Homework

- HW1 mean: 17.4
- HW2 mean (among turned-in hws): 16.5
- If I have a fitness of 0.9 and you have a fitness of 1.0, are you 10% better?

Also, equation page for midterm is on the web site for study

Testing for neutrality versus selection

1. Types of selection
2. Frequency dependent selection
3. Synonymous versus non-synonymous substitutions
4. Within-species versus between-species comparisons

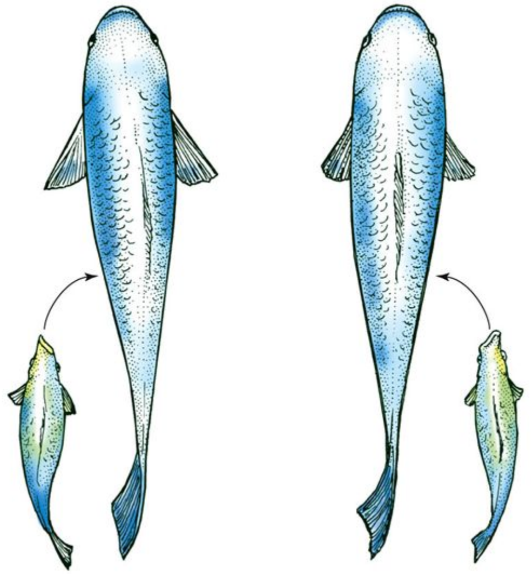
Terms for types of selection

- Purifying selection:
    - Selection against a bad new variant
    - Preserves the original sequence
- Directional selection:
    - Selection for a good new variant
    - Changes to a new sequence
- Balancing selection:
    - Selection to maintain multiple alleles
    - Overdominance, frequency dependent selection

Frequency-dependent selection

- Sometimes having a rare trait is an advantage
- This behaves like overdominance

