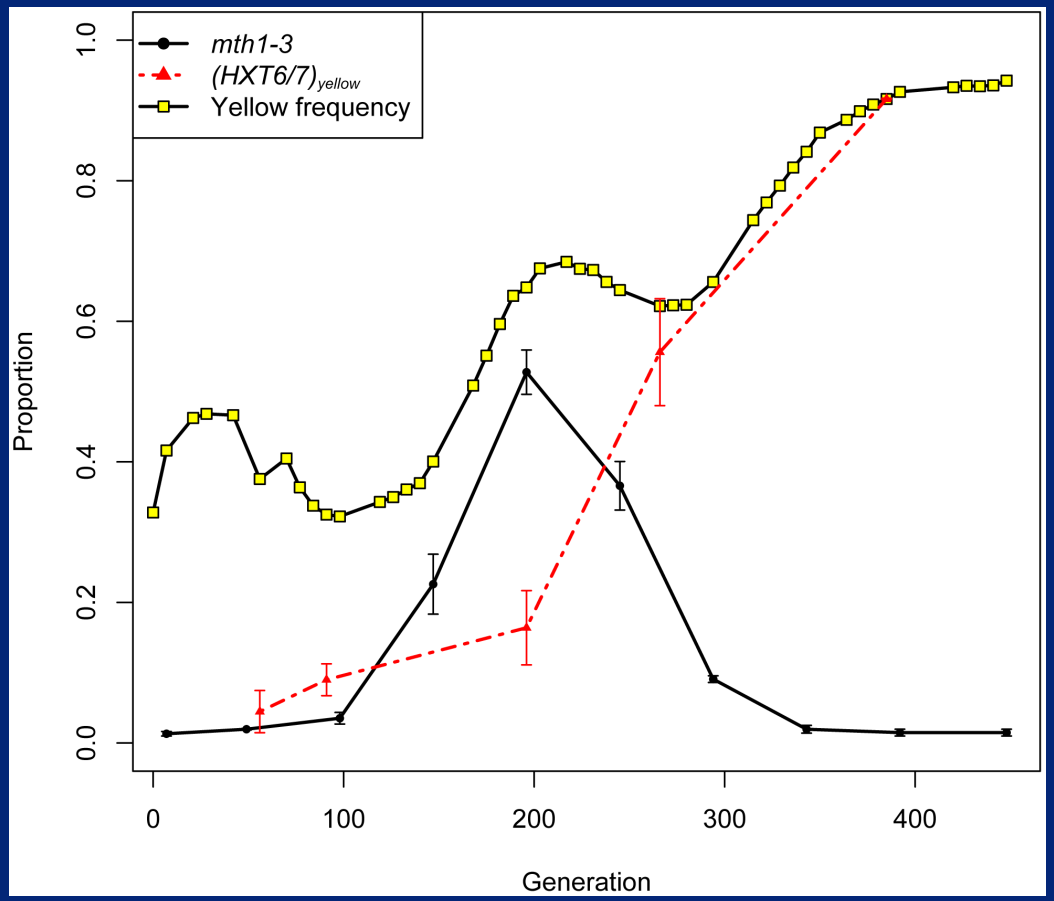
Summary

- Linkage disequilibrium
- Calculating the disequilibrium coefficient
- Recombination
- Selection in the presence of recombination

One-minute responses

- Epistasis is confusing: diagrams? (2x)
- Shorten the introduction a bit.



From Kvitek and Sherlock (2011) PLOS Genet.

Kvitek and Sherlock experiment

- $E.\ coli$ competition in a chemostat
- $mth1$ mutations and $HXT6/7$ mutations both arose several times independently, but never in the same cells (double mutant not fit?)
- Lineages carrying them competed, with $mth1$ taking an early lead but $HXT6/7$ winning in the end
- Authors attribute this to sign epistasis

Linkage equilibrium

- If there is no association between alleles at different loci:
  - Frequency of a haplotype is product of frequencies of its alleles
  - pAB = pA x pB
- This is the normal case for unlinked loci
- Could be disrupted by:
  - Selection
  - Population subdivision with different allele frequencies in the two populations
  - Non-random mating

Linkage equilibrium

- Linked loci can also be in equilibrium if:
  - Recombination has sufficiently reshuffled alleles
  - Recurrent mutation has sufficiently reshuffled alleles (requires a big population)
- Vulnerable to selection, subdivision, non-random mating as before
- Also able to cause lasting disequilibrium between linked loci:
  - Genetic drift
  - Mutation

