Overview

- How much of the genome is functional?
- Selection at multiple unlinked loci
- Interactions among loci
- A first look at linkage

One-minute responses

- Include recent examples of neutrality tests. *Not until after my grant deadline–sorry!*
- What is the scale of diversity–what number would be considered diverse?
- What sorts of questions will be on the exam?
- In-class question: what happens to Tajima's D with directional selection?

Scale of diversity

- Biological populations vary from maybe 10^3 to 10^{11}
- μ per site varies from 10^{-3} to 10^{-9}
- $\theta = 4N_e\mu$ can therefore vary a LOT
- Comparison most useful among fairly related organisms

Prado-Martinez et al. (2013) Nature 499

"Genome-wide patterns of heterozygosity [...] reveal a threefold range in single nucleotide polymorphism (SNP) diversity. Non-African humans, Eastern lowland gorillas, bonobos, and Western chimpanzees show the lowest genetic diversity (0.8×10^3 heterozygotes/bp). In contrast, Central chimpanzees, Western lowland gorillas, and both orangutan species show the greatest ($1.62.4 \times 10^3$ heterozygotes/bp). "

What sort of questions will be on the exam?

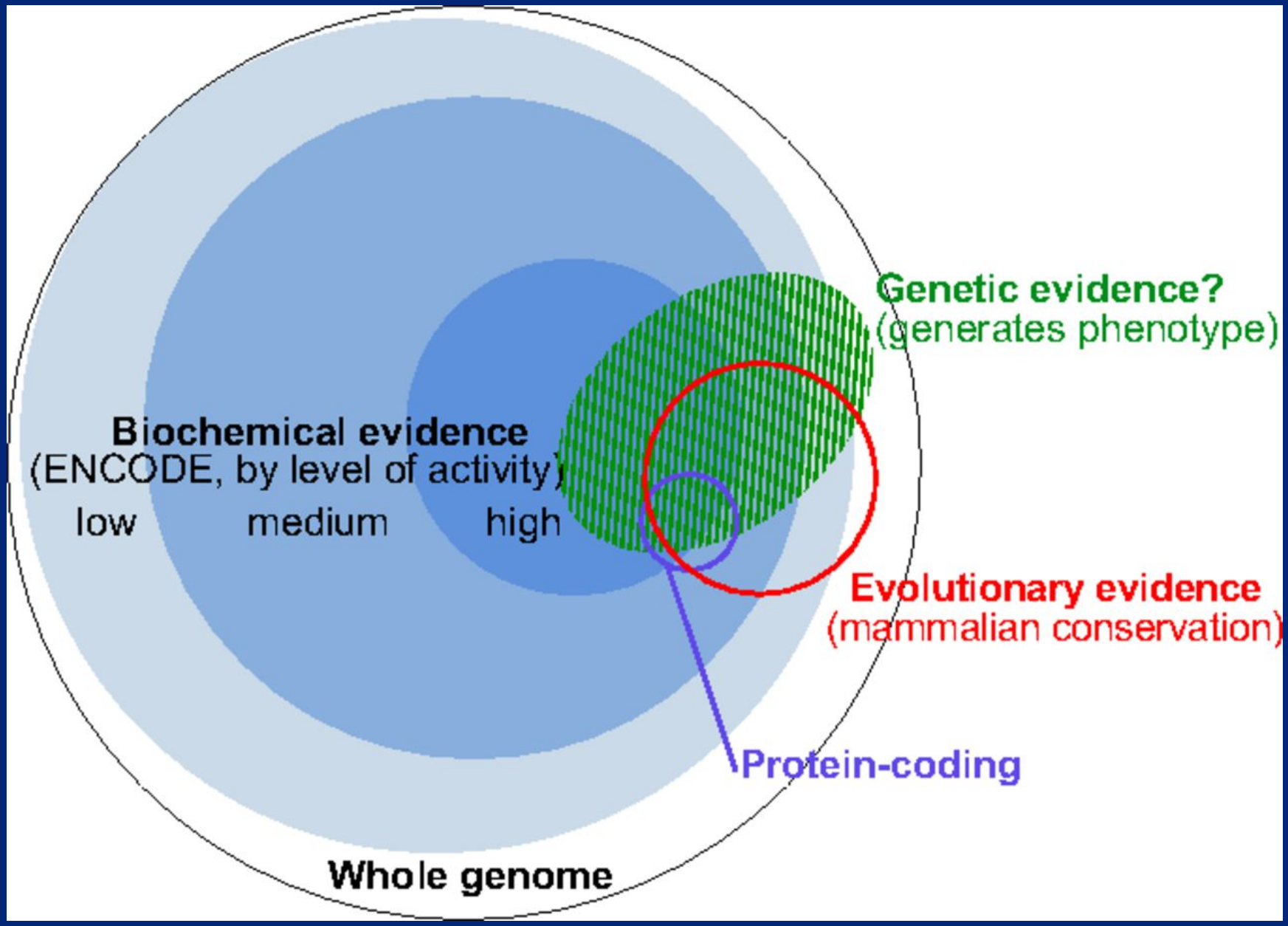
- Similar to the short HW problems
- Mix of calculation and explanation
- Exam will include a formula reminder page; I'll provide a copy in advance so you can look it over

