

VILNIUS UNIVERSITY  
MATEMATIKOS INFORMATIKOS FAKULTETAS  
PROGRAMŲ SISTEMŲ KATEDRA

Bioinformatika  
**Pirmas laboratorinis darbas**

Ataskaitą parengė:  
Vilius Macijauskas

## Turinys

1. Atstumo skaičiavimo funkcija .....	3
2. Atstumo skaičiavimo rezultatai .....	4
1. Kodonų atstumai .....	4
2. Dikodonų atstumai .....	5
3. Medžiai .....	5
3. Kodonų atstumų medis .....	5
4. Dikodonų atstumų medis .....	6
4. Rezultatai .....	6
5. Programos kodas.....	7

## 1. Atstumo skaičiavimo funkcija

Suskaičiuojame kodonų bei dikodonų dažnius kiekvienai sekai (1 bei 2 pav.). Lyginame juos atimdami dažnius vienodų kodonų/dikodonų lyginamoms sekoms, o skirtumas įdedamas į modulį, kad būtų panaikinamas ženklas. Taip gaunamas nuokrypis (kiek viena seka skiriasi dažniais nuo kitos) (3 pav.). Nuokrypiai sudedami ir gaunamas atstumas tarp dviejų sekų.

```
def findCodonsFrequency(sequences):
    frequencies = {}
    codonCount = 0
    i = 0

    for codon in ALL_CODONS:
        frequencies[codon] = 0

    for clipping in sequences:
        while i < len(clipping) - 2:
            codon = clipping[i:i + 3]
            frequencies[codon] += 1
            codonCount += 1
            i += 1

    for codon in ALL_CODONS:
        frequencies[codon] = round(frequencies[codon] / codonCount, 3)

    return frequencies
```

1 pav. „findCodonsFrequency“ metodas

```
def findDicodonsFrequency(sequences):
    frequencies = {}
    dicodonCount = 0
    i = 0

    for dicodon in ALL_DICODONS:
        frequencies[dicodon] = 0

    for clipping in sequences:
        while i < len(clipping) - 5:
            dicodon = clipping[i:i + 6]
            frequencies[dicodon] += 1
            dicodonCount += 1
            i += 1

    for dicodon in ALL_DICODONS:
        frequencies[dicodon] = round(frequencies[dicodon] / dicodonCount, 3)

    return frequencies
```

2 pav. „findDicodonsFrequency“ metodas



```
def compareFrequencies(frequencies1, frequencies2, codons):  
    distance = 0  
    for codon in codons:  
        distance += abs(frequencies1[codon] - frequencies2[codon])  
    return distance
```

