# Mathematical Theory and Scientific Understanding

"All models are wrong, some are useful." – George Box

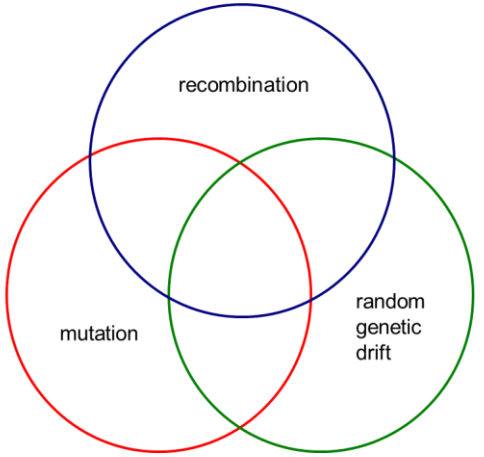
"No theory should fit all the facts because some of the facts are wrong." – Niels Bohr

"What I cannot build, I cannot understand." – Richard Feynman

"You can't just make stuff up. We have a choice to make and the choice is clear."
– Barack Obama, comment on Sarah Palin.

# Population Genetics and Evolutionary Hypotheses:

The Population-genetic Environment




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

- The critical nature of the population-genetic environment.

  Multicellular eukaryotes provide a population-genetic environment that promotes several types of evolution that are difficult, if not impossible, to achieve in most unicellular species.

  With their large population sizes, unicellular species are capable of evolutionary changes that are unattainable by multicellular species.

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

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

**Neutral Constructive Evolution**: Can Complex Structures Arise by Neutral Processes Rather Than Being Promoted For Their Selective Advantages?

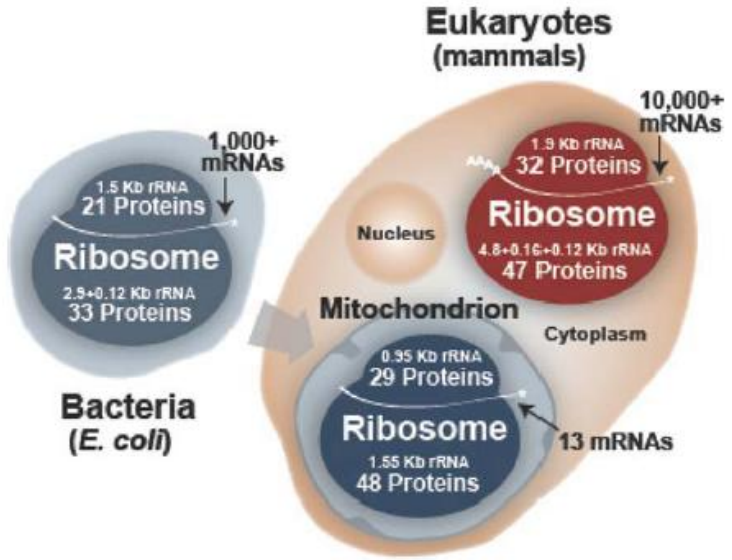



**Figure 3.** Ribosome complexity in bacteria and eukaryotes. The cartoon on the left summarizes the complexity of the ribosome of *Escherichia coli*, on the right, the human cytoplasmic and mitochondrial ribosomes. In each case, the number of proteins comprising the small and large ribosomal subunits is provided, as is the approximate size and number of ribosomal RNA (rRNA) species and the number of messenger RNAs (mRNAs) translated.

Gain in ribosomal proteins in eukaryotes.

Stoltzfus (J. Mol. Evol., 1999)
Archibald et al. (IUBMB Life, 2011)

Fortuitous interaction with B suppresses deleterious mutational effects in A, making A dependent on B.

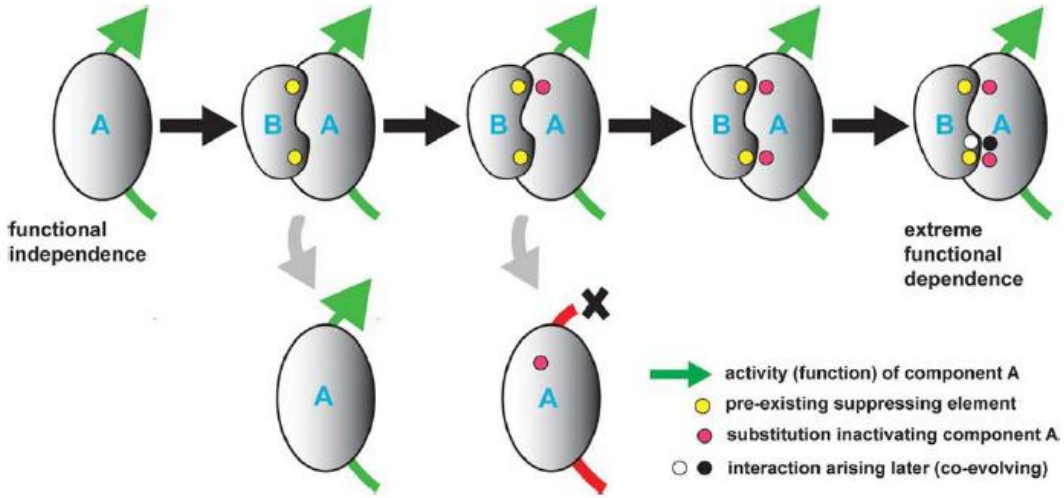



**Figure 1.** Constructive neutral evolution of biochemical complexity. Schematic depicts (i) a generic enzymatic reaction carried out by cellular component A, (ii) fortuitous (and presuppressing) neutral interactions (yellow dots) with component B, (iii) mutation in A (red dot) that inactivates its activity but that is suppressed by existing interaction with B, (iv) additional mutation in A that is also presuppressed by interaction with B, and (v) coevolving A:B interaction arising later. At stage (ii), A is able to function whether or not B is present and interacting with it, but at stages (ii) and beyond, A is not able to function in the absence of B.

Archibald et al. (IUBMB Life, 2011)

**Road map**:

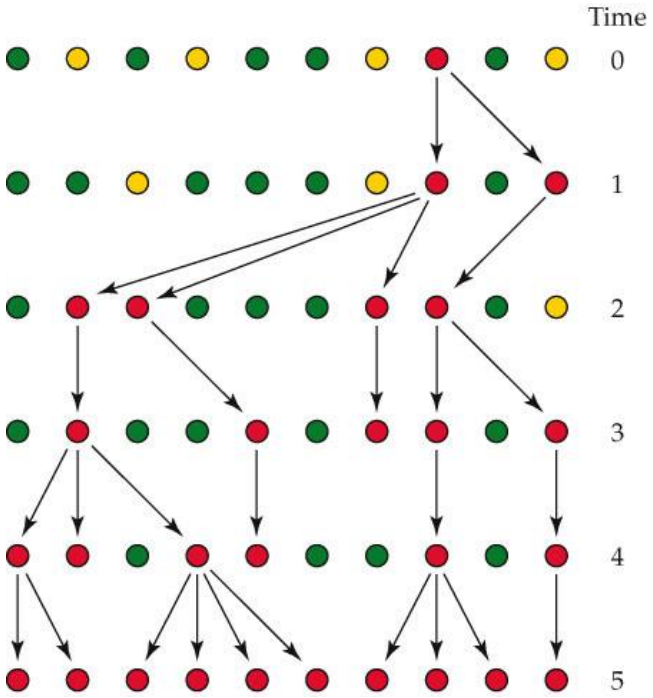
- Demystifying the process of random genetic drift.
- The definition of effective population size.
- Neutral evolution: joint operation of drift and mutation.
- Simple selection models involving single mutations.
- Selection involving epistatically interacting mutations.

Random genetic drift: not a mystical process.

# Random Genetic Drift at a Neutral Locus

- Neutral models serve as formal null hypotheses for interpreting evolutionary observations.

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




Probability of loss in first generation = $[\ 1 - \ 1/(2N)\ ]^{2N} \approx [e^{-1/(2N)}]^{2N} = e^{-1} = 0.368$.

Probability of eventual fixation = $1/(2N)$.

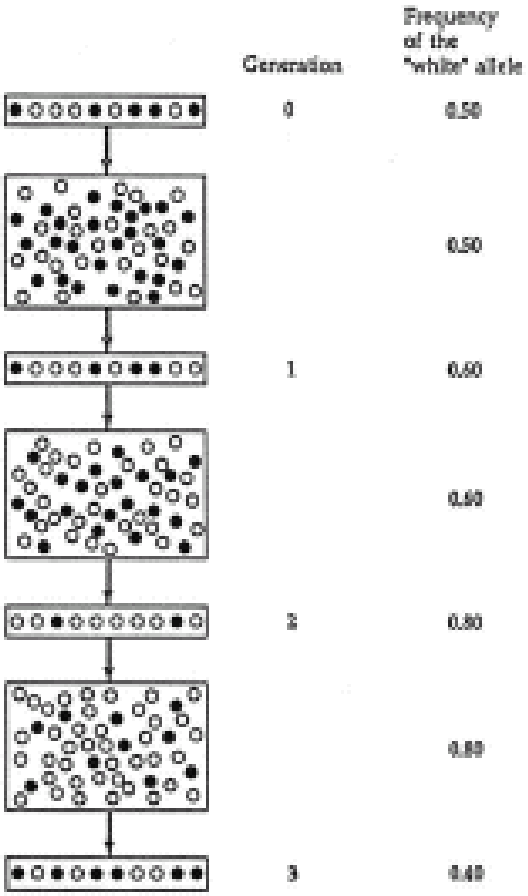
Probability of eventual loss = $1 - (1/2N)$.

$N$ = a measure of population size

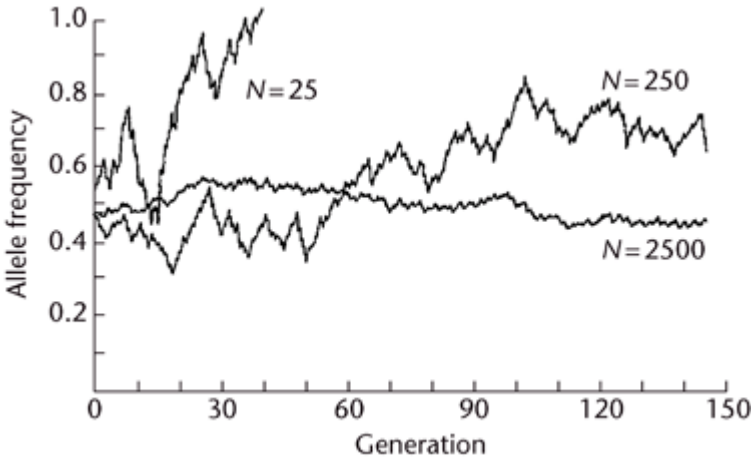
- Almost always, there are no more than two variants per nucleotide site, so everything can be studied as a one-dimensional process.

