

# Laboratorio 03a: Árboles Filogenéticos

Frank Roger Salas Ticona

26 de septiembre de 2024

## 1. Introducción

El algoritmo UPGMA (Unweighted Pair Group Method with Arithmetic Mean) y Neighbor Joining son dos de los métodos más utilizados en la construcción de árboles filogenéticos. UPGMA asume tasas de evolución constantes entre todas las ramas, lo que lo hace ideal para datos ultramétricos. Por otro lado, Neighbor Joining es un método más flexible que no requiere esta suposición, lo que lo convierte en una opción robusta para datos con diferentes tasas de evolución. En este trabajo, se implementan ambos algoritmos y se comparan sus rendimientos en distintos conjuntos de datos.

## 2. Implementación

Se utilizó C puro para implementar los algoritmos UPGMA y Neighbor Joining, lo que permitió una manipulación eficiente de las estructuras de datos y un control total sobre el proceso de cálculo. La implementación se optimizó para manejar grandes volúmenes de datos, asegurando que las ejecuciones fueran rápidas y efectivas. En el [link](#) del repositorio de GitHub se encuentran los archivos con la implementación completa.

## 3. Resultados

### 3.1. Experimentos

Se probaron los algoritmos de construcción de árboles filogenéticos utilizando cinco secuencias (S1 a S5), calculando primero la matriz de distancias mediante la Ecuación 1 para ajustar la tasa de sustitución. Las distancias resultantes fueron utilizadas como entrada para UPGMA y Neighbor Joining, generando árboles filogenéticos que evidencian la agrupación correcta de las secuencias según sus relaciones evolutivas.

Se realizaron pruebas con los apellidos de los alumnos y del profesor, calculando la distancia entre ellos basada en las diferencias de letras. Se aplicaron

```

| -0.00000 || 0.82396 || 0.67346 || 0.55715 |
| 0.82396 || -0.00000 || 0.38312 || 0.82396 |
| 0.67346 || 0.38312 || -0.00000 || 1.34382 |
| 0.55715 || 0.82396 || 1.34382 || -0.00000 |

| -0.00000 || 0.74871 || 0.55715 |
| 0.74871 || -0.00000 || 1.08389 |
| 0.55715 || 1.08389 || -0.00000 |

| -0.00000 || 0.86044 |
| 0.86044 || -0.00000 |

(((0, 1), 4), (2, 3))
Cluster 1:
Cluster (0.28, 0.28)
├─ Cluster (0.06, 0.06)
│   ├── {ATTGCCATT}
│   └── {ATGCCATT}
└─ {ACTGACC}
Cluster 2:
Cluster (0.19, 0.19)
├─ {ATCCAATTTT}
└─ {ATCTTCTT}
daerean 雨 Lab 03 (master *) ▶ |

```

Figura 1: Algoritmo UPGMA corriendo con las secuencias de ejemplo.

```

Qmatrix:
| 0.00000 || -3.98887 || -2.44163 || -3.11200 || -4.01807 |
| -3.98887 || 0.00000 || -2.67425 || -3.34461 || -3.55284 |
| -2.44163 || -2.67425 || 0.00000 || -4.77949 || -3.66519 |
| -3.11200 || -3.34461 || -4.77949 || 0.00000 || -2.32447 |
| -4.01807 || -3.55284 || -3.66519 || -2.32447 || 0.00000 |

New Distance Matrix:
| -0.00000 || 0.12026 || 0.55715 || 0.44084 |
| 0.12026 || -0.00000 || 0.55715 || 0.67346 |
| 0.55715 || 0.55715 || -0.00000 || 0.89233 |
| 0.44084 || 0.67346 || 0.89233 || -0.00000 |

Qmatrix:
| 0.00000 || -2.22859 || -2.01058 || -2.24319 |
| -2.22859 || 0.00000 || -2.24319 || -2.01058 |
| -2.01058 || -2.24319 || 0.00000 || -2.22859 |
| -2.24319 || -2.01058 || -2.22859 || 0.00000 |

New Distance Matrix:
| -0.00000 || 0.17644 || 0.50432 |
| 0.17644 || -0.00000 || 0.89233 |
| 0.50432 || 0.89233 || -0.00000 |

Qmatrix:
| 0.00000 || -1.57309 || -1.57309 |
| -1.57309 || 0.00000 || -1.57309 |
| -1.57309 || -1.57309 || 0.00000 |

New Distance Matrix:
| -0.00000 || 0.61011 |
| 0.61011 || -0.00000 |

(((0, 4), 1), (2, 3))
Cluster 1:
Cluster (0.28, -0.11)
├─ Cluster (0.44, -0.00)
│   ├── {ATTGCCATT}
│   └── {ACTGACC}
└─ {ATGCCATT}
Cluster 2:
Cluster (0.23, 0.16)
├─ {ATCCAATTTT}
└─ {ATCTTCTT}
daerean 雨 Lab 03 (master *) ▶ |

