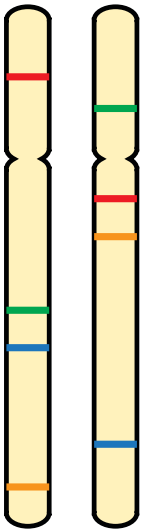


Chance and determinism in experimental evolution

Greg Lang, Lehigh University
www.glanglab.com

Princeton Plasma Physics Laboratory
26 February 2014

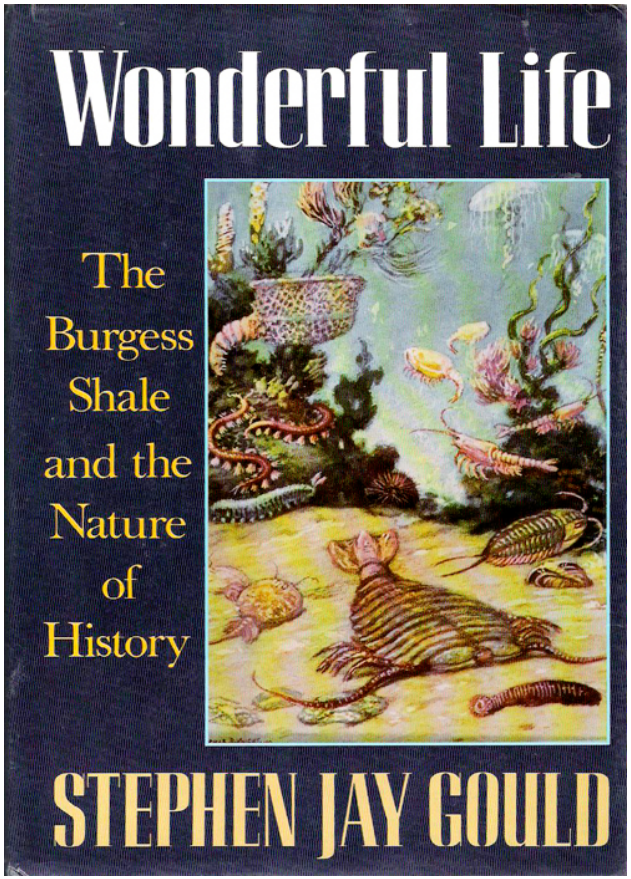
A canonical example of parallelism in evolution

Surface

Cave population “A”

Cave population “B”

Gould's Gedankenexperiment



"our origin is the product of massive historical contingency, and we would probably never rise again even if life's tape could be replayed a thousand times."

Stephen Jay Gould, *Wonderful Life*

"evolutionary routes are many, but the destinations are limited."

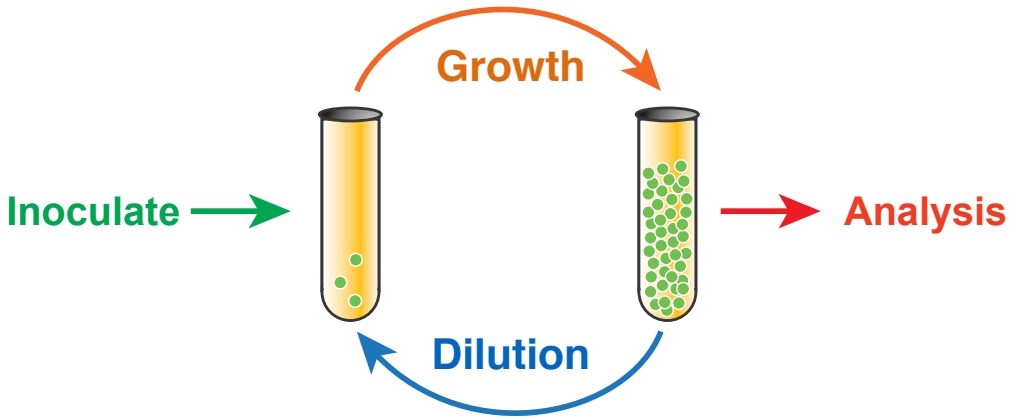
Simon Conway Morris, *Life's Solution*

A laboratory system for studying evolution

Yeast as a model system

- Rapid growth
- Haploid or diploid
- Sexual or asexual
- Small genome size
- Model system

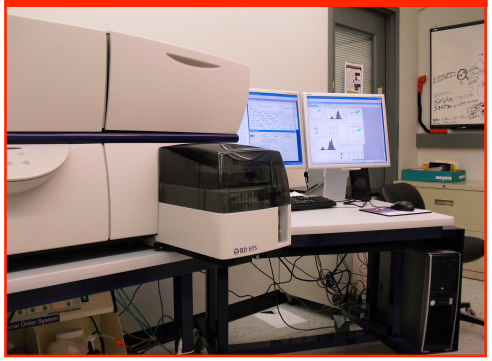
Yeast



Automation

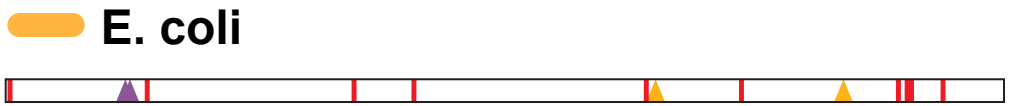
"Frozen fossil record"

Flow Cytometry

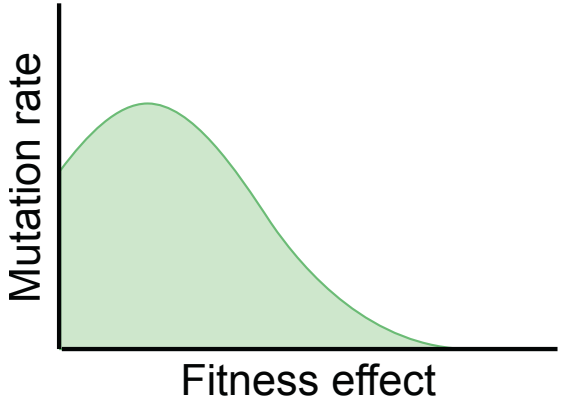


Sequencing

How much parallelism should we expect?

