1. (5 points) Suppose that you encounter an alternative genetic code which, like ours, uses four nucleotide bases, encodes amino acids with triplets, and has three stop codons. In the alternative code, however, every codon which is not a stop codon encodes a different amino acid. Estimate $K_A$ and $K_S$ under the alternative code for the alignment below. (In case you need it, the Jukes-Cantor correction formula is $K = -\frac{3}{4}\log_e(1 - \frac{4}{3}p)$. An exact expression or numerical approximation are both acceptable answers.)

<table>
<thead>
<tr>
<th>TGG</th>
<th>TAT</th>
<th>CAT</th>
<th>CAC</th>
<th>CCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGG</td>
<td>TAC</td>
<td>CAG</td>
<td>CAT</td>
<td>CAA</td>
</tr>
</tbody>
</table>

A nucleotide change is synonymous or nonsynonymous only with reference to the genetic code. In the extreme case of a genetic code in which all changes are nonsynonymous, as here, all changes and all sites must be nonsynonymous. We consider the definition of $K_A$,

$$K_A = -\frac{3}{4}\log_e(1 - \frac{4}{3} \frac{\text{# nonsynonymous changes}}{\text{# nonsynonymous sites}})$$

We observe 6 nonsynonymous changes and 15 nonsynonymous sites, giving:

$$K_A = -\frac{3}{4}\log_e(1 - \frac{4}{3} \frac{6}{15}) \approx 1.02.$$ 

Now we consider $K_S$:

$$K_S = -\frac{3}{4}\log_e(1 - \frac{4}{3} \frac{\text{# synonymous changes}}{\text{# synonymous sites}})$$

Because all sites and changes are nonsynonymous there are, indubitably, no synonymous changes, nor synonymous sites. Thus $K_S$ is 0 or undefined, depending on your worldview.

2. (5 points) Suppose that you observe a length of double-stranded DNA, whose base composition is unusual, in that A/T pairs far outnumber G/C pairs. Give a plausible explanation for the A/T enrichment, assuming that the DNA does not code for a protein.

Such a bias, broadly speaking, may arise from either neutral or selective causes. A common neutral cause of A/T bias is heterogeneity in substitution rates, such that the net rate of substitution into A/T pairs exceeds that into G/C pairs. As for selective causes: DNA may well be under selection even if it does not code for a protein. Perhaps this locus codes for an rRNA or a regulatory element whose function depends, for some reason, on having more A/T pairs than G/C.