

BTE

Phylogeny

Pre-revolutionary state of living state or species

DNA based approach \Rightarrow study evolution

Out of Africa Hypothesis: Most ancestor lived in there
200000 years ago.

Leaves \Rightarrow existing
Internal vertices \Rightarrow Ancestors

Roots \Rightarrow oldest

No common Ancestor \Rightarrow Rootless \Rightarrow connect both
Tree \Rightarrow cycleless graph

Tree Distance: between nodes

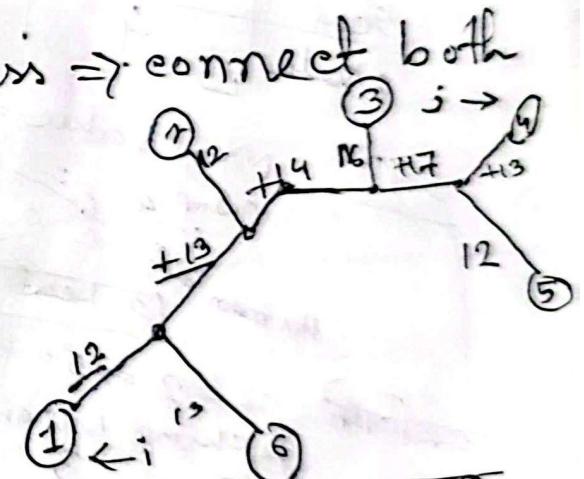
$$(12 + 13 + 14 + 17 + 12) = 68$$

Distance Matrix: $n \times n$

d_{ij} = edit distance

d_{ij} = tree distance

[between
2 species]
 $i \& j$

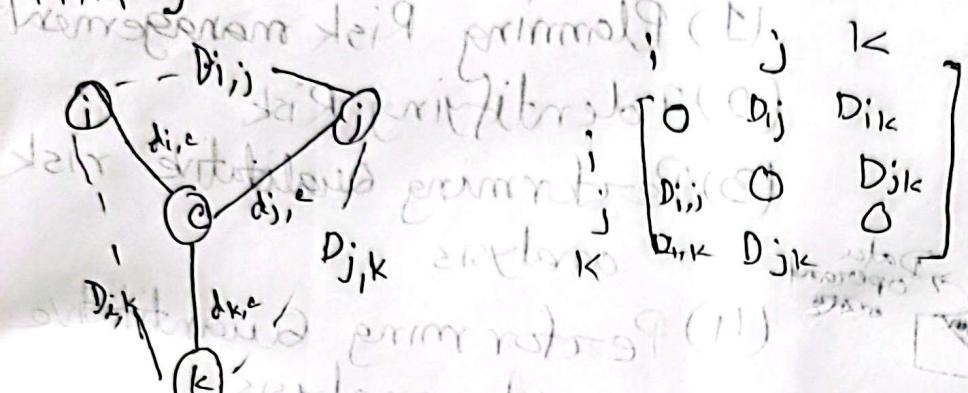


[edge value \Rightarrow evolution
mutation
go down]

= Time estimate of
one to another
species

Editing Distance Matrix

~~transformation~~ fitting means $D_{i,j} = d_{i,j}$



center vertex C

$$[i] + [1] + [11] = 2(d_{ic} + d_{jc} + d_{kc}) = D_{ij} + D_{jk} + D_{ik}$$

$$= d_{ic} + d_{jc} + d_{kc} = \frac{1}{2}(D_{ij} + D_{jk} + D_{ik})$$

$$\Rightarrow d_{ic} + d_{jc} = D_{ij}$$

$$\text{d}_{k \rightarrow i} = \frac{1}{2} (\text{D}_{ik} + \text{D}_{ik} - \text{D}_{ij})$$

$$d_{ik} = \frac{1}{2} (D_{ik} + D_{ij} - D_{jk})$$

$$d_{jk} = \frac{1}{2}(d_{ij} + d_{ik} - d_{ik})$$

Ex: Evolutionary Tree: 6x6 Distance Matrix:

	Chimp	Human	Seal	Whale
Chimp	0	3	7	5
Human	3	0	0	2
Seal	6	7	0	0
Whale	4	5	2	8

neighbours of each other
- common parent

$$\text{Solt!} \quad \text{Min value} = 2 \quad d_{\text{seal,clip}} = D_{\text{seal,whale}} + D_{\text{seal,chimp}} - D_{\text{whale,chimp}}$$

Human 3 Leaf structure

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

For cto e

Chimp to seal = 6
 6 stages, 3 more
 Human to whale = 5 stages,
 some nodes complete
 + D Chimp & Human