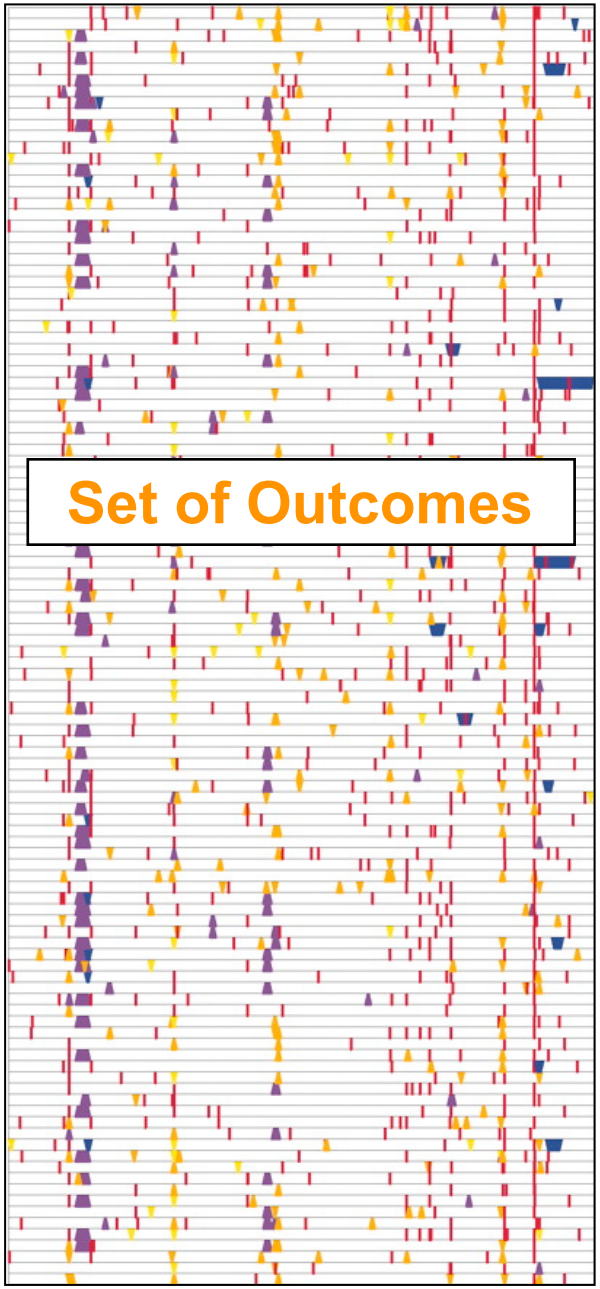


Set of Possibilities

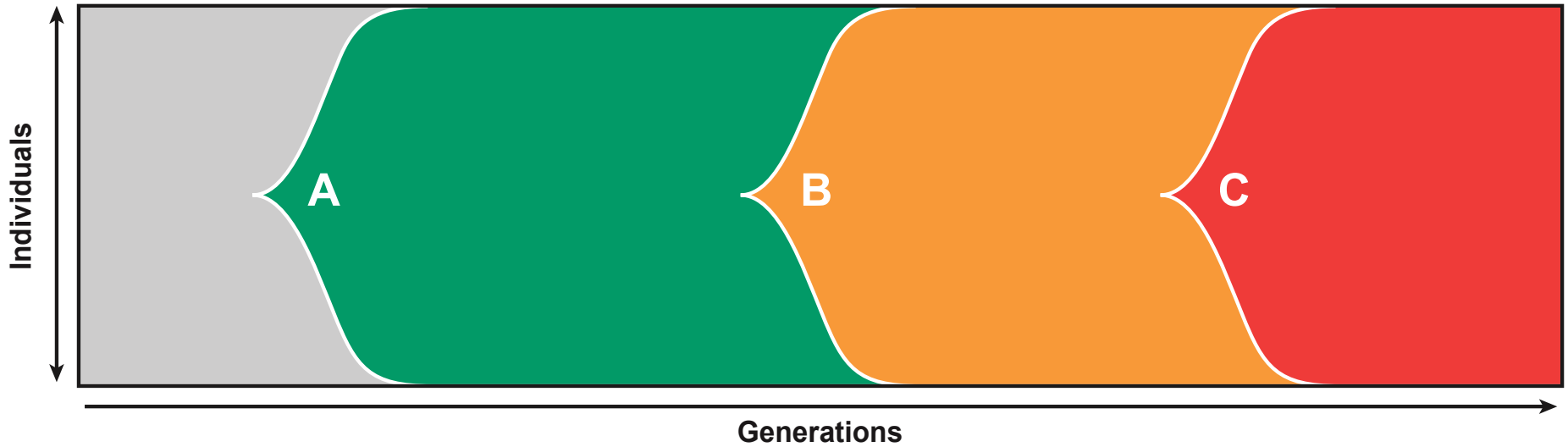


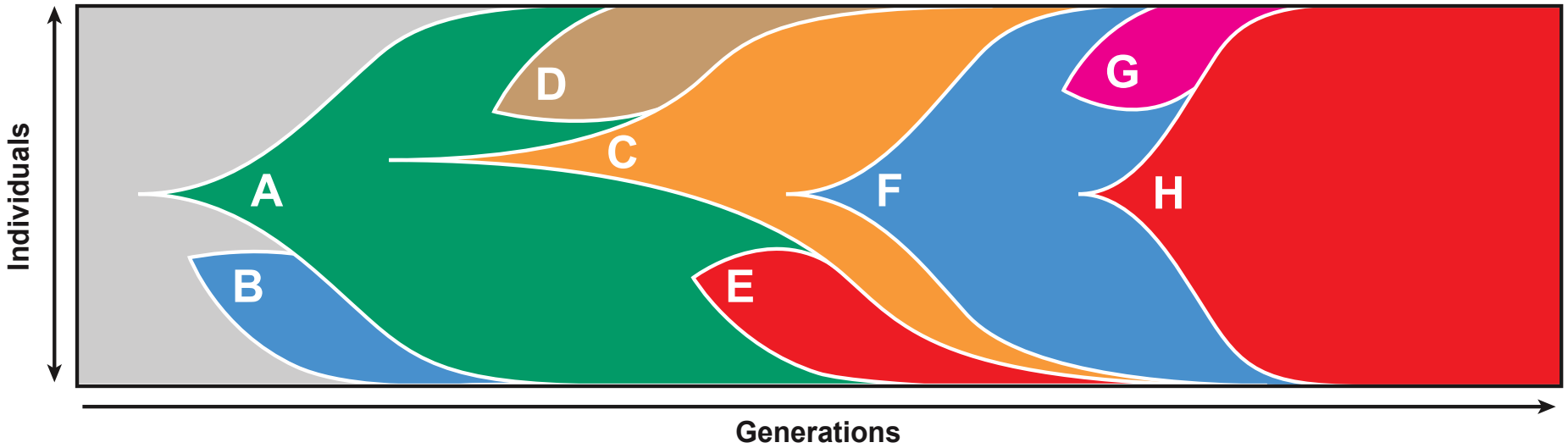
Evolutionary Dynamics

Mutation, Drift, Selection,
Genetic Hitchhiking

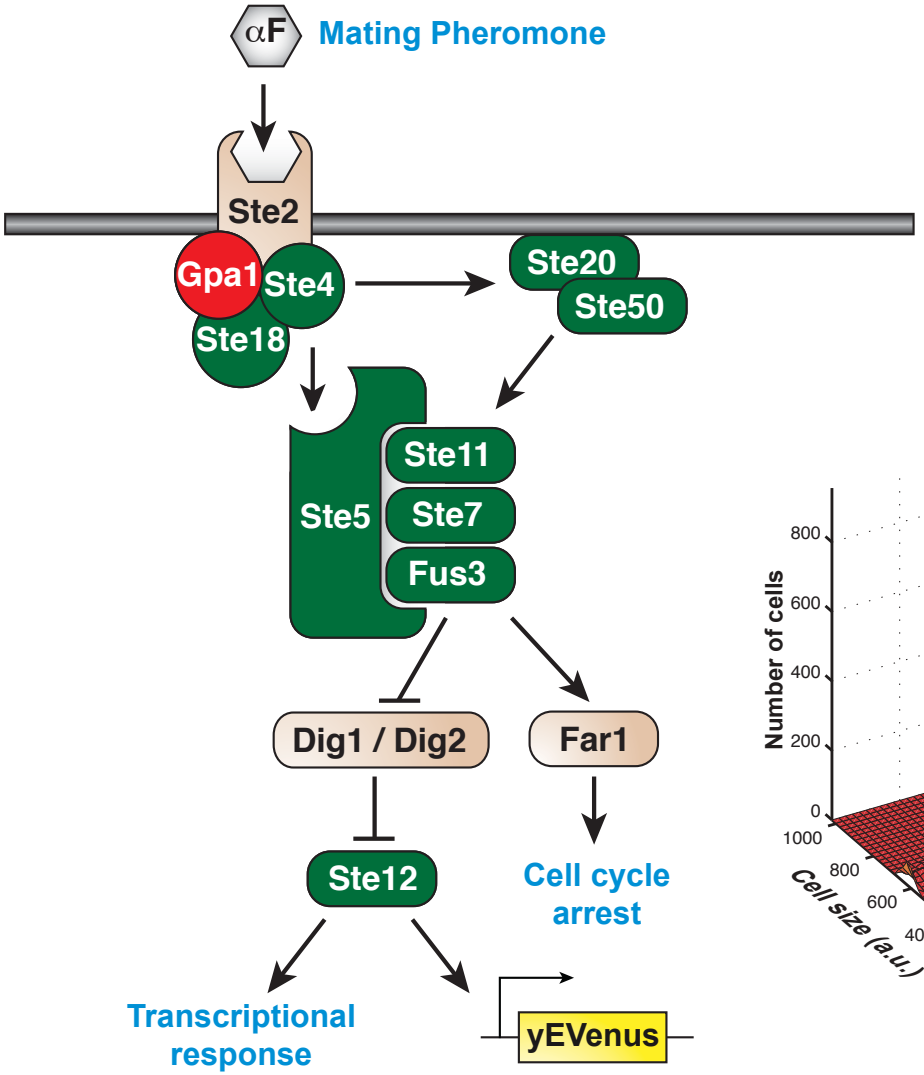


Chance in adaptation



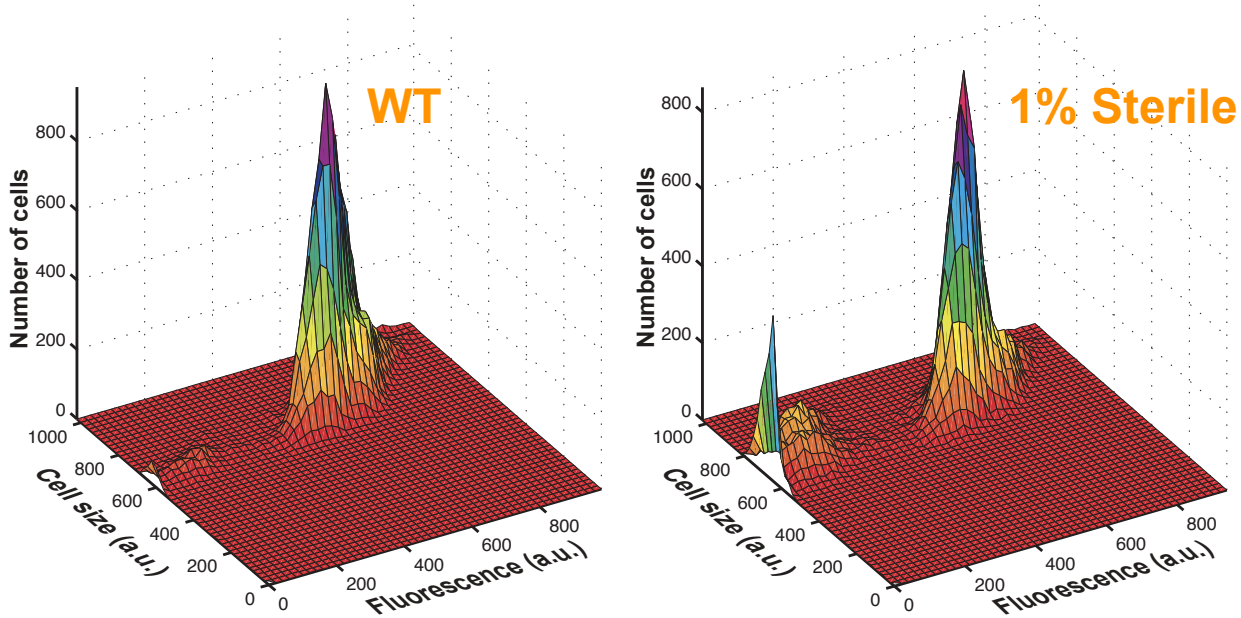


Sterility provides a fitness advantage



Selective Advantage of Sterility

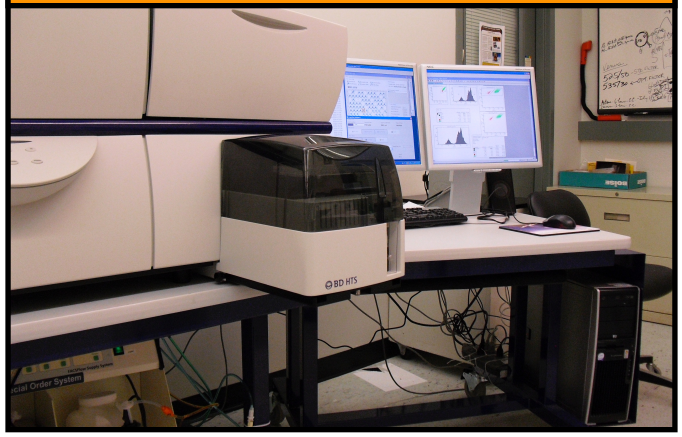
- "**R**" allele of *GPA1* = 0.6%
- "**B**" allele of *GPA1* = 1.5%



Tracking sterile mutations in evolving populations

Robotics for transfers

96-well flow cytometry



Experimental setup

- Two strains
  - 0.6%
  - 1.5%
- Two bottlenecks
  - Small: $N \sim 10^4$
  - Large: $N \sim 10^6$
- 148 replicates each (592 total populations)

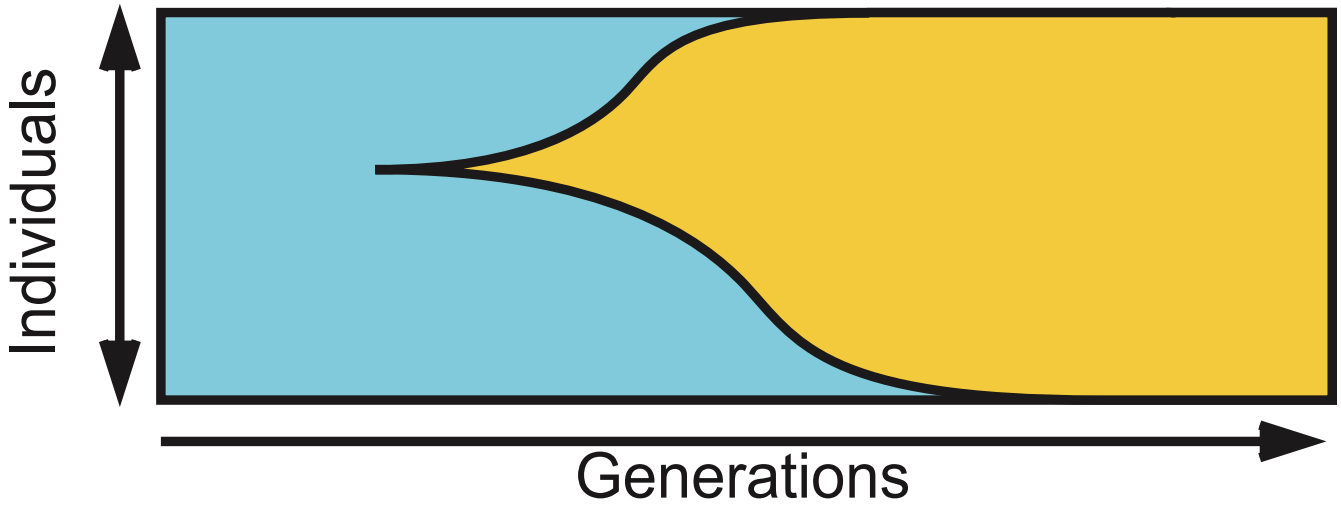
Experimental details

- Daily (or twice daily) transfers in YPD
- Flow cytometry every 40 generations
- 1000 generations

