Homework Exercise 1

1)

What data types would you use to represent the following: (explain your choices)

  a) A DNA sequence
  b) Names/symbols of genes associated with a disease
  c) RNA sequences of all the transcripts expressed in a tissue
  d) Frequencies of amino acids in a given protein sequence
  e) Unique amino-acids appearing in a short peptide

2)

A)

Two strains of bacteria are grown in a lab on a petri dish. Strain A doubles (X2) its population every day, while strain B triples it (X3). The experiment starts with a colony of 10,000 bacteria of type A and 100 of type B on the same petri dish. When the total population (of both strains together) exceeds 10^8 bacteria, growth is reduced to X1.5 for strain A and X1.8 for strain B. When it exceeds 10^{10}, growth stops. What will be the number of bacteria during each day of the experiment? How many days will it take until growth stops? What will be the proportion between the two strains at the end of the experiment?

B)

In a second experiment, a bacteriophage is introduced into the dish. When the total population exceeds 10^6 bacteria, it starts spreading and kills 40% of the cells every day (of both strains). Every day after the bacteriophage becomes active, 10% of type-A bacteria will develop immunity to the bacteriophage and retain normal growth rate. Strain B has a lower mutation rate, so only 1% of its cells will develop immunity every day. Run the simulation again and determine the fraction of the 4 populations at the end of the experiment (type A/B with/without immunity).

3)

A)

Find the reverse complement of the following DNA sequence and translate the output DNA sequence into protein sequence:

```
CTACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACA
CTGAAGGAAGTGACACAATCACCCTCCCATGCAGAATAAAACAAATTATAAACATGTGGCAGAAAGTAGGAAAA
GCAATGTATGCCCCTCCCATCAGTGGACAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGA
```

```
TGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTG
AATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTG
CAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGG
CGCAGCCTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGC
TGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCAT
```

B)

What is the frequency of each codon in the obtained sequence?

C)

[**Optional**]

Does the distribution of codons seem uniform? Are there codons that appear dramatically more (or less) than expected (given their amino-acid frequency in that sequence)?

Bonus Questions

1)

Explain the *is* statement and *id* built-in function. Use it to explain what variables in Python are.

Provide an explanation for the following code:

```
a = [1, 2, 3]
b = [1, 2, 3]
c = None

print(a == b)
print(a is b)
print(3 == 3.0)
print(3 is 3.0)
print(3 is 3)
print(c is None)

x = 10 * [[]]
x[2].append(5)
print(x)
```