SFB 680
Molecular Basis of
Evolutionary Innovations

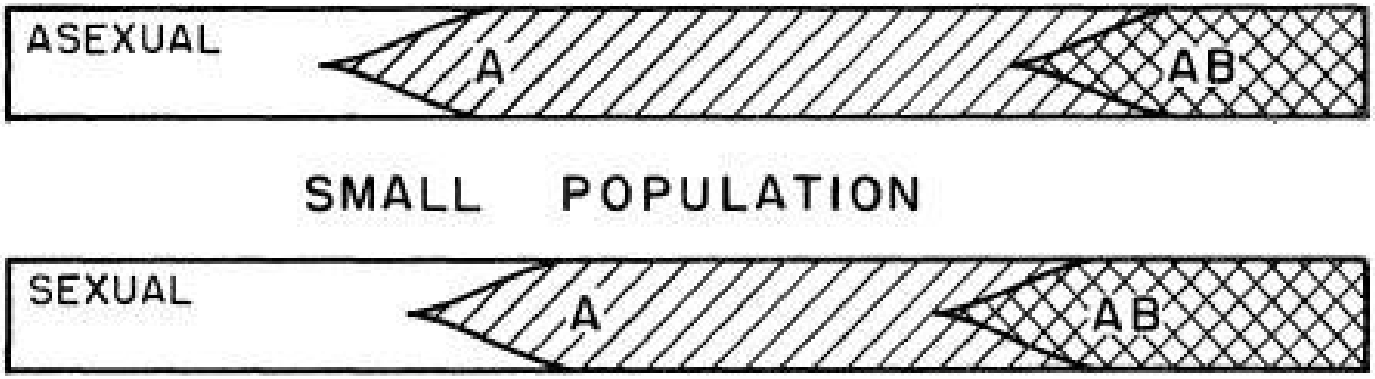
Clonal interference in large populations

Joachim Krug
Institute of Theoretical Physics, University of Cologne

- Introduction: Muller-Fisher hypothesis, Wright-Fisher model
- Clonal interference and the rhythm of microbial adaptation
  PNAS **104**, 18135 (2007)
- House-of-cards model in finite and infinite populations
  arXiv:0711.1989

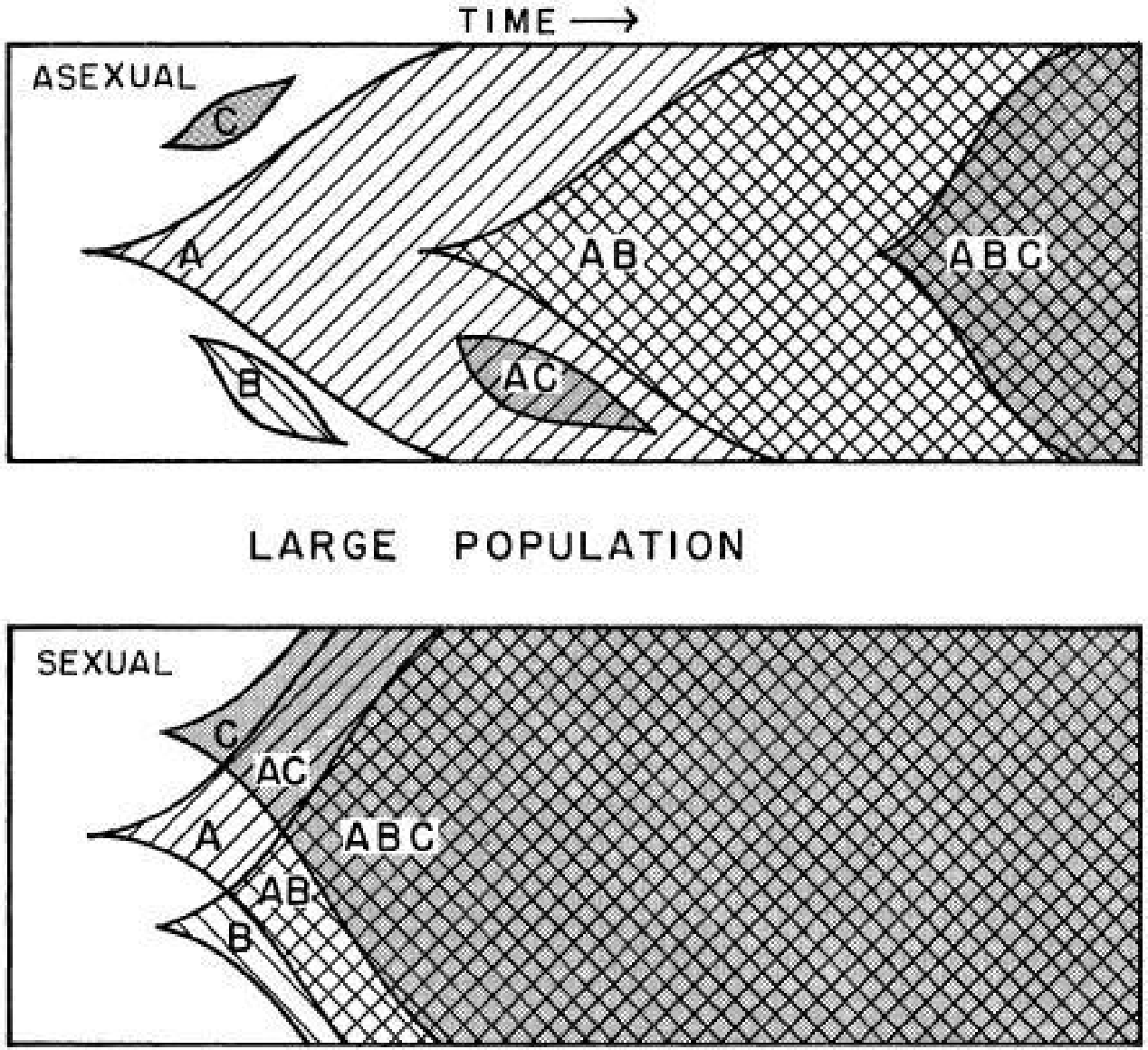
Joint work with Su-Chan Park

The Muller-Fisher hypothesis for the advantage of sex



Muller 1932; Crow & Kimura 1965

The Muller-Fisher hypothesis for the advantage of sex

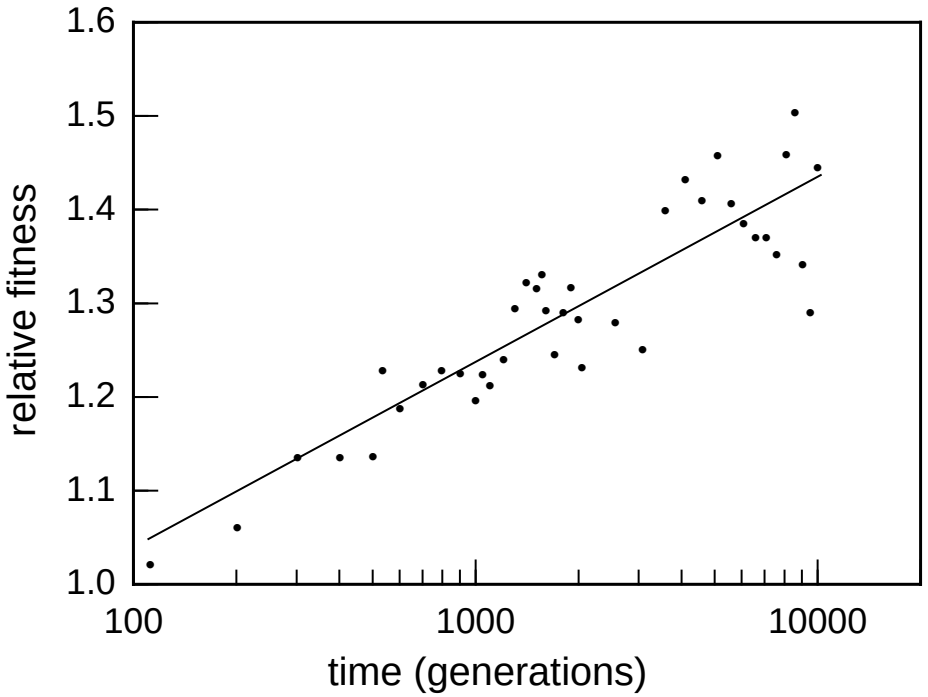


Experimental evolution with microbial populations

S.F. Elena, R.E. Lenski, Nature Reviews Genetics **4**, 457 (2003)

