

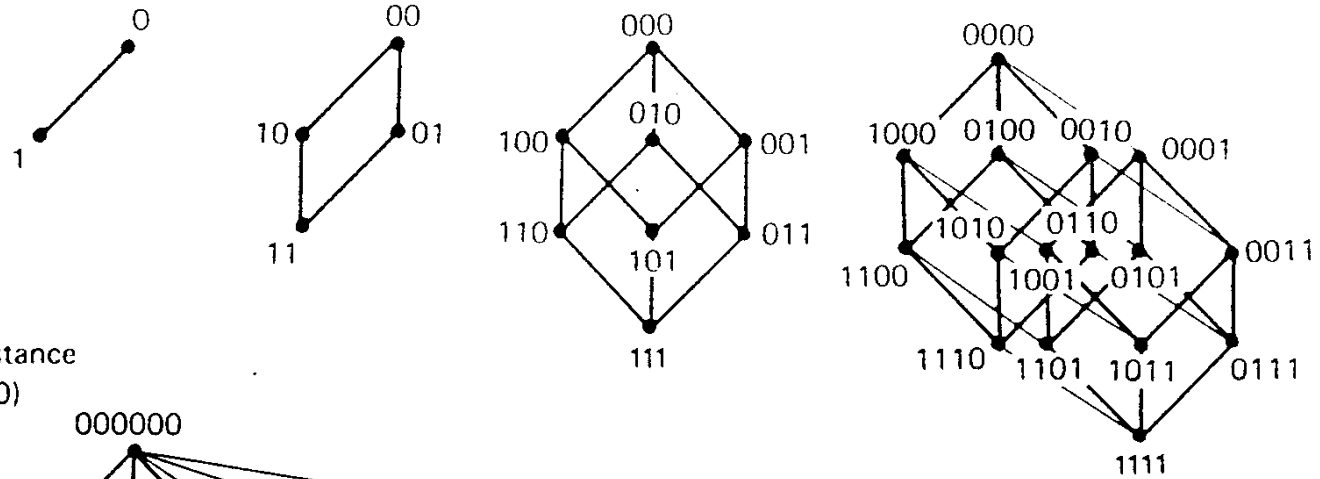
Statistical physics of biological evolution: Fitness landscapes

Joachim Krug

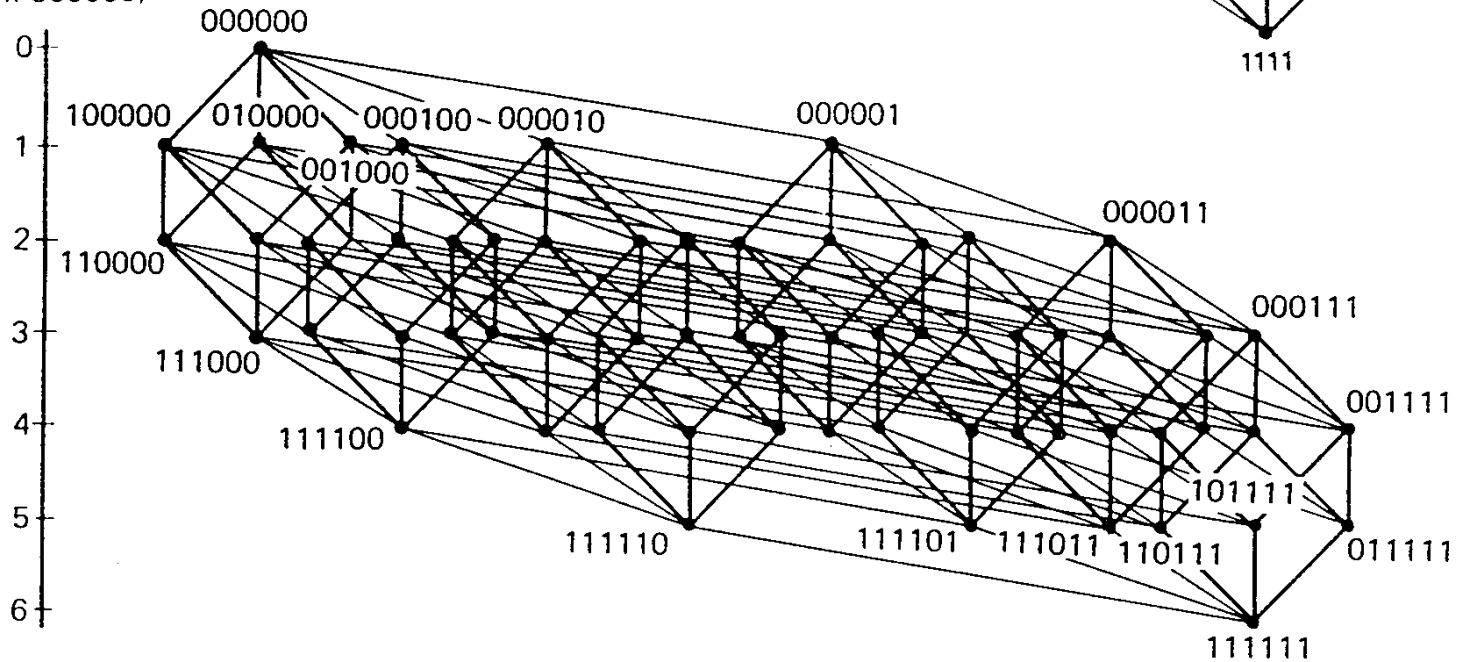
Institute of Biological Physics, University of Cologne

