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.**
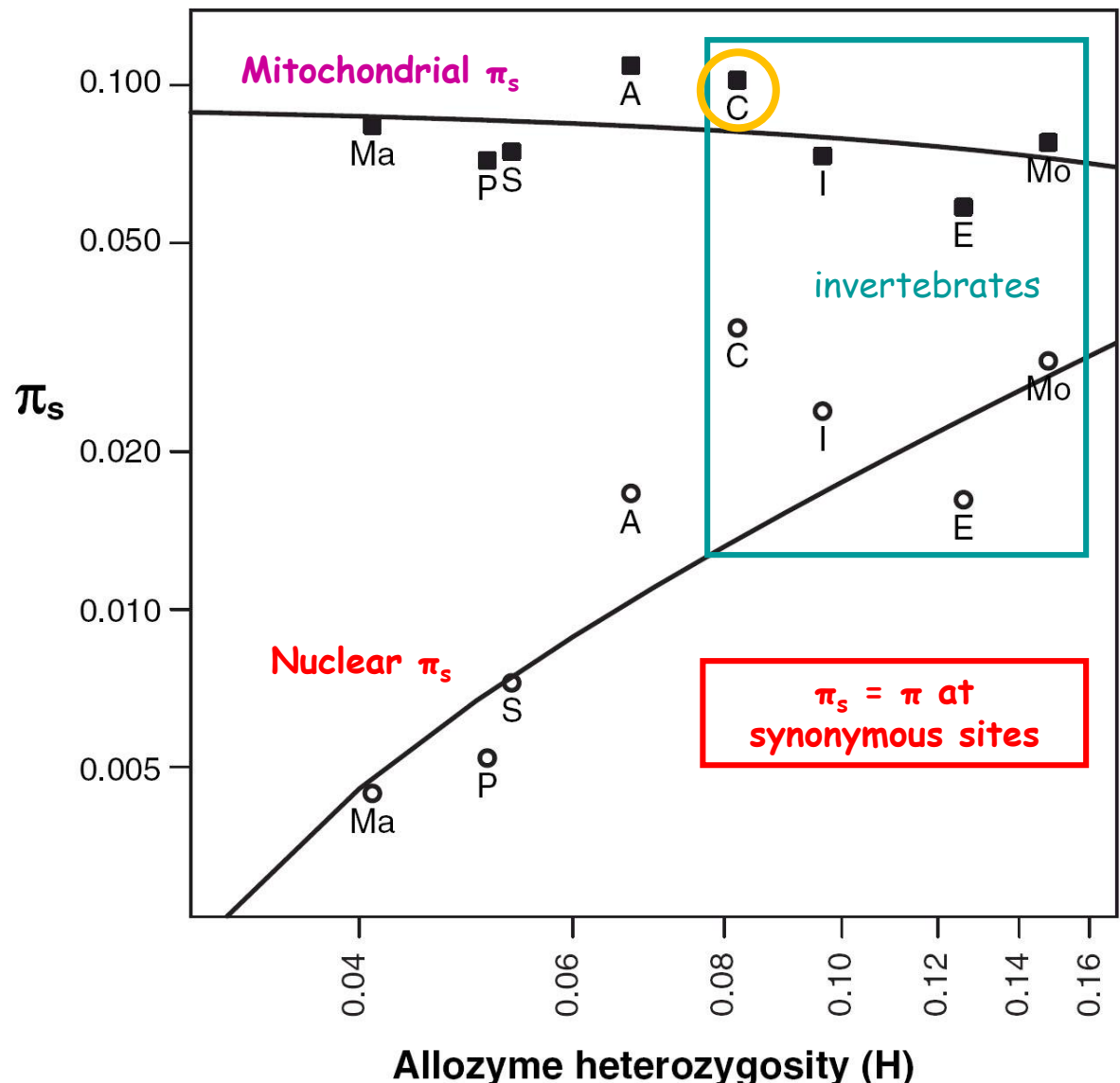
Why should we care?