# The Features of Random Genetic Drift Are Dictated By The Effective Number of Individuals, $N_e$, in the Population

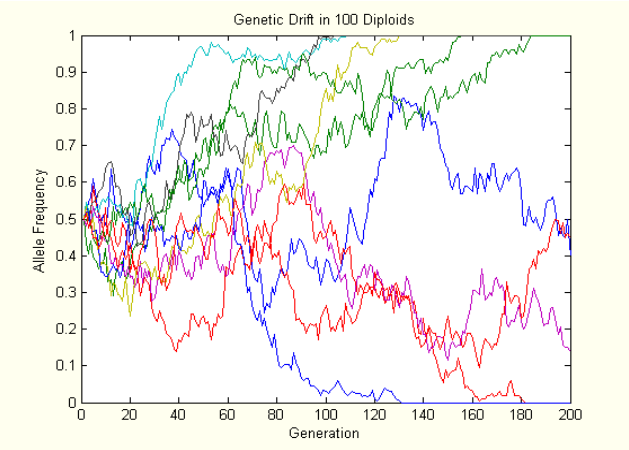
1) Sampling of finite numbers of gametes results in allele-frequency fluctuations.




2) The magnitude of fluctuations declines with increasing population size.




3) Each evolutionary trajectory is unique.



## THE WRIGHT – FISHER MODEL

Sewall Wright (1889 – 1988)

Assumes: 1) a diploid monoecious (hermaphroditic) population;
2) constant population size of *N* individuals;
3) discrete generations;
4) an effectively infinite pool of gametes.

Usually assumes two alleles, but this is not essential.

Let the initial numbers of alleles **B** and **b** be *i* and (2*N*-*i*).

The probability that there are *j* copies of **B** and (2*N*-*j*) copies of **b** in the next generation is given by the binomial sampling distribution.

$$P_{ij} = \binom{2N}{j} (i/2N)^j [1 - (i/2N)]^{2N-j}$$

The Transition-matrix Approach yields the entire probability distribution of allele frequencies at any future point in time, given the starting condition,

$$\mathbf{x}(t+1) = \mathbf{P}\mathbf{x}(t) \longrightarrow \mathbf{x}(t) = \mathbf{P}^t\mathbf{x}(0)$$

**P** is a (2N+1) x (2N+1) transition matrix, the elements of which are the $P_{ij}$ .

**x**(*t*) is the vector of probabilities that the population is in state *i* = 0, 1, 2, …, 2*N*.

For a newly arisen mutation, **x**(0) has only a single non-zero element, $x_1 = 1$ as the starting frequency.

KIMURA'S (1955) DIFFUSION-THEORY APPROACH: treats allele frequency as a continuous variable.

Probability density at time *t*, given an initial frequency $p_0$ and population size *N*,

$$\varphi(p_t|p_0) = p_0(1-p_0)\sum_{i=1}^{\infty} i(2i+1)(i+1)\cdot$$
$$F(1-i, i+2, 2, p_0)\cdot F(1-i, i+2, 2, p_t)\cdot e^{-i(i+1)t/(4N)}$$

where *F*(……) is the confluent hypergeometric function.

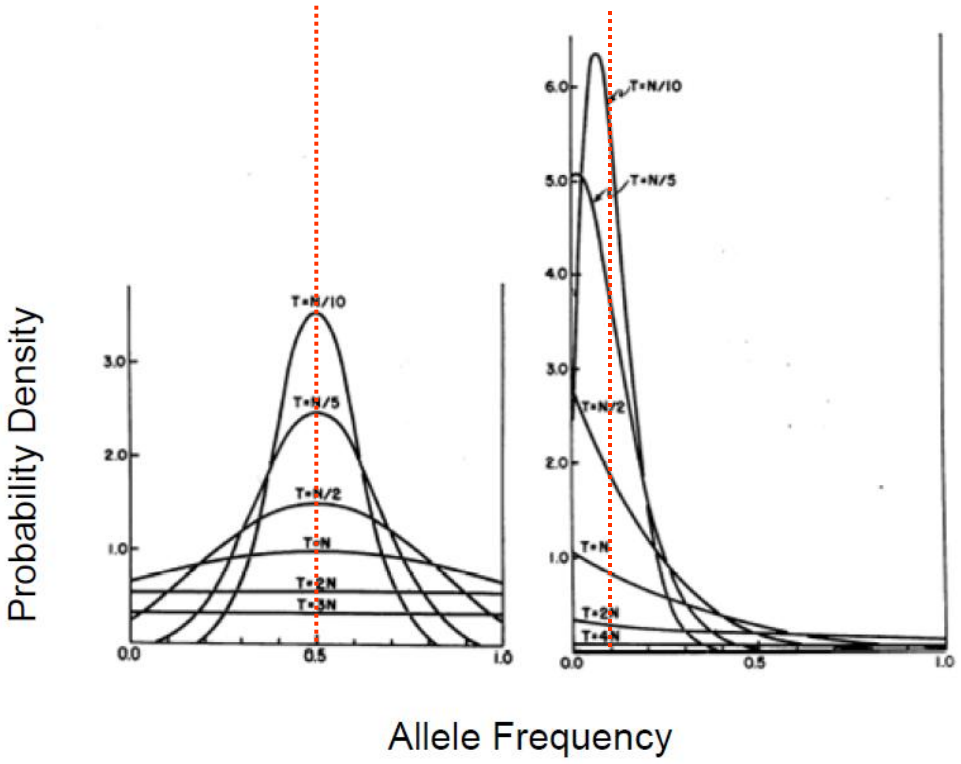
Probability of fixation by time *t*,

$$p_f(p_0, t) = p_0 + p_0(1-p_0)\sum_{i=1}^{\infty}(2i+1)(-1)^i$$
$$\cdot F(1-i, i+2, 2, p_0)\cdot e^{-i(i+1)t/4N}$$

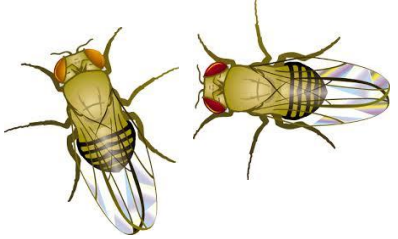
Motoo Kimura (1924 – 1994)

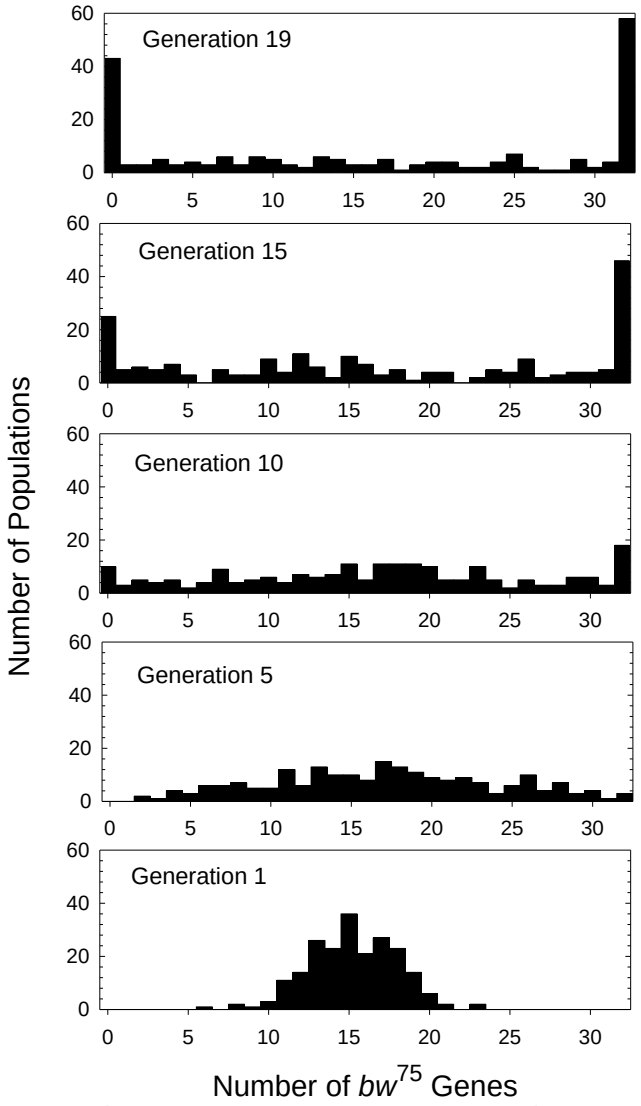
# Under Neutrality, the Probability Distribution of Allele Frequencies Scales with *t/N* Generations

Starting allele frequency: 0.5 0.1



# Buri's Big Drift Experiment





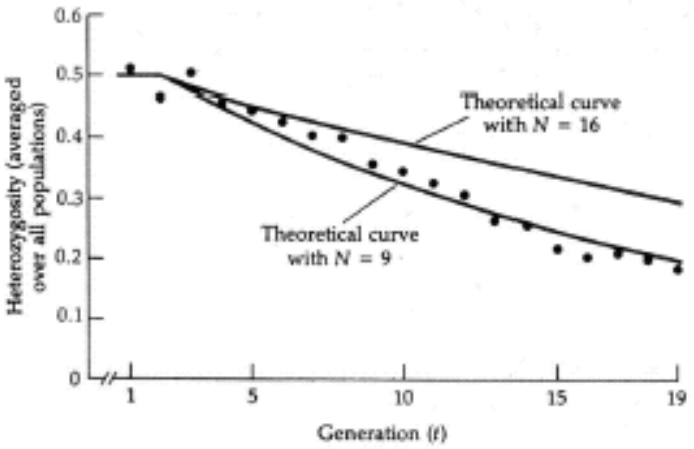

**Experimental demonstration of genetic drift.** The number of copies of an allele, $bw^{75}$, in each of many replicate populations of *Drosophila melanogaster* maintained in the laboratory at 16 flies for 19 generations. In each population, the frequency of the allele fluctuated, so the variation in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

Heterozygosity = 2p(1-p)

After *t* generations at population size *N*:

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

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

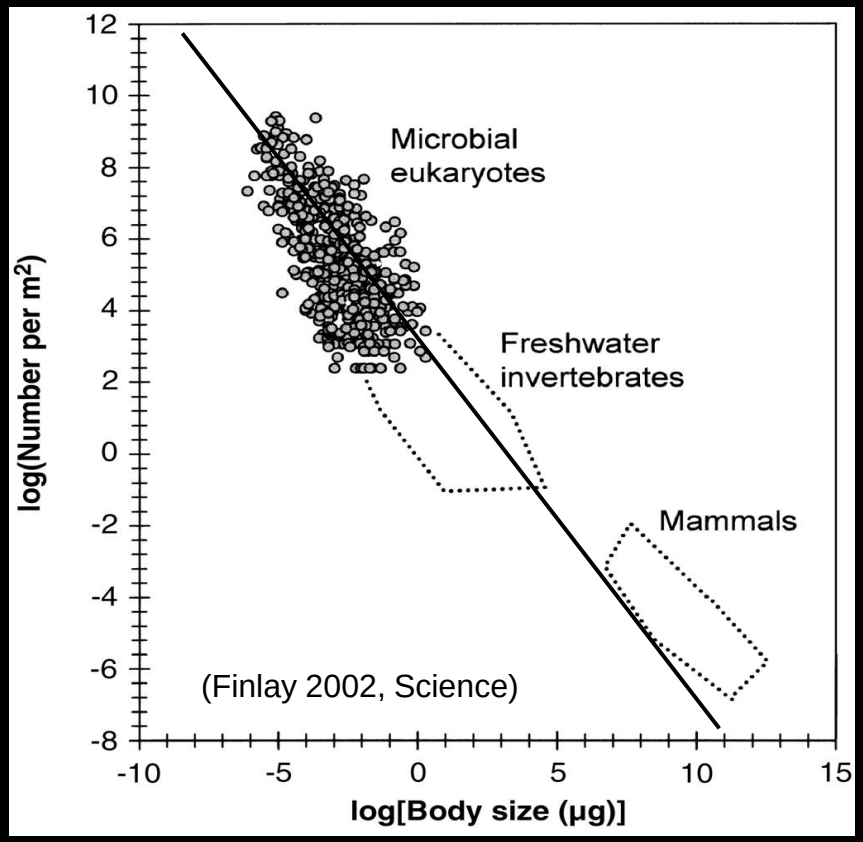


## Most violations of the demographic assumptions of the Wright-Fisher model cause $N_e$ to be much smaller than the actual number of breeding adults:

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

Frankham (1995, Genet. Res.) – $N_e / N$ averages ~0.1 in natural populations of vertebrates, not including the effects of temporal variation in the environment.