BSSP X, Bangalore, June 19, 2019

Hypercubes for $L = 1 - 6$



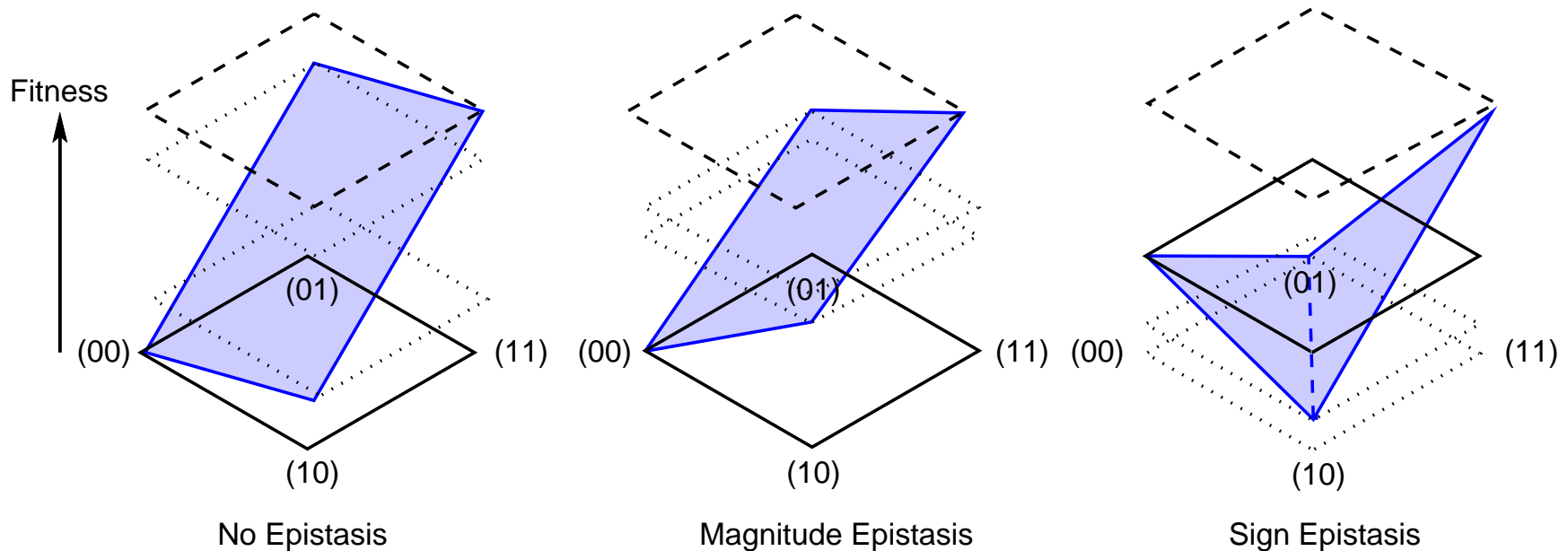
Hamming Distance
(from 000000)



Types of epistasis for two loci

- Encode the genotype by a pair $\tau = (\tau_1, \tau_2)$ of binary variables $\tau_i \in \{0, 1\}$, where $\tau_i = 0/1$ implies absence/presence of the i 'th mutation.
- The fitness landscape for the two-locus system can then be written as

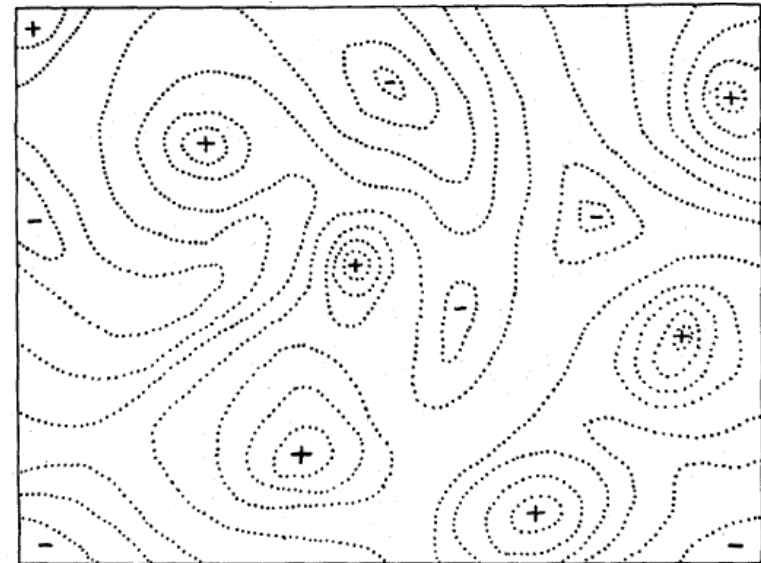
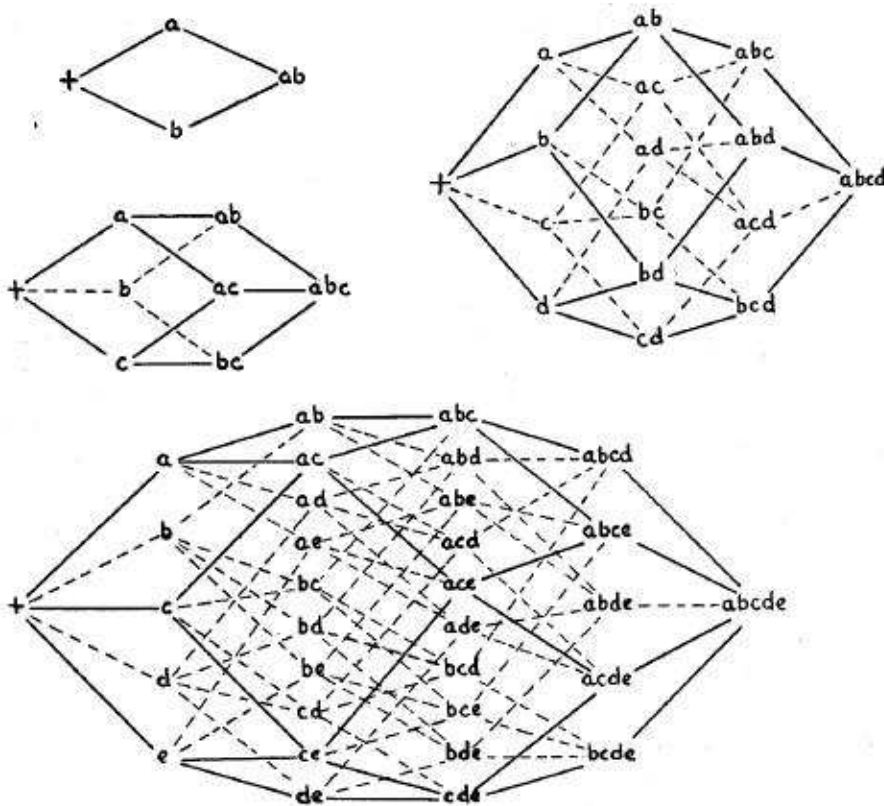
$$f(\tau_1, \tau_2) = f_0 + s_1\tau_1 + s_2\tau_2 + \varepsilon_{12}\tau_1\tau_2$$



