

**Topic:** Python control structures  
**Techniques:** Use of for loops and if statements  
**Reading:** lecture notes  
**Collaboration Policy:** The lab should be completed **working in pairs**  
**Submission Deadline:** Monday, 4 February 2019

Students: \_\_\_\_\_

To receive full credit for this lab, you must provide Python code to solve at least four of the following five questions.

**1 What percentage of codons across all primary reading frames are ATG?**

If this were completely random, we'd expect  $1/64 = 1.5625\%$  of the triples.

We observe 1.196% for guinea pig and 0.978% for human.

**2 If two consecutive nucleotides match each other, how often is the next nucleotide that same nucleotide?**

If nucleotides were completely random, we'd expect 25%;

we observe 28.392% in guinea pig and 30.620% in human.

**3 How many times does a motif of the form CC?AT occur within the sequence? (where '?' could be anything)**

For guinea pig, 111 times; for humans, 132 times.

**4 When the motif CC?AT does occur, what percentage of the time is the middle nucleotide an A? (a so-called “cat” box CCAAT)**

For guinea pig, 27.027%; for humans, 21.212%.

**5 The pattern CCAAT is known as a “cat” box. What are the relative percentage of bases immediately following the pattern CCAA in the dna?**

We find the following for guinea pig:

A: 28.431% C: 31.373% G: 10.784% T: 29.412%

and the following for human:

A: 39.416% C: 29.197% G: 10.949% T: 20.438%