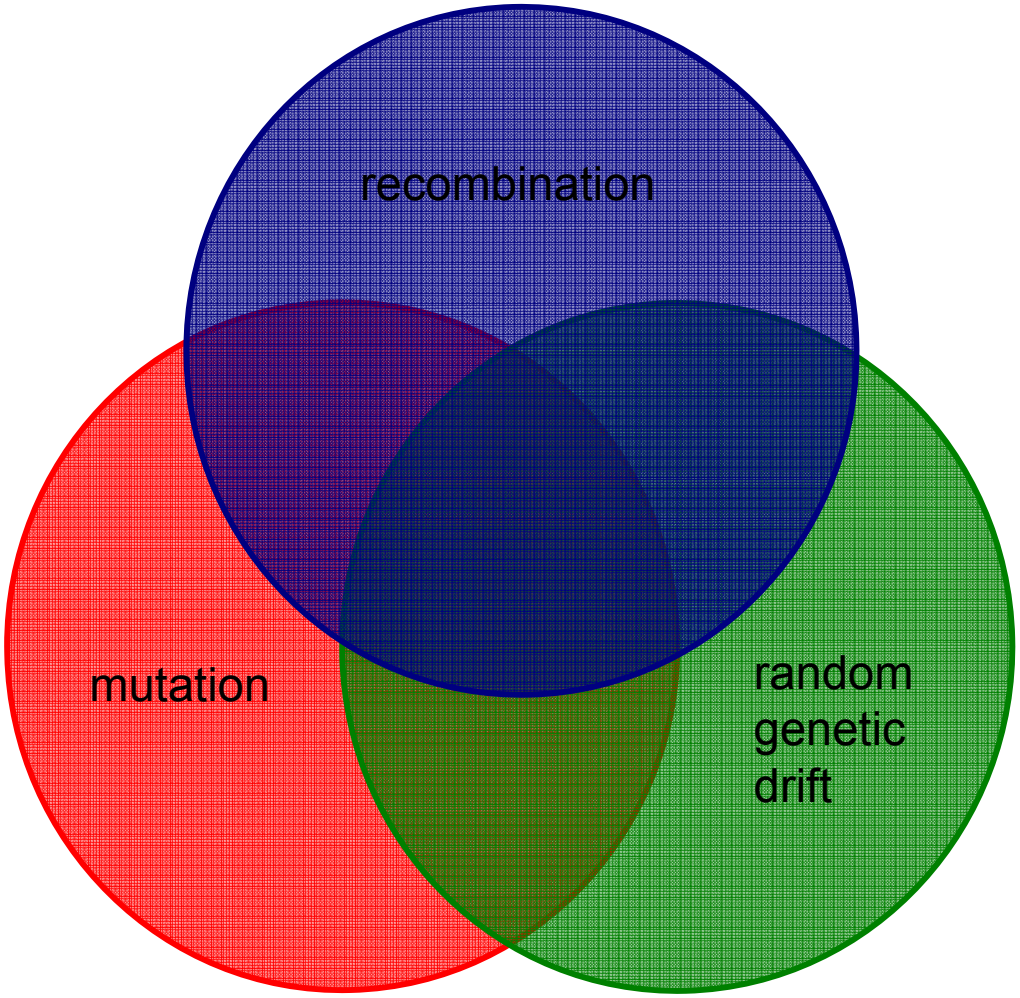
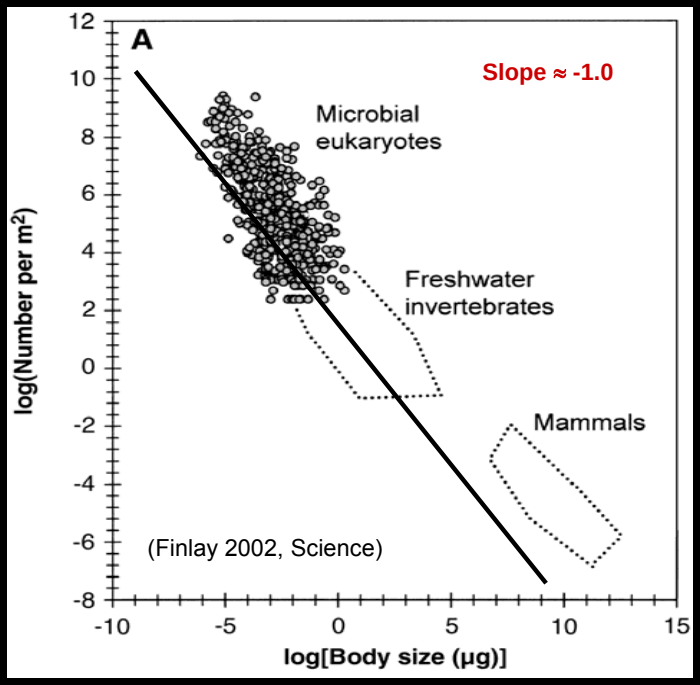
Nothing in evolution makes sense except
in the light of population genetics.

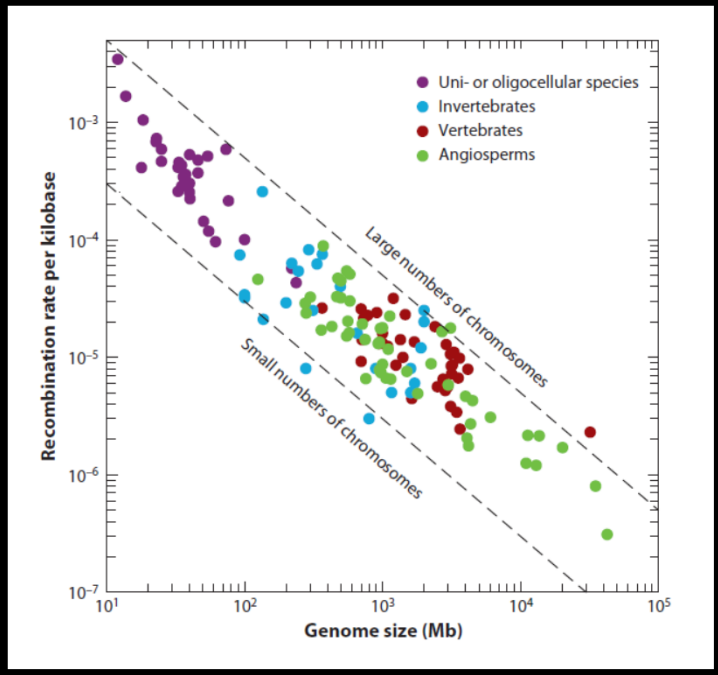
The Population-genetic Environment



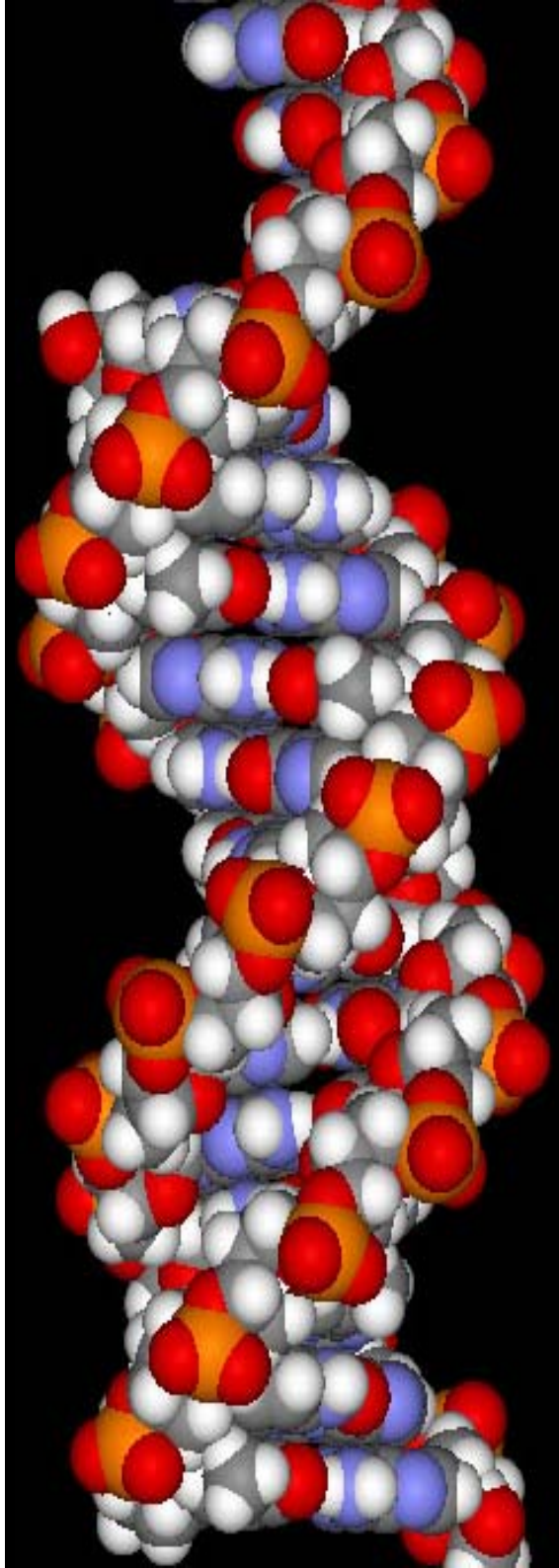
Scaling of Population Size and Recombination Rate with Organism Size



Reduction in absolute population size



Reduced recombination per physical distance



Evolution of the Mutation Rate

- The mutation rate scales across phylogenetic groups, among tissues, and among polymerases within cells in ways that are consistent with population-genetic theory.

- No evidence that mutation rates have been optimized to maximize the long-term rate of adaptive evolution.

- No evidence that mutation rates cannot be reduced below their current levels, i.e., the efficiency of replication has not reached the limits of molecular perfection.

- **The Drift Barrier to Mutation-rate Reduction**: Once the selective advantage of lowering the mutation rate is less than the power of drift, $1/(2N_e)$, the mutation rate has reached its minimum possible value.

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

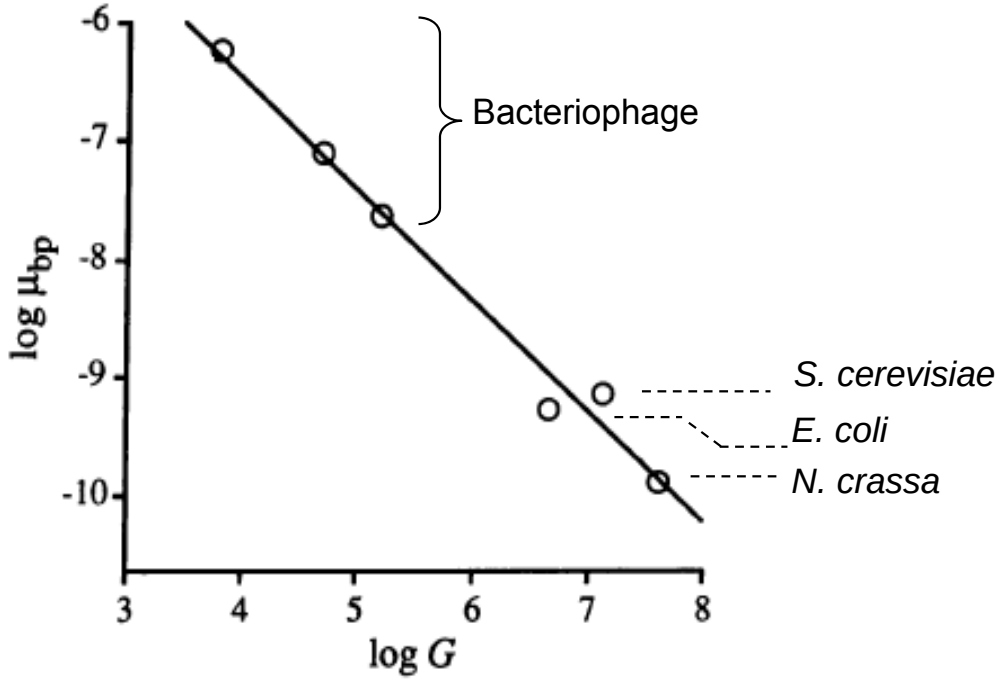
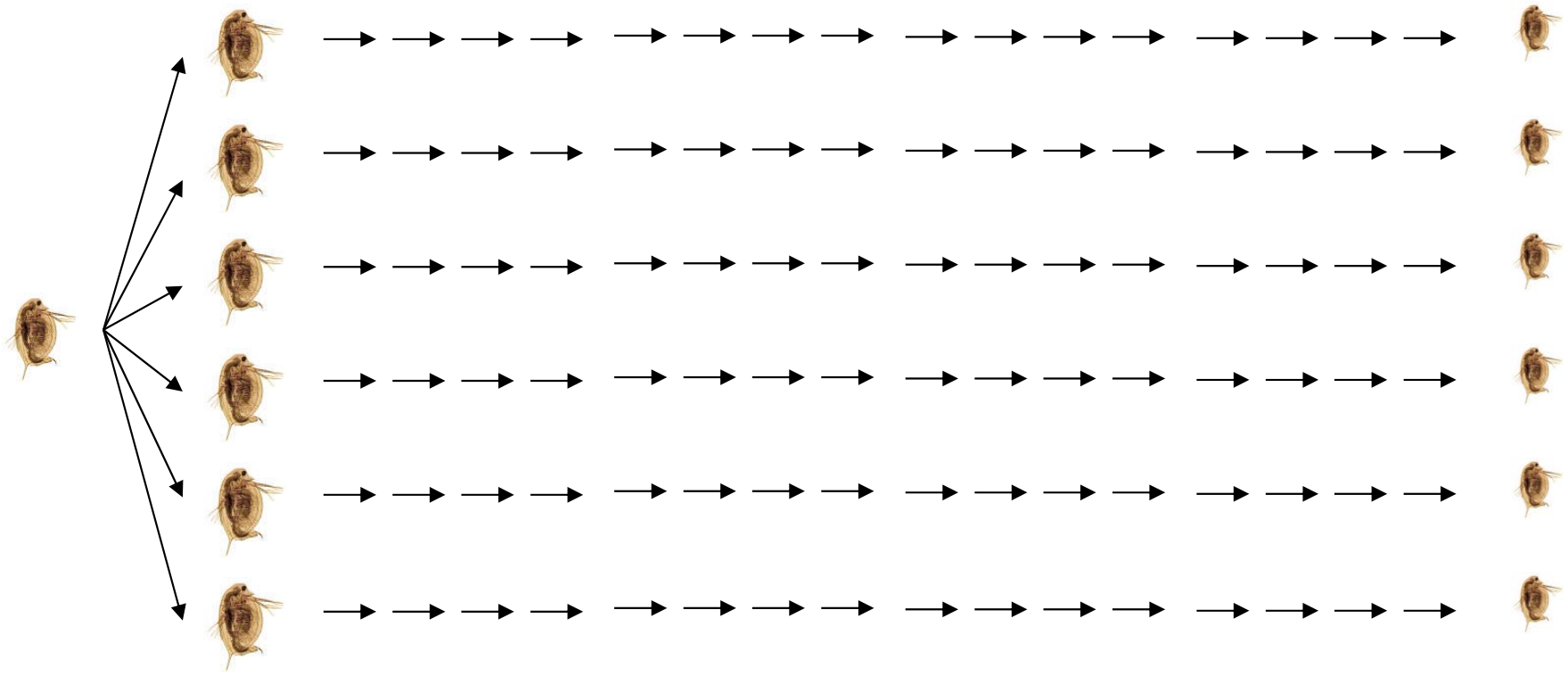


FIG. 1. Average mutation rate μ_{bp} per base pair as a function of genome size G in bp. The logs of the rates for each organism were averaged and all 13 values are included. Phages T2 and T4 were treated as a single organism.

