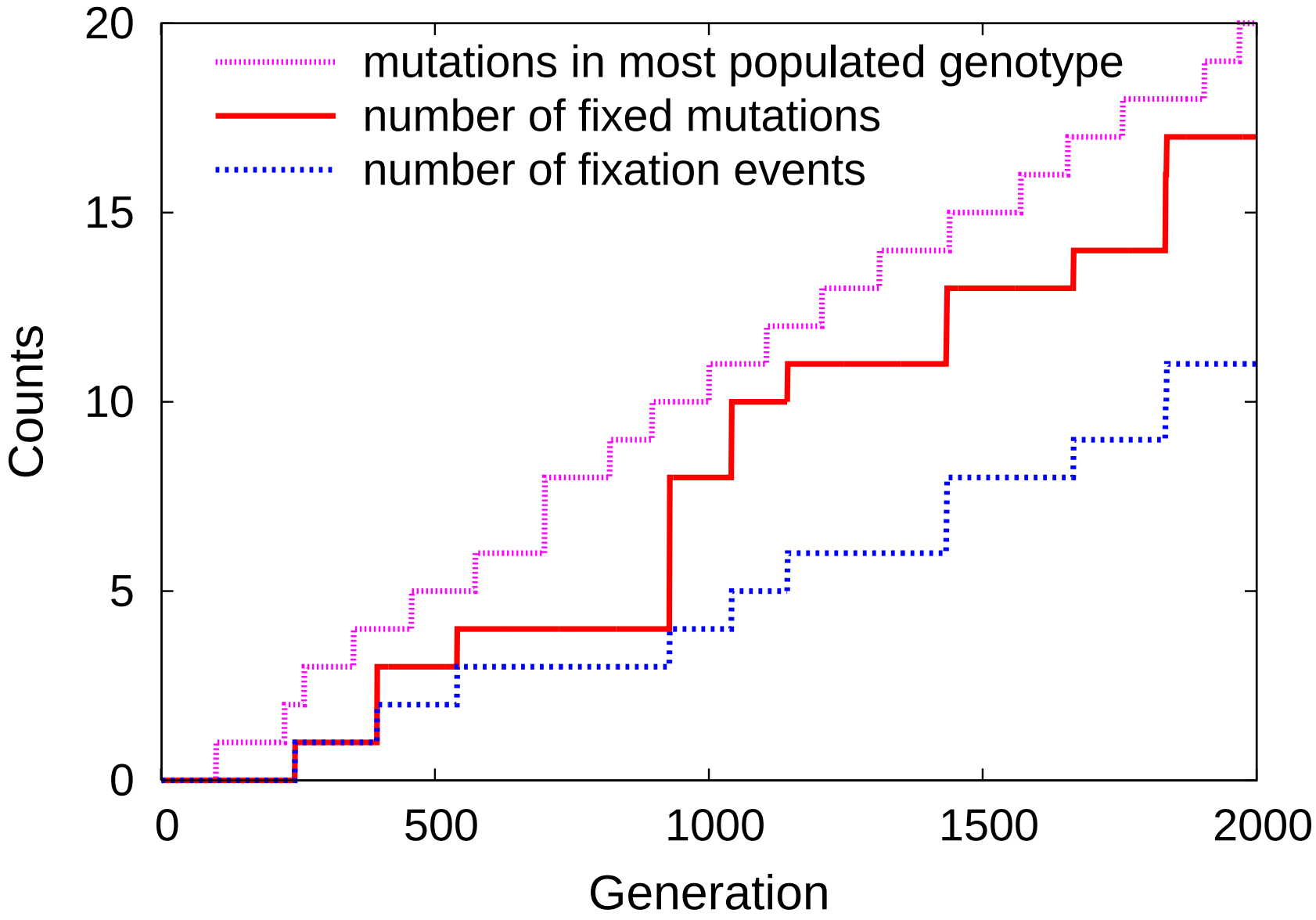
J.H. Gillespie, Genetics **134**, 971 (1993)

Fixation of multiple mutations

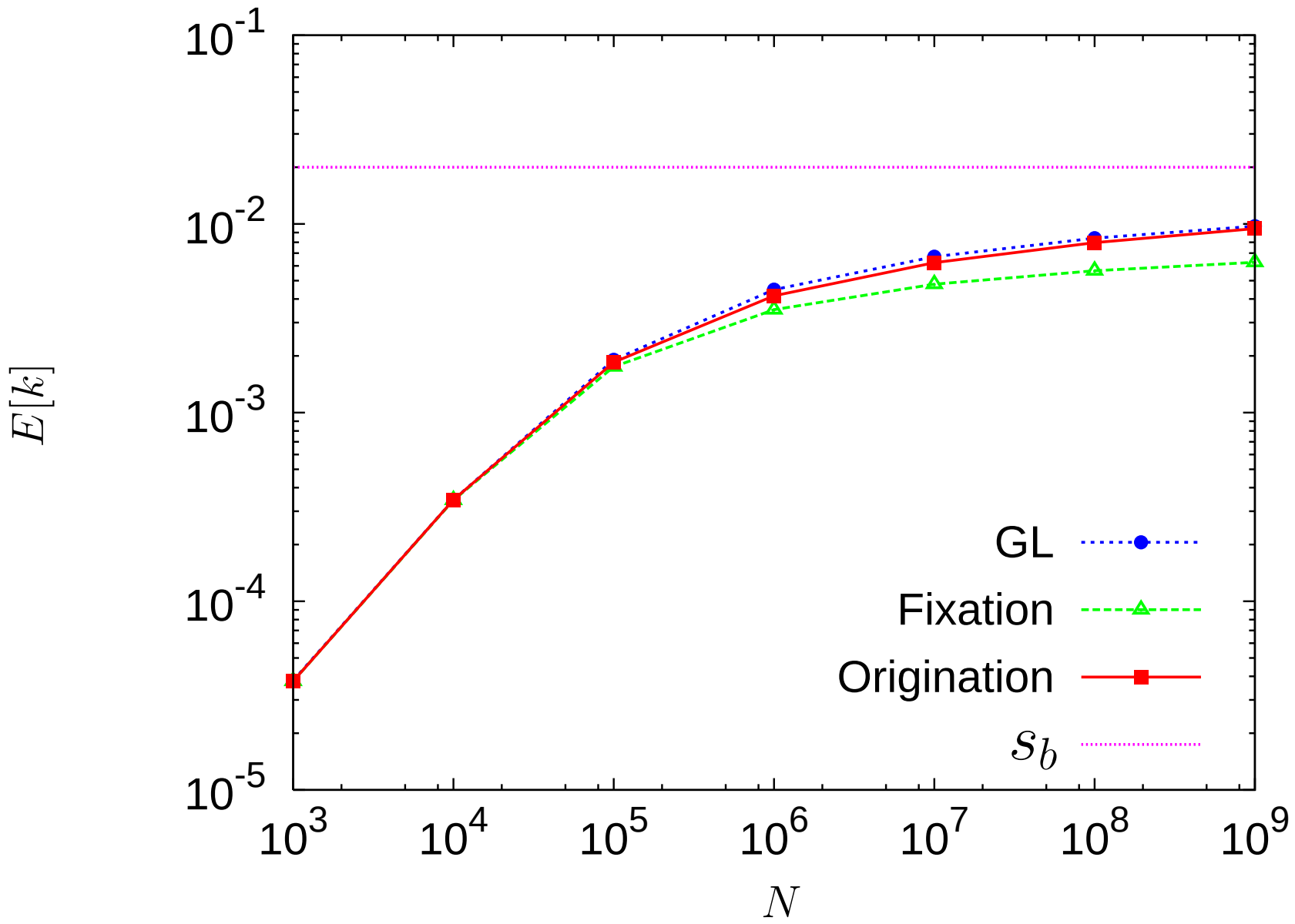
Fixation: Change in the genotype of the **most recent common ancenstor**



