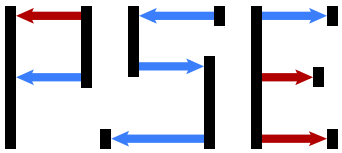


PROBABILISTIC STRUCTURES
IN EVOLUTION

DFG SPP 1590

COLLABORATIVE RESEARCH CENTER | SFB 680
Molecular Basis of
Evolutionary Innovations

Mutational pathways in complex fitness landscapes

Joachim Krug
Institute for Theoretical Physics, University of Cologne

joint work with Arjan de Visser & group (Wageningen)

Modeling Tumor Evolution: Initiation, Growth and Progression
ZiF Bielefeld, September 14, 2016

Cancer as an evolutionary process

- Cancer initiation and resistance evolution typically requires multiple mutational steps

- These mutations can interact
  - directly in terms of their phenotypic effects, or
  - indirectly through the clonal dynamics

- Direct interactions are called epistatic and can be encoded in a multidimensional fitness landscape de Visser & Krug, Nat. Rev. Gen. 2014

Cancer as an evolutionary process

- Cancer initiation and resistance evolution typically requires multiple mutational steps

- These mutations can interact
  - directly in terms of their phenotypic effects, or
  - indirectly through the clonal dynamics

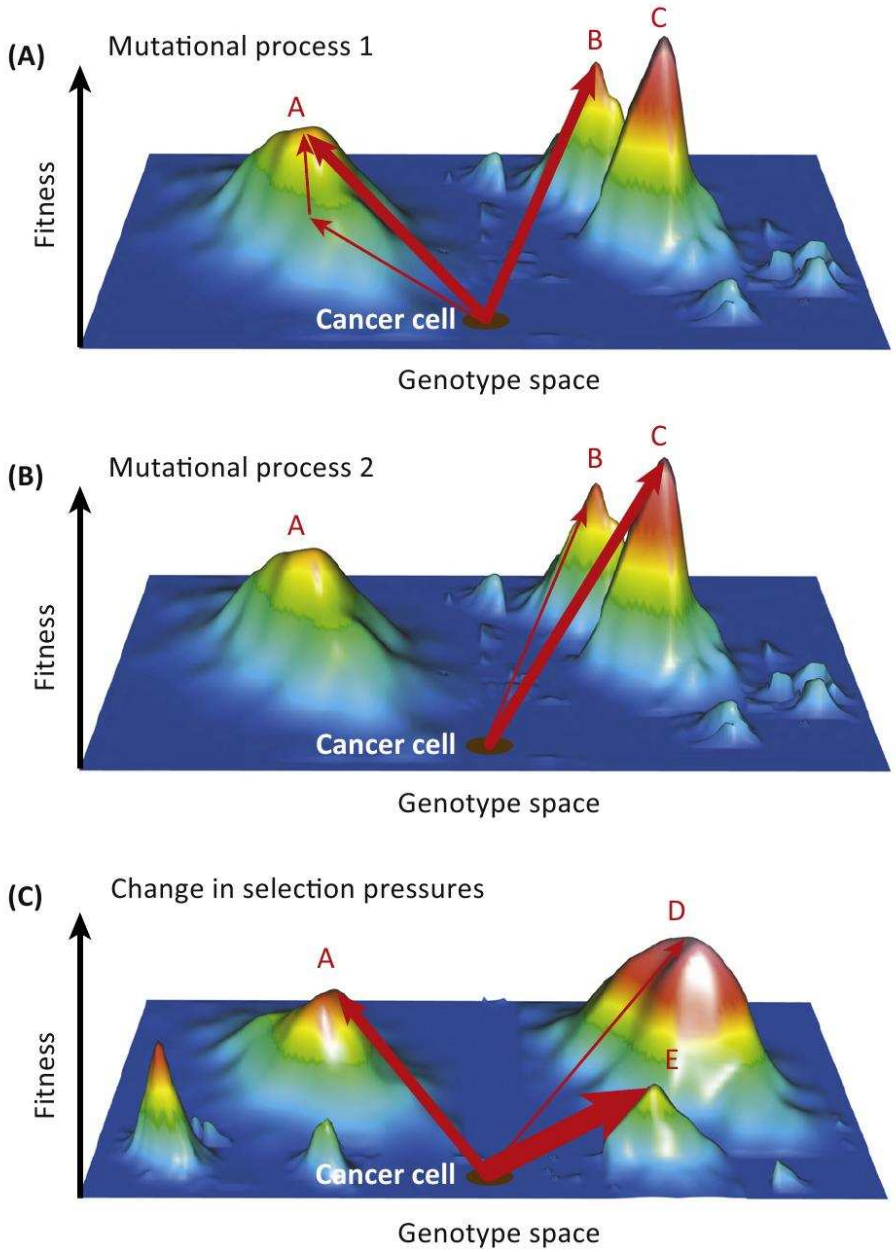
- Direct interactions are called epistatic and can be encoded in a multidimensional fitness landscape de Visser & Krug, Nat. Rev. Gen. 2014

- The focus of this talk will be on epistatic interactions and their consequences for predictability

- Empirical examples will be drawn from antibiotic resistance evolution in bacteria

The fitness landscape metaphor

- Multiple fitness peaks shape the evolutionary process S. Wright 1932
- Only a subset of peaks are accessible by mutational pathways
- How to turn this picture into a quantitative, predictive tool?



Lipinski et al.,
Trends in Cancer 2016

Predictability of a single mutational step

The probability of parallel evolution

H.A. Orr, Evolution 2005

- n beneficial single step mutations are available from the initial genotype
- Each mutant is characterized by its selective advantage $s_i > 0$
- The fixation probability for the i'th mutant is $2s_i$ (Haldane 1927), hence the probability that the i'th mutant is the first to fix is given by

