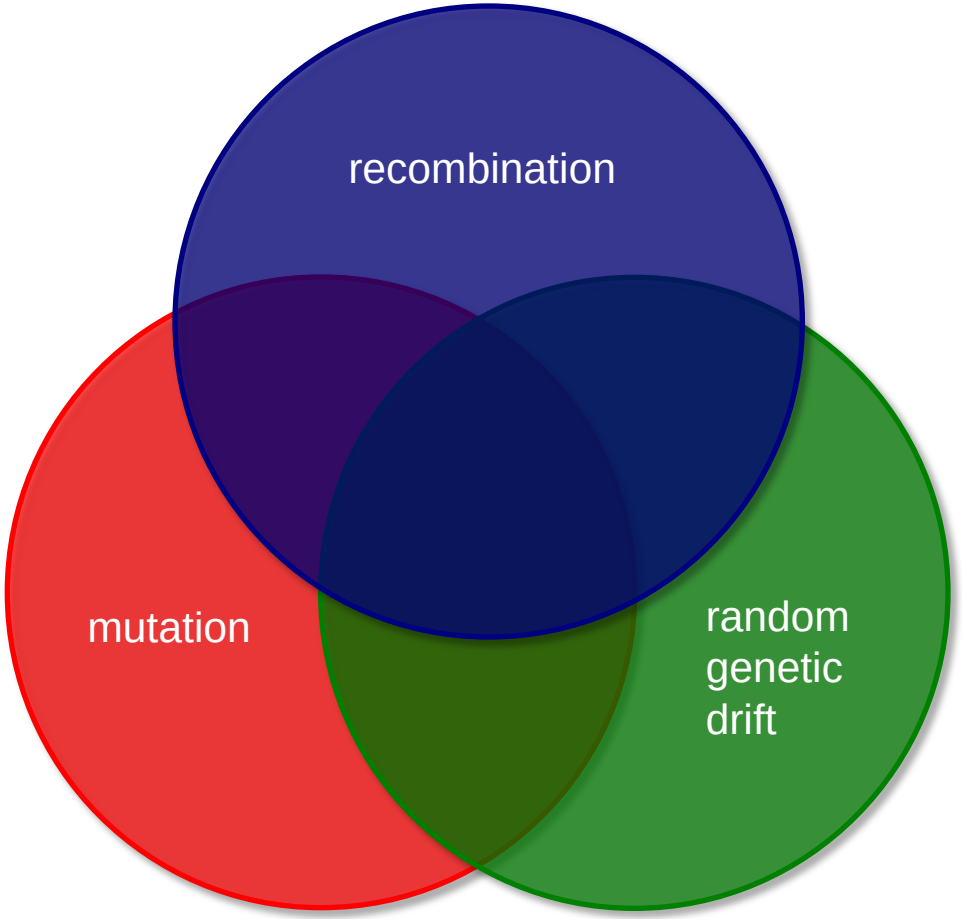
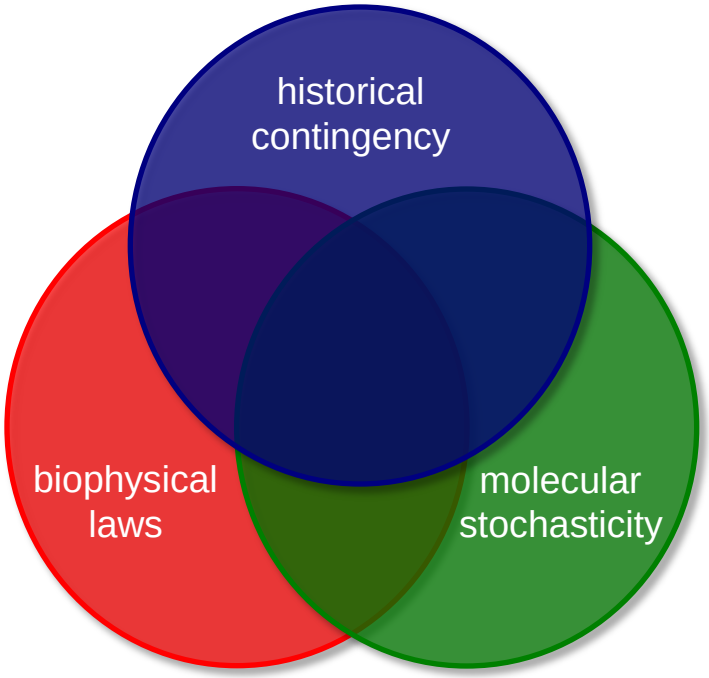
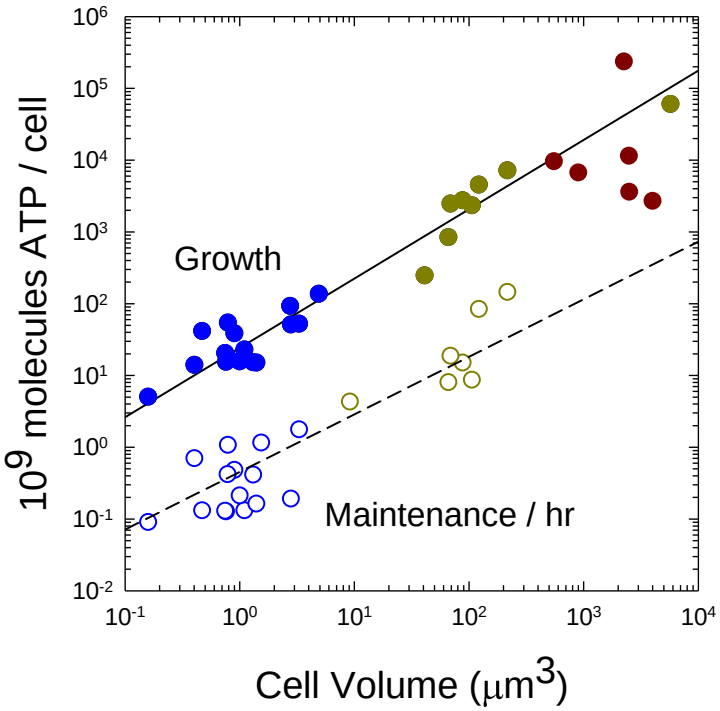
The Population-genetic Environment



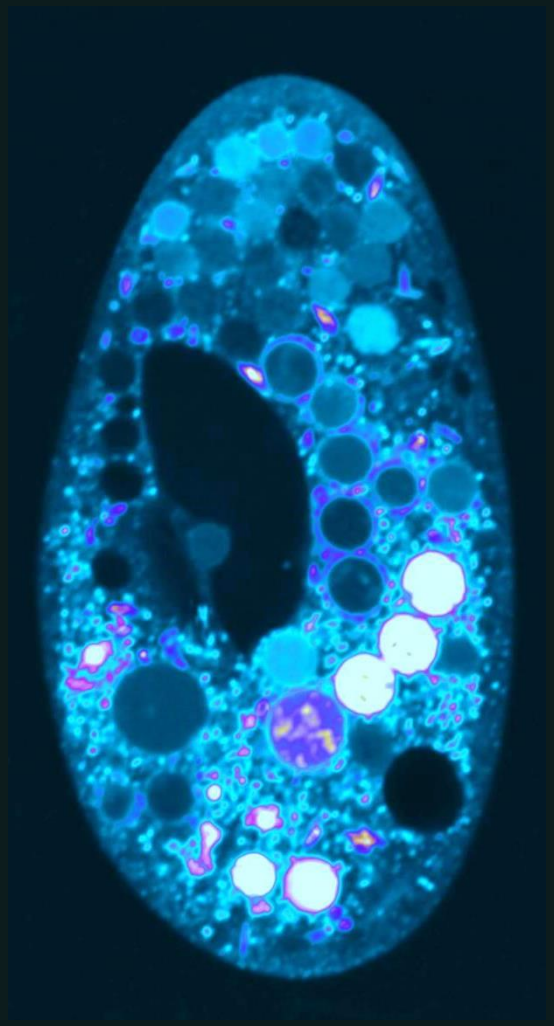
The Cellular Environment





- It takes ~24 x 10^9 ATP units to build 1 μm^3 of cell volume, across the Tree of Life.
- What are the energetic costs / gains of each subcellular embellishment?

Evolution of Subcellular Features



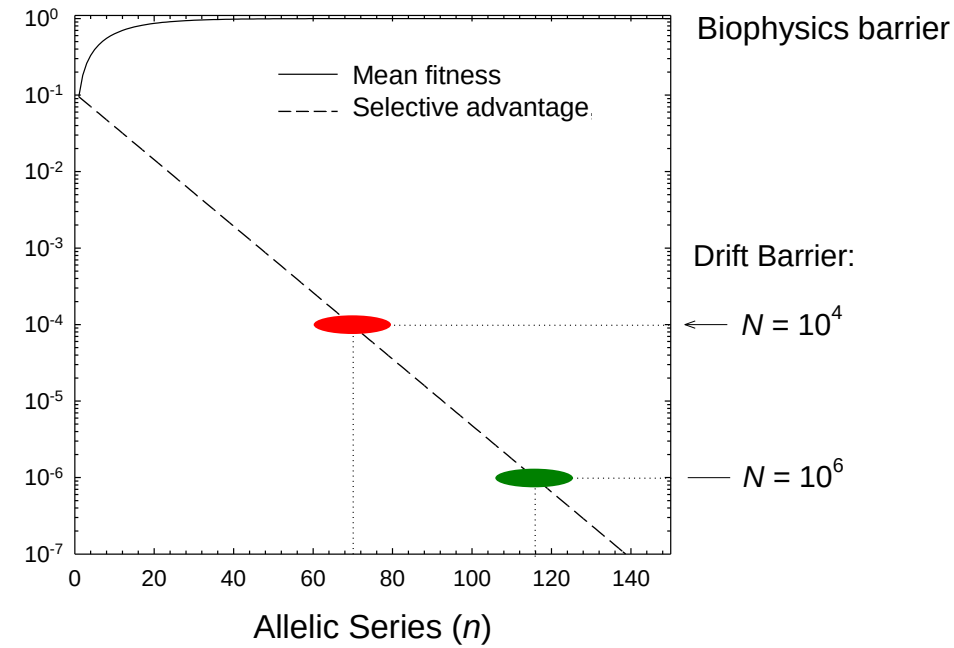
- How does the efficiency and structure of molecular features vary across phylogenetic groups?

- Do cellular adaptations hit the Biophysics Barrier – the absolute limits of molecular perfection?

