Phylogenetics:
Distance Methods

COMP 571

Luay Nakhleh, Rice University

Outline

- Evolutionary models and distance corrections
- Distance-based methods

Evolutionary Models and Distance Correction

Pairwise Distances

- Calculating the distance between two sequences is important for at least two reasons:
  - it's the first step in distance-based phylogeny reconstruction
  - models of nucleotide substitution used in distance calculation form the basis of likelihood and Bayesian phylogeny reconstruction methods

Pairwise Distances

- The distance between two sequences is defined as the expected number of nucleotide substitutions per site.

Pairwise Distances

- If the evolutionary rate is constant over time, the distance will increase linearly with the time of divergence.
- A simplistic distance measure is the proportion of different sites between two sequences, known as the p distance.

The p Distance

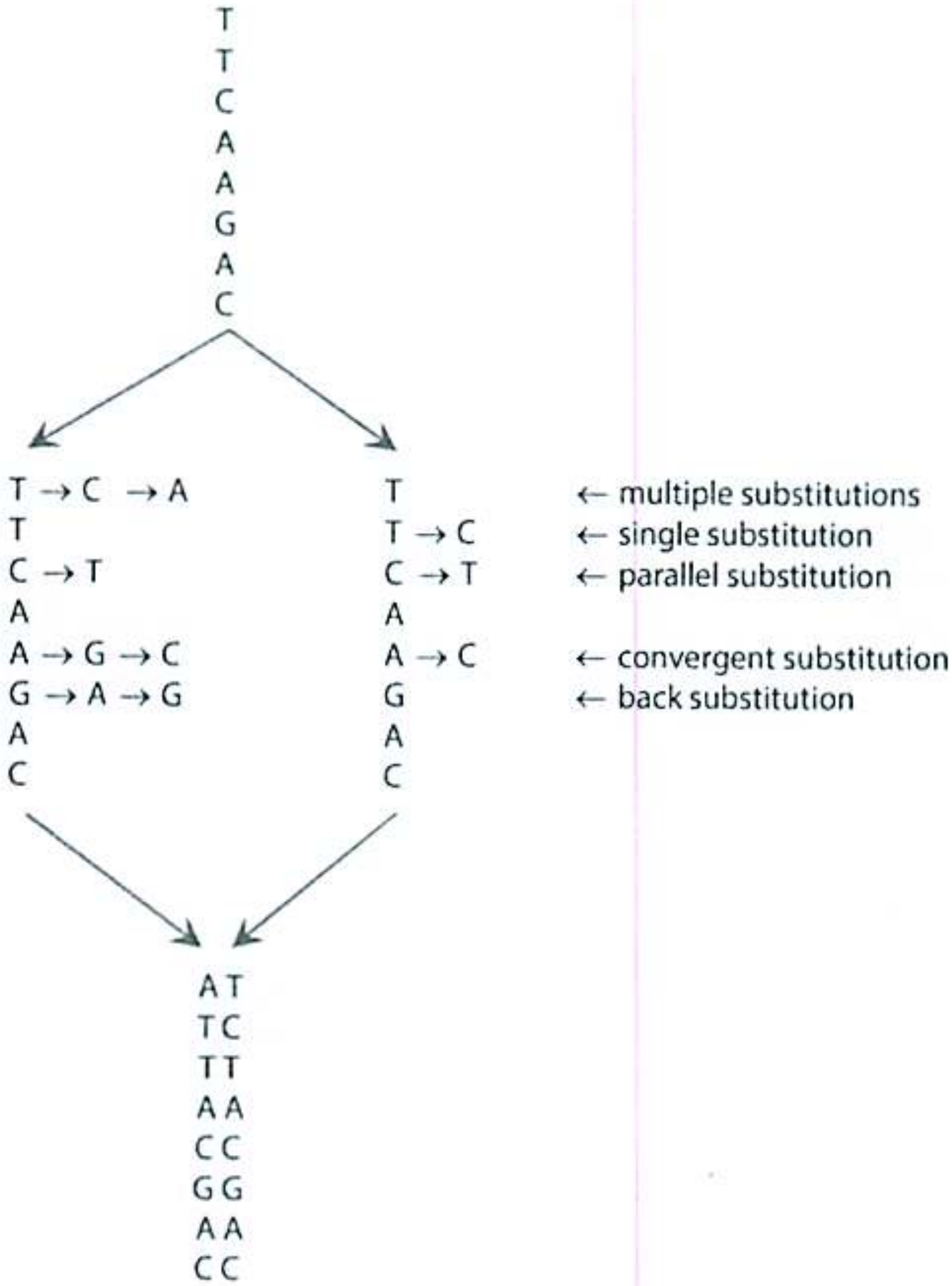
$$p = \frac{D}{L}$$

D : the number of positions at which two sequences differ

L : the length of each of the two sequences

The p Distance

- Due to back or parallel substitutions, the p distance often underestimates the number of substitutions that have occurred (the p distance works fine for very similar sequences, say, with $p < 5\%$).



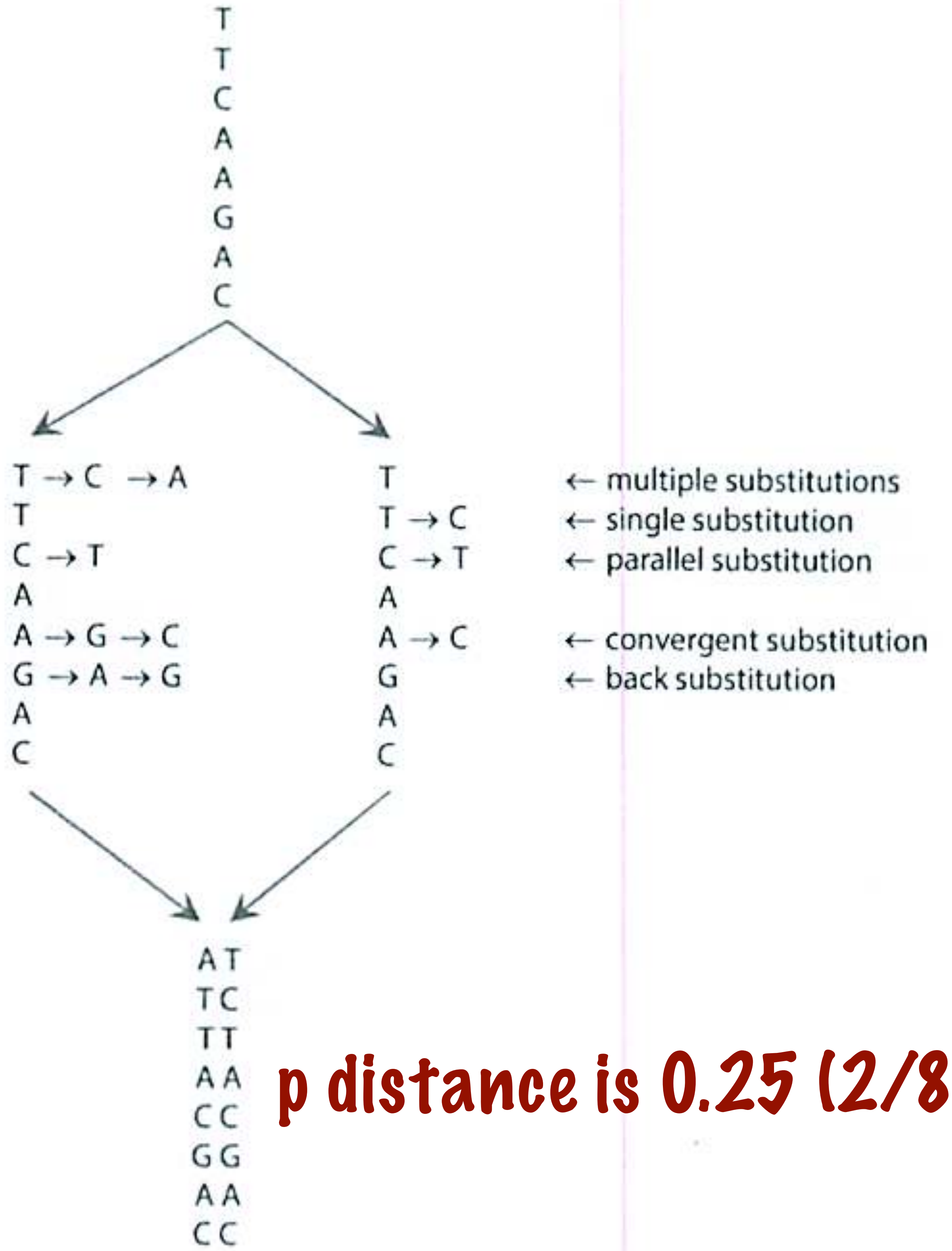
← multiple substitutions

← single substitution

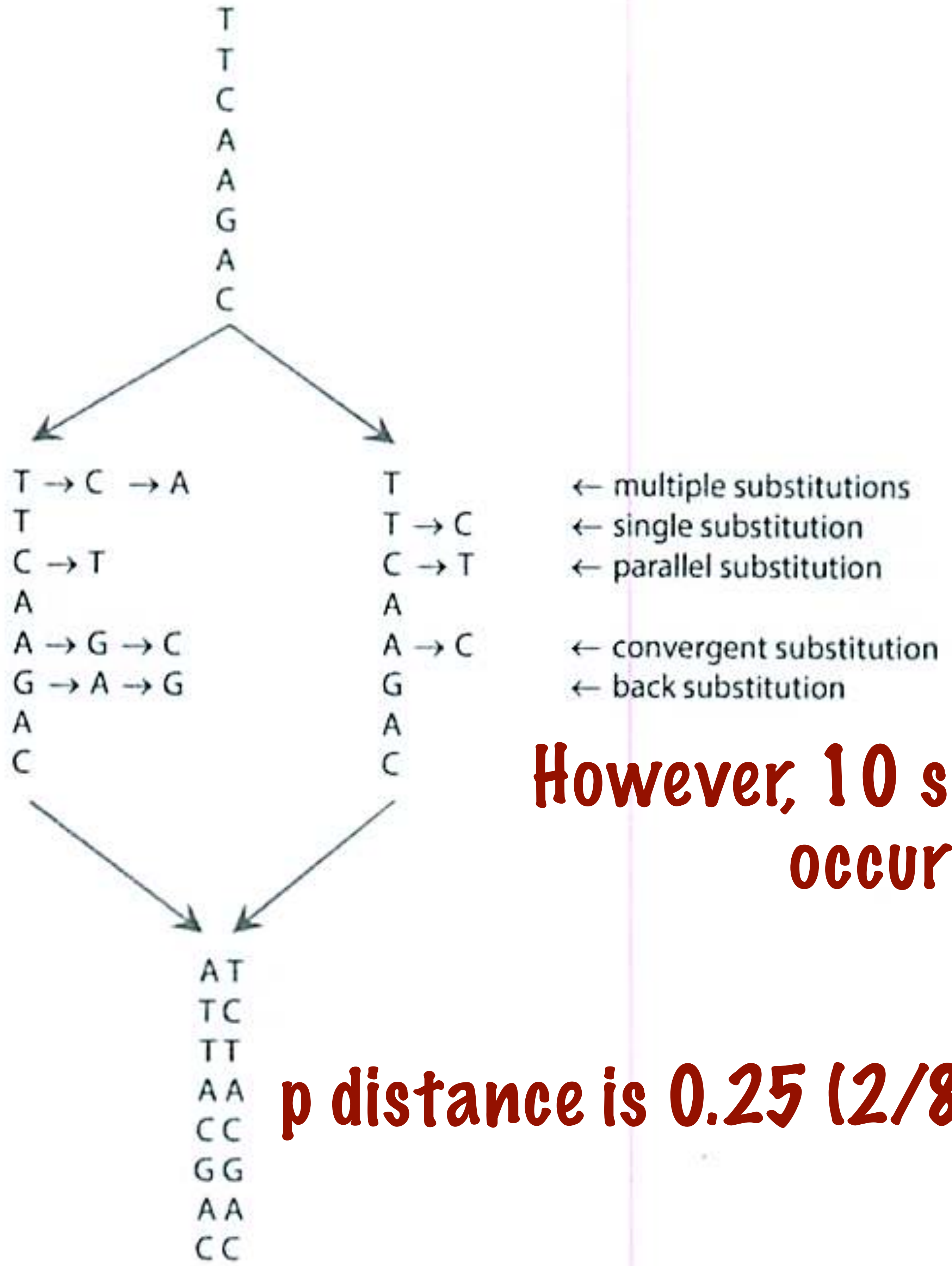
← parallel substitution

← convergent substitution

← back substitution



p distance is 0.25 (2/8)



However, 10 substitutions occurred!

p distance is 0.25 (2/8)

Models of Sequence Evolution

- To estimate the "actual" number of substitutions, we need a probabilistic model to describe changes between nucleotides over evolutionary time.
- Continuous-time Markov chains are commonly used for this purpose.