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

$$\pi \approx \theta \approx 4N\mu \qquad \text{(at a diploid locus)}$$

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

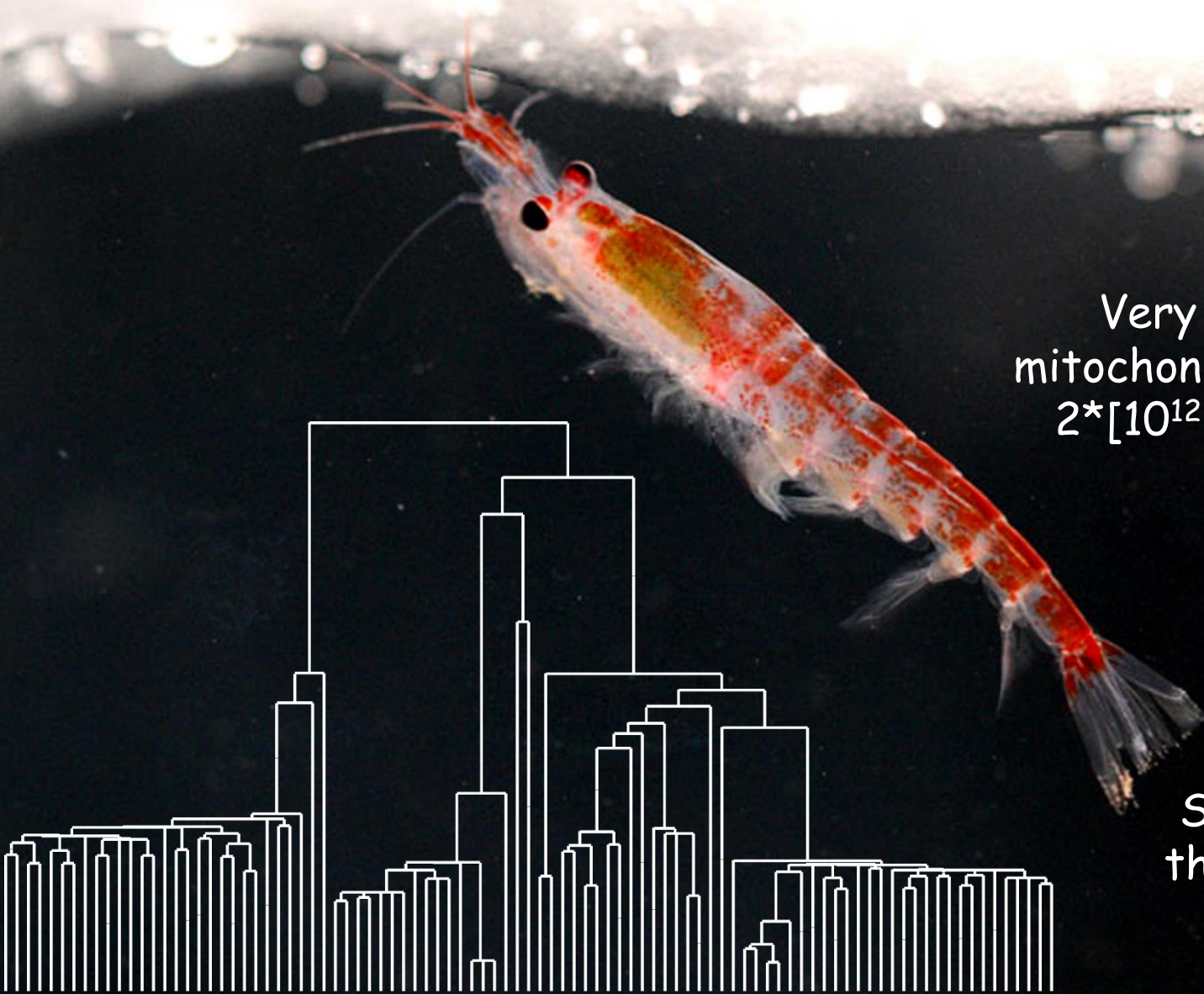Key link in Southern Ocean ecosystem

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

$1-4 \times 10^8$ metric tons

Human biomass, for comparison?

$7 \times 10^9 \times 0.04$ m.t. $= 2.8 \times 10^8$ m.t.

Very conservatively, mitochondrial $\theta = 2N_f\mu = 2*[10^{12}]*[5*10^{-9}] = 10^4$

But $\pi_s \approx 0.03$!

Recall the crustacean mean = 0.1

So why is π_s less than $\theta/333,333$?

