Calculate the probability that two amino acids a and b will appear aligned purely by random chance, eab.

* The chance of seeing two of aa aligned, assuming all differences in the sequences were randomly generated, is eaa=pa2.
* The chance of seeing a and b aligned is eab=2papb

The 2 is because there are two ways of selecting two different items by chance (ab or ba). The matrix e is our random model.

If we now compare these two probabilities, from the evolutionary model and the random model, we can see whether it is more likely to see two specific amino acids aligned by chance or by an evolutionary process.

If we set the score to be qab/eab, this will give us what we want. Greater than 1 and the alignment is likely from evolution, between 0 and 1 and it is likely by random chance. If we take the log of this score, then we end up with a negative score for chance and a positive score for the evolutionary model. This is called the log-odds score, sab = log(qab/eab), and this is what we will be calculating

**Question 1A: Provide the answer to your calculation of eaa**

0.0196

**Question 1B: Provide the answer to your calculation of eab**

0.0616

**Question 1C: Provide the answer to your calculation of sab**

-0.60200451209

**Question 1D: Provide an explanation for how the calculation of substitution scores works**

**Question 2A: Enter your Python code for calculating seqPairs. This should be submitted to Coder Quiz in the format seqPairs = MY\_ANSWER**

seqPairs = math.factorial(numSeqs) / (2 \* math.factorial(numSeqs - 2))

**Question 2B: Enter your Python code for calculating aaPairs. This should be submitted to Coder Quiz in the format aaPairs = MY\_ANSWER**

aaPairs = columns \* seqPairs

**Question 2C: Enter your Python code for calculating eab where a == b . This should be submitted to Coder Quiz in the format eab = MY\_ANSWER**

eab = p[a] \* p[a]

**Question 2D: Enter your Python code for calculating eab where a!= b. This should be submitted to Coder Quiz in the format eab = MY\_ANSWER**

eab = 2 \* p[a] \* p[b]

**Question 3B: Submit the alignment if the gap penalty is set to -8 . To differentiate the two rows of the alignment separate them by a comma e.g. MADMAN,MAD-AM.**

THISLINE-, ISALIGNED

**Question 3C: Submit the alignment if the gap penalty is set to -4 . To differentiate the two rows of the alignment separate them by a comma e.g. MADMAN,MAD-AM.**

THIS-LI-NE-, --ISALIGNED

**Question 4A: How many cells would the S matrix (in the alignGlobal function) contain when aligning HQ659871.1 and JX416721.1?**

2345868

**Question 4B: If you leave the DNA substitution matrix untouched, what is a biologically sensible gap penalty?**

-4

**Question 4C: What steps did you take to determine 4B?**

Good alignment is achieved by choosing a good combination of gap penalty and substitution matrix, and now the substitution matrix is untouched. The biological sensible gap penalty are the gap penalties makes the good alignment.

if the gap penalty is very negative, more so than any score in the substitution matrix, then it will never be good score-wise to place gaps. As a result, the sequences will be lined up with no gaps, regardless of how bad the result is. Conversely, a too-small gap penalty may lead to an alignment where there is little, or no identity overlap between the sequences.

**Question 4D: Given the original DNA substitution matrix and a gap penalty of -5, at what position is the first ATG codon in reading frame +2 of seqB (JX416721.1)?**

92