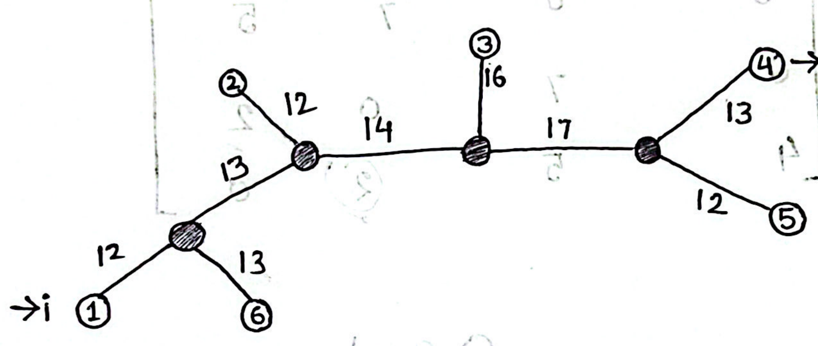
Bioinformatics Engineering

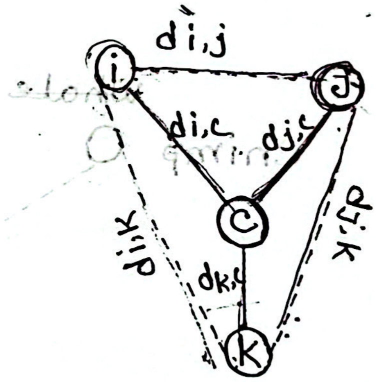
Distance in tree

$d_{ij}(T)$ = Tree distance between i and j



$d_{1,4} = 12+13+14+17+13$

$= 69$



Distance matrix:

$$d_{j,c} = \frac{d_{ij} + d_{j,k} - d_{i,k}}{2}$$

$$d_{i,c} = \frac{d_{i,j} + d_{i,k} - d_{j,k}}{2}$$

$$d_{k,c} = \frac{d_{ik} + d_{jk} - d_{ij}}{2}$$

Observe:

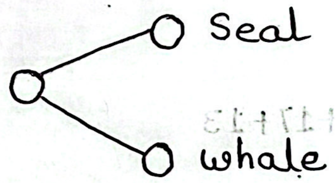
$d_{i,k} = d_{kc} + d_{ic}$

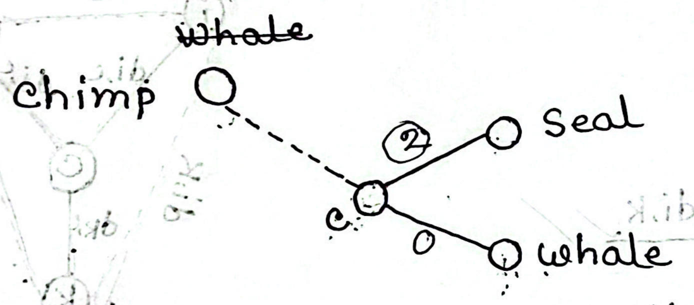
$d_{i,j} = d_{ic} + d_{jc}$

$d_{jk} = d_{jc} + d_{kc}$

Distance matrix

$$D = \begin{array}{c|cccc} & \text{chimp} & \text{human} & \text{seal} & \text{whale} \\ \hline \text{chimp} & 0 & 3 & 6 & 4 \\ \text{human} & 3 & 0 & 7 & 5 \\ \text{seal} & 6 & 7 & 0 & 2 \\ \text{whale} & 4 & 5 & 2 & 0 \end{array}$$