The draft model: Locus "A" is selected, locus "B" is neutral

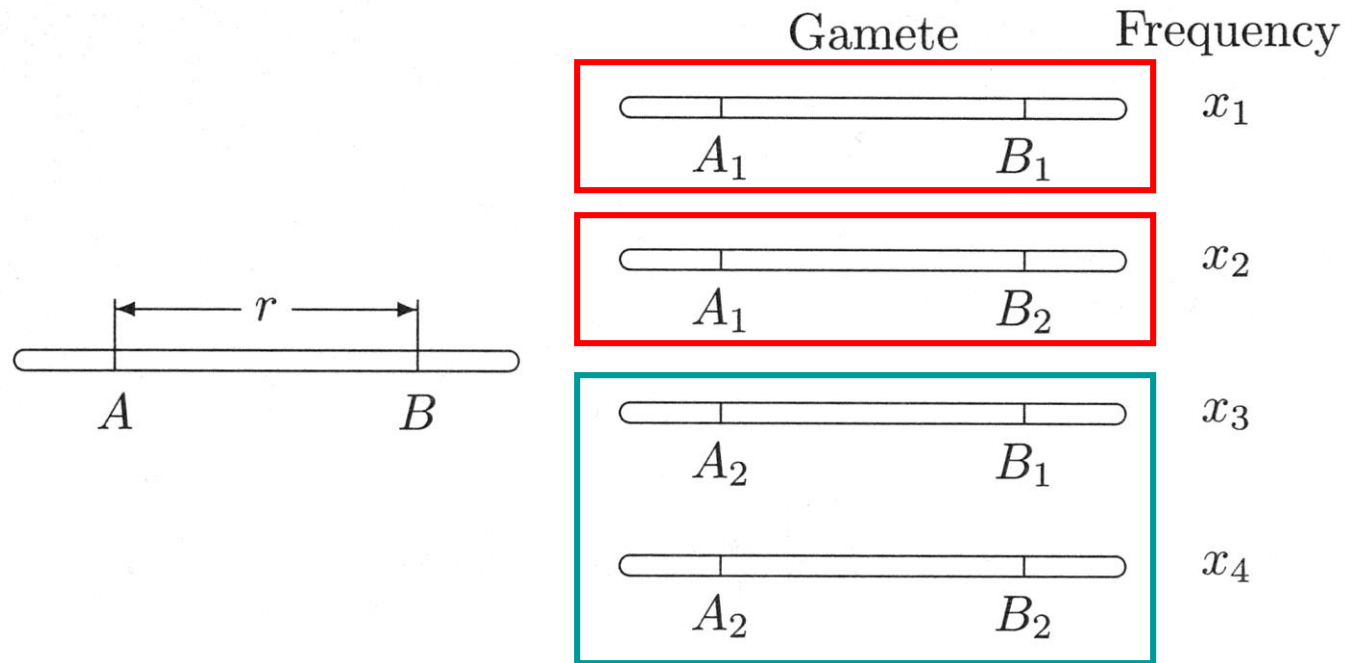**Figure 4.1:** The chromosome on the left shows the position of the A and B loci. The right side illustrates the four possible gametes with their frequencies.

Step 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?

Step 2: The lucky B-allele "hitches a ride" with A_1 …

$p(\mathbf{B_1})$

2pq ≈ 0

r/s = 0

2pq ≈ 0.4

r/s = 0.05

… as long as it remains in *linkage disequilibrium* with A_1 …

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

Initially A_1B_1

2pq = 0.42

r/s = 0.05

2pq ≈ 0.2

Initially A_1B_2

r/s = 0

2pq ≈ 0

p_2

(Gillespie gets confused by his notation here.)

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.

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.
($10^{-5} * 10^4 = 0.1$)