Issues:

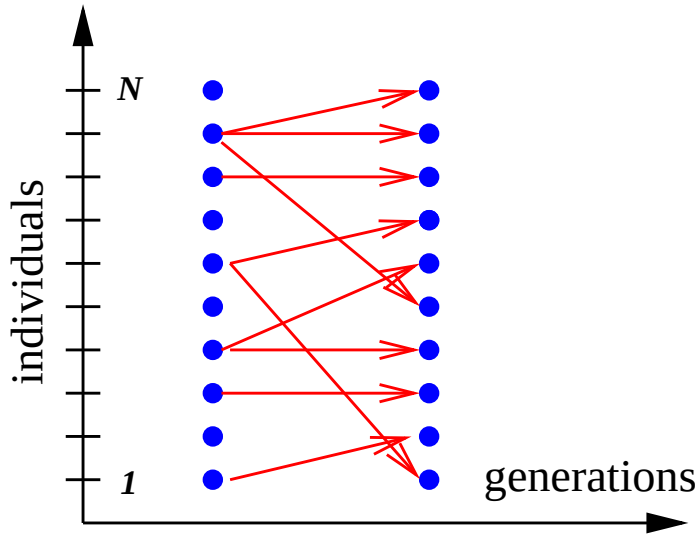
- Speed of adaptation
- Statistics of adaptive events
- Fitness advantage of fixed beneficial mutations
- Structure of the fitness landscape
- Deterministic vs. stochastic evolution



fit: Sibani, Brandt & Alstrøm (1998)

Evolution of asexual populations

Basic model: Wright-Fisher sampling of a finite population of size N



- Each individual choses 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
- Two distinct sources of fluctuations $(\sim 1/N, U)$

Fixation

- In the absence of mutations $(U = 0)$ the population becomes genetically homogeneous (monomorphic) for $t \to \infty$

- 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)

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

- Under strong selection ($N|s| \gg 1$) deleterious mutations ($s < 0$) cannot fix, while beneficial mutations ($s > 0$) fix with probability $\pi(s) = 1 - e^{-2s}$

- Mean time to fixation of a beneficial mutation: $t_{\text{fix}} \approx \ln N / s$

Mutation and fitness models

- Infinite sites approximation: Each mutation creates a new genotype

- 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 $p(s)$

- Standard choices for beneficial mutations $(s > 0)$:

$p(s) = s_b^{-1} e^{-s/s_b}$ this work — J.H. Gillespie, 1983; H.A. Orr, 2003

$p(s) = \delta(s - s_b)$ — Rouzine et al., 2003; Desai & Fisher, 2007

- House of cards model: — J.F.C. Kingman, 1978

  Fitness of offspring w' is chosen randomly and independently from a probability distribution $g(w')$

A criterion for clonal interference

C.O. Wilke, Genetics **167**, 2045 (2004)

- Probability of beneficial mutations U_b per individual and generation
- Beneficial mutations arise in the population at rate NU_b and fix with probability $\pi(s_b) \approx 2s_b$ when $s_b \ll 1$.
- Compare typical time to fixation $t_{\text{fix}} \approx \ln N / s_b$ to the time interval between fixed beneficial mutations $t_{\text{mut}} = 1/(2NU_b s_b)$
- Beneficial mutations interfere when $t_{\text{fix}} \gg t_{\text{mut}}$ or

$$2NU_b \ln N \gg 1$$

$\Rightarrow$ clonal interference is inevitable for large N **if** U_b is constant

- Deleterious mutations with probability U_d and strength s_d reduce supply of beneficial mutations by e^{-U_d/s_d} (ignored in the following)

The rate of adaptation

- Population mean fitness $\overline{w}(t) = N^{-1}\sum_i w_i(t)$

- Rate of adaptation H.A. Guess, 1974

$$R = \lim_{t\to\infty}\frac{1}{t}\langle \ln \overline{w}\rangle = \langle \ln(1+s)\rangle + \frac{1}{N}\langle \sum_i (w_i/\overline{w} - 1)\ln(w_i/\overline{w})\rangle$$

  is finite for finite N

- In general $R = E[r]\ln(1+E[s]) \approx E[r]E[s]$ C.O. Wilke, 2004

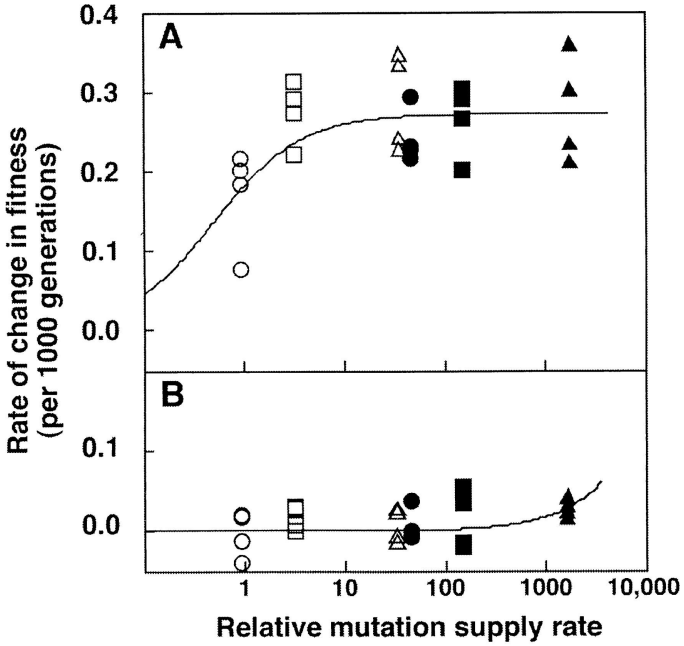
  $E[r]$: rate of substitution $E[s]$: expected selection coefficient of fixed mutations

- For small populations $E[r] = 2s_bU_bN$ and $E[s] = 2s_b \Rightarrow R = 4s_b^2U_bN$

