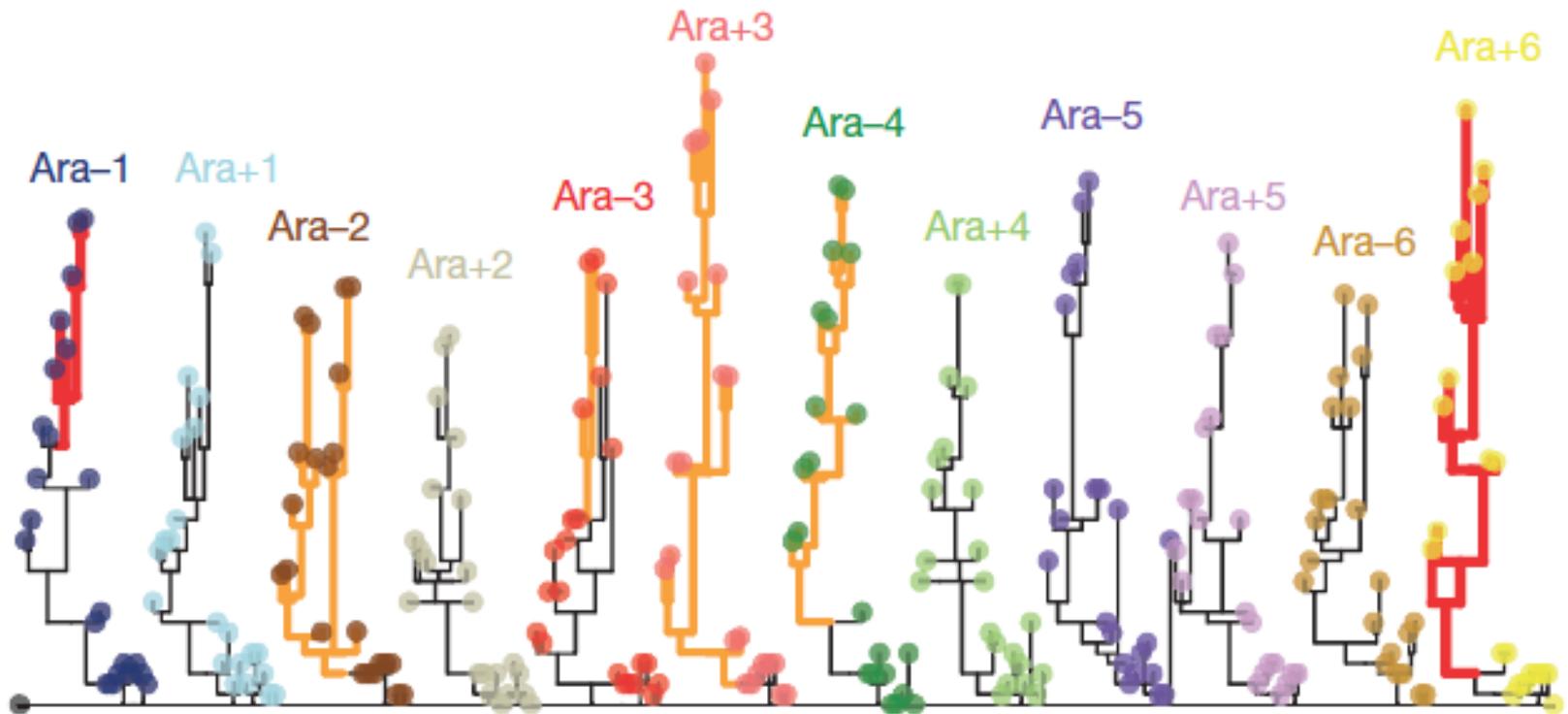


Dynamics of Adaptation and Genome Evolution in a Long-Term Experiment

Richard Lenski



Acknowledgments, among many more

- Jeff Barrick
- Zack Blount
- Chris Borland
- Tim Cooper
- Vaughn Cooper
- Carla Davidson
- Michael Desai
- Arjan de Visser
- Santiago Elena
- Phil Gerrish
- Ben Good
- Tanush Jagdish
- Jihyun Kim
- Bruce Levin
- Rohan Maddamsetti
- Elizabeth Ostrowski
- Noah Ribeck
- Peg Riley
- Michael Rose
- Danny Rozen
- Dominique Schneider
- Paul Sniegowski
- Olivier Tenaillon
- Mike Travisano
- Caroline Turner
- Sébastien Wielgoss
- Mike Wiser
- Bob Woods

Keeping it all going

- Neerja Hajela
- Lynette Ekunwe
- Sue Simpson

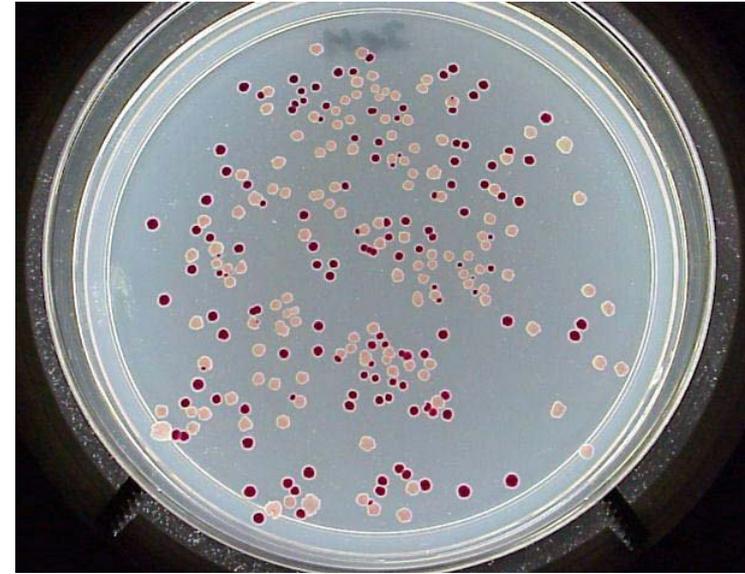


Motivating questions

- **Dynamics of adaptation by natural selection**
 - Is improvement slow and gradual? Or are there periods of rapid change and stasis? How long can fitness increase?
- **Repeatability of evolution**
 - Will replicate populations evolve along similar adaptive paths? Or will they find different solutions to identical environments?
- **Integration of phenotypic and genomic evolution**
 - How are phenotypic and genomic evolution coupled, both dynamically and functionally?

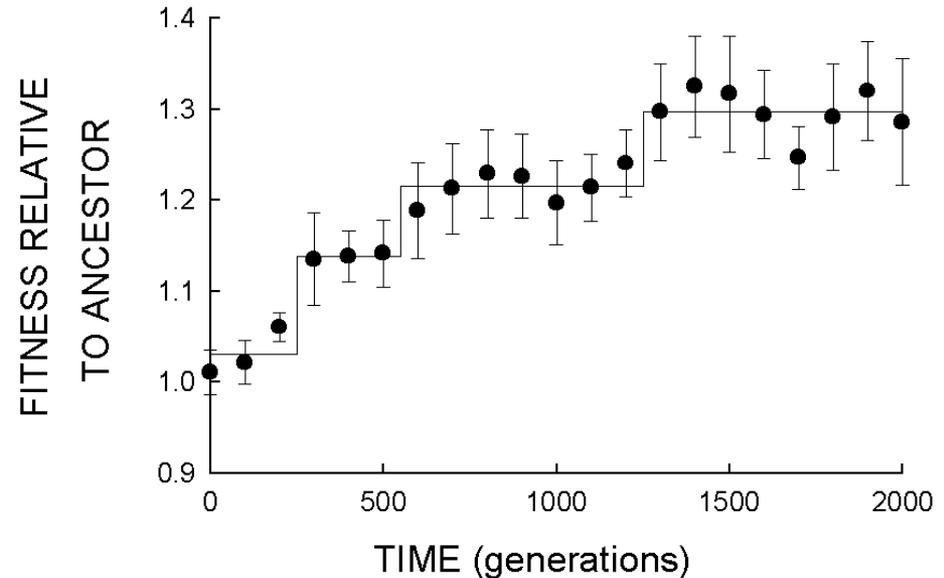
The experiment

- 12 populations started from same *E. coli* strain except for genetic marker used to distinguish them in competition assays.
- Propagated daily by 1:100 transfer in liquid medium with limiting glucose.
- (Citrate also present, but *E. coli* cannot grow on citrate in oxic environment.)
- $\text{Log}_2 100 \approx 6.6$ cell generations per day.
- Started 1988, now 67,000+ generations.
- Asexual and no HGT, but each population has had billions of mutations.
- Frozen “fossil record” includes ancestors and samples every 100 or 500 generations.



Fitness trajectories

- Short-term, individual population: Step-like sweeps of beneficial mutations of large effect.



- Short-term, individual population: Step-like sweeps of beneficial mutations of large effect.
- Long-term, grand mean across the populations: Gradually decelerating as fitness improves.

