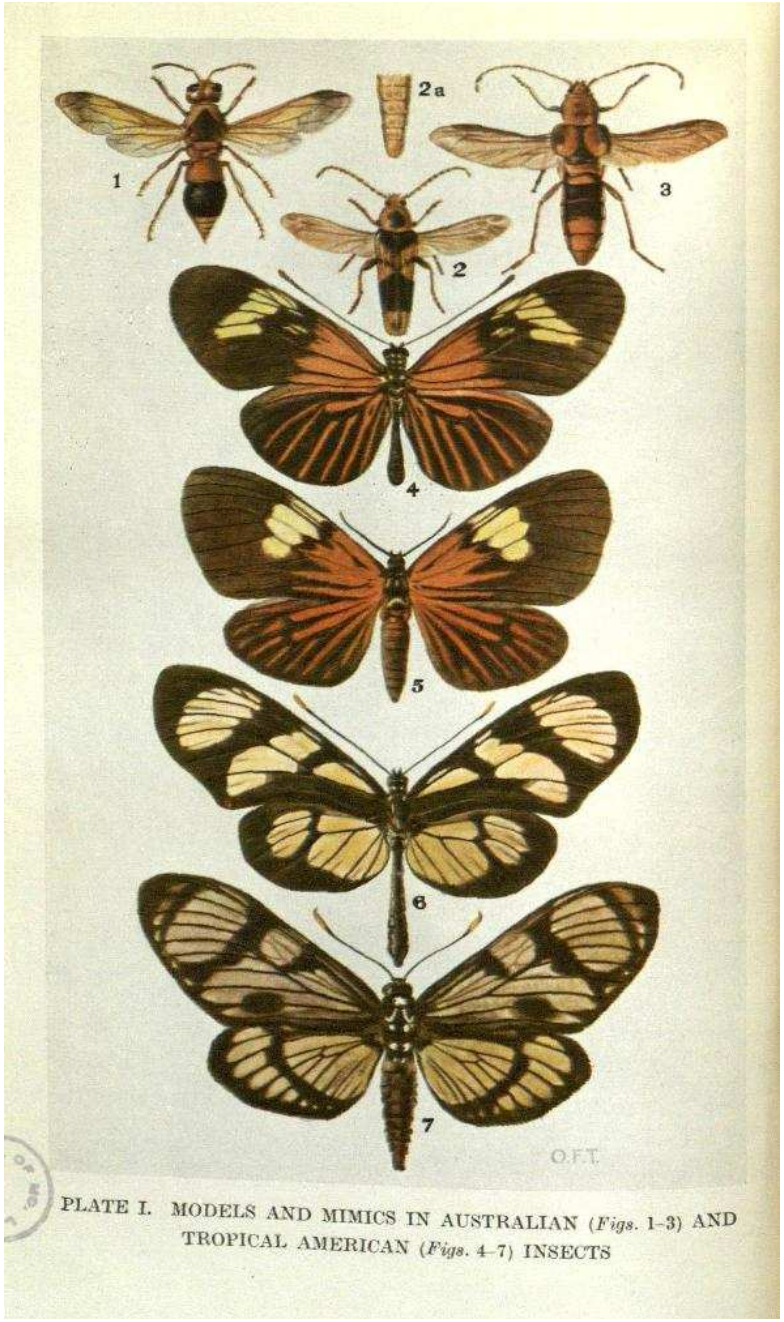


PLATE I. MODELS AND MIMICS IN AUSTRALIAN (*Figs.* 1–3) AND TROPICAL AMERICAN (*Figs.* 4–7) INSECTS

R.A. Fisher:

The Genetical Theory of Natural Selection

"Nothing in biology makes sense except in the light of evolution."

T. Dobzhansky

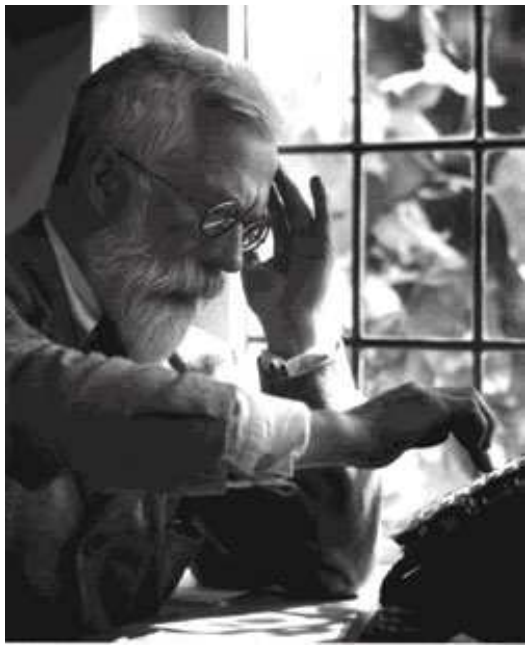
SFB 680
Molecular Basis of
Evolutionary Innovations

Evolution of sexual and asexual populations on rugged fitness landscapes

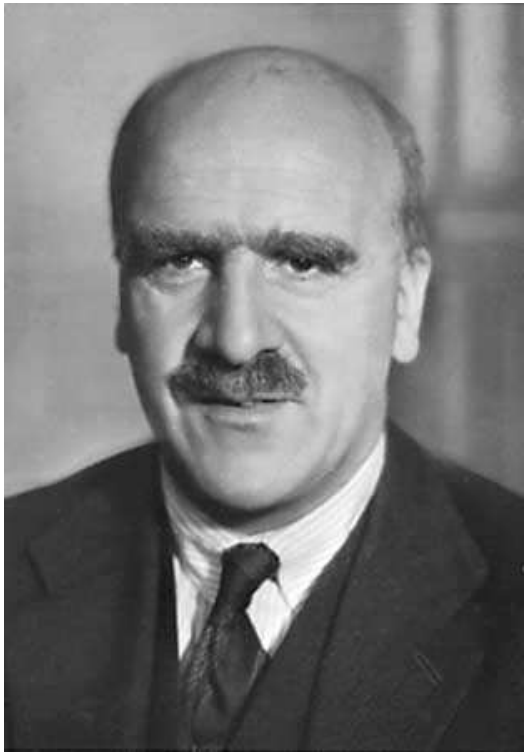
Joachim Krug
Institut für Theoretische Physik, Universität zu Köln

- Introduction to fitness landscapes
- Statistical topography of rugged fitness landscapes
  with Jasper Franke, Alexander Klözer and Arjan de Visser
- Dynamics of adaptation on rugged fitness landscapes
  with Kavita Jain and Su-Chan Park

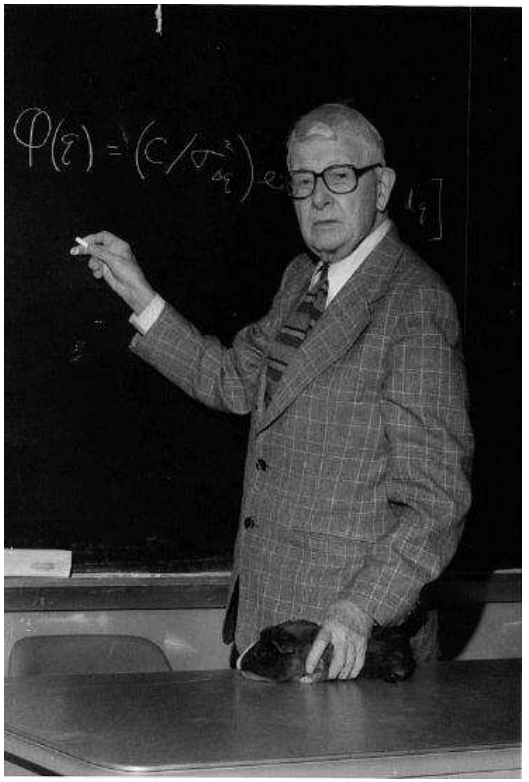
The modern evolutionary synthesis



R.A. Fisher
(1890-1962)



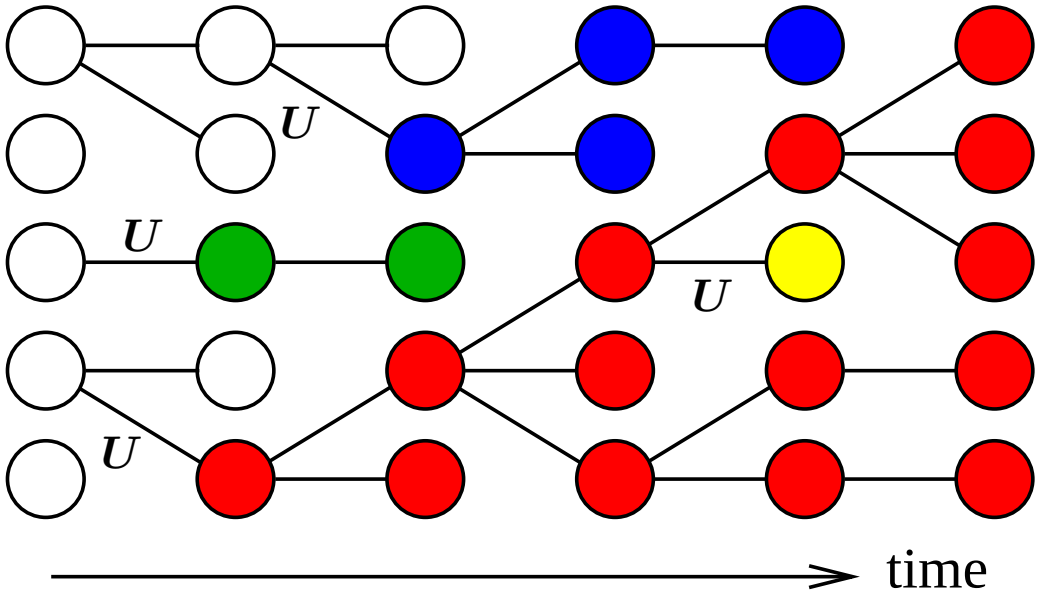
J.B.S. Haldane
(1892-1964)



S. Wright
(1889-1988)

The Wright-Fisher model for asexual reproduction

Review: S.-C. Park. D. Simon, JK, arXiv:0910.0219



- 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

Fitness landscapes

- The fitness of an individual depends on its type (= genotype or phenotype)
- Here we describe the genotype by a binary sequence

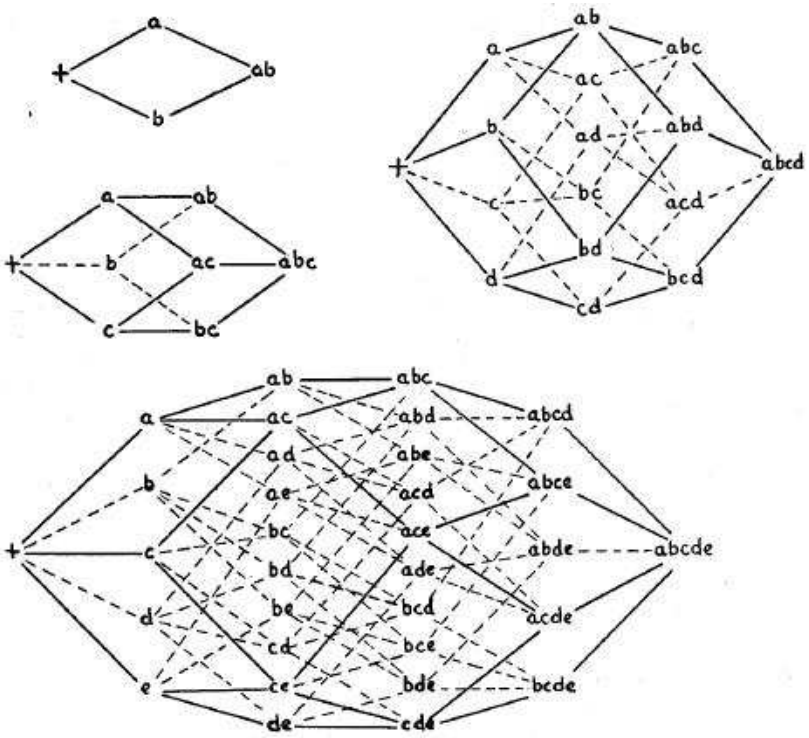
$$\sigma = (0101100...10100010101010011111)$$