$$\pi_i = \frac{s_i}{\sum_{j=1}^{n} s_j}$$

  and the same mutation is fixed in two replicate populations with probability

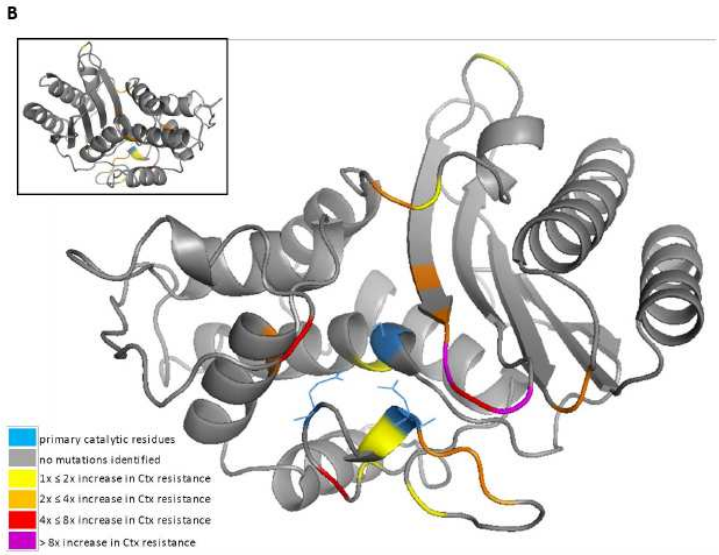
$$P_2 = \sum_{i=1}^{n} \pi_i^2$$

- This quantity is determined by the distribution of beneficial fitness effects

The TEM-1 β-lactamase enzyme

M.F. Schenk, I.G. Szendro, JK, J.A.G.M. de Visser, PLoS Genet. 2012

- β-lactamase confers resistance against penicillin to *E. coli*

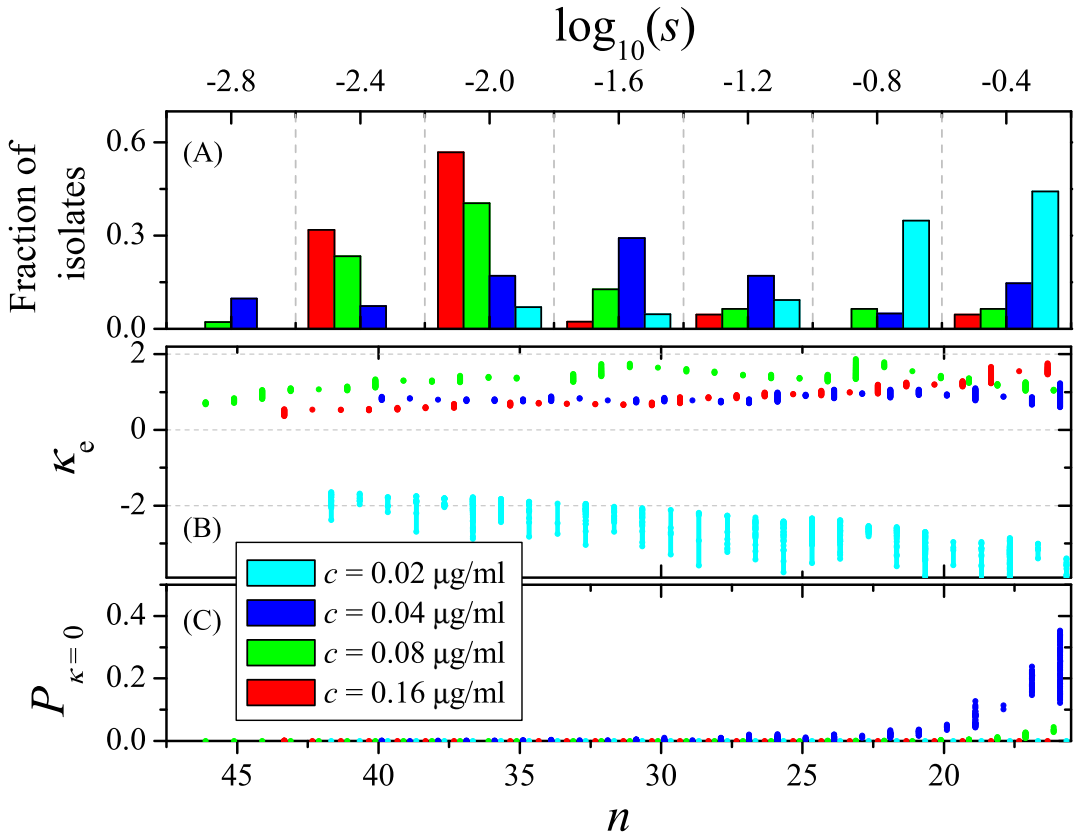


- 48 out of 2583 point mutations increase resistance against cefotaxime

- Colony survival translated into fitness using branching process simulations

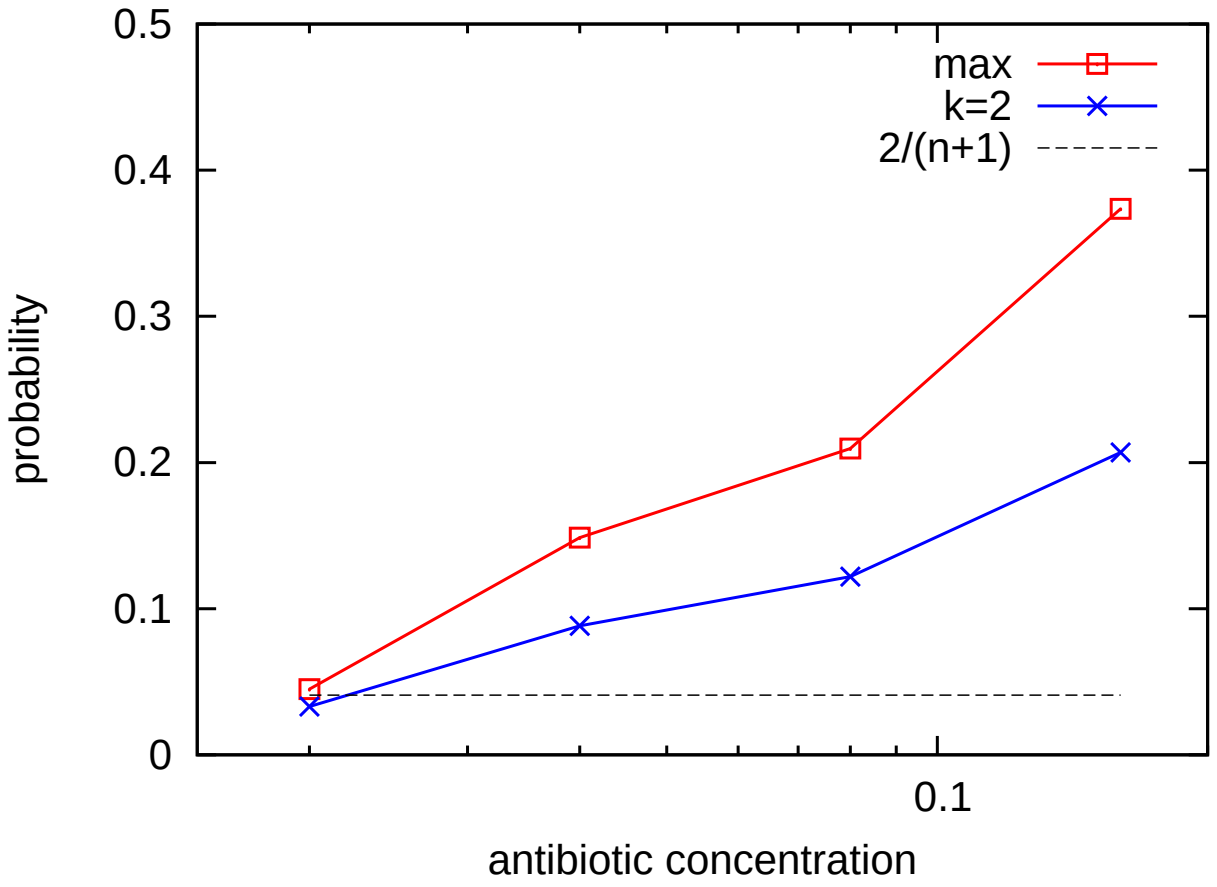
The TEM-1 β-lactamase enzyme

M.F. Schenk, I.G. Szendro, JK, J.A.G.M. de Visser, PLoS Genet. 2012



- Data analysis based on extreme value theory reveals a heavy-tailed distribution of effect sizes with extreme value index $\kappa \sim 1$

