Evolutionary Thinking 2022
TA session
week 2 – Tree building methods

Jilong Ma
aujilongm@birc.au.dk

Outline

1. Recap

1. The learning outcome of this today (20 minutes)

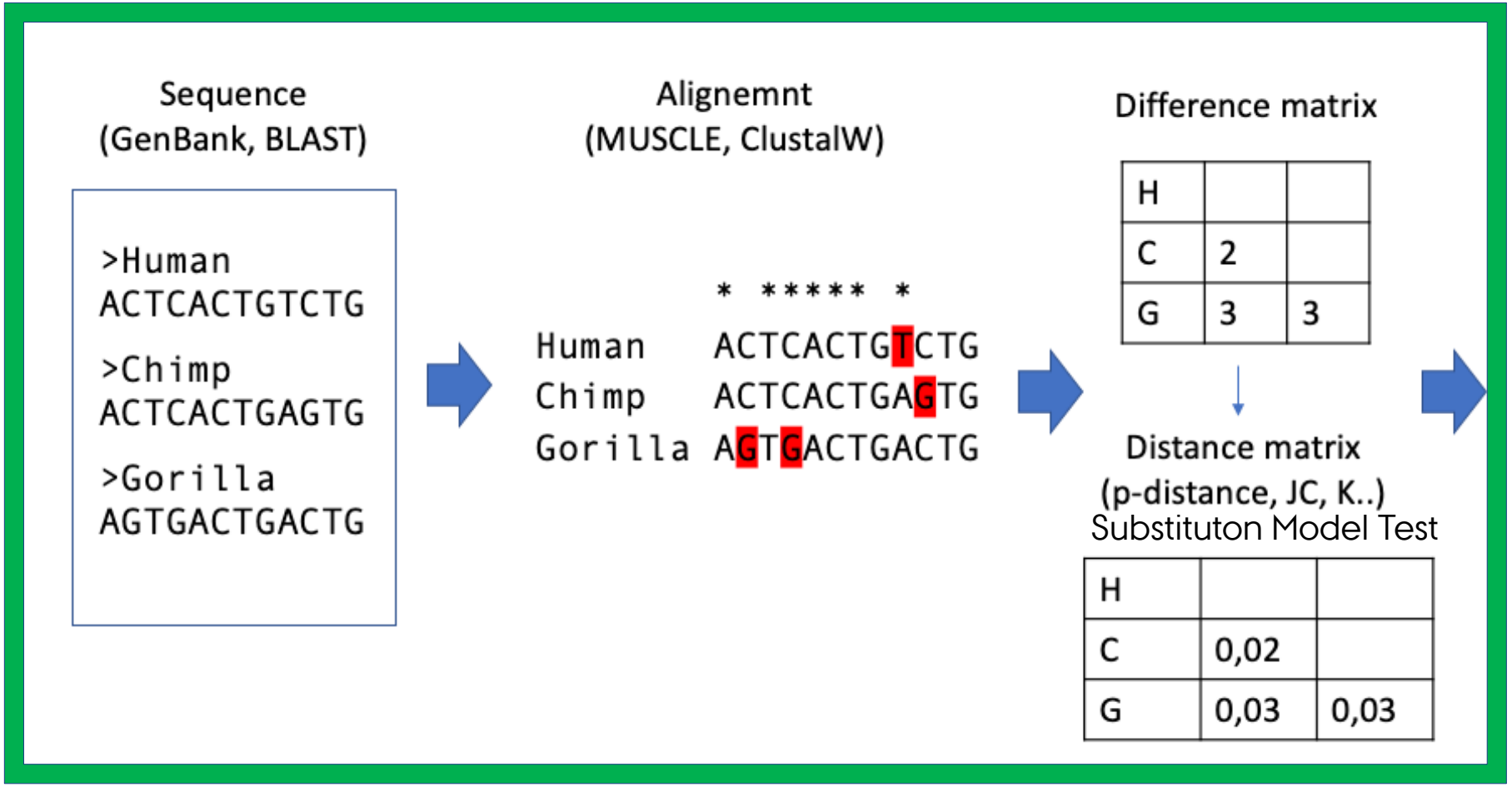
Tree building methods

Distance-based phylogeny tree building algorithms

UPGMA, NJ tree

2. Working on MEGA exercises

Recap



Distance based phylogeny tree building

Recap

Sequence
(GenBank, BLAST)

```
>Human
ACTCACTGTCTG
>Chimp
ACTCACTGAGTG
>Gorilla
AGTGACTGACTG
```