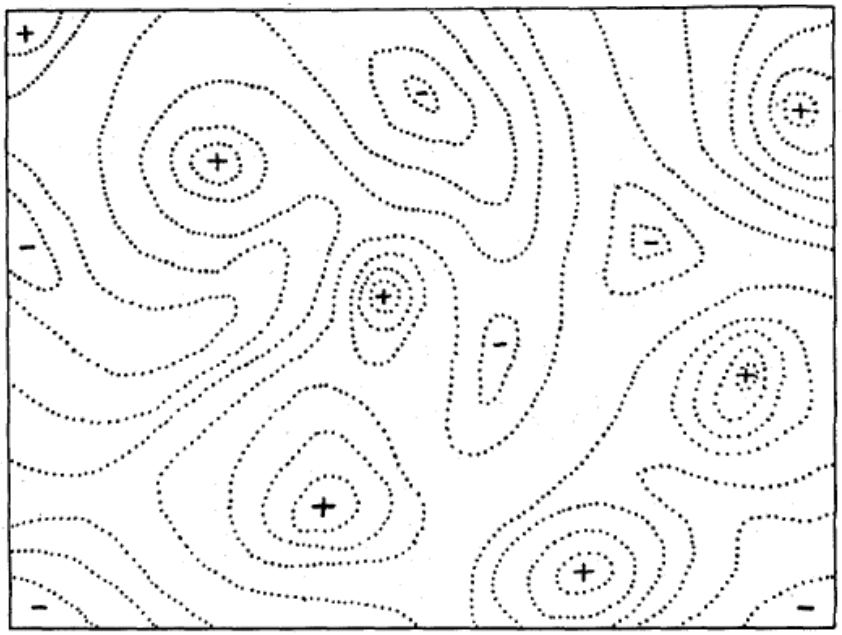
  of L loci, each of which can carry one of two alleles 0 or 1

- A fitness landscape is a function $w(\sigma)$ on the set of 2^L genotypes sequences $\sigma = (\sigma_1, ..., \sigma_L)$
- Epistasis implies interactions between different loci in their effect on fitness
- In the absence of epistatic interactions $w(\sigma) = \prod_{i=1}^{L} \omega_i(\sigma_i)$ and the fitness landscape has a unique optimum

Visualizing fitness landscapes

S. Wright, Proc. 6th Int. Congress of Genetics (1932)

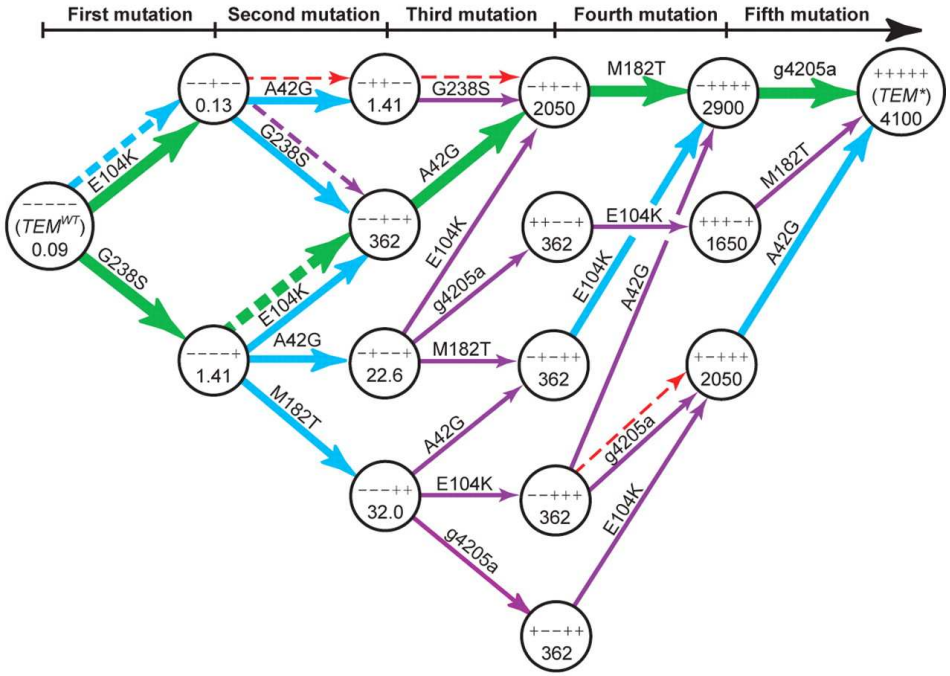



+
+
+

Empirical fitness landscapes

Example 1: The TEM1 β-lactamase resistance landscape

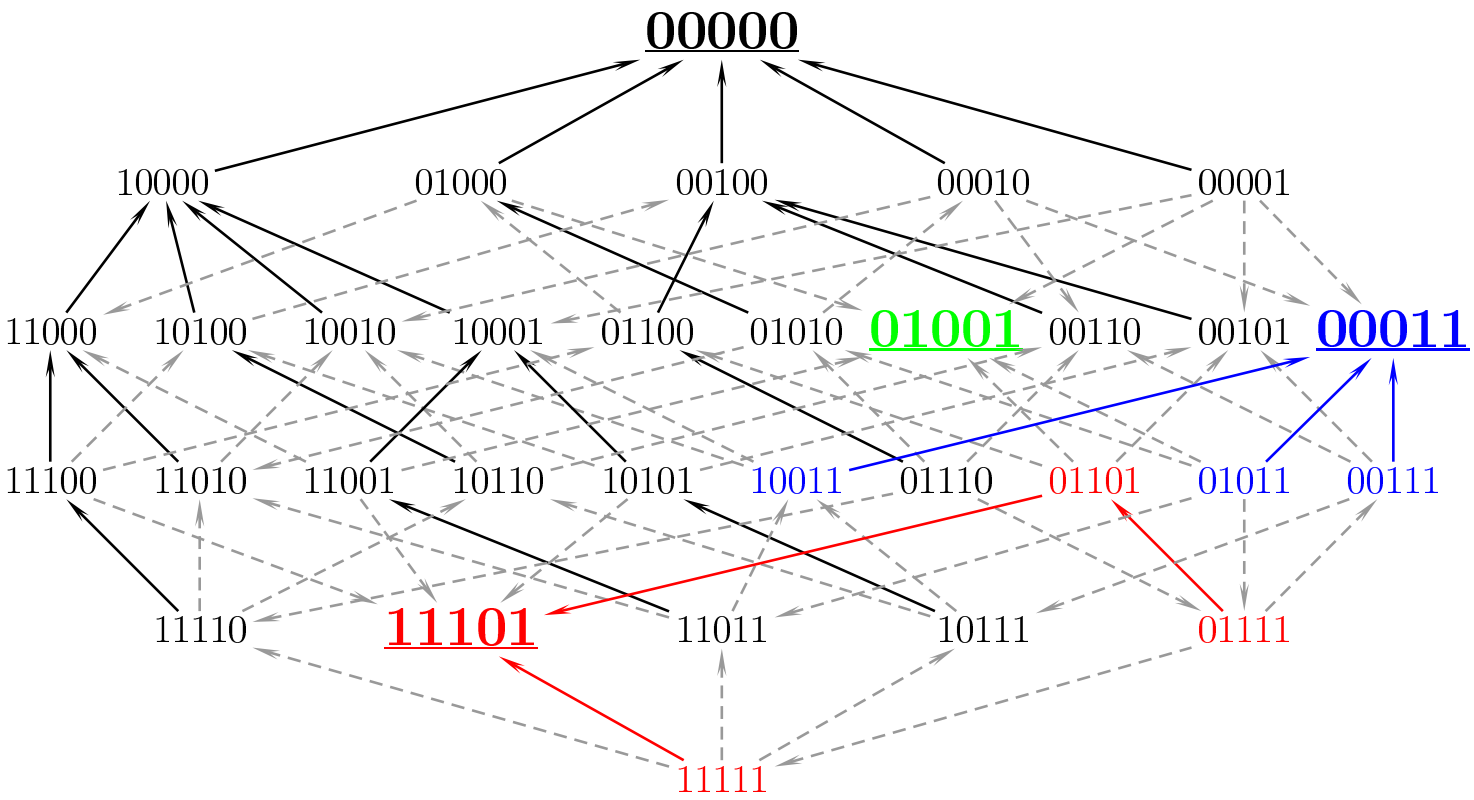
D.M. Weinreich, N.F. Delaney, M.A. De Pristo, D.L. Hartl, Science **312**, 111 (2006)



- 5 mutations in the β-lactamase enzyme confer resistance to novel antibiotic
- 102 out of L!=120 paths from the wildtype to the fivefold mutant contain downhill steps $\Rightarrow$ evolution can follow only a small number of paths and is highly predictable

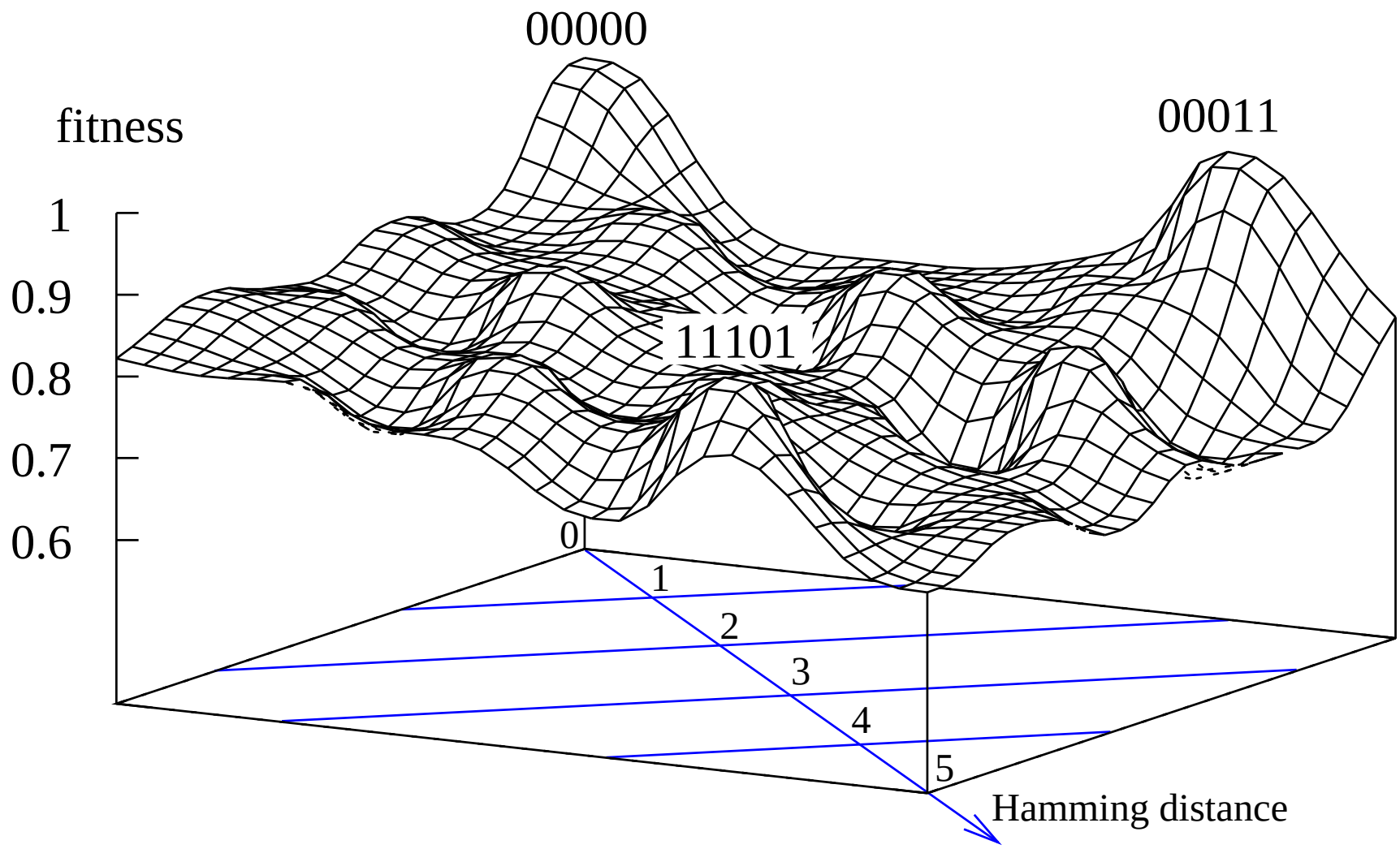
Example 2: The *Aspergillus niger* fitness landscape