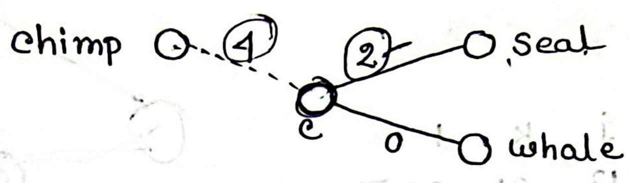


$$d_{seal,c} = \frac{d_{seal,whale} + d_{seal,chimp} - d_{chimp,whale}}{2}$$

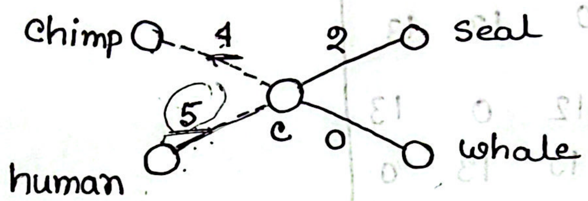
$$= \frac{2+6-4}{2} = 2$$

$$d_{whale,c} = \frac{D_{seal,whale} + D_{chimp,whale} - D_{seal,chimp}}{2}$$

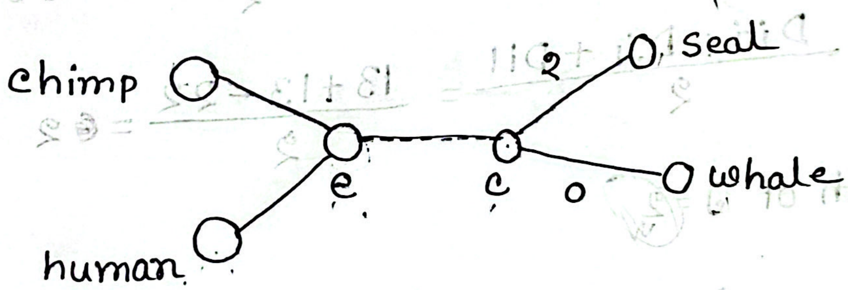
$$= \frac{2+4-6}{2} = 0$$



$6 - 2 = 4$



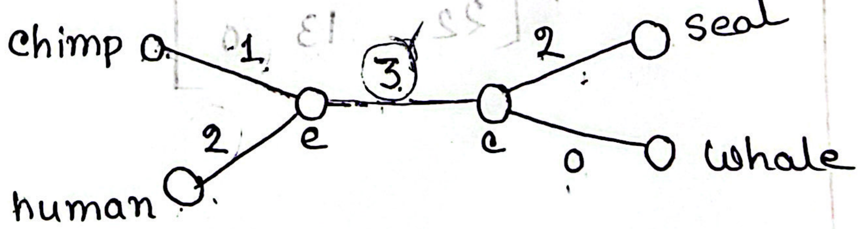
$$D' = \begin{array}{c|ccc} & \text{chimp} & \text{human} & c \\ \hline \text{chimp} & 0 & 3 & 4 \\ \text{human} & 3 & 0 & 5 \\ c & 4 & 5 & 0 \end{array}$$



$$d_{chimp, e} = \frac{D_{chimp, human} + D_{chimp, c} - D_{human, c}}{2}$$

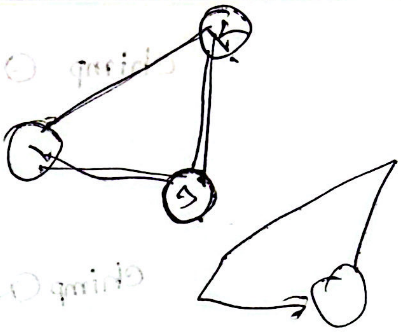
$$= \frac{3 + 4 - 5}{2} = 1$$

$1 + 3 \quad 2 = 6$



Limb Length

Example



$$D = \begin{array}{c|cccc} & i & j & k & l \\ \hline i & & & & \\ j & & & & \\ k & & & & \\ l & & & & \end{array}$$

$$D = \begin{array}{cc} & \begin{array}{cccc} i & j & k & l \end{array} \\ \begin{array}{c} i \\ j \\ k \\ l \end{array} & \begin{bmatrix} 0 & 13 & 21 & 22 \\ 13 & 0 & 12 & 13 \\ 21 & 12 & 0 & 13 \\ 22 & 13 & 13 & 0 \end{bmatrix} \end{array}$$

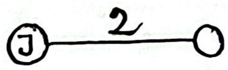
possible leaf keeping J common :-

$$(i,j,k) \rightarrow d_j = \frac{D_{ij} + D_{jk} - D_{ik}}{2} = \frac{13+12-21}{2} = \cancel{6}\,2$$

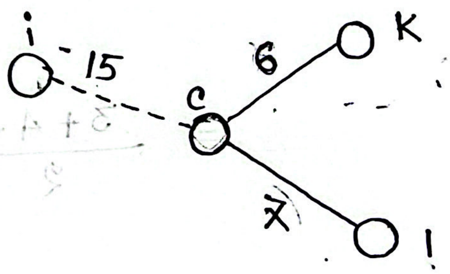
$$(J,k,l) \rightarrow d_j = \frac{D_{kj} + D_{jl} - D_{lk}}{2} = \frac{12+13-\cancel{2}13}{2} = \cancel{2}\,6$$

$$(i,j,l) \rightarrow d_j = \frac{D_{ij} + D_{jl} + D_{il}}{2} = \frac{13+13-22}{2} = \cancel{6}\,2$$

So, limb length of J = 2



$$D^{trim} = \begin{array}{cc} & \begin{array}{ccc} i & k & l \end{array} \\ \begin{array}{c} i \\ k \\ l \end{array} & \begin{bmatrix} 0 & 21 & 22 \\ 21 & 0 & 13 \\ 22 & 13 & 0 \end{bmatrix} \end{array}$$



$$d_{kc} = \frac{D_{kl} + D_{ik} - D_{il}}{2}$$

$$= \frac{13 + 21 - 22}{2}$$

$$= 6$$

$$D^{Bold} = \begin{array}{c|cccc} & i & j & k & l \\ \hline i & 0 & 11 & 21 & 22 \\ j & 21 & 0 & 10 & 11 \\ k & 21 & 10 & 0 & 13 \\ l & 22 & 11 & 13 & 0 \end{array}$$

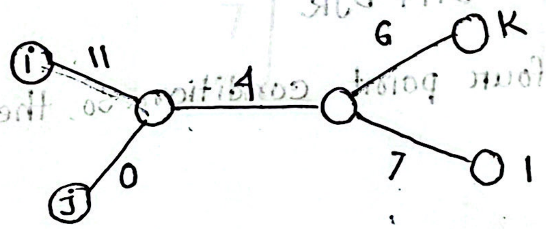
$$D_j^{Bold} = \frac{D_{ij}^{Bold} + D_{jk}^{Bold} - D_{ik}^{Bold}}{2} = \frac{11 + 10 - 21}{2} = 0$$

