

# Bioinformatics

## Neighbor Joining

MSc. Vicente Machaca Arceda

Universidad Nacional de San Agustín de Arequipa

June 23, 2021

# Table of Contents

- 1 Introduction
  - Objectives
  - Methods

- 2 Neighbor joining
  - Method
  - Example

# Table of Contents

- 1 Introduction
  - Objectives
  - Methods
- 2 Neighbor joining
  - Method
  - Example

# Introduction

## Objectives

- Understand and implement Neighbor joining algorithm to build phylogenetics trees.

# Table of Contents

- 1 Introduction
  - Objectives
  - **Methods**
  
- 2 Neighbor joining
  - Method
  - Example

# Phylogenetics

## Methods

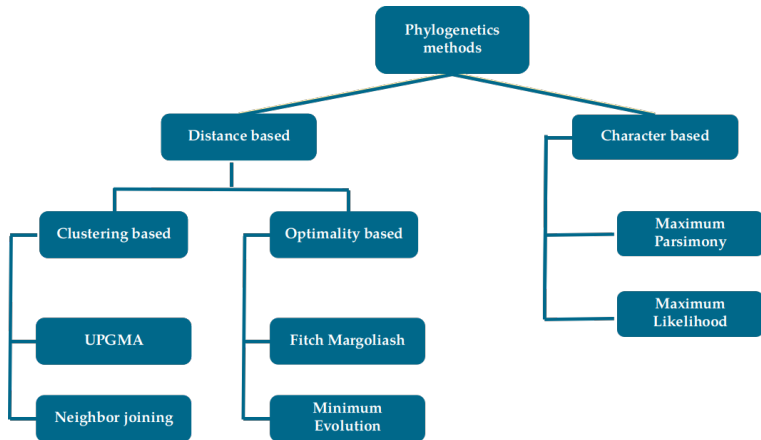


Figure: The most used methods to build phylogenetic trees.

# Table of Contents

- 1 Introduction
  - Objectives
  - Methods
  
- 2 Neighbor joining
  - **Method**
  - Example

# Neighbor joining

## Definition

The UPGMA method uses unweighted distances and assumes that all taxa have constant evolutionary rates.

Neighbor joining (NJ) does not assume the taxa to be equidistant from the root. It corrects for unequal evolutionary rates between sequences by using a conversion step.



# Neighbor joining

## Definition

NJ requires the calculations of “r-values” and “transformed r-values” using:

$$d'_{AB} = d_{AB} - \frac{1}{2}(r_A + r_B) \quad (1)$$

- $d'_{AB}$  is the converted distance between  $A$  and  $B$ .
- $d_{AB}$  is the actual evolutionary distance between  $A$  and  $B$ .
- $r_A$  (or  $r_B$ ) is the sum of distances of  $A$  (or  $B$ ) to all other taxa.

# Neighbor joining

## Definition

The  $r$ -values are needed to create a modified distance matrix:

$$r_i = \sum d_{ij} \quad (2)$$

- $i$  and  $j$  are two different taxa.

The transformed  $r$ -values ( $r'$ ) are used to determine the distances of an individual taxon to the nearest node:

$$r'_i = \frac{r_i}{n-2} \quad (3)$$

- $n$  is the total number of taxa.

# Neighbor joining

## Definition

Assuming  $A$  and  $B$  form a node called  $U$ , the distance  $A$  to  $U$  is determined by the following formula:

$$d_{AU} = \frac{d_{AB} + (r'_A - r'_B)}{2} \quad (4)$$

# Table of Contents

- 1 Introduction
  - Objectives
  - Methods
- 2 Neighbor joining
  - Method
  - **Example**

# Neighbor joining

## Example

	A	B	C
B	<b>0.40</b>		
C	<b>0.35</b>	<b>0.45</b>	
D	<b>0.60</b>	<b>0.70</b>	<b>0.55</b>

The first step of the NJ method is  $r$ -value and  $r'$ -value calculation. According to Eq. 2 and 3.

# Neighbor joining

## Example

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

Eq. 2 and 3.

$$r_i = \sum d_{ij}; r'_i = \frac{r_i}{n-2}$$

For node A:

$$r_A = d_{AB} + d_{AC} + d_{AD} = 0.4 + 0.35 + 0.6$$

$$r_A = 1.35$$

$$r'_A = \frac{1.35}{4-2}$$

# Neighbor joining

## Example

	A	B	C
B	<b>0.40</b>		
C	<b>0.35</b>	<b>0.45</b>	
D	<b>0.60</b>	<b>0.70</b>	<b>0.55</b>

Eq. 2 and 3.

$$r_i = \sum d_{ij}; r'_i = \frac{r_i}{n-2}$$

Do the same for nodes  $B, C$  and  $D$ :

# Neighbor joining

## Example

	A	B	C
B	<b>0.40</b>		
C	<b>0.35</b>	<b>0.45</b>	
D	<b>0.60</b>	<b>0.70</b>	<b>0.55</b>

$$r_A = 1.35; r'_A = 0.675$$

$$r_B = 1.55; r'_B = 0.775$$

$$r_C = 1.35; r'_C = 0.675$$

$$r_D = 1.85; r'_D = 0.925$$



# Neighbor joining

## Example

	A	B	C
B	<b>0.40</b>		
C	<b>0.35</b>	<b>0.45</b>	
D	<b>0.60</b>	<b>0.70</b>	<b>0.55</b>

Calculate  $d'_{ij}$  using Eq. 1.

$$d'_{AB} = d_{AB} - \frac{1}{2}(r_A + r_B)$$

# Neighbor joining

## Example

	A	B	C
B	<b>0.40</b>		
C	<b>0.35</b>	<b>0.45</b>	
D	<b>0.60</b>	<b>0.70</b>	<b>0.55</b>

$$d'_{AB} = -1.05$$

$$d'_{AC} = -1.00$$

$$d'_{AD} = -1.00$$

$$d'_{BC} = -1.00$$

$$d'_{BD} = -1.00$$

$$d'_{CD} = -1.05$$

# Neighbor joining

## Example

The old distance matrix.