# Universal Scaling of Population Size and Organism Size

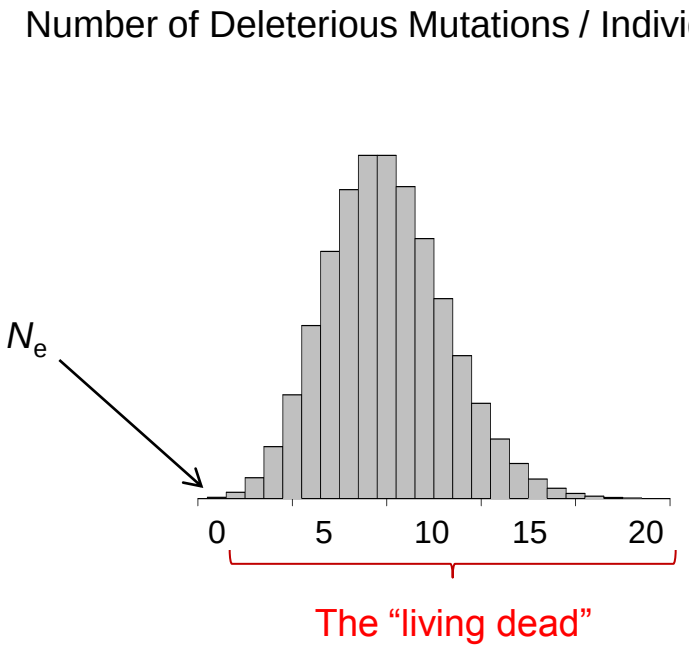



- An inverse scaling (slope ≈ -1) between numbers and sizes of individuals implies that, *on average*, each species harbors the same total amount of biomass.

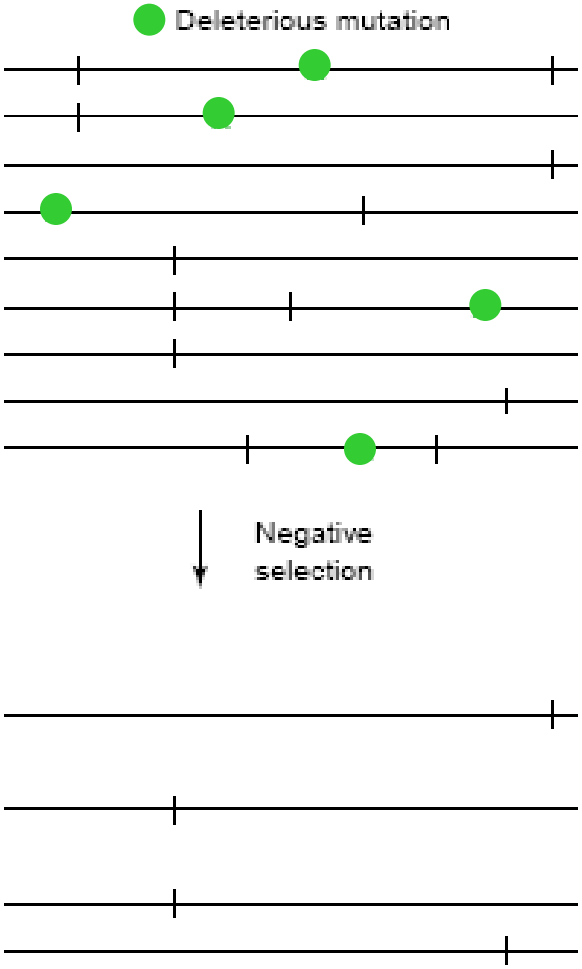
- There is an approximately two orders of magnitude range of variation above and below the regression.

# The Effects of Chromosomal Linkage on the Effective Population Size

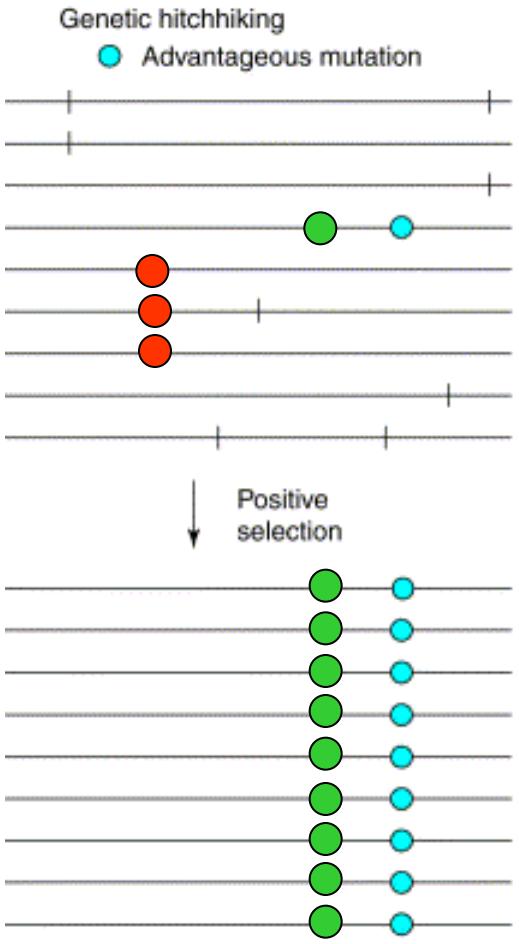
- In an obligately asexual population, the effective population size ($N_e$) is roughly equal to the number of individuals in the class with the fewest deleterious mutations.



# Linked Deleterious Mutations Reduce the Efficiency of Selection by Reducing the Effective Population Size: the Background-selection Model



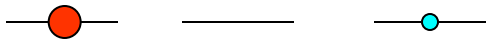
Genetic Hitch-hiking Via Selective Sweeps Causes the Genetic Effective Size of a Species to be Much Smaller Than the Actual Census Size




a background beneficial mutation lost from the population

a background deleterious mutation fixed in the population

With free recombination, the outcome would be:



# Reduced Levels of Variation in Regions of Low Recombination in Humans

Measures of Nucleotide Heterozygosity

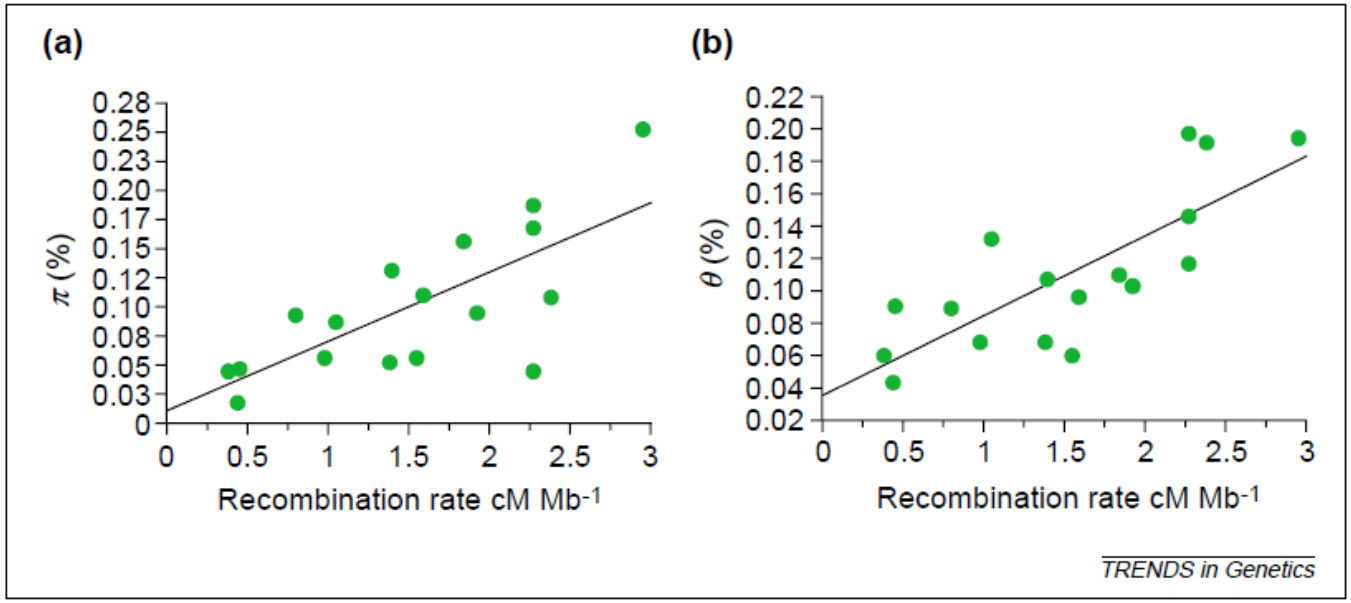



**Fig. 2.** (a) Scatterplot of nucleotide diversity versus recombination rate for the 17 loci in Table 1 for which sample size is greater than ten. Linear regression $R^2 = 0.54$, $P < 0.001$. (b) Scatterplot of proportion of segregating sites versus recombination rate for the seventeen loci in Table 1 for which sample size is greater than ten. Linear regression $R^2 = 0.63$, $P < 0.001$. Recombination rate estimates are from Ref. 11. Recombination rates for X-linked genes have been multiplied by two-thirds to account for the fact that the X chromosome spends two-thirds of its time in females where it recombines, and one-third of its time in males where no recombination occurs. Both $\pi$ and $\theta$ for X-linked genes have been multiplied by four-thirds, to account for the fact that the effective population size of the X chromosome is three-quarters that of the autosomes.

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

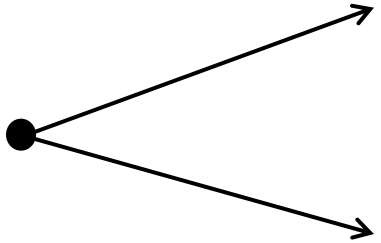
The Classical (Kimura) Neutral Model:

the joint operation of random genetic drift and mutation.

- Rate of sequence divergence between isolated species.
- Steady-state level of variation within populations.

## Divergence between species

Under neutrality, the cumulative rate of fixation of neutral mutations is equal to the per-generation mutation rate, independent of the population size.



