**Ataskaita**

**3 laboratorinis darbas**

**Darbą atliko: Aleksandra Kondratjeva**

## Santrauka

[Programos veikimas ir skaičiavimo algoritmas 7](#_Toc152769698)

[Programos skaičiavimų rezultatai 7](#_Toc152769699)

## Darbo eiga

1.Išnagrinėjau prieinamą atitinkamo geno informaciją.

a) Atlikau paieška "Gene" duomenų bazėje NCBI portale http://www.ncbi.nlm.nih.gov/. Panaudojau užklausą ALB[title] AND human[Organism] .

b) Baltymo prieigos kodas: **„NCBI Reference Sequence: NP\_000468.1”**

c) Seką išsaugota fasta formatu.

d) Pagal užklausą "serum albumin"[Protein name] AND mammals[Organism] "Gene" duomenų bazėje NCBI rastas 61 rezultatas, iš viso skirtingų organizmų 40:

1. Homo sapiens (human)
2. Mus musculus (house mouse)
3. Rattus norvegicus (Norway rat)
4. Bos taurus (cattle)
5. Canis lupus familiaris (dog)
6. Sus scrofa (pig)
7. Equus caballus (horse)
8. Macaca mulatta
9. (Rhesus monkey)
10. Ovis aries (sheep)
11. Oryctolagus cuniculus (rabbit)
12. Felis catus (domestic cat)
13. Meriones unguiculatus (Mongolian gerbil)
14. Bos mutus (wild yak) Heterocephalus glaber (naked mole-rat)
15. Macaca fascicularis (crab-eating macaque)
16. Camelus dromedarius (Arabian camel)
17. Capra hircus (goat)
18. Cricetulus griseus (Chinese hamster)
19. Echinops telfairi (small Madagascar hedgehog)
20. Equus asinus (ass)
21. Fukomys damarensis (Damara mole-rat)
22. Ornithorhynchus anatinus (platypus)
23. Panthera tigris (tiger)
24. Pongo abelii (Sumatran orangutan)
25. Pteropus alecto (black flying fox)
26. Trichechus manatus latirostris (Florida manatee)
27. Ictidomys tridecemlineatus (thirteen-lined ground squirrel)
28. Chinchilla lanigera (long-tailed chinchilla)
29. Myotis brandtii (Brandt's bat)
30. Myotis lucifugus (little brown bat)
31. Myotis davidii
32. Dipodomys ordii (Ord's kangaroo rat)
33. Miniopterus natalensis
34. Castor canadensis (American beaver)
35. Phascolarctos cinereus (koala)
36. Urocitellus parryii (Arctic ground squirrel)
37. Vombatus ursinus (common wombat)
38. Marmota flaviventris (yellow-bellied marmot)
39. Grammomys surdaster
40. Arvicanthis niloticus (Nile rat)

2.Gavau ŽSA ir panašių sekų žinduoliuose sekų palyginį. Tolimesnei analizei naudojau ŽSA baltymo seką.

a-c) Suradau 35 į ŽSA panašias žinduolių sekas ("mammals"), naudojau „SwissProt“ duomenų bazę. Atmetus atitikmenis, kurie apima mažiau nei 80 % sekos ilgio, gavau 34 sekas. Papildomai atmetus sekas, kurios nėra albuminai (vitaminai, fetoproteinai, afaminai), gavau 16 sekų kurias išsaugojau tolimesniam naudojimui (**manual\_result/seqdump.fasta**).

Toliau, naudodama MAFFT server ir JalView programą, gavau palyginį (**manual\_result/mafft\_L-INS-i.fasta** failas) bei filogenetinį medį.

Iš filogenetinio medžio matome, jog panašiausias serumo albumino baltymas į žmogaus yra **Pongo abelii**.