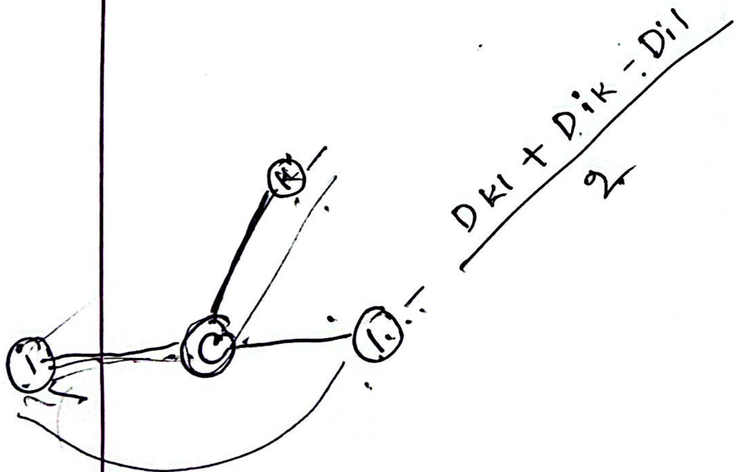
$$\Rightarrow \frac{D_{ij}^{Bold} + D_{jk}^{Bold} - D_{ik}^{Bold}}{2} = 0$$

$$\Rightarrow D_{ij}^{Bold} + D_{jk}^{Bold} - D_{ik}^{Bold} = 0$$

$$\Rightarrow 11 + 10 = 21$$

Final tree,





Four point condition

$$
D = \begin{array}{c|cccc}
 & i & j & k & l \\
\hline
i & 0 & 13 & 21 & 22 \\
j & 13 & 0 & 12 & 13 \\
k & 21 & 12 & 0 & 13 \\
l & 22 & 13 & 13 & 0
\end{array}
$$

1. $D_{ij} + D_{kl} = 13 + 13 = 26$ ✓

2. $D_{ik} + D_{jl} = 21 + 13 = 34$ ✓

3. $D_{il} + D_{jk} = 22 + 12 = 34$ ✓

$D_{ij} + D_{kl} < D_{ik} + D_{jl} = D_{il} + D_{jk}$

This satisfy the four point condition. So, the matrix is additive matrix.

Example

$$D = \begin{array}{c|ccccc} & i & j & k & l & m \\ \hline i & 0 & 3 & 4 & 5 & 6 \\ j & 3 & 0 & 5 & 6 & 7 \\ k & 4 & 5 & 0 & 7 & 8 \\ l & 5 & 6 & 7 & 0 & 9 \\ m & 6 & 7 & 8 & 9 & 0 \end{array}$$

The possible 4x4 matrix from the above matrix:

$$D_1 = \begin{array}{c|cccc} & i & j & k & l \\ \hline i & 0 & 3 & 4 & 5 \\ j & 3 & 0 & 5 & 6 \\ k & 4 & 5 & 0 & 7 \\ l & 6 & 7 & 8 & 9 \end{array} \qquad D_2 = \begin{array}{c|cccc} & i & j & k & m \\ \hline i & 0 & 3 & 4 & 6 \\ j & 3 & 0 & 5 & 7 \\ k & 4 & 5 & 0 & 8 \\ m & 6 & 7 & 8 & 0 \end{array}$$

$$D_3 = \begin{array}{c|cccc} & i & j & l & m \\ \hline i & 0 & 3 & 5 & 6 \\ j & 3 & 0 & 6 & 7 \\ l & 5 & 6 & 0 & 9 \\ m & 6 & 7 & 9 & 0 \end{array} \qquad D_4 = \begin{array}{c|cccc} & i & k & l & m \\ \hline i & 0 & 4 & 5 & 6 \\ k & 4 & 0 & 7 & 8 \\ l & 5 & 7 & 0 & 9 \\ m & 6 & 8 & 9 & 0 \end{array}$$

$$D_5 = \begin{array}{c|cccc} & j & k & l & m \\ \hline j & 0 & 5 & 6 & 7 \\ k & 5 & 0 & 7 & 8 \\ l & 6 & 7 & 0 & 9 \\ m & 7 & 8 & 9 & 0 \end{array}$$

From D_1,

1. $D_{ij} + D_{kl} = 3+7=10$

2. $D_{ik} + D_{jl} = 4+6=10$

3. $D_{il} + D_{jk} = 5+5=10$

From D_2,

1. $D_{ij} + D_{km} = 3+8=11$

2. $D_{ik} + D_{jm} = 4+7=11$

3. $D_{im} + D_{jk} = 6+5=11$

From D_3,

1. $D_{ij} + D_{lm} = 3+9=12$

2. $D_{il} + D_{jm} = 5+7=12$

3. $D_{im} + D_{jl} = 6+6=12$

From D_4,

1. $D_{ik} + D_{lm} = 4+9=13$

2. $D_{il} + D_{km} = 5+8=13$

3. $D_{im} + D_{kl} = 6+7=13$

From D_5,

1. $D_{jk} + D_{lm} = 5+9=14$

2. $D_{jl} + D_{km} = 6+8=14$

3. $D_{jm} + kl = 7+7=14$

UPGMA: Unweighted pair group method with arithmatic mean.