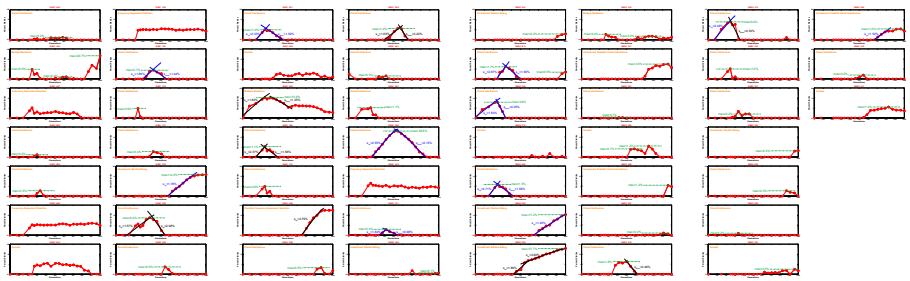
The Goal

Watch sterile strains arise and move through populations

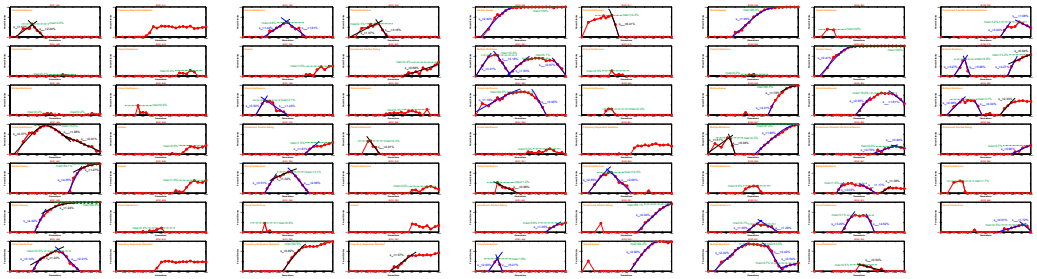
The simplest case



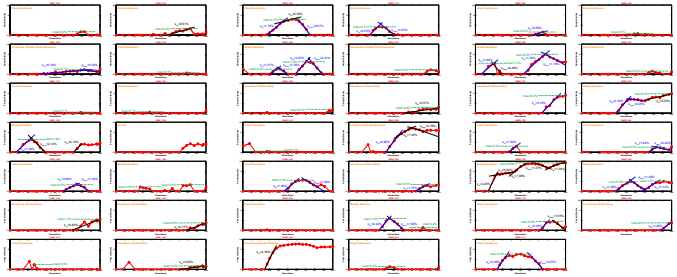
Trajectories of beneficial mutations



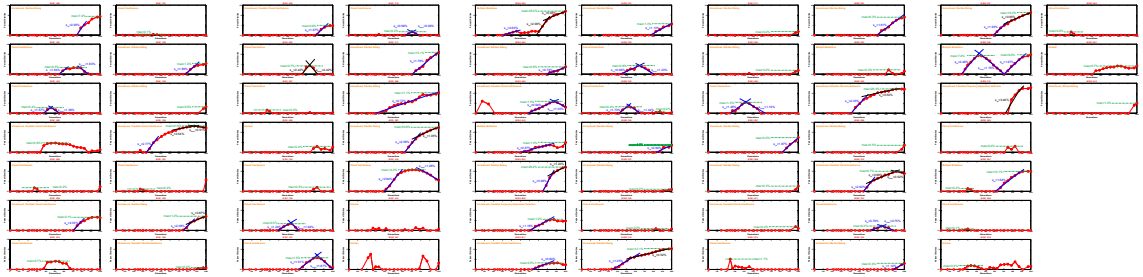
S_{ste} = 0.6%
Small N_e



S_{ste} = 1.5%
Small N_e

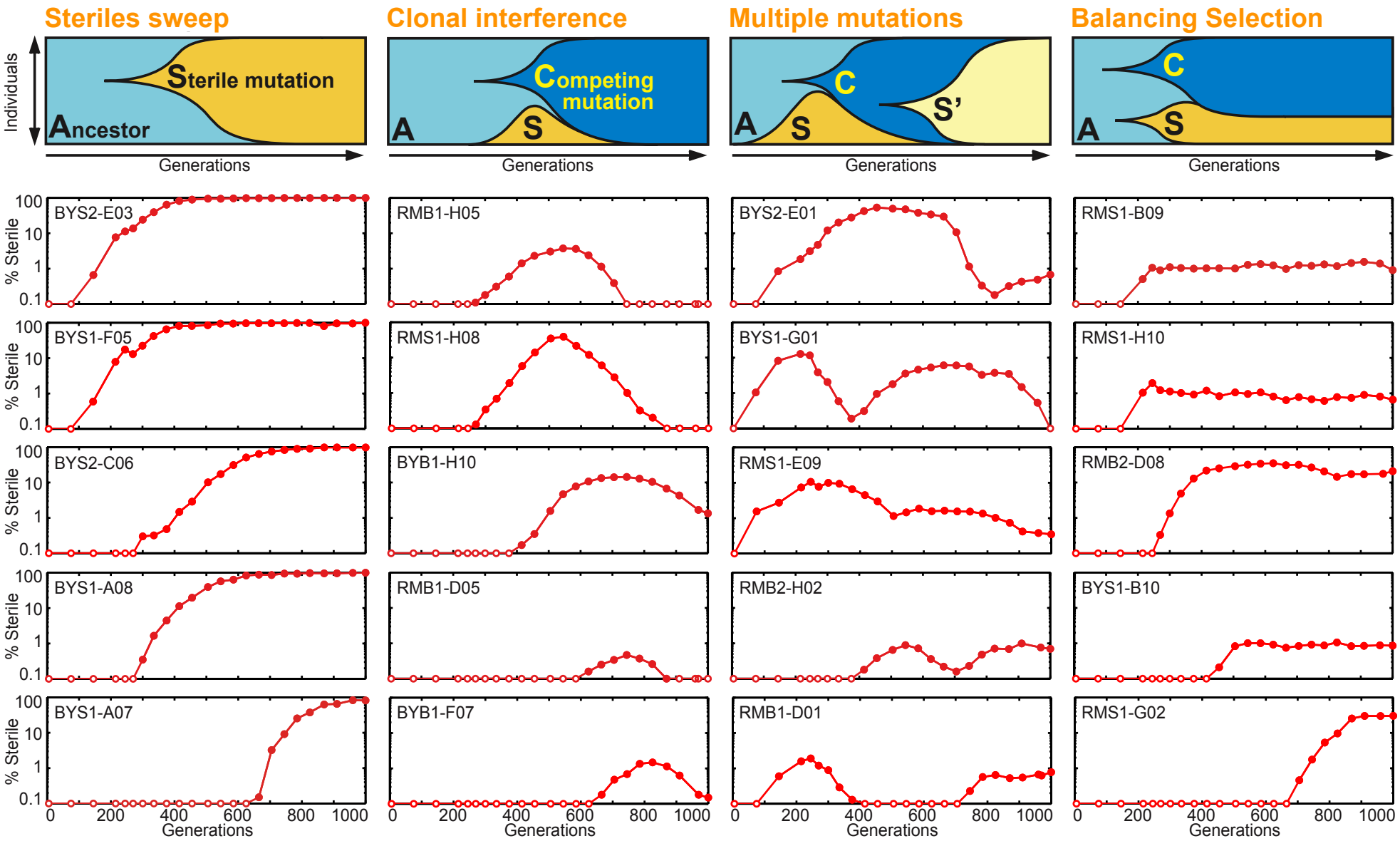


S_{ste} = 0.6%
Big N_e

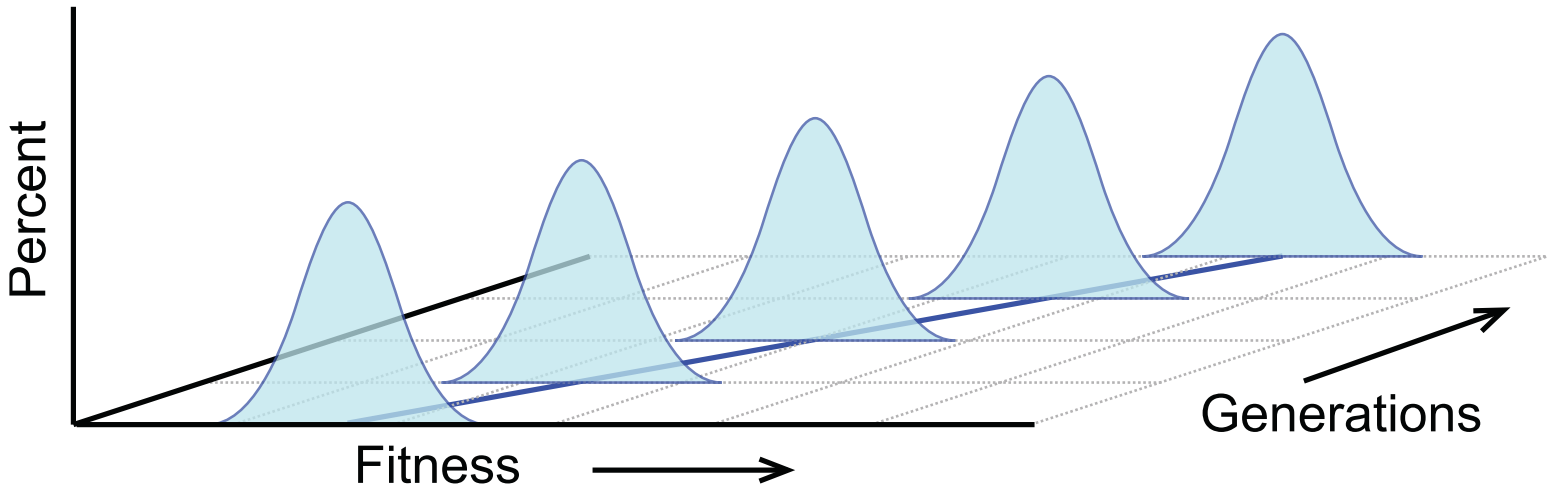


S_{ste} = 1.5%
Big N_e

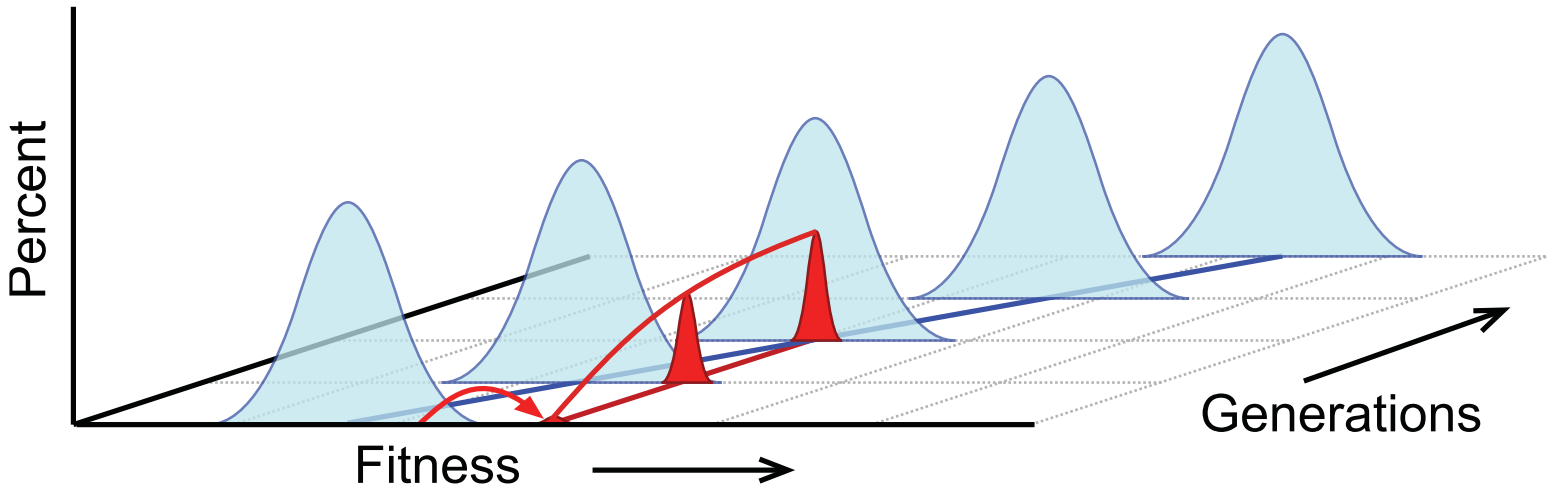
The fates of beneficial mutations



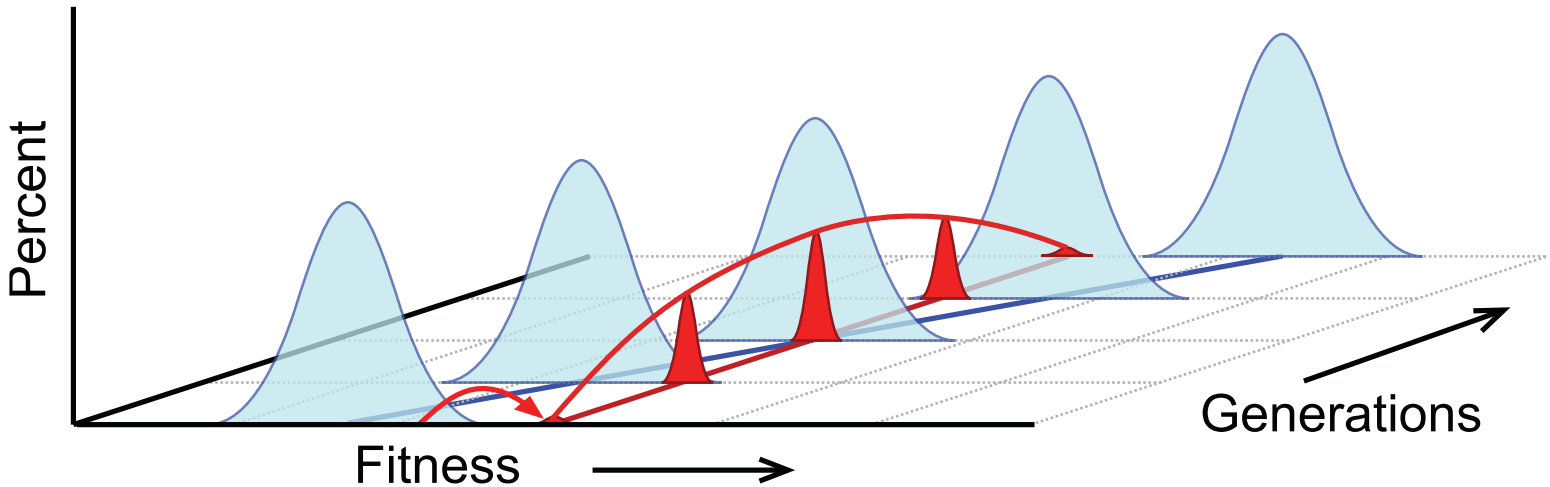
Genetic variation governs the fate of new mutations



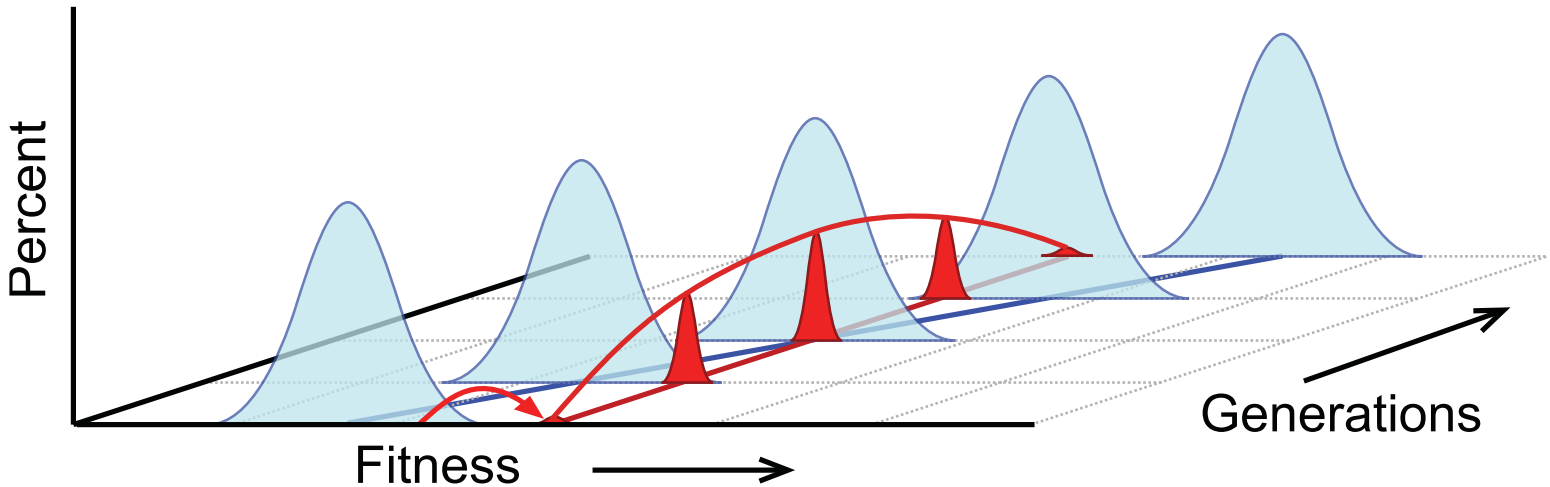
Genetic variation governs the fate of new mutations

