

1. For the homologous sequences #1 and #2 shown below, give arguments either FOR or AGAINST them being under strong functional constraint. (Tip: On a test, you would be given a genetic code table.)

#1      5' ... C G T T A C C G T T C G A G A A C T T G T G C A ... 3'  
 #2      5' ... C G A T A C C A T T T G A T A A T C T G T G T A ... 3'

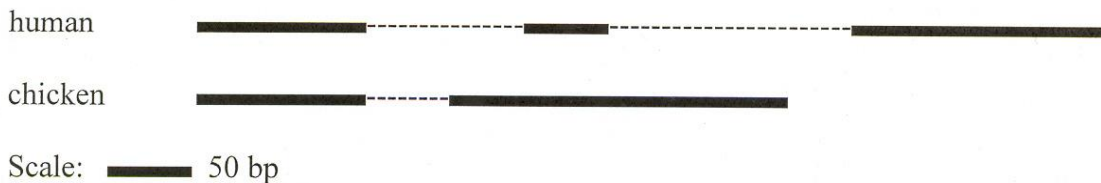
2. Explain why you AGREE or DISAGREE with each of the following statements.

- i) Four-fold degenerate sites evolve more rapidly than other nucleotide sites.
- ii) Indels are expected to occur less frequently than point mutations.
- iii) In a DNA sequence, the 4 bp motif GACT is expected to occur **twice** as frequently as the 8 bp stretch AAGAGGAG.
- iv) The “stick insect” wing study (discussed in class) provides a good example of Kimura’s neutral theory.

3. For protein Y which is 200 amino acids long, approximately how many non-synonymous sites are expected in its mRNA? (This question was given in class when we discussed Table 1.5 at the end of Topic 1).

4. For growth hormone gene G, the rates of non-synonymous and synonymous substitution are 1.34 and 3.79 substitutions per site per 10<sup>9</sup> years, respectively (Table 4.1 in text). If protein G contains 189 amino acids, what % amino acid identity is expected between the mouse and human proteins (divergence time about 80 million years ago)? Show your calculations and reasoning.

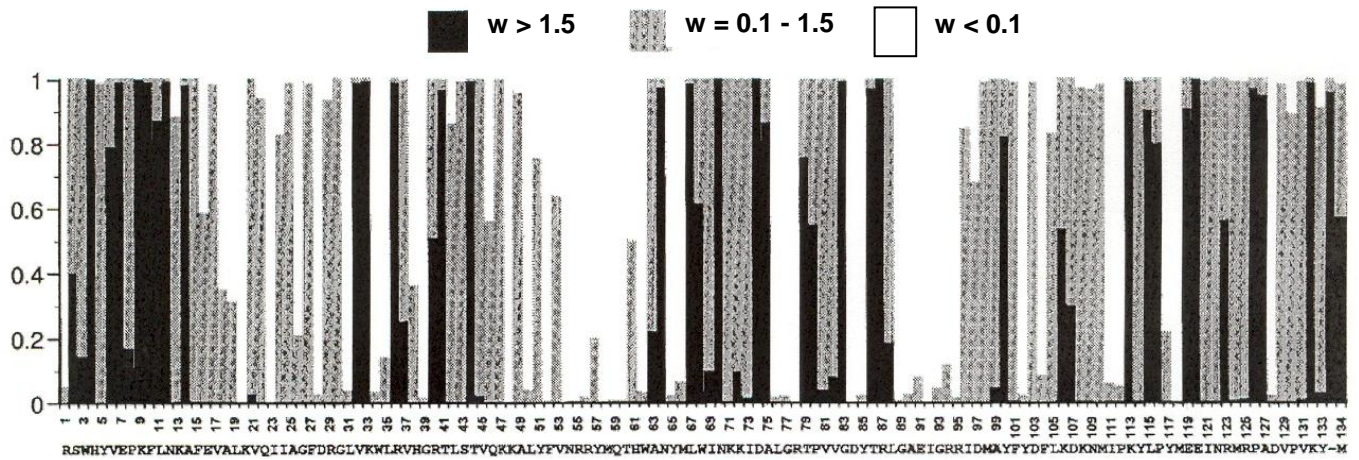
5. Suppose that gene X encodes an essential protein with ligand L binding domains in its carboxy terminal region. The gene X structure in human and chicken are shown below (black lines = exons, dotted lines = introns). Assume that the human protein X has two L domains (25 amino acids each) whereas the chicken protein has just one.



Use the dot matrix method to show the profiles expected when these human and chicken sequences are compared at the DNA level (box 1) and at the protein level (box 2). Be sure to label the axes clearly. Below, give your reasoning (based on molecular evolutionary principles) for features shown.

6. The coding sequence for enzyme E was compared from 25 species of the genus *Geewhiz*. In the figure below, the symbol “w” represents the **ratio of non-synonymous to synonymous substitution rates**, and the protein E sequence (134 amino acids in one-letter code) from

*G.rufescens* is shown at the bottom. (Adapted from Yang et al. *Mol Biol Evol.* 17: 1446, 2000). How do you interpret this information? In your answer, present a model consistent with these data.



7. The dot matrix plot below shows a comparison of gene A (and its flanking sequences) from organisms X and Y. How do you interpret these data? In your answer, include a schematic (drawn approximately to scale) of the gene A regions from these two organisms, and clearly label important features.

