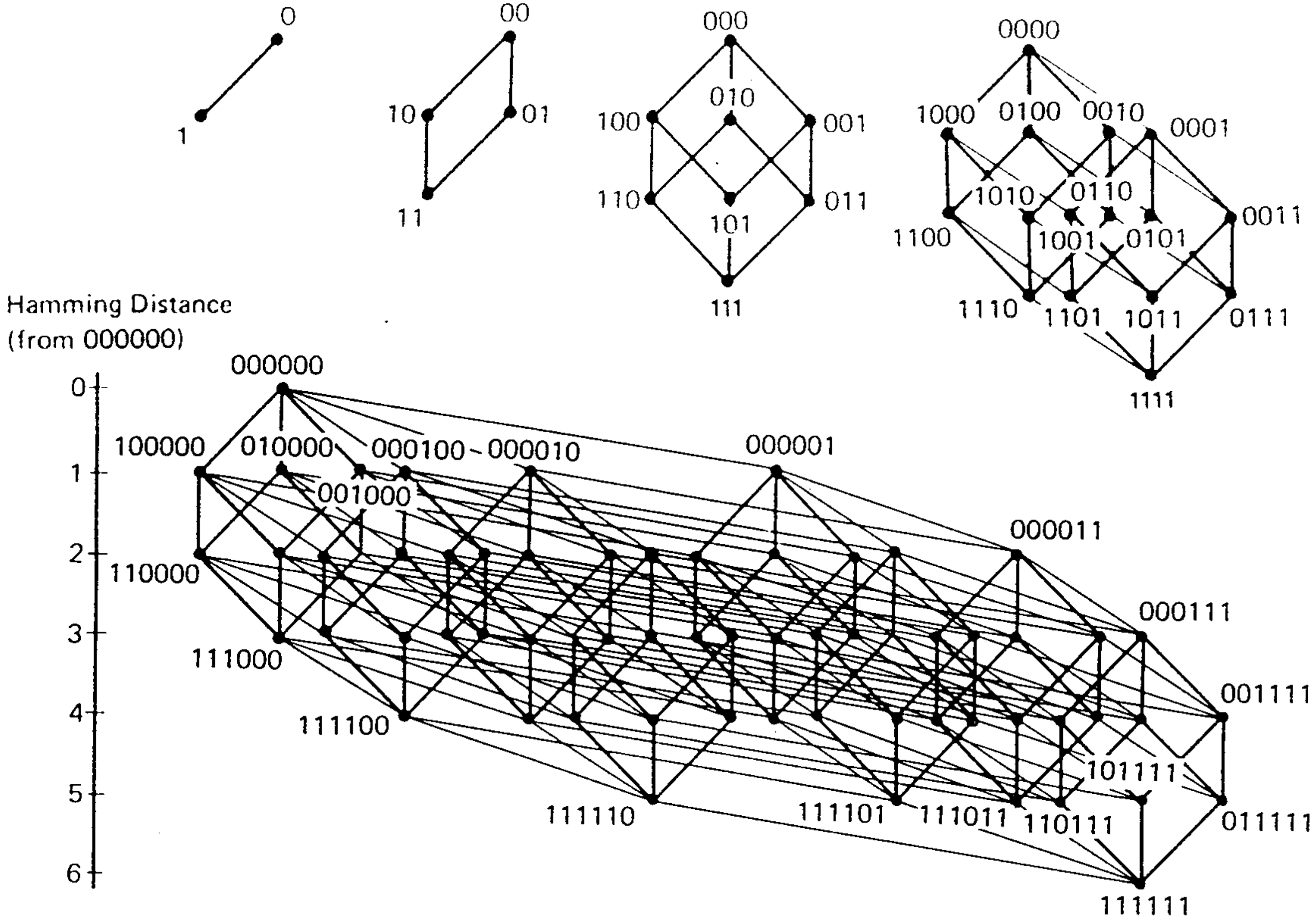
Fitness landscapes in theory and application

Joachim Krug
Institute of Biological Physics
University of Cologne

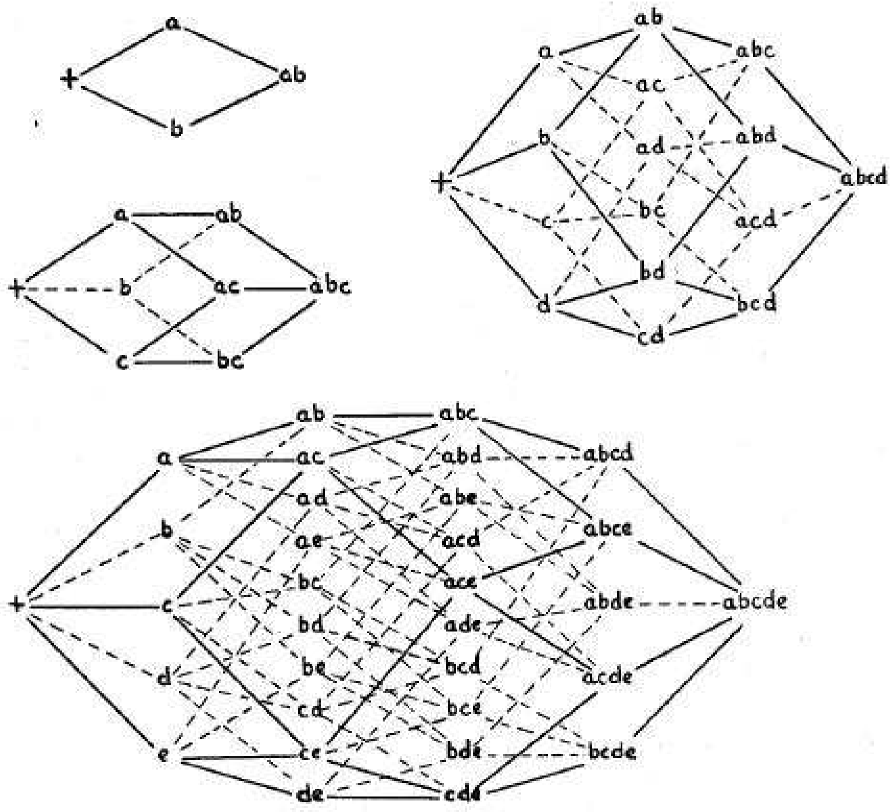
Third Bangalore School on Population Genetics and Evolution,
March 5-16, 2018

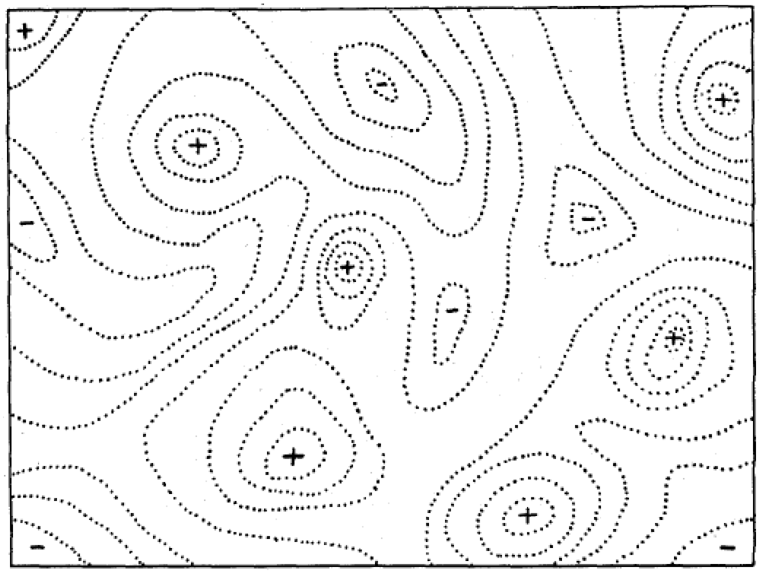
Hypercubes for $L = 1 - 6$



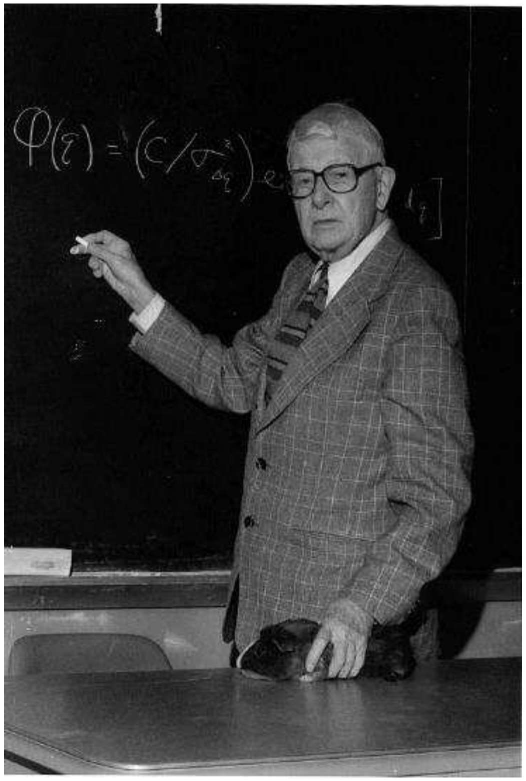
The roles of mutation, inbreeding, cross-breeding and selection in evolution

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



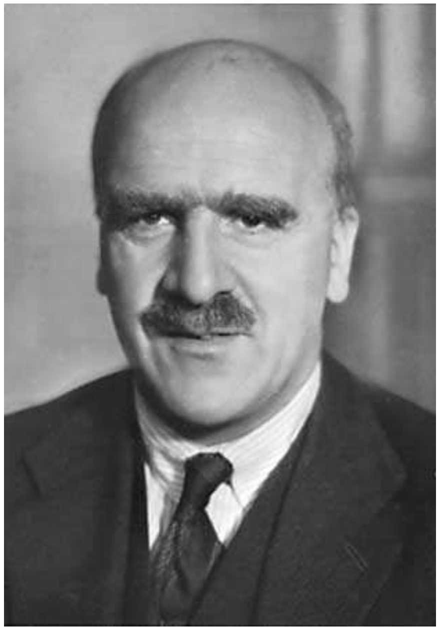


“The two dimensions of figure 2 are a very inadequate representation of such a field.”



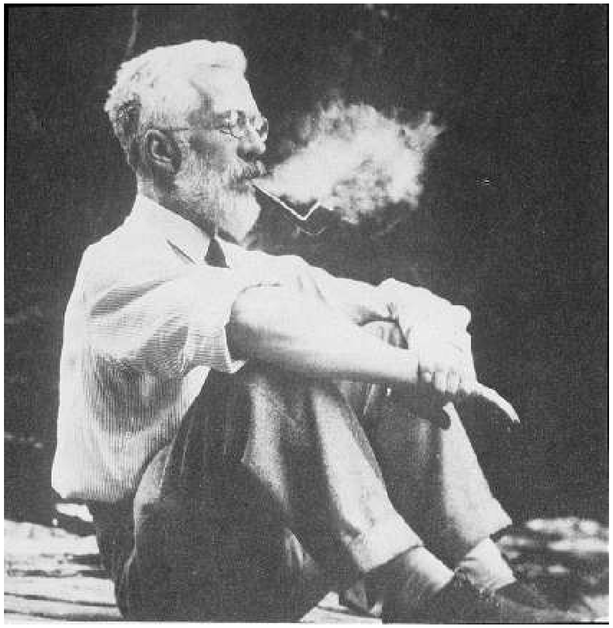
S. Wright

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.



J.B.S. Haldane

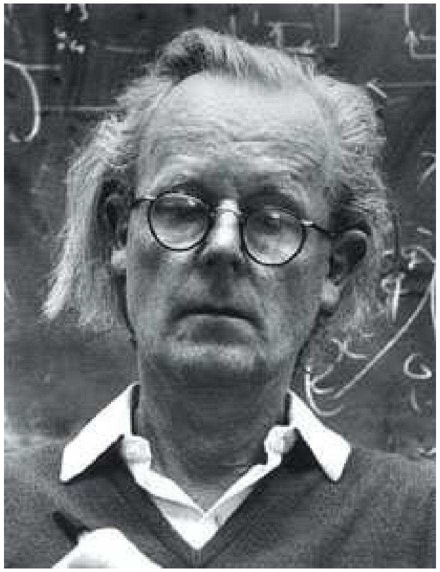
We may consider m genes. In this case any population can be represented by a point in m-dimensional space, all populations being represented by the points of a regular orthotope, or hypercube. Each of the 2^m apices of this figure represents a homozygous population. Clearly the condition for stability of any such population is that no change in a single factor should yield a more viable type. In other words, no adjacent apices can both represent stable populations. The maximum number of stable populations is thus 2^{m-1}, represented by the vertices of the polytope arising from the omission of alternate vertices of the regular orthotope. Proc. Camb. Philos. Soc. 27:137 (1931)



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.