Repeatability measures



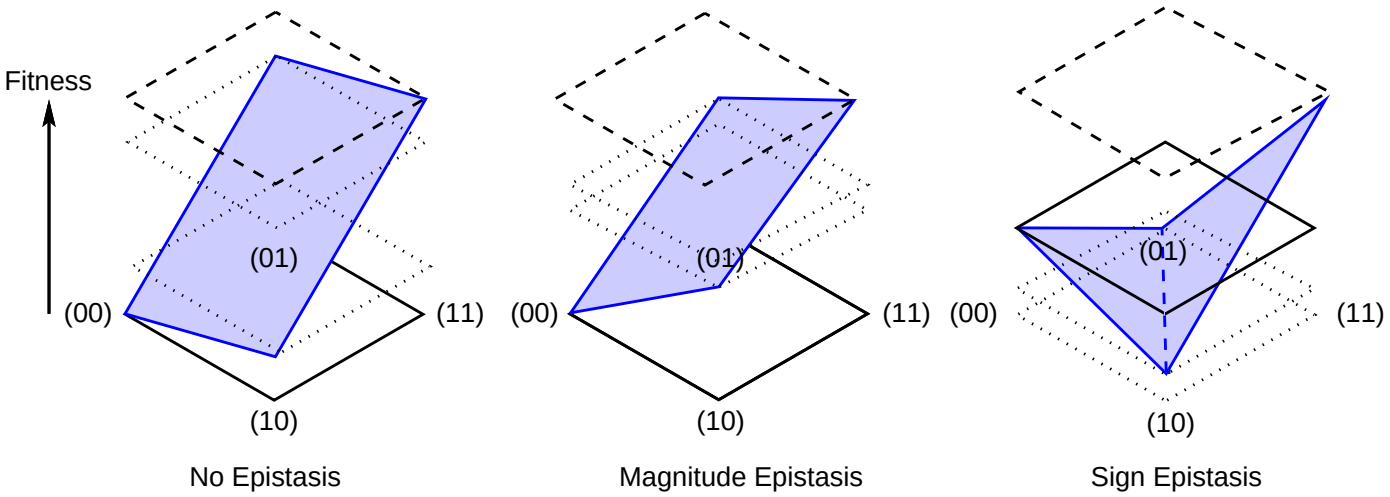
- P_2 and $P_{\max} = \max_i \pi_i > P_2$ increase with antibiotic concentration
- $P_2 = \frac{2}{n+1}$ predicted for effect size distributions with exponential-like tails

Mutational pathways

Epistasis and sign epistasis

- General setting: L diallelic haploid loci τ_i at which a mutation can be present ($\tau_i = 1$) or absent ($\tau_i = 0$).
- A genotypic fitness landscape is a function on the set of 2^L genotypes
- Epistasis implies interactions between the effects of different mutations
- Sign epistasis: Mutation at a given locus is beneficial or deleterious depending on the state of other loci

Weinreich, Watson & Chao 2005

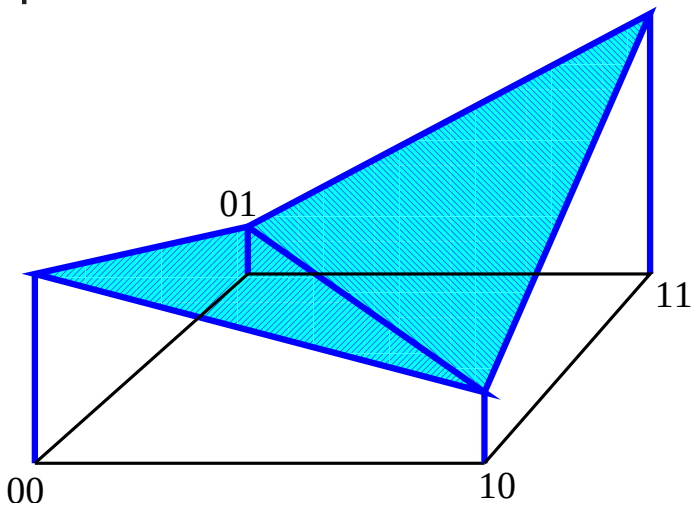


Complex fitness landscapes

- A genotypic fitness landscape is complex/rugged if it has multiple fitness maxima

- The existence of reciprocal sign epistasis is a necessary condition for the existence of multiple peaks

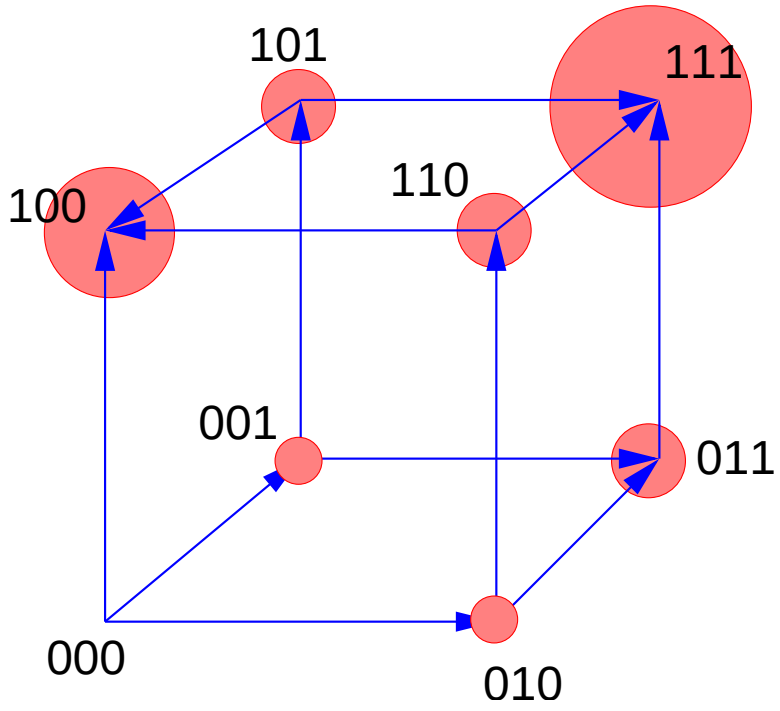
Poelwijk et al., JTB 2011



- Multi-peakedness is guaranteed if all instances of pairwise sign epistasis are reciprocal

Crona et al., JTB 2013

Mutational pathways



- $L = 3$ mutational steps from wildtype 000 to adapted type 111
- Mutations can occur in $3 \times 2 \times 1 = 3! = 6$ different orders corresponding to 6 possible pathways
- Only a subset of pathways are “uphill” (= increasing in fitness)

SSWM dynamics

- SSWM = Strong Selection/Weak Mutation Gillespie 1983, Orr 2002

- Weak mutation: Each new mutation goes to fixation or is lost before the next one arrives