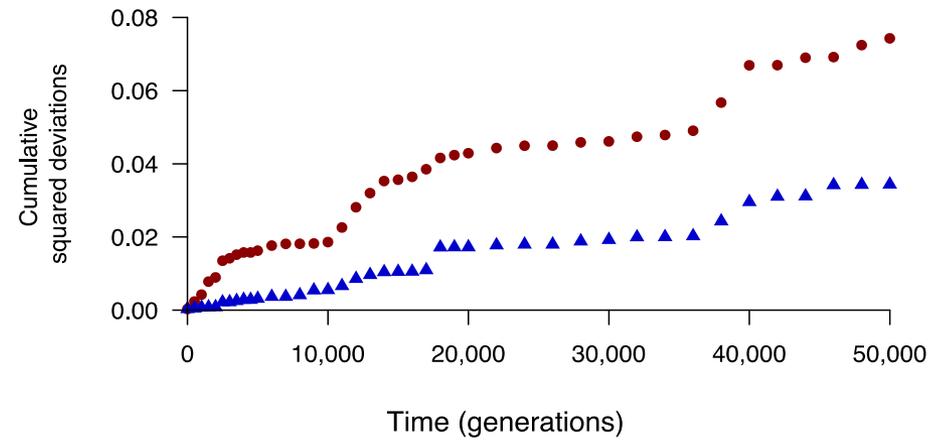
Asymptotic or not?

- Two models – both with just two free parameters – fit well. Correlations to grand mean of 0.969 and 0.981.

$$\bar{w} = (bt + 1)^a$$

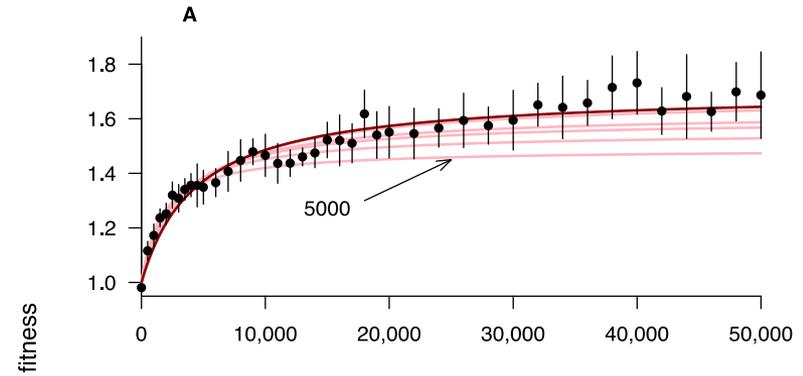
$$\bar{w} = 1 + at/(t + b)$$

- Two models – both with just two free parameters – fit well. Correlations to grand mean of 0.969 and 0.981.
- But **non-asymptotic power law** fits much better based on squared deviations ...

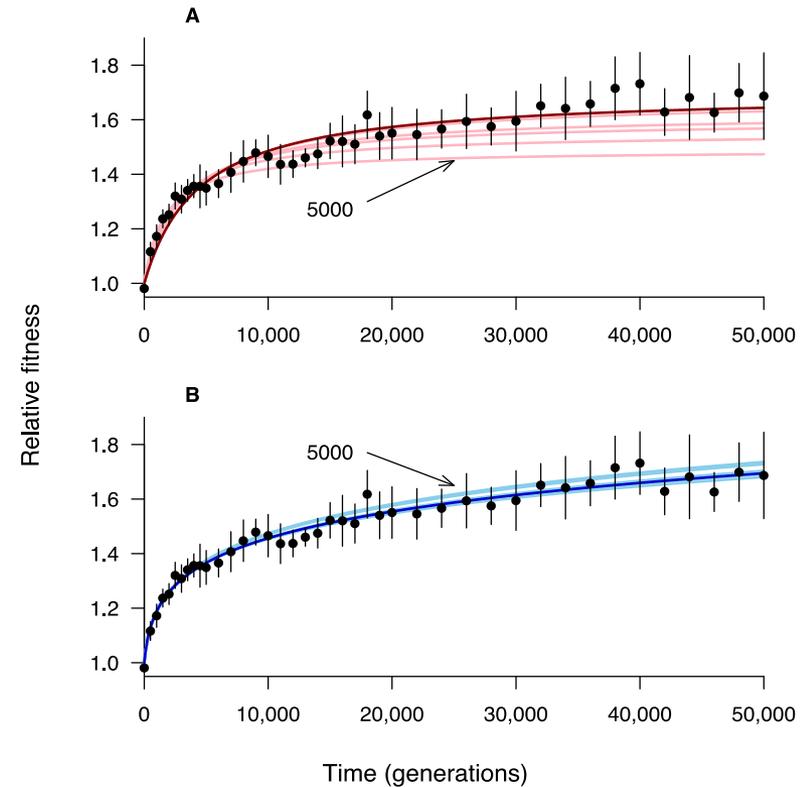


Power-law predictions

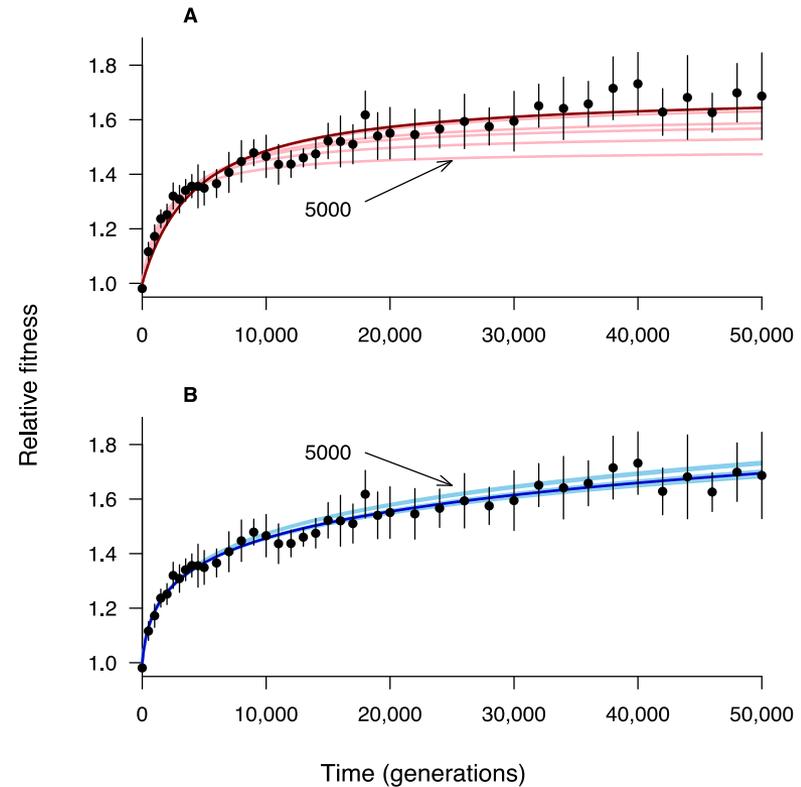
- When data is truncated, **asymptotic model** consistently underestimates future trajectory ...



- When data is truncated, **asymptotic model** consistently underestimates future trajectory ...
- While **power-law model** accurately predicts trajectory out to 50K given only first 5K of data.

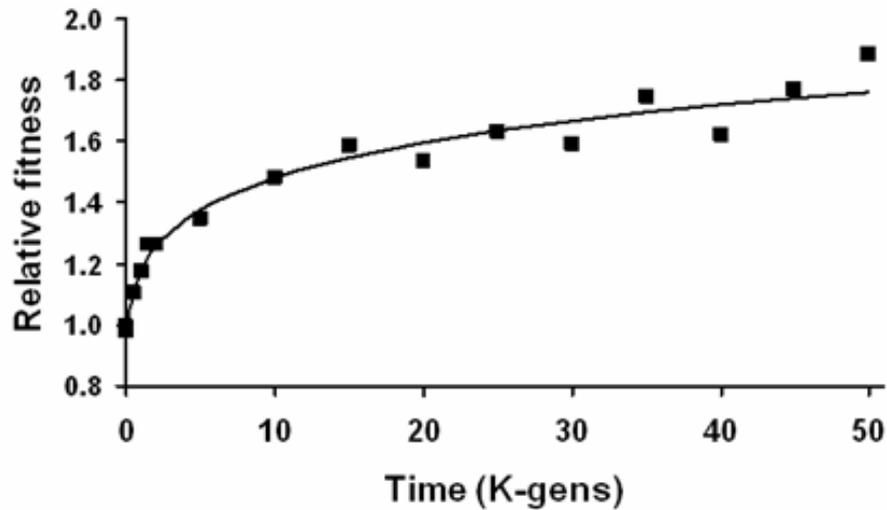


- When data is truncated, **asymptotic model** consistently underestimates future trajectory ...
- While **power-law model** accurately predicts trajectory out to 50K given only first 5K of data.
- Power-law model does *not* predict “crazy” values even extrapolated to $50,000^2$ generations (>1 million yr).
- Predicted doubling time (~23 min) very fast for minimal medium, but similar to *E. coli* in rich medium.

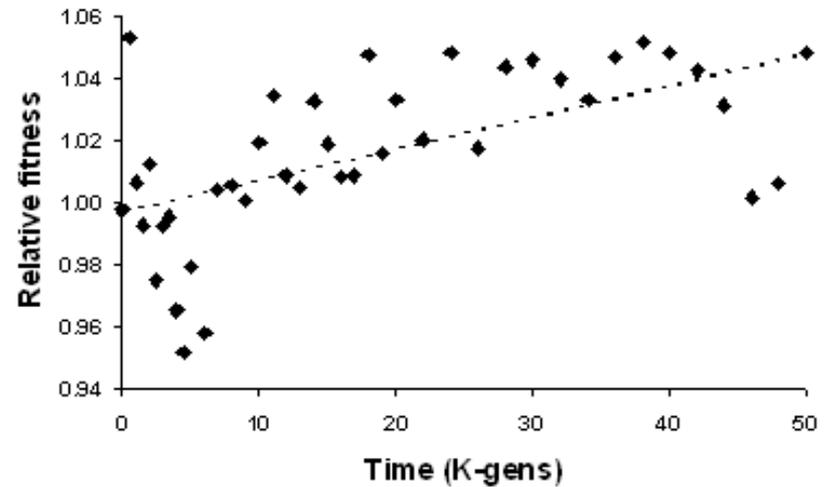


How repeatable?