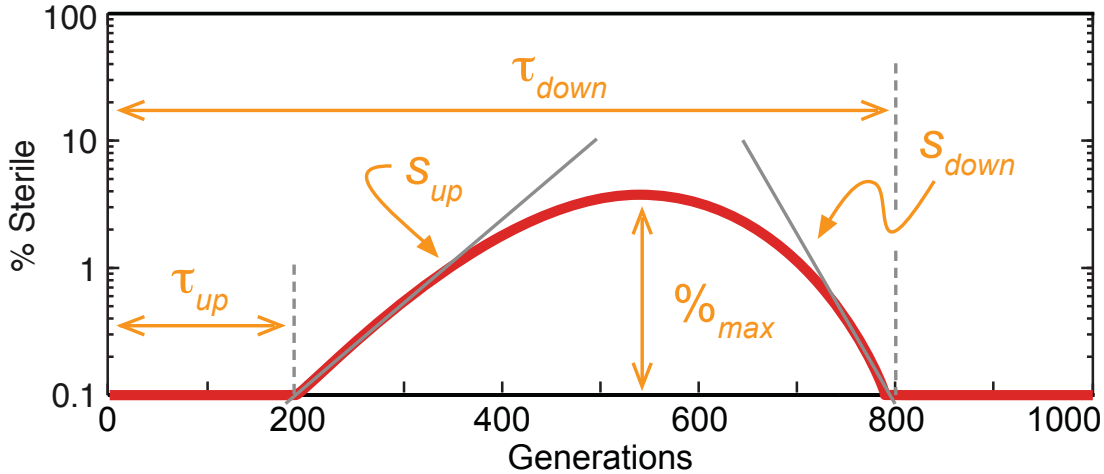


Genetic variation governs the fate of new mutations

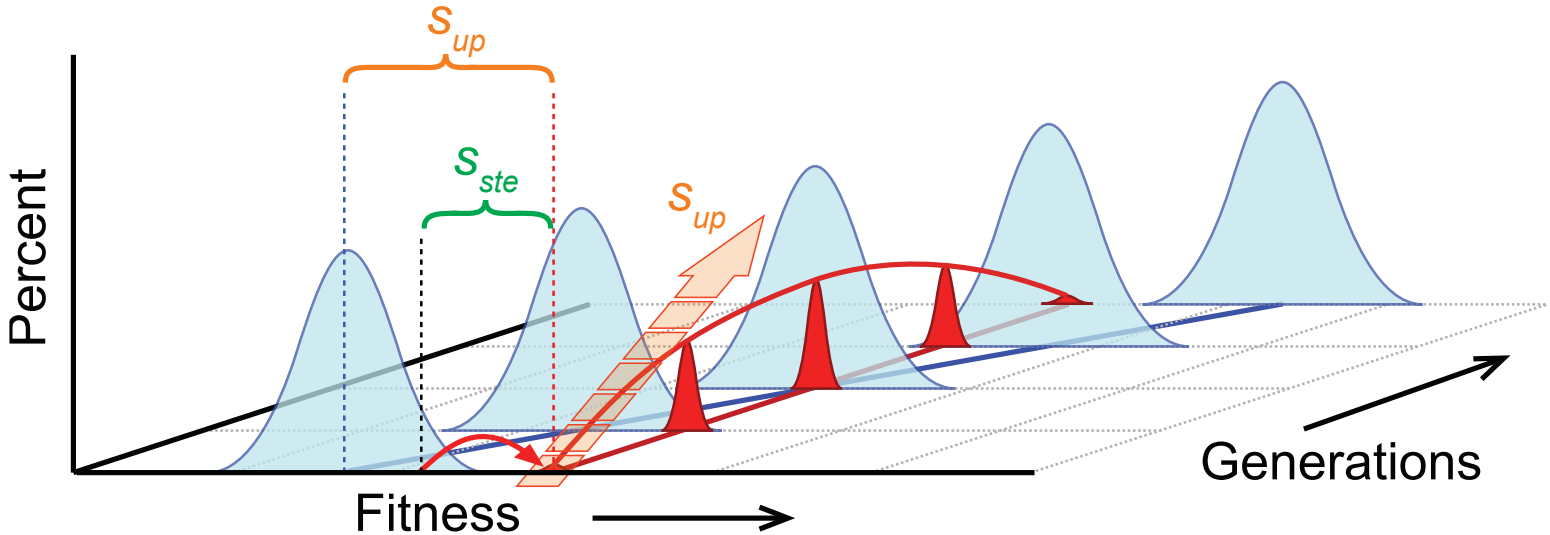


Genetic variation governs the fate of new mutations



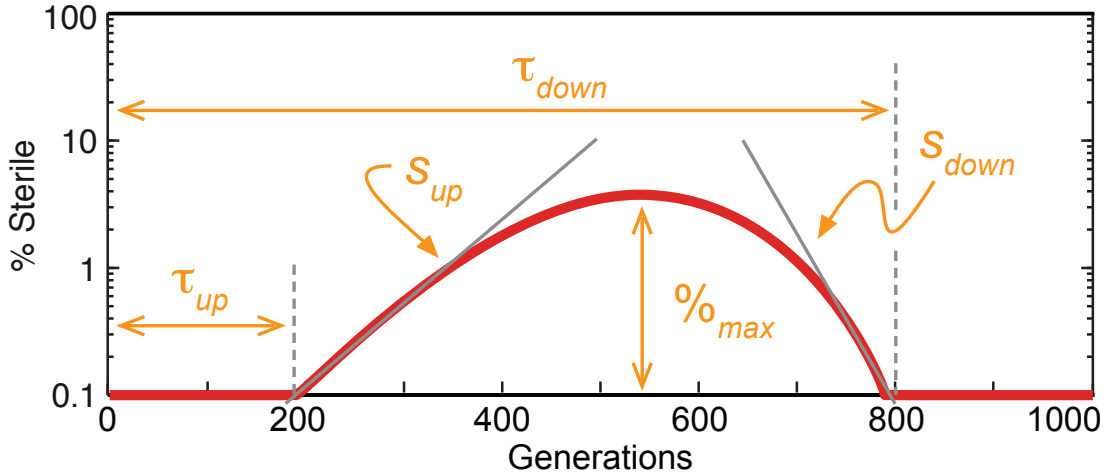


Genetic variation governs the fate of new mutations

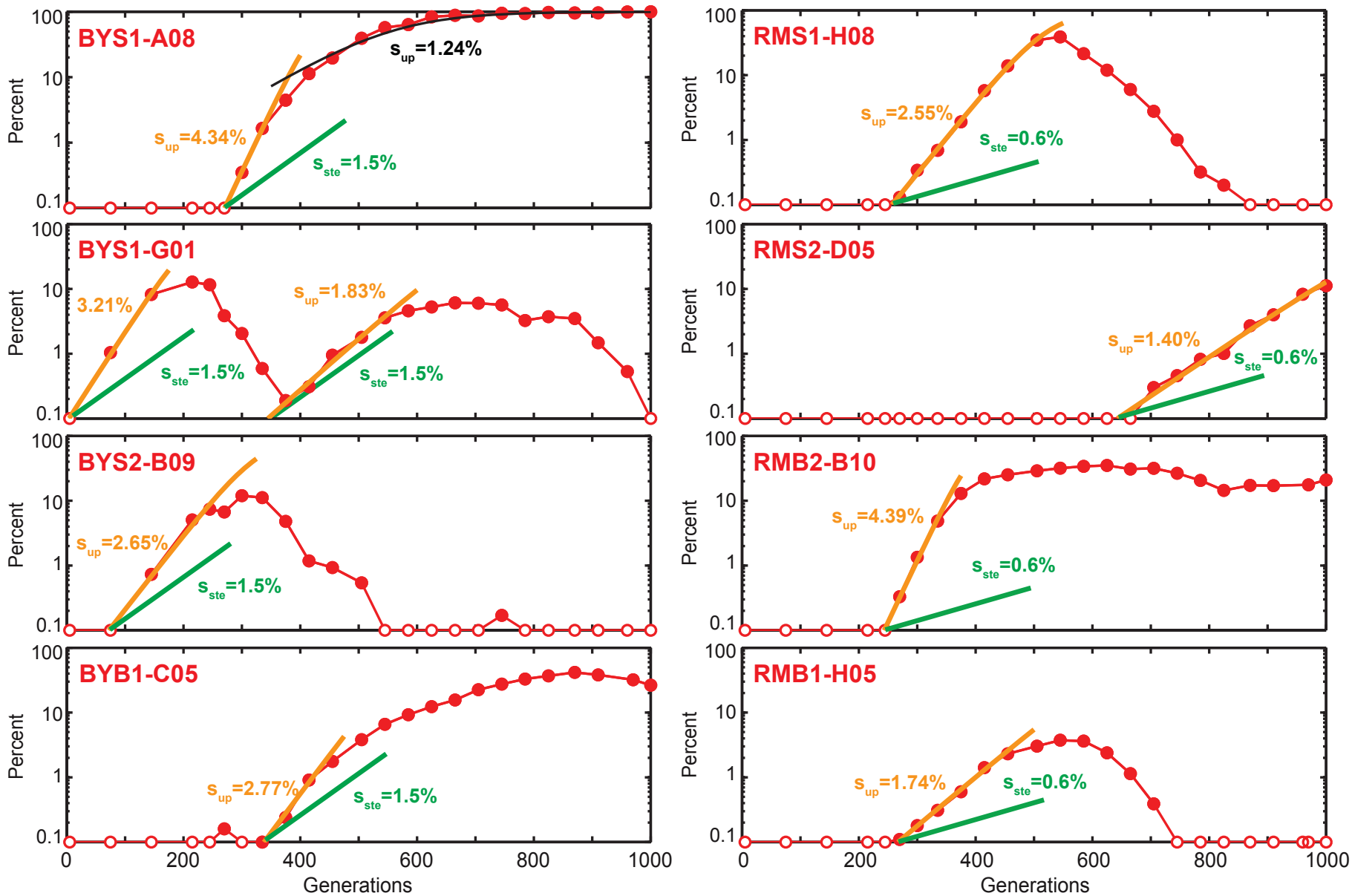


Background variation

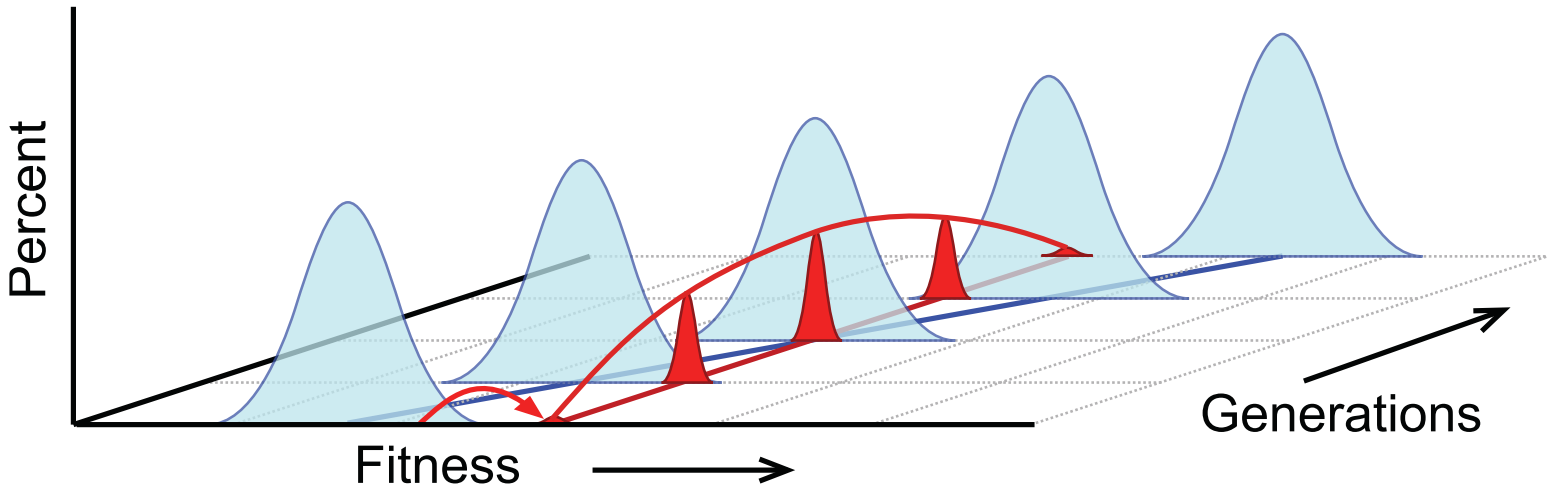
background fitness = $s_{up} - s_{sterile}$