Models of Sequence Evolution

- The nucleotide sites are assumed to be evolving independently of each other.
- Substitutions at any particular site are described by a Markov chain, with the four nucleotides to be the states of the chain.

Models of Sequence Evolution

- Besides the Markovian property (next state depends only on the current state), we often place constraints on substitution rates between nucleotides, leading to different models of nucleotide substitution.

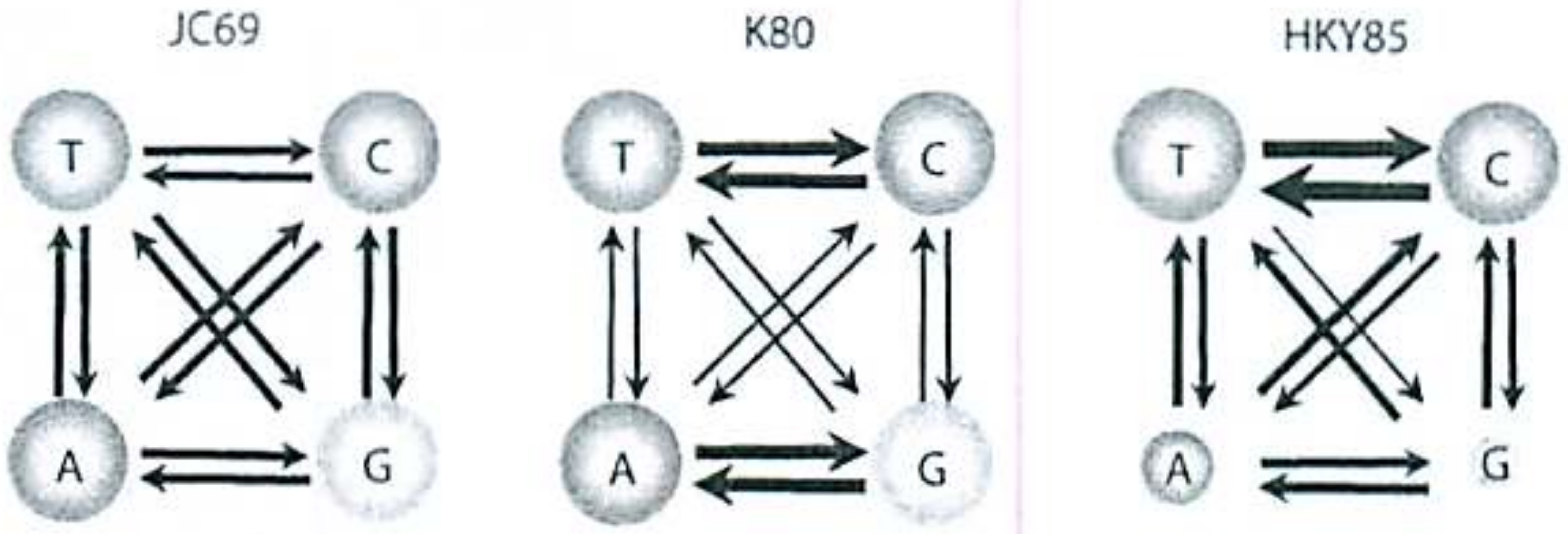


Fig. 1.2 Relative substitution rates between nucleotides under three Markov chain models of nucleotide substitution: JC69, K80, and HKY85. The thickness of the lines represents the substitution rates, while the sizes of the circles represent the steady-state distribution.

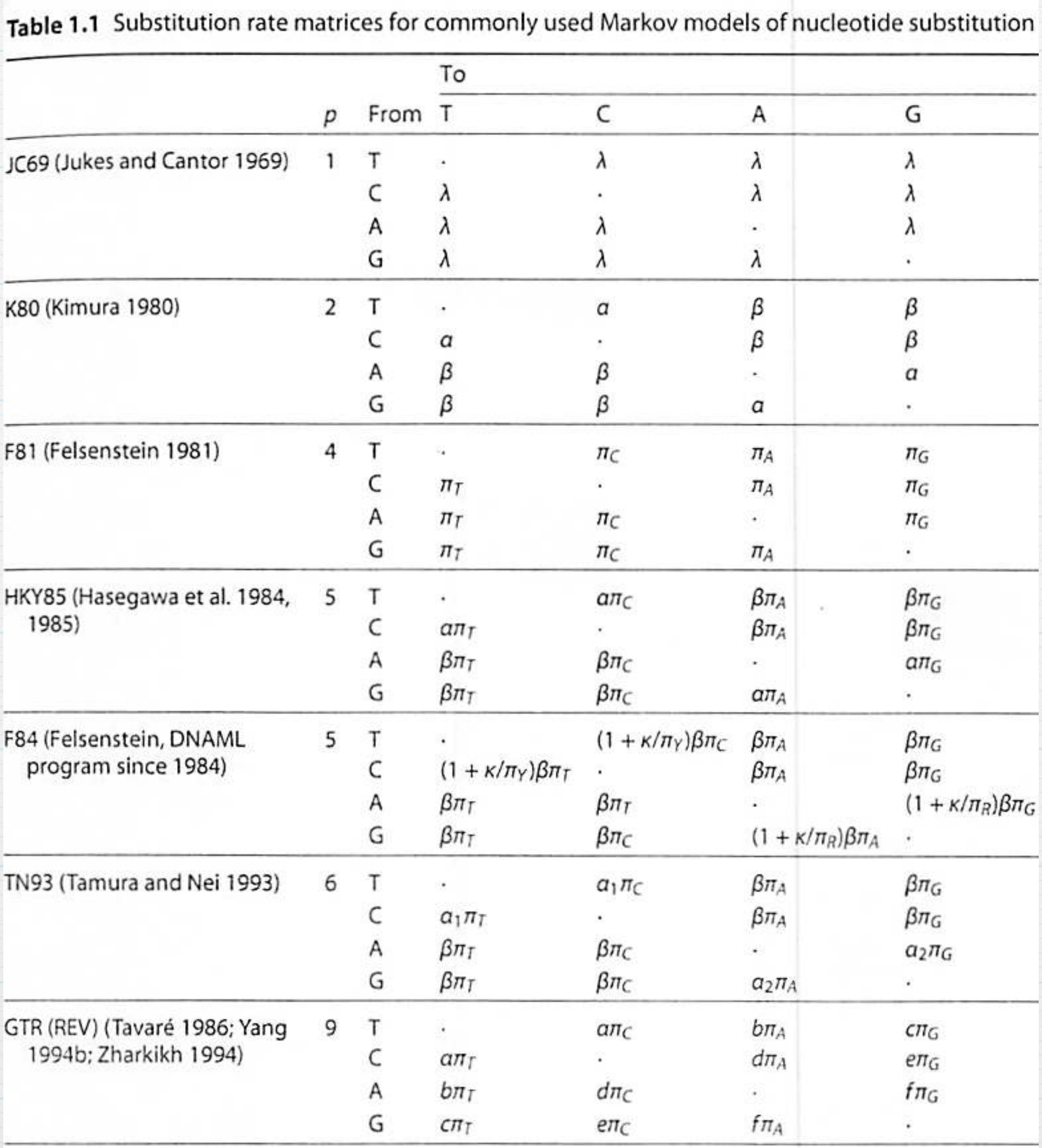


Table 1.1 Substitution rate matrices for commonly used Markov models of nucleotide substitution

			To			
	p	From	T	C	A	G
JC69 (Jukes and Cantor 1969)	1	T	·	λ	λ	λ
		C	λ	·	λ	λ
		A	λ	λ	·	λ
		G	λ	λ	λ	·
K80 (Kimura 1980)	2	T	·	α	β	β
		C	α	·	β	β
		A	β	β	·	α
		G	β	β	α	·
F81 (Felsenstein 1981)	4	T	·	π_C	π_A	π_G
		C	π_T	·	π_A	π_G
		A	π_T	π_C	·	π_G
		G	π_T	π_C	π_A	·
HKY85 (Hasegawa et al. 1984, 1985)	5	T	·	$\alpha\pi_C$	$\beta\pi_A$	$\beta\pi_G$
		C	$\alpha\pi_T$	·	$\beta\pi_A$	$\beta\pi_G$
		A	$\beta\pi_T$	$\beta\pi_C$	·	$\alpha\pi_G$
		G	$\beta\pi_T$	$\beta\pi_C$	$\alpha\pi_A$	·
F84 (Felsenstein, DNAML program since 1984)	5	T	·	$(1+\kappa/\pi_Y)\beta\pi_C$	$\beta\pi_A$	$\beta\pi_G$
		C	$(1+\kappa/\pi_Y)\beta\pi_T$	·	$\beta\pi_A$	$\beta\pi_G$
		A	$\beta\pi_T$	$\beta\pi_T$	·	$(1+\kappa/\pi_R)\beta\pi_G$
		G	$\beta\pi_T$	$\beta\pi_C$	$(1+\kappa/\pi_R)\beta\pi_A$	·
TN93 (Tamura and Nei 1993)	6	T	·	$\alpha_1\pi_C$	$\beta\pi_A$	$\beta\pi_G$
		C	$\alpha_1\pi_T$	·	$\beta\pi_A$	$\beta\pi_G$
		A	$\beta\pi_T$	$\beta\pi_C$	·	$\alpha_2\pi_G$
		G	$\beta\pi_T$	$\beta\pi_C$	$\alpha_2\pi_A$	·
GTR (REV) (Tavaré 1986; Yang 1994b; Zharkikh 1994)	9	T	·	$a\pi_C$	$b\pi_A$	$c\pi_G$
		C	$a\pi_T$	·	$d\pi_A$	$e\pi_G$
		A	$b\pi_T$	$d\pi_C$	·	$f\pi_G$
		G	$c\pi_T$	$e\pi_C$	$f\pi_A$	·