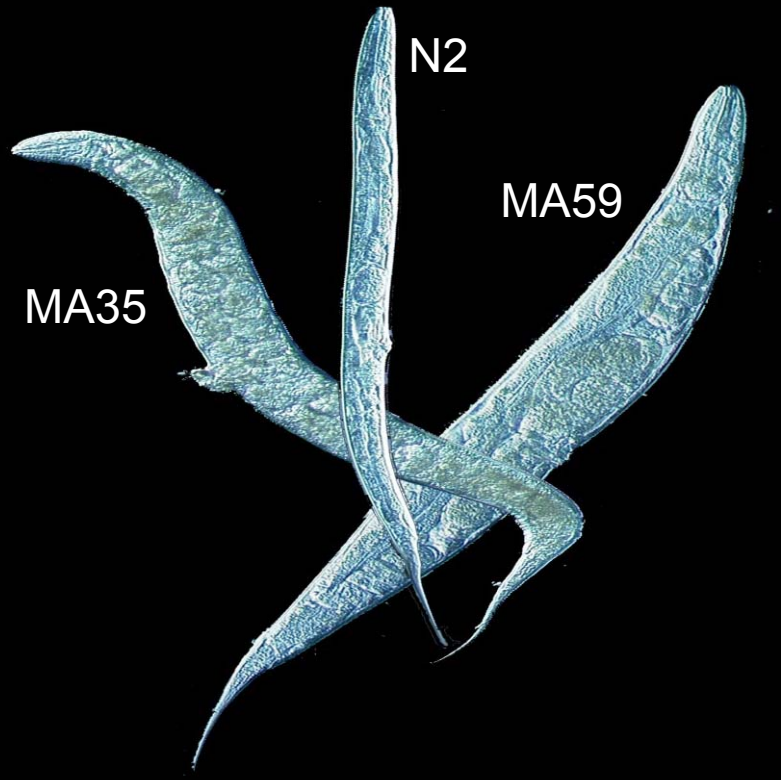
Mutation-accumulation experiment. Starting with a single stem mother, sublines are maintained by single-progeny descent, preventing selection from removing spontaneous mutations. This protocol is continued for hundreds of generations with dozens of lines.



Advantage – essentially no selection bias; allows a genome-wide perspective of the entire molecular mutation profile, from substitutions to large deletion/duplications.

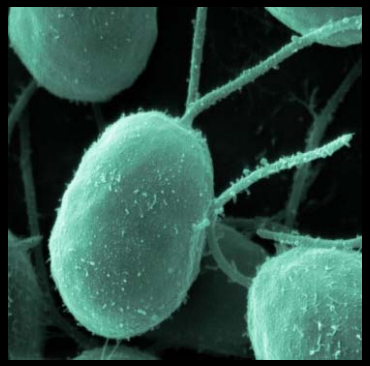
Disadvantage – labor intensive; line / investigator loss.

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



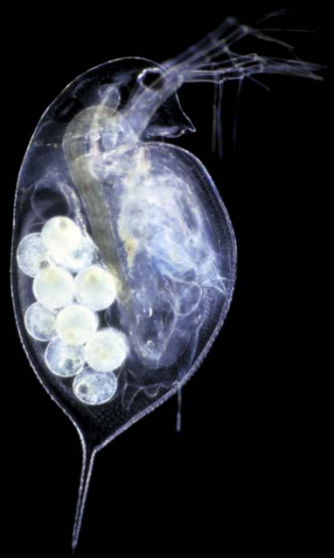
Recent and Current Eukaryotic Targets of Study

Arabidopsis

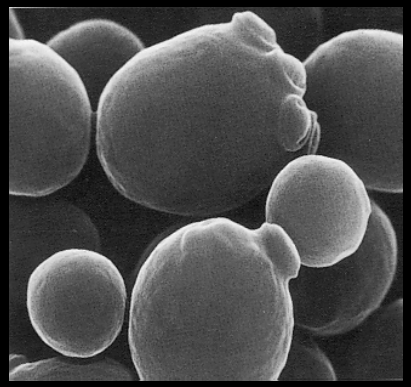


Chlamydomonas

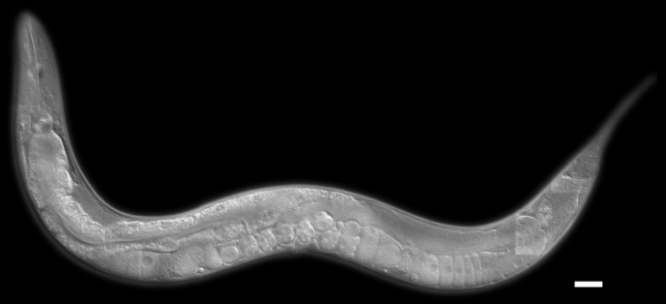
Phaeodactylum



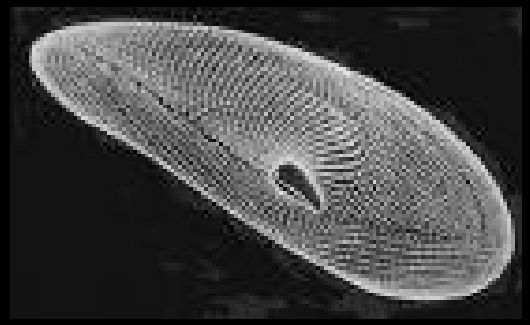
Daphnia



Saccharomyces

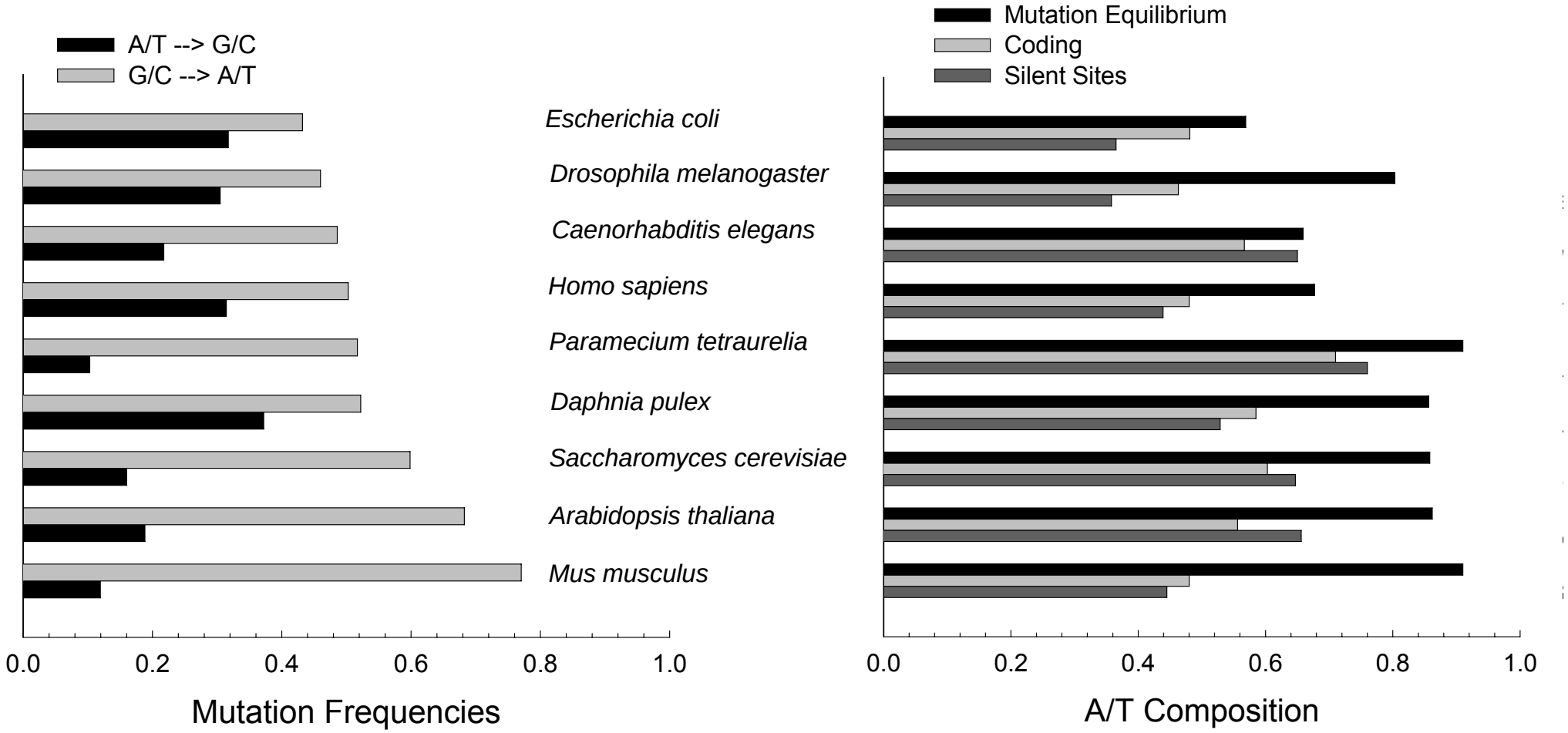


Caenorhabditis



Paramecium

- All genomes have substantial mutation bias towards A/T production.
- Genome-wide nucleotide compositions are not in mutation equilibrium.
- This universal deficit of A/T must be a result of selection and/or biased gene conversion.



- **A permanent resource for the life sciences community**.

Key
Lynch
Cooper
Winkler
Soares
Fuqua
Brun
Velicer
Rainey

	Phylum	Species / Strain	Mb	G/C %	Start Date	Notes
	Tenericutes	*Mesoplasma florum L1*	0.8	27	11-Jan	Closely related to pathogen
	Spirochaetes	*Brachyspira hyodysenteriae WA1*	3.0	27		Swine dysentary
	Archaea	*Methanococcus voltae A3*	1.9	29		Methanogen
	Proteobacteria	*Campylobacter jejuni RM1221*	1.8	30		Food contamination
	Archaea	*Methanocaldococcus jannaschii DSM 2661*	1.8	31		Methanogen
	Firmicutes	*Staphylococcus epidermidis ATCC 12228*	2.6	32	11-Jan	Antibiotic detection
	Firmicutes	*Bacillus cereus ATCC 14579*	5.4	35		Food contamination
	Firmicutes	*Oenococcus oeni PSU-1*	1.8	38		Wine fermentation
	Proteobacteria	*Vibrio fisherii*	4.3	38	11-Sep	Squid symbiont
	Firmicutes	*Streptococcus pneumoniae JJA*	2.1	40	11-Dec	Pneumonia
	Proteobacteria	*Photorhabdus luminescens subsp. laumondii*	5.7	43	11-Dec	Nematode symbiont
	Firmicutes	*Bacillus subtilis subsp. subtilis str. 168*	4.2	44	10-Dec	Soil bacterium
	Proteobacteria	*Shewanella putrefaciens CN-32*	4.7	45		Marine bacterium
	Proteobacteria	*Vibrio cholerae*	4.1	48	11-Sep	Cholera
	Proteobacteria	*Teredinibacter turnerae CS30*	5.2	51	12-Nov	Mollusk symbiont
	Cyanobacteria	*Thermosynecoccus elogatus*	2.5	54	11-Dec	Cell fusion, thermophile
	Proteobacteria	*Serratia proteamaculans 568*	5.5	55		Pneumonia association
	Proteobacteria	*Agrobacterium vitis S4*	6.3	58		Crown gall disease (grapes)
	Proteobacteria	*Agrobacterium tumefaciens str. C58*	5.7	59	11-Jun	Tumor-inducing bacteria (plants)
	Proteobacteria	*Rhizobium sp. NGR234*	6.9	62		Nitrogen fixation
	Proteobacteria	*Pseudomonas fluorescens Pf-5*	7.1	63	11-Feb	Disease related
	Euryarchaeota	*Haloferax volcanii*	4.0	66	11-Aug	High salt growth
	Deinococcus-The	*Deinococcus radiodurans R1*	3.2	67	11-Feb	Radiation tolerant
	Proteobacteria	*Burkolderia cenocepacia HI2424*	7.8	67	11-Sep	Cystic fibrosis pathogen
	Proteobacteria	*Caulobacter crescentus NA1000*	4.0	67	11-Jan	Synchronized growth
	Proteobacteria	*Rhodobacter sphaeroides ATCC 17025*	4.5	68	11-Dec	Phototrophic bacteria
	Proteobacteria	*Myxococcus xanthus DK 1622*	9.1	69	11-Dec	High gene duplications
	Actinobacteria	*Kineococcus radiotolerans SRS30216*	5.0	74	11-Feb	Radiation tolerant

Major contributors:

<u>All US tax payers for the past 15 years, and the next five</u>.

Indiana University: Matthew Ackerman, Yves Brun, Jamie Choi, Nicole Coffey, Tom Doak, Clay Fuqua, Pat Foster, Ignasi Lucas, Rohan Maddamsetti, Sam Miller, Sarah Schaack, Amanda Seyfert, Haixu Tang, Abe Tucker, Greg Velicer, Emily Williams.

University of New Hampshire: Kelley Thomas, Vaughn Cooper, Shilpa Kulkarni, Krystalynne Morris, Kazufusa Okamoto, Way Sung.

University of Utah: Joe Dickinson.

Max Planck Institute, Tubingen: Detlef Weigel, Korbinian Schneeberger, Stephan Ossowski, Norman Warthmann.

Universidade Federal do Rio de Janeiro: Marcus Senra, Carlos Soares.

Harvard University: Daniel Hartl, Christian Landry, Eric Dopman.

