Epistasis: Evolutionary Consequences and Metabolic Origins

Sergey Kryazhimskiy

Section of Ecology, Behavior and Evolution
Division of Biology
University of California San Diego

Evolutionary Cell Biology Program
KITP

20 October 2015

What fraction of possibilities has life explored?

Daniel Fisher

Number of bacteria on Earth

$$\mathbf{10^{30}}$$

Number of cell divisions since origin of life

$$10^{30} \times (3.5 \times 10^{9}) \times (2.6 \times 10^{4}) \approx \mathbf{10^{44}}$$

Number of genomes explored by life

$$10^{44} \times 10^{-4} \times 10^{6} \approx \mathbf{10^{46}}$$

Number of possible genomes

$$4^{1,000,000} \approx \mathbf{10^{600,000}}$$

Do evolutionary outcomes depend on identity and order of mutations?

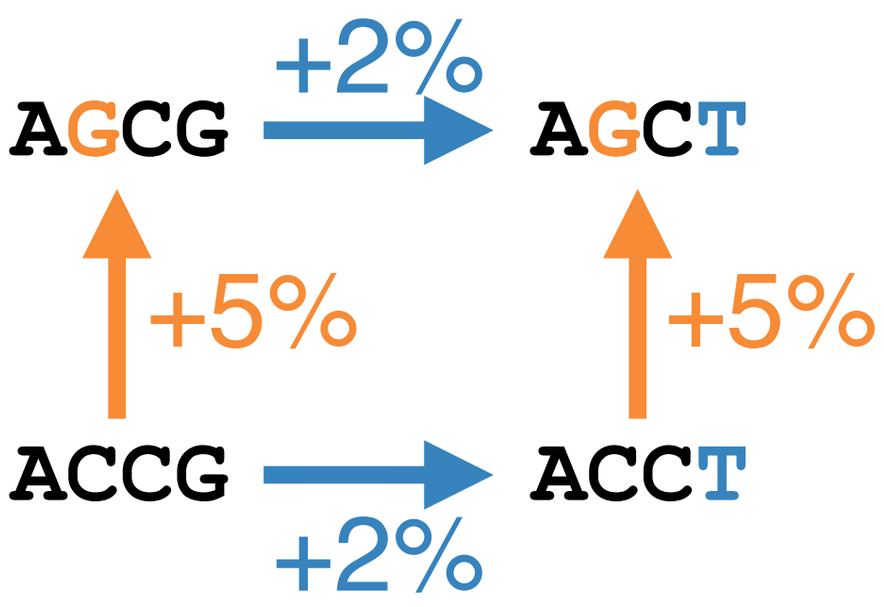
Effects of mutations are context-independent (no epistasis)

One fittest genotype

Effects of mutations are context-dependent (epistasis)

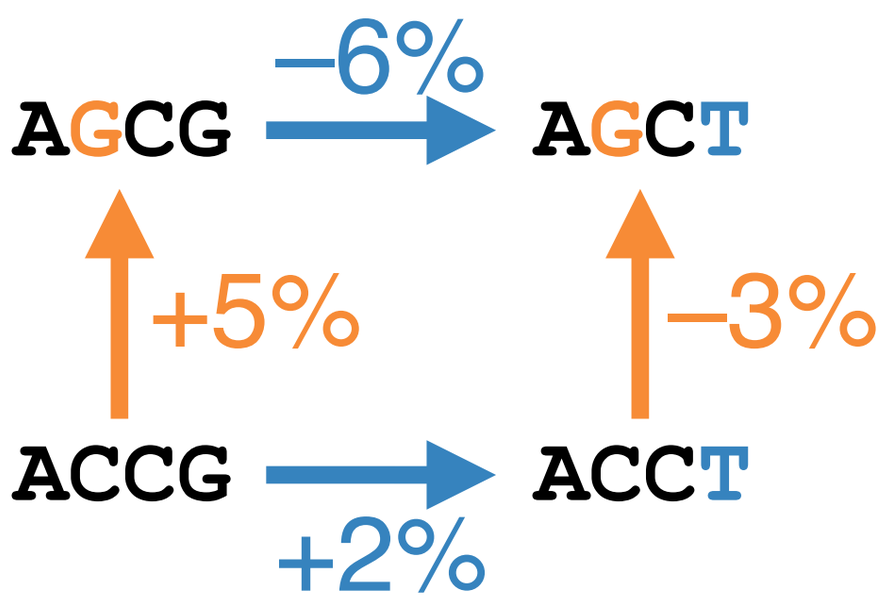
Potentially multiple “fitness peaks”

Do evolutionary outcomes depend on identity and order of mutations?



All mutational paths lead to the same final genotype

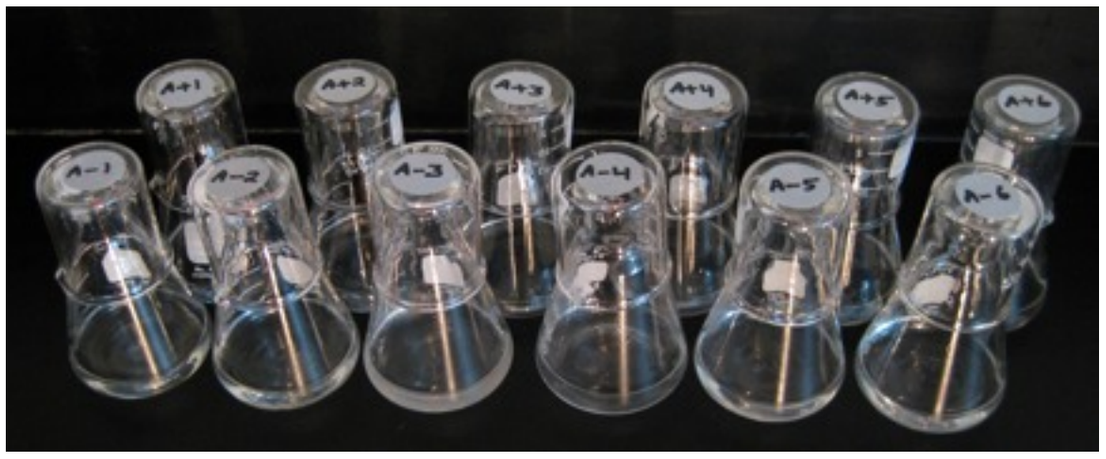
no historical contingency
smooth fitness landscape

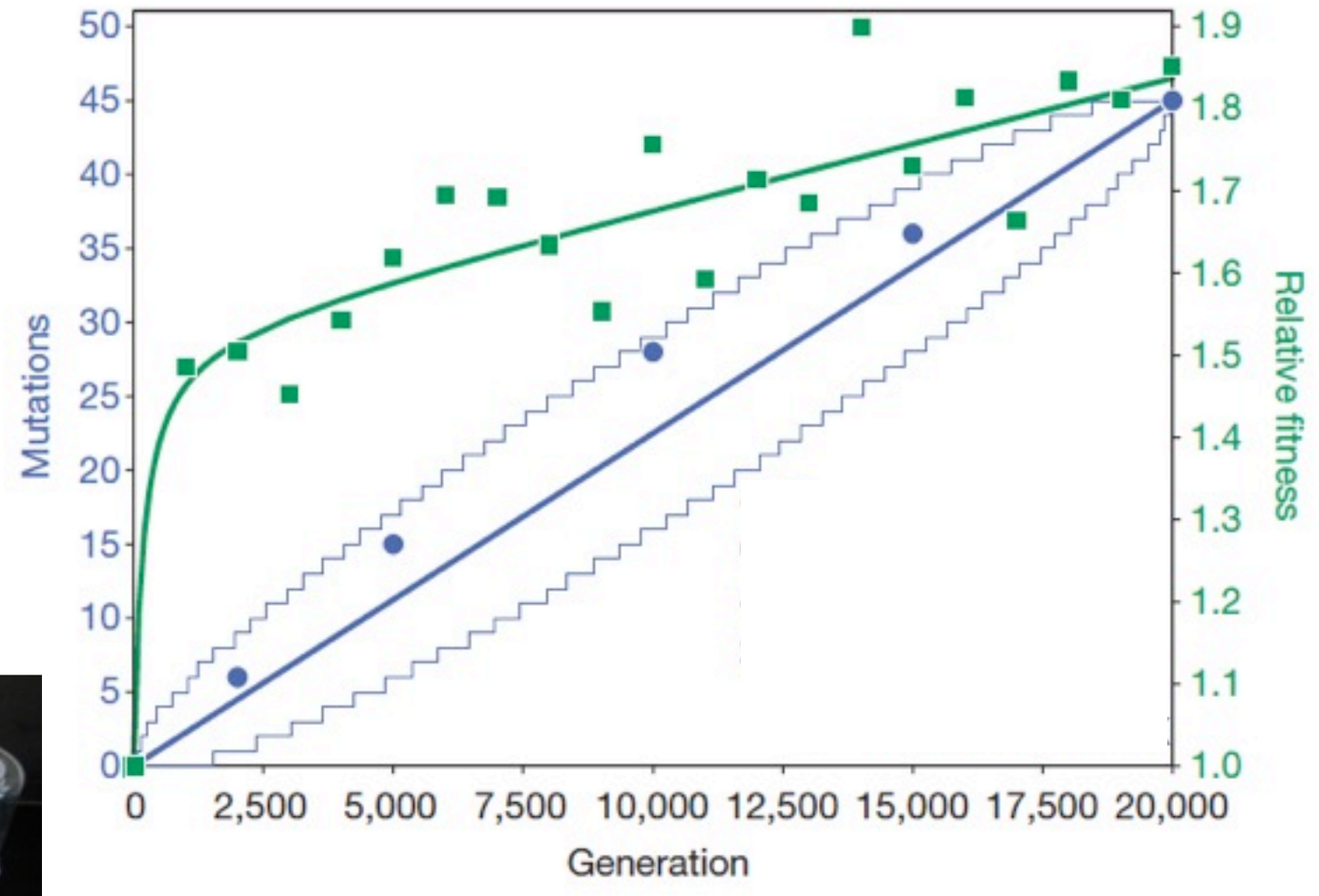


Evolutionary outcomes may be path-dependent

historical contingency
rugged fitness landscape

Evidence for historical contingency due to epistasis

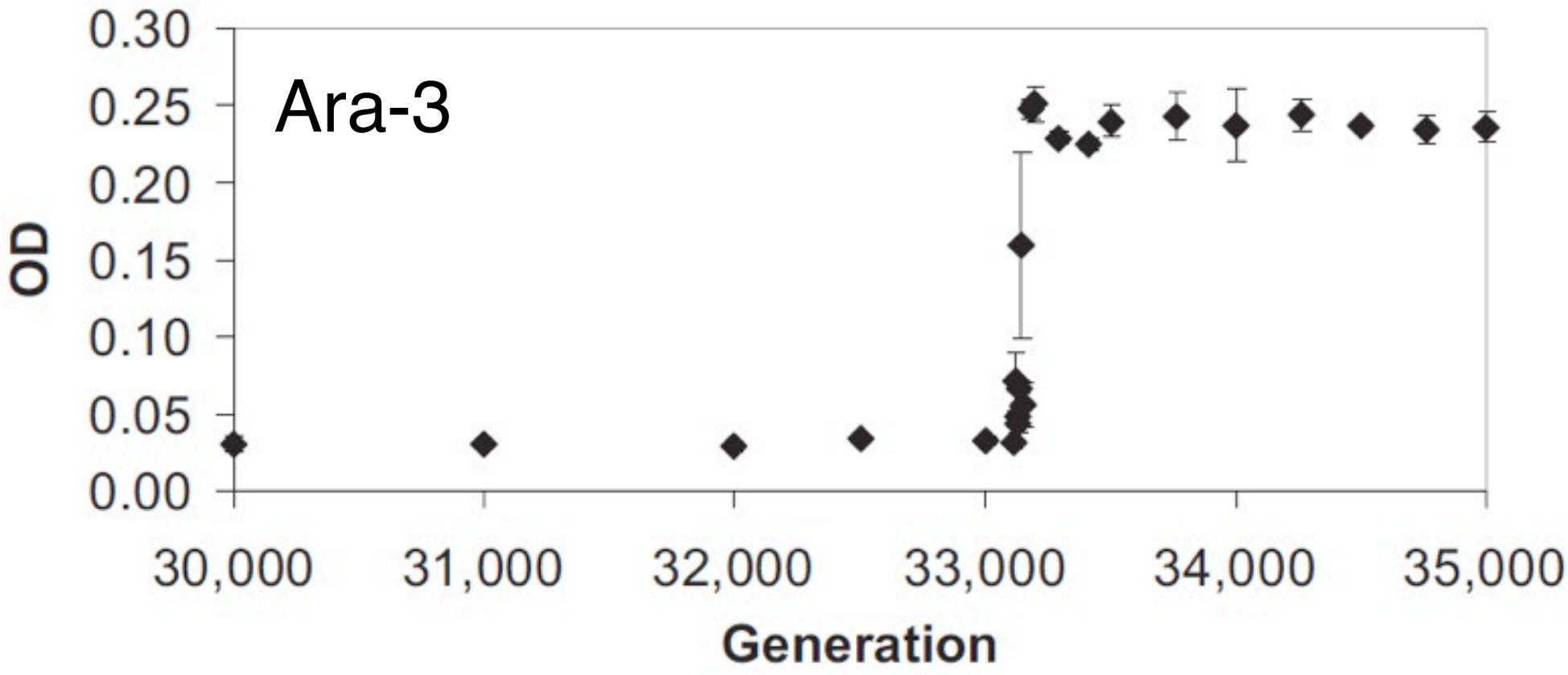


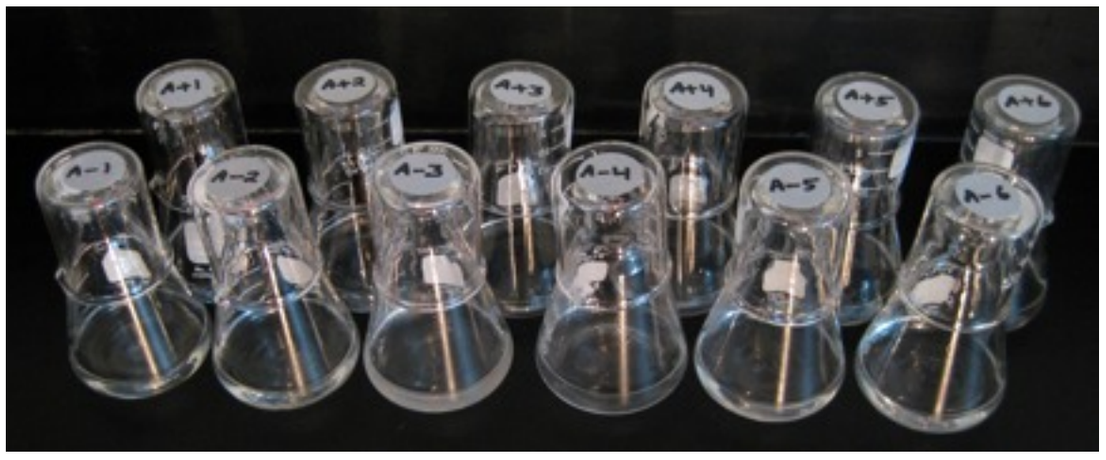


Richard Lenski's LTEE

Barrick et al, *Nature* 2009

Evidence for historical contingency due to epistasis





Richard Lenski's LTEE

Cit++ phenotype:

ability to metabolize citrate in the presence of oxygen

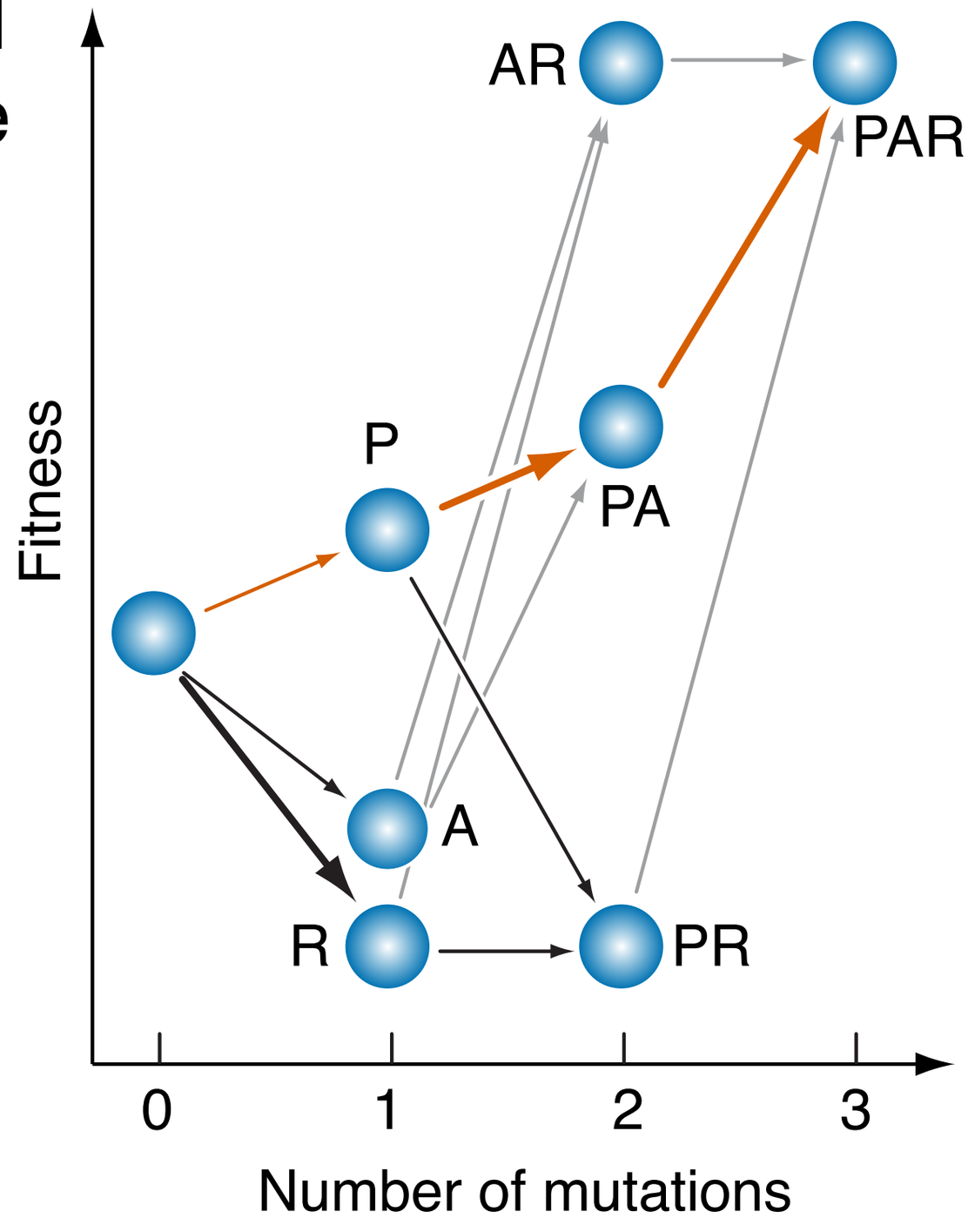
Blount et al, *PNAS* 2008

Evidence for historical contingency due to epistasis

Mutations involved in Cit++ phenotype

- **P**otentiating
- **A**ctualizing
- **R**efining

Richard Lenski's LTEE

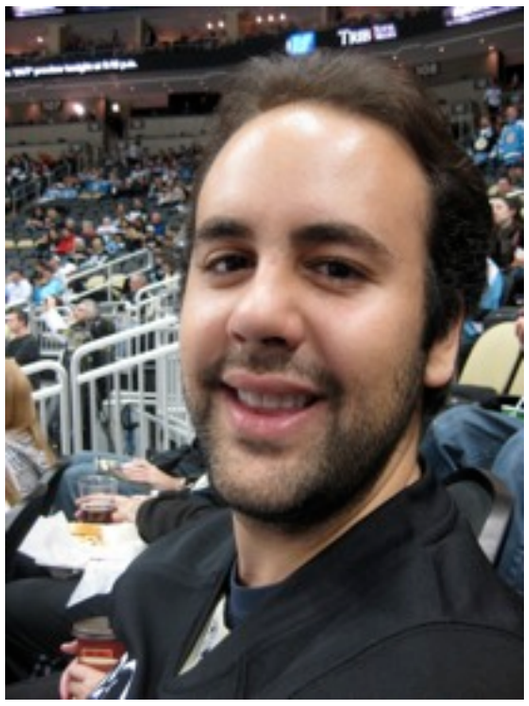


Talk outline

1. How typical is historical contingency?

How does adaptation depend on the initial genotype?

Michael Desai



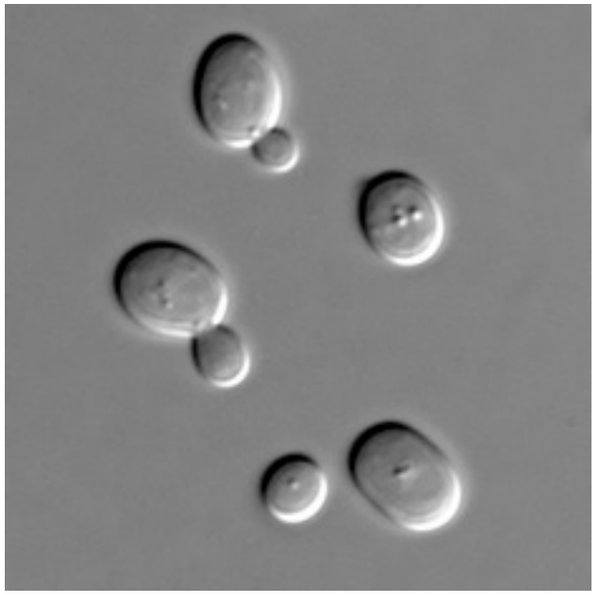
Dan Rice

Elizabeth Jerison

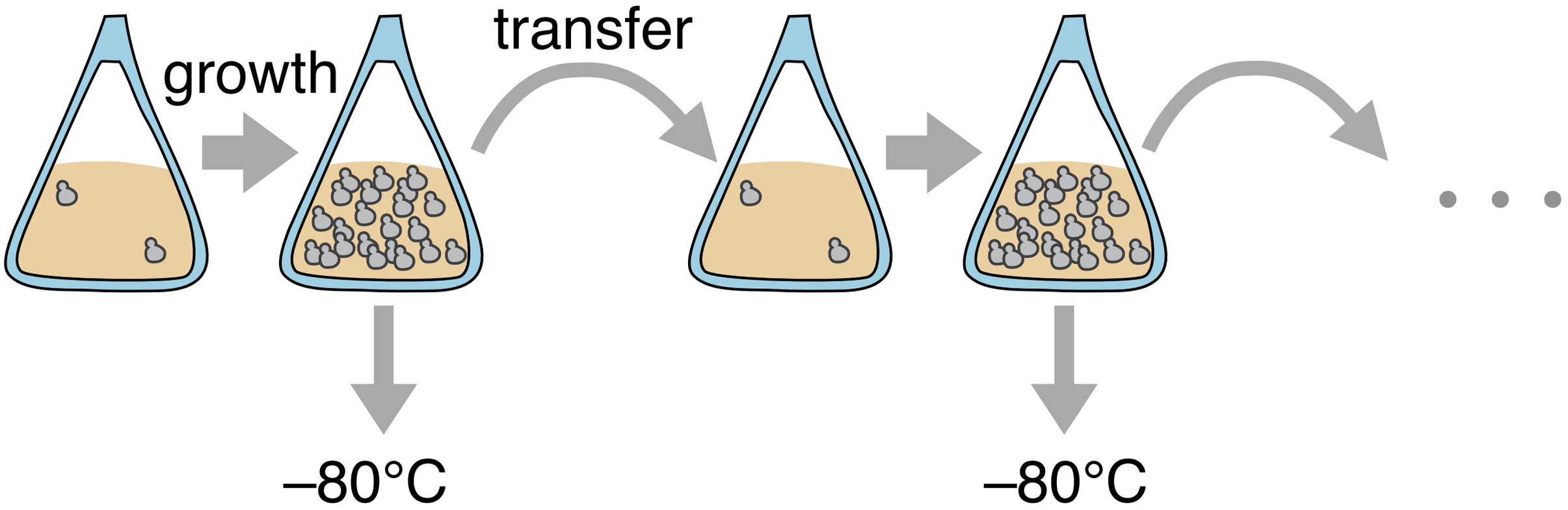
2. What is the metabolic basis of epistasis?

