**BIOINFORMATIKOS ANTROS UŽDUOTIES ATASKAITA**

Visa užduotis buvo daryta pagal dėstytojo pateiktus punktus. Taip kaip reikėjo daryti su Linux sistema, tai dariau su Windows, nes neišeidavo viską padaryti Linux‘e dėl techninių kliūčių, todėl gali trukti kai kurių failų. Todėl buvo atsiųstos ir/ar naudotos šios programos:

* Cygwin64 Terminal (buvo naudota mafft to dėka)
* <http://etetoolkit.org/treeview/>
* <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch>
* <https://github.com/niu-lab/gclust> (buvo naudota Linux‘e, nes Command Prompt‘e negalėjo suprasti ./gclust)
* <https://bioinf.shenwei.me/seqkit/download/> (tam, kad galima būtų naudoti Command Prompt)

File‘ų pavadinimai ir paaiškinimai:

* NC\_045515.fasta – pirmoje dalyje atsiųstas nukleotidas.
* seq.fasta – antroje dalyje gautas genomas.
* MN514967\_1.fasta – kupranugario virusas atsiųstas antroje dalyje.
* finalSeq.fasta – sujungta seq.fasta, NC\_045515.fasta ir MN514967\_1.fasta seka, kuri gauta antroje dalyje.
* sortedFinalSeq.fasta – surūšiuotas genomus mažėjančia ilgio tvarka trečioje dalyje.
* sortedFinalSeq.fasta.clustering.out – po cluster trečioje dalyje.
* finalSeq.seqkit.extract.out – ištraukti id iš rinkinio trečioje dalyje.
* D3W8N4.fasta – atsiųstas ketvirtoje dalyje proteinas.
* alignedSeq.fasta – gauta seka po tblastn ketvirtoje dalyje.
* translatedSeq.fasta – po translate gauta seka ketvirtoje dalyje.
* filteredSeq.fasta – po filtravimo gauta seka ketvirtoje dalyje.
* MafftAligned.fasta – po lygiavimo su mafft ketvirtoje dalyje.
* tree.txt – sugeneruotas medis ketvirtoje dalyje.
* bioinfor2.ipynb – penktos dalies kodas, kur yra pateiktas medis prieš kupranugario viruso prijungimą ir po. Per šį failą geriausiai bus pamatyti tuos medžius negu čia darydama copy paste, nes užims porą puslapių.

Klausimai ir atsakymai:

* How did the Covid-19 evolve, what path through hosts was taken?
  + Galime matyti, kad Covid-19 yra artimas šiksnosparnių ir pangolin koronovirusams. Situacija tokia pati tirk pridėjus, tiek ne kupranugario viruso šaknies. Negalima iš šių medžių pasakyti, kad Covid-19 yra giminingas camel virusui.

A picture containing timeline

Description automatically generated

* Would it be different interpretation if out-group is not used?
  + Taip, nes matytųsi tada skirtingai giministės ryšiai.
* What about Urbani SARS origin?
  + Matosi, kad Urbani SARS yra labiau giminingas SARS viruso variantams, negu šikšnosparnio ir kupranugario virusams. Ryškaus ryšio tarp Urbani SARS ir Palm Civet su Batcoranovirus nėra.

A picture containing diagram

Description automatically generated

* Is the Palm Civet origin evident?
  + Taip, medžiose yra matoma, kad yra giminingas šiknosparnio virusui.
* Why we exclude taxid 2697049?
  + Nes jeigu to nepadarysime, tada visos paieškos rezultatai būtų įvairūs Covid-19 viruso variantai.