J.A.G.M. de Visser, S.C. Park, JK, American Naturalist **174**, S15 (2009)



- 5 individually deleterious marker mutations in different chromosomes
- 3 local fitness optima at which populations can get stuck

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



Statistical topography of fitness landscapes

Two measures of landscape ruggedness

Number of local fitness optima

- A genotype σ is a local optimum if $w(\sigma) > w(\sigma')$ for all one-mutant neighbors σ'
- Local optima present obstacles to the adaptation of asexual and sexual populations
- In the absence of epistasis there is a single global optimum

Number of selectively accessible paths

- A path of single mutations connecting two genotypes $\sigma \rightarrow \sigma'$ with $w(\sigma) < w(\sigma')$ is selectively accessible if fitness increases monotonically along the path
- In the absence of epistasis all paths to the global optimum are accessible

The house-of-cards/random energy model

- In the house-of-cards model fitness is assigned randomly to genotypes

  Kingman 1978, Derrida 1980

- Probability of local optimum is $1/(L+1) \Rightarrow$ expected number of optima

$$\langle n_{\rm opt} \rangle = \frac{2^L}{L+1}$$

- Variance of the number of optima

  Macken & Perelson 1989

$$\langle n_{\rm opt}^2 \rangle - \langle n_{\rm opt} \rangle^2 = \frac{2^L(L-1)}{2(L+1)^2} \quad \rightarrow \quad \frac{1}{2}\langle n_{\rm opt} \rangle \quad \text{for} \quad L \rightarrow \infty$$

- Fluctuations are sub-Poissonian because optima cannot be nearest neighbors

- Distribution of $n_{\rm opt}$ is asymptotically normal

  Baldi & Rinott 1989

The number of accessible paths in the house-of-cards model

- Relabel genotypes such that the global optimum is located at $\sigma^{(0)} = ..0000....$

- What is the mean number of shortest, selectively accessible paths $n_{\rm acc}$ from the antipodal genotype $\sigma^{(1)} = ...11111...$ to $\sigma^{(0)}$?

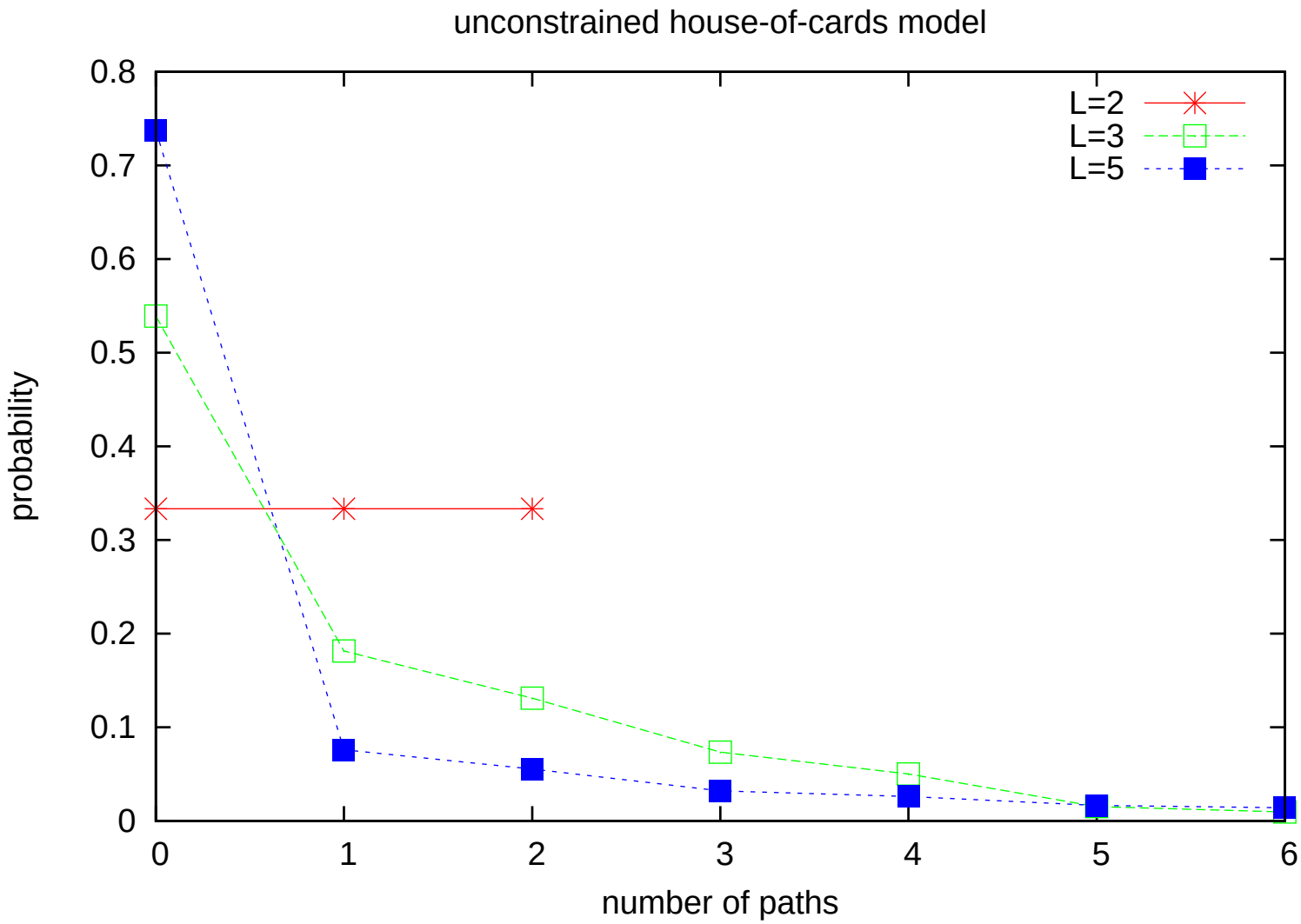
- A given path consists of L independent, identically distributed fitness values $w_1, ..., w_L = w(\sigma^{(1)})$, and is accessible iff

$$w_1 > w_2 .... > w_L$$

- Since all $L!$ permutations of the L random variables are equally likely, the probability $p_{\rm acc}$ for this event is $1/L!$

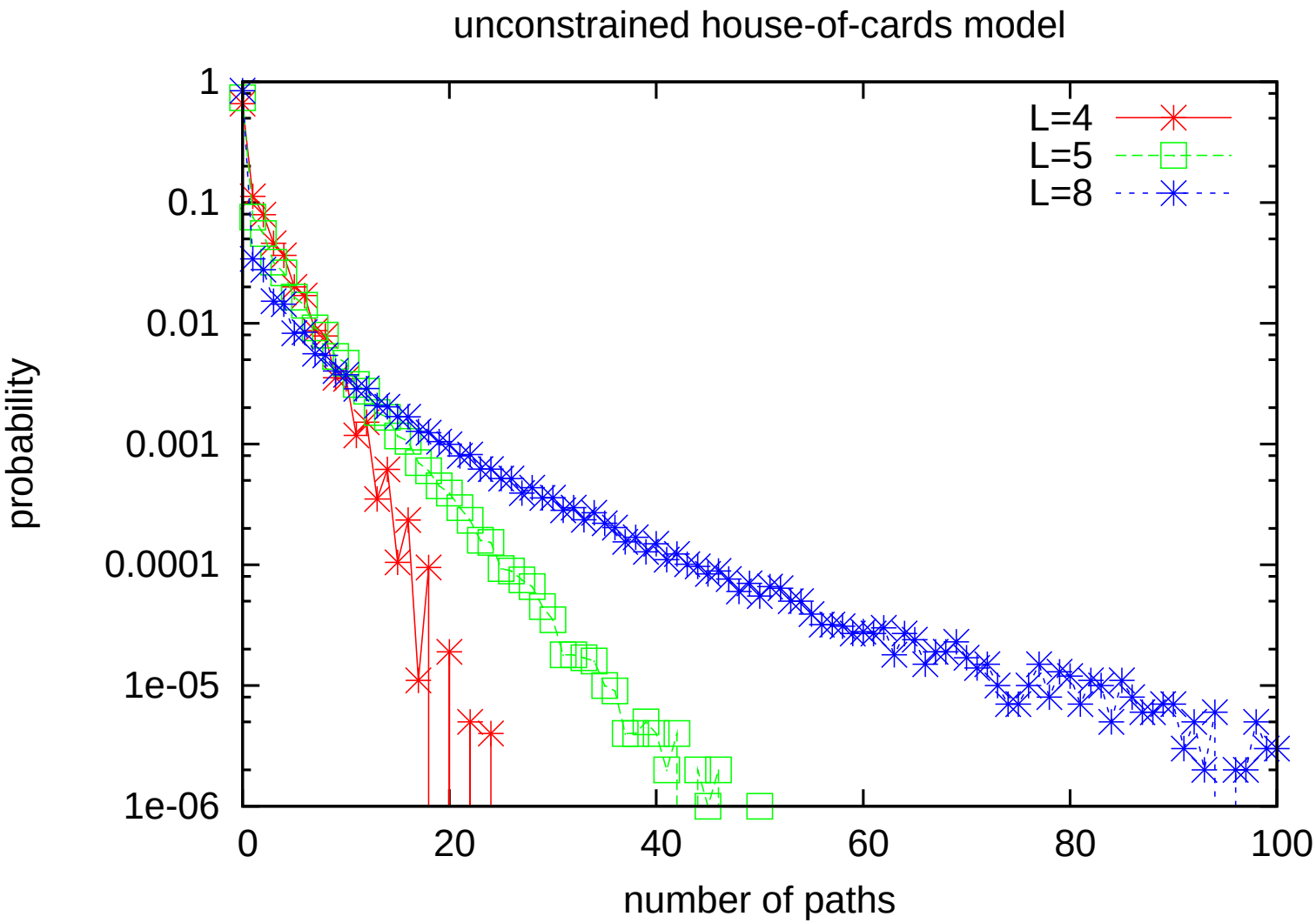
$$\Rightarrow \langle n_{\rm acc} \rangle = p_{\rm acc} \times L! = \frac{1}{L!} \times L! = 1$$