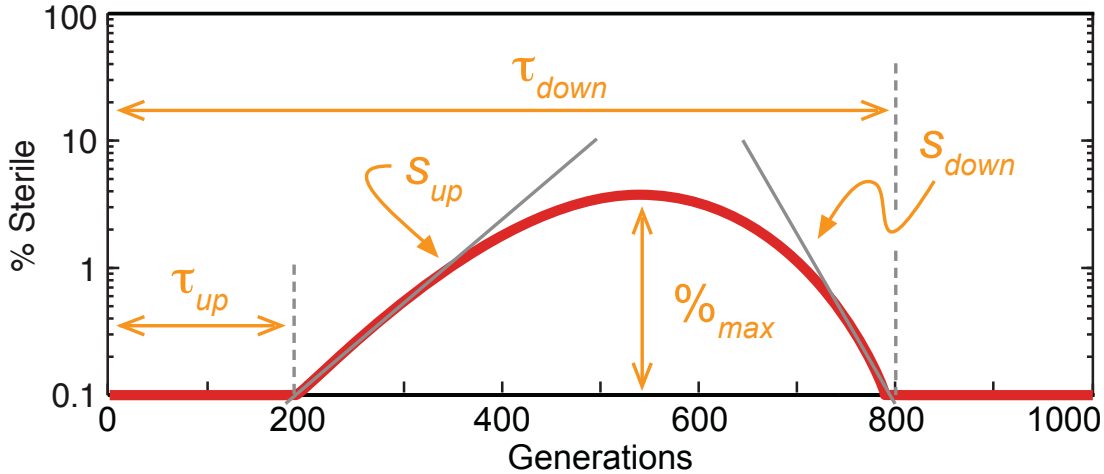


The most successful mutations are the luckiest

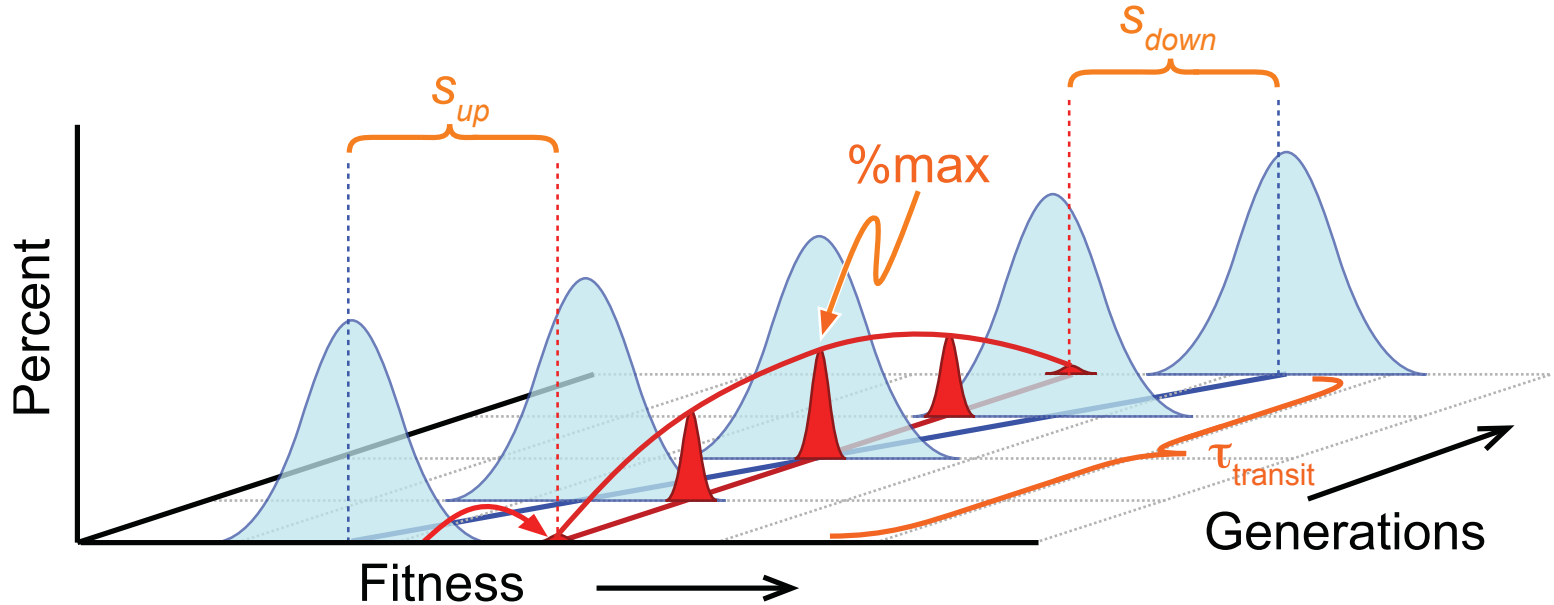


Inferring the speed of adaptation from trajectories



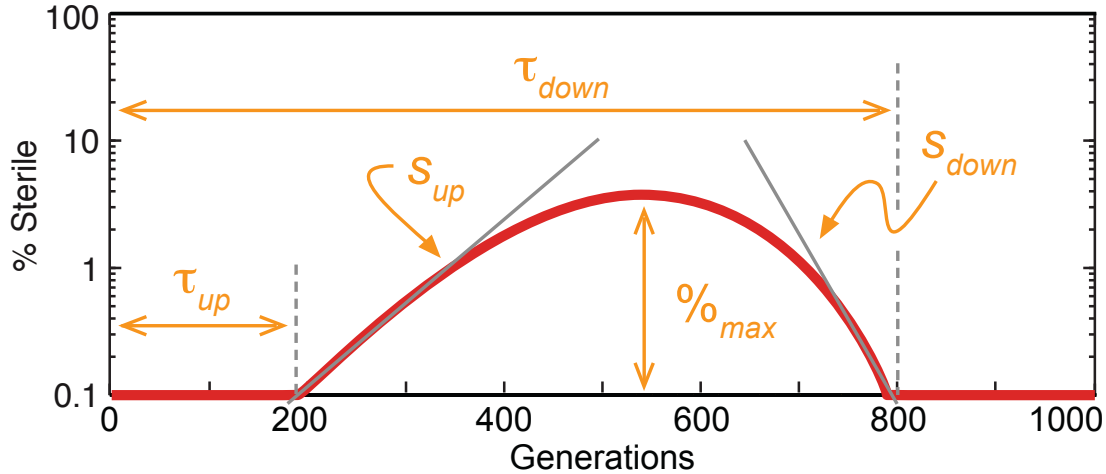


Inferring the speed of adaptation from trajectories

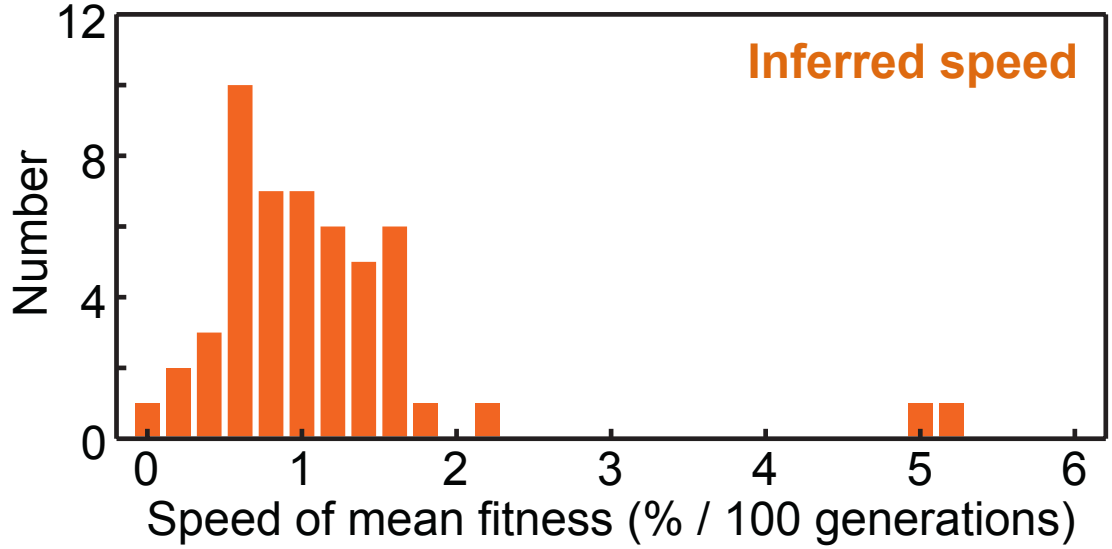


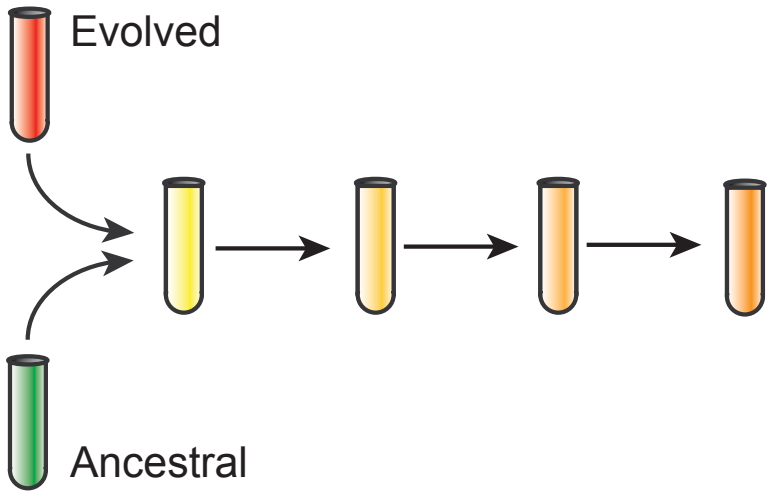
Speed of adaptation

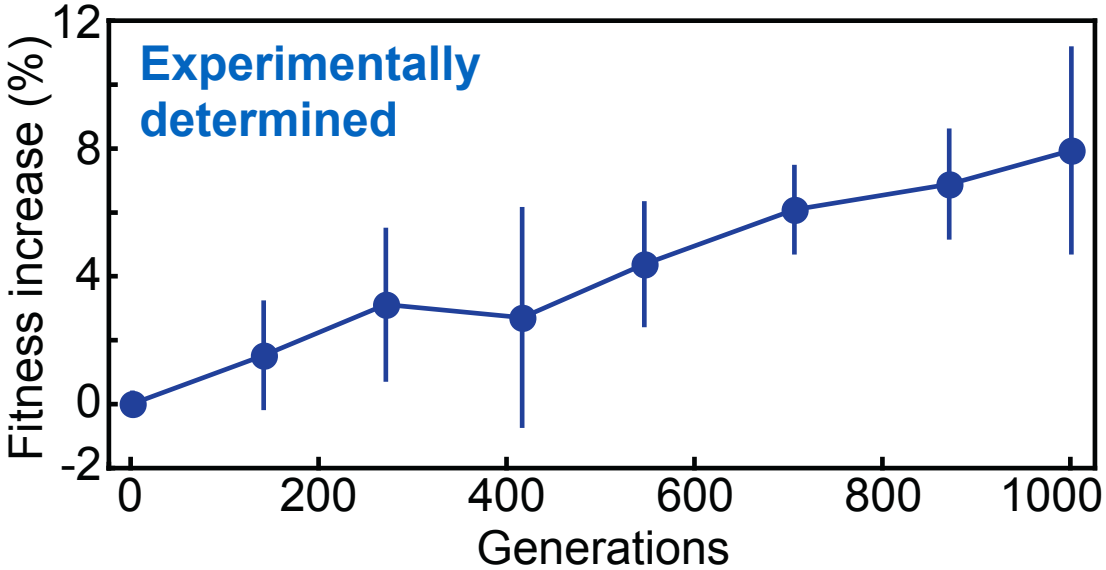
$$\text{Average speed} = \frac{s_{up} + s_{down}}{\tau_{transit}}$$



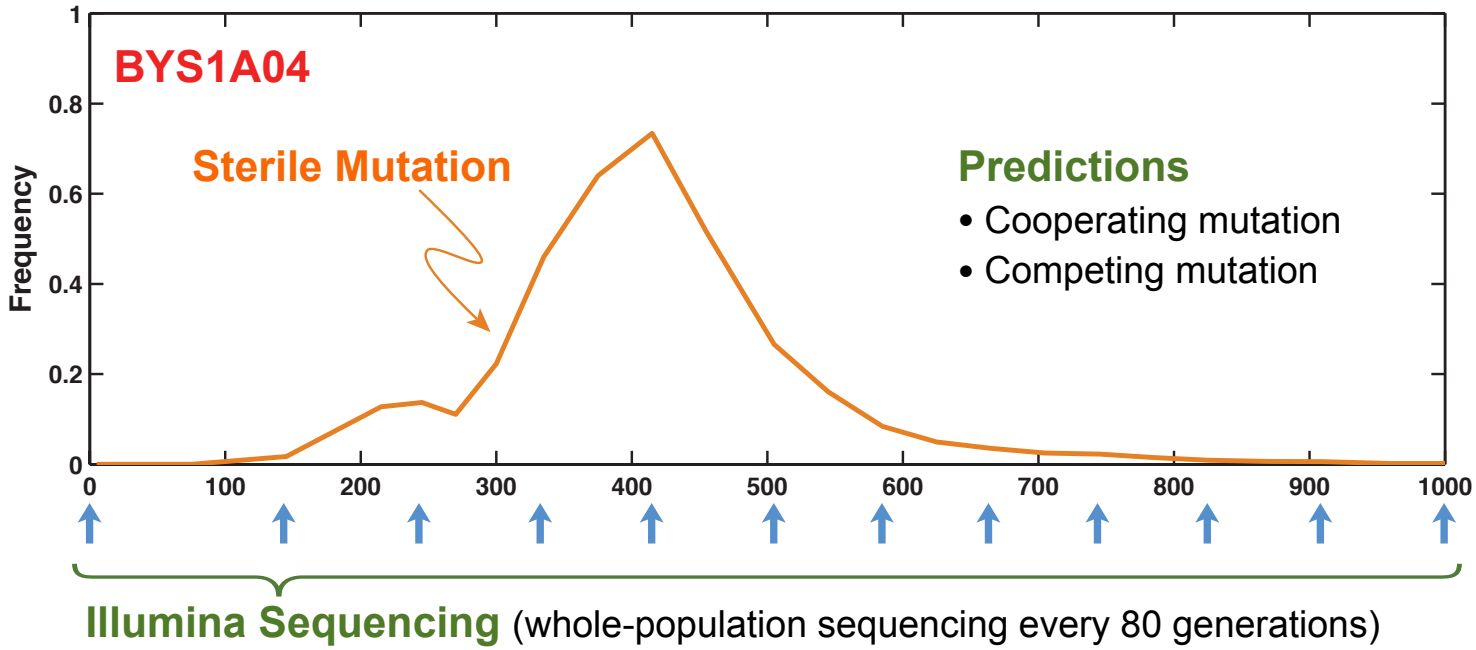
Inferred speed of adaptation is consistent with empirical data