The Jukes-Cantor (JC) Model

- Some evolutionary models have been constructed specifically for nucleotide sequences
- One of the simplest such models is that Jukes-Cantor (JC) model
- It assumes all sites are independent and have identical mutation rates
- Further, it assumes all possible nucleotide substitutions occur at the same rate α per unit time

The Jukes-Cantor (JC) Model

- A matrix Q can represent the substitution rates:

	A	C	G	T
A	-3α	α	α	α
C	α	-3α	α	α
G	α	α	-3α	α
T	α	α	α	-3α

math requirement: each row sums to 0

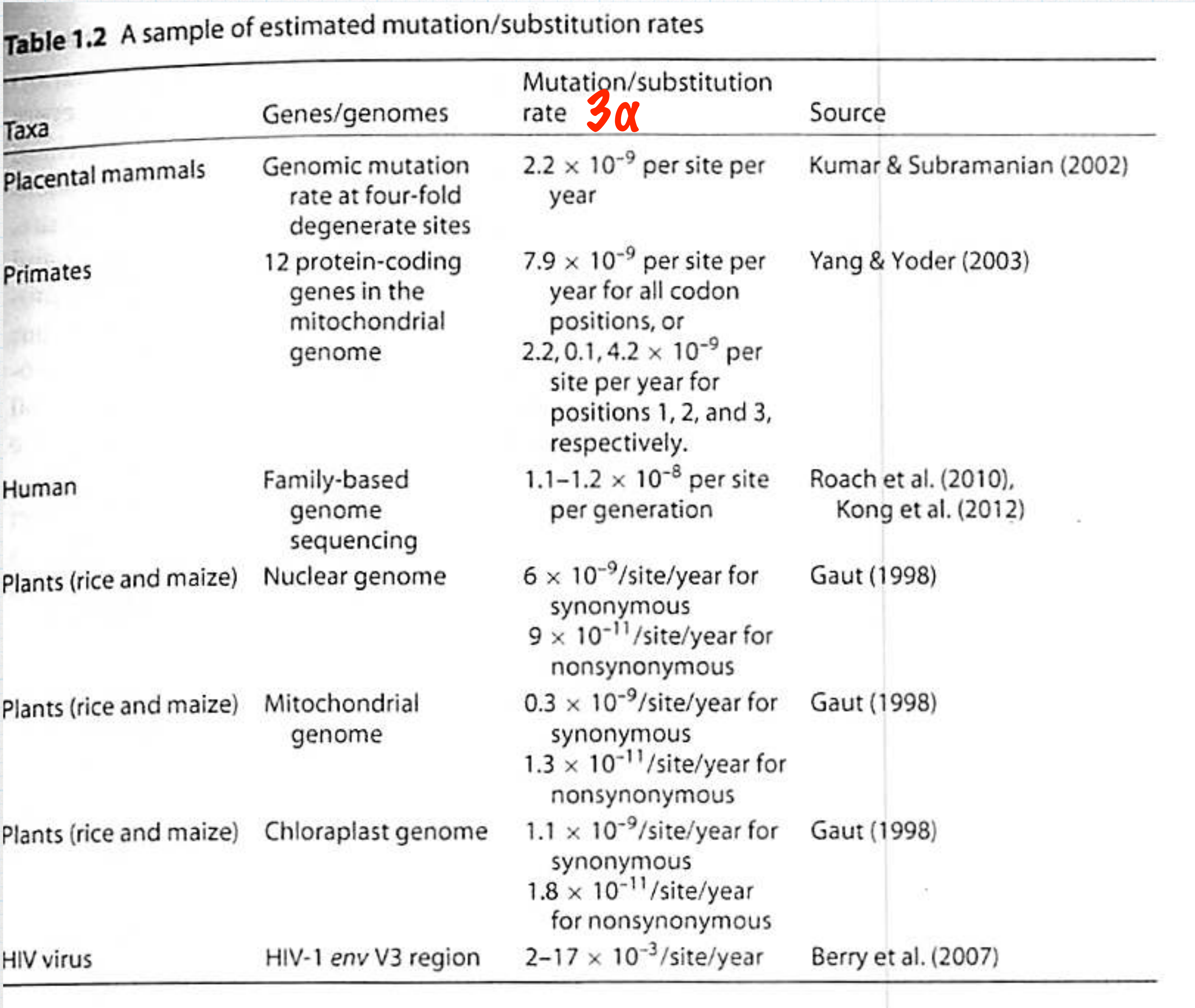


Table 1.2 A sample of estimated mutation/substitution rates

Taxa	Genes/genomes	Mutation/substitution rate 3α	Source
Placental mammals	Genomic mutation rate at four-fold degenerate sites	2.2 × 10^{-9} per site per year	Kumar & Subramanian (2002)
Primates	12 protein-coding genes in the mitochondrial genome	7.9 × 10^{-9} per site per year for all codon positions, or 2.2, 0.1, 4.2 × 10^{-9} per site per year for positions 1, 2, and 3, respectively.	Yang & Yoder (2003)
Human	Family-based genome sequencing	1.1–1.2 × 10^{-8} per site per generation	Roach et al. (2010), Kong et al. (2012)
Plants (rice and maize)	Nuclear genome	6 × 10^{-9}/site/year for synonymous 9 × 10^{-11}/site/year for nonsynonymous	Gaut (1998)
Plants (rice and maize)	Mitochondrial genome	0.3 × 10^{-9}/site/year for synonymous 1.3 × 10^{-11}/site/year for nonsynonymous	Gaut (1998)
Plants (rice and maize)	Chloraplast genome	1.1 × 10^{-9}/site/year for synonymous 1.8 × 10^{-11}/site/year for nonsynonymous	Gaut (1998)
HIV virus	HIV-1 *env* V3 region	2–17 × 10^{-3}/site/year	Berry et al. (2007)

The Jukes-Cantor (JC) Model

- To relate the Markov chain model to sequence data, we need to calculate the probability that given the nucleotide i at a site now, it will become nucleotide j time t later.
- This is known as the transition probability, denote by $p_{ij}(t)$.

The Jukes-Cantor (JC) Model

- Continuous-time Markov chain theory tells us that

$$P(t) = e^{Qt} = I + Qt + \frac{1}{2!}(Qt)^2 + \frac{1}{3!}(Qt)^3 + \cdots$$

The Jukes-Cantor (JC) Model

- For Jukes-Cantor, this results in

$$p_{ii}(t) = \frac{1}{4} + \frac{3}{4}e^{-4\alpha t}$$

$$p_{ij}(t) = \frac{1}{4} - \frac{1}{4}e^{-4\alpha t} \qquad i \neq j$$

We always estimate αt; it is impossible to tell α and t values separately from two sequences!

The Jukes-Cantor (JC) Model

- Given a sequence where every nucleotide is i, then the proportion of nucleotide j after time period t is $p_{ij}(t)$.
- To get αt, solve $p = 3\left(\frac{1}{4} - \frac{1}{4}e^{-4\alpha t}\right)$
- $3\alpha t$ mutations would be expected during a time t for each sequence site on each sequence (call this d_{JC})
- this yields

$$d_{JC} = -\frac{3}{4}\ln\left(1 - \frac{4}{3}p\right)$$

The Jukes-Cantor (JC) Model

- This corrected distance, d_{JC}, can be obtained as

$$d_{JC} = -\frac{3}{4}\ln\left(1 - \frac{4}{3}p\right)$$

- To obtain a value for the corrected distance, substitute p with the observed proportion of site differences in the alignment

The Kimura 2-Parameter Model

- One “improvement” over the JC model involves distinguishing between rates of transitions and transversions

- Rates α and β are assigned to transitions and transversions, respectively

- When this is the only modification made, this amounts to the Kimura two-parameter (K2P) model, and has the rate matrix

	A	C	G	T
A	$-2\beta-\alpha$	β	α	β
C	β	$-2\beta-\alpha$	β	α
G	α	β	$-2\beta-\alpha$	β
T	β	α	β	$-2\beta-\alpha$

The Kimura 2-Parameter Model

- The K2P model results in a corrected distance, d_{K2P}, given by

$$d_{K2P} = -\frac{1}{2}\ln(1-2P-Q) - \frac{1}{4}\ln(1-2Q)$$

where P and Q are the observed fractions of aligned sites whose two bases are related by a transition or transversion mutation, respectively

- Notice that the p-distance, p, equals P+Q
- The transition/transversion ratio, R, is defined as $\alpha/2\beta$

The HKY85 Model

- Hasegawa, Kishino, and Yano (1985)
- Allows for any base composition $\pi_A:\pi_C:\pi_G:\pi_T$
- Has the rate matrix

$$\begin{array}{c|cccc}
 & A & C & G & T \\
\hline
A & (-2\beta-\alpha)\,\pi_A & \beta\pi_C & \alpha\pi_G & \beta\pi_T \\
C & \beta\pi_A & (-2\beta-\alpha)\pi_C & \beta\pi_G & \alpha\pi_T \\
G & \alpha\pi_A & \beta\pi_C & (-2\beta-\alpha)\pi_G & \beta\pi_T \\
T & \beta\pi_A & \alpha\pi_C & \beta\pi_G & (-2\beta-\alpha)\pi_T
\end{array}$$

Choice of a Model of Evolution

Model	Base composition	R=1?	Identical transition rates?	Identical transversion rates?	Reference
JC	1:1:1:1	no	yes	yes	Jukes and Cantor (1969)
F81	variable	no	yes	yes	Felsenstein (1981)
K2P	1:1:1:1	yes	yes	yes	Kimura (1980)
HKY85	variable	yes	no	no	Hasegawa et al. (1985)
TN	variable	yes	no	yes	Tamura and Nei (1993)
K3P	variable	yes	no	yes	Kimura (1981)
SYM	1:1:1:1	yes	no	no	Zharkikh (1994)
GTR	variable	yes	no	no	Rodriguez et al. (1990)

Rates Across Sites

- To allow for varying mutation rates across sites, the Gamma distribution can be applied
- If it is applied to the JC model with Γ parameter a, the corrected distance equation becomes

$$d_{JC+\Gamma} = \frac{3}{4}a\left[\left(1 - \frac{4}{3}p\right)^{-\frac{1}{a}} - 1\right]$$