$$D_{C_1,C_2} = \frac{\sum_{i\in C_1} \sum_{j\in C_2} D_{i,j}}{|C_1|.|C_2|}$$

Example:

$$D = \begin{array}{c|cccc} & i & j & k & l \\ \hline i & 0 & 3 & 4 & 3 \\ j & 3 & 0 & 4 & 5 \\ k & 4 & 4 & 0 & 2 \\ l & 3 & 5 & 2 & 0 \end{array}$$

$D_{i,j} = 3$

$3/2 = 1.5$

$D_{ij/kl} = 4$

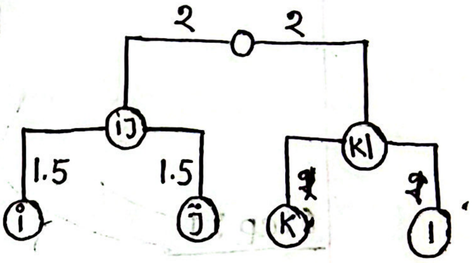
$4/2 = 2$

$D_{kl} = \cancel{4}\,2$

$2/2 = 1$

$2/2 = 1$

$D_{ij} = 3 = 1.5$



$$\Rightarrow D' = \begin{array}{c|ccc} & i & j & kl \\ \hline i & 0 & 3 & 3.5 \\ j & 3 & 0 & 4.5 \\ kl & 3.5 & 4.5 & 0 \end{array}$$

$$\longrightarrow D'_{i,kl} = \frac{D_{ik} + D_{il}}{2} = \frac{4+3}{2} = 3.5$$

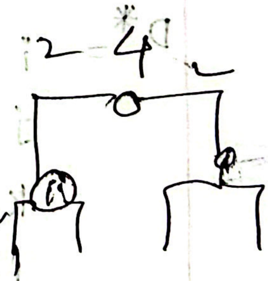
$$D'_{j,kl} = \frac{D_{jk} + D_{jl}}{2} = \frac{4+5}{2} = 4.5$$

$$\Rightarrow D'' = \begin{array}{c|cc} & ij & kl \\ \hline ij & 0 & 4 \\ kl & 4 & 0 \end{array}$$

$$; \longrightarrow D''_{ij,kl} = \frac{D_{i,kl} + D_{j,kl}}{2} = \frac{3.5+4.5}{2} = 4$$

$D_{kl} = 4$

$4/2 = 2$



$D_{ij} =$

$D_{kl} =$ $\quad D_{i\,kl} =$

$D_{j\,kl} =$

$$D = \begin{array}{c|c} i & \\ j & \\ k & \end{array}$$

Neighbor Joining Algorithm

Example

$$D = \begin{array}{c|cccc} & i & j & K & l \\ \hline i & 0 & 13 & 21 & 22 \\ j & 13 & 0 & 12 & 13 \\ K & 21 & 12 & 0 & 13 \\ l & 22 & 13 & 13 & 0 \end{array}$$

Formula

$$D^*_{ij} = (n-2) \times \boxed{D_{ij}} - TD_D(i) - TD_D(j)$$

TD = The sum of the dis: from one leaf to another leaf

Step: 1

Total Distance,

$i = 0+13+21+22 = 56$

$j = 13+0+12+13 = 38$

$K = 21+12+0+13 = 46$

$l = 22+13+13+0 = 48$

$D_{ij} = (4-2) \times 13 - 56 - 38 = -68$

$D_{iK} = (4-2) \times 21 - 56 - 46 = -60$

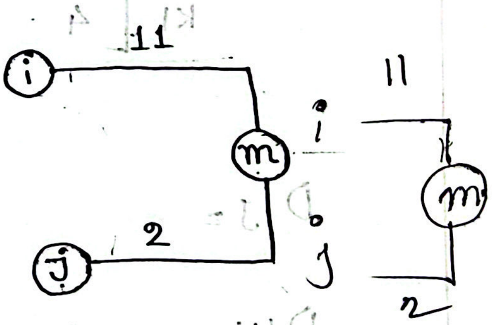
$D_{il} = (4-2) \times 22 - 56 - 48 = -60$

$D_{jK} = (4-2) \times 12 - 38 - 46 = -60$

$D_{jl} = (4-2) \times 13 - 38 - 48 = -60$

$D_{Kl} = (4-2) \times 13 - 46 - 48 = -68$

$$D^* = \begin{array}{c|cccc} & i & j & K & l \\ \hline i & 0 & -68 & -60 & -60 \\ j & -68 & 0 & -60 & -60 \\ K & -60 & -60 & 0 & -68 \\ l & -60 & -60 & -68 & 0 \end{array}$$



$$\Delta = \frac{\text{Total Distance}_D(i) - \text{Total Distance}_D(j)}{n-2}$$

$$= \frac{56-38}{4-2}$$

$$= \frac{18}{2} = 9$$

Length (i) $= \frac{1}{2}(D_{ij} + \Delta)$

$= \frac{1}{2}(13+9)$

$= \frac{1}{2} \times 22$

$= 11$

Limb length (j) $= \frac{1}{2}(D_{ij} - \Delta)$

$= \frac{1}{2} \times (13-9)$

$= \frac{1}{2} \times 4$

$= 2$

Step: 2

m i j

$$D'_{mk} = \frac{D_{ik} + D_{jk} - D_{ij}}{2} = \frac{21+12-13}{2} = 10$$

$$D'_{ml} = \frac{D_{il} + D_{jl} - D_{ij}}{2} = \frac{22+13-13}{2} = 11$$

$$D' = \begin{array}{c|ccc} & m & k & l \\ \hline m & 0 & 10 & 11 \\ k & 10 & 0 & 13 \\ l & 11 & 13 & 0 \end{array}$$