Figure 7.10 Two hereditary forms of an African cichlid fish

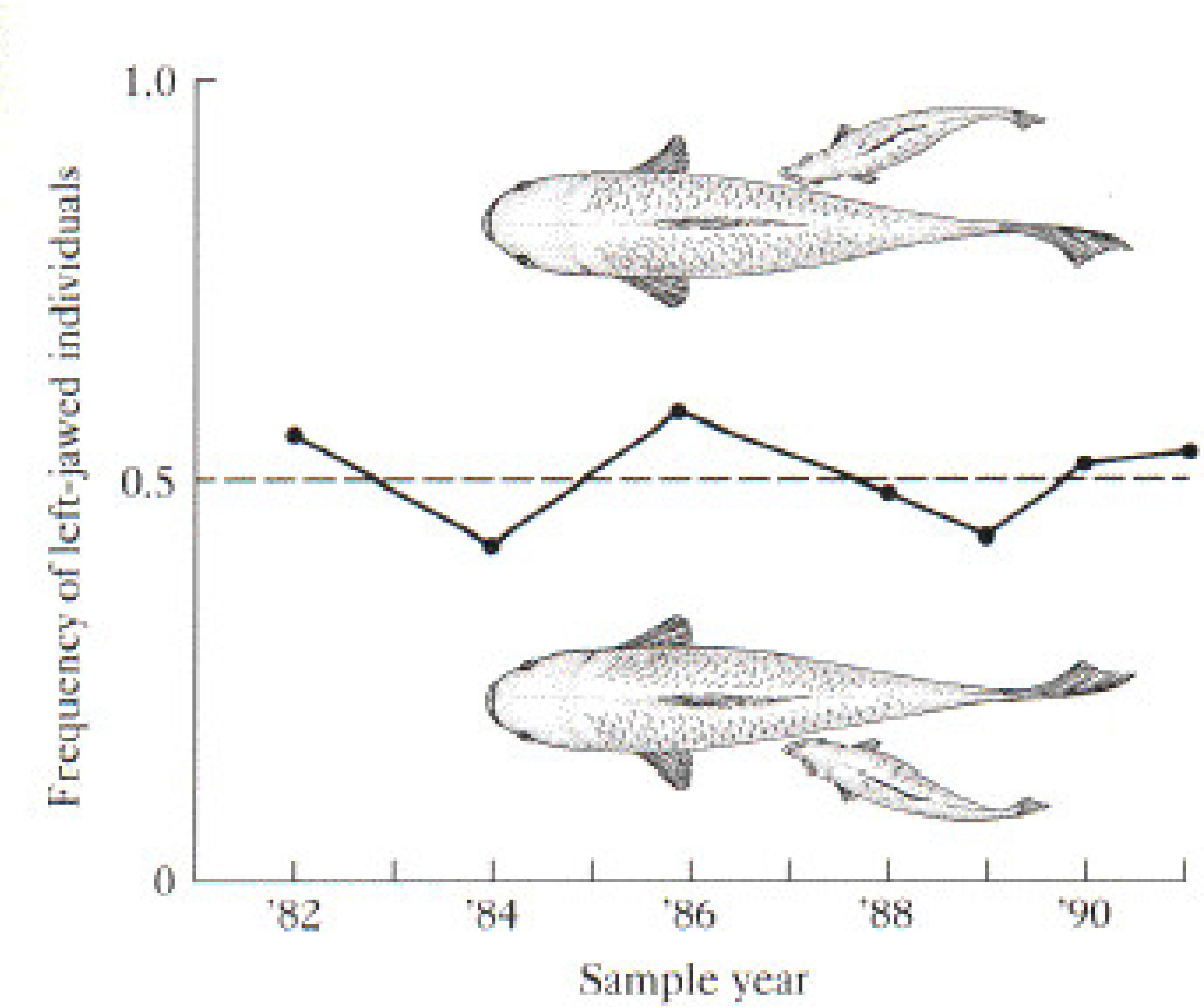


“Right-jawed” *Perissodus* attack prey from the left rear side

“Left-jawed” *Perissodus* attack prey from the right rear side

ANIMAL BEHAVIOR 9e, **Figure 7.10**

Frequency-dependent selection



Frequency-dependent selection examples

- Rare individual can exploit an underused resource
- Rare individual is sexually attractive
- Rare individual has different disease susceptibility than others, so doesn't catch common diseases
- Rare individual does not fit predator's expectations

Why look for selected genes?

- Understand an organism's recent history:
  - Which genes were selected as humans changed rapidly?
- Find genes important to a function:
  - Which genes are selected when we treat malaria with drugs?
  - Which genes were selected in domestication of plants or animals?
- Identify non-functioning genes:
  - Which apparent genes are non-selected (thus probably non-used)?

Retrospective tests of selection

- It would be ideal to measure selection directly
- We usually can't
- Gene sequences can provide indirect measures of selection

Synonymous versus non-synonymous substitutions

Within a protein-coding gene:

- *Synonymous* (silent) substitutions don't change the protein sequence
- *Non-synonymous* (coding) substitutions do change it
- Selection mostly acts on coding substitutions
- Silent substitutions mainly reflect the mutation rate

Silent vs. coding substitutions

Why can't we just count silent and coding substitutions?

- The genetic code gives more coding than silent targets:
  - Most 1st position changes are coding
  - All 2nd position changes are coding
  - Most 3rd position changes are silent

- Therefore, we count substitutions PER TARGET POSITION

- I will present an oversimplified method; real methods have to deal with multiple hits in the same codon

Examples

ATG – Methionine
No other codon means methionine, so this codon contributes 3 coding targets.
GTT – Valine
GTC – Valine
GTA – Valine
GTG – Valine
No other codon means valine, so this codon contributes 2 coding targets and 1 silent target.

Examples

GAT – Asparagine
GAC – Asparagine
GAA – Glutamic acid
GAG – Glutamic acid

This codon contributes 2.67 coding targets and 0.33 silent target (one-third of the 3rd position changes are silent).

