

# HW1

---

Make sure that running the commands in google colab returns the right answer(s) in the correct data types. Your answers should be as accurate as possible. Make sure that everything is commented and documented, or points will be deducted. You should submit your homework as hw01\_<netid>.pdf on Gradescope

## Questions

---

### 1. Sequence Analysis

Write code in python that would execute the tasks assigned below. Remember, each question should have an associated code for answering the questions on a generic (in-frame) nucleic acid sequence represented as letters A,G,C, or T. We do not want to see if you can answer the questions biologically, we want to see if you can outline the process of answering these questions computationally. (Note: use the standard codon library, assume the first 3 nucleotides begin the translated region/open reading frame) sequence =

```
"ATGTACAACATAGTGATGGAGACGGAGCCATTGAAGCCGCCGGGCCCCGCAGCAA  
ACCTTCGGGGGGCGGCGGCGGCAACTCCACCGCGGCATTCGTUAG"
```

- A. Calculate number of nucleotides in the given sequence.**
- B. Calculate number of codon in the given sequence (if it is all translated)?**
- C. Check whether the given sequence starts with 'ATG'.**
- D. Check whether the sequence ends inframe. (the length of the sequence is divisible by 3)**
- E. What percentage of sequence is GC ?**
- F. Print only the first and last codon.**
- G. Using the dictionary provided below, what amino acid does the second codon represent?**

```
geneticcode = {
    'ATA': 'I', 'ATC': 'I', 'ATT': 'I', 'ATG': 'M',
    'ACA': 'T', 'ACC': 'T', 'ACG': 'T', 'ACT': 'T',
    'AAC': 'N', 'AAT': 'N', 'AAA': 'K', 'AAG': 'K',
    'AGC': 'S', 'AGT': 'S', 'AGA': 'R', 'AGG': 'R',
    'CTA': 'L', 'CTC': 'L', 'CTG': 'L', 'CTT': 'L',
    'CCA': 'P', 'CCC': 'P', 'CCG': 'P', 'CCT': 'P',
    'CAC': 'H', 'CAT': 'H', 'CAA': 'Q', 'CAG': 'Q',
    'CGA': 'R', 'CGC': 'R', 'CGG': 'R', 'CGT': 'R',
    'GTA': 'V', 'GTC': 'V', 'GTG': 'V', 'GTT': 'V',
    'GCA': 'A', 'GCC': 'A', 'GCG': 'A', 'GCT': 'A',
    'GAC': 'D', 'GAT': 'D', 'GAA': 'E', 'GAG': 'E',
    'GGA': 'G', 'GGC': 'G', 'GGG': 'G', 'GGT': 'G',
    'TCA': 'S', 'TCC': 'S', 'TCG': 'S', 'TCT': 'S',
    'TTC': 'F', 'TTT': 'F', 'TTA': 'L', 'TTG': 'L',
    'TAC': 'Y', 'TAT': 'Y', 'TAA': '_', 'TAG': '_',
    'TGC': 'C', 'TGT': 'C', 'TGA': '_', 'TGG': 'W'}
```

## 2. Fahrenheit to Celsius conversion

**A. Convert the temperature 23 degrees in Fahrenheit to Celsius.**

## 3. Gene Sets

Below are two lists of genes in Rice. One is a list of genes that responds to *Nitrogen* and the other is *Carbon*.

Nitrogen = Os1, Os2, Os3, Os5

Carbon = Os1, Os3, Os4, Os5

**A. Create 2 sets of genes, one for Nitrogen and one for Carbon.**

**B. Using python functions, determining which genes are present in both sets.**

**C. Using python functions, determining which gene(s) are present only in Nitrogen Set.**

**D. Create a new set called Allgenes that contains genes from both sets**