**Tutorial**

**Background**

Deoxyribonucleic acid (DNA) is a molecule composed from two polynucleotide chains that coil around each other to form a double helix. This structure contains the genetic code for all known organisms. Each polynucleotide is made up a number of monomeric units called nucleotides. Each one of these nucleotides is a combination of a sugar (deoxyribose), a phosphate group and one of four nitrogen-containing nucleobases (cytosine [C], guanine [G], adenine [A] or thymine [T]). It is the sequence of these four nucleobases that encodes genetic information.

Ribonucleic acid(RNA) strands are created using DNA strands as a template in a process called transcription, where DNA bases are exchanged for their corresponding bases except in the case of thymine (T), for which RNA substitutes uracil (U). Under the genetic code, these RNA strands specify the sequence of amino acids within proteins in a process called translation.

**DNA Match**

DNA profiling is a forensic technique used in criminal investigations. It compares the criminal suspects’ profiles to DNA evidence in an attempt to assess the likelihood of their involvement in a crime.

There has been a robbery at the bank, where thousands of pounds sterling has been stolen. Fortunately, the robbers must have been cut, as they left drops of blood on the floor. The forensics team have used the blood to isolate and mark up a DNA sample of the perpetrators. Following reviews of recent CCTV footage, police have arrested a couple of suspects they believe may have committed the crime.

Let’s use the DNA Analysis tool to test whether the suspects’ DNA matches with the DNA sample we have acquired!

1. First, let’s get our tool up and running. In a new terminal, making sure you are in the same directory as the DNA Analysis Tool, run the program with python using the command:

**$ python dna\_analysis.py**

1. Select the **DNA Match** option from the menu and enter the suspect DNA file:

**$ suspect1.txt**

1. A number of codons from the DNA sample can now be entered in a list format:

**$ GGG**

1. After entering a codon from our sample, the minimum number of matches in the suspect’s DNA needs to be selected. As we are only testing for a single codon, we’ll make the number of matches needed only two (**2**).
2. The tool will calculate the number of matches and print a message to the command line. This message can then be saved in an output file by entering the name of the file:

**$ suspect1.1c.2m.txt**

1. For the second suspect, we have a more accurate sample with a higher number of codons. Use the **DNA Match** option again, this time with the DNA file **suspect2.txt**, the sample codons **GGG, GTA, CAC**, a minimum of ten (**10**) matches and name the output file **suspect2.3c.10m.txt** so the investigation message can be printed there.
2. View both output files to see which suspect the police should continue investigating.

**DNA Replication**

DNA replication is the biological process of producing two identical replicas of DNA from a single original DNA molecule. Replication is important in all living organisms as it is essential for biological inheritance through the division of cells.

During the replication process, the complementary strands of a double helix DNA molecule are separated using the enzyme helicase. Each separate strand acting as a template for the production of its counterpart. The enzyme DNA polymerase synthesises the new strands by adding nucleotides that complement each strand. Generally there is a near perfect fidelity for DNA replication due to cellular proofreading and error-checking mechanisms.

We can use the DNA Replication option to complete a double helix DNA from a single strand input.

1. Select the DNA Replication option from the menu and enter the single DNA strand file we want to complete:

**$ test.txt**

1. The complementary strand is printed to the command line and written to an output file of our choosing:

**$ replication\_test.txt**

1. View the replication file to see how both the original and complementary strand fit together nicely.

**DNA Transcription**

Transcription is the process where a particular segment of DNA is transcribed into RNA (mRNA) by the enzyme RNA polymerase. RNA is formed as a chain of nucleotides in a single strand folded onto itself, rather than the paired double strand of DNA. mRNA in particular conveys genetic information that directs the synthesis of specific proteins. Upon binding and separate of DNA strands, RNA polymerase adds RNA nucleotides which are complementary to the DNA strand before breaking its hydrogen bond and letting the newly synthesized strand free. This mRNA strand serves as a template for protein synthesis via translation.

We can use the DNA Transcription tool to convert a DNA strand into its complementary RNA strand.

1. Select the DNA Transcription option from the menu and enter the DNA file we want to transcribe to RNA:

**$ test.txt**

1. The RNA strand is printed to the command line and written to an output file of our choosing:

**$ transcribed\_rna.txt**

1. View the transcribed file to see that the base Uracil (U) is used in RNA instead of Thymine (T), as in DNA.

**mRNA Translation**

Translation is the process in which ribosomes synthesise proteins using an RNA strand generated from the transcription of DNA. Messenger RNA is decoded in the ribosome to produce a specific amino acid chain (polypeptide). The polypeptide folds into the active protein and performs a function in the cell. The translation process occurs across 3 main phases. First, the ribosome assembles around the target mRNA and the first transfer RNA (tRNA) is attached at the start codon. Each complementary tRNA anticodon sequences to an mRNA codon. The tRNAs carry specific amino acids that are added to the polypeptide chain as the mRNA passes through the ribosome. When the peptidyl tRNA encounters a stop codon, the ribosome folds the polypeptide into its final structure.

We can use the mRNA Translation tool to “synthesize” a polypeptide of amino acids from an mRNA strand of bases. The ribosome will search the mRNA sequence for a start codon (AUG) to begin translating each subsequent codon into an amino acid. This process continues until one of three stop codons are encountered (UAA, UAG, UGA).

1. Select the mRNA Translation option from the menu and enter the mRNA file we created from the last section:

**$ transcribed\_rna.txt**

1. Once translated, the amino acid sequence is printed to the command line and can be written to an output file:

**$ translated\_AA.txt**

1. Check the output file and view the sequence of amino acids synthesised.

**Random DNA Generator**

This tool can generate a DNA strand of a given number of bases, with each base being one of the 4 main bases at random (A, T, G, C).

1. Select the DNA generator option from the menu and enter the value **100** for the total number of base units.
2. The random DNA strand can be written to an output file given:

**$ random\_dna.txt**

1. Check the output file and see if you can spot any start codons within the base sequence.

**Random DNA Mutation**

During the process of DNA replication there can be errors in encoding the nucleotide bases. These often happen due to the proofreading of the DNA Polymerases. The error of a different base is a mutation and can significantly affect the protein product. Mutations can change the properties of the coded amino acid polarity, produce a stop codon too early and ultimately lead to genetic diseases.

This tool adds an additional random nucleotide base to a random place within the base sequence of a file that is inputted by the user.

1. Select the random Mutation option from the menu and enter in a DNA strand input file:

**$ test.txt**

1. The mutated DNA strand can be written to an output file:

**$ mutated\_dna.txt**

1. Now we want to transcribe this mutated DNA strand to RNA using the Transcribe DNA option from the menu and entering our previous output file:

**$ mutated\_dna.txt**

1. Name the output file:

**$ mutated\_rna.txt**

1. Then translate the RNA file to its amino acid sequence by selecting the Translation tool from the menu and inputting the mutated RNA file and name the output file:

**$ mutated\_rna.txt**

**$ mutated\_AA.txt**

1. We can now view the mutated\_AA.txt file to see if the mutation has any effect on the amino acids produced in the chain, the length of the chain or none at all. See if you can figure out where the mutation has occurred based on the resulting amino acid sequence.