Overview

- Finishing up frequency-dependent selection
- Selection vs. drift
- Tests for selection:
  - dN/dS
  - HKA

One-minute responses

- It would be helpful if you coud walk through the gene frequency graphs, labeling the most common or increasing frequency genotypes represented by each trend

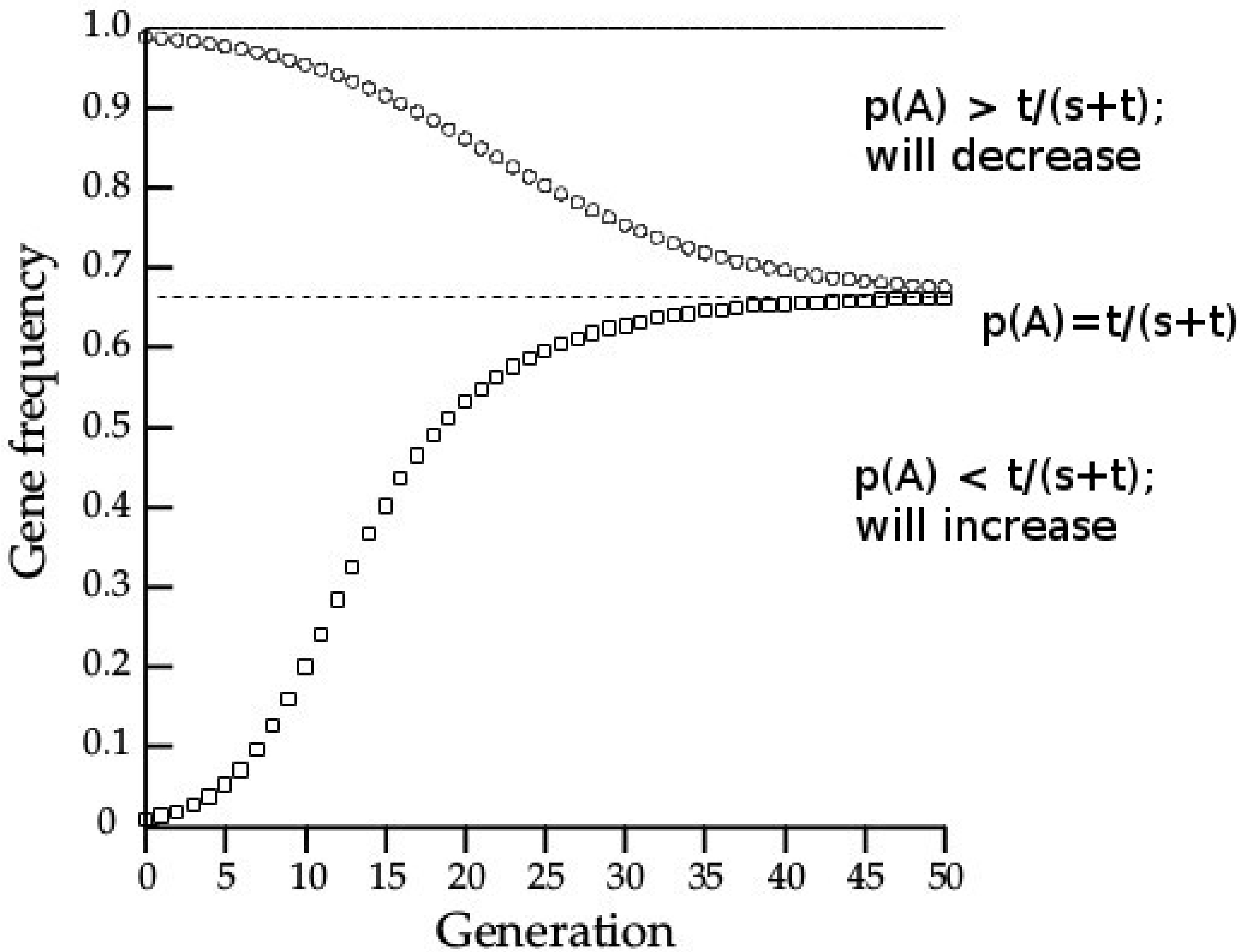


Figure 2.5: Convergence of initial gene frequencies from $p_A = 0.99$ and $p_a = 0.01$ to equilibrium when the fitnesses of *AA*, *Aa*, and *aa* are 0.85 : 1 : 0.70

.