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

several softish

Leaf Length Theorem

HW BTE

$$D =$$

$$\begin{bmatrix} i & j & k & l \\ i & 0 & 13 & 21 & 22 \\ j & 13 & 0 & 12 & 13 \\ k & 21 & 12 & 0 & 13 \\ l & 22 & 13 & 13 & 0 \end{bmatrix}$$

Non
Additive
Matrix
Negative Rang

Finding Neighbour leaves

i & j are neighbours, but $d_{ij} = 13 > d_{jkl} = 12$

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

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

$$(j,k,l) \rightarrow d_j = \frac{d_{jk} + d_{jl} - d_{kl}}{2} = \frac{12 + 13 - 13}{2} = 6$$

For Limb Length Theorem,

Given an additive Matrix D and a leaf j, $\text{LIMBLENGTH}(j) = \min\left(\frac{D_{i,j} + D_{j,k} - D_{i,k}}{2}\right)$

bald matrix

* Tree Construction based

modification in j row & column

$$j - \min(\text{limblength}) \rightarrow$$

modification in j

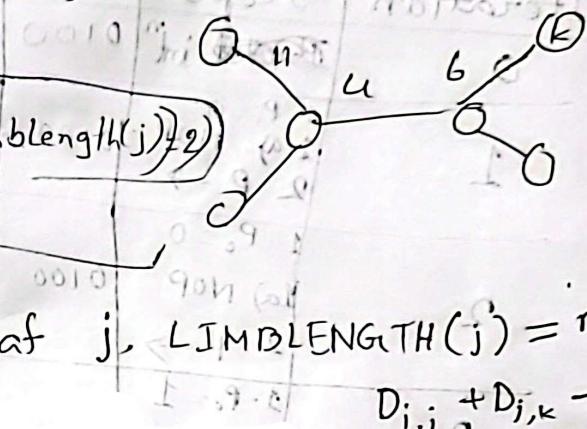
j row & column cancelled

using 3 leaf Calculation

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

$$\Rightarrow D_{ij} + D_{jk} = D_{ik}$$

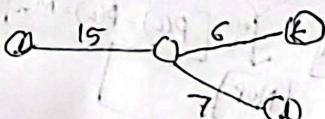
[i to k path
j comparing]



over all leaves i & k.

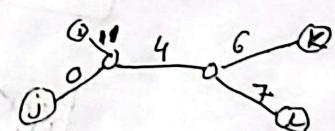
$$D^{\text{bald}} = \begin{bmatrix} i & j & k & l \\ i & 0 & 11 & 21 & 12 \\ j & 13 & 0 & 10 & 11 \\ k & 21 & 10 & 0 & 13 \\ l & 22 & 11 & 13 & 0 \end{bmatrix}$$

$$D^{\text{Trim}} = \begin{bmatrix} i & k & l \\ i & 0 & 21 & 22 \\ k & 21 & 0 & 13 \\ l & 22 & 13 & 0 \end{bmatrix}$$



; to intermediate = 15

$$i-j = 11$$



4 Point Combination satisfy: $D_{ij} + D_{kl}$ Result same
according to D_{bald}

$$(1) D_{ij} + D_{kl}, \quad (2) D_{ik} + D_{jl}, \quad (3) D_{il} + D_{jk}$$

$$\text{Now, } (1) + (2) = 26$$

$$(2) 21 + 13 = 34$$

$$(3) 12 + 12 = 34$$

So, that is an additive matrix

UPGMA			
i	j	k	l
0	3	4	3
3	0	4	5
4	4	0	2
3	5	2	0

$D =$

smallest node $j = j'$

2nd smallest node k, l

i	j	k	l
0	3	3.5	
3	0	4.5	
2.5	4.5	0	

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

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

$$D' = \begin{matrix} ij \\ ij \\ kl \end{matrix} \begin{bmatrix} D_{ij} & D_{ik} & D_{il} \\ D_{kj} & D_{kl} & D_{jl} \\ D_{li} & D_{lj} & D_{kl} \end{bmatrix}$$

$$D'_{ijkl} = \frac{D_{ij} + D_{kl}}{2}$$

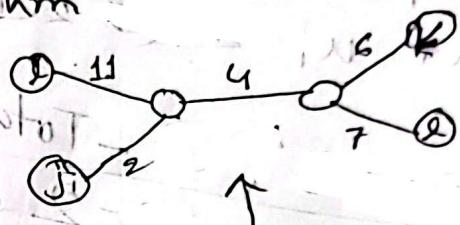
$$\frac{D_{ij} + D_{kl}}{2}$$

$$\frac{3.5 + 4.5}{2} = 4$$

Averaging Theorem can't do Ultra Matrix
Always do with Double Tree

Neighbor Joining Algorithm

$$D = \begin{matrix} i & j & k & l & \text{Total} \\ i & 0 & 13 & 21 & 22 & 56 \\ j & 13 & 0 & 12 & 23 & 38 \\ k & 21 & 12 & 0 & 13 & 46 \\ l & 22 & 23 & 13 & 0 & 48 \end{matrix}$$