```

Figura 2: Algoritmo Neighbor Joining corriendo con las secuencias de ejemplo.

los algoritmos UPGMA y Neighbor Joining para construir un árbol filogenético. La visualización de los árboles resultantes reveló que algunos clústeres son similares, indicando posibles relaciones entre los apellidos.

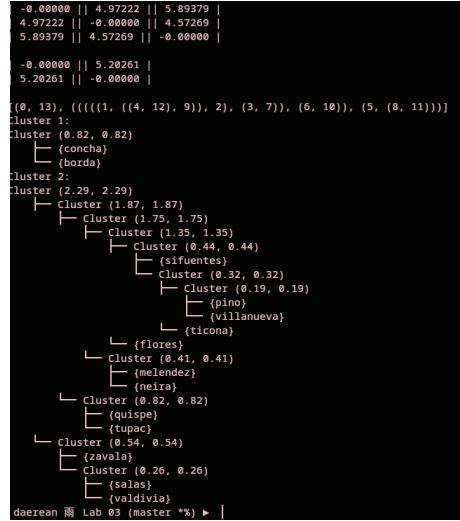


Figura 3: Algoritmo UPGMA corriendo con los apellidos de los integrantes de la clase.

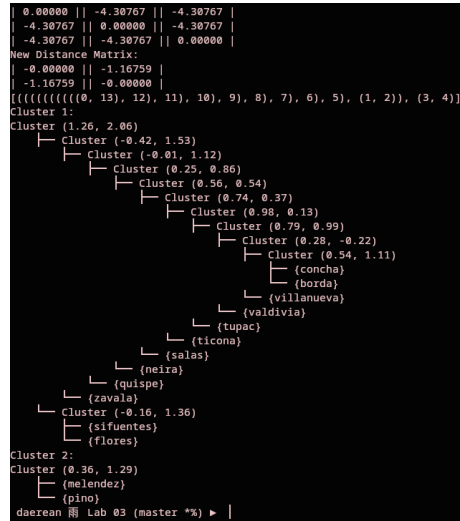


Figura 4: Algoritmo Neighbor Joining corriendo con los apellidos de los integrantes de la clase.

## 4. Análisis y Discusión

Se analizan los árboles filogenéticos generados a partir de las distancias entre secuencias y apellidos. Los algoritmos UPGMA y Neighbor Joining mostraron eficacia en la identificación de clústeres, evidenciando relaciones evolutivas consistentes. La métrica de distancia basada en diferencias de caracteres demostró ser adecuada, resaltando la sensibilidad de los métodos a la variabilidad de los datos.

## 5. Conclusiones

Los resultados validan a UPGMA y Neighbor Joining como métodos robustos para construir árboles filogenéticos, evidenciando agrupaciones coherentes entre secuencias y apellidos. Este enfoque tiene aplicaciones tanto en bioinformática como en estudios genealógicos, sugiriendo direcciones futuras para investigaciones interdisciplinarias.