3 pav. „compareFrequencies” metodos

## 2. Atstumo skaičiavimo rezultatai

### 1. Kodonų atstumai

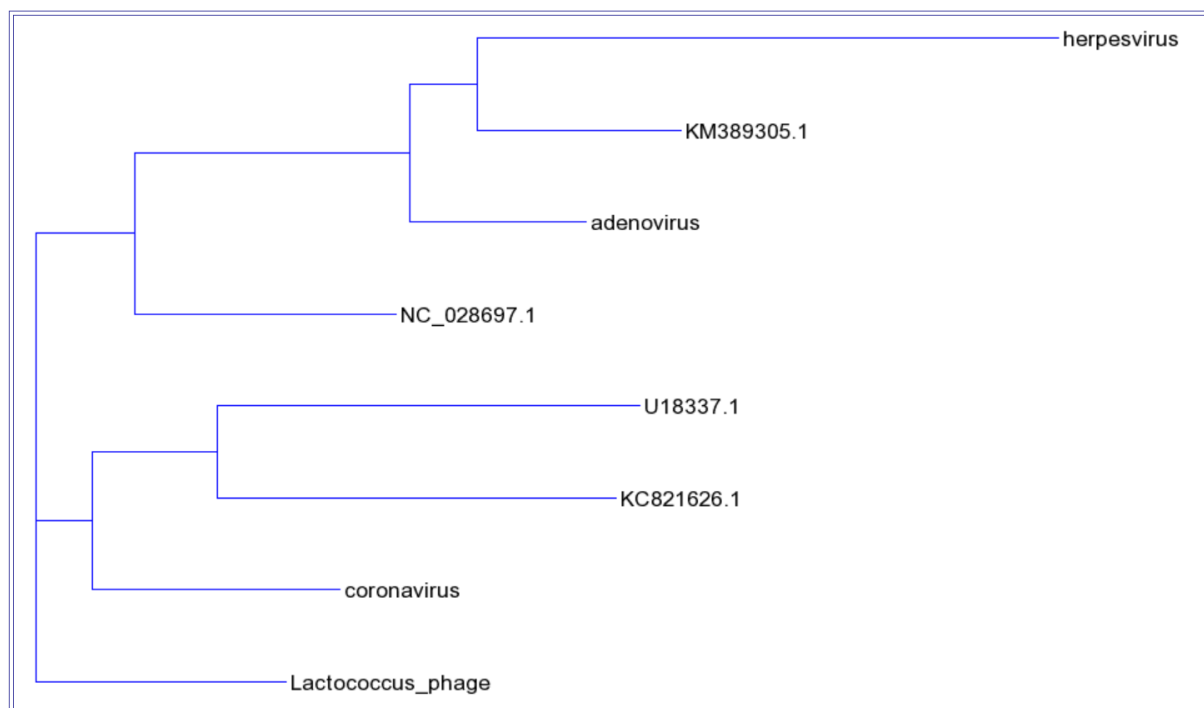
	Lactococcus_phage	KM389305.1	NC_028697.1	KC821626.1	coronavirus	adenovirus	U18337.1	herpesvirus
Lactococcus_phage	0.0	0.426	0.279	0.363	0.328	0.419	0.431	0.67
KM389305.1	0.426	0.0	0.421	0.607	0.464	0.277	0.599	0.394
NC_028697.1	0.279	0.421	0.0	0.46	0.333	0.334	0.522	0.603
KC821626.1	0.363	0.607	0.46	0.0	0.401	0.602	0.412	0.825
coronavirus	0.328	0.464	0.333	0.401	0.0	0.389	0.385	0.666
adenovirus	0.419	0.277	0.334	0.602	0.389	0.0	0.588	0.361
U18337.1	0.431	0.599	0.522	0.412	0.385	0.588	0.0	0.805
herpesvirus	0.67	0.394	0.603	0.825	0.666	0.361	0.805	0.0