What kinds of epistasis should we expect to observe?

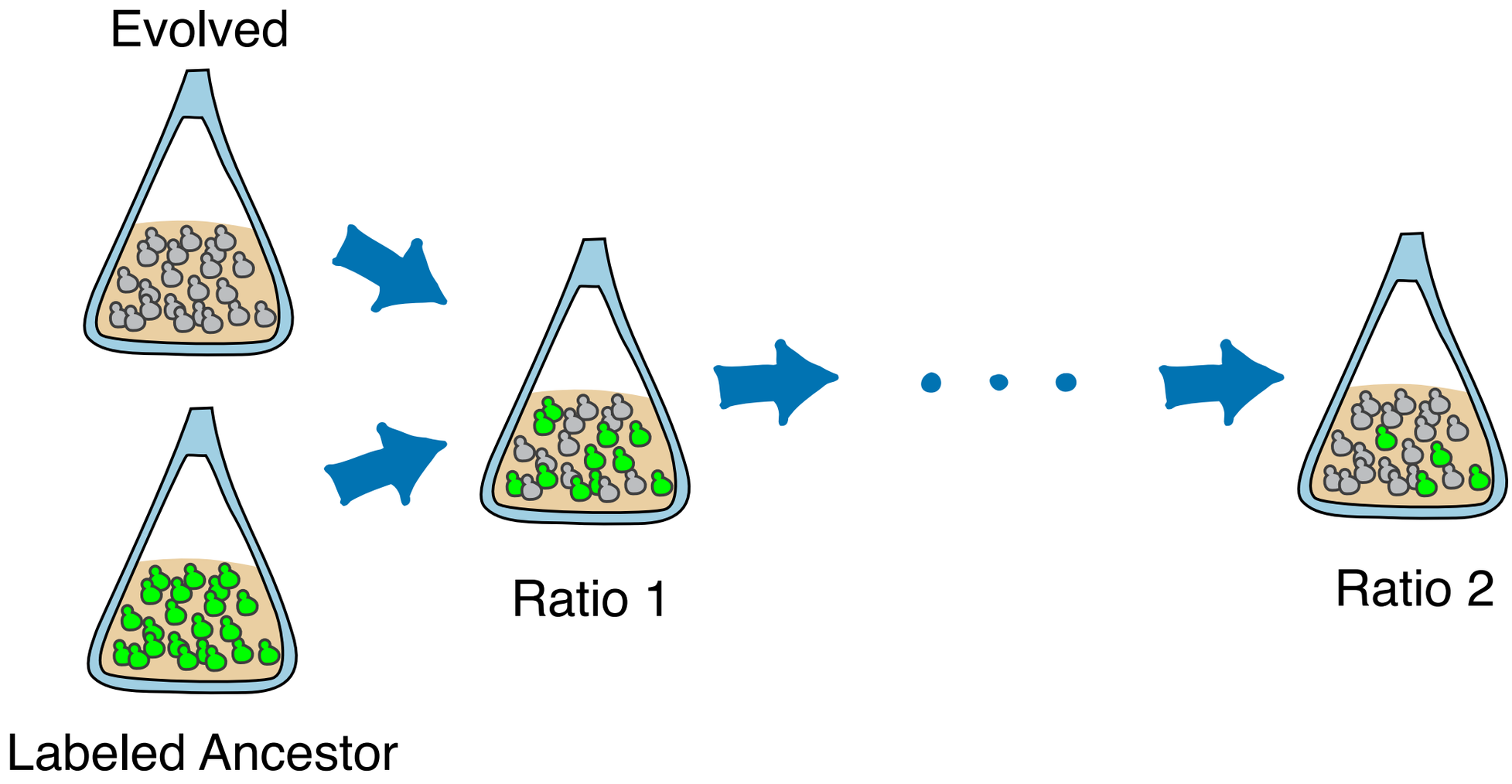
The basics of evolution experiments



Saccharomyces cerevisiae

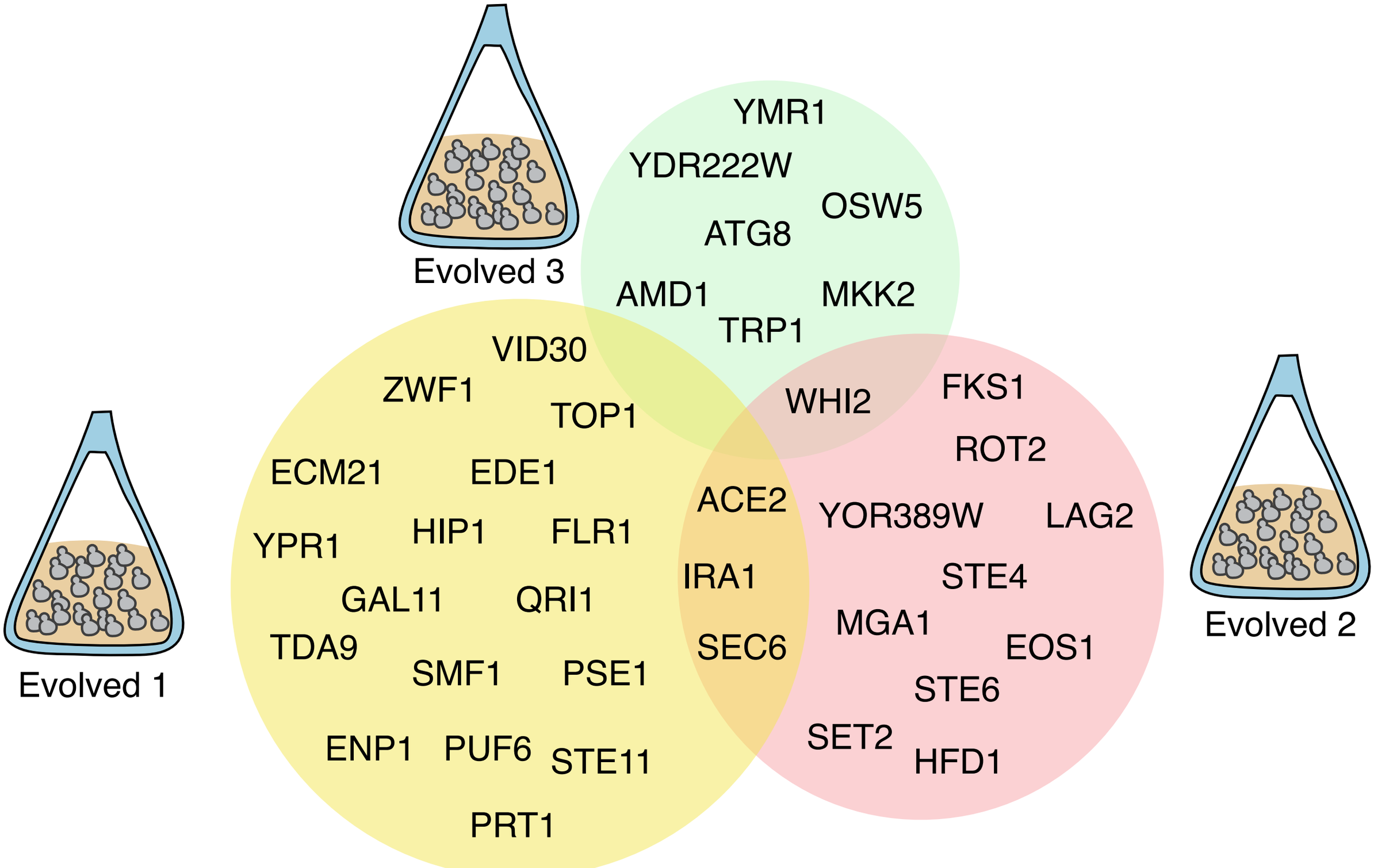


Measure fitness by direct competition with ancestor

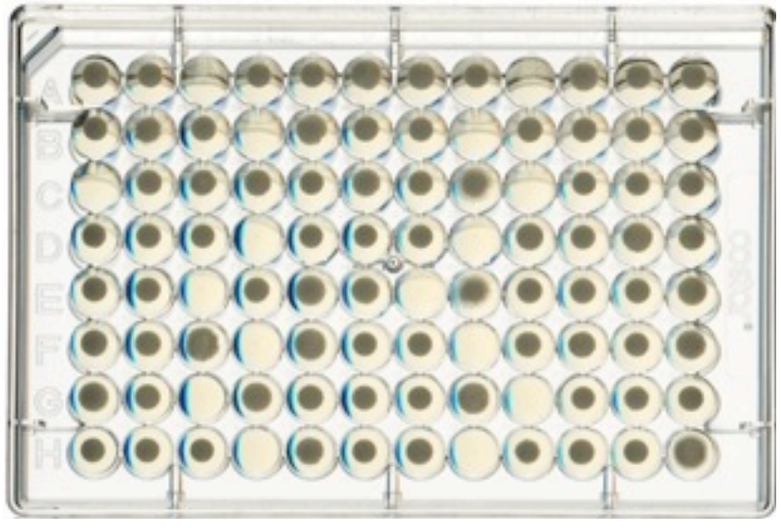


$$\text{Fitness} \propto \ln(\text{Ratio 2}) - \ln(\text{Ratio 1})$$

Identify adaptive mutations by parallel evolution in replicate lines



Experimental evolution in hundreds of parallel populations



Maintain

Transfer

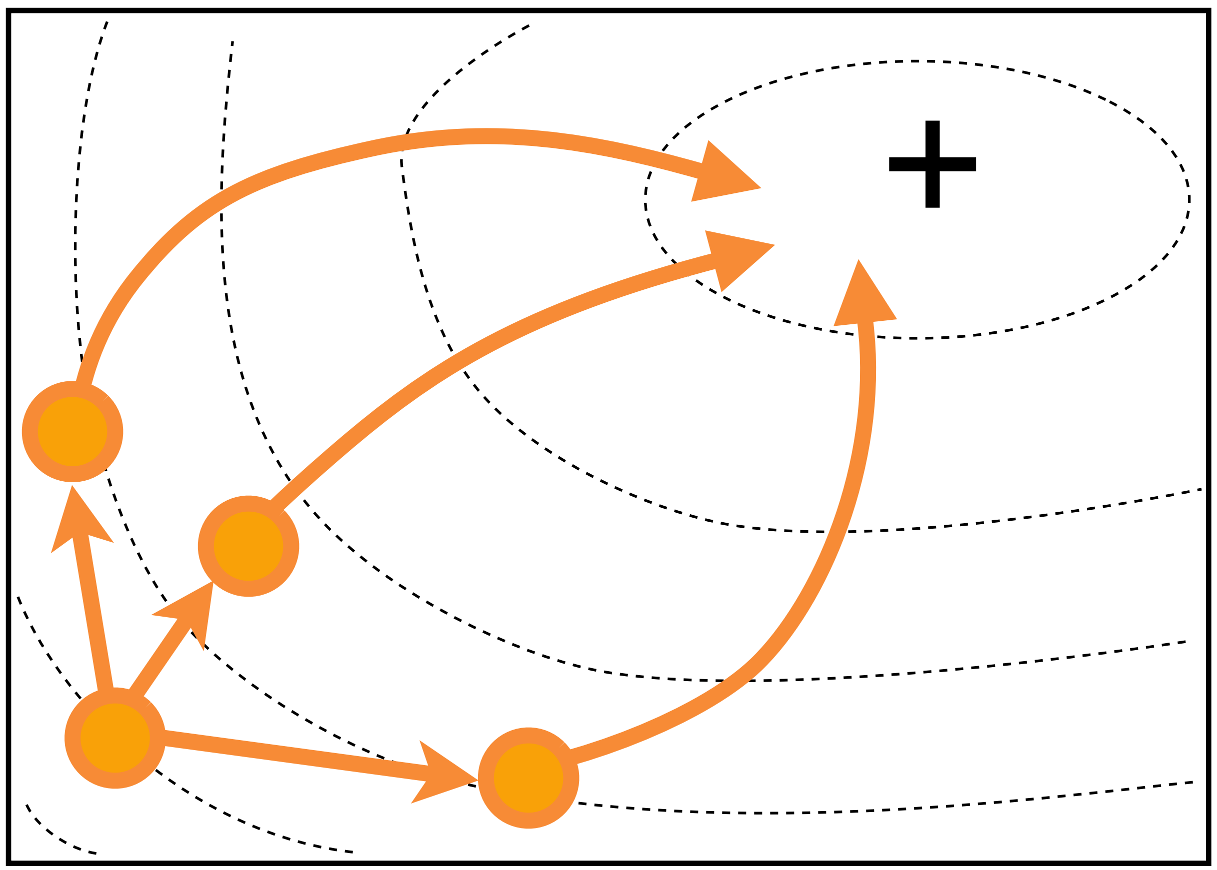
Measure fitness

Find mutations

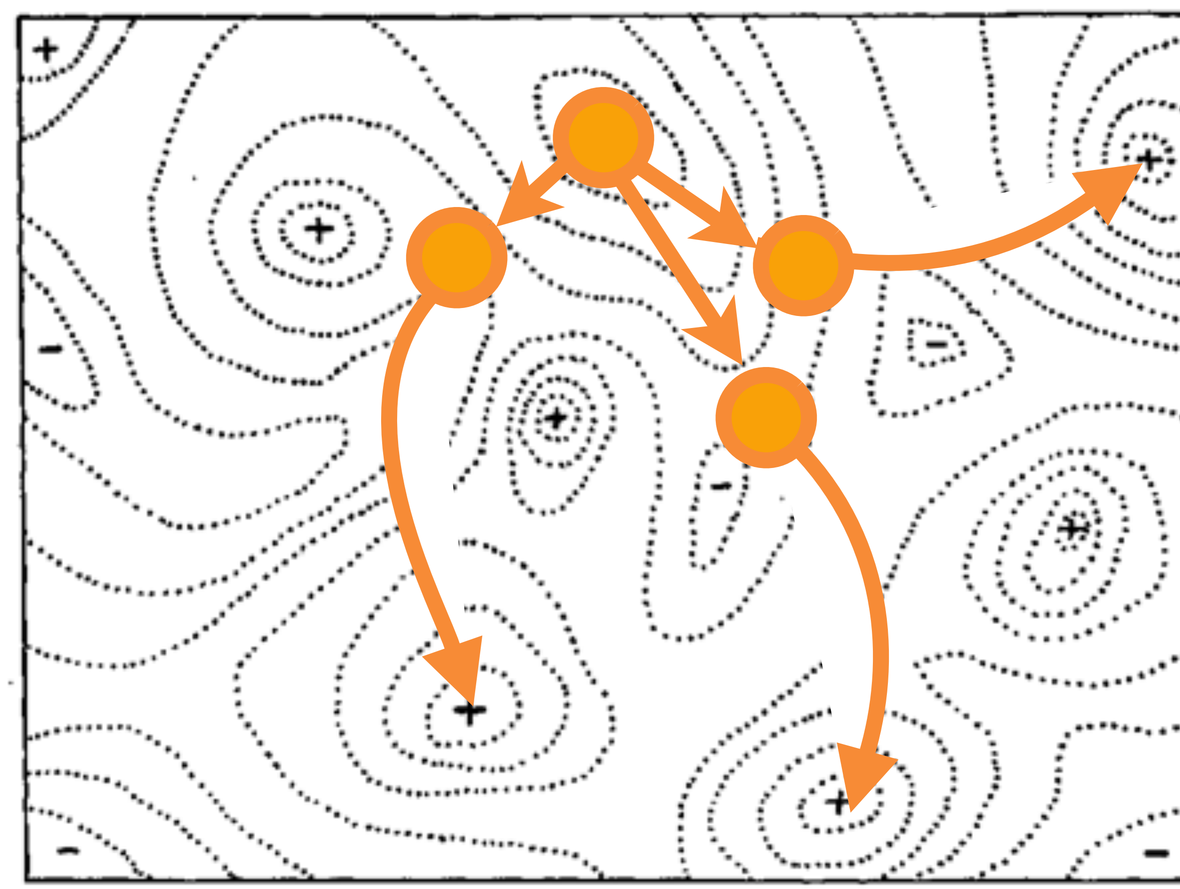
How does prior evolution affect future evolutionary outcomes?

Mutations are beneficial in all backgrounds

Mutations are beneficial in specific backgrounds



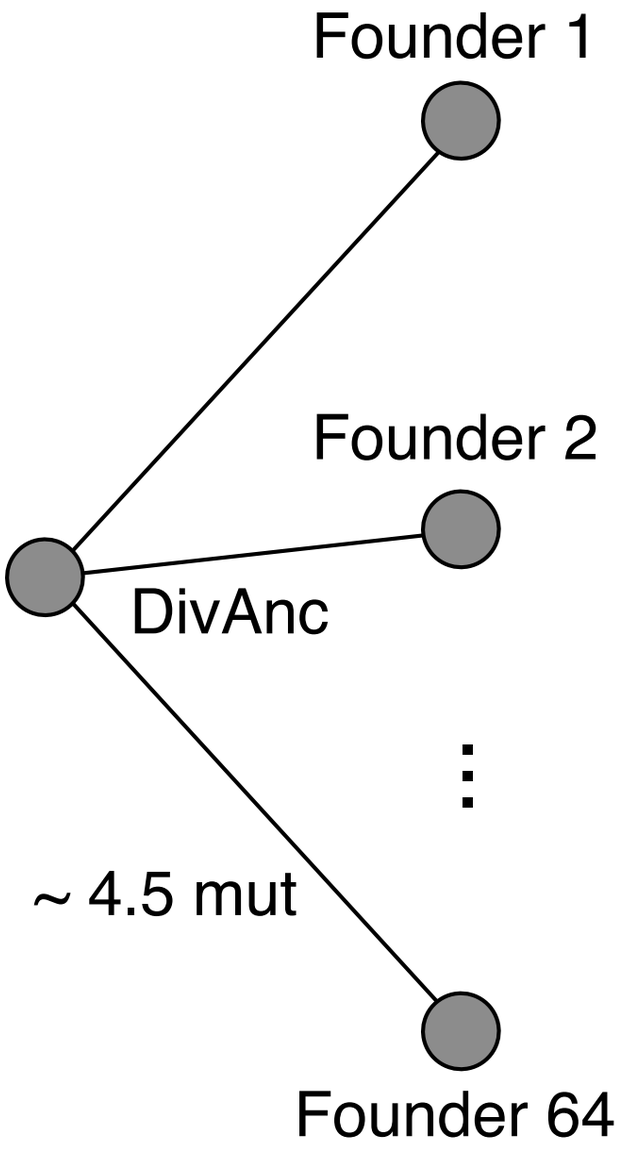
smooth fitness landscape



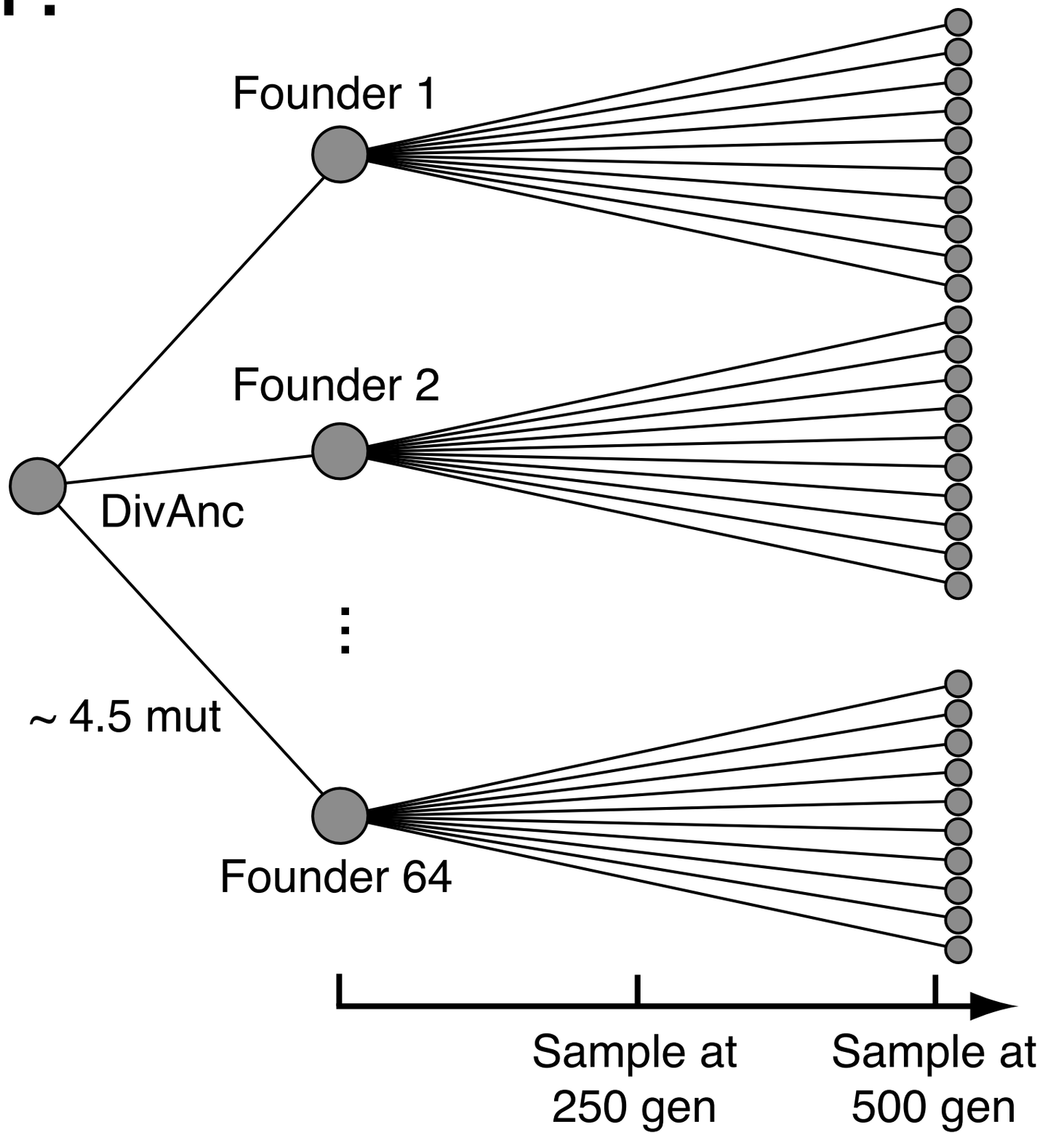
rugged fitness landscape

Wright, *6th Int Congr Genet*, 1932

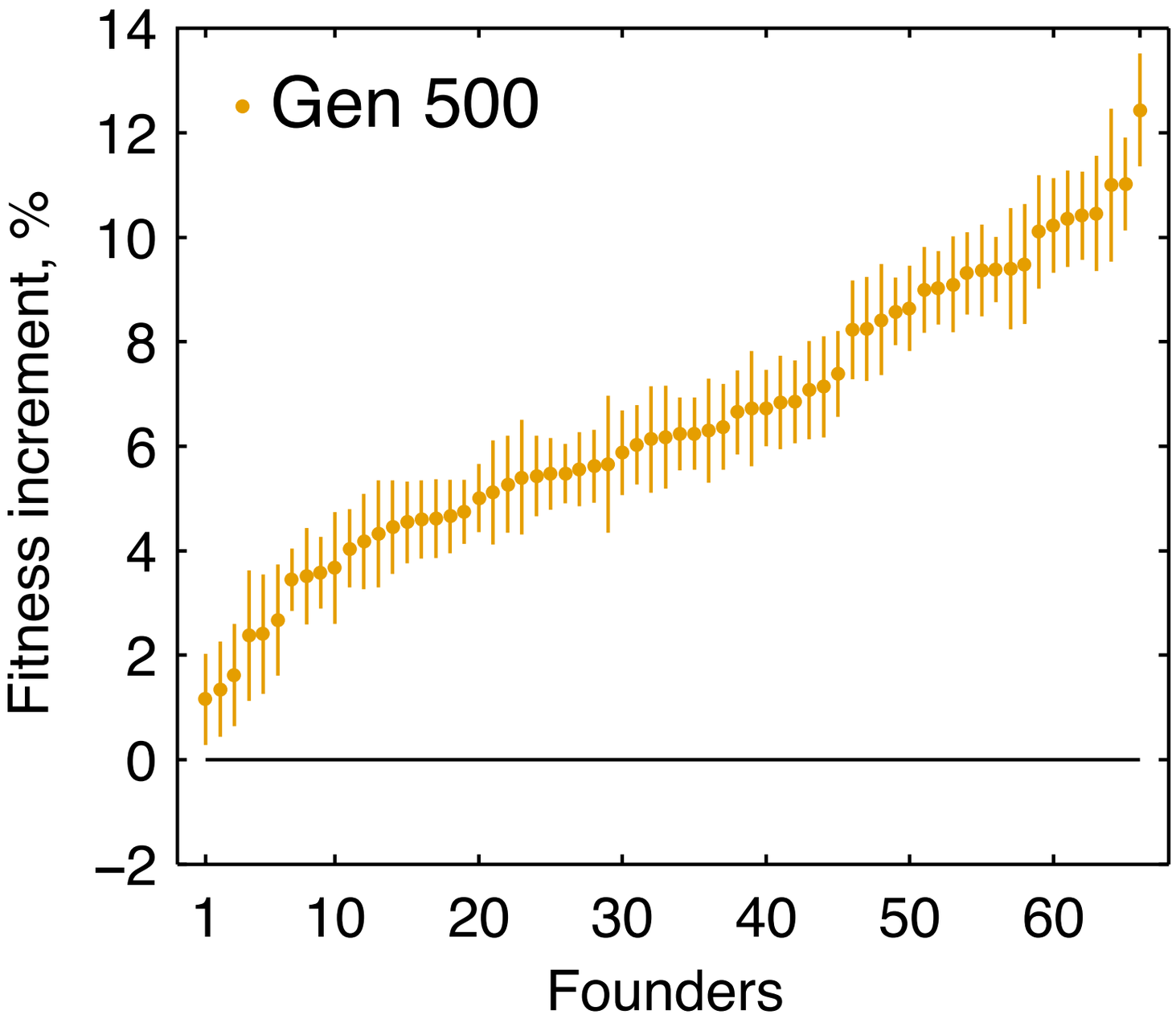
Does initial genotype affect further evolution?