Tajima's D and directional selection

- Consider a locus with repeated sweeps of favorable alleles
- What little diversity it shows will be mostly recent, rare alleles
- $D < 0$ as a result; number of variable sites higher than heterozygosity

ENCODE controversy

- ENCODE study mapped:
  - transcription
  - transcription factor binding
  - chromatin structure
  - histone modification
- "These data enabled us to assign biochemical functions for 80% of the genome"
- (1.5% of the genome is coding sequence)
- ENCODE Project Consortium (2012) Nature 489: 57-74.



From Kellis et al. (2014) PNAS 111: 6131-6138

Could 80% of the genome be under selection?

Based on Kellis et al. (2014)

- Arguments for:
  - Pervasive evidence of biochemical activity
  - GWAS for phenotypes often lands in areas lacking known functional elements
- Arguments against:
  - Haldane argument: can a population afford selection on very many loci?
  - Much of the genome is repeats: they may be "active" but are they meaningful?
  - Lack of conservation–only 5% of genome strongly conserved in mammals
  - Low N_e of large mammals makes very weak selection ineffective

Haldane's argument: "Genetic Load"

- Haldane argued that the cost of a harmful allele to a population is nearly independent of s:
  - Every copy added by mutation must eventually be removed by selection (a "selective death")
  - Strongly harmful alleles hurt a few individuals a lot, then are gone
  - Weakly harmful alleles hurt each individual less, but hang around longer
- How many "selective deaths" can a population handle?
- Depends on reproductive excess

Weaknesses in this argument

- Hard selection:
  - Regardless of competition, unfit genotype tends to die (or fail to reproduce)
  - Too much of this threatens the population's survival
- Soft selection:
  - In the absence of competition, all genotypes are viable
  - "Unfit" genotypes have a competitive disadvantage in the presence of fitter ones
  - Does not reduce population viability
- bad alleles tend to be rare: drift may do some of selection's dirty work for it

Another question about genetic load

- How do different loci under selection interact?
- Some forms of interaction allow us to detroy many bad alleles with a single selective death
- This could reduce the fitness cost of having many loci under selection

Selection on more than one gene

Note: The next several slides discuss TWO LOCI in a HAPLOID

- Two loci, one with alleles A/a, one with alleles B/b
- We know the fitness at each locus separately:

Genotype	A	a	B	b
Fitness	1.0	0.4	1.0	0.2

- If the loci don't interact at all, we can multiply the fitnesses:

Genotype	AB	Ab	aB	ab
Fitness	1.0	0.2	0.4	0.08

How realistic is that multiplication?