$N$ = absolute population size
$u$ = mutation rate / gene (or nucleotide site) / generation

1. For any locus, $2Nu$ mutations enter a diploid population each generation.

2. Each new neutral mutation has a fixation probability equal to its initial frequency, $1/(2N)$.

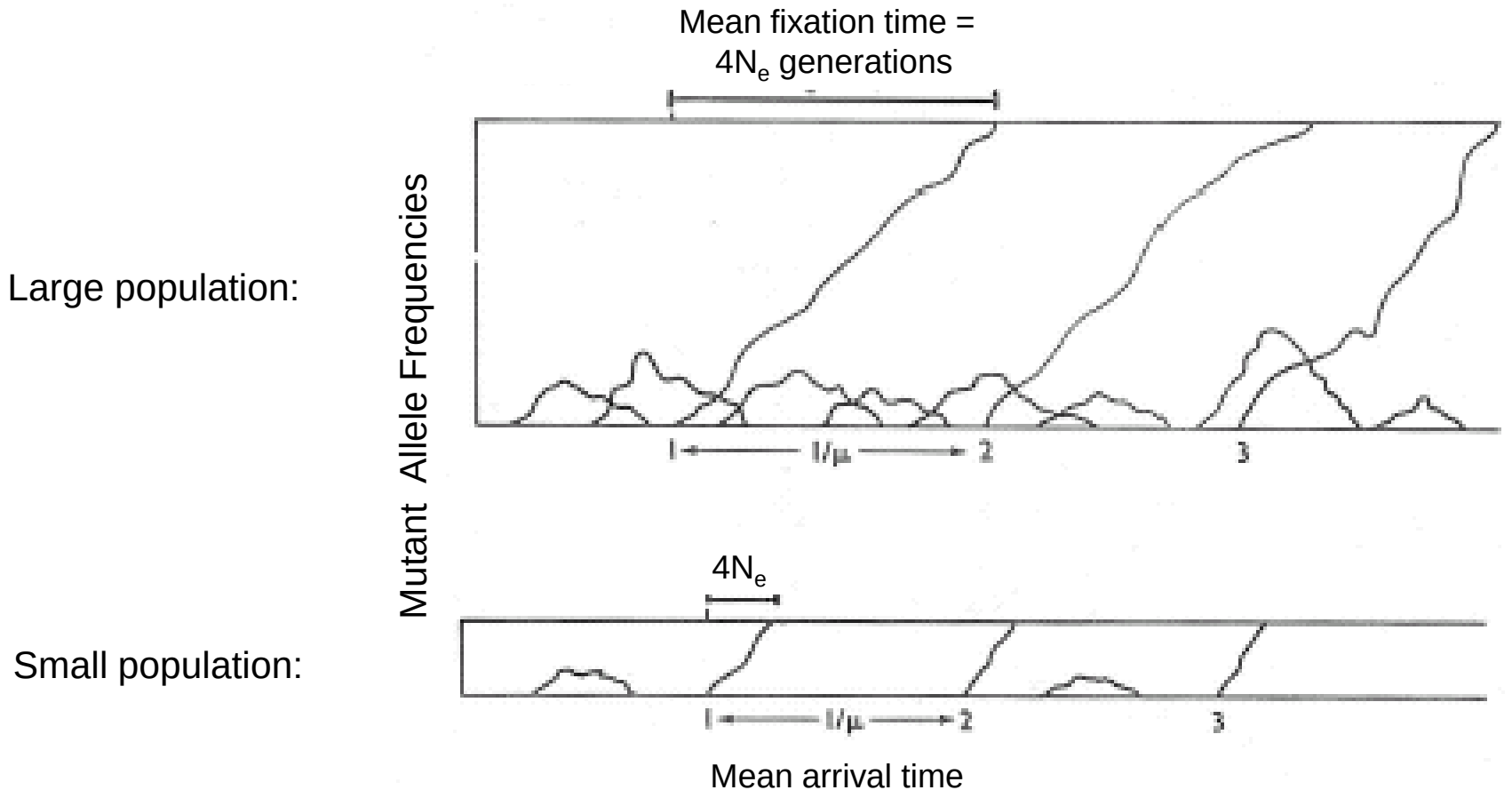
3. The average number of substitutions per locus per generation is equal to the product,

$$2Nu \times 1/(2N) = u.$$

4. Universally true regardless of the breeding system, effective population size, ploidy level, degree of linkage, mode of gene action, etc.

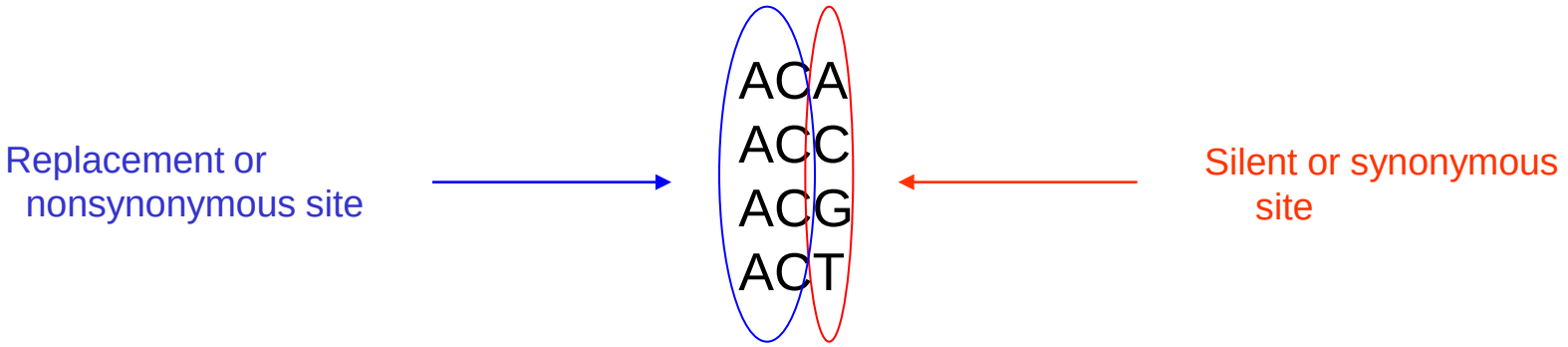
# Within-population variation at neutral sites

Because the sojourn times of neutral mutations increase with $N_e$, the steady-state heterozygosity does as well.



# Estimation of the Relative Power of Evolutionary Forces From Polymorphism Data

- Nucleotide sites in coding DNA can be subdivided into classes thought to be neutral vs. under selection.



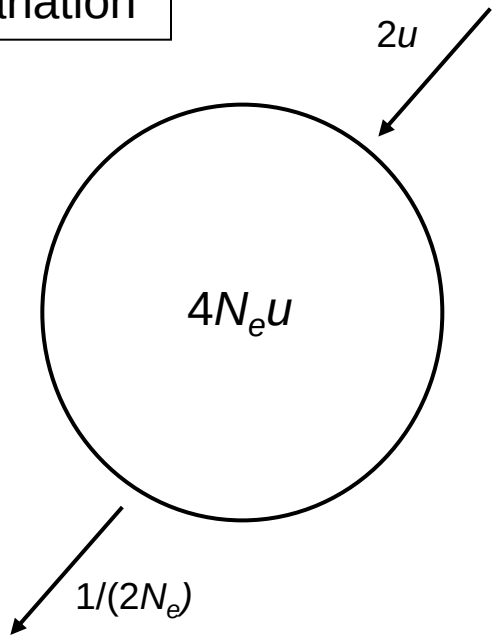

The Four Threonine Codons

# Estimation of the Effective Population Size From Silent-site Variation

$H_t$ = fraction of sites that differ between two randomly sampled alleles (nucleotide heterozygosity); e.g., 2p(1-p);

$N_e$ = the effective size of a population;

$u$ = base-substitution mutation rate per nucleotide site / generation.




- For a pair of identical nucleotides, $2u$ is the rate at which one or the other mutates to a new state.
- $1/(2N_e)$ is the rate of loss of heterozygosity by random genetic drift for a diploid autosomal locus.

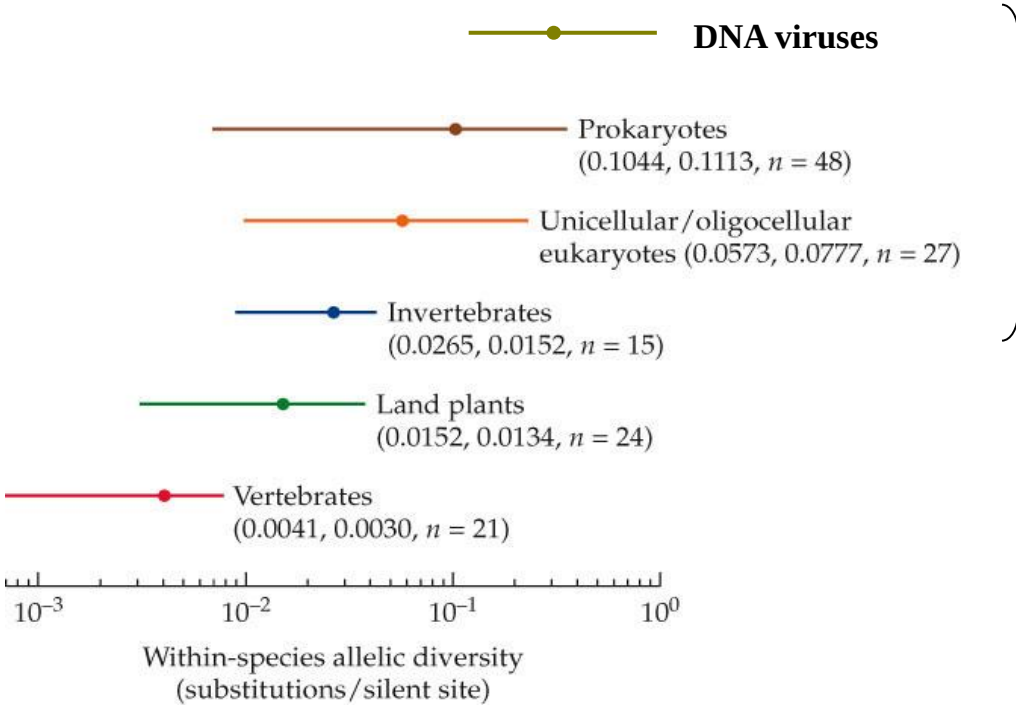
The recursion equation for average heterozygosity per site: $H_t = H_{t-1}[1 - (1/(2N_e))] + (1 - H_{t-1})(2u)$,

The equilibrium solution is obtained by setting: $H_t = H_{t-1}$,

$$H \approx 2u / [\, 2u + (1/2N_e)\,] = 4N_e u / (1 + 4N_e u) \approx 4N_e u.$$

- At mutation-drift equilibrium, the expected nucleotide heterozygosity at neutral sites provides an estimate of the ratio of the power of mutation to the power of random genetic drift.

Estimates of $4N_e u$ from population surveys of silent-site variation: the power of drift exceeds that of mutation (per nucleotide site) in perhaps all species, i.e., $4N_e u < 1.0$.