- Clonal interference decreases $E[r]$ but increases $E[s]$

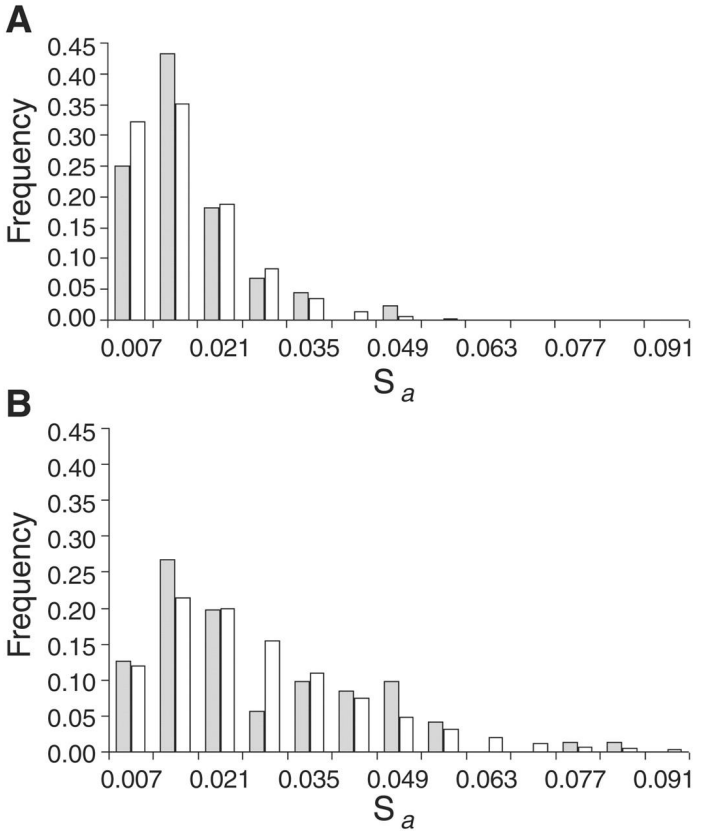
Experimental evidence for clonal interference (*E. coli*)

Rate of adaptation



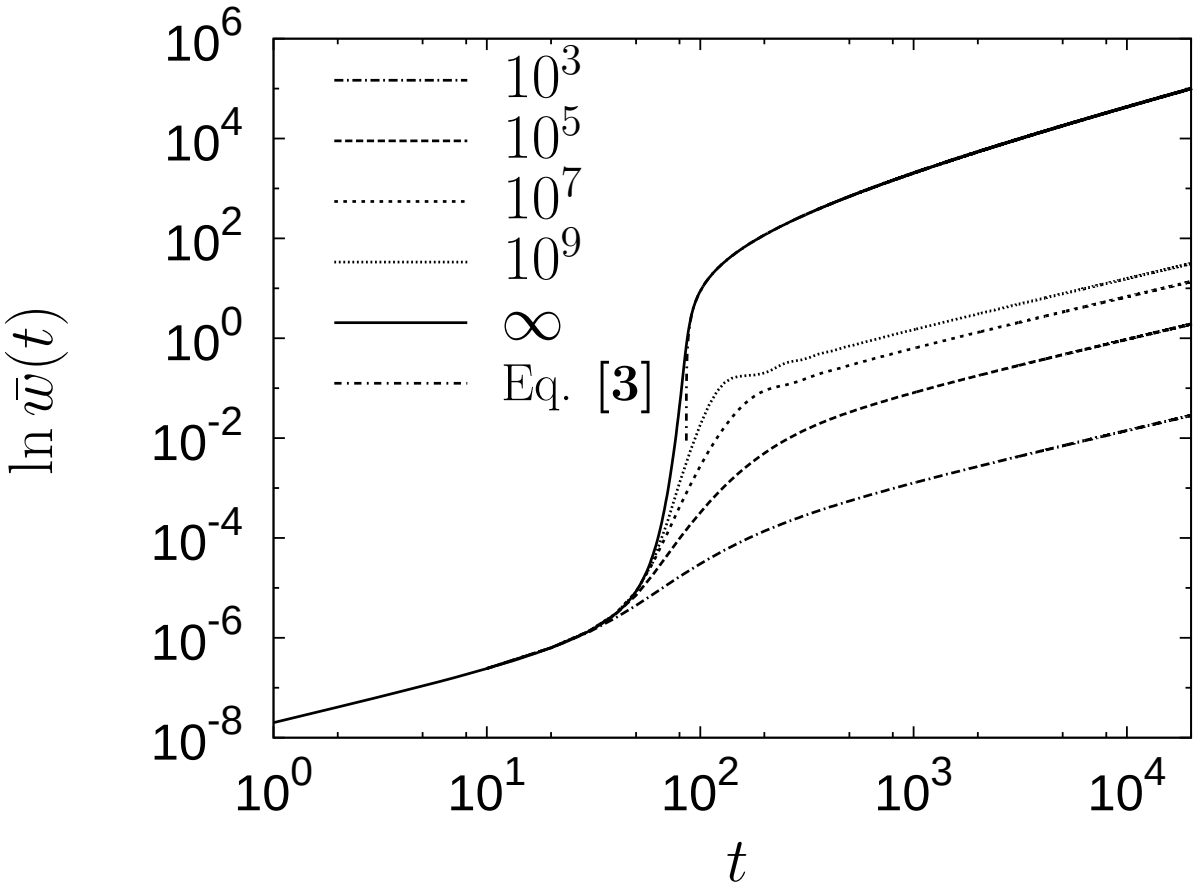
de Visser et al., Science **283** (1999)

Distribution of mutational effects



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

Finite vs. infinite populations [$U_b = 10^{-6}, s_b = 0.02$]



- Top curve: Infinite population limit

$$\ln \overline{w} \approx t[\ln(s_b t) - 1] + \frac{1}{2}\ln(2\pi U^2 t) + \frac{1}{s_b} \sim t \ln t \;\Rightarrow\; R = \infty$$

The Gerrish-Lenski theory of clonal interference

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

Fixation of a beneficial mutation requires

- Survival against genetic drift with probability $\pi(s) = 1 - e^{-2s} \rightarrow$ contenders

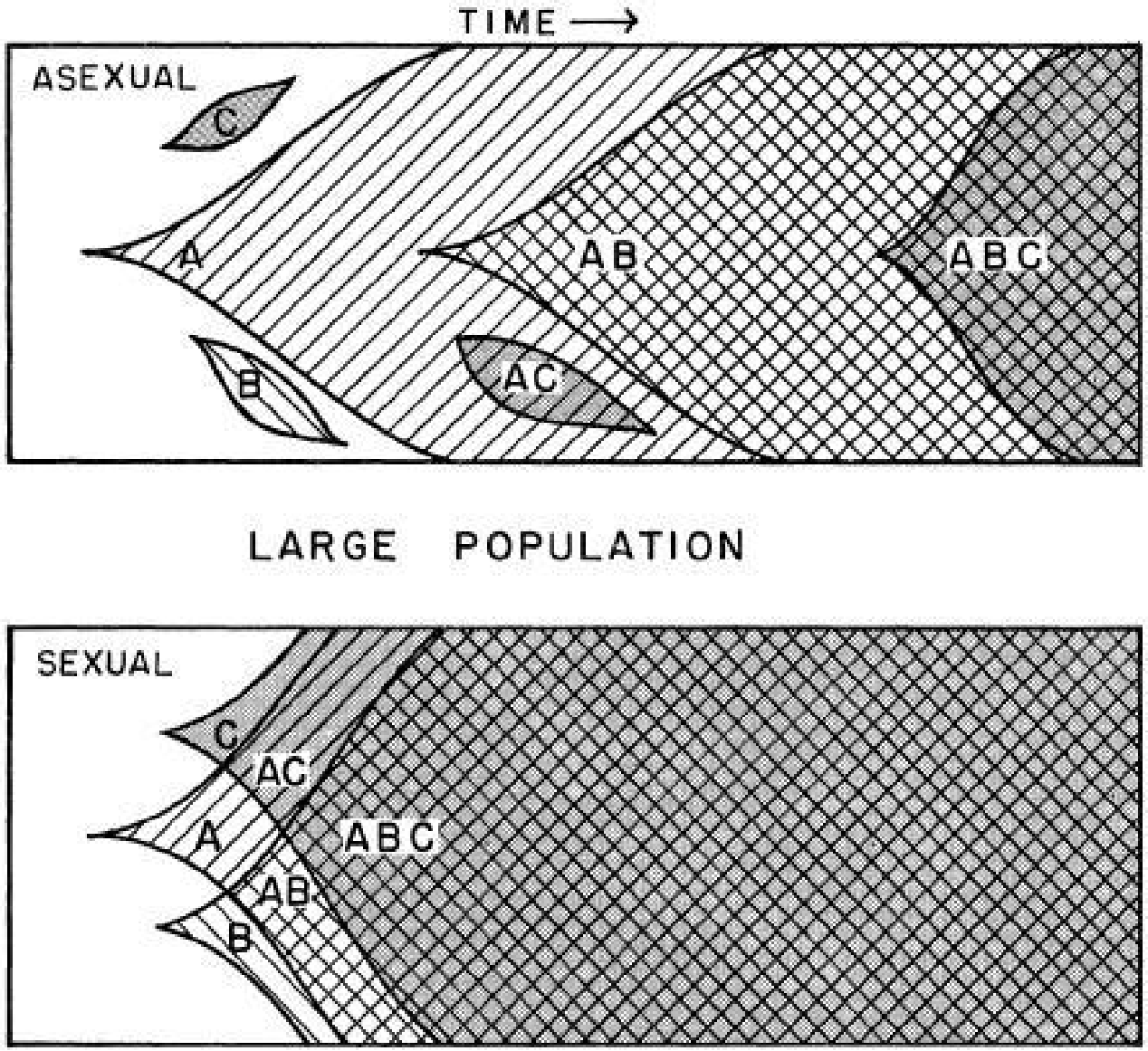
- Survival against clonal competition:
  Probability that no superior mutation s' arises and survives genetic drift during time to fixation of s is $\pi(s)\exp[-\lambda(s)]$ with

