

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Bioinformatika
1 Labaratorinis

Darbą atliko:
Rimgaudas Vėjelis

Vilnius, 2025

1. Kodo repozitorija: <https://github.com/RimgaudasV/Bioinformatics>

2. Atstumo matricos skaičiavimas

Atstumo matricai sudaryti buvo naudojama klasikinė Euklido atstumo formulė

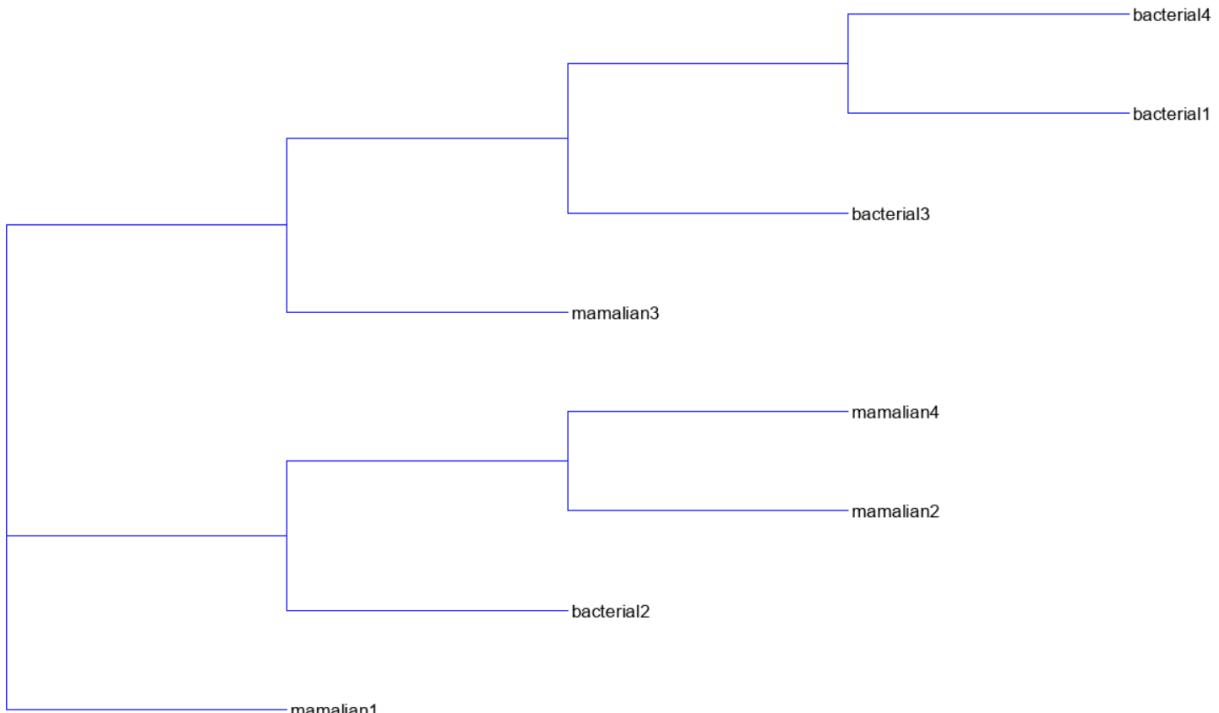
Kiekvieno failo seka buvo lyginama su kito failo seka. Šioje formulėje pi yra vieno failo kodonu/dikodonu dažnis, qi – kito, o i atitinka amino rugštį, kuri yra palyginama šiuo metu.

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

1 pav. Paprastoji Euklido formulė

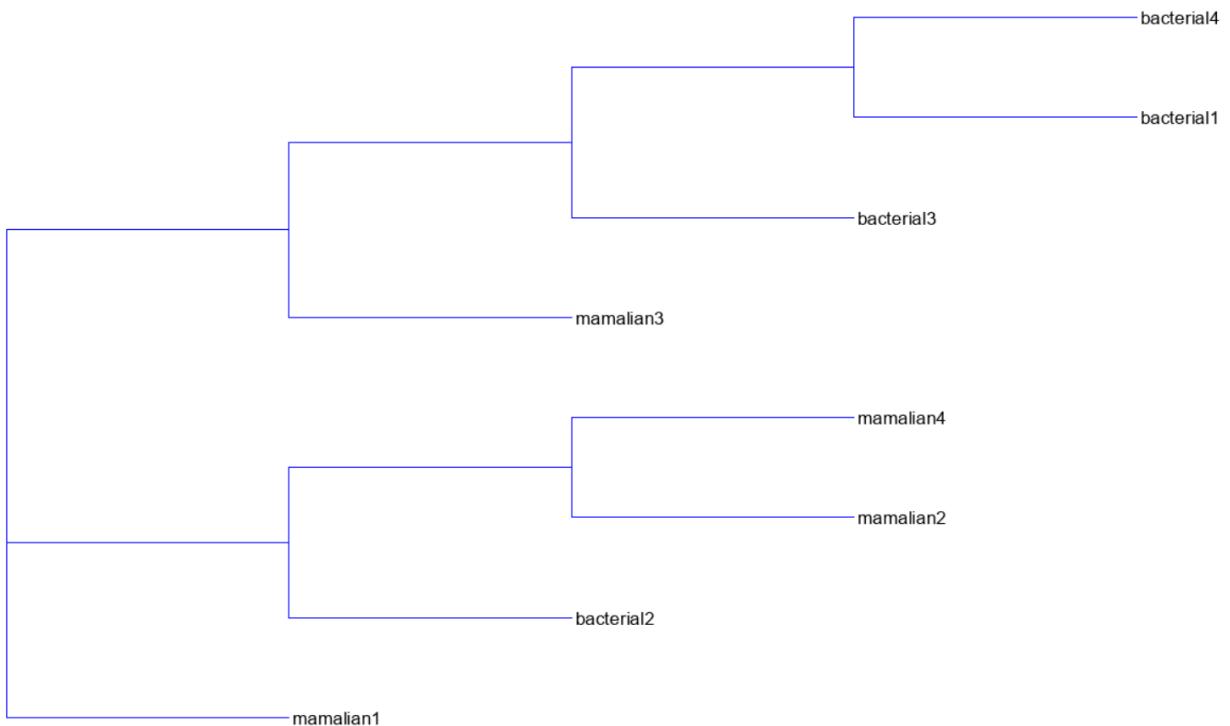
Taigi d yra atstumas tarp dviejų lyginamų sekų. Iš d sudaroma matrica kur matomas sekos palyginimas su kita seka – pagrindinė matrices istrižainė yra 0.0 kadangi seka lyginama su savimi.

3. Kodonų medis



2 pav. Medis, rodantis kodonų klasterizavimą neighbour joining metodu

4. Dikodonų medis



3 pav. Medis, rodantis dikodonų klasterizavimą neighbour joining metodu

5. Išvada/rezultatai

Dikodonų analizė sumažina atstumų skirtumus, bet medžių topologija lieka identiška, todėl medžiai yra visiškai vienodi.

Kodonų ir dikodonų dažniai skiriasi tarp žinduolinių ir bakterinių virusų, todėl dažniausiai virusai klasterizuojasi pagal šeimininkus (išimtys: bacteria2 ir mamalian3 klaseterizuoja ne pagal šeimininką). Labiausiai išsiskiria mamalian1 virusas, kuris nėra panašus nei į bakterinius virusus nei į žinduolinius virusus.