Models of Protein-sequence Evolution

- Models that we just described can be modified to apply to protein sequences

- For example, the JC distance correction for protein sequences is

$$d_{JCprot} = -\frac{19}{20} \ln \left(1 - \frac{20}{19} p \right)$$

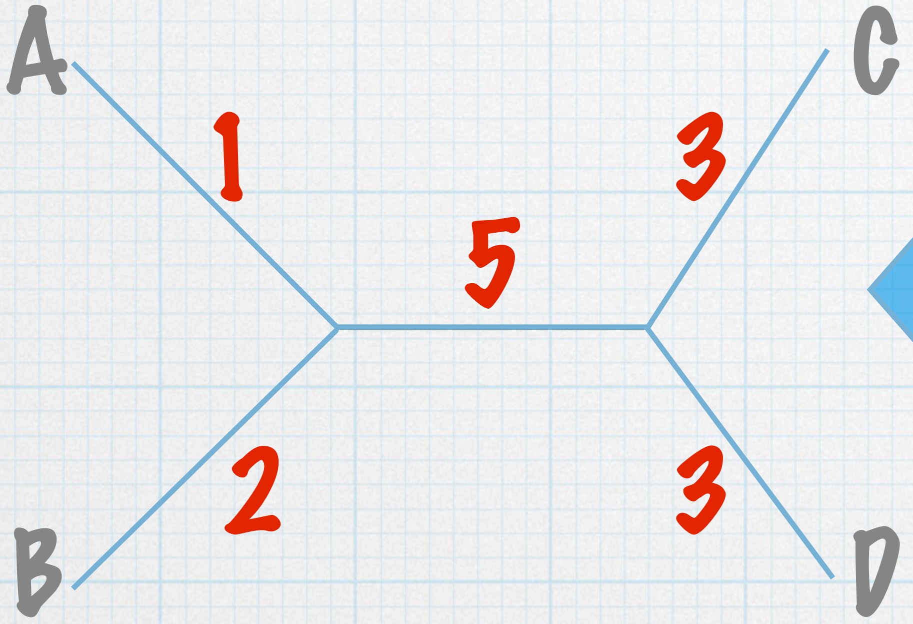
- However, the more common practice is to use empirical matrices, such as the JTT (Jones, Taylor, and Thornton) matrix

Distance-based Methods

Distance-based Methods

- Reconstruct a phylogenetic tree for a set of sequences on the basis of their pairwise evolutionary distances
- Derivation of these distances involve equations such as the ones we saw before (distance correction formulas)
- Problems with distances include
  - Wrong alignment leads to incorrect distances
  - Assumptions in the evolutionary models used may not hold
  - Formulas for computing distances are exact only in the limit of infinitely long sequences, which means the true evolutionary distances cannot always be recovered exactly

Additivity



	A	B	C	D
A	0	3	9	9
B		0	10	10
C			0	6
D				0

The Distance-based Phylogeny Problem

- Input: Matrix M of pairwise distances among species S
- Output: Tree T leaf-labeled with S, and consistent with M

The Least-squares Problem

- Input: Distance matrix D, and weights matrix w
- Output: Tree T with branch lengths that minimizes

$$LS(T) = \sum_{i=1}^{n} \sum_{j \neq i} w_{ij} (D_{ij} - \underset{\uparrow}{d_{ij}})^2$$

The distances defined by the tree T

Distance-based Methods

- The least-squares problem is NP-complete
- We will describe three polynomial-time heuristics
  - Unweighted pair-group method using arithmetic averages (UPGMA)
  - Fitch-Margoliash
  - Neighbor joining

The UPGMA Method

- Assumes a constant molecular clock, and a consequence, infers ultrametric trees
- Main idea: the two sequences with the shortest evolutionary distance between them are assumed to have been the last to diverge, and must therefore have arisen from the most recent internal node in the tree. Furthermore, their branches must be on equal length, and so must be half their distance

The UPGMA Method

1. Initialization
    1. n clusters, one per taxon
2. Iteration
    1. Find two clusters X and Y whose distance is smallest
    2. Create a new cluster XY that is the union of the two clusters X and Y, and add it to the set of clusters
    3. Remove the two clusters X and Y from the set of clusters
    4. Compute the distance between XY and every other cluster in the set
    5. Repeat until one cluster is left

The UPGMA Method

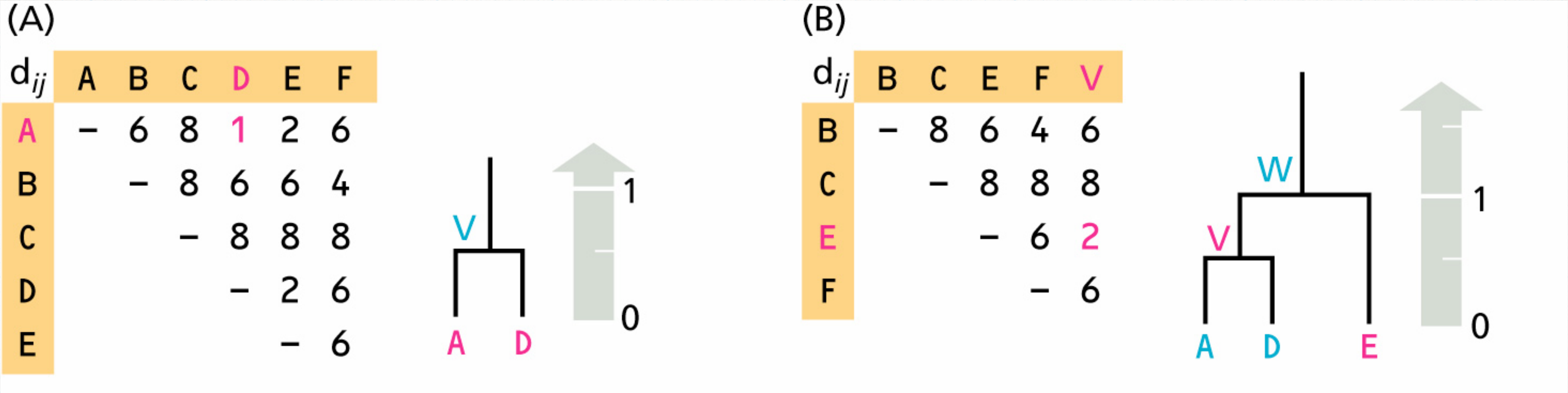
Q1: What is the distance between two clusters X and Y?

$$d_{XY} = \frac{1}{N_X N_Y} \sum_{i \in X, j \in Y} d_{ij}$$

Q2: When creating a new cluster Z, how do we compute its distance to every other cluster, W?

$$d_{ZW} = \frac{N_X d_{XW} + N_Y d_{YW}}{N_X + N_Y}$$

UPGMA: An Example



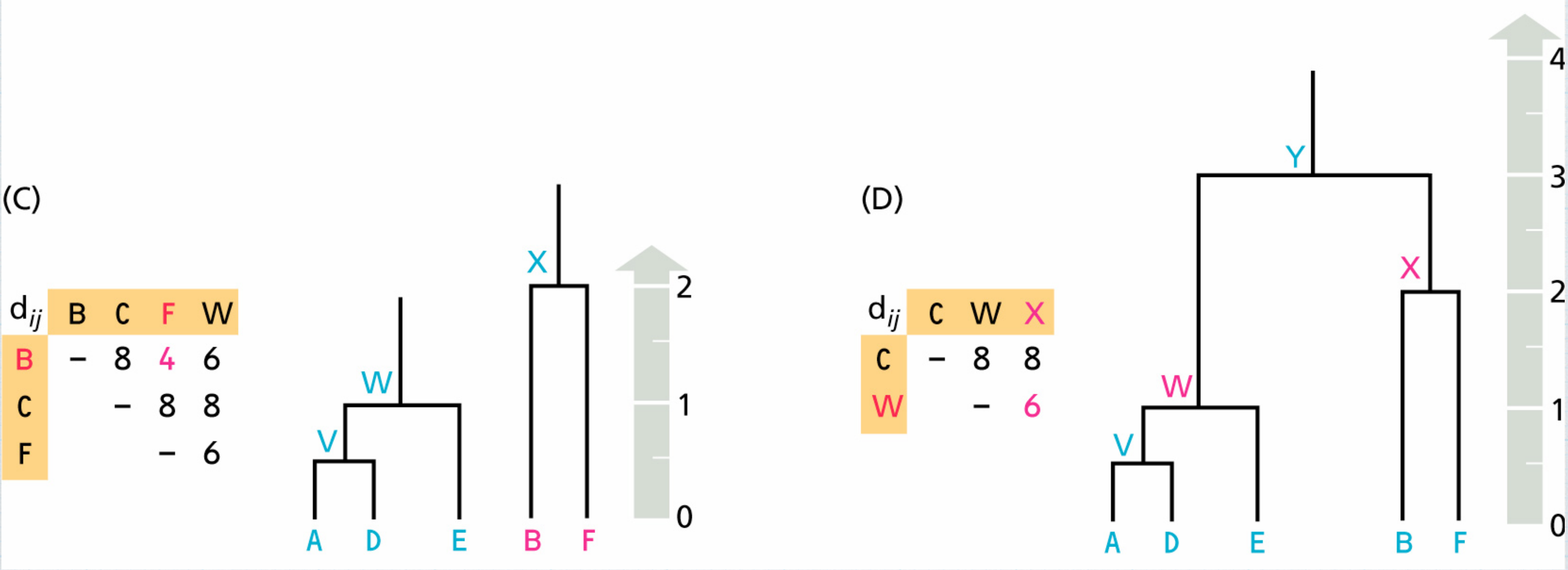
(A)

d_{ij}	A	B	C	D	E	F
A	–	6	8	1	2	6
B		–	8	6	6	4
C			–	8	8	8
D				–	2	6
E					–	6

(B)

d_{ij}	B	C	E	F	V
B	–	8	6	4	6
C		–	8	8	8
E			–	6	2
F				–	6

UPGMA: An Example



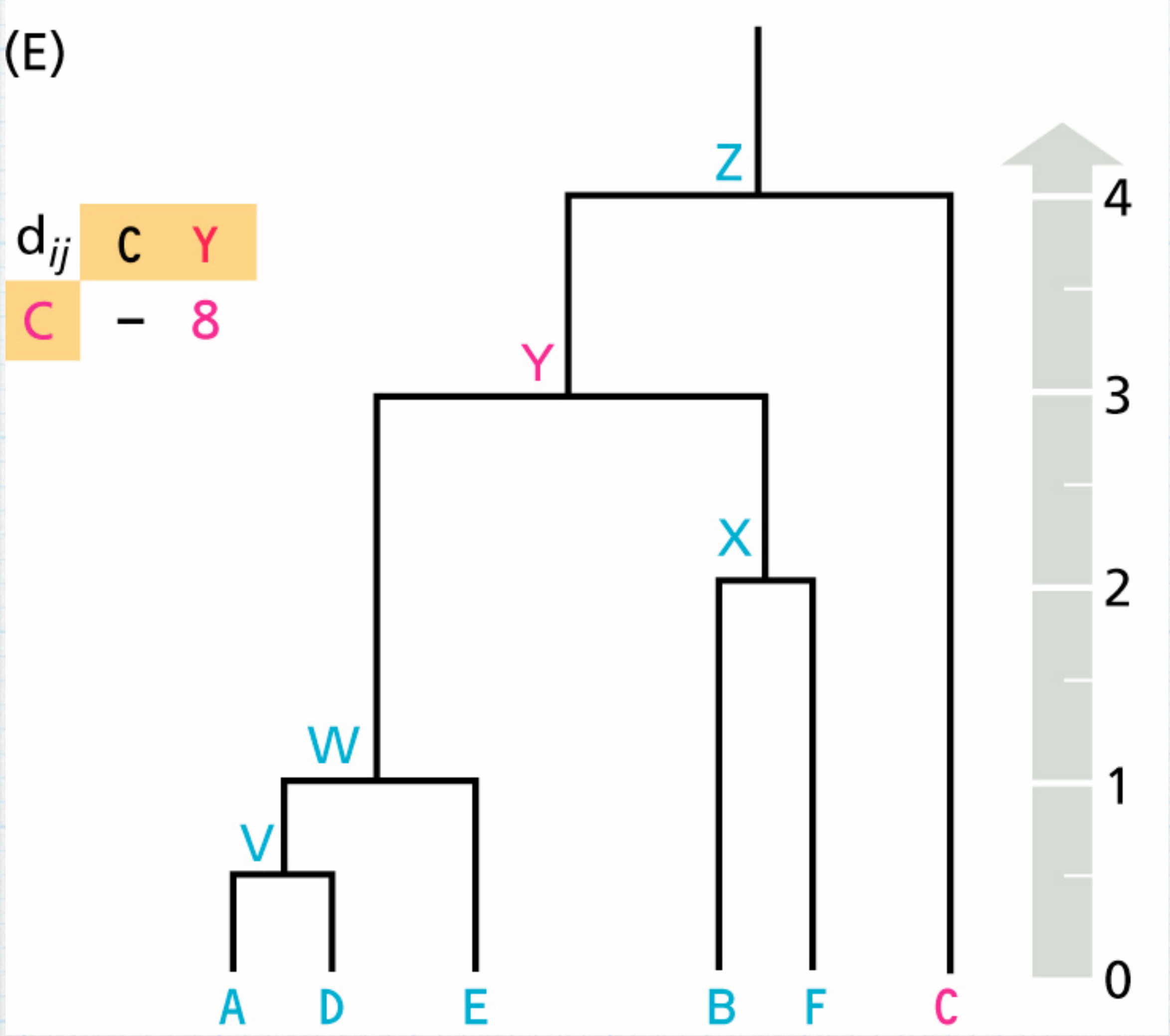
(C)