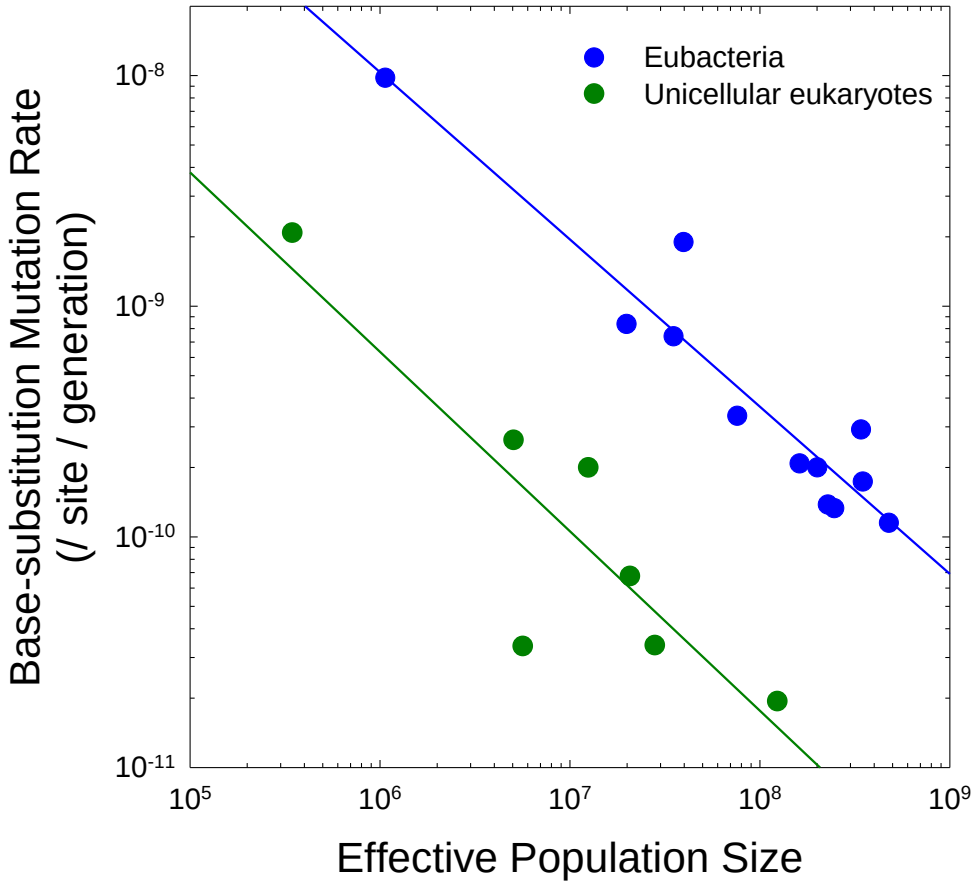



downward bias
caused by selection
for efficient translation

- The ratio of the power of mutation to drift increases in organisms of smaller size, although not nearly as rapidly as expected based on absolute population sizes.

# 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.*



# Approximate Ranges of Effective Population Sizes

Factoring out the mutation rate *(u)* from $N_e u$:

| | | | |
|---|---|---|---|
| Prokaryotes | $5 \times 10^7$ | to | $2 \times 10^9$ |
| Unicellular eukaryotes | $5 \times 10^6$ | to | $4 \times 10^8$ |
| Invertebrates | $6 \times 10^5$ | to | $2 \times 10^6$ |
| Vertebrates | 30,000 | to | 100,000 |
| Annual plants | 40,000 | to | 500,000 |
| Trees | ~10,000 | | |

**Implications**:

- Most deleterious mutations with selection coefficients $<10^{-5}$ or so are expected to accumulate in the genomes of large multicellular eukaryotes in an effectively neutral fashion.

- Likewise, beneficial mutations with advantages $<10^{-5}$ cannot be promoted by selection in mutlicellular eukaryotes.

- In contrast, prokaryotes are typically capable of promoting / eradicating mutations with selection coefficients as small as $10^{-9}$, a level that is unmeasurable in the laboratory.

# Summary of Main Observations on Effective Population Size

- Natural variation at silent sites tells us that average long-term effective population sizes differ by at least five orders of magnitude between the largest multicellular and the smallest unicellular species.
  - This is about twenty orders of magnitude less than the disparity in absolute population sizes.

- If the downward bias in estimates of prokaryotic $N_e$ induced by selection (on silent sites) is as much as tenfold, then the upper limit to $N_e$ is still just ~$10^{10}$.

- The additional reduction in $N_e$ relative to absolute population size ($N$) appears to be consistent with the significant variance-reducing effects of hitch-hiking events in large populations – once $N$ gets very large, the primary stochastic effects are associated with draft rather than drift.

# Incorporating Natural Selection

"In recent years, there has been some tendency to revert to more or less mystical conceptions revolving about such phrases as "emergent evolution" and "creative evolution." The writer must confess to a certain sympathy with such viewpoints philosophically but feels that they can have no place in an attempt at scientific analysis of the problem." Wright (1931)

- Simply invoking natural selection to explain the origin and maintenance of every aspect of biodiversity is conceptually no different than invoking an intelligent designer.

- Confidence in an adaptational hypothesis should mean that simpler hypotheses based on drift and mutation can be rejected.

Incorporating Natural Selection: Kimura's (1965) diffusion approximation for the fixation probability of a newly arisen mutation.

Let $s$ = selection coefficient  Fitnesses – AA: 1  Aa: 1 + s  aa: 1 + 2s

$N$ = actual population size  defines the initial frequency, $p_0 = 1/(2N)$ for a diploid
$N_e$ = effective population size  defines the efficiency of selection

Fixation probability for a newly arisen mutation:

$$p_f \cong (4sN_e p_0) / (1 - e^{-4Nes})$$

For a beneficial mutation:

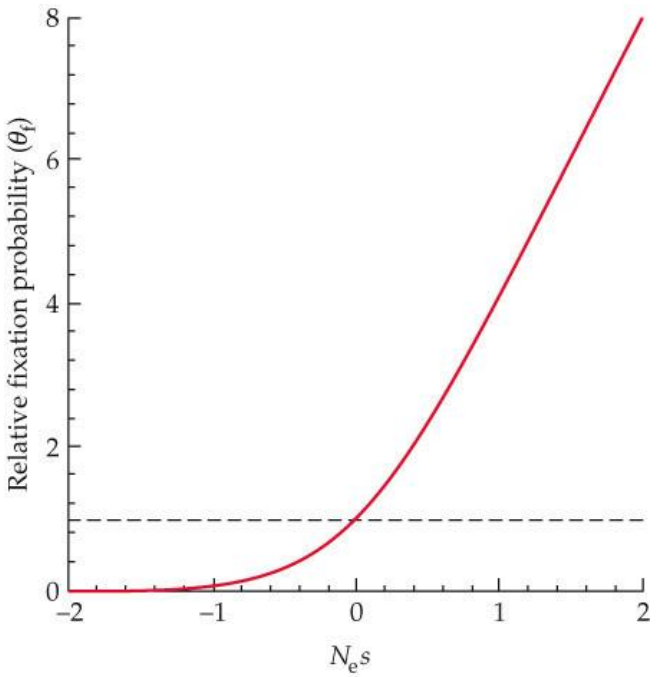
as $N_e \to \infty$, $p_f \to 2sN_e / N$

as $N_e \to 0$, $p_f \to 1/(2N)$

Fixation probability relative to neutrality:

$$\theta_f \cong 4N_e s / (1 - e^{-4Nes})$$

as $4N_e s \to 0$, $\theta_f \to 1$



# The Maximum Rate of Adaptive Evolution: single-site changes.

- As $N_e \rightarrow \infty$, the probability of fixation $p_f \rightarrow 2sN_e/N$
- Number of new mutations entering the population per generation = $2Nu$
- Long-term rate of fixation = $4N_e su$

If most molecular evolution reflected adaptive mutations, larger populations would evolve more rapidly, assuming the mutation rate does not scale more negatively than with the inverse of $N_e$.

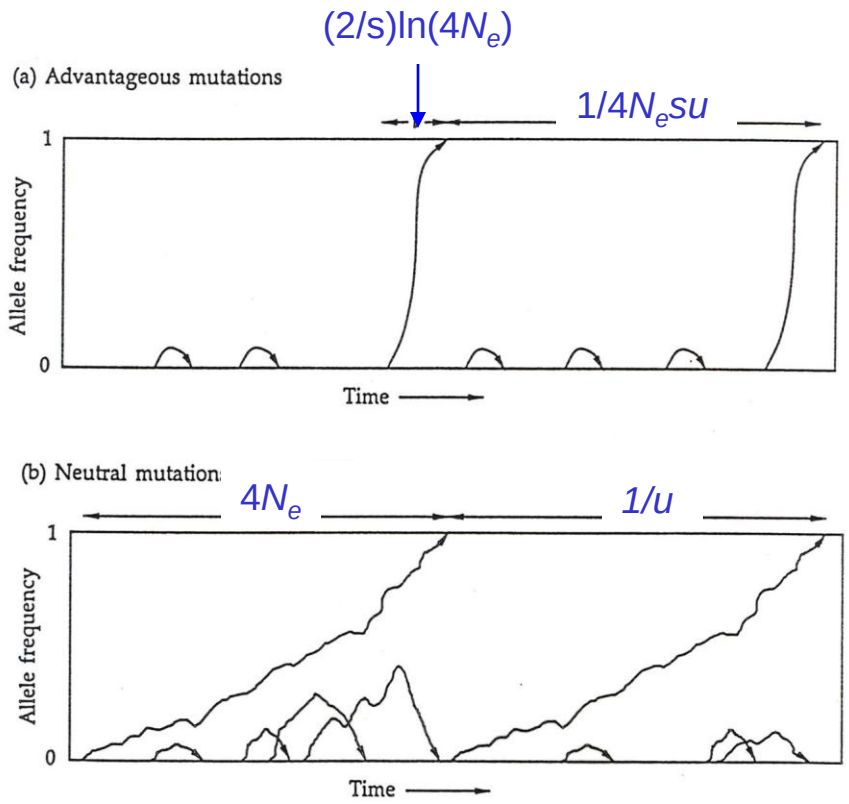
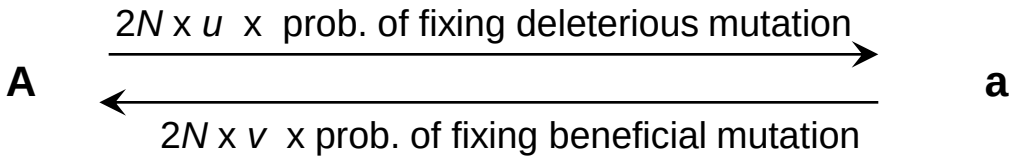



Figure 6. Dynamics of gene substitution for (a) advantageous and (b) neutral mutations. Advantageous mutations are either quickly lost from the population or quickly fixed, so that their contribution to genetic polymorphism is small. The frequency of neutral alleles, on the other hand, changes very slowly by comparison, so a large amount of transient polymorphism is generated. $\bar{t}$ is the conditional fixation time and $1/\alpha$ is the mean time between consecutive fixation events. From Nei (1987).