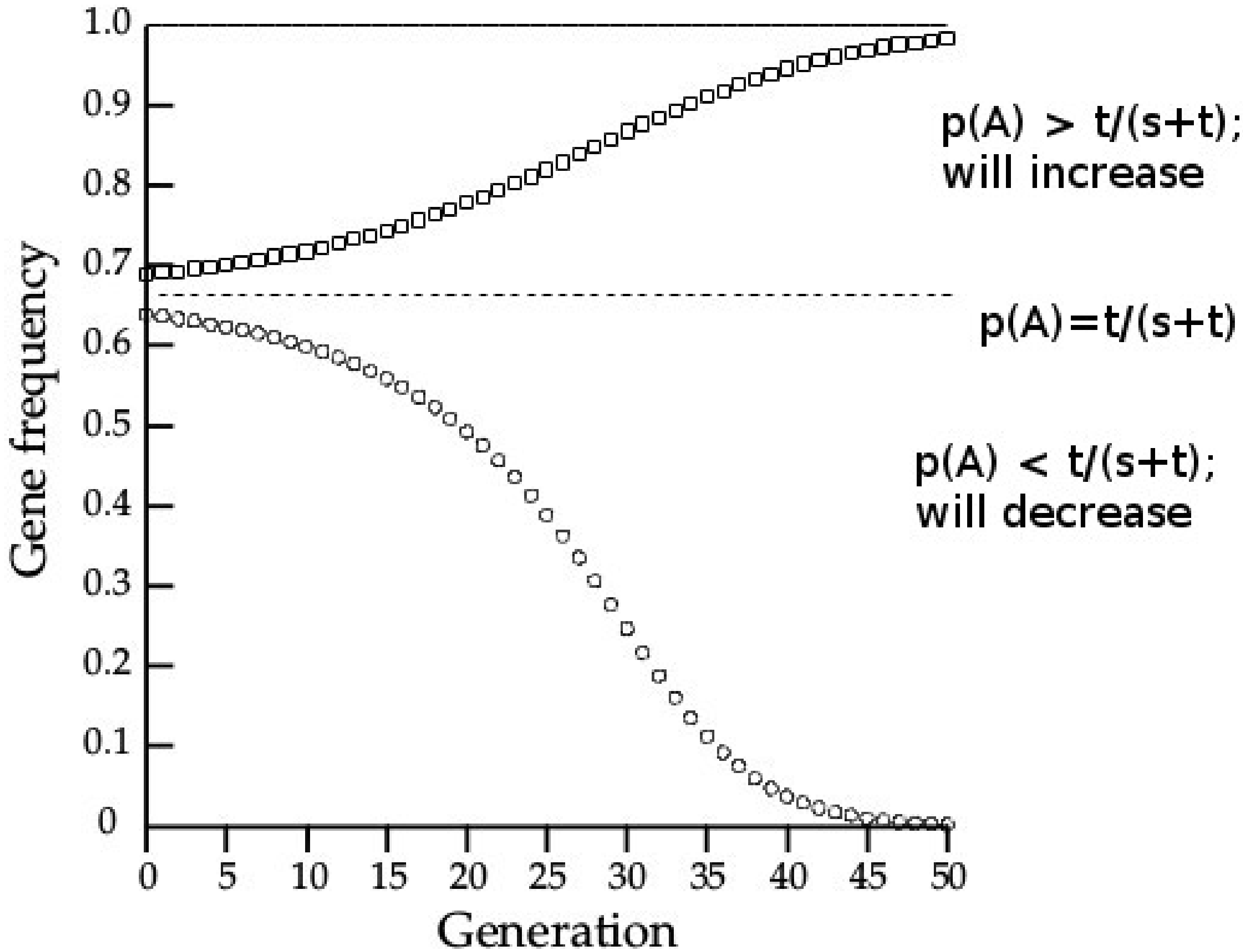


Figure 2.6: Gene frequencies in successive generations when fitnesses of *AA*, *Aa*, and *aa* are underdominant (1.15 : 1 : 1.3) and the initial gene frequency is 0.65 (circles) or 0.68 (squares).