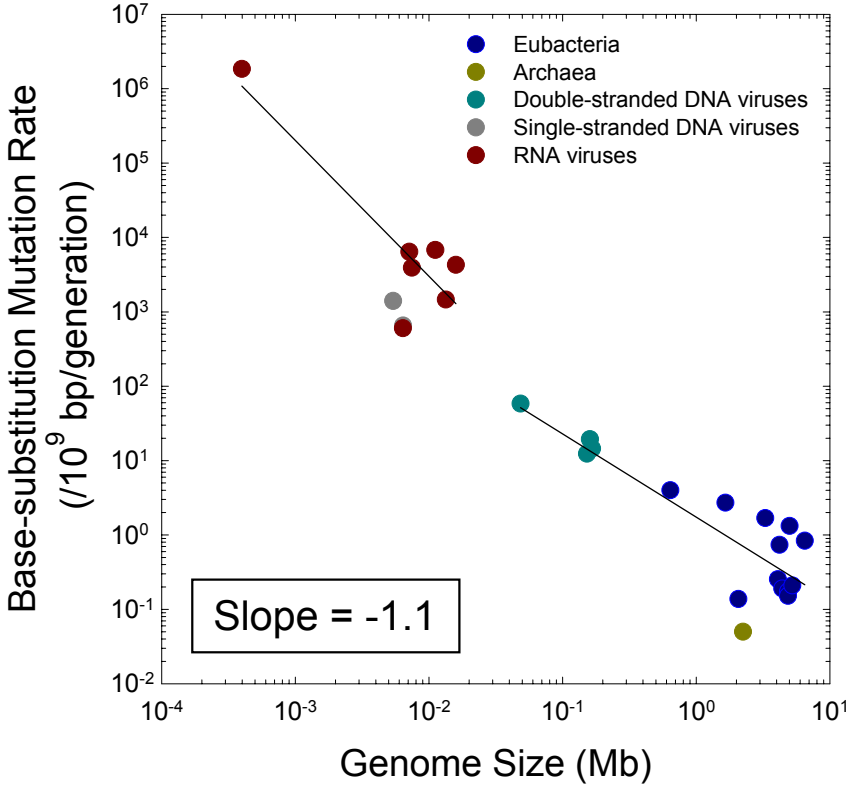
University of Florida: Charles Baer.

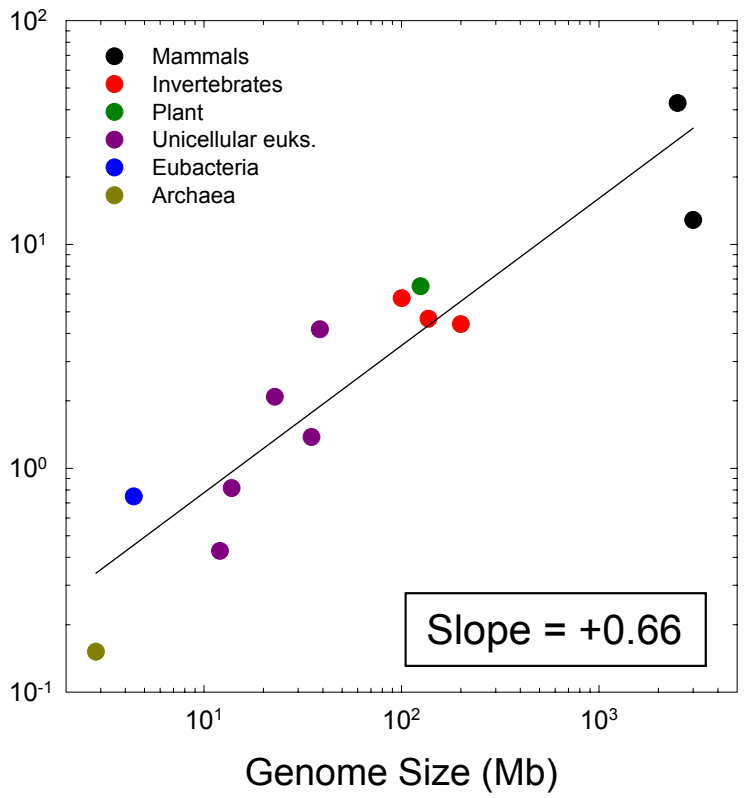
Oregon State University: Dee Denver.

University of Minnesota: Ruth Shaw.

- The *average* number of mutations per genome per generation is roughly constant in **noneukaryotic microbes**, in accordance with Drake's hypothesis.

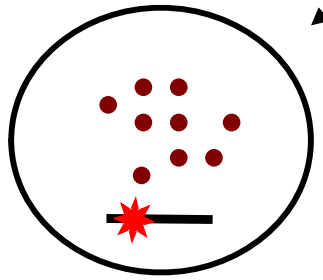
- The mutation rate per nucleotide site increases with genome size in **eukaryotes**, yielding a dramatic increase in the genome-wide mutation rate per generation.





Trends in Genetics, 2010

Asexual Populations: the selective disadvantage of a weak mutator allele = the increase in the genome-wide deleterious mutation rate



ΔU, increase in genome-wide rate of deleterious mutation

s, rate of removal by selection

Excess number of mutations at equilibrium = $\Delta U / s$

X

Effect / mutation = s

Total effect on fitness = ΔU

Sexual Populations: the selective disadvantage of a mutator allele is much smaller, $2s \cdot \Delta U$, because recombination prevents the buildup of linked mutations.

Sexual populations: The induced selection coefficient on a mutator / antimutator allele

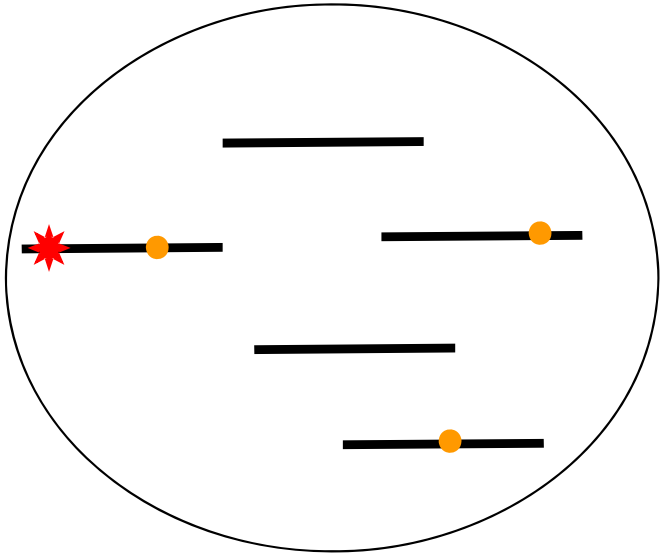
$\cong$ the excess genomic mutation rate to deleterious alleles

x the average deleterious effect of a heterozygous mutation

x 2 generations of association

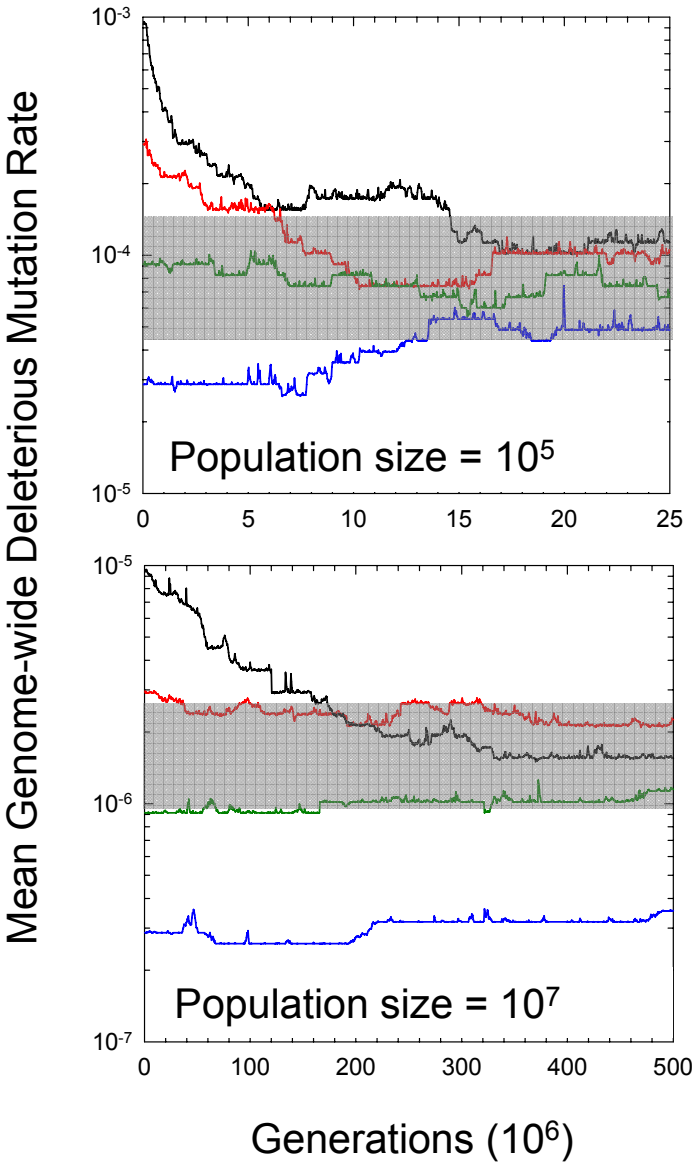
$= 2 \cdot \Delta U \cdot s.$

- the heterozygous effect of a deleterious mutation (s) $\cong$ 0.001 to 0.01;
- the genomic mutation rate to deleterious alleles (U) $\cong$ 0.001 to 1.0;
- small modifications to the mutation rate (ΔU) will be $<< 10^{-4}$;
- the selective advantage of a weak antimutator allele will often be $< 10^{-6}$.



Transient Effects
of Induced Mutations

Quasi-equilibrium Mutation Rates Resulting From Deleterious-mutation Load

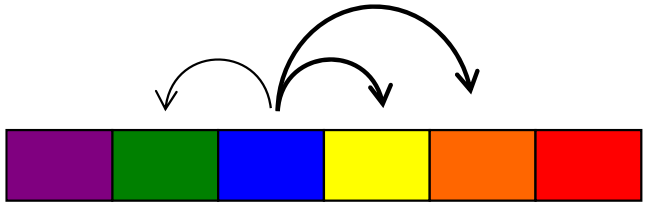


Effective selection for antimutators

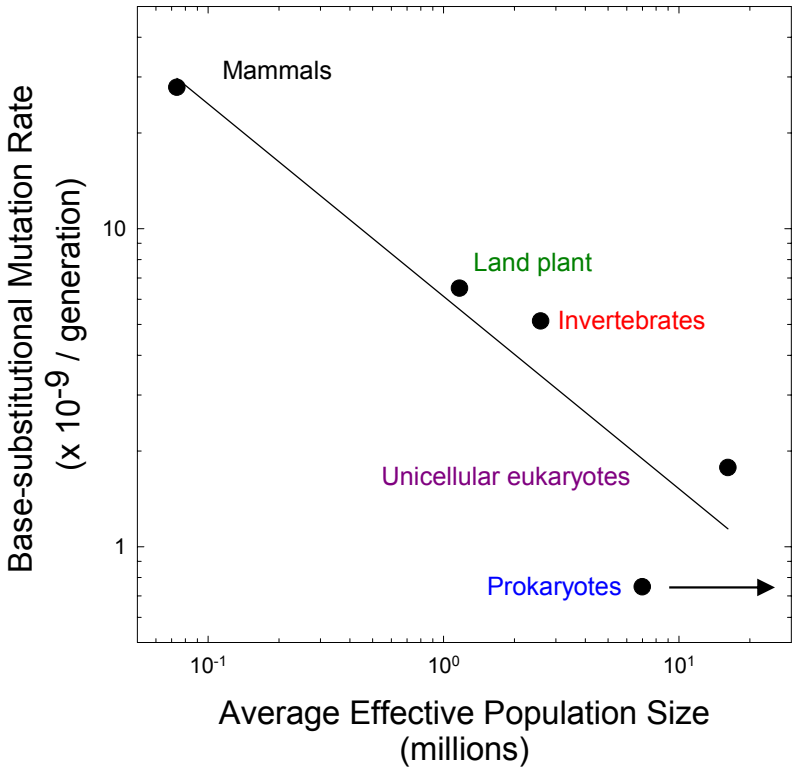
DRIFT BARRIER

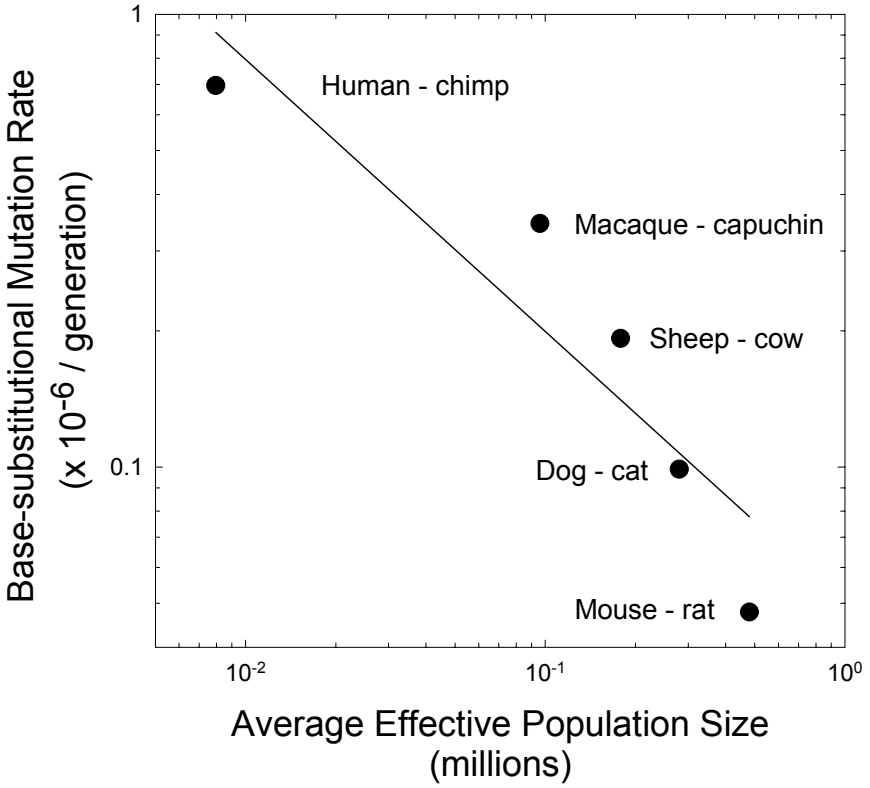
Biased production of mutators

- Equilibrium mutation rate is inversely proportional to the effective population size.
- Asymmetry of rate of approach to equilibrium.