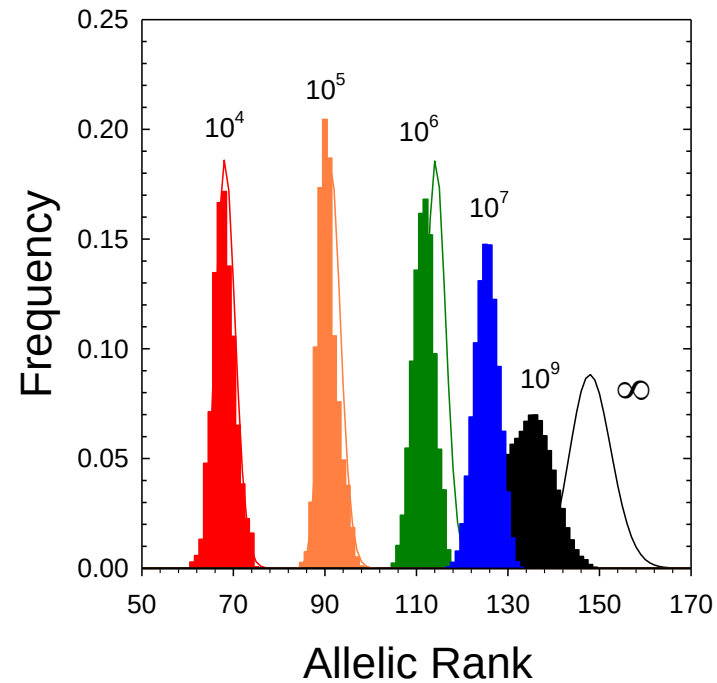
- **<u>The Drift Barrier to Achieving Adaptive Perfection</u>**: Once the selective advantage of improving a trait is less than the power of drift, $1/(2N_e)$, where N_e is the effective population size, no further improvement in fitness can be sustained.

The Drift-barrier Hypothesis for a Single Trait

Asymptotically Increasing Perfection in an Allelic Series

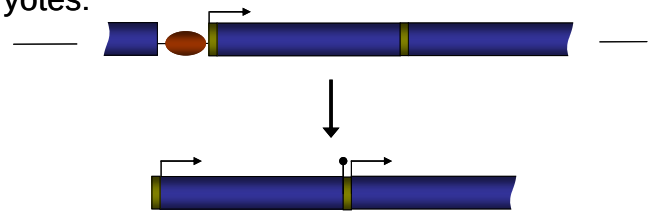


Equilibrium Allele-frequency Distributions with Increasing Population Sizes

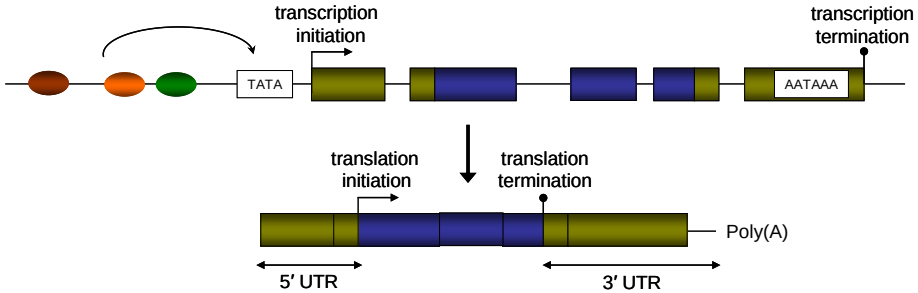


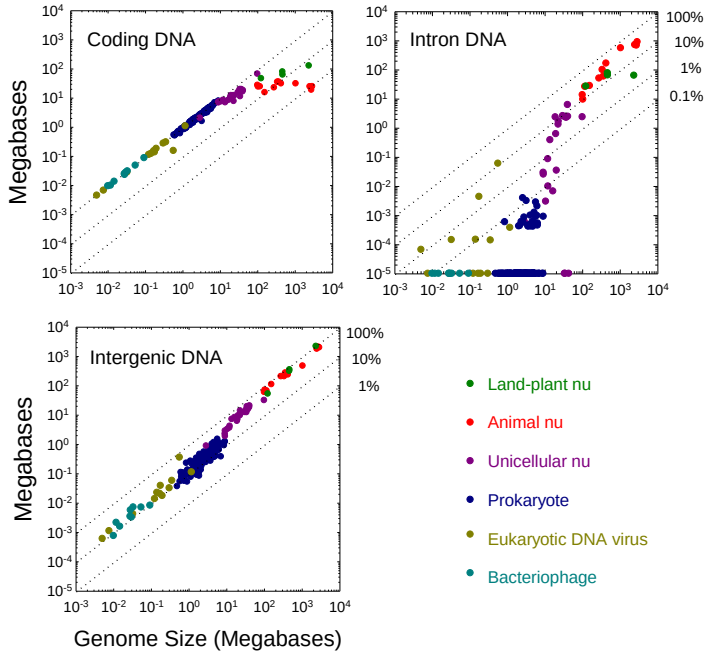
The Origin of Gene-structure Complexity by Nonadaptive Mechanisms

Prokaryotes:



Eukaryotes:





- Nearly all embellishments to gene structure impose weak mutational disadvantages. While these can be efficiently removed by selection in prokaryotes with large effective population sizes, they can accumulate in an effectively neutral fashion in eukaryotes experiencing relatively high levels of random genetic drift.

Can these general principles help explain structural features of proteins and cellular diversity?

Effectively Neutral Evolution at the Level of Cellular Features?

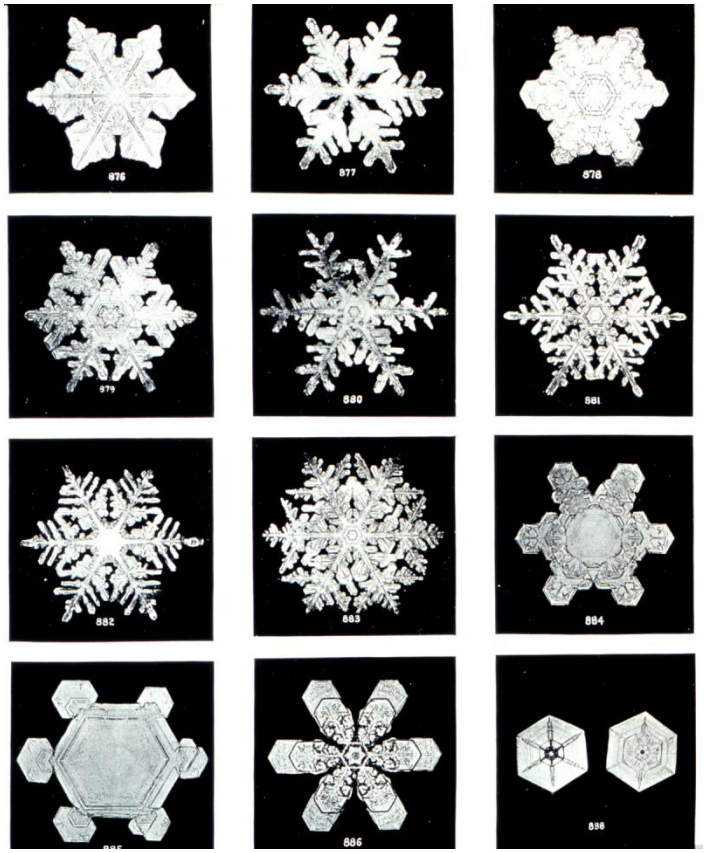
- Complete rewiring of regulatory pathways (transcription factors and their binding sites) in different yeast species – ribosomal proteins; mating type; galactose utilization.

- Enzyme reaction rates are orders of magnitude less than the diffusion limit, and enzyme promiscuity is the rule.

- Replication fidelity is reduced in species with smaller effective population sizes.

- Variation in the multimeric nature of proteins is independent of organismal complexity.

Mesmerizing Beauty, Diversity, and the Adaptationist Paradigm

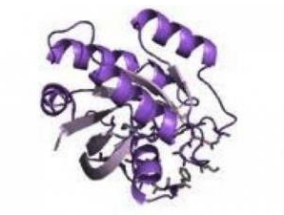


"….. from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved." Charles Darwin

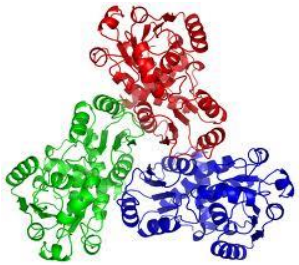
The Origin of Variation in Molecular Complexes:

Driven by adaptive processes unique to individual lineages?

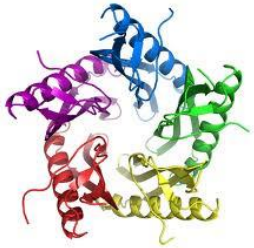
Or a consequence of biased mutation pressure and biophysical factors?



monomer



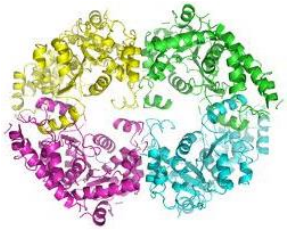
trimer



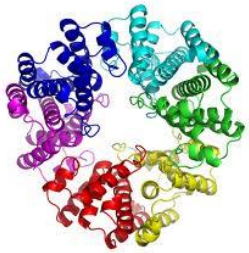
pentamer

heptamer

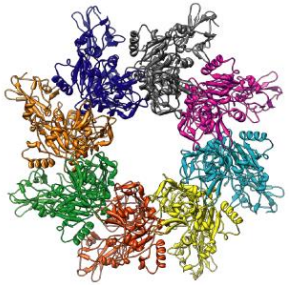
dimer



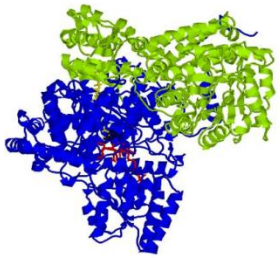
tetramer



hexamer



octamer

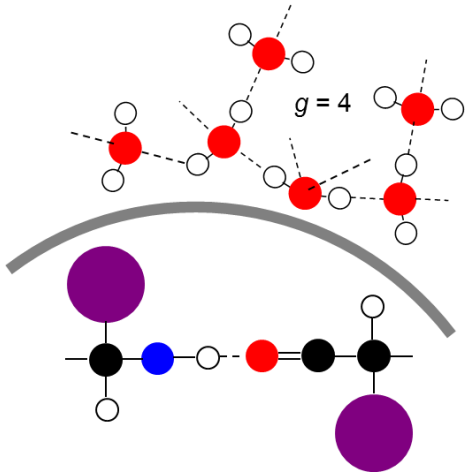


- **Potential advantages to complex formation**:
  - increased structural diversity,
  - reduced surface area increases productive encounter rate with substrate,
  - reduced problems of folding single large proteins,
  - reduced vulnerability to denaturation and/or engagement in promiscuous interactions,
  - reduced molecular motion at the catalytic site increases substrate specificity,
  - increased flexibility for allosteric regulation,

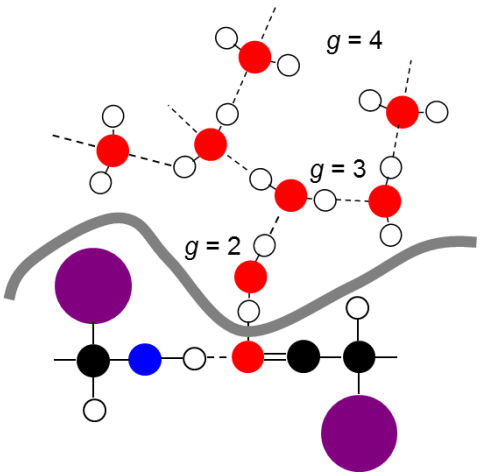
- **Compensation for structural deficiencies in monomeric subunits?**

- **Proteins with an affinity to oligomerize can also come at a cost:**
  - Elevated production levels necessary for a critical encounter rate for successful multimerization.
  - Problems with harmful interactions between heterotypic molecules in heterozygotes in the establishment phase.
  - Concatenation into indefinite filaments – human disorders involving the production of inappropriate protein aggregates include Alzheimer's, Parkinson's, sickle-cell anemia, and amyotrophic lateral sclerosis (ALS).

