Mathematical Theory and Scientific Understanding

“Make things as simple as possible, but no simpler.” – Albert Einstein

“All models are wrong, some are useful.” – George Box

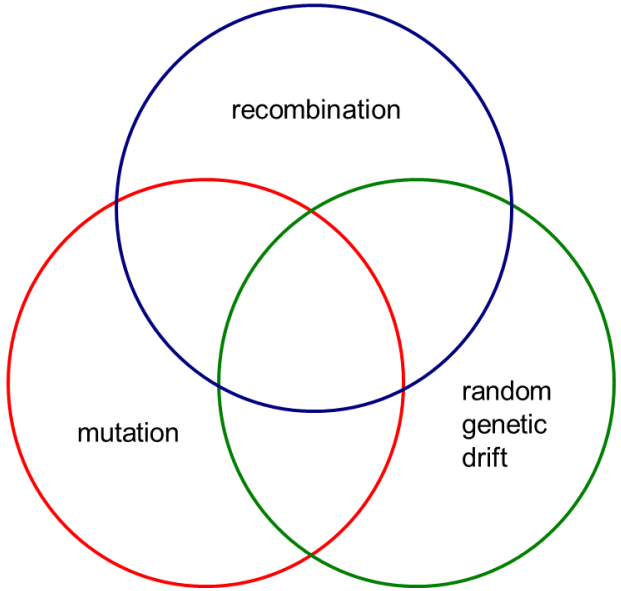
“No theory should fit all the facts because some of the facts are wrong.” – Niels Bohr

“What I cannot build, I cannot understand.” – Richard Feynman

Population Genetics and Evolutionary Hypotheses

- The general principles of population genetics are so well established that the credibility of any proposed scenario for an evolutionary observation must remain in doubt until it has can be shown to be theoretically feasible.

The Population-genetic Environment



- The types of evolution that can occur within a species depend critically on the mutation rate, effective population size, and degree of linkage in the genome – these vary by orders of magnitude among species.

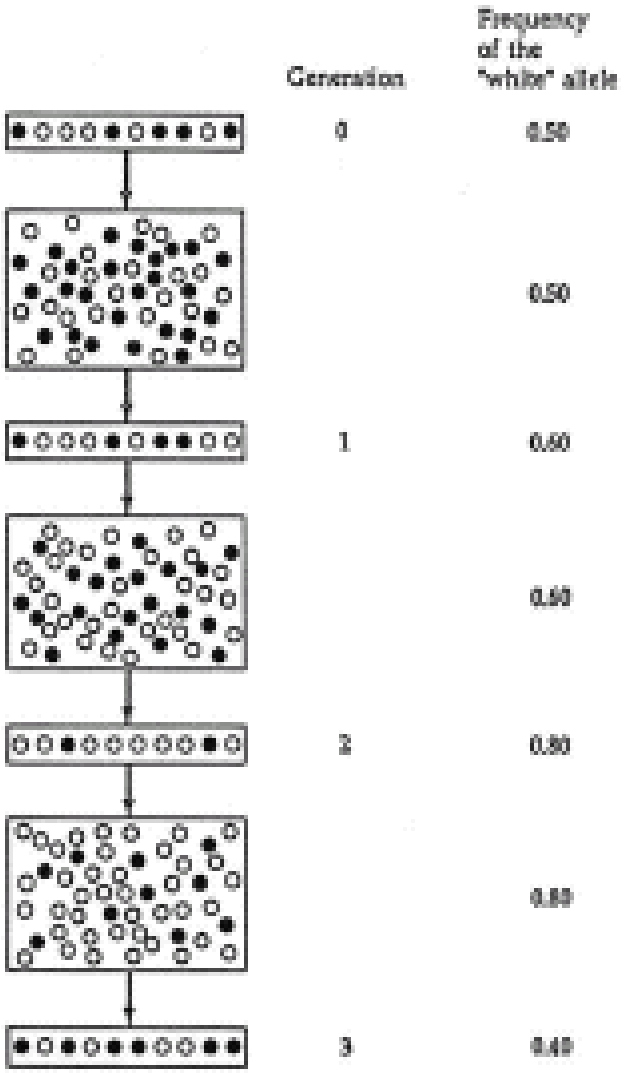
  As a consequence, there are certain kinds of evolution that are difficult, if not impossible to achieve in multicellular species with relatively small population sizes, but readily attainable in microbes, and vice versa.

- Biologists almost always assume that every feature of the organism has been molded by natural selection and nothing else.

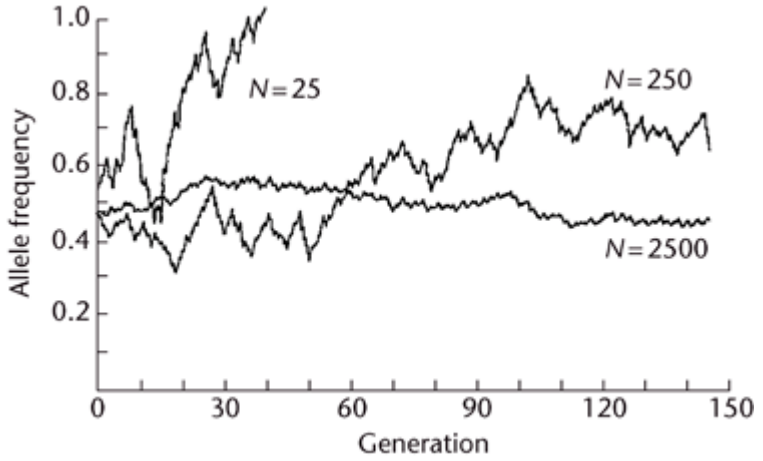
- It remains unclear as to whether natural selection is a necessary or sufficient condition for the origin of cellular complexity.

Random Genetic Drift at a Neutral Locus is Inversely Proportional to the Effective Population Size, N_e

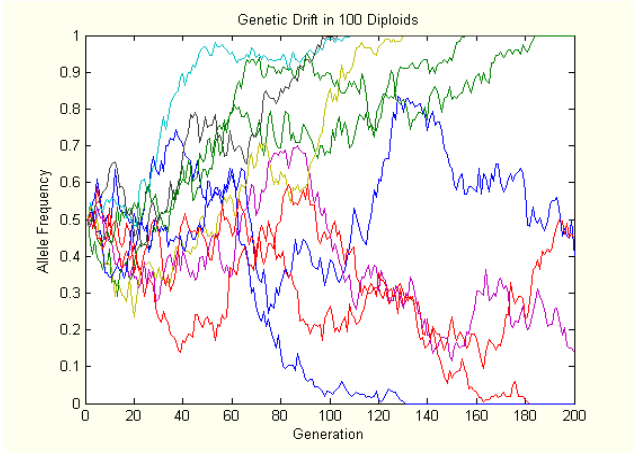
1) Sampling of finite numbers of gametes results in allele-frequency fluctuations.

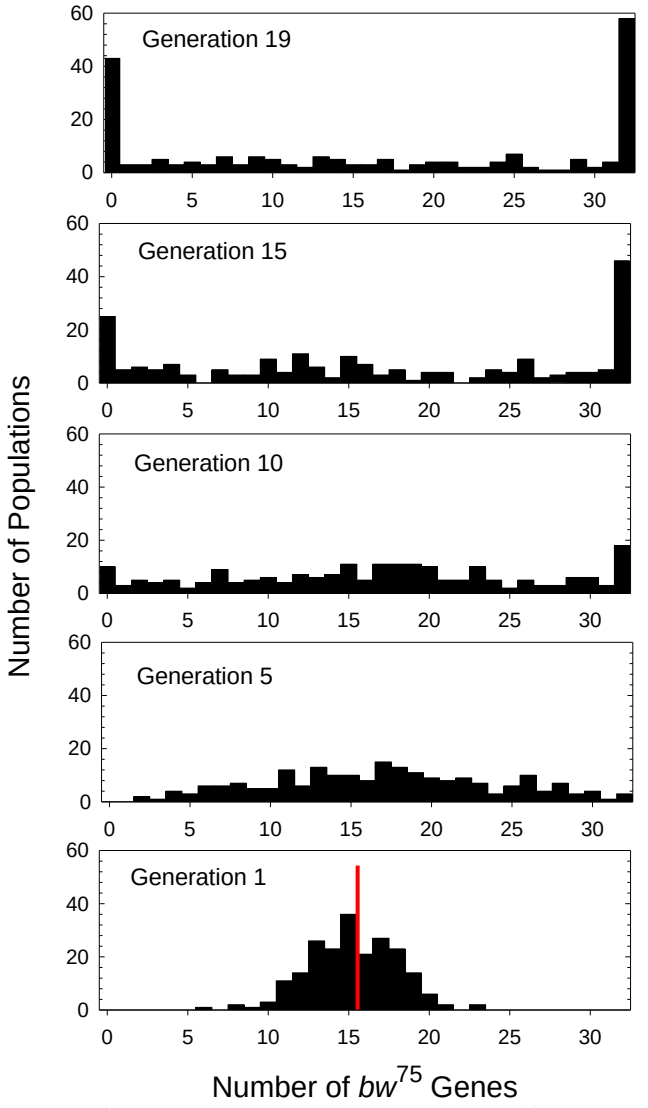


2) The magnitude of fluctuations declines with population size.