Total Distance,

m = 0 + 10 + 11 = 21

k = 10 + 13 + 0 = 23

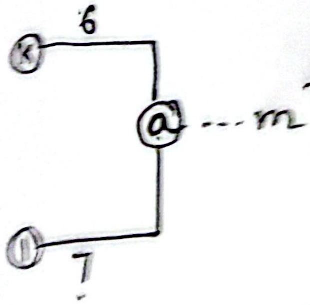
l = 11 + 13 + 0 = 24

$$\Rightarrow D'^{*} = \begin{array}{c|ccc} & m & k & l \\ \hline m & 0 & -34 & -34 \\ k & -34 & 0 & -34 \\ l & -34 & -34 & 0 \end{array}$$

$D'^{*}_{mk} = (3-2) \times 10 - 21 - 23 = -34$

$D'^{*}_{ml} = (3-2) \times 11 - 21 - 24 = -34$

$D'^{*}_{kl} = (3-2) \times 13 - 23 - 24 = -34$



$$\Delta = \frac{TD_D(k) - TD_D(l)}{n-2} = \frac{23-24}{3-2} = -1.$$

Limb Length (k) $= \frac{1}{2}(D'_{kl} + \Delta)$

$= \frac{1}{2} \times (\cancel{-34 + (-1)}\ 13 - 1)$

$= 6$

(l) $= \frac{1}{2}(D'_{kl} - \Delta)$

$= \frac{1}{2} \times 14$

$= 7$

Step : 3

$$D'' = \begin{array}{c|cc} & m & a \\ \hline m & 0 & 4 \\ a & 4 & 0 \end{array}$$

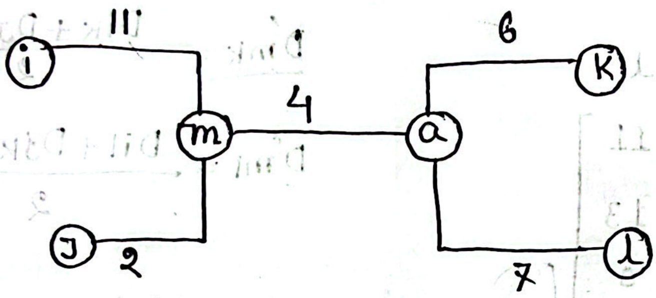
$$D'_{am} = \frac{D'_{km} + D'_{im} - D'_{kl}}{2}$$

$$= \frac{10+11-13}{2}$$

$$= 4$$

a m

k'