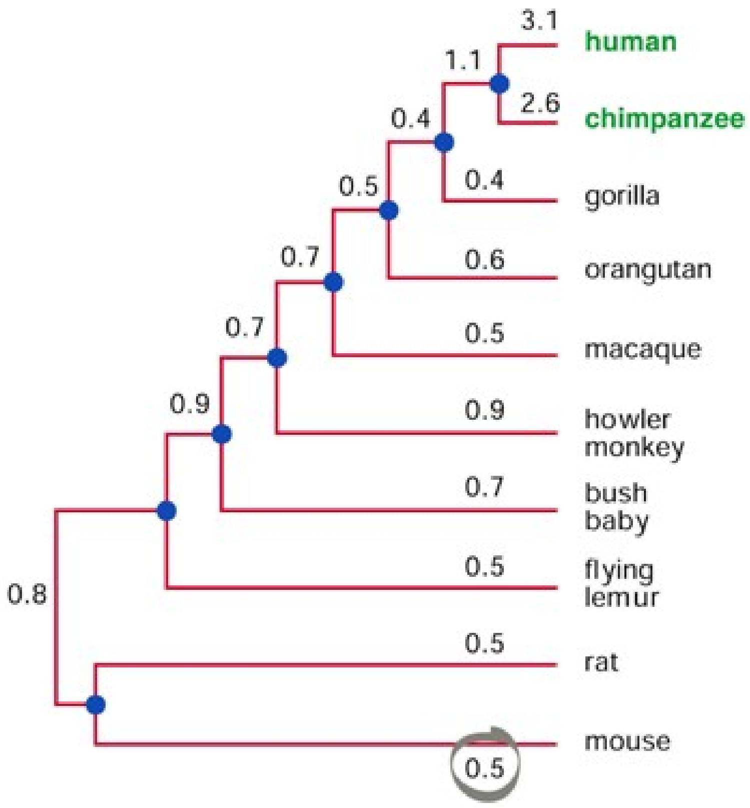
Synonymous versus non-synonymous substitutions

- D_s – number of synonymous changes per synonymous site
- D_n – number of nonsynonymous changes per nonsynonymous site
- (You may prefer the words ”silent” and ”coding”)

Synonymous versus non-synonymous substitutions

- Masatoshi Nei proposed $\omega = D_n/D_s$ as a test for selection
- $\omega = 1$ indicates neutrality
- $\omega < 1$ indicates purifying selection
- $\omega > 1$ indicates balancing selection or directional selection

ω in the *BRCA1* gene



Assumptions of this test

- Test makes several assumptions:
    - Gene has many sites under selection
    - Not a mix of multiple kinds of selection
    - Only coding changes are important to natural selection

- Despite these limitations, Nei's test has been powerful in finding selected genes:
    - Pseudogenes are often recognized by $\omega \approx 1$
    - Interesting genes have been found by scanning for $\omega >> 1$

- A major limitation is that it can't detect selection on control regions

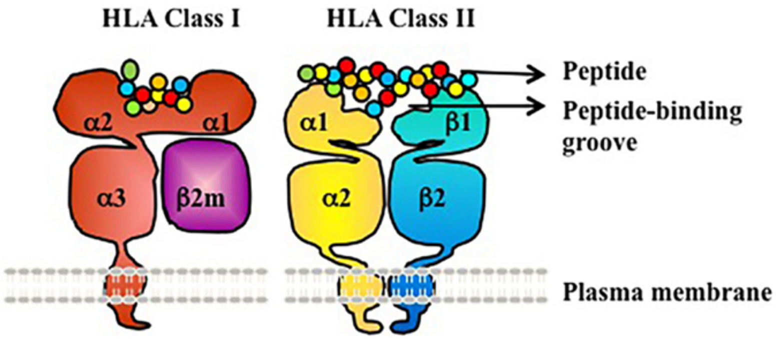
Terminology confusion

- This test is common and well accepted
- However, it has many names in the literature:
    - ω
    - D_n/D_s
    - dN/dS
    - k_N/k_S
    - Nei's test of selection
    - Nei's test of neutrality
- These are all the exact same test

Synonymous versus non-synonymous substitutions

In HLA:

- Antigen-binding region, $\omega \approx 3$
- Elsewhere in the gene, $\omega << 1$