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

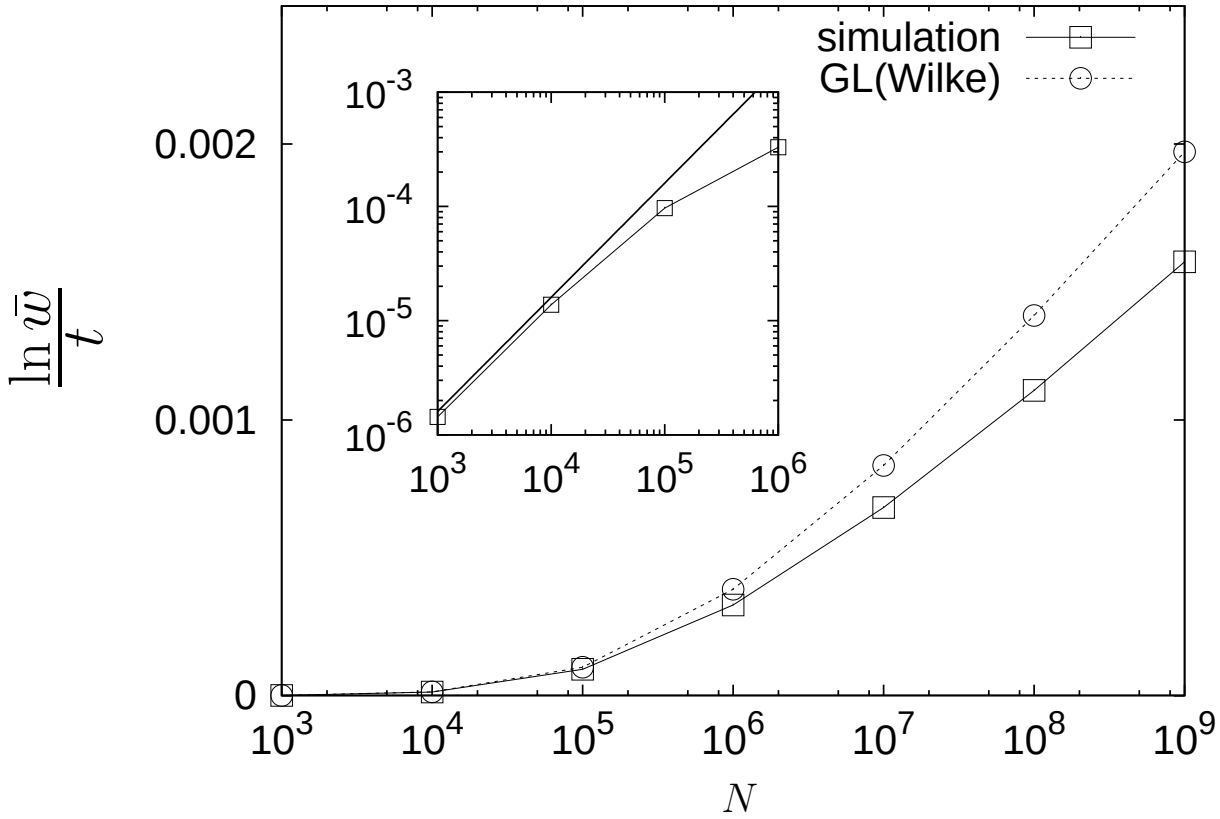
  $\Rightarrow$ analytic expression for the rate of adaptation

Key assumption of GL theory: All mutations occur relative to the current wildtype, which is replaced by fixation of the most fit of the contending mutations $\Rightarrow$ no multiple mutations, adaptation is a **renewal process**

The Gerrish-Lenski approximation illustrated



GL-theory vs. simulations: Rate of adaptation



- Transition from “periodic selection” to clonal interference at $N \approx 10^4$
- Predicted asymptotics: $R_{\mathrm{GL}} \to s_b^2 \ln(NU_b) \approx 0.0028$ at $N = 10^9$
- True asymptotics: $R \to s_b \ln(NU_b) = 50 \times R_{\mathrm{GL}}$!

Extremal statistics estimates

- Largest selection coefficient in one generation $s_{\max} = s_b \ln(NU_b)$
- Associated fixation time

$$t_{\text{fix}} \approx \frac{\ln N}{s_{\max}} \to \frac{1}{s_b} \quad \text{for } N \to \infty$$