Disequilibrium

Haplotype	Observed	Expected
AB	58	
Ab	32	
aB	2	
ab	8	
Total	100	

- Is this a haploid or diploid? Doesn't matter here–

Disequilibrium coefficient D

Haplotype	Observed	Expected
AB	0.58	0.54
Ab	0.32	0.36
aB	0.02	0.06
ab	0.08	0.04
Total	1.0	1.0

- D is defined as:
  $pAB - pApB = 0.04$
- Why do we look only at pAB and not the others?
- What's special about AB anyway?
- What does the sign of D mean?

Unlovable qualities of D

- D ranges from $+1/2$ to -1/2
- Most loci can never achieve those limits
- What happens if we try to give our example D=0.5?

Haplotype	Expected	D=0.5
AB	0.54	1.04
Ab	0.36	-0.14
aB	0.06	-0.44
ab	0.04	0.54
Total	1.0	1.0

- The more extreme the allele frequencies, the narrower the range of D

Normalize D

- Define D'=D/max(D)
- For positive D, max(D) is the smaller of pApb and papB
- For negative D, max(D) is the larger of pApB and papb
- This ranges from +1 to -1 for all loci
- Alas, still not indifferent to allele frequencies (Hedrick (1987) Genetics 117)

Another try: r and r^2

- We could compute the correlation coefficient between A and B:

$$r = \frac{D}{\sqrt{pA\ pa\ pB\ pb}}$$

- Alternatively, we might prefer its square:

$$r^2 = \frac{D^2}{pA\ pa\ pB\ pb}$$

- This is also often (wrongly) called the correlation coefficient

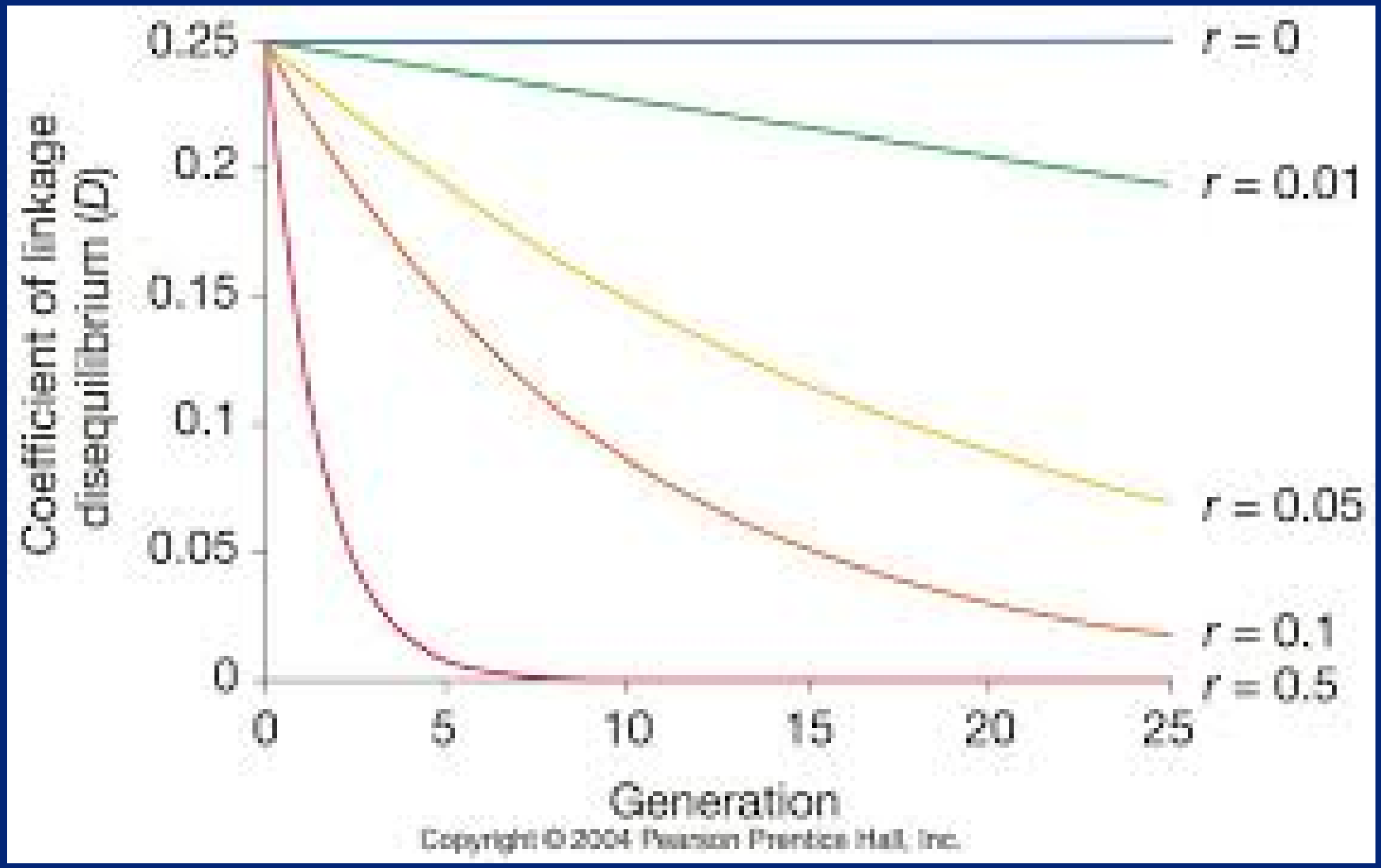
- These measures don't completely get rid of allele frequency either....

