

Homework 1 CSCI 451

Nicholas Rust

due: 12 September 2019

1.
 - Yes you can visit a gene bank and check a number of protein sequences.
 - The MET is the first amino acid.
2. There are usually 2 copies of nuclear DNA in a eukariotic cell. There are between 100 and 1000 copies of mitochondrial DNA in a eukariotic cell, about 500 per human cell.
3. Genes are currently supposed to be somewhat randomly distributed throughout the genome, however, there are theories that some groups are clumped together rather than completely randomly distributed.
4. It is mentioned that there is a codon bias for *D. melanogaster* where the more commonly expressed genes use the more readily available codons in the cell. This does not appear to be true for most organisms.
5. ATG is the start translation sequence, so ACTTGTC is the 5' UTR. TAG is the end sequence used here, meaning the 3' UTR is GTCATG.

```
$ python Zalg.py
Testing Z algorithm output on string 'aabaabcaabaabaabc'
[0, 1, 0, 3, 1, 0, 0, 6, 1, 0, 7, 1, 0, 3, 1, 0, 0]

Testing Z algorithm exact pattern matching on string aabaabca$aabaabcaabaabcaa
ccbbcabbbcbabcbabcbabcccbabbbbcacaaacbbbaacbbcbcaacbaabaabcbabcbabccabb
bcabcbbbbcccccaabaabcbabcbabcbabccbbbabaaa

There is an exact match at index 11
There is an exact match at index 18
There is an exact match at index 81
There is an exact match at index 123
```

Figure 1: Z Algorithm Output