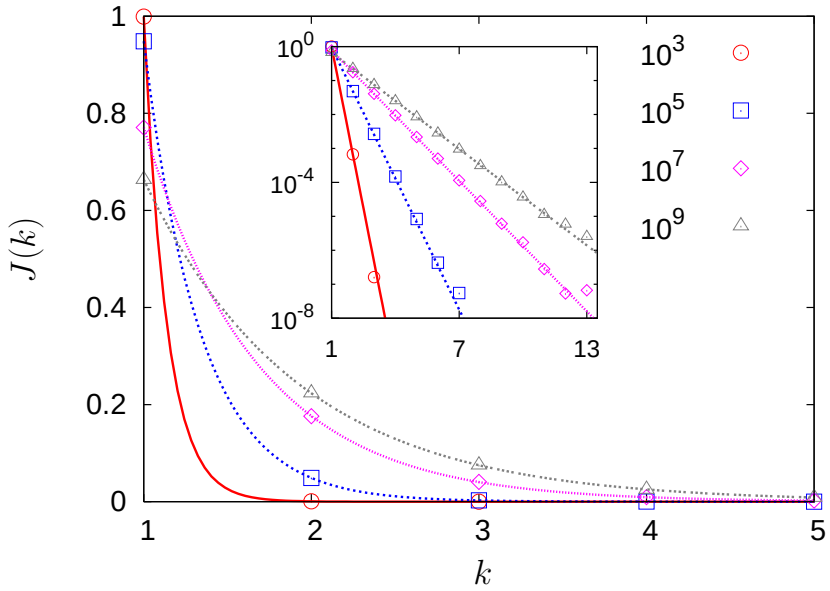
Mutation and fixation processes [$N = 10^9$]

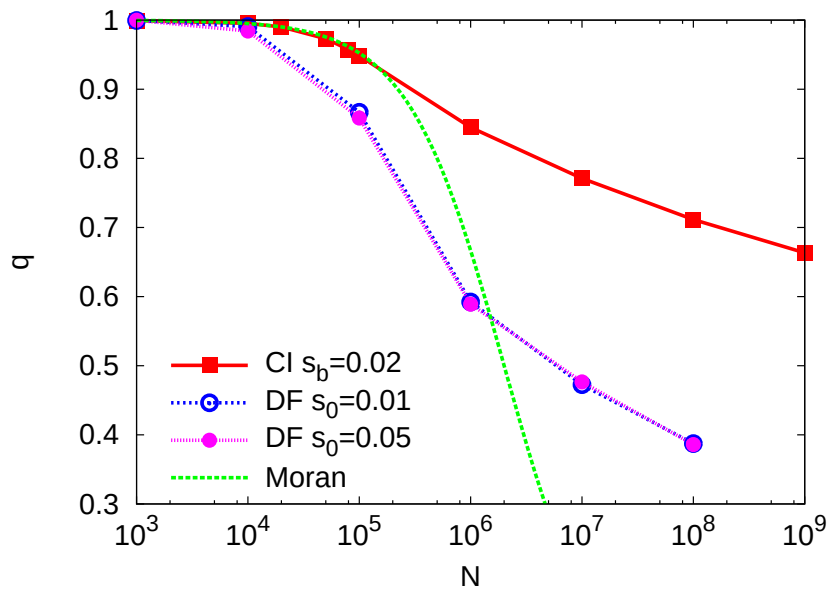


Rate of substitution: Simulations



Distribution of the number of simultaneously fixed mutations





- Geometric distribution $J(k) = q(1-q)^{k-1}$ with mean $1/q$
- Geometric distribution with $q(N) = 2/(2+NU)$ is exact in the neutral case (Watterson, 1982)
- Stronger effect of multiple mutations for $P_b(s) = \delta(s - s_0)$

The rhythm of microbial adaptation

P.J. Gerrish, Nature **413**, 299 (2001)