3) Each evolutionary trajectory is unique.





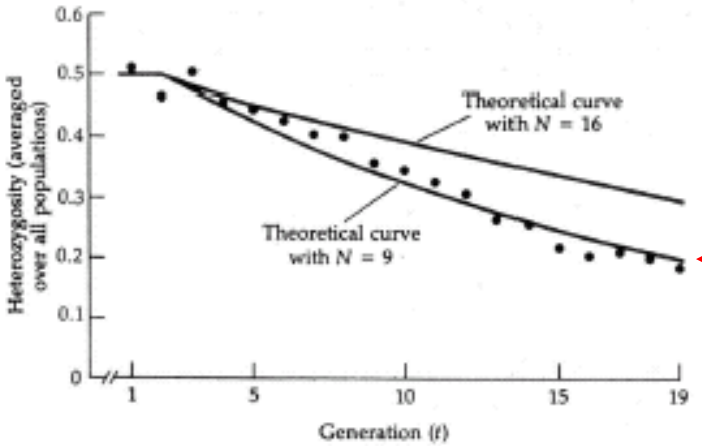
Experimental demonstration of genetic drift. The number of copies of an allele, bw^{75}, in each of many replicate populations of *Drosophila melanogaster* maintained in the laboratory at 16 flies for 19 generations. In each population, the frequency of the allele fluctuated, so the variation in gene frequency increased. After about 12 generations all gene frequency classes have become about equally frequent. (From Buri 1956)

Buri's Big Drift Experiment

Heterozygosity after *t* generations at population size *N*:

$$H_t = H_0 \times [1 - (1/2N)]^t$$

$$\approx H_0 \times e^{-t/(2N)}$$

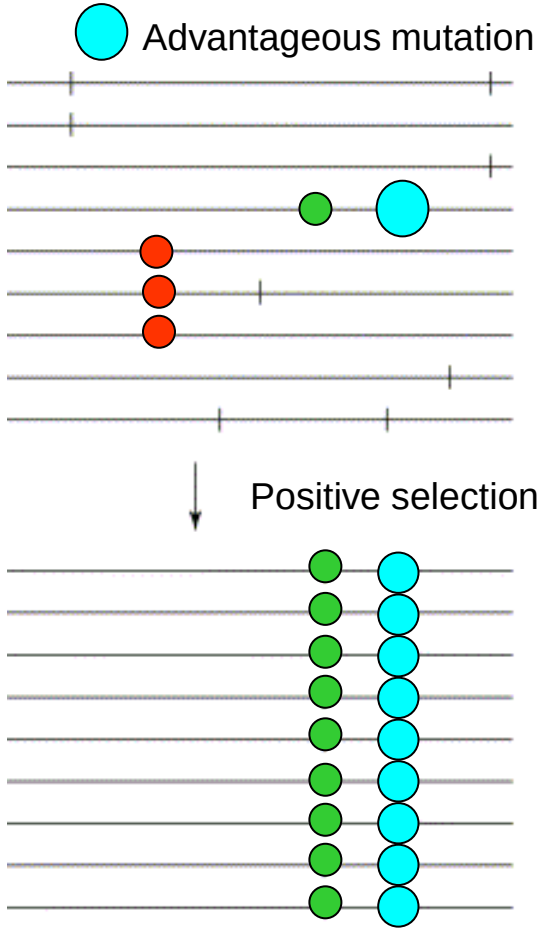


population behaves genetically as though it effectively contains N_e = 9 individuals

Most Demographic Deviations From the Standard Model Cause N_e to be << the Census Number

- Variation in gamete production due to selection.
- Population subdivision and variation in productivity among subpopulations (spatial ecological variation).
- Uneven sex ratio.
- Temporal variation in population size.

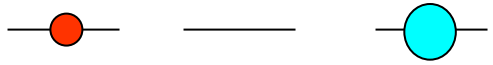
Genetic Hitch-hiking Via Selective Sweeps Depresses N_e Below the Actual Census Size



background deleterious mutation fixed in the population

background beneficial mutation lost from the population

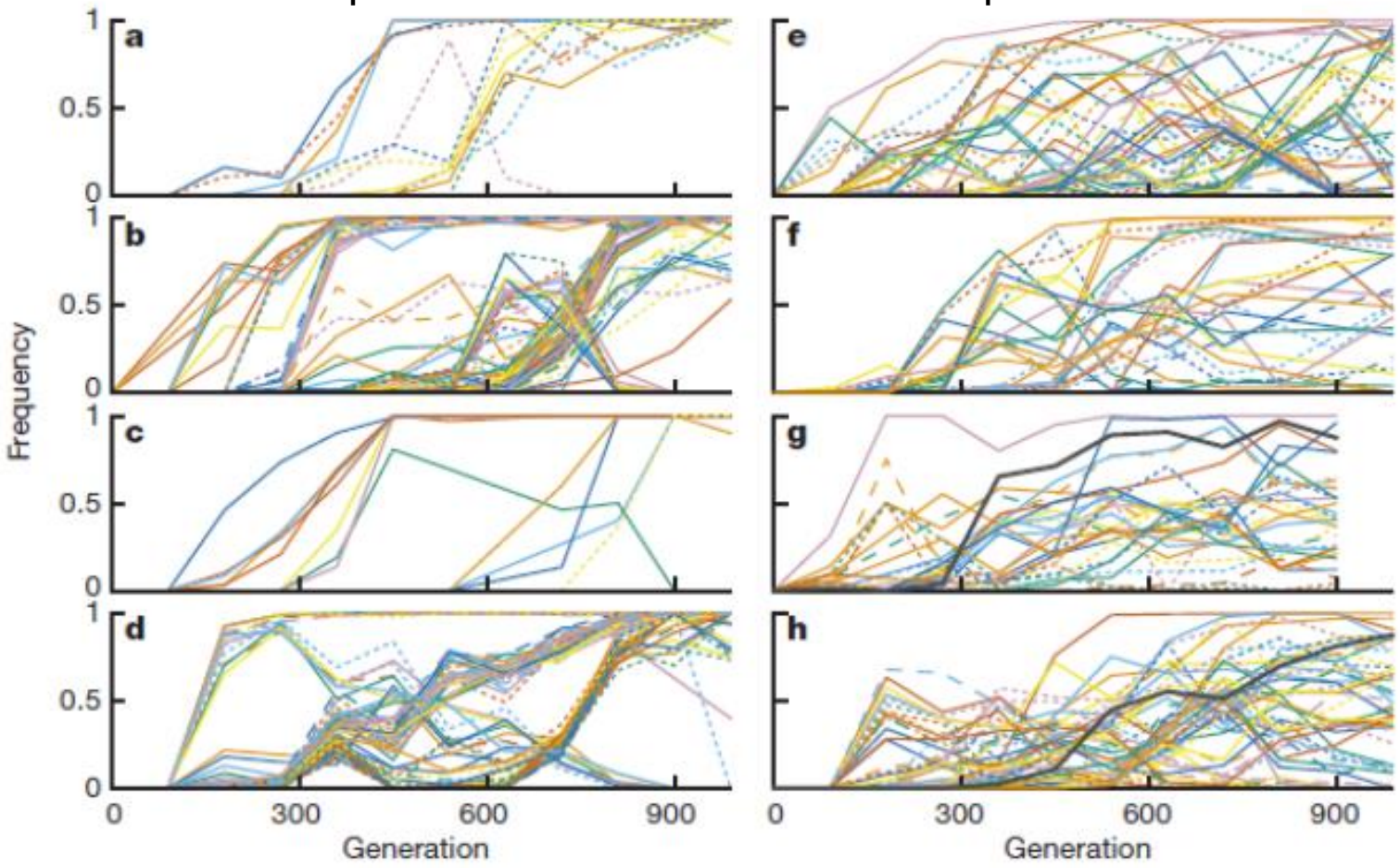
With free recombination, the outcome would be:



Allele-Frequency Trajectories for Mutations in Replicate Experimental Yeast Populations