Distribution of the number of accessible paths



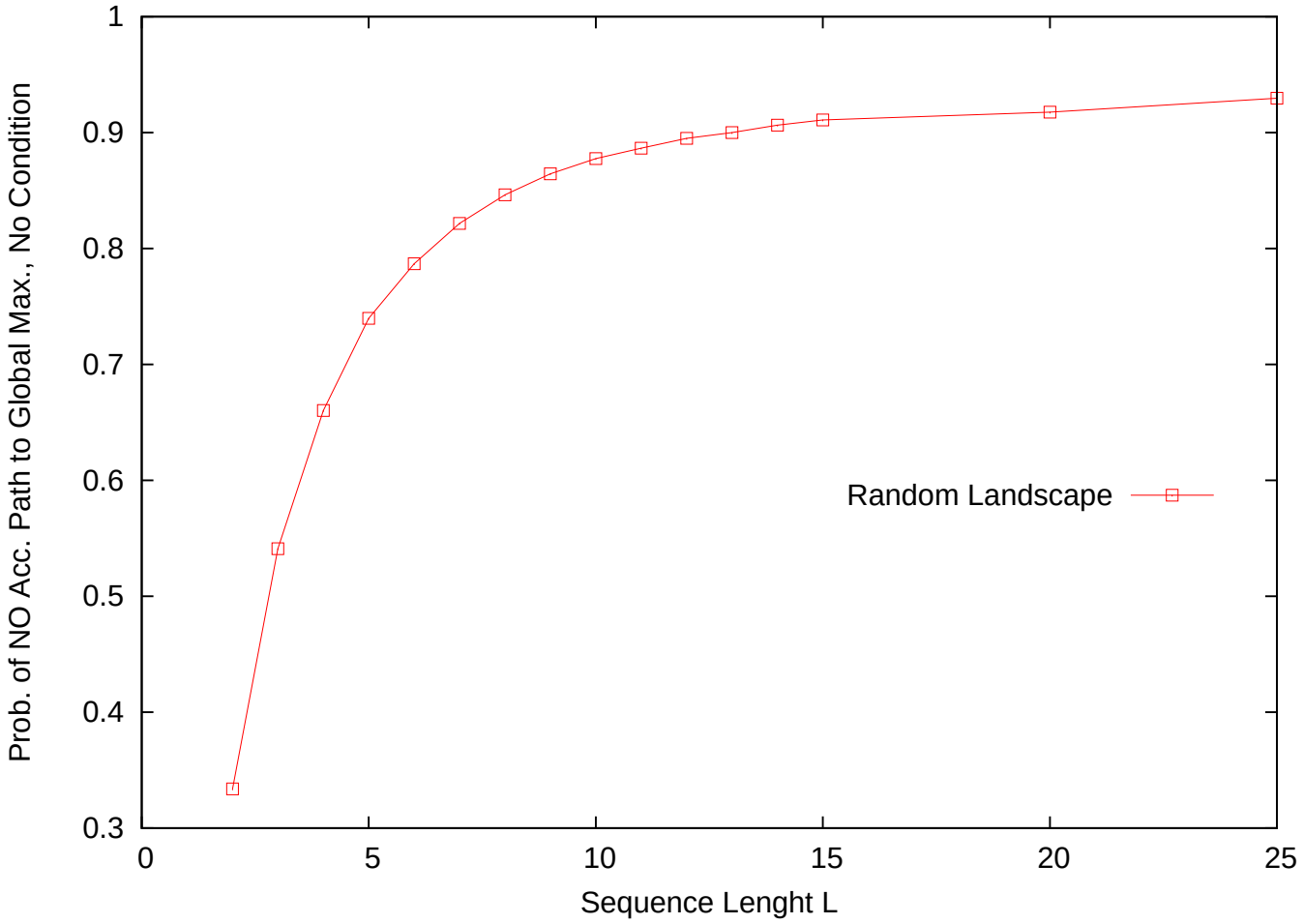
- Condensation of probability at $n_{\mathrm{acc}} = 0$

Distribution of the number of accessible paths



- Roughly exponential decay for large n_{acc}

The probability of no accessible path



- Conjecture: $\lim_{L\to\infty} P_{\mathrm{acc}}^{(L)}(0) = 1$

Models with tunable ruggedness

Kauffman's LK-model

Kauffman & Levin 1987

- Each locus interacts randomly with $K \leq L-1$ other loci:

$$\ln w(\sigma) = \sum_{i=1}^{L} f_i(\sigma_i | \sigma_{i_1}, ..., \sigma_{i_K})$$

f_i: Uncorrelated RV's assigned to each of the 2^{K+1} possible arguments

- $K = 0$: Non-epistatic $\quad K = L-1$: House-of-cards

Rough Mt. Fuji landscapes

Aita et al. 2000

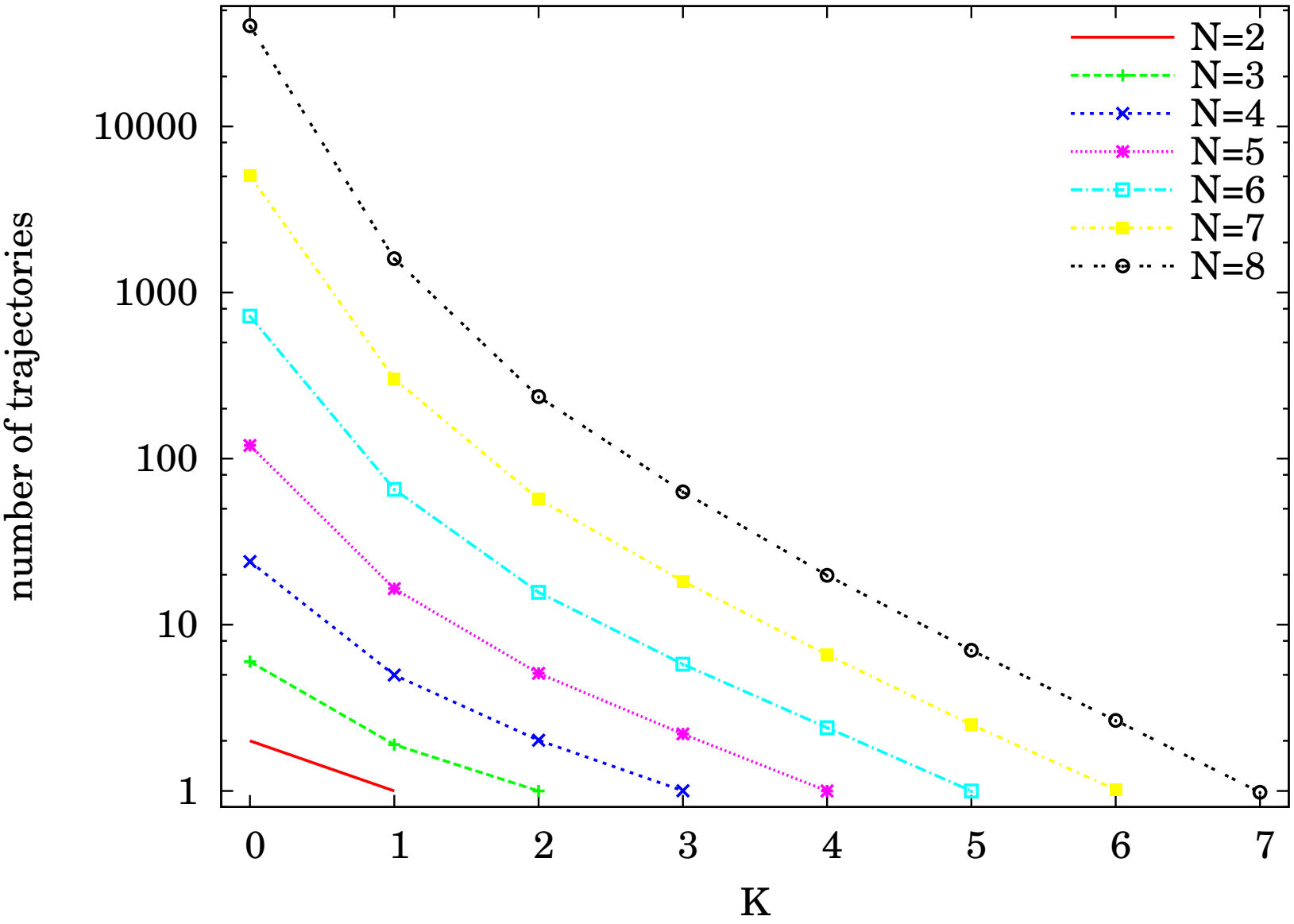
- Non-epistatic ("Mt. Fuji") landscape perturbed by a random component:

$$\ln w(\sigma) = -cd(\sigma, \sigma^{(0)}) + \eta(\sigma)$$

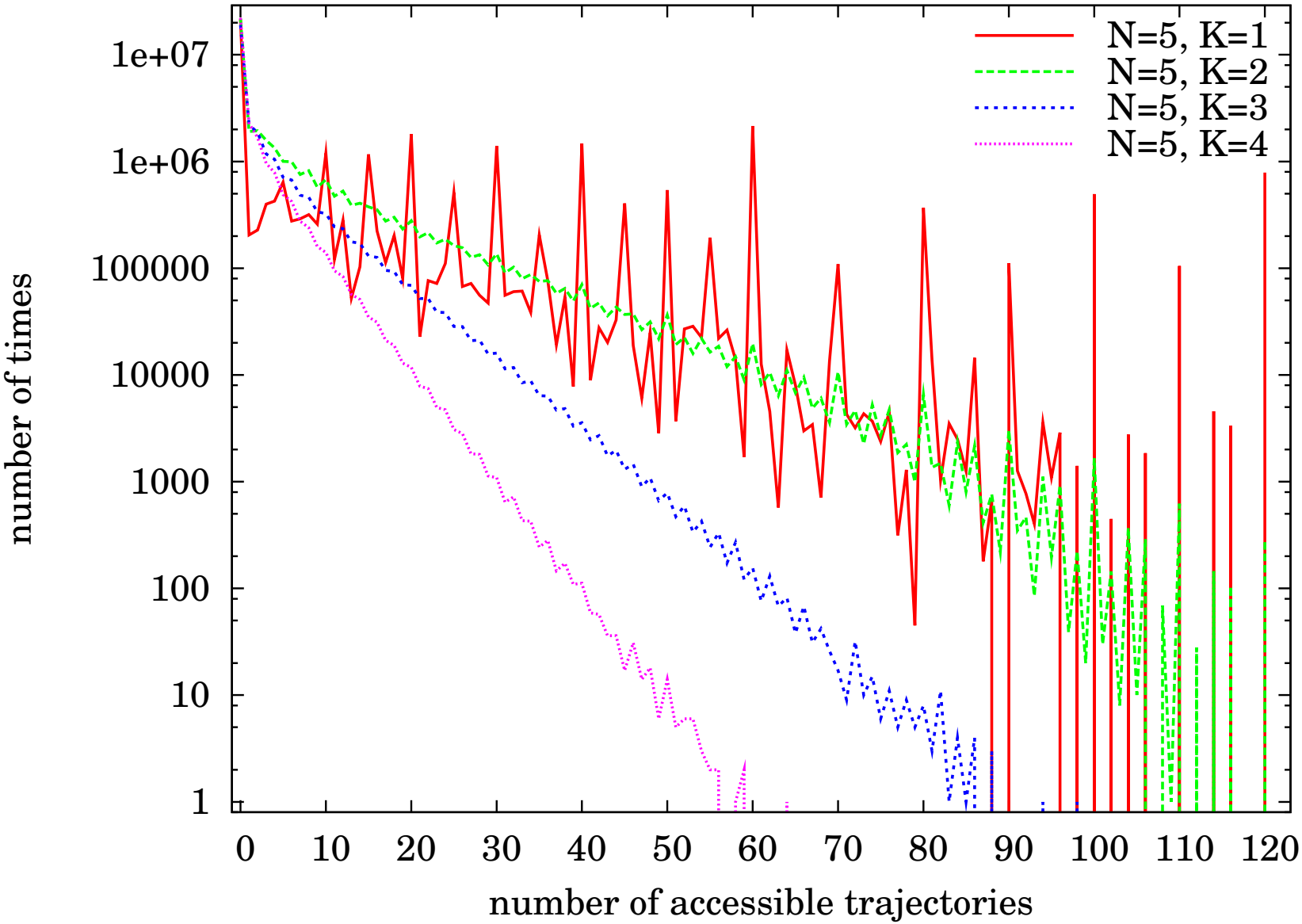
η: (Gaussian) RV's with unit variance $\quad d(\sigma, \sigma')$: Hamming distance