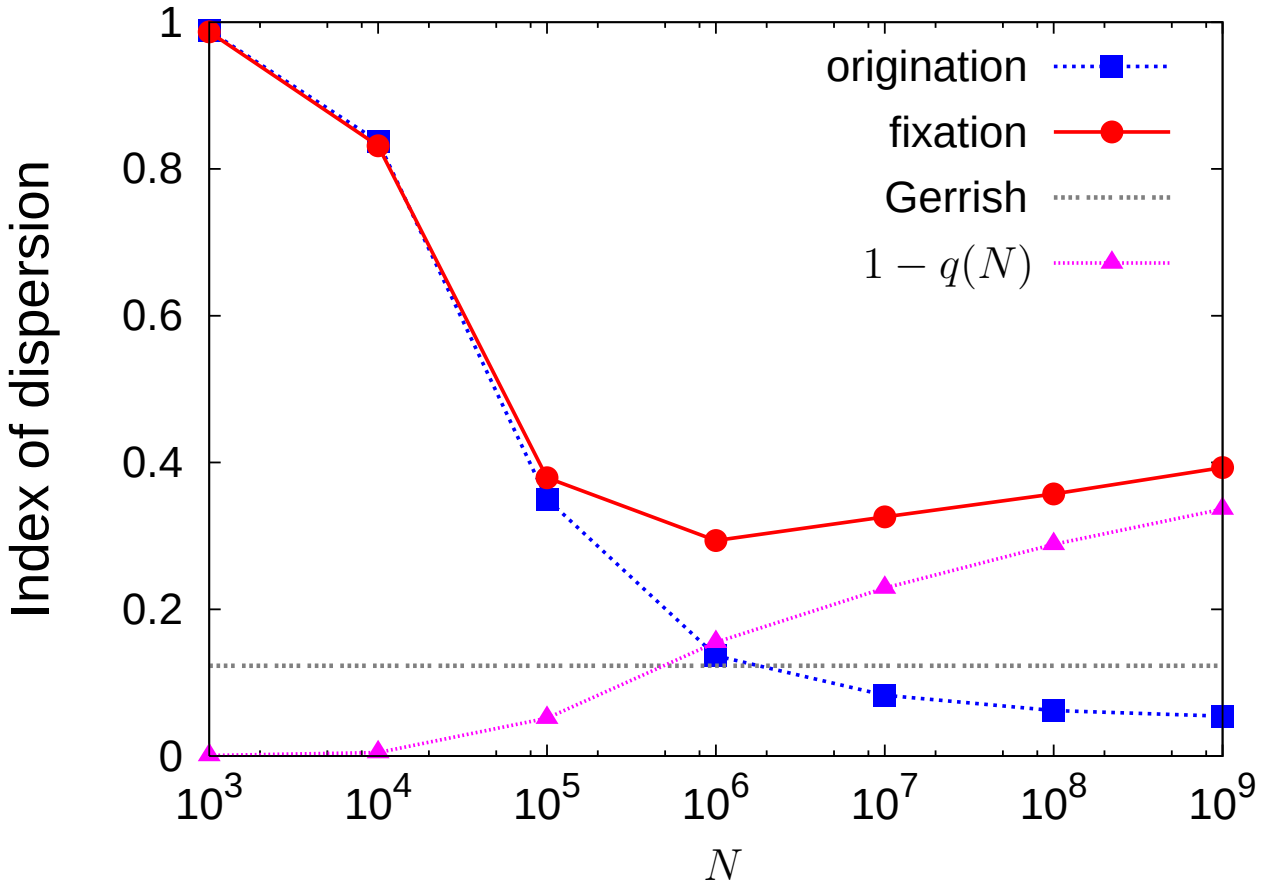
- Statistics of the number $n(t)$ of substitution events up to time t ?
- The index of dispersion of the substitution process is the ratio of the variance of $n(t)$ to the mean:

$$I = \frac{Var[n(t)]}{E[n(t)]}$$

- In the periodic selection regime $I = 1$ (Poisson statistics)
- GL-theory predicts a universal, sub-Poissonian limit in the clonal interference regime:

$$\lim_{N\to\infty} I(N) = 2e^{-\gamma} - 1 \approx 0.123$$

The rhythm of origination and fixation



- Origination process becomes regular $(I \to 0)$ for large N
- Index of dispersion of fixation process $I \approx 1 - q \to 1$ for $N \to \infty$

Fitness landscapes and epistasis

- **So far:** Fitness effects of different beneficial mutations are independent
- Epistasis implies interactions between the effects of different mutations
- General setting: Genome of L binary loci (sites) $i = 1, \ldots, L$ at which a mutation can be present ($\sigma_i = 1$) or absent ($\sigma_i = 0$)
- A fitness landscape is a function $w(\sigma)$ on the set of 2^L genotype sequences $\sigma = (\sigma_1, \ldots, \sigma_L)$
- In the absence of epistatic interactions $w(\sigma) = \prod_{i=1}^{L} \omega_i(\sigma_i)$
- **What is the effect of epistasis on asexual and sexual adaptation?**
- **How epistatic are real fitness landscapes?**

The house-of-cards model

S.C. Park, JK, JSTAT (2008) P04014

- Infinite sites model with mutant fitnesses w drawn randomly and independently from mutation distribution $g(w)$
  $\Rightarrow$ maximally epistatic fitness landscape

- In the limit $N \to \infty$ the population fitness distribution evolves according to

$$f_{t+1}(w) = (1-U)\frac{w f_t(w)}{\overline{w}_t} + U g(w) \qquad \overline{w}_t : \text{ mean fitness}$$

- Mutation-selection balance for $g(w)$ with bounded support Kingman (1978)

- For unbounded $g(w) \sim \exp[-(w/w_0)^\beta]$ mean fitness grows as

$$\overline{w}_t \approx C_\beta w_0 (1-U) t^{1/\beta} \qquad 1-U : \text{ mutational load}$$

Finite populations and records

- At long times beneficial mutations are rare events:

$$U_b(t) = U\ \mathrm{Prob}[w > \overline{w_t}] = U \int_{\overline{w_t}}^{\infty} dw\, g(w) \to 0 \ \text{ for } \ t \to \infty$$

- For $U \ll 1$ the effect of deleterious mutations can be neglected as well $\Rightarrow$ approximation by a **diluted record process** w_t^{DRP}, in which mutants of fitness $w' > w$ replace current genotype w with the **fixation probability**

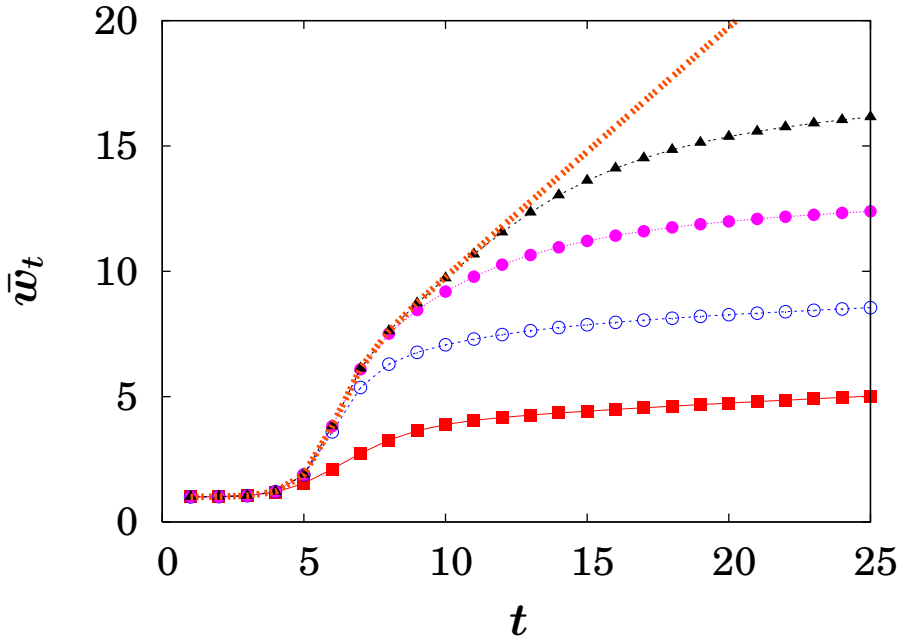
$$\pi(s) = 1 - e^{-2s}, \quad s = w'/w - 1.$$