Is disequilibrium statistically significant?

- r^2n, where n is the number of sampled chromosomes, has a χ^2 distribution and can be directly used for a statistical test
- One degree of freedom, but why?

Recombination causes disequilbrium to decrease with time

- If no recombination or mutation:
  - In an infinite population, disequilibrium persists
  - In a finite population, it wanders at random (a kind of drift)
- With mutation:
  - Disequilibrium increases when a novel allele arises
  - The same allele arising many times can reduce disequilibrium, but it takes a huge θ to see this
- With recombination:
  - Disequilibrium decays over time
  - The speed of this depends on the amount of recombination (rate times distance)



(the r in this graph is recombination rate times distance)

Where did that graph come from?

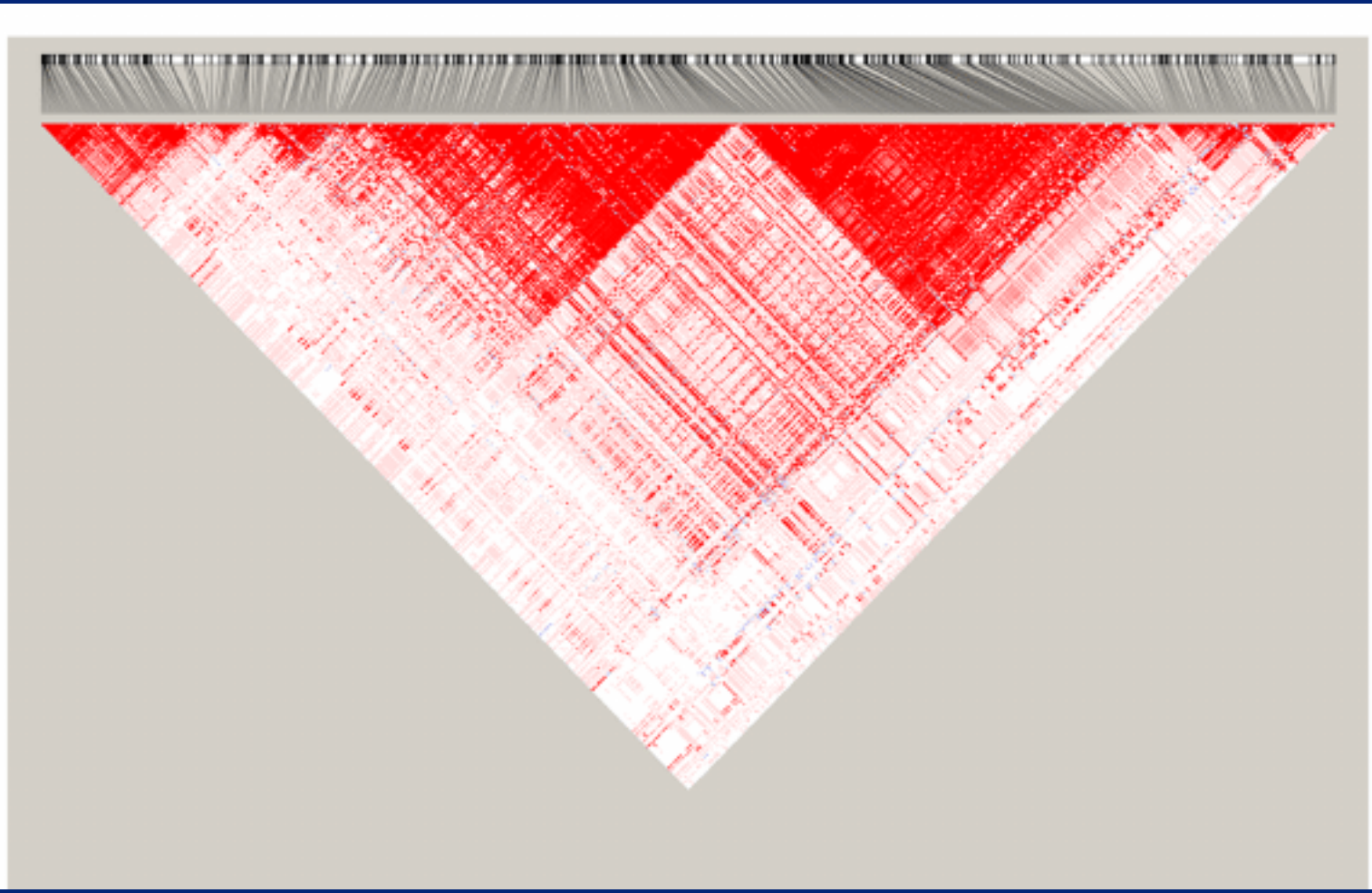
- Call the recombination fraction between the two loci c
- (In what proportion of meioses are these loci separated by a recombination?)
- In one generation:
  $pAB_1 = (1-c)pAB_0 + c\,pA\,pB$
  $D_1 = (1-c)D_0$
- Over n generations:
  $D_n = (1-c)^n D_0$
- This goes to zero with increasing n