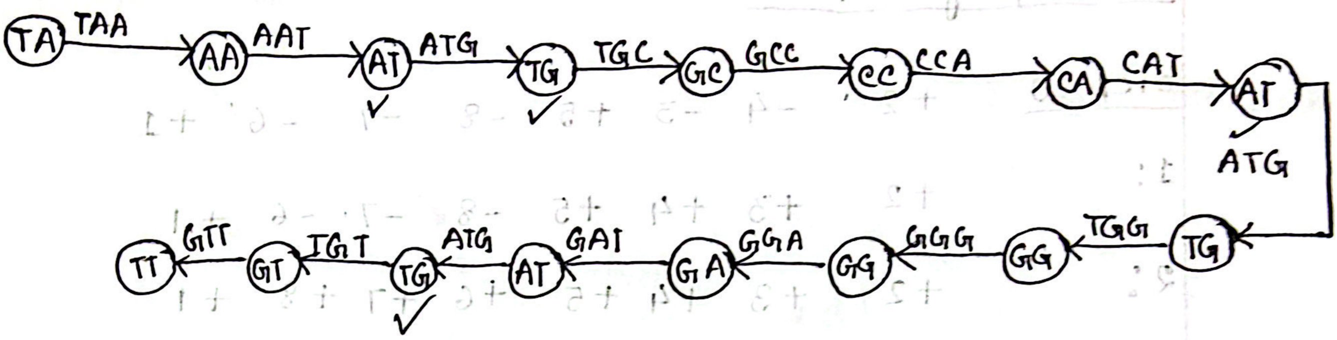


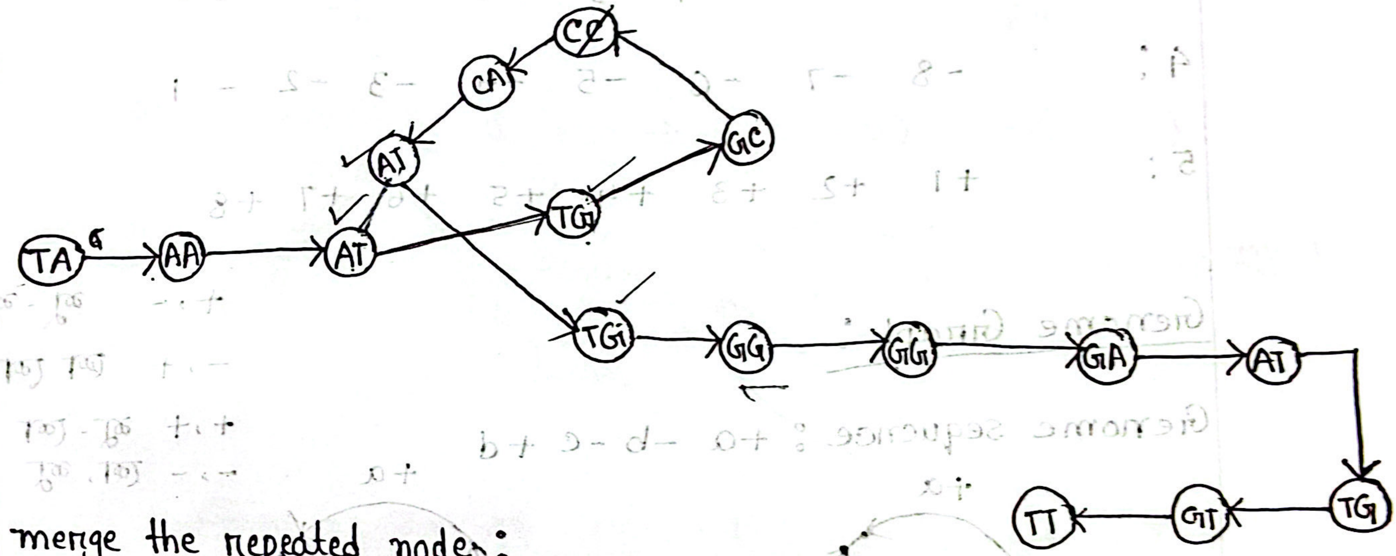
De Bruijn Graph

T A A T G C C A T G G G A T G T T

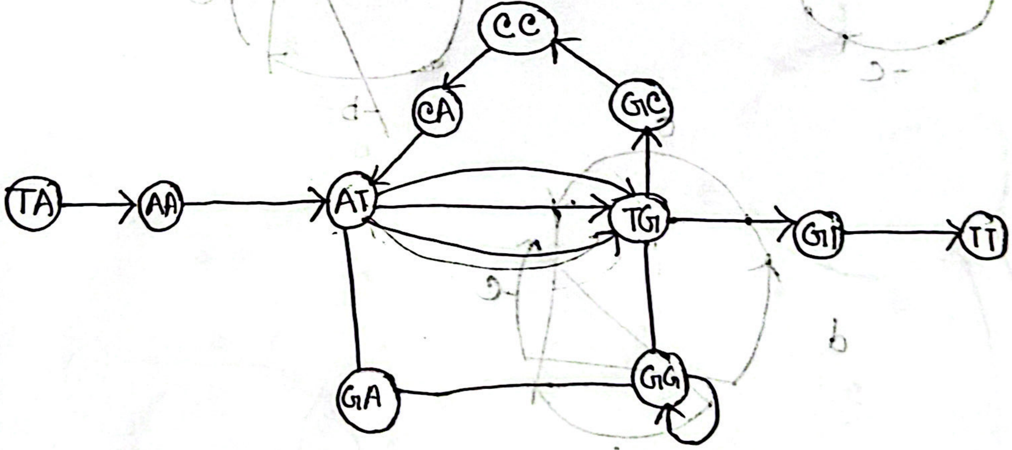
k-merse (K=3):

TAA
AAT
ATG
TGC
GCC
CCA
CAT
ATG
TGG
GGG
GGA
GAT
ATG
TGT





merge the repeated node:



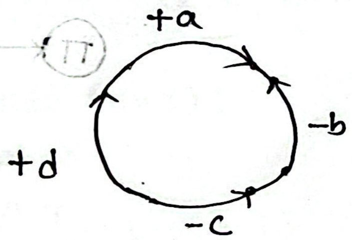
Sorting By Reversal

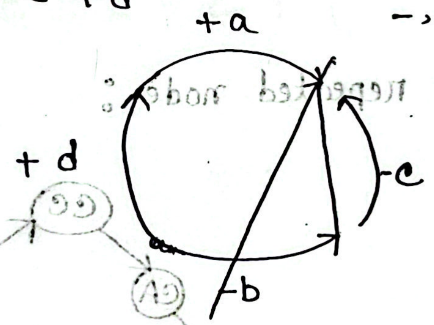
Step : 0	+2	−4	−3	+5	−8	−7	−6	+1
1:	+2	+3	+4	+5	−8	−7	−6	+1
2:	+2	+3	+4	+5	+6	+7	+8	+1
3:	−2	−3	−4	−5	−6	−7	−8	+1
4:	−8	−7	−6	−5	−4	−3	−2	−1
5:	+1	+2	+3	+4	+5	+6	+7	+8