- $c = 0$: House-of-cards $\quad c \to \infty$: Non-epistatic

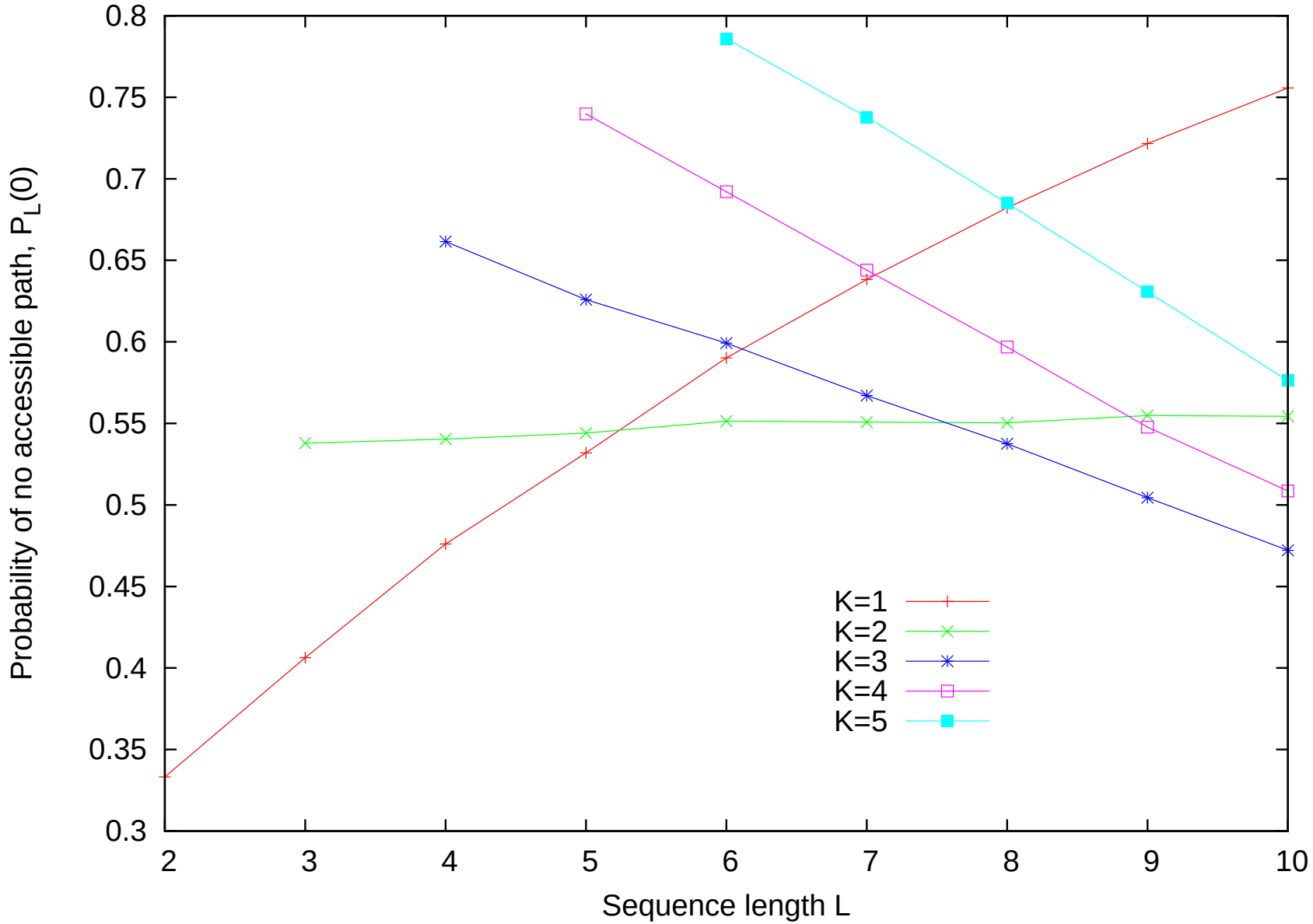
Mean number of paths in the Kauffman model



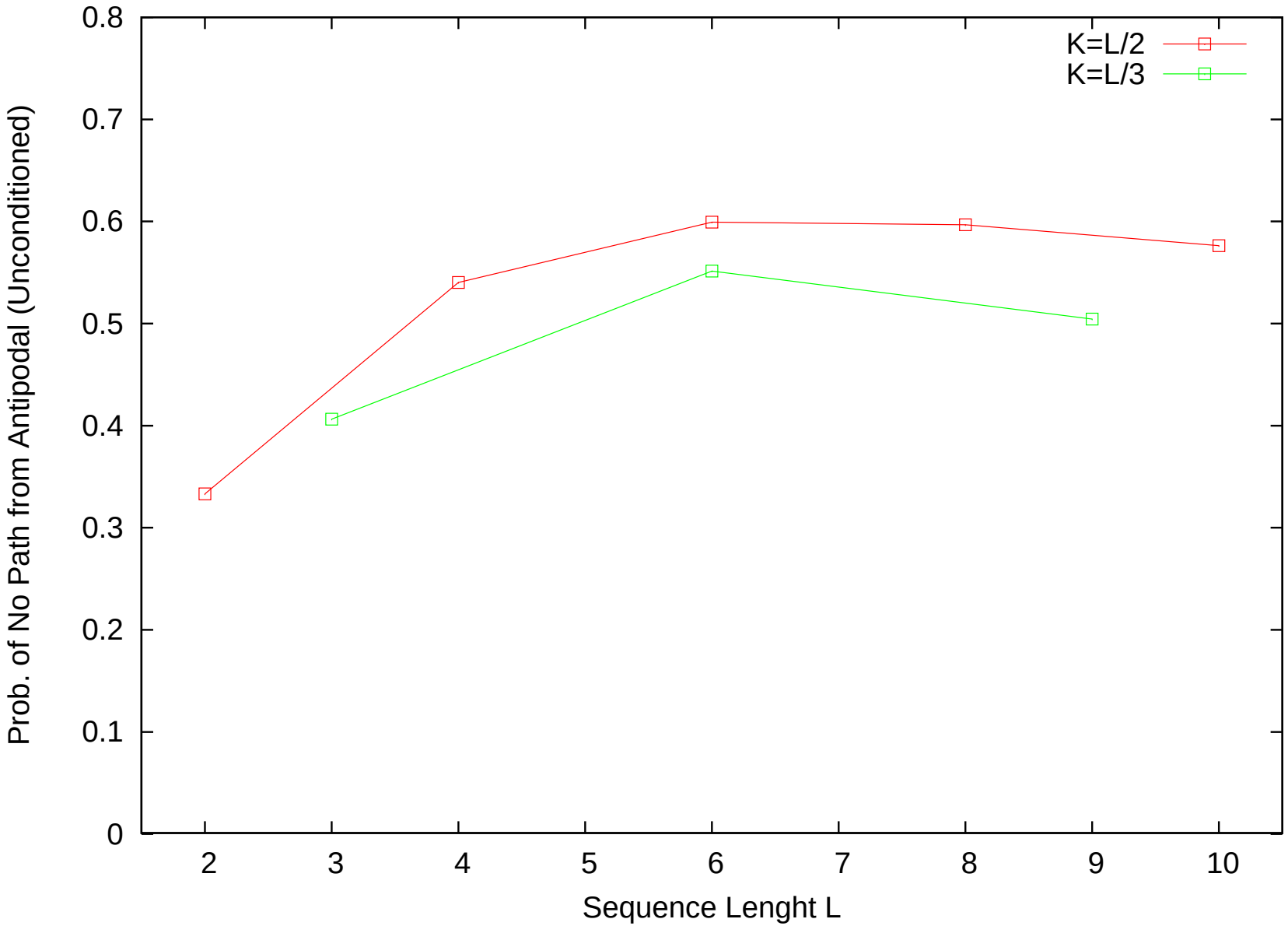
Distribution of the number of paths in the Kauffman model (L=5)



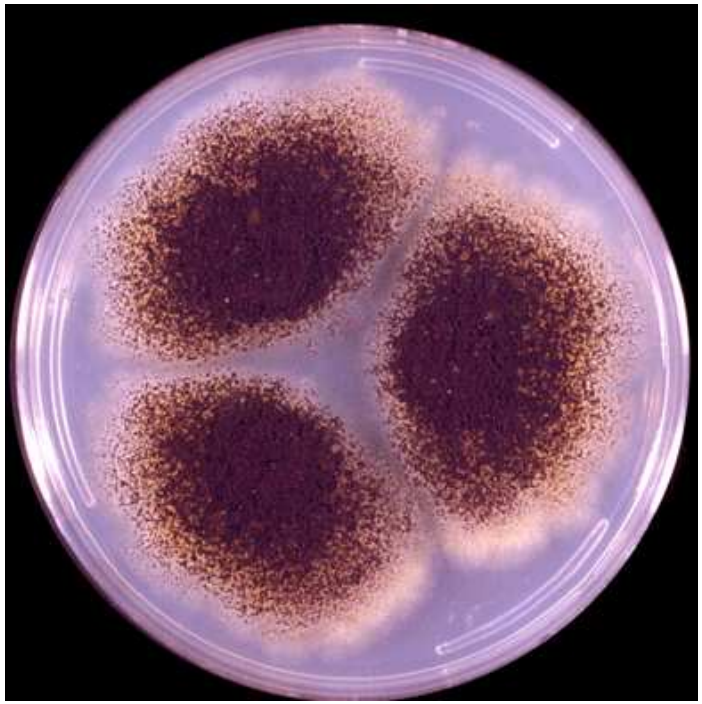
Probability of no adaptive path in the Kauffman model

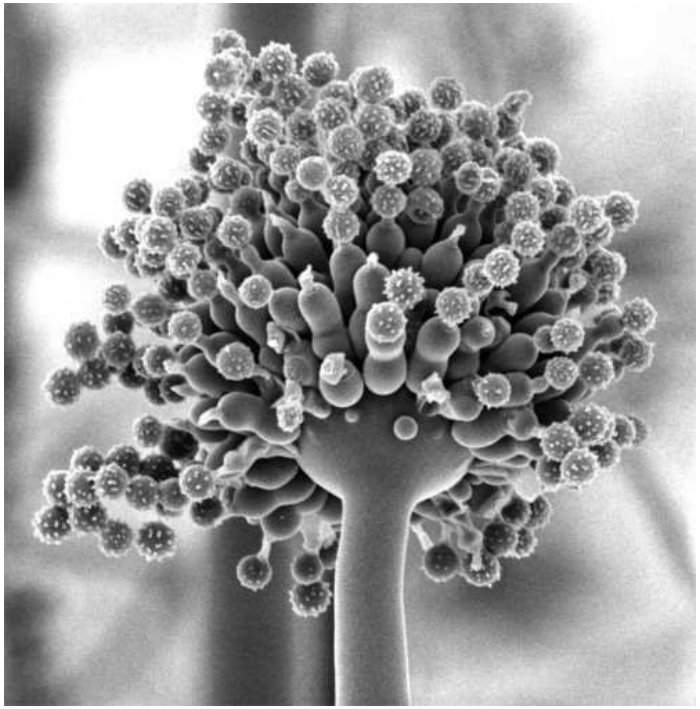


Probability of no adaptive path in the Kauffman model



Application to an empirical fitness landscape





Aspergillus niger (courtesy of Mycology Online & N.D. Read)

Empirical fitness landscape for *Aspergillus niger*

- 8 marker mutations residing in different chromosomes
  (1 spore color mutation, 5 auxotrophies, 2 resistances)

- 186 out of $2^8 = 256$ possible combinations were isolated among ~ 2500 strains
  J.A.G.M. de Visser et al., Evolution **51**, 1499 (1997)

- Fitness (= growth rate) was measured for two replicates per strain

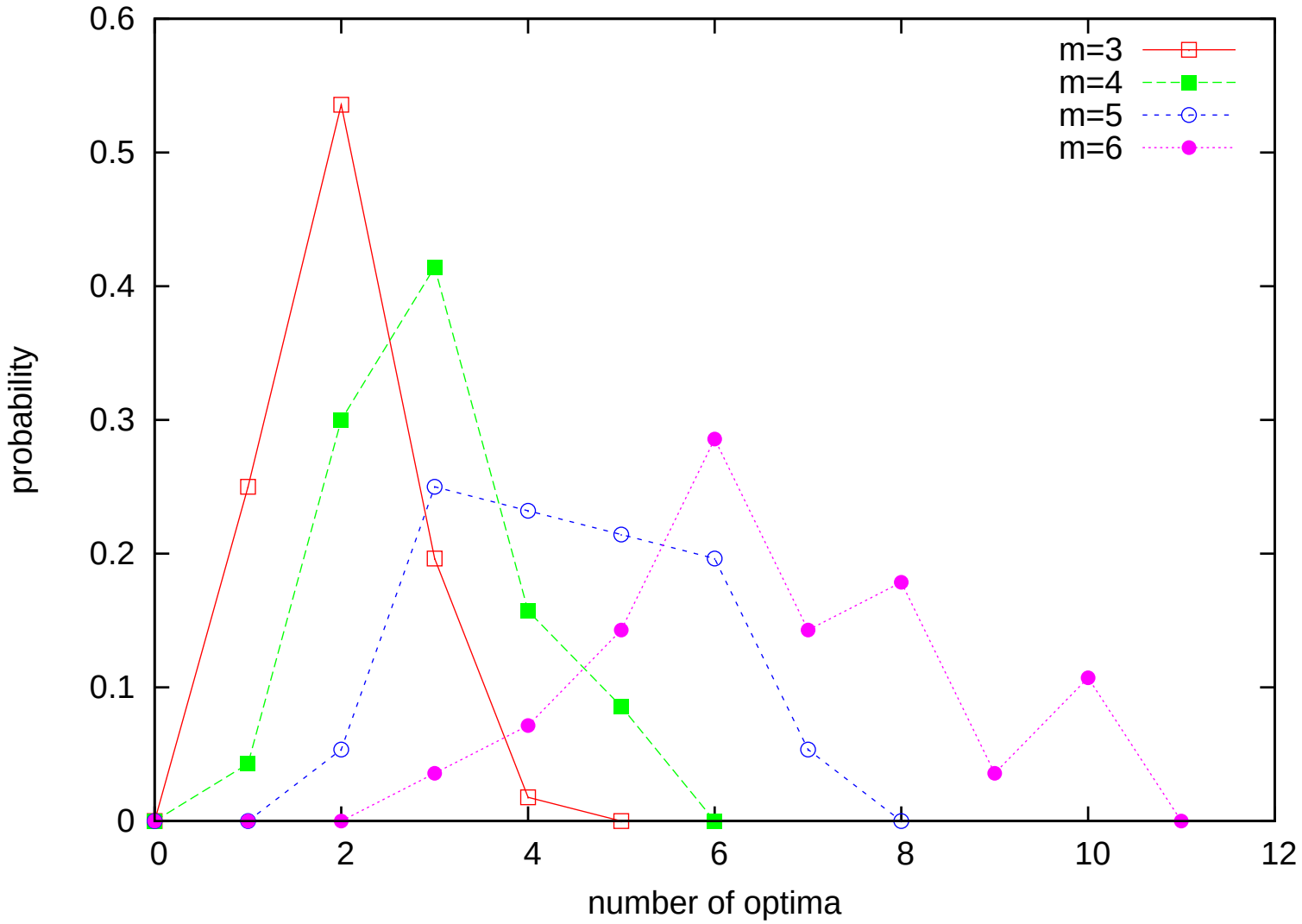
- Data includes 2 complete sets of $2^5 = 32$ combinations analyzed previously
  J.A.G.M. de Visser et al., Am. Nat. **174**, S15 (2009)