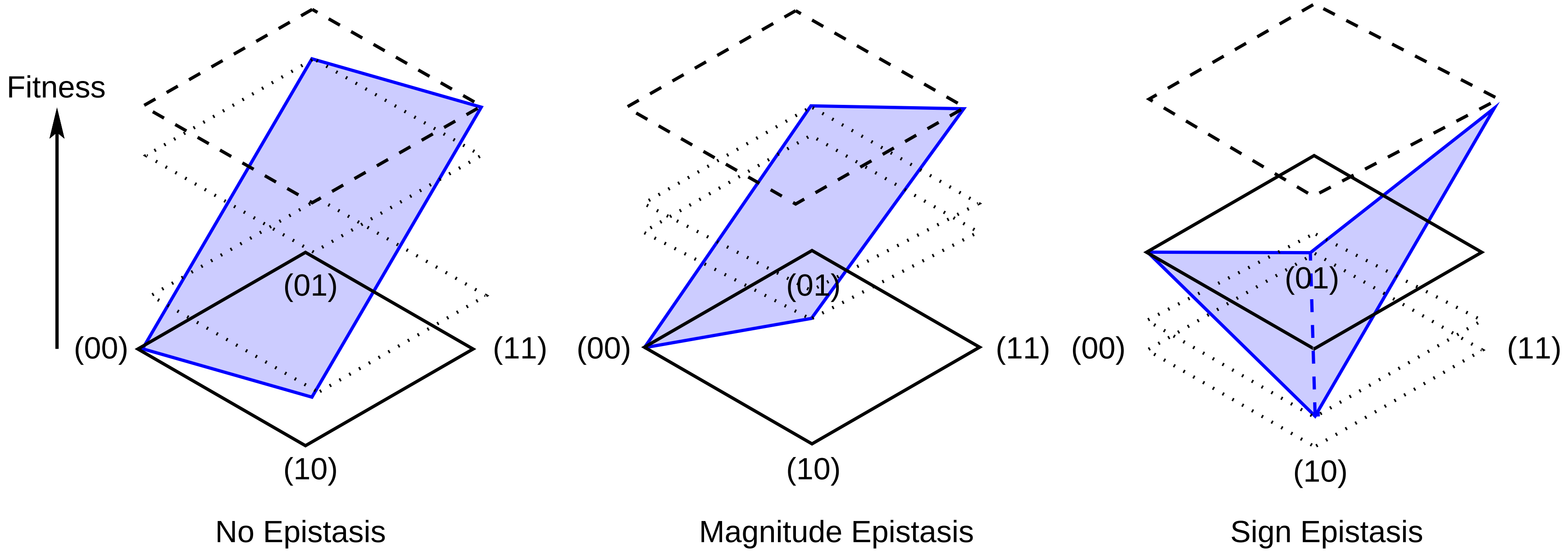
Nature 225:563 (1970)

Fitness landscape for a pair of 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

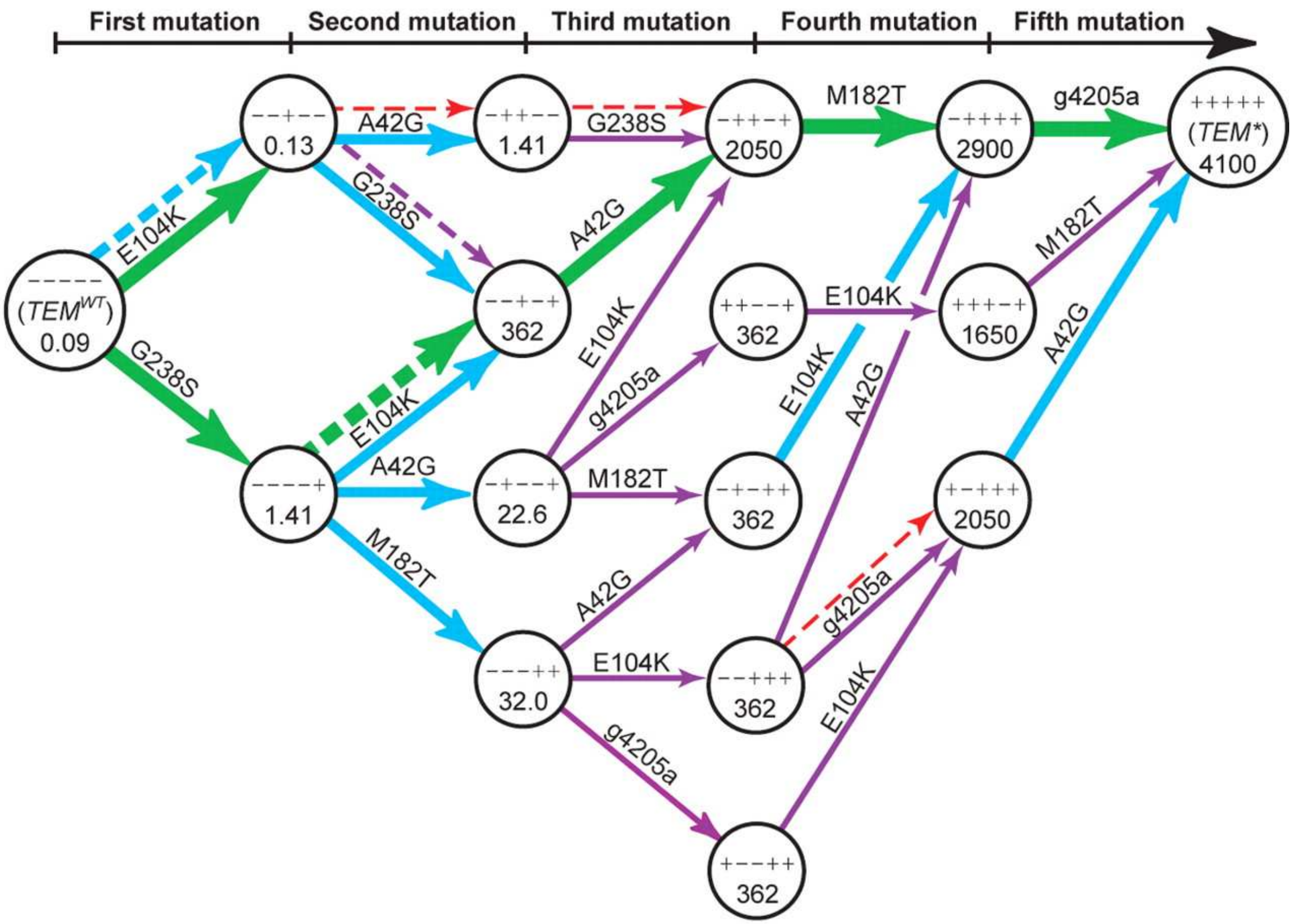
$$f(\tau_1, \tau_2) = f_0 + s_1\tau_1 + s_2\tau_2 + \varepsilon_{12}\tau_1\tau_2$$