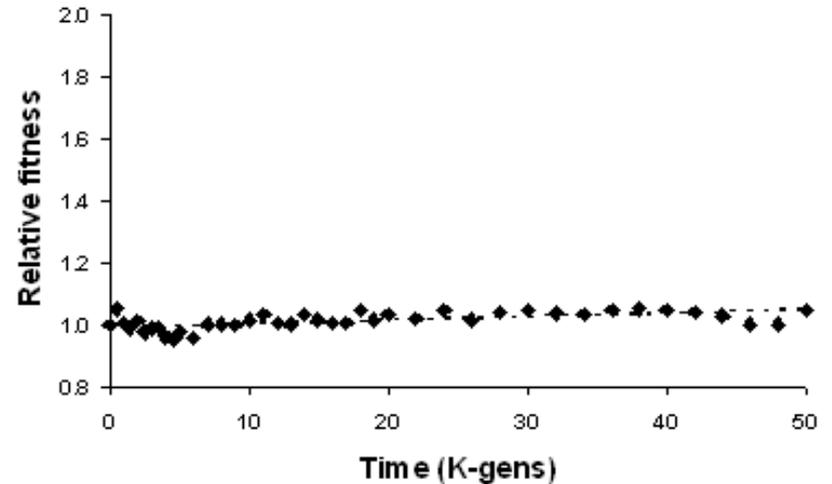
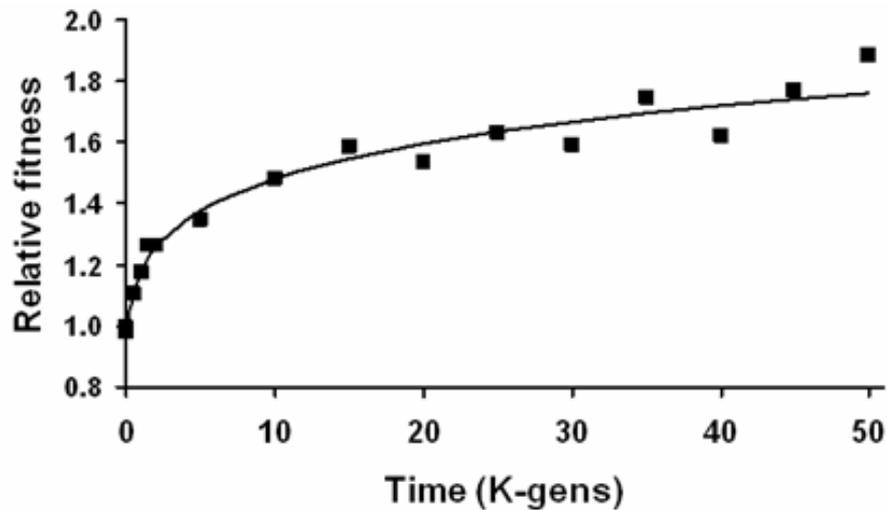
- Lower left: Fitness of one line *relative to ancestor*.



- Lower left: Fitness of one line *relative to ancestor*.
- Upper right: Fitness of two lines *relative to one another*.

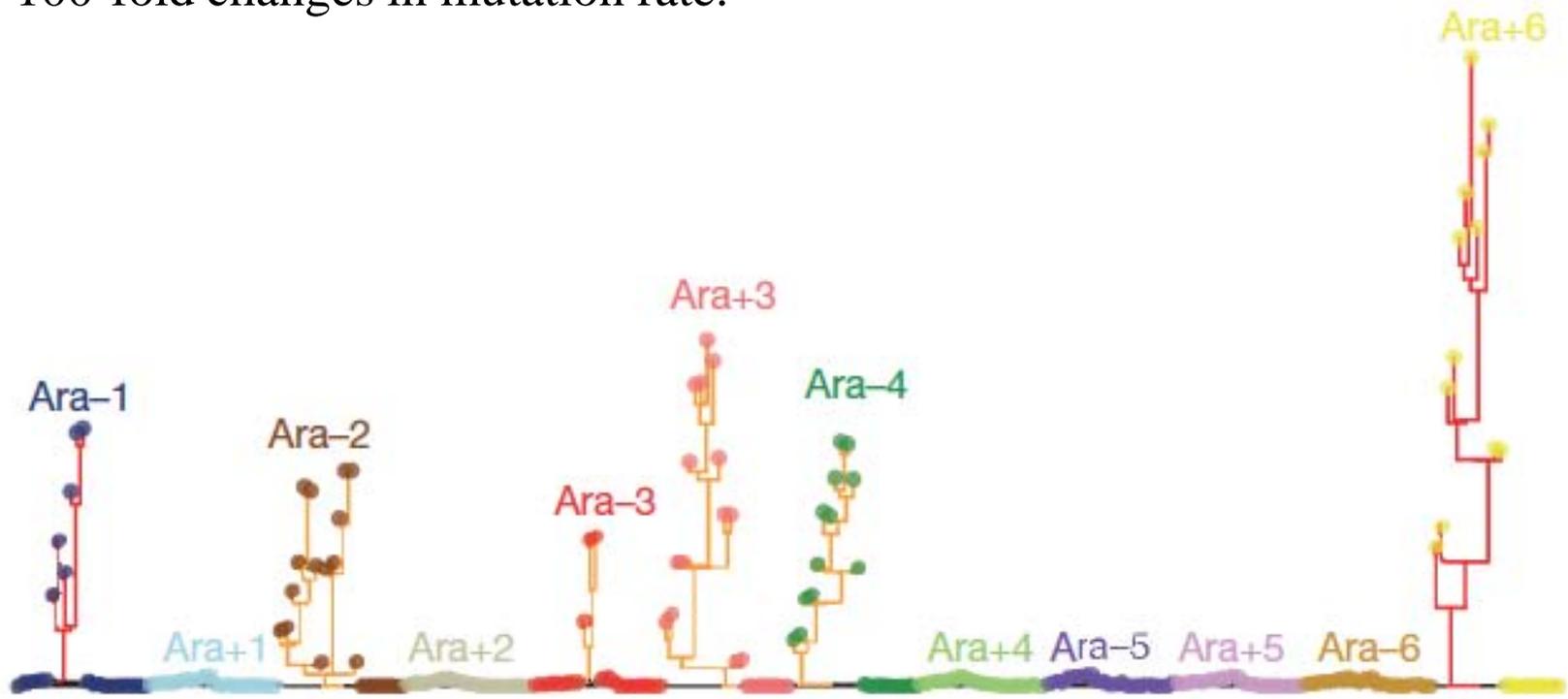


- Lower left: Fitness of one line *relative to ancestor*.
- Upper right: Fitness of two lines *relative to one another*.
- Lower right: Fitness of two *rescaled* as at left.