- Here we analyze the complete data set and assign fitness 0 to missing strains ("lethals")

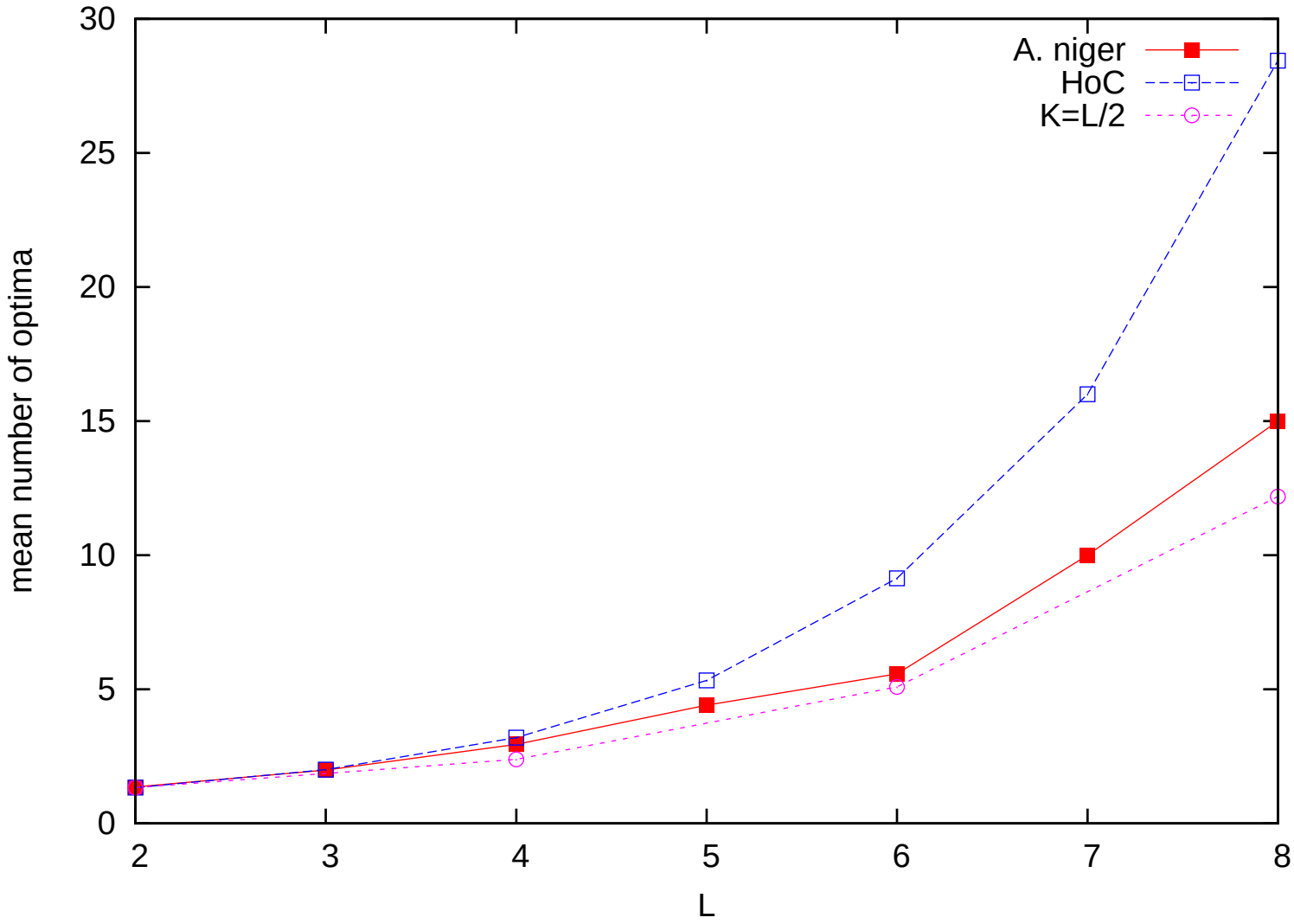
- Effects of lethals are relatively minor

- **Key idea:** Treat the $\binom{8}{m}$ subgraphs including all possible combinations of $m = 2, 3, 4, 5, 6$ mutations as ensembles of m-locus fitness landscapes

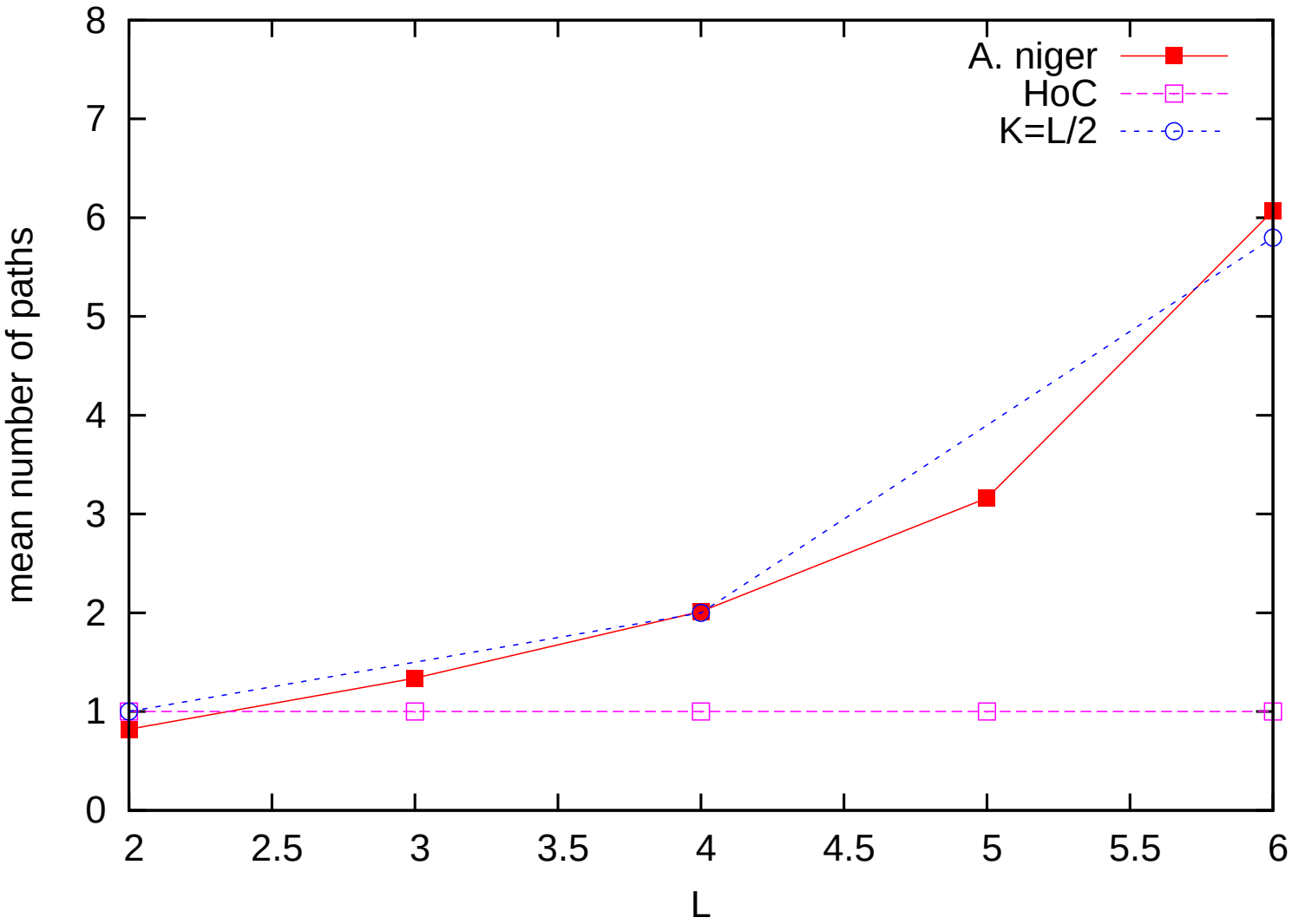
Probability distribution of the number of optima



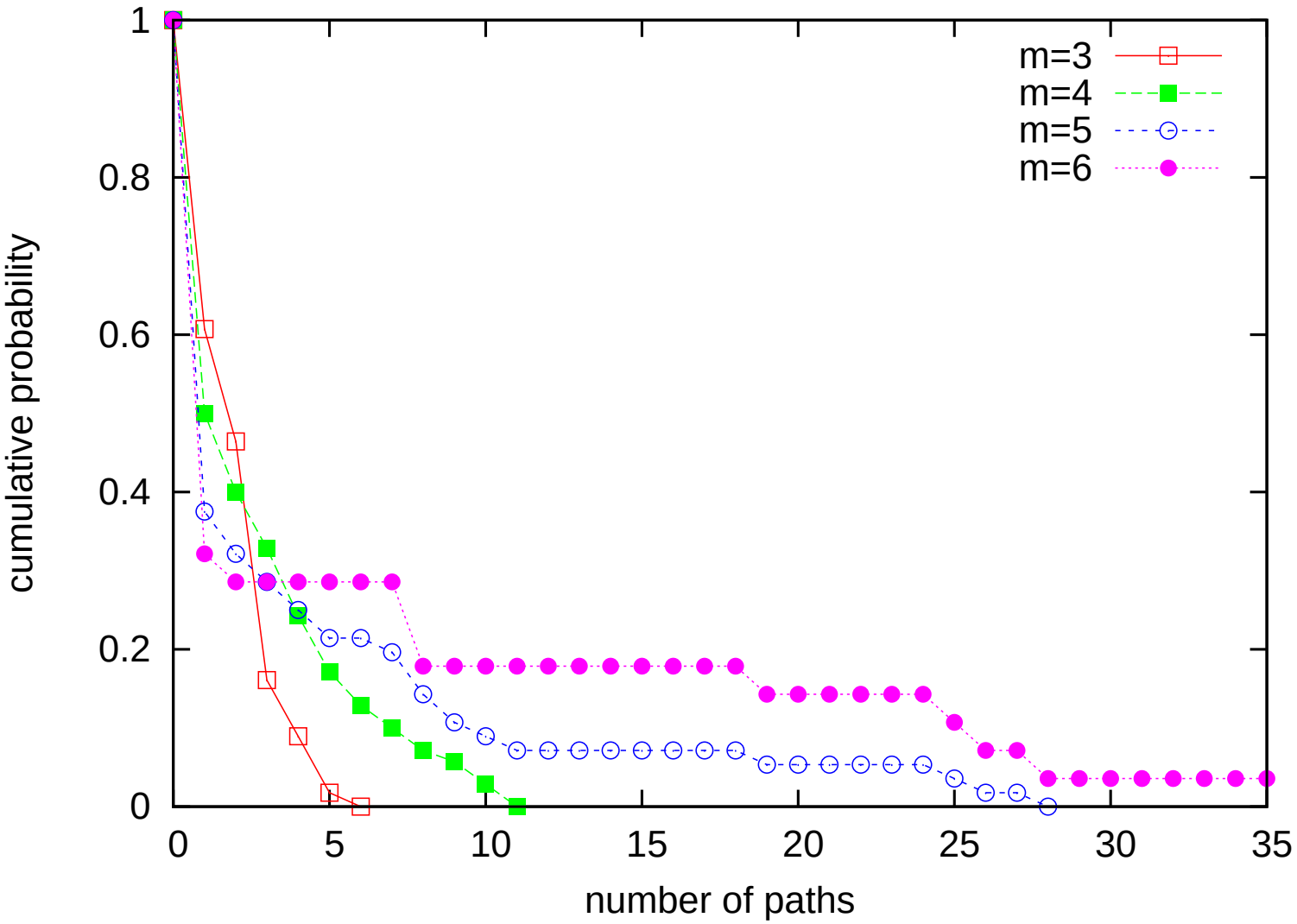
Mean number of optima compared to Kauffman model