- GL-theory suppresses multiple mutations

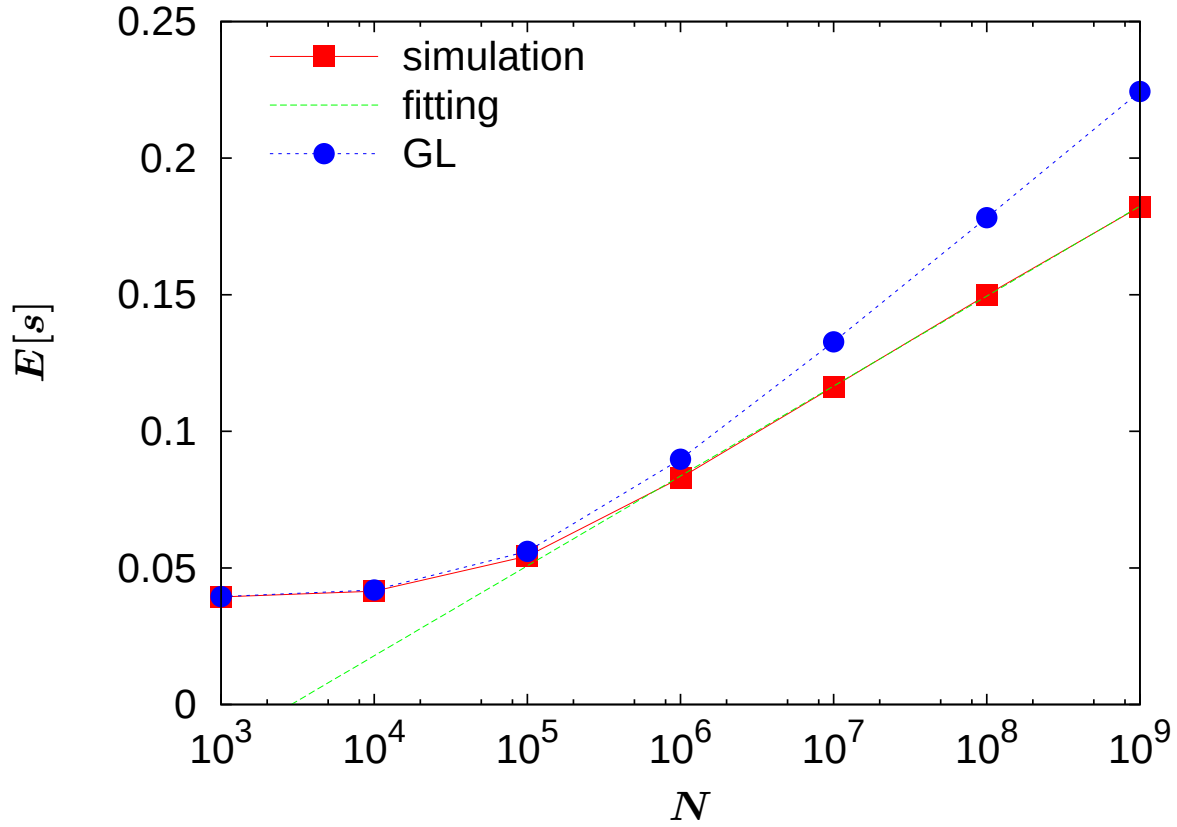
$$\Rightarrow \quad E[r] \to s_b, \quad R \to s_{\max} E[r] = s_b^2 \ln(NU_b)$$

- In the presence of multiple mutations $\lim_{N\to\infty} E[r] = \ell$

  ℓ: Maximum number of mutations per individual and generation

- Here $\ell = 1 \quad \Rightarrow \quad R \to s_{\max} = s_b \ln(NU_b)$

GL-theory vs. simulations: Mean mutational effect



- Fit: $E[s] = A \ln N - B$ with $A = 0.014$ and $B = 0.11$
- Expected asymptotics: $E[s] \to s_b \ln(NU_b) \quad \Rightarrow \quad A \to s_b = 0.02$, $B \to 0.276$

The rhythm of microbial adaptation

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

- GL-theory predicts universal, sub-Poissonian fluctuations of the number of substitution events $n_s(t)$ up to time t:

$$\frac{\langle (n_s - \langle n_s \rangle)^2 \rangle}{\langle n_s \rangle} \to 2e^{-\gamma} - 1 \approx 0.123 \quad \text{for} \quad t \to \infty \quad \text{(index of dispersion)}$$

- **But:** When mutations are not restricted to the wild-type, the notion of a substitution event becomes ambiguous, because multiple mutations can be fixed at the same time (Gillespie, 1993)

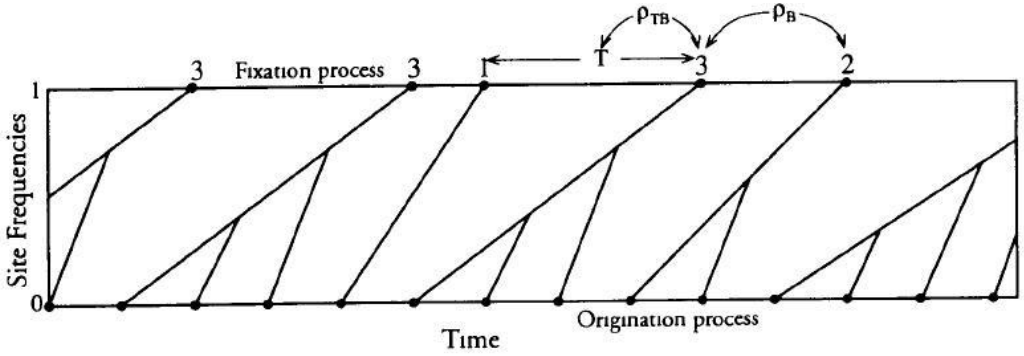
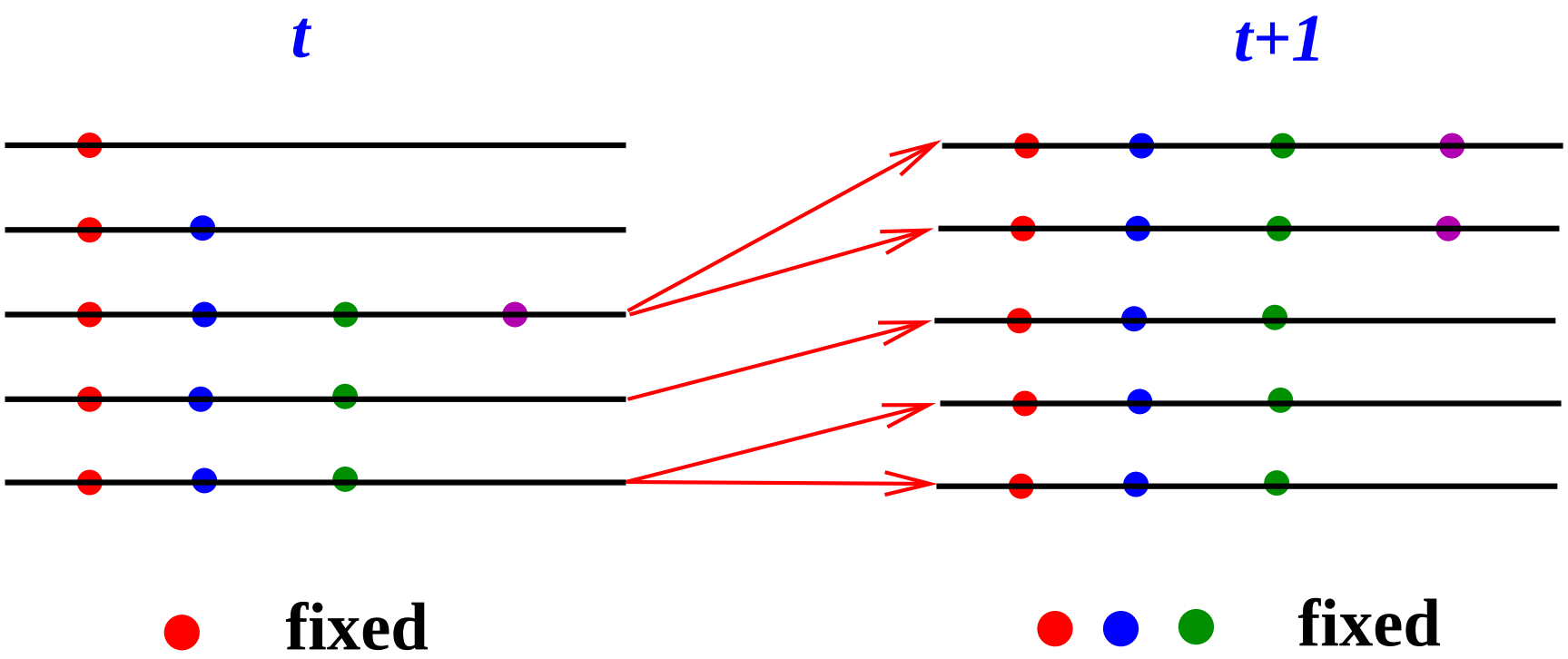