# Mutation, Drift, and Selection: Relative Incidence of Alternative States

**Optimal** **Suboptimal**




- *u* and *v* are the mutation rates to deleterious and advantageous alleles, so *v/u* is the mutation bias in the direction of the favorable allele.
- The ratio of fixation probability for a favorable relative to a deleterious allele (the selection bias) is $e^S$, where $S = 4N_e s$.

At equilbrium: $P_A \times u \times p_f(-s) = P_a \times v \times p_f(+s)$

*For an equilibrium (well-adapted) population, the number of beneficial mutations fixing each generation must equal the number of deleterious mutations fixing.*

Ratio of time a site is fixed for ancestrally beneficial and deleterious alleles:

$$\frac{\tilde{P}_A}{\tilde{P}_a} = \left(\frac{v}{u}\right) e^S$$

# Summary of Main Observations on the Efficiency of Natural Selection

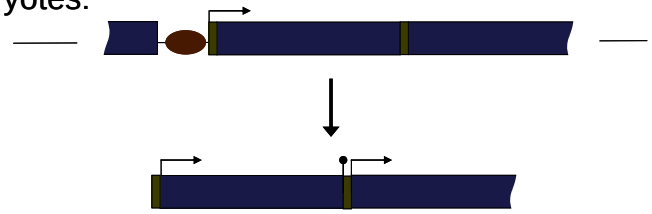
- The probability of fixation of a beneficial mutation is no greater than twice its selective advantage, $2s$.

- If $s$ is substantially smaller than the power of random genetic drift, $1/(2N_e)$, the mutation will evolve in an effectively neutral manner, with fixation probability equal to its initial frequency.

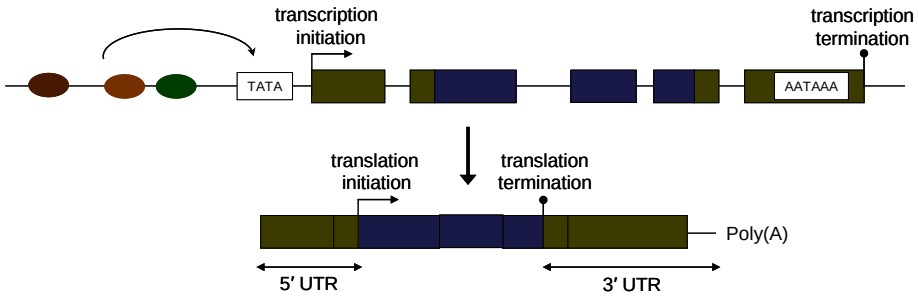
- Over a long evolutionary time period, the relative times spent in alternative states depend on the ratio of mutation and selection pressures in both directions – if mutation is sufficiently biased, it will govern the evolved states of a population regardless of the alternative selective advantages.

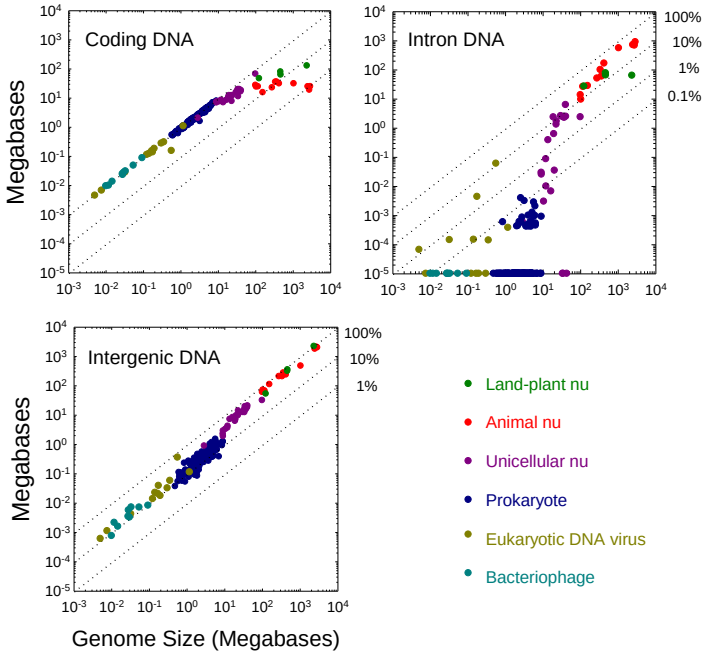
# The Origin of Gene-structure Complexity by Nonadaptive Mechanisms

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.

# The Passive Emergence/Deterence of Gene-structural Complexity by Nonadaptive Mechanisms

• **Mutational opposition** – genes with increased structural complexity involving *n* critical nucleotide sites have an elevated degenerative mutation rate (*nu*).

*n* ≈ 30 for introns; ~10 for transcription factor binding sites, ~4 for 5′ UTRs; *u* = mutation rate per nucleotide site.

• **Drift facilitation** – the effective number of gene copies per locus in a species (*N*) dictates the efficiency of natural selection – the power of random drift ~1/*N*.

• If $nu << 1/N_e$, a mutationally harmful embellishment can establish by drift.

If $nu >> 1/N_e$, the evolution of gene-architectural complexity is inhibited.

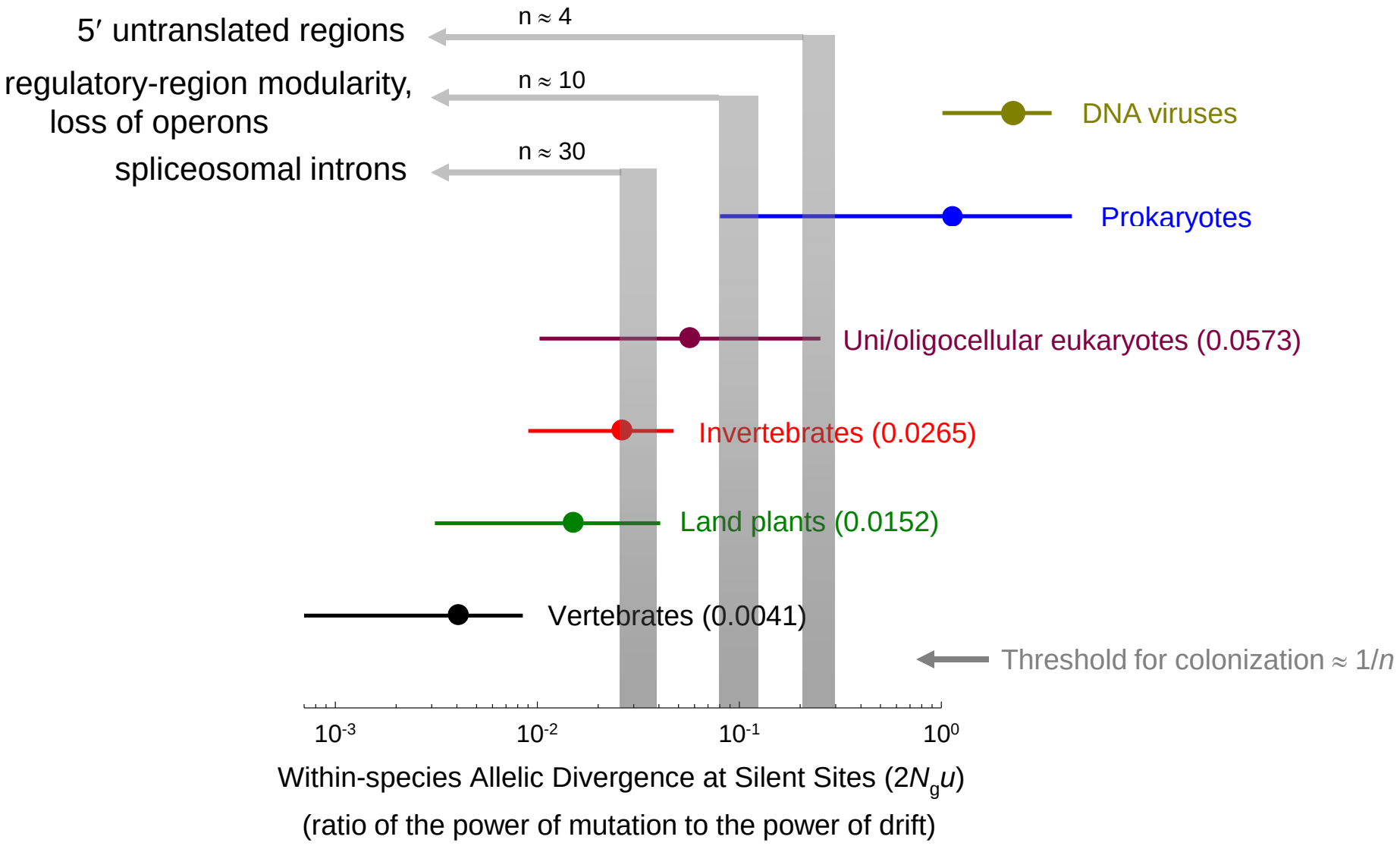
A key determinant in genome evolution is the ratio of these opposing forces:

$nu / (1/N_e) = N_e u \cdot n$

$N_e u$ ← Estimated from molecular biology

$n$ ← Estimated from population surveys

# Population-genetic Environments Permissive to the Passive Emergence of Cellular Complexities



# The Role of Genetic Drift and Mutation In Genomic Evolution

- The qualitative predictions of population-genetic theory combined with quantitative estimates of $4N_e u$ (from silent-site polymorphism data) jointly point to the effectiveness of large population size as a barrier to the evolution of complex features of gene and genomes.

- One need not invoke any direct selective advantage of the complexity of eukaryotic gene or genome structure to explain its *origin*.

- The expansion of gene architectural complexity by semi-neutral processes opens up new types of previously inaccessible evolutionary pathways.

How do complex adaptations requiring more than one mutation evolve?

The existence of apparent “compensatory pathogenic deviations” in sister taxa suggests that evolution may commonly proceed through deleterious intermediates:

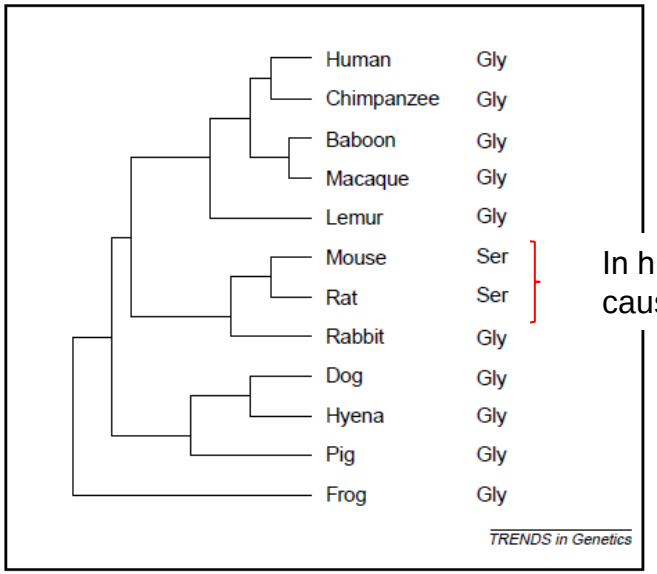



**Figure 1.** An example of fixed differences of disease-associated mutation (FDDAM). Amino acids at position 491 of the androgen receptors of 11 mammals and one non-mammalian vertebrate are shown. G491S leads to the complete androgen insensitivity syndrome in humans [2]. The tree topology follows [16] and the branches are not drawn to scale. The species names and GenBank accession numbers are: human (*Homo sapiens*) NP_000035; chimpanzee (*Pan troglodytes*), O97775; baboon (*Papio hamadryas*), O97960; macaque (*Macaca fascicularis*), O97952; lemur (*Eulemur fulvus*), O97776; mouse (*Mus musculus*), NP_038504; rat (*Rattus norvegicus*), AAA40734; rabbit (*Oryctolagus cuniculus*), P49699; dog (*Canis familiaris*), Q9TT90; hyena (*Crocuta crocuta*), AAM96904; pig (*Sus scrofa*), AAG40566; frog (*Xenopus laevis*), AAC97386. Abbreviation: Gly, glycine.