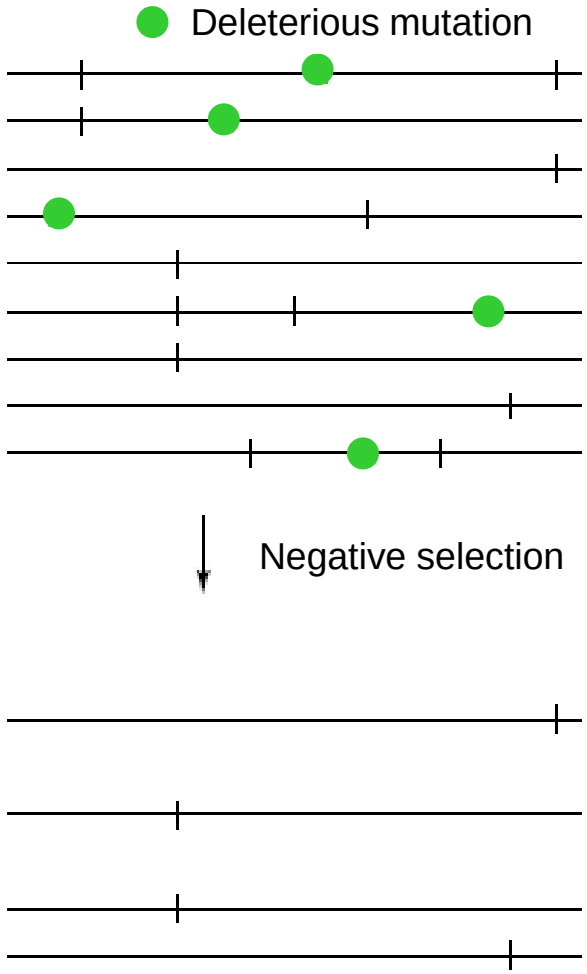
Asexual Populations:

Sexual Populations:



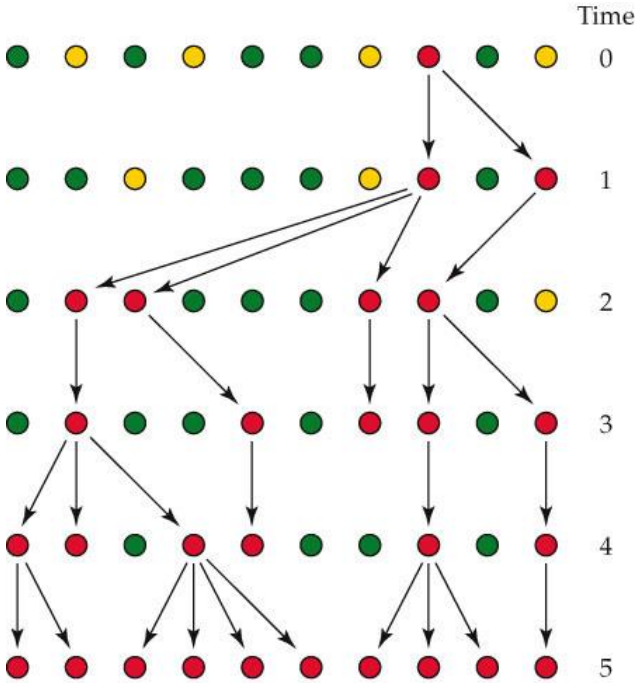
McDonald et al. (2016)

Selection Against the Constant Background Rain of Deleterious Mutations Further Depresses N_e



The Concept of Effective Neutrality

- Even non-neutral mutations will behave in an effectively neutral fashion provided the population size is sufficiently small that the strength of selection is overwhelmed by the stochastic fluctuations induced by genetic drift.



Selective advantage (or disadvantage) of mutant allele = s

Fitnesses – A: 1 a: 1 + s

Power of random genetic drift in a haploid population = $1 / N_e$

Tomoko Ohta

- Provided $s << 1/ (2N_e)$, which means $2N_es <<1$, selection is rendered ineffective by the noise from random genetic drift.

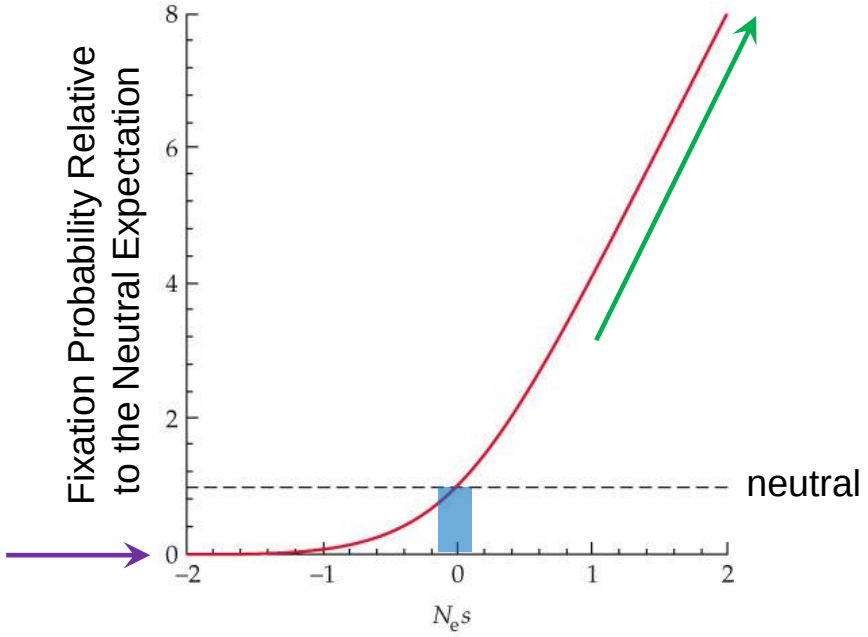
Probability of Fixation of a New Mutation

$$\phi_f(1/2N) \simeq \frac{1 - e^{-2(N_e/N)s}}{1 - e^{-4N_e s}}$$

N = absolute population size

N_e = effective population size

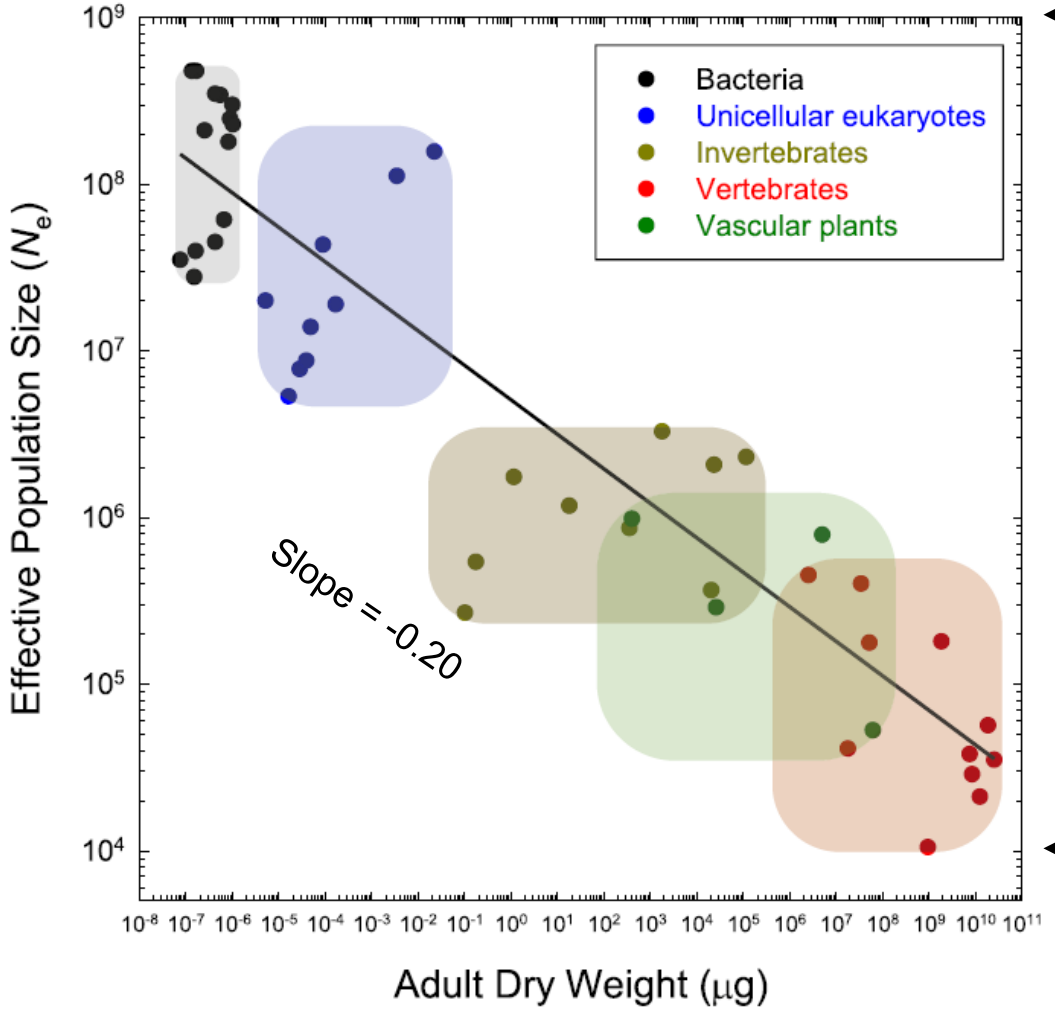
s = selective advantage / disadvantage / generation ($0 \leq |s| \leq 1$)



Motoo Kimura

- If the new mutation is effectively neutral, $|N_e s| << 1$, the probability of fixation ≈ the initial frequency, 1/(2N).
- If the new mutation is strongly deleterious, $N_e s << 0$, the probability of fixation ≈ 0.
- If the new mutation is strongly advantageous, $N_e s >> 1$, the probability of fixation ≈ $2s(N_e/N)$.

