NAME:_____

BIO/CMPSC 300
Chapter 10 Genome Annotation II
Fall 2017

1. Why is codon usage a poor predictor of the exact point where an exon and intron are joined? Why is the 5′ splice site consensus (GT) also a poor predictor?

2. Suppose you use the sliding-window algorithm described to analyze codon bias. At several points in a DNA sequence, you notice that you see a high score in your first window and a low score in your second window. But, when you slide the window by one or two nucleotides, you get low scores in both windows. How would you explain this pattern? How might you want to account for it in deciding where your exon-intron boundaries are? (Hint: think reading frames!)

3. Why are CpG islands considered valuable for gene prediction? Where would you expect to find one with respect to a eukaryotic transcription unit? What other elements might you look for in connection with the CpG island to increase the strength of a gene prediction?

4. How could alignment of a sequence with orthologous sequences (using BLAST) contribute to the prediction of exons and introns? How could expression data (e.g., mRNA sequences) contribute?