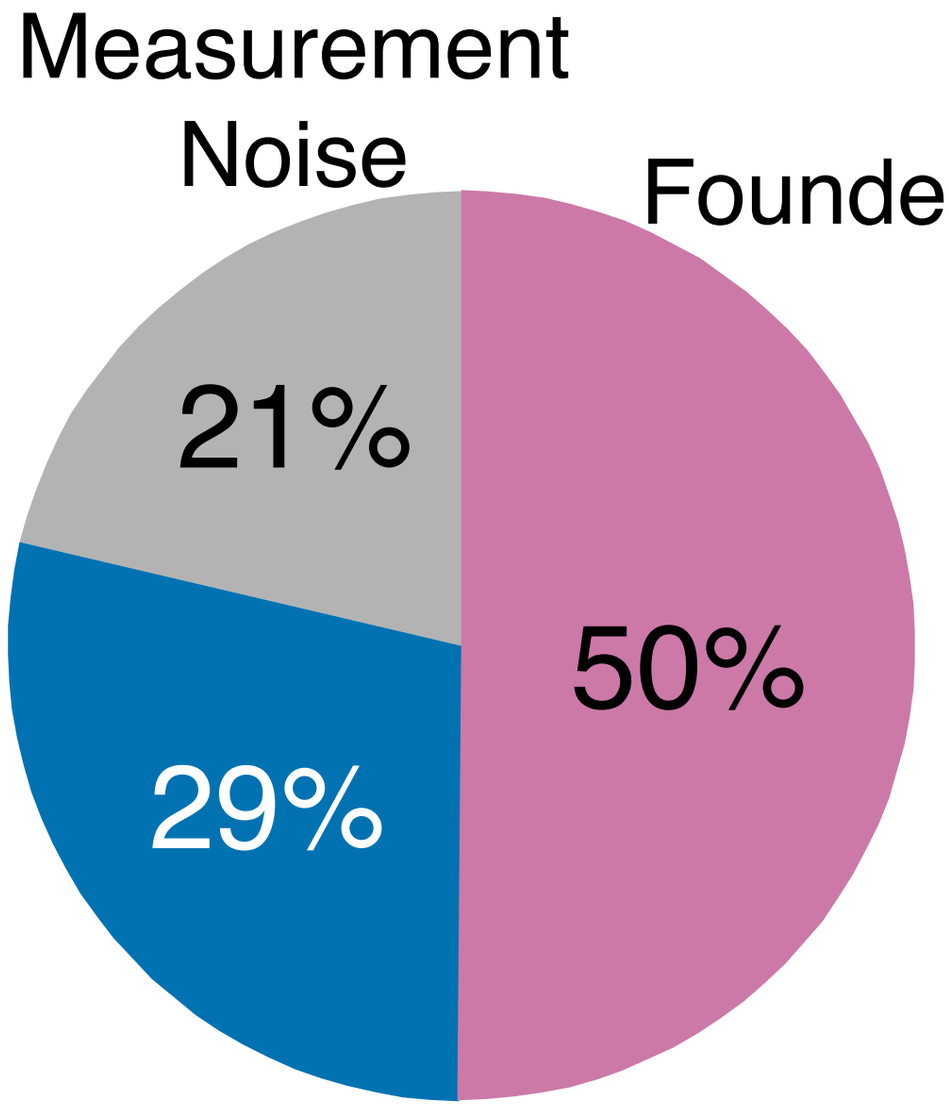


Does initial genotype affect further evolution?

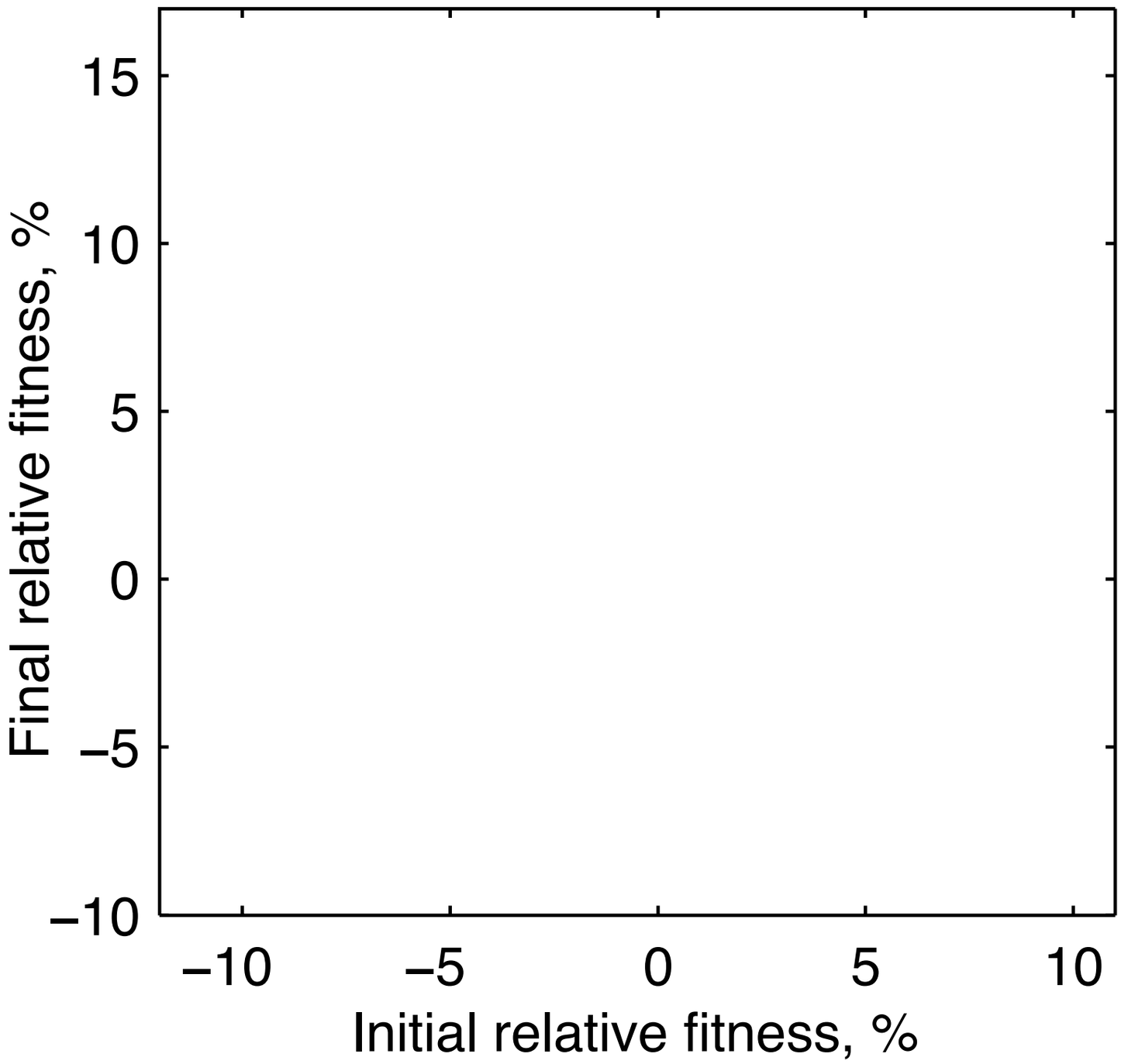


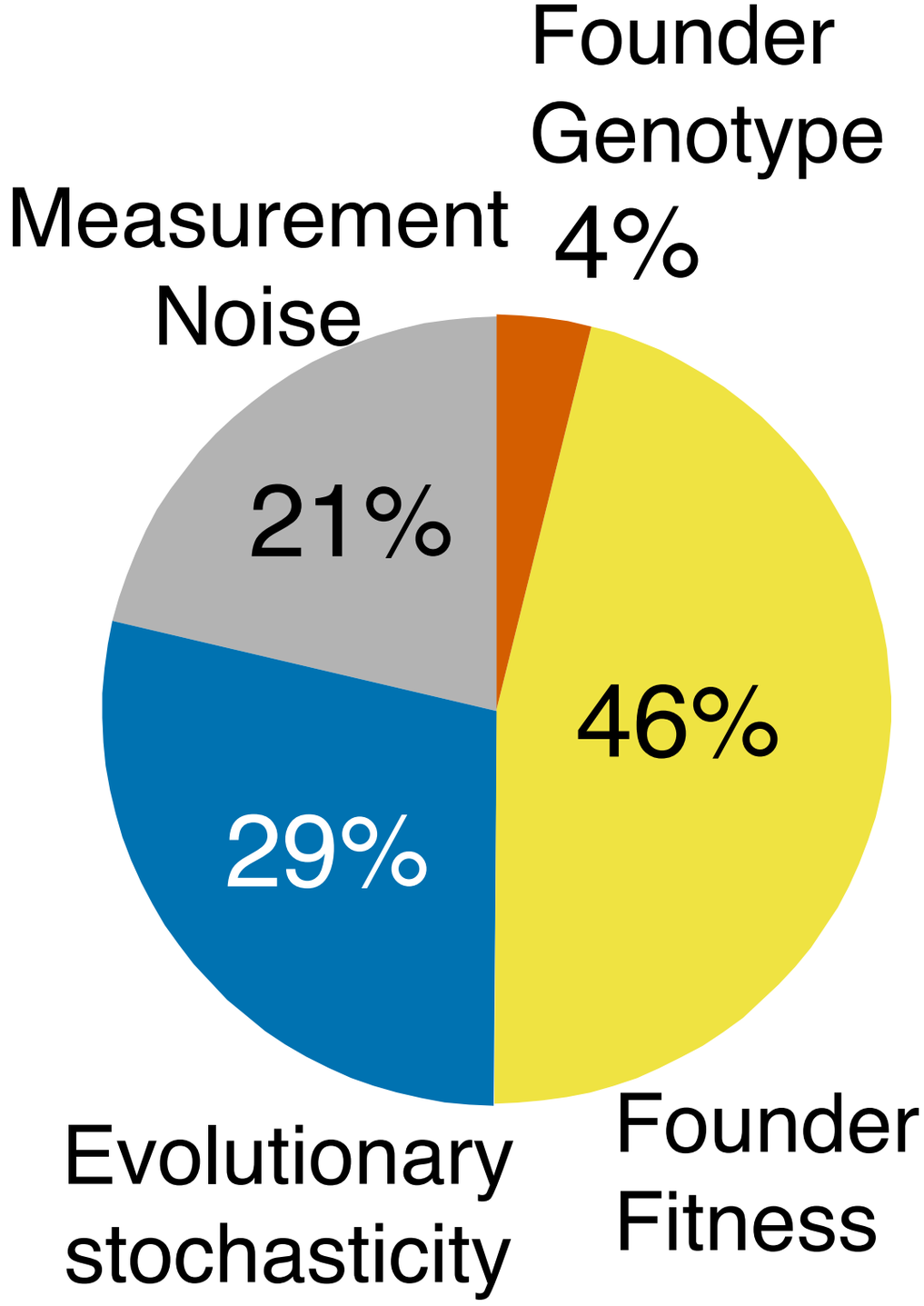
Rate of adaptation varies among Founders





Adaptation rate declines with Founder fitness

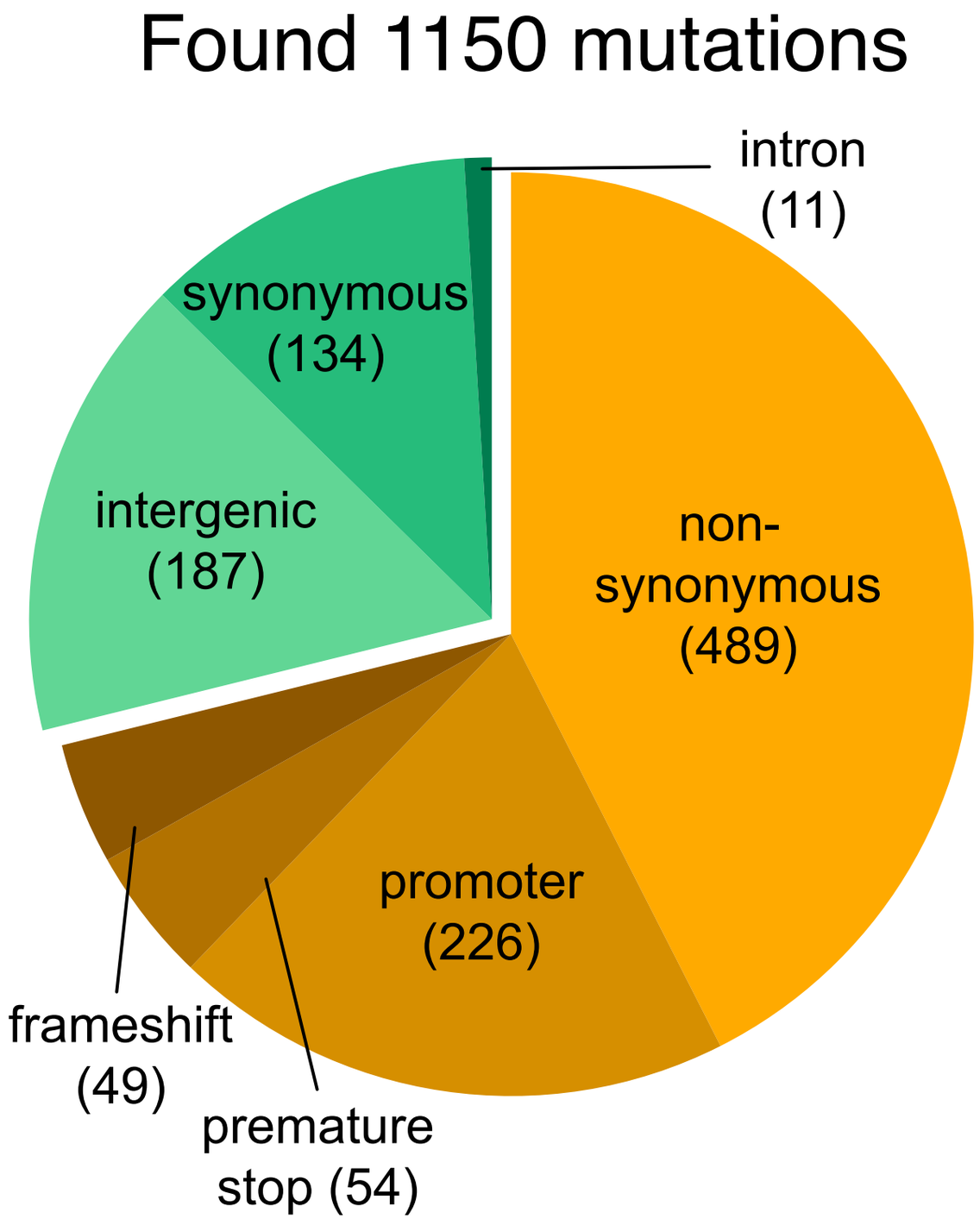




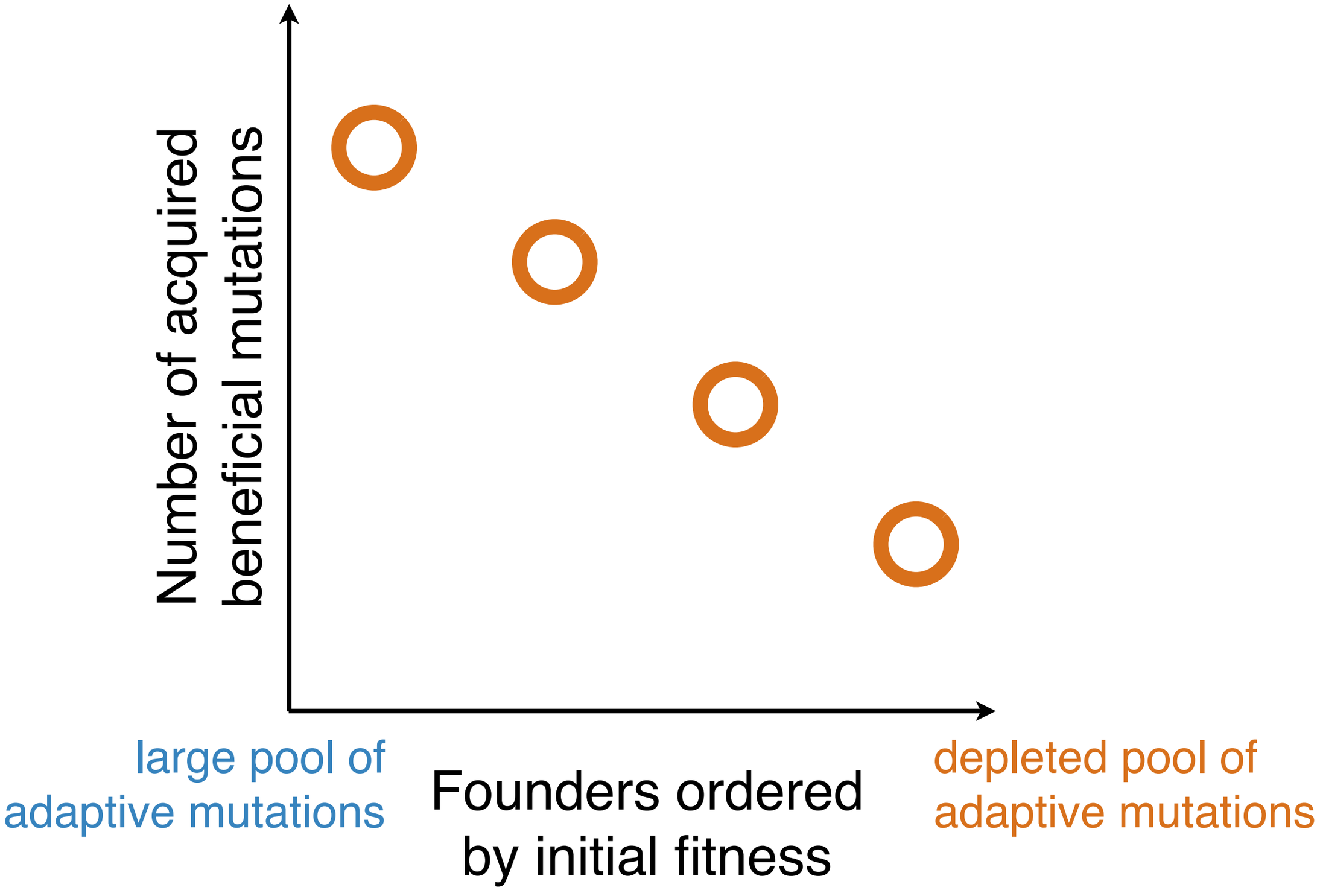
Kryazhimskiy *et al*, *Science* 2014

Why does “adaptability” decline with fitness?