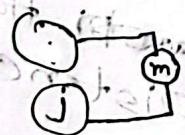


$$D^* = \begin{matrix} ij & kl \\ ij & kl \\ ik & jl \\ ik & jl \\ jl & ik \end{matrix} \begin{bmatrix} 0 & -68 & -60 & -60 \\ -68 & 0 & -60 & -60 \\ -60 & -60 & 0 & -68 \\ -60 & -60 & -68 & 0 \end{bmatrix} \rightarrow \begin{matrix} m & k & l & i & j \\ m & 0 & 10 & 11 & 21 \\ k & 10 & 0 & 13 & 23 \\ l & 11 & 13 & 0 & 24 \end{matrix}$$

$$\text{Step } 1 = \frac{\text{Total } D(ij)}{n=4}$$

$$D_{ij}^* = (n-2) \times D_{ij} - \text{Total } D(i) - \text{Total } D(j)$$

Position $\min(D^*) \rightarrow ij, kl$ share common parent



$$\text{imblength}(i) = \frac{1}{2} (D_{ij} + \Delta) = \frac{1}{2} (13 + 9) = 11$$

$$\text{imblength}(j) = \frac{1}{2} (D_{ij} - \Delta) = \frac{1}{2} (13 - 9) = 2$$

$$\Delta = \frac{\text{Total Distance } D(i) - \text{Total } D(j)}{n-2} = \frac{13 + 9}{4-2} = 11$$

Step-2

$$\text{for } D \rightarrow D'_{mk} = \frac{D_{ik} + D_{jk} - D_{ij}}{2}; D'_{ml} = \frac{D_{il} + D_{jl} - D_{ij}}{2}$$

$$\text{for } D^* \rightarrow D^*_{mk} = (3-2)/n - 21 - 23 = -34$$

$$D' = m \begin{bmatrix} m & K & l \\ 0 & -34 & -34 \\ -34 & 0 & -34 \\ -34 & -34 & 0 \end{bmatrix}$$

[common Parent consideration
m, K, l]

original leaf \rightarrow node CTC

min नाही स्वयंसेवा सुटी pair up बरते के (K & l)

$$\Delta = \frac{\text{Total } D'(K) - \text{Total } D'(l)}{2n-2} = \frac{23-24}{1} = -1$$

2x2 matrix
Iteration
at 2CTI

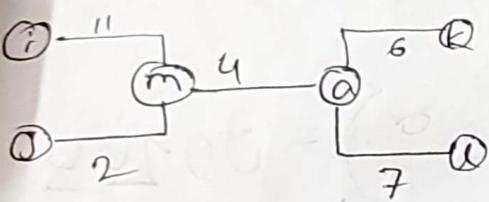
$$\text{Limblength}(K) = \frac{1}{2} (D'_{KK} + \Delta) = 6$$

$$\text{Limblength}(l) = \frac{1}{2} (D'_{ll} - \Delta) = 7$$

$$D'' = m \begin{bmatrix} m & a \\ 0 & u \\ a & 0 \end{bmatrix}$$

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

$$= \frac{10 + 11 - 13}{2} = 9$$



Alignment vs Distance Matrix

character-based Tree Reconstruction
parsimony score \rightarrow sum (Hamming Distance)

Accc
ACCA ACCG
ATCG ATCA
less Parsimonious
Score = 6

Edge Value

- (1) 200 million years Evolution
- (2) 2023 mutation
- 2021 mismatch \rightarrow Hamming distance

$\boxed{ACCA} \rightarrow 2$

More Score
 $\boxed{ACCA} \quad \boxed{ATCC}$
 $\boxed{ATCG} \quad \boxed{ACCG}$
Score = 5

Small Parsimony

Tree T with each leaf labeled by

Weighted Small Parsimony

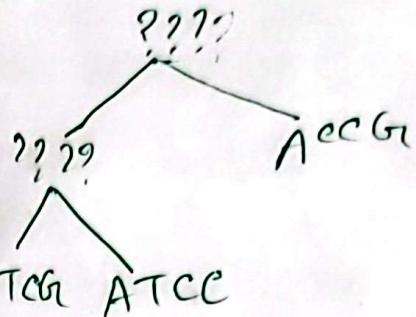
$$d_H(v, w) = 0 \text{ if } v=w$$

$$d_H(v, w) = 1 \text{ otherwise}$$

scoring Matrix

	A	C	G	T
A	0	1	1	1
C	1	0	1	1
G	1	1	0	1
T	1	1	1	0

Ex:



Parent 1st char = child's 1st char

2	0	1	2
---	---	---	---

2	1	1	2
---	---	---	---

(min)

Sankoff's Algo:

$s_t(v)$ = min parsimony score of the subtree rooted at vertex v if v has character t .