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
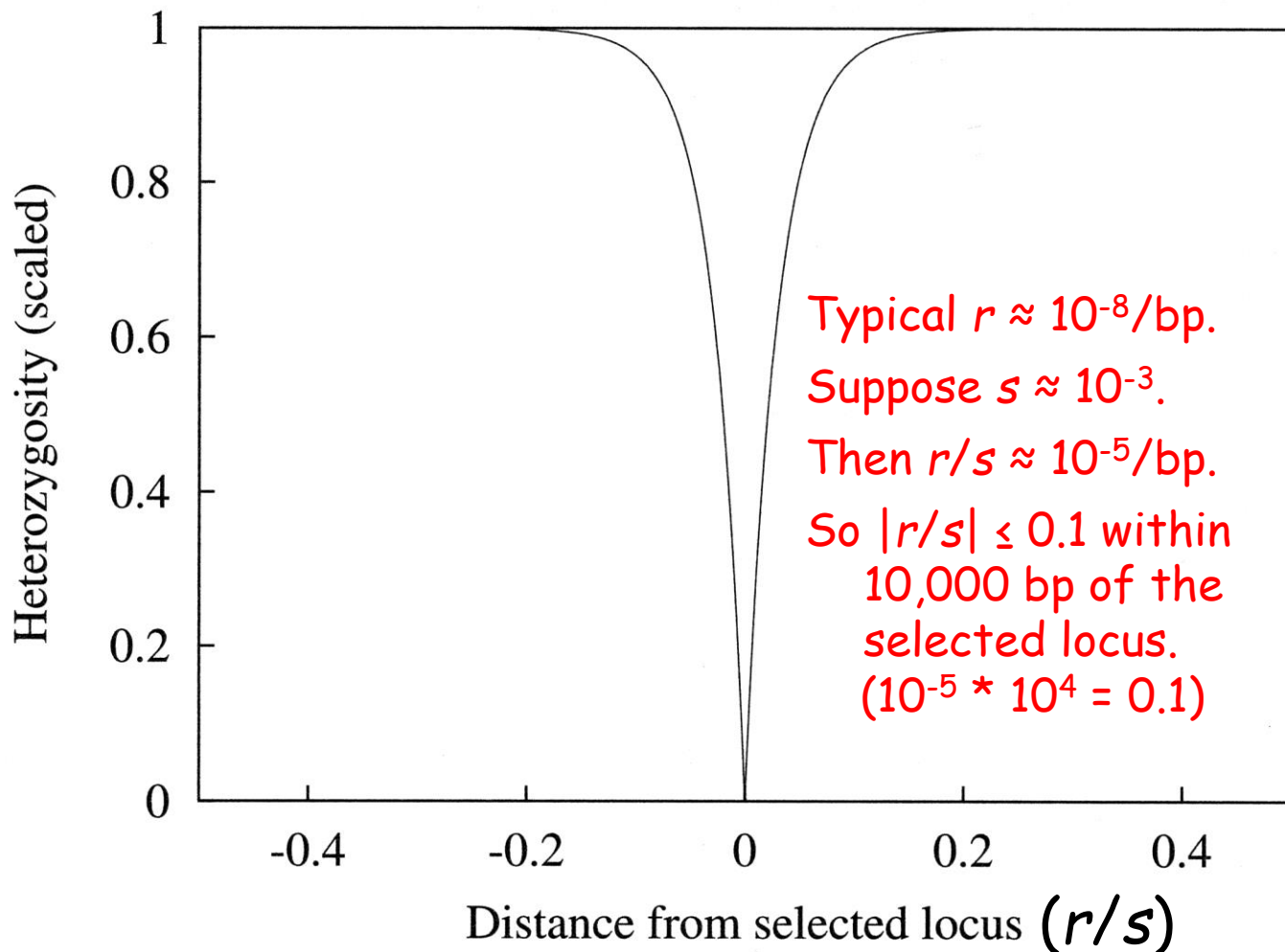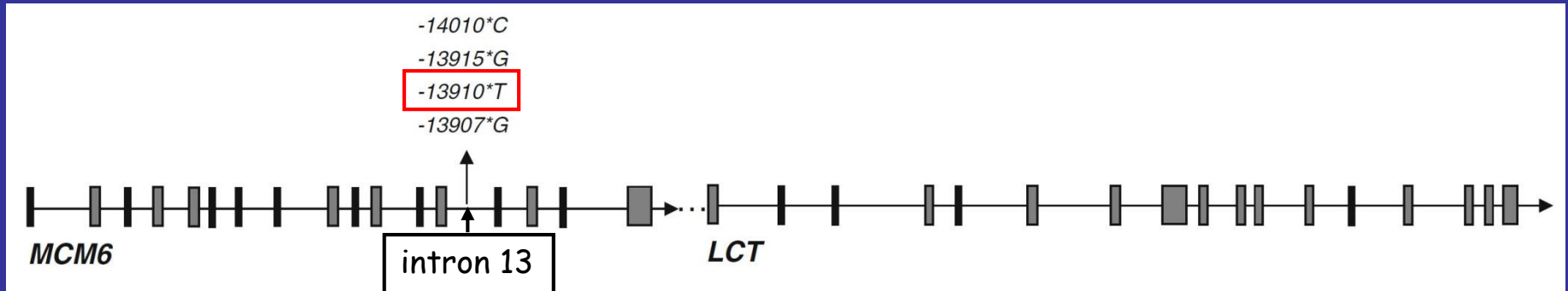
Lactase in Utah, again!