- Strong selection: The fixation probability of a mutation of selective advantage s in a population of size N is Kimura 1963

$$p_{\mathrm{fix}}(s,N) \approx \frac{1-\exp[-2s]}{1-\exp[-2Ns]} \approx 2s$$

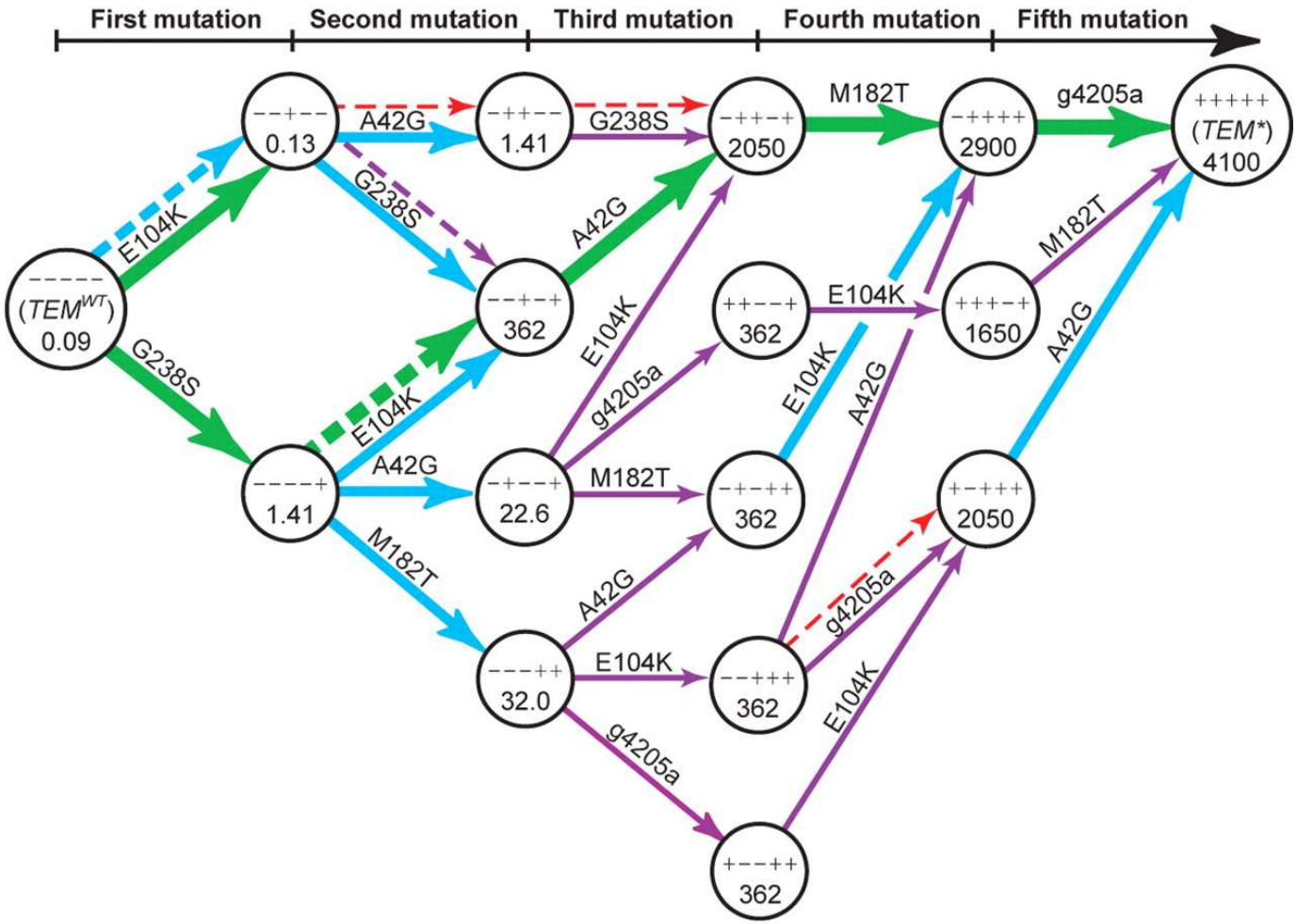
for $0 < s \ll 1$ and $p_{\mathrm{fix}} = 0$ for $s \leq 0$, provided $N|s| \gg 1$

- Under these conditions the evolution is restricted to uphill (= fitness monotonic) mutational pathways, which are called accessible

- The weight of an accessible path in the SSWM regime is the product of the normalized fixation probalities $\pi_i = \frac{s_i}{\sum_j s_j}$

