

Sequence Alignment

Jyoti Kataria
Bioinformatics Analyst, Medgenome Labs Ltd.
M.Sc. Biotechnology (IIT Bombay)





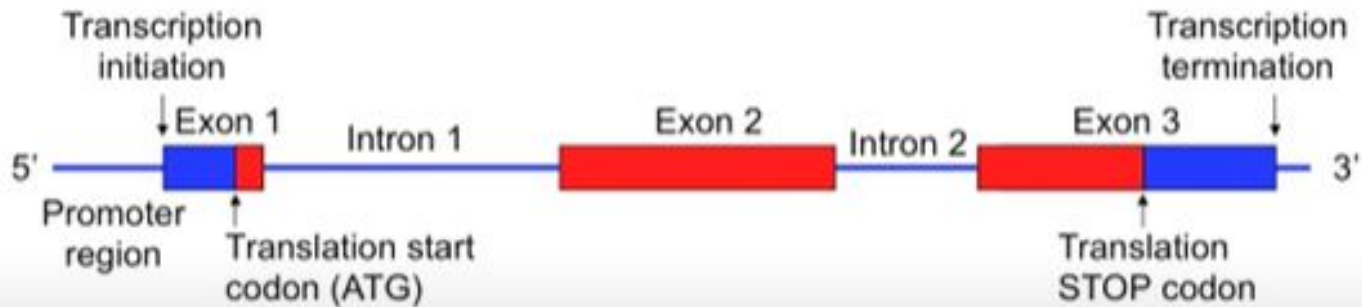
Inclass Activity 1 :

Biological databases trip (10-15 mins)

Take the assigned databases from the google sheet, find out the answers to the following three questions :-

1. What does it have?
2. How is the data stored inside it?
3. Why this database is important or Who all would be the potential users for the same?

Gene Structure





Quick Recap

DNA & RNA

Central Dogma

Gene

Introns & Exons

Databases

Dna replication

Transcription

Translation

Codons

1

Sequences you should be comfortable with :

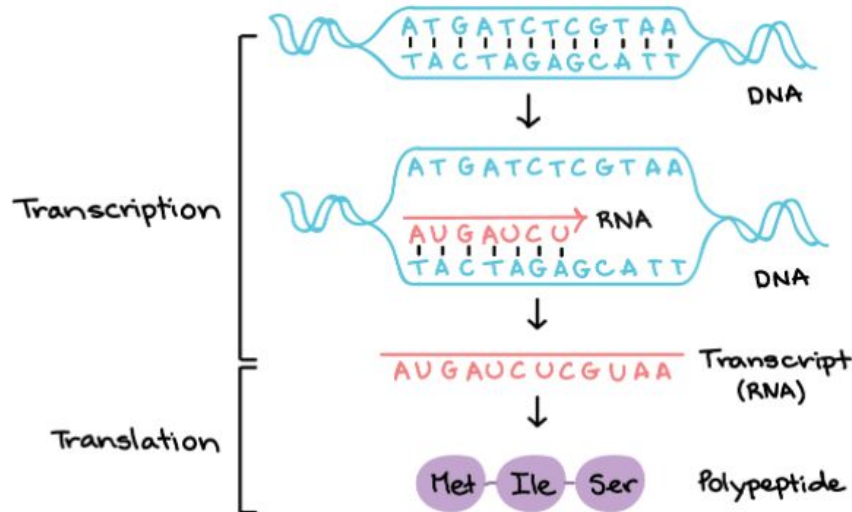
TTGGATTTCCCCAGGAGGAGTTTGGCAACCAGTTCCAAAAGGCTGAAACCATCCCTG
 TCCTCCATGAGATGATCCAGCAGATCTTCAATCTCTTCAGCACAAAGGACTCATCTG
 CTGCTTGGGATGAGACCCTCCTAGACAAATTCTACACTGAACTCTACCAGCAGCTGA
 ATGACCTGGAAGCCTGTGTGATACAGGGGGTGGGGGTGACAGAGACTCCCCTGATGA
 AGGAGGACTCCATTCTGGCTGTGAGGAAATACTTCCAAAGAATCACTCTCTATCTGA