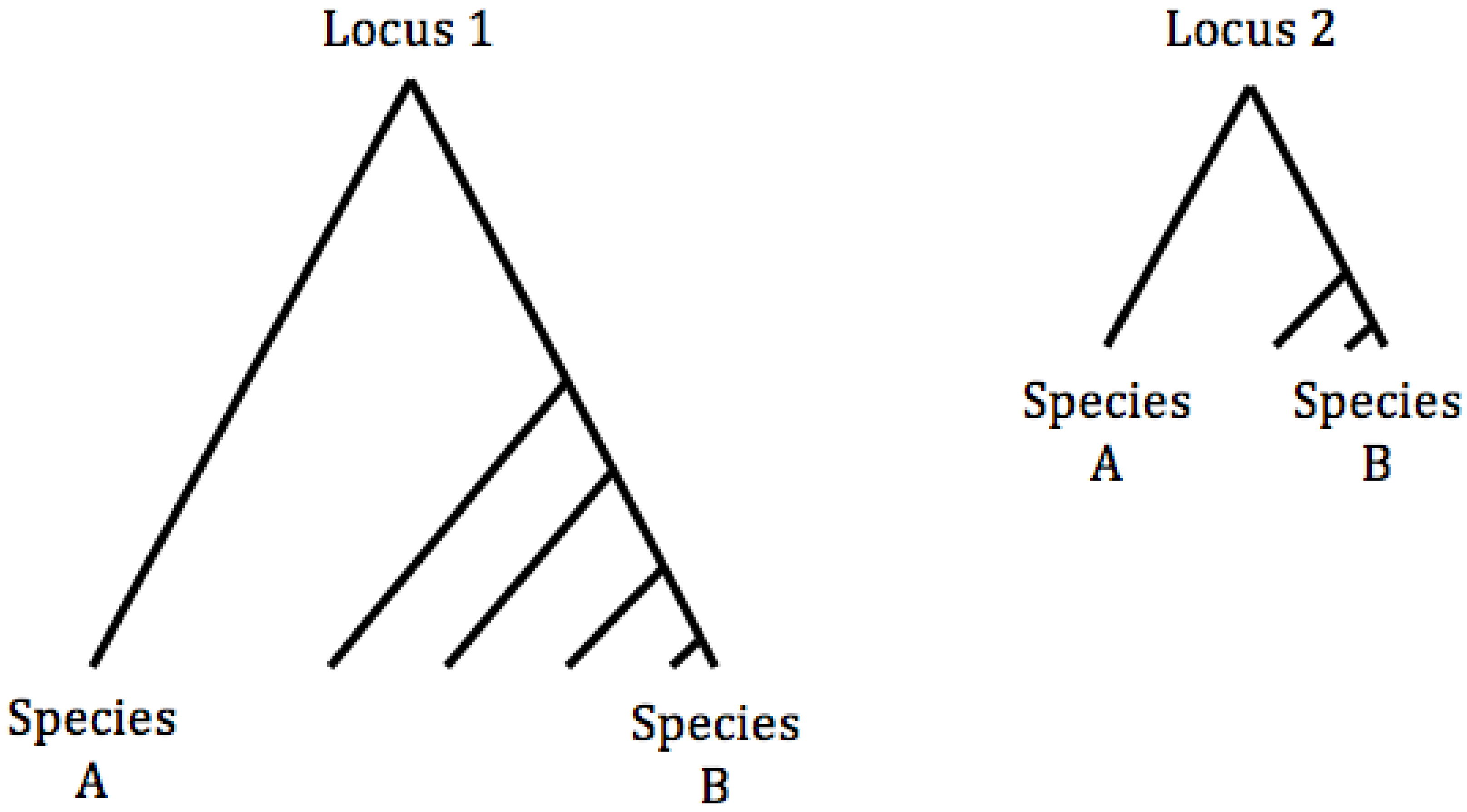
What could this be?

- Initially interpreted as overdominance
- Frequency-dependent selection (rare allele advantage) looks the same and cannot be ruled out
- Not high mutation rate: mutation should affect D_s and D_n equally
- Could it be rapid directional selection?

What could this be?