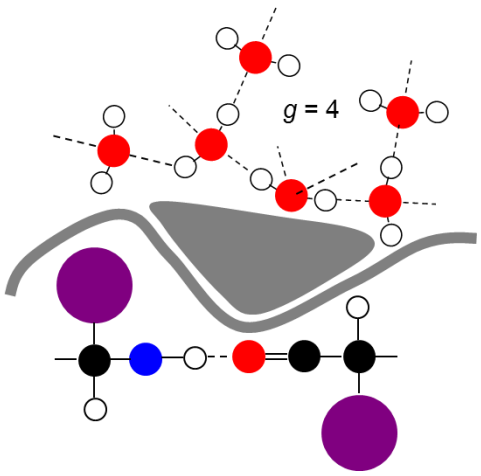
Can Nonadaptive Processes Lead to the Evolution of Protein Complexity?



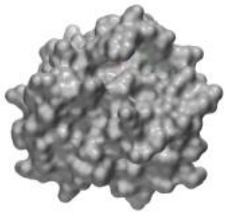
Well-protected



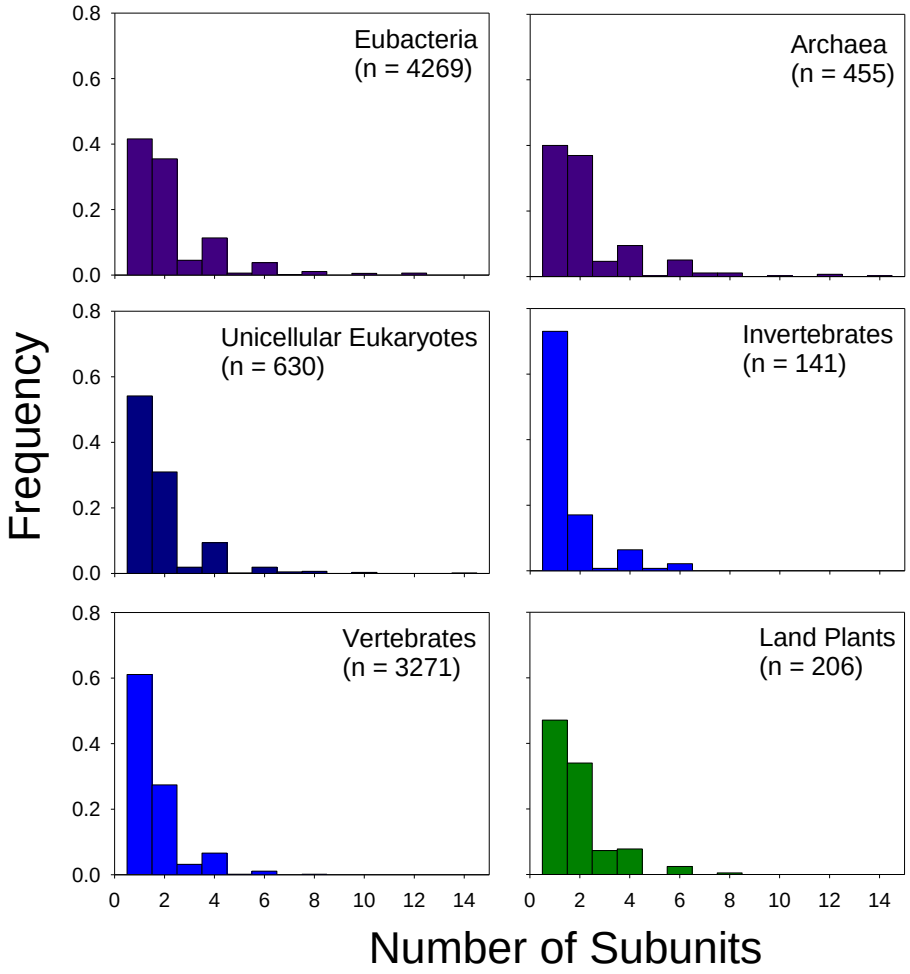
Exposed



Tension relieved



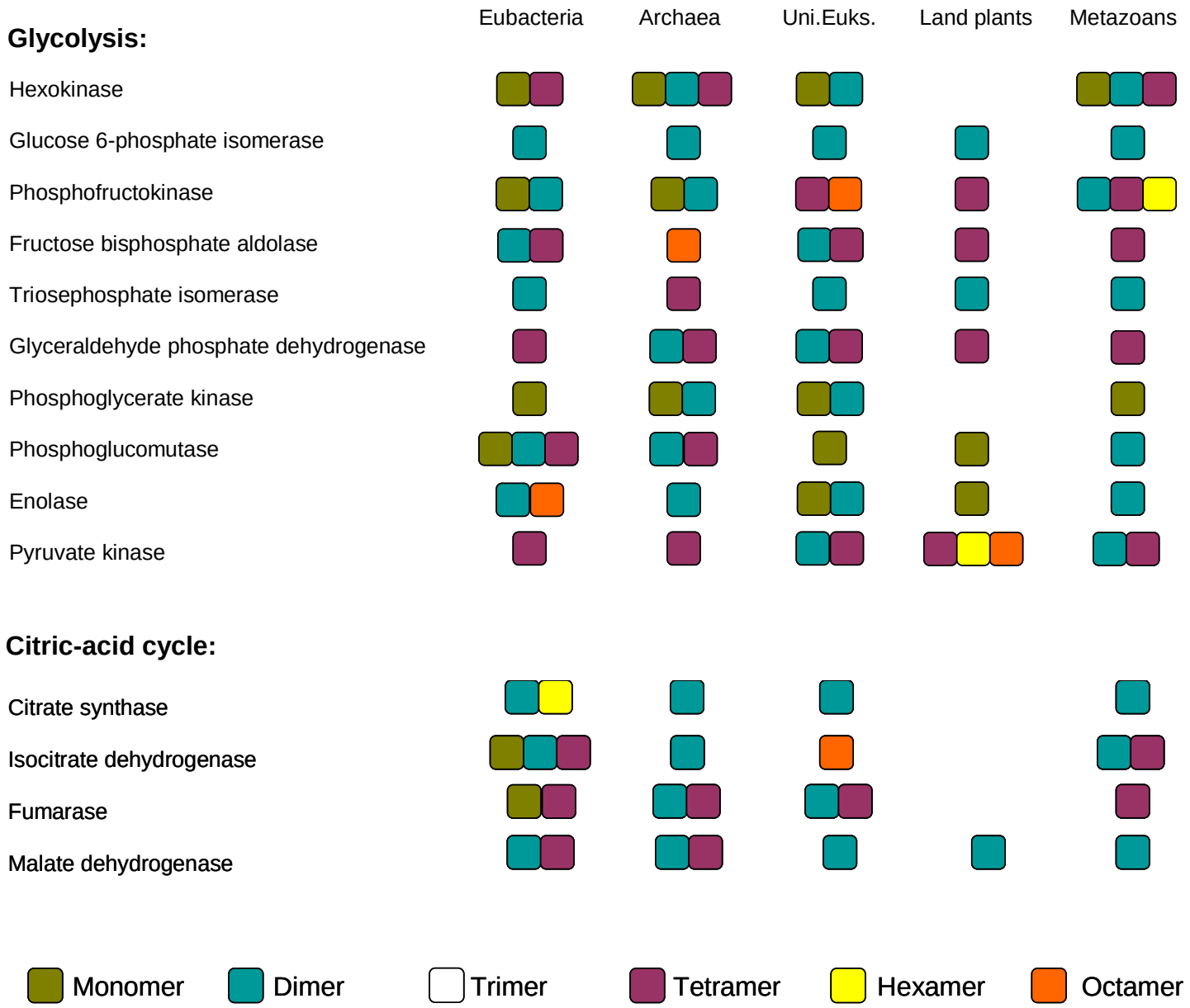
Distribution of Homomeric Types: approximate constancy across the Tree of Life.



- Roughly two thirds of proteins are multimeric, independent of phylogenetic lineage.
- Roughly two thirds of multimers are dimers.
- ~15% are tetramers, most of which are “dimers of dimers,” most likely arising via an intermediate dimeric state.
- Odd-mers are greatly under-represented.