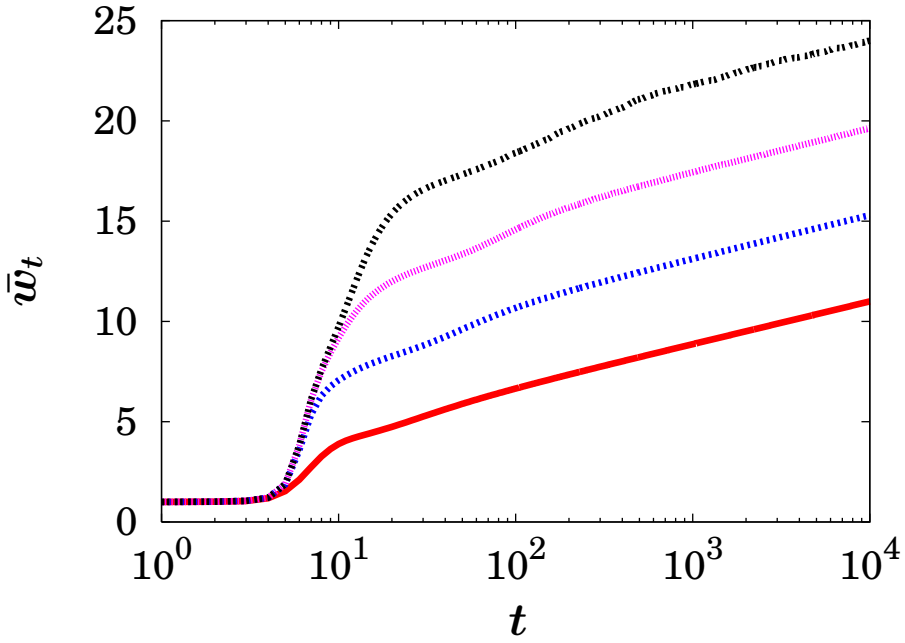
- To leading order $\overline{w_t^{\mathrm{DRP}}}$ is equal to the largest fitness value encountered up to time t [=standard record process], with corrections that can be systematically computed

- Deleterious mutations rescale the fitness according to

$$\overline{w_t} \approx (1-U)\overline{w_t^{\mathrm{DRP}}} \approx (1-U)\ln(NUt) \ \text{ for } \ g(w) = e^{-w}$$

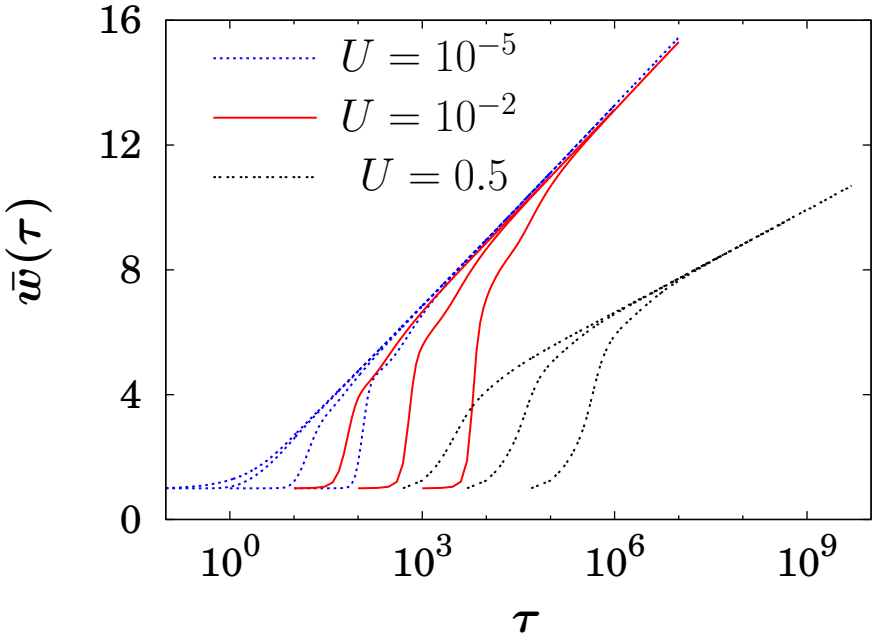
Simulations: Finite vs. infinite populations