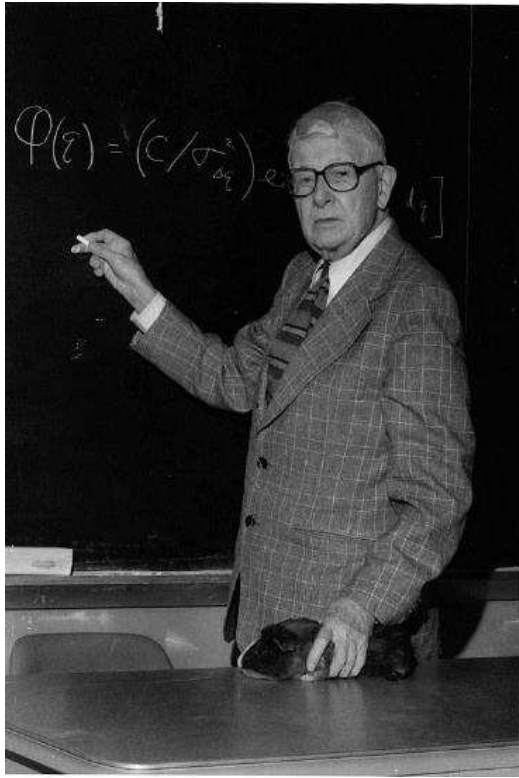
History of the concept

The roles of mutation, inbreeding, cross-breeding and selection in evolution

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

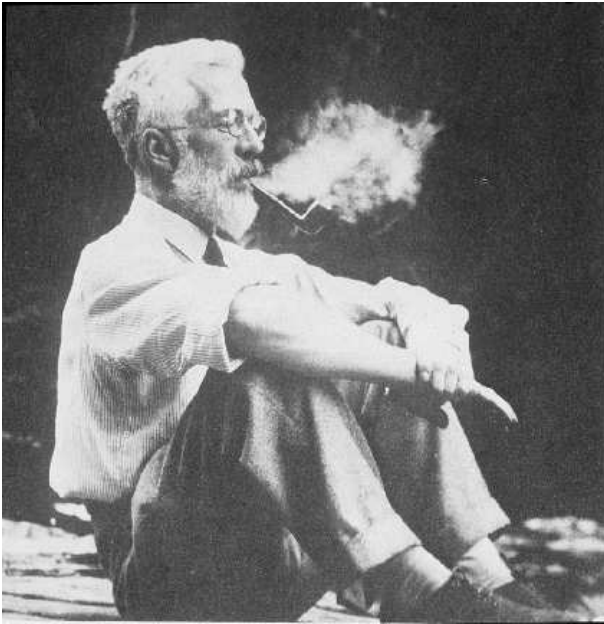


“The two dimensions of figure 2 are a very inadequate representation of such a field.”



S. Wright

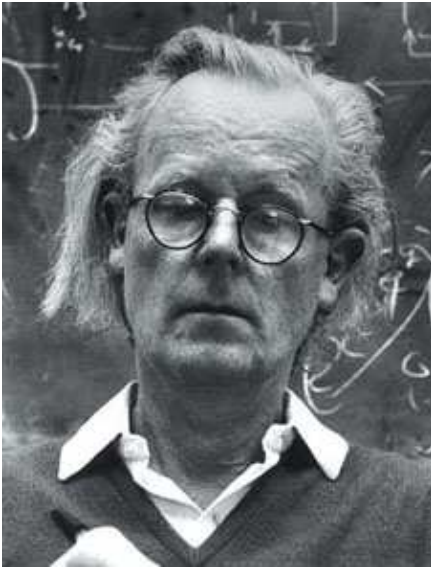
In a rugged field of this character, selection will easily carry the species to the nearest peak, but there will be innumerable other peaks that will be higher but which are separated by “valleys”. The problem of evolution as I see it is that of a mechanism by which the species may continually find its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



Ronald A. Fisher

“In one dimension, a curve gives a series of alternate maxima and minima, but in two dimensions two inequalities must be satisfied for a true maximum, and I suppose that only about one fourth of the stationary points will satisfy both. Roughly I would guess that with n factors only 2^{-n} of the stationary points would be stable for all types of displacement, and any new mutation will have a half chance of destroying the stability. This suggests that true stability in the case of many interacting genes may be of rare occurrence, though its consequence when it does occur is especially interesting and important”

Fisher to Wright, 31.5.1931



John Maynard Smith

The model of protein evolution I want to discuss is best understood by analogy with a popular word game. The object of the game is to pass from one word to another of the same length by changing one letter at the time, with the requirement that all the intermediate words are meaningful in the same language. Thus **WORD** can be converted into **GENE** in the minimum number of steps, as follows:

WORD → **WORE** → **GORE** → **GONE** → **GENE**

This is an analogue of evolution, in which the words represent proteins.

