



GENOME DEFINITIONS*

Used for Genome Analysis Assignment.

* Short definition of terms. Please note that all of these have a lot more nuanced detail about them.

**I'm using the terms gene and protein interchangeably here. *They are not the same*, but for the analysis level we are doing, I don't want you to worry about the differences.

Nucleotide: Building blocks of DNA (and RNA) molecules. For DNA, there are 4: A, T, C, and G.

Codon: A group of 3 nucleotides. Together, those 3 letters code for a specific amino acid.

Amino Acid: The building blocks of a protein and is usually represented as a single letter. An amino acid can be created via several different codons. A protein is represented as a string of amino acids.


Start Codon: Not every codon in a DNA sequence is part of a gene. The specific codon that begins the gene is ATG.

Stop Codon: signifies the end of the gene. There are three stop codons: TAG, TAA, and TGA.

Gene/Protein:** The series of nucleotides between the start and stop codon. These strings of nucleotides->codons->amino acids are what we aim to determine within the complete DNA sequence.

Genome: The entire DNA sequence of an organism.

Codon Bias: As an amino acid can be created via multiple different codons, this is the analysis to determine which of the codons an organism seems to use more frequently in their DNA structure to create a specific amino acid.



Reading Frame: there are three ways to determine codons in a DNA sequence. Reading Frame 1, 2, or 3 specifies which nucleotide the first codon starts with. In this instance:

TAAGCTATGGC

Reading Frame 1 would have codons TAA, GCT, ATG

Reading Frame 2 would have codons AAG, CTA, TGG

Reading Frame 3 would have codons AGC, TAT, GGC

ORF (Open Reading Frame): A portion of DNA that has a small number of nucleotides between the start and stop codon will not likely code for a functional protein. Therefore, the ORF is a threshold for where to consider a string of amino acids to be a viable protein. This is generally around 100 amino acids but can vary quite a bit based upon the organism.