Alignemnt
(MUSCLE, ClustalW)

```
          *  *****  *
Human     ACTCACTGTCTG
Chimp     ACTCACTGAGTG
Gorilla   AGTGACTGACTG
```

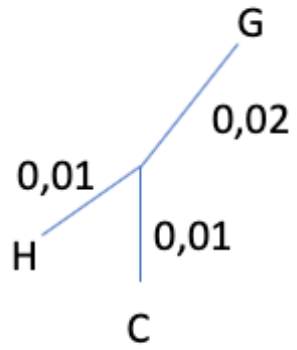
Difference matrix

H		
C	2	
G	3	3

Distance matrix
(p-distance, JC, K..)
Substituton Model Test

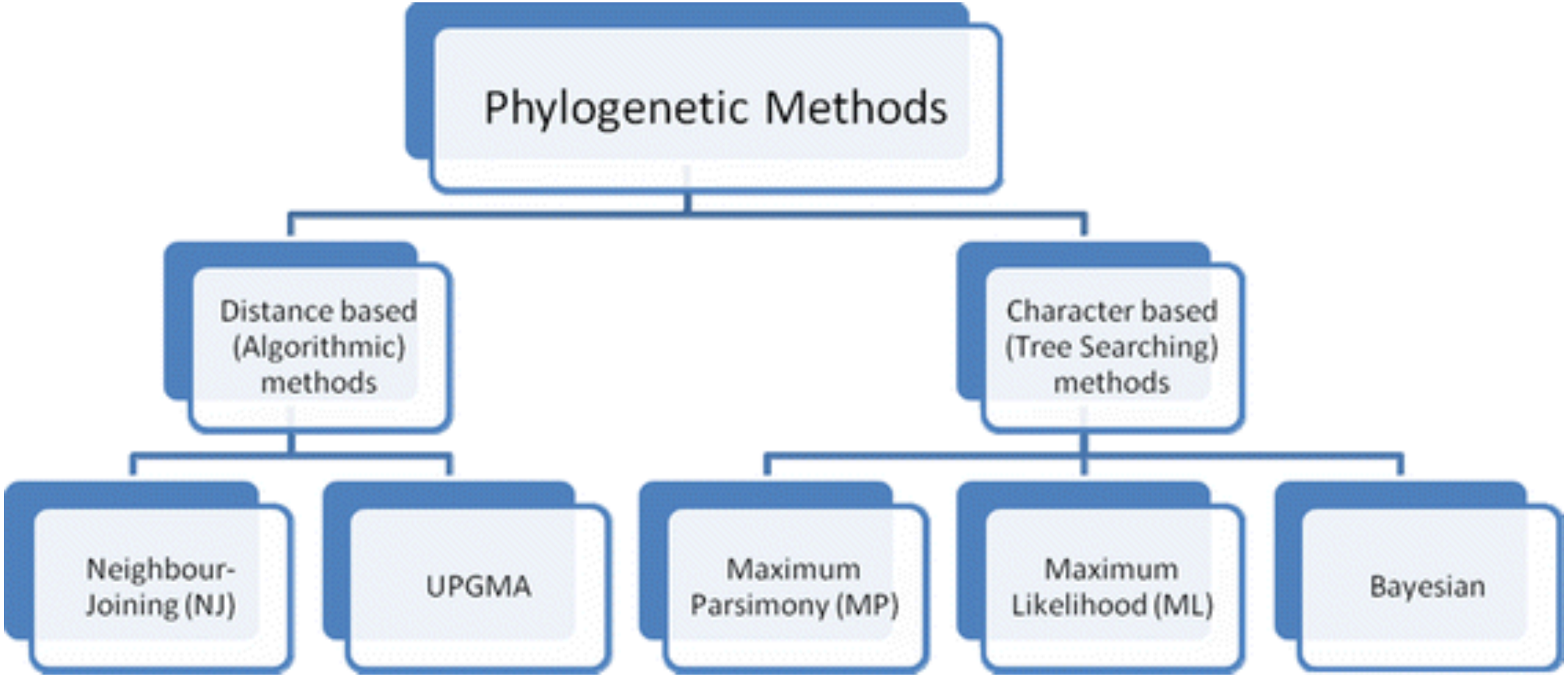
H		
C	0,02	
G	0,03	0,03

Tree building
(UPGMA, Neighbour-Joining...)