Quick demo of PopG goes here

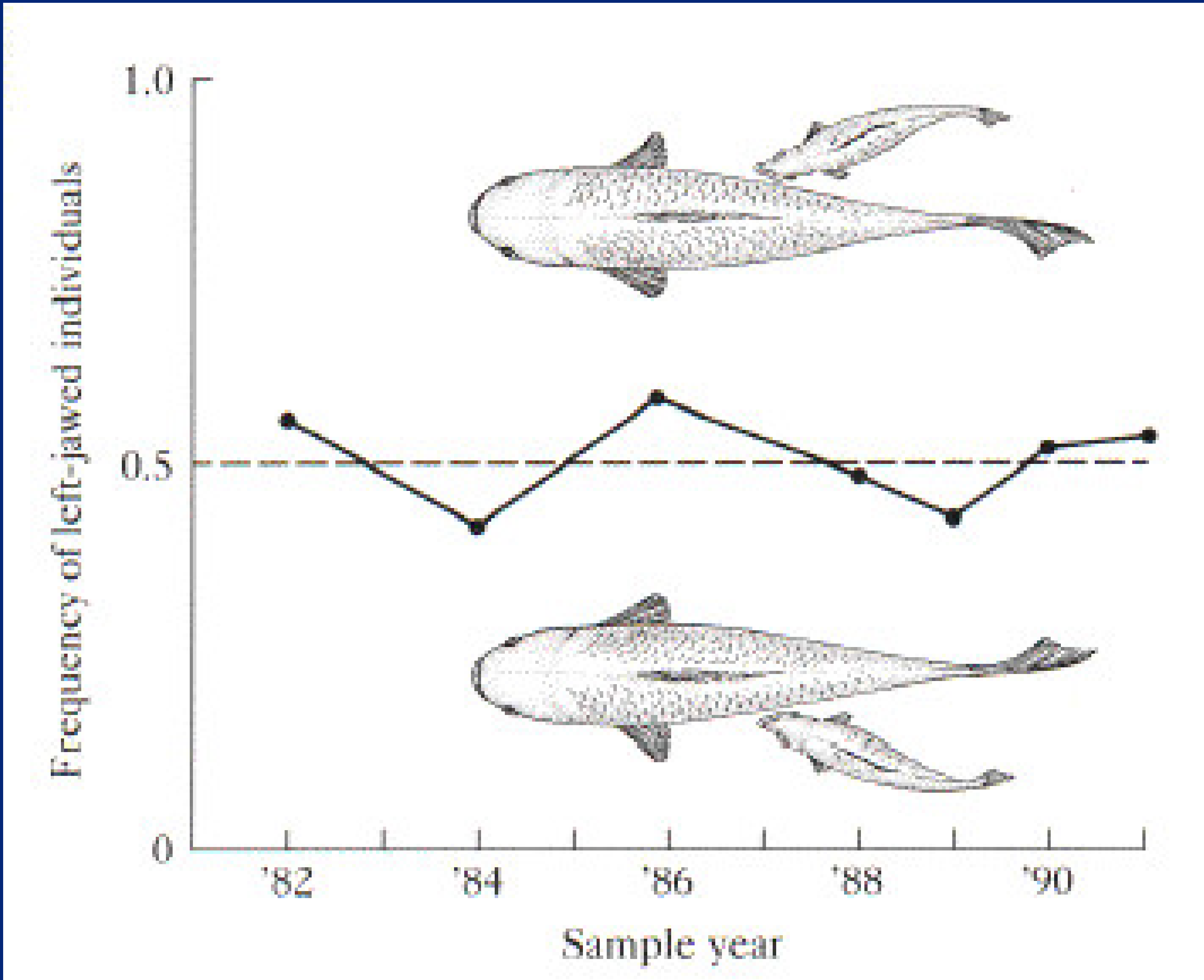
PopG: http://evolution.gs.washington.edu/popgen/popg.html

Frequency dependent selection

- The fitness of a phenotype depends on its rarity
- Rare alleles favored:
  - Rare type has less competition for resources
  - Rare type suffers less from parasites, pathogens, or predators
  - Rare type is sexually attractive
- Common alleles favored:
  - Rare type is sexually unattractive
  - Rare type catches predator's attention