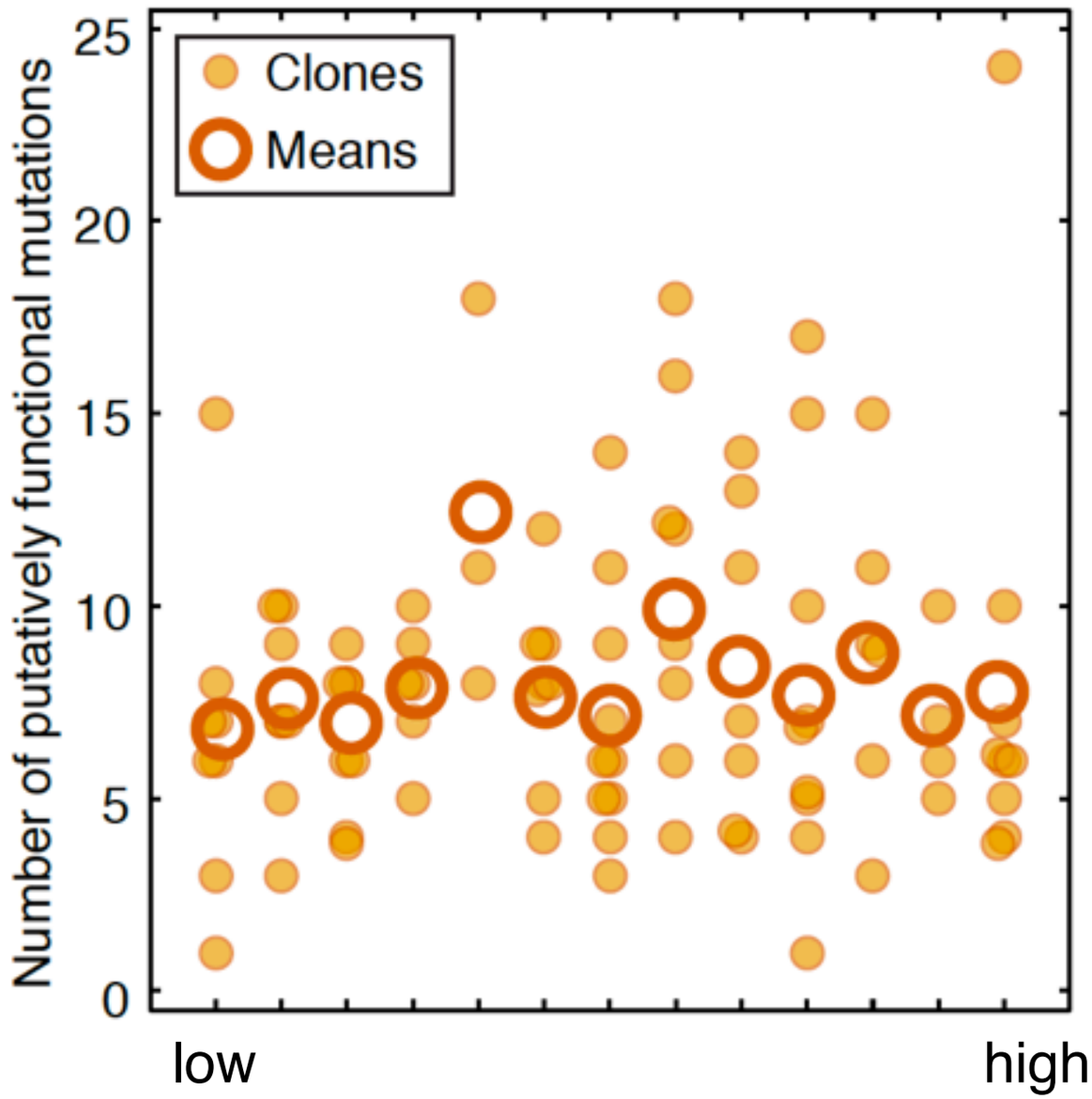
Sequenced full genomes of 104 adapted clones



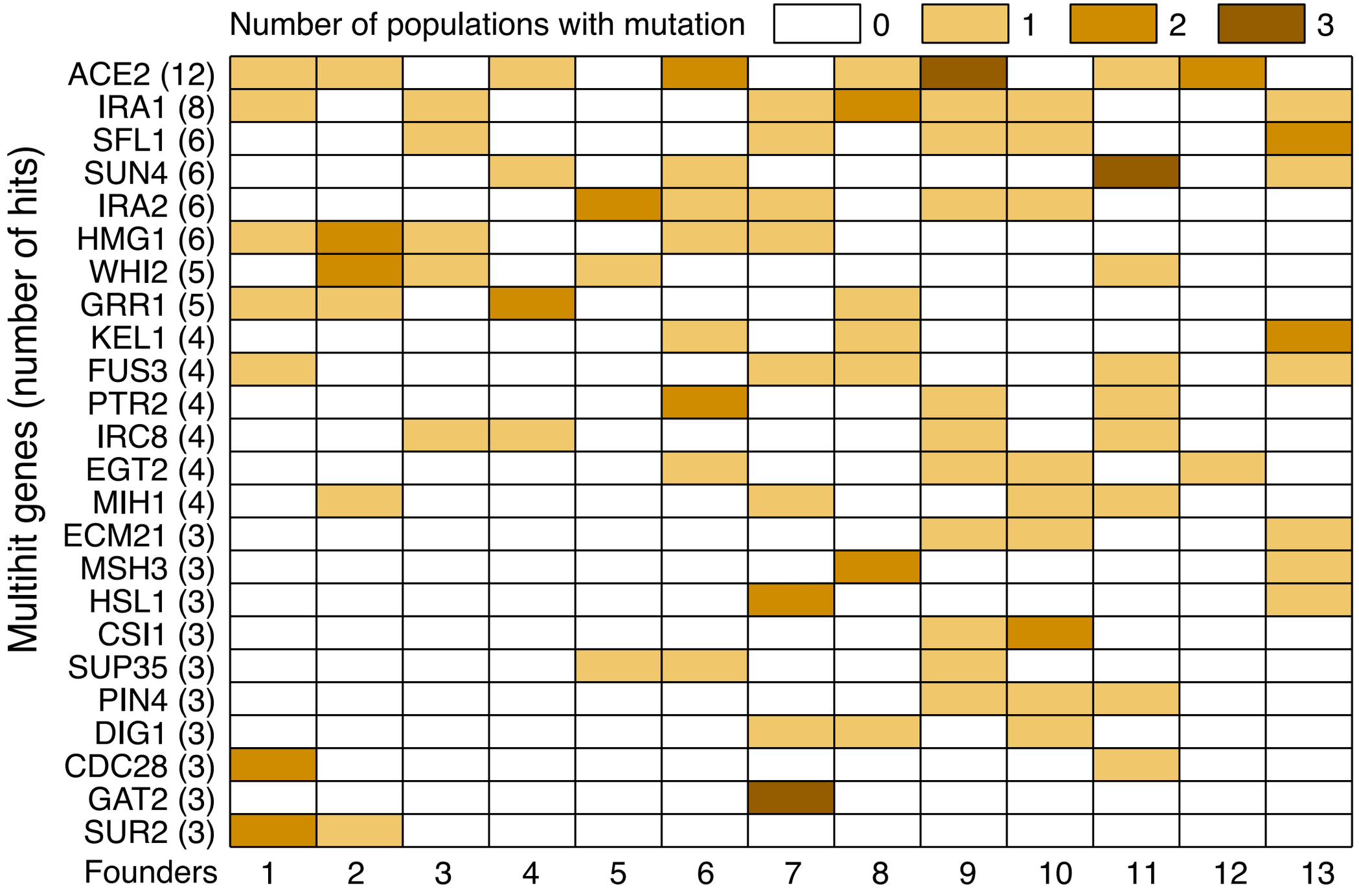
Hypothesis: “Running out” of beneficial mutations



No support for the running out of beneficial mutations



Hypothesis: Different Founders acquire different mutations

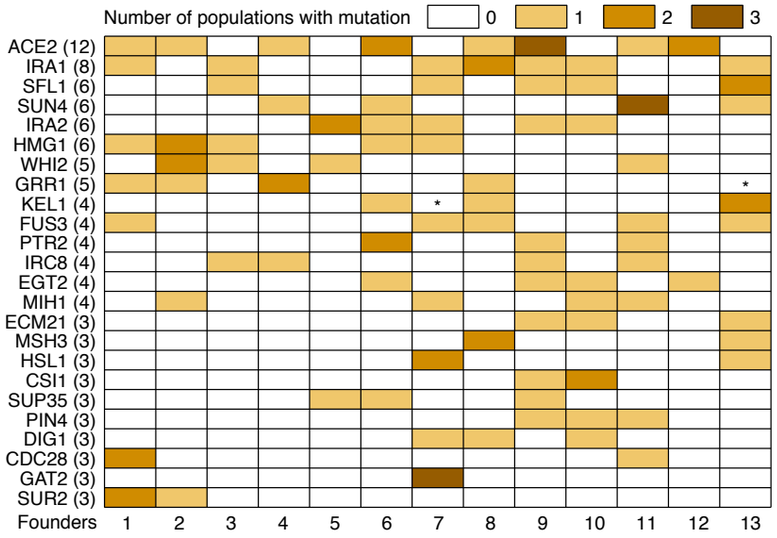


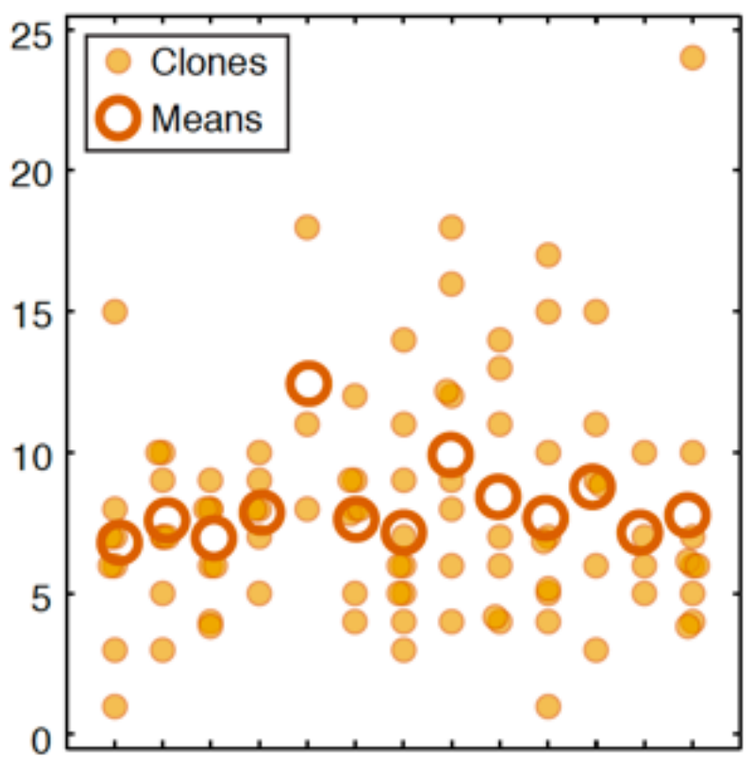
Consider all data together

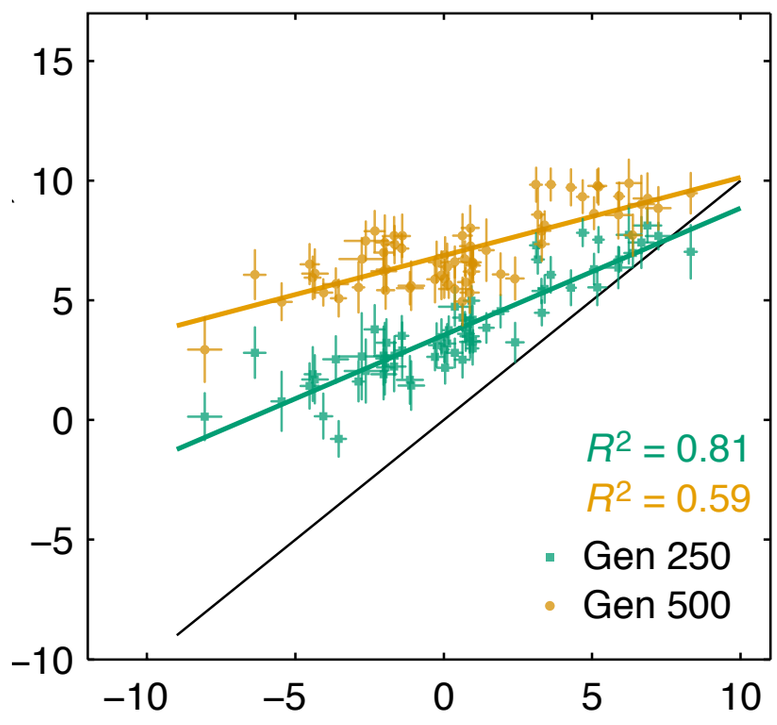
All Founders acquire mutations from the same pool

All Founders acquire same number of mutations

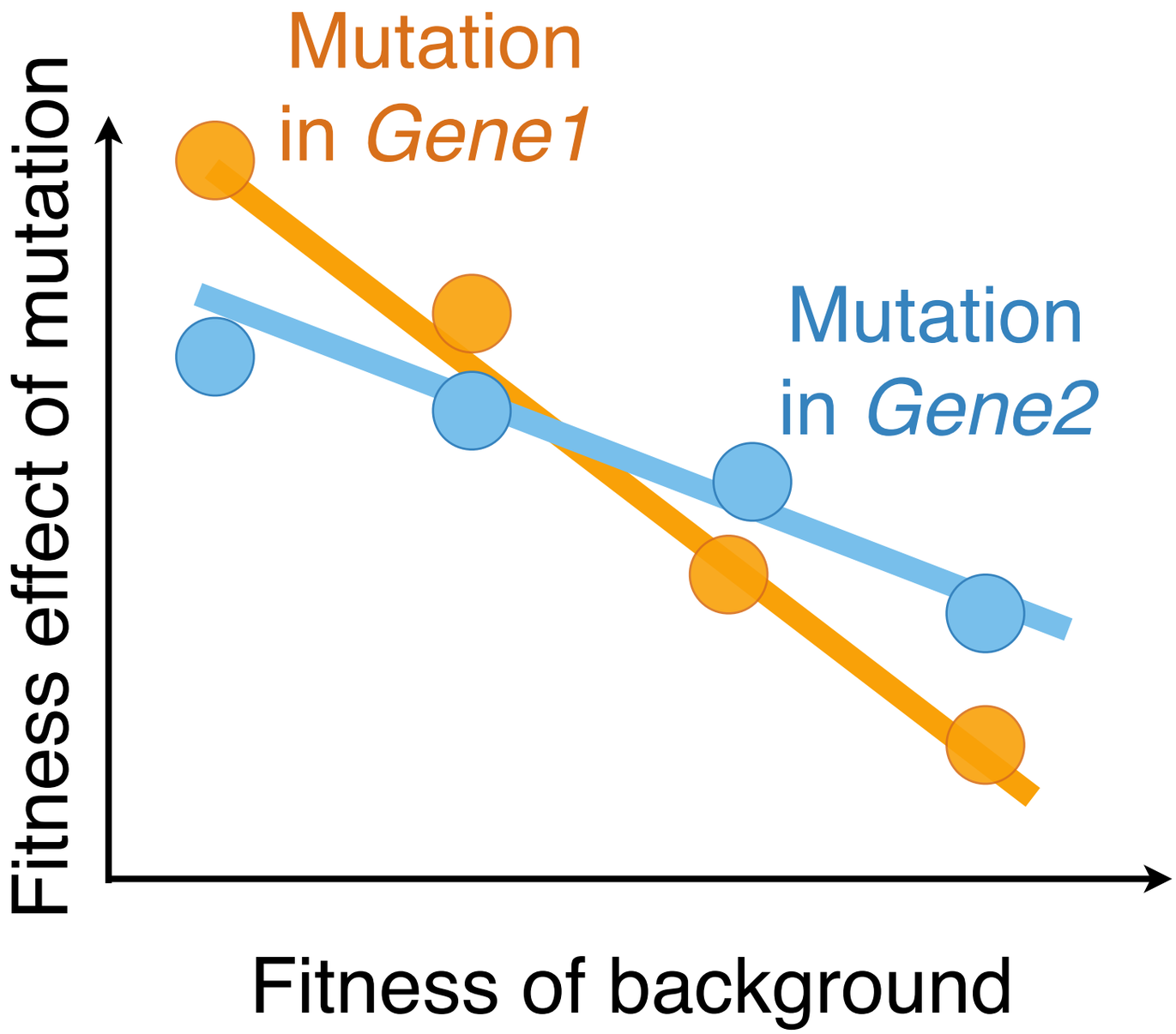
Fitter Founders adapt slower



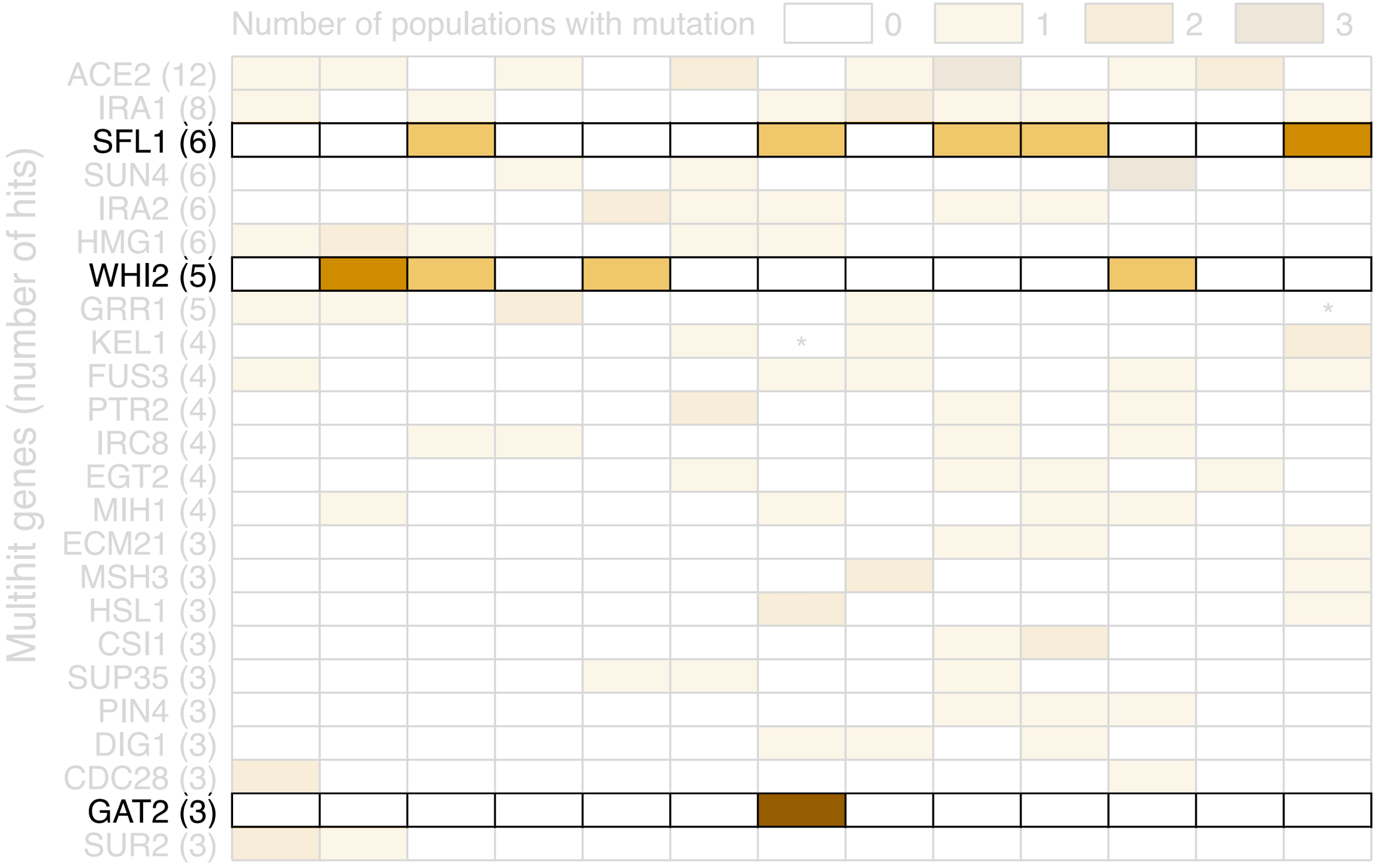




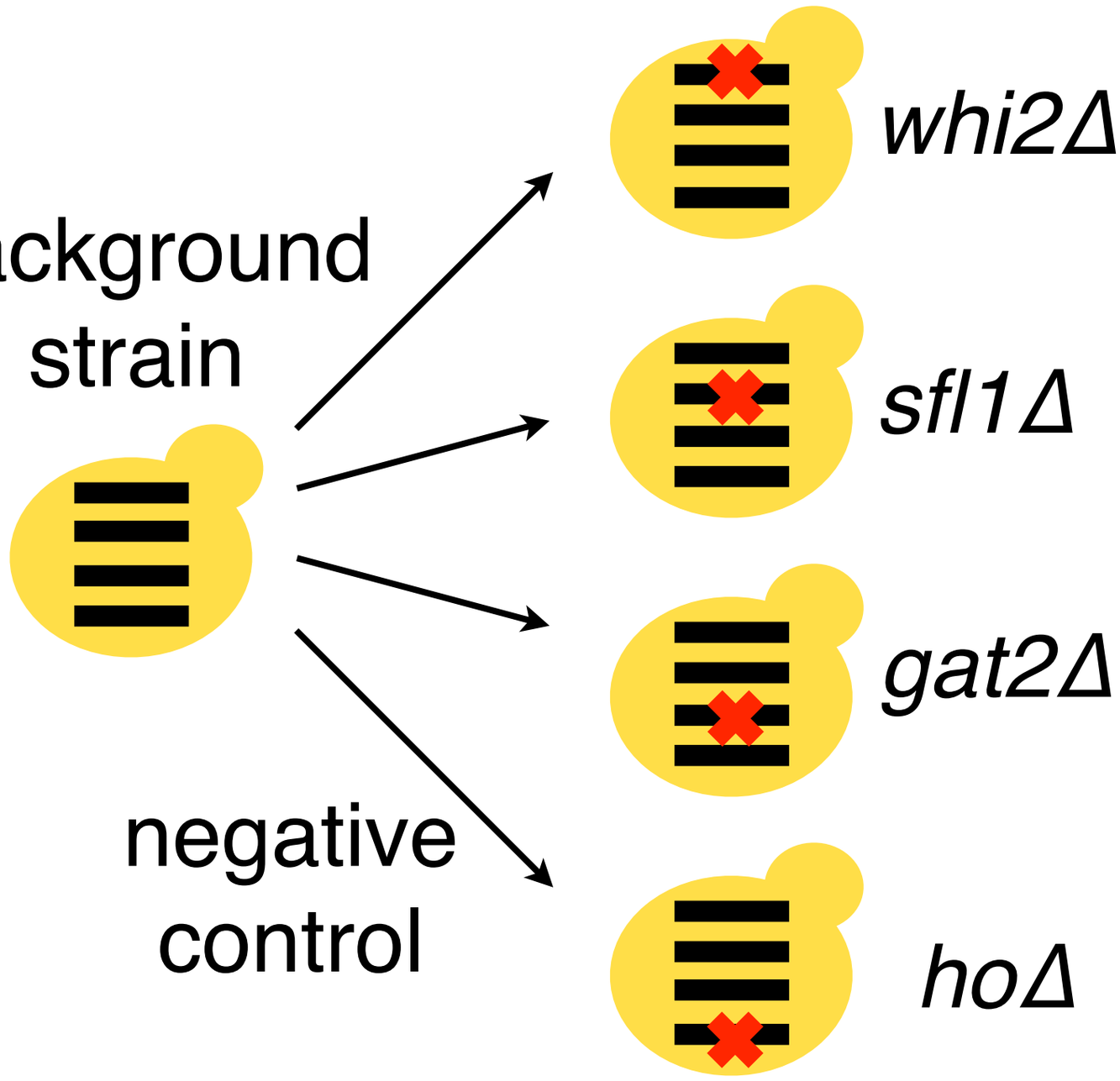
“Diminishing returns” epistasis



Test diminishing returns epistasis by allele replacement



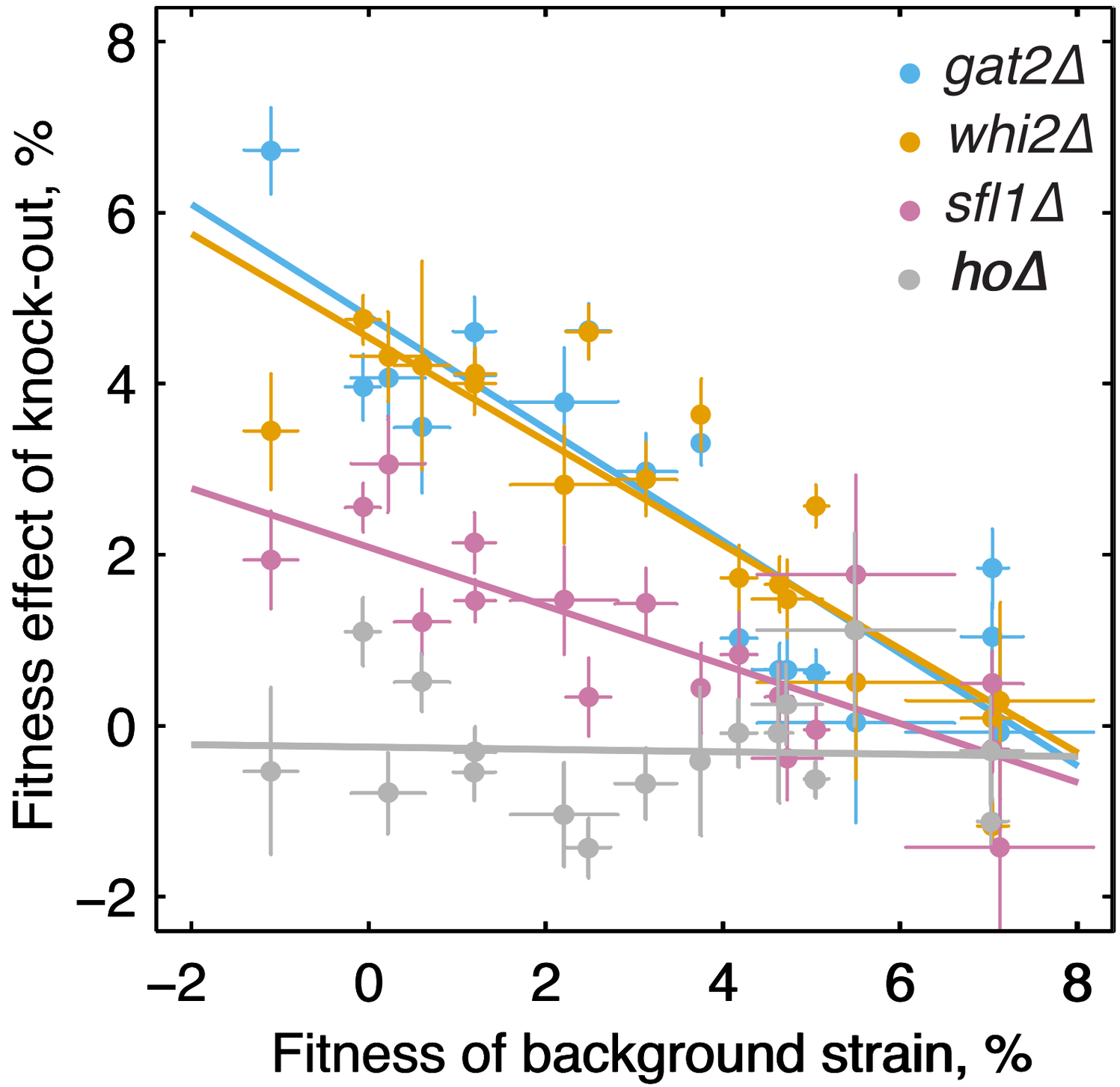
Knock-out genes in different genetic backgrounds