The score is based on its children

$$s_t(\text{parent}) = \min \{ s_i(\text{left child}) + s_{i,t} \} +$$

$$\min \{ s_j(\text{right child}) + s_{j,t} \}$$

Begin at leaves: If leaf has char in ques, score = 0
else, score = ∞

S	A	T	G	C
A	0	3	4	9
T	3	0	2	9
G	4	2	0	9
C	9	4	4	0

$$\rightarrow s_A(v) = \min \{ s_i(u) + s_{i,A} \}$$

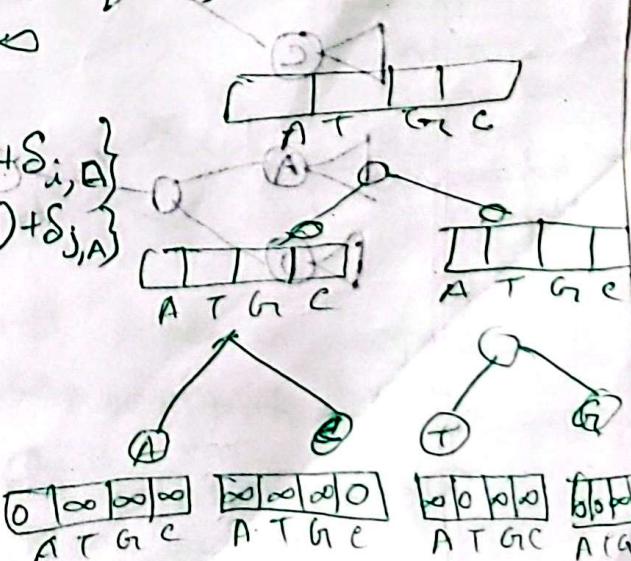
$$\min \{ s_i(u) + s_{i,A} \}$$

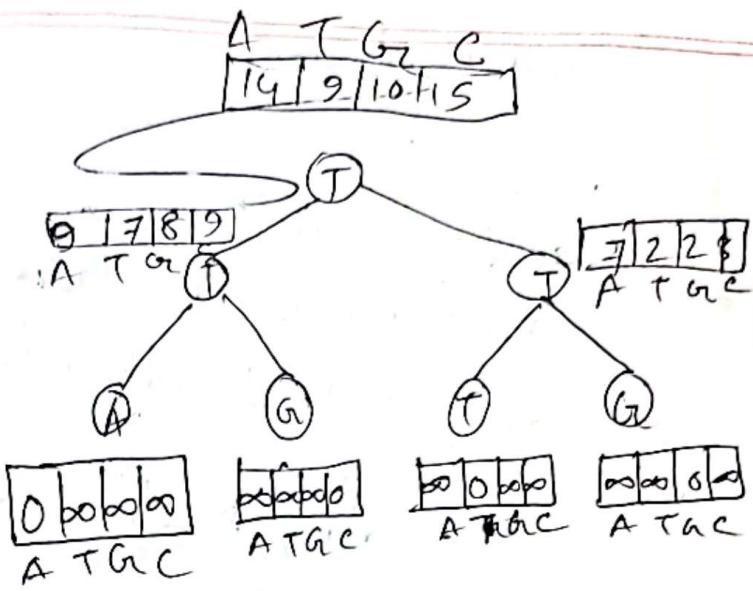
$$= s_A(u) = 0$$

$s_i(u)$ = Leftside child value

	$s_i(u)$	$s_{i,A}$	sum
A	0	0	0
T	∞	3	∞
G	∞	4	∞
C	∞	9	∞

	$s_i(u)$	$s_{i,A}$	sum
A	∞	∞	∞
T	∞	∞	∞
G	∞	∞	∞
C	∞	∞	∞



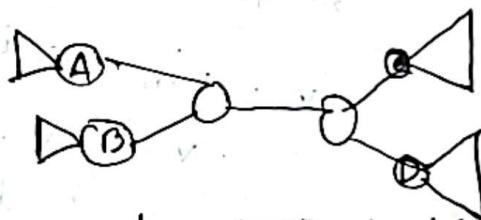
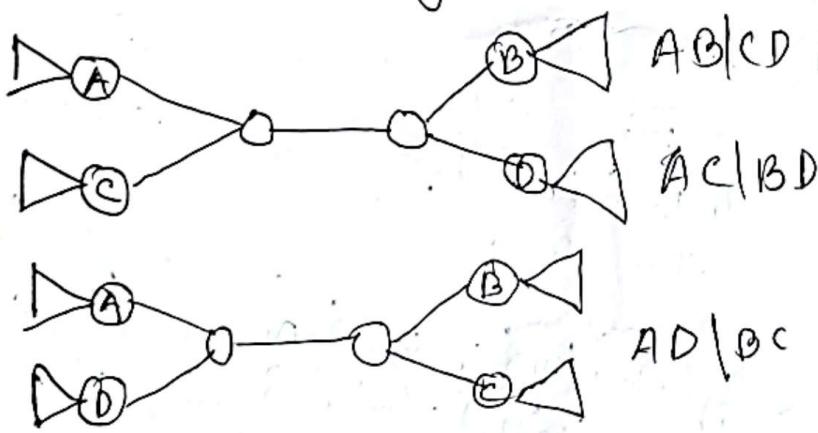


Large Parsimony

Input $n \times m$ matrix
 $m \Rightarrow n$ species, [m - char string]

Output

Nearest Neighbour
 Interchange



edge कोडे subtree
 शुल्ला Rearrange
 रात्रे।
 Rearrange करते रात्रे।
 यहाँ इसी most parsimonious
 tree प्रौङ्गली है (But
 not guaranteed)

Polytomy (intermediately)

Tree has a node that has atleast 2 child nodes