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. (And sorted by their number of 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.

```
cons  aaggaggcgacattccgcttcaggcattcctatctaaacagaccaacgtaAgggtacaatgcctaacccagacgtttcaactctggctgttattcctcgat

01    ....................................................................................................
02    ....................................................................................................
03    ....................................................................................................
04    ....................................................................................................
05    ....................................................................................................
06    ....................................................................................................
07    ....................................................................................................
08    ....................................................................................................
09    ....................................................................................................
10    ....................................................................................................
11    ....................................................................................................
12    ....................................................................................................
13    ....................................................................................................
14    ....................................................................................................
15    ....................................................................................................
16    ....................................................................................................
17    ....................................................................................................
18    ....................................................................................................
19    ....................................................................................................
20    ....................................................................................................
21    ....................................................................................................
22    ....................................................................................................
23    ....................................................................................................
24    ....................................................................................................
25    ....................................................................................................
26    ....................................................................................................
27    ...........................................t.......................................................
28    ...........................................t.......................................................
29    ...........................................................c........................................
30    ...........................................................g........................................
31    gg.a.ca.ag.g.gt.....................................................................................
32    ..................................................G.................................................
33    ..................................................G.................................................
34    ..................................................G.................................................
35    ..................................................G.................................................
36    ..................................................G.................................................
37    ..................................................G..a.gt.....t.........gac.c.tgtct.....a..gc..t.t..c
38    ...........c.....ccgga....gat..at..gg..c.....tc.gGaaa.g..ccttt...tg......c...t.t.........g.....t...
39    ....g....g..c.....ccgga....gat..at..gg..c.....tc.gGaaa.g..ccttt...tg......c...t.t.........g.....t...
40    gg.a..at.gt.c.t..tcc...agtag.t.cat..g.....t..ttccgG..a.gt.....t.........gac.c.tgtct.................
41    ...........c.t..tcc...agtag.t.cat..g.....t.gttccgG..a.gt.....t.........gac.c.tgtct.....a..gc..t.t..c
42    ...........c.t..tcc...agtag.t.cat..g.....t.gttccgG..a.gt.....t.........gac.c.tgtct.....a..gc..t.t..c
43    ..aa..at.gt.c.t..tcc...agtag.t.cat..g.....t.g.tc.gG..a.gt.....t.........gac.c.tgtct.........g..t.t..c
44    ..aa..at.gt.c.t..tcc...agtag.t.cat..g.....t...ttc.gG..acgt.....t.........gac.c.tgtct.....a..gc..t.t..c
45    ..aa..at.gt.c.t..tcc...agtag.t.cat..g.....t.gttc.gG..a.gt.....t.........gac.c.tgtct.....a..gc..t.t..c
46    ....g....g..c.....ccgga....gat..at..gg..c.....tc.gGaaa.g..ccttt...tg......cg.gt.t..ctata.ccg.c..ctcg.
47    gg.a..at.gt.c.t..tcc...agtag.t.cat..g.....t.gttccgG..a.gt.....t.........gac.c.tgtct.....a..gc..t.t..c
48    gg.a..at.gt.c.t..tcc...agtag.t.cat..g.....t.gttccgG..a.gt.....t.........gac.c.tgtct.....a..gc..t.t..c
49    gg.a..at.gt.c.t..tcc...agtag.t.cat..g.....t.gttccgG..a.gt.....t.........gac.c.tgtct.....a..gc..t.t..c
50    .........g..c..tatccgga....g.tc.atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t..ctata.ccg.c..ctcg.
51    gg.a.ca.ag.g.gtta.ccgga....g.t..atc.g.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t...ct...a..gc..t.t..c
52    gg.a.ca.ag.g.gtta.ccgga....g.t..atc.g.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t..ctata.ccg.c..ctcg.
53    gg.a.ca.ag.g.gtta.ccgga....g.t..atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t..ctata.ccg.c..ctcg.
54    gg.a.ca.ag.g.gtta.ccgga....g.t..atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t..ctata.ccg.c..ctcg.
55    gg.a.ca.ag.g.gtta.ccgga....g.t..atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t..ctata.ccg.c..ctcg.
56    gg.a.ca.ag.g.gtta.ccgga....g.t..atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t..ctata.ccg.c..ctcg.
57    gg.a.ca.ag.g.gtta.ccgga....g.t..atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggtg..cg.gt.t..ctata.ccg.c..ctcg.
58    gg.a.ca.ag.g.gtta.ccgga....g.t..atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t..ctata.ccg.ca.ctcg.
59    gg.a.ca.ag.g.gtta.ccgga....g.tc.atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t..ctata.ccg.c..ctcg.
60    gg.a.ca.ag.g.gtta.ccgga....g.t..atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggtg..cg.gt.t..ctata.ccg.c..ctcg.
```

