

## Assignment 1

7/Mar/2020

1. Write a code to accept any DNA sequence (of varying lengths) and produce as output the corresponding RNA strand synthesized and protein strand synthesized?

Sample DNA sequence:

```
gtttcattataccagtttagatctatcgacagggcggtgagtggtgcttactcacggct
ggcatgtaggtaacagtagtggggaagcgtaacatctgaggcctgactcacatatagagt
gtcgaccaaggggtgaagcatcatagccatacaggcccctagcgaaacgacctagtcta
aagacacacgagaatgaaacccgtggacttggttacagcgtaataatctggtcagagctg
gtccggcgctggcgatgtaccttacgccactgcaaaccggcttgcagagaacatctggg
tacattcccgtgtcatgtcaaagcaggtgattcccgcgaaaaacaattaacgacgcattt
gctattgacgaagtcctagttctccgaattgagcgggagacatatgatgtcgagactgca
ggaaccgaattatcctgtccgcagatccaatagctcacagaggtaaggggagtgatgg
tgccctagggtgtttgaacg
```

2. Is Taxonomy a constant nomenclature for a species? Give reasons for your answer.
3. Give the taxonomy classification of humans. Are there any other species with Genus – Homo? Mention them.
4. What observations did Charles Darwin make during his voyage across Galapagos islands?
5. What are the roles of repressors and activators in a gene?
6. What do you mean by alternative splicing and alternative initiation?
7. What kind of biological data analysis can help in testing and in finding solutions for treatment options for combating COVID19?

Deadline for submission: 14/Mar/2020

The program should be uploaded. Execution/evaluation will be done on a specified date.

Handwritten solutions have to be submitted for all other questions.

Assignment can be submitted at TA office – B2 301.

Solutions : ( I am writing them here for the time being , then later , I shall make the final hard copy for submission :

**Solution 1 :** Code file to be submitted later :

**Solution 2 :**

Taxonomy is not a constant nomenclature for a species.

This is because taxonomy of a species can get changed in the following circumstances.

1. New species are discovered that have to be placed in the tree structure of the species.

This might lead to some new groups forming and that could potentially lead to changing of the taxonomic nomenclature.

2. New links are discovered between the currently existing species. In this case , their positions might need to be changed in the species tree. This could lead to change in the taxonomical nomenclature of the species.

3. Some new information may be discovered about the past of the species in question itself.

This may lead to potential changes in the species tree and potential changes in the taxonomic nomenclature of the species.

**Solution 3 :**

Given the taxonomic classification of the humans, Yes, there are other species with the Genus Homo. Some of them are listed as under:

1. Homo Neanderthalensis

2. Homo Floresiensis

3. Homo Erectus

There are many others too .

**Solution 4 :**

The observations made by Charles Darwin in the Galapagos islands were :

1. In each island, there were unique birds, species and tortoises and no two islands had exactly the same kind of species, though similar to those found in the nearest island.

2. Though the species were different in different islands, based on the environment which was

being provided by the island, the species could adapt to that environment. Eg : In the islands where vegetations were found at a much lower level, the neck of the tortoises were smaller, while in those islands, where vegetations were slightly at a higher level, the tortoises had a longer neck ( We infer from this that the species can adapt to the changing environment )

3. As a population grows in size, the resources become limited. This results in a competition among the individuals of a population for the resource. This leads to the “survival of the fittest “ I.e Individuals with traits, or variations which allow them to best adapt to the environment are most likely to survive and reproduce , and also pass on these favorable traits to the next generation.

#### **Solution 5 :**

In one line, Repressors inhibit the transcription and activators that stimulate the activity of RNA polymerase. This is explained in detail as under :

Repressors and activators are proteins produced in the cell. Both repressors and activators regulate gene expression by binding to specific DNA sites *adjacent* to the genes they control. In general, the activators bind to the promoter site, while repressors bind to operator regions. Repressors prevent transcription of a gene in response to an external stimulus, whereas activators increase the transcription of a gene in response to an external stimulus.

#### **Solution 6 :**

Basically , both activators and repressors can be encoded by the same gene. Two mechanisms have been proposed to account for this. They are alternative splicing and alternative initiation. Their explanation is as follows :

##### **1. Alternative Splicing :**

- a. A cell can splice an RNA transcript differently and thereby make different polypeptide chains from the same gene—a process called alternative RNA splicing. A substantial proportion of animal genes (estimated to be 90% in humans ) produce multiple proteins this way. It is rare in single celled budding yeasts but very common in flies.
- b. In some cases , alternate splicing occurs because there is an intron sequence ambiguity. The standard spliceosome mechanism for removing intron sequences is unable to distinguish clearly between two or more alternate pairings of 5' and 3' splice sites, so that different choices are made by chance on different individual transcripts.
- c. The process is shown in detail via the diagram below:
  - i. { draw the diagram from the slides }

##### **2. Alternative Initiation of Translation:**

- a. The process of initiation occurs after RNA polymerase binds to a specific Dna sequence called a promoter. This is a relatively straightforward procedure which is followed by the second Phase called the Transcription.
- b. These new methods have only been recently discovered. Recently, high throughput assays have identified thousands of mammalian transcripts with translation initiation occurring at non canonical start codons, upstream of and within protein coding regions.
- c. While these newly discovered modes of translation initiation have largely been studied in isolation , it is likely that several act on the same mRNA and exquisite coordination is necessary to maintain 'normal' translation.
- d. The procedure of alternate initiation is showed in the following diagram
  - i. { draw the diagram from the slides }

### **Solution 7:**

There are some pretty good machine learning and AI tools that have assisted researchers to predict the virus, it's structure and it's spreading methods. Artificial Intelligence in fact tops the technologies used to handle this coronavirus outbreak.

Tech companies are developing applications to help people confirm their movements during the period of the outbreak as a safety measure and to avoid further spread. Eg data from trains is a great example of passenger screening where applications check the movement and contact of the people.

Predictive analytics and big data have become a powerful tool for combating this deadly disease outbreaks.

There are startups that have developed intelligent systems that sift through data about people to determine the chances of disease occurrence.

There are AI systems that gave an outbreak alert with regards Coronavirus as early as December 2019.

In a similar fashion, Natural Language Processing is used by startups where they analyse human languages and use the information to assist in forecasting disease outbreaks. Even ML is being used to help give updated information about disease outbreaks.

Eg : There is a startup which is working to use focussed analytics to develop tech that will inform doctors about molecules capable of fighting against the coronavirus. It is currently developing a database of molecular information that medical researchers can use in combating this pandemic.

Eg: Harvard Medical School is spearheading the efforts to find solutions for coronavirus by using machine learning technology to review data and information from various sources including patient records, social media , public health data. As a result of NLP, researchers can search online for information about the virus and understand the current location of the outbreak.

Thus we can conclude that there are a variety of ways through which data science and AI and ML can help fight the coronavirus.