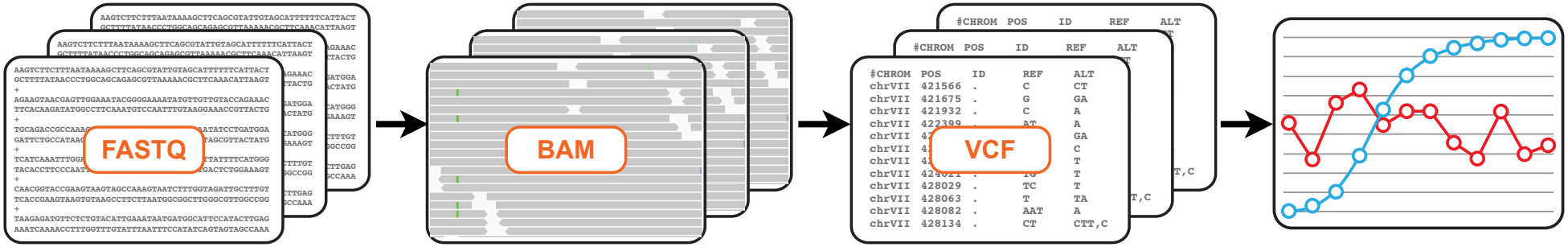




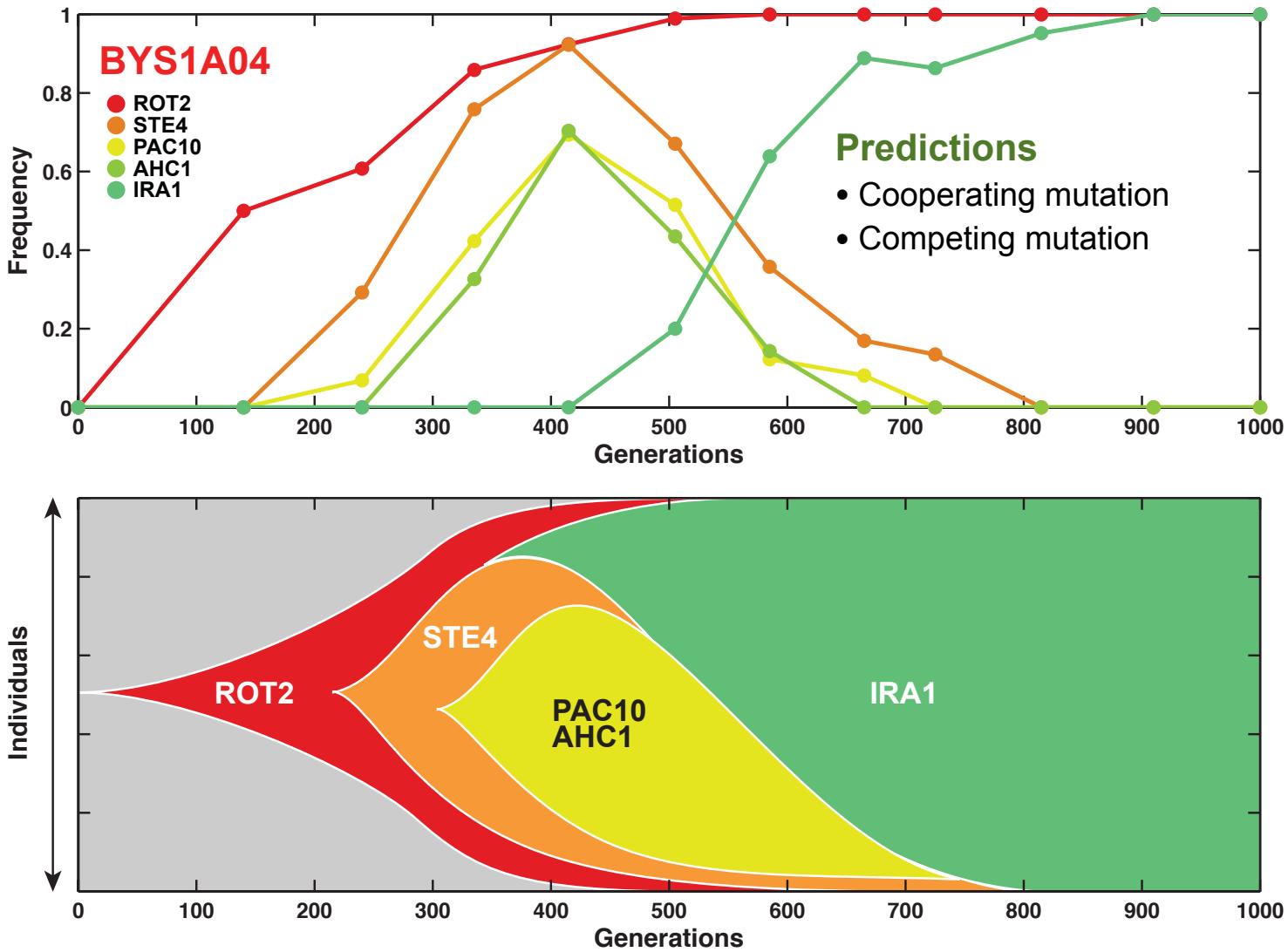


Sequencing to reveal dynamics

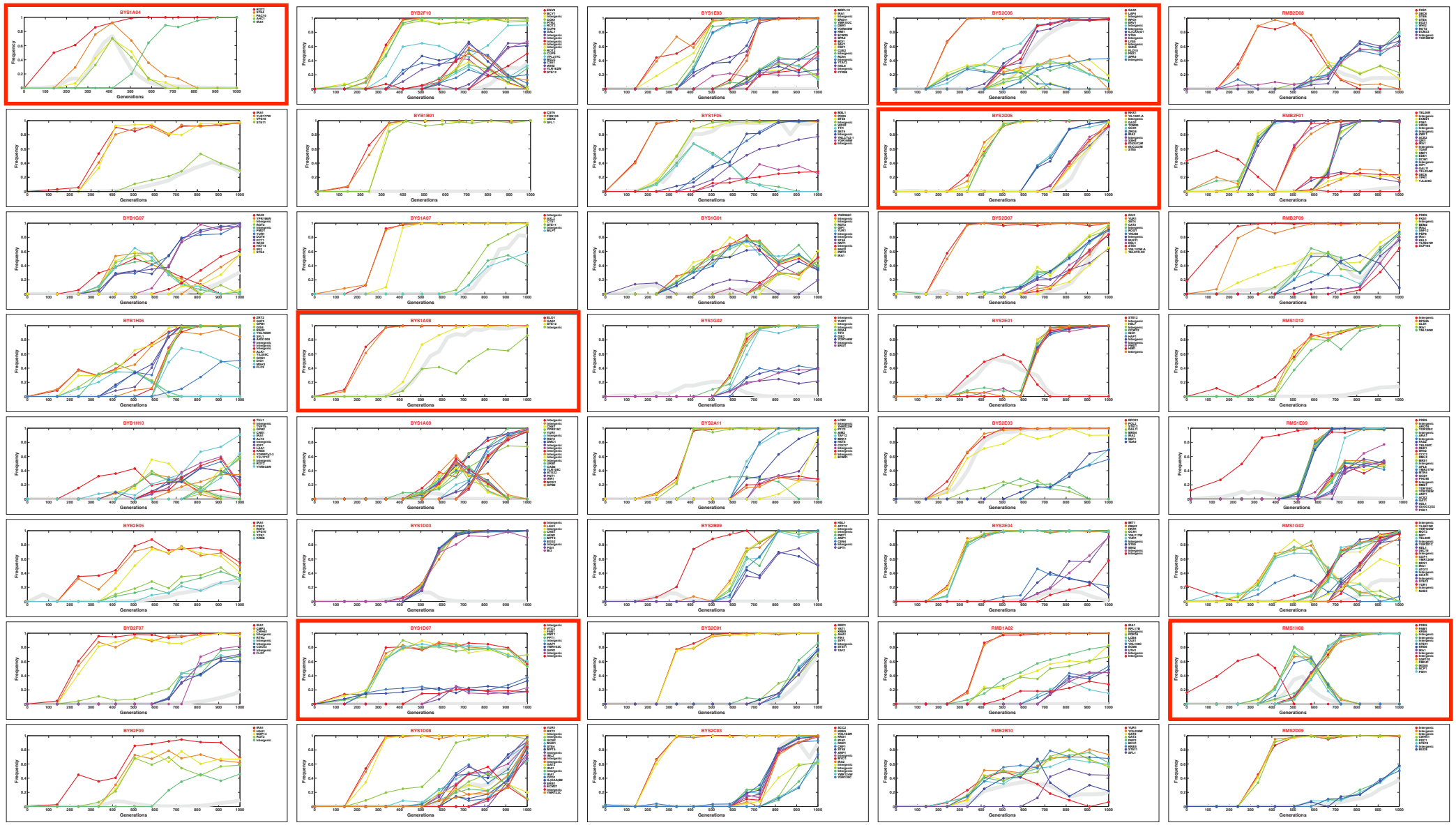




The molecular dynamics of adaptation

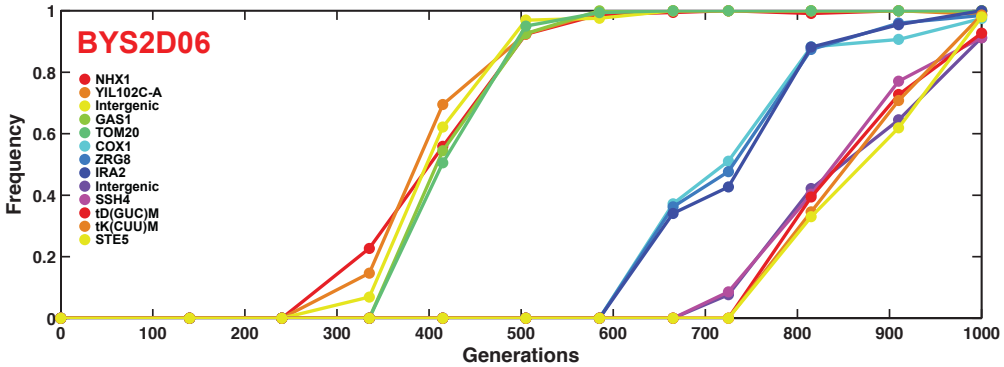


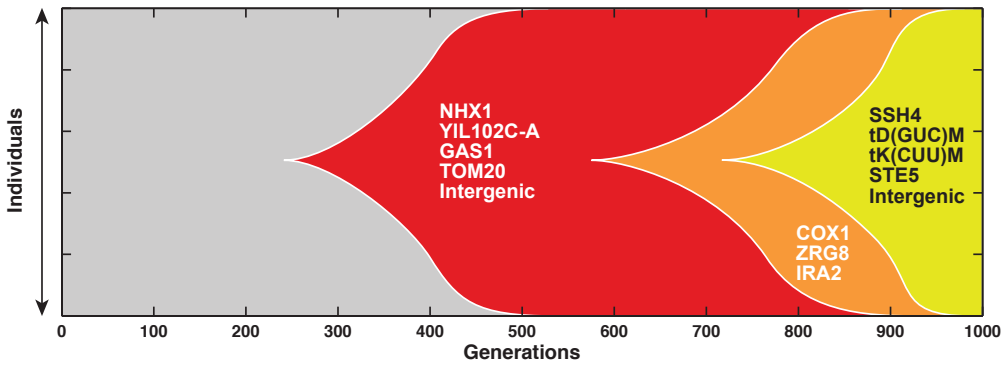
Identification of 1020 mutations in 40 populations