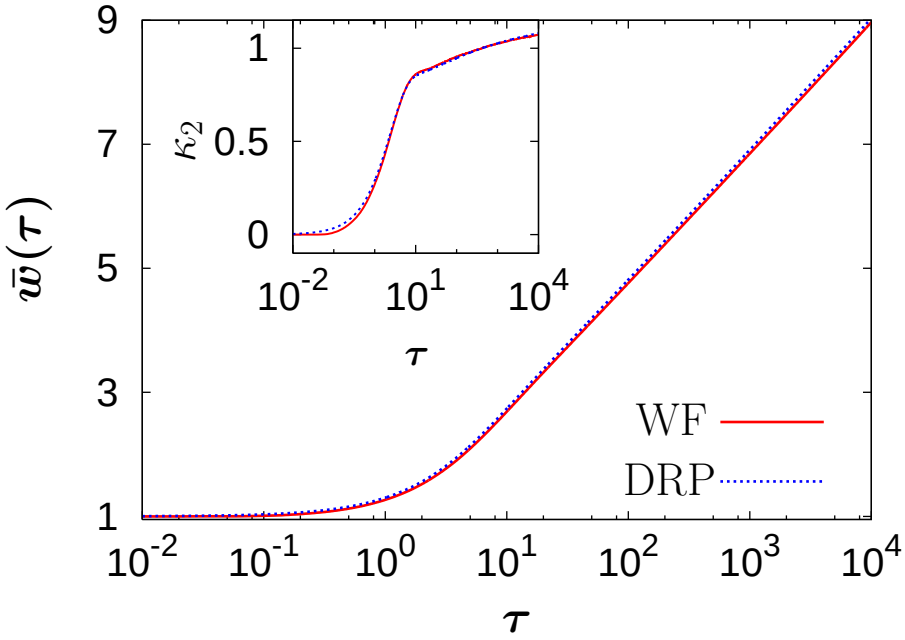




$U = 0.01, N = 10^3, 10^5, 10^7, 10^9, \infty$

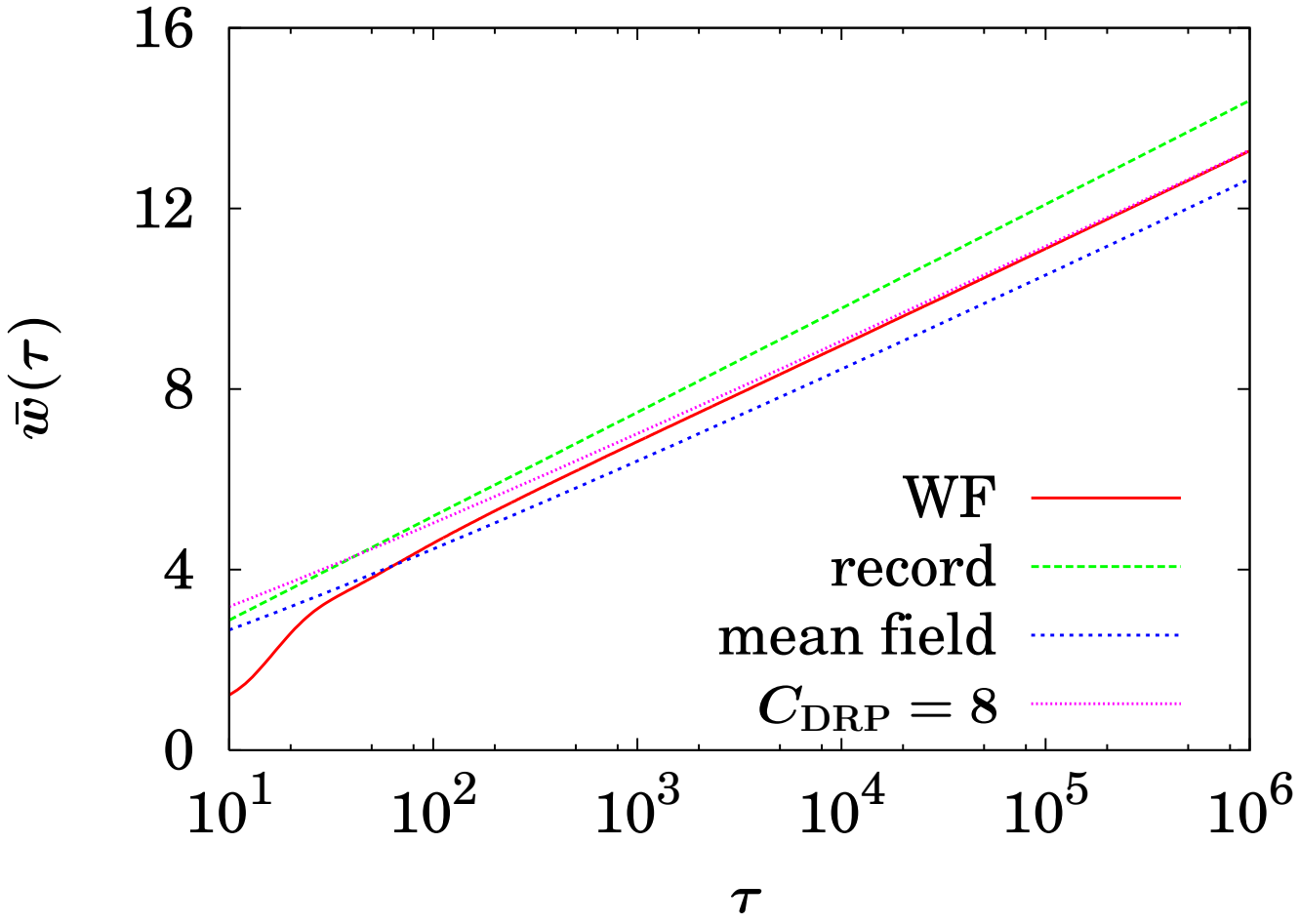
Comparison to the diluted record process





- scaled time $\tau = NUt$
- fitness variance $\kappa_2 \to \text{const.}$

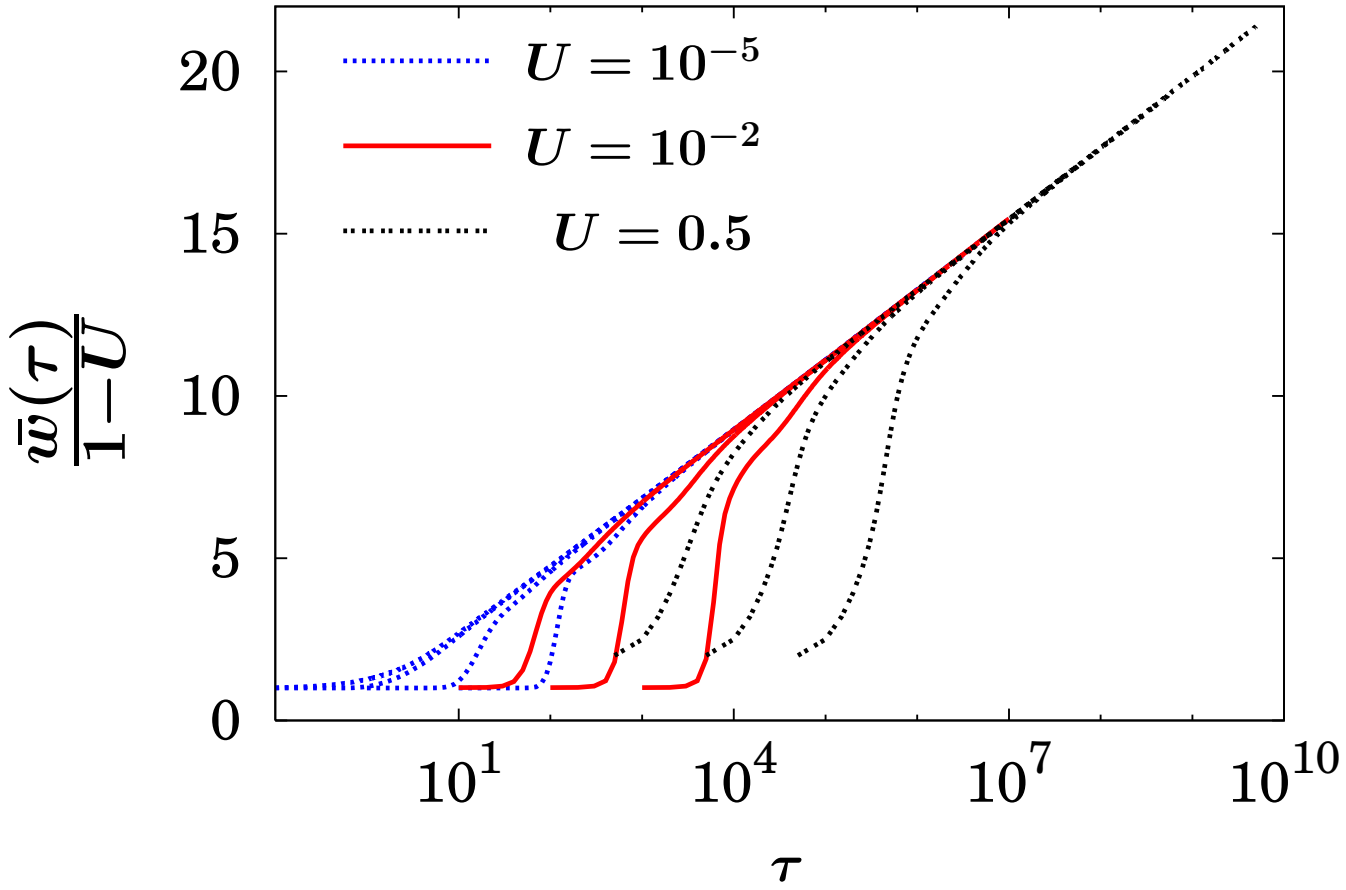
Diluted record process: Bounds and approximations