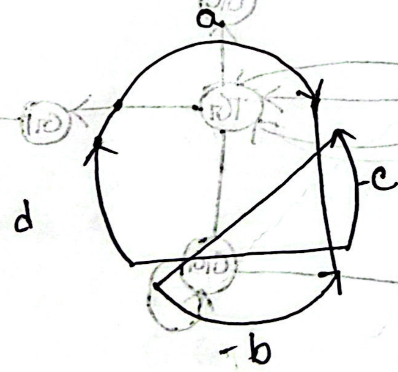
Genome Graph :

+, − মু - মু
−, + লে লে
+, + মু - লে
−, − লে, মু

Genome sequence : +a −b −c +d

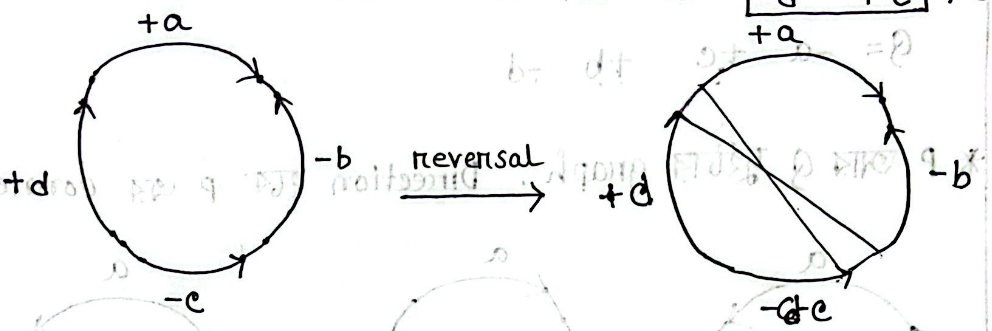




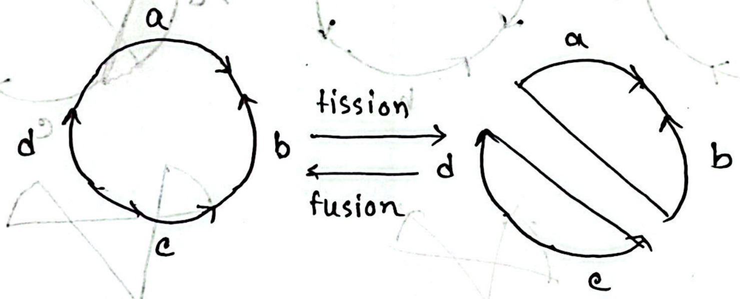


Genome e block operation :

+a −b −c +d $\xrightarrow{\rho(3,4)}$ +a −b $\boxed{-d \quad +c}$ → 3,4 change



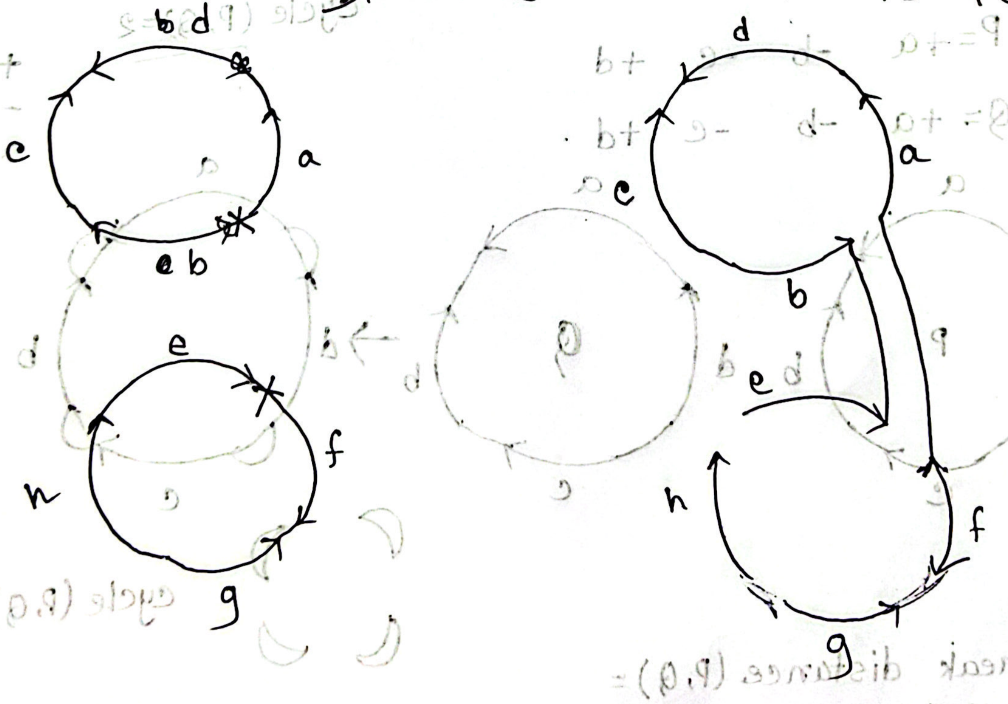
Fission : +a −b −c +d → (+a −b) (−c +d)



(Loop হবে)