The Per-generation Mutation Rate Is Inversely Proportional to the Average Effective Population Size of a Lineage



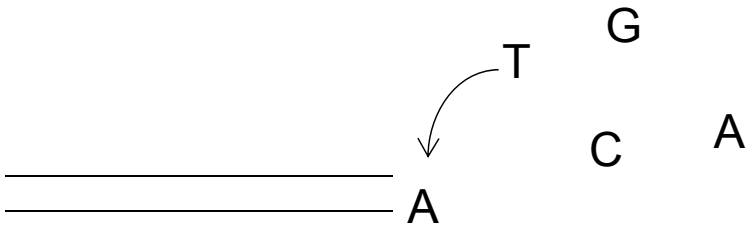


Nuclear genes: Trends in Genetics, 2010

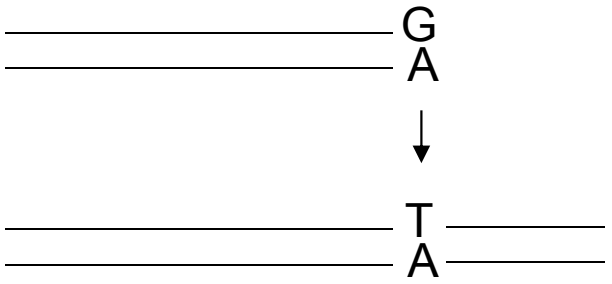
Mitochondrial genes: Piganeau and Eyre-Walker (2009)

The Three Molecular Lines of Defense Against Mutation

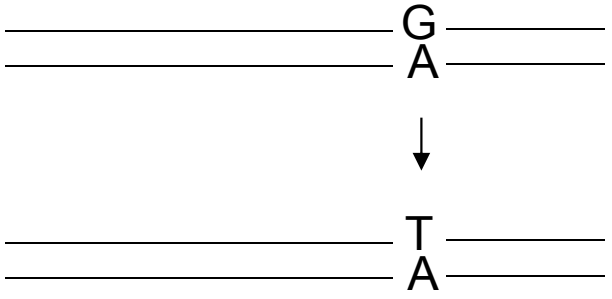
1) Polymerase base-incorporation fidelity:



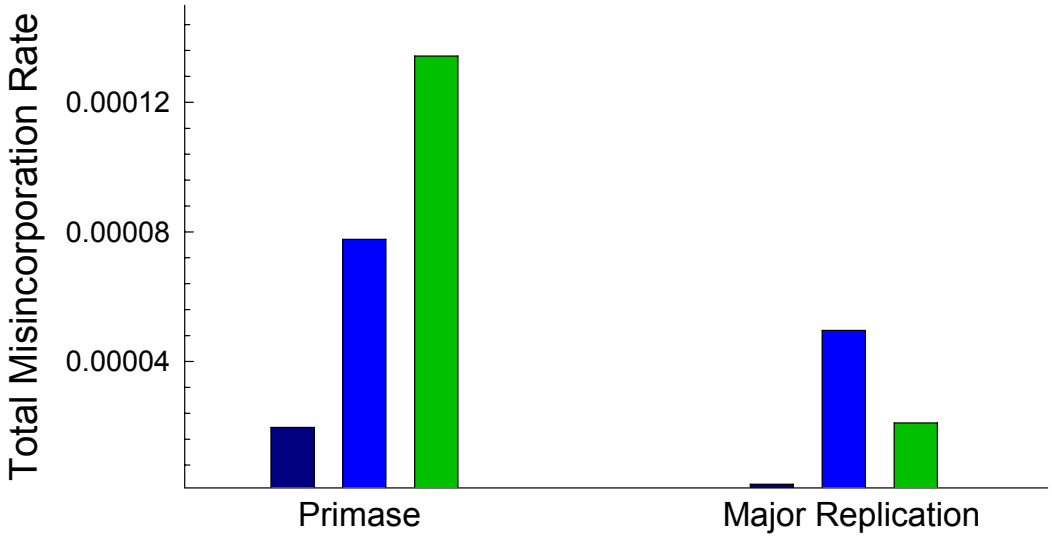
2) Polymerase proofreading:

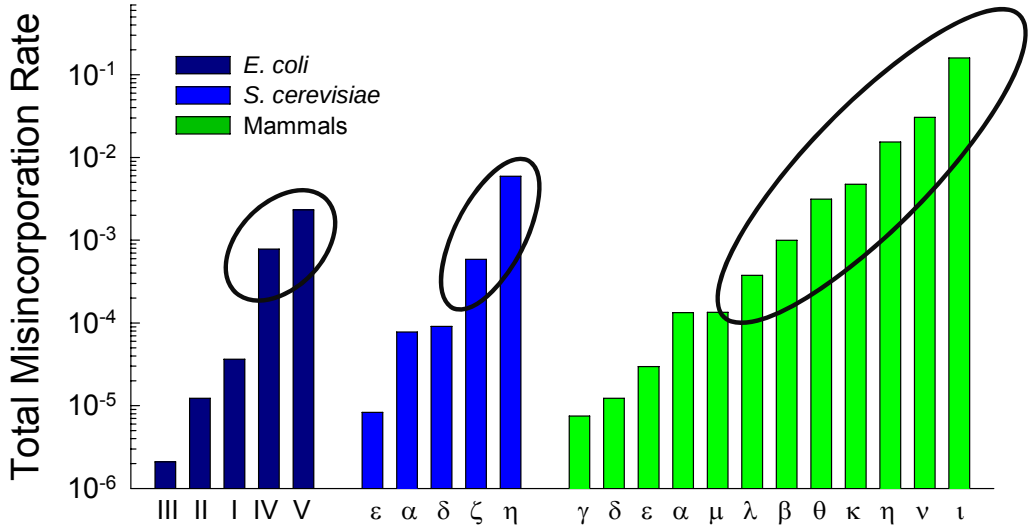


3) Post-replicative mismatch repair:



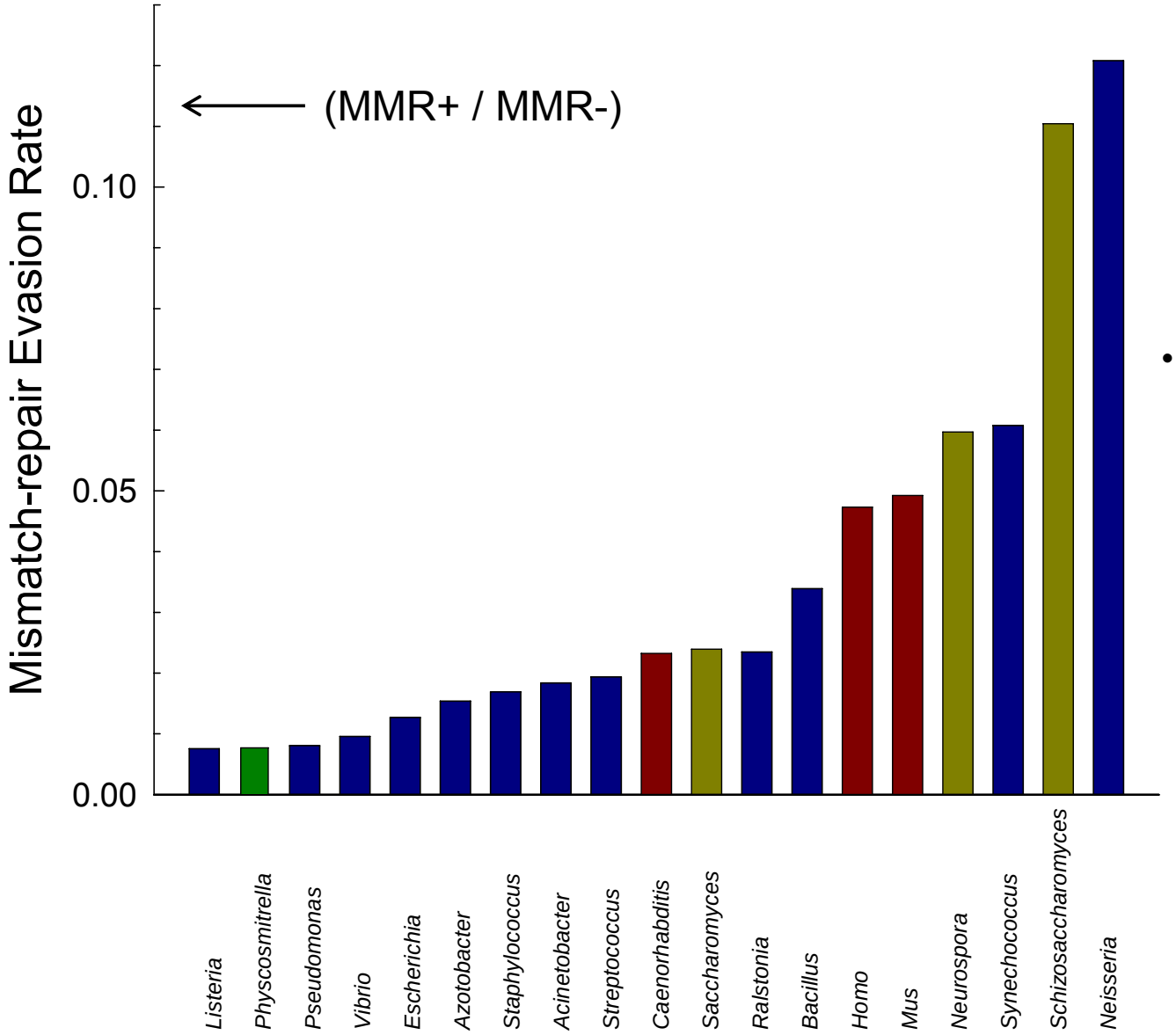
Polymerase Error Rates Are Magnified in Eukaryotes





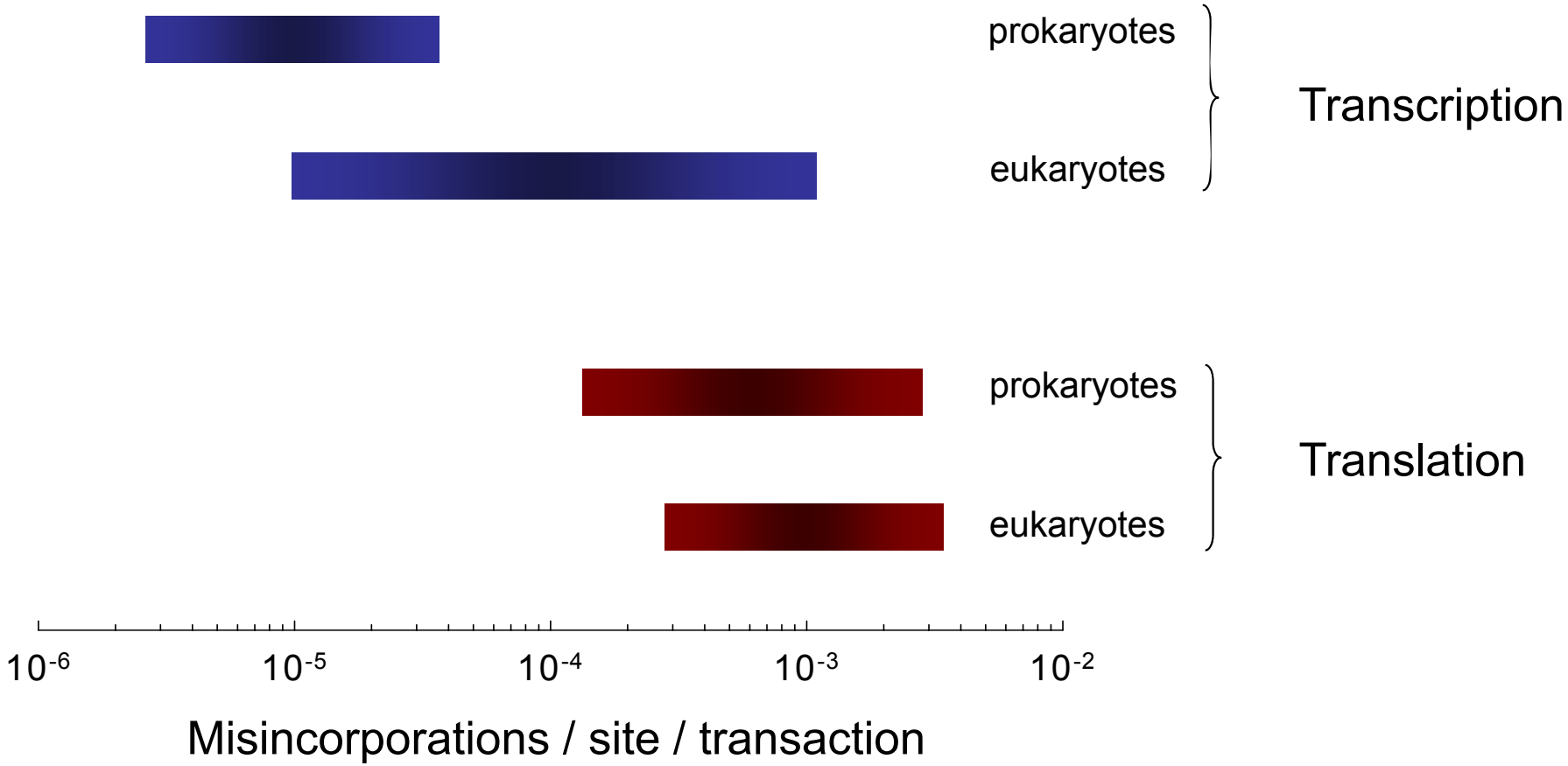
Polymerases used in DNA repair are highly error prone, consistent with the drift hypothesis: enzymes involved in fewer nucleotide transactions experience less selection for fidelity.

The Efficiency of the Third Line of Defense – Mismatch Repair – is Much Lower Than That at the Polymerization and Proof-reading Steps



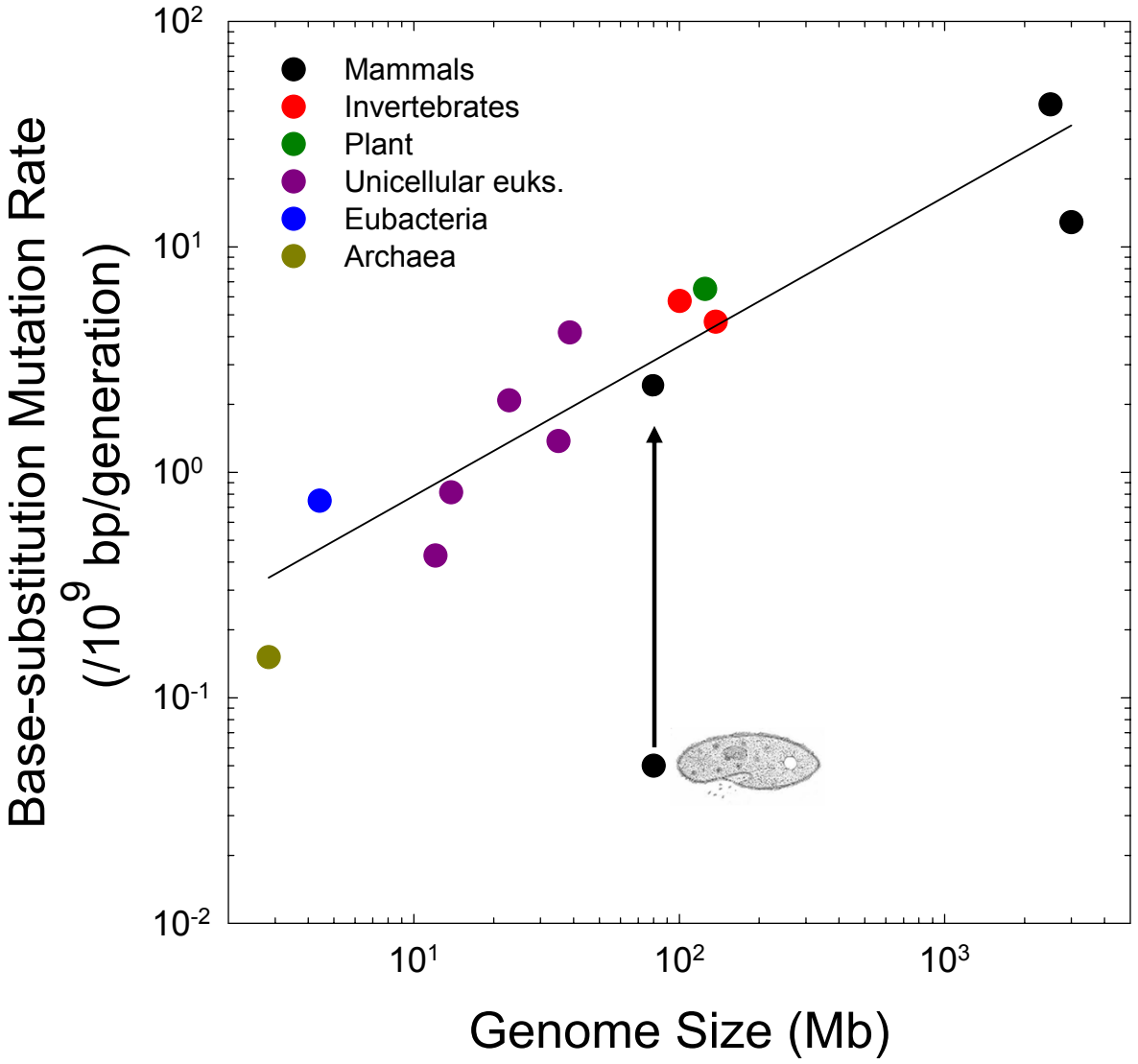
- The fidelity of this downstream repair pathway is >100x lower than that for the upstream polymerase, consistent with the drift hypothesis.