Nature 225:563 (1970)

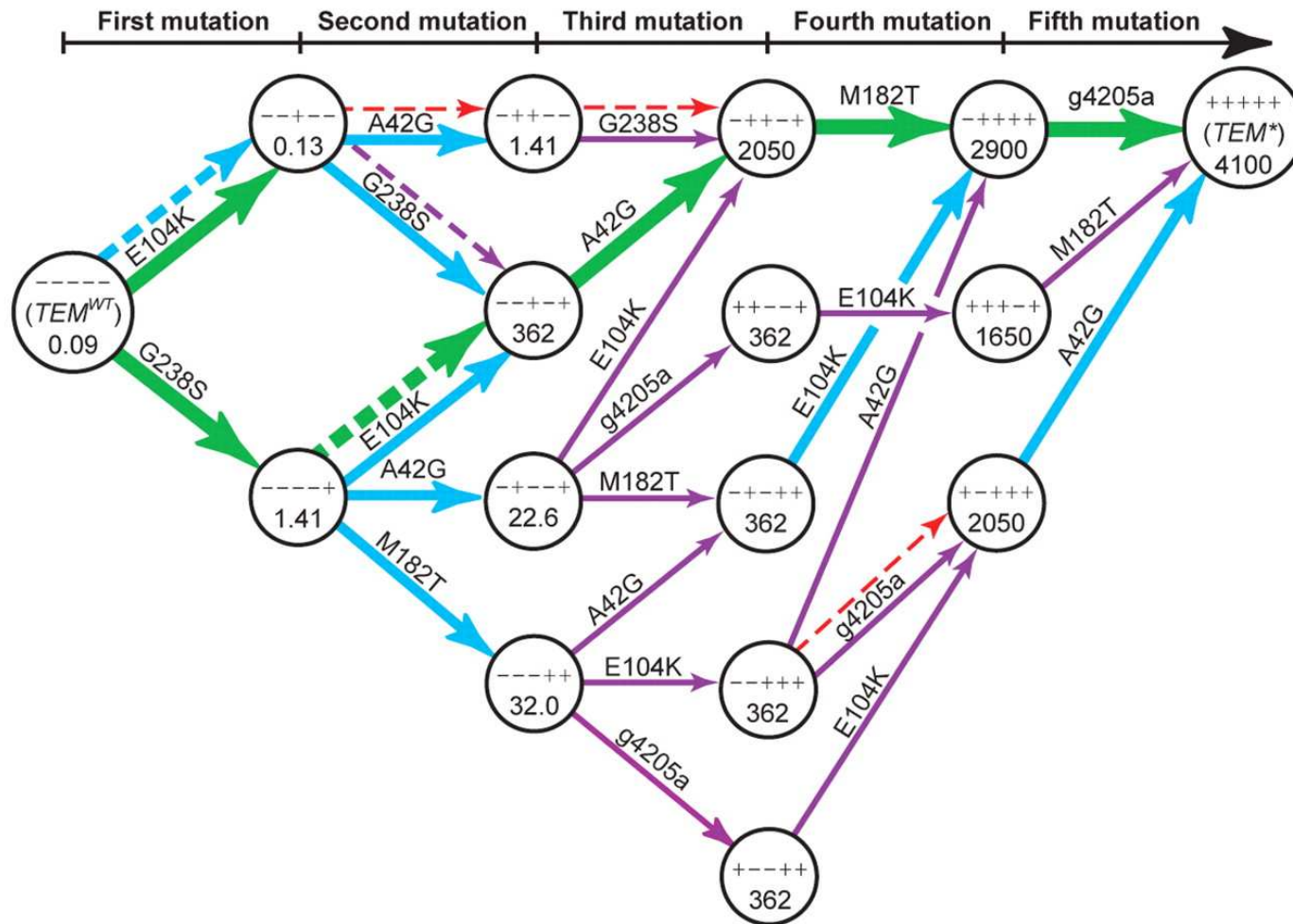
Empirical fitness landscapes

J.A.G.M. de Visser, JK, Nature Reviews Genetics **15**, 480-490 (2014)



“Darwinian evolution can follow only very few mutational paths to fitter proteins”