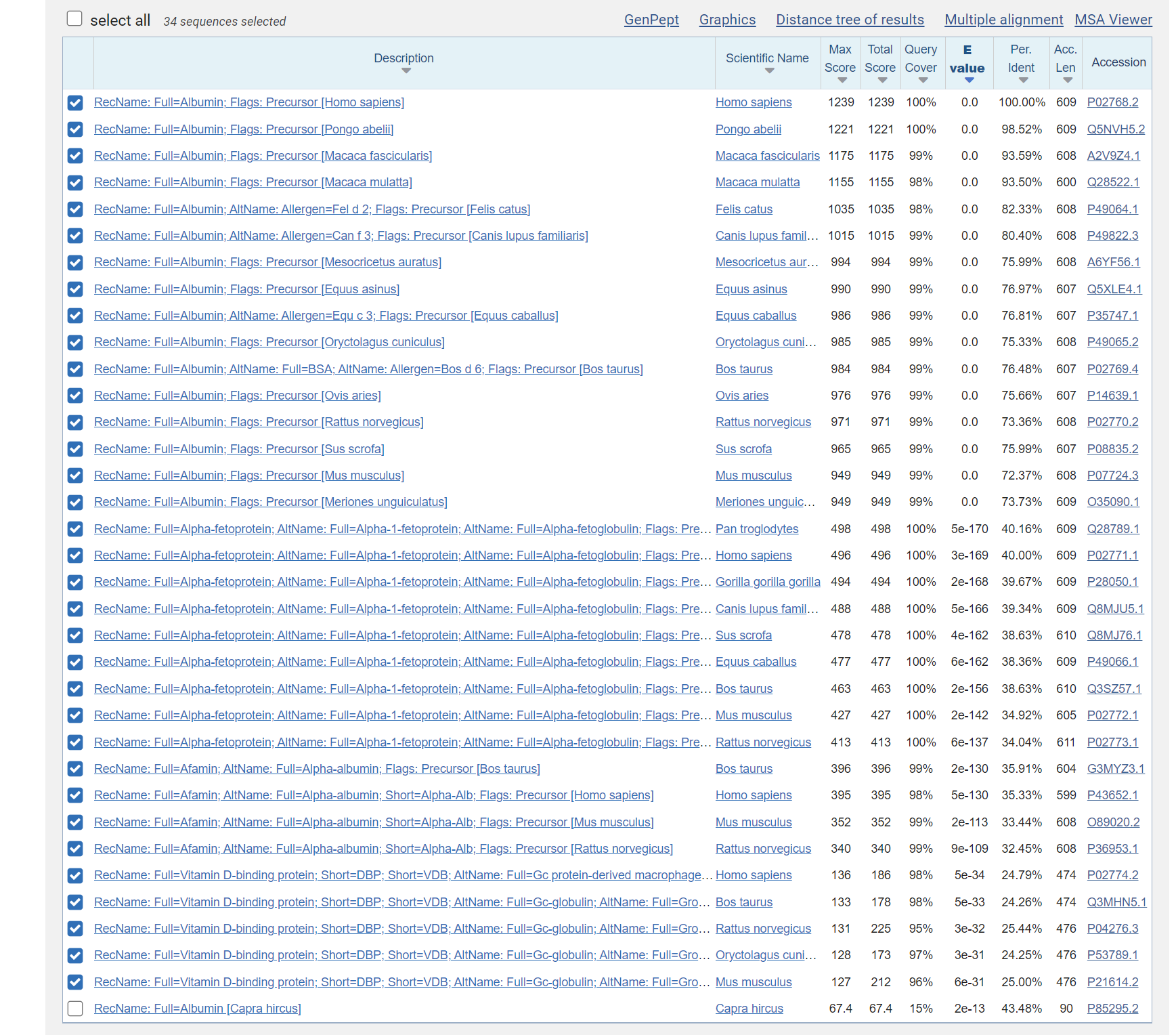


Figure 1. Surastos sekos

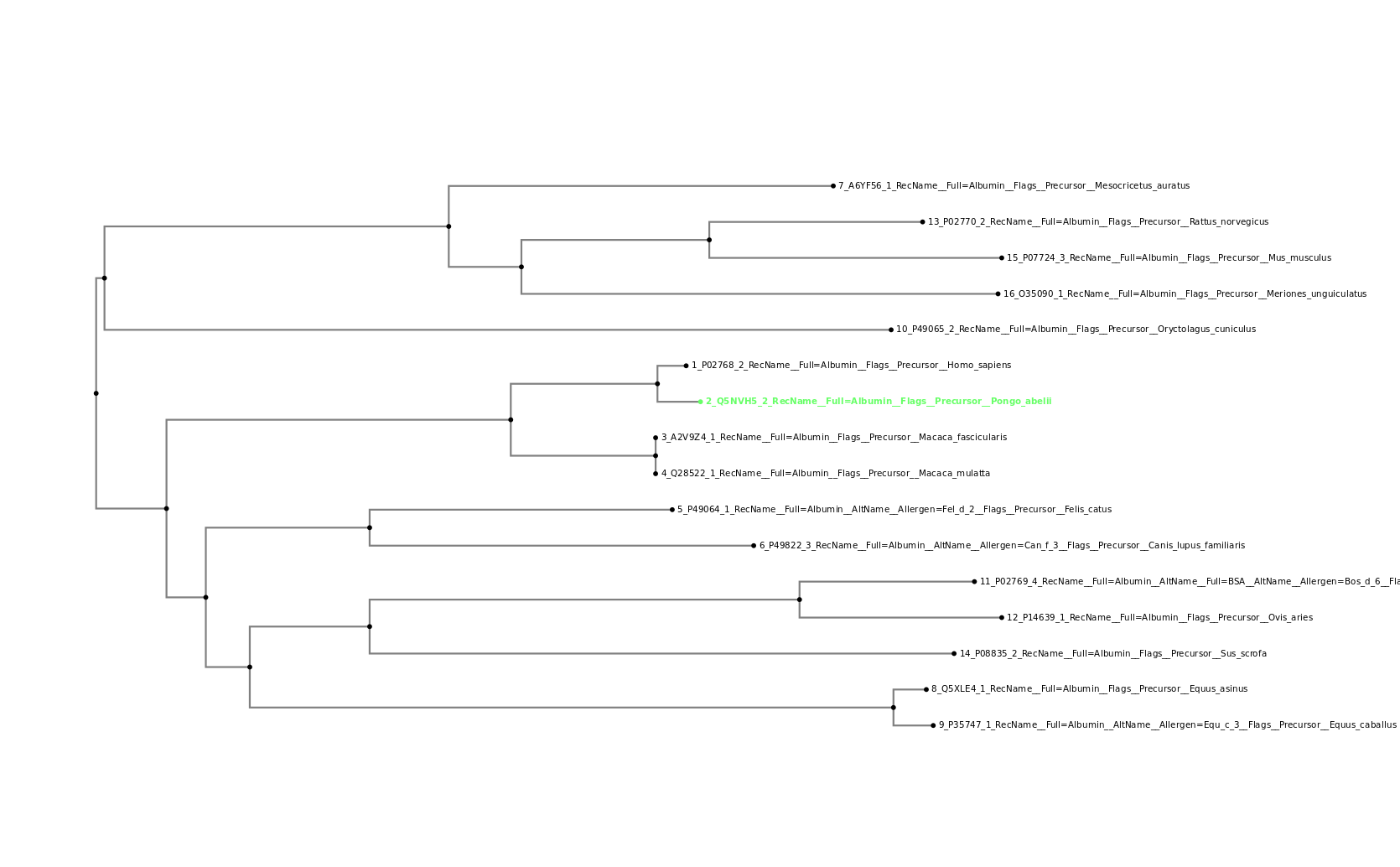


Figure 2. Filogenetinis medis

Iš filogenetinio medžio matome, jog panašiausias serumo albumino baltymas į žmogaus yra **Pongo abelii**.

A screenshot of a computer

Description automatically generated

Figure 3. JalView

1. Iš akies sunku tiksliai nustatyti, kuri seka būtų kiek įmanoma skirtinga nuo kitų žinduolių sekų, tačiau galime daryti prielaidą, kad šis fragmentas (**NPNLPRLVRPEVDVMCTAFH**), turintis žemą ir nepastovų konservatyvumą yra tinkamiausias.

A graph with numbers and a bar chart

Description automatically generated

Figure 4. Konservatyvumas nepanašiausios sekos

1. Taip pat, galime daryti prielaidą ir rasti fragmentą, kuris galėtų būti panašiausias į kitas žinduolių sekas (**ECCHGDLLECADDRADLAKY**).

A yellow rectangle with stars

Description automatically generated

Figure 5. Konservatyvumas panašiausios sekos

## Programos veikimas ir skaičiavimo algoritmas

* **Baltymų sekų paieška naudojant BLAST.** Nuskaitoma žmogaus albumino seka iš FASTA failo. Tada atliekama BLAST paieška „SwissProt“ duomenų bazėje, ribojant paiešką tik iki žindolių. Rezultatai išsaugomi blast.xml faile.