Transcription and Translation Error Rates Are Orders of Magnitude Higher Than Replication Error Rates

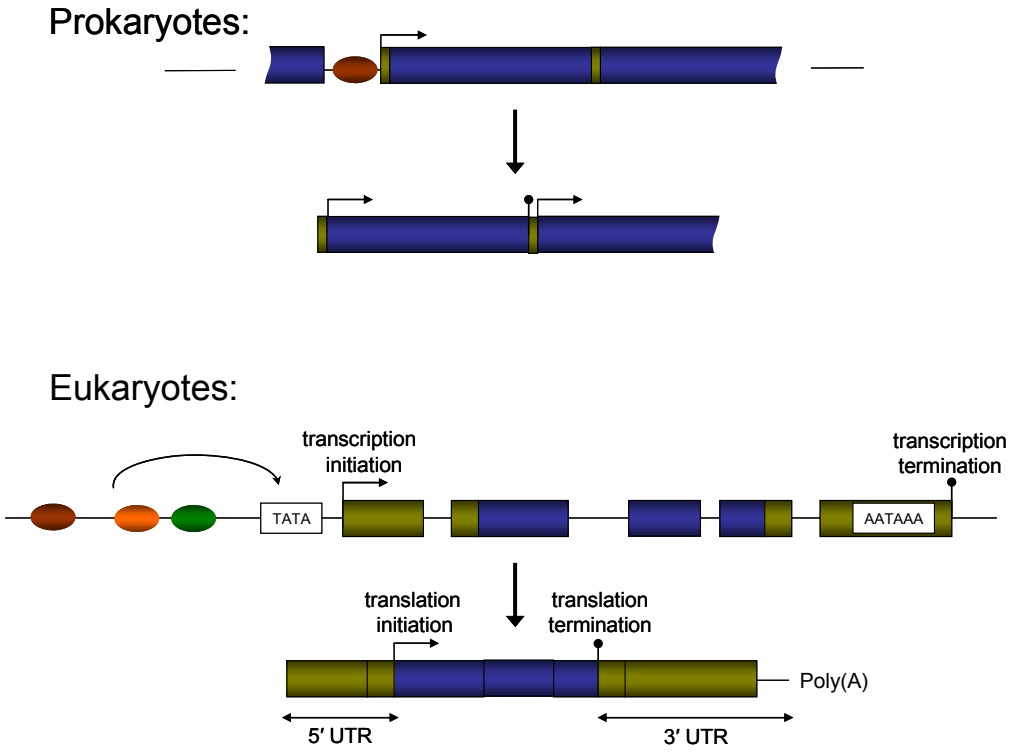


- Because products of transcription and translation are more transient than inherited germ-line mutations, selection to reduce error rates at these levels is less efficient.

Paramecium Has the Lowest Known Mutation Rate



The Origin of Gene-structure Complexity by Nonadaptive Mechanisms



- 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.

- Subsequent to passive establishment, such alterations can sometimes provide the substrate for adaptive evolution.

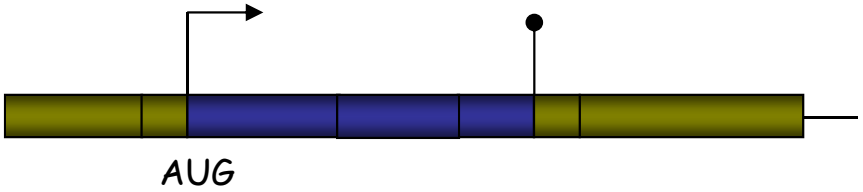
The Mutational Cost of Genomic Embellishments

- **The selective disadvantage of a mutational hazard** – alleles with increased structural complexity involving *n* key nucleotide sites have elevated mutation rates to defective alleles = *nu*, where *u* = mutation rate per nucleotide site.

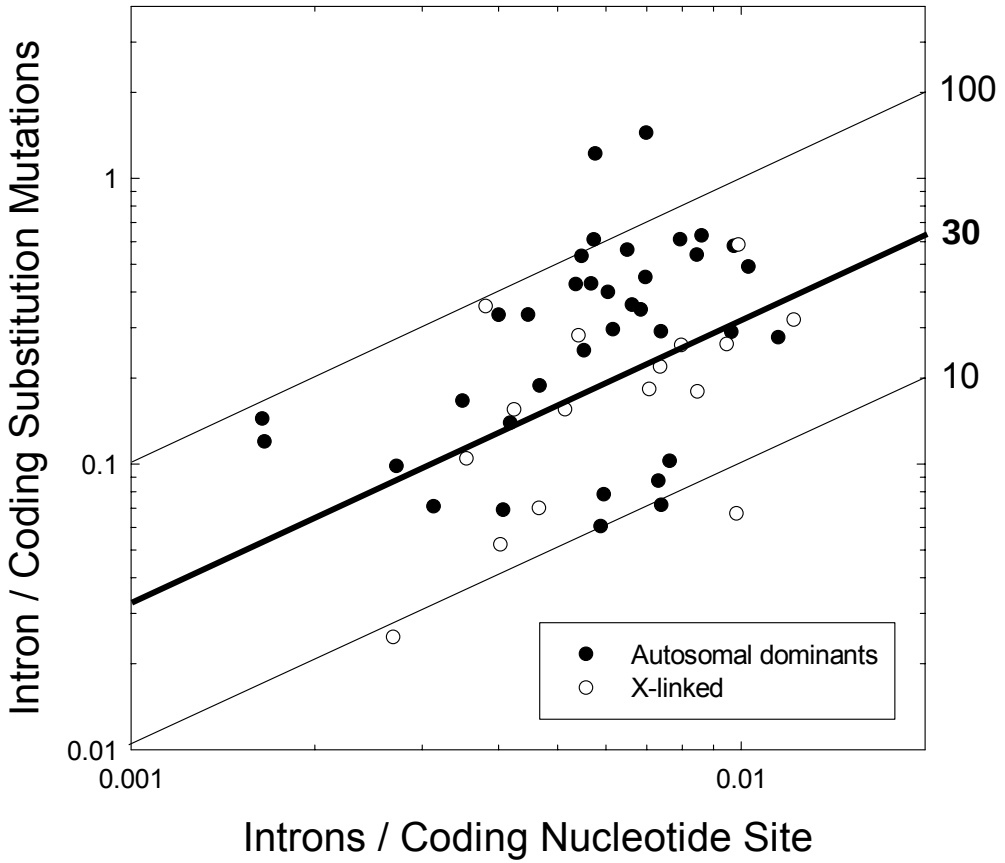
n ≈ 30 for introns

4 for 5′ UTRs

?? nonfunctional DNA



Cost of an Intron – equivalent to adding 10 to 100 essential nucleotides to a gene.



Results from surveys of *de novo* defective alleles for monogenic human genetic disorders.

e.g., Alagille syndrome, neurofibromatosis type 1, Waardenburg syndrome, retinoblastoma, tuberous sclerosis, adrenoleukodystrophy, Alport syndrome, hemophilia A, HPRT deficiency

Ratio of relative mutation rates to relative target numbers = effective cost of an intron in units of key nucleotide sites.

• About 8% of human deaths are caused by introns – exceeds the total from accidents and war.

2010, PNAS

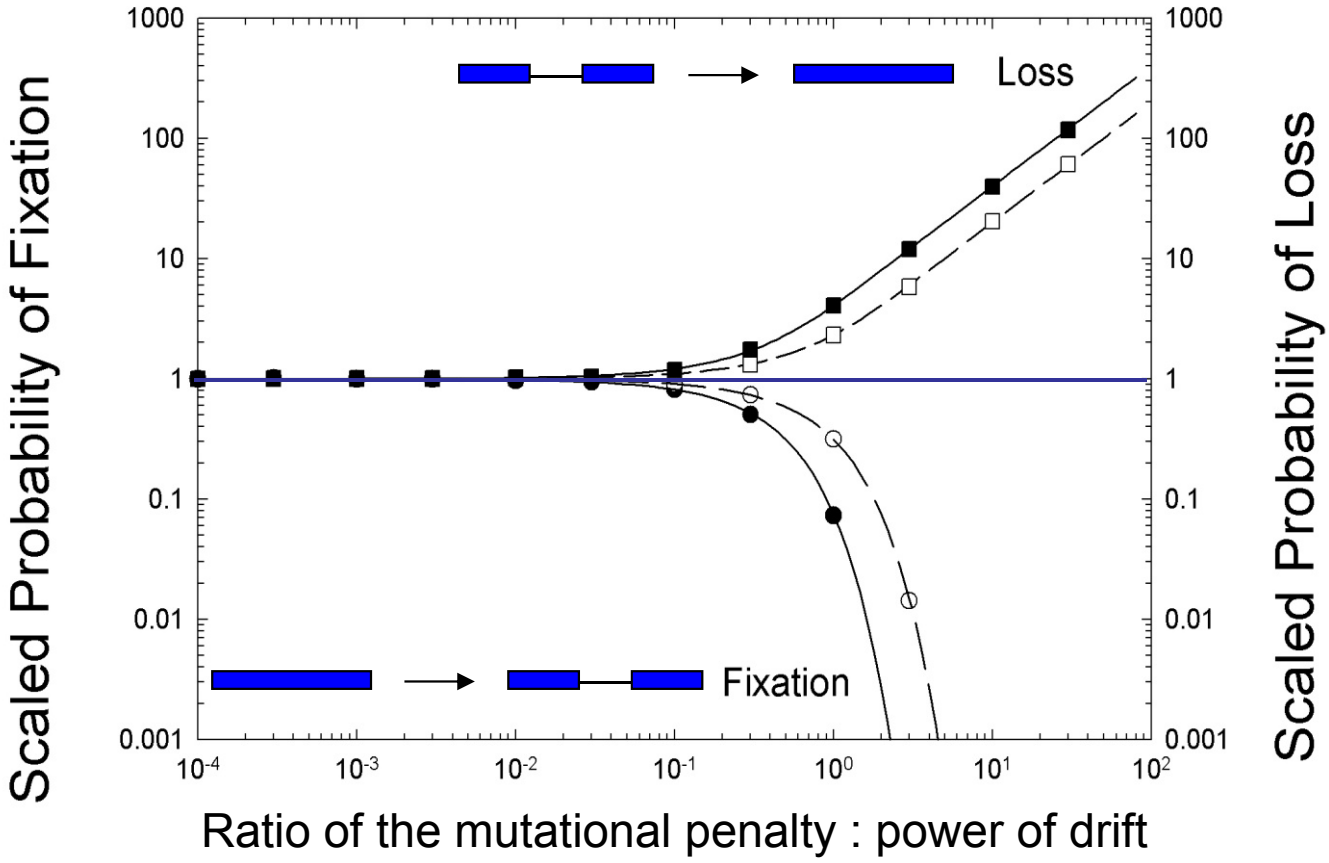
The probability of fixation of a mutationally harmful intron declines with increasing population size (N).

Probability of intron fixation = $2s / (e^{4Ns} - 1)$
Probability of intron loss = $2s / (1 - e^{-4Ns})$

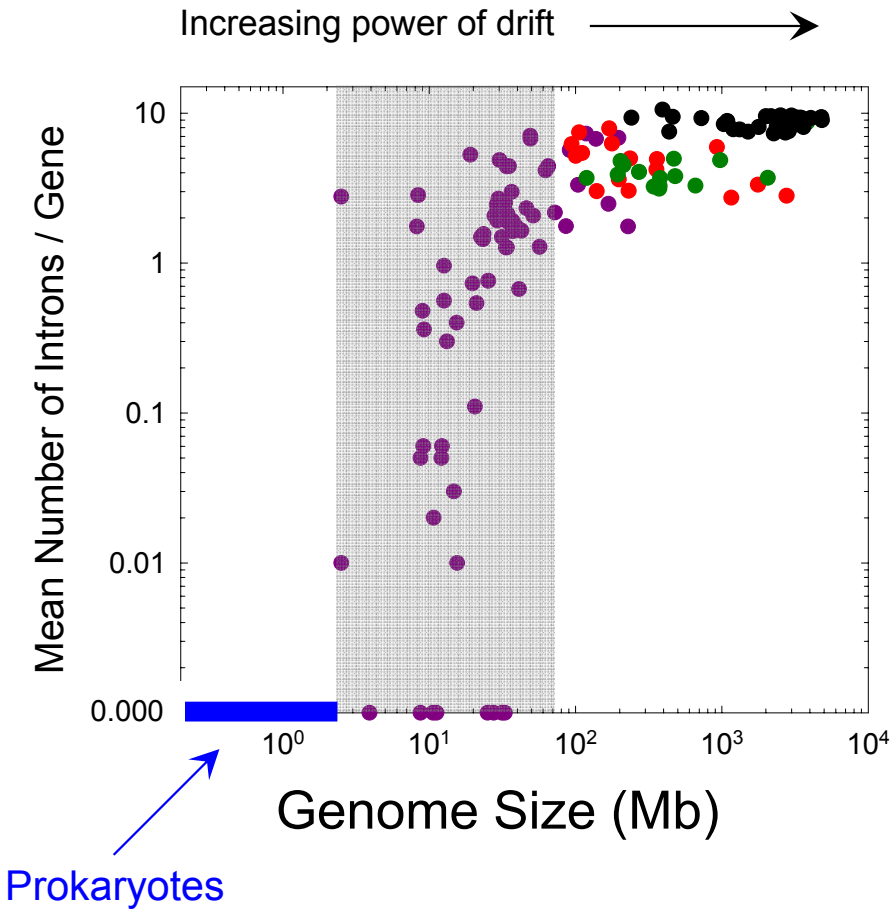
$s = nu$ = excess mutation rate to defective alleles

Effective Neutrality:

Probability = 1/(2*N*) →

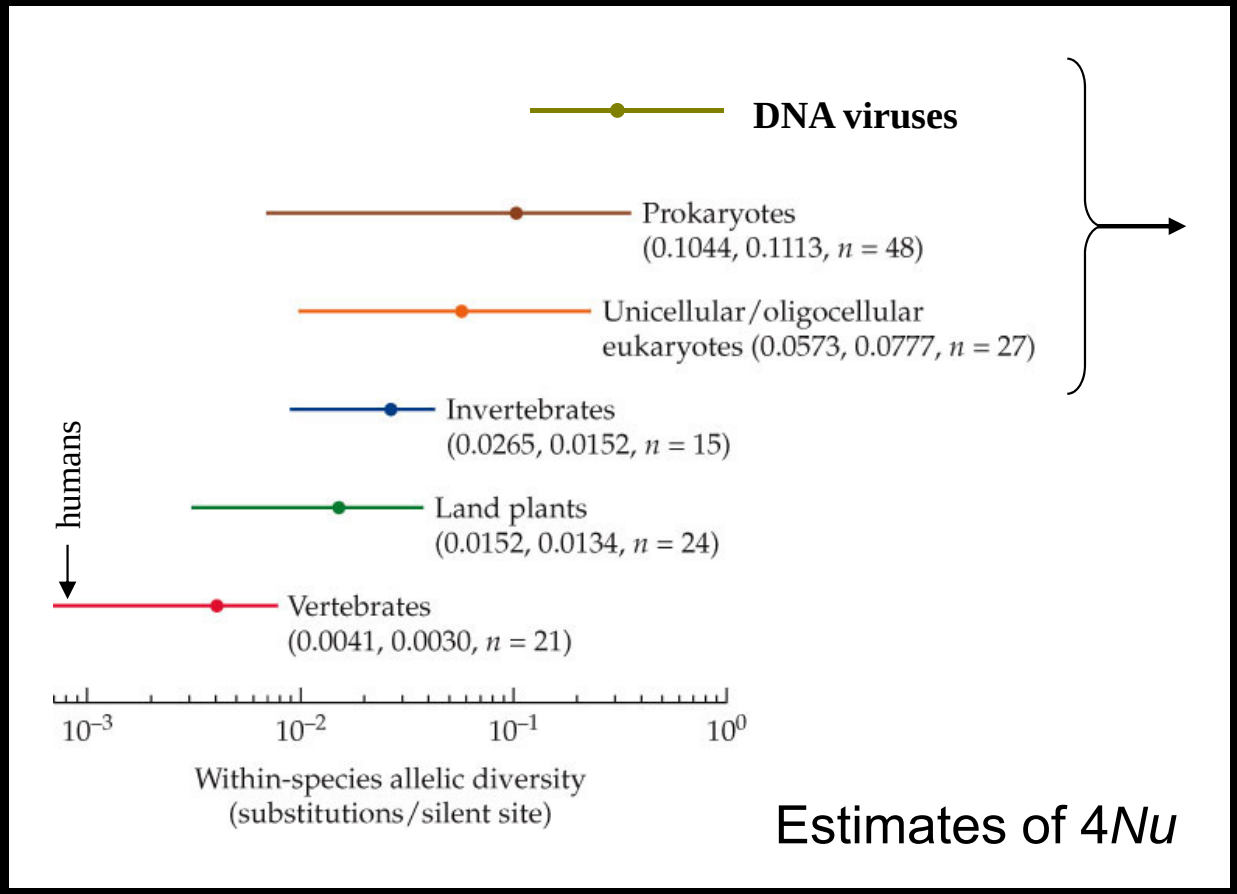


Threshold Behavior for Intron Colonization

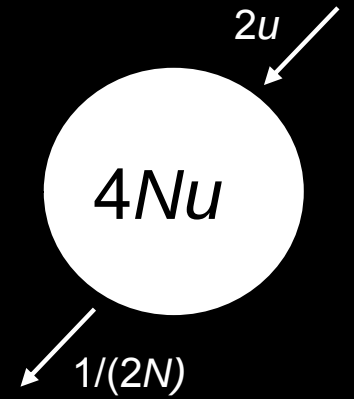


- Unicellular eukaryotes
- Invertebrates
- Plants
- Vertebrates

Estimates of the ratio of the power of mutation ($2u$) to the power of random genetic drift ($1/2N$) from standing population-level nucleotide heterozygosity at silent sites.



Estimates of $4Nu$

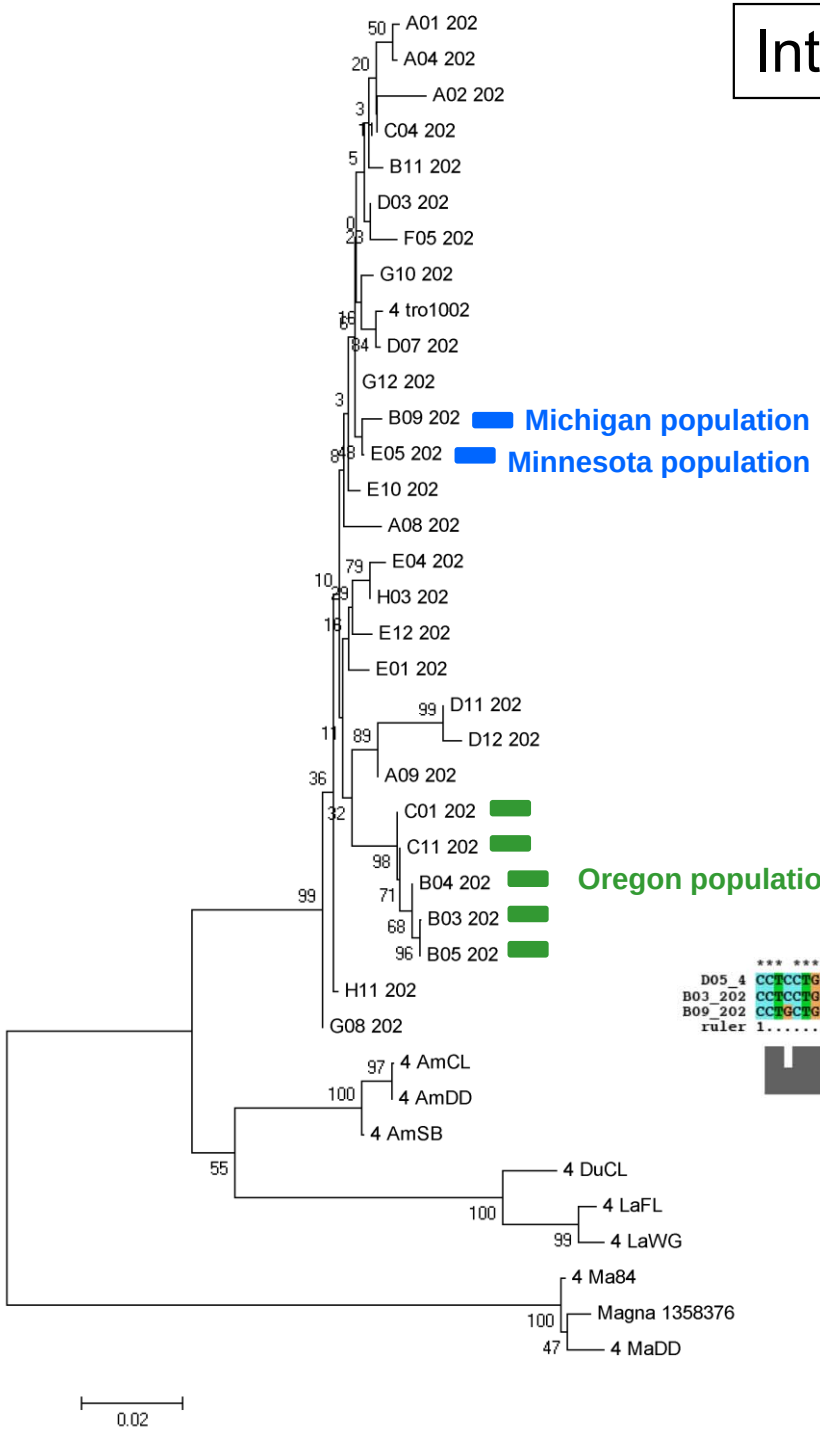


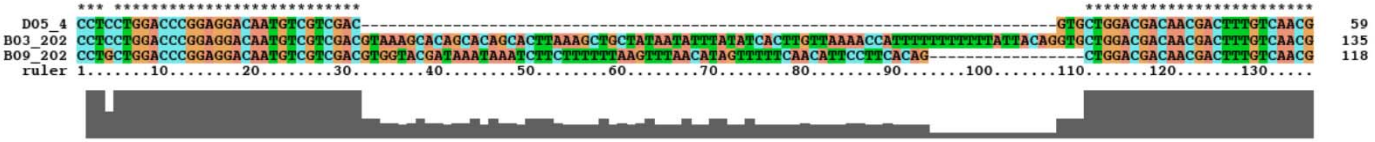
At equilibrium, average allelic divergence at neutral sites =

ratio of the power of mutation to the power of random genetic drift.

Intron gain/loss in *Daphnia pulex*

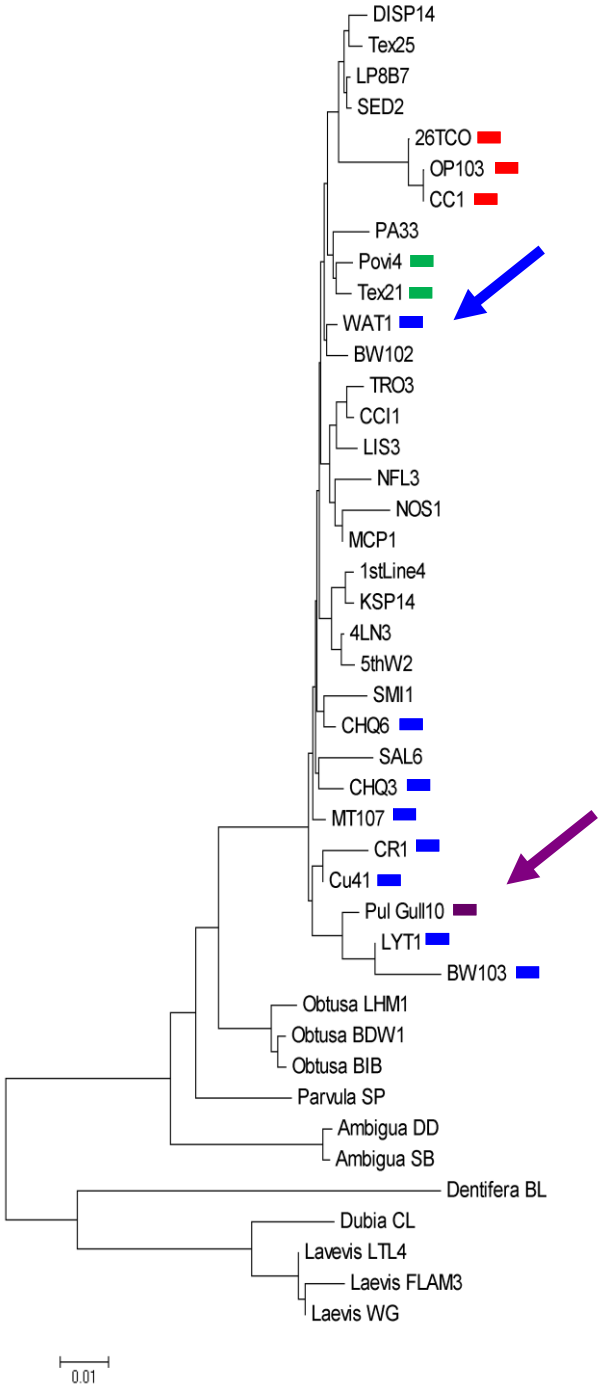
- Dozens of cases of intron gain.
- Several cases of parallel gains at the same site.
- Few cases of intron loss.
- Most of the newly gained introns are in a set of isolated Oregon populations that have been through a prolonged population bottleneck.





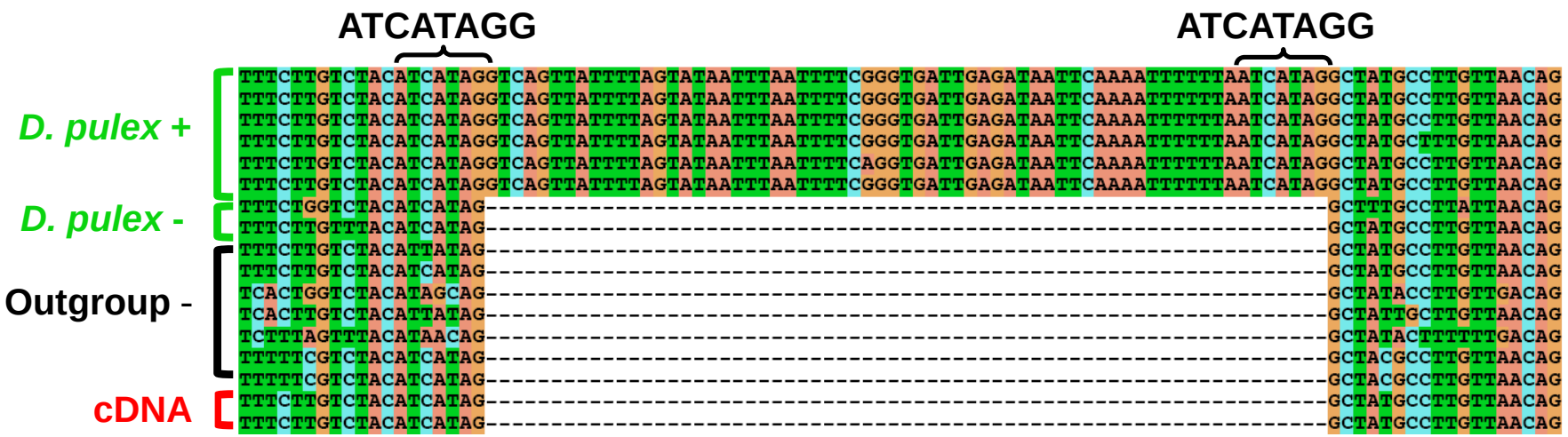
Bootstrap tree of a locus showing two newly gained introns at nearly identical sites.

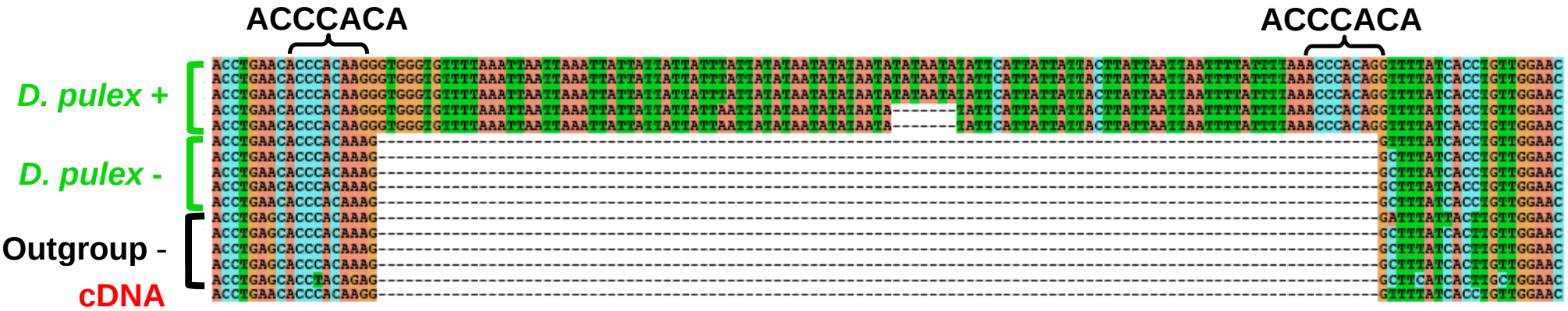
Li, Tucker, et al., 2009, Science



- Four independently gained intron sequences at the same site.
- One apparent intron replacement.
- One apparent gene conversion.