Data from: 3D Complex.org (Levy et al. 2008)

Known Oligomerization Structures for the Enzymes of Central Metabolism



Enzymes with Identical Multimeric States Need Not Have the Same Structural Basis

Dihydrodipicolinate synthase (involved in lysine synthesis)

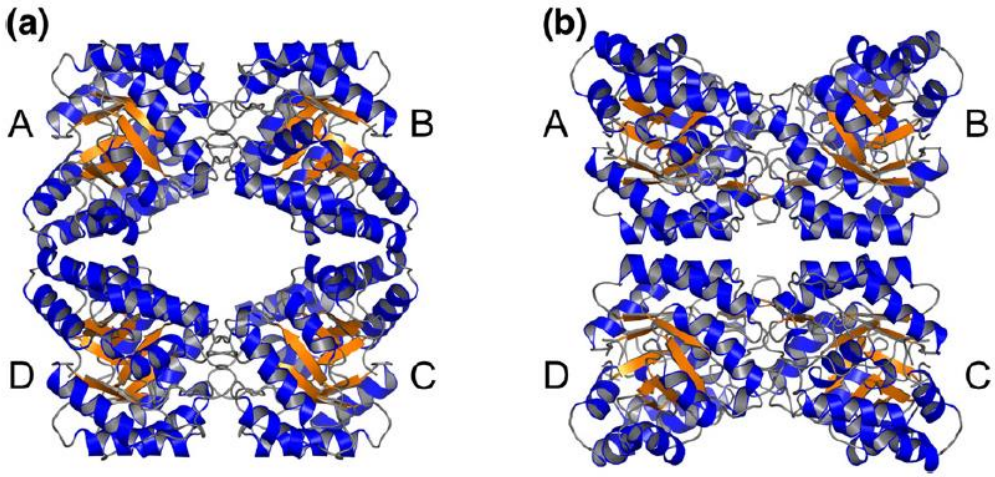
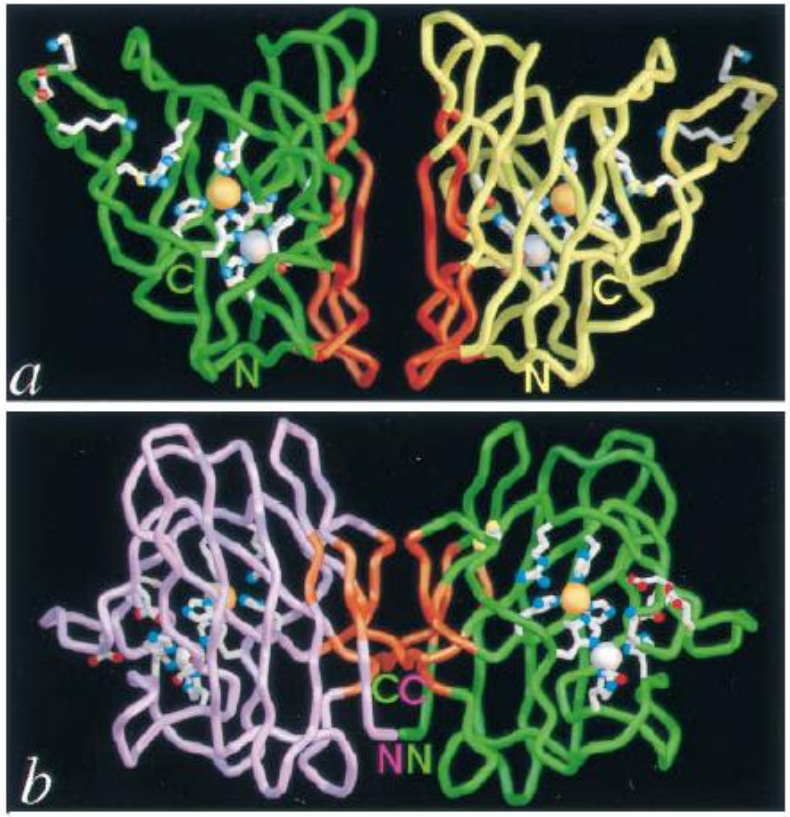


Fig. 1. The X-ray crystal structures of DHDPS from (a) *E. coli*[18,19] and (b) *N. sylvestris*.[23] Each enzyme is a homotetramer of $(\beta/\alpha)_8$-barrels composed of two tight-dimer units (A–B and C–D), but the arrangement of the two dimeric units is different.

Both species make homotetramers, but the dimer-dimer interfaces are completely nonoverlapping, face to face in the former, and back to back in the latter (Griffin et al. 2008).

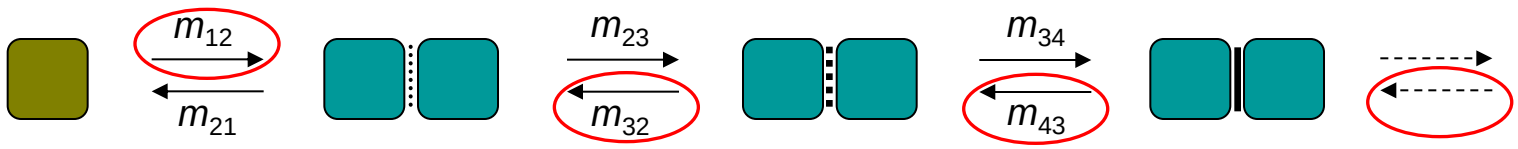
Cu,Zn Superoxide Dismutase:

Dimer interfaces in *Photobacterium* (above) and cow (below) are constructed from diametrically opposite beta-barrel elements (Bourne et al. 2008).



- Dayhoff et al. (2010) estimate that about two-thirds of protein families containing homomers exhibit phylogenetic variation in the binding interfaces.

Evolution of a Dimeric Structure

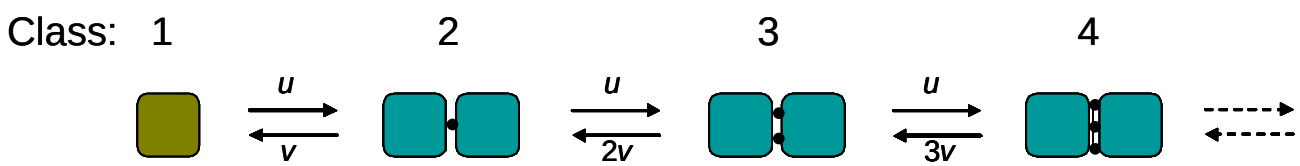


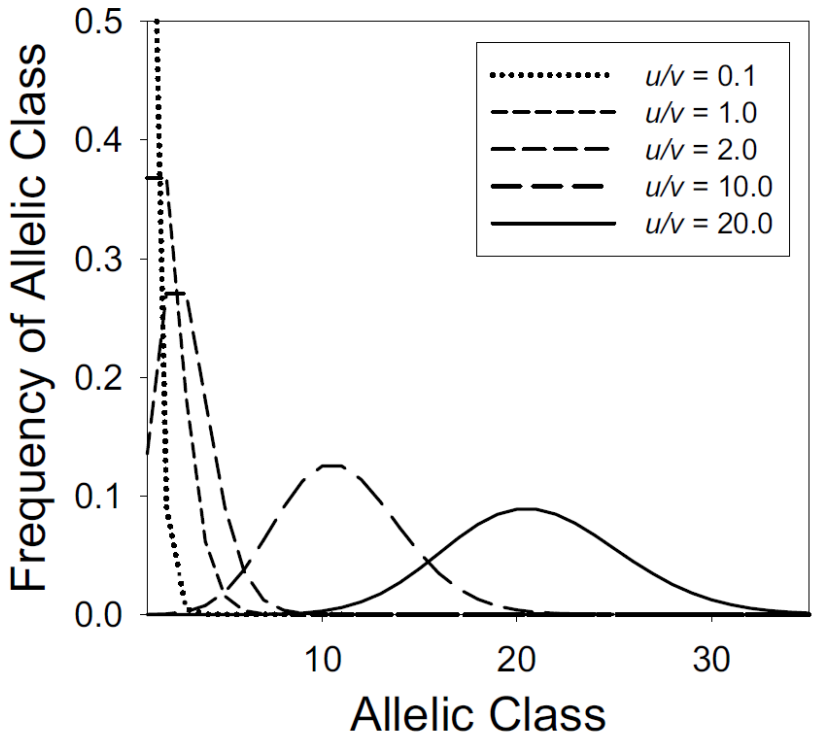
- Each transition rate is equal to the product of the number of relevant mutations arising per generation and the fixation probability.

- At steady state, the flux rate must be equal in both directions. This means that the net rate of establishment of dimers from monomers must equal the reverse rate.

- The equilibrium probability of each state is simply proportional to the product of the total set of transition rates towards the state from both directions.

The Neutral Expectation: the steady-state distribution of alternative allelic states is Poisson, a simple function of the ratio of upward and downward mutation rates, independent of population size.





$$\widetilde{P}_i = \frac{(u/v)^{i-1}}{(i-1)!C}$$

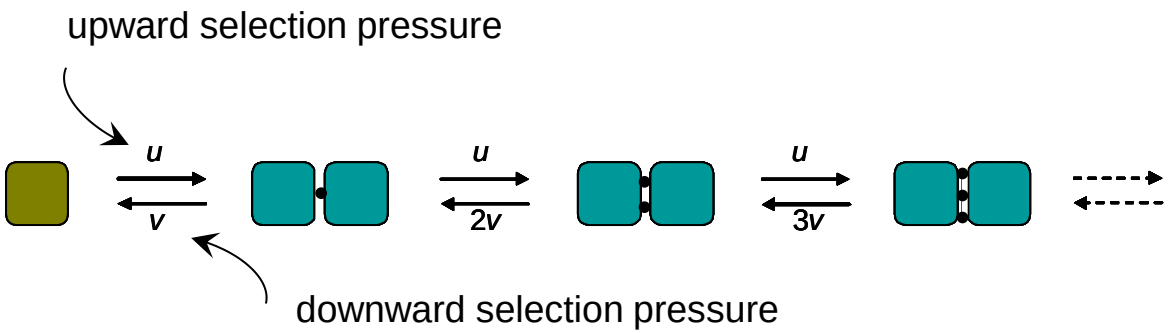
$$C = \sum_{i=0}^{\infty} (u/v)^i / i! = e^{u/v}$$

u/v is the mutation bias.

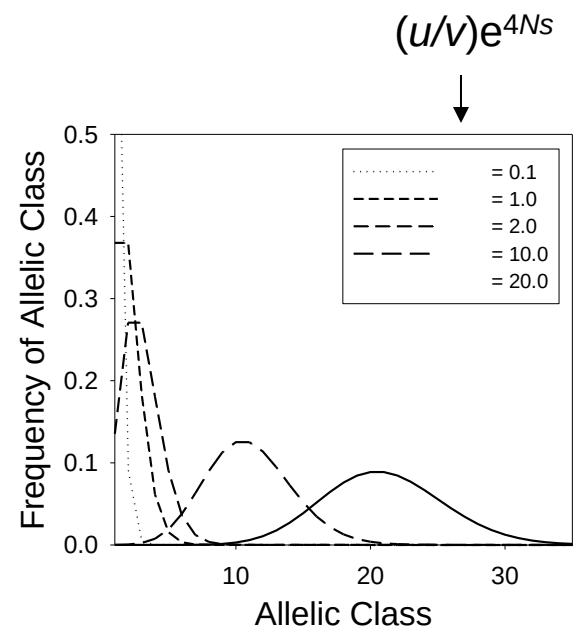
Expected frequency of monomers = $e^{-u/v}$

Adding in Selection:

- s is the selective advantage (or disadvantage) of each incrementing allele.
- e^{4Ns} is the ratio of fixation probabilities for beneficial vs. deleterious mutations.
- $4Ns$ is the ratio of the power of selection to random genetic drift.