But we still see disequilibrium in humans!

- New mutations start out in disequilibrium
- Population subdivision creates disequilibrium
- Selection on multi-locus genotypes can also create disequilibrium
- Drift can cause short-range disequilibrium

Approximate scale of disequilibrium due to drift

- Disequilibrium will be seen at scales where $4Nc < 1$
- c is recombination rate times distance
- This predicts disequilibrium at about 2500 bp scale
- Real human data has much longer disequilibrium tracts
  - Recombination hotspots separated by cold zones?
  - Selection?
  - Population subdivision?



The triangular disequilibrium graph

- Discontinuities often interpreted as hot spots
- They might be:
  - Hotspots definitely make discontinuities
  - Presence of a recombination deep in the coalescent tree can too
  - Need more statistics–overreliance on pictures here

HLA example

- HLA loci are tightly linked on chromosome 6
- High disequilibrium across most of the gene family (but not HLA-DP)
- Many of the alleles are old–found worldwide–and maintained by some kind of balancing selection
- Recombinations in this region are observed in pedigrees
- Why doesn't the disequilibrium decay?

From Bardi et al. (2012) Rev Bras Hematol Hemoter 34.

- In their Brazilian-European sample: p(A*01)=0.1022
  p(B*08)=0.0558
  p(DRB*03)=0.0986

- We would therefore expect the frequency of the haplotype with all three of these to be 0.1022 * 0.0558 * 0.0986 = 0.00056

- Observed value: 0.0270

Possible explanations for this

- This specific combination has a fitness advantage
- The haplotype is linked to a variant that has a fitness advantage
- The haplotype comes from a population which had much higher frequencies of these alleles, and hasn't yet been broken up
- This haplotype can't recombine, or its recombinants don't live
- A bottleneck eliminated most copies of the other haplotypes, but this one got lucky
- Genetic drift jackpot?

Selection on two linked loci with recombination

- If the loci interact (epistasis):
  - No simple equations for what happens
- If alleles with a favorable interaction start out togther, recombination reduces population fitness
- Predicted to lead to strongly interacting loci being clustered in the genome
  - Not much evidence for this in eukaryotes
  - HLA genes might be a counterexample–they are tightly linked in many/most mammals

Friday

- Hitchhiking
- Gene surfing
- “Genetic draft”

One-minute responses

- Please:
  - Tear off a slip of paper
  - Give me one comment or question on something that worked, didn't work, needs elaboration, etc.