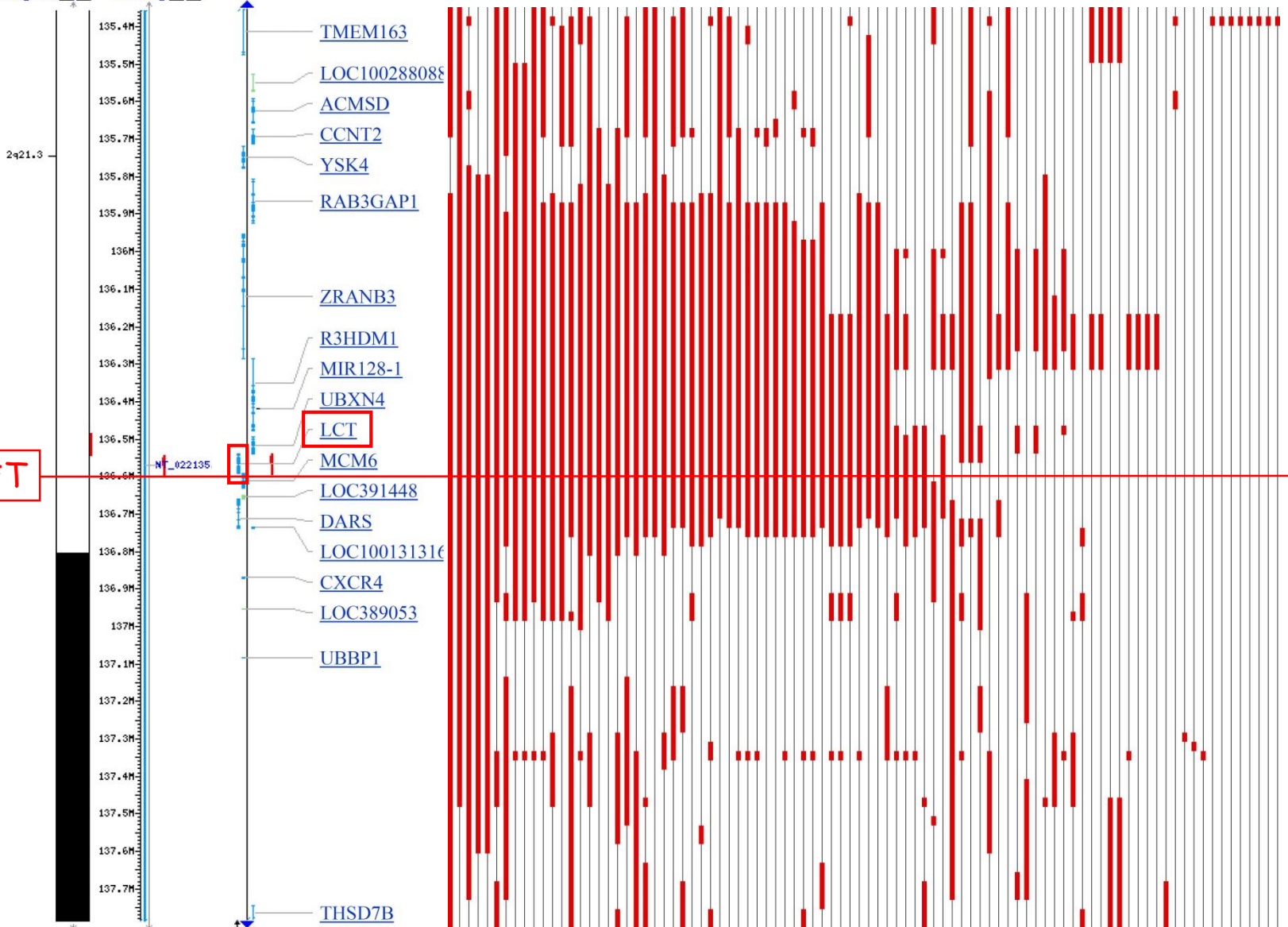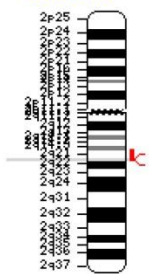Homozygous intervals of chromosome 2 in 90 Utahns

Homo sapiens (human) **Build 37.1 (Current)**
Chromosome: 1 | 2 | 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
Query: LCT [clear]
Master Map: Genes On Sequence
Region Displayed: 135,360K-137,790K bp

Region Shown:
135,360K
137,790K

out
zoom
in
You are here:
Ideogram

LCT-13190*T

2.4 Mbp
19 genes

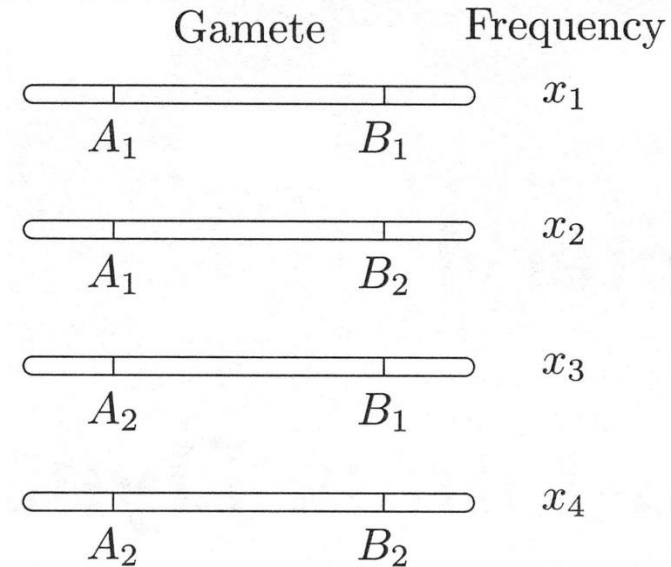# Overall heterozygosity is "drafted down" at HapMap SNP loci.