Empirical fitness landscapes

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

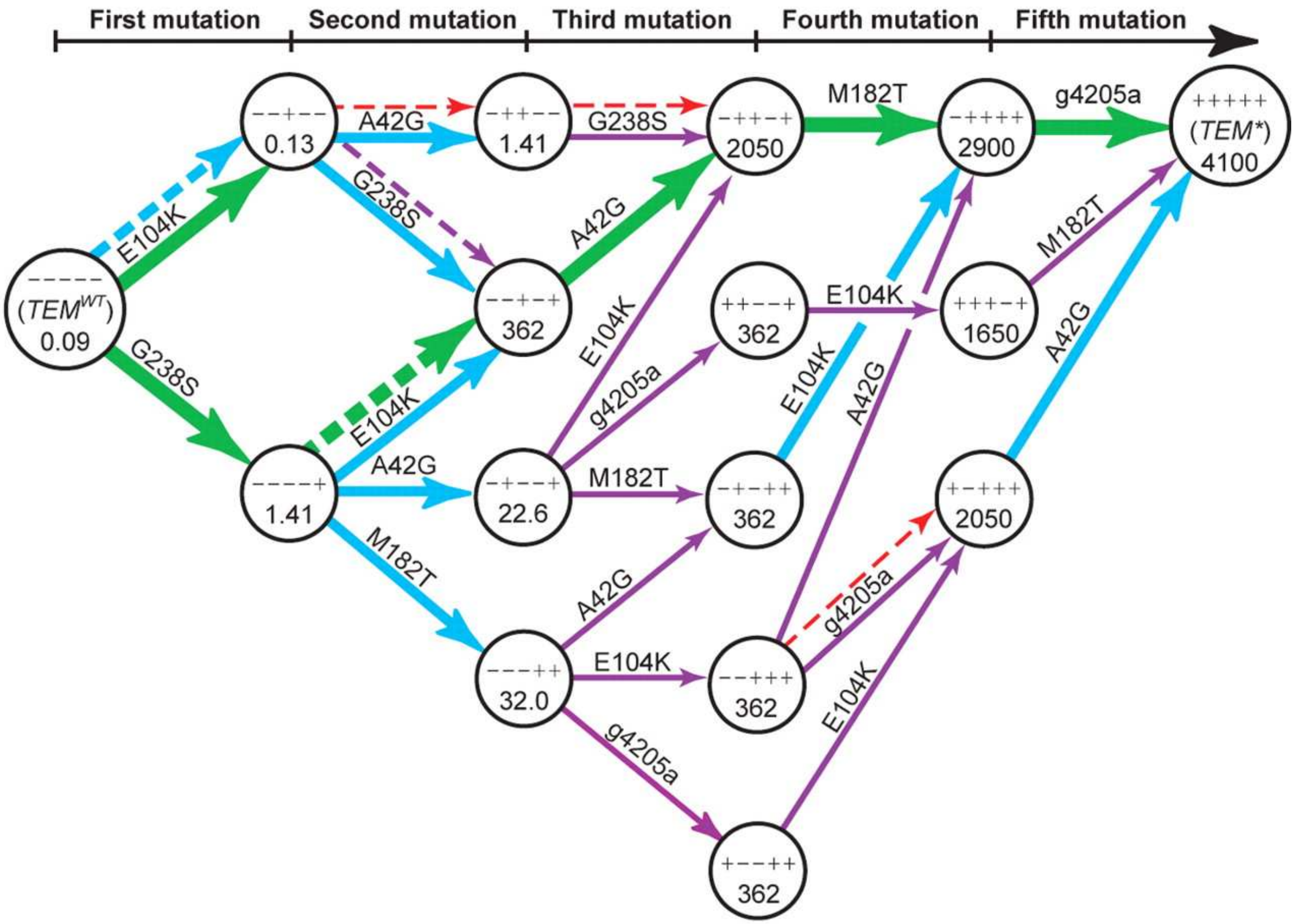
D.M. Weinreich et al., Science 2006



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

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

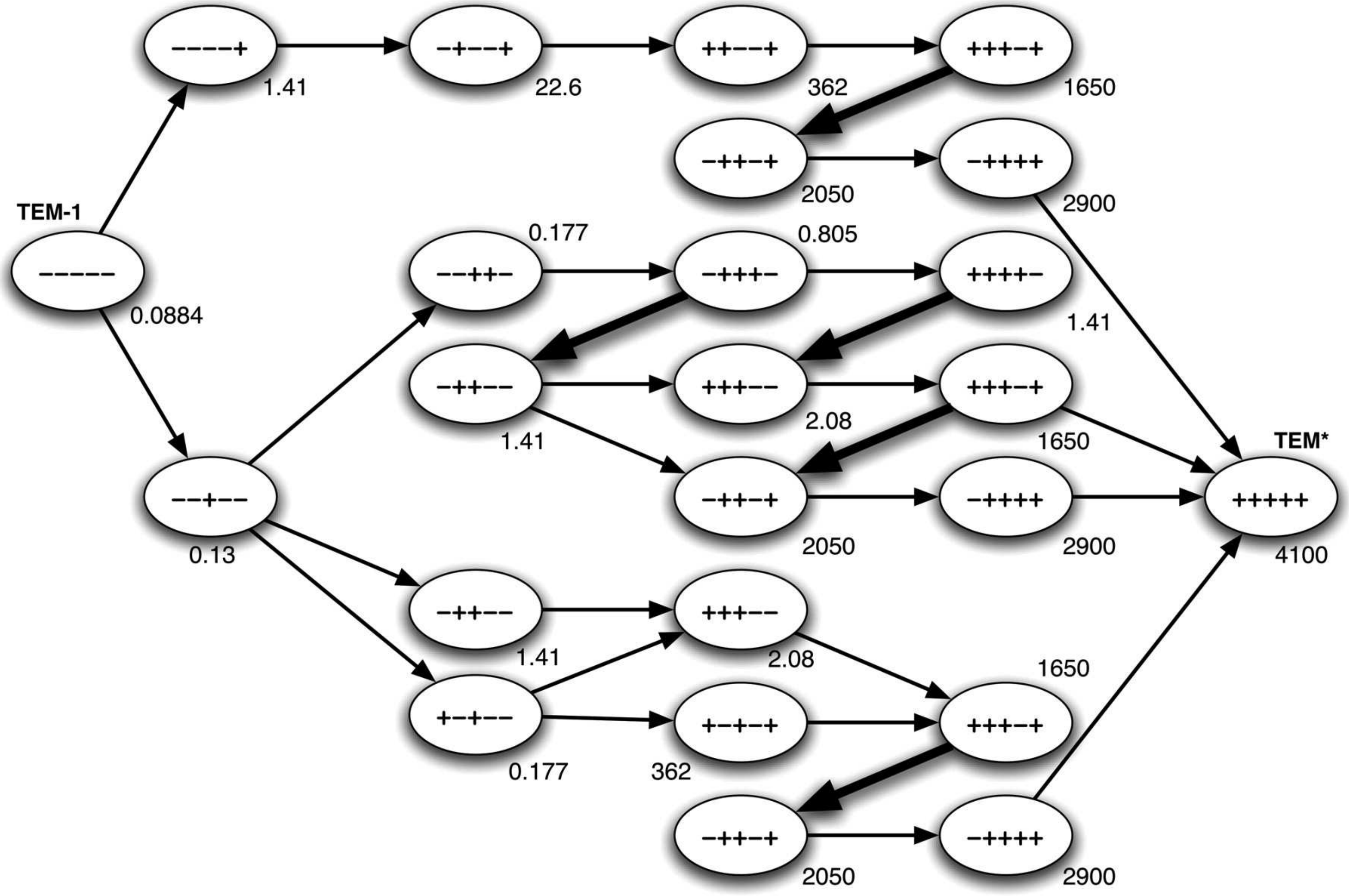
D.M. Weinreich et al., Science 2006



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

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

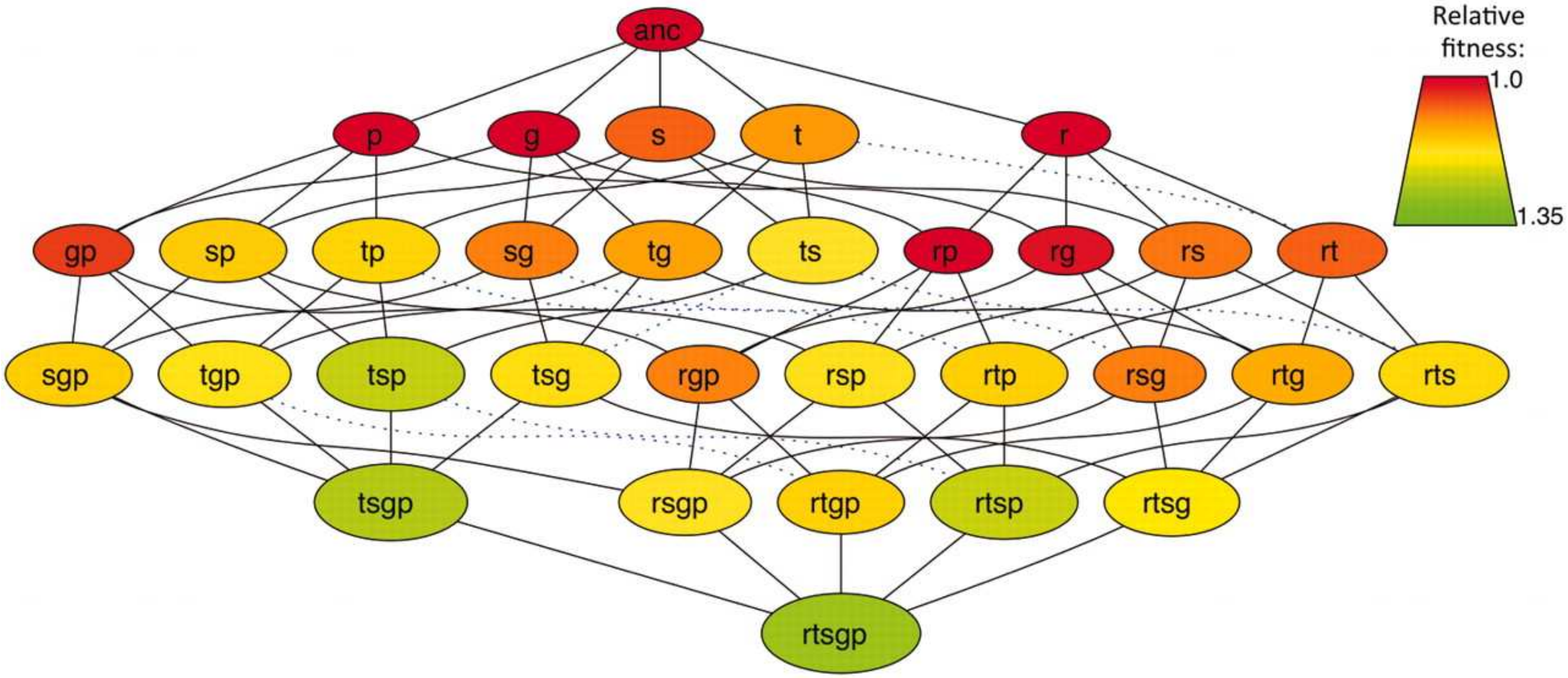
De Pristo et al. 2007



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

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

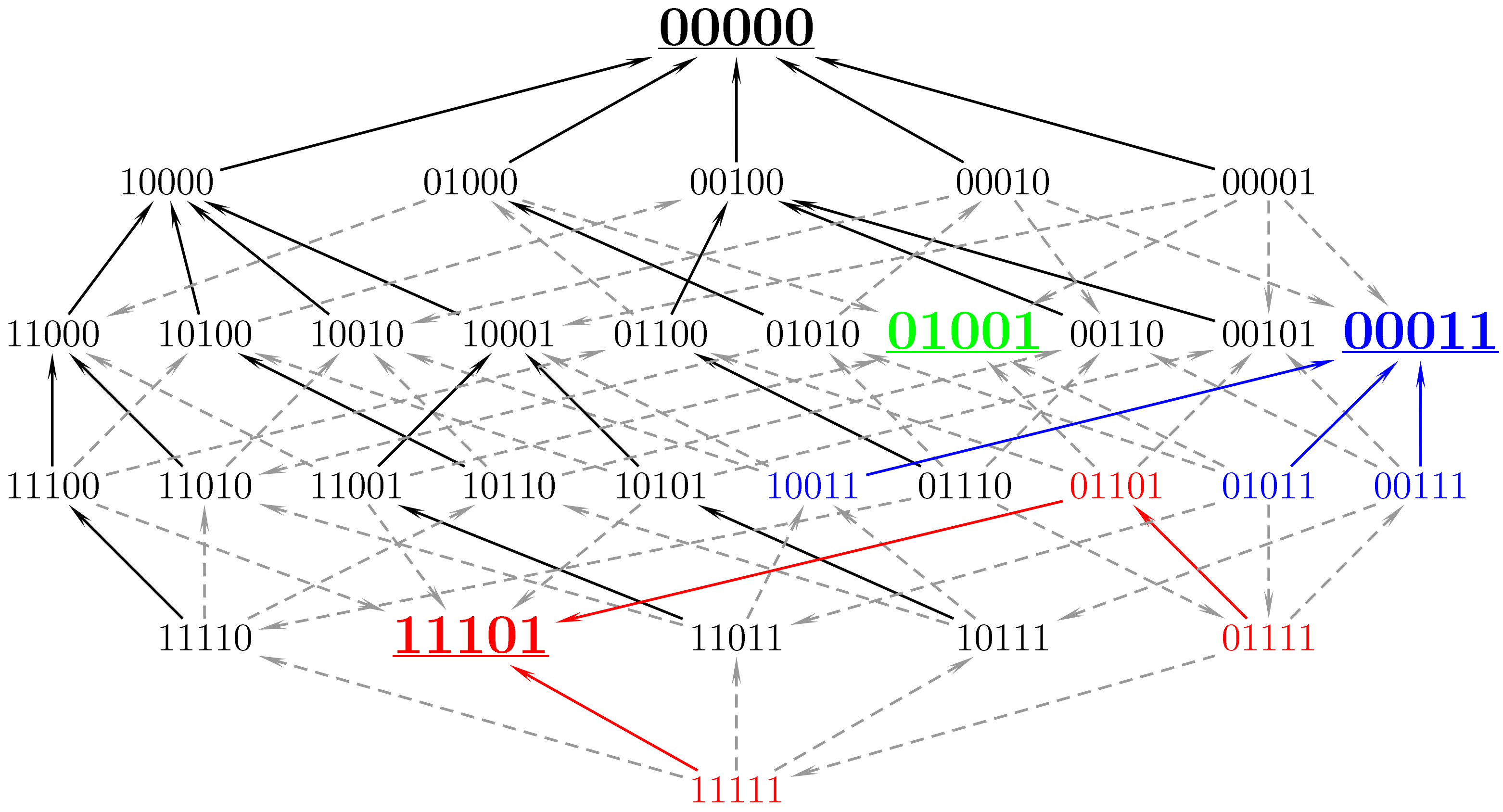
A.I. Khan et al., Science **332** (2011) 1193



- 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

J.A.G.M. de Visser et al., Evolution **51**, 1499 (1997)

J. Franke et al., PLoS Comp. Biol. **7** (2011) e1002134

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

- 186 out of $2^8 = 256$ possible combinations were isolated among ~ 2500 segregants

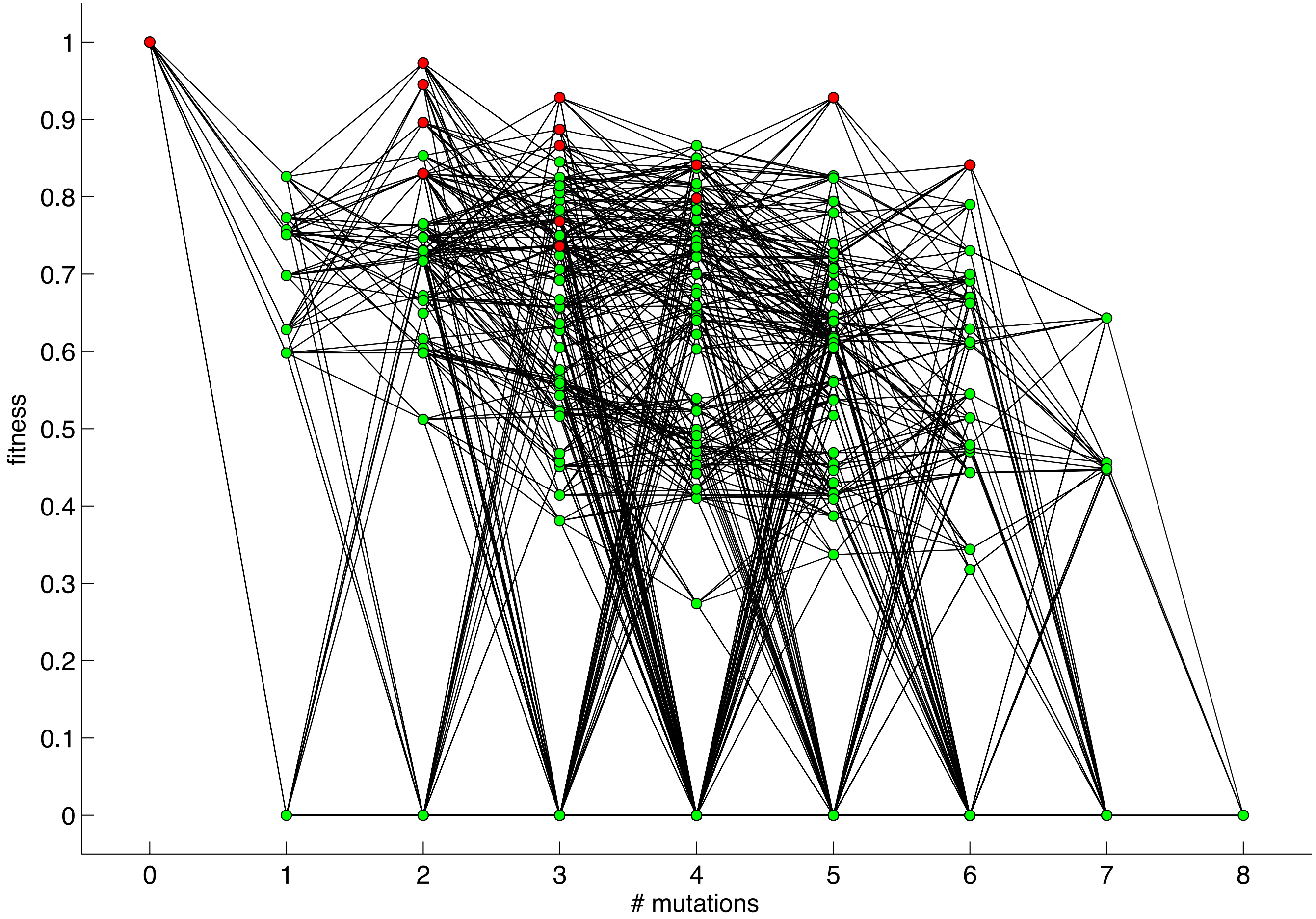
- Fitness (= growth rate) was measured for two replicates per strain

- Fitness relative to wild type falls in the range $w_{\min} = 0.274 \leq w \leq 1$

- Likelihood of missing more than one strain with fitness $> w_{\min}$ is < 5 %
  $\Rightarrow$ assign zero fitness to missing strains

- Lethals are mostly associated with lysine deficiency (62 out of 70)

The *Aspergillus niger* fitness landscape



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

Summary of empirical data sets

Szendro et al., JSTAT P01005 (2013)

ID	System (*organism*/gene)	L	Available combinations	Fitness (proxy)	Direction of mutations	Known effects
A	*Methylobacterium extorquens*	4	16/16	Growth rate	Beneficial	Combined
B	*Escherichia coli*	5	32/32	Fitness	Beneficial	Combined
C-D	Dihydrofolate reductase	4	16/16	Resistance/ Growth rate	Beneficial	Individual/ Combined
E	β-lactamase	5	32/32	Resistance	Beneficial	Combined
F	β-lactamase	5	32/32	Resistance	Beneficial	Combined
G	*Saccharomyces cerevisiae*	6	64/64	Growth rate	Deleterious	Individual
H	*Aspergillus niger*	8	186/256	Growth rate	Deleterious	Individual
I-J	Terpene synthase	9	418/512	Enzymatic specificity	–	–

Measures of epistasis

Local fitness maxima

Haldane 1931, Wright 1932

- A genotype σ is a local maximum if $f(\sigma) > f(\sigma')$ for all one-mutant neighbors σ'
- In the absence of sign epistasis there is a single global maximum
- Reciprocal sign epistasis is a necessary but not sufficient condition for the existence of multiple fitness peaks Poelwijk et al. 2011

Selectively accessible paths

Weinreich et al. 2005

- A path of single mutations connecting two genotypes $\sigma \longrightarrow \sigma'$ with $f(\sigma) < f(\sigma')$ is selectively accessible if fitness increases monotonically along the path
- In the absence of sign epistasis all direct paths to the global optimum are accessible, and vice versa

- Roughness to slope ratio r/s: Fit the most general non-epistatic model $f_{\mathrm{lin}}(\sigma) = a^{(0)} + \sum_{j=1}^{L} a_j^{(1)} \sigma_j$ and define slope s and roughness r through

$$s = \frac{1}{L}\sum_{j=1}^{L} |a_j^{(1)}|, \quad r^2 = 2^{-L}\sum_{\sigma}(f(\sigma) - f_{\mathrm{lin}}(\sigma))^2$$

Aita et al. 2001

- Fourier spectra: Any fitness landscape can be decomposed into epistatic interactions of different orders

Stadler 1996; Neidhart et al., JTB 2013

$$f(\sigma) = a^{(0)} + \sum_{j=1}^{L} a_j^{(1)}\sigma_j + \sum_{\substack{j,k=1 \\ j>k}}^{L} a_{jk}^{(2)}\sigma_j\sigma_k + \ldots + a^{(L)}\sigma_1\sigma_2\ldots\sigma_L$$

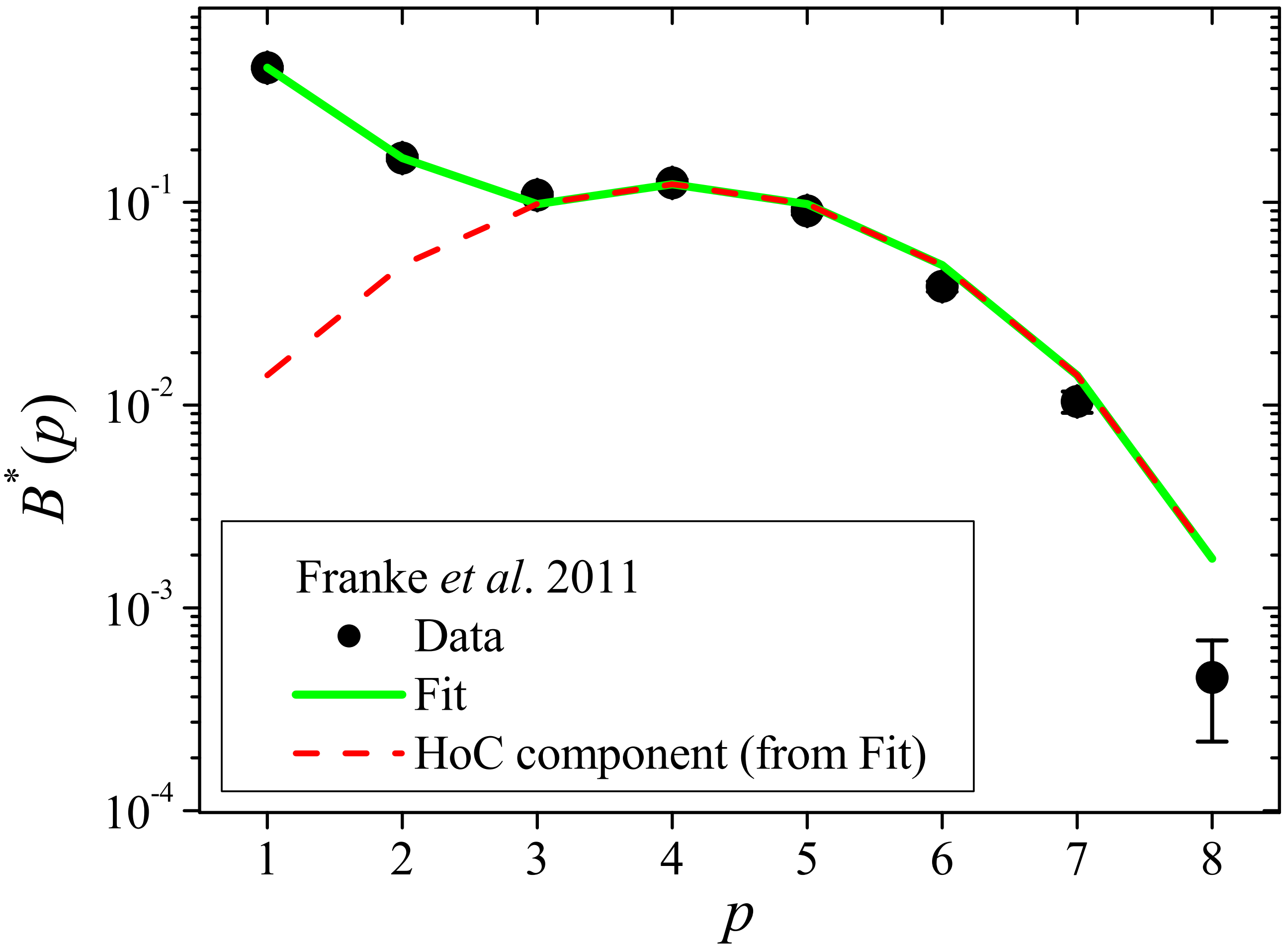
$\Rightarrow$ weight of epistatic interactions of order n is quantified by

$$B_n = \frac{b_n}{\sum_{l=1}^{L} b_l} \text{ with } b_l = \sum_{j=1}^{\binom{L}{l}} (a_j^{(l)})^2, \;\; n = 2, \ldots, L$$