- The distribution is again Poisson, but now the key parameter is $(u/v)e^{4Ns}$.
- The effects of selection, drift, and mutation bias cannot be disentangled from observations on the steady-state distribution alone.



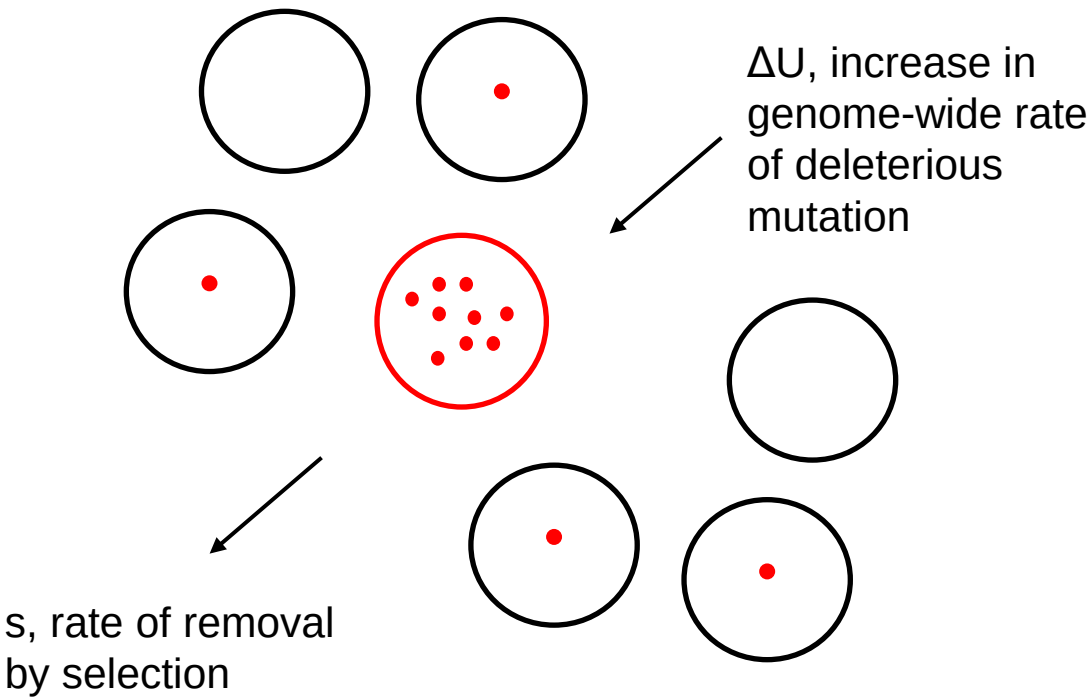
General Conclusions on Multimer Evolution

- Substantial phenotypic variation can arise among lineages, **even when selection and mutation is operating in an identical manner in all lineages.**

- The most common molecular state is not necessarily the optimum – **even with *negative* selection against multimers, they will still be common provided the mutational bias towards binding affinity is sufficiently large.**

- If the ratio of the power of selection and drift is < 1.0, the phenotypic distribution is entirely driven by mutation bias – **effective neutrality**.

The Magnitude of Selection Operating to Improve Replication Fidelity



Excess number of mutations at equilibrium = ΔU / s

↓

X

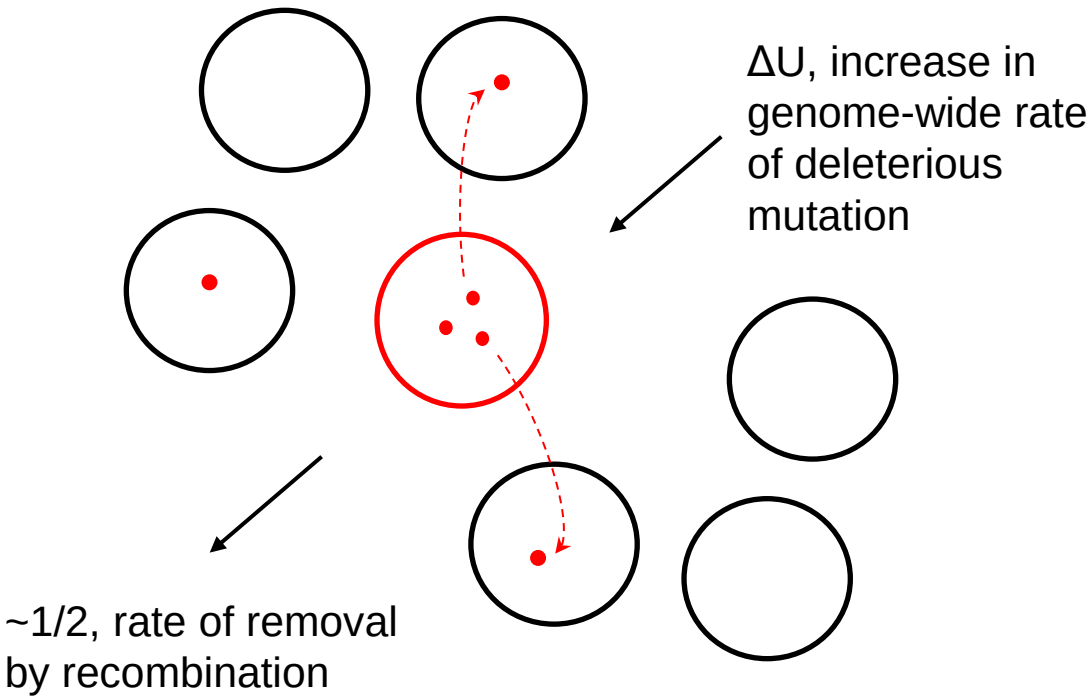
Effect / mutation = s

↓

Total effect on fitness = ΔU

- Selective disadvantage of a mutator in an asexual population
  = increase in genome-wide deleterious mutation rate

The Force of Selection to Improve Replication Fidelity is Greatly Reduced in Sexual Populations



Excess number of mutations at equilibrium = ΔU / (1/2)

↓

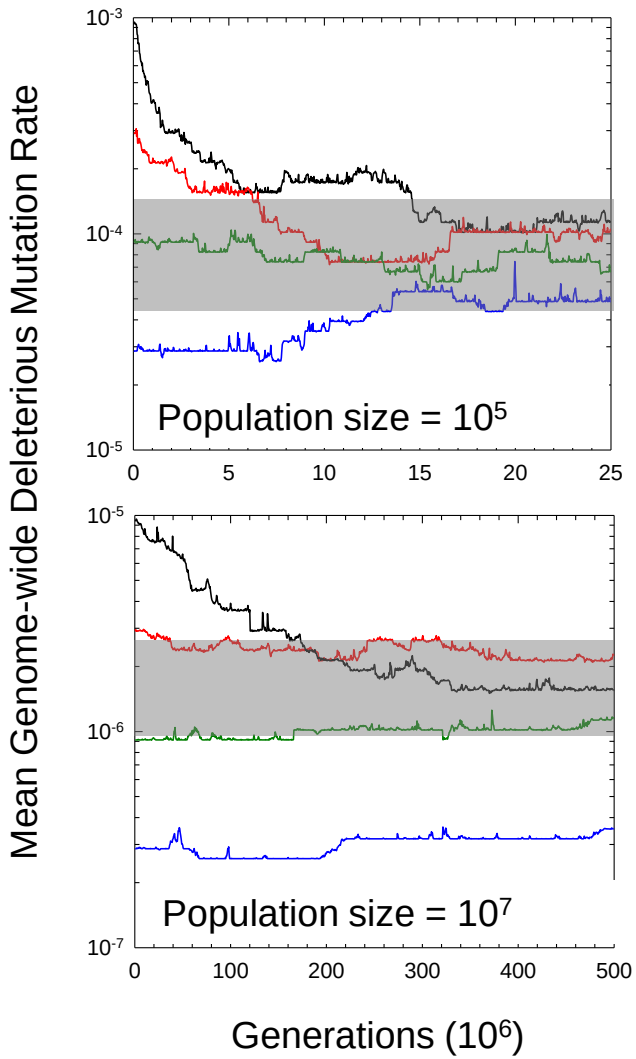
X

Effect / mutation = s

↓

Total effect on fitness = 2 s ΔU

Quasi-equilibrium Mutation Rates Resulting From Deleterious-mutation Load

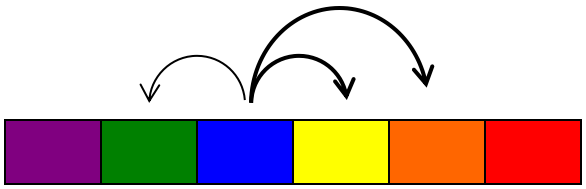


Effective selection for antimutators

DRIFT BARRIER

Biased production of mutators

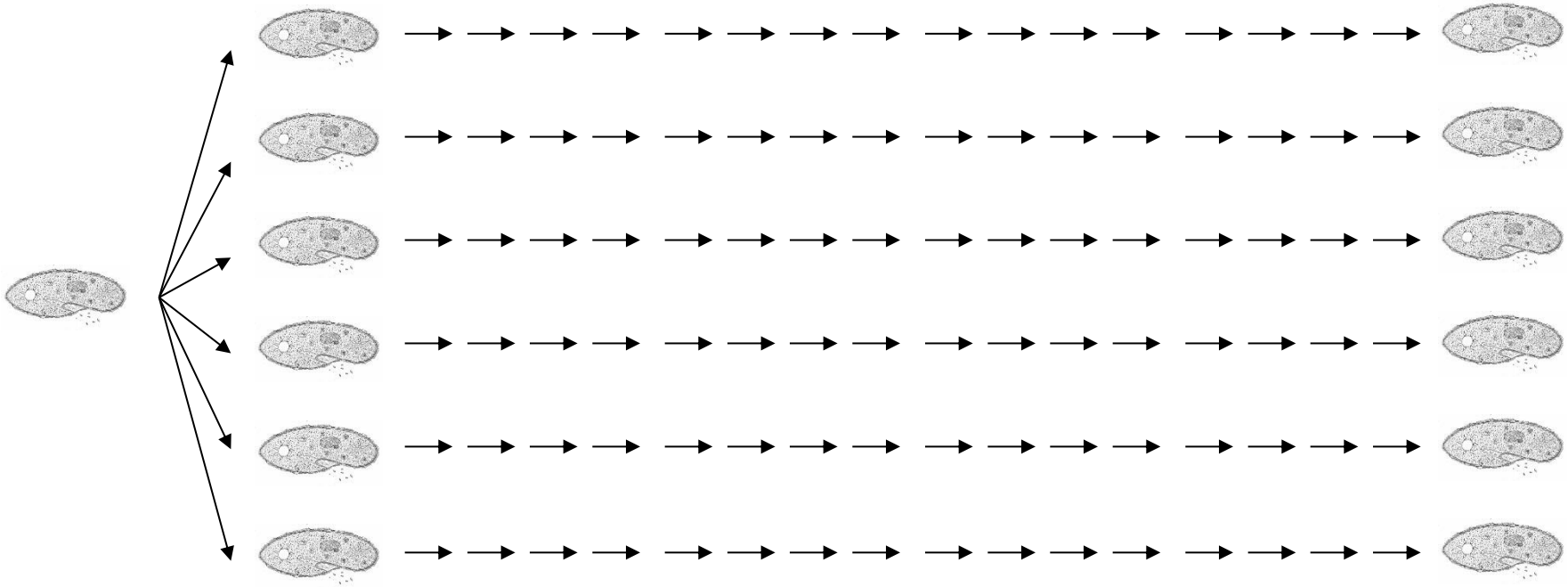
- Equilibrium mutation rate is expected to be inversely proportional to the effective population size.