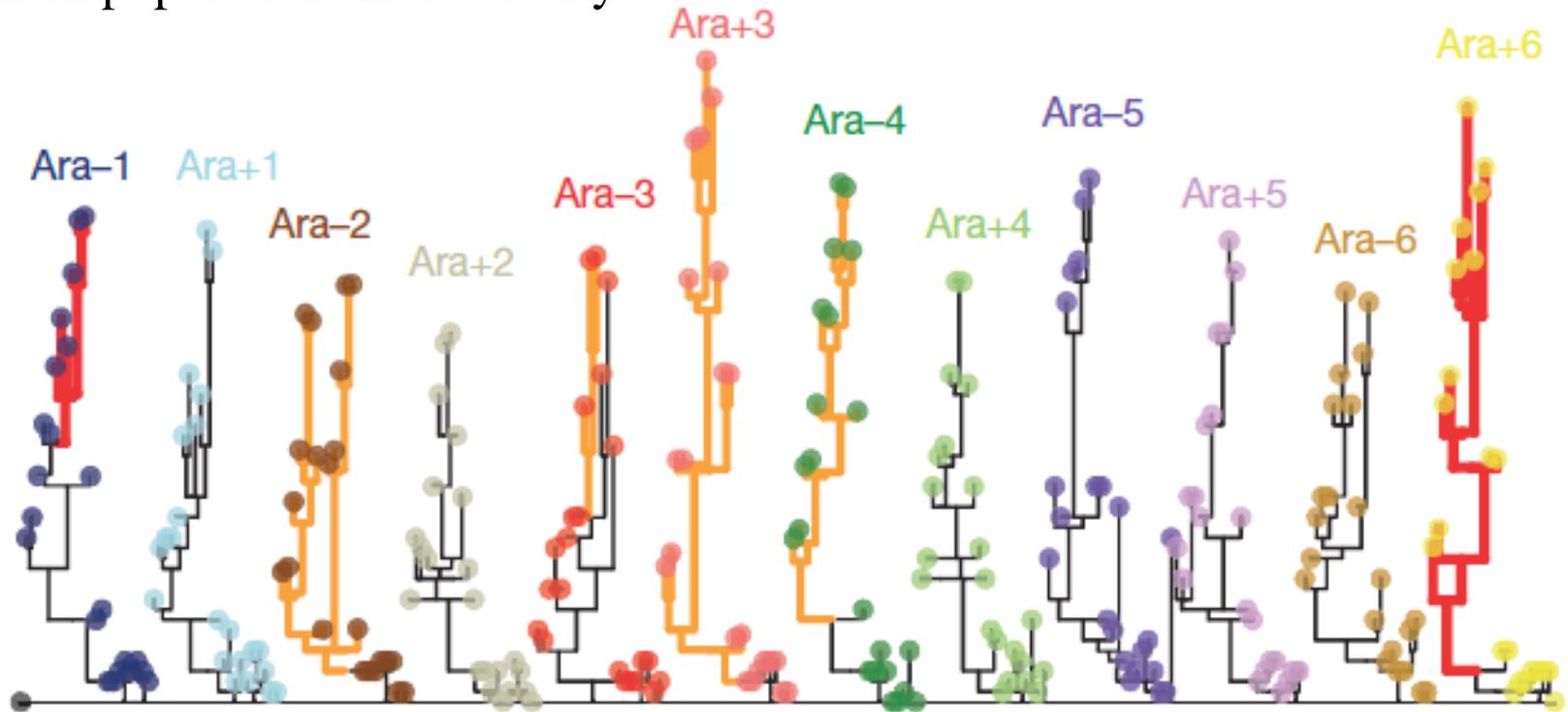


Tempo and mode of genome evolution

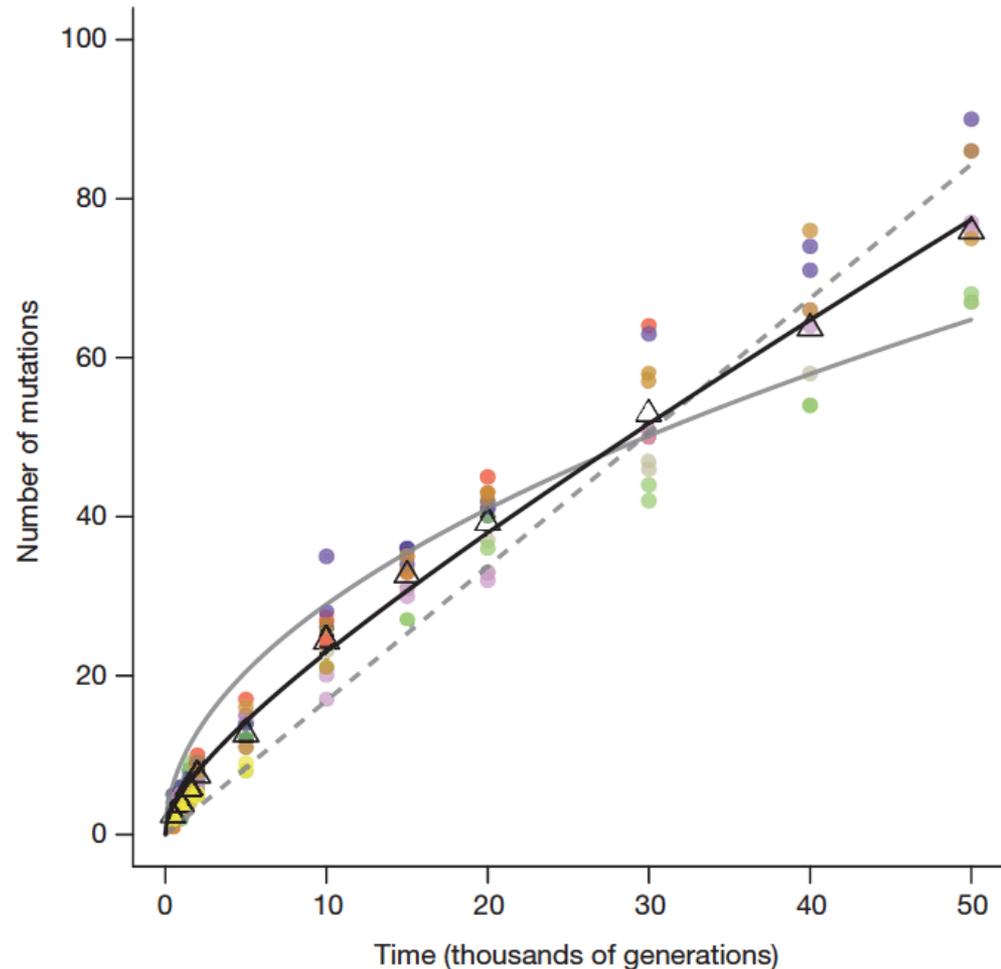
- Sequenced **264 whole genomes** (12 populations x 11 time points x 2 clones).
- Branch lengths differ greatly due to ~100-fold changes in mutation rate.



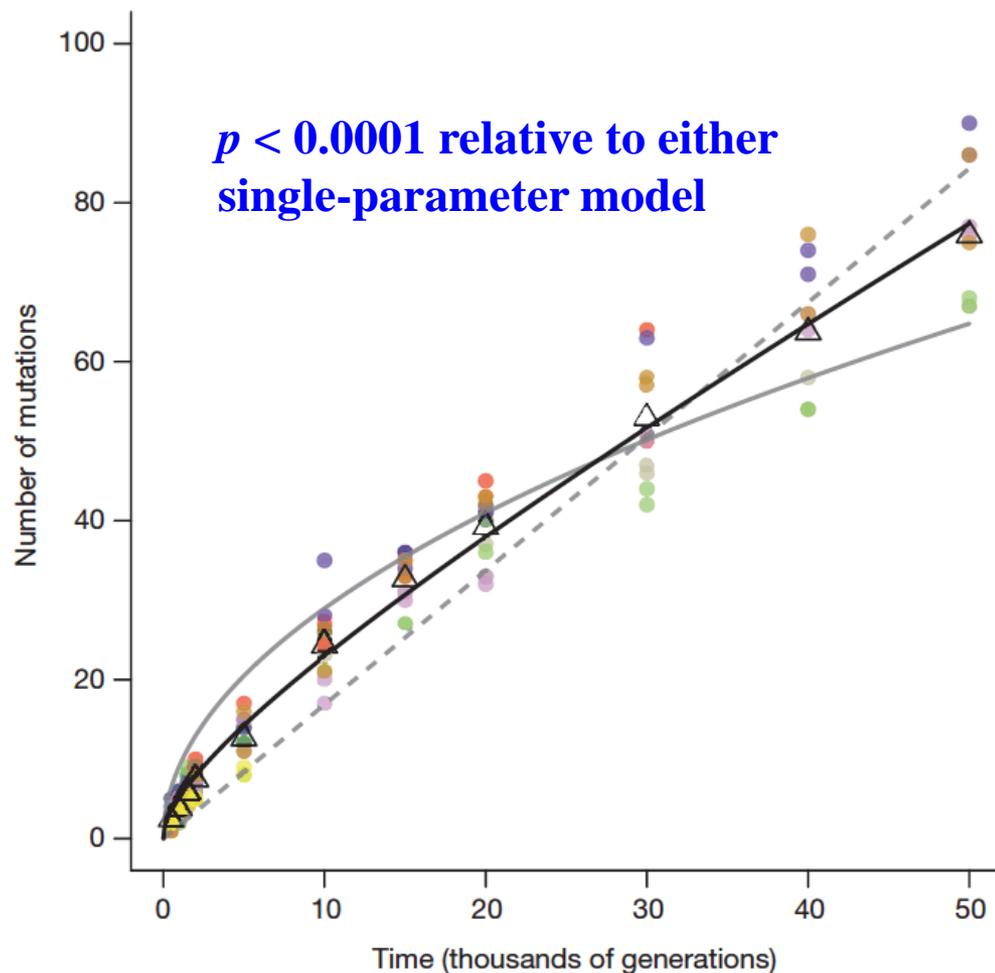
- Sequenced **264 whole genomes** (12 populations x 11 time points x 2 clones).
- **Branch lengths *rescaled*** to show all 12 populations more clearly.



- Sequenced **264 whole genomes** (12 populations x 11 time points x 2 clones).
- **Branch lengths *rescaled*** to show all 12 populations more clearly.
- Mutations in **nonhypermutable populations** increase steadily.
- **Beneficial mutations** predicted by dynamical model that gives a power-law trajectory for fitness (solid grey) – too curved.
- **Neutral hitchhikers and drifters** predicted to increase at constant rate (dashed grey) – too straight.

