

Assignment 2

- Write a script which takes as input the below sequence and displays the frequency of each codon and di-nucleotide in it. Also provide a protein translation of this sequence.

>gi|143612989|gb|EP656924.1| marine metagenome

```
AATTCGAAGACATAATATTTTTCCATTACCTCTAAAATTATGTTTTAAGAAAGGCTTTATATAGCAAAGAATCACCGGGGTATAATTTCTCTTTCCGTTTTCAAAA
TTAATTGCTAATTCCGAATCACCTATATTATAAATATACTGATGTAAGCCAATATTTTGATCATAGTCTTCATGATCAGATTTCAGAAAAAAGTATATTTCTAAAC
TTCTTGCTAAGGAAGAGTAGAAGAGTTAGCAAGTTCAACAACCTGGTATGCATTATTCGCTAGCTTCCATTGGTTACCTTCATCGTGCTTTTGAATAATAGTTG
CAGGCTCTATTAAGTCATTAGATAAAATATCTCTTAGATTATATTCATCGATGCTGATATGGTTTTAAGTTCATCAAGCGTTGATGAGGAACCATTTTCAATGT
CATTTAATTTTTTTTCAGAAATTTAGATCTTCTTGAAAGTTCTTCTAAAGATATACTGAAGCTGTTTCTAAAATATTTAATTAAGAACCAATACCTTTTTCATAA
CTTGTAAGTCTAATGCAAACCTCACTTGCAGAAATTGGGTTAGAGATAGCTGCCAATTCTTGCTGAGCATCGCCGCTAAGACTGCTGCCAAATGTAAGAGCAA
GAATAAGCCCAGGTTTTTTTGAGCCCCGCCCTAGTGGTAAAAGAATGAGGTACAAATGGAGTTATGTAGCACGAATCCCCTGTATTTCATAATGCCAACACACTT
TTTTCCATTTTCATCAACATAATAGTAATTTACATCACCAATAAAATATGTGAATTGATGCATAAAATGCCCATTTCCACTGTACAGATTTATTATTTGGATCA
TTATCATCAACAATACATAATTCCTCGATCCATTCTGGTCTAAATGTAGTAGATGAAGTTATGACAGTGTACAGATAATCATAATAGGGTTTTCCAGCCCTTTCC
ATTATTCTAGAGCTTTTATTTGAATCGGATGACCTCATAATTTTTATCCCGCTTGGACAATCATCTTTTATAACAAAGAAATCTCTCTCATTAAATGGGGCACTTAT
CACATGCAGTTTTAATAAAGTCTGAGGATATTGGTTTTGTTCTGATAAATAATCTTCTAAATCTTTAATATCAAAACCAAGTTCGTTTGCTGCATCTTCTGGTGC
CCTTTTAAATCATTTAATAAACCACGCAAATTATGAGCAGATTTTTTTAATTCAGCATCATTGGATTTCATTTAATAATTATTACATAAATCAATACAATTTTGAA
TAAAAATTAAACAAACAAAAGTAATTTATTGAGCTAGATAGCCACCATCAACAGGCAAACATGCTCCAGTTATAAAAGATGCTTCATCAGATGCTAAAAATA
AAGCAGCATATGCT
```

- Now download the genomic sequence of *Mycoplasma genitalium* G37 in *.fna format from the ftp site of NCBI and identify the number of open reading frames encoded in the genome. The output of the script should be all the ORFs present in a simple fasta format. (ORFs are defined as those which have a start codon 'ATG' and any of the three stop codons 'UAA', 'UAG' and 'UGA') (ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria/Mycoplasma_genitalium_uid97/)