Negative Scaling of N_e with Organism Size Defines the Range of Mutations Discernible by Selection



Maximum $N_e = 10^9$

All mutations with absolute effects $>10^{-9}$ are visible to selection.

For random genetic drift not to play a role in evolution, all mutations must have fitness effects $<10^{-9}$ and / or $>10^{-4}$, with nothing in between.

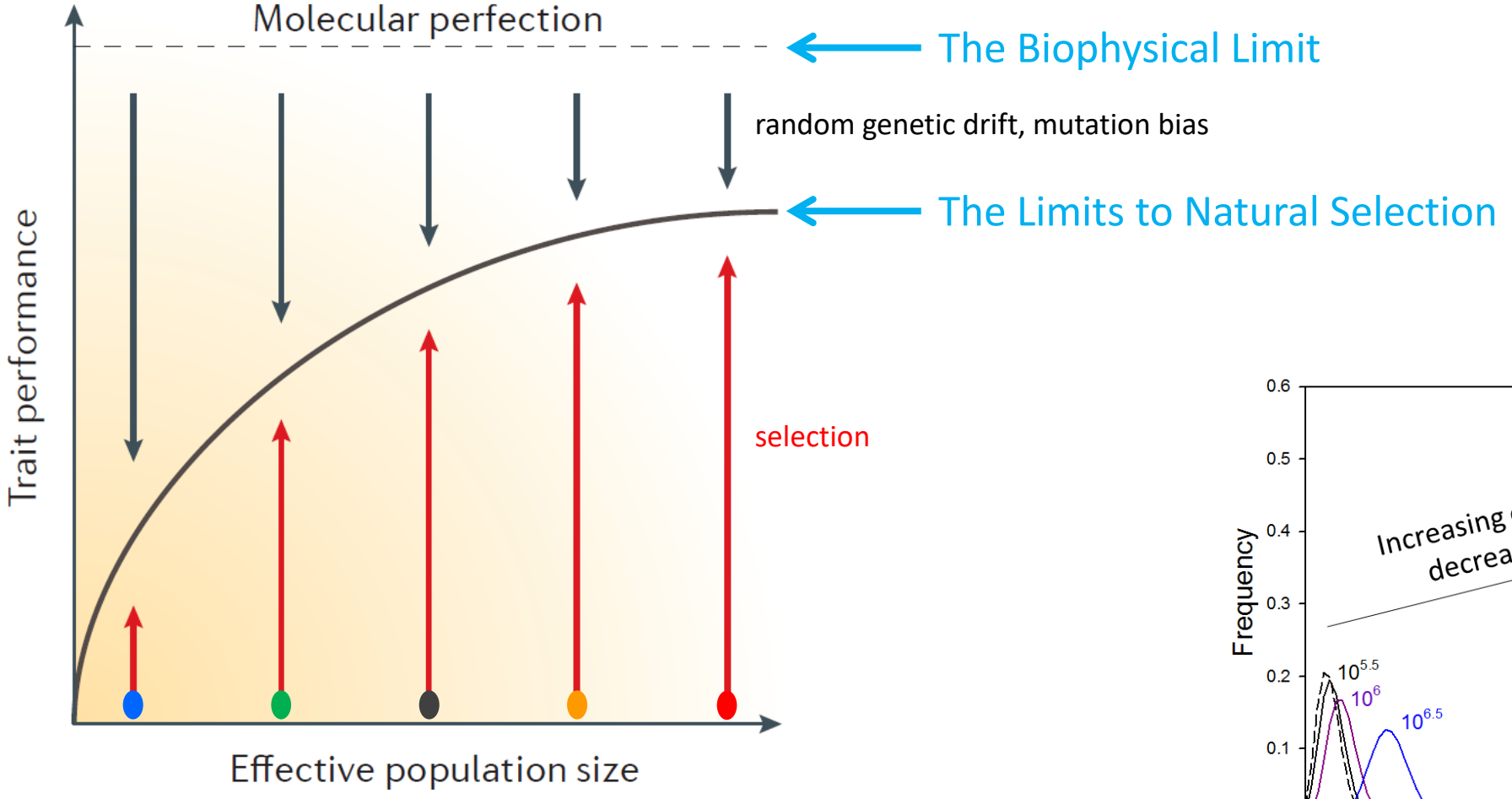
$N_e = 10^4$

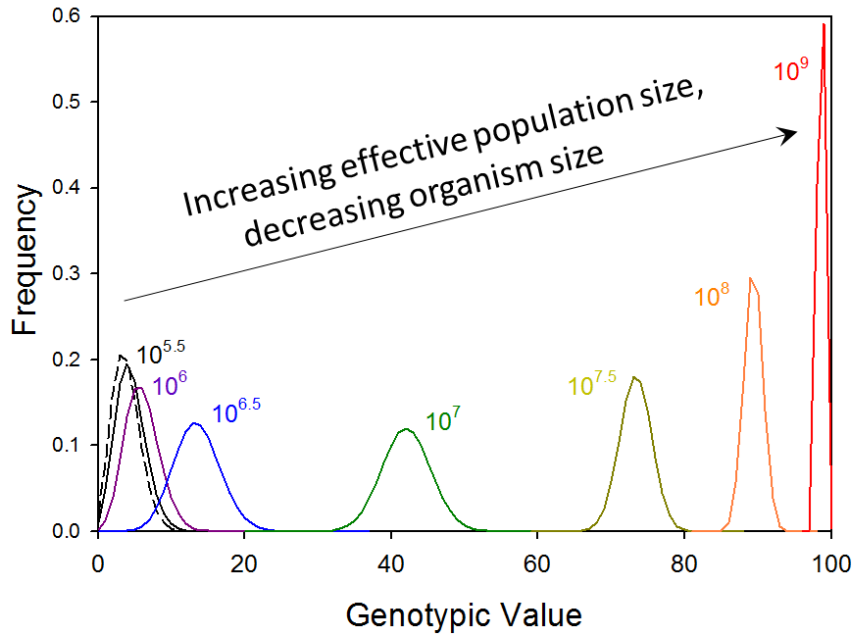
All deleterious mutations with effects $<10^{-4}$ are free to fix;

mutations with advantages $<10^{-4}$ are invisible to selection.

Lynch and Trickovic 2020

The Drift-Barrier Hypothesis





Drift Barriers in Biology

- Evolution of senescence.
- Marginal stability of protein folding and binding-interface strength in multimeric enzymes.
- Reduced enzyme catalytic capacities relative to the diffusion limit.

- Increase in mutation rates with decreased effective population sizes.
- Passive expansion of mutationally/energetically harmful genomic DNA with population-size reduction.
- Reduction in maximum growth rate with increasing eukaryotic cell / body size.

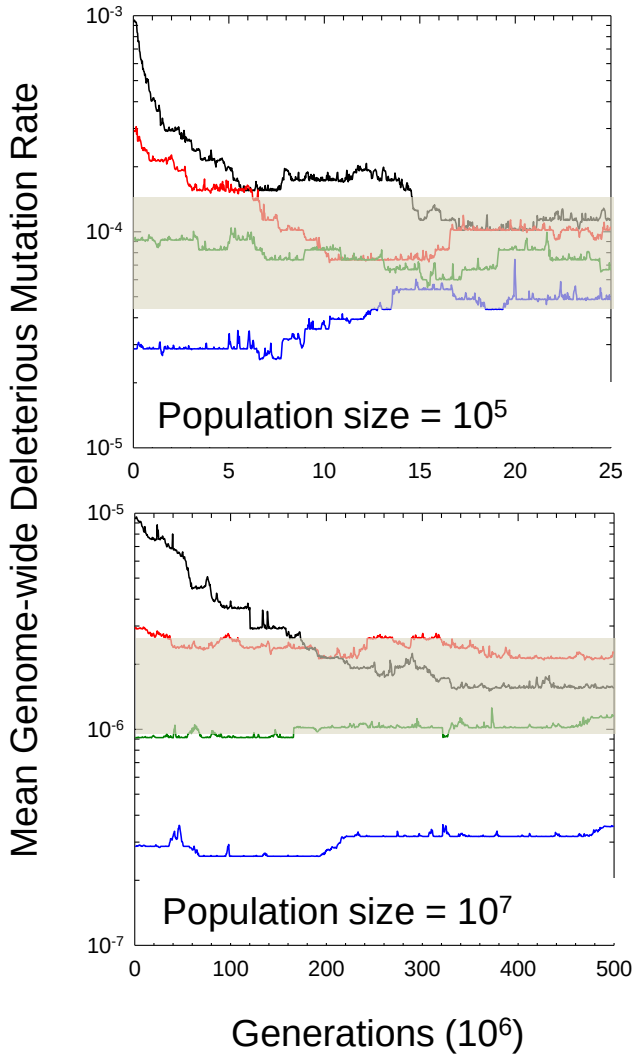
Evolution of Mutation Rates

- Because most mutations are deleterious, we expect natural selection to relentlessly drive mutation rates in a downward direction.