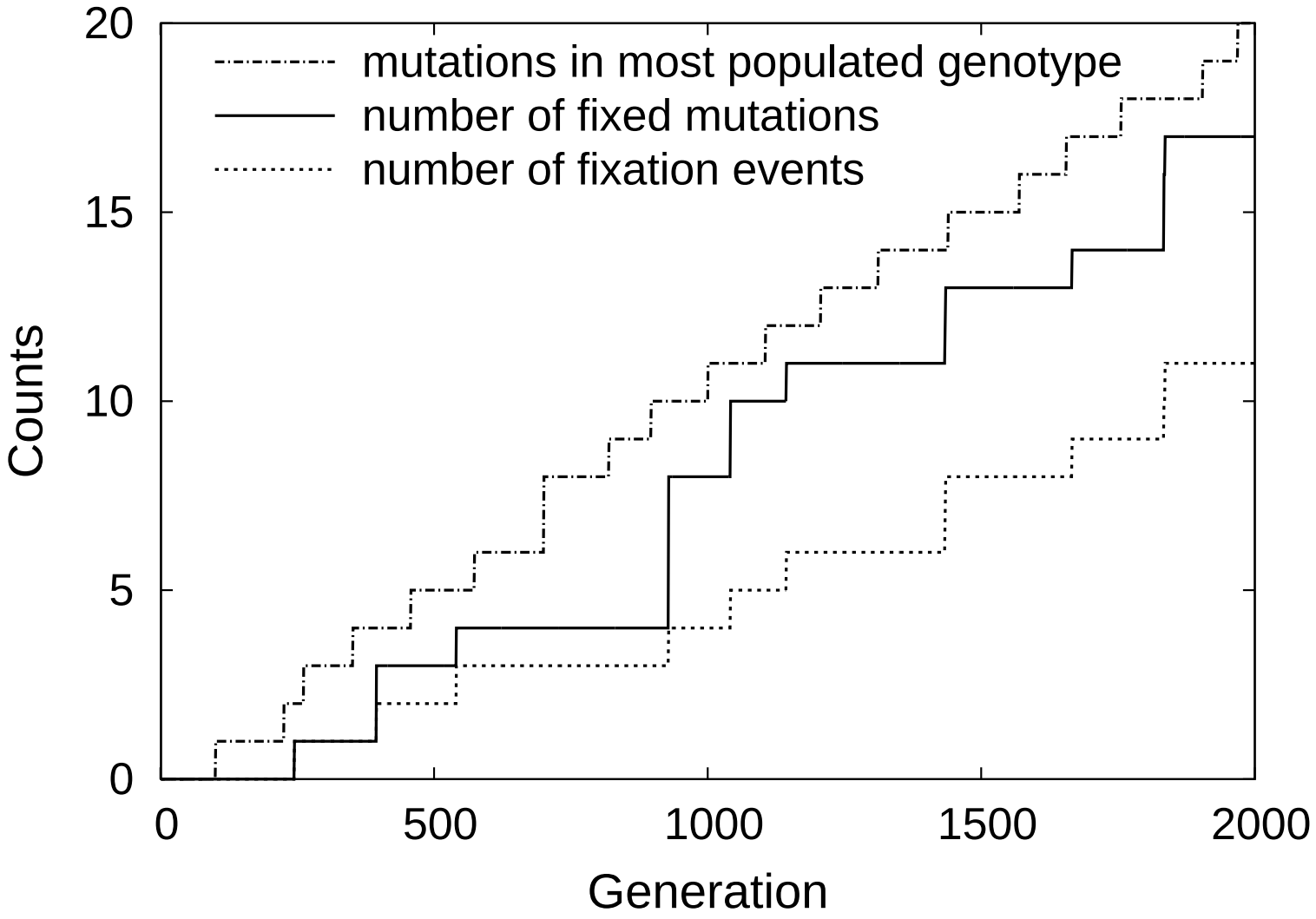
FIG. 1. A diagram of the trajectories of mutations that ultimately fix in the population.

Fixation of multiple mutations

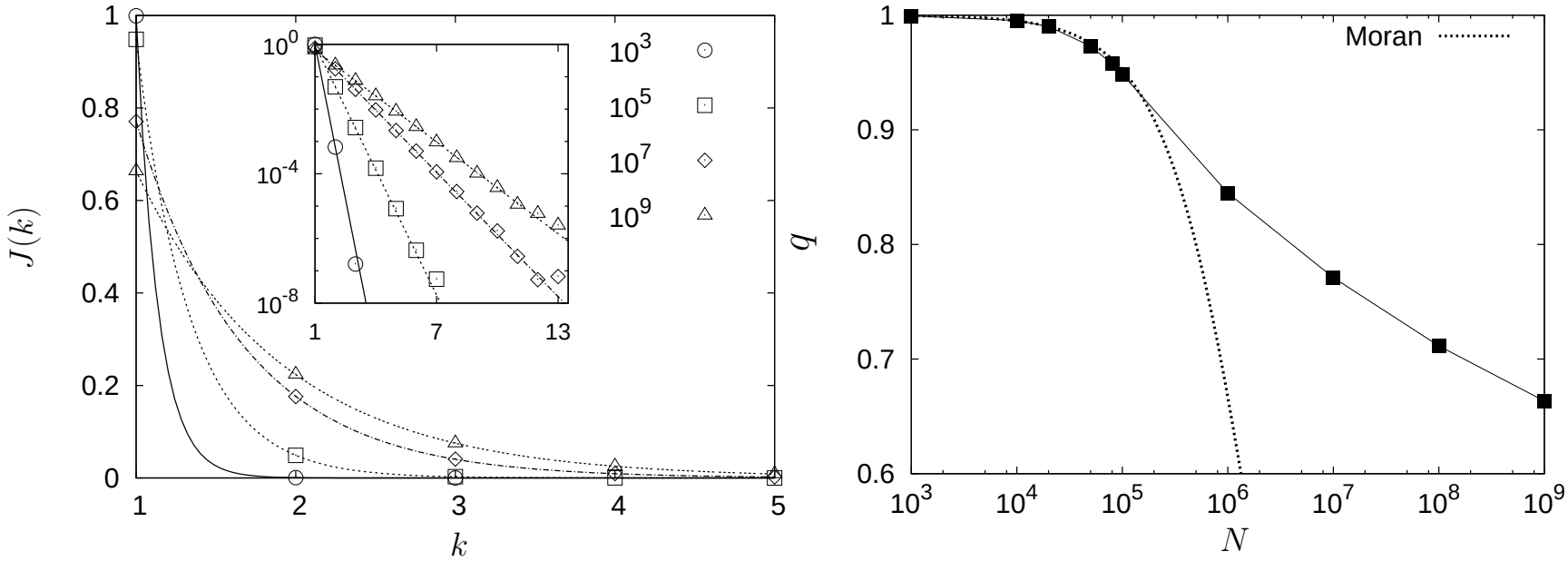
Fixation: Change in the genotype of the **most recent common ancenstor**



Mutation and fixation processes ($N = 10^9$)

