

Bioinformatika

1 laboratorinis darbas

Gustas Juškevičius

4 kursas 5 grupė

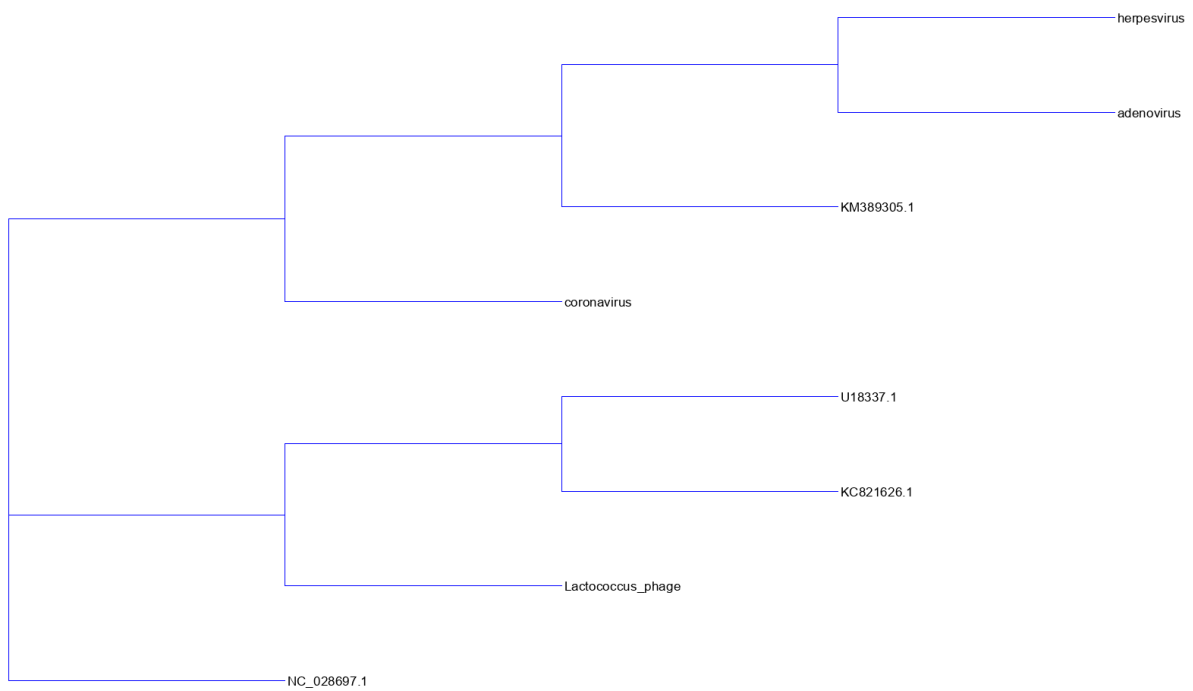
Atstumo matricos skaičiavimas:

1. Suskaičiuojame, kiek per visas sekas yra amino rūgščių.
Amino rūgštys: A, R, N, D, C, E, Q, G, H, I, L, K, M, F, P, S, T, W, Y, V, *.
2. Suskaičiuojame, kiek sekoje yra amino rūgščių ir padaliname iš viso amino rūgščių skaičiaus.
3. Konvertuojame visas amino rūgščių dažnių reikšmes į vektorių (taip su visomis 8 fasta sekomis)
4. Imame po 2 vektorius ir skaičiuojame jų atstumą, remdamiesi Euklidiniu atstumu

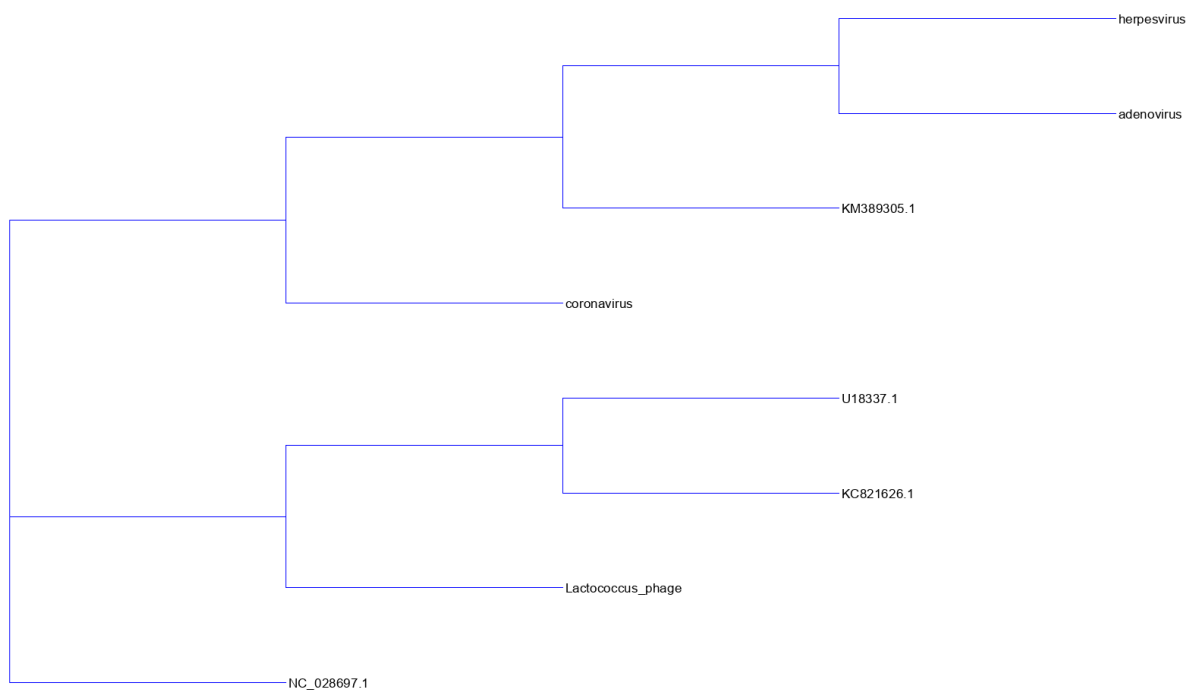
$$d(X, Y) = \|X - Y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

5. Visas reikšmes (iš viso 64) išrašome Phylip formatu.

Kodonų medis



Dikodonų medis



Išvados:

Medžiai gauti vienodi, tad tai gali rodyti, kad virusai turi panašias genominės struktūras.

Iš medžio matome, kad kai kurie virusai yra artimai susiję (pvz., koronavirusas ir Lactococcus phage), o kiti, pavyzdžiui, herpesvirus ir adenovirus, atrodo, yra labiau atskirti. Tai gali rodyti, kad skirtingi virusai turėjo skirtingas evoliucines keliones.

Dažniausiai pasirodantis kodonai yra Leucinas (L), Serinas (S) ir Isoleucinas (I)

Dekodonai dažniausi yra LL ir SS