Short Direct Motifs Associated with Intron Origins

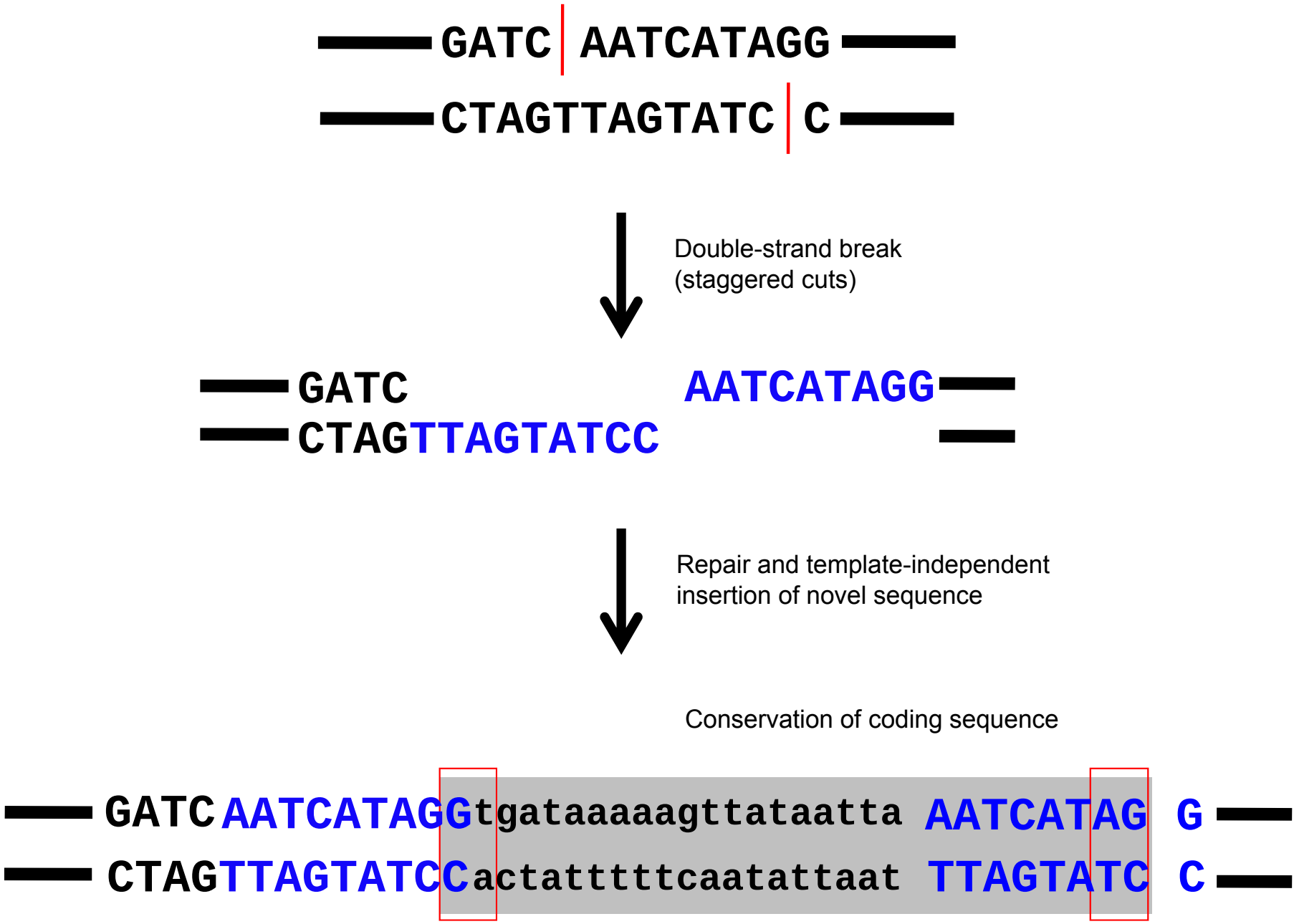




- ~60% of the newly gained introns have short repeats.
- These short repeats are 5 to 22 bp long.
- Each intron gain has a unique repeat.

Li, Tucker et al., 2009, Science

The Double-Strand Break Model for the Origin of Introns



OPEN ACCESS Freely available online PLoS GENETICS

Nonsense-Mediated Decay Enables Intron Gain in *Drosophila*

Ashley Farlow[9], Eshwar Meduri[9], Marlies Dolezal, Liushuai Hua, Christian Schlötterer*

Institut für Populationsgenetik, Veterinärmedizinische Universität Wien, Wien, Austria

J Mol Evol
DOI 10.1007/s00239-010-9391-6

Evaluation of Models of the Mechanisms Underlying Intron Loss and Gain in *Aspergillus* Fungi

Lei-Ying Zhang · Yu-Fei Yang · Deng-Ke Niu

The Origin of Gene-structure Complexity by Nonadaptive Mechanisms

- **The Drift-barrier Hypothesis**: once a molecular / cellular feature has been refined to the point where the fitness advantages of any further beneficial mutations are smaller than the power of drift, the limit to molecular perfection has been reached.

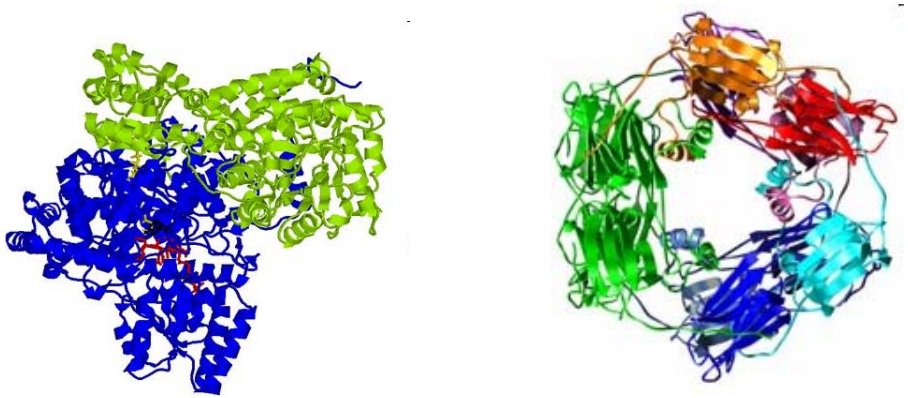
- **The Mutational-hazard Hypothesis**: mutation pressure can drive the establishment of more complex gene structure and genomic complexity, provided the mutational cost of the embellishment is less than the power of drift.

- Can these general principles help explain higher-order features of organismal diversity – protein characteristics and cellular infrastructure?

How Did The Complex Cellular Features of Eukaryotes Become Established?

- Most cellular components and pathways are assembled from protein subunits derived from the same gene or from related loci arising via gene duplication, rather than from products of unrelated genes:

the flagellum
the nuclear pore complex
the cytoskeleton
the proteasome
chaperones
ion channels
nucleosomes



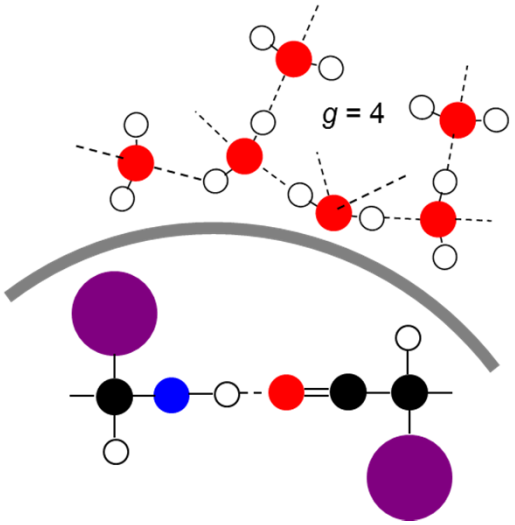
- Potential advantages to complex formation:

increased structural size and diversity,
reduced problems of folding single large proteins,
increased flexibility for allosteric regulation,

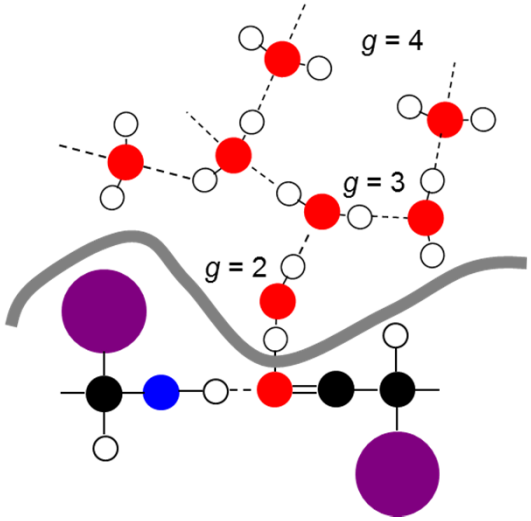
***compensation for structural deficiencies in monomeric subunits*.**

- Proteins with a propensity to oligomerize can also come at a cost:
  - Human disorders involving the production of inappropriate protein aggregates include Alzheimer's, Parkinson's, and amyotrophic lateral sclerosis (ALS).

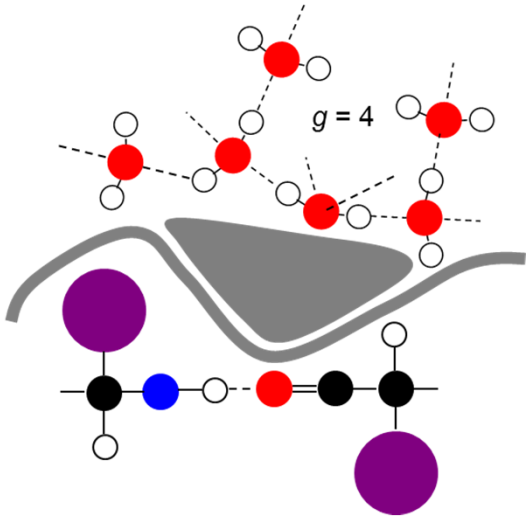
Can Nonadaptive Processes Lead to the Evolution of Protein Complexity?



Well-wrapped

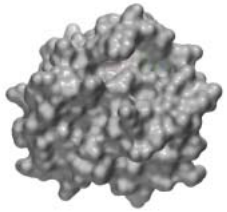


Exposed

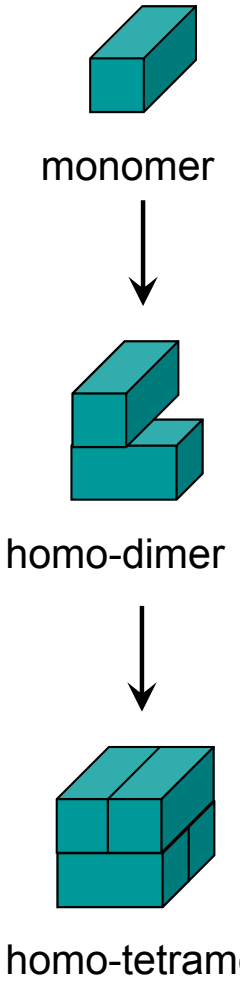


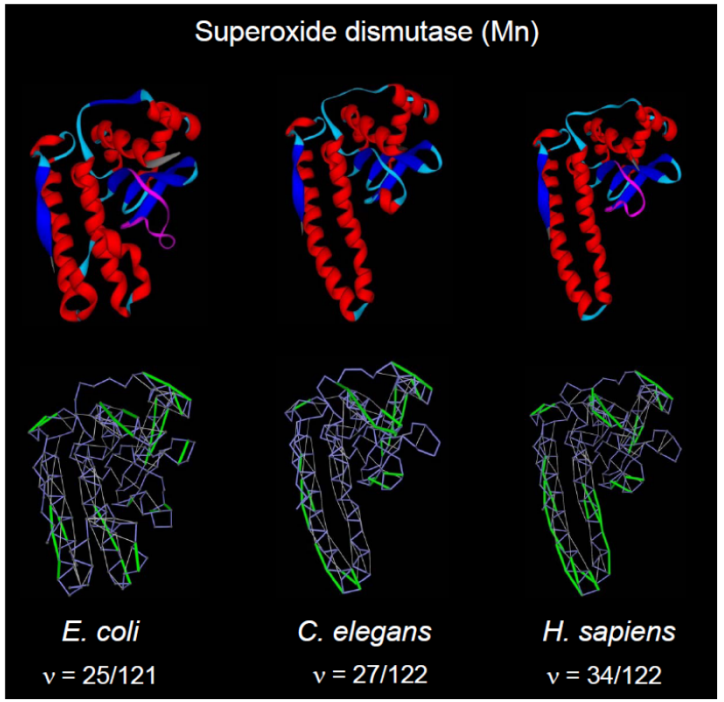
Tension relieved

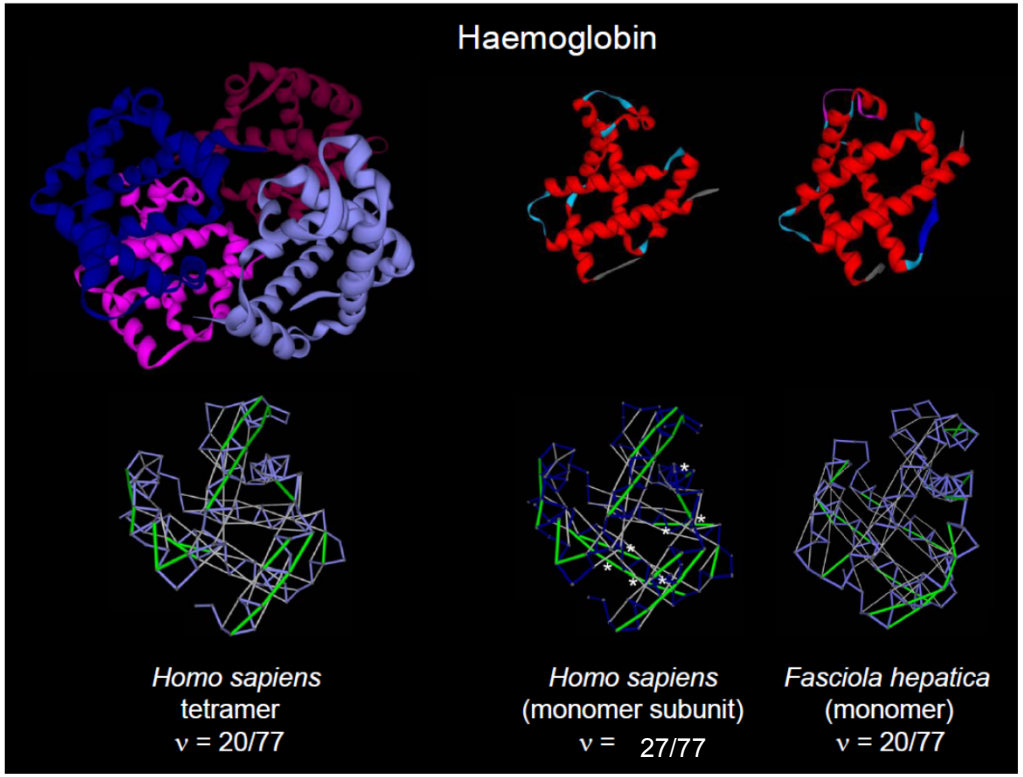
Exposed backbone hydrogen bonds and/or the appearance of hydrophobic surface residues often reduce protein functionality, increasing stickiness and the potential for protein-protein interactions.