- Local epistasis measures: Count fraction of 2-locus motifs that display 'simple' sign epistasis f_s and reciprocal sign epistasis f_r

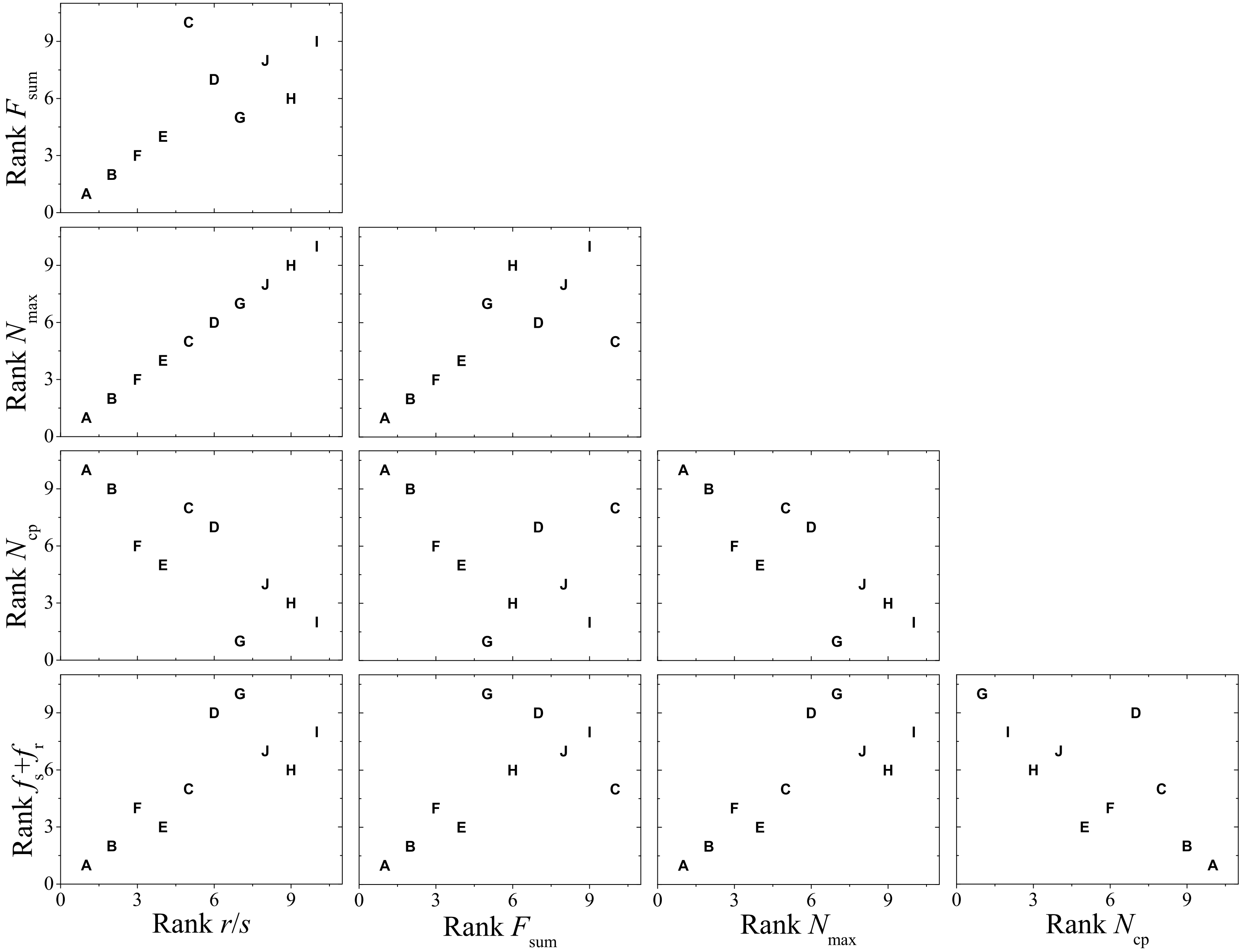
Fourier spectrum of the *Aspergillus niger* landscape

J. Neidhart, I.G. Szendro, JK, JTB 332:218 (2013)

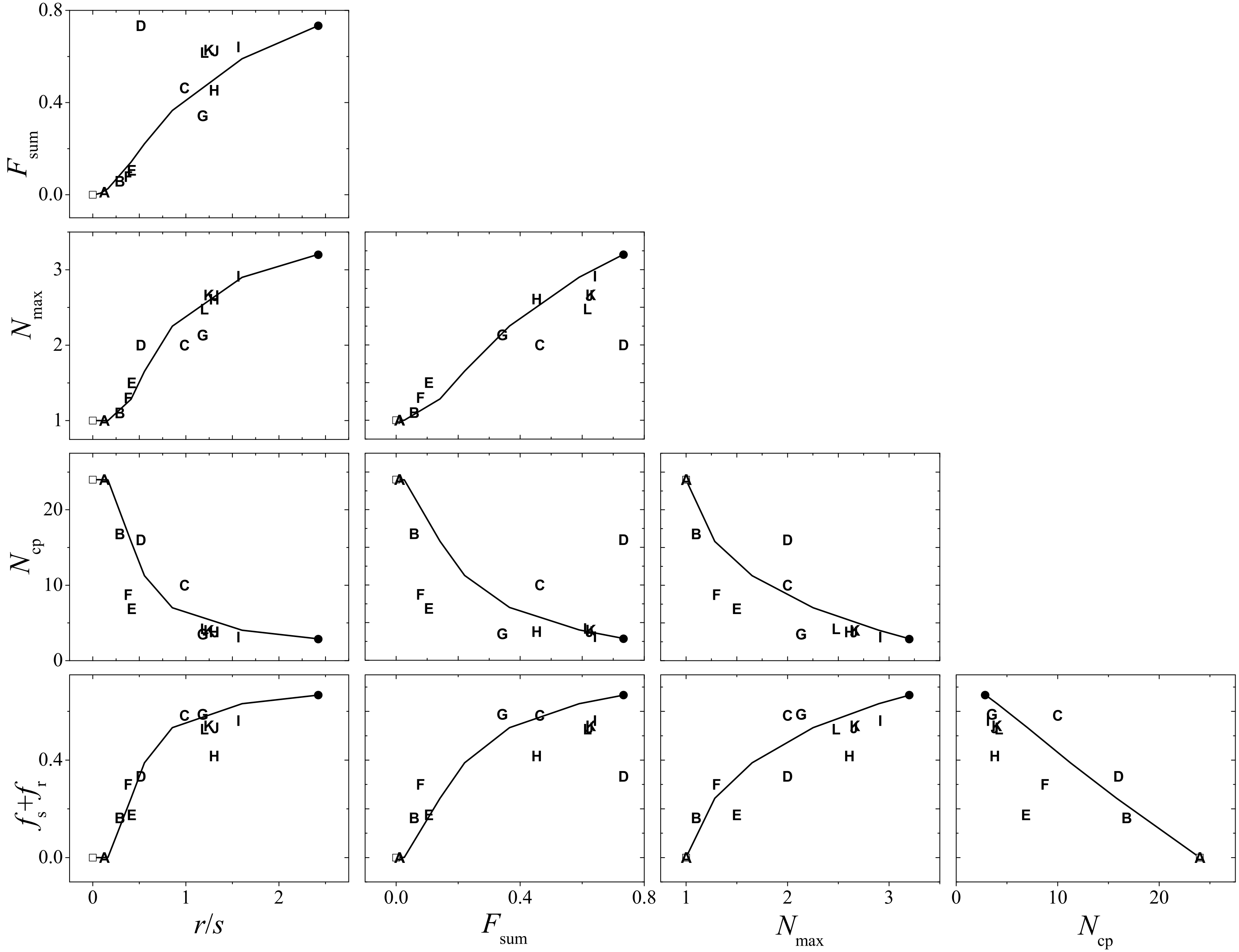


- Dashed line shows spectrum of an uncorrelated random landscape $\tilde{B}_p = 2^{-L}\binom{L}{p}$

Rank correlation between epistasis measures



Parametrization by a rough Mount Fuji model



Patterns of epistasis in an antibiotic resistance enzyme

Patterns of epistasis

- Comparative studies of empirical fitness landscapes reveal generic features but also characteristic differences

- In particular, the choice of the subset of mutations used to construct a landscape biases the patterns of epistasis:
  - singly beneficial vs. singly deleterious mutations
  - mutations chosen for individual or collective effects
  - mutations in the same gene or different genes
  - mutations occurring along an adaptive trajectory

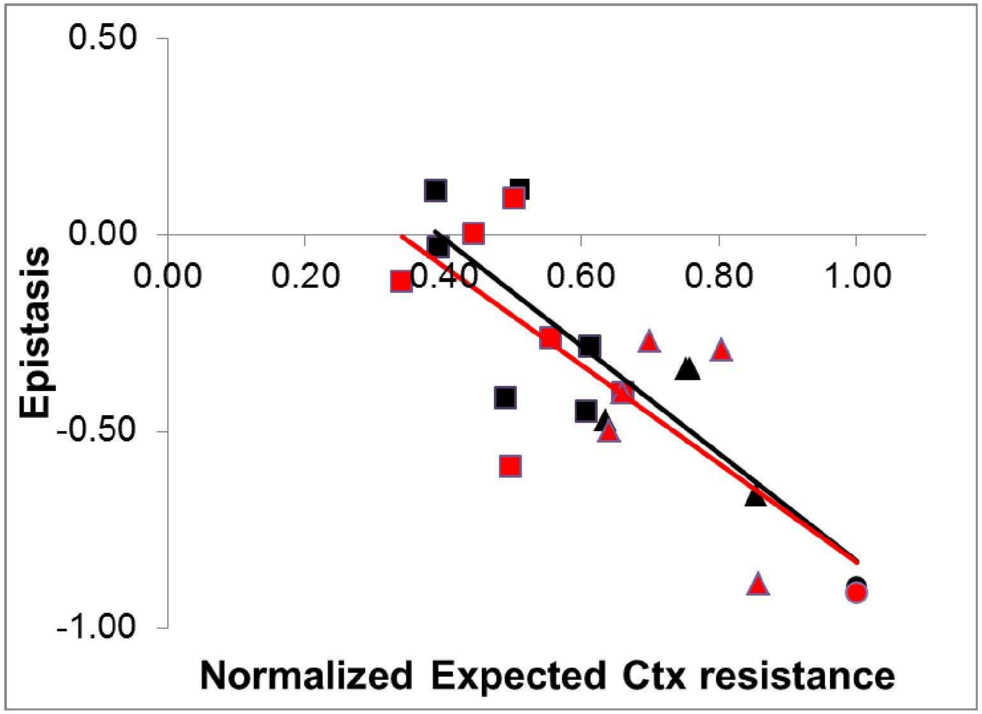
A case study:

M.F. Schenk et al., Mol. Biol. Evol. 2013

- Comparative analysis of two subsets of 4 mutations each chosen from the pool of 48 individually beneficial mutations in TEM-1 β-lactamase

- Mutations chosen according to effect on resistance (weakly vs. strongly beneficial)

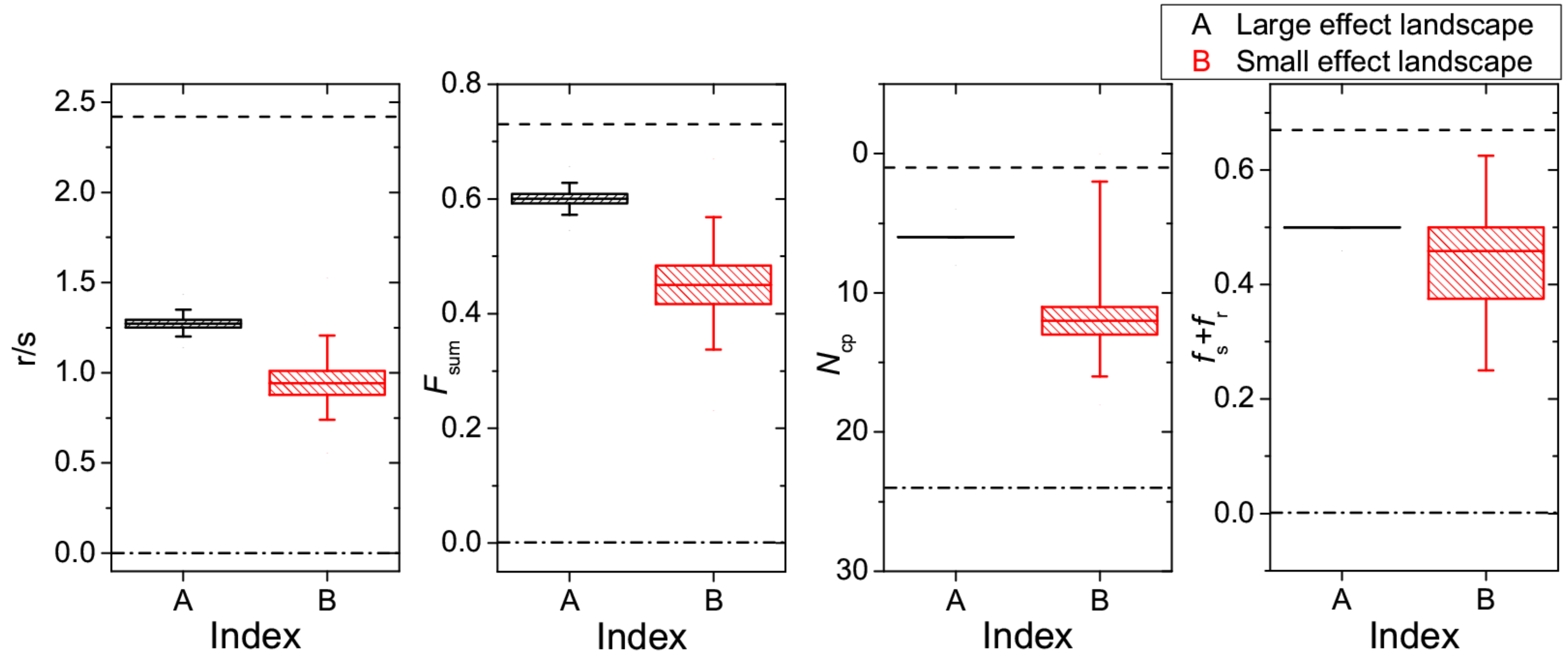
Diminishing returns epistasis



- Resistance of multiple mutants is lower than expected, and the deviation increases with effect strength
- Generic pattern that appears also in multicellular organisms