But why so modestly?

smoothed mean heterozygosity

LCT

4.3 The following program, written in Python, will print out the ratio of the final to starting heterozygosities at the B locus.

Gamete	Frequency


```
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.)

What if recombination is *mutagenic?*
How could we *test* for this?

Heterozygosity

0.012

0.009

0.006

0.003

0

0   0.02   0.04   0.06   0.08   0.1
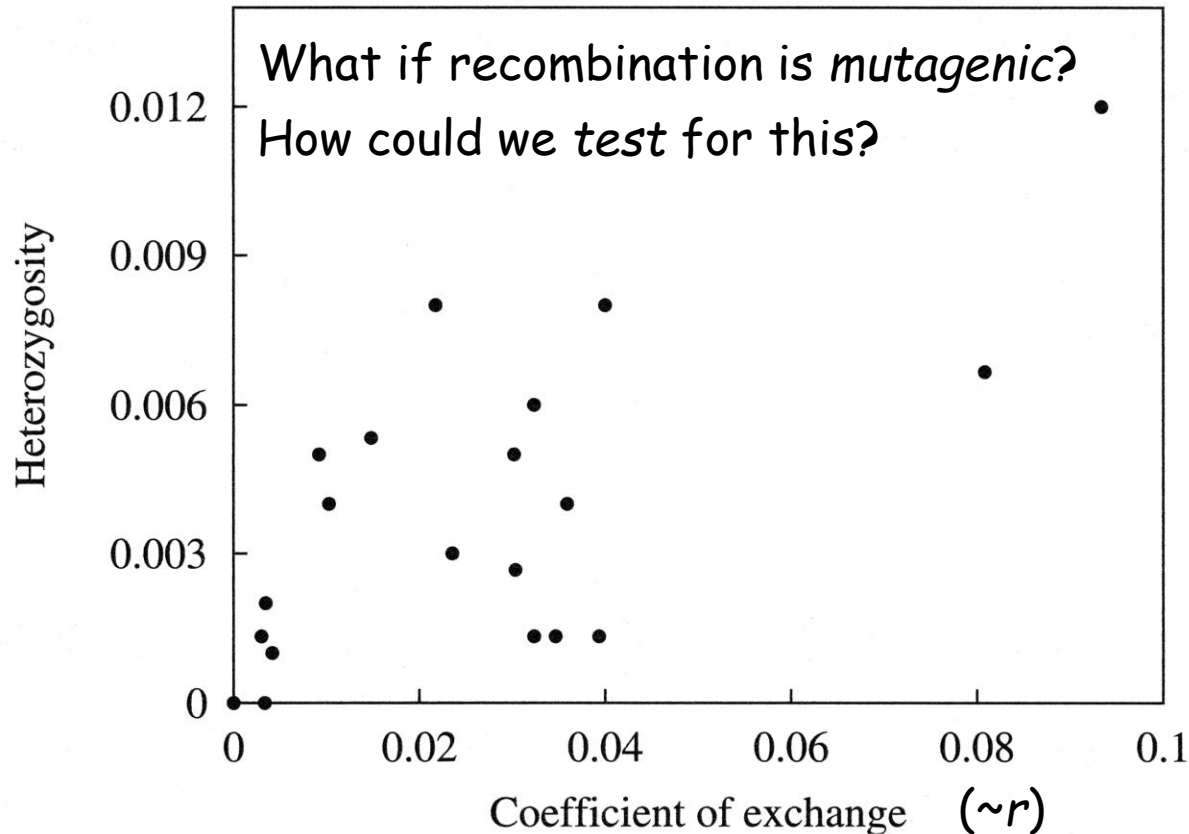
Coefficient of exchange   (~r)

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 (ρ)?

1. The rate of environmental change.

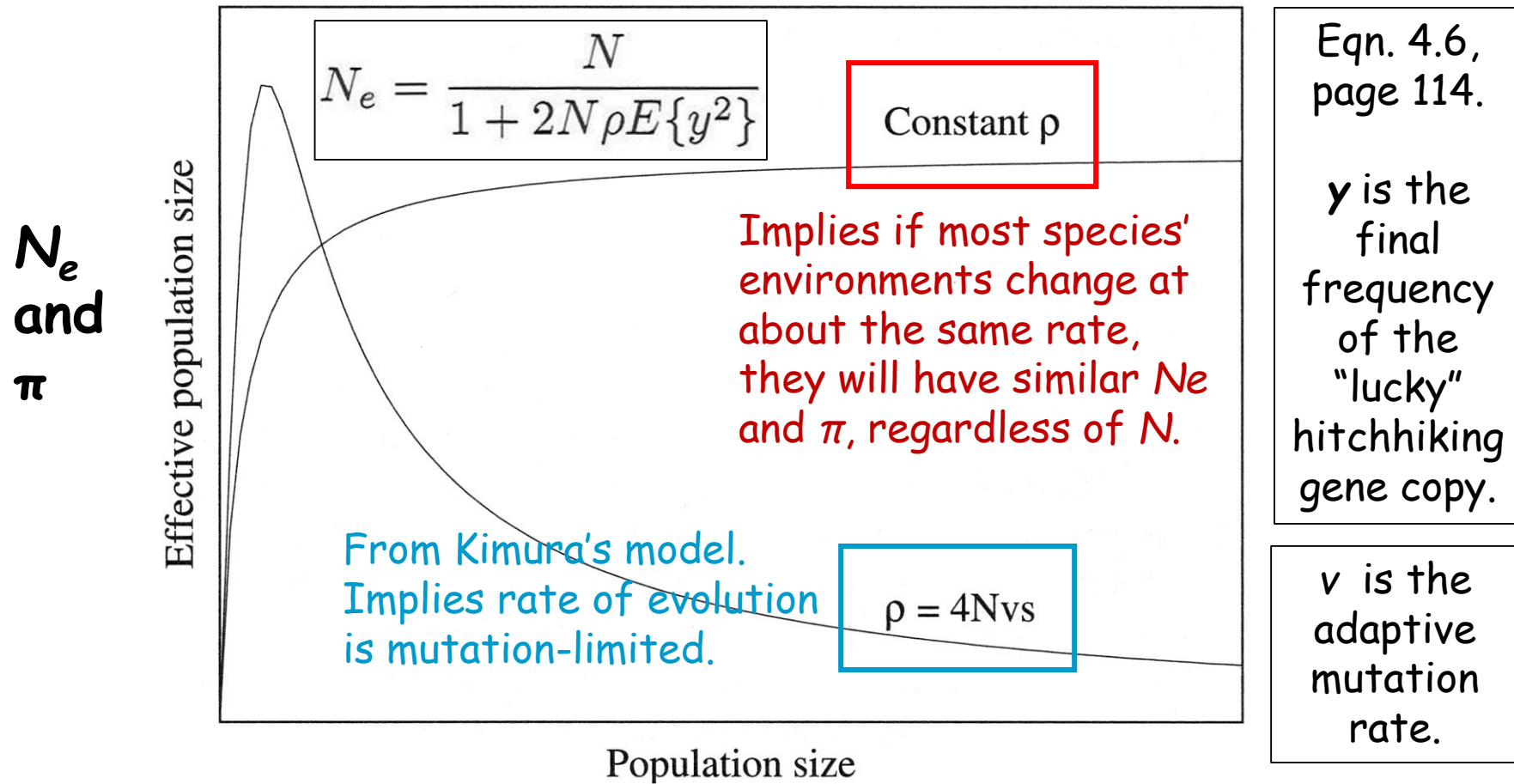2. The rate at which adaptive mutations occur in the population.

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

N_e and π

Effective population size

Constant ρ

Implies if most species' environments change at about the same rate, they will have similar Ne and π, regardless of N.

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

ρ = 4Nvs

Population size

Eqn. 4.6, page 114.

y is the final frequency of the "lucky" hitchhiking gene copy.

v is the adaptive mutation rate.

Figure 4.6: The relationship between the population size and the effective population size under genetic draft.

Mitochondria have no recombination, so $E\{y^2\} = 1$