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

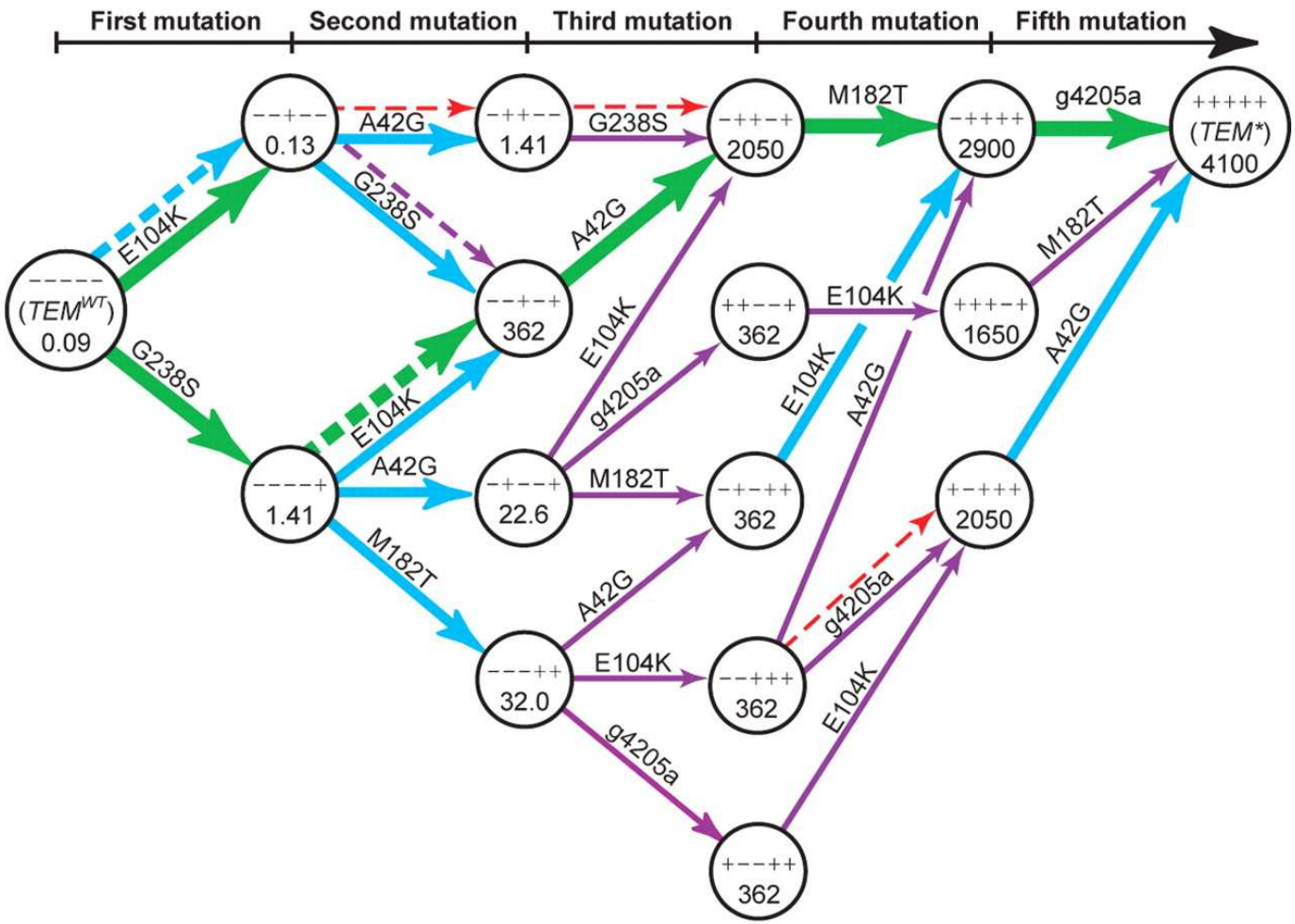
D.M. Weinreich et al., Science 2006



- 5 mutations increase resistance of TEM-1 β-lactamase by $\sim 10^5$

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

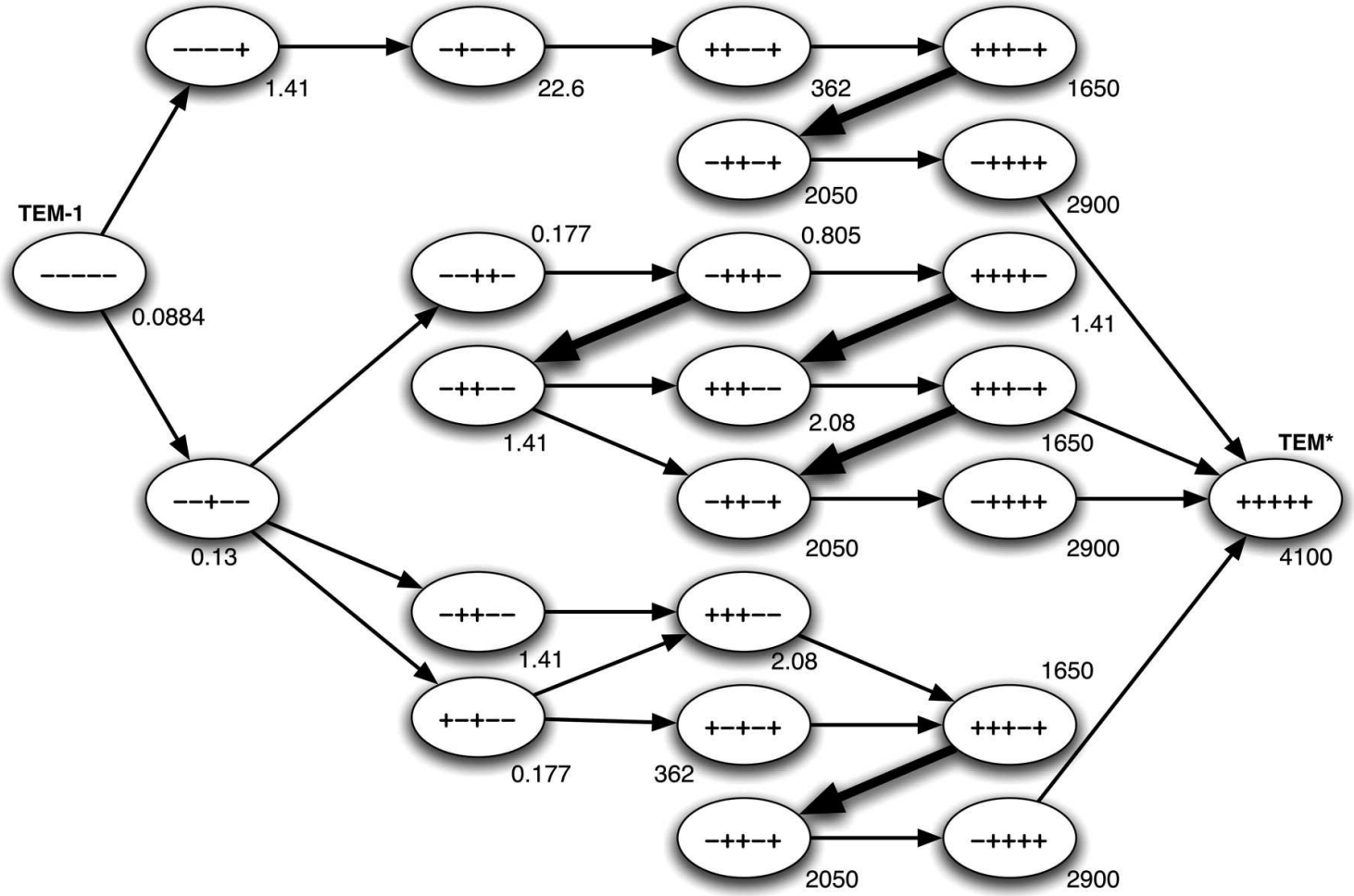
D.M. Weinreich et al., Science 2006



- 18 out of $5! = 120$ directed mutational pathways are increasing...

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

D.M. Weinreich et al., Science 2006



- ...and 27 out of 18651552840 undirected pathways

De Pristo et al. 2007

Accessibility and predictability

- Pathways are accessible if fitness/resistance increases monotonically
- Existence of a small but nonzero fraction of accessible pathways implies high (retrospective) predictability

Accessibility and predictability

- Pathways are accessible if fitness/resistance increases monotonically
- Existence of a small but nonzero fraction of accessible pathways implies high (retrospective) predictability

Questions for mathematical theory

- How does accessibility depend on the genetic interactions and on the boundary conditions of the paths?
- How typical is it that a small but nonzero fraction of pathways are accessible?
- These questions can be addressed systematically using probabilistic models of fitness landscapes

Null model: House-of-Cards

J. Franke, A. Klözer, J.A.G.M. de Visser & JK, PLoS Comp. Biol. 2011

- In the house-of-cards model fitness is assigned randomly to genotypes, for example, from a uniform distribution Kingman 1978, Kauffman & Levin 1987

- Then the probability that a given path is accessible is $1/L!$ and hence the expected number of accessible paths is $L! \times \frac{1}{L!} = 1$ which suggests very high predictability.

Null model: House-of-Cards

J. Franke, A. Klözer, J.A.G.M. de Visser & JK, PLoS Comp. Biol. 2011

- In the house-of-cards model fitness is assigned randomly to genotypes, for example, from a uniform distribution Kingman 1978, Kauffman & Levin 1987

- Then the probability that a given path is accessible is $1/L!$ and hence the expected number of accessible paths is $L! \times \frac{1}{L!} = 1$ which suggests very high predictability.

- This is however misleading, because most landscape realizations do not possess a single accessible path.

