VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

PROGRAMŲ SISTEMŲ BAKALAURO STUDIJŲ PROGRAMA

TREČIOJO LABORATORINIO DARBO ATASKAITA

Bioinformatika

III – asis laboratorinis darbas

Atliko: 4 kurso 3 grupės studentas Lukas Orliukas

Vilnius – 2023

Turinys

[1. Darbo eiga 3](#_Toc152867214)

[2. Darbo kodas 5](#_Toc152867215)

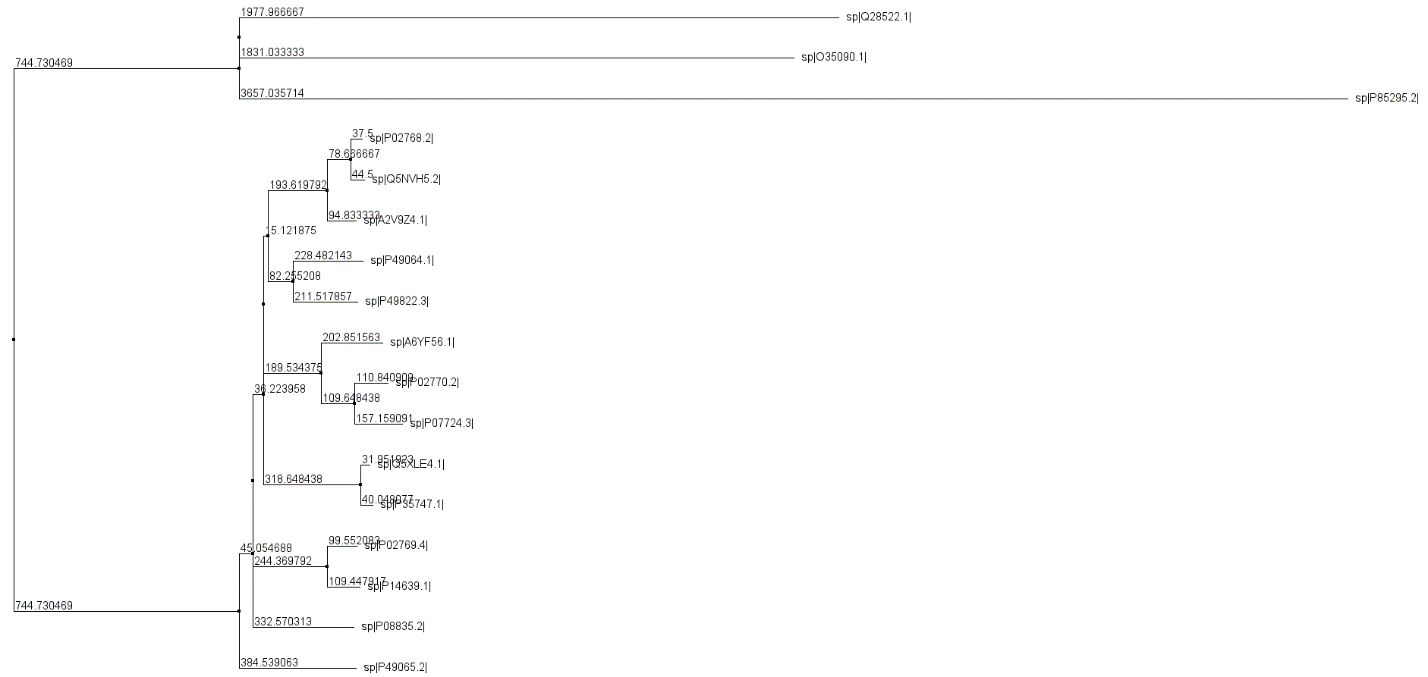
# Darbo eiga

* 1. Baltymo prieigos kodas yra NP\_000468.1
  2. Šio baltymo seka:

>NP\_000468.1 albumin preproprotein [Homo sapiens]