18 backgrounds
×
4 genes
=
72 strains

Measure fitness of knock-out strains

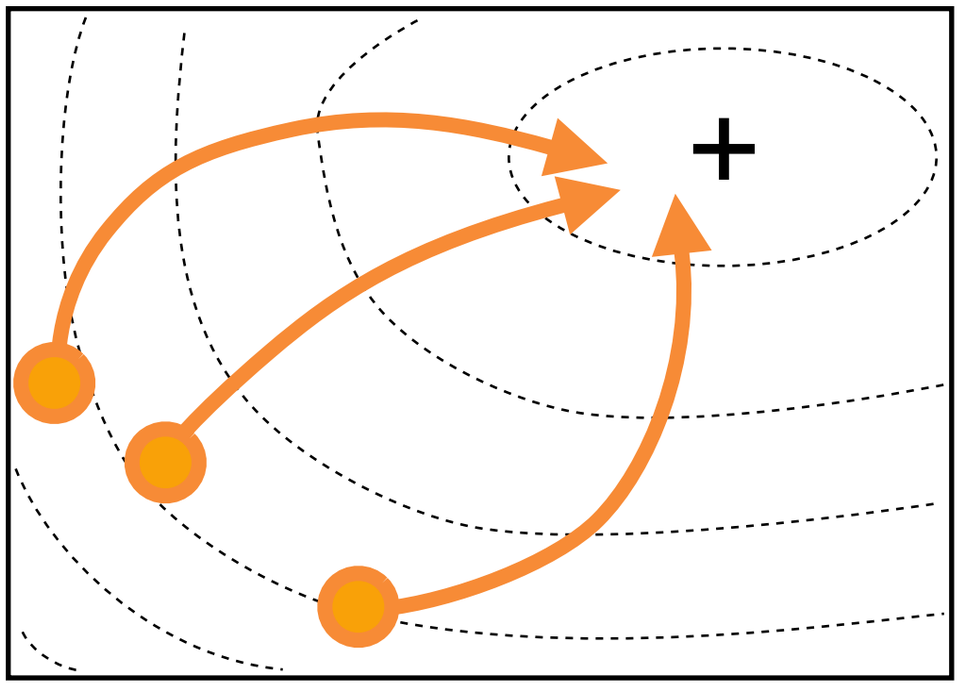
Diminishing returns epistasis supported



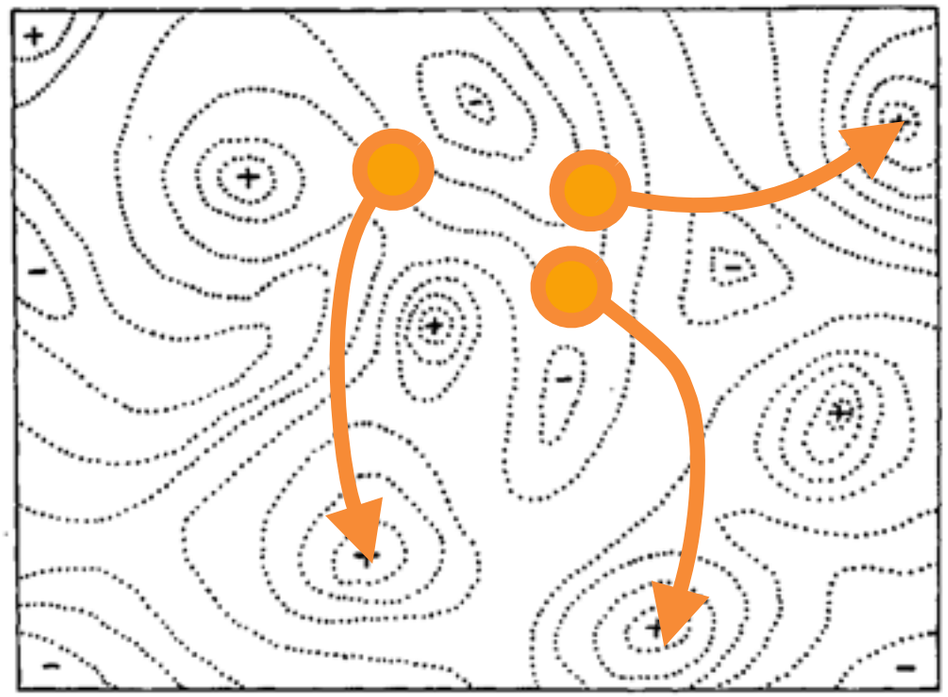
Kryazhimskiy *et al*, *Science* 2014

Fitness landscape structure

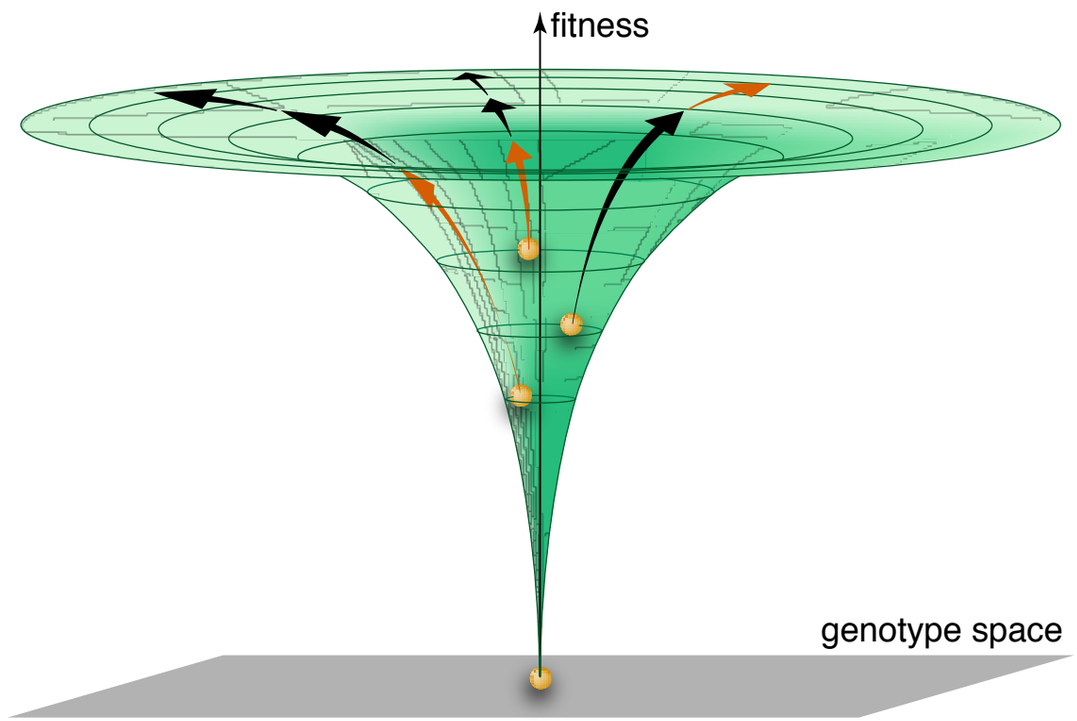
Classic smooth



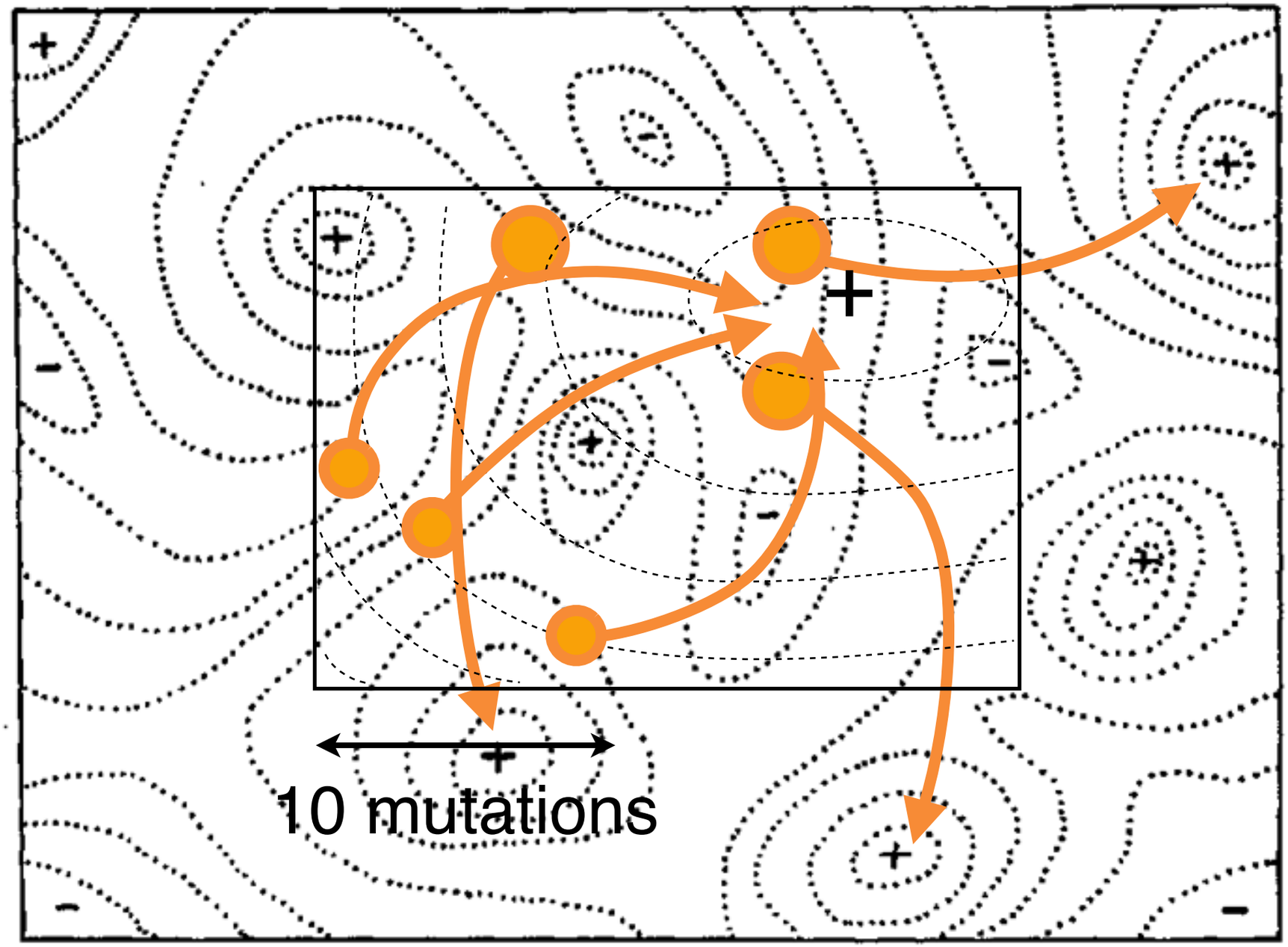
Classic rugged



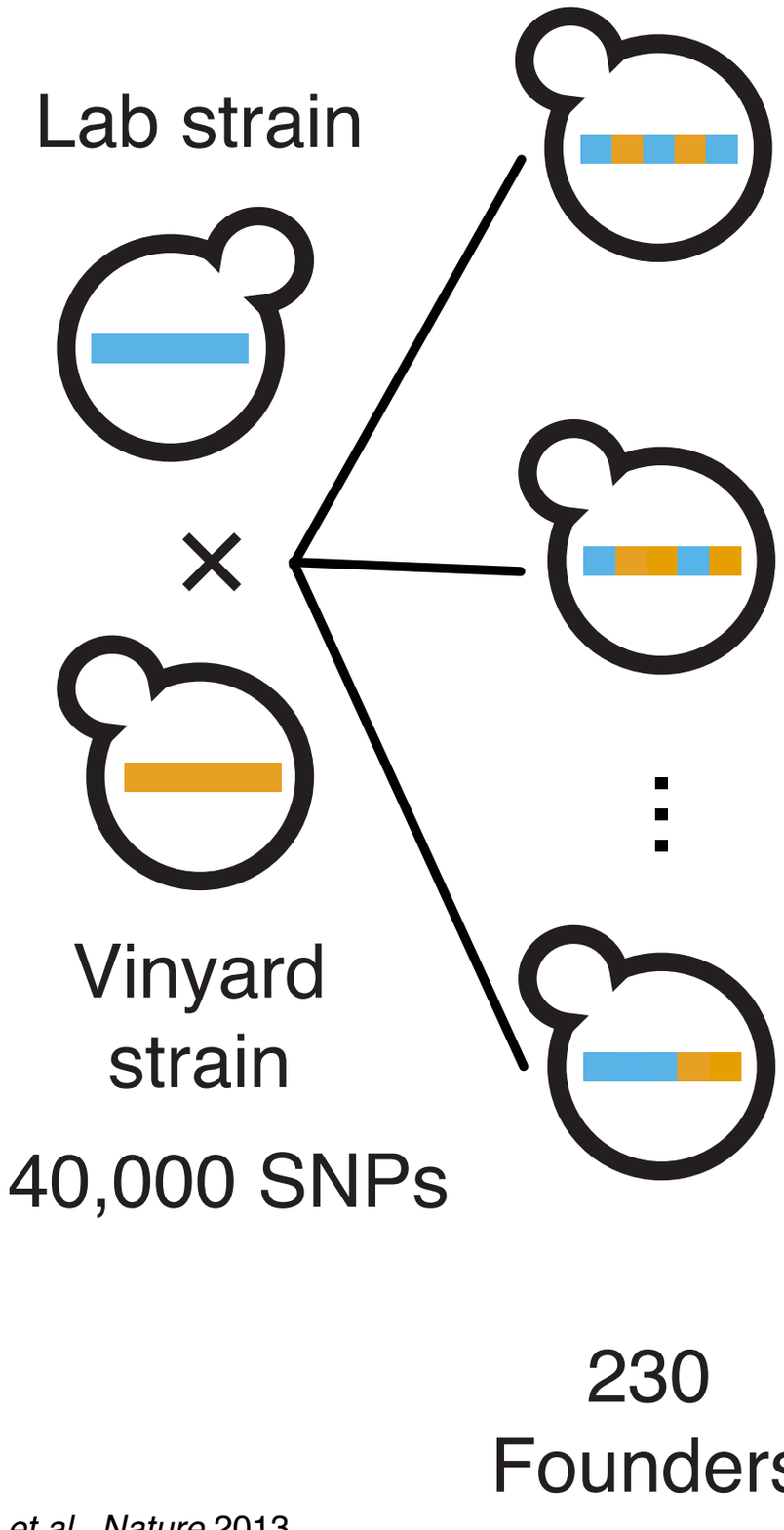
Diminishing returns



Ruggedness likely depends on scale of genetic divergence



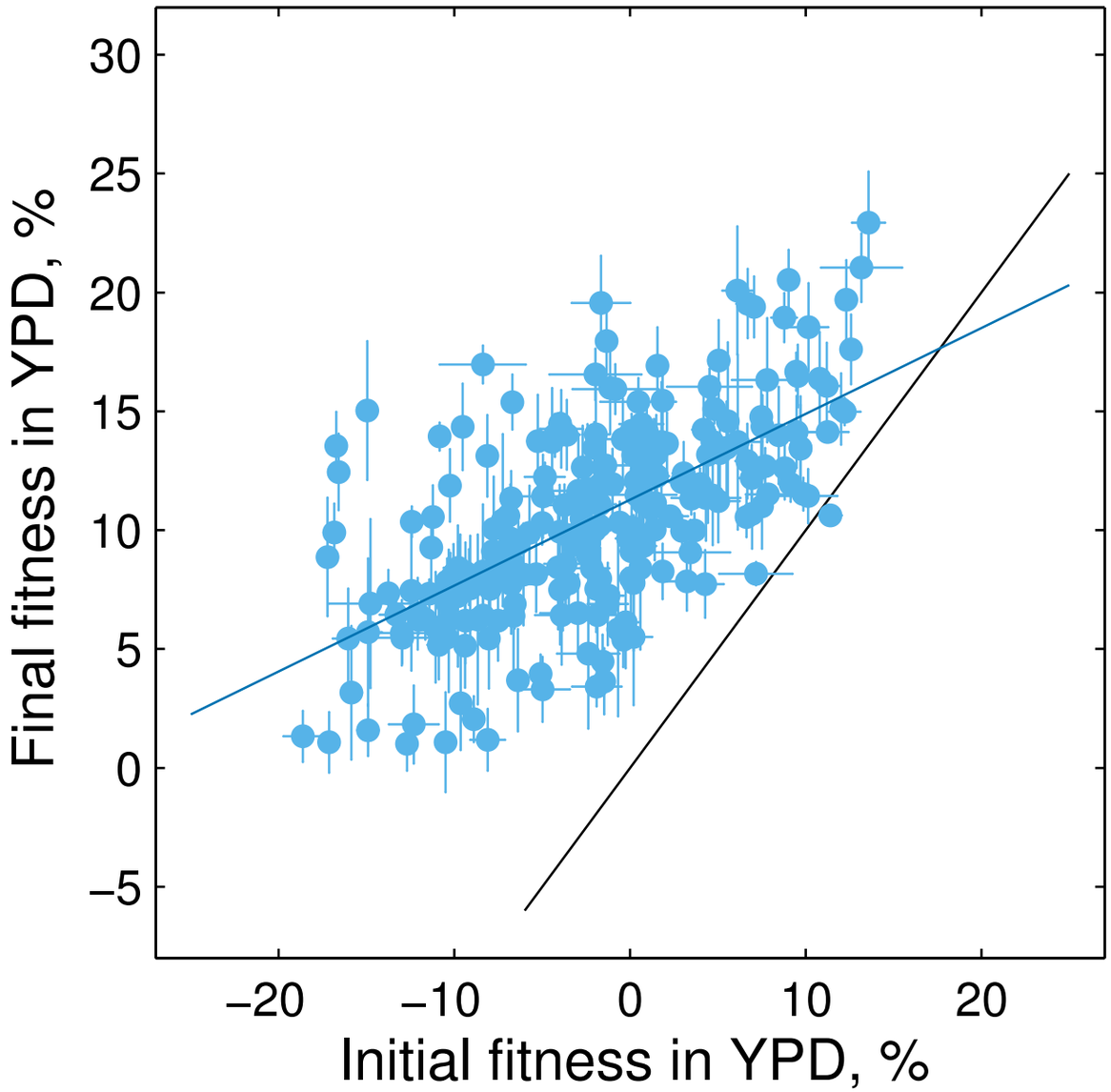
Adaptation from divergent strains



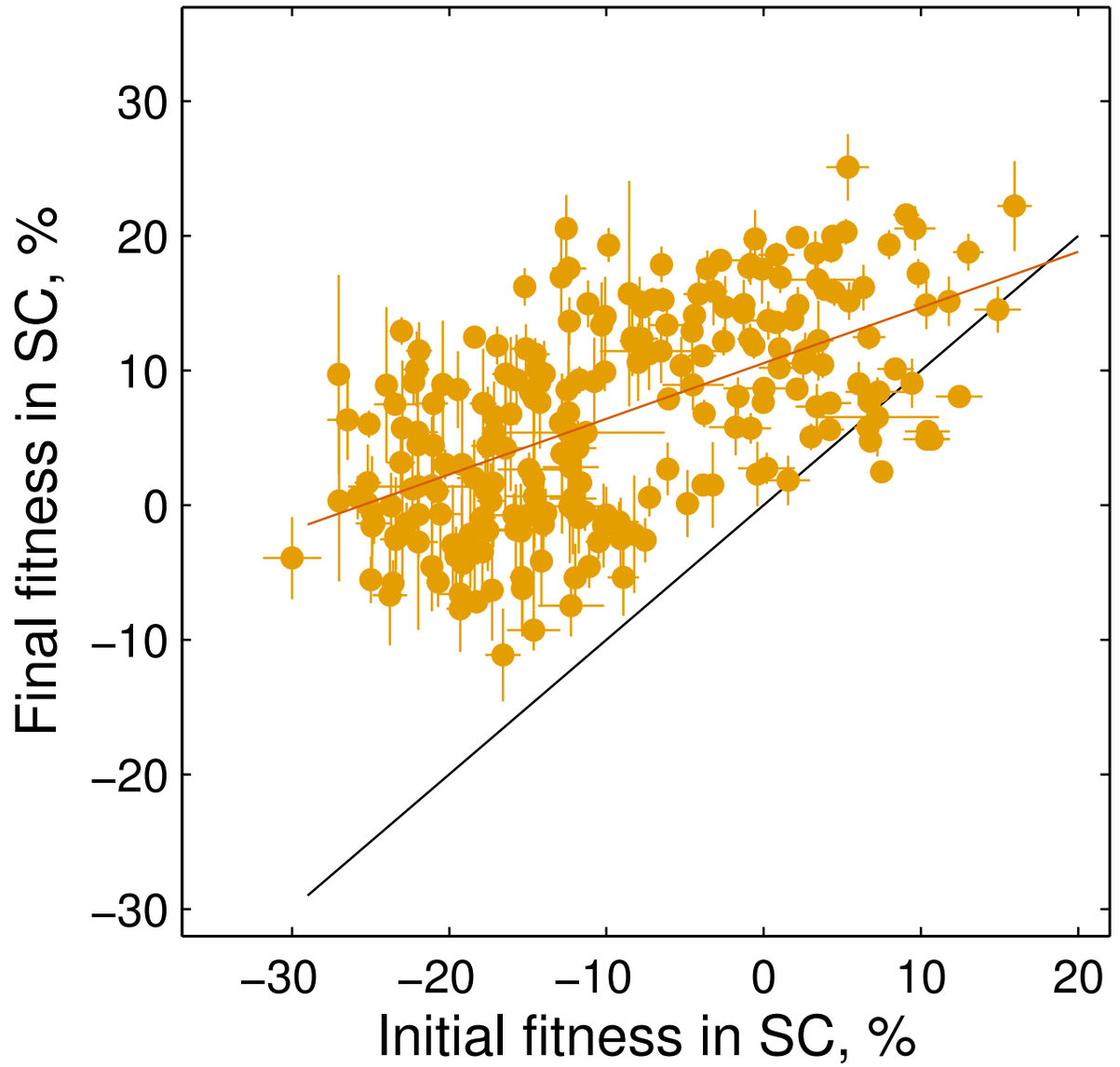
Bloom *et al*, *Nature* 2013

Observe similar decline in adaptability with fitness among divergent strains

Evolved in YPD + 30°C

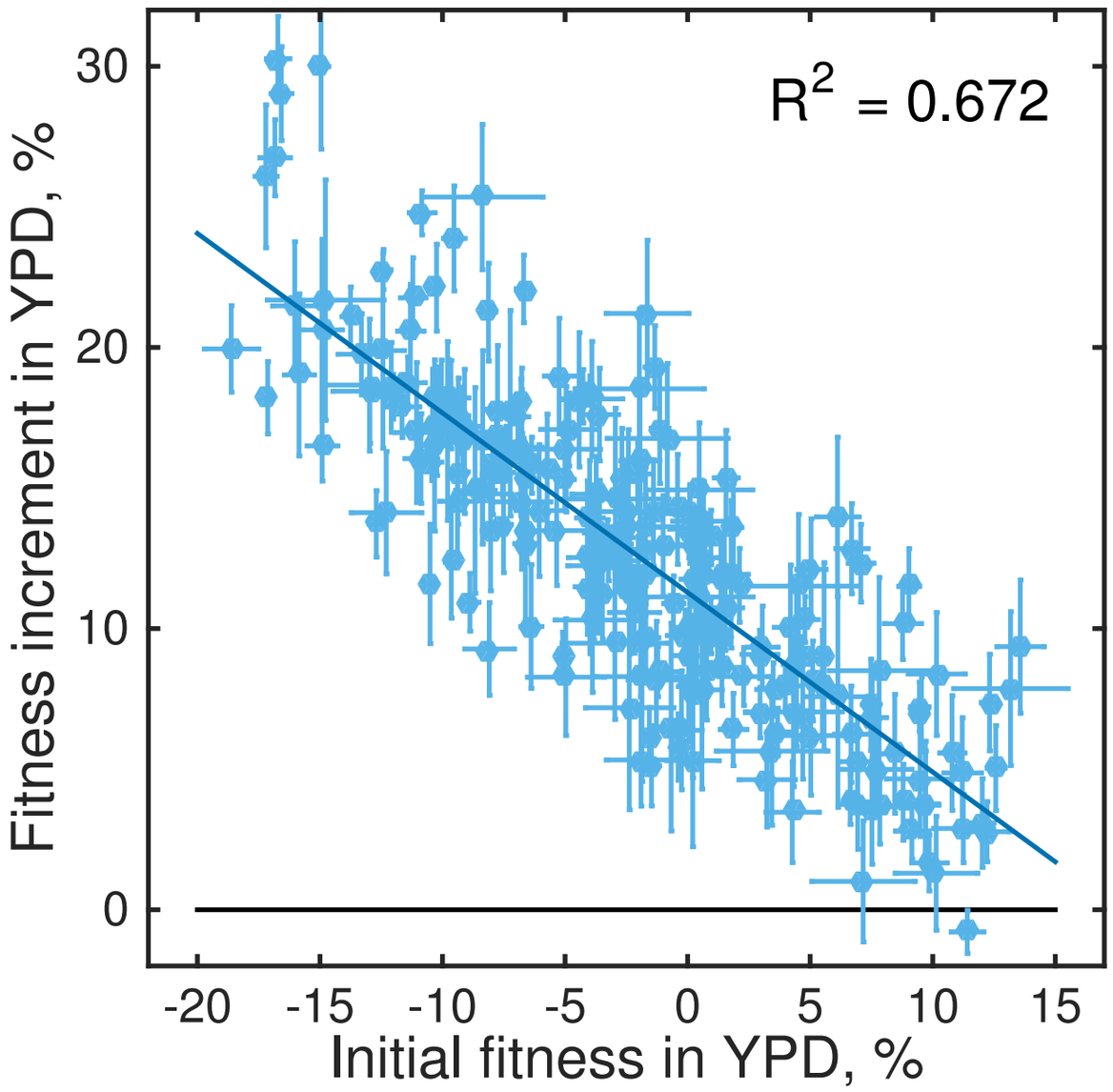


Evolved in SC + 37°C