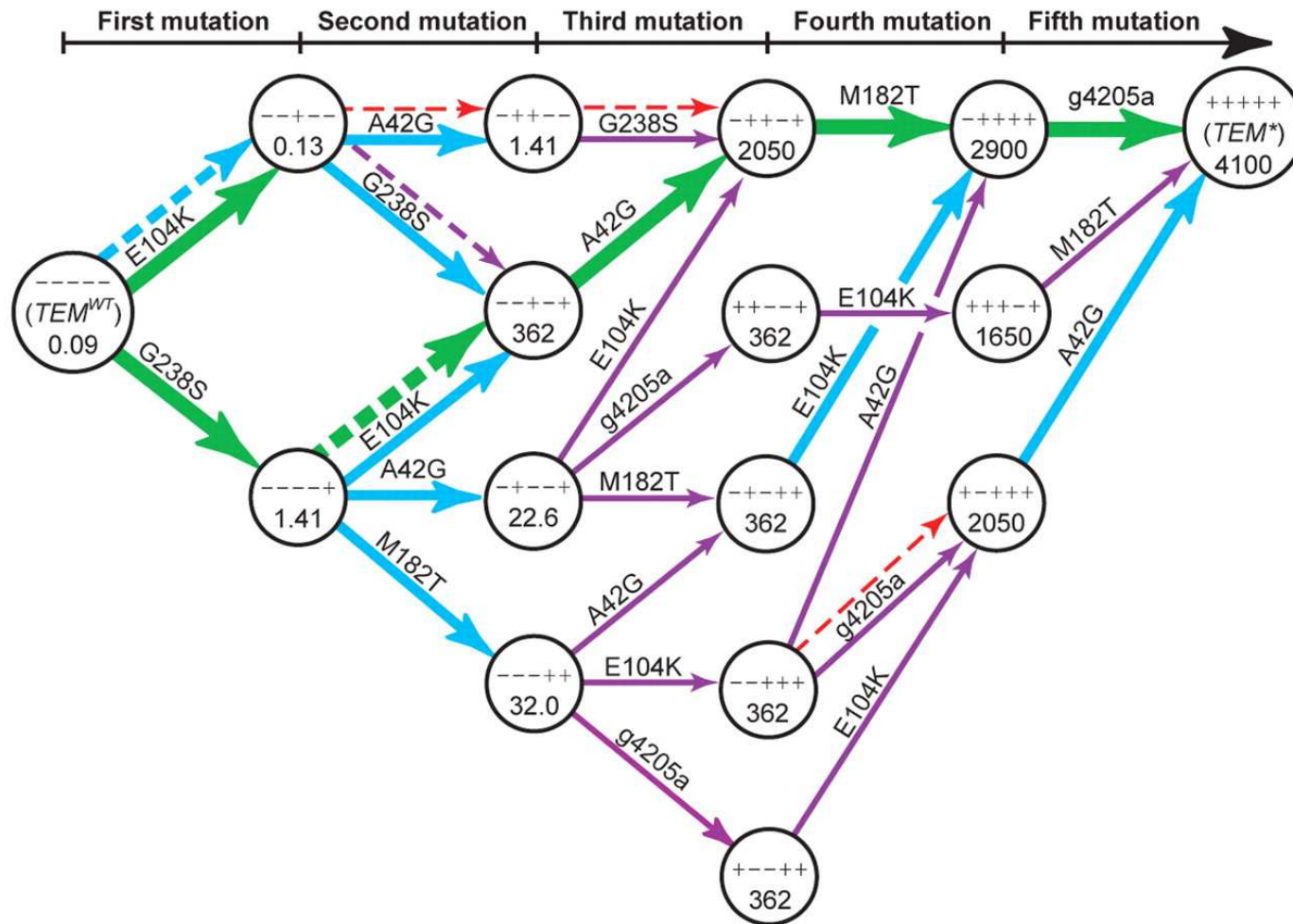
D.M. Weinreich et al., Science 2006



- 5 mutations increase resistance of TEM-1 β -lactamase against cefotaxime

“Darwinian evolution can follow only very few mutational paths to fitter proteins”