- Accessibility is determined primarily by initial fitness and transitions sharply from high to low at a threshold fitness $\sim \log L/L$ Hegarty & Martinsson 2014

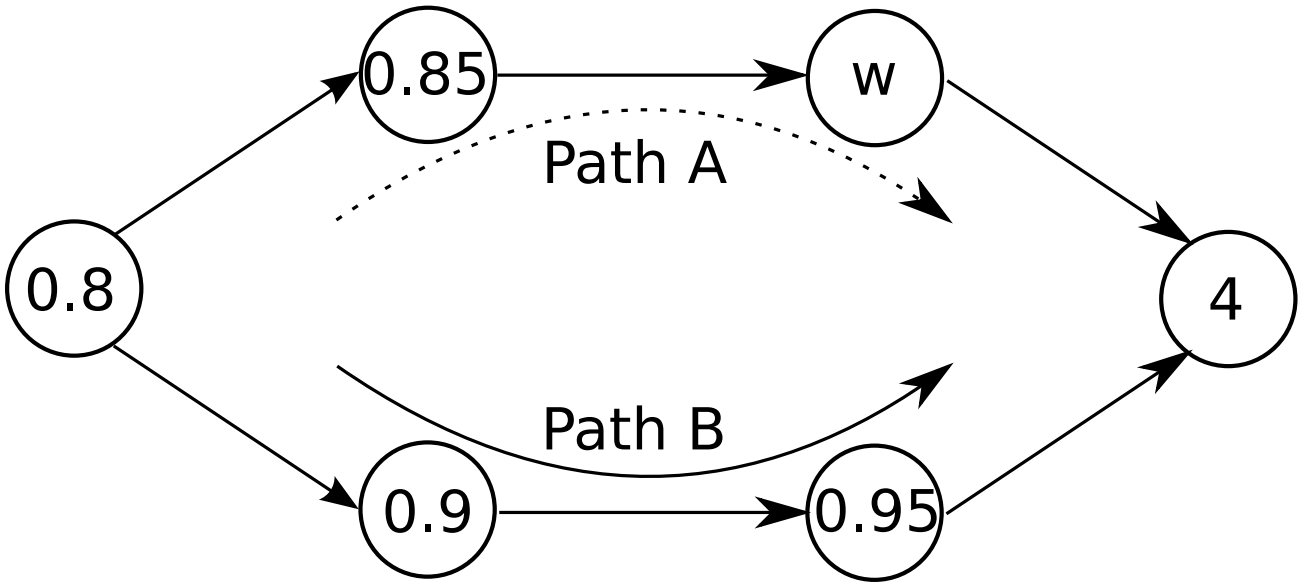
- As a consequence, conditioned on accessibility (or low initial fitness) the typical number of paths is of order $L \ll L!$

Paths to evolutionary rescue

Jan Schmidt

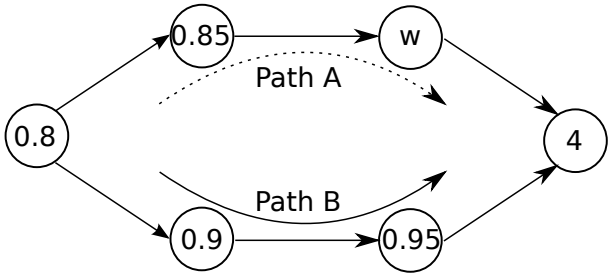
- A population with negative absolute fitness that is destined for extinction can be rescued by *de novo* beneficial mutations
- If this requires multiple mutational steps there can be different rescue paths
- Evolutionary rescue is the process underlying the evolution of drug resistance
- Path weights for evolutionary rescue can be determined using path-resolved branching process theory Iwasa et al. 2003; Bauer & Gokhale 2015
- SSWM weights and the weights of rescue paths can differ substantially
- In particular, evolutionary rescue does not require fitness to be monotonically increasing

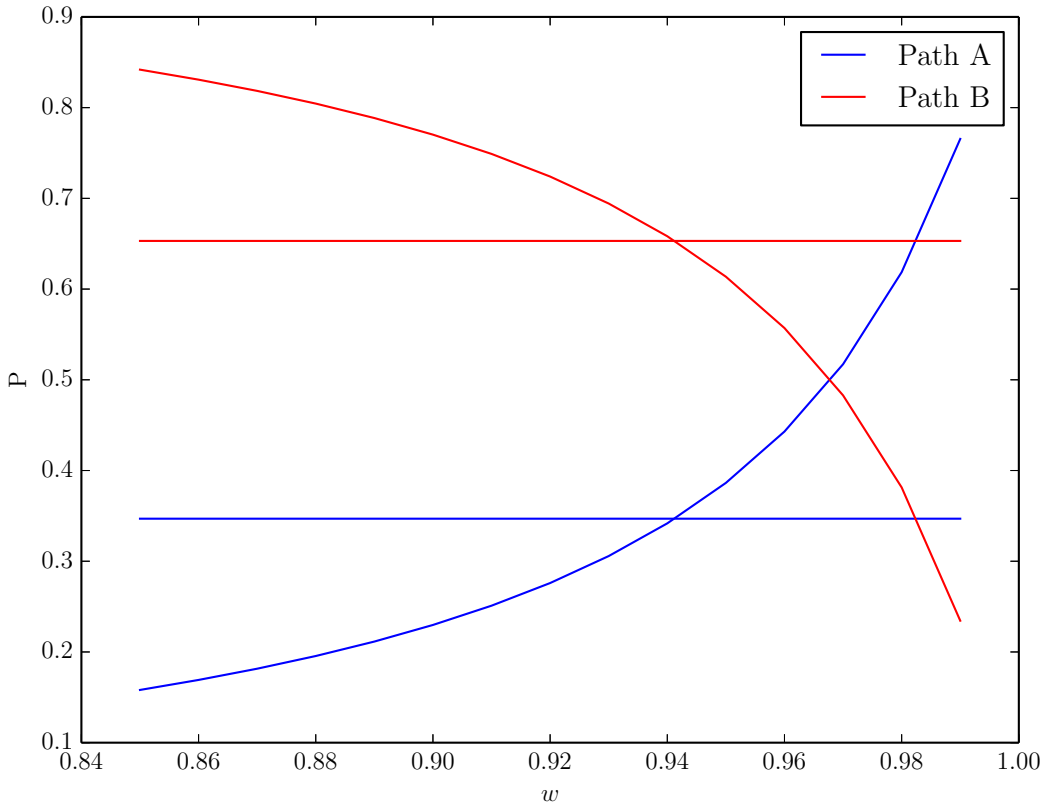
Example: Two competing accessible paths



- Initial and all intermediate types have multiplicative fitness/reproductive ratio < 1
- SSWM weights are determined by the first step and do not depend on the intermediate fitness value w
- By contrast, rescue occurs preferentially along path A as $w \to 1$

