

Ataskaita

3 laboratorinis darbas

Darbą atliko: Aleksandra Kondratjeva

Santrauka

Programos veikimas ir skaičiavimo algoritmas.....	7
Programos skaičiavimų rezultatai.....	7

Darbo eiga

1. Išnagrinėjau prieinamą atitinkamo geno informaciją.

a) Atlikau paiešką "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 naudoju ŽSA baltymo seką.

a-c) Suradau 35 į ŽSA panašias žinduolių sekas ("mammals"), naudoju „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**.

<input type="checkbox"/> select all 34 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer										
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Homo sapiens]	Homo sapiens	1239	1239	100%	0.0	100.00%	609	P02768.2	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Pongo abelii]	Pongo abelii	1221	1221	100%	0.0	98.52%	609	Q5NVH5.2	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Macaca fascicularis]	Macaca fascicularis	1175	1175	99%	0.0	93.59%	608	A2V824.1	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Macaca mulatta]	Macaca mulatta	1155	1155	98%	0.0	93.50%	600	Q28522.1	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; AltName: Allergen=Fel d 2; Flags: Precursor [Felis catus]	Felis catus	1035	1035	98%	0.0	82.33%	608	P49064.1	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; AltName: Allergen=Can f 3; Flags: Precursor [Canis lupus familiaris]	Canis lupus famil...	1015	1015	99%	0.0	80.40%	608	P49822.3	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Mesocricetus auratus]	Mesocricetus aur...	994	994	99%	0.0	75.99%	608	A6YF56.1	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Equus asinus]	Equus asinus	990	990	99%	0.0	76.97%	607	Q5XLE4.1	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; AltName: Allergen=Equ c 3; Flags: Precursor [Equus caballus]	Equus caballus	986	986	99%	0.0	76.81%	607	P35747.1	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Cryotolagus cuniculus]	Cryotolagus cuni...	985	985	99%	0.0	75.33%	608	P49065.2	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; AltName: Full=BSA; AltName: Allergen=Bos d 6; Flags: Precursor [Bos taurus]	Bos taurus	984	984	99%	0.0	76.48%	607	P02769.4	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Ovis aries]	Ovis aries	976	976	99%	0.0	75.66%	607	P14639.1	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Rattus norvegicus]	Rattus norvegicus	971	971	99%	0.0	73.36%	608	P02770.2	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Sus scrofa]	Sus scrofa	965	965	99%	0.0	75.99%	607	P08835.2	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Mus musculus]	Mus musculus	949	949	99%	0.0	72.37%	608	P07724.3	
<input checked="" type="checkbox"/>	RecName: Full=Albumin; Flags: Precursor [Meriones unguiculatus]	Meriones unguic...	949	949	99%	0.0	73.73%	609	Q35090.1	
<input checked="" type="checkbox"/>	RecName: Full=Alpha-fetoprotein; AltName: Full=Alpha-1-fetoprotein; AltName: Full=Alpha-fetoglobulin; Flags: Precursor [Pan troglodytes]	Pan troglodytes	498	498	100%	5e-170	40.16%	609	Q28789.1	
<input checked="" type="checkbox"/>	RecName: Full=Alpha-fetoprotein; AltName: Full=Alpha-1-fetoprotein; AltName: Full=Alpha-fetoglobulin; Flags: Precursor [Homo sapiens]	Homo sapiens	496	496	100%	3e-169	40.00%	609	P02771.1	
<input checked="" type="checkbox"/>	RecName: Full=Alpha-fetoprotein; AltName: Full=Alpha-1-fetoprotein; AltName: Full=Alpha-fetoglobulin; Flags: Precursor [Gorilla gorilla gorilla]	Gorilla gorilla gorilla	494	494	100%	2e-168	39.67%	609	P20850.1	