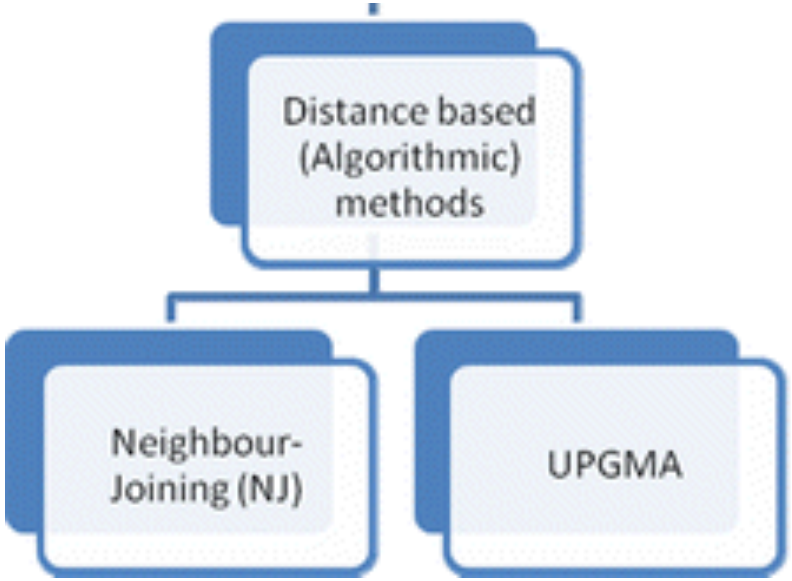


Distance based phylogeny tree building

Learning outcome of today



Learning outcome of today



Learning outcome of today

Conceptually, what are MP, ML and Bayesian doing ? (3 minutes)

Learning outcome of today

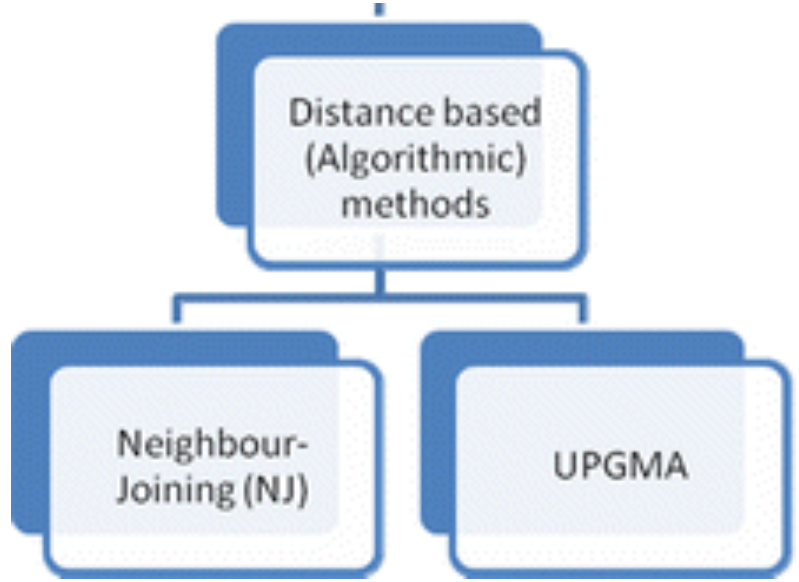
Conceptually, what are MP, ML, and Bayesian doing?