- Mutation rates evolve to be inversely proportional to N_e, ranging from 10^{-11} / nucleotide site / cell division in some microbes to 10^{-8} in vertebrates, in accordance with the drift-barrier hypothesis.

- Infrequently used DNA polymerases have highly elevated error rates, in accordance with the drift-barrier hypothesis.

Quasi-Equilibrium Mutation Rates Resulting From Deleterious-Mutation Load

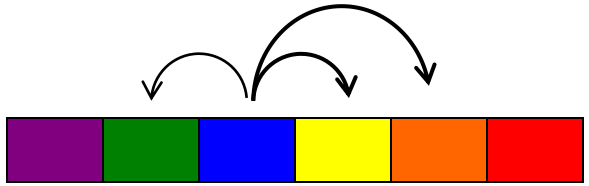


Effective selection for antimutators

DRIFT BARRIER

Biased production of mutators

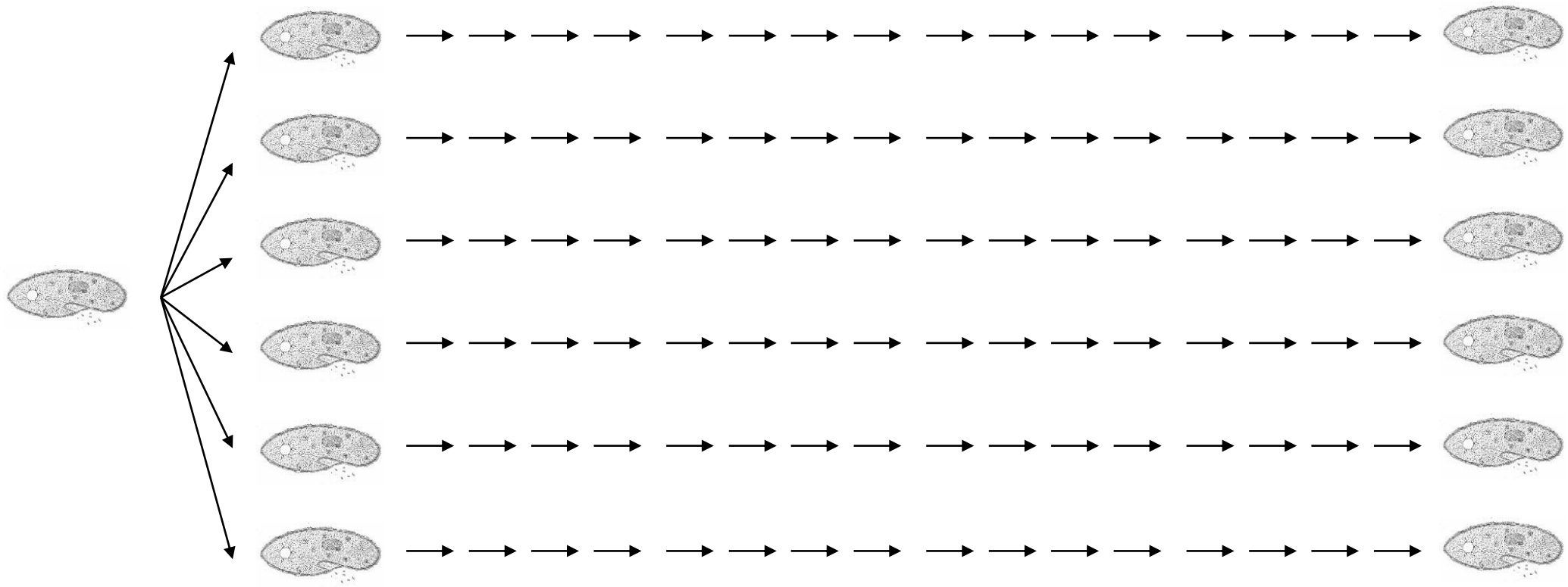
- Equilibrium mutation rate is expected to be inversely proportional to the effective population size.



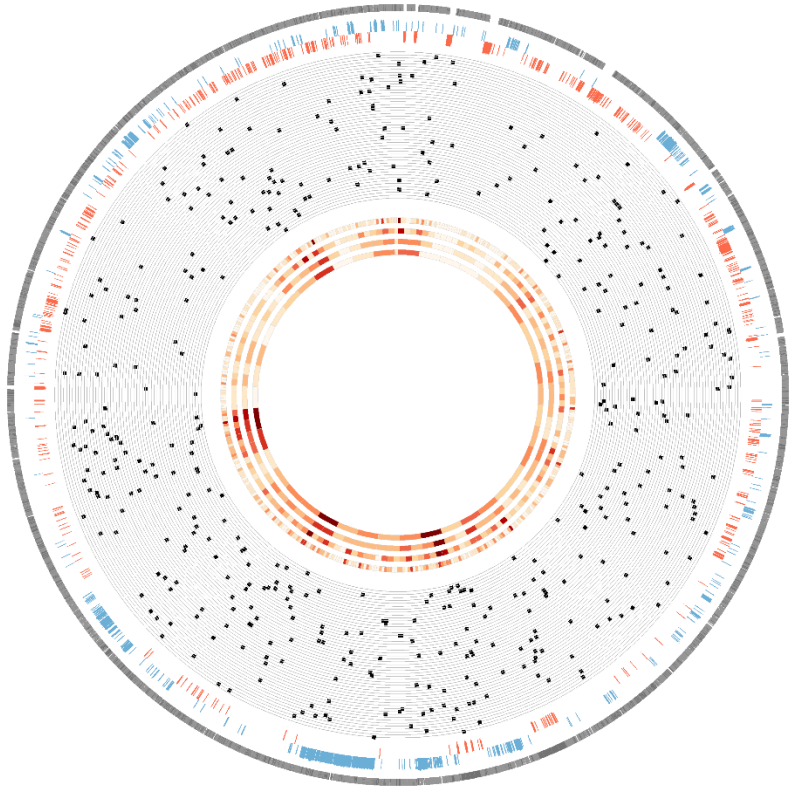
Mutation-rate classes

Analysis of Genome Stability with a Mutation-accumulation Experiment:

- Starting with a single stem cell, sublines are maintained by single-progeny descent, preventing selection from removing spontaneous mutations.
- Continue for thousands of cell divisions.
- Quantify and characterize mutations by whole-genome sequencing ~50 lines.



Mutation in Small vs. Large Genomes

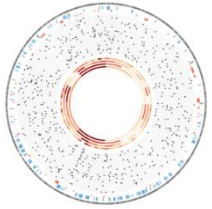


Index:

Outer Rings
- Gene Density
- High G/C Region
- High A/T Region

Intermediate Rings
- Mutations

Inner Rings
- Mutation Density
  - Window Size (1k, 5k, 25k, 100k)



Bacillus subtilis 3610

Genome size: 4,214,598 bp
GC content: 43.5%

50 lines - 450 mutations - 5000 generations

Mutation Rate : 3.27×10^{-10}/site/gen.

Mesoplasma florum L1

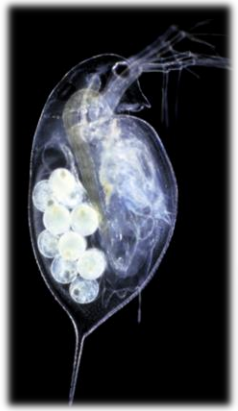
Genome size: 793,224 bp
GC content: 27.0%

50 lines – 599 mutations - 2000 generations

Mutation Rate : 1.14×10^{-8}/site/gen.