- Multiplication says: either factor could kill you, and they don't interact at all
- More realistic for unrelated functions:
  - Nearsightedness and atherosclerosis–seems reasonable
  - Cystic fibrosis and emphysema–not so reasonable
- More realistic for small fitness effects:
  - Critter with huge problems from one locus might not be able to handle problems at a second locus

Epistasis

- Definition: phenotype produced by alleles at one locus depends on alleles at another locus
- Also often means: fitness of genotype at one locus depends on alleles at another locus
- Often talked about in terms of "wild type" and "mutant"
- Magnitude epistasis: the direction of selection doesn't change, but its intensity does
  - Negative epistasis: double mutant worse than expected
  - Positive epistasis: double mutant better than expected
- (This rapidly becomes confusing if there is no clear "wild type")

Negative epistasis

- Negatively epistatic:

Genotype	AB	Ab	aB	ab
Fitness	1.0	0.2	0.4	0.0 expected 0.08

- With these fitnesses, selection will remove a and b faster when both are present

- Also negatively epistatic:

Genotype	AB	Ab	aB	ab
Fitness	1.0	1.2	1.4	1.5 expected 1.68

- With these fitnesses, selection will increase a and b slower when both are present

Negative epistasis example

Khan et al. (2011) Science 332: 1193-1196.

- Experimental evolution of $E.\ coli$
- Studied all pairwise combinations of five best favorable mutations
- For 4 out of 5, double mutant not as good as single mutants would predict
- Exerts a "drag" on improvement in fitness
- Another way of saying this: For 4/5 of the mutations studied, the higher the organism's current fitness, the less gain for adding another favorable mutation

Positive epistasis

- Double mutant better than expected
- Positively epistatic:

Genotype	AB	Ab	aB	ab
Fitness	1.0	0.2	0.4	0.18 expected 0.08

- This slows down removal of a and b from the population
- Also positively epistatic

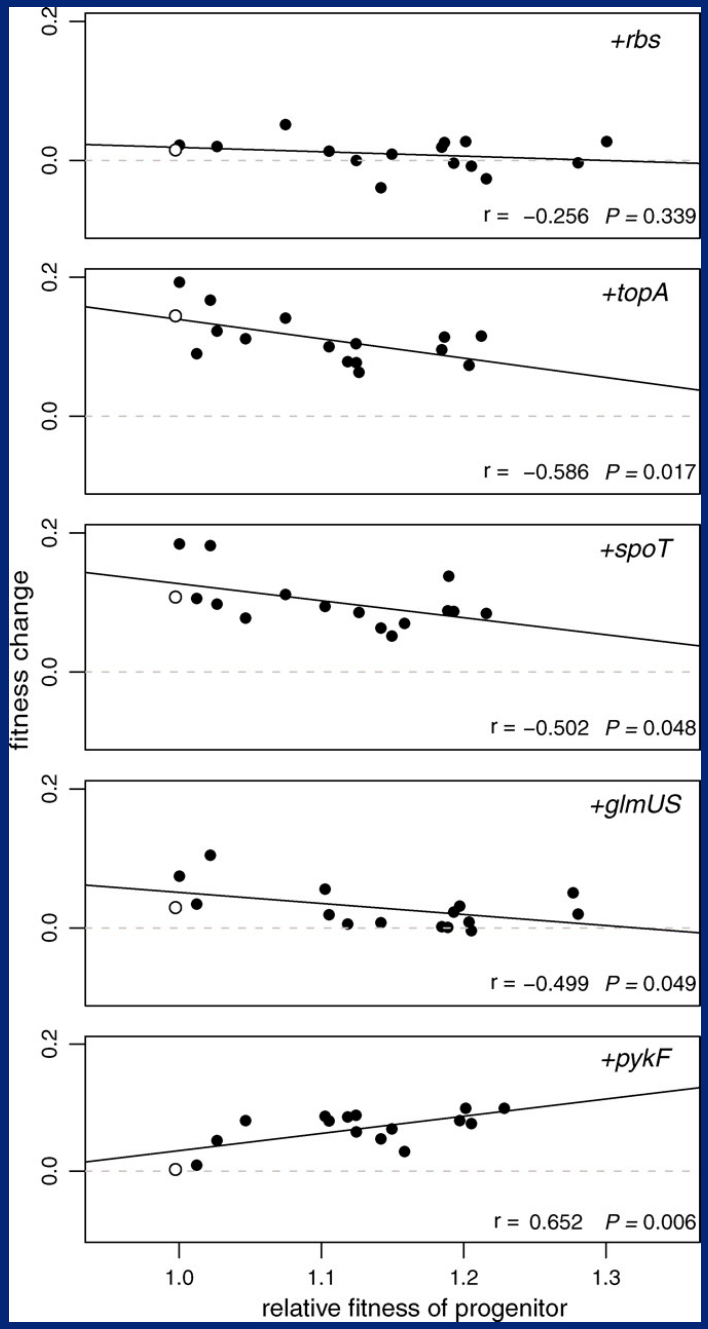
Genotype	AB	Ab	aB	ab
Fitness	1.0	1.2	1.4	1.8 expected 1.68

