

CSE 449 – Bioinformatics

Pranta Sarker

Lecturer

Department of CSE

North East University Bangladesh



Outline

- FASTA
- BLAST
- Point mutation
- Why point mutation occurred
- Algorithm of point mutation



FASTA



- Pairwise sequence alignment tool.
- Input: nucleotides or protein sequences.
- Output: compare the sequence with existing database.
- Format: text-based can be written with the help of text editor or word processor.
- A FASTA file starts with the '>' symbol.



BLAST

- Basic Local Alignment Search Tool or BLAST.
- Pairwise sequence alignment tool.
- Input: nucleotides or protein sequences, known as a “query” sequence.
- Output: compare the sequence with existing database, known as “subject” sequence.
- Format: text-based can be written with the help of text editor or word processor.
- A BLAST file starts with the ‘>’ symbol as same as the FASTA file format.

Point mutation

- When mutation happened on a single nucleotides –

ATTGC
ATTCC

- Also known as the hamming distance, $d_H(LS, RS)$.
- In the same species that point mutation is known as a special name Single Nucleotide Polymorphism or SNP.
- Mutation in DNA is an evolutionary process; the DNA replication errors cause *substitutions*, *insertions*, and *deletions* of nucleotides, leading to “edited” DNA texts.

Point mutation

- In point mutation, sometimes it is occurred that a nucleotide might replace/substitute another nucleotide, such as *A* is substituted by *T*. For example –

ATATATAT
TATATATA

- They are very different from the perspective of Hamming distance.
- However, they are very similar if one simply moves the second string over one place to align $(i + 1)$ st letter in *ATATATAT* against the i th letter in *TATATATA*. Therefore,

ATATATAT –
–TATATATA

Edit Distance

- Unlike Hamming distance, Edit Distance allows one to compare strings of different lengths.
- For example, we have required operations to change *ATCCGAT* to *TGCATAT*.

$$\begin{cases} \min\{(i-1, j), (i, j-1), (i-1, j-1)\} + 1, & \text{if } m[i][j] \neq n[i][j] \\ (i-1, j-1) & \end{cases}$$

Edit Distance



		<i>T</i>	<i>G</i>	<i>C</i>	<i>A</i>	<i>T</i>	<i>A</i>	<i>T</i>
	0	1	2	3	4	5	6	7
<i>A</i>	1	0+1=1	1+1=2	2+1=3	3	3+1=4	5	5+1=6
<i>T</i>	2	1	1+1=2	2+1=3	3+1=4	3	3+1=4	5
<i>C</i>	3	1+1=2	1+1=2	2	2+1=3	3+1=4	3+1=4	4+1=5
<i>C</i>	4	2+1=3	2+1=3	2	2+1=3	3+1=4	4+1=5	4+1=5
<i>G</i>	5	3+1=4	3	2+1=3	2+1=3	3+1=4	4+1=5	5+1=6
<i>A</i>	6	4+1=5	3+1=4	3+1=4	3	3+1=4	4	4+1=5
<i>T</i>	7	6	4+1=5	4+1=5	3+1=4	3	3+1=4	4



Any Questions??