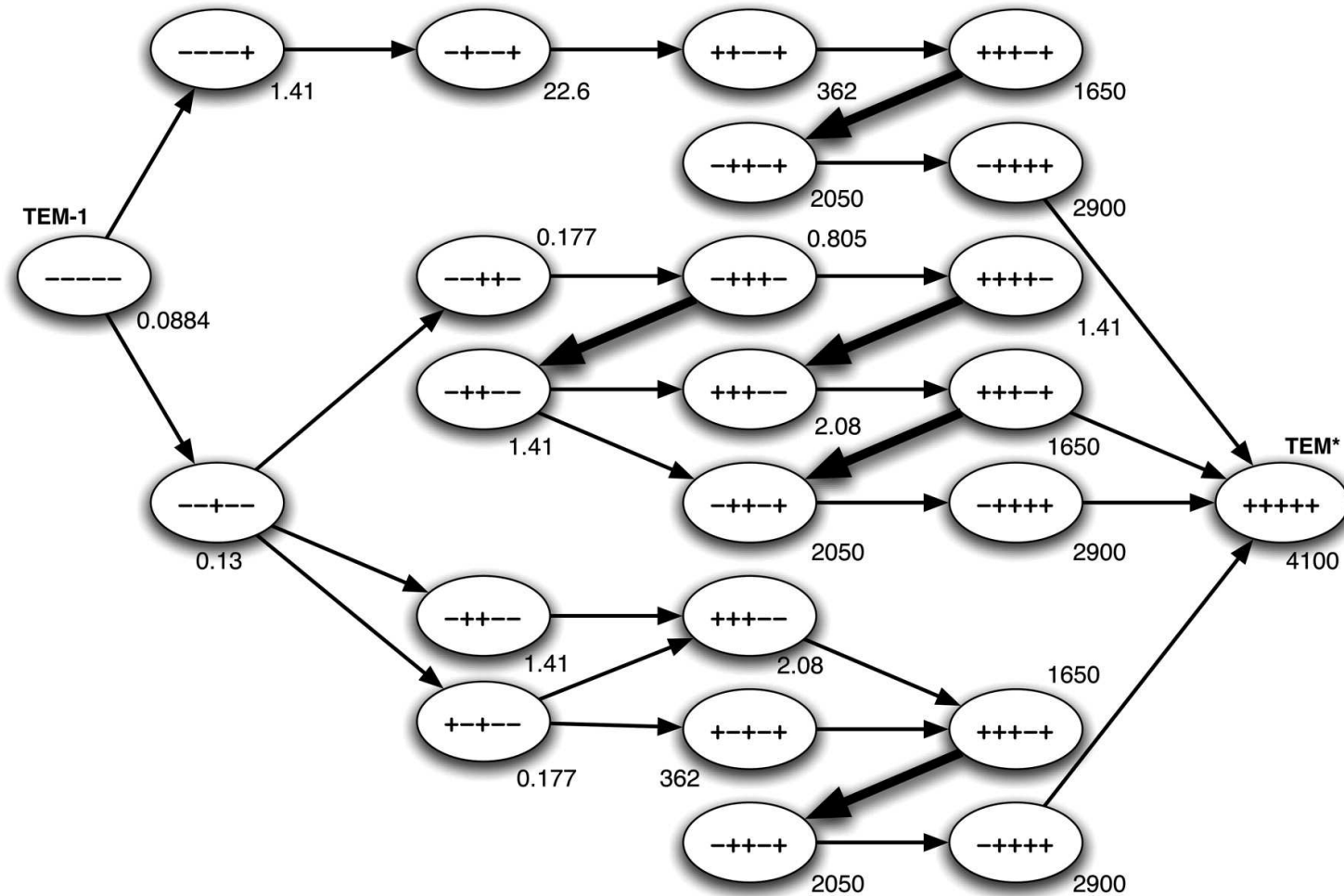
D.M. Weinreich et al., Science 2006



- 18 out of $5! = 120$ direct mutational pathways are accessible...

“Darwinian evolution can follow only very few mutational paths to fitter proteins”