In humans, the Gly → Ser mutation at this site in the androgen receptor gene causes males to have female external genitalia.

- Sequencing of the mouse genome revealed 160 situations in which a mouse gene is fixed for a variant that is a clear pathogenic mutation in humans.
- Kulathinal et al. (2004, Science) obtained a nearly identical estimate in flies when considering mutations known to cause a dramatic fitness loss in *D. melanogaster*.

Gao and Zhang (2003, Trends Genetics)

# How do complex adaptations requiring more than one mutation become established?

- The role of population size:

  Larger populations provide more mutational targets.

  Selection is more efficient in large populations, so what happens if the first-step changes are deleterious? The usual view is that adaptation via such a pathway requires a small enough population size to enable drift across an adaptive valley.

- The role of the mutation rate ($u$):

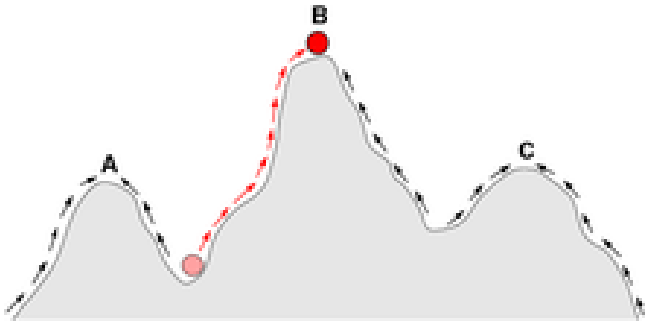
  If $n$ mutations are required, does the rate of establishment of the adaptation scale as $u^n$?

  What if mutations are spatially clustered?

- The role of recombination:

  The usual view is that recombination accelerates the rate of evolution, but just as recombination can bring two independent mutations together, it can also separate them.

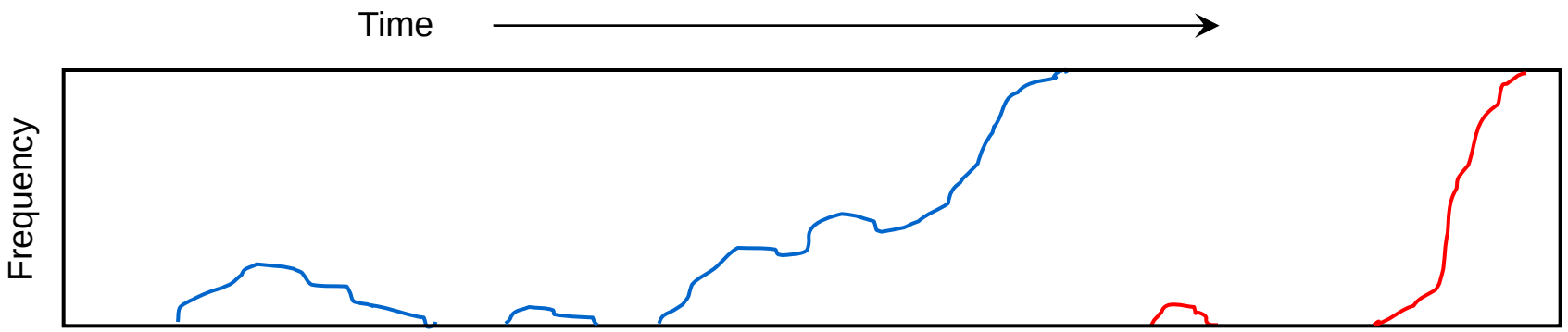
A common but incorrect view: selection cannot take a population from one adaptive peak to another, unless the population size is small enough to allow maladaptive drift across the fitness valley.



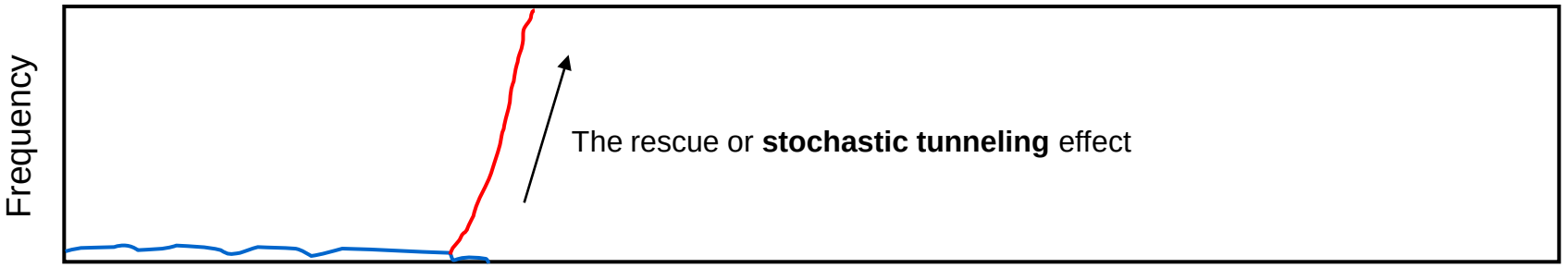

“Can a population at one peak get to another through natural selection alone? NO!”

“Is there a way to cross the valley in fitness? Perhaps through genetic drift a subset of the population can find itself in the range of the alternative attractor.”

# Two Contrasting Views of Evolutionary Dynamics




- Under the **sequential-fixation model**, adaptation proceeds in a stepwise fashion, resulting in fixation of the intermediate state, which can necessitate a sojourn through a mean-population fitness bottleneck.




- In larger populations – intermediate-state alleles need never be fixed, and if deleterious, are kept at low frequencies by selection-mutation balance, serving as launching pads for the final adaptation.

## When does the sequential-fixation model break down?

- Assuming first-step neutrality, the mean time to fix the first (neutral) mutation = $4N_e$ generations.

- During this period, the first mutation's average frequency is 0.5, so an expected $(0.5 \cdot 2N \cdot u) = Nu$ second-step mutations will arise per generation.

- Thus, conditional on a first-step mutation destined to fixation, the expected number of double mutants during the intervening period is $(4N_eNu)$,

  Assuming $N_e \approx N$, this number is less than one only if the population size $N < (4u)^{-1/2}$.

**Example**. With $u = 10^{-9}$, $N < 16{,}000$ is required for the sequential model to hold.

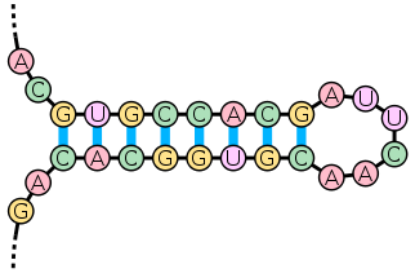
If the mutation rate is higher, the critical population size is even smaller.

Evolution of a Complex Adaptation Through Deleterious Intermediates: two alternative adaptive states having equivalent fitness.

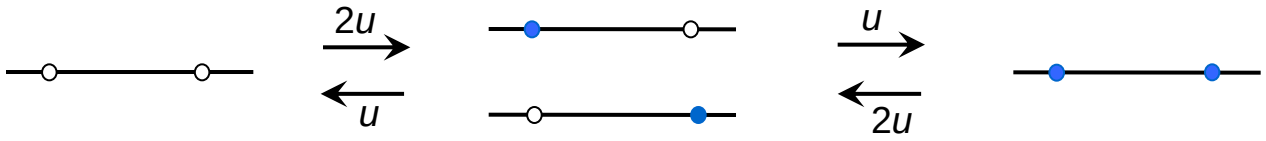
| Gamete type: | **ab** | **Ab** | **aB** | **AB** |
|---|---|---|---|---|
| Fitness: | 1 | 1 - s | 1 - s | 1 |

Compensatory RNA stem-pair mutations:

A=U ⟷ G··U / A··C ⟷ G=C



Evolution by Compensatory Mutations: sequential pathway (population size small enough that the intermediate state can become fixed).




Fitness: 1 1-*s* 1

Rate of fixation of the first (deleterious) mutation = $r_1 = 2Nu \times 2 \times p_f(-s)$

Rate of fixation of the second (advantageous) mutation = $r_2 = 2Nu \times 1 \times p_f(+s)$

Time to establishment of the compensatory change = sum of the reciprocal of the rates of the steps

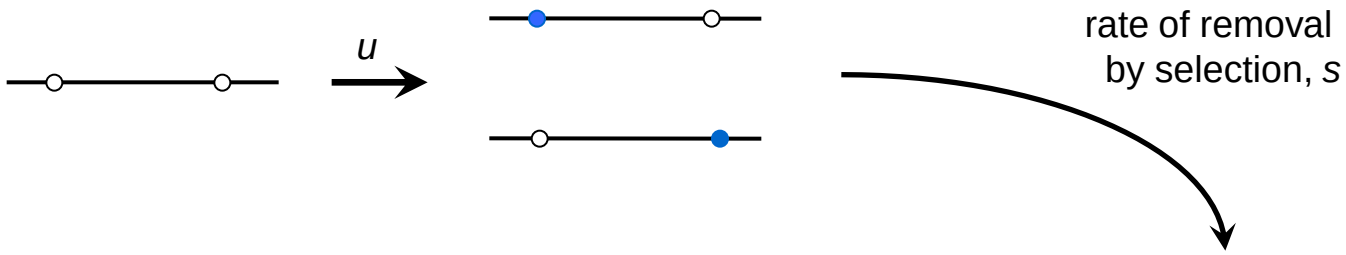
$= (1 / r_1) + (1 / r_2)$

$\approx 1 / [4Nup_f(-s)]$ because the rate of the second step is much greater than the first.

Rate of establishment of compensatory change = $4Nup_f(-s)$, which approaches zero as the population size increases (because of the reduced probability of fixation of a deleterious mutation).

Evolution by Compensatory Mutations: *stochastic tunneling pathway* (population size large enough that intermediate state never becomes fixed on its own).

Maintenance of intermediate-step deleterious alleles by selection-mutation balance:




- Equilibrium frequency for each of the intermediate types = $u / s$
- Equilibrium number of copies of intermediate alleles = $2 \times 2N \times u / s$
- Rate of mutation to second-step alleles = $u$
- Probability of fixation of neutral second-step allele = $1 / (2N)$

Rate of establishment of compensatory change = $(4Nu/s) \times u / (2N) = 2u^2 / s$, independent of population size.