Frequency dependent selection

- Rare-favored often resembles overdominance
  - Multiple alleles maintained in population
  - Alleles can be very old
- Common-favored often resembles underdominance
  - Rarer allele tends to be lost
  - Don't expect to see these within a single population
- The math may be the same as overdominance/underdominance or not, depending on *how* fitness depends on frequency


1.0
0.5
0
Frequency of left-jawed individuals
'82
'84
'86
'88
'90
Sample year

Selection varying with time

- Directional selection that switches back and forth might be able to maintain variability
  - If it switches too fast, it won't do anything
  - If it switches too slowly, alleles will be lost between switches
- This has been proposed as a reason for high variation in natural popuations, but is it really plausible?
- Possible examples:
  - Seasonal variation in micro-organisms
  - Host cell switching in HIV

When will selection overcome drift?

- Often stated rule of thumb: when $4N_e s >> 1$
- What is the s in that formula??
- Does this apply to a very rare allele?

Rare allele is good in the heterozygote

- Call the advantage of the allele in the heterozygote s (fitness of heterozygote is $1 + s$)
- The homozygote is so rare initially that its fitness doesn't matter
- Approximate chance to survive the early period is $2s$
- Alleles die early or not at all
- *Question: does population size matter? If so, how?*

Rare allele is good only in the homozygote

- Approximate formula predicts probability 0
- Obviously real answer must be greater than $\frac{1}{2N}$ (the answer for a neutral allele)
- Algebraic solution not available, though you can simulate it

The one diploid case you can solve

- Solving these exactly involves considering the chance that 17 gene copies in this generation give rise to 14 in the next generation....
- A diffusion approximation which assumes that N is quite large and s is quite small is more tractable
- The feasible case is multiplicative fitness:

Genotype	AA	Aa	aa
Fitness	1	$1 + s$	$(1 + s)^2$

- This is tractable because each a contributes the same benefit whether it is in Aa or aa

The one diploid case you can solve

- Fixation probability for multiplicative fitnesses:

$$\frac{1 - e^{-4Nsp}}{1 - e^{-4Ns}}$$

- p is starting allele frequency of favored allele
- s is selection coefficient from previous slide
- When s approaches 0, this approaches p