Schoustra et al., Proc. Roy. Soc. B 2016

Large effect landscape is consistently more rugged



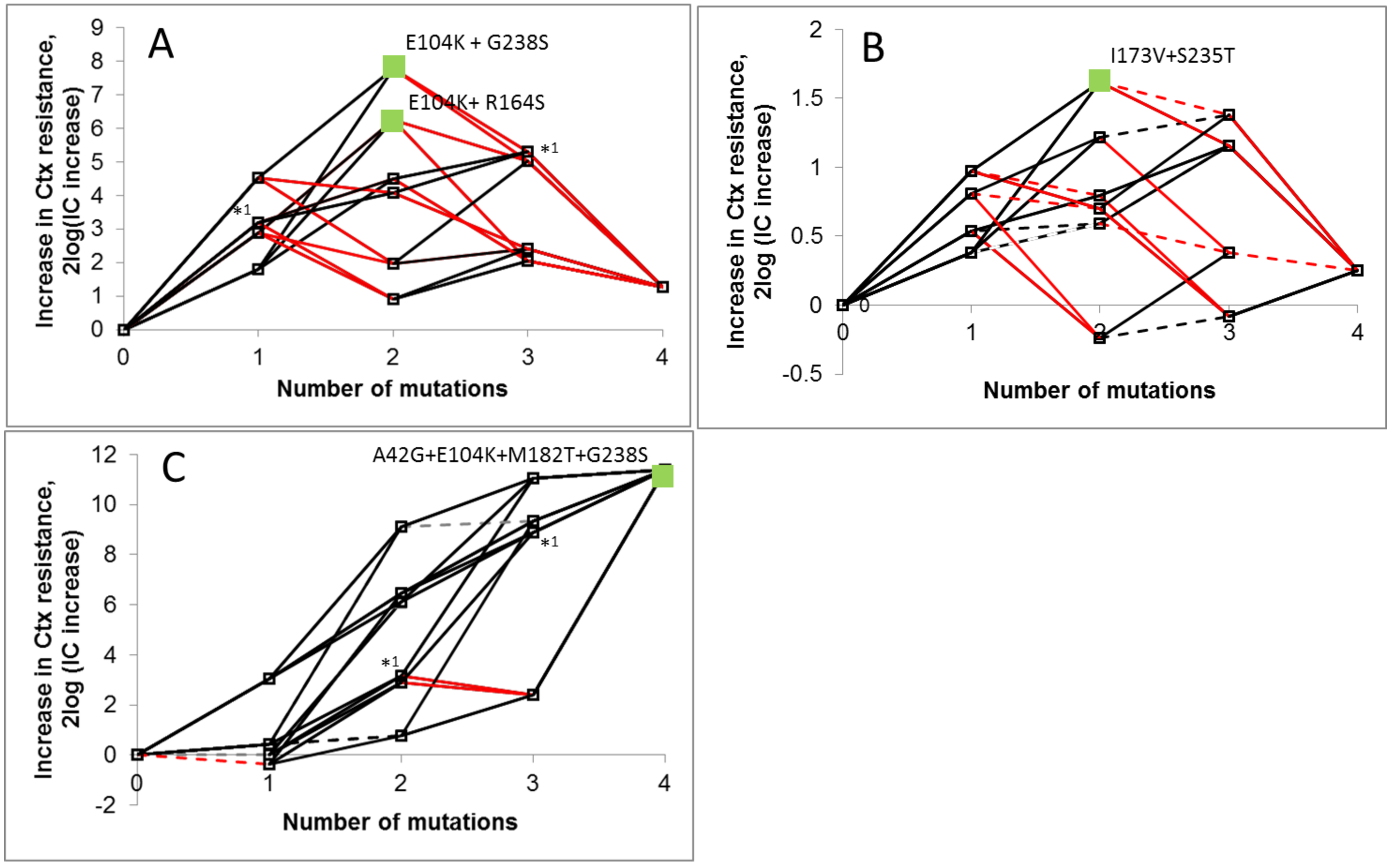
r/s: Roughness-to-slope ratio

F_{sum}: Relative weight of interactions

N_{cp}: Number of accessible paths

$f_s + f_r$: Fraction of sign-epistatic pairs

Mutations chosen for individual vs. collective effect



A: Large effect

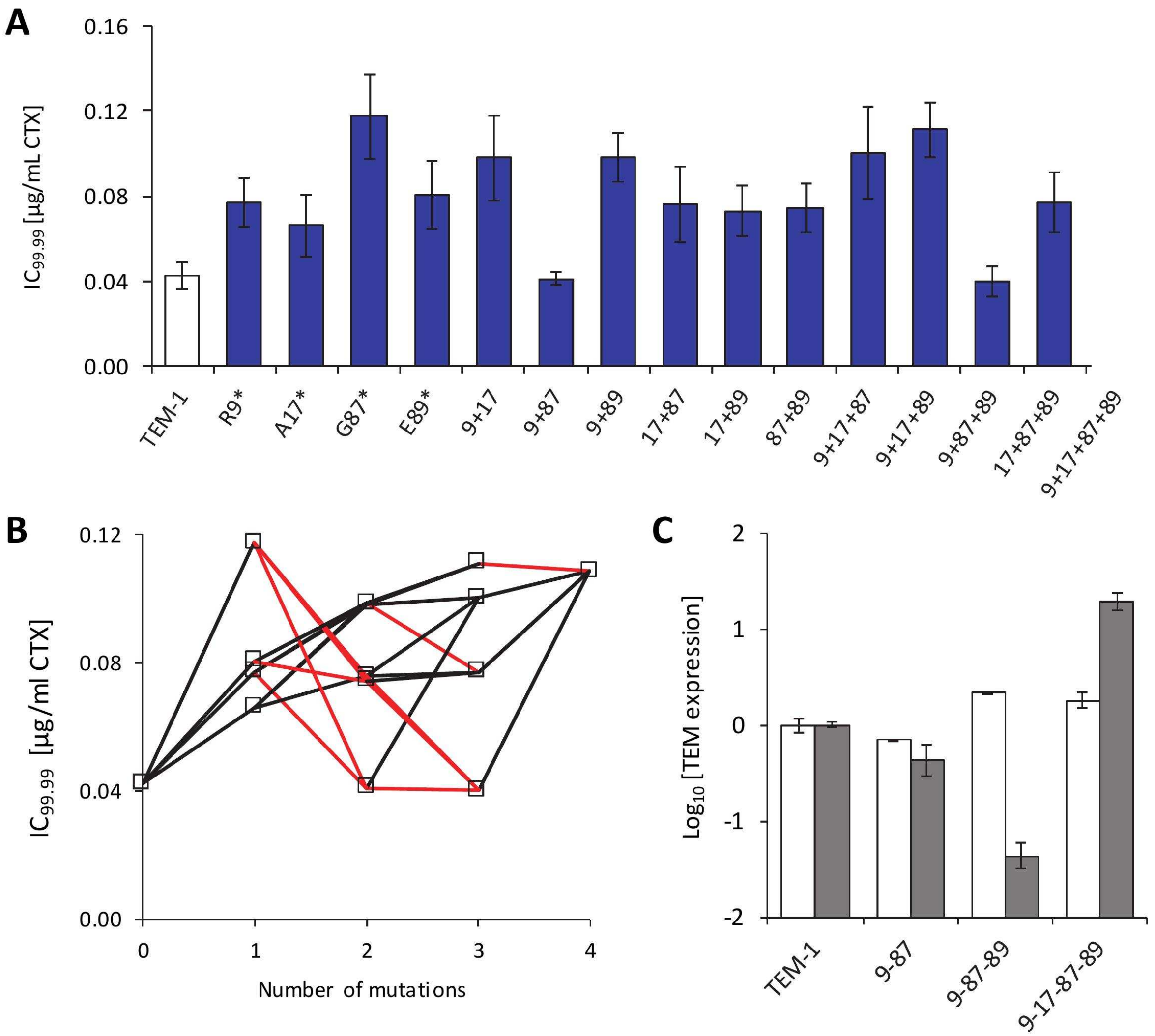
B: Small effect

C: Weinreich 2006

- Mutations chosen for individual effect interact more strongly and negatively than mutations chosen “with hindsight” because of their collective effect

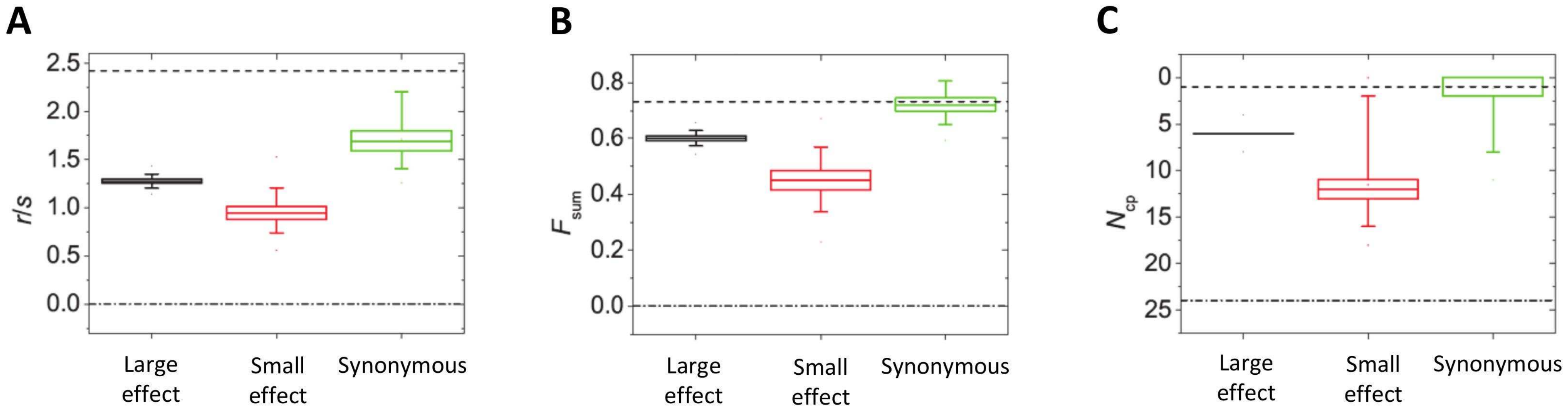
A landscape of synonymous mutations

M.P. Zwart et al. (under review)



The synonymous landscape is most rugged...

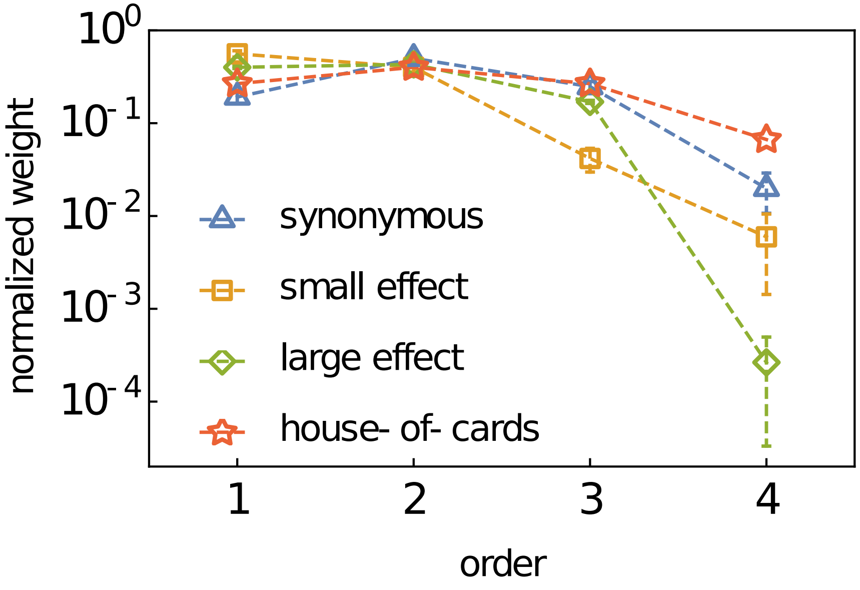
M.P. Zwart et al. (under review)



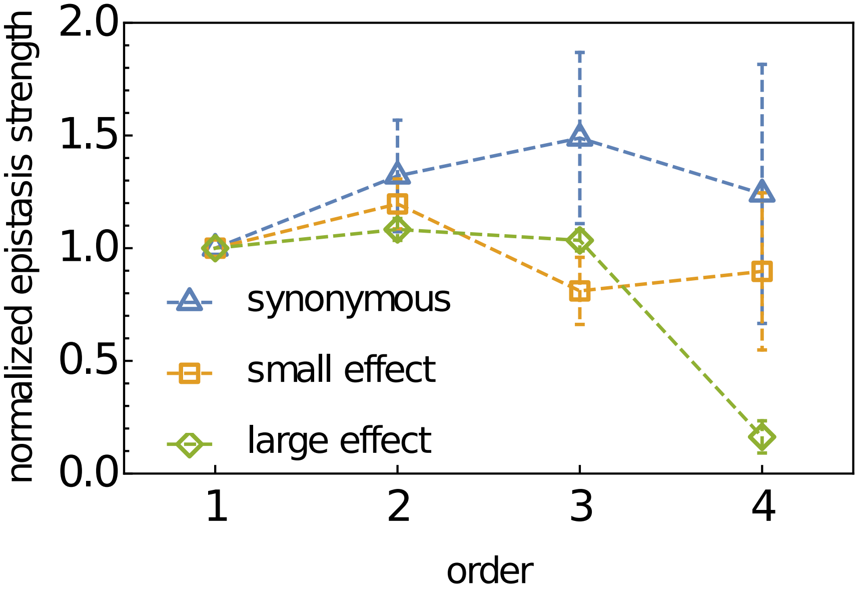
- For two of three measures the landscape is indistinguishable from an uncorrelated random landscape