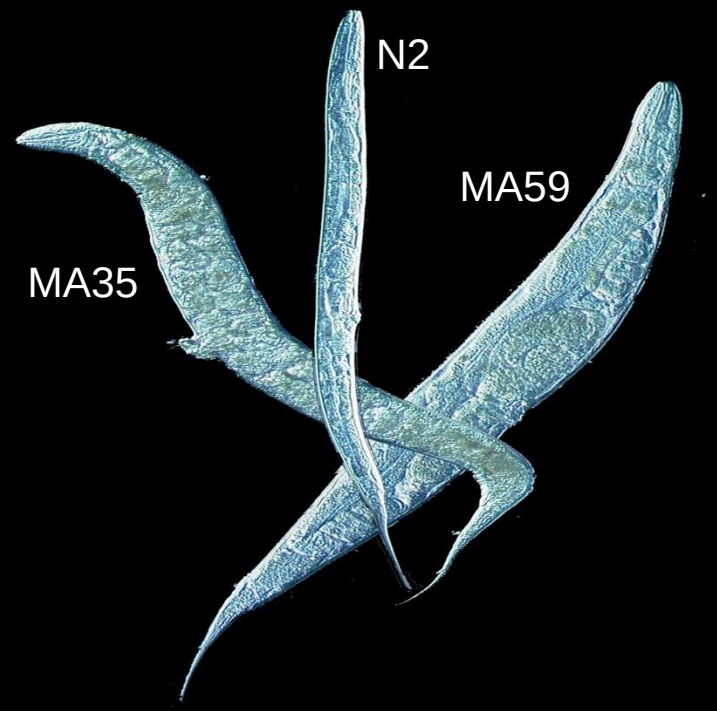
Mutation-rate classes

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.
- Characterize by whole-genome sequencing.

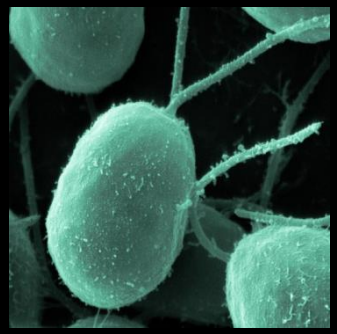


Extreme Morphological Divergence in MA lines of *C. elegans*



Recent and Current Eukaryotic Targets of Study

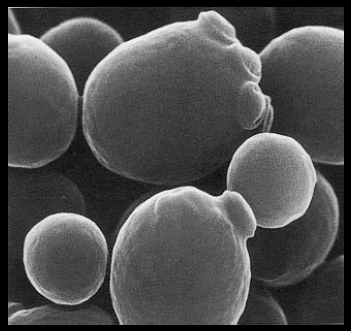
Arabidopsis



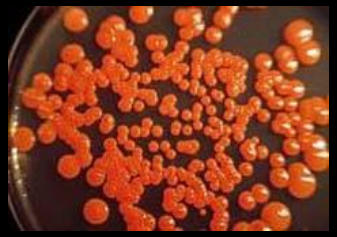
Chlamydomonas

Phaeodactylum

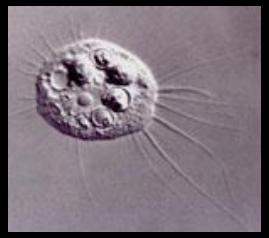
Dictyostelium



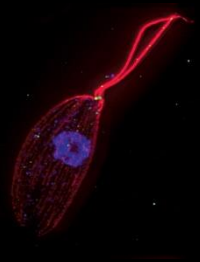
Saccharomyces



Rhodotorula

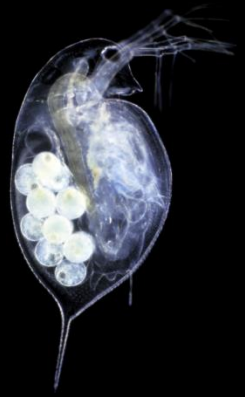


Ichthyosporean



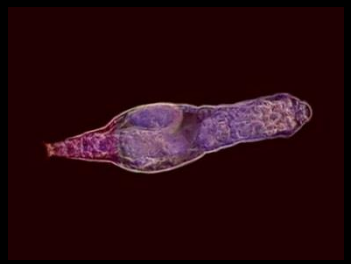
Naegleria

Paramecium

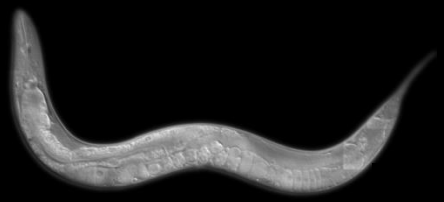


Daphnia

Drosophila



Adineta



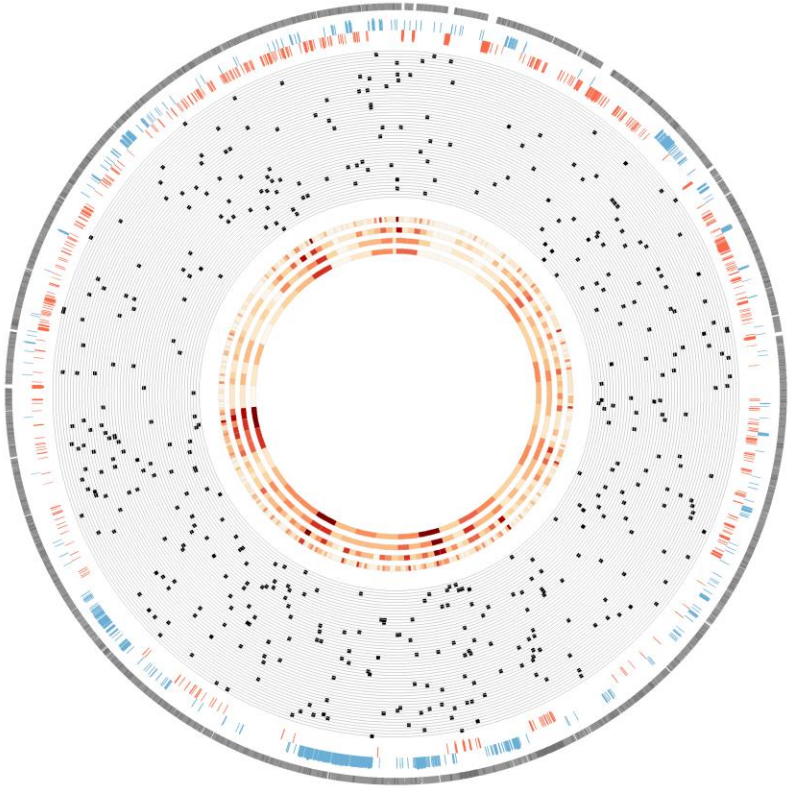
Caenorhabditis

Mutation-accumulation Studies in Prokaryotes

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

* = concurrent study with mismatch-repair deficient lines

Mutation in Small vs. Large Genomes



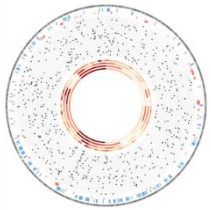
Index:

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.

Drake's (1991) Conjecture:
A Constant Rate of Mutation per Genome per Cell Division in Microbes

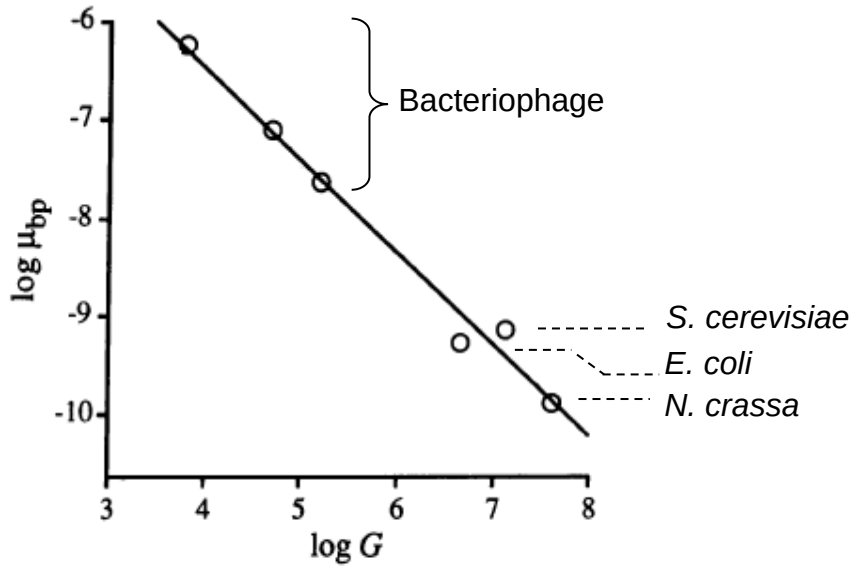
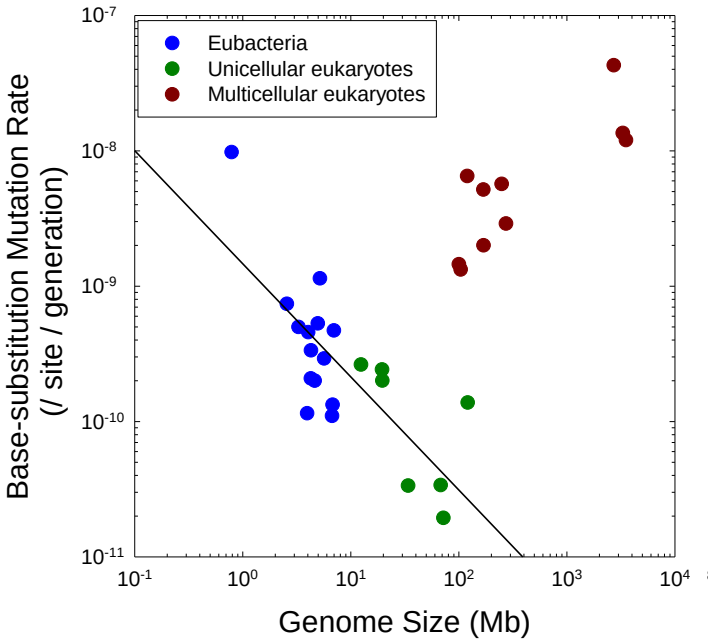


