What is “Genetic Draft”?  

It’s not a fundamental “force” like mutation, selection, and drift.  

It’s an effect of mutation at a selected locus, that reduces variation at nearby (linked) loci, thereby reducing the apparent $N$.  

Why should we care?  

Neutral theory predicts that at mutation-drift equilibrium, DNA polymorphism should be proportional to $N$:  

$$\pi \approx \theta \approx 4N\mu$$ (at a diploid locus)  

But it’s not! (Not even approximately, especially for mitochondria.)  

If we can figure out why the neutral theory fails, we’ll learn something important about genetics and ecology (i.e., mutation and selection).
Fig. 1. Average allozymic, nuclear DNA, and mtDNA diversity in eight animal taxa. $x$ axis: allozyme average heterozygosity. $y$ axis: circles, nuclear DNA average synonymous diversity (kendall test: $\tau = 0.87, P < 0.05$); squares, mtDNA average synonymous diversity (kendall test: $\tau = -0.14$, not significant). Ma: Mammalia (allozymes: 184 species; nuclear: 30 species; mtDNA: 350 species); S: Sauropsida (reptiles and birds: 116, 20, 378); A: Amphibia (61, 4, 96); P: Pisces (bony fish and cartilaginous fish: 183, 22, 270); I: Insecta (156, 73, 511); C: Crustacea (122, 2, 78); E: Echinodermata (sea stars and urchins: 15, 14, 47); and Mo: Mollusca (46, 9, 125). The nuclear averages of the little-represented Amphibia (four species) and Crustacea (two species) are shown but were not used for the statistical test.
The most abundant animal on Earth?

*Euphausia superba*

Key link in Southern Ocean ecosystem

$10^{14} - 10^{15}$ post-larval individuals

$1\text{-}4 \times 10^8$ metric tons

Human biomass, for comparison?

$7 \times 10^9 \times 0.04 \text{ m.t.} = 2.8 \times 10^8 \text{ m.t.}$
Very conservatively, mitochondrial $\theta = 2N_f \mu$ 
$= 2 \times 10^{12} \times 5 \times 10^{-9} = 10^4$.

But $\pi_s \approx 0.03$!

Why is $\theta$ 333,333 times larger than $\pi_s$?
The draft model: Locus “A” is selected, locus “B” is neutral

Step 0: The population is fixed for $A_2$, polymorphic for $B_1$ and $B_2$.
Frequency of $B_1 = p_B$, frequency of $B_2 = q_B = (1-p_B)$.

Step 1: A mutation to the selectively favored allele $A_1$ occurs.
But on which genetic background?
$B_1$?
$B_2$?

Figure 4.1: The chromosome on the left shows the position of the $A$ and $B$ loci. The right side illustrates the four possible gametes with their frequencies.
Step 2: The lucky B-allele “hitches a ride” with $A_1$ …

... as long as it remains in linkage disequilibrium with $A_1$ …

... and as a consequence, B-locus variation tends to be reduced.

$2pq = 0.42$

$2pq \approx 0.4$

$2pq \approx 0.2$

$2pq \approx 0$

Figure 4.5: The frequency of the $B_2$ allele under different hitchhiking scenarios. For the upper two curves, the $A_1$ allele in initially linked to the $B_1$ allele; in the bottom two, it is linked to the $B_2$ allele. $s = 0.2$ for all trajectories.

(Gillespie gets confused by his notation here.)
An important consequence: hitchhiking "sweeps away" variation near the selected locus

Typical $r \approx 10^{-8}$/bp.
Suppose $s \approx 10^{-3}$.
Then $r/s \approx 10^{-5}$/bp.
So $|r/s| \leq 0.1$ within 10,000 bp of the selected locus.
($10^{-5} \times 10^4 = 0.1$)

**Figure 4.4:** The ratio of the final to initial heterozygosity at a neutral locus as a function of the distance from the selected locus as measured by $r/s$. Negative values of $r/s$ are left of the selected locus, positive values are to the right.
LCT region and putative lactase persistence mutations

Ingram et al. (2009) Human Genetics 124:579-591
Lactase in Utah, again!

The first 26 chromosomes share the consensus sequence (as do 60 others omitted to save space).

The other 34 chromosomes are shown as differences from the consensus.

These 101 variable sites are embedded in a region of roughly 140,000 base pairs.

In what sense are the consensus chromosomes "younger" than the others?

Are their mutations younger? Or is it just the combination of mutations?

Enlightening exercise:
Identify and describe a few recombination events.
Homozygous intervals of chromosome 2 in 90 Utahans

LCT-13190*T

2.4 Mbp
19 genes
Overall heterozygosity is “drafted down” at HapMap SNP loci.

But why so modestly?
4.3 The following program, written in Python, will print out the ratio of the final to starting heterozygosities at the $B$ locus.

```python
s, r, N = 0.1, 0.001, 5000
eps = 1.0 / (2 * N)
x1, x2 = eps, 0.0
x3, x4 = 0.5 - x1, 0.5
while x1 + x2 < 1.0 - eps:
    p1 = x1 + x2
    q1 = 1.0 - p1
    wBar1 = 1.0 - q1 * s / 2.0
    wBar3 = 1.0 - p1 * s / 2.0 - q1 * s
    wBar = 1.0 - q1 * s
    rWD = r * (1 - s / 2.0) * (x1 * x4 - x2 * x3)
    x1 = (x1 * wBar1 - rWD) / wBar
    x2 = (x2 * wBar1 + rWD) / wBar
    x3 = (x3 * wBar3 + rWD) / wBar
    x4 = (x4 * wBar3 - rWD) / wBar
    p2 = x1 + x3
    q2 = 1.0 - p2
    print 2.0 * p2 * q2 / 0.5
```

Exercise: Do this for $s = 0.001, 0.01, 0.1$
$r = 0.0001, 0.003, 0.001$
And consider initial B locus allele frequencies other than 0.5!
So, what are the predictions of this model?

Less variation where sweeps are more frequent.

Less variation where recombination rates are lower (Loci are strongly affected where $|r/s| < 0.1$).

**Figure 4.3:** The observed silent heterozygosity on the X chromosome of *Drosophila melanogaster* as a function of the local rate of recombination. The data are from Begun and Aquadro (1992).
What determines the frequency of adaptive sweeps ($p$)?

1. The rate of environmental change.
2. The rate at which adaptive mutations occur in the population.

But if most species' environments change at about the same rate, they will have similar $N_e$ and $\pi$, regardless of $N$.

From Kimura's model. Implies rate of evolution is mutation-limited.

\[ N_e = \frac{N}{1 + 2N \rho E\{y^2\}} \]

But if most species' environments change at about the same rate, they will have similar $N_e$ and $\pi$, regardless of $N$.

$\rho = 4Nv$s

**Figure 4.6:** The relationship between the population size and the effective population size under genetic draft.
But is the rate of environmental change similar for all genetic loci, in addition to being similar for all species?

\( \pi \) varies overall by roughly a factor of 100, which is really not enough. Most values fall between 0.002 and 0.02, and are much smaller than \( 4N\mu \! \).
Summary

Genetic “draft” is the effect of a selective sweep at one locus on the variation at nearby (linked) loci.

On average it reduces variation at those linked loci.

In this way it reduces the apparent or effective $N$.

Gillespie calls it a “new stochastic force in evolution”, arising from the stochasticity of the initial conditions.

That is, when and where (on which chromosome) a favored mutation occurs.

Whether we consider it a new force, or just an effect, it’s important!

But apparently we don’t yet understand the whole story ...