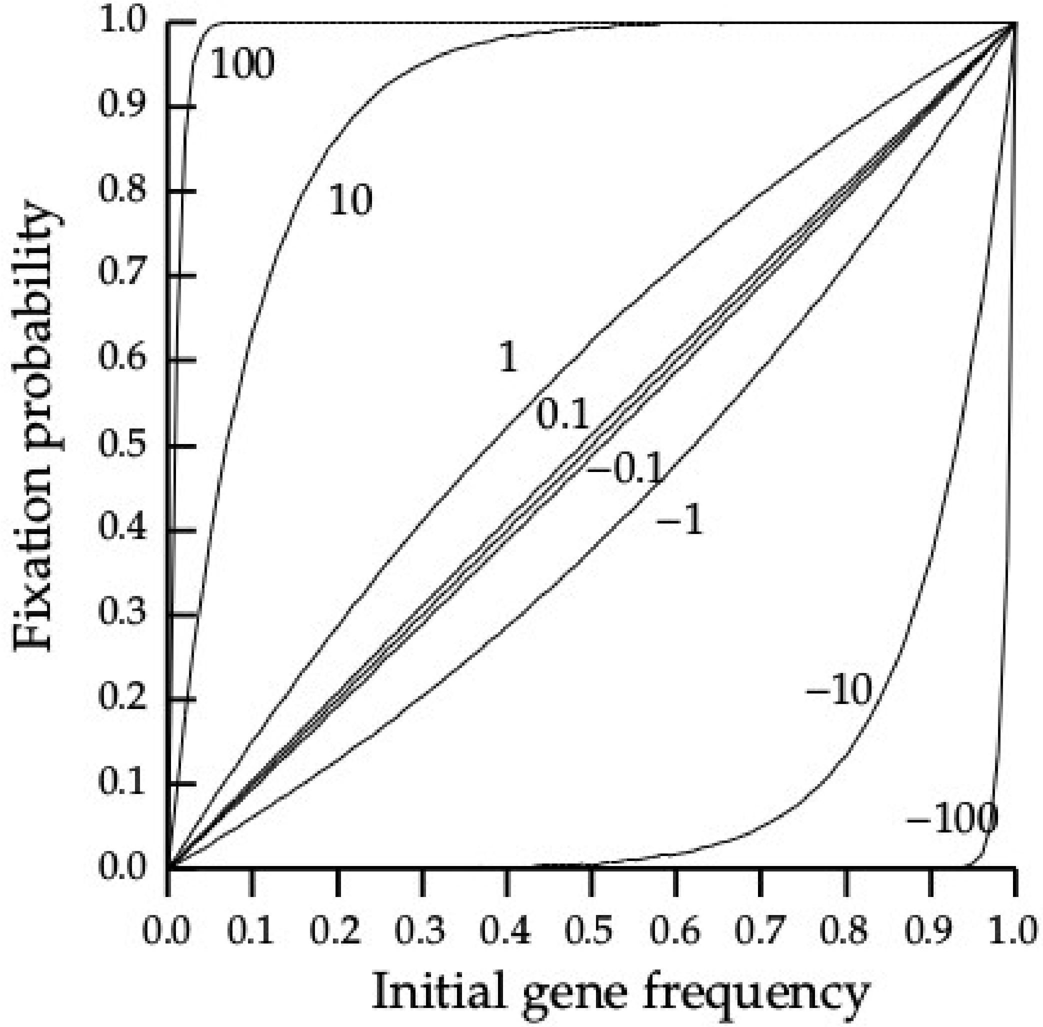


Figure 7.3: Probability of fixation of an allele with multiplicative fitnesses. Results from the diffusion approximation for various values of $4Ns$ and p are shown. The values of $4Ns$ are shown next to the nine curves, except for the diagonal, which has $4Ns = 0$.

Very rough results

- When $|4Ns| >> 1$ selection definitely makes a big difference
- When $|4Ns| << 1$ selection is ineffectual
- There is a wide murky range in the middle, and if the allele frequency is very extreme, selection has trouble even in a big population
- These results are for multiplicative:
  - Rare dominant close to multiplicative
  - Rare recessive much more influenced by drift

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)?

Testing for selection: dN/dS

- Mutations in protein coding sequence can be:
  - Nonsynomymous (coding): amino acid change
  - Synonymous (silent): no amino acid change
- Synonymous more likely to be neutral or nearly neutral

The Standard Genetic Code

First Position (5' end)	Second Position				Third Position (3' end)
	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met Start	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

dN/dS

Concept:

- Count positions that *could* have a silent or coding change
- What proportion actually did?
- $dN =$ nonsynonymous mutations per nonsynonymous site
- $dS =$ synonymous mutations per synonymous site
- dN/dS is a measure of selection:
  - ≈ 1 for no selection
  - < 1 for purifying selection
  - > 1 for diversifying or ongoing directional selection

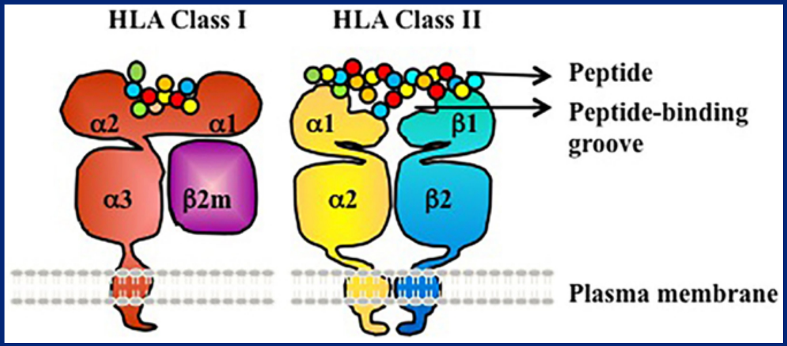
dN/dS

- Also known as:
  - ω (omega)
  - D_n/D_s
  - k_N/k_S
  - Nei's test of selection
  - Nei's test of neutrality
- Standard software for this is PAML package

dN/dS varying across a gene

In HLA loci:

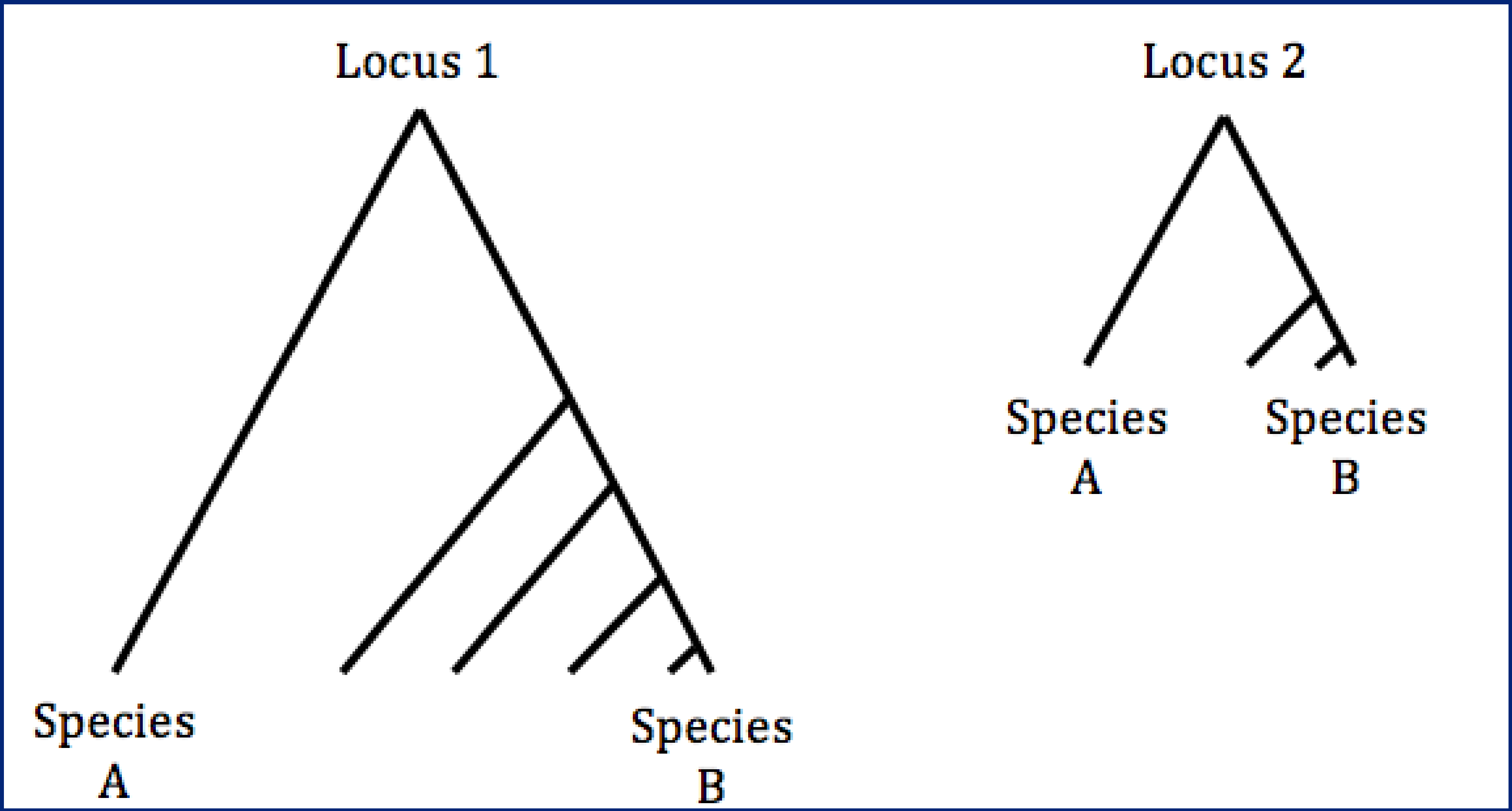
- Antigen-binding region, $\omega \approx 3$
- Elsewhere in the gene, $\omega << 1$