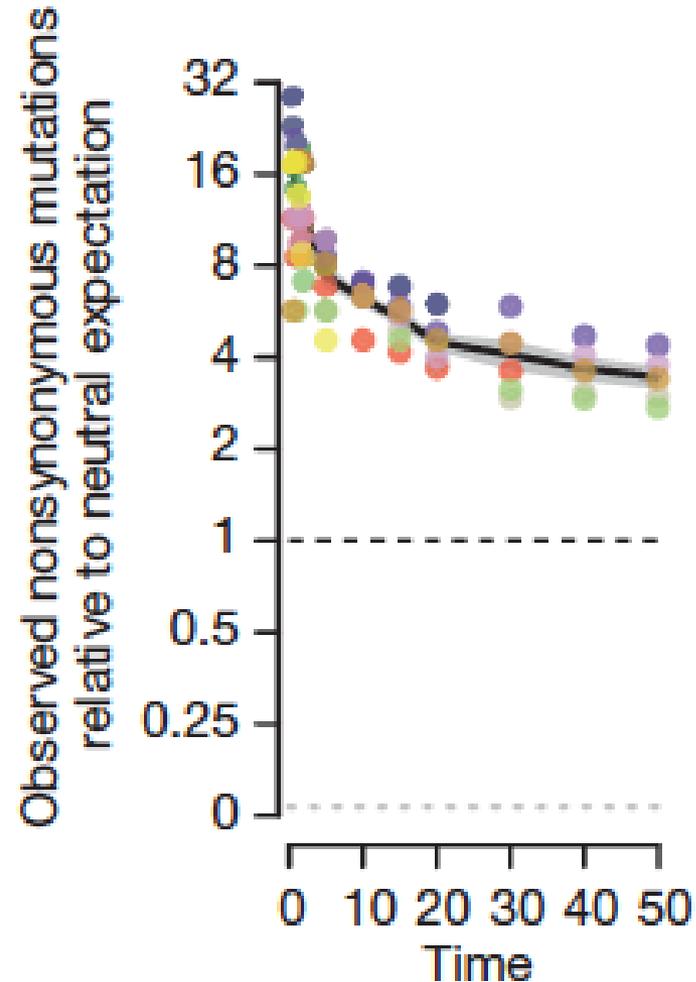


- Sequenced **264 whole genomes** (12 populations x 11 time points x 2 clones).
- **Branch lengths *rescaled*** to show all 12 populations more clearly.
- Mutations in **nonhypermutable populations** increase steadily.
- **Beneficial mutations** predicted by dynamical model that gives a power-law trajectory for fitness (solid grey) – too curved.
- **Neutral hitchhikers and drifters** predicted to increase at constant rate (dashed grey) – too straight.
- Data fits model with **both drivers and passengers** (black).



Excess nonsynonymous mutations

- From the **genomic data**, we can also quantify the **ratio of nonsynonymous to synonymous mutations** in protein-coding genes (adjusted for sites at risk).



Parallel genetic evolution

- Very few mutations are *identical* at the level of nucleotides.
- Extensive parallelism at the level of genes. Just 57 genes that comprise ~2% of the protein-coding genome have 50% of all nonsynonymous changes in the nonhypermutable lineages.

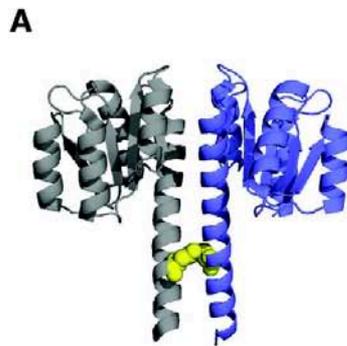
- Very few mutations are *identical* at the level of nucleotides.
- Extensive parallelism at the level of genes. Just 57 genes that comprise ~2% of the protein-coding genome have 50% of all nonsynonymous changes in the nonhypermutable lineages.
- Randomization test indicates this parallelism is significant ($p < 10^{-143}$).

- Very few mutations are *identical* at the level of nucleotides.
- Extensive parallelism at the level of genes. Just 57 genes that comprise ~2% of the protein-coding genome have 50% of all nonsynonymous changes in the nonhypermutable lineages.
- Randomization test indicates this parallelism is significant ($p < 10^{-143}$).

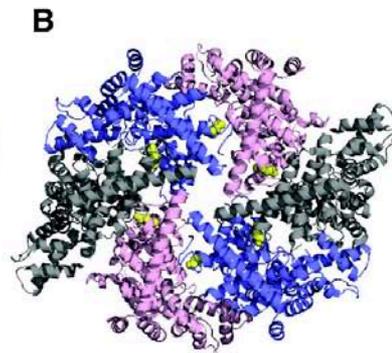


- Very few mutations are *identical* at the level of nucleotides.
- Extensive parallelism at the level of genes. Just 57 genes that comprise ~2% of the protein-coding genome have 50% of all nonsynonymous changes in the nonhypermutable lineages.
- Randomization test indicates this parallelism is significant ($p < 10^{-143}$).
- Among the genes with strongest parallelism are ones with *core metabolic and regulatory functions* including pyruvate kinase, peptidoglycan synthesis, DNA supercoiling, and stringent response.
- **Hypothesis: Optimization by fine-tuning** the regulation and expression of important functions in this simple, predictable environment.

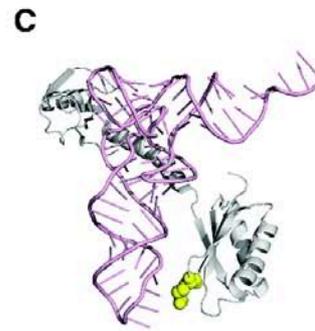
- In 10 of these 57 genes, some of the parallel changes are at the level of the *amino-acid* sequence. In 8 cases, the substitutions occur within 8 Å of a bound protein, RNA, or ligand.
- Consistent with fine-tuning (as opposed to loss of function).



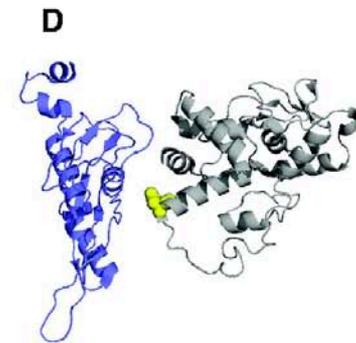
atoC: 1NY5



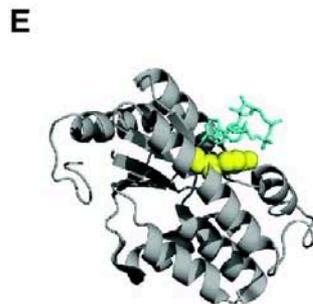
hflB: 2CE7



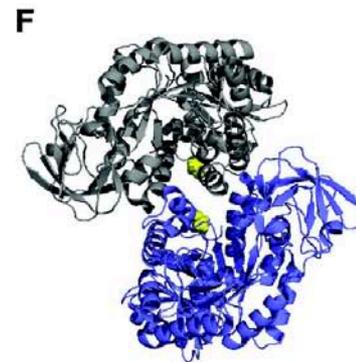
infC: 5LMQ



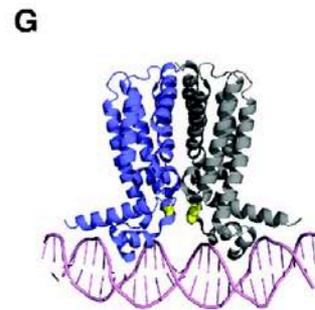
rpsD: 3J9Y



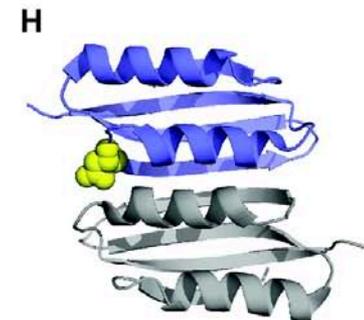
nadR: 1LW7



pykF: 4YNG



yijC: 3LSR



spoT: 3IBW

- Also frequent parallelism involving *gene deletions*.
- Many deleted genes were horizontally acquired (in the distant past) including remnants of prophages and plasmid-derived toxin-antitoxin modules.
- Other deleted genes encode host-specific adaptations including production of extracellular capsules and fimbriae.
- Average genome size declined by ~1.4% over 50K generations.
- *Hypothesis: Optimization by eliminating* unused and perhaps costly functions.

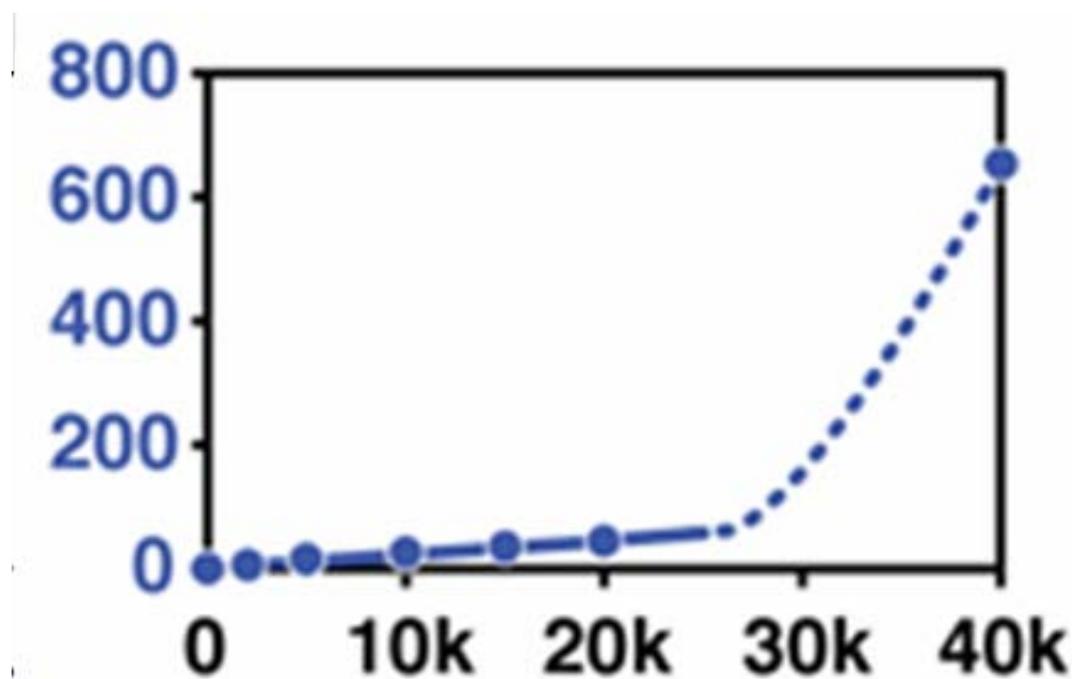
Evolving complexity

- Despite the intended simplicity of the experiment, some fascinating complications have evolved.
- Changing mutation rates and mechanisms including:
 - Hypermutability, as already noted,
 - And subsequent compensatory changes.