record process: upper bound

mean field approximation: lower bound

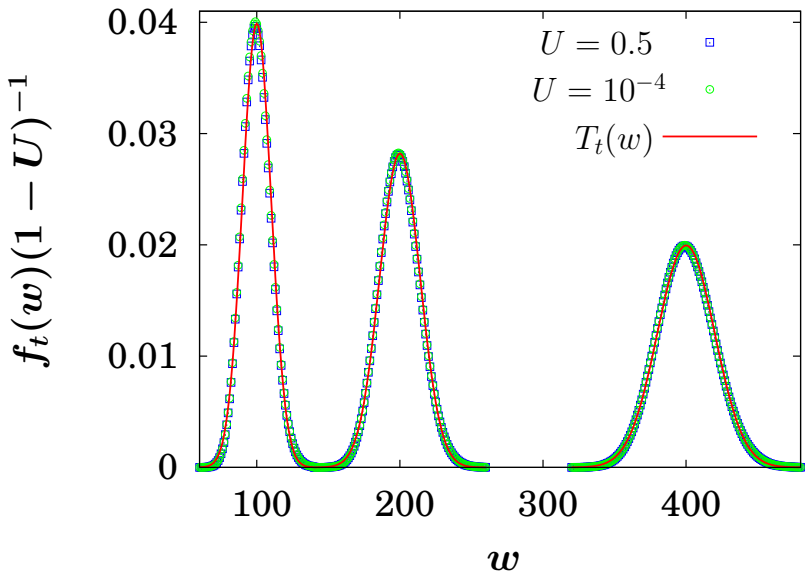
Finite populations at arbitrary U



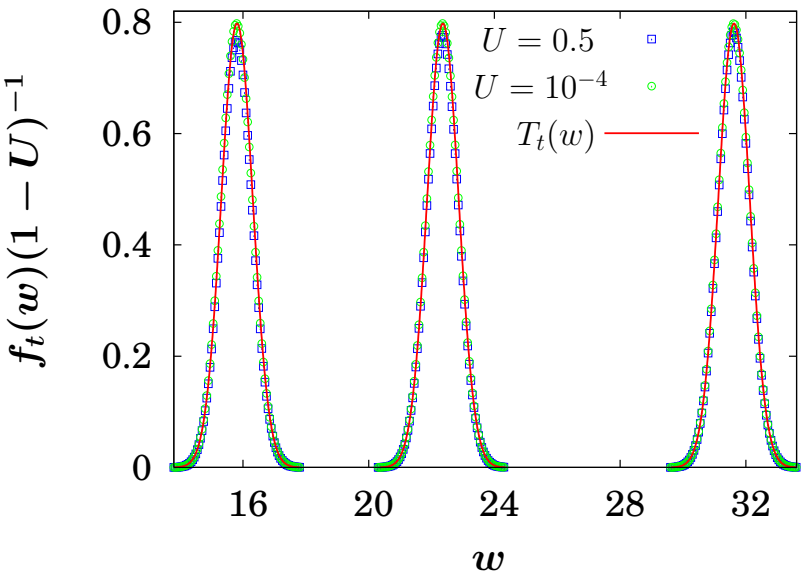
$N = 10^3, 10^4, 10^5$

Bimodality of fitness distribution

exponential $g(w)$



Gaussian $g(w)$



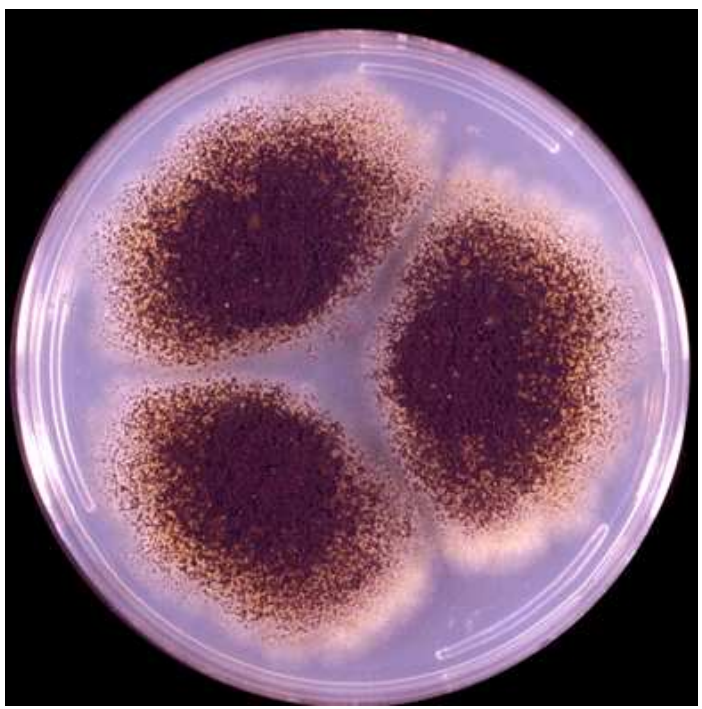
- Asymptotic decomposition

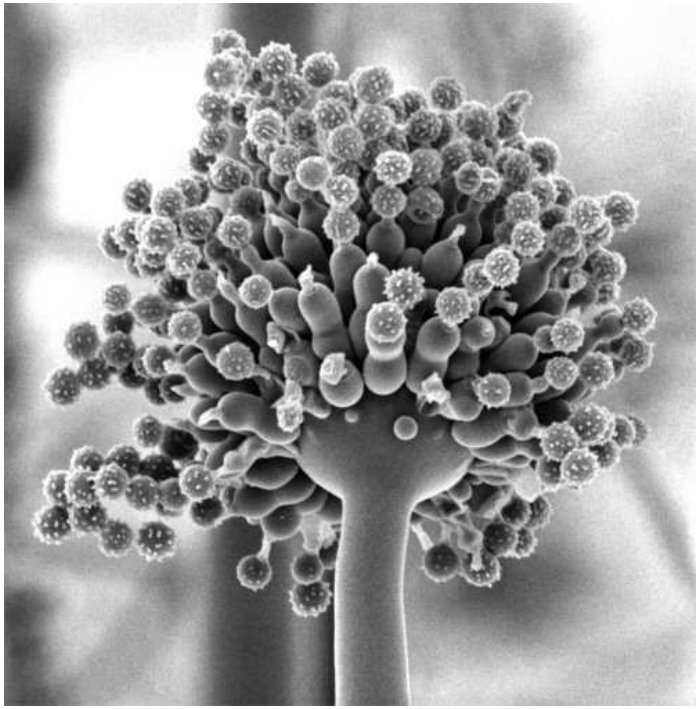
$$f_t(w) \approx Ug(w) + (1-U)T_t(w)$$

with a “traveling wave” contribution $T_t(w)$ holds for finite and infinite populations