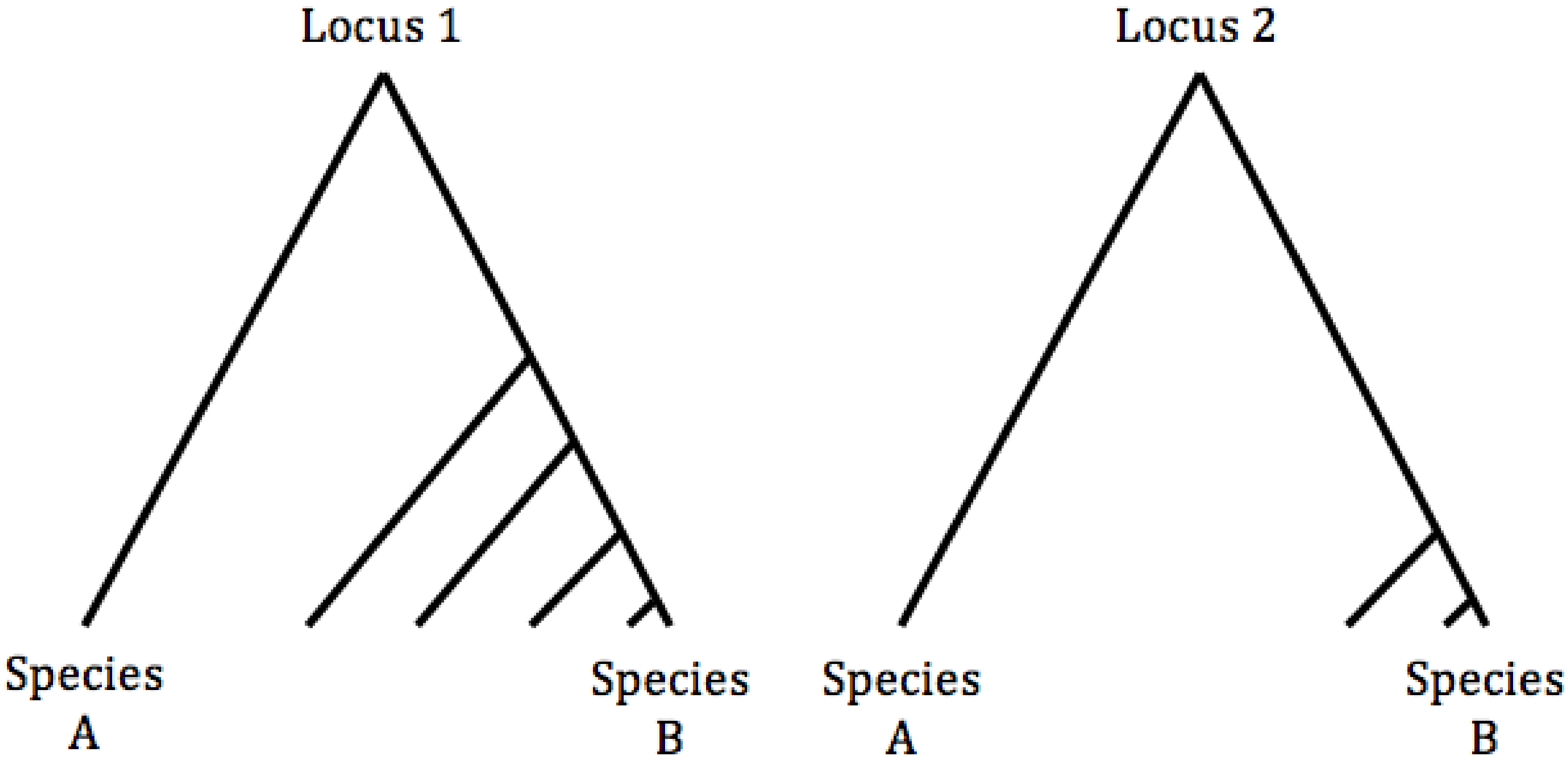
- Could it be rapid directional selection?
- Ruled out by comparison with other primates:
  - Directional selection should cause species to become dissimilar
  - Humans, chimps and gorillas share some identical HLA alleles
- Test for selection by comparing species

Hudson, Kreitman and Aguade (HKA)



Two loci evolving in the same way (though with different mutation rates)

Hudson, Kreitman and Aguade (HKA)



Two loci evolving in different ways–at least one is under selection

Hudson, Kreitman and Aguade (HKA)

- If variation is neutral, polymorphism within species and divergence between species both depend on μ
- Selection can disrupt this:
  - Bad variants may persist in a population but won't be fixed between species
  - Variants that are good in just one species will rapidly fix there
- HKA compares within-species and between-species differences at two regions
- Pick one region that is probably neutral (junk DNA) and compare a possibly interesting region to it

HKA example

	Gene1	Gene2
Differences between species	100	180
Differences within species	25	20

Is the ratio of between to within the same in both genes?