# Evolution by Compensatory Mutations by Stochastic Tunneling: *second-step mutation is beneficial*.

Identical to previous analysis, except:

Probability of fixation of second-step allele ≈ $2(N_e/N)s_b$

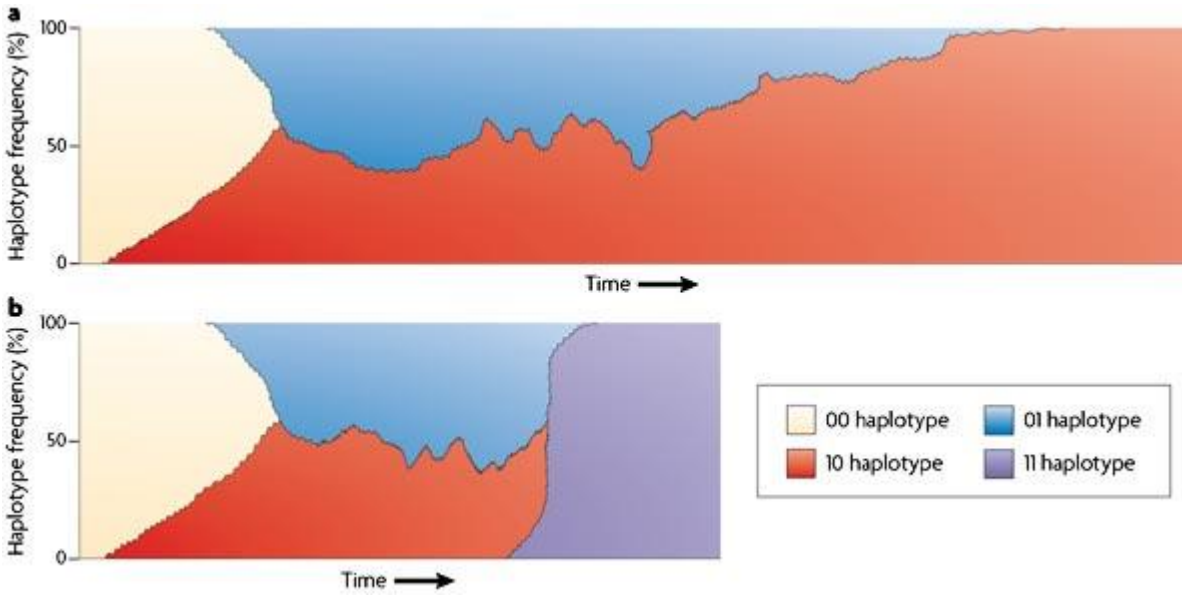
$$\overline{r}_e \simeq (2u/s_d)(2Nu)[2s_b(N_e/N)] = \frac{8N_e u^2 s_b}{s_d}$$

The rate of establishment (by a peak shift) now increases linearly with the effective population size,

*contrary to the classical view that such transitions are only possible in small populations.*

# Effects of Recombination on Adaptive Evolution

# The Classical View That Recombination Enhances the Rate of Adaptive Evolution

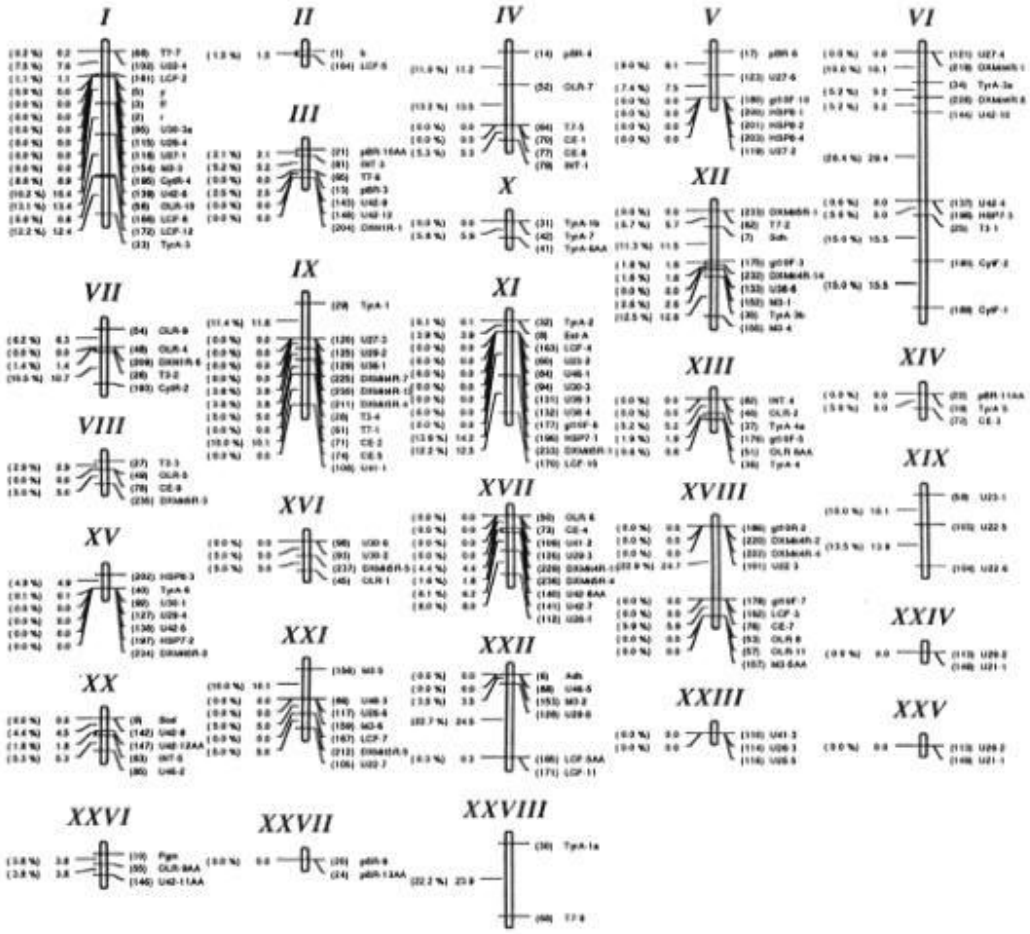



**No recombination**:
the two single-site mutations "compete" for fixation, and the final two-site adaptation must evolve by sequential substitutions.

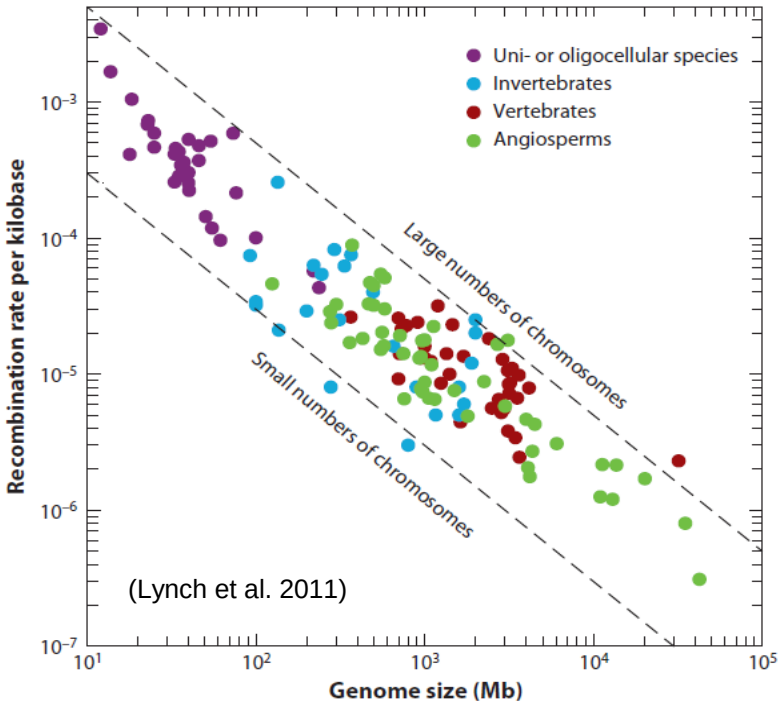
**With recombination**:
the emergence and fixation of the final adaptation is accelerated.

# Recombination Rates Are Measured with Genetic Linkage Maps

one centiMorgan = 1% probability of recombination per meiosis



# Large Genomes (in Species with Relatively Small $N_e$) Have Low Rates of Recombination per Physical Distance



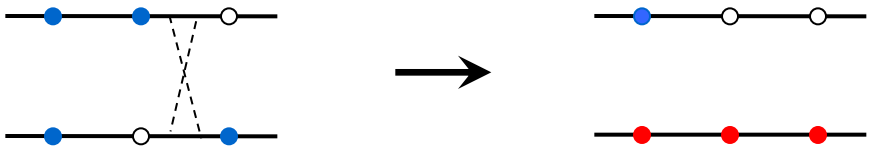

(Lynch et al. 2011)

- Virtually all of the interspecific variation in the recombination rate per physical distance is explained by variation in genome size and chromosome number.
- This is a simple and apparently nearly absolute constraint of the physical aspects of meiosis.
- Low recombination rates inhibit the promotion of beneficial mutations and removal of deleterious mutations when effects are additive – *selective sweeps and background selection*.
- Low recombination rates might enhance the promotion of beneficial mutations when effects are epistatic, e.g., when two mutations are required to produce the adaptation.

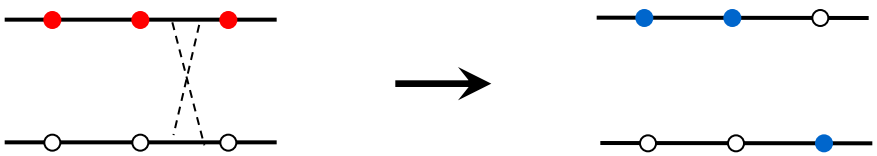
If selection operates to increase the recombination rate for the benefit of long-term evolutionary progress, why do essentially all organisms have ~1 crossover per chromosome arm per meiotic division?

# How Does Recombination Influence the Evolution of Complex Traits?

Recombination can facilitate the arrival of an adaptive combination,



but it also inhibits the fixation of the adaptive allele,



# Influence of the Recombination Rate When Intermediate-state Alleles Are Neutral

- There is a threshold recombination rate (*r*) beyond which the rate of adaptation is inhibited.
  - This is approximately equal to the selective advantage (*s*) of the adaptive allele,.

    At this point, the rate of advancement of the adaptation (*s*) is exceeded by the rate of breakdown (*r*), and the net advantage is $s - r$.

- The "optimal" recombination rate is equal to half the selective advantage:

  1) the rate of production of **AB** alleles by **Ab/aB** heterozygotes is proportional to *r*;

  2) the net selective advantage of **AB** is reduced to ($s - r$);

  3) the product $r(s-r)$ is maximized when $r = s/2$.

- The effects of recombination are diminished in populations of small size where evolution occurs sequentially, with each mutation becoming fixed before the next arrives.

- There is no generally optimal recombination rate, as this is context dependent.

- If the intermediate state is deleterious, recombination can impose a complete barrier to establishment of a new complex-adaptation.


NATURAL SELECTION
THE MECHANISM OF EVOLUTION