Empirical fitness landscapes for *Aspergillus niger*

J.A.G.M. de Visser, S.C. Park, JK, arXiv:0807.3002

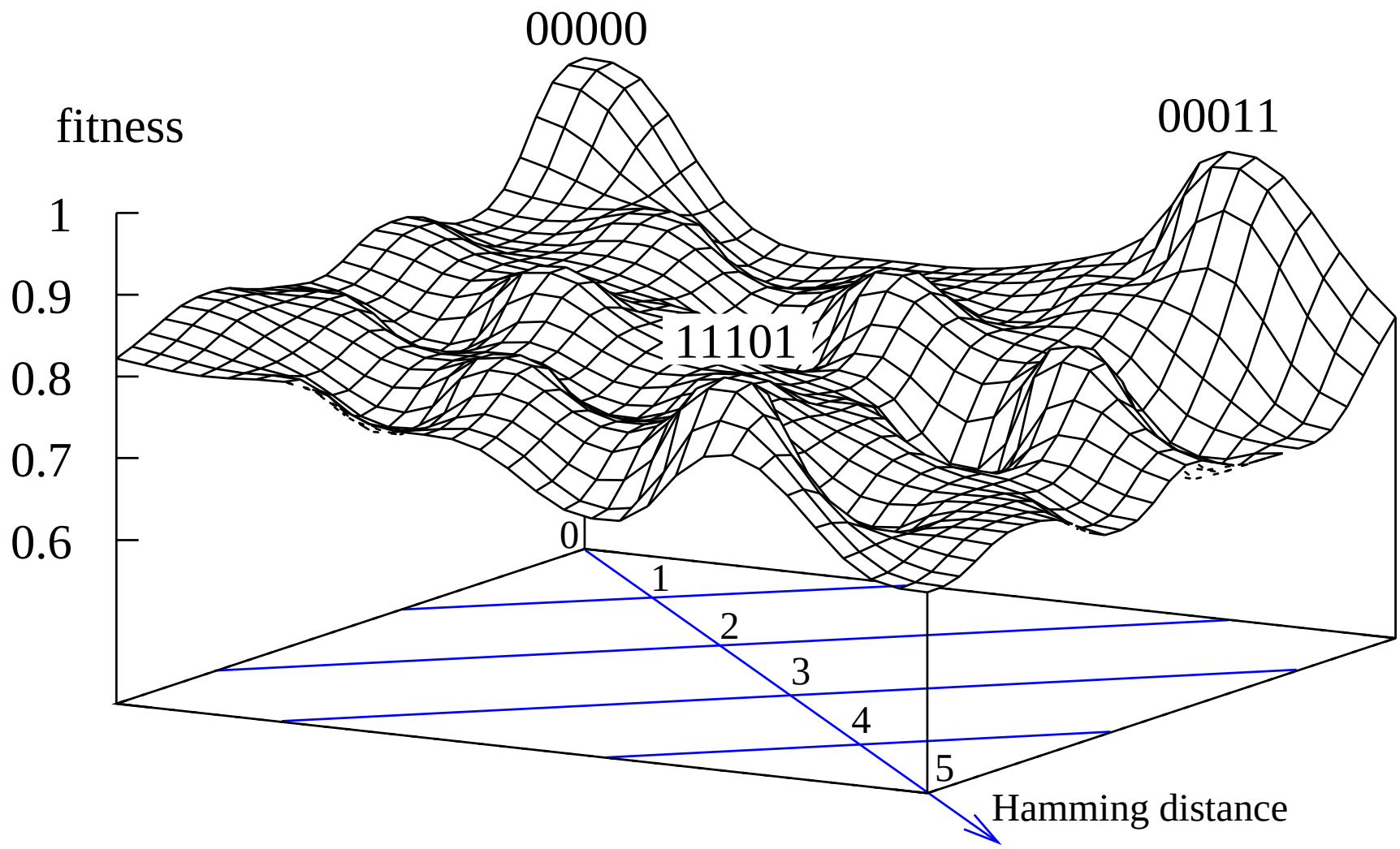




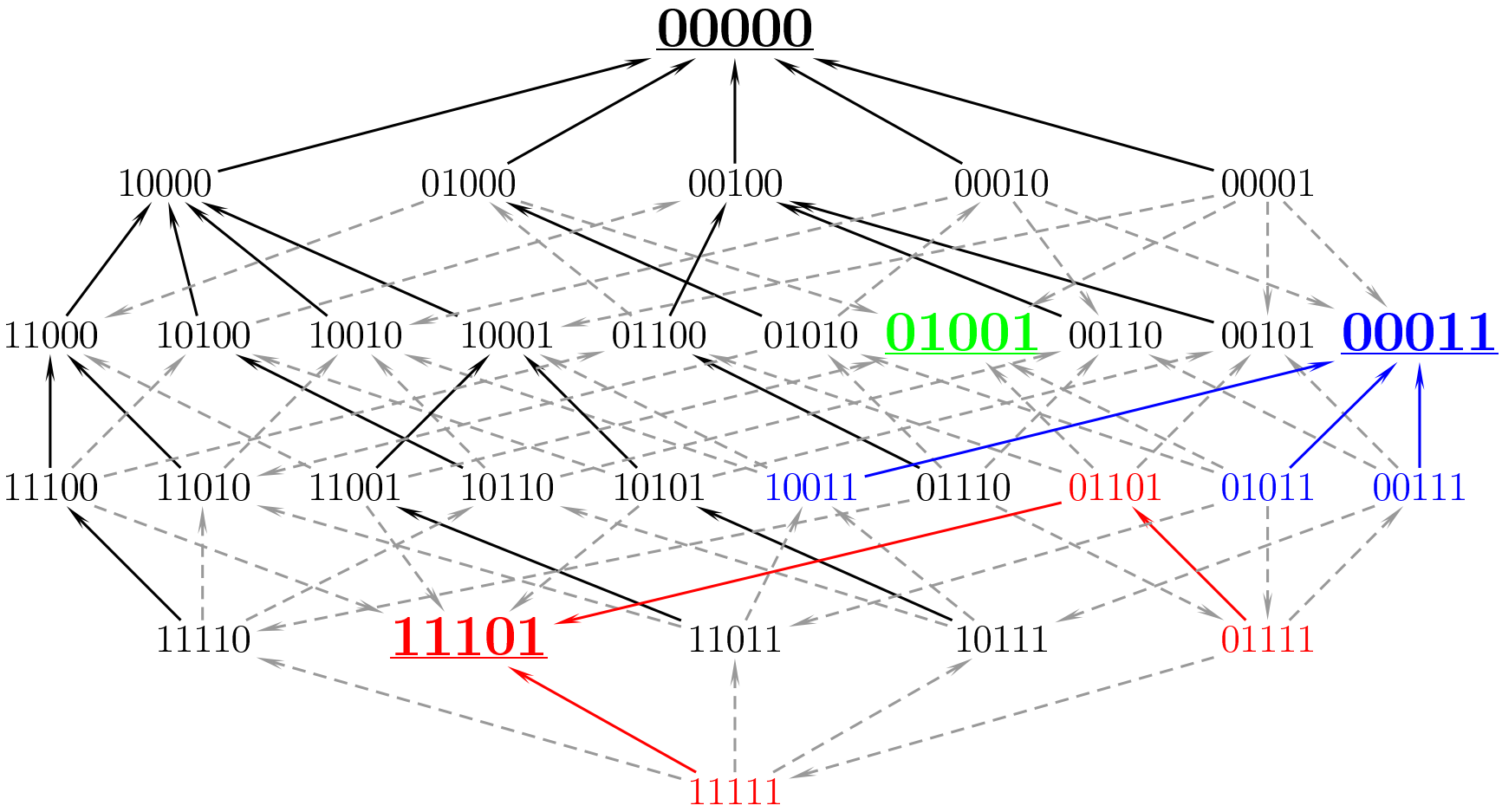
Figures courtesy of Mycology Online & N.D. Read (Edinburgh)