<input checked="" type="checkbox"/>	RecName: Full=Alpha-fetoprotein; AltName: Full=Alpha-1-fetoprotein; AltName: Full=Alpha-fetoglobulin; Flags: Precursor [Canis lupus familiaris]	Canis lupus famil...	488	488	100%	5e-166	39.34%	609	Q8MJU5.1	
<input checked="" type="checkbox"/>	RecName: Full=Alpha-fetoprotein; AltName: Full=Alpha-1-fetoprotein; AltName: Full=Alpha-fetoglobulin; Flags: Precursor [Sus scrofa]	Sus scrofa	478	478	100%	4e-162	38.63%	610	Q8MJ76.1	
<input checked="" type="checkbox"/>	RecName: Full=Alpha-fetoprotein; AltName: Full=Alpha-1-fetoprotein; AltName: Full=Alpha-fetoglobulin; Flags: Precursor [Equus caballus]	Equus caballus	477	477	100%	6e-162	38.36%	609	P49066.1	
<input checked="" type="checkbox"/>	RecName: Full=Alpha-fetoprotein; AltName: Full=Alpha-1-fetoprotein; AltName: Full=Alpha-fetoglobulin; Flags: Precursor [Bos taurus]	Bos taurus	463	463	100%	2e-156	38.63%	610	Q3S257.1	
<input checked="" type="checkbox"/>	RecName: Full=Alpha-fetoprotein; AltName: Full=Alpha-1-fetoprotein; AltName: Full=Alpha-fetoglobulin; Flags: Precursor [Mus musculus]	Mus musculus	427	427	100%	2e-142	34.92%	605	P02772.1	
<input checked="" type="checkbox"/>	RecName: Full=Alpha-fetoprotein; AltName: Full=Alpha-1-fetoprotein; AltName: Full=Alpha-fetoglobulin; Flags: Precursor [Rattus norvegicus]	Rattus norvegicus	413	413	100%	6e-137	34.04%	611	P02773.1	
<input checked="" type="checkbox"/>	RecName: Full=Afamin; AltName: Full=Alpha-albumin; Flags: Precursor [Bos taurus]	Bos taurus	396	396	99%	2e-130	35.91%	604	G3MYZ3.1	
<input checked="" type="checkbox"/>	RecName: Full=Afamin; AltName: Full=Alpha-albumin; Short=Alpha-A1b; Flags: Precursor [Homo sapiens]	Homo sapiens	395	395	98%	5e-130	35.33%	599	P43652.1	
<input checked="" type="checkbox"/>	RecName: Full=Afamin; AltName: Full=Alpha-albumin; Short=Alpha-A1b; Flags: Precursor [Mus musculus]	Mus musculus	352	352	99%	2e-113	33.44%	608	Q89020.2	
<input checked="" type="checkbox"/>	RecName: Full=Afamin; AltName: Full=Alpha-albumin; Short=Alpha-A1b; Flags: Precursor [Rattus norvegicus]	Rattus norvegicus	340	340	99%	9e-109	32.45%	608	P36953.1	
<input checked="" type="checkbox"/>	RecName: Full=Vitamin D-binding protein; Short=DBP; Short=VDB; AltName: Full=Gc protein-derived macrophage... [Homo sapiens]	Homo sapiens	136	186	98%	5e-34	24.79%	474	P02774.2	
<input checked="" type="checkbox"/>	RecName: Full=Vitamin D-binding protein; Short=DBP; Short=VDB; AltName: Full=Gc-globulin; AltName: Full=Gro... [Bos taurus]	Bos taurus	133	178	98%	5e-33	24.26%	474	Q3MHN5.1	
<input checked="" type="checkbox"/>	RecName: Full=Vitamin D-binding protein; Short=DBP; Short=VDB; AltName: Full=Gc-globulin; AltName: Full=Gro... [Rattus norvegicus]	Rattus norvegicus	131	225	95%	3e-32	25.44%	476	P04276.3	
<input checked="" type="checkbox"/>	RecName: Full=Vitamin D-binding protein; Short=DBP; Short=VDB; AltName: Full=Gc-globulin; AltName: Full=Gro... [Cryotolagus cuniculus]	Cryotolagus cuni...	128	173	97%	3e-31	24.25%	476	P53789.1	
<input checked="" type="checkbox"/>	RecName: Full=Vitamin D-binding protein; Short=DBP; Short=VDB; AltName: Full=Gc-globulin; AltName: Full=Gro... [Mus musculus]	Mus musculus	127	212	96%	6e-31	25.00%	476	P21614.2	
<input type="checkbox"/>	RecName: Full=Albumin [Capra hircus]	Capra hircus	67.4	67.4	15%	2e-13	43.48%	90	P85295.2	