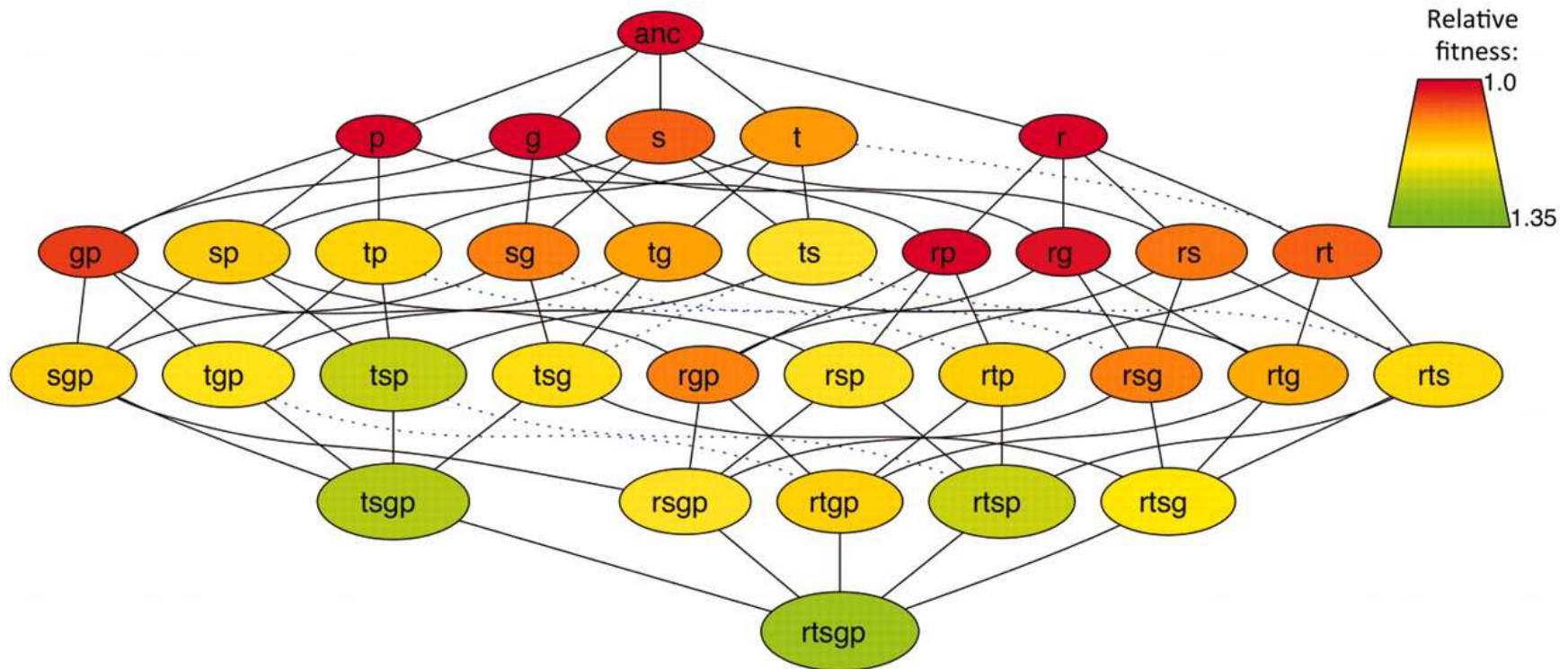
De Pristo et al. 2007



- ...and 27 out of 18651552840 directed and undirected pathways

Five mutations from Lenski's long-term evolution experiment with *E. coli*

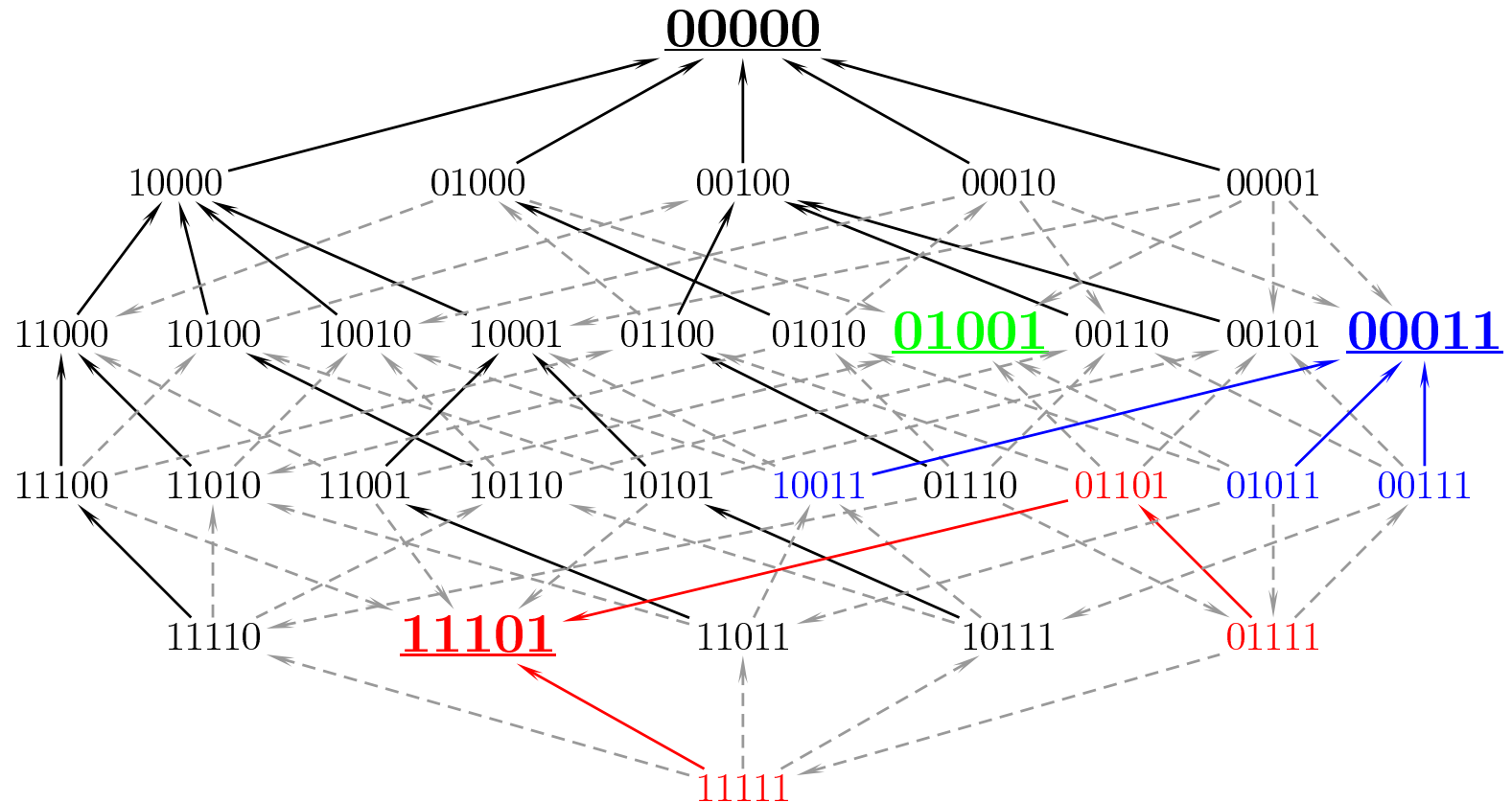
A.I. Khan et al., Science **332** (2011) 1193



- single fitness peak, 86 out of $5! = 120$ pathways are accessible

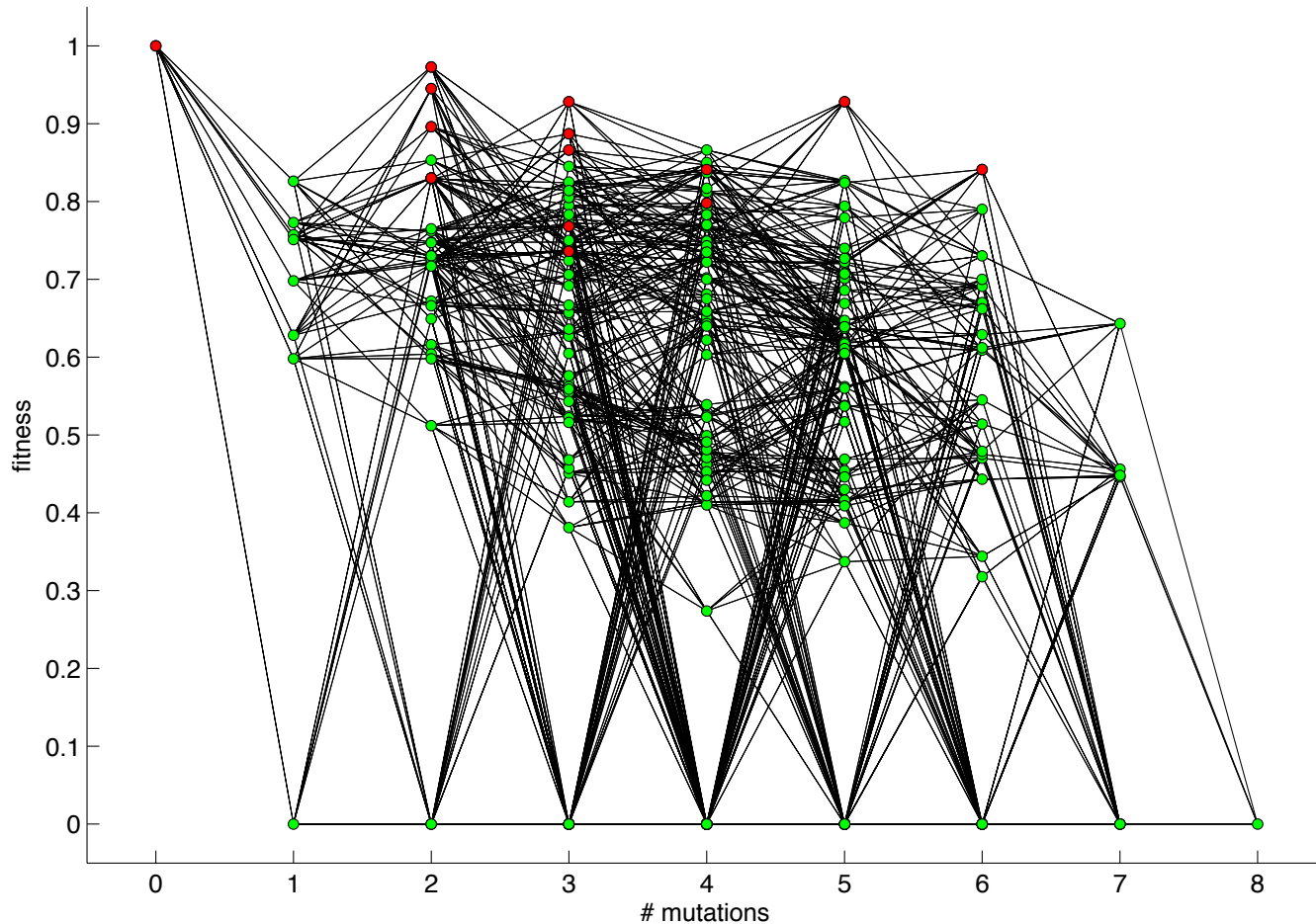
The *Aspergillus niger* fitness landscape