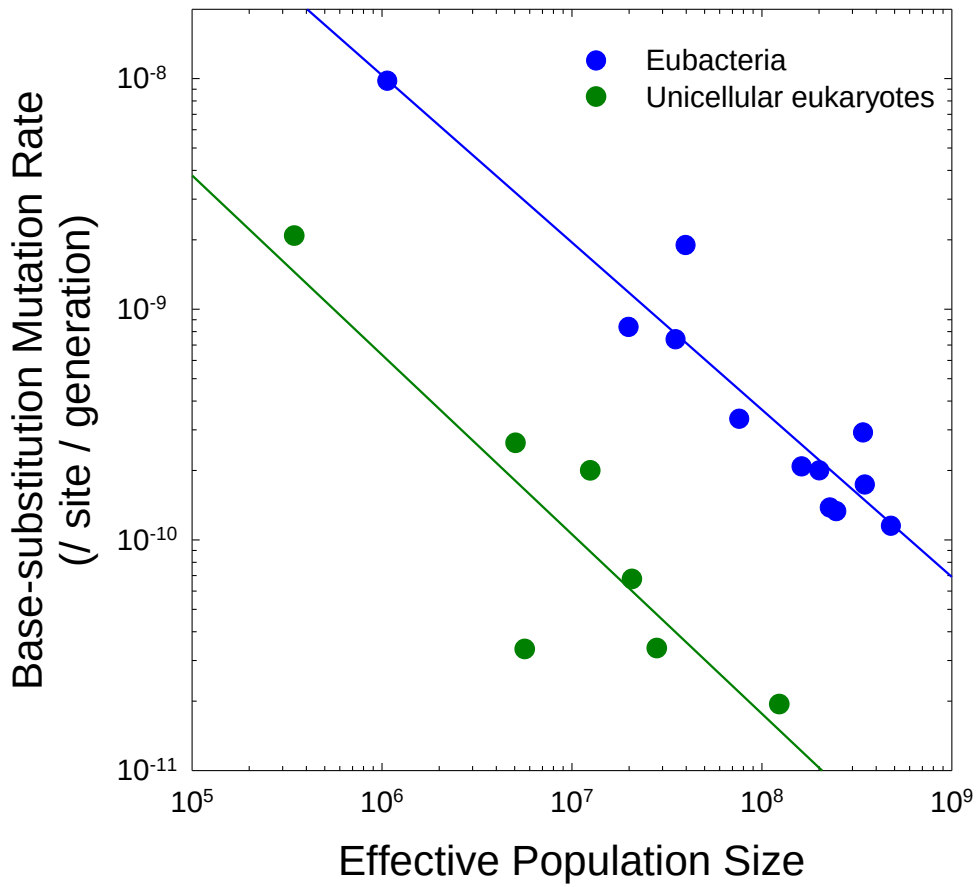
FIG. 1. Average mutation rate μ_{bp} per base pair as a function of genome size G in bp. The logs of the rates for each organism were averaged and all 13 values are included. Phages T2 and T4 were treated as a single organism.



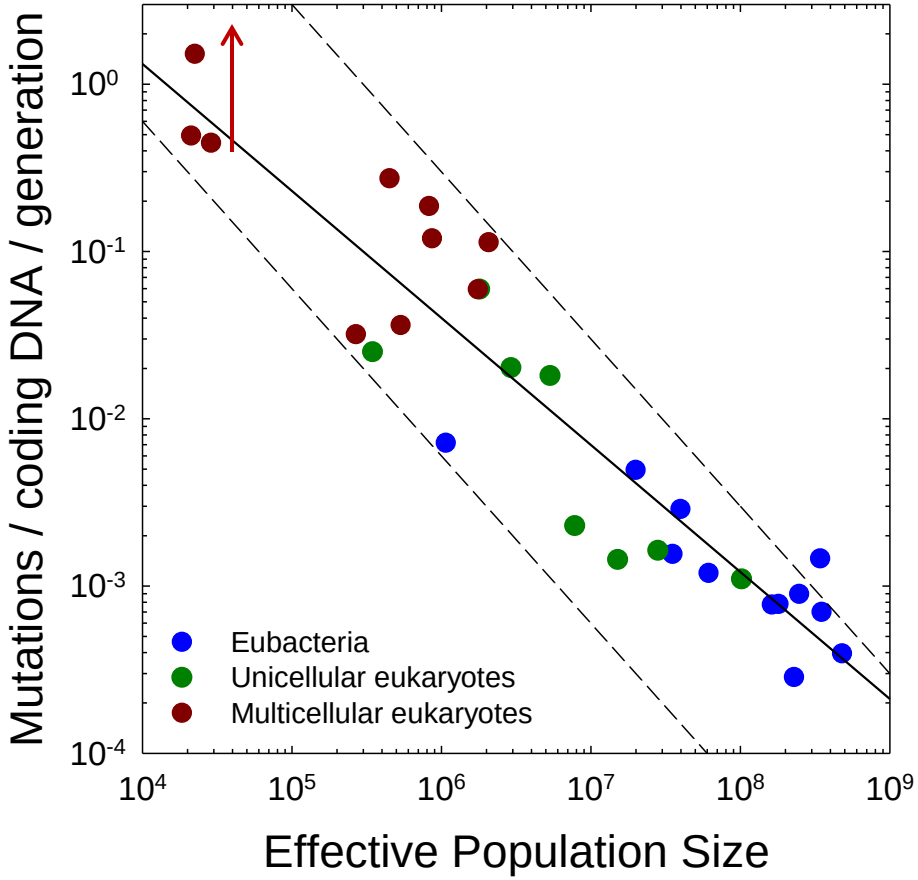
"Because this rate is uniform in such diverse organisms, it is likely to be determined by deep general forces."

The Mutation Rate / Nucleotide Site Is Inversely Proportional to the Average Effective Population Size of a Species

For a given magnitude of genetic drift, selection is capable of driving the mutation rate down further in eukaryotes than prokaryotes.



A Universal Inverse Scaling Between the Genome-wide Deleterious Mutation Rate (uG_e) and N_e Across the Tree of Life



- The mutation rate per nucleotide site scales inversely with both the effective population size and the amount of functional DNA in the genome (the total target size for deleterious mutations).

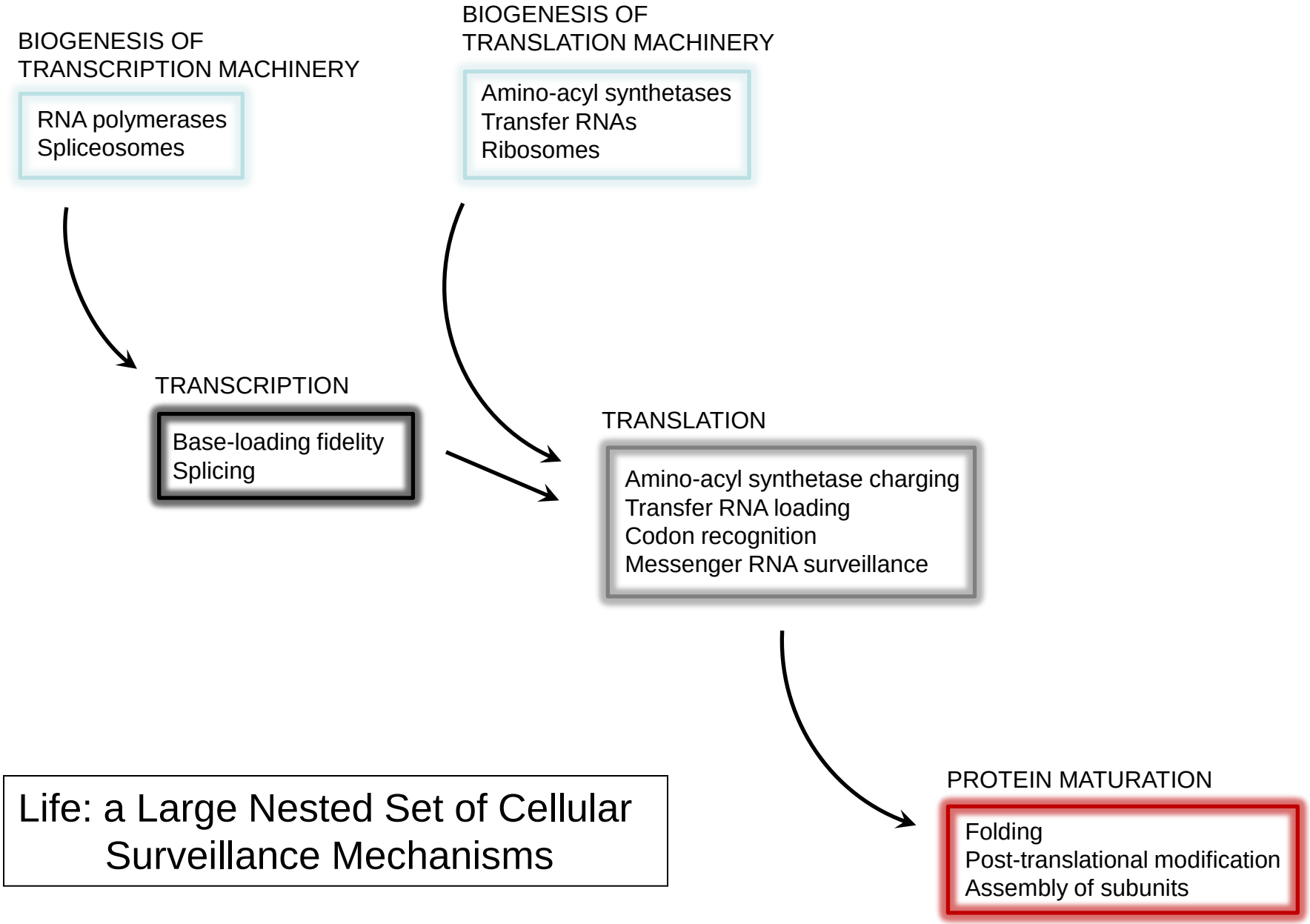
$uG_e \sim 1 / N_e \quad \rightarrow \quad u \sim 1 / (G_e \cdot N_e)$

u = mutation rate / site / generation

G_e = amount of functional DNA (sites)