2

human . AAGACUUCGGAUCUGGCG . ACA
 mouse aUACACUUCGGAUG - CACC . AAA
 orc . AGGUCUUC - GCACGGGGCAgCCA

3


FAEELQKRERFLLEREQLLFRHENALSKIKGVEEEVLTRFQIIKEQHDAE
 VEHLTEVLKEKNKETKRLRSSFDALKELDNLTKKQLNEASEENRKIDIQA
 KRVQARLDNLQRKYEFMTIQRLKGSSHAVHEMKSLLKQEKAPVSKTYKVPL
 NGQVYELLTVFMDWISDHLSKVKHEESGMDGKKPQLKFASQRNDIQEKC
 VKLLPLMTEQLQWMPFVNILHEPFVKFIYWSLRQLDAGAQHSTMTSTLR
 RLGEDIFKGVVTKGIQDNSPQHSVENKPKTAAPFKSSNLPLRFLSTLIVL





Alignment

- Arranging sequences of DNA, RNA or protein to dig out some relationships based on their similarity/dissimilarity
- Relationships could be functional, structural or evolutionary



Why compare sequences? Do more..

- Determine the evolutionary constraints at work
- Find mutations in a population or family of genes
- Find similar looking sequence in a database
- Find secondary/tertiary structure of a sequence of interest – molecular modeling using a template (homology modeling)

Alignment

Example 1. Multiple Sequence Alignment with a Single Nucleotide Polymorphism (SNP) in Sequence 3 at position 3.

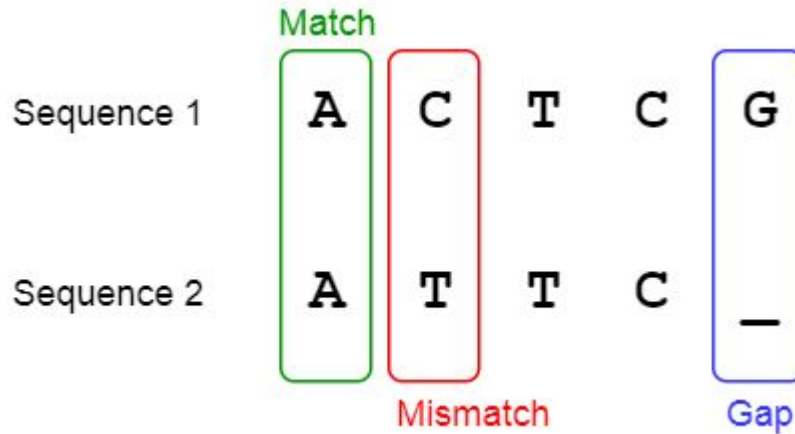
```
      1 2 3 4 5 6 7 8 <- Nucleotide Positions
Seq1 G T C G C A A A
Seq2 G T C G C A A A
Seq3 G T G G C A A A
```

Example 2. Multiple Sequence Alignment with an insertion in Sequence 1 at position 6.

```
      1 2 3 4 5 6 7 8
Seq1 G T C G C A A A
Seq2 G T C G C - A A
Seq3 G T C G C - A A
```




Scoring: Good or bad ? Match, Mismatch and Gap





Lets align manually (with our brains on)!

Part 1

$X = \text{ACGCTGAT}$ and $Y = \text{CAGCTAT}$.

```
AC-GCTGAT
 |  ||   ||
-CAGC-TAT
```

```
A C - G C T G A T
 |  | | |  | |
- C A G C T - A T |
```

If we set a scoring scheme as *match score* = 1, *mismatch score* = 0 and *gap penalty* = 0, then the overall score for the above alignment will be,



Lets align manually (with our brains on)!

Part 2

X=GGTCTGATG and Y=AAACGATC

```
CTGAT (in X)
|  |||
C-GAT (in Y)
```

If we set a scoring scheme as *match score* = 1, *mismatch score* = 0 and *gap penalty* = 0, then the overall score for the above alignment will be,

```
Score = nMatch x 1 + nMismatch x 0 + nGap x 0
       = 4x1 + 0x0 + 1x0
       = 4
```



Practical Session : Global and Local Alignment



Quick Recap

Pairwise Alignment

MSA

Indels

SNPs