* **Sekų filtravimas**. Iš BLAST rezultatų išrenkamos tik tos sekos, kurios yra albumino ir turi „query coverage“ ne mažiau nei 80%.
* **Palyginio generavimas naudojant MAFFT.**
* **Sekų panašumo ir unikalumo analizė**. Visi esminiai žingsniai atliekami *`find\_sequences`* funkcijoje. Paieškai naudojau fiksuoto dydžio langą (15 simboliu), kuris perstumiamas per visą žmogaus albumino seką ir palyginamas su kitų sekų atitinkamomis dalimis. Jei langelyje atsirasdavo gap‘ų, pasiimdavau tiek papildomu simbolių, kiek iš viso buvo gap‘ų. Sekų palyginimui naudojau BLOSUM62 matricą, kuri suteikia taškus kiekvienai aminorūgščių porai pagal jų pakeitimo tikimybę. Kad gaučiau galutinį rezultatą, skaičiavau bendrą balą kiekvienam žmogaus sekos fragmentui, sudedant kiekvieno žinduolio ir žmogaus atitinkamo fragmento palyginimo rezultatą. Galiausiai, grąžinamas labiausiai unikalus ir labiausiai panašus fragmentas, kartu su jų atitinkamais balais.

## Programos skaičiavimų rezultatai

Most unique sequence: RLVRPEVDVMCTAFH 493.0

Most similar sequence: ECCHGDLLECADDRA 1350.0  
  
Galime teigti, kad gauti rezultatai atitinka konservatyvumo grafiką ir prielaidos buvo teisingos. Matome, jog seka **` RLVRPEVDVMCTAFH`** yra dalis ` **NPNLPRLVRPEVDVMCTAFH`** sekos ir seka **` ECCHGDLLECADDRA`** yra dalis ` **ECCHGDLLECADDRADLAKY`** sekos.