N_e = effective population size

Eubacteria
Unicellular eukaryotes
Multicellular eukaryotes



Life: a Large Nested Set of Cellular Surveillance Mechanisms

Selection on the Replication Error Rate in Sexual Populations:

the selective disadvantage of a mutator allele is $\Delta u \cdot 2 \cdot G_e \cdot s$

Mutations remain linked to a mutator allele for an average of 2 generations

Number of nucleotides in the genome subject to selection

Heterozygous effect of a deleterious mutation

Selection on the Transcription Error Rate:

selective disadvantage of a transcriptional mutator is $\Delta u \cdot 1 \cdot T_e \cdot s \cdot d$

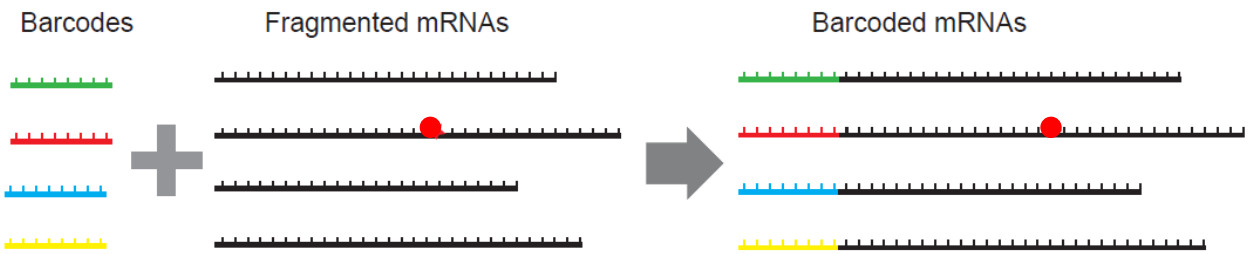
The pool of errors remains associated with the mutator for <1 generation

Number of nucleotides in the transcriptome subject to selection

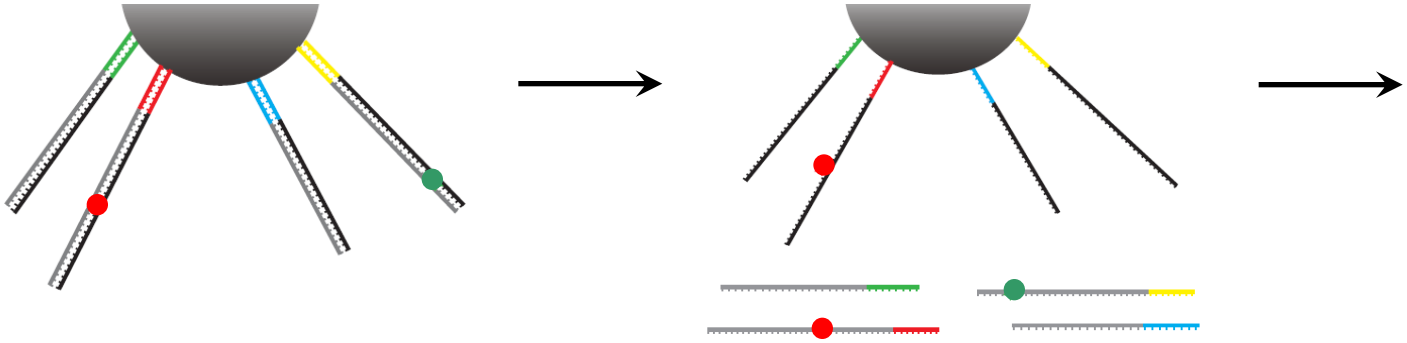
Heterozygous effect of a deleterious mutation

Dilution effect (<< 1.0)

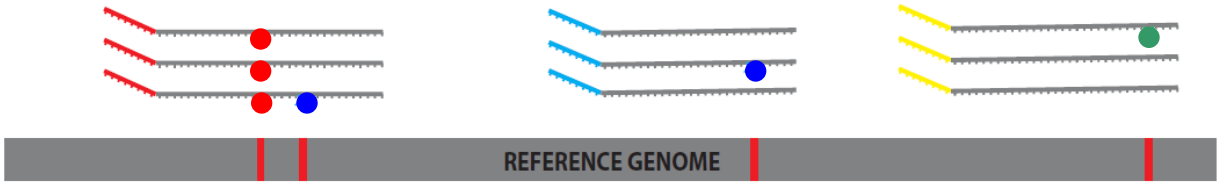
Estimation of the *in vivo* Transcription-error Rate From an RNA Library (Gout et al., PNAS, 2013)



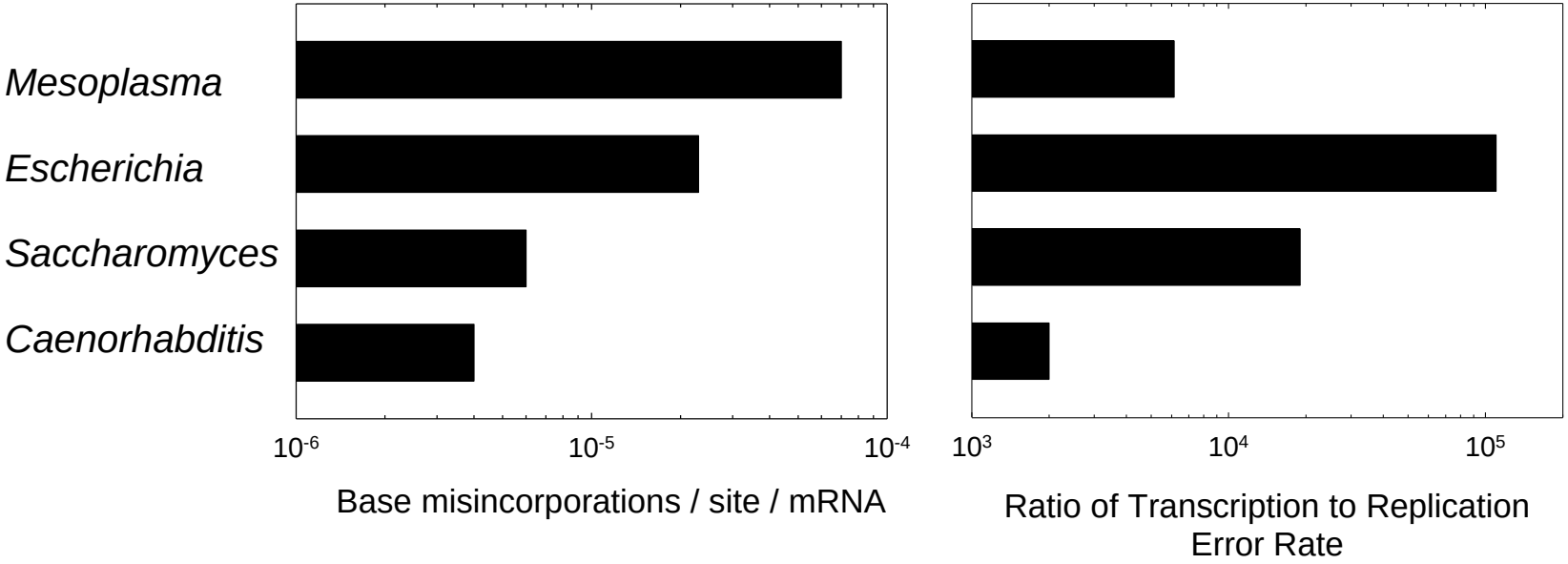
Capture fragments on beads; reverse transcribe; isolate cDNAs; repeat to obtain replicates:



Sequence to high depth; sort into uniquely coded families; search for consistent errors;



Transcription-error Rates Are Orders of Magnitude Higher Than Replication-error Rates



- ~1 to 5% of transcripts contain errors

“Must we geneticists become bacteriologists, physiological chemists and physicists, simultaneous with being zoologists and botanists? Let us hope so.”

H. J. Muller (American Naturalist, 1922)

Collaborators:

Indiana University:
Matthew Ackerman, Tom Doak, Pat Foster, Jean-Francois Gout, Matthew Hahn, Nate Keith, Iain Konigsberg, Jay Lennon, Weiyi Li, Hongan Long, Jake McKinlay, Sam Miller, Ron Pearson, Dan Schrider, Joe Shaw, Way Sung, Emily Williams, Sen Xu

University of New Hampshire:
Vaughn Cooper, Marcus Dillon, Kelley Thomas

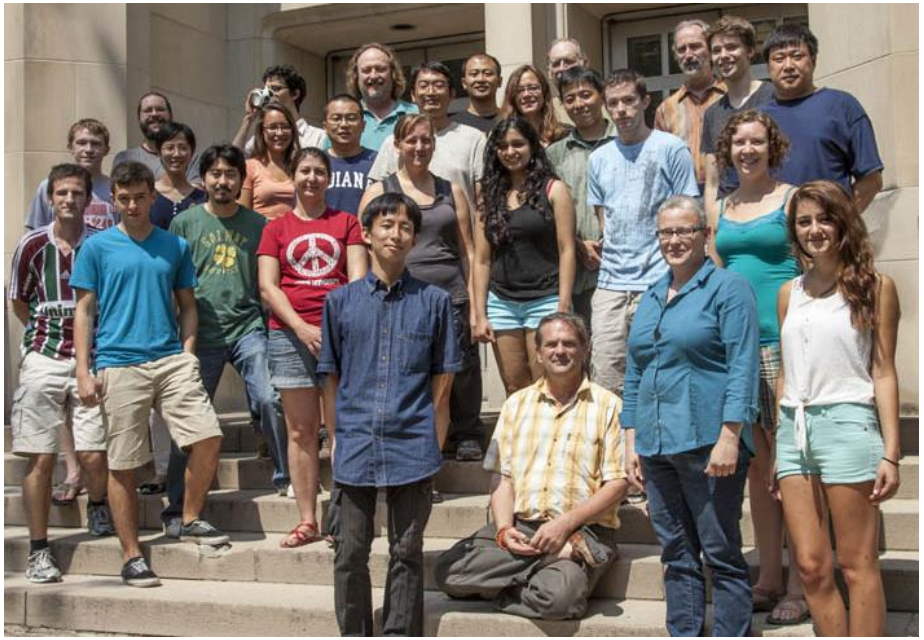
Hacettepe University:
Sibel Kucukyildirim

Universidade Federal do Rio de Janeiro:
Marcus Senra, Carlos Suarez

University of Chicago:
Allan Drummond

University of Houston:
Becky Zufall

Arizona State University:
Reed Cartwright, David Winter



Desks, Salary:

Indiana University
Department of Biology

NATIONAL INSTITUTES OF HEALTH

DEPARTMENT OF DEFENSE
UNITED STATES OF AMERICA

NSF