Mutation-accumulation Studies Across the Tree of Life

Group	Species	Genome Size (Mb)	G/C %
Bacteria:			
Acidobacteria	*Acidobacterium capsulatum*	4.1	61.0
Actinobacteria	*Kineococcus radiotolerans*	5.0	74.2
Actinobacteria	*Mycobacterium smegmatis*	7.2	65.2
Actinobacteria	*Mycobacterium* sp.	7.2	65.2
Alpha-proteobacteria	*Agrobacterium tumefaciens*	5.7	59.0
Alpha-proteobacteria	*Caulobacter crescentus*	4.0	67.2
Alpha-proteobacteria	*Rhodobacter sphaeroides*	4.5	68.2
Beta-proteobacteria	*Burkolderia cenocepacia*	7.8	66.8
Beta-proteobacteria	*Janthinobacterium* sp.	6.0	61.1
Gamma-proteobacteria	*Photorhabdus luminescens*	5.7	42.8
Gamma-proteobacteria	*Pseudomonas fluorescens**	7.1	63.3
Gamma-proteobacteria	*Shewanella putrefaciens*	4.7	44.5
Gamma-proteobacteria	*Teredinibacter turnerae*	5.2	50.9
Gamma-proteobacteria	*Vibrio cholerae**	4.1	47.5
Gamma-proteobacteria	*Vibrio fischeri**	4.3	38.3
Cyanobacteria	*Synechococcus elongatus*	2.7	55.5
Deino-Thermus	*Deinococcus radiodurans**	3.2	66.6
Firmicute	*Bacillus subtilis**	4.2	43.5
Firmicute	*Staphylococcus epidermidis*	2.6	32.0
Flavobacteria	*Flavobacterium* sp.	6.1	34.1
Lactobacillale	*Lactobacillus* sp.	2.9	46.4
Planctomycete	*Gemmata obscuriglobus*	9.2	67.2
Tenericute	*Mesoplasma florum*	0.8	27.0
Archaea:			
Euryarchaeota	*Haloferax volcanii*	4.0	65.5



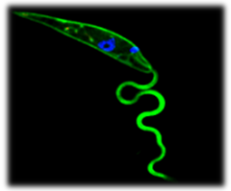
Daphnia

Drosophila

Caenorhabditis

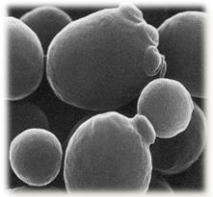
Arabidopsis

Chlamydomonas

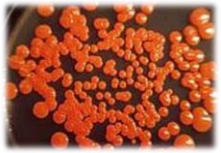


Leishmania

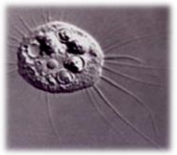
Dictyostelium



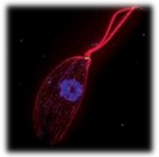
Saccharomyces



Rhodotorula



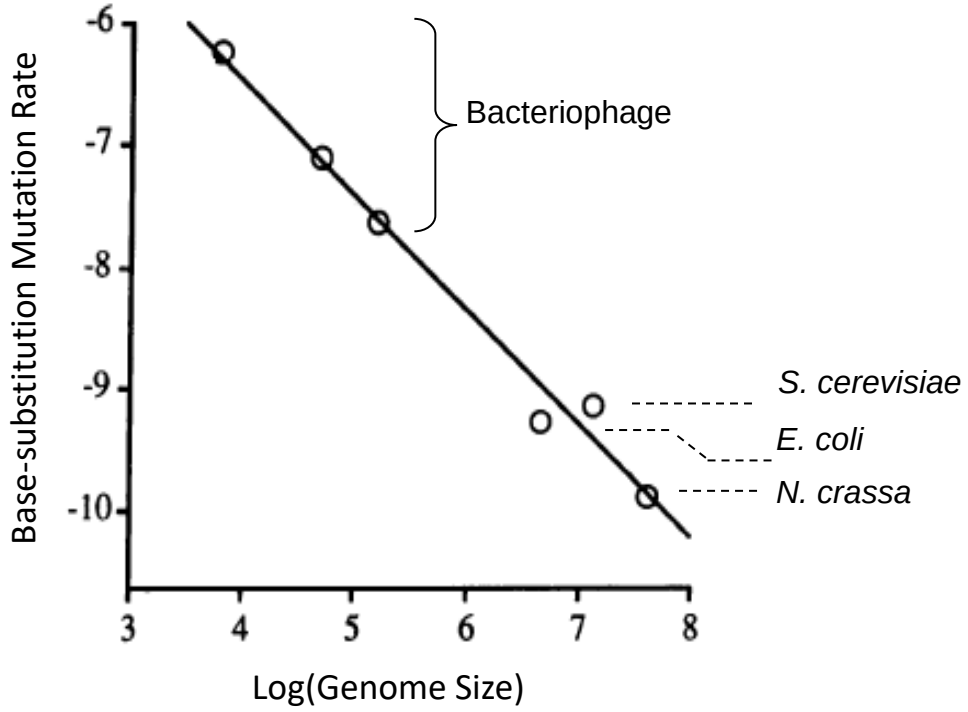
Ichthyosporean



Naegleria

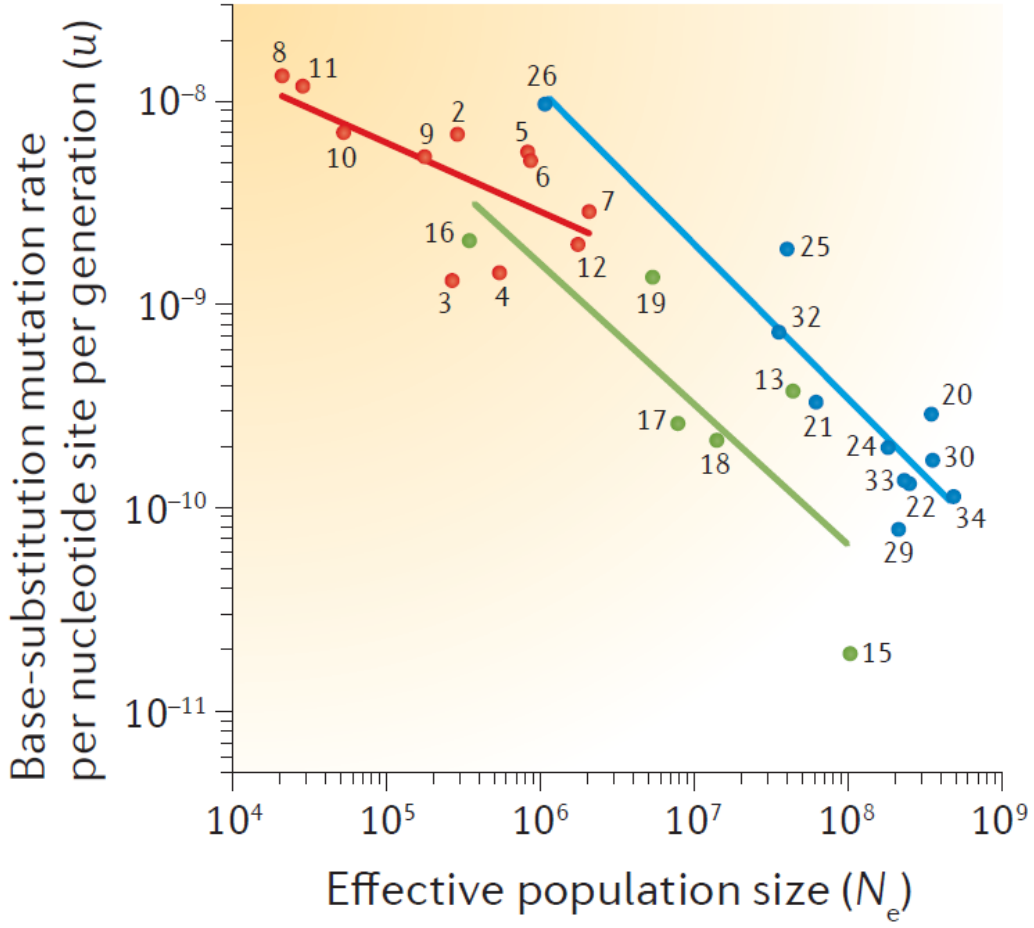
Paramecium

Drake's (1991) Law for Mutation-Rate Evolution Revisited: constant number of mutations / genome = 0.003.



"Because this rate is uniform in such diverse organisms, it is likely to be determined by deep general forces."

