Bioinformatics
Phylogenetic trees

MSc. Vicente Machaca Arceda

Universidad Nacional de San Agustín de Arequipa

June 3, 2021

Table of Contents

vmachacaa@unsa.edu.pe

Table of Contents

vmachacaa@unsa.edu.pe

Introduction
Objectives

vmachacaa@unsa.edu.pe

- Understand the importance of phylogenetic trees.

Introduction
Objectives

- Understand the importance of phylogenetic trees.
- Understand and implement UPGMA.

Table of Contents

vmachacaa@unsa.edu.pe

Phylogenetics
Definition

Phylogenetics is the study of the evolutionary history of living organisms using tree like diagrams to represent pedigrees of these organisms [1].
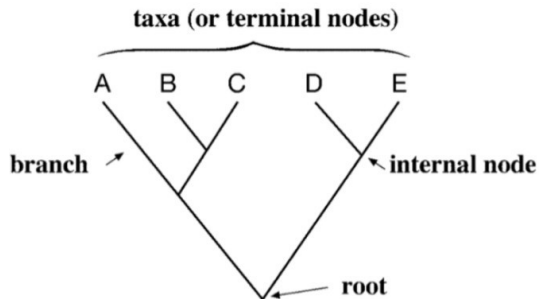
Phylogenetics
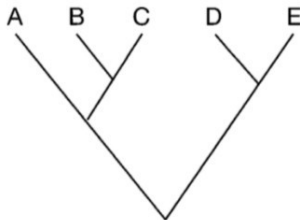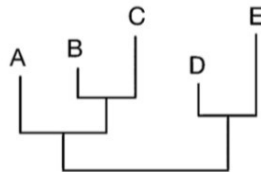## Example

Figure: A typical bifurcating phylogenetic tree showing root, internal nodes, terminal nodes and branches. Source: [1]

Phylogenetics
Example

Figure: Newick format of tree representation. Source: [1]

vmachacaa@unsa.edu.pe

Phylogenetics
Methods
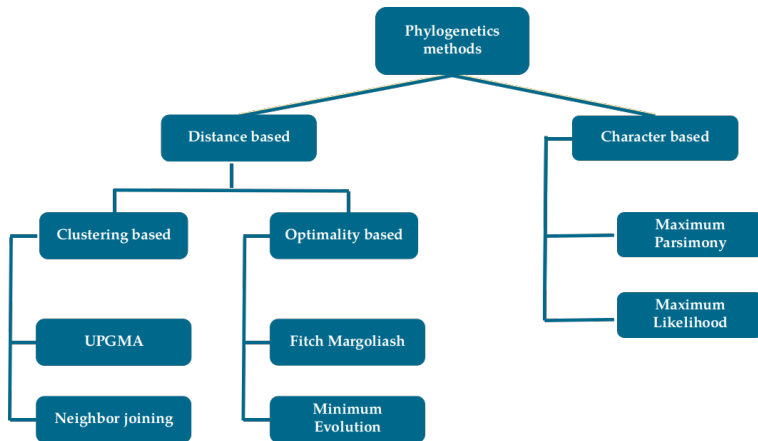
Figure: The most used methods to build philogenetic trees.

Table of Contents

UPGMA
Unweighted Pair Group Method Using Arithmetic Average

	A	**B**	**C**
B	**0.40**		
C	0.35	**0.45**	
D	**0.60**	**0.70**	**0.55**

Figure: Distances between sequences (score).

Choose the smallest distance *(A, C)*. Then computer the cluster.

$$\frac{AC}{2} = \frac{0.35}{2} = 0.175$$

vmachacaa@unsa.edu.pe
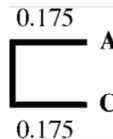
UPGMA
Unweighted Pair Group Method Using Arithmetic Average

Join *C, A* columns.

	A	B	C
B	0.40		
C	0.35	0.45	
D	0.60	0.70	0.55

Figure: Distances between sequences (score).

	A-C	B
B	$\frac{0.4 + 0.45}{2} = 0.425$	
D	$\frac{0.55 + 0.6}{2} = 0.575$	0.70

Figure: New matrix

UPGMA
Unweighted Pair Group Method Using Arithmetic Average

Choose the smallest distance *(B, A-C)*. Then computer the cluster.

	A-C	B
B	$\frac{0.4 + 0.45}{2} = 0.425$	
D	$\frac{0.55 + 0.6}{2} = 0.575$	0.70

Figure: Distances between sequences (score).
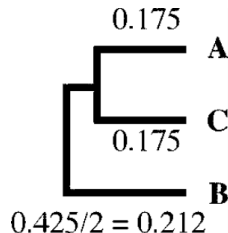
UPGMA
Unweighted Pair Group Method Using Arithmetic Average

Join *AC, B* columns

	A-C	B
B	$\frac{0.4 + 0.45}{2} = 0.425$	
D	$\frac{0.55 + 0.6}{2} = 0.575$	0.70

Figure: Distances between sequences (score).

	B-A-C
D	$\frac{0.7 + 0.6 + 0.55}{3} = 0.617$

Figure: New matrix

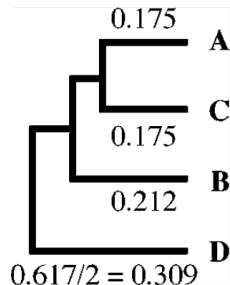# UPGMA
Unweighted Pair Group Method Using Arithmetic Average

Choose the smallest distance *(B-A-C)*. Then computer the cluster.

	B-A-C
D	$\dfrac{0.7 + 0.6 + 0.55}{3} = 0.617$

Figure: Distances between sequences (score).

0.175   **A**

0.175   **C**

0.212   **B**

$0.617/2 = 0.309$ **D**

UPGMA
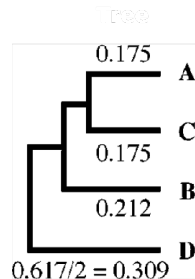Unweighted Pair Group Method Using Arithmetic Average

The estimated distances don't match the actual evolutionary distances shown, which illustrates the failure of UPGMA. However, owing to its fast speed of calculation, it has found extensive usage.

	A	B	C
B	0.42		
C	0.35	0.42	
D	0.62	62	0.62

0.212 + 0.212

0.309 + 0.309

0.175 + 0.175

0.175 — A

0.175 — C

0.212 — B

0.617/2 = 0.309 — D

vmachacaa@unsa.edu.pe

Questions?

References I

J. Xiong, *Essential bioinformatics*. Cambridge University Press, 2006.

X. Tang, C. Wu, X. Li, Y. Song, X. Yao, X. Wu, Y. Duan, H. Zhang, Y. Wang, Z. Qian *et al.*, "On the origin and continuing evolution of sars-cov-2," *National Science Review*, 2020.

S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of molecular biology*, vol. 215, no. 3, pp. 403–410, 1990.

vmachacaa@unsa.edu.pe