MKWVTFISLLFLFSSAYSRGVFRRDAHKSEVAHRFKDLGEENFKALVLIAFAQYLQQCPFEDHVKLVNEVTEFAKTCVADESAENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNECFLQHKDDNPNLPRLVRPEVDVMCTAFHDNEETFLKKYLYEIARRHPYFYAPELLFFAKRYKAAFTECCQAADKAACLLPKLDELRDEGKASSAKQRLKCASLQKFGERAFKAWAVARLSQRFPKAEFAEVSKLVTDLTKVHTECCHGDLLECADDRADLAKYICENQDSISSKLKECCEKPLLEKSHCIAEVENDEMPADLPSLAADFVESKDVCKNYAEAKDVFLGMFLYEYARRHPDYSVVLLLRLAKTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQNLIKQNCELFEQLGEYKFQNALLVRYTKKVPQVSTPTLVEVSRNLGKVGSKCCKHPEAKRMPCAEDYLSVVLNQLCVLHEKTPVSDRVTKCCTESLVNRRPCFSALEVDETYVPKEFNAETFTFHADICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVMDDFAAFVEKCCKADDKETCFAEEGKKLVAASQAALGL

* 1. Pagal nurodyta užklausa atlikus paieška gautas 61 rezultatas, tačiau į tai įeina: albumin, serum albumin ir serum albumin – like. Kadangi vėliau kode naudojamas raktažodis albumin, kadangi su juo grąžinamos sekos atlikus BLAST paieška ir taip pat surasta baltymo seka, pateikta c dalyje (kuri buvo kaip ir patvirtinta dėstytojo paskaitos metu, kad turėtų būti teisinga). Darau prielaida, kad tai, ko prašo užduotis dabar yra vadinama tiesiog albuminu, nes pagal pateikta užklausos pavyzdį pirmiausia pateikiami albumin ir rasti tik keli serum albumin. Tada iš 61 rezultato lieka 25, kurių pavadinimas yra ALB arba Alb. Abiejų aprašymas nurodo, kad tai albuminai.
  2. Atliekama BLAST paieška su nurodytais parametrais: žinduolių sekos, „swissprot“ duomenų bazė.
  3. Po atliktos BLAST paieškos išrenkamos sekos su bent 80% padengimu ir tik albuminai ir jie išsaugomi į failą.
  4. Gautiems albuminams su Jalview surandamas palyginys, naudojant MAFT, ir „Neighbour joining“ metodu, bei naudojant „BLOSUM62“ suskaičiuojamas filogenetinis medis, kurį galima matyti 1 pav.



1 pav. Filogenetinis medis

Pagal gautą palyginį ir medį į žmogaus albuminą (kodas medyje yra sp|P02768.2|) panašiausias yra sumatrinio orangutano [Pongo abelii] (kodas medyje yra sp|Q5NVH5.2|) albuminas.

* 1. Iš MAFT palyginio grafiko ieškota originaliausia žmogaus albumino seka. Surasta seka buvo pozicijose 140-160: VRPEVDVMCTAFHDNEETFLK. Sekos konservatyvumas matomas 2 pav.



2 pav. Originaliausios sekos konservatyvumas

* 1. Iš MAFT palyginio grafiko ieškota panašiausia į kitų žinduolių žmogaus albumino seka. Surasta seka buvo pozicijose 429-449: NALLVRYTKKVPQVSTPTLVE. Sekos konservatyvumas matomas 3 pav.

A yellow bars on a white background

Description automatically generated

3 pav.Panašiausios sekos konservatyvumas

* 1. Pateikta programa įgyvendina BLAST paieška, paieškos rezultatų atrinkimą, suskaičiuoja palyginį ir filogenetinį medį.
  2. Parenka fragmentus:
     1. Originaliausia žmogui seka: DVMCTAFHDNEETFL
     2. Panašiausia su kitais žinduoliais seka: KKVPQVSTPTLVEVS

1. Fragmento paieškai buvo naudojama „biopython“ viename iš paketų esanti „information\_content“ funkcija. Su ja buvo apskaičiuojama reikšmė kiekvienam pasirinkto ilgio palyginio fragmentui ir ieškoma didžiausios bei mažiausios reikšmės einat per palyginį.

# Darbo kodas

Programa yra pateika faile 3lab.py.