Evaluation of the Drift-Barrier Hypothesis

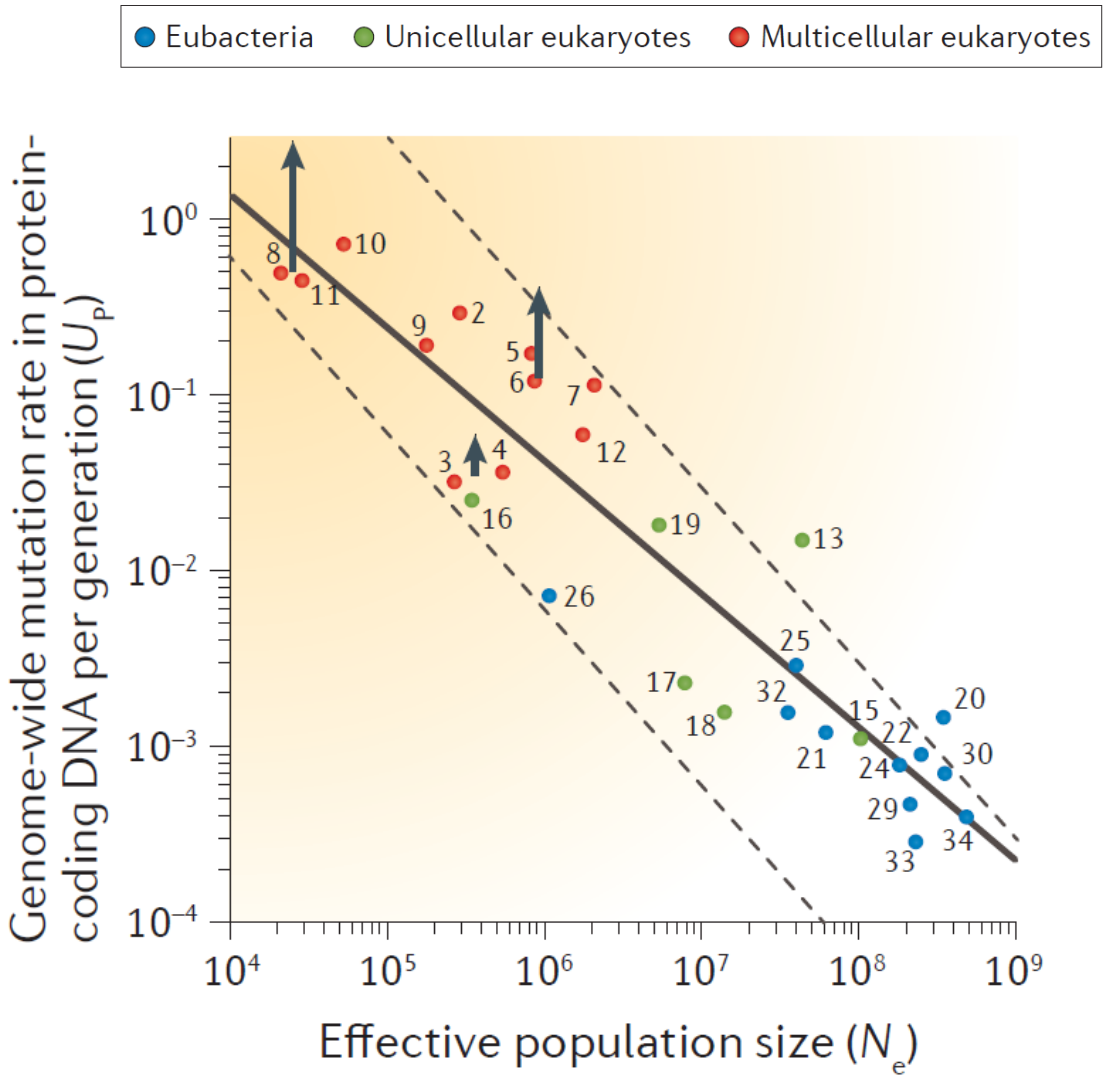


- The mutation rate per nucleotide site scales inversely with the effective population size.

- For a given N_e, **unicellular eukaryotes** have lower mutation rates per nucleotide site than **bacteria** because there are more functionally significant genomic sites, and hence stronger selection to maximize replication fidelity.

Eubacteria Unicellular eukaryotes Multicellular eukaryotes

Inverse Scaling Between the Genome-wide Deleterious Mutation Rate and the Effective Population Size of a Species

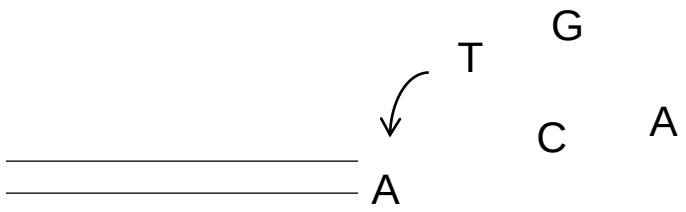


- The only trait for which we have a comprehensive theory for the evolution of mean phenotypes across the Tree of Life in mechanistic terms.

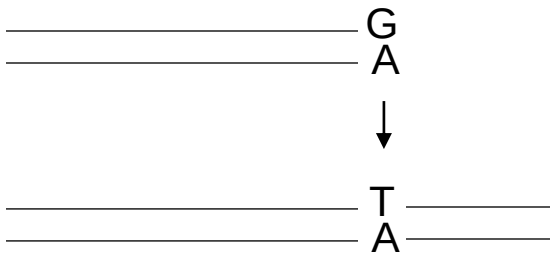
- This pattern goes against the grain of biophysical hypotheses, e.g., speed vs. efficiency, as the most rapidly growing species are the least error-prone.

The Three Molecular Lines of Defense Against Mutation

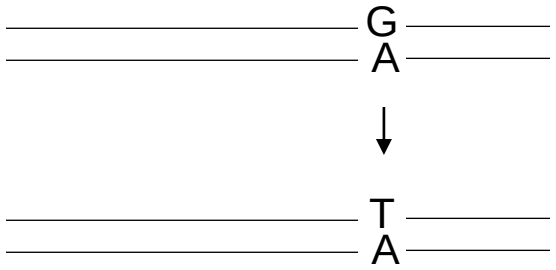
1) Polymerase base-incorporation fidelity:



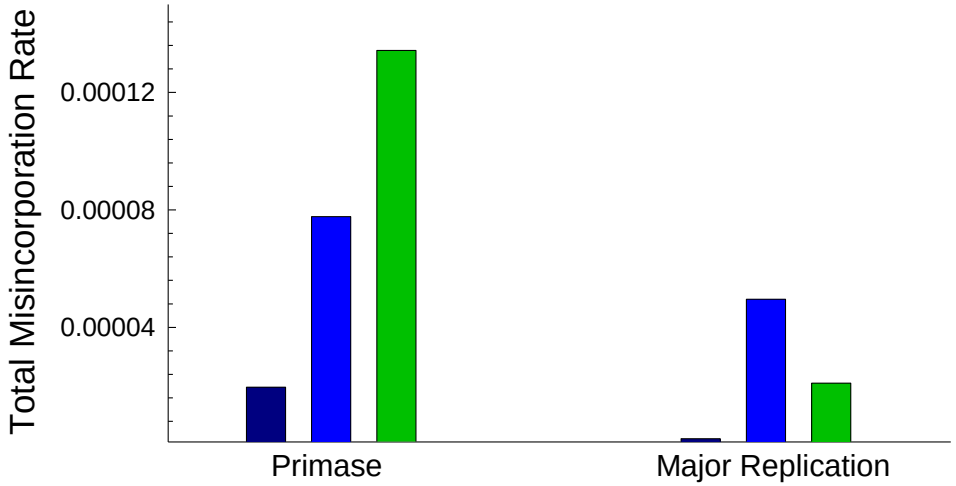
2) Polymerase proofreading:

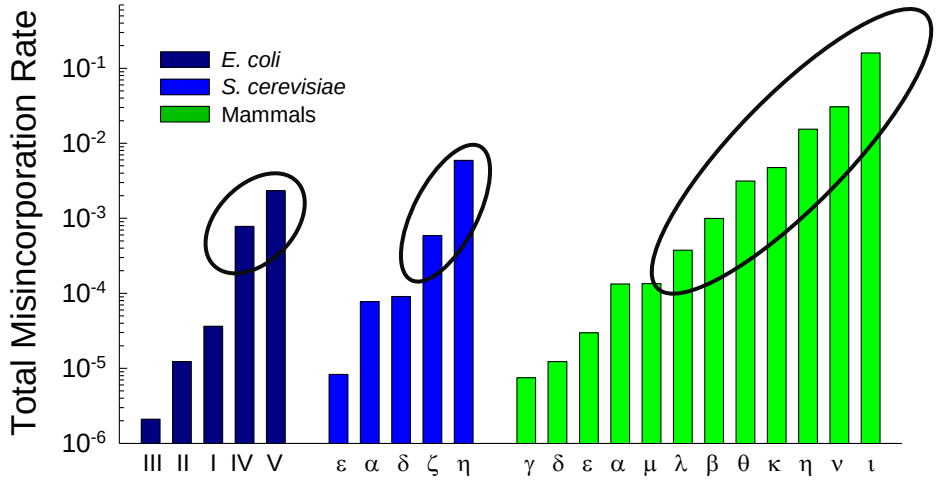


3) Post-replicative mismatch repair:



Polymerase Error Rates Are Magnified in Enzymes Involved in Fewer Nucleotide Transactions





Polymerases used in DNA repair are highly error prone, consistent with the drift hypothesis:

enzymes involved in fewer nucleotide transactions experience less selection for fidelity.