d_{ij}	B	C	F	W
B	–	8	4	6
C		–	8	8
F			–	6

(D)

d_{ij}	C	W	X
C	–	8	8
W		–	6

UPGMA: An Example



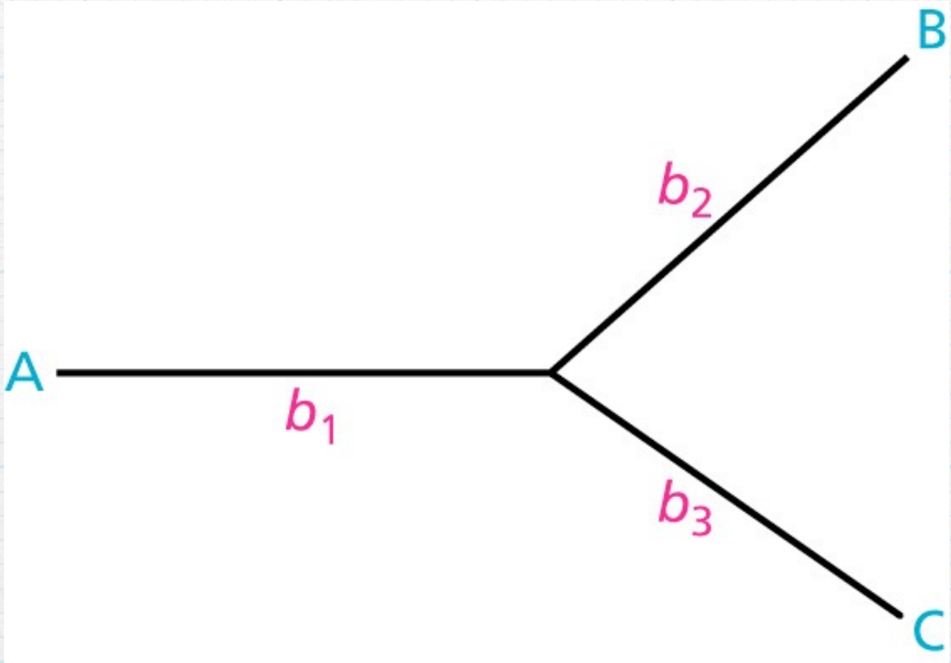
(E)

d_{ij}	C	Y
C	–	8

The Fitch-Margoliash Method

The method is based on the analysis of a three-leaf tree (triplet)

$$d_{AB} = b_1 + b_2 \qquad d_{AC} = b_1 + b_3 \qquad d_{BC} = b_2 + b_3$$



$$b_1 = \frac{1}{2}(d_{AB} + d_{AC} - d_{BC})$$

$$b_2 = \frac{1}{2}(d_{AB} + d_{BC} - d_{AC})$$

$$b_3 = \frac{1}{2}(d_{AC} + d_{BC} - d_{AB})$$

The Fitch-Margoliash Method

- Trees with more than three leaves can be generated in a stepwise fashion similar to that used in UPGMA
- At every stage, three clusters are defined, with all sequences belonging to one of the clusters
- The distance between clusters is defined by a simple arithmetic average of the distances between sequences in the different clusters

The Fitch-Margoliash Method

- At the start of each step, we have a list of sequences not yet part of the growing tree and of clusters representing each part of the growing tree

- The distances between all these sequences and clusters are calculated, and the two most closely related are selected as the first two clusters of a three-leaf tree

- A third cluster is defined that contains the remainder of the sequences, and the distances to the other two are calculated

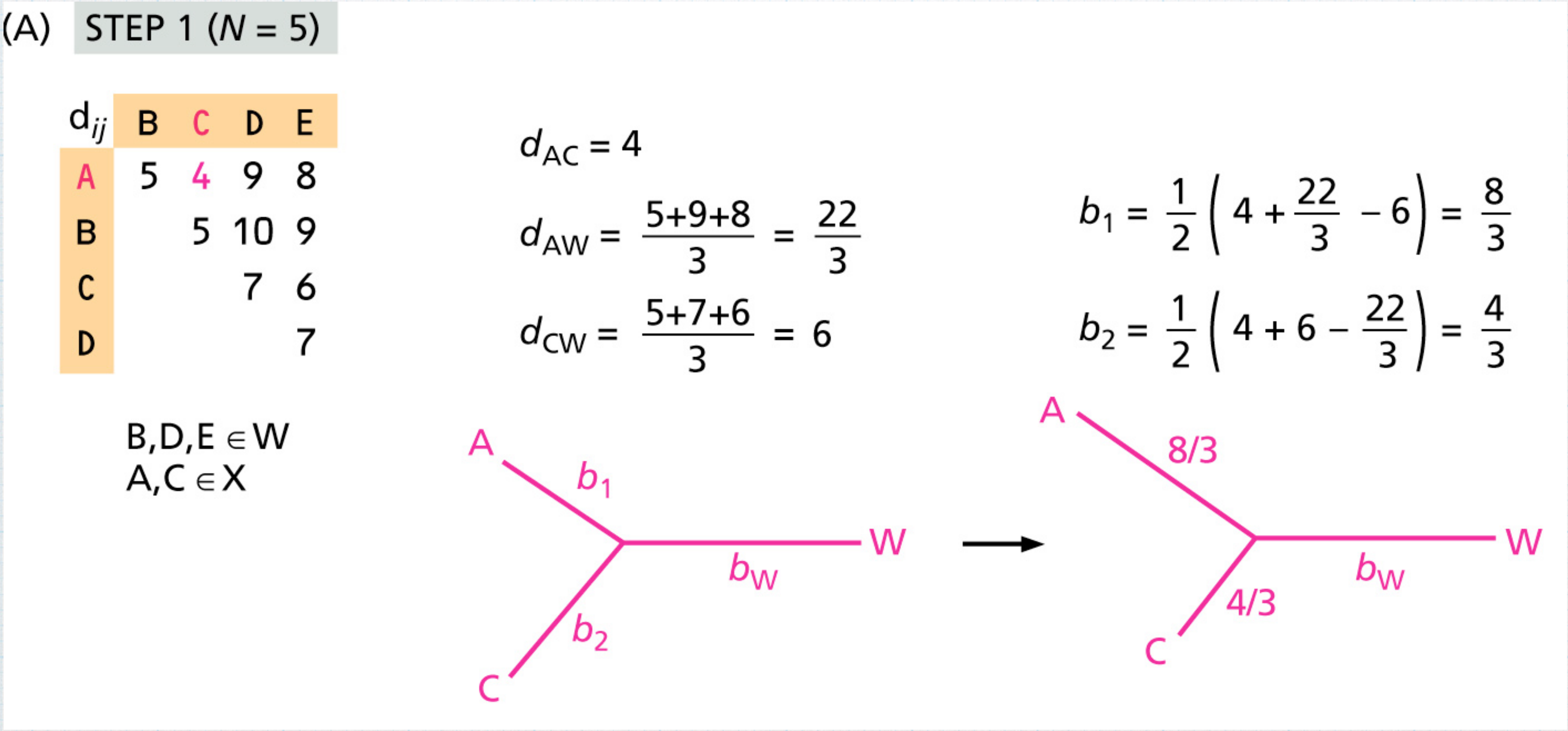
The Fitch-Margoliash Method

- Using the equations described, one can then determine the branch lengths from this third cluster to the other two clusters and the location of the internal node that connects them
- These two clusters are then combined into a single cluster with distances to other sequences again defined by simple averages

The Fitch-Margoliash Method

- There is now one less sequence (cluster) to incorporate into the growing tree
- By repetition of these steps, this technique is able to generate a single tree in a similar manner to UPGMA
- The trees produced by UPGMA and Fitch-Margoliash are identical in terms of topology, yet differ in the branch lengths assigned

Fitch-Margoliash: An Example



(A) STEP 1 ($N = 5$)