Limitations of dN/dS

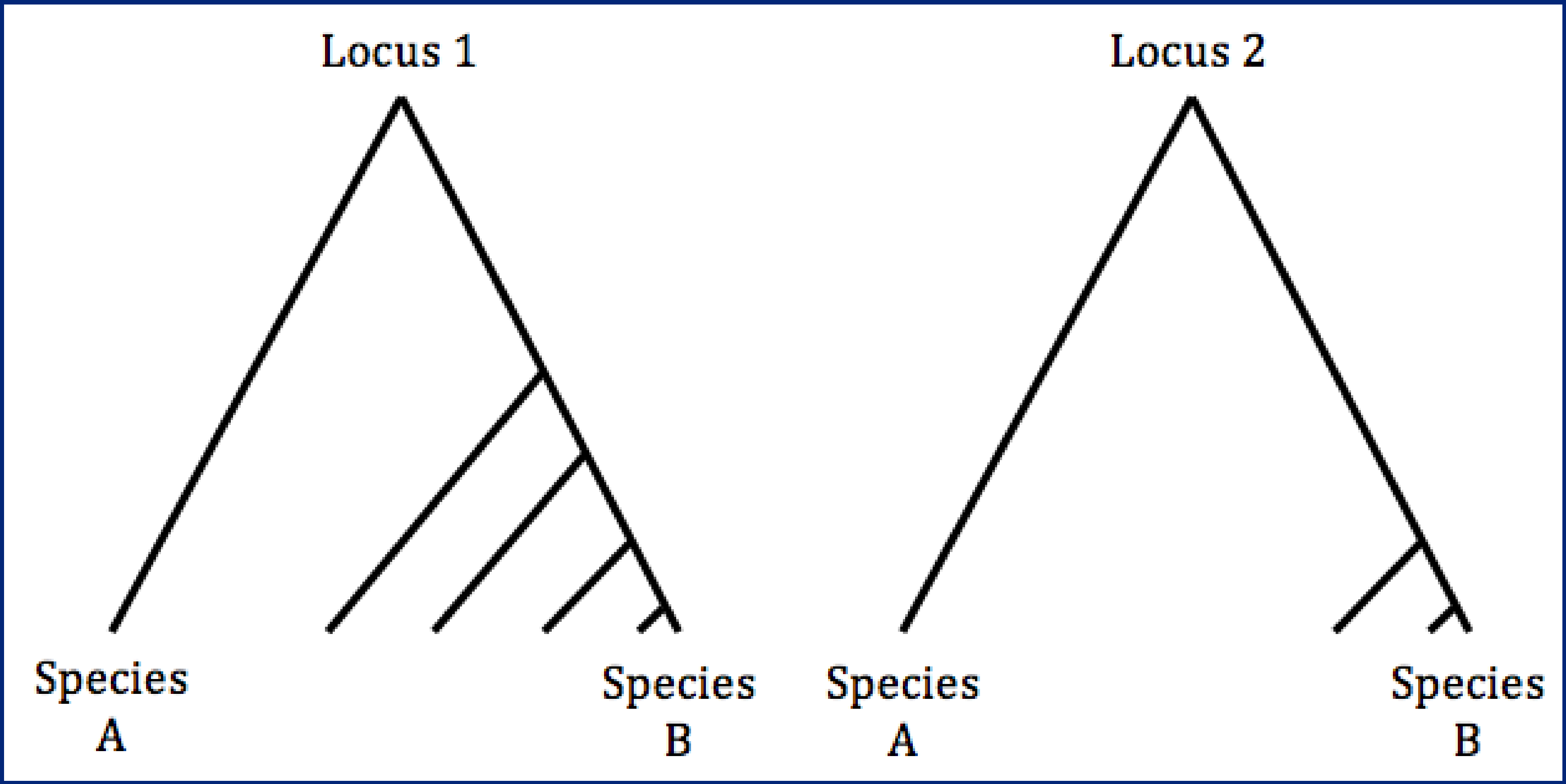
- Coding sequences only: not promoters, enhancers, non-coding RNA loci, etc.
- Needs lots of sequences
- Needs lots of selected sites
- Different selection in different regions of same gene can confuse test
- Assumes silent substitutions are neutral:
  - Codon bias?
  - DNA binding proteins?
  - Splice sites?

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 differences or 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

Friday

- McDonald-Kreitman test
- Tajima's D
- How much of the genome is functional?
  - ENCODE project
  - "Genetic load"

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.