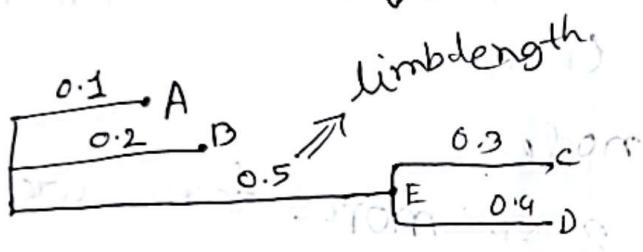
Polytomy \Rightarrow Tree has a node that has at least 3 child nodes

— Convert unrooted tree into rooted tree

Newick Notation ((a,b),(c,d,e),f,g,h)



Dissolve (a,b)



limb length



A tree with Polydony

Polytony harder calculation

opposite

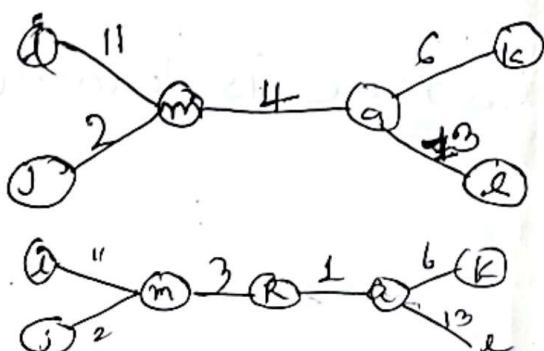
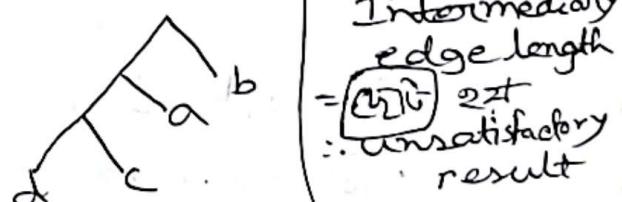
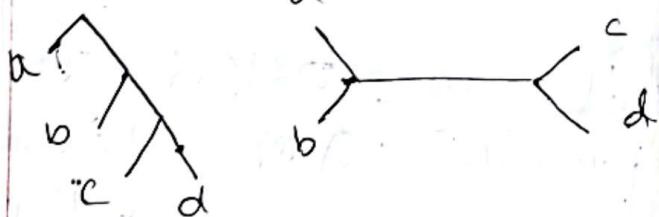


Fully resolved

edge अंतीम

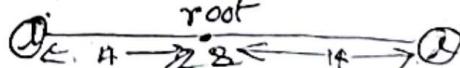
How to make a unrooted tree to rooted tree

$$((a,b), (c,d)) = (ab, cd)$$



Find longest leaf-leaf path

(1) longest path is $l = 11 + 4 + 8 = 23$



(11) Place the roof

Rooting tree

Blank node

merge

contraction



note

edge

2020 22/21

Refinement
node merge process



phylogenetic tree

একটি Rooted মর্গেন

৩মা A নড়ুন root

২মা ২২/২১=outgroup

Phylogenetic

tree

node

Consensus tree → 2 or more trees are given
on same leaf set,

How do Biologists Assemble Genomic Puzzles
from Millions of Pieces? [Graph Algorithm]