MP: Finding the topology requires the smallest evolutionary changes (substitutions)

ML: Finding the phylogeny maximize P(Data | Phylogeny)

Bayesian: Finding the Phylogeny has the highest P(Phylogeny | Data)

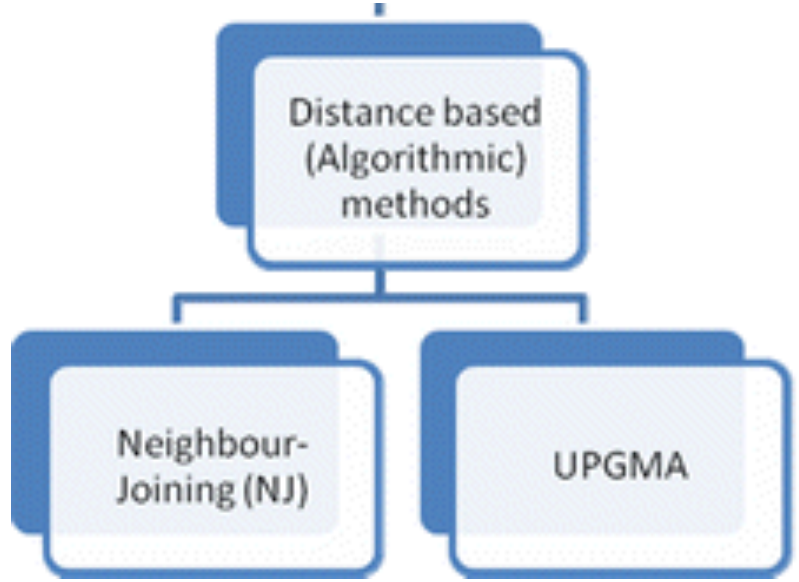
Learning outcome of today



UPGMA
(unweighted pair-group method with arithmetic means)

	A	B	C	D
A				
B	5			
C	2	3		
D	1	4	6	

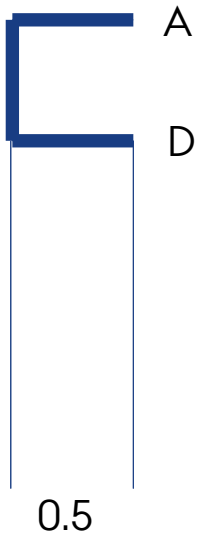
Learning outcome of today



UPGMA
(unweighted pair-group method with arithmetic means)

	A	B	C	D
A				
B	5			
C	2	3		
D	1	4	6	

Min d = d_AD = 1



Learning outcome of today

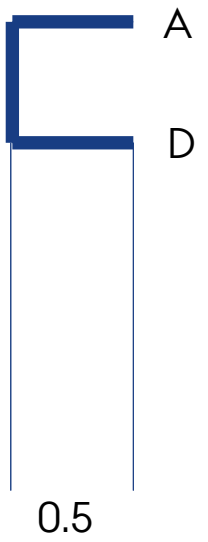
UPGMA
(unweighted pair-group method with arithmetic means)

	A	B	C	D
A				
B	5			
C	2	5		
D	1	4	6	

	AD	B	C
AD			
B	4.5		
C	3	5	

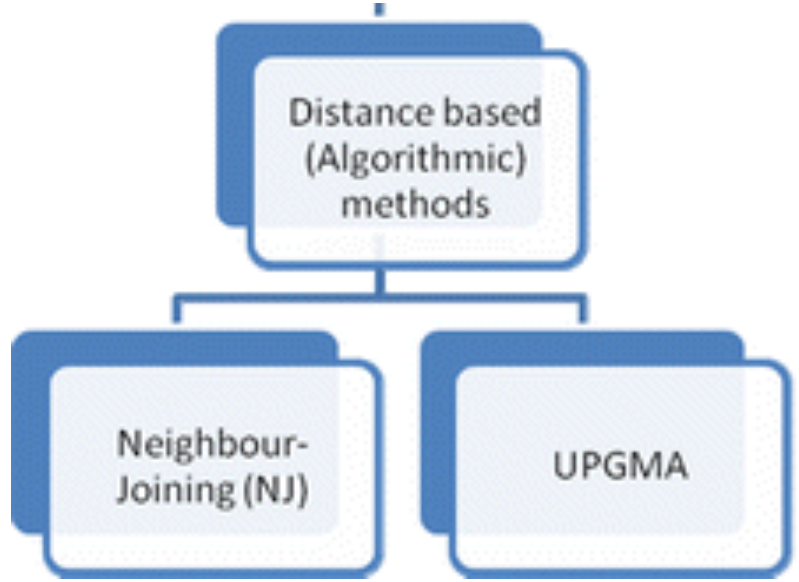
$$d(AD)B = (dAB + dBD) / 2$$
$$= (5 + 4) / 2 = 4.5$$