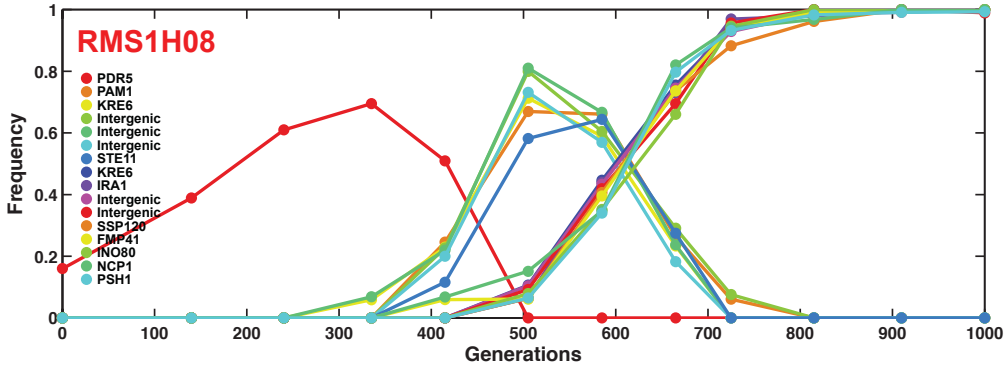
The molecular dynamics of adaptation

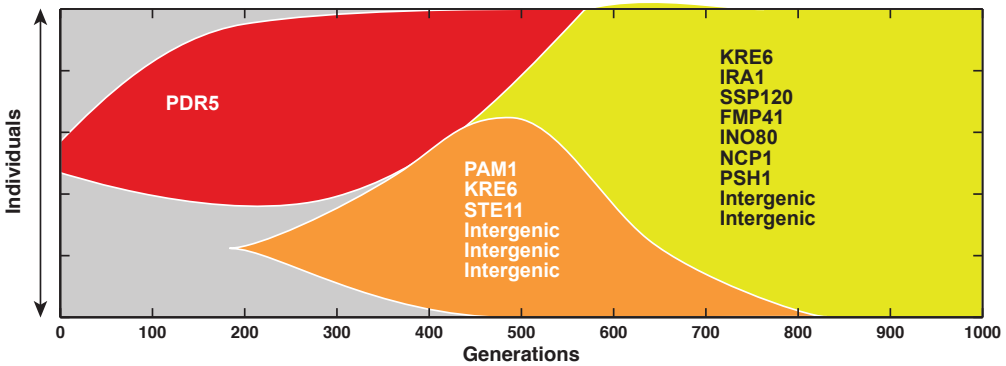
Mutational Cohorts & Multiphase Sweeps



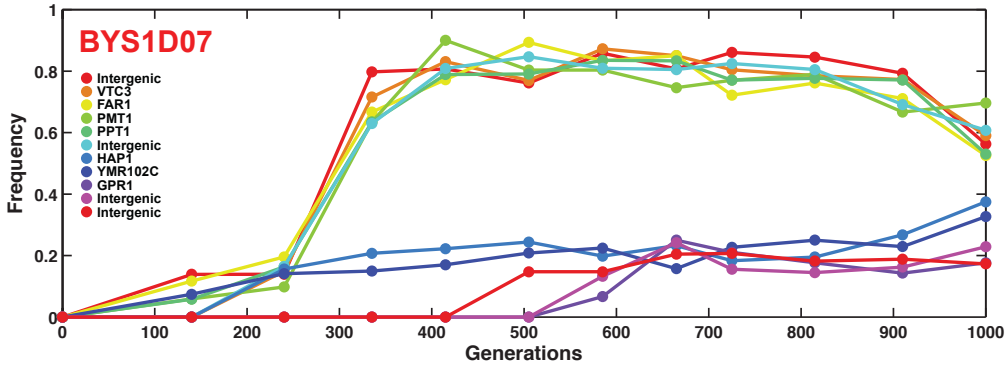


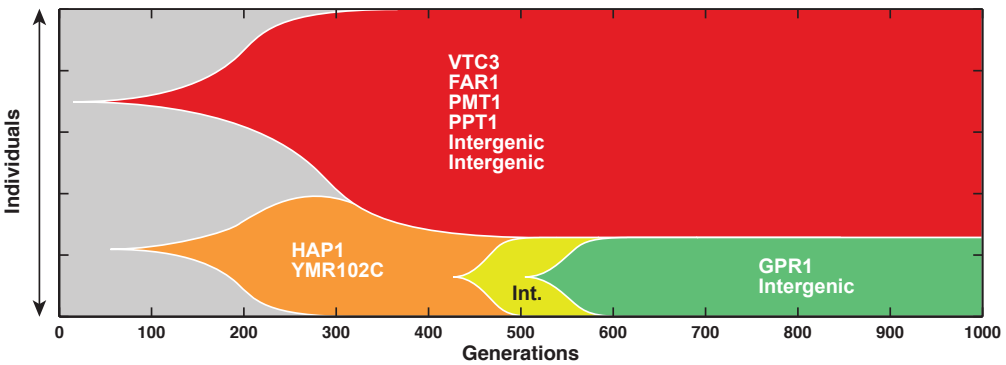
Recurrent Mutations



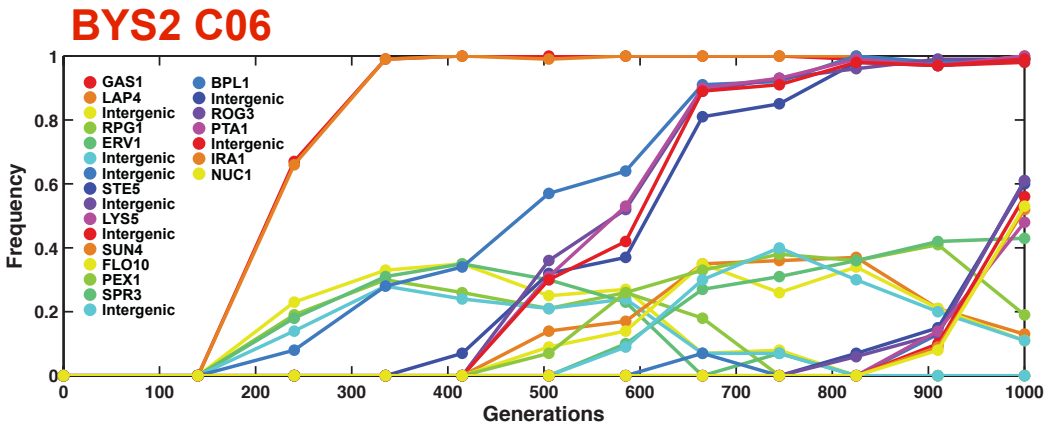


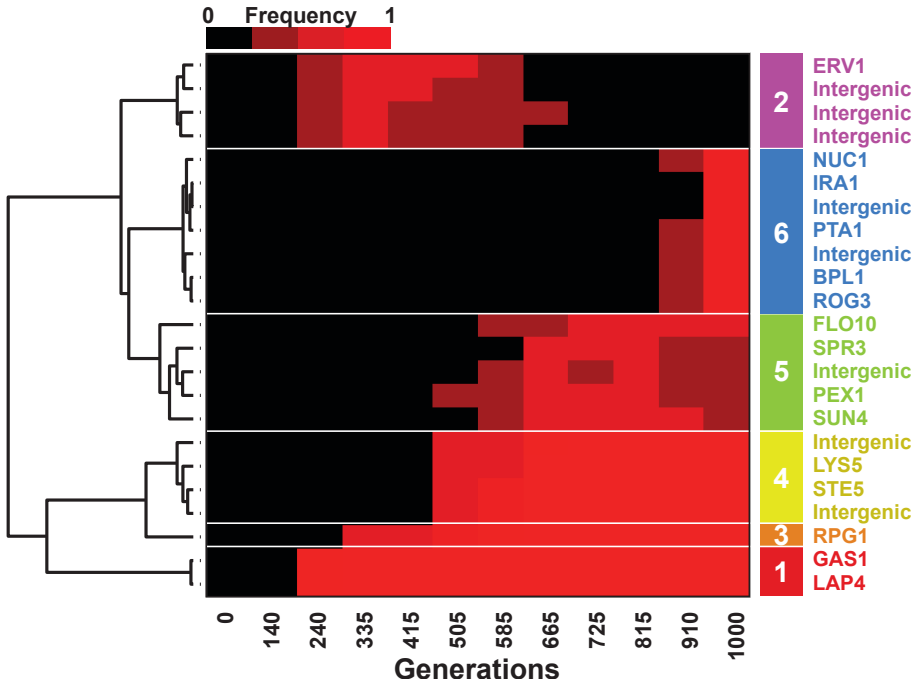
Sweeps within subpopulations

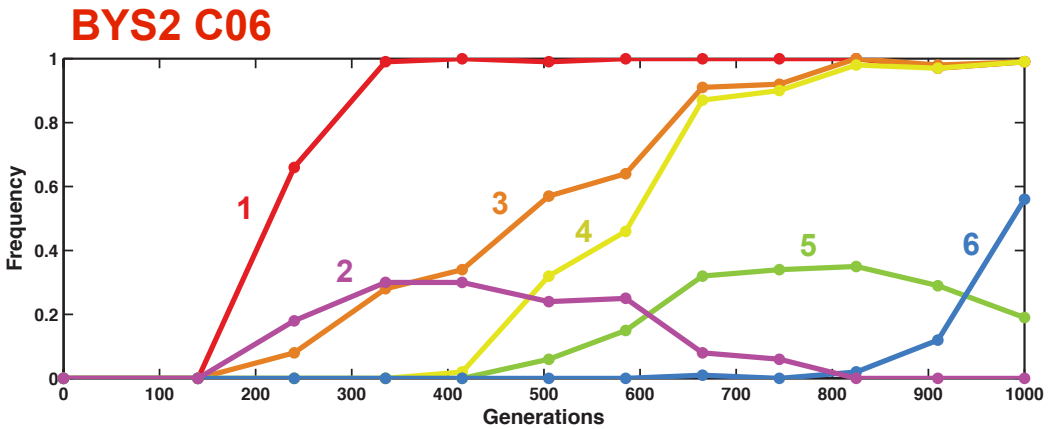


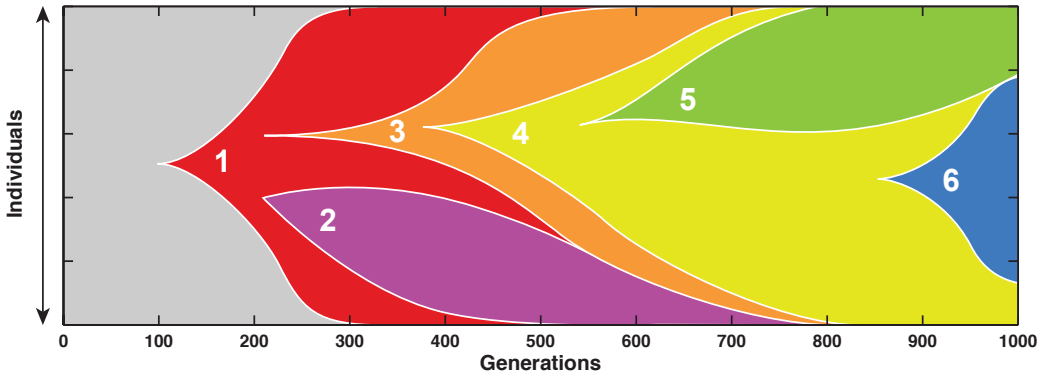


Clustering mutations and assigning lineages









Recurrent mutations drive adaptation

Yeast genome contains 5799 genes
We found 723 mutations in coding sequence

Hits	Exp.	Obs.
1	638	470
2	40	56
3+	2	**24**

0.6% of the yeast genome by size (74 kb / 12,463 kb)
14.0% of the observed mutations (141/1020)

Recurrent mutations drive adaptation

Hits	Exp.	Obs.
1	638	470
2	40	56
3+	2	**24**

Recurrent mutations drive adaptation

Hits	Exp.	Obs.
1	638	470
2	40	56
3+	2	24

Gene	Hits	Fixed	Biological Process
IRA1	21	10	Negative regulator of Ras
ROT2	11	2	Cell wall biogenesis
YUR1	11	5	Cell wall biogenesis
ACE2	9	4	Cytokinesis
STE11	9	1	Mating
STE12	9	2	Mating
PDR5	8	5	Multidrug transport
WHI2	7	2	General stress response
STE4	6	1	Mating
IRA2	5	3	Negative regulator of Ras
KRE6	4	1	Cell wall assembly
SFL1	4	1	Regulation of flocculation genes
STE5	4	3	Mating
ANP1	3	1	Protein glysosylation
CNE1	3	2	Protein folding
GAS1	3	3	Cell wall assembly
GCN1	3	1	Regulation of translation
GPB1	3	1	Negative regulator of Ras
GPB2	3	1	Negative regulator of Ras
KEG1	3	0	Cell wall assembly
KRE5	3	1	Cell wall assembly
RPO31	3	0	RNA ploymerase III transcription
SET4	3	2	unknown
YJL171C	3	0	unknown

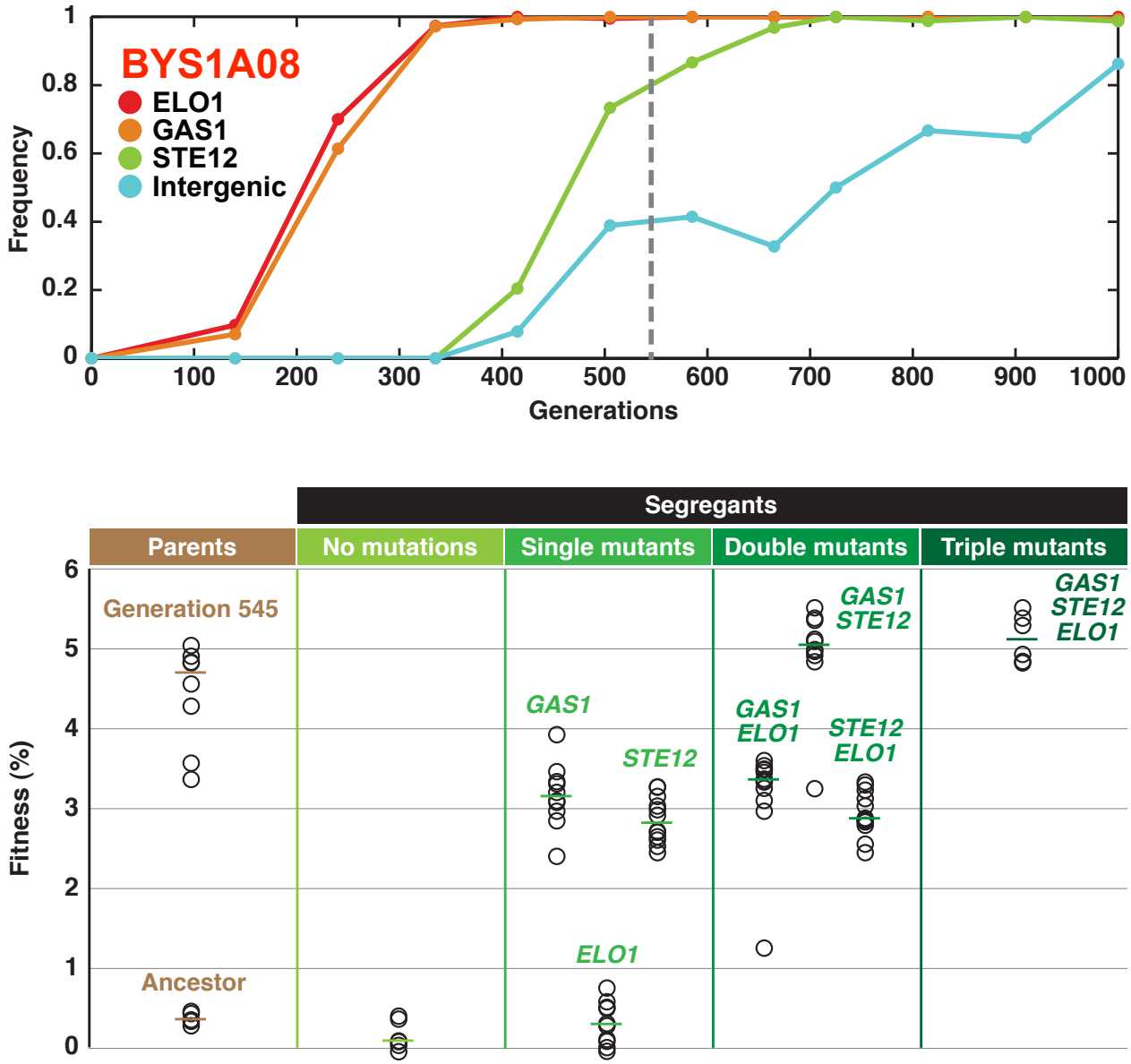
<1% synonymous (1/141)