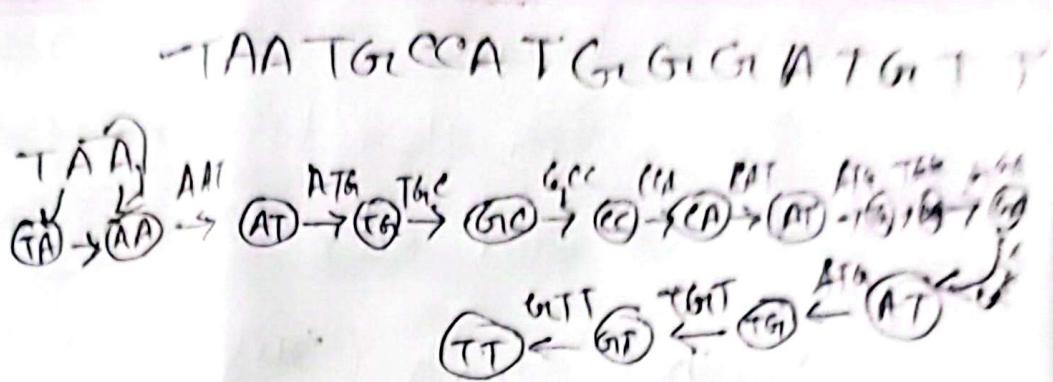
Eulerian Path → এটি Path → একটি node এখানে যেকোনো visit করা না

Hamiltonian Path → এটি edge এখানে যেকোনো visit করা না
এটি sequence এখানে Graph আবৃত্তি। [Graph theory
Page 90, 122]

Constructing de Bruijn Graph when Genome is known
[Graph theory
Page 132, 185,
190 - 194]

K-mers ($K=3$)

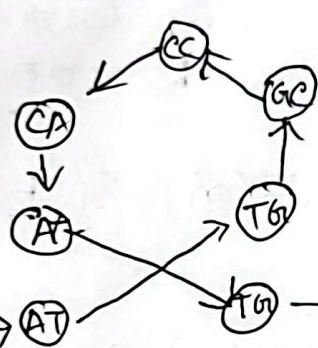
TAA
AAT
ATG
TGC
GCC
CCA
CAT
ATG
TGTG



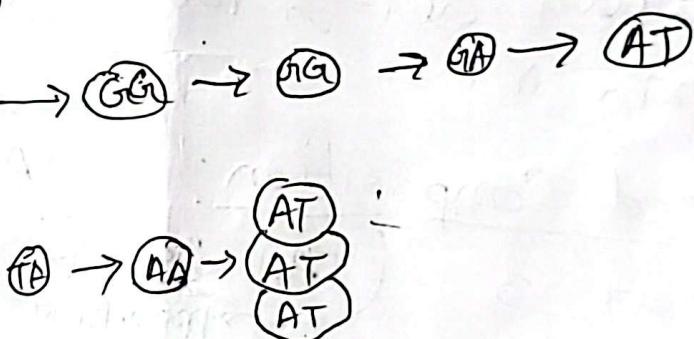
Repeated = AT

छाँटे AT का छाँटना

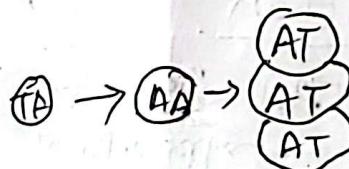
Ans.