$$d(AD)C = (dAC + dCD) / 2$$
$$= (2 + 6) / 2 = 3$$

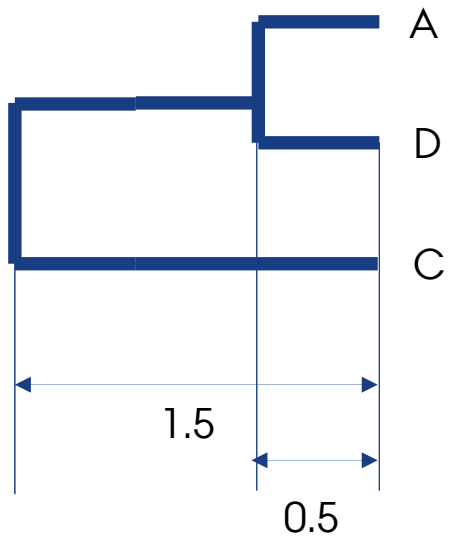


Learning outcome of today

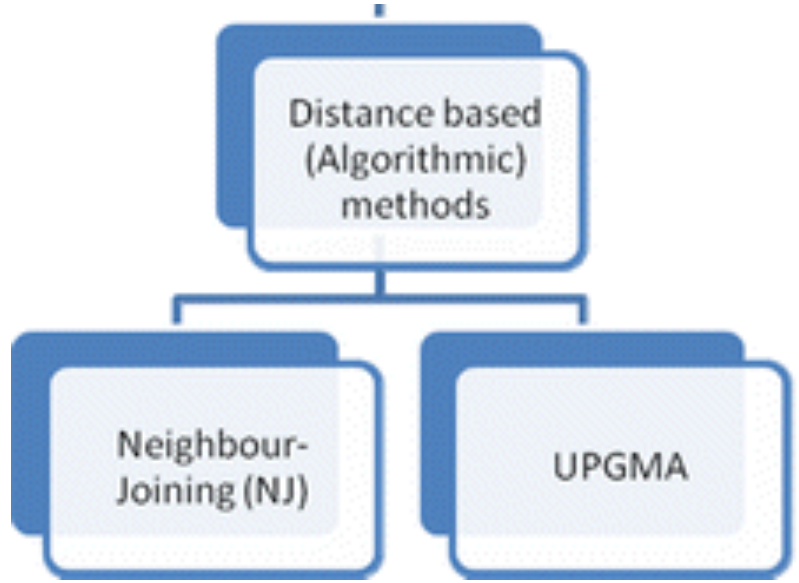
UPGMA
(unweighted pair-group method with arithmetic means)



	AD	B	C
AD			
B	4.5		
C	3	5	



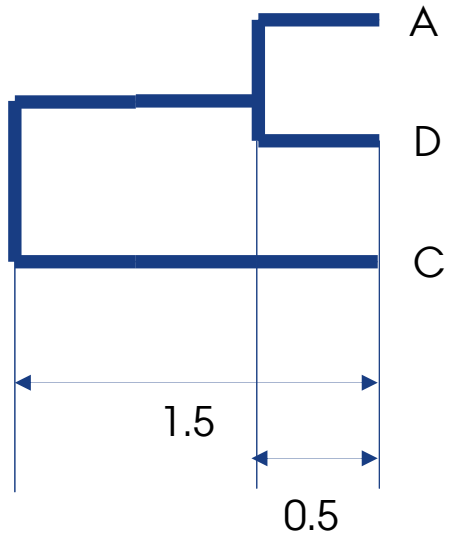
Learning outcome of today



UPGMA
(unweighted pair-group method with arithmetic means)