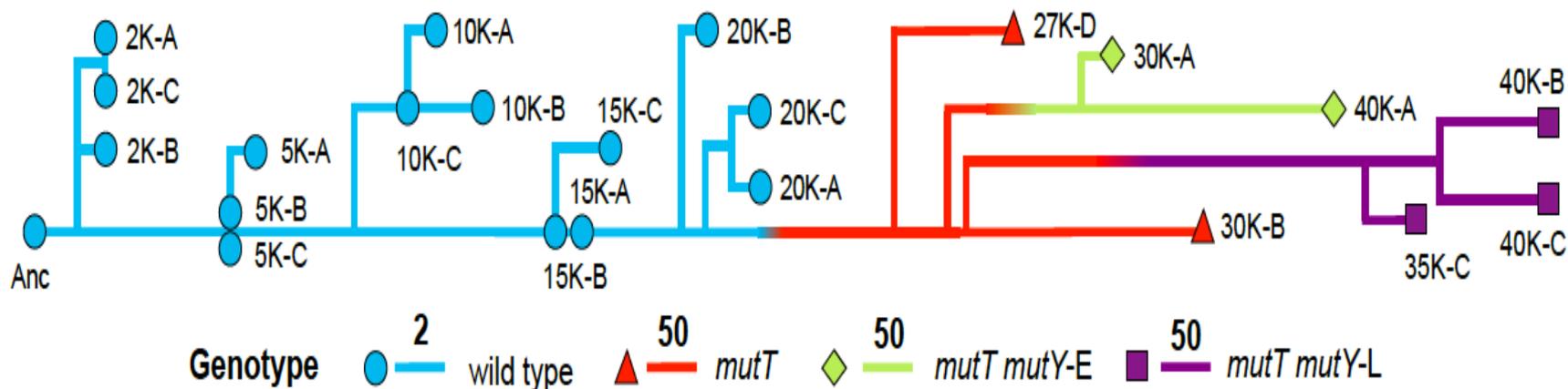
- Despite the intended simplicity of the experiment, some fascinating complications have evolved.
- Changing mutation rates and mechanisms including:
 - Hypermutable, as already noted,
 - And subsequent compensatory changes.
- Niche construction via secreted metabolites and crossfeeding interactions leading to:
 - Sometimes transient coexistence,
 - And sometimes sustained coexistence.

Changing mutation rates and mechanisms

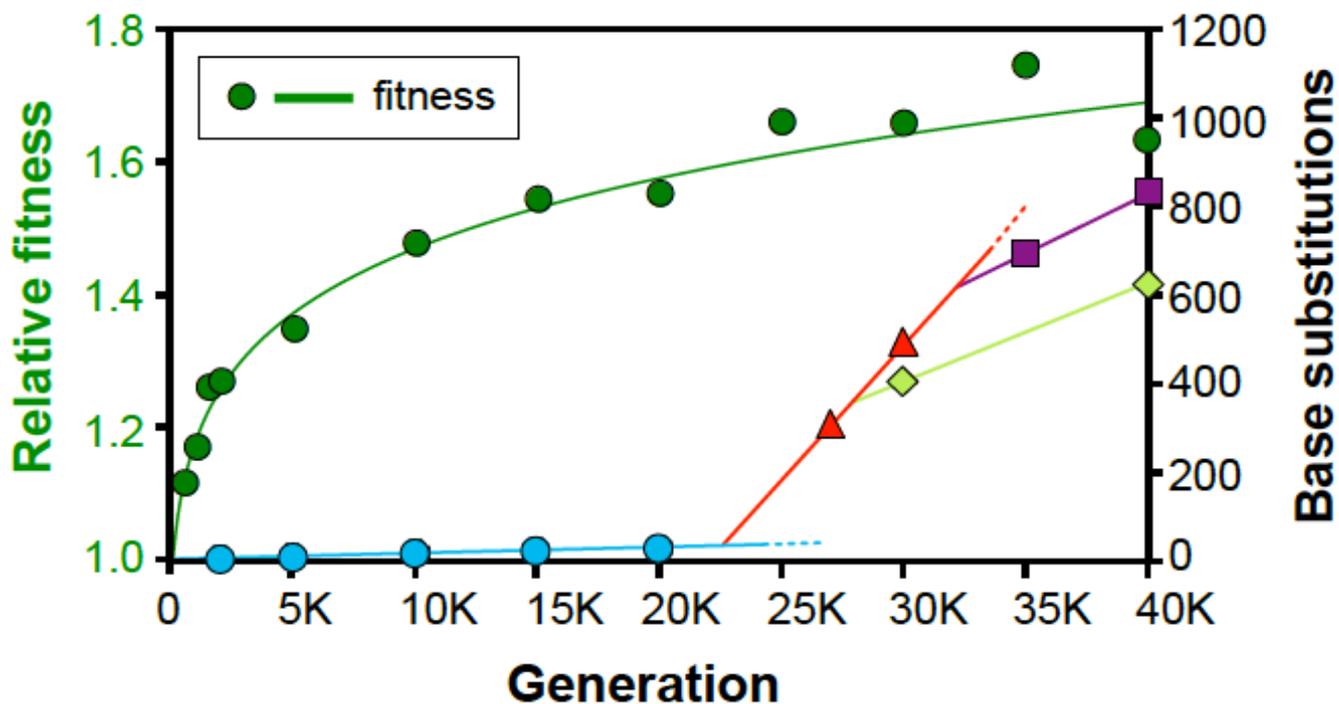
- Between 20K and 30K generations, population below evolved hypermutability, with 646 mutations at 40K vs. only 45 at 20K.
- Many of these late mutations were synonymous, and most bore the *mutT* signature of A·T→C·G transversions.



- Mutation dynamics became even more complex and interesting when sequenced more clones from this population.
- Branch lengths varied significantly among 40K clones.
- Further mutation-rate increase on one branch? Or compensatory reduction in the rate on the other branch?



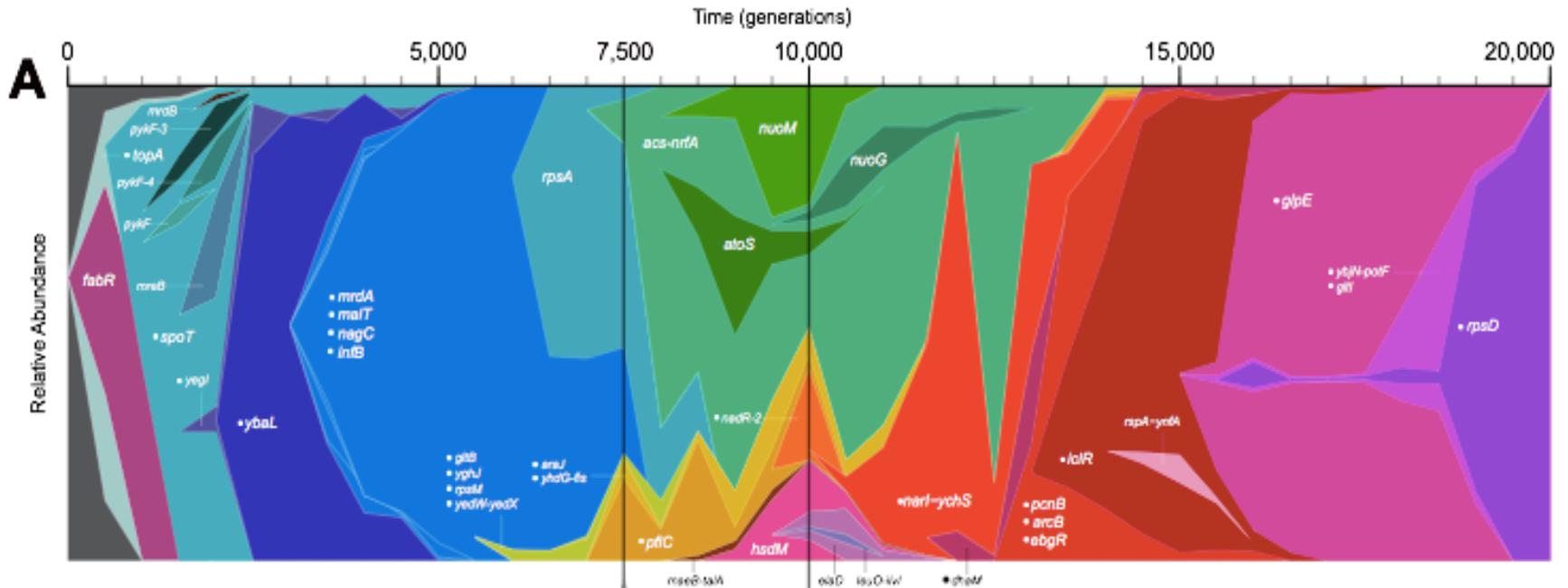
- In fact, the rate of mutation accumulation had declined along *both* branches (but starting at different times).
- Anti-mutators **reduced genetic load** and thereby increased fitness by $\sim 0.5\%$. (Example of parallelism *within* population.)



Evolving complexity

- Despite the intended simplicity of the experiment, some fascinating complications have evolved.
- Changing mutation rates and mechanisms including:
 - Hypermutable, as already noted,
 - And subsequent compensatory changes.
- Niche construction via secreted metabolites and crossfeeding interactions leading to:
 - Sometimes transient coexistence,
 - And sometimes sustained coexistence.