Evolution of Recombination Rates

- Evolutionary consequences of recombination:

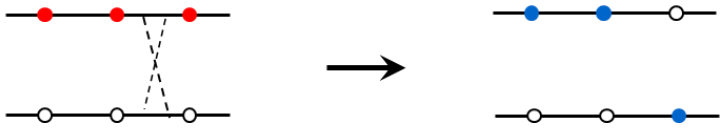
  1) reduces background-selection and hitch-hiking effects, allowing for more efficient natural selection on individual sites;

  2) can create novel genotypes by merging mutations from different genomes, reducing the waiting time for the arrival of multiple mutations in single individuals;

  3) destroys favorable combinations of mutations prior to fixation.

Recombination can facilitate the arrival of an adaptive combination,

but it also inhibits the fixation of the adaptive allele,



Inverse Scaling of the Recombination Rate / Physical Distance and Genome Size is a Natural Outcome of the "One Crossover / Chromosome Arm" Rule

Decreasing population size →

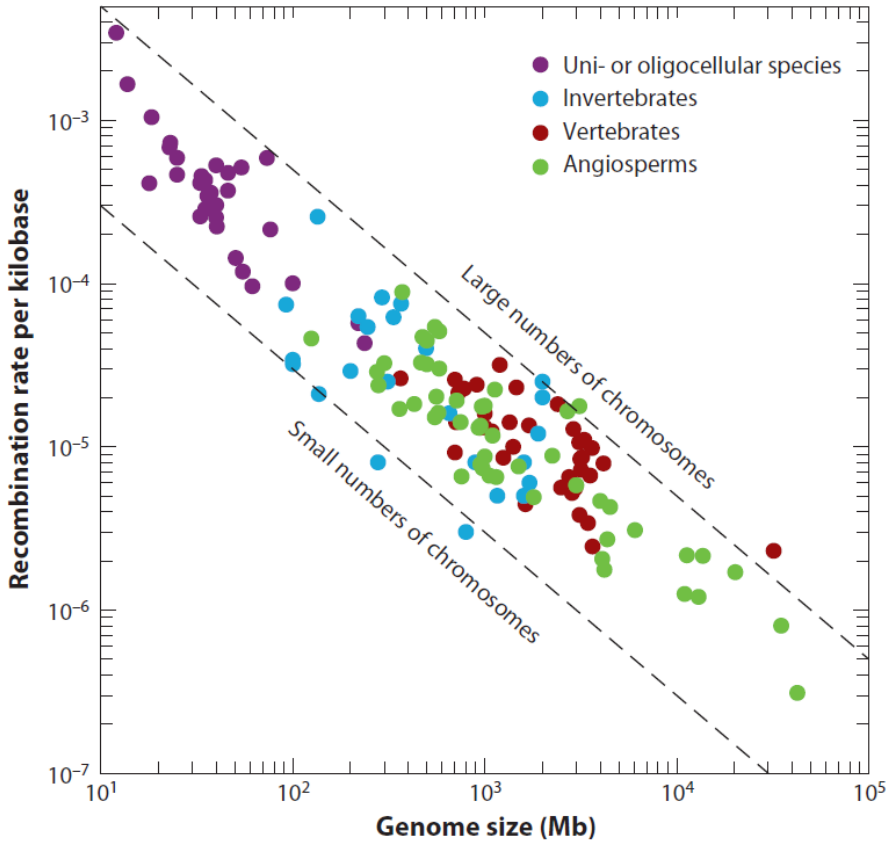


Figure 1

A compilation of estimates of the average amount of recombination per unit of physical distance in eukaryotic genomes, derived from 137 meiotic genetic maps. The diagonal lines have slopes of −1.

- Virtually all variation in the recombination rate among species is explained by variation in genome size and chromosome number.
- Large genomes (in species with relatively small N_e) have low rates of recombination / physical distance.
- This reflects a near-absolute constraint of the physical aspects of meiosis.

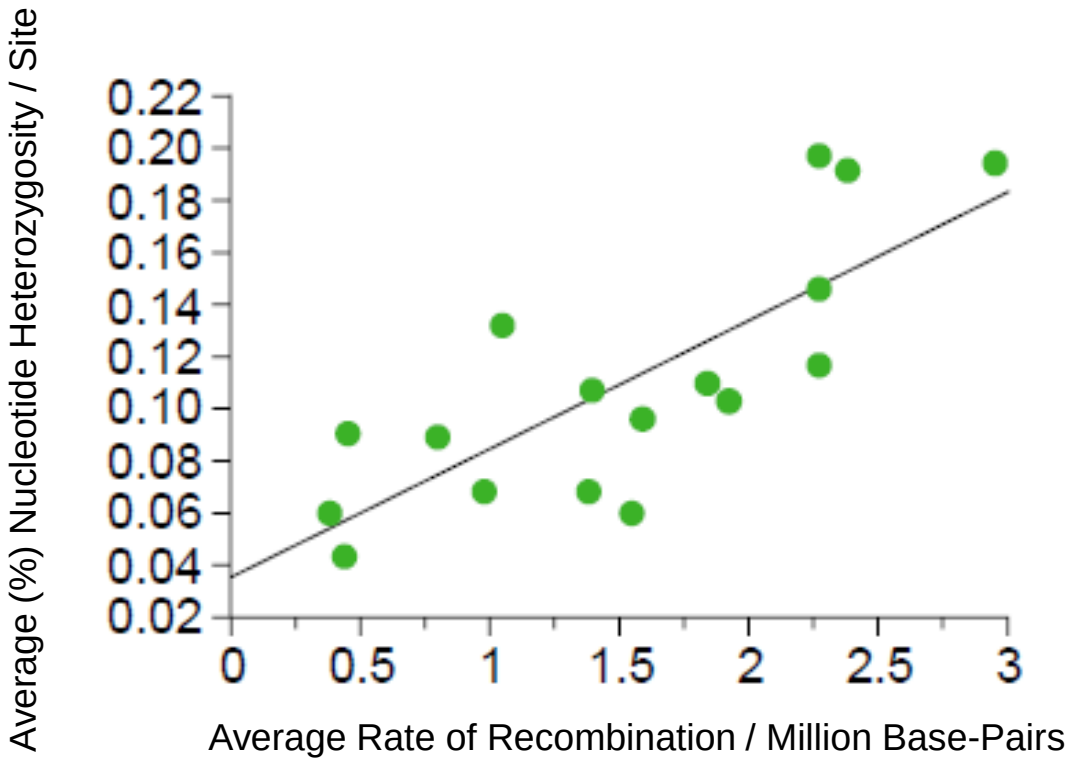
Relative Magnitudes of Recombination (*c*) and Mutation (*u*) Rates Per Nucleotide Site

- Many Bacteria are just as recombinationally active as Eukaryotes.

Estimates of *c/u,* ratio of the recombination rate to the mutation rate per base pair.

Animals:		
Homo sapiens	0.6	Ptak et al. (2004)
Chorthippus parallelus	2.5	Ibrahim et al. (2002)
Drososphila sps.	3.8	Hey and Wakeley (1997)
		Machado et al. (2002)
Land plants:		
Arabidopsis sps.	0.7	Wright et al. (2003)
Brassica nigra	0.3	Lagercrantz et al. (2002)
Cryptomeria japonica	3.0	Kado et al. (2003)
Pinus taeda	0.3	Brown et al. (2004)
Zea mays	1.6	Tenaillon et al. (2004)
Bacteria:		
Neisseria gonorrheae	1.0	Posada et al. (2000)
Neisseria meningitidis	4.8	Feil et al. (2001)
Pseudomonas syringae	0.3	Sarkar and Guttman (2004)
Staphylococcus aureus	6.5	Feil et al. (2001)
Streptococcus pneumoniae	8.9	Feil et al. (2001)

Reduced Levels of Variation in Regions of Low Recombination Tell Us That Linkage Magnifies the Power of Genetic Drift



Data for humans, from Nachman (2001, Trends in Genetics)

What Occurs in Evolution is Dictated by What Natural Selection Can and Cannot Do

- The population-genetic environment evolves – increases in organism size induce declines in population size and rates of recombination, leading to an increase in the power of drift, which in turn encourages the evolution of increased mutation rates.

- These covarying aspects of the population-genetic environment modify the ways in which evolution by natural selection can proceed in different phylogenetic lineages.

    - Natural selection's search for perfection is limited by the granularity of mutational effects, rates of mutation and recombination, and the power of random genetic drift.

- Because mutations with selective effects $<< 1/N_e$ are overwhelmed by drift, small organisms with higher N_e are capable of utilizing a wider range of mutational effects in adaptive evolution. Larger organisms, with correspondingly smaller N_e, have a reduced capacity for evolutionary fine-tuning and hence are constrained to more coarse-grained evolution.

- Owing to the limited reach of natural selection, at all levels of biological organization, we expect mean phenotypes to scale with N_e, such that organisms under identical selection pressures may nonetheless undergo predictable patterns of divergence.