- This speeds up fixation of a and b

Positive epistasis example

Khan et al. (2011)

- The fitter a strain was, the more it benefited from a $pykF$ mutation


+rbs
r = −0.256 P = 0.339
+topA
r = −0.586 P = 0.017
+spoT
r = −0.502 P = 0.048
+glmUS
r = −0.499 P = 0.049
+pykF
r = 0.652 P = 0.006
fitness change
0.2
0.0
1.0
1.1
1.2
1.3
relative fitness of progenitor

Sign epistasis

- Whether an allele is good or bad depends on alleles at another locus

Genotype	AB	Ab	aB	ab
Fitness	1.0	0.2	0.4	1.2

- Whether a is good or bad depends on whether you have B or b
- Example: gene that synthesizes an antibiotic and gene that makes you immune to the antibiotic
- Behaves like underdominance: the outcome will depend on starting allele frequencies
- Population may be unable to optimize fitness if it starts in the wrong place

Epistasis in diploids

- Same idea: just more genotypes to track
- Two loci controlling pigment in mice:

Genotype	A?B?	aaB?	A?bb	aabb
Phenotype	Brown	White	White	White

- Assume white mice have a fitness disadvantage in the wild
- What kind of epistasis is this?

Linkage

- What if the two loci were completely linked?
- We know the fitness at each locus separately:

Genotype	A	a	B	b
Fitness	1.0	0.4	1.0	0.2

- If the loci don't interact at all, we can multiply the fitnesses:

Genotype	AB	Ab	aB	ab
Fitness	1.0	0.2	0.4	0.08

- What if only Ab and aB exist in the population, and there is no recombination or new mutation?

A diploid example

Genotype	AA	Aa	aa	BB	Bb	bb
Fitness	1.0	0.8	0.4	1.0	0.9	0.8

- If only Ab and aB in population, and no recombination, what will happen?

Implications of linkage

- Linkage between loci:
  - Can cause them to behave as a singe locus
  - Interferes with the action of selection
- Recombination may exist to allow more efficient selection
- This is surprisingly controversial

Summary

- Epistasis is interaction between phenotypes (or fitnesses) of two loci
- Magnitude epistasis
  - Positive epistasis: double mutant better than expected
  - Negative epistasis: double mutant worse than expected
- Sign epistasis: direction of one locus' fitness slope depends on other ocus
- Magnitude epistasis changes speed of evolution
- Sign epistasis can behave like underdominance–outcome depends on initial frequencies
- Linkage between loci can interfere with selection on them

Wednesday

- Linkage disequilibrium
- Selection on linked loci with recombination

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.