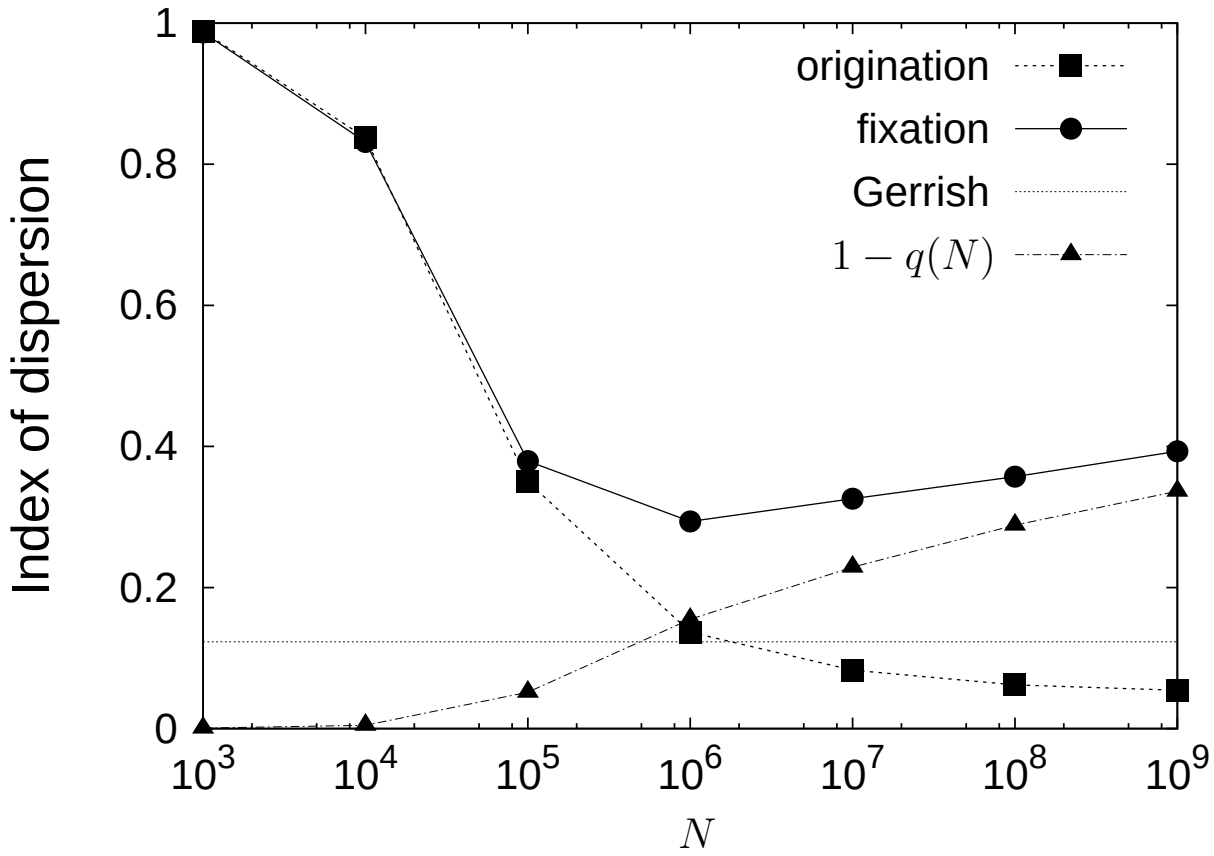


Distribution of the number of simultaneously fixed mutations



- Data are well fitted by a geometric distribution: $J(k) = q(1-q)^{k-1}$
- $1/q$: mean number of simultaneously fixed mutations, $q(N) \to 0$ for $N \to \infty$
- Geometric distribution with $q(N) = 2/(2+NU)$ is exact in the neutral case (Watterson, 1982)

The rhythm of origination and fixation



- $E[r] \to 1$: Origination process becomes regular for large N
- Index of dispersion of fixation process $\approx 1 - q \to 1$ for $N \to \infty$

The house of cards model

- Mutant fitness $w > 0$ is drawn independently and randomly from probability distribution $g(w) = e^{-w}$ $\Rightarrow$ maximally epistatic fitness landscape

- In the limit $N \to \infty$ the population fitness distribution evolves according to (Kingman, 1978)

$$f_{t+1}(w) = (1-U)\frac{w f_t(w)}{\bar{w}(t)} + U g(w)$$

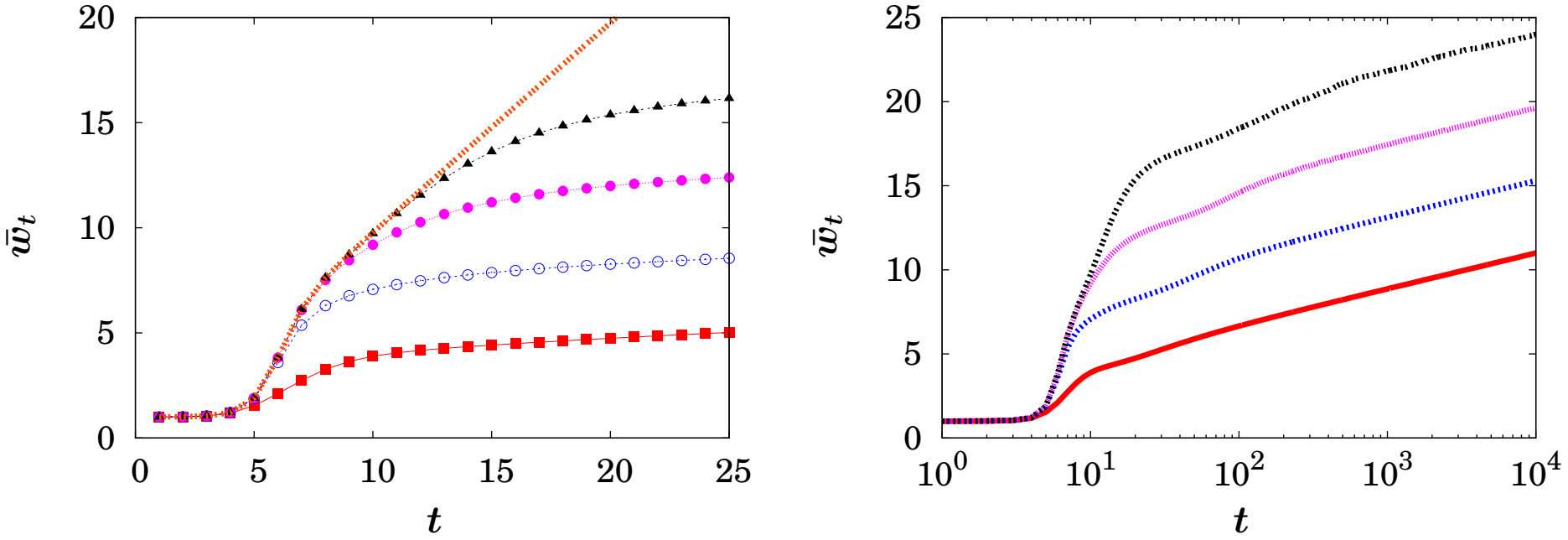
$$\Rightarrow \quad \bar{w}(t) \approx w_0(1-U)t \text{ for large } t$$

- Finite population asymptotics: $\bar{w}(t) \to (1-U)m(\tau)$ with $\tau = NUt$ and $m(\tau)$ is the solution of

$$\frac{dm}{d\tau} = \frac{C}{me^m} \text{ with } C \approx 8 \Rightarrow m \sim \ln(\tau) - \mathcal{O}(\ln(\ln\tau))$$