Example: Two competing accessible paths





Landscapes of antibiotic resistance

Patterns of epistasis

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

  Szendro et al., JSTAT 2013; de Visser & Krug, Nat. Rev. Genet. 2014

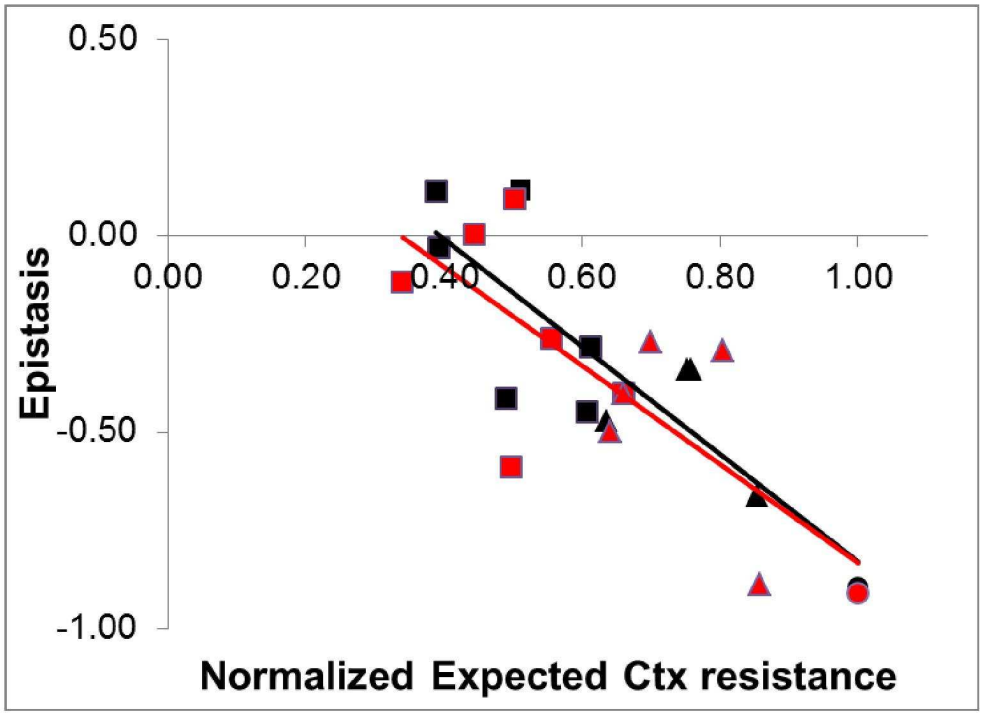
- In particular, the choice of the subset of mutations used to construct a landscape biases the patterns of epistasis:
  - singly beneficial vs. singly deleterious mutations
  - mutations chosen for individual or collective effects
  - mutations in the same gene or different genes
  - mutations occurring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

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

- Comparative analysis of two subsets of 4 mutations each chosen from a 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

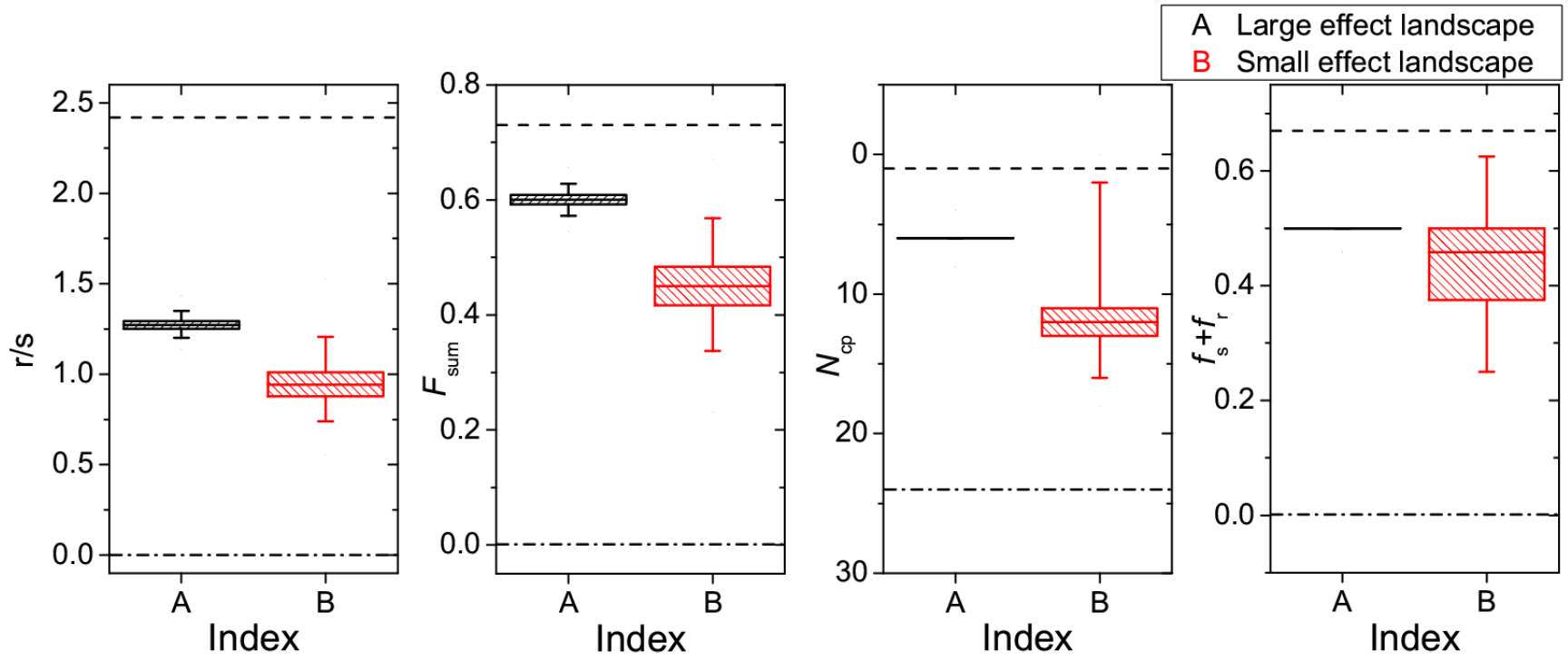


- Resistance of multiple mutants is lower than expected assuming multiplicative effects, and the deviation increases with effect strength