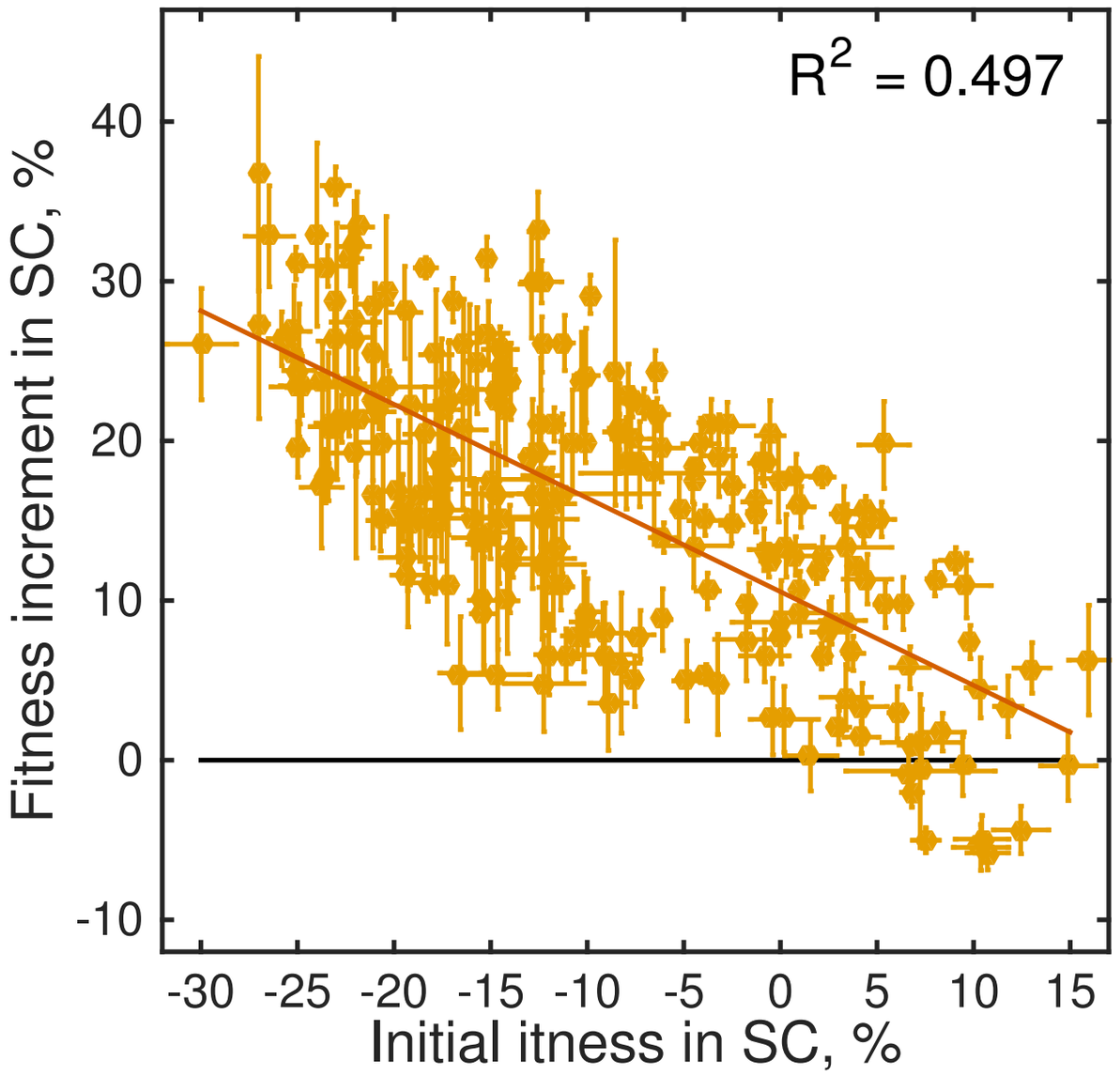


Observe similar decline in adaptability with fitness among divergent strains

Evolved in YPD + 30°C

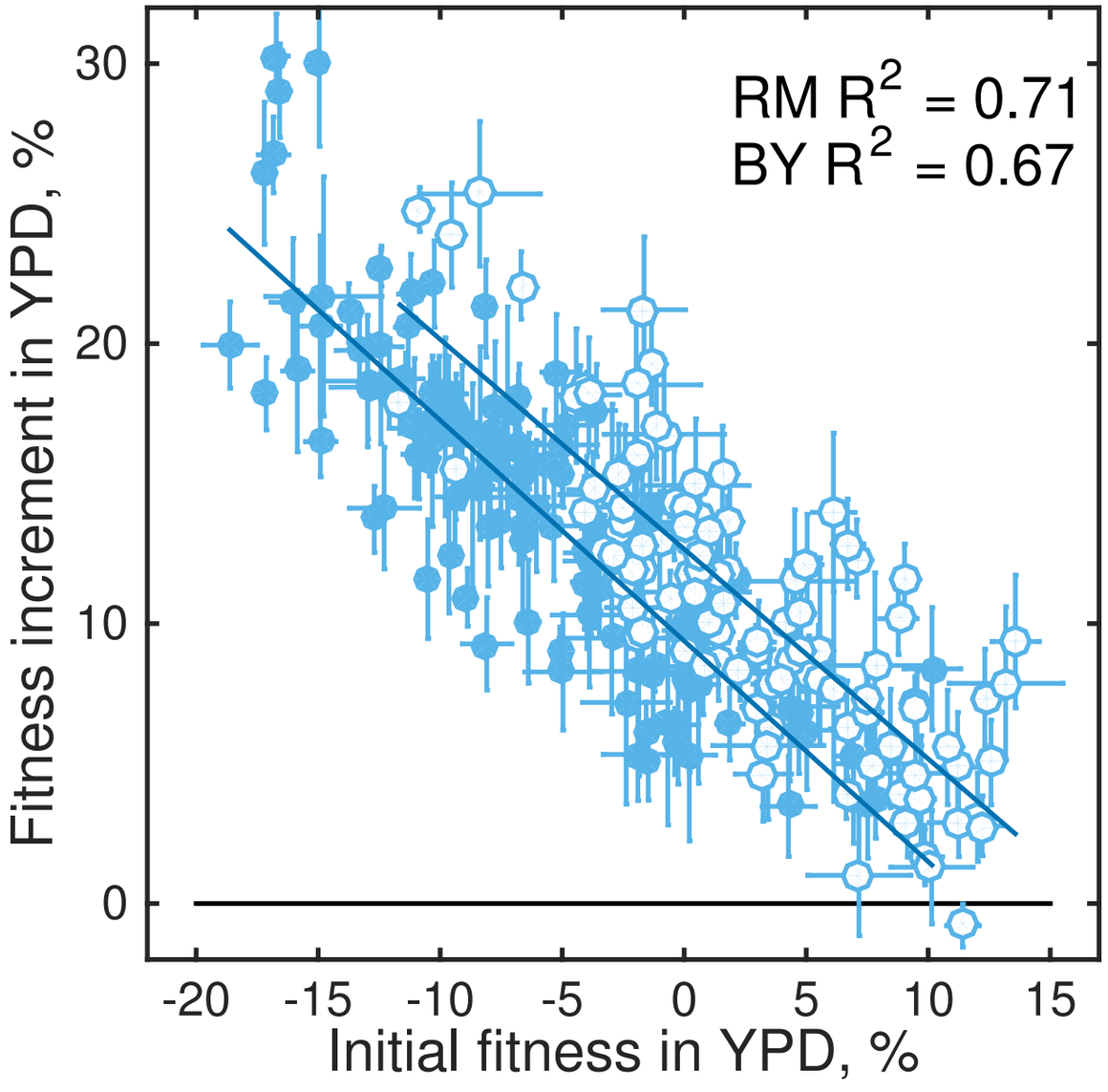


Evolved in SC + 37°C

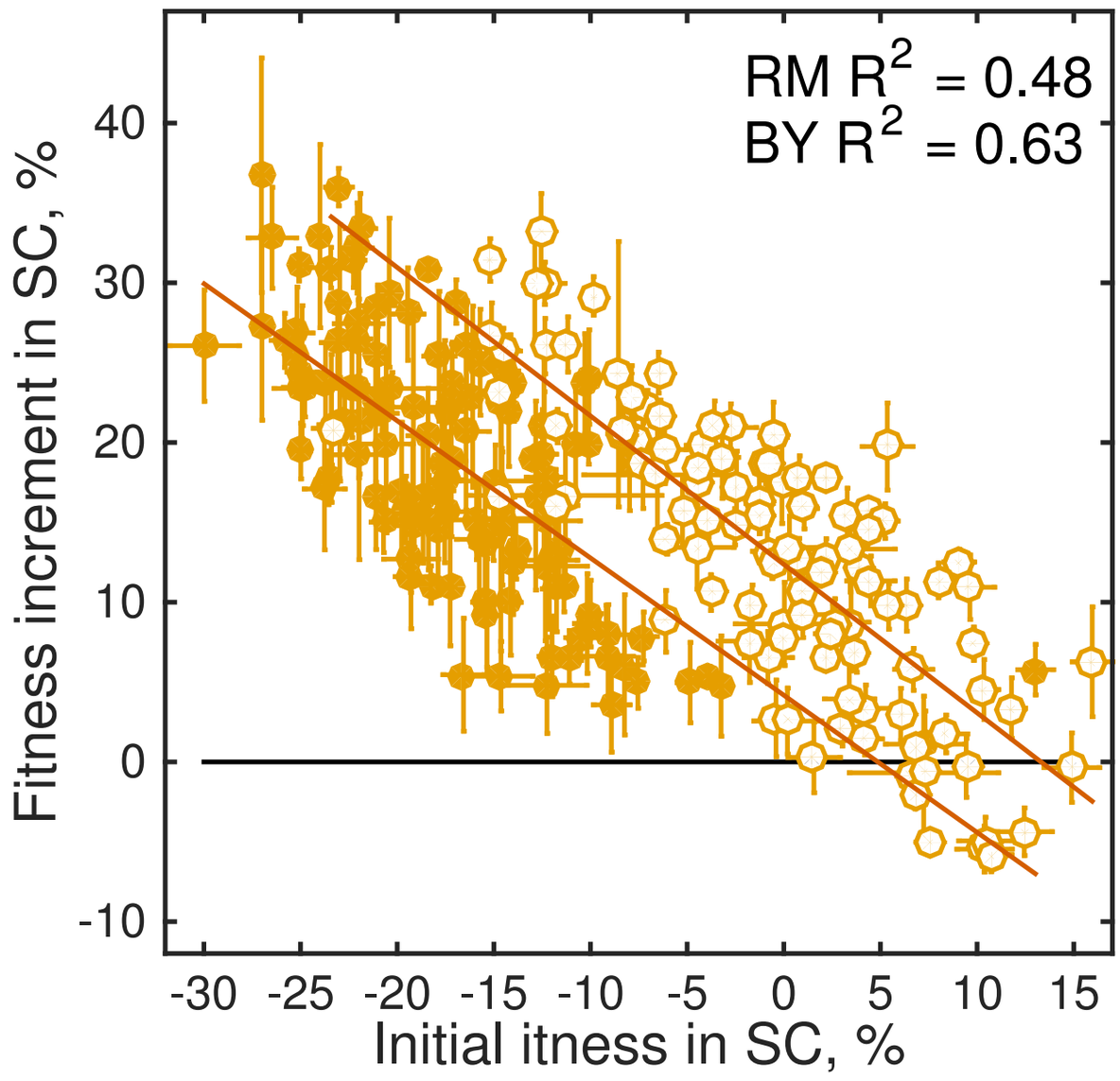


SNP at KRE33 locus affects adaptability

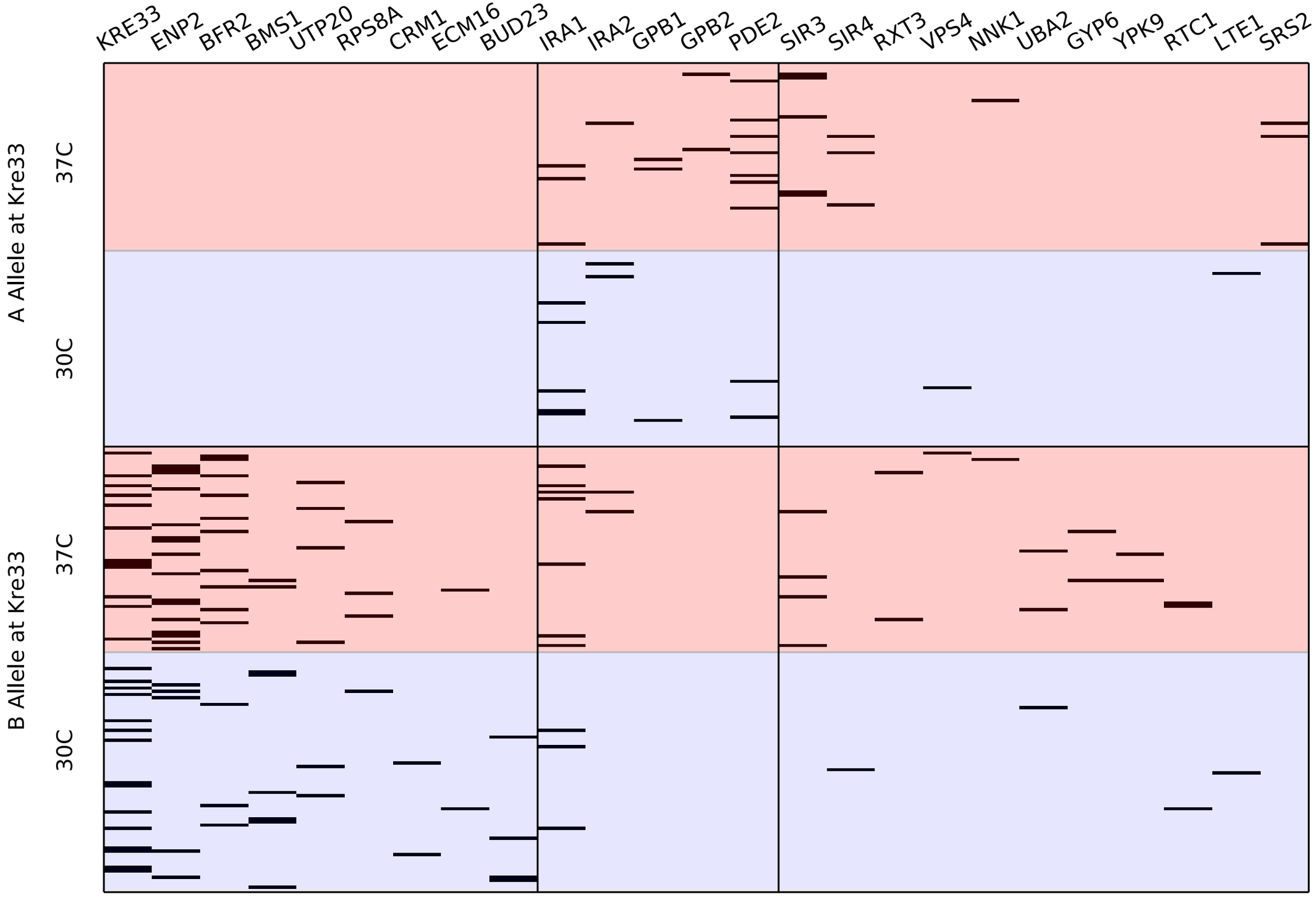
Evolved in YPD + 30°C



Evolved in SC + 37°C



SNP at KRE33 locus affects the pool of adaptive mutations



Conclusions

"Rule of declining adaptability" = rate of adaptation declines with initial fitness

Negative "diminishing returns" epistasis is at least partially responsible

Pool of adaptive mutations is common to most closely related genotypes

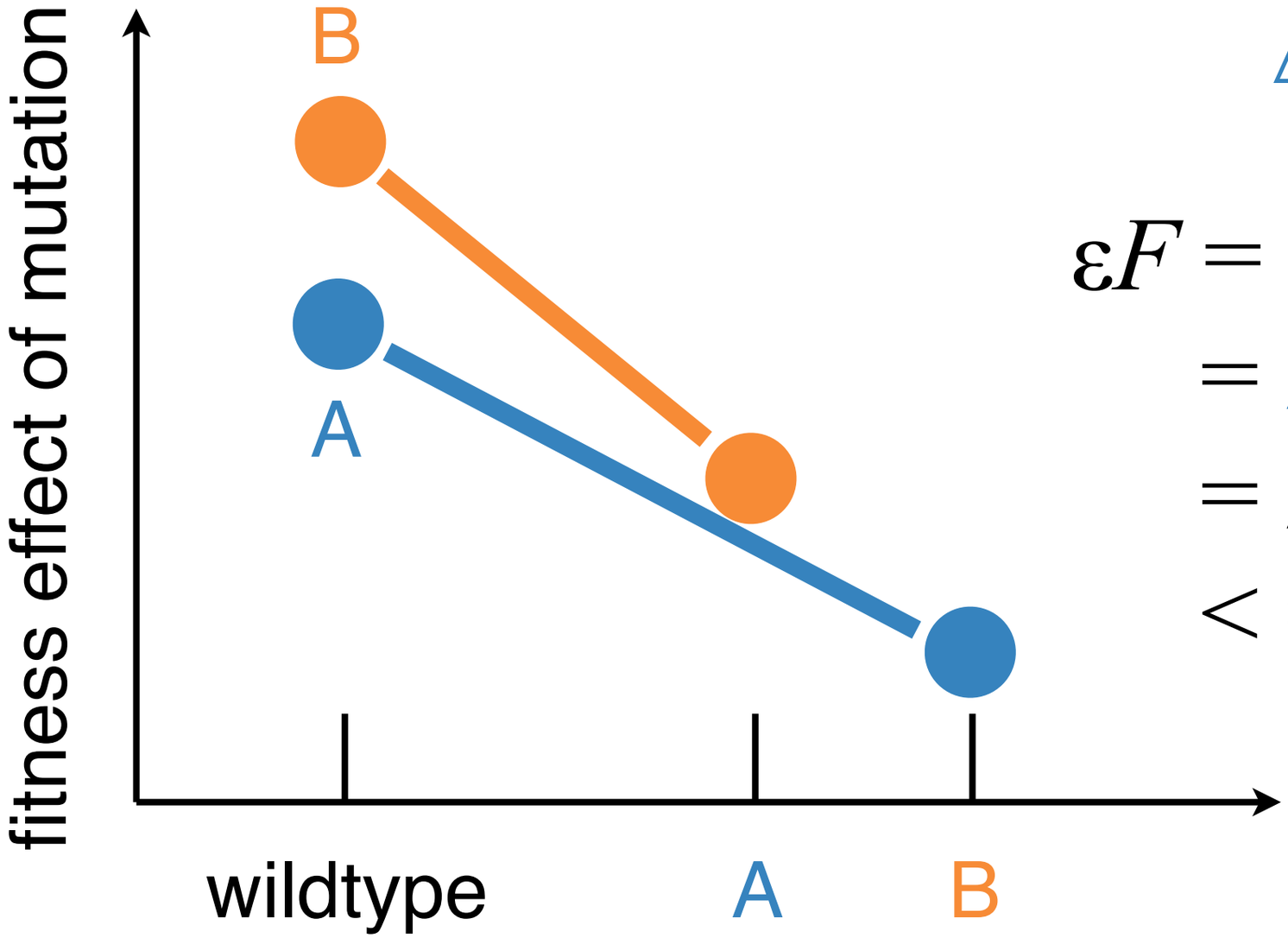
Some (rare) mutations dramatically change the pool of adaptive mutations

Part 2.

Where does epistasis come from?