HKA example

	Gene1	Gene2
Differences between species	100	180
Differences within species	25	20
Ratio	4:1	9:1

What could this mean? Assume that Gene1 is a probably neutral pseudogene.

HKA example

	Gene1	Gene2
Differences between species	100	180
Differences within species	25	20
Ratio	4:1	9:1

- Gene2 diverges among species unusually fast for the amount of polymorphism (raw genetic material for divergence) that it possesses.
- Strong directional selection fixing favorable mutations at Gene2
- Gene2 might be involved in the difference between the species

Another HKA example

	Gene1	Gene2
Differences between species	100	120
Differences within species	25	95

- Again, assume Gene1 is neutral.
- (This test only compares genes; it can't tell us if our baseline gene is neutral or not.)

Another HKA example

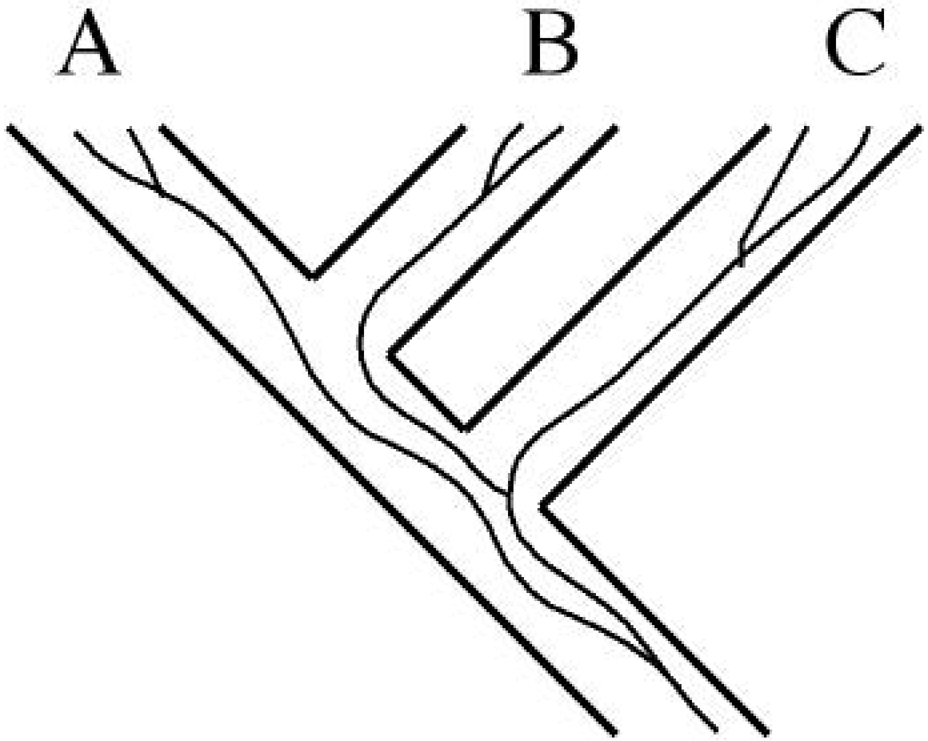
	Gene1	Gene2
Differences between species	100	120
Differences within species	25	95
Ratio	4:1	1.2:1

- Gene2 has too much polymorphism for its amount of divergence.
- This may represent:
  - Weakly harmful alleles waiting to be eliminated by selection
  - Overdominant alleles kept in polymorphism
  - Frequency dependent selection

HKA assumptions

- This test makes some assumptions
  - The "neutral" comparison gene is really neutral
  - Mutation rate constant for each gene (doesn't need to be equal between genes)
  - No large changes in population size
  - We are not in an "ancestral polymorphism" case where the divergence time of the two genes is greatly different
- Measure statistical significance with a χ^2 test

Ancestral polymorphism?