	AD	B	C
AD			
B	4.5		
C	3	5	

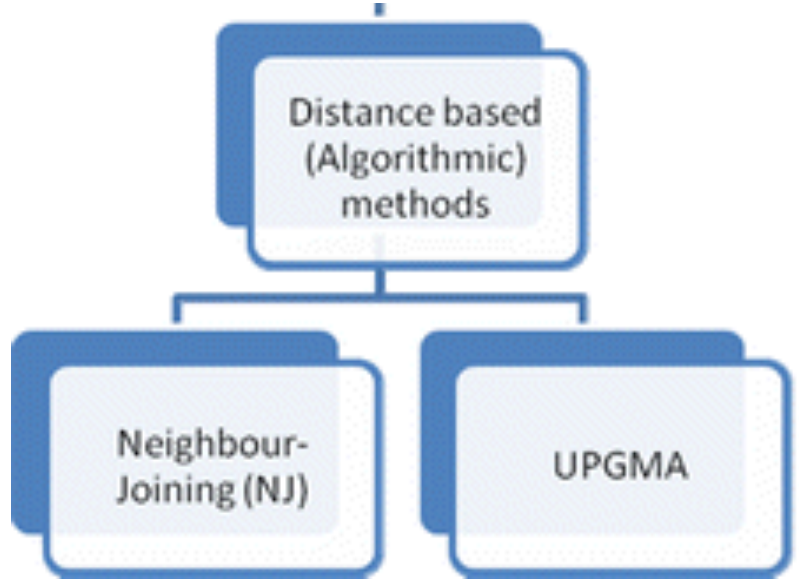
	ADC	B
ADC		
B	4.75	

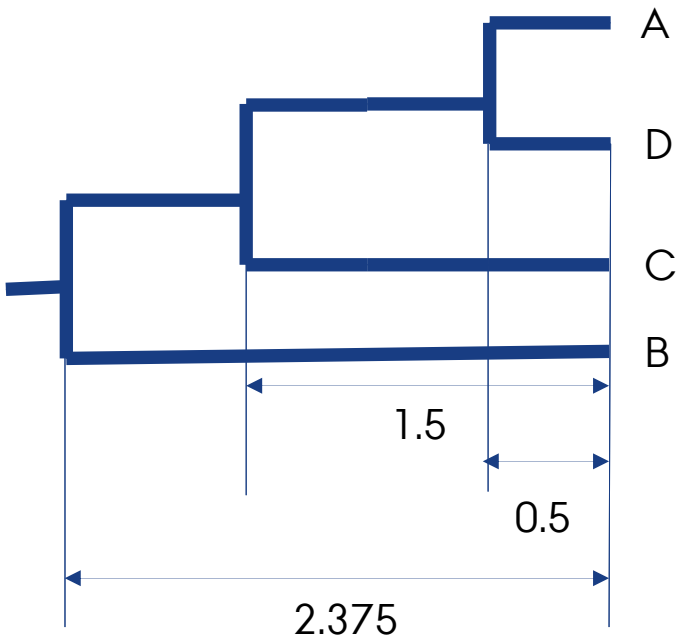


Learning outcome of today

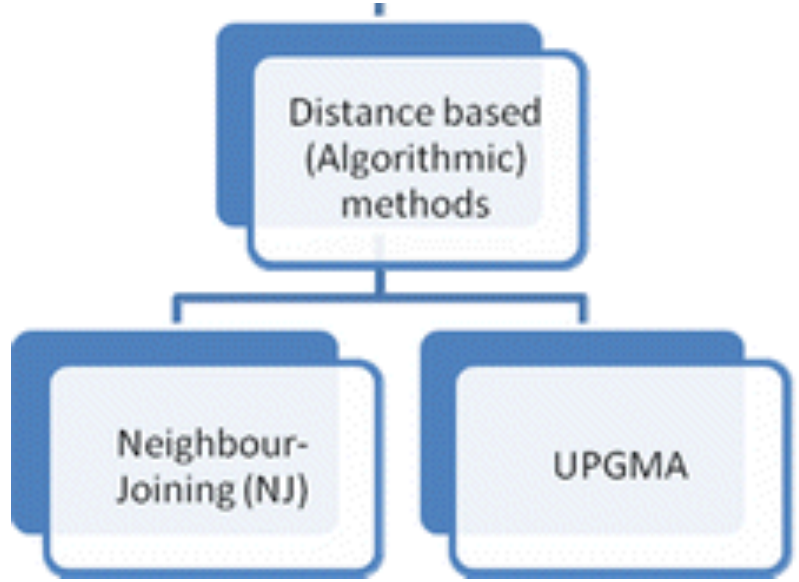
UPGMA
(unweighted pair-group method with arithmetic means)