Epistasis between two mutations

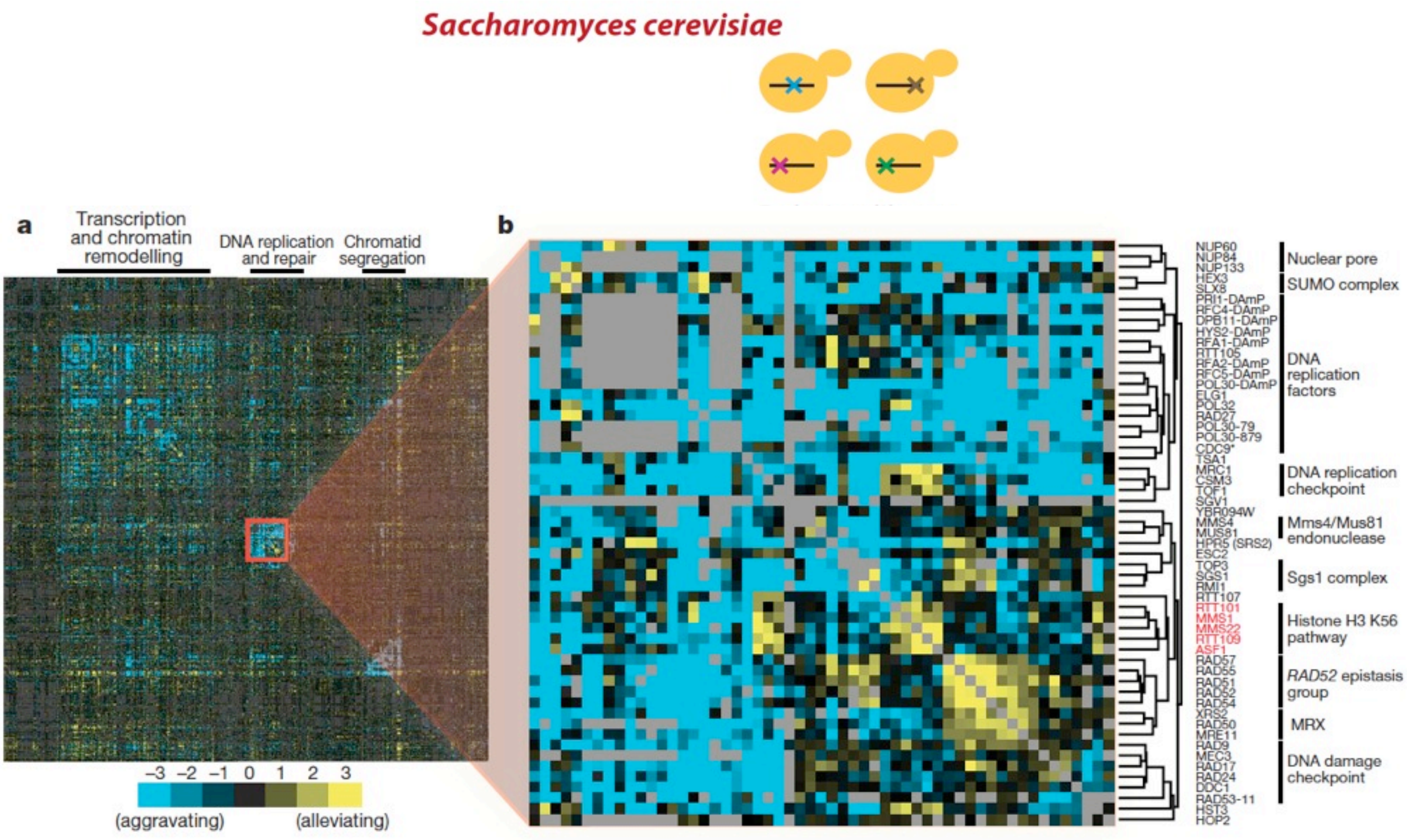
"Diminishing returns" epistasis between two beneficial mutations



$$\Delta F_{\mathrm{B|A}} < \Delta F_{\mathrm{B|WT}}$$
$$\Delta F_{\mathrm{A|B}} < \Delta F_{\mathrm{A|WT}}$$

$$\begin{aligned}
\varepsilon F &= \Delta F_{\mathrm{B|A}} - \Delta F_{\mathrm{B|WT}} \\
&= \Delta F_{\mathrm{A|B}} - \Delta F_{\mathrm{A|WT}} \\
&= \Delta F_{\mathrm{AB}} - \Delta F_{\mathrm{A}} - \Delta F_{\mathrm{B}} \\
&< 0
\end{aligned}$$

Systematic measurements of epistasis between gene knock-out mutations



Collins *et al*, *Nature* 2007

We measured epistasis. So what?

Do we have a null expectation for epistasis?

Defining genetic interaction

Ramamurthy Mani*, Robert P. St.Onge†, John L. Hartman IV‡, Guri Giaever§, and Frederick P. Roth*¶‖

Properties Expected of an Ideal Definition for Identifying Functional Relationships. Gene function can be defined at multiple levels of specificity (27). By definition, there are few genes that hold any given specific function, and gene pairs sharing a specific function should then also be rare. Therefore, if interaction (either synthetic or alleviating) is to be an ideal indicator of specific functional relationships, the vast majority of gene pairs should be noninteracting. An ideal definition for interaction should then yield a distribution of observed double-mutant fitness values that closely approximates the expected distribution over most gene pairs.

Do we have a null expectation for epistasis?

Quantitative analysis of fitness and genetic interactions in yeast on a genome scale

Anastasia Baryshnikova[1,2,10], Michael Costanzo[1,10], Yungil Kim[3,4], Huiming Ding[1], Judice Koh[1], Kiana Toufighi[1], Ji-Young Youn[1,2], Jiongwen Ou[5], Bryan-Joseph San Luis[1], Sunayan Bandyopadhyay[3], Matthew Hibbs[6], David Hess[7], Anne-Claude Gingras[8], Gary D Bader[1,2], Olga G Troyanskaya[9], Grant W Brown[5], Brenda Andrews[1,2], Charles Boone[1,2] & Chad L Myers[3]

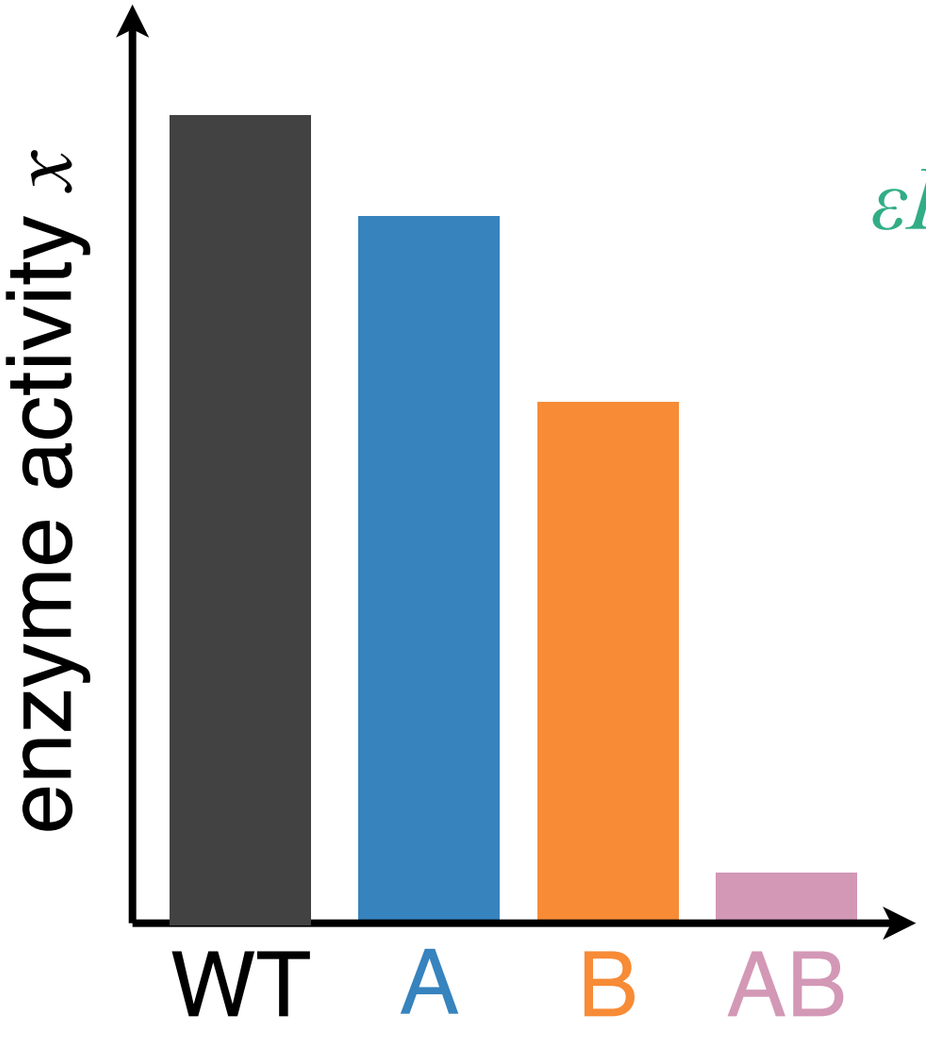
size is related to the corresponding query mutation. From **Eq. 1** above, we assume that observed colony area is a function of the single mutations fitness defects as well as time:

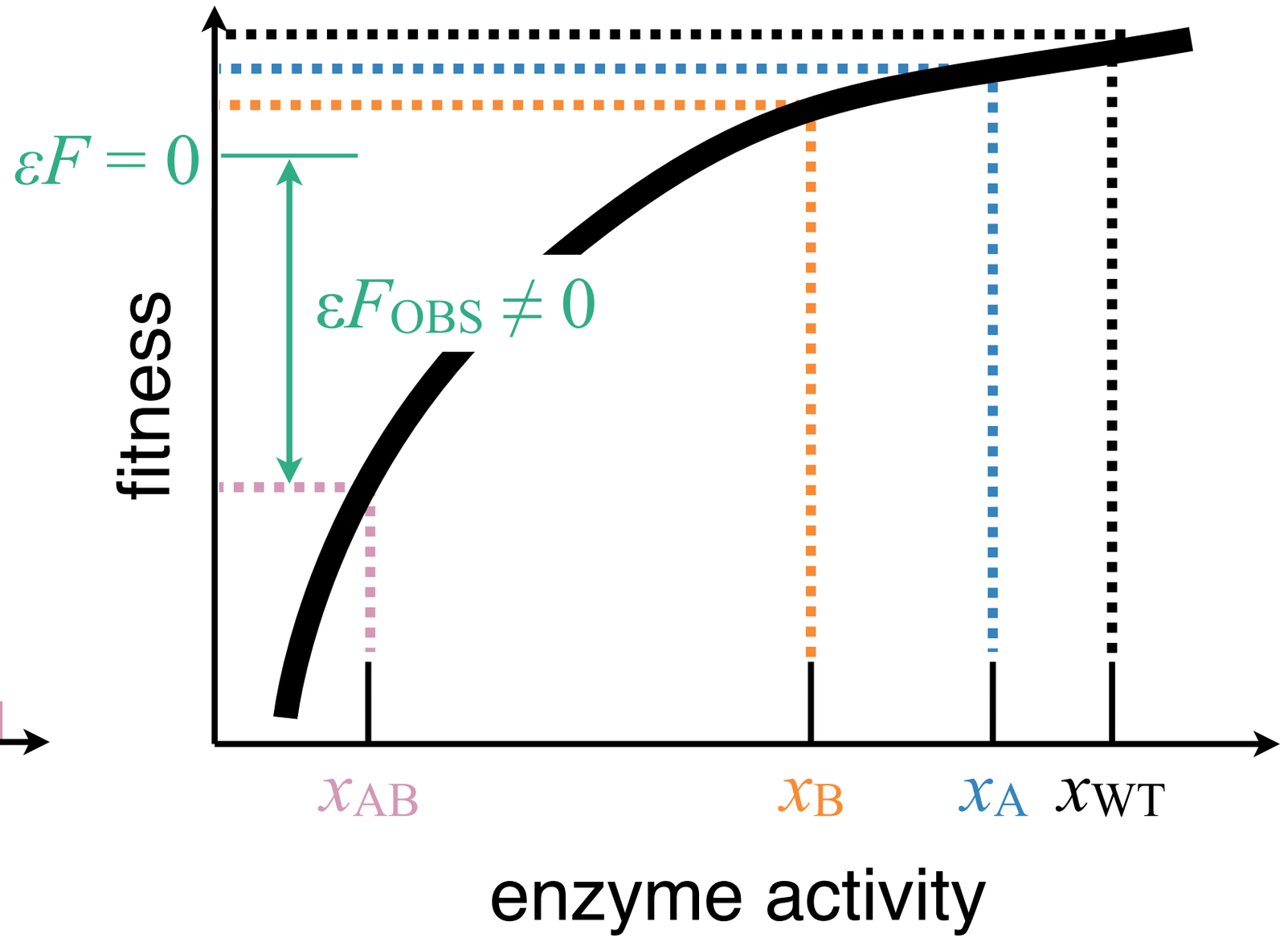
$$C_{ij} = \alpha \cdot i_j \cdot f_j \cdot t \cdot s_{ij} \cdot e \qquad \text{Eq. 2}$$

as in most cases $\varepsilon_{ij} \approx 0$ because genetic interactions are rare. Due to the

Is “no epistasis” a biologically meaningful expectation?

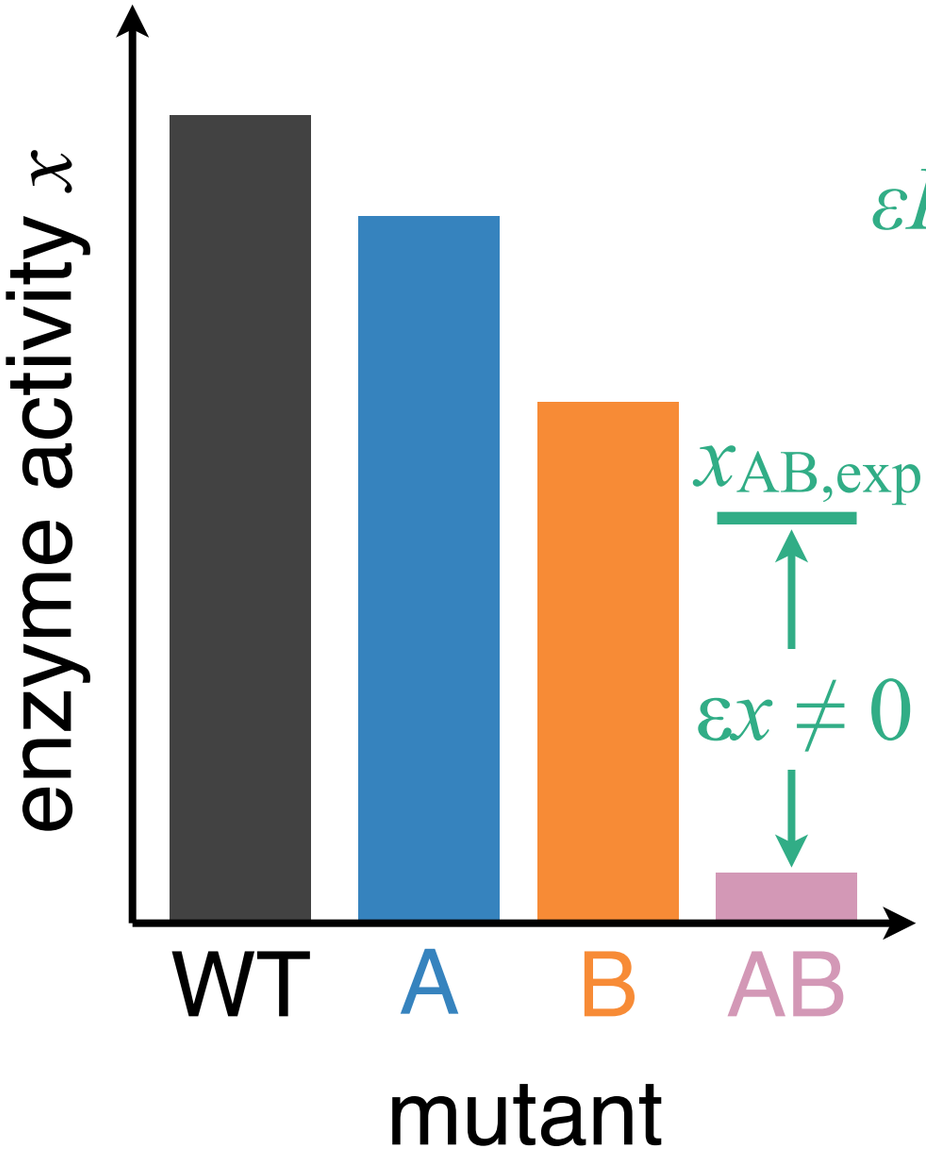
Simple example: epistasis between mutations in the same enzyme



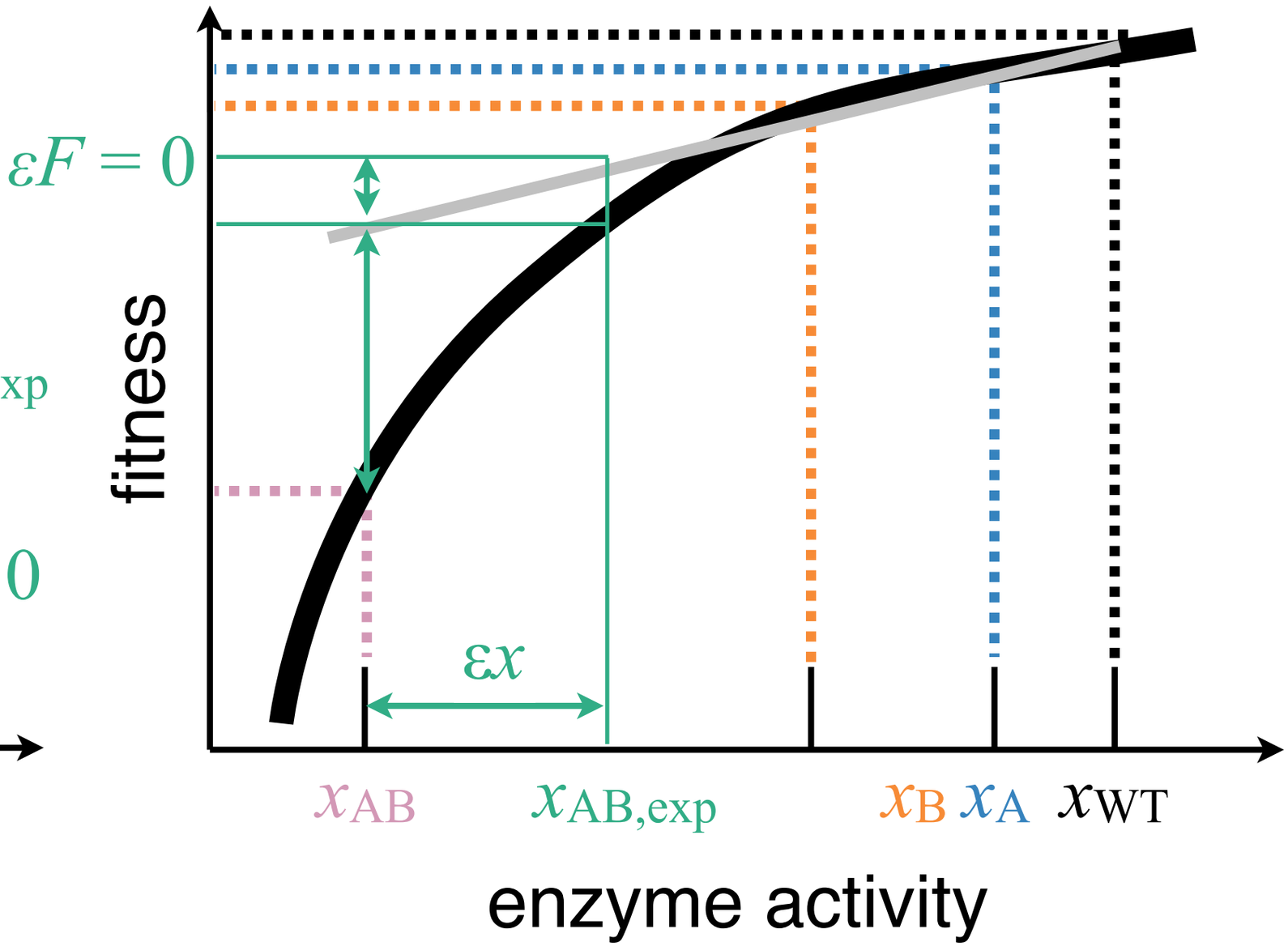


Within-protein epistasis arises from two contributions

1. Activity is not additive



2. Fitness is a non-linear function of activity



Within-protein epistasis arises from two contributions

$F = F(x)$ Fitness is function of activity

$\delta^{\mathrm{A}}x,\ \delta^{\mathrm{B}}x,\ \delta^{\mathrm{AB}}x$ Effects of mutations on activity

$\varepsilon^{\mathrm{AB}}x = \delta^{\mathrm{AB}}x - \delta^{\mathrm{A}}x - \delta^{\mathrm{B}}x$ Epistasis for activity

$\delta^{\mathrm{A}}F,\ \delta^{\mathrm{B}}F,\ \delta^{\mathrm{AB}}F$ Effects of mutations on fitness

$\varepsilon^{\mathrm{AB}}F = \delta^{\mathrm{AB}}F - \delta^{\mathrm{A}}F - \delta^{\mathrm{B}}F$ Epistasis for fitness

$$\varepsilon^{\mathrm{AB}}F = C{\cdot}\varepsilon^{\mathrm{AB}}x + H{\cdot}\delta^{\mathrm{A}}x{\cdot}\delta^{\mathrm{B}}x$$

“Propagation of epistasis”

Epistasis between enzymes also arises from two contributions

$F = F(x_1,x_2)$ Fitness is function of activities x_1, x_2

$\delta^{\mathrm{A}}x_i$, $\delta^{\mathrm{B}}x_i$, $\delta^{\mathrm{AB}}x_i$ Effects of mutations on activity i

$\varepsilon^{\mathrm{AB}}x_i = \delta^{\mathrm{AB}}x_i - \delta^{\mathrm{A}}x_i - \delta^{\mathrm{B}}x_i$ Epistasis for activity i

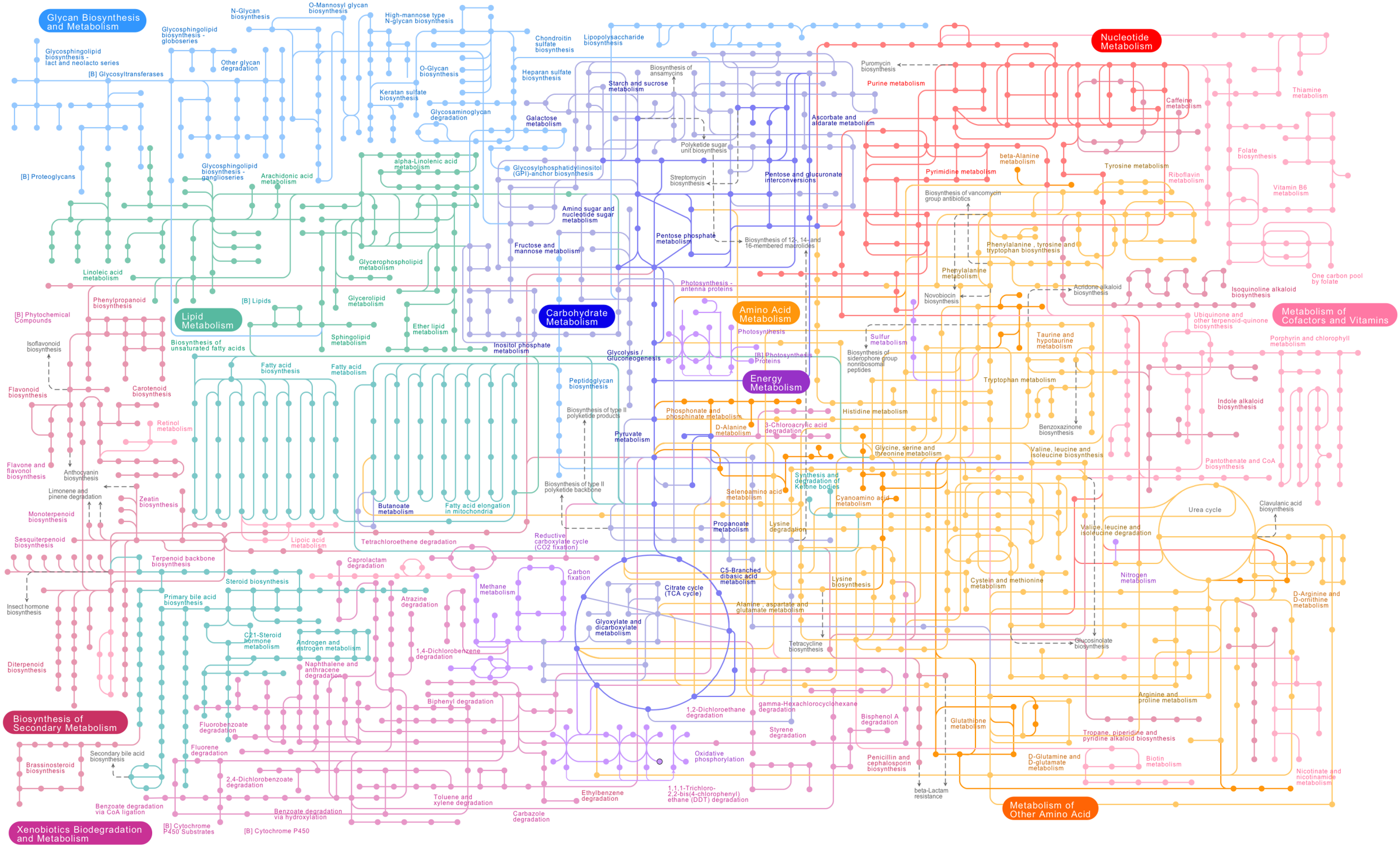
$\delta^{\mathrm{A}}F$, $\delta^{\mathrm{B}}F$, $\delta^{\mathrm{AB}}F$ Effects of mutations on fitness

$\varepsilon^{\mathrm{AB}}F = \delta^{\mathrm{AB}}F - \delta^{\mathrm{A}}F - \delta^{\mathrm{B}}F$ Epistasis for fitness

$$\varepsilon^{\mathrm{AB}}F = \boldsymbol{C}\cdot\varepsilon^{\mathrm{AB}}\boldsymbol{x} + (\delta^{\mathrm{A}}\boldsymbol{x})^{\mathrm{T}}\cdot\boldsymbol{H}\cdot\delta^{\mathrm{B}}\boldsymbol{x}$$

“Propagation of epistasis”

What is function F?