Mean number of paths compared to Kauffman model



Cumulative distribution of the number of paths



Dynamics of adaptation on rugged fitness landscapes

In a rugged field of this character, selection will easily carry the species to the nearest peak, but there will be innumerable other peaks that will be higher but which are separated by “valleys”. The problem of evolution as I see it is that of a mechanism by which the species may continually find its way from lower to higher peaks in such a field. S. Wright (1932)

Evolutionary regimes for asexuals in rugged fitness landscapes

K. Jain, JK, Genetics **175**, 1275 (2007)

Parameters:
Population size N, mutation probability μ per locus, sequence length L

- $LN\mu$: Number of mutants produced per generation
- $LN\mu \ll 1$: Mutations are rare $\Rightarrow$ population occupies a single site in sequence space and performs an uphill **adaptive walk** which terminates at local fitness optima
- $1 \ll LN\mu \ll L$: Stochastic regime with competing clones
- $LN\mu > L$: Locally deterministic evolution within a shell of size $d_{\rm eff} \sim \ln N/|\ln \mu|$ but valleys broader than this cannot easily be crossed

Evolutionary regimes for asexuals in rugged fitness landscapes

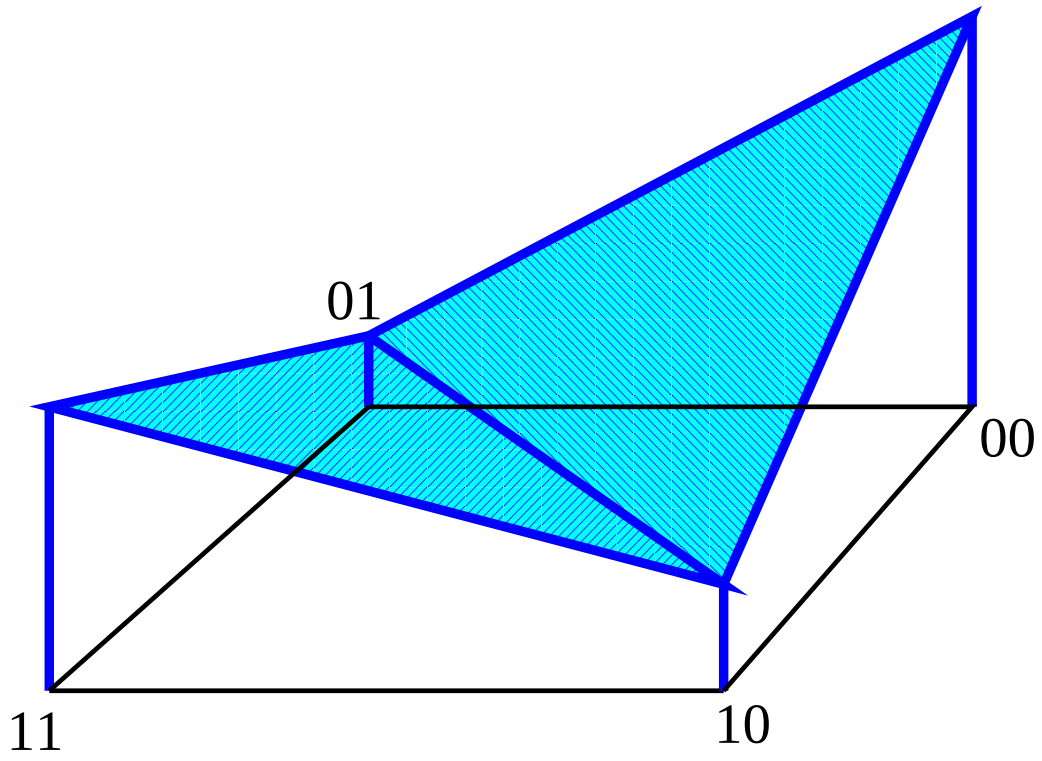
K. Jain, JK, Genetics **175**, 1275 (2007)

Parameters:
Population size N, mutation probability μ per locus, sequence length L

- $LN\mu$: Number of mutants produced per generation
- $LN\mu \ll 1$: Mutations are rare $\Rightarrow$ population occupies a single site in sequence space and performs an uphill **adaptive walk** which terminates at local fitness optima
- $1 \ll LN\mu \ll L$: Stochastic regime with competing clones
- $LN\mu > L$: Locally deterministic evolution within a shell of size $d_{\rm eff} \sim \ln N/|\ln \mu|$ but valleys broader than this cannot easily be crossed

Can sexual reproduction speed up the escape from local peaks?

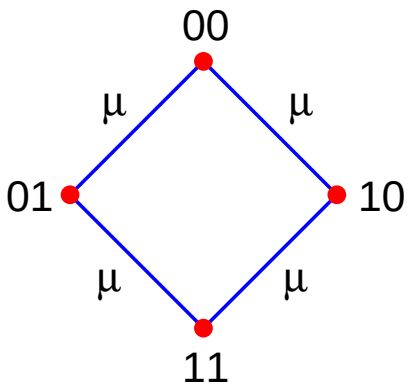
A two-locus rugged fitness landscape



- Population starts at lower fitness peak (11)
- How long does it need to get to the global optimum (00)?
- Does it get there at all?

Infinite population dynamics with recombination

S.C. Park, JK, arXiv:1001.1348



w(00)=1

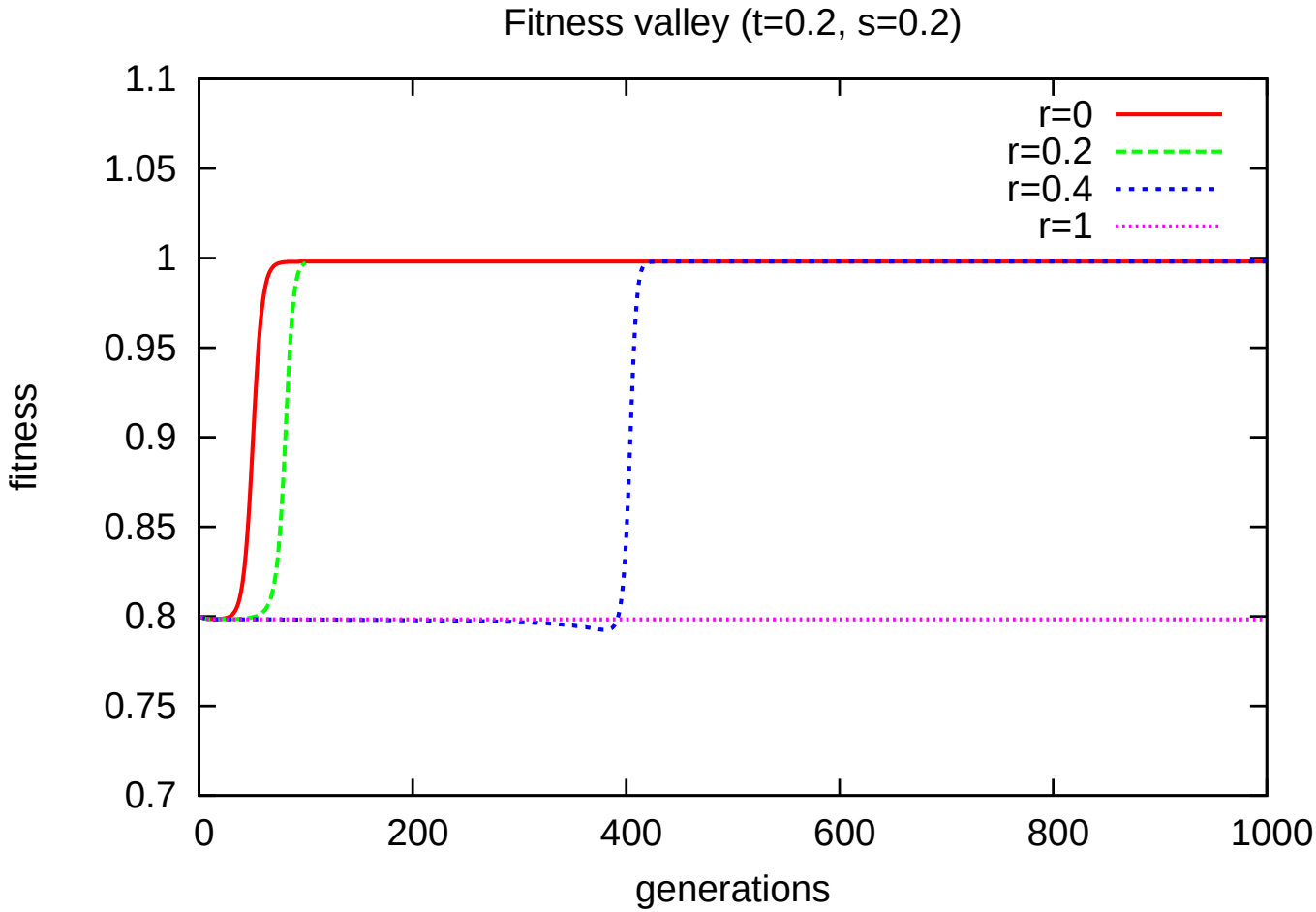
w(10)=w(01)=1−t−s

w(11)=1−t

- Recombination r exchanges genotypes $(00)\times(11) \;\Leftrightarrow\; (10)\times(01)$
- Genotype frequencies $X_0 = P(00),\ X_1 = P(10) = P(01),\ X_2 = P(11)$

$$
\begin{aligned}
\bar{w}X_0' &= (1-\mu)^2 w_{00}X_0 + 2\mu(1-\mu)w_{10}X_1 + \mu^2 w_{11}X_2 - (r/2\bar{w})(1-2\mu)^2 D \\
\bar{w}X_1' &= [1-2\mu(1-\mu)]w_{10}X_1 + \mu(1-\mu)[w_{00}X_0 + w_{11}X_2] + (r/2\bar{w})(1-2\mu)^2 D \\
\bar{w}X_2' &= (1-\mu)^2 w_{11}X_2 + 2\mu(1-\mu)w_{10}X_1 - (r/2\bar{w})(1-2\mu)^2 D \\
D &= w_{00}w_{11}X_0X_2 - w_{10}^2X_1^2 \quad \text{linkage disequilibrium}
\end{aligned}
$$