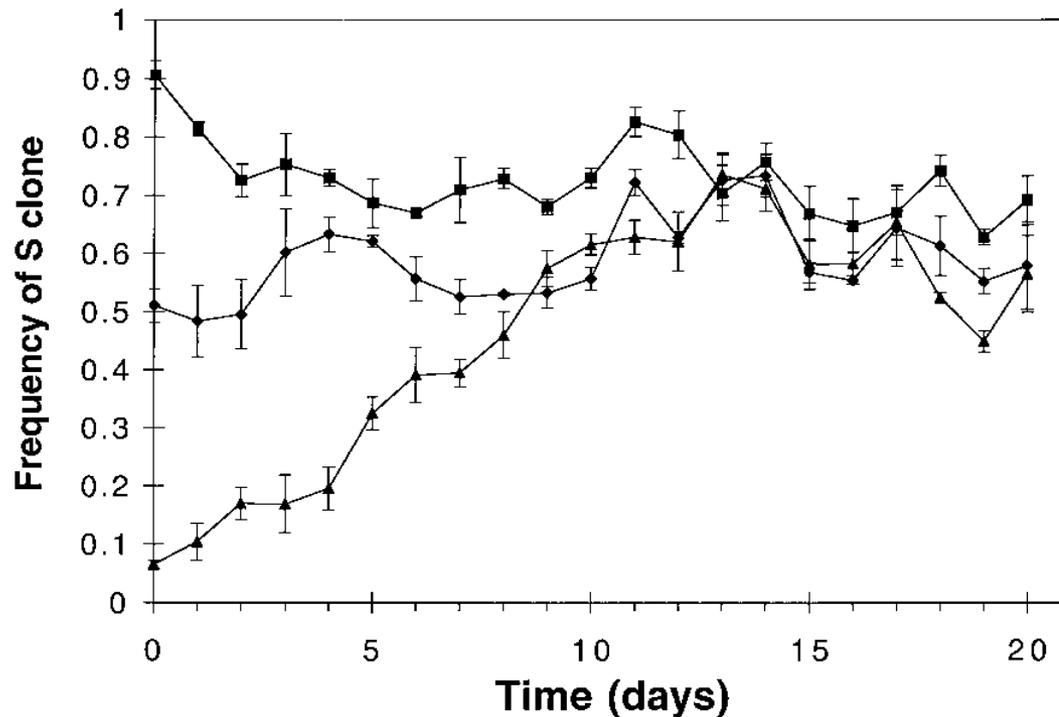
Transient coexistence



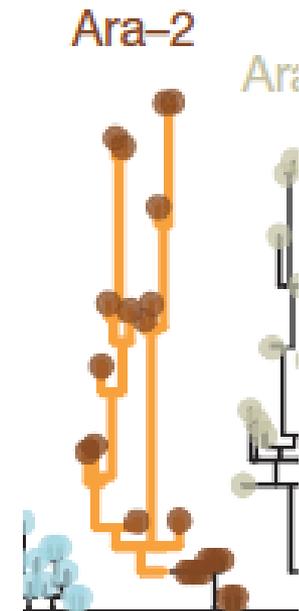
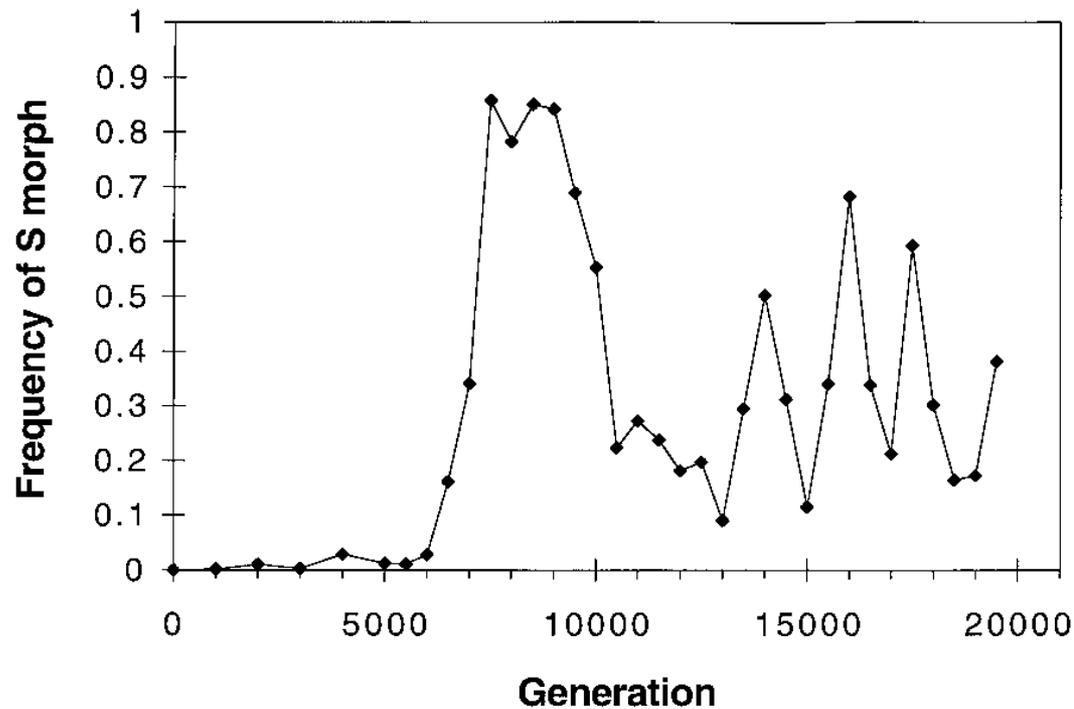
- Using frozen samples, tracked several dozen mutations in this population (before hypermutability evolved).
- Rapid selective sweeps—some single mutations, others involving two or more mutations—both early (before 5K) and late (after 15K).
- Quasi-stable coexistence of two lineages from ~6K to ~14K. Oscillations probably driven by beneficial mutations in one or the other lineage, before “red” clade finally drove “green” clade extinct.

Sustained coexistence

- In another population, we saw large (L) and small (S) colony-morphology variants after ~6,000 generations.
- When L and S clones compete at different starting ratios, they converge on intermediate frequencies → stable equilibrium on *ecological* time scale.

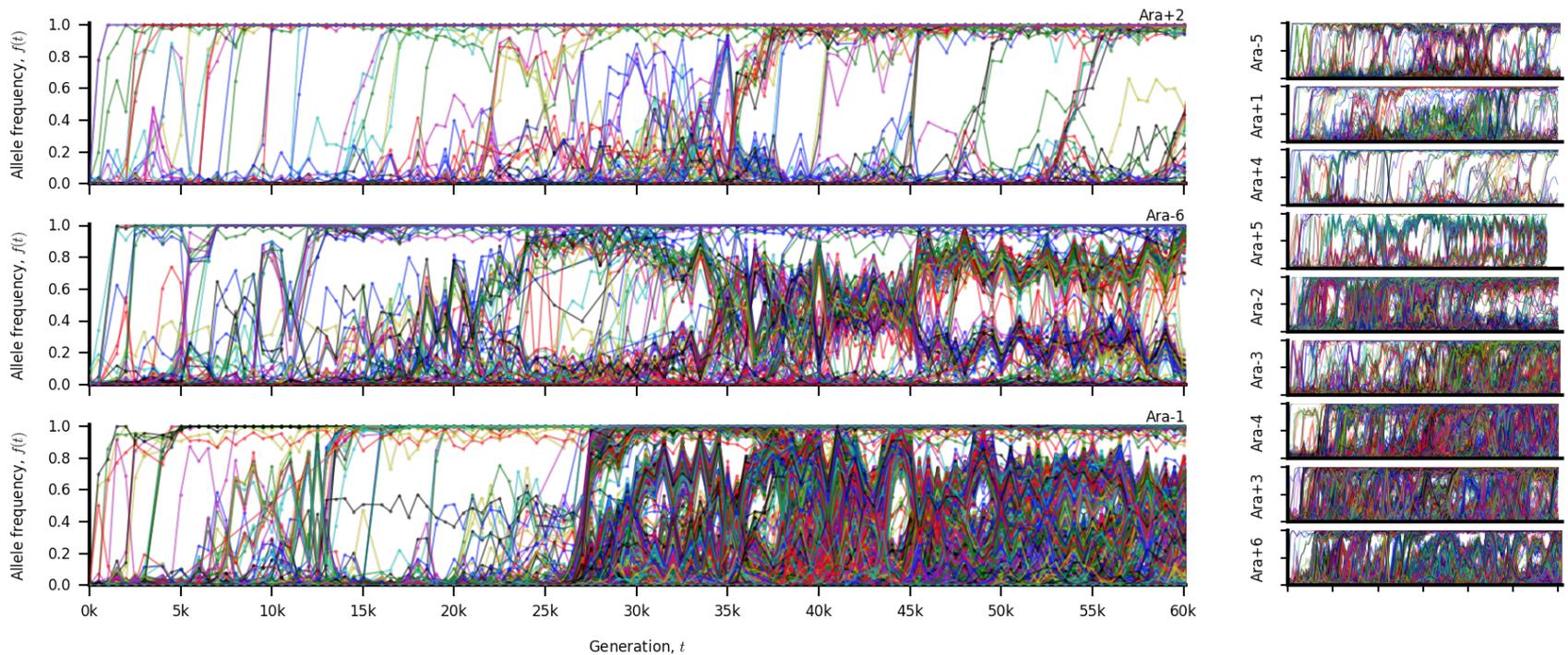


- But over *evolutionary* time, their relative abundances fluctuate as they continue to evolve and adapt to the environment and to one another.
- Still, neither has driven the other extinct even after 50K gens.