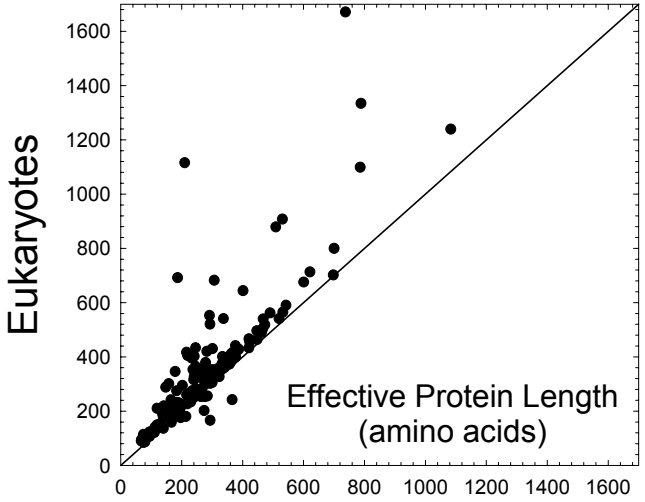
The Origin of Protein Complexes as a Means for Relieving Backbone Deficiencies or Hiding Hydrophobic Surface Residues

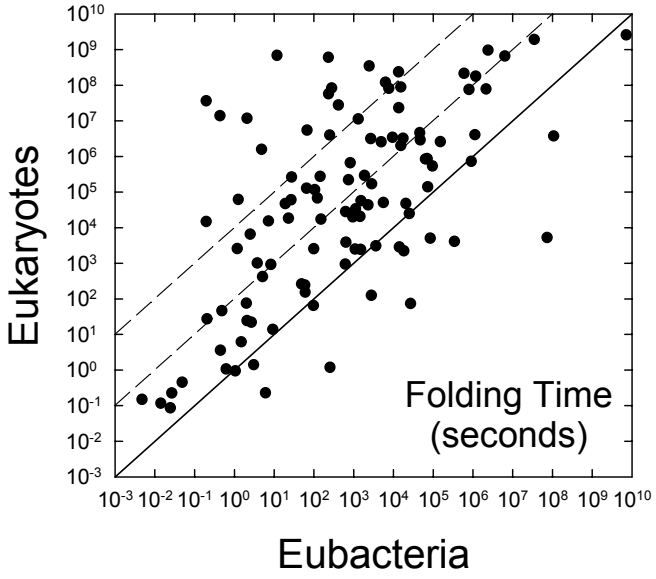


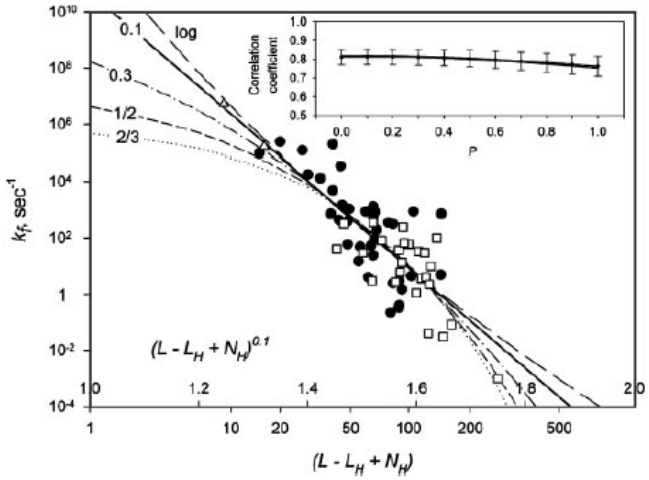




- Expansions of linker regions cause eukaryotic proteins to be longer than their Eubacterial orthologs.
- This substantially increases the folding time, and the propensity for aggregation.







Assumes alpha helices fold essentially instantaneously.

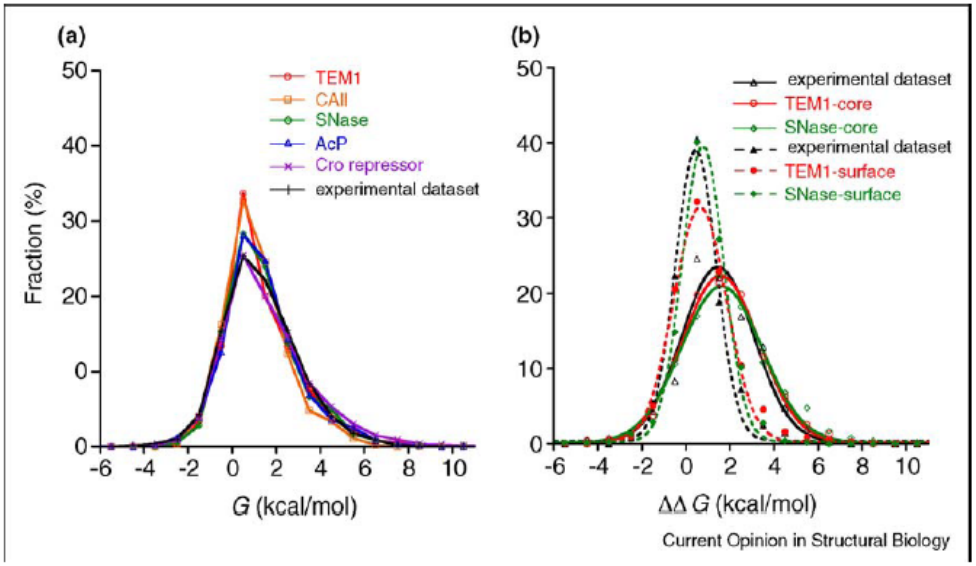
L = total chain length
L_H = number of alpha helices
N_H = weighted number of residues per helix
(1 to 3)

(Ivankov and Finkelstein, PNAS, 204)

With Jean-Francoi Gout.

Proteins Are Typically on the Margin of Stability, and Most Mutations Alter Aspects of Folding And Stability Rather Than Modifying the Functional Core

- The average stability of proteins is on the order of ΔG = 10 kcal/mol.
- ΔΔG for single amino-acid substitution mutations is often > 2 kcal/mol, which is near the point at which protein stability is compromised.
- Mutations to surface residues are less destabilizing than those to the core.
- Smaller proteins are less destabilized by mutations, presumably because of the relative reduction in core size.



The universal distribution of stability effects of mutations [31]. ΔΔ*G* values are presented in histograms using 1 kcal/mol bins. **(a)** The predicted ΔΔ*G* values by FoldX for all possible mutations in many proteins (shown are few characteristic examples), and the experimentally measured ΔΔ*G* values for 1285 mutations, all give similar asymmetric distributions with larger destabilizing shoulders (ΔΔ*G* > 0). **(b)** Separated ΔΔ*G* distributions of core and surface residues. Residues were divided according to their accessible surface area (ASA) values, and the ΔΔG values for all possible mutations were arranged in histograms and fitted to a single Gaussian.

Tokuriki et al., 2007, J. Mol. Biol.; Tokuriki and Tawfik, 2009, Curr. Opin. Struct. Biol.

Evolution of the “Marginal Stability” of Proteins to the Drift Barrier

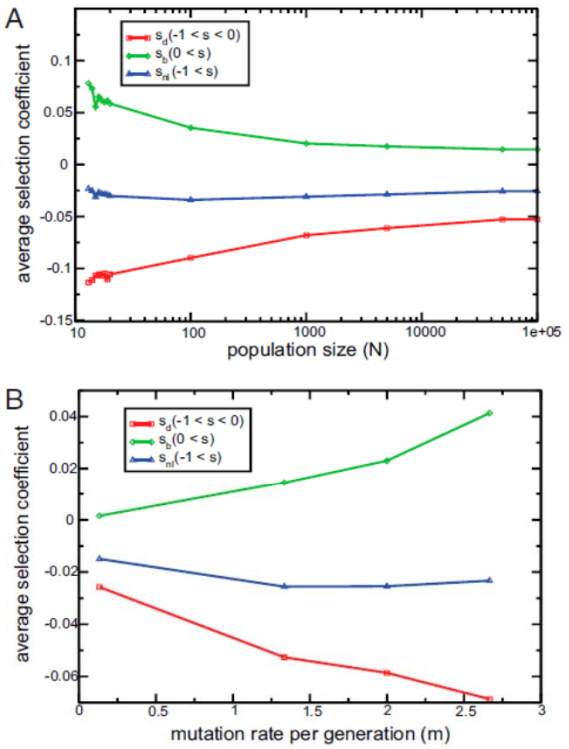


Fig. 4. Dependence of selection coefficients on population size (N) and mutation rate (m). (A) The strength of mutations decreases with N and (B) increases with m. Regardless of N and m, mutations altering protein-folding thermodynamics have substantial fitness effects (e.g., ~2% among all non-lethal mutations). Parameters are m = 1.3 (A) and $N = 10^5$ (B).

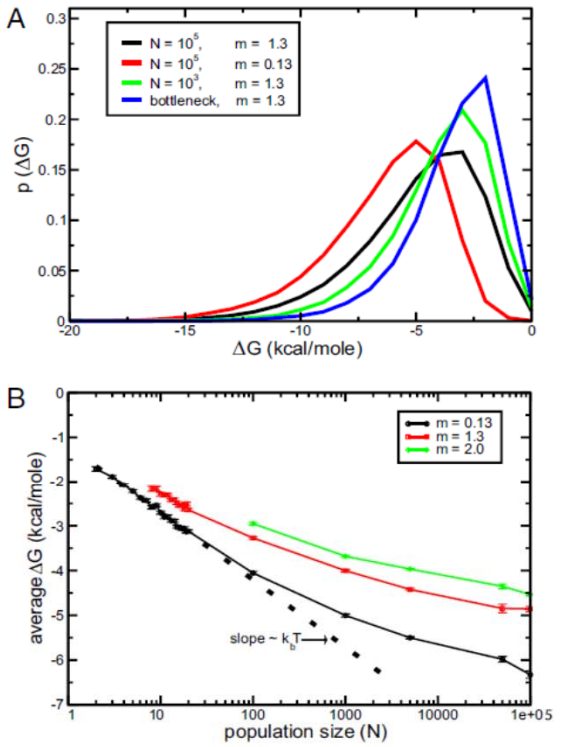


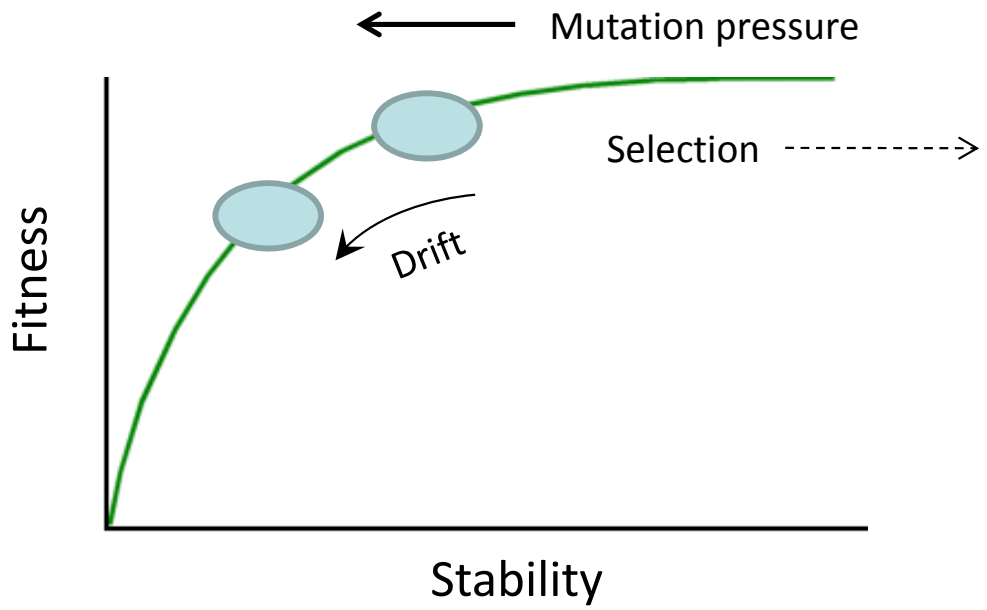
Fig. 5. Protein stability increases with population size (N) and decreases with mutation rate (m). (A) The distribution of stabilities [$p(\Delta G)$] of all proteins in populations with different N and m. Genetic drift and mutational load shift the distribution of evolved stabilities to $\Delta G = 0$. Deep population bottlenecks also decrease protein stability. (B) Dependence of ΔG (averaged over the population and >10 replicates) on N. For small Nm, the population is mostly monoclonal, and ΔG decreases as $\Delta G \sim k_b T \ln N$. The dotted line is fit from the interval $2 \leq N \leq 20$ (slope = 0.65). Values of all parameters are stated on the axes. Error bars are 1 SEM.

- Evolution towards a neutral mutational landscape as the population size increases and the mutation rate decrease.

Wylie and Shakhnovich (2011, PNAS)

Selection-mutation Balance and the Margin of Stability

- One potential explanation for marginal stability is that overly rigid proteins will compromise protein function.
- However, this argument is inconsistent with a number of observations indicating that proteins engineered to have higher stability have normal enzyme function.



Near-neutrality and the Passive Emergence of Organismal Complexity

- The population-genetic environment of multicellular species provides a setting that is conducive to the evolution of gene, genomic, protein and cellular features that may be essentially unattainable in unicellular species, many of which are inherently deleterious but effectively neutral.

- Such alterations impose internal selective challenges that result in changes in the intracellular environment, including the emergence of novel protein structures, numerous surveillance mechanisms, etc., not seen in prokaryotes.

- The *nonadaptive* forces that allow the establishment of new forms of genomic and cellular resources also provide the substrate for natural selection to operate in novel ways.

Drift, Mutation Pressure, and the Origins of Complex Cellular Features

- Extension of the mutational-hazard hypothesis: the evolution of complex cellular features by effectively neutral processes.

- The origin of multimeric protein complexes: homomers and heteromers.

- Extension of the drift-barrier hypothesis: the evolution of layered mechanisms of cellular surveillance, and the myth of the adaptiveness of robustness.