...and displays a lot of high order epistasis

M.P. Zwart et al. (under review)



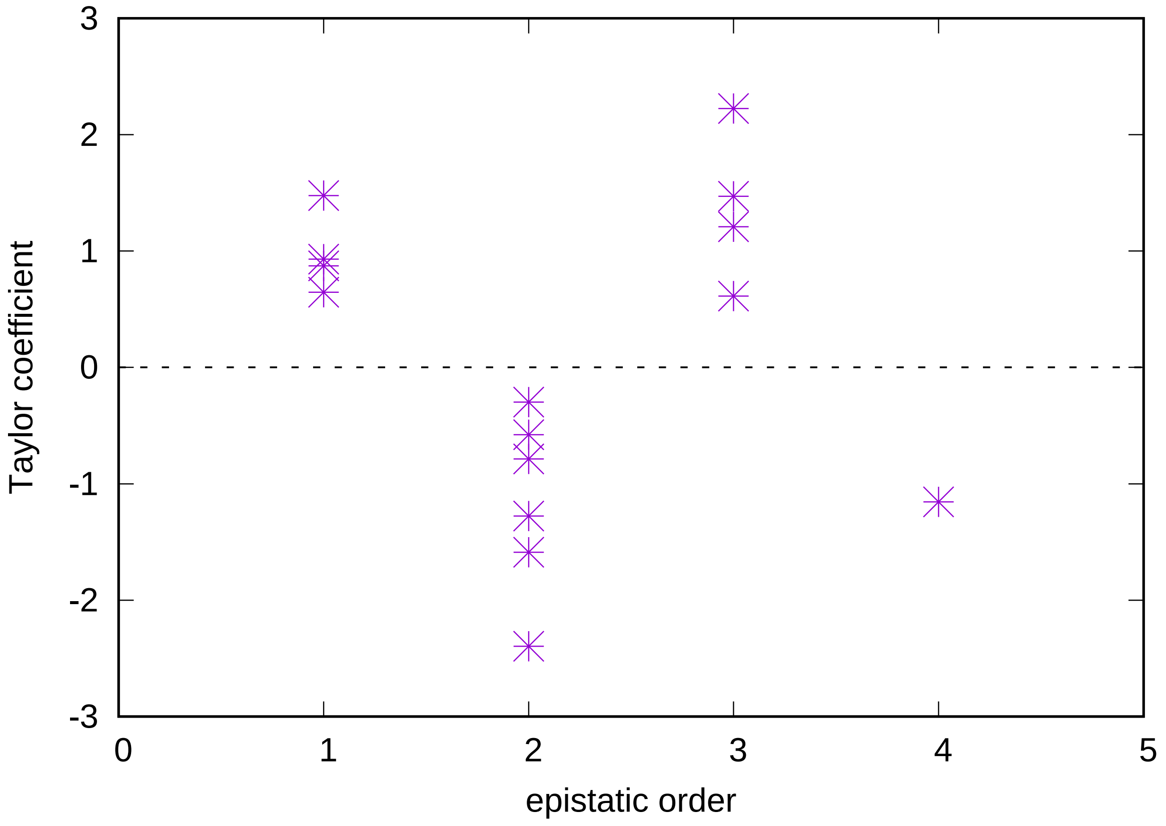
Fourier



Taylor

Taylor coefficients for the synonymous landscape

M.P. Zwart et al. (under review)



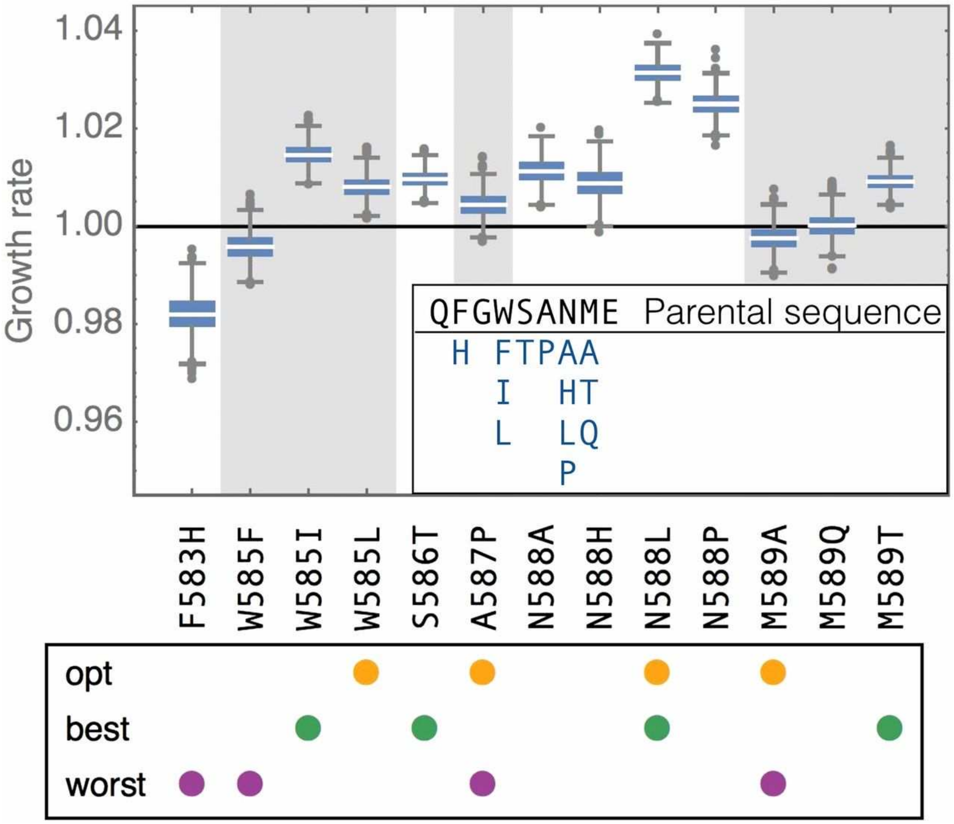
- Alternating sign pattern is induced by conditioning on positive single mutant effects

Patterns of epistasis in a large intragenic fitness landscape

Bank et al., PNAS 2016

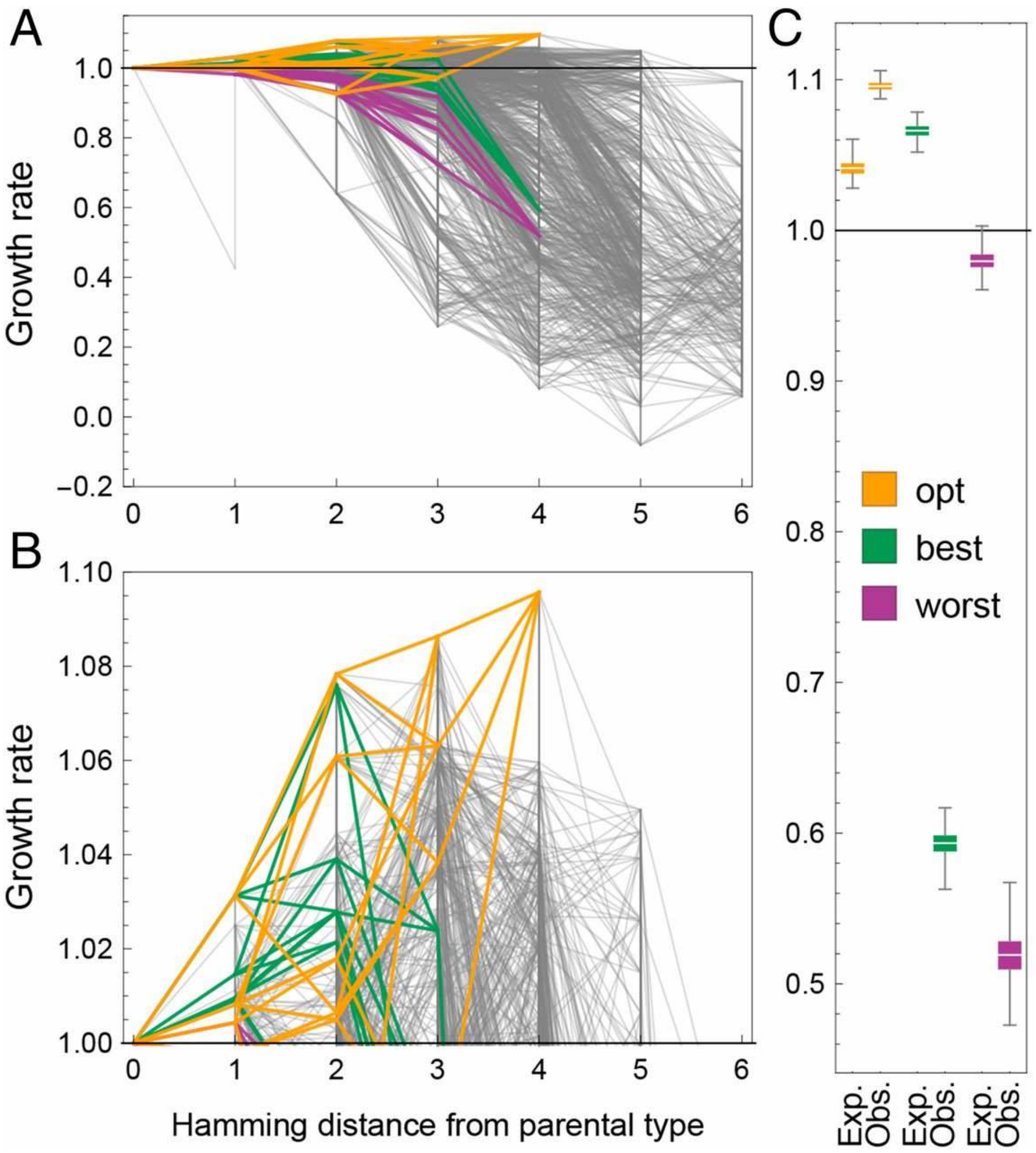
A large intragenic fitness landscape

- All combinations of 13 mutations at 6 positions of the Hsp90 heat-shock protein in yeast