	ADC	B
ADC		
B	4.75	





Learning outcome of today



Neighbor-joining

Understand
the formula (5.11) and (5.12). Dan Graur P107
With drawing examples

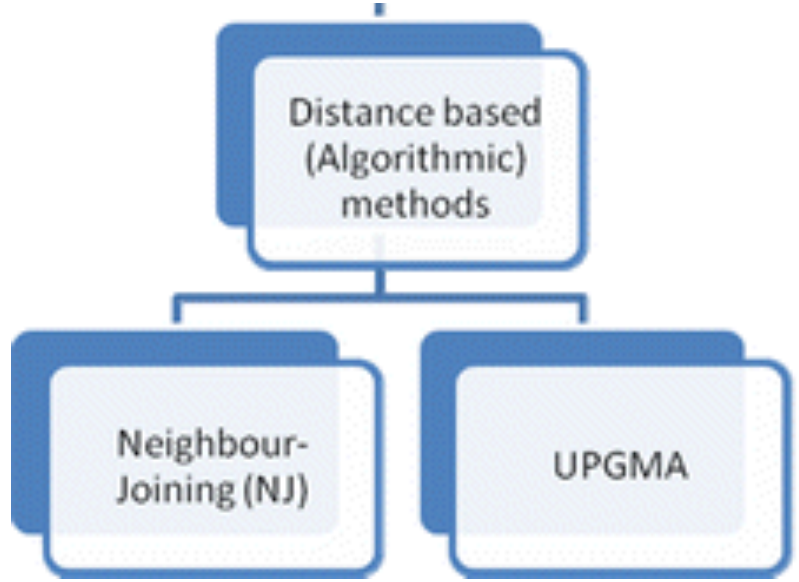
Total tree length

$$S_0 = \frac{1}{N-1}\sum_{i<j} d_{ij}$$

Total tree length after connecting 1 and 2

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

Learning outcome of today



Difference between UPGMA and NJ (5 min)

UPGMA

1. Rooted Tree

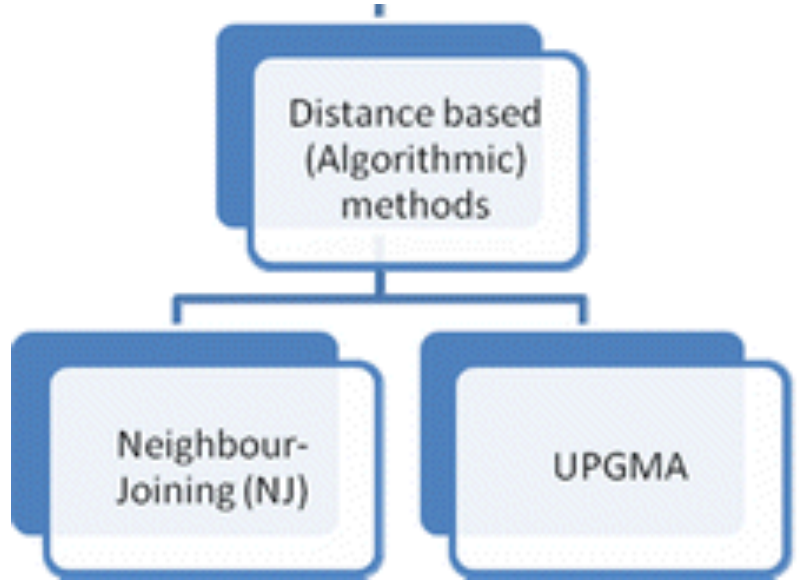
2. Branch tips come out equal
(equal distance from the root, equal rates of evolution)

NJ

1. Unrooted Tree

2. Branch length varies and proportional to the amount of change
(Allow unequal rates of evolution)

Learning outcome of today



Difference between UPGMA and NJ (5 min)

UPGMA

1. Rooted Tree

2. Branch tips come out equal
(equal distance from the root, equal rates of evolution)

NJ

1. Unrooted Tree

2. Branch length varies and proportional to the amount of change
(Allow unequal rates of evolution)

MEGA exercises

1. Finding the best-fitted substitution models

2. Building phylogeny trees with different methods

AARHUS
UNIVERSITY