- 7 marker mutations known to be individually deleterious
- Fitness measurements of 186 strains, including 2 complete sets of $2^5 = 32$ combinations of $L = 5$ binary mutations J.A.G.M. de Visser et al. (1997)

The *A. niger* fitness landscape: An artist's impression



The *A. niger* fitness landscape: Arrow graph



- Ruggedness: Several local fitness maxima (underlined)
- Most paths $11111 \rightarrow 00000$ are selectively inaccessible

Effect of recombination on adaptation

- Free recombination: Offspring choses each locus at random from one of the two parents; e.g.,

$$\left.\begin{array}{l} 11101 \\ 10100 \end{array}\right\} \Rightarrow \quad 11101 \quad 10101 \quad 11100 \quad 10100$$

  with equal probability

- Recombination occurs with probability r per individual and generation

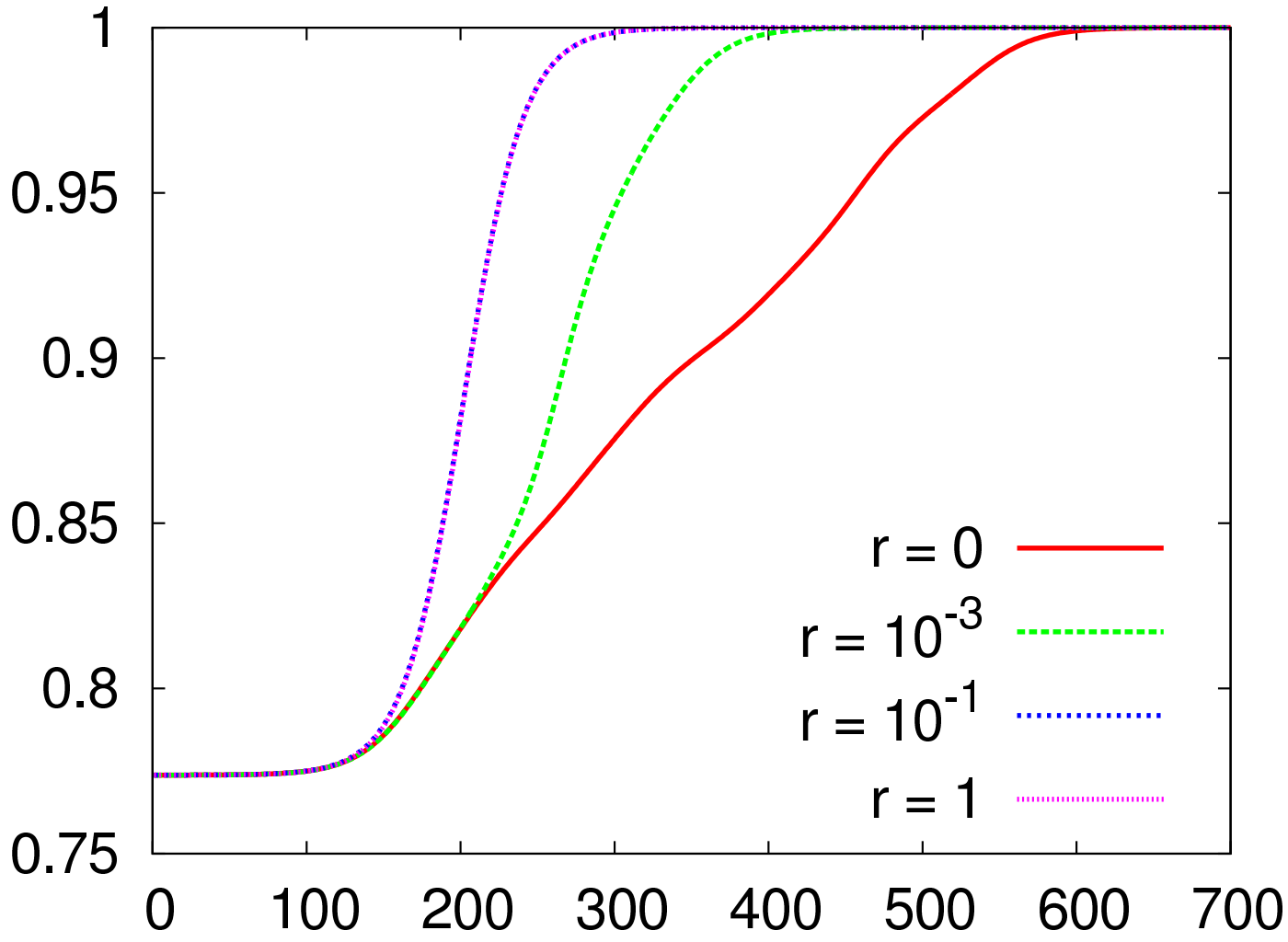
Expectation from two-locus models:

I. Eshel, M.W. Feldman, Th. Pop. Biol. **1**, 88 (1970)

- Recombination speeds up (slows down) adaptation in the presence of negative (positive) epistasis

- Transition from a lower to a higher fitness peak can be completely suppressed by recombination

Multiplicative landscape: Fisher-Muller-effect



$U = 10^{-5}, N = 10^{7}$

Infinite populations in the *A. niger* landscape