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



- Combinations of 8 individually deleterious marker mutations (one out of $\binom{8}{5} = 56$ five-dimensional subsets shown)

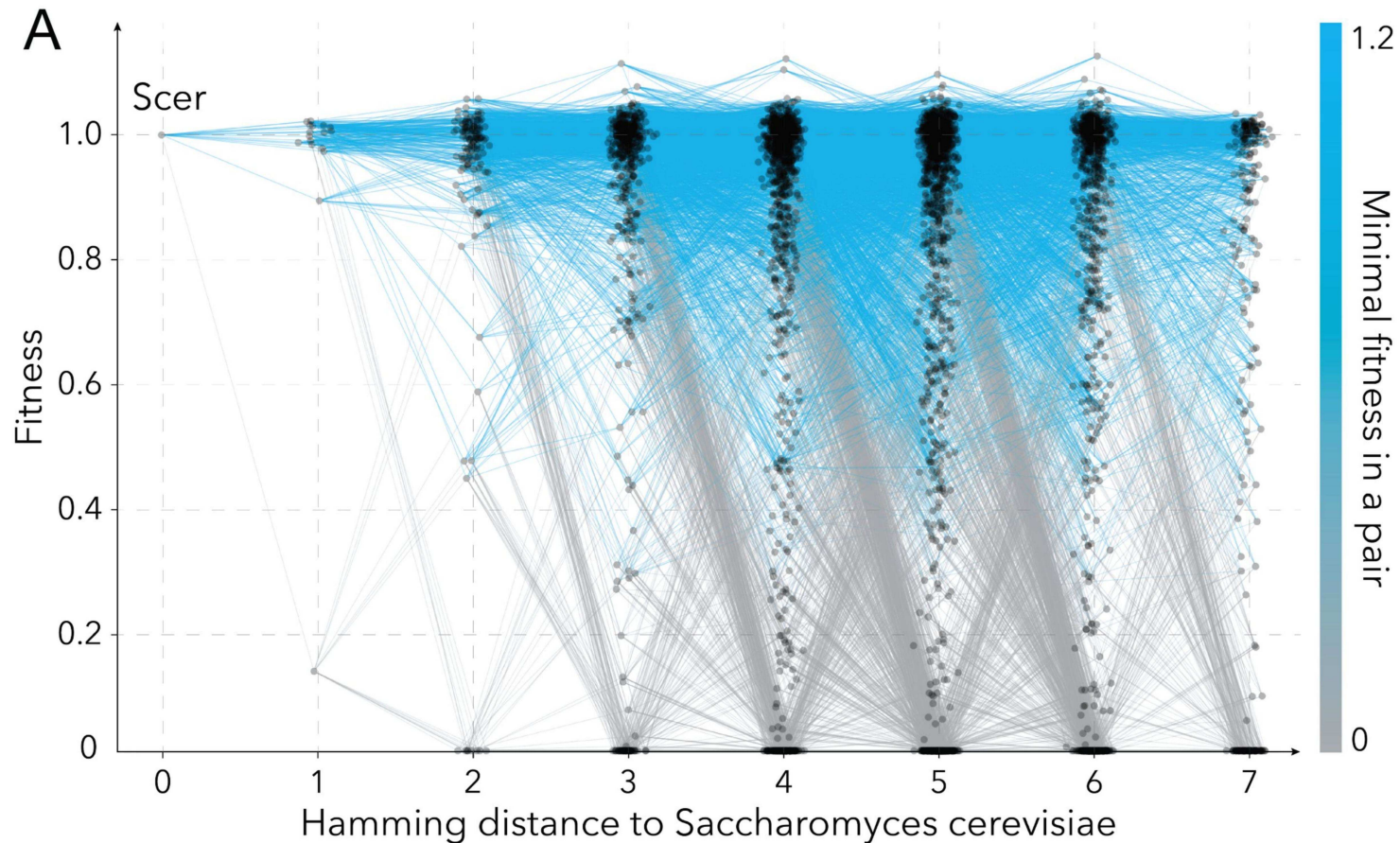
The *Aspergillus niger* fitness landscape



- Average fitness declines roughly linearly with the number of mutations
- Local fitness maxima marked in red

A macroevolutionary fitness landscape for yeast

V.O. Pokusaeva et al., PLoS Genetics (2019)



- 176879 nucleotide sequences/26364 amino acid sequences of a segment of the His3 gene involved in histidine synthesis