	A	B	C
B	<b>0.40</b>		
C	<b>0.35</b>	<b>0.45</b>	
D	<b>0.60</b>	<b>0.70</b>	<b>0.55</b>

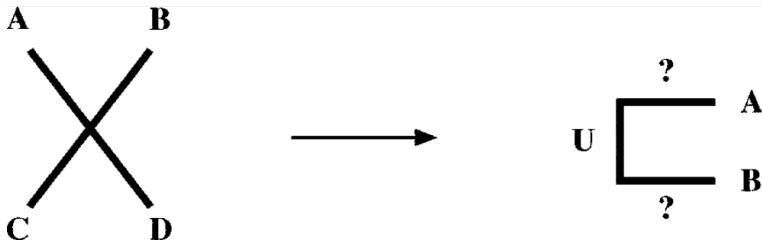
The new distance matrix.

	A	B	C
B	<b>-1.05</b>		
C	<b>-1</b>	<b>-1</b>	
D	<b>-1</b>	<b>-1</b>	<b>-1.05</b>

# Neighbor joining

## Example

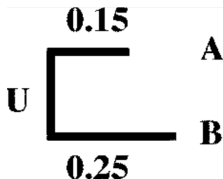
The pair of taxa with the shortest distances in the new matrix are separated (either  $AB$  or  $CD$ ).



Then, use Eq. 4 to calculate  $d_{AU}$  and  $d_{BU}$ .

# Neighbor joining

## Example



Eq. 4:

$$d_{AU} = \frac{d_{AB} + (r'_A - r'_B)}{2} = 0.15$$

$$d_{BU} = \frac{d_{AB} + (r'_B - r'_A)}{2} = 0.25$$

# Neighbor joining

## Example

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

	U	C
C	0.20	
D	0.45	0.55

$$d_{CU} = \frac{(d_{AC} - d_{UA}) + (d_{BC} - d_{UB})}{2} = \frac{(0.35 - 0.15) + (0.45 - 0.25)}{2}$$

$$d_{CU} = 0.2$$

$$d_{DU} = \frac{(d_{AD} - d_{UA}) + (d_{BD} - d_{UB})}{2} = \frac{(0.6 - 0.15) + (0.7 - 0.25)}{2}$$

$$d_{DU} = 0.45$$

# Neighbor joining

## Example

Based on the reduced distance matrix, a new set of  $r$ - and  $r'$ -values are calculated (Use Eq. 2 and 3).

	U	C
C	0.20	
D	0.45	0.55

$$r_C = ?; r'_C = ?$$

$$r_D = ?; r'_D = ?$$

$$r_U = ?; r'_U = ?$$

# Neighbor joining

## Example

Based on the reduced distance matrix, a new set of  $r$ - and  $r'$ -values are calculated (Use Eq. 2 and 3).

	U	C
C	0.20	
D	0.45	0.55

$$r_C = 0.75; r'_C = 0.75$$

$$r_D = 1.00; r'_D = 1.00$$

$$r_U = 0.65; r'_U = 0.65$$



# Neighbor joining

## Example

Now compute new corrected distances (Use Eq. 1).

	<b>U</b>	<b>C</b>
<b>C</b>	<b>0.20</b>	
<b>D</b>	<b>0.45</b>	<b>0.55</b>

$$d'_{CU} = ?$$

$$d'_{DU} = ?$$

$$d'_{CD} = ?$$

# Neighbor joining

## Example

Now compute new corrected distances (Use Eq. 1).

	<b>U</b>	<b>C</b>
<b>C</b>	<b>0.20</b>	
<b>D</b>	<b>0.45</b>	<b>0.55</b>

$$d'_{CU} = -0.500$$

$$d'_{DU} = -0.375$$

$$d'_{CD} = -0.325$$

# Neighbor joining

## Example

Old matrix.

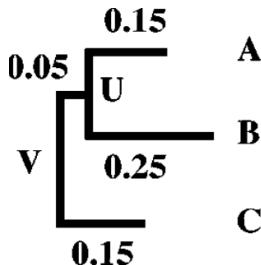
	U	C
C	0.20	
D	0.45	0.55

New matrix.

	U	C
C	-0.5	
D	-0.375	-0.325

# Neighbor joining

## Example



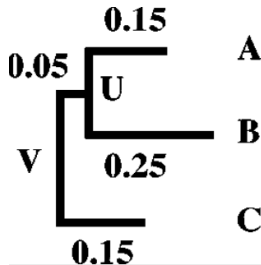
The pair of taxa with the shortest distances in the new matrix are separated ( $CU$ ). Then, use Eq. 4 to calculate  $d_{CV}$  and  $d_{UV}$ .

$$d_{CV} = \frac{d_{CU} + (r'_C - r'_U)}{2} = 0.15$$

$$d_{UV} = \frac{d_{CU} + (r'_U - r'_C)}{2} = 0.05$$

# Neighbor joining

## Example



Because  $D$  is the last branch to be decomposed from the star tree, there is no need to convert to  $r$  and  $r'$ .

$$d_D = \frac{(d_{DU} - d_{UV}) + (d_{DC} - d_{CV})}{2}$$

$$d_D = \frac{(0.45 - 0.05) + (0.55 - 0.15)}{2}$$

$$d_D = 0.4$$

# Questions?