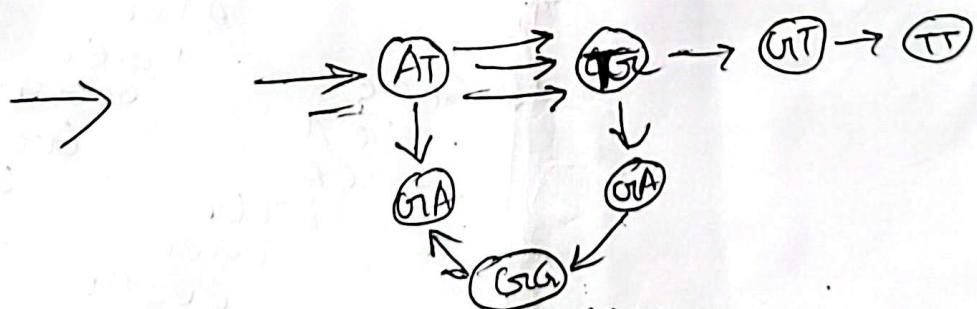
Gr Gr Gr
Gr Gr A
Gr AT
AT G
T GT
G TT



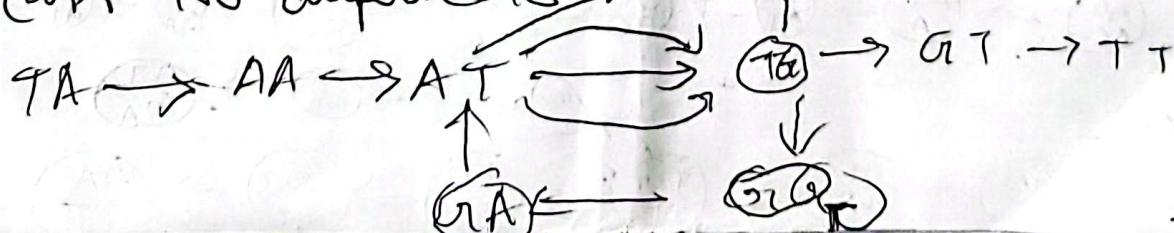
अब AT साझाकरि,



एटी + TGT साझाकरि, (Same)



(कोई No duplicate)



Reverse
can't be done
in DeBruijn

In Eulerian, রেখাগুলি একটি বিশিষ্ট ক্ষয়। ক্ষয়ের
আইন অসম্ভব Hamiltonian Path স্থাপিত করা।

$K=3$ રાને edge-3 વાળે Node-2 રાની ડ્રાઇવ.

Original string (એકો એંબે K-mers પણ હશે એવા)

(શાચાડી)

TADTGCCA

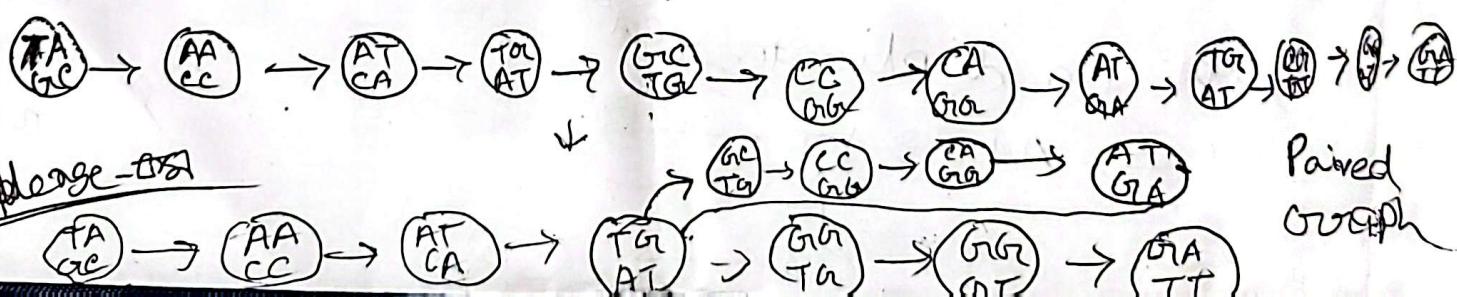
TAA
AAT
ATG
TGC
GCC
CCP

* How to Reconstruct,
Graph ଏ. ଯନାଲେ,
ଏଇ Graph ଅନୁଭାୟୀ
Reconstruct କିମ୍ବା ଏହା
ଦେଖିବା

Eulerian Path - ଏଲ୍ୟୁରିଆନ ପଥ - ମାଧ୍ୟମରେ, ଏକ Eulerian path ଦ୍ୱାରା ପରିଷ୍ରମ କରିବାକୁ ବିଶ୍ଵାସ କରିବାକୁ ପାଇଲା.