Coming down the pipeline ...

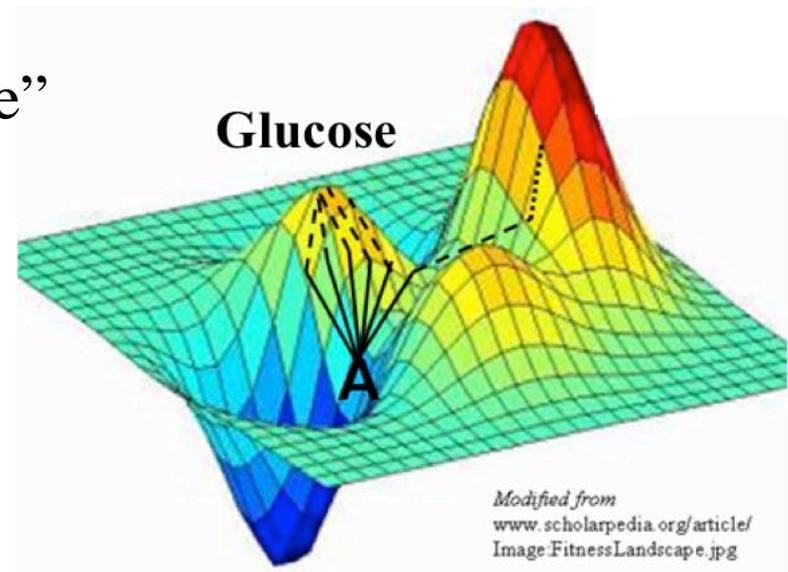
- Metagenomic sequencing of all 12 populations at 500-generation resolution
- Work with Ben Good and Michael Desai



Partial summary

- Most *observed* mutations (excluding in hypermutators) are beneficial.
- Despite the increasing complexity, substantial parallel evolution, both genotypically and phenotypically.

- Most *observed* mutations (excluding in hypermutators) are beneficial.
- Despite the increasing complexity, substantial parallel evolution, both genotypically and phenotypically.
- Populations climbing “Mt. Glucose” in similar, but not identical, ways.

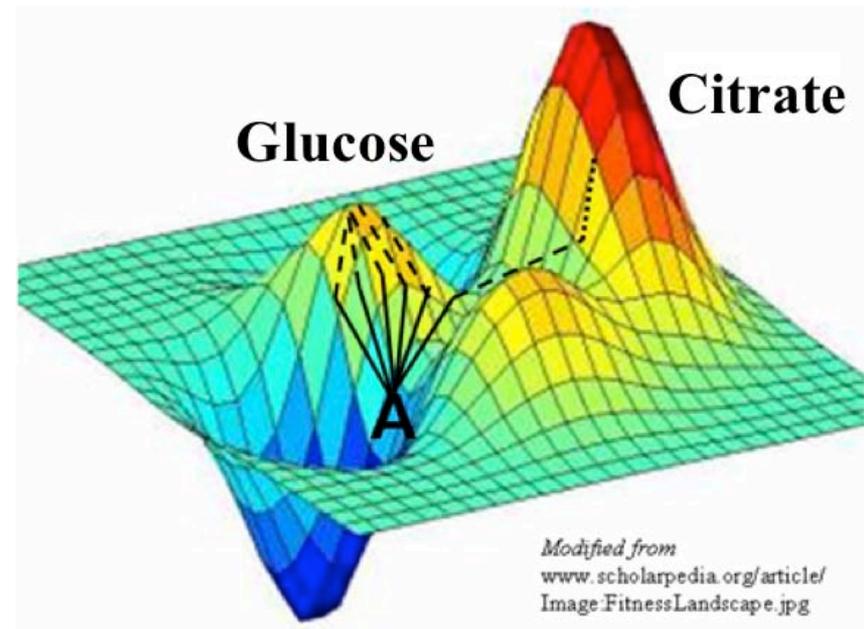


Until a striking *exception*

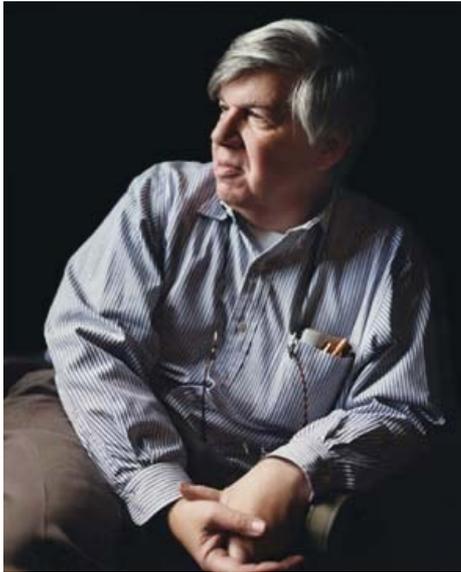
- After 30,000 gens. one population evolved ability to use the citrate that had been present throughout.
- Diagnostic feature of *E. coli* is it cannot transport citrate into cell under oxic conditions.



- After 30,000 gens. one population evolved ability to use the citrate that had been present throughout.
- Diagnostic feature of *E. coli* is it cannot transport citrate into cell under oxic conditions.
- An extremely rare mutation? Or historical contingency dependent on particular evolutionary path?



- “I call this experiment ‘**replaying life’s tape.**’ You press the rewind button and ... go back to any time and place in the past ... Then let the tape run again and see if the repetition looks at all like the original ... The bad news is that **we can’t possibly perform the experiment.**”



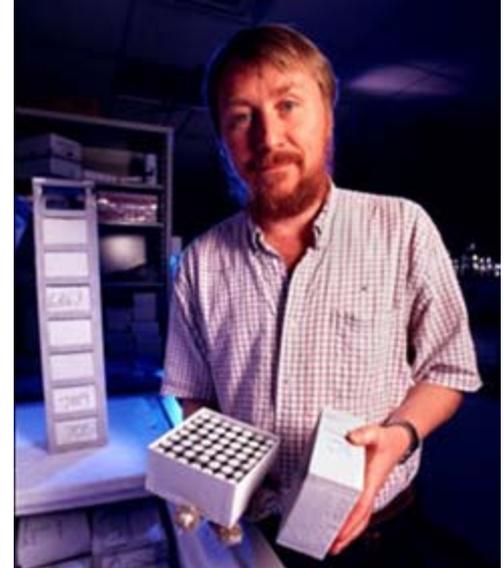
Kathy Chapman via Wikipedia

Stephen Jay Gould,
1989, *Wonderful Life*



Mark A. Wilson via
Wikipedia

- “I call this experiment ‘**replaying life’s tape.**’ You press the rewind button and ... go back to any time and place in the past ... Then let the tape run again and see if the repetition looks at all like the original ... The bad news is that **we can’t possibly perform the experiment.**”
- But we can using our frozen fossil record of viable bacteria.



Replay experiment

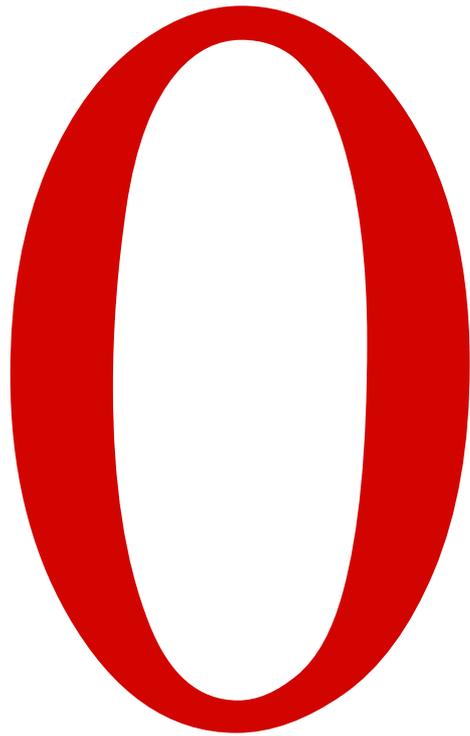
- Zack looked for mutations to Cit^+ in Cit^- clones sampled at various generations from population that evolved Cit^+ .
- Tested over **40 trillion** cells.



Zachary Blount

0

- Zack looked for mutations to Cit^+ in Cit^- clones sampled at various generations from population that evolved Cit^+ .
- Tested over **40 trillion** cells.
- ***Zero*** of 10 trillion ancestral cells gave Cit^+ mutant.



- Zack looked for mutations to Cit^+ in Cit^- clones sampled at various generations from population that evolved Cit^+ .
- Tested over **40 trillion** cells.
- *Zero* of 10 trillion ancestral cells gave Cit^+ mutant.
- *So, did Zack get **angry?***

- No, Zack's experiment was a *smashing success!*



- Neither the ancestor nor any early-generation clone produced a single Cit⁺ mutant. But clones from generation 20,000 onward produced 19 additional Cit⁺ mutants.

- Neither the ancestor nor any early-generation clone produced a single Cit⁺ mutant. But clones from generation 20,000 onward produced 19 additional Cit⁺ mutants.
- **By virtue of its particular path up Mt. Glucose, one lineage stumbled upon a route that led to Mt. Citrate.**