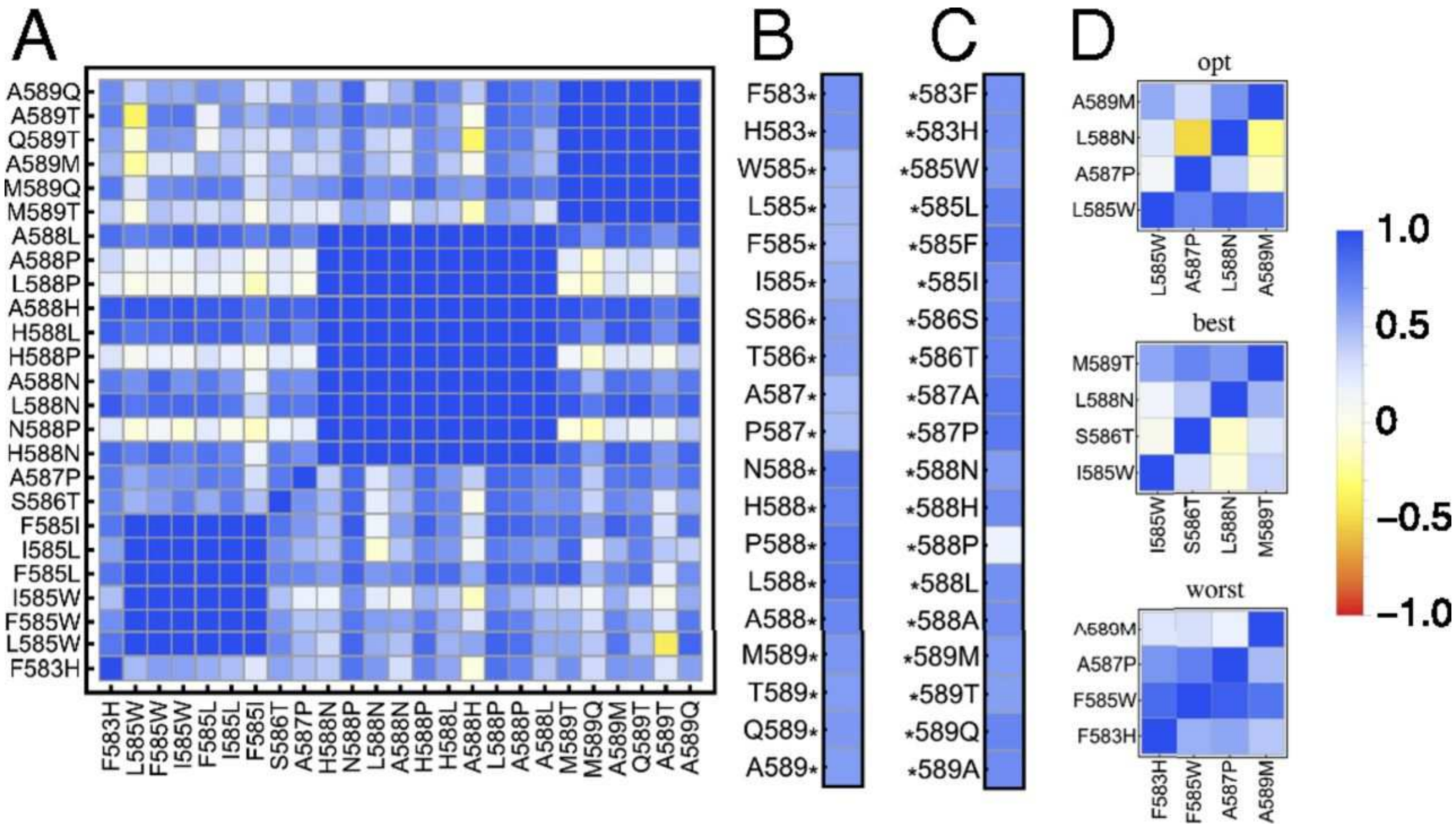


- Growth rate at elevated salinity measured for 640 mutants

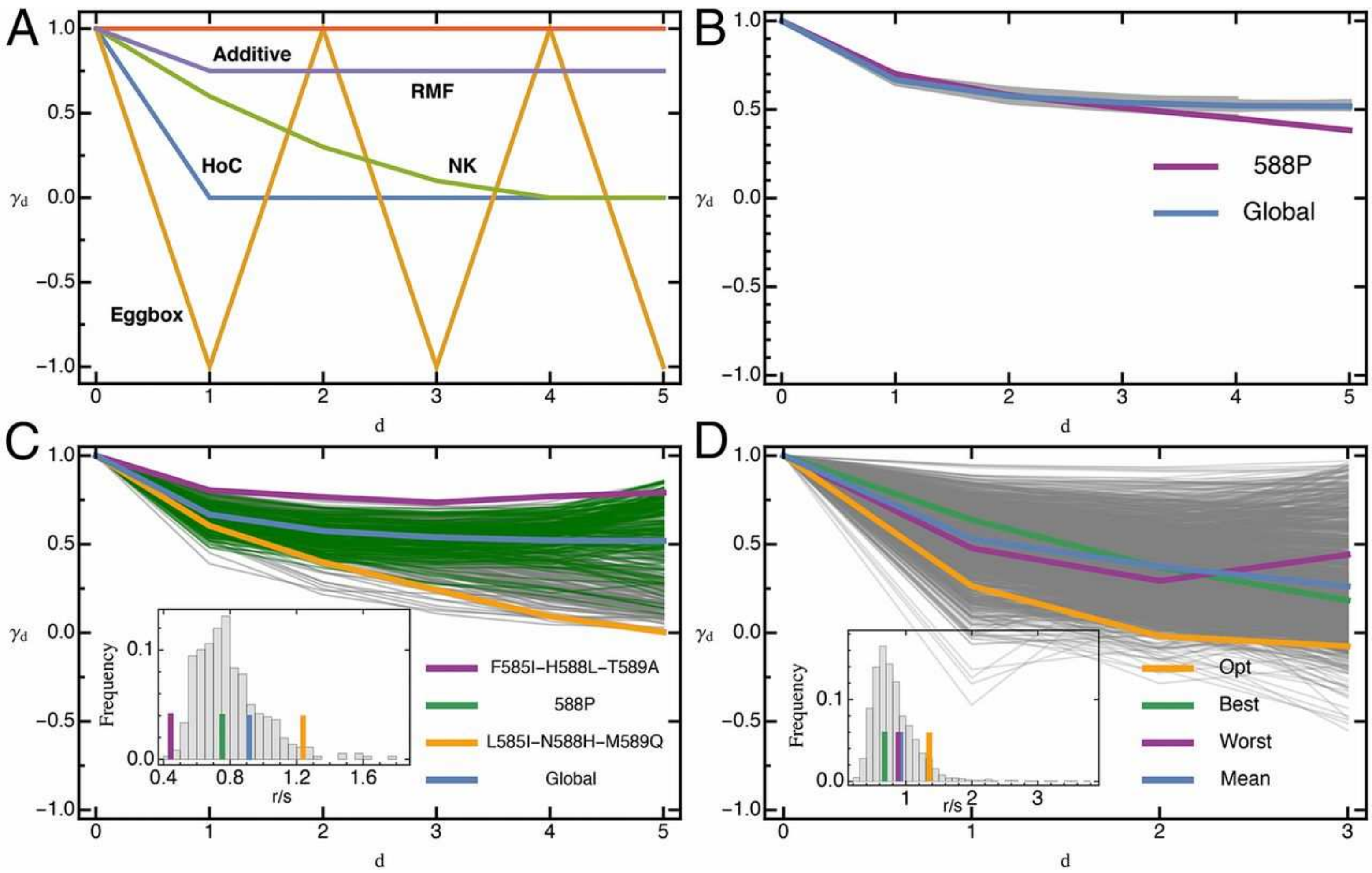
Global view and three focal landscapes



Pairwise epistasis

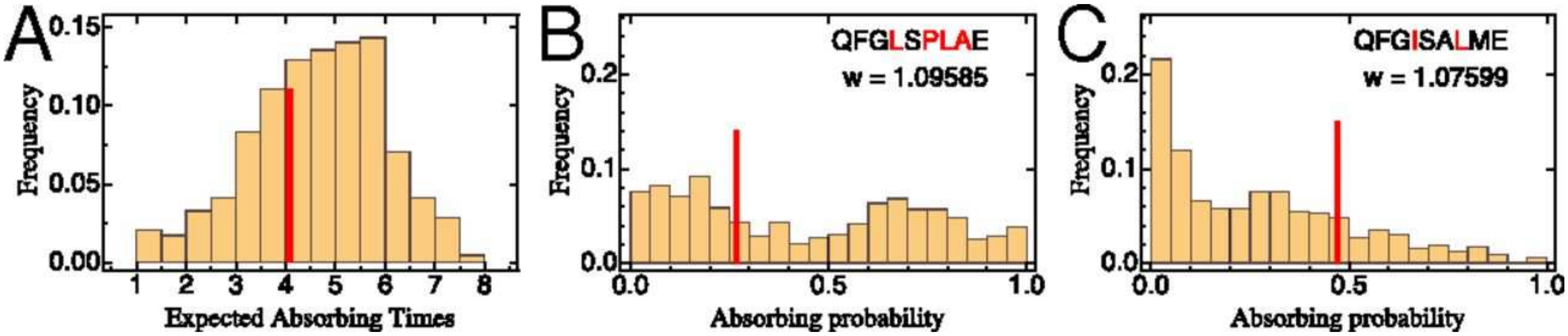


γ-statistics



- Correlation between selection coefficients of the same mutation on two backgrounds that are distance d apart Ferretti et al., JTB 2016

Adaptive walks

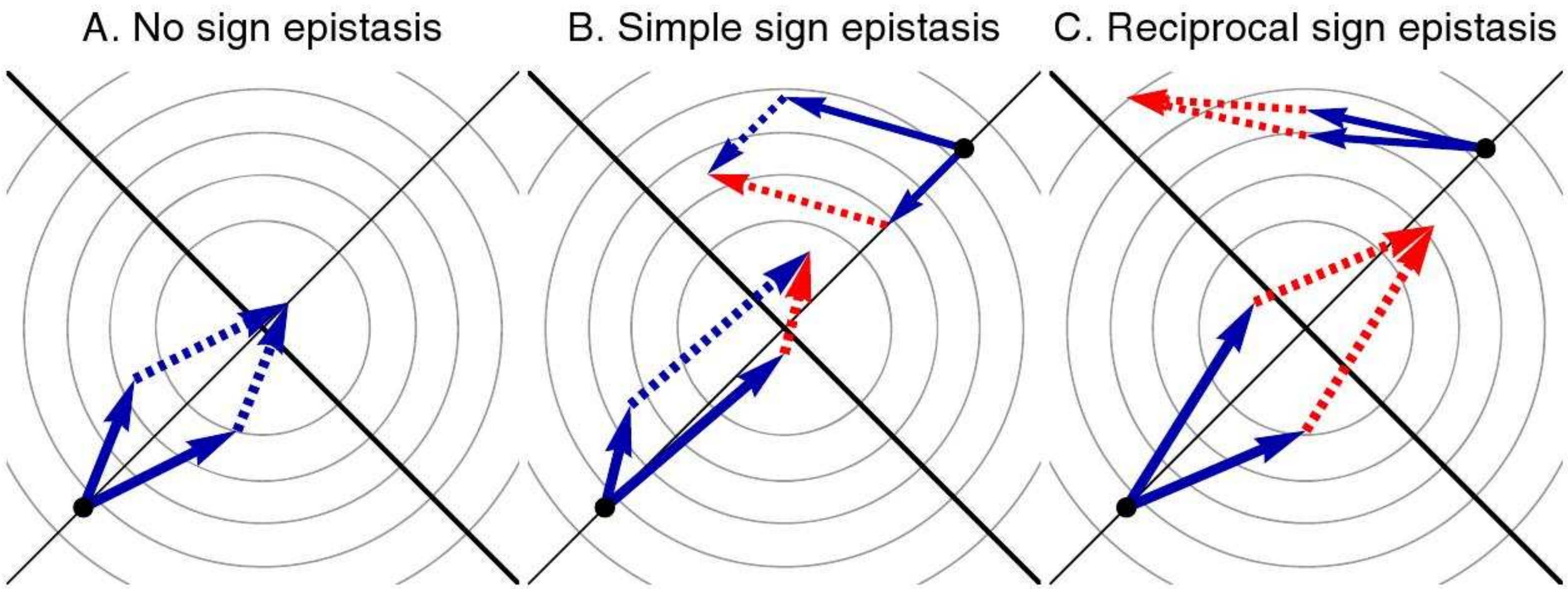


- QFGLSPLAE: Global fitness maximum
- QFGISALME: Local fitness maximum at distance 2 from parental sequence

Genotype-phenotype-fitness maps

Sign epistasis in Fisher’s geometric model

Blanquart et al., Evolution (2014)

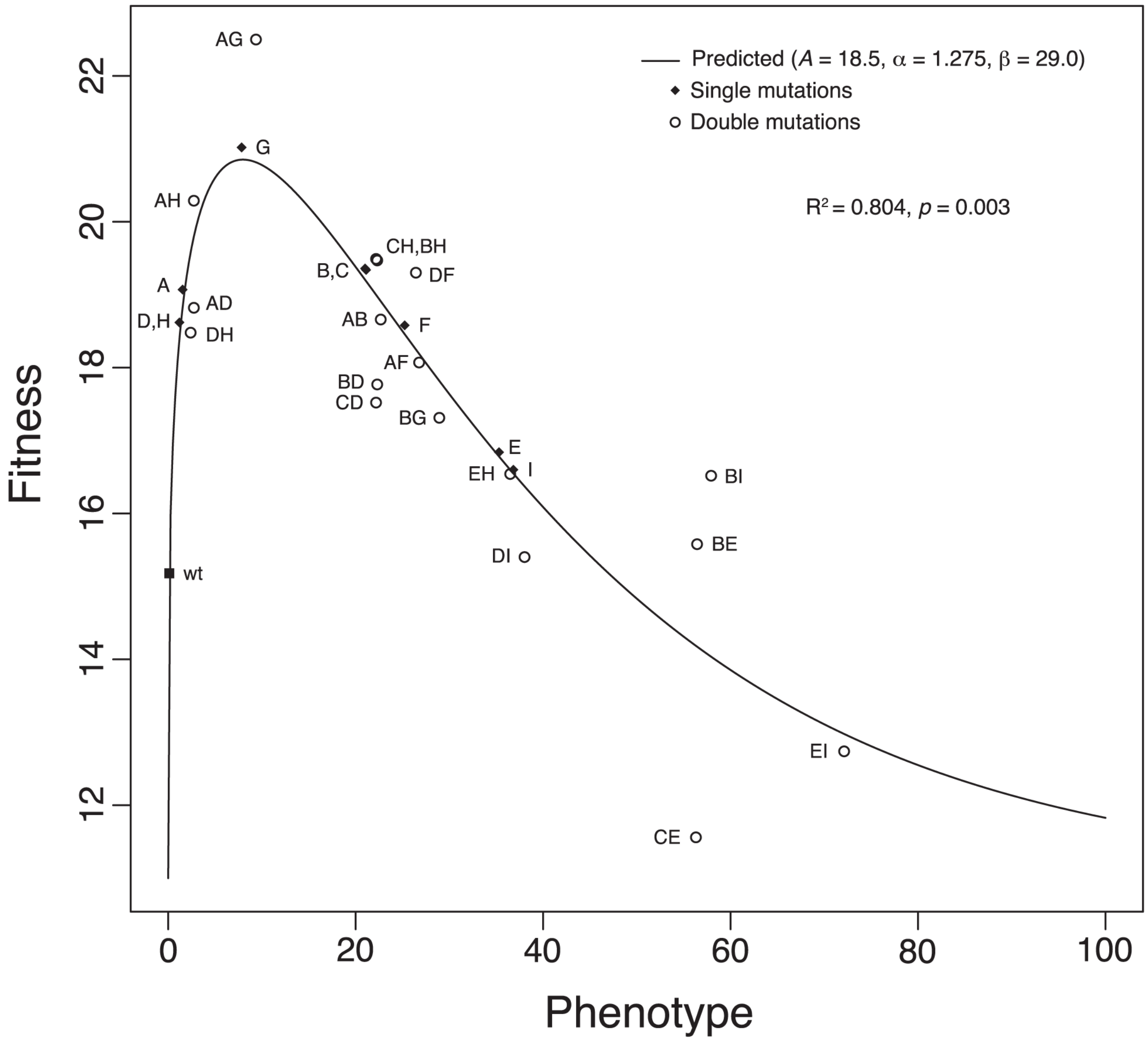


Two distinct mechanisms related to

- the overshooting of the phenotypic optimum or
- (for $n > 1$) the curvature of fitness isoclines (antagonistic pleiotropy)

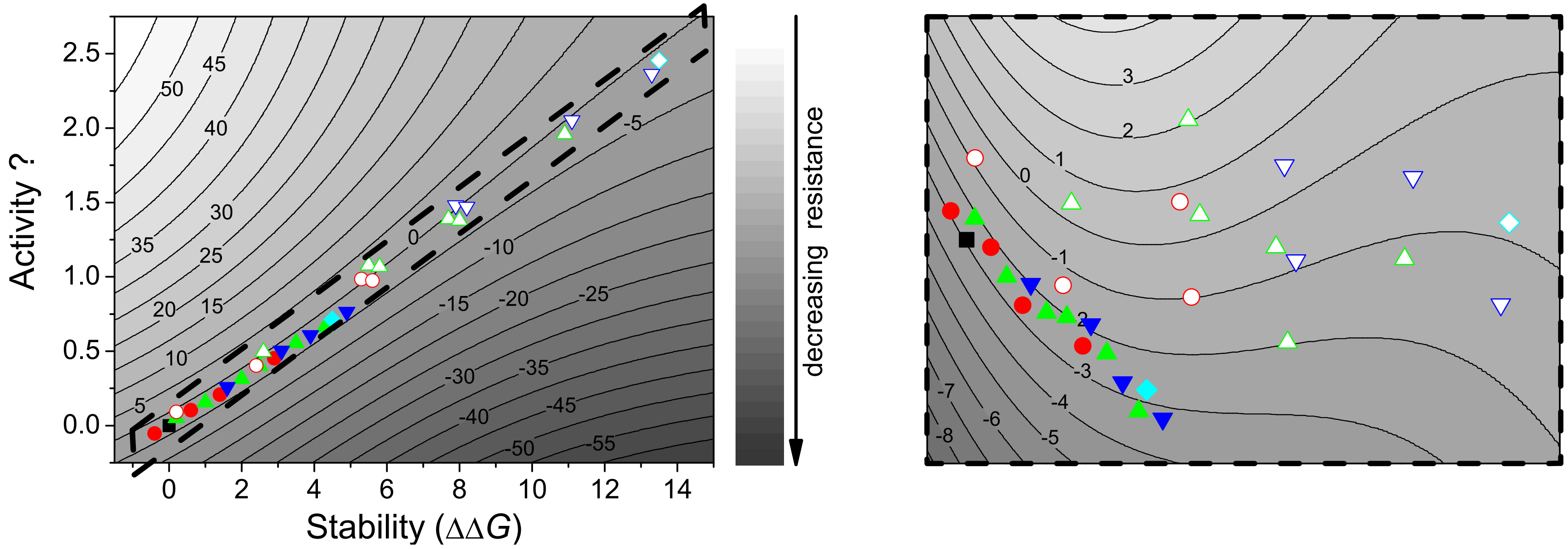
One-dimensional example: Bacteriophage ID 11