Translocation : −a +b +c −d → −a +f −g +h

+e +f −g +h → +e +b +c −d

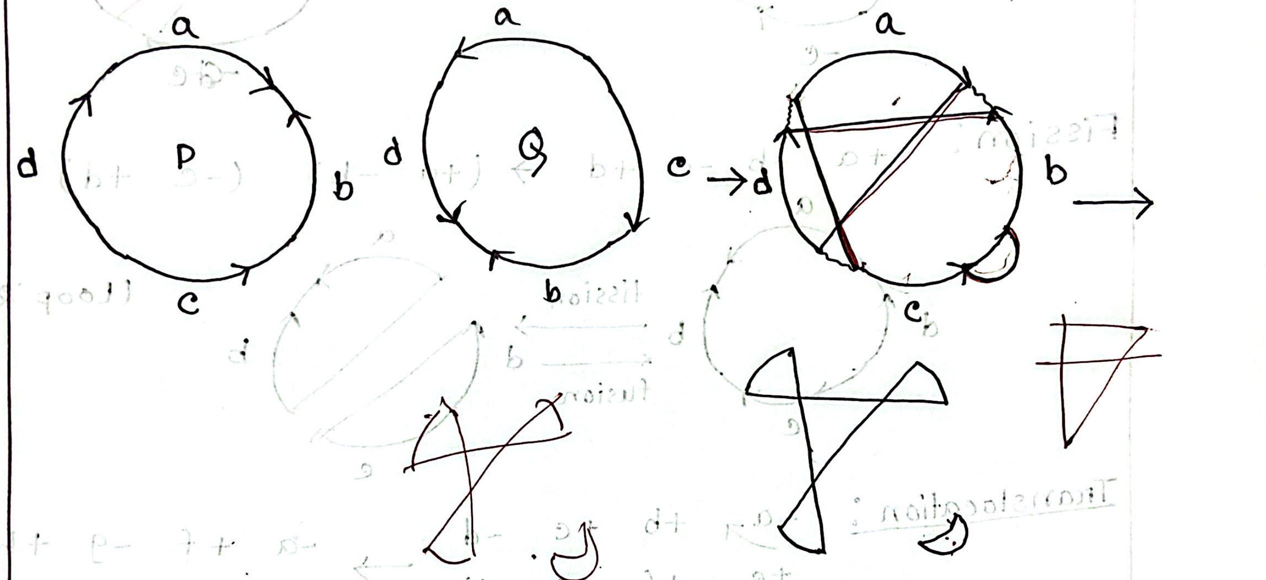


Break-point analysis graph

1. $P = +a \ -b \ -c \ +d$

$Q = -a \ +c \ +b \ -d$

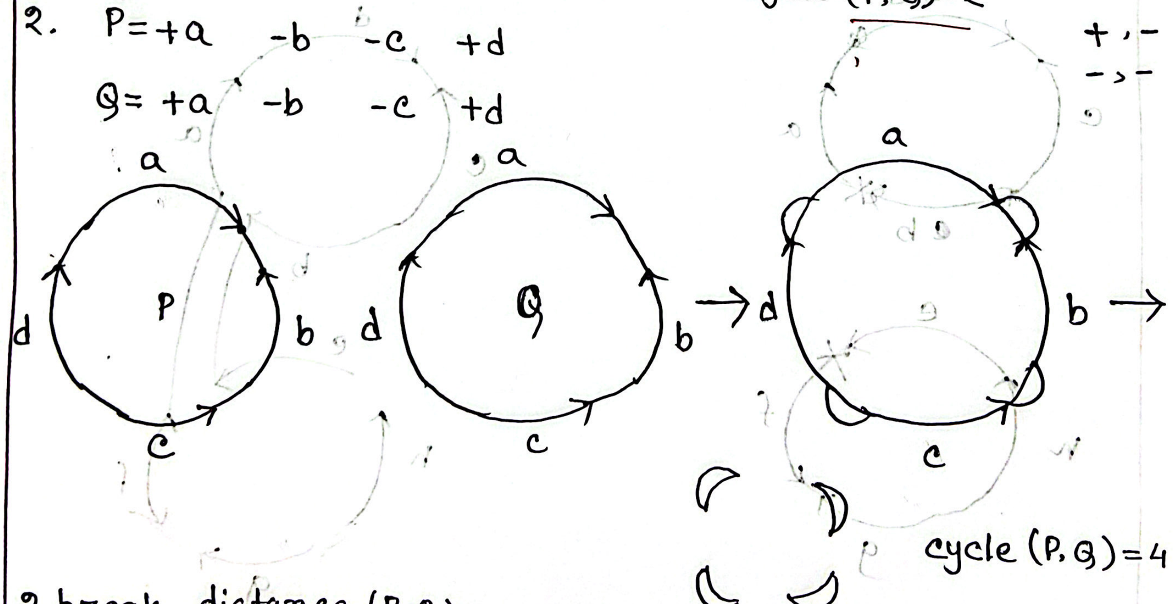
* P আর Q দুইটার graph. Direction হবে P এর connection হবে Q এর



cycle (P, Q) = 2

2. $P = +a \ -b \ -c \ +d$

$Q = +a \ -b \ -c \ +d$



cycle (P, Q) = 4

2 break distance (P, Q) =
no of blocks − no of cycle = 4 − 2 = 2