## 2. Dikodony atstumai

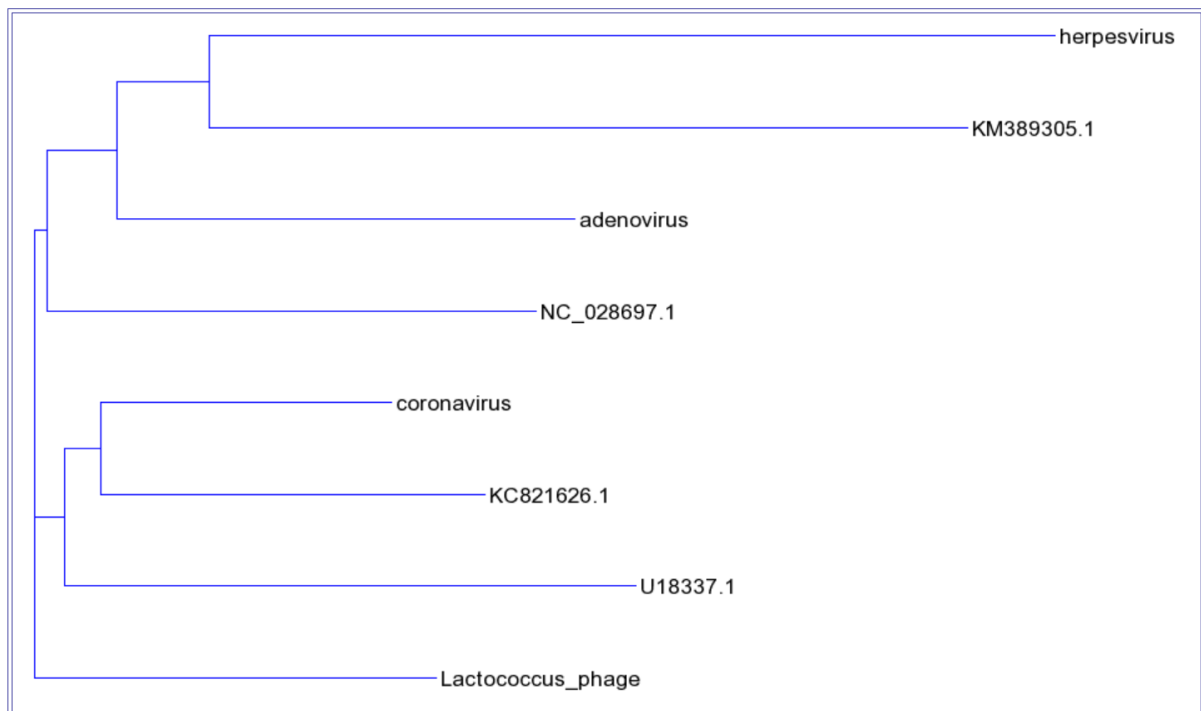
	<b>Lactococcus_phage</b>	<b>KM389305.1</b>	<b>NC_028697.1</b>	<b>KC821626.1</b>	<b>coronavirus</b>	<b>adenovirus</b>	<b>U18337.1</b>	<b>herpesvirus</b>
<b>Lactococcus_phage</b>	0.0	1.689	1.169	1.044	0.983	1.22	1.332	1.839
<b>KM389305.1</b>	1.689	0.0	1.812	1.823	1.72	1.735	1.837	2.062
<b>NC_028697.1</b>	1.169	1.812	0.0	1.219	1.048	1.323	1.469	1.912
<b>KC821626.1</b>	1.044	1.823	1.219	0.0	0.867	0.867	1.25	1.929
<b>coronavirus</b>	0.983	1.72	1.048	0.867	0.0	1.099	1.179	1.782
<b>adenovirus</b>	1.22	1.735	1.323	1.27	1.099	0.0	1.5	1.739
<b>U18337.1</b>	1.332	1.837	1.469	1.25	1.179	1.5	0.0	2.089
<b>herpesvirus</b>	1.839	2.062	1.912	1.929	1.782	1.739	2.089	0.0

## 3. Medžiai

### 3. Kodonų atstumų medis



#### 4. Dikodony atstumų medis



## 4. Rezultatai

- Ar skiriasi kodonų ir dikodonų dažnis tarp žinduolių ir bakterijų virusų, kaip klasterizuojasi virusai?

Kodonų ir dikodonų dažnis realiai nesiskiria tarp žinduolių ir bakterijų. Tai bakterijos ar žinduolio virusas klasterizavimuisi įtakos neturėjo. Žinduolių: herpesvirusas bei adenovirusas pateko į vieną klasterį su bakterijų virusais: NC\_028697.1 bei KM389305.1, o kitame klasteryje žinduolių: koronavirusas bei *Lactococcus*\_phage pateko kartu su bakterijų: KC821626.1 ir U18337.1.

- Gal kažkuris virusas labai išsiskyrė?

Herpes virusas, nes jis labiausiai nutolęs nuo kitų virusų pagal kodonų atstumus.

- Kokie kodonai/dikodonai labiausiai varijuoja?

Kodonų bei dikodonų pasikartojimo skaičius visose sekose susumavus (10 dažniausiai pasikartojusių):

<b>Kodonai</b>	<b>Dikodonai</b>
('TTT', 972)	('AAAAAA', 45)
('AAA', 946)	('AAAAAT', 36)
('TTA', 829)	('TTAATG', 34)
('ATG', 784)	('TTAAAA', 32)
('AAT', 757)	('TTTAAT', 32)
('TAT', 750)	('GAAAAA', 32)
('TGT', 732)	('AAAAAG', 31)
('ATT', 728)	('TTTTAA', 31)
('TTG', 722)	('TGGTAA', 31)
('TAA', 684)	('TAAAAA', 30)

## 5. Programos kodas

Google colab nuoroda:

<https://colab.research.google.com/drive/18QG5F4pEFE40R9jfrdQNtdX4z0f3760r?usp=sharing>