Rokyta et al., PLOS Genetics 2011



Two-dimensional example: TEM-1 β-lactamase

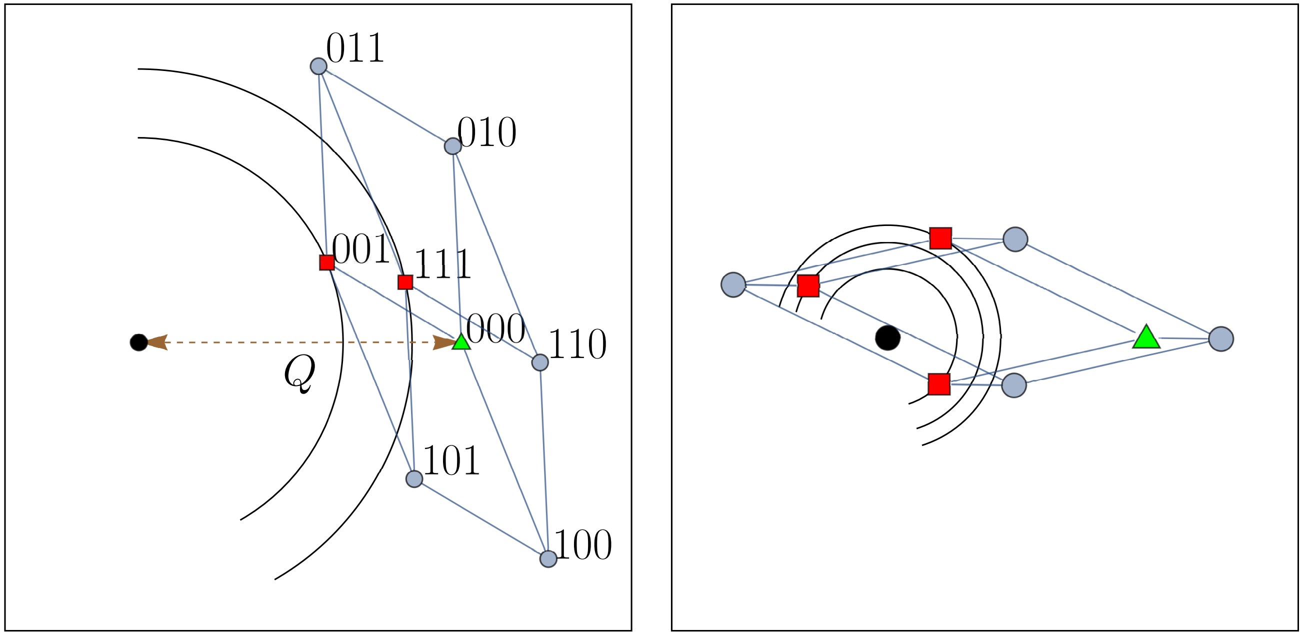
M.F. Schenk et al., Mol. Biol. Evol. (2013)



- Genotypic landscapes constructed from two sets of mutations increasing resistance against cefotaxime
- Data of large (open) and small (filled) effect landscapes are well described by a two-dimensional phenotype-fitness map without an optimal phenotype

Genotype-phenotype-fitness map for multiple loci

Hwang et al., Genetics 2017



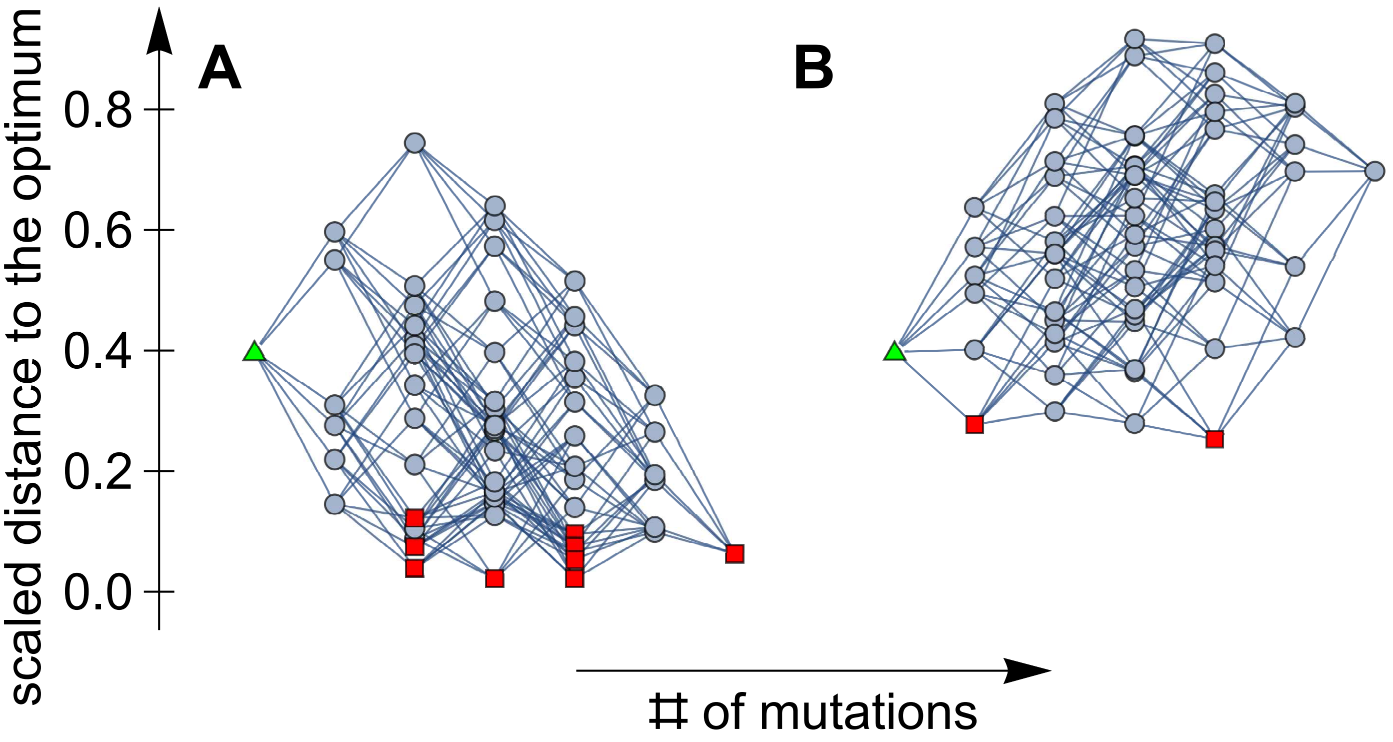
- The mapping $\tau \rightarrow \vec{z}(\tau) = \vec{Q} + \sum_{i=1}^{L} \tau_i \vec{m}_i$ projects L-dimensional hypercube onto n-dimensional phenotype space

- Figure shows the wild type phenotype (green triangle) and genotypic fitness maxima (red squares) for $L = 3, n = 2$

Coexistence of different landscape structures

- Coexistence leads to a large heterogeneity of landscape structures that has been observed in previous work

Blanquart et al. 2014; Blanquart and Bataillon 2016



Two landscape realizations for $q = 0.5, L = 6, n = 2$

Diminishing returns epistasis in *Aspergillus nidulans*

S. Schoustra et al., Proc. Roy. Soc. B (2016)

- 244 beneficial mutants of *A. nidulans* collected from the boundary of growing colonies in complex (rich) or minimal (poor) medium

- Generated 55 pairwise combinations between mutations of similar effect using sexual crosses

- Goal: Quantify the dependence of pairwise epistatic interaction

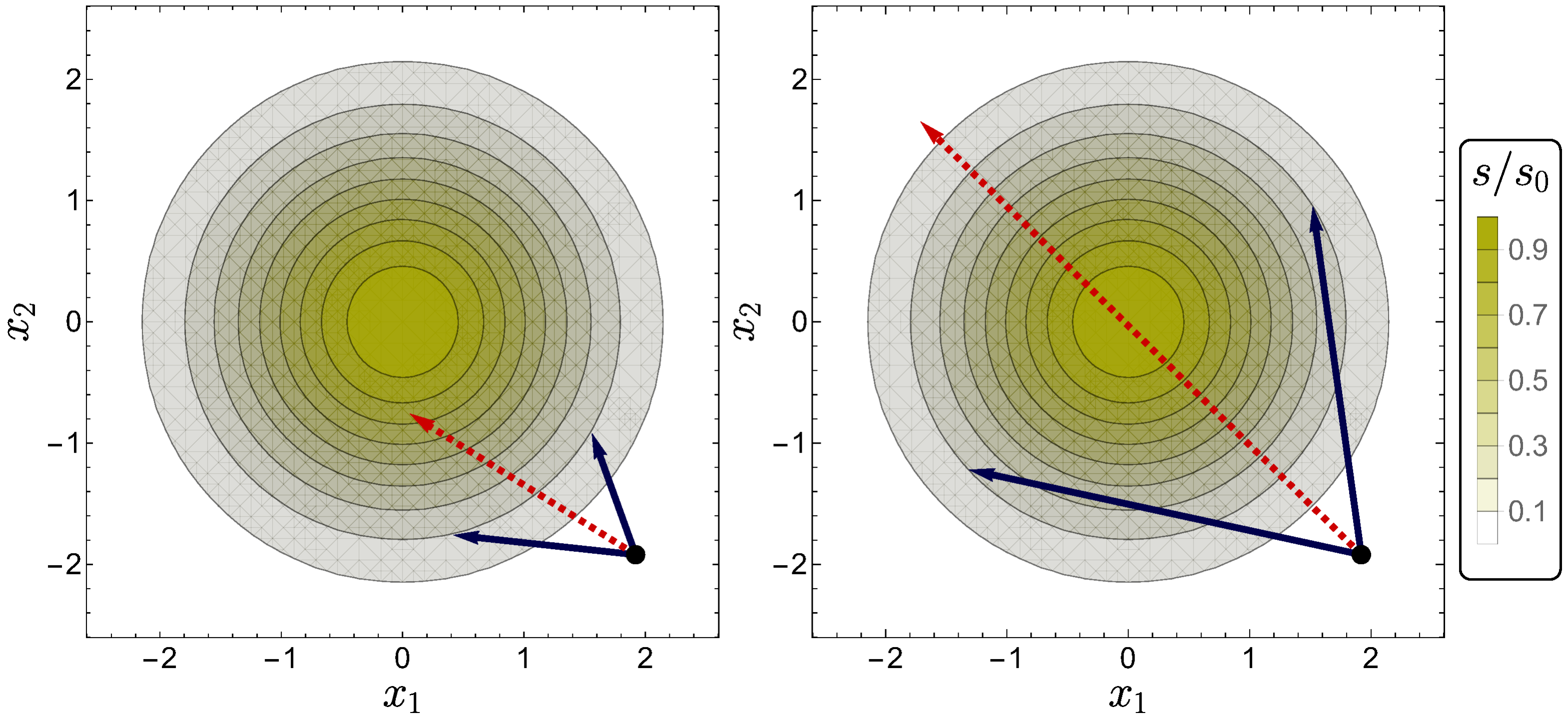
$$\varepsilon_{ab} = \Delta f_{ab} - (\Delta f_a + \Delta f_b)$$

  on the strength $s \approx \Delta f_a \approx \Delta f_b$ of single mutations

- Since double mutant fitness is determined by measuring the growth rate of colonies containing all four types, it can be detected only if

$$\Delta f_{ab} > \max\{\Delta f_a, \Delta f_b\} = s \quad \text{or} \quad \varepsilon_{ab} > -s$$

Diminishing returns epistasis from FGM

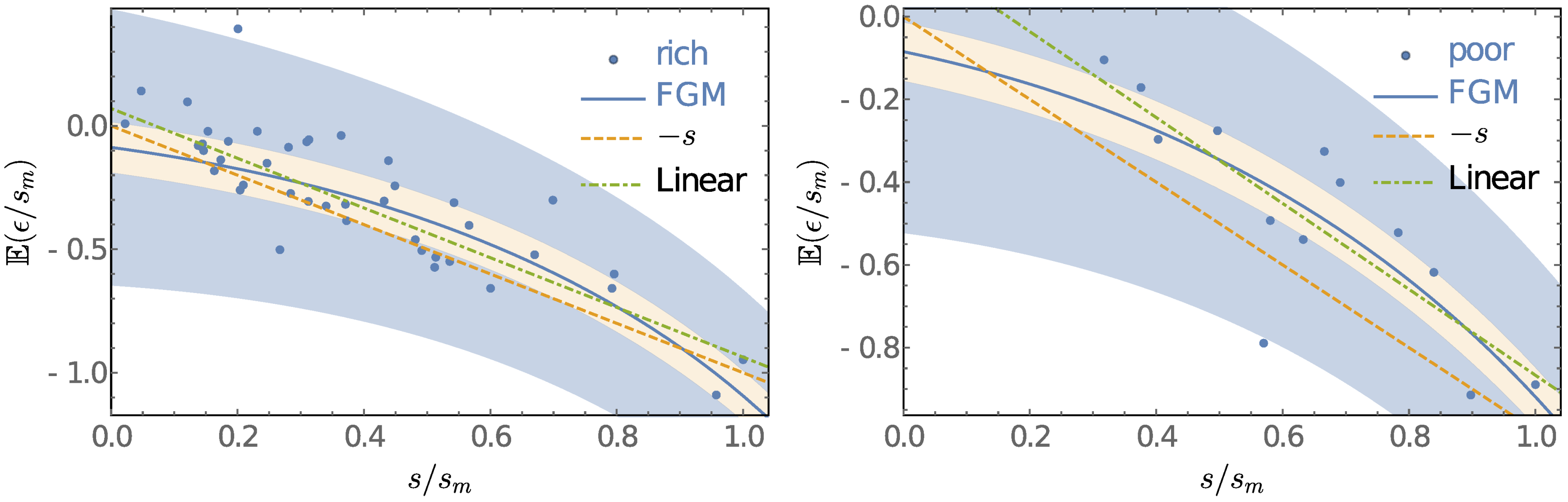


magnitude epistasis

sign epistasis

- Pairs of mutations with the same fitness effect differ widely in their epistatic interactions
- FGM contains a mechanism of intrinsic variability

Fit of FGM to data



- Measurement error (inner pink region) is insufficient to explain variability
- Crowding of data points around the line $\varepsilon = -s$; outliers below this line originate from a tradeoff between germination and growth
- FGM parameters: $d/\sigma = 6.89,\ n = 19.3,\ s_0/s_m = 1.41$ (rich)
  $d/\sigma = 9.81,\ n = 34.8,\ s_0/s_m = 1.62$ (poor)