d_{ij}	B	C	D	E
A	5	4	9	8
B		5	10	9
C			7	6
D				7

B,D,E $\in$ W
A,C $\in$ X

$$d_{AC} = 4$$

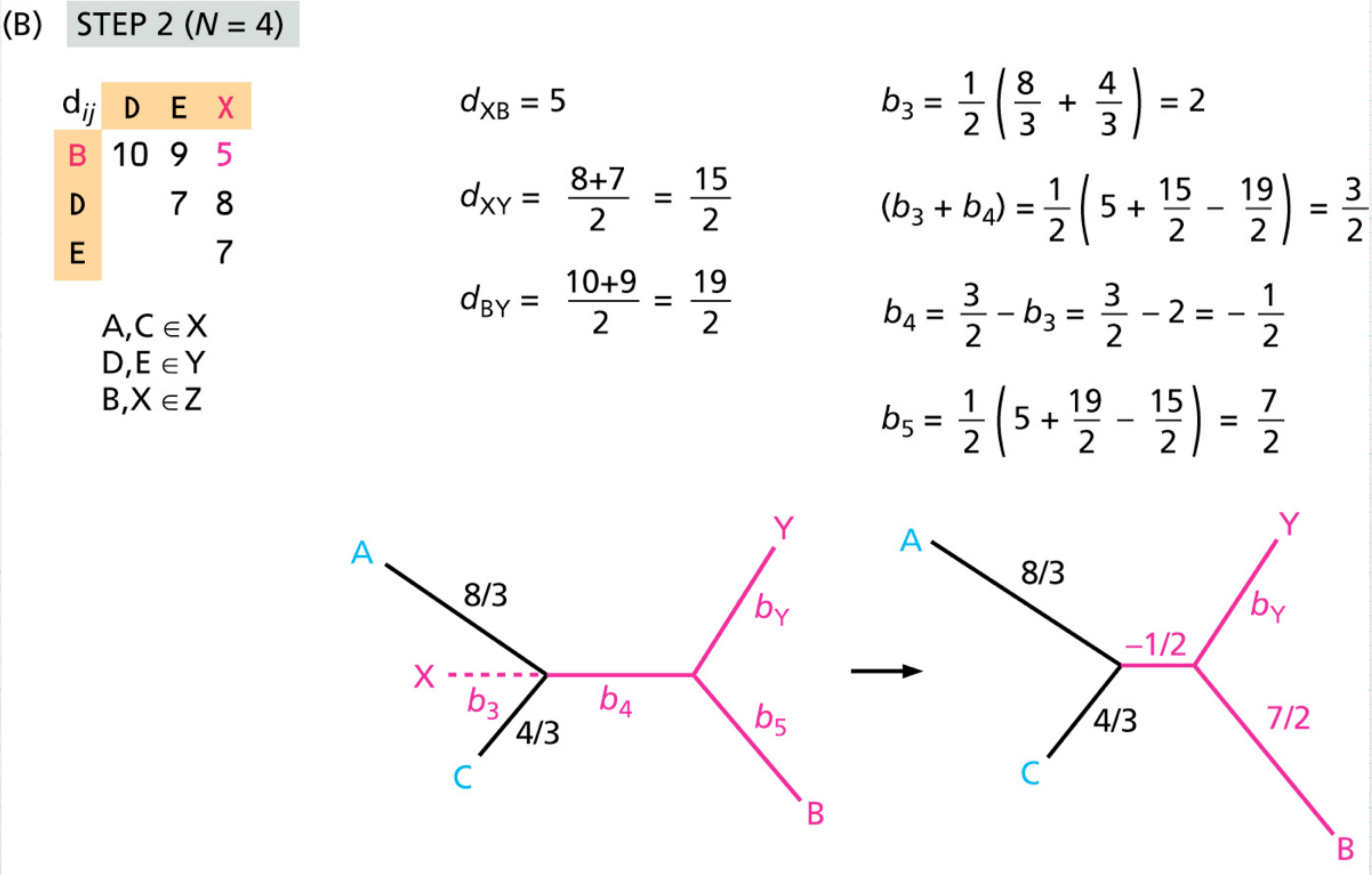
$$d_{AW} = \frac{5+9+8}{3} = \frac{22}{3}$$

$$d_{CW} = \frac{5+7+6}{3} = 6$$

$$b_1 = \frac{1}{2}\left(4 + \frac{22}{3} - 6\right) = \frac{8}{3}$$

$$b_2 = \frac{1}{2}\left(4 + 6 - \frac{22}{3}\right) = \frac{4}{3}$$

Fitch-Margoliash: An Example



(B) STEP 2 ($N = 4$)

d_{ij}	D	E	X
B	10	9	5
D		7	8
E			7

A,C $\in$ X
D,E $\in$ Y
B,X $\in$ Z

$$d_{XB} = 5$$

$$d_{XY} = \frac{8+7}{2} = \frac{15}{2}$$

$$d_{BY} = \frac{10+9}{2} = \frac{19}{2}$$

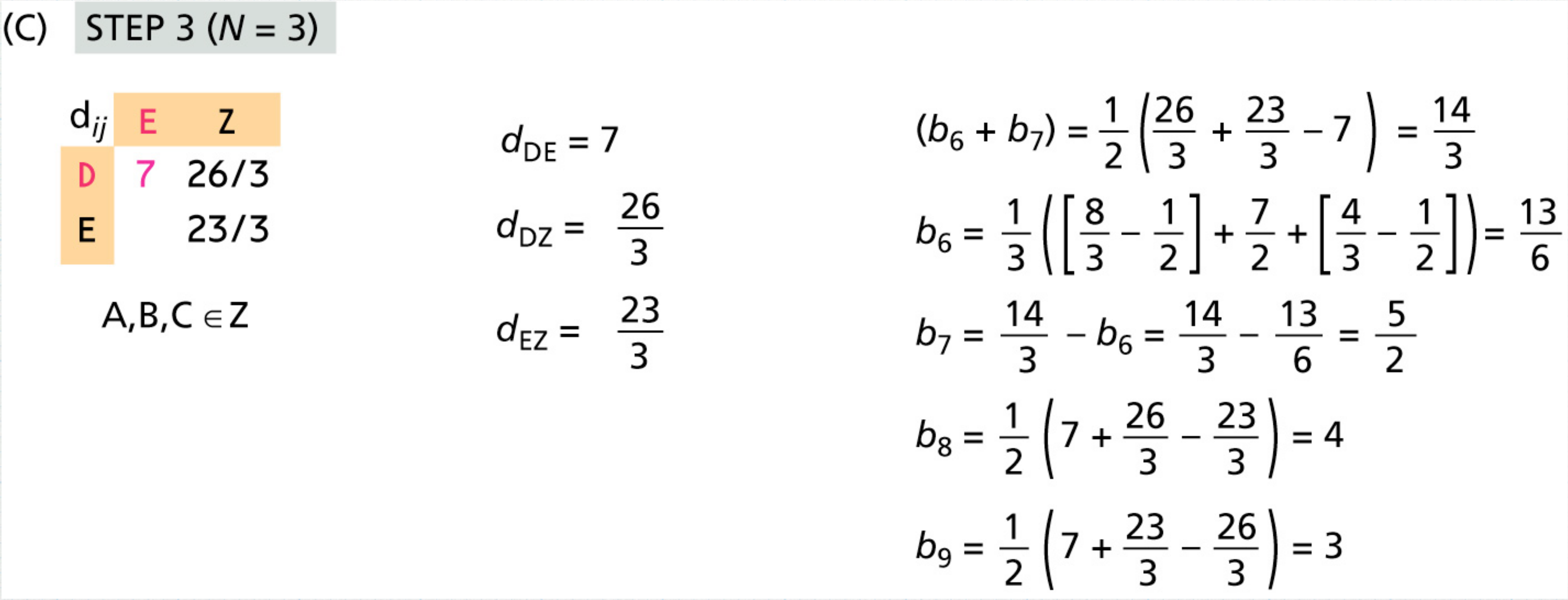
$$b_3 = \frac{1}{2}\left(\frac{8}{3} + \frac{4}{3}\right) = 2$$

$$(b_3 + b_4) = \frac{1}{2}\left(5 + \frac{15}{2} - \frac{19}{2}\right) = \frac{3}{2}$$

$$b_4 = \frac{3}{2} - b_3 = \frac{3}{2} - 2 = -\frac{1}{2}$$

$$b_5 = \frac{1}{2}\left(5 + \frac{19}{2} - \frac{15}{2}\right) = \frac{7}{2}$$

Fitch-Margoliash: An Example



(C) STEP 3 (N = 3)

d_{ij}	E	Z
D	7	26/3
E		23/3

A,B,C ∈ Z

$$d_{DE} = 7$$

$$d_{DZ} = \frac{26}{3}$$

$$d_{EZ} = \frac{23}{3}$$

$$(b_6 + b_7) = \frac{1}{2}\left(\frac{26}{3} + \frac{23}{3} - 7\right) = \frac{14}{3}$$

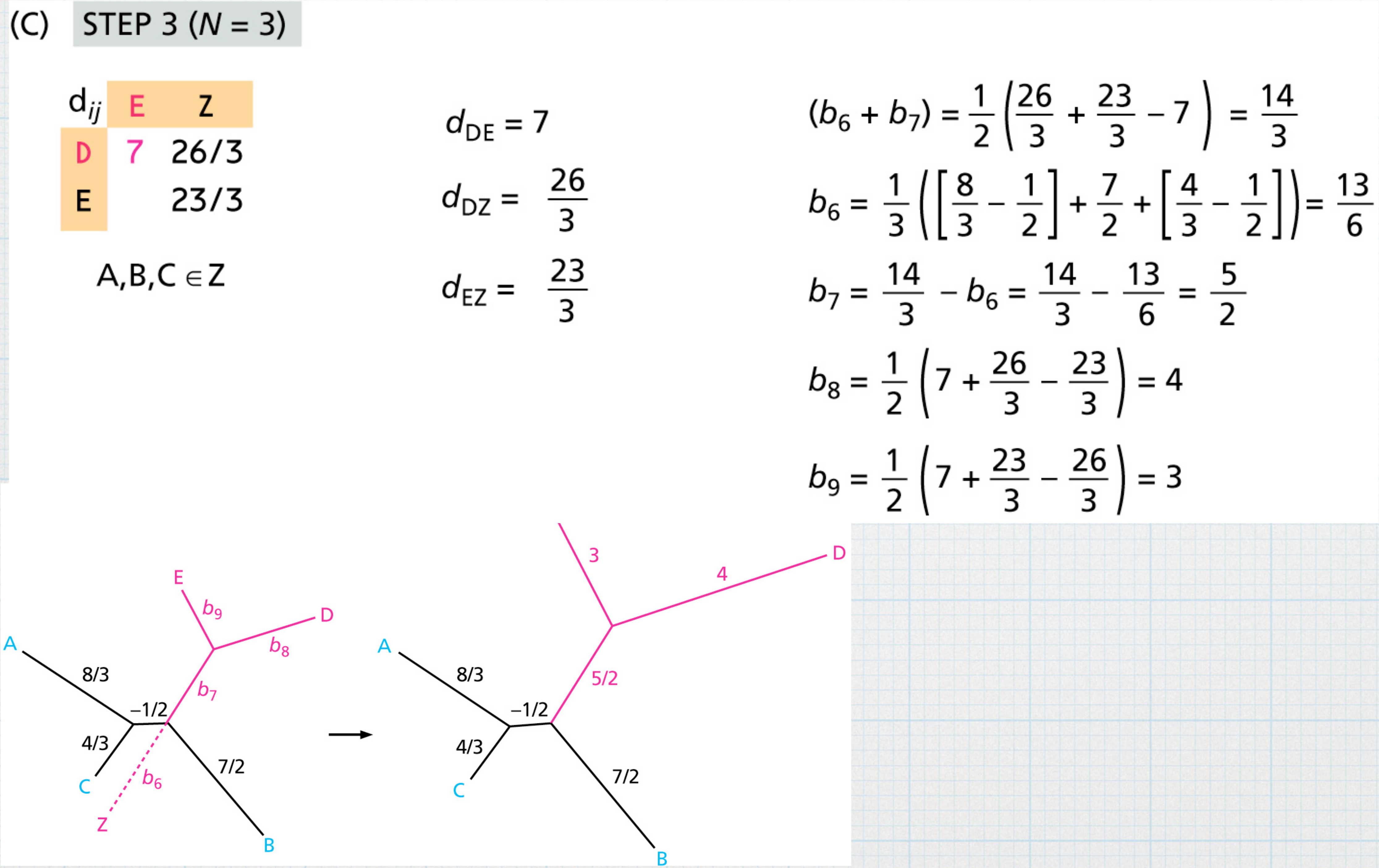
$$b_6 = \frac{1}{3}\left(\left[\frac{8}{3} - \frac{1}{2}\right] + \frac{7}{2} + \left[\frac{4}{3} - \frac{1}{2}\right]\right) = \frac{13}{6}$$

$$b_7 = \frac{14}{3} - b_6 = \frac{14}{3} - \frac{13}{6} = \frac{5}{2}$$

$$b_8 = \frac{1}{2}\left(7 + \frac{26}{3} - \frac{23}{3}\right) = 4$$

$$b_9 = \frac{1}{2}\left(7 + \frac{23}{3} - \frac{26}{3}\right) = 3$$

Fitch-Margoliash: An Example



(C) STEP 3 (N = 3)

d_{ij}	E	Z
D	7	26/3
E		23/3

A,B,C $\in$ Z

$$d_{DE} = 7$$

$$d_{DZ} = \frac{26}{3}$$

$$d_{EZ} = \frac{23}{3}$$

$$(b_6 + b_7) = \frac{1}{2}\left(\frac{26}{3} + \frac{23}{3} - 7\right) = \frac{14}{3}$$