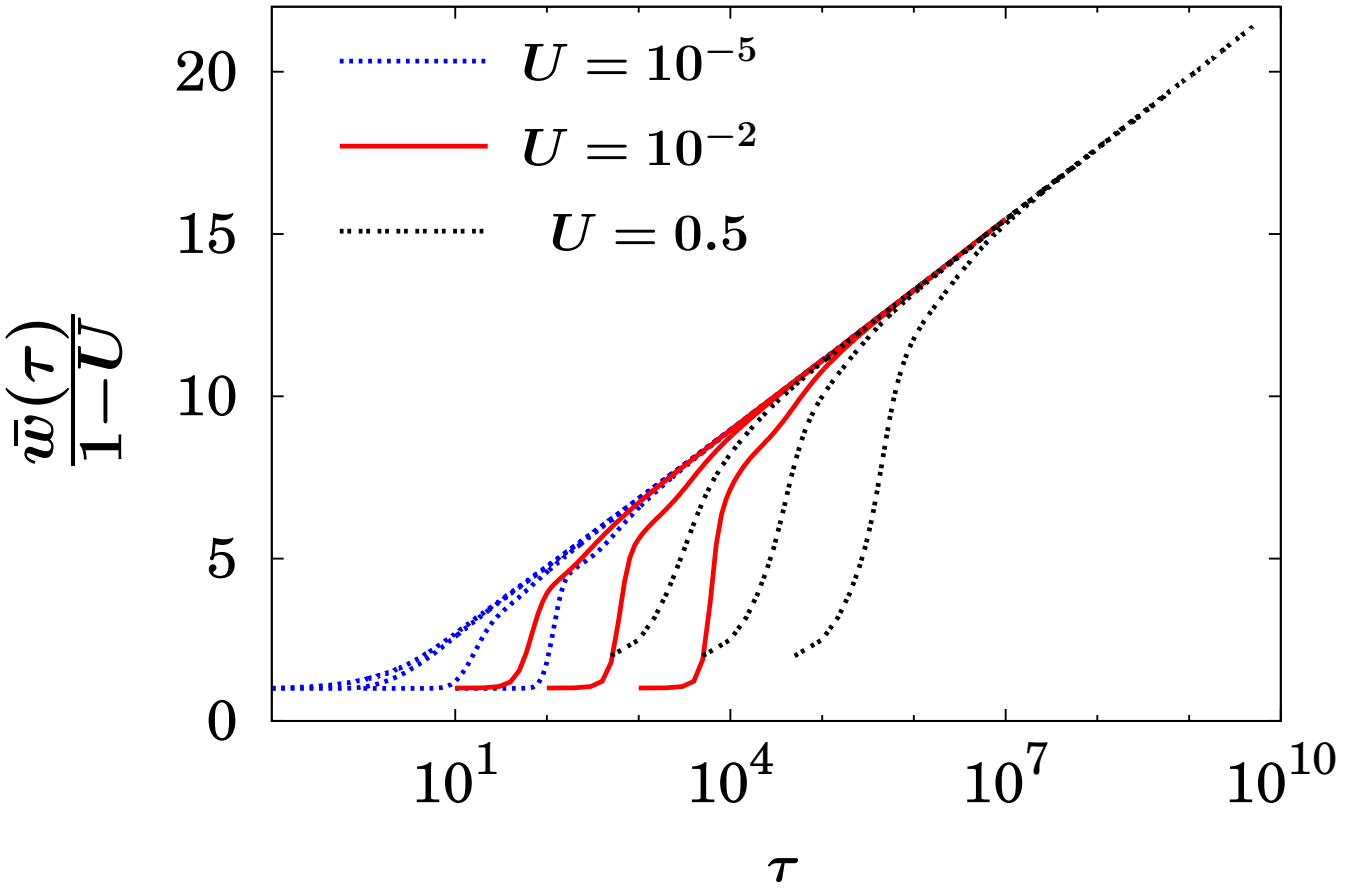
- Clonal interference is irrelevant asymptotically because $U_b \to 0$, $U_d \to U$

Finite vs. infinite populations



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

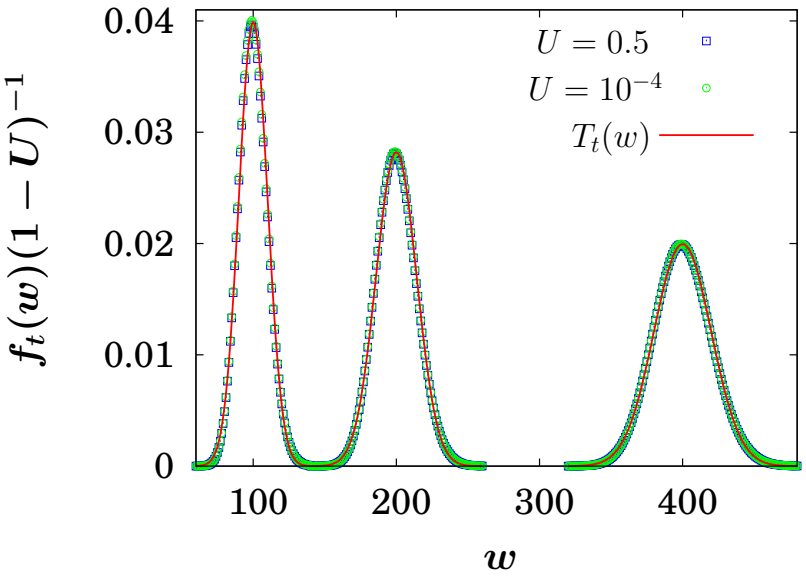
Asymptotics for finite populations



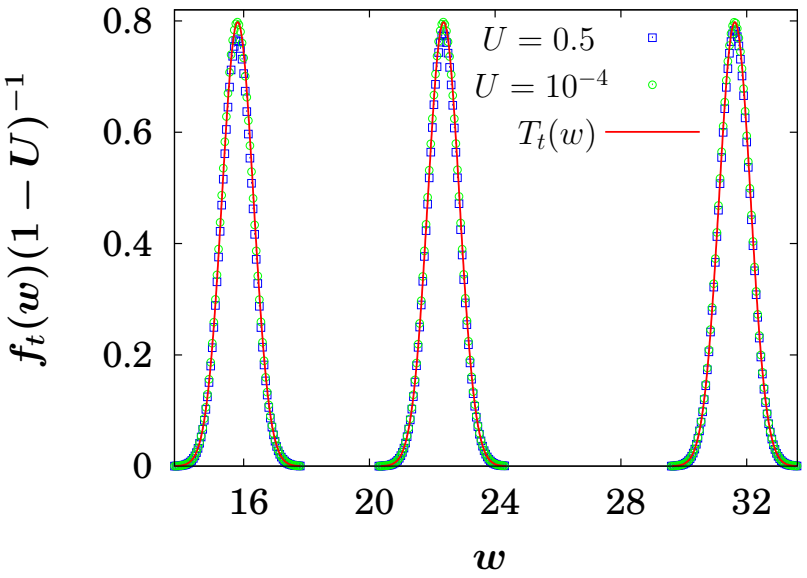
$N = 10^3, 10^4, 10^5$

Bimodality of fitness distribution ($N = \infty$)

exponential $g(w)$



Gaussian $g(w)$



- Asymptotic decomposition:

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

$T_t(w)$: broadening or sharpening “traveling wave”, independent of U

Summary

Multiplicative model

- Gerrish-Lenski theory of clonal interference works surprisingly well for reasonable population sizes
- Multiple mutations have a qualitative effect on the temporal statistics of substitution events
- How large is a large population? (in the sense of $N \to \infty$)

House of cards model

- Clonal interference is asymptotically irrelevant in a rugged fitness landscape
- Asymptotic expression for fitness available from records statistics