18% (106/582) for all other coding mutations

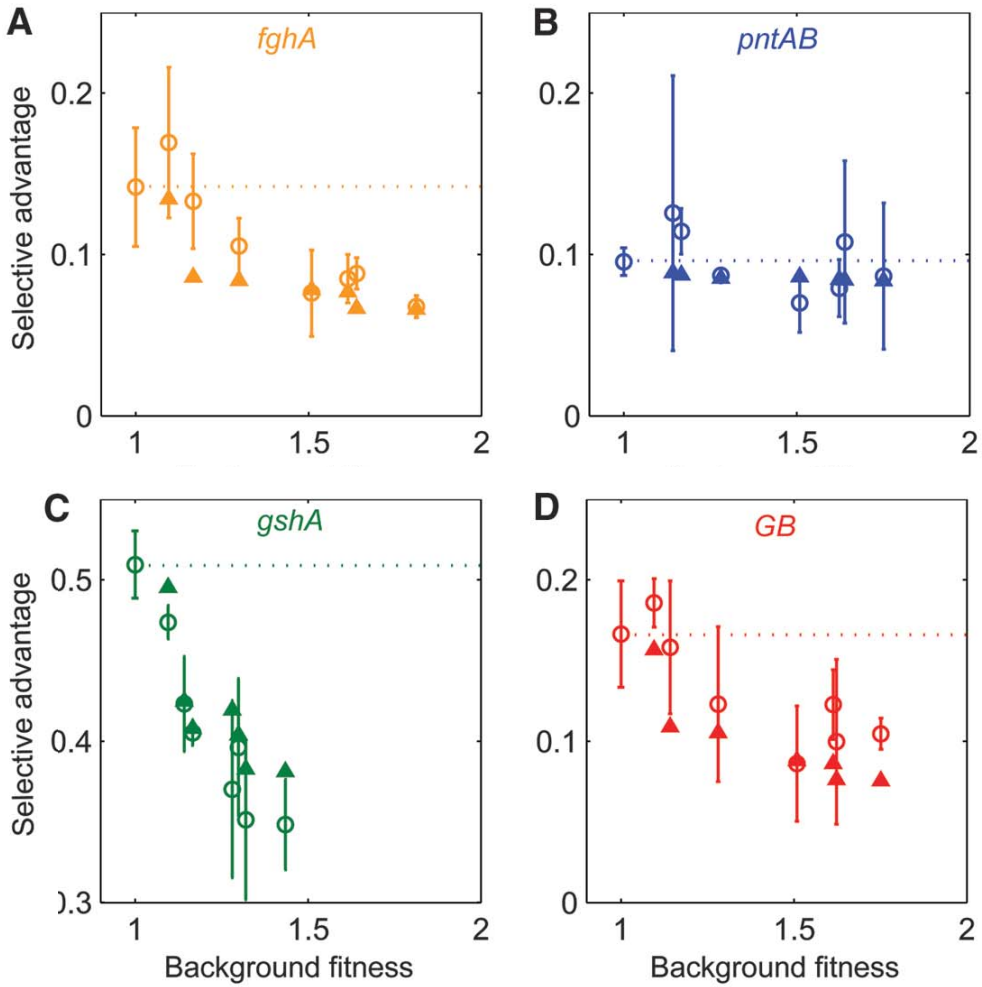
Gene	Amino Acid	Change	Fixed
IRA1	60	A->D	No
IRA1	163	L->P	Yes
IRA1	229	E->K	Yes
IRA1	400	K->*	Yes
IRA1	592	Y->S	No
IRA1	631	Frameshift	Yes
IRA1	761	Q->*	Yes
IRA1	822	Y->*	Yes
IRA1	1185	Frameshift	Yes
IRA1	1311	A->P	No
IRA1	1432	Q->K	No
IRA1	1553	T->P	No
IRA1	1846	P->L	No
IRA1	2685	E->*	No
IRA1	2689	W->*	No
IRA1	2723	Frameshift	Yes
IRA1	2723	Frameshift	No
IRA1	2723	Frameshift	Yes
IRA1	2723	Frameshift	No
IRA1	2972	Frameshift	No
IRA1	3034	E->*	Yes

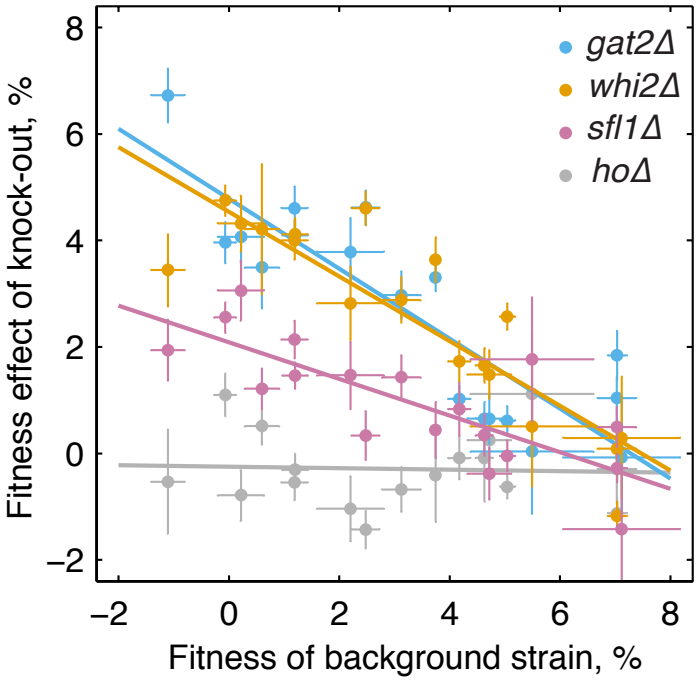
Gene	Amino Acid	Change	Fixed
KRE6	390	Y->S	Yes
KRE6	478	I->L	No
KRE6	515	S->C	No
KRE6	638	A->V	No
GAS1	177	A->T	Yes
GAS1	195	P->L	Yes
GAS1	259	F->L	Yes

Deconstructing fitness effects of individual mutations



Diminishing returns epistasis is a general phenomena





Chou et al. Science 2011
Kryazhimskiy et al. BioRxiv 2014

Summary

Adaptation is bound by the chance effects of mutation, drift, and hitchhiking

Unpredictability of genomic evolution

Selection acts deterministically on a small number of driver loci

Only a few traits under selection, but many mutations will give the desired phenotype

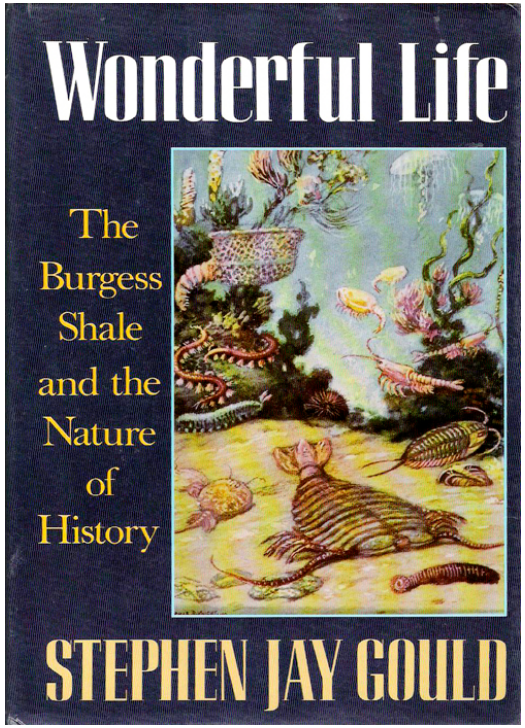
Lack of specific epistatic interactions

The identity of fixed mutations does not constrain future evolution

*****The number of phenotypic solutions is strongly limited such that replicate populations will converge on the same phenotype, each by way of a unique evolutionary trajectory.***

Caveat: It has been observed over long time scales that chance events can affect subsequent evolutionary change.

Gould's Gedankenexperiment



"our origin is the product of massive historical contingency, and we would probably never rise again even if life's tape could be replayed a thousand times."

Stephen Jay Gould, *Wonderful Life*

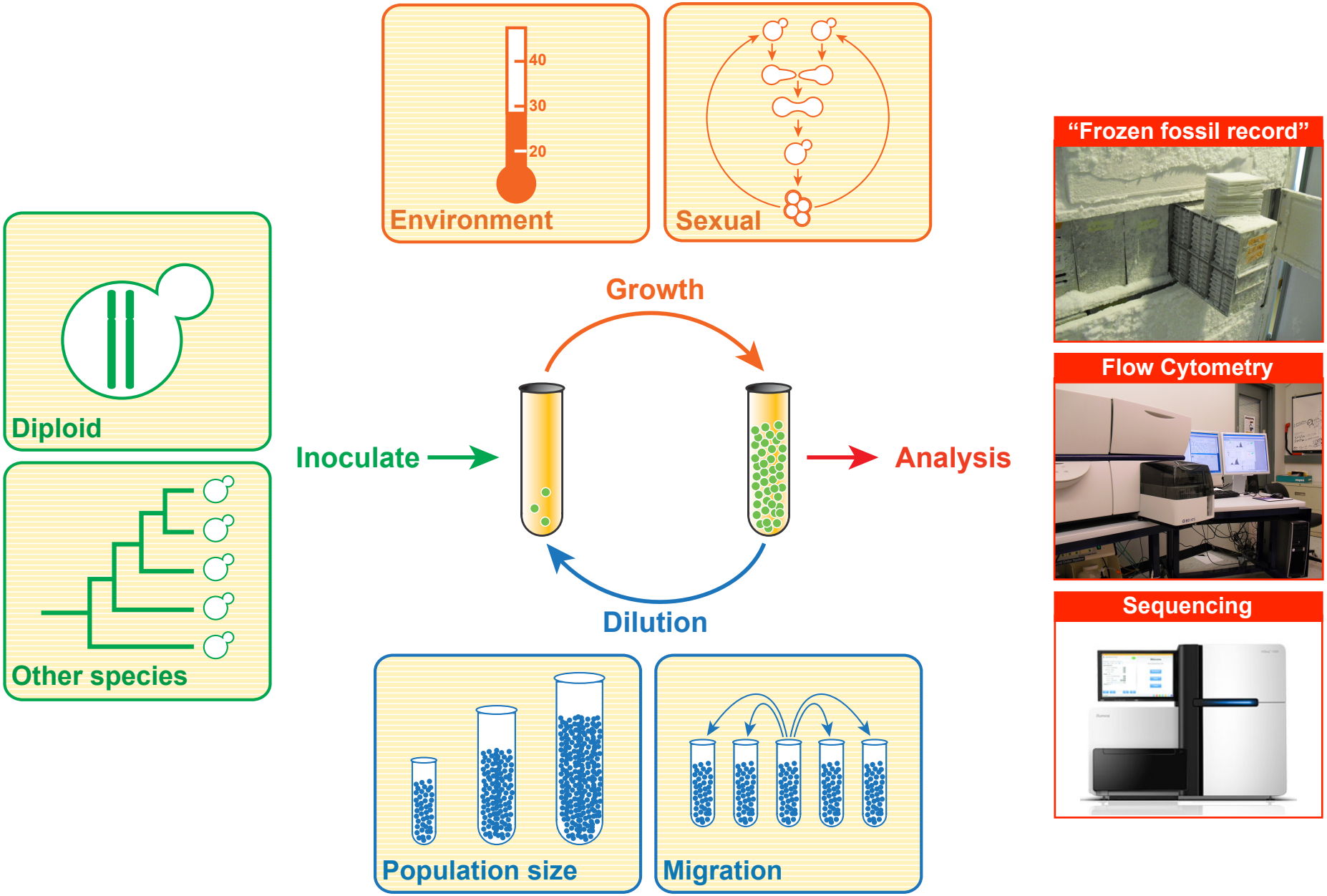
"evolutionary routes are many, but the destinations are limited."

Simon Conway Morris, *Life's Solution*

Genotypic evolution is not predictable.

Phenotypic evolution is predictable over short time scales, but over longer time scales, evolutionary outcomes are conditioned by specific epistatic interactions that either permit (or prohibit) particular evolutionary paths.

Versatile system for studying evolution



Acknowledgements

Princeton University

David Botstein
Mark Hickman

Harvard University

Michael Desai
Dan Rice

Washington University

George Weinstock
Erica Sodegren

Sequencing

Lance Parsons (Princeton)
Pat Gibney (Princeton)
Andrew Levy (Wash U.)
Tracie Miner (Wash U.)

Core Facilities

Tina DeCoste (Princeton)
Donna Storton (Princeton)
Jessica Buckles (Princeton)
Wei Wang (Princeton)

Funding

NIH