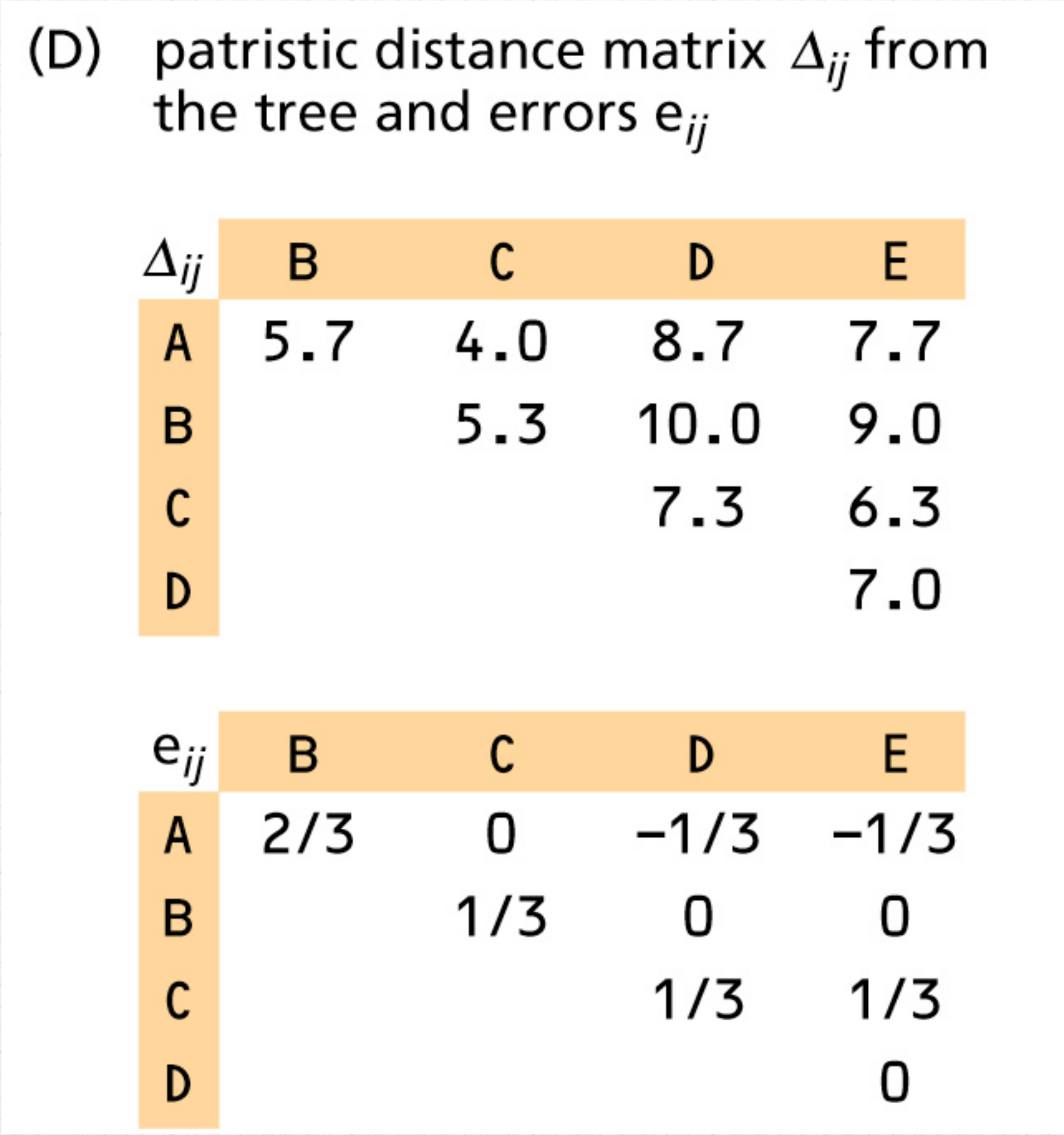
$$b_6 = \frac{1}{3}\left(\left[\frac{8}{3} - \frac{1}{2}\right] + \frac{7}{2} + \left[\frac{4}{3} - \frac{1}{2}\right]\right) = \frac{13}{6}$$

$$b_7 = \frac{14}{3} - b_6 = \frac{14}{3} - \frac{13}{6} = \frac{5}{2}$$

$$b_8 = \frac{1}{2}\left(7 + \frac{26}{3} - \frac{23}{3}\right) = 4$$

$$b_9 = \frac{1}{2}\left(7 + \frac{23}{3} - \frac{26}{3}\right) = 3$$

Fitch-Margoliash: An Example



(D) patristic distance matrix Δ_{ij} from the tree and errors e_{ij}

Δ_{ij}	B	C	D	E
A	5.7	4.0	8.7	7.7
B		5.3	10.0	9.0
C			7.3	6.3
D				7.0

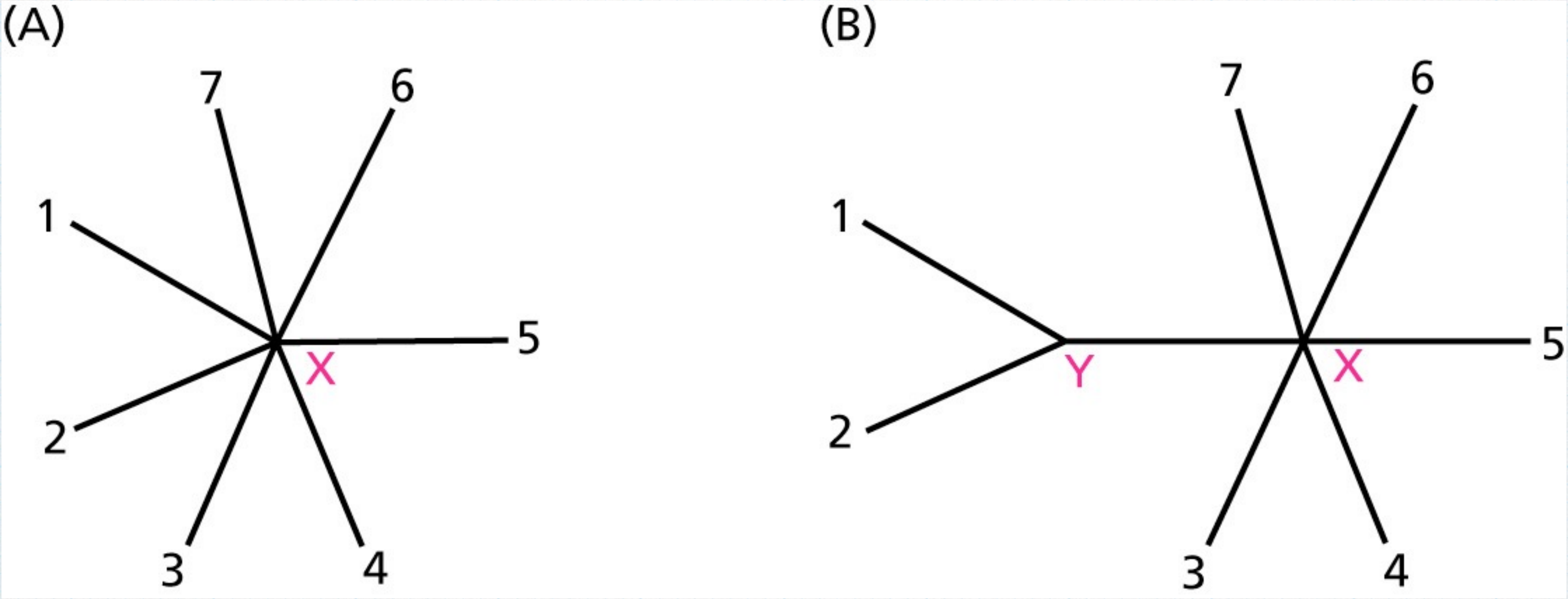
e_{ij}	B	C	D	E
A	2/3	0	−1/3	−1/3
B		1/3	0	0
C			1/3	1/3
D				0

The NJ Method

- The basis of the method lies in the concept of minimum evolution, namely that the true tree will be that for which the total branch length, S, is shortest
- Neighbors in a phylogenetic tree are defined by a pair of nodes that are separated by just one other node
- Pairs of tree nodes are identified at each step of the method (just like with UPGMA and Fitch-Margoliash) and used to gradually build up a tree

The NJ Method:

Deriving the Neighbor-joining Equations



$$S = \sum_{i=1}^{N} b_{iX} = \frac{1}{N-1} \sum_{i<j}^{N} d_{ij}$$

$$S_{12} = b_{1Y} + b_{2Y} + b_{XY} + \sum_{i=3}^{N} b_{iX}$$

b_{ef} : the length of the branch between nodes e and f

The NJ Method:

Deriving the Neighbor-joining Equations

- We need to convert the equation into a form that uses the sequence distances d
- This can be achieved as

$$S_{12} = \frac{1}{2(N-2)} \sum_{i=3}^{N} (d_{1i} + d_{2i}) + \frac{1}{N-2} \sum_{3 \leq i < j}^{N} d_{ij} + \frac{d_{12}}{2}$$

and simplified further into

$$S_{12} = \frac{2d_{sum} - U_1 - U2}{2(N-2)} + \frac{d_{12}}{2}$$

where

$$U_1 = \sum_{i=1}^{N} d_{1i} \qquad U_2 = \sum_{i=1}^{N} d_{2i} \qquad d_{sum} = \sum_{i<j}^{N} d_{ij}$$

The NJ Method:

Deriving the Neighbor-joining Equations

- Every pair of sequences i and j, if separated from the star node, produce a tree of total branch length S_{ij}

- According to the minimum evolution principle, the tree that should be chosen is that with the smallest S_{ij}

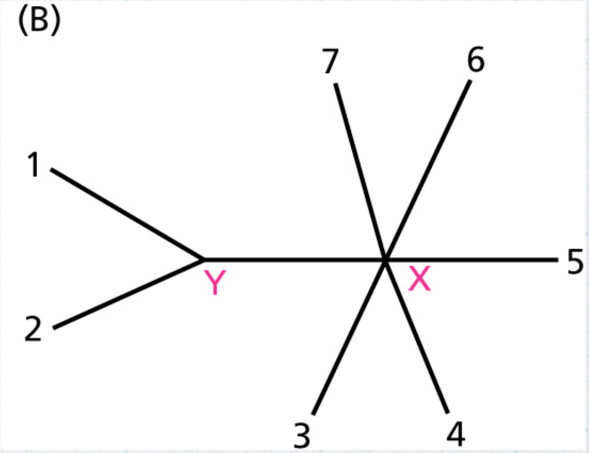
- This is equivalent to finding the pair of sequences with the smallest value of the quantity δ_{ij} defined by

$$\delta_{ij} = d_{ij} - \frac{U_i + U_j}{N - 2}$$

The NJ Method:

Deriving the Neighbor-joining Equations

- Once this pair has been found, the distances to the new node Y must be calculated



$$b_{iY} = \frac{1}{2}\left(d_{ij} + \frac{U_i - U_j}{N-2}\right)$$

and

$$b_{jY} = d_{ij} - b_{iY}$$

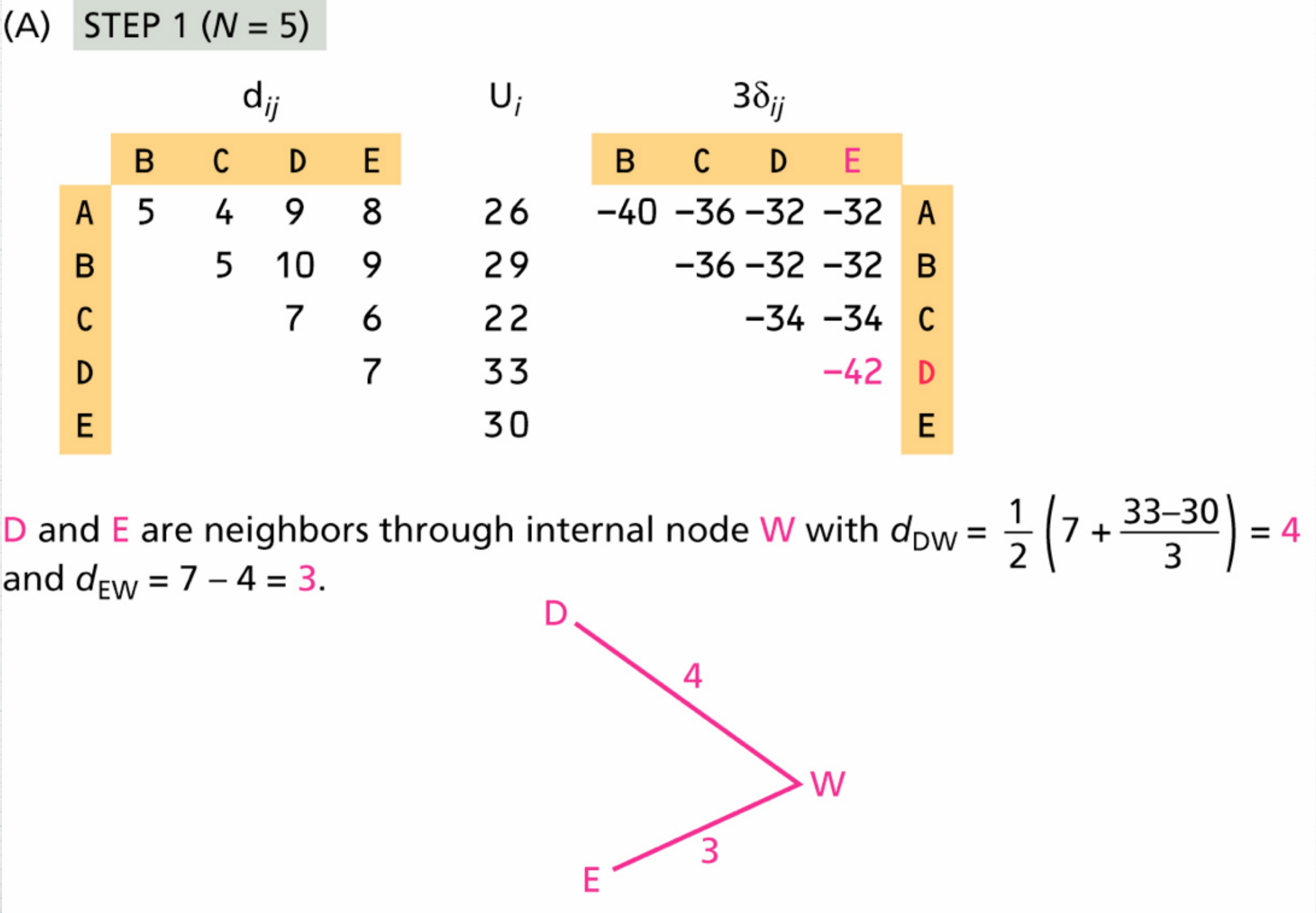
- To calculate the distances from Y to every other sequence k:

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

The NJ Method:
Deriving the Neighbor-joining Equations

- To add more nodes, we now repeat the process, starting with the star tree formed by removing sequences i and j, to leave a star tree with node Y as a new leaf
- Note that at each step, the value of N in the formulas decreases by 1

NJ: An Example



(A) STEP 1 ($N = 5$)

d_{ij} and U_i

	B	C	D	E	U_i
A	5	4	9	8	26
B		5	10	9	29
C			7	6	22
D				7	33
E					30

$3\delta_{ij}$

B	C	D	E	
−40	−36	−32	−32	A
	−36	−32	−32	B
		−34	−34	C
			−42	D
				E

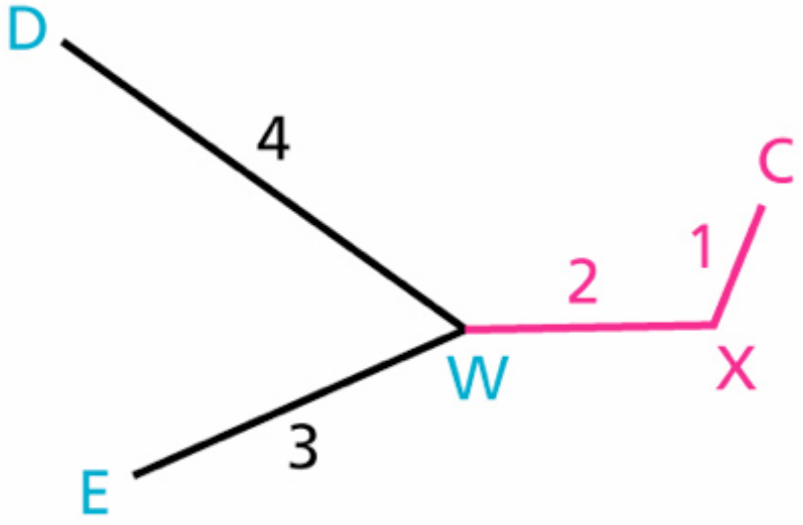
D and E are neighbors through internal node W with $d_{DW} = \frac{1}{2}\left(7 + \frac{33-30}{3}\right) = 4$ and $d_{EW} = 7 - 4 = 3$.

NJ: An Example

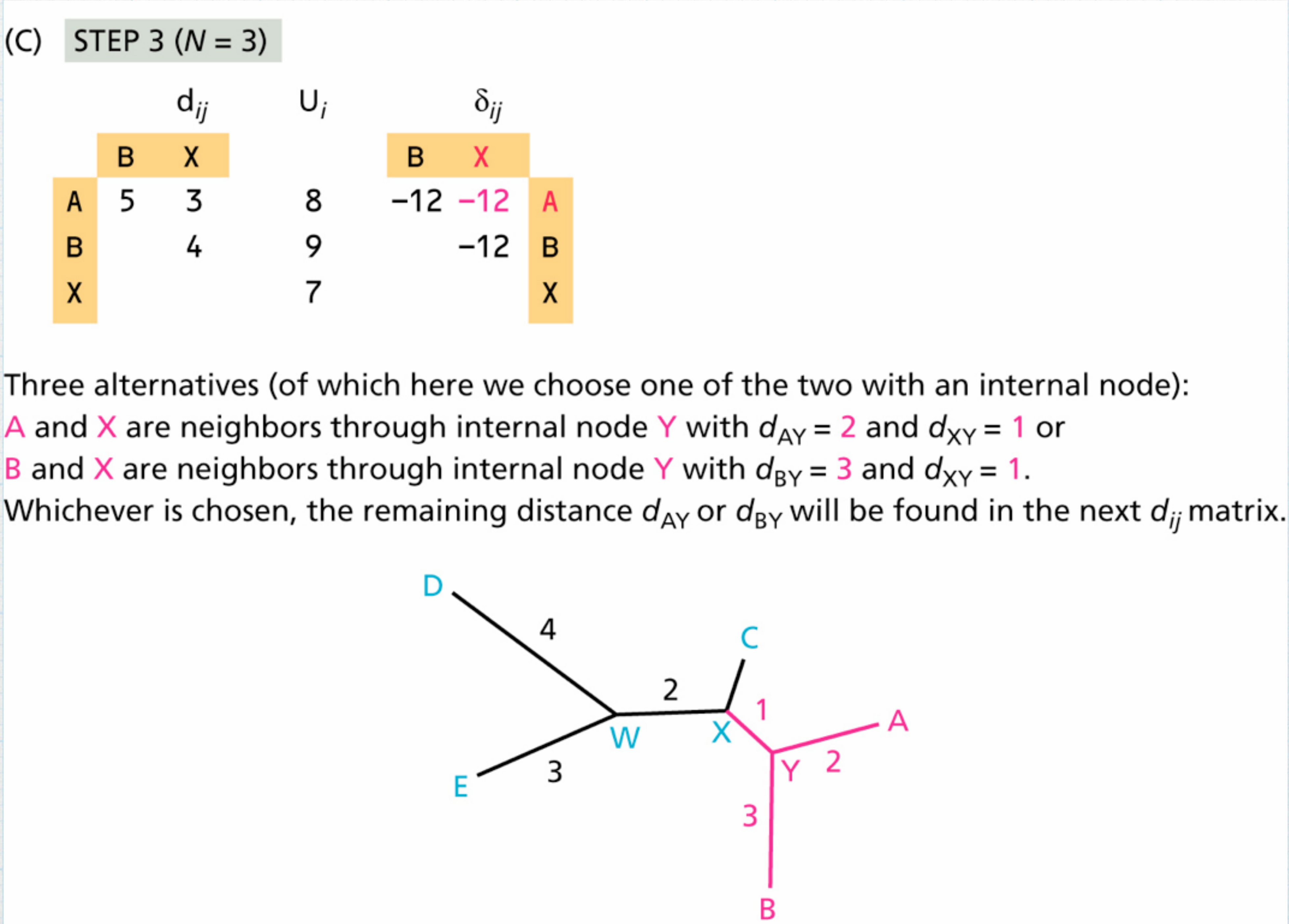
(B) STEP 2 ($N = 4$)

	d_{ij}			U_i	$2\delta_{ij}$			
	B	C	W		B	C	W	
A	5	4	5	14	−20	−18	−18	A
B		5	6	16		−18	−18	B
C			3	12			−20	C
W				14				W

C and W are neighbors through internal node X with $d_{CX} = \frac{1}{2}\left(3 + \frac{12-14}{2}\right) = 1$ and $d_{WX} = 3 - 1 = 2$.



NJ: An Example



(C) STEP 3 (N = 3)

d_{ij}	B	X	U_i
A	5	3	8
B		4	9
X			7

B	X	δ_{ij}
−12	−12	A
	−12	B
		X

Three alternatives (of which here we choose one of the two with an internal node):
A and X are neighbors through internal node Y with $d_{AY} = 2$ and $d_{XY} = 1$ or
B and X are neighbors through internal node Y with $d_{BY} = 3$ and $d_{XY} = 1$.
Whichever is chosen, the remaining distance d_{AY} or d_{BY} will be found in the next d_{ij} matrix.

Acknowledgments

- Materials are from
  - ‘Understanding Bioinformatics’, by Zvelebil and Baum
  - ‘Molecular Evolution: A Statistical Approach”, by Yang

Questions?