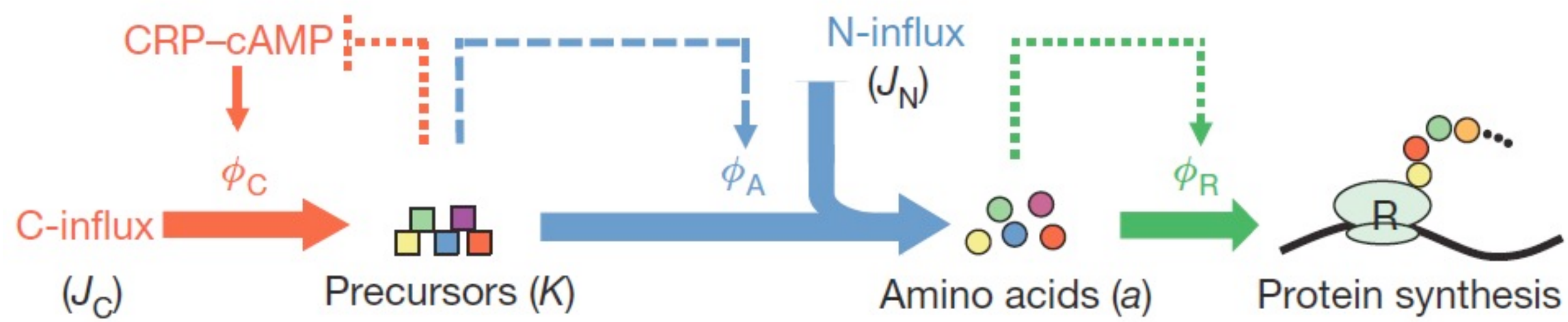


$F(x_1, x_2, \ldots, x_n)$ must be impossibly complex?

Bacterial growth laws offer a simple "coarse" description of metabolism

$$F = \frac{C}{1/x_{\mathrm{C}} + 1/x_{\mathrm{N}} + 1/x_{\mathrm{T}}}$$

Terry Hwa



T. Hwa's model predicts epistasis between mutations that affect different cellular processes

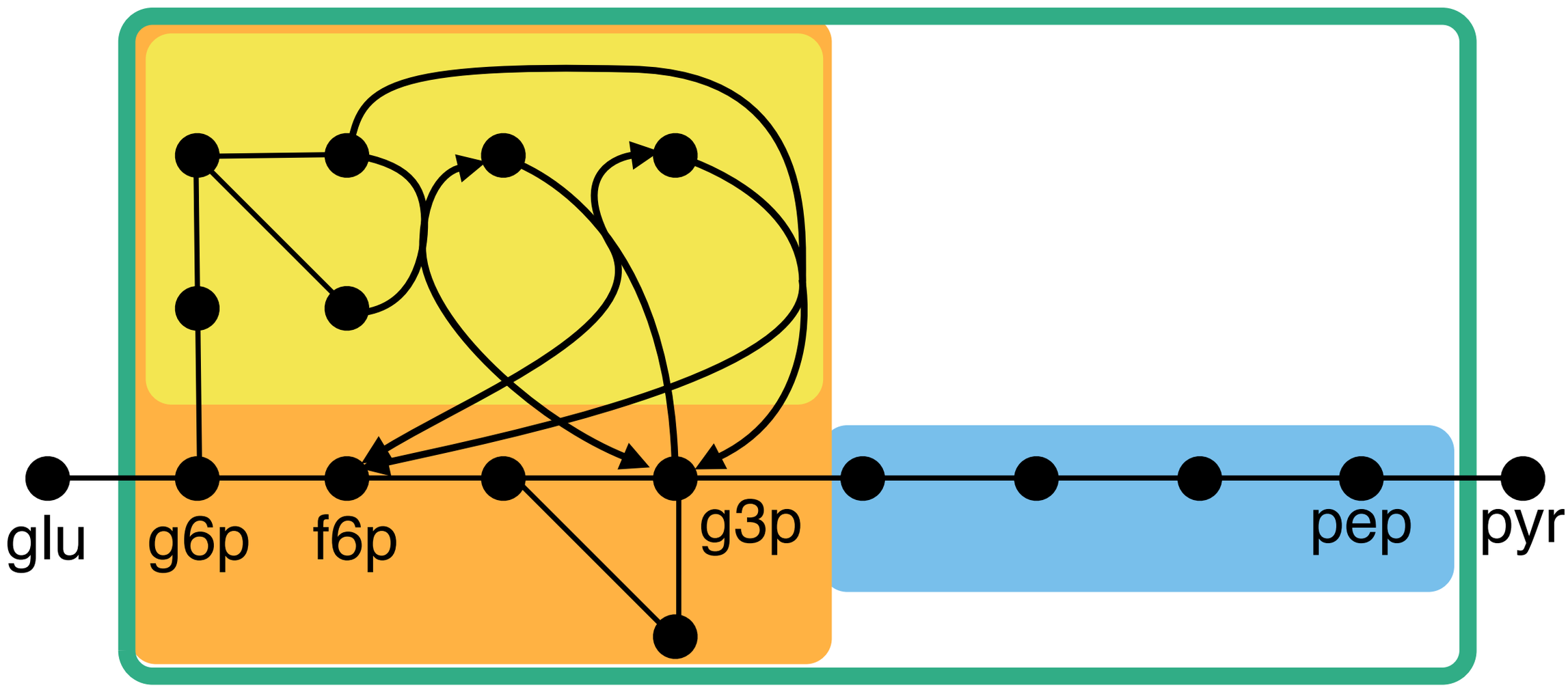
$$F = \frac{C}{1/x_{\mathrm{C}} + 1/x_{\mathrm{N}} + 1/x_{\mathrm{T}}}$$

Mutation A reduces only x_{C}
Mutation B reduces only x_{N}

$$\varepsilon^{\mathrm{AB}} F > 0$$

What about epistasis between mutations that affect the same process?

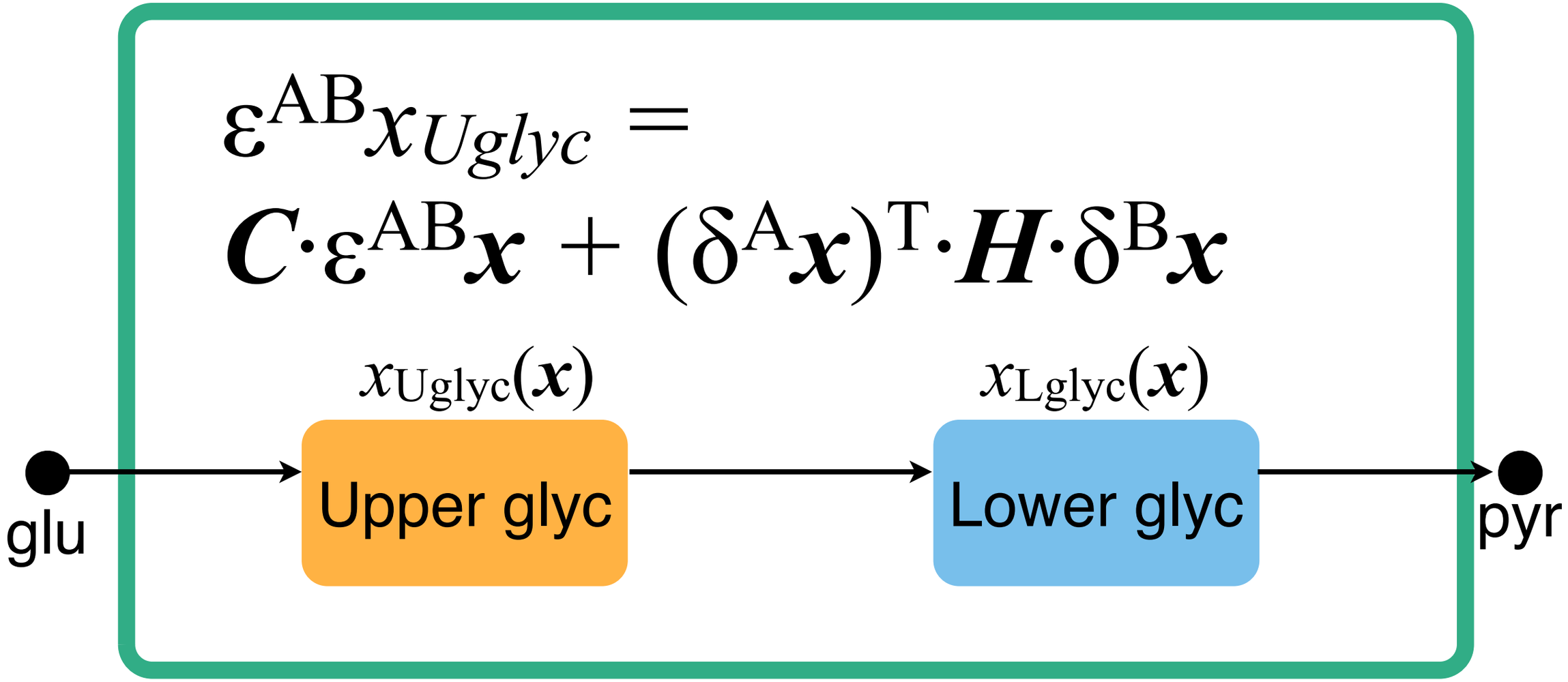
1. Break down metabolic networks into pathways
2. Characterize pathways by effective parameters
3. Propagate epistasis



What about epistasis between mutations that affect the same process?

1. Break down metabolic networks into pathways
2. Characterize pathways by effective parameters
3. Propagate epistasis

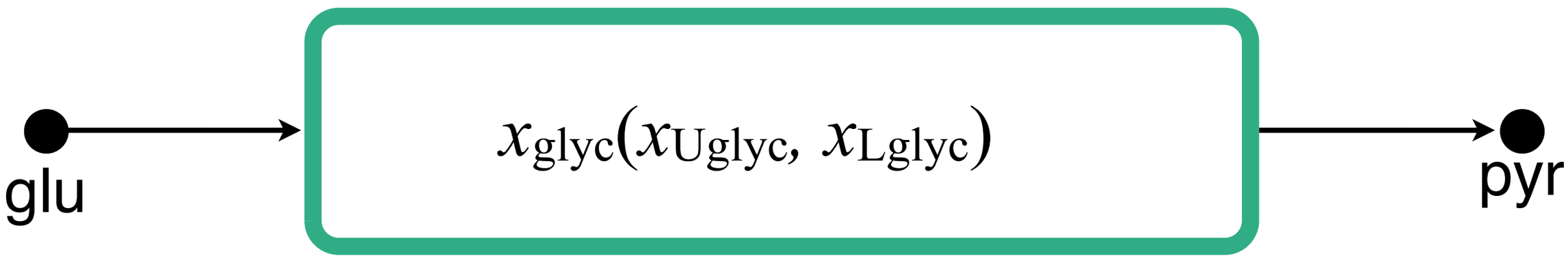
$$\varepsilon^{AB}x_{Uglyc} =$$
$$\boldsymbol{C}\cdot\varepsilon^{AB}\boldsymbol{x} + (\delta^{A}\boldsymbol{x})^{T}\cdot\boldsymbol{H}\cdot\delta^{B}\boldsymbol{x}$$



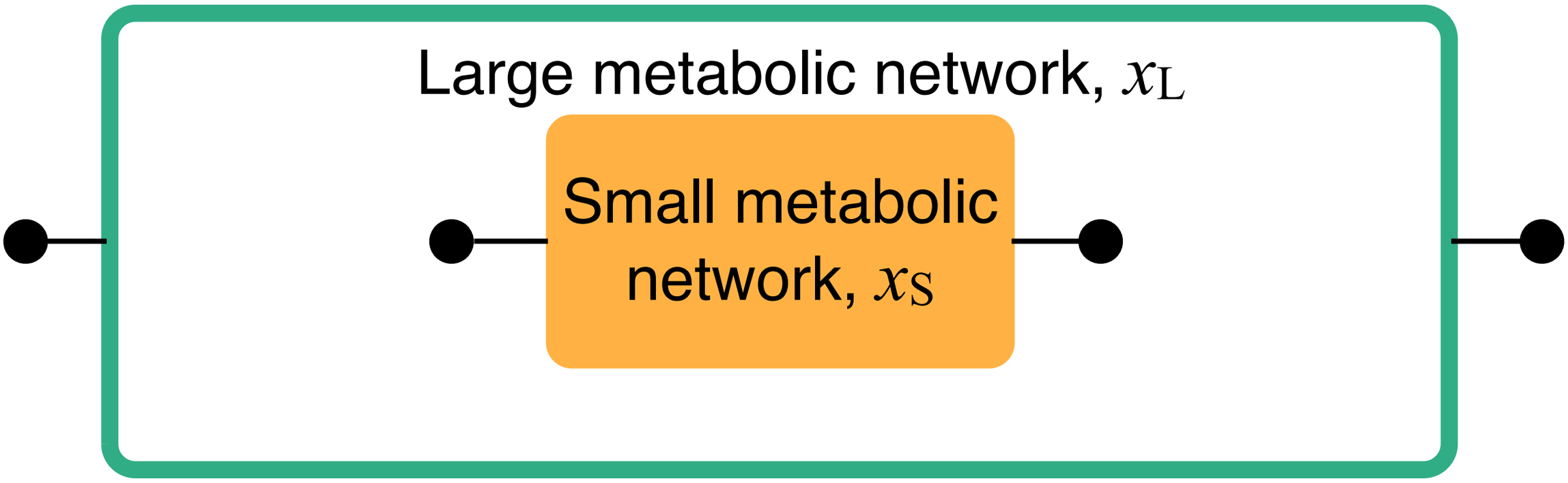
What about epistasis between mutations that affect the same process?

1. Break down metabolic networks into pathways
2. Characterize pathways by effective parameters
3. Propagate epistasis

$$\varepsilon^{\mathrm{AB}}x_{glyc} = \boldsymbol{C}\cdot\varepsilon^{\mathrm{AB}}\boldsymbol{x} + (\delta^{\mathrm{A}}\boldsymbol{x})^{\mathrm{T}}\cdot\boldsymbol{H}\cdot\delta^{\mathrm{B}}\boldsymbol{x}$$



Some properties of propagation of epistasis can be derived for metabolic networks with unsaturated reactions

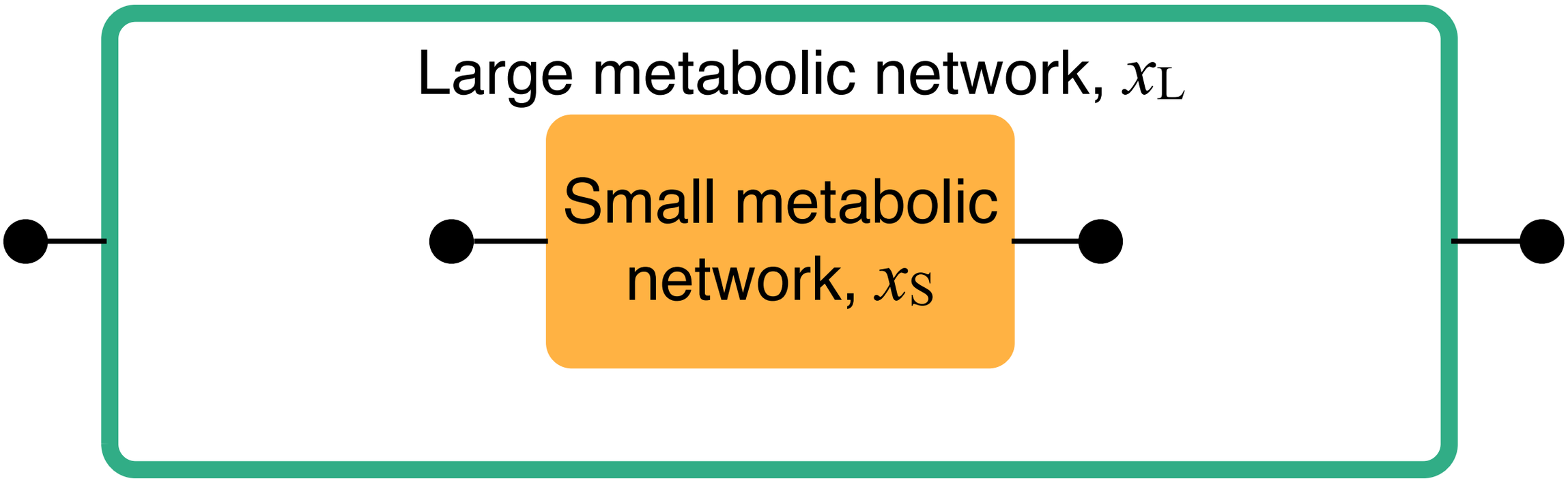


Mutations A and B reduce only x_{S} with $\varepsilon^{\mathrm{AB}} x_{\mathrm{S}} \neq 0$

$$\varepsilon^{\mathrm{AB}} x_{\mathrm{L}} = C \cdot \varepsilon^{\mathrm{AB}} x_{\mathrm{S}} + \delta^{\mathrm{A}} x_{\mathrm{S}} \cdot H \cdot \delta^{\mathrm{B}} x_{\mathrm{S}}$$

$$C > 0 \qquad H < 0$$

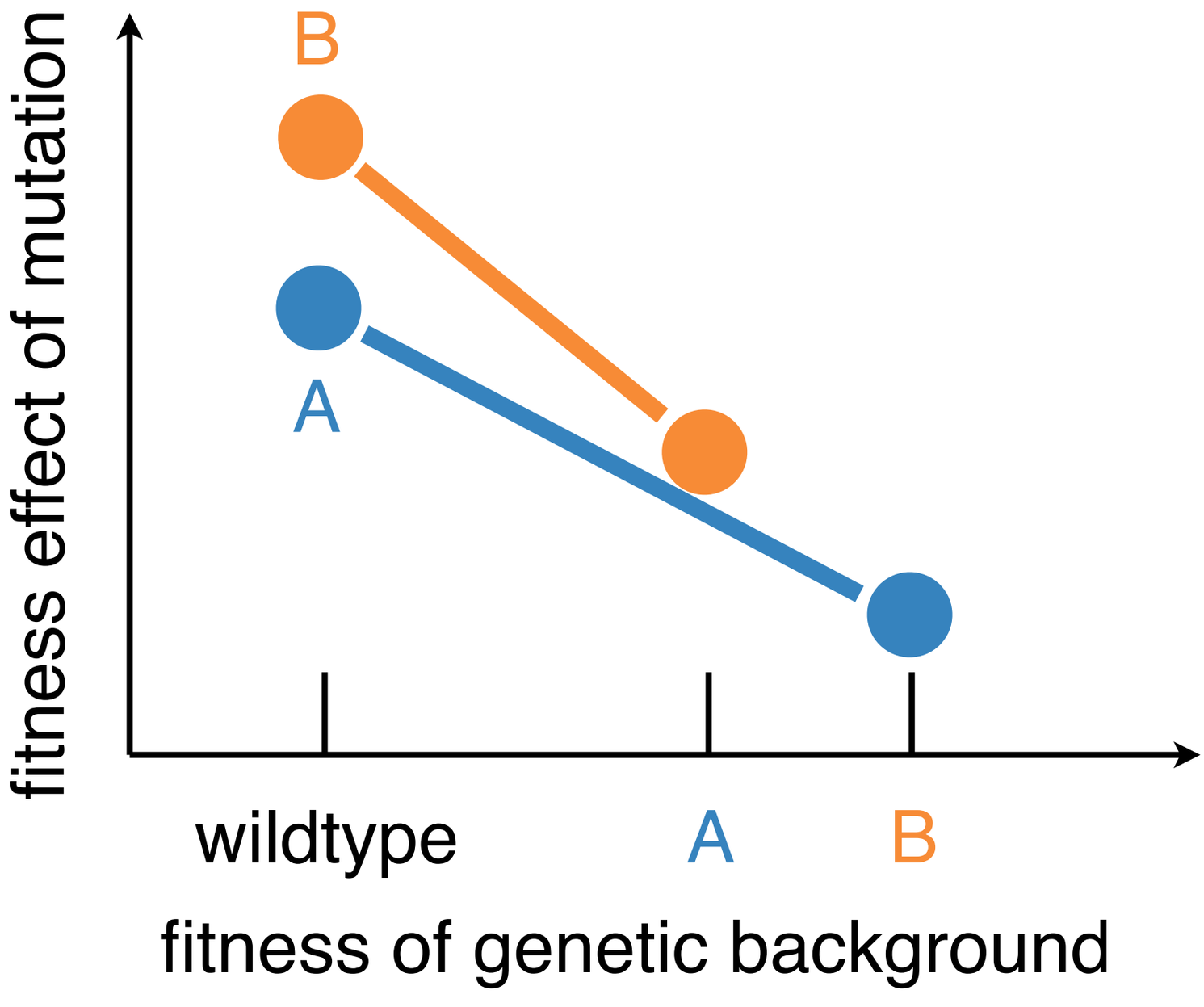
Some properties of propagation of epistasis can be derived for metabolic networks with unsaturated reactions



Mutations A and B reduce only x_{S} with $\varepsilon^{\mathrm{AB}} x_{\mathrm{S}} \neq 0$

Negative epistasis tends to accumulate but positive epistasis does not

Can propagation of epistasis explain prevalence of negative epistasis for fitness in evolution experiments?



Number of mutations indistinguishable among Founders