Using Codon Alignment to Find Gene Mutations

指導教授：謝孫源
專題成員：魏喬浩
開發工具：Python

一、簡介：

DNA mutations can be caused by different factors, and some of them can be detected by analyzing the codons. Our DNA is made up by four different nucleotides – cytosine (C), guanine (G), adenine (A) and thymine (T); which can be translated into proteins, which possesses the same nucleotides, except for thymine, which is replaced by Uracil (U). A codon is simply a triplet (a combination of three) of these nucleotides, and depending on its combination, it will synthesize a different protein. This protein may determine whether or not a gene will be expressed and in the case a simple nucleotide is changed, all the sequence will be affected, producing a different protein or not producing any at all.
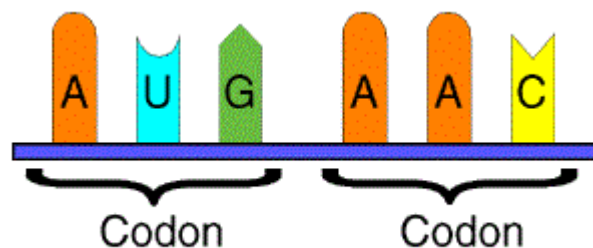
Mutations can occur either by substitution, insertion or deletion, among others. In this project, I will focus more on those resulting by substitution. As it name implies, a substitution mutation exchanges one DNA base (nucleotide) for another – like switching and T to a C. Those substitutions can cause one of the following:

1. Change a codon to a different one and this will cause a small change in the

protein produced, as it occurs to those with sickle cell anemia.

2. Have a silent mutation, meaning that the change was not a big problem and the same protein is produced.

3. Change the coding codon to a "stop" codon, resulting in an incomplete protein.

CTGG**A**G
CTGG**G**G

二、測試結果：

This program will take different sequences of DNA from different individuals and print their codon alignment and the estimate nonsynonymous substitutions per site (dN) and synonymous substitutions per site (dS). With the ratio of these two values we can get an indicator of selective pressure acting on a protein-coding genes. As these mutations will not always result in negative changes (such as silent mutations), a database should be used in order to compare results and determine whether the obtained changed codons represent or not a problem to the individual.

```
GTGATTTTCGTTGCCGGTCTGGGAGGCATT...ATC
GTGATTTTCGTTGCCGGTCTGGGAGGCATT...ATC
GTGATTTTCGTTGCCGGTCTGGGAGGCATT...ATC
GTGATTTTCGTTGCCGGTCTGGGAGGCATC...ATC
GTGATTTTCGTGGCCGGTCTGGGAGGCATT...ATC
GTGATTTTCGTGGCCGGTCTGGGAGGCATT...ATC
GTGATTTTCGTGGCCGGTCTGGGAGGCATT...ATC
GTGATTTTCGTGGCCGGTCTGGGAGGCATT...ATC
```