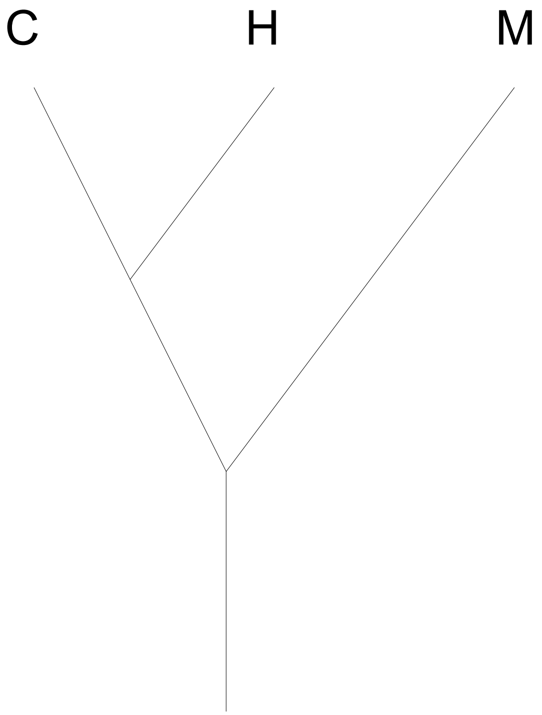


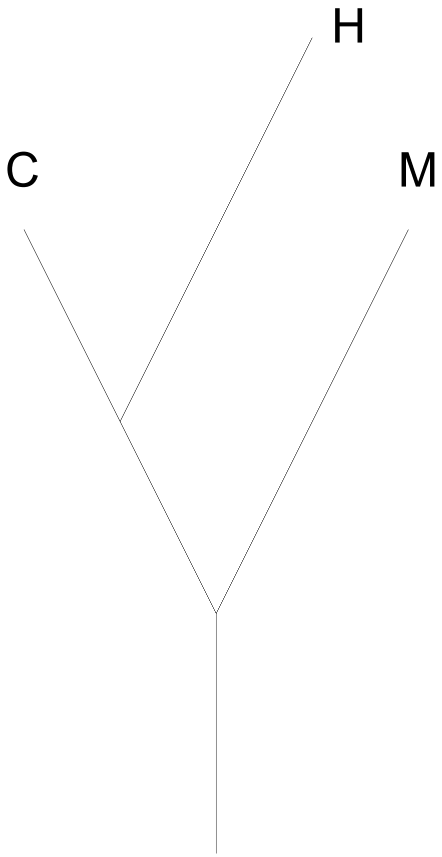
MacDonald and Kreitman

- A similar concept to HKA
- Under neutrality:
- D_s(within species)$/D_s$(between species)$=D_n$(within species)$/D_n$(between species)
- Deviation from this indicates some kind of selection
- Not used as frequently (I don't know why)

Humans and chimpanzees

- Andy Clark and co-workers compared humans and chimpanzees using mouse as the outgroup.
- They looked for genes with accelerated evolution in human compared to chimp and mouse


C
H
M


H
C
M

Brainstorm

- What could cause a long branch?
- If all human genes showed long branches, what could that mean?
- If only certain human genes showed long branches, what could that mean?

Accelerated evolution in the human lineage

Some ideas:

- Adaptive evolution in humans
- Deterioration in humans due to fixing bad mutations (bottlenecks?)
- Weaker selection on humans (technology?)
- Increased mutation rate in humans
- Decreased mutation rate in chimpanzees
- Shorter generation time in humans than chimpanzees

Humans and chimpanzees

Gene categories whose evolution has accelerated in human evolution (Clark et al. 2003):

- Senses
- Digestion and food metabolism
- Reproduction, especially spermatogenesis
- Immune system and tumor suppression
- NOT brain function

Flaws in this comparison?

- Significant changes from one big mutation
- Coding regions only
- Some “mutations” are really polymorphisms, and their frequency depends on population size
  - Chimp long-term population size is larger than human, so this does not explain away human-specific increases
- Some false positives likely due to large number of comparisons

One-minute responses

- Tear off a half-sheet of paper
- Write one line about the lecture:
  - Was anything unclear?
  - Did anything work particularly well?
  - What could be better?
- Leave at the back on your way out