Figure 1. Surastos sekos

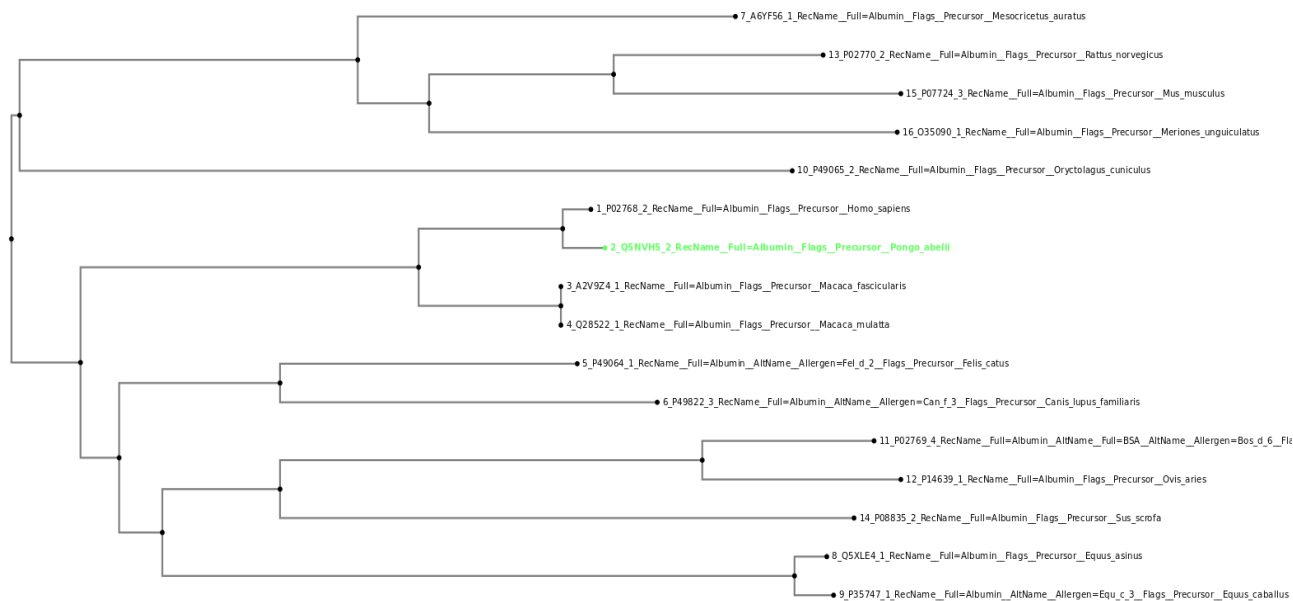


Figure 2. *Filogenetinis medis*

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

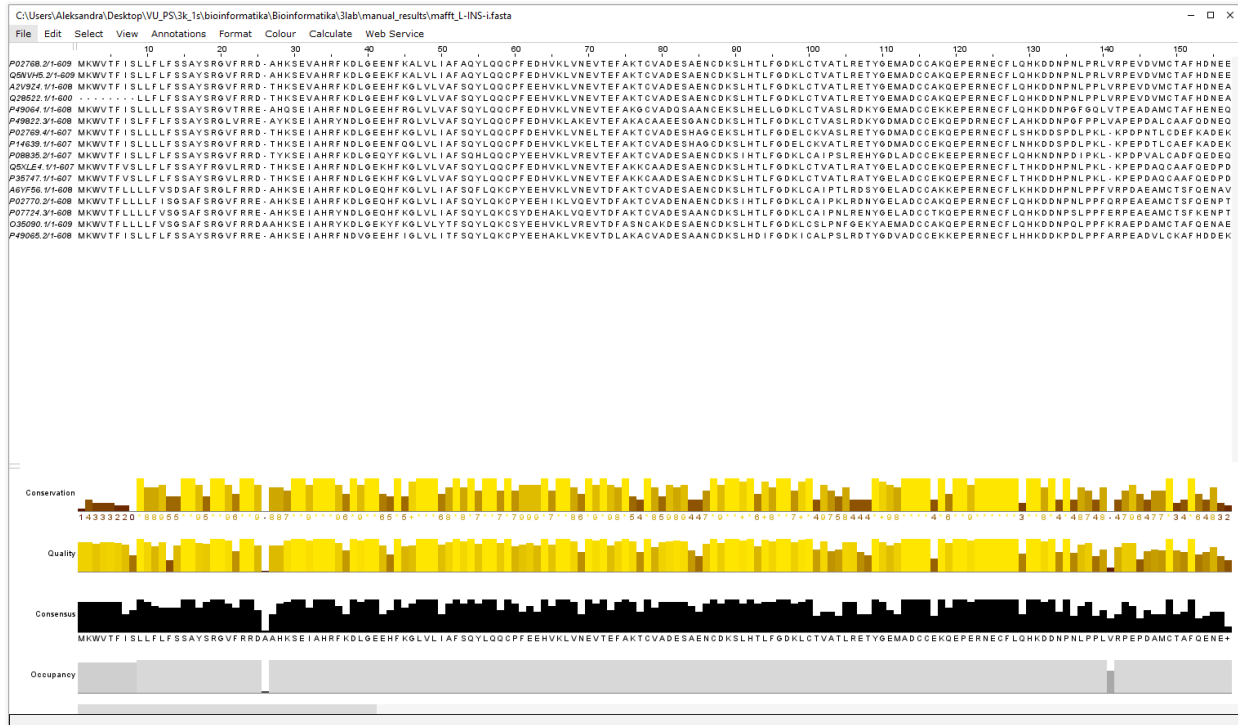


Figure 3. JalView

- a) 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.

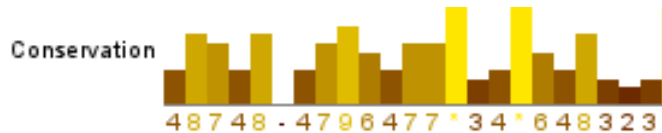


Figure 4. Konservatyvumas nepanašiausios sekos

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



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 simbolių), 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: RLV~~R~~PEVDVMCTAFH 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.