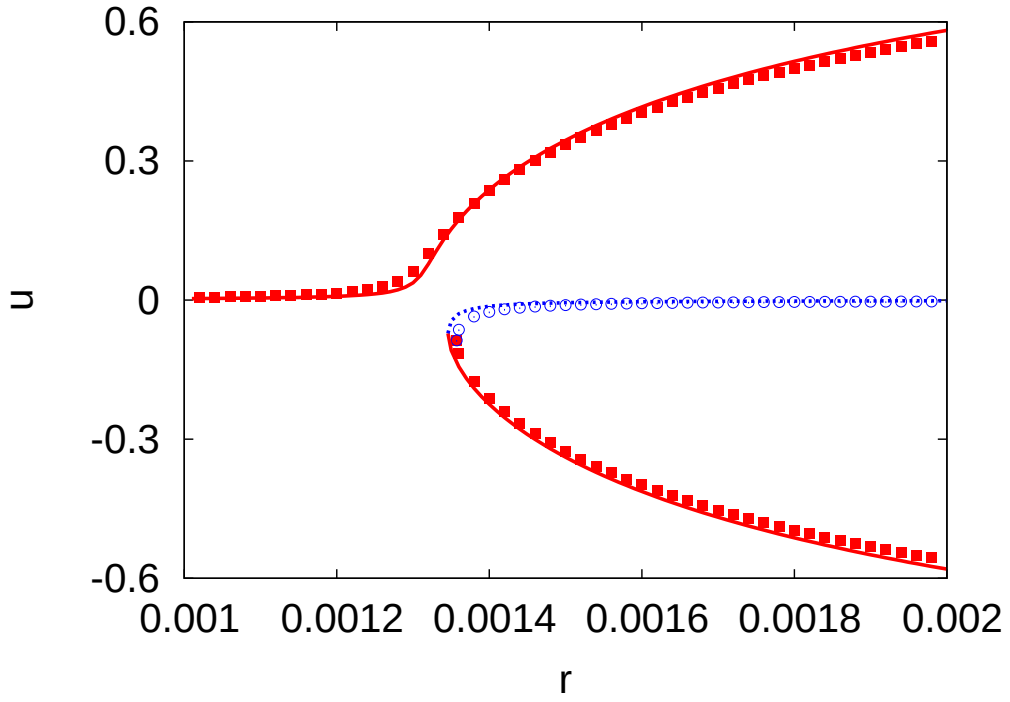
Effect of recombination on the speed of adaptation



- Population remains at the initial genotype (11) forever (up to stochastic escape) when $r > r_c(\mu, t, s)$

Bistability

- Stationary values of $u = P(00) - P(11)$ for $t = 10^{-6}, s = 10^{-2}, \mu = 10^{-3}$:

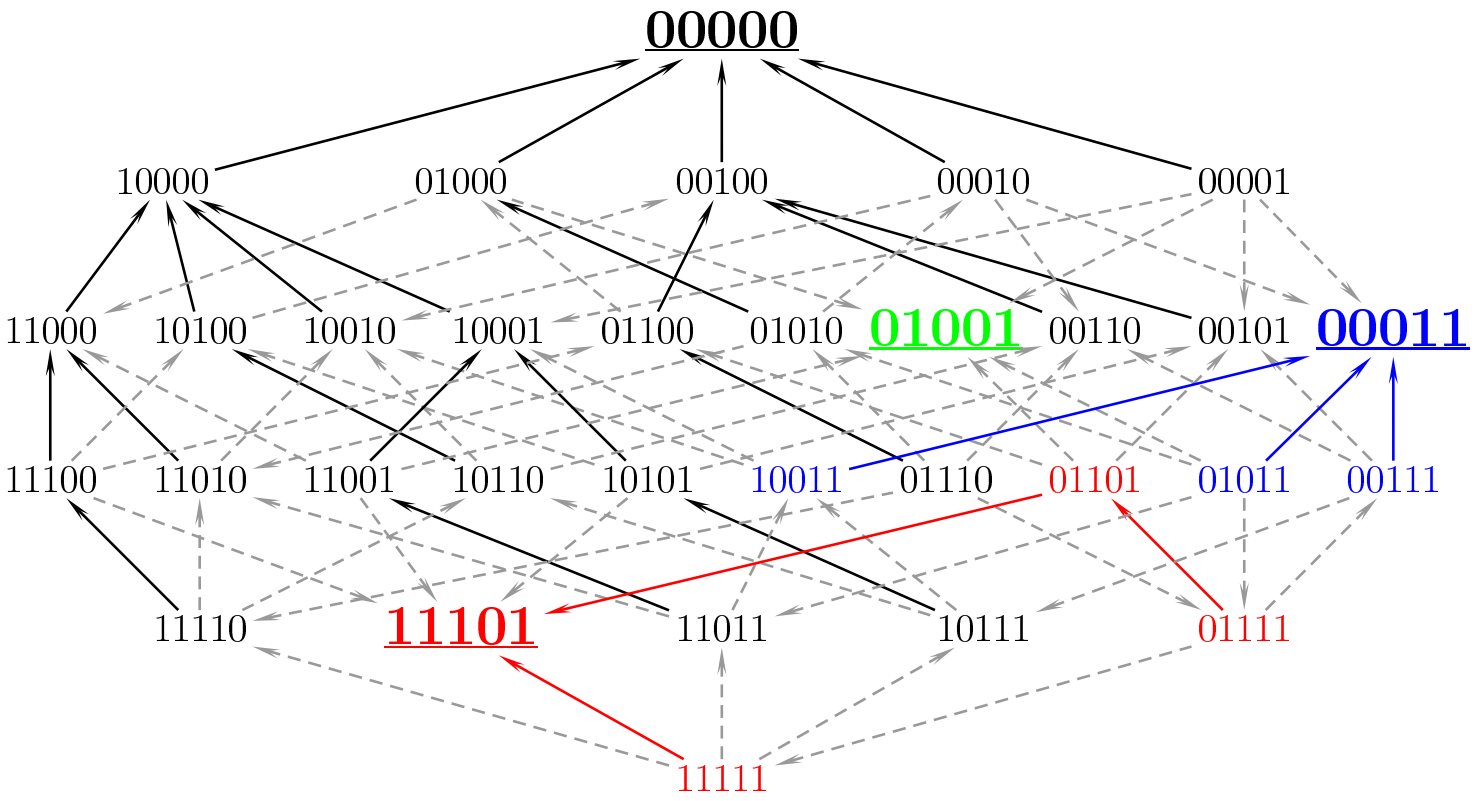


- For $t \ll \mu \ll s \ll 1$ stationary values are solutions of the "Landau" equation $t - (r_0 - r)u - ru^3 = 0$ with $r_0 = 8\mu^2/s$.

- For $\mu \to 0$ the bifurcation occurs at $r_c = t$.

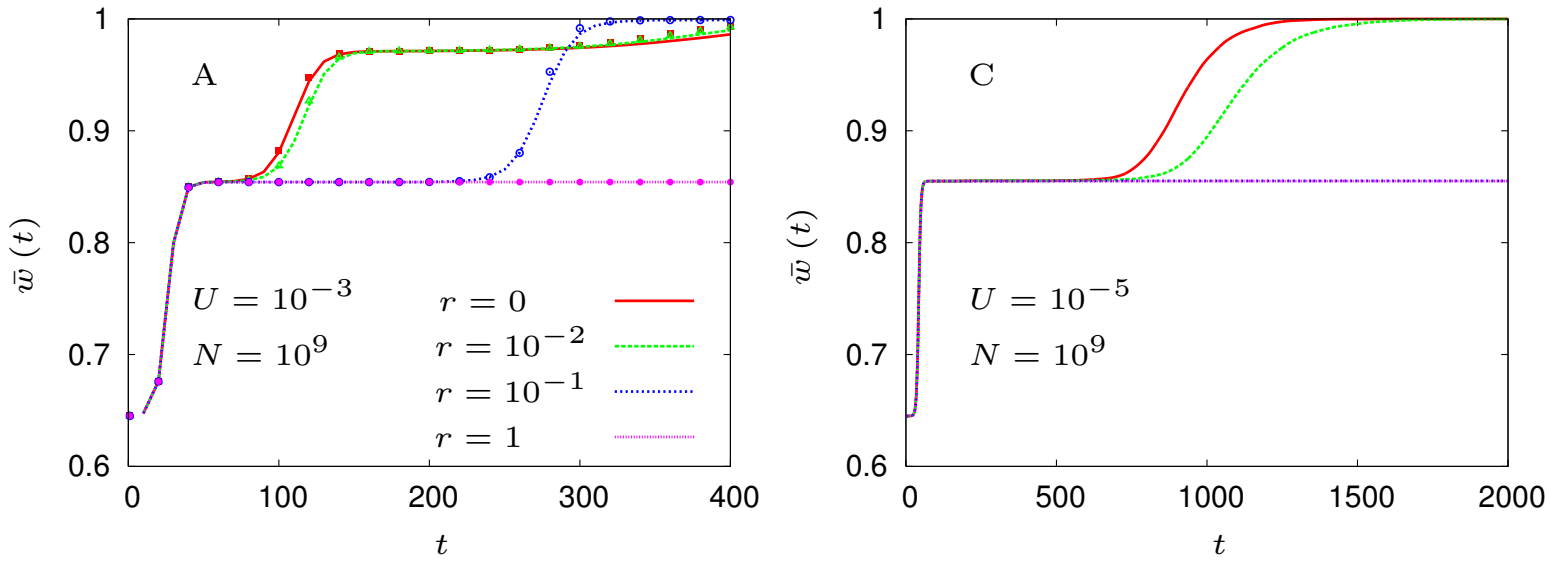
Adaptation in the *A. niger* landscape

J.A.G.M. de Visser, S.C. Park, JK, American Naturalist **174**, S15 (2009)

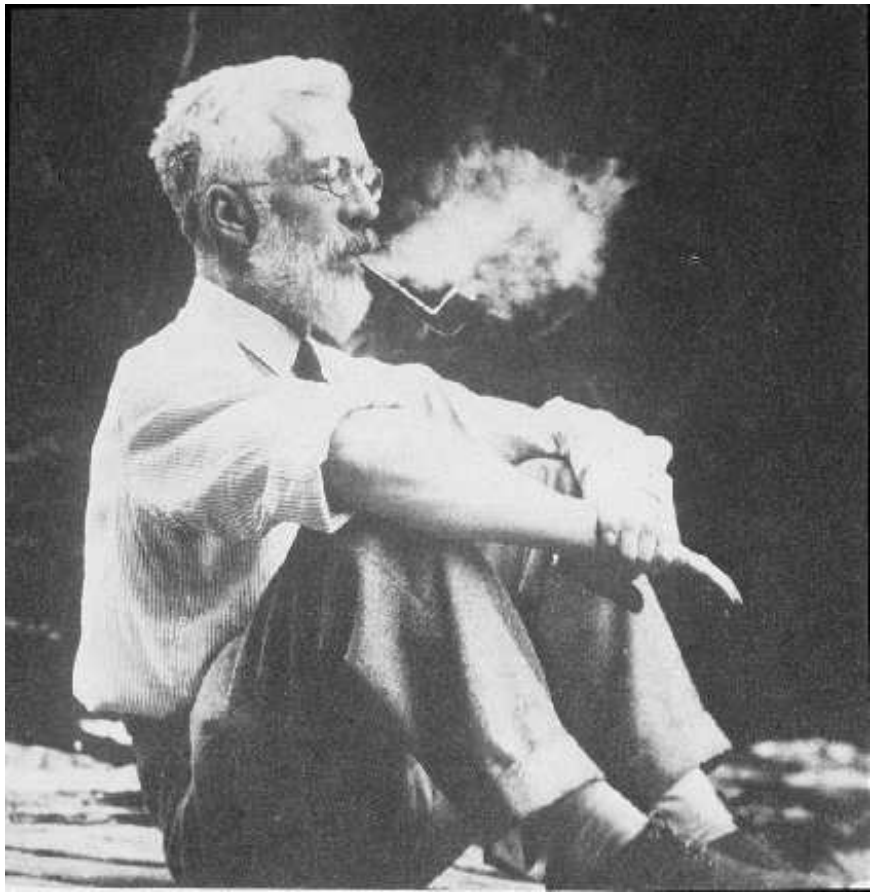


- How does the population move from the fivefold mutant (11111) to the global optimum (00000)?

Adaptation in the *A. niger* landscape



- Recombination delays/suppresses escape from the local peak (11101)
- At intermediate recombination rate $(r = 0.1)$ the secondary local peak (00011) is bypassed
- This effect disappears for smaller U or N



“ In some ways some of us have overtaken Fisher; in many, however, this brilliant, daring man is still far in front.” W.D. Hamilton (1999)