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

Krill live "naked" in the Southern Ocean which is always changing (e.g. in temperature).

So their rate of adaptation could be mutation-limited.

$\rho = 2N_f vs$

So ρ could increase and N_e decrease with N !

But is the rate of environmental change similar for all *genetic loci,* in addition to being similar for most *species?*

Π 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 4Nμ!

TABLE 2. Estimates of nucleotide diversity (π; heterozygosity at the nucleotide level)

DNA or gene region	Organism	Method	n	bp	π
mtDNA	Human	R	100	16,500	0.004
mtDNA	D. melanogaster	R	10	11,000	0.008
β-Globin	Human	R	50	35,000	0.002
Growth hormone	Human	R	52	50,000	0.002
Notch gene region	D. melanogaster (1)	R	37	60,000	0.005
White locus region	D. melanogaster (2)	R	38	45,000	0.011
Factor IX	Human (3)	S	22	2460	0.0002
Adh locus (C)	D. melanogaster	S	11	765	0.006
Prochymosin (C)	Bovine (4)	S	8	1146	0.004
Growth hormone (C)	Pig (4)	S	6	651	0.007
Class I MHC (HLA-A)	Human (5)	S	5	274	0.043
Class I MHC (H2-K)	Mouse (5)	S	4	273	0.077

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**!

Apparently we don't yet understand the whole story.

For example, could it help explain the evolution of recombination rates?