- Generic pattern that appears also in multicellular organisms

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

Large effect landscape is consistently more rugged



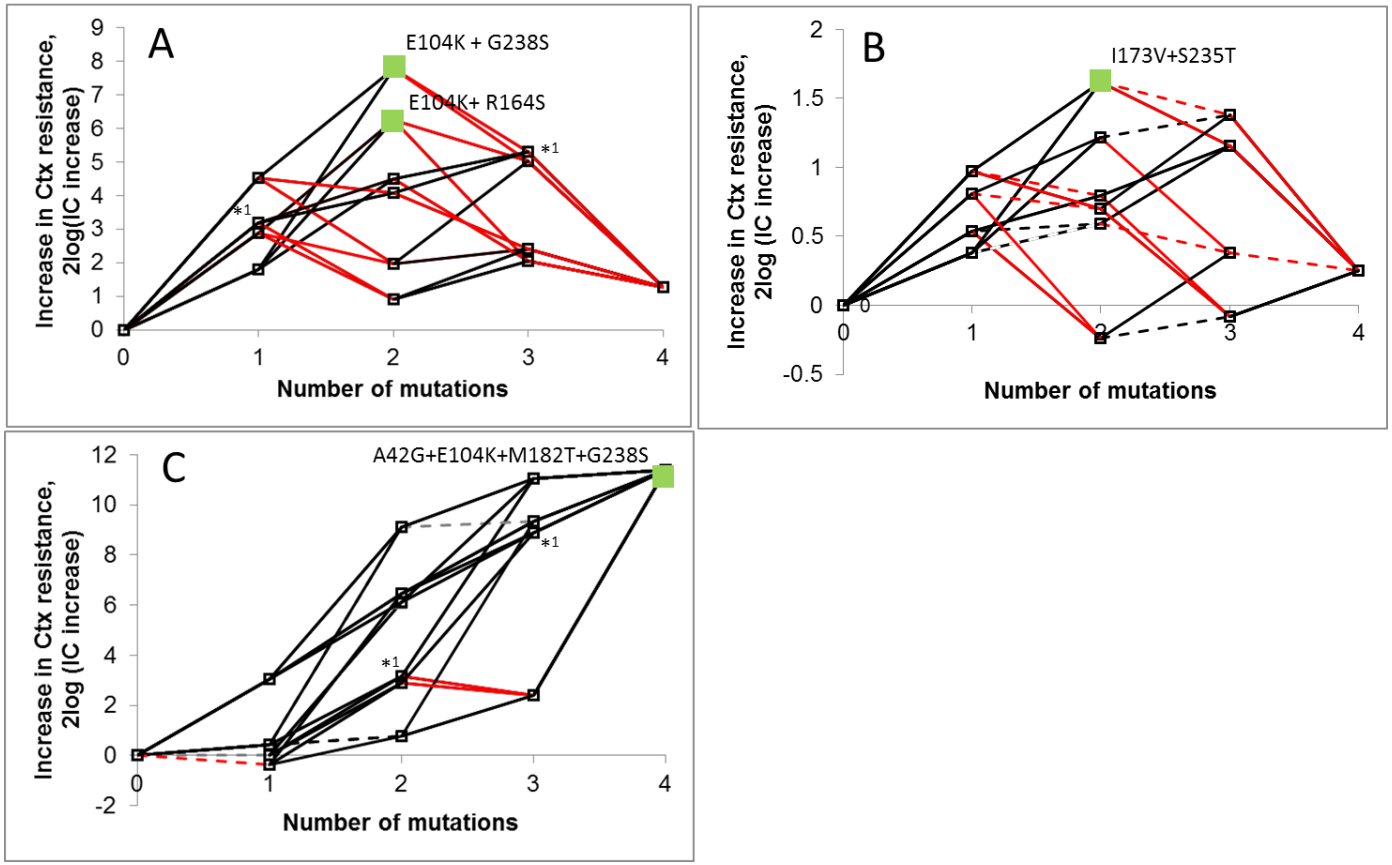
r/s: Roughness-to-slope ratio

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

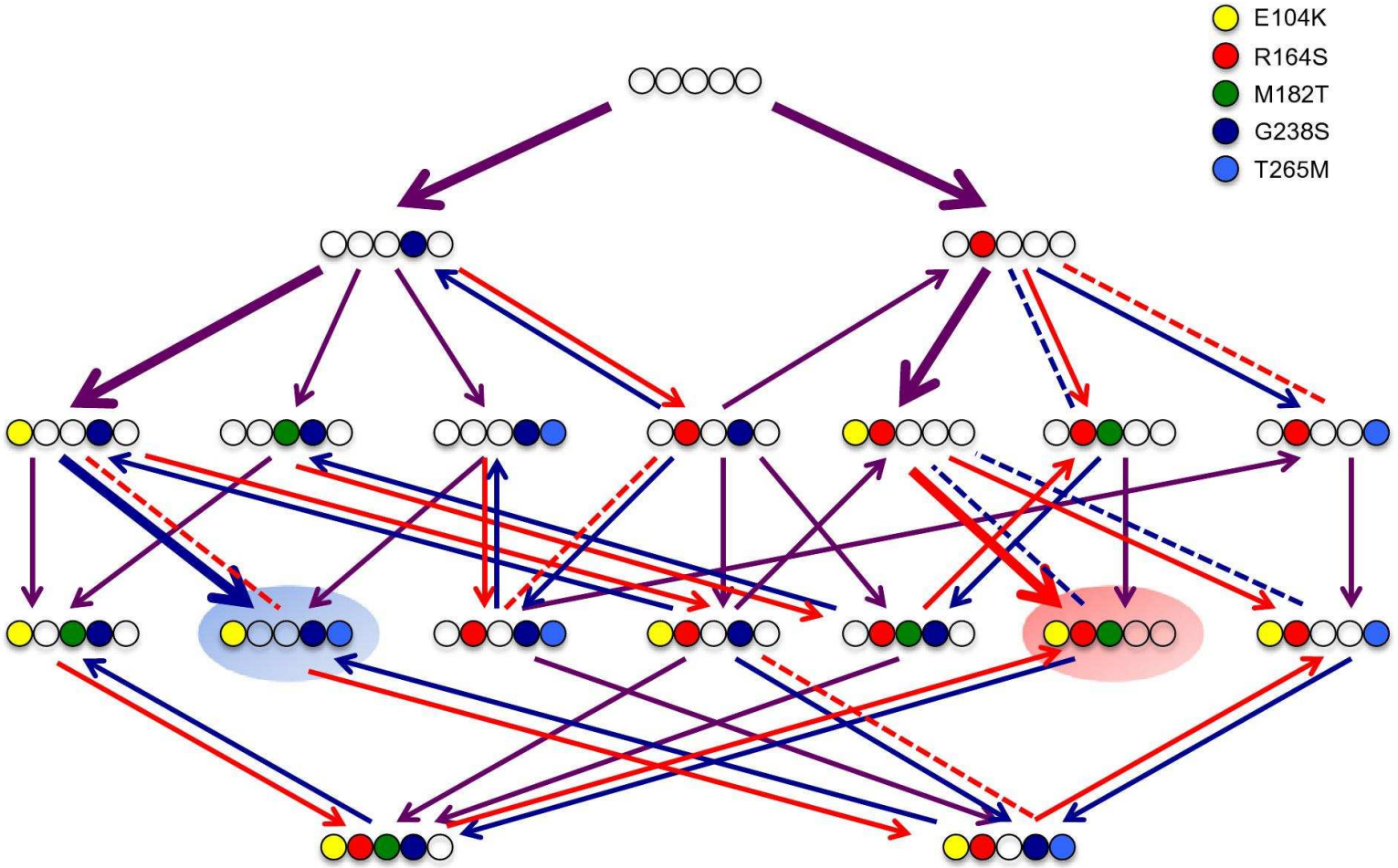


A: Large effect B: Small effect C: Weinreich 2006

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

Resistance landscapes for two different drugs

M.F. Schenk et al., Evol. Appl. (2015)



Arrows point to increasing resistance against cefotaxime, ceftazidime or both

Summary

- Empirical fitness landscapes are beginning to provide insights into the genetic constraints underlying evolutionary processes

- The weight and predictability of mutational pathways is determined by an interplay of landscape structure and population dynamics

- (Sign) epistatic interactions appear to be common across many different systems, but their mechanistic/phenotypic basis remains to be elucidated

Summary

- Empirical fitness landscapes are beginning to provide insights into the genetic constraints underlying evolutionary processes
- The weight and predictability of mutational pathways is determined by an interplay of landscape structure and population dynamics
- (Sign) epistatic interactions appear to be common across many different systems, but their mechanistic/phenotypic basis remains to be elucidated

Thanks to

- Jasper Franke, Johannes Neidhart, Stefan Nowak, Jan Schmidt, Benjamin Schmiegelt, Ivan Szendro (Cologne)
- Martijn Schenk, Sijmen Schoustra, Arjan de Visser, Mark Zwart (Wageningen)