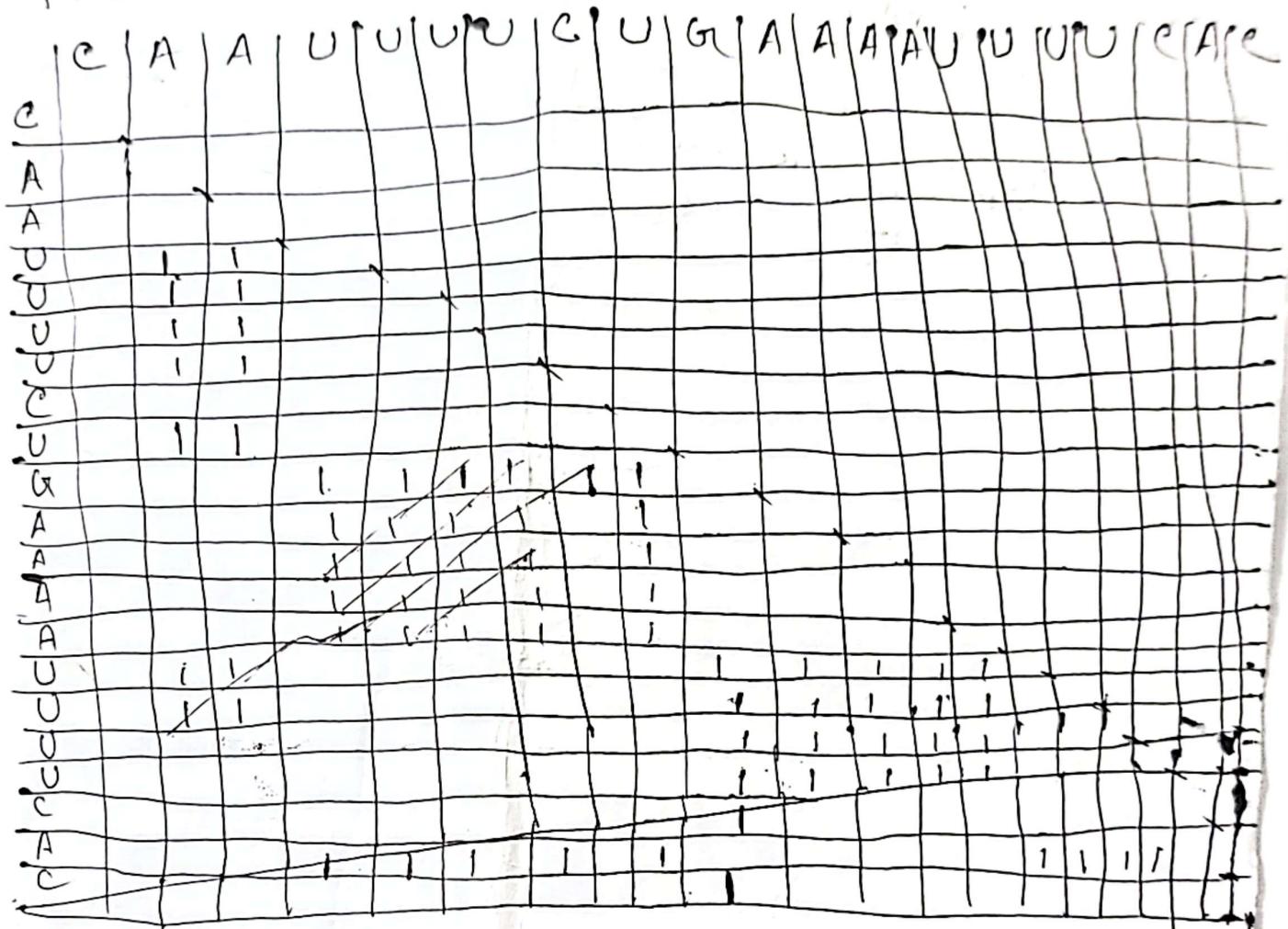
Paired Composition: TAATGC CATGGGATGTT
 $K\text{-mar} = 3$ (TAA এবং কার এটা এ রেটি গ্রাফ ক্ষেত্ৰ-
 সম্পর্কিত রেটি sequence create কৰ)

TAA	GCC
AAT	CCA
ATG	CAT
TTC	ATG
GCC	TGCG
CCA	GGGG
CAT	GGGA
ATG	CAAT
TAG	ATG
GGGG	TGCT
	GGTT



RNA Secondary Structure Prediction with Pseudocode

RNA: C A A U U U U C U G A A A A U U U U U C A C



Primary Diagonal
Lower Triangular Matrix $\text{matrix}^{(1)}$

ফিলার ম্যাট্রিস Match 2nd

Match

A U
C G

G U

* 2ndary Diagonal - ২ শাখা একাবি

(1) কাঁচা গাতে আছে

(৩টা পাই)

পাই এক stem^{index (8,10)}

1টোৱ 2ndary

ফিলার

কাঁচা নথেকে ১০- Diagonal

2ndary Diagonal কুরুব

index (6,15)-৭ (stem) :-

Helix
should
be:
connected



Staircase
(loop)

Sequence: CAAU~~UUCG~~¹⁵ GAAA UUU~~CAC~~
~~((((()))))~~

प्राप्ति

प्रैटि

Bracket

प्राप्ति

प्राप्ति

प्राप्ति

प्राप्ति (100P)

Gap loop

Stem

