

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ KATEDRA

Bioinformatikos

Antrasis laboratorinis darbas

Atliko: 4 kurso 5 grupės studentas
Aurelijus Banelis

Vilnius – 2012

1. Išnagrinėkite prieinamą atitinkamo geno informaciją.

a) Ieškokite "Gene" duomenų bazėje NCBI portale <http://www.ncbi.nlm.nih.gov/>.
Naudokite užklausą "serum albumin"[Protein name] AND human[Organism].

[http://www.ncbi.nlm.nih.gov/gene?term=%22serum%20albumin%22\[Protein%20name\]
%20AND%20human\[Organism\]](http://www.ncbi.nlm.nih.gov/gene?term=%22serum%20albumin%22[Protein%20name]%20AND%20human[Organism])

b) Koks yra baltymo prieigos kodas? (Prasideda raidėmis NP). Parsisiųskite ir pateikite seką fasta formatu.

NP_000468.1

```
>gi|4502027|ref|NP_000468.1| serum albumin preproprotein [Homo sapiens]  
MKWVTFISLLFLFSSAYSRGVFRRDAHKSEVAHRFKDLGEENFKALVLIAFAQYLQQCFEDHVKLVNEV  
TEFAKTCVADESAENCDSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNECFLQHKDDNPPLRLV  
RPEVDVMCTAFHDNEETFLKKYLYEIARRHPYFYAPELLFFAKRYKAAFTECCQAADKAACLLPKLDEL  
DEGKASSAKQRLKCASLQKFGERAFAKAWAVARLSQRFPKAEFAEVSKLVTDLT KVHTECCHGDLLECADD  
RADLAKYICENQDSISSKLKECCEKPLLEKSHCIAEVENDEMPADLPSLAADFVESKDVCKNYAEAKDVF  
LGMFLYEYARRHPDYSVVLRLAKTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQNLIKQNCLEFE  
QLGEYKFQNALLVRYTKKVPQVSTPTLVEVSRNLGKVGSKCKHPEAKRMPCAEDYLSVVLNQLCVLHEK  
TPVSDRVTCKCTESLVNRRPCFSALEVDETYVPKEFNAETFTFHADICTLSEKERQIKKQTALVELVKHK  
PKATKEQLKAVMDDFAAFVEKCKADDKETCFAEEGKKLVAASQAALGL
```

c) Kelioms žinduolių rūšims yra surinkta informacija apie serumo albumino geną - atsakykite naudodami užklausą "serum albumin"[Protein name] AND mammals[Organism] " "Gene" duomenų bazėje NCBI. Kelioms paukščių rūšims yra prieinama tokia informacija (lotyniškai paukščiai - aves, žinduoliai - mammals)

19 žinduolių rūšių

3 paukščių rūšims

2. Gaukite ŽSA¹ ir panašių sekų žinduoliuose sekų palyginį. Tolimesnei analizei naudokite ŽSA baltymo seką.

a) suraskite visas patikimai į ŽSA panašias žinduolių sekas.
(Naudokite swissprot duomenų bazę bei 'Entrez Query' langelį).
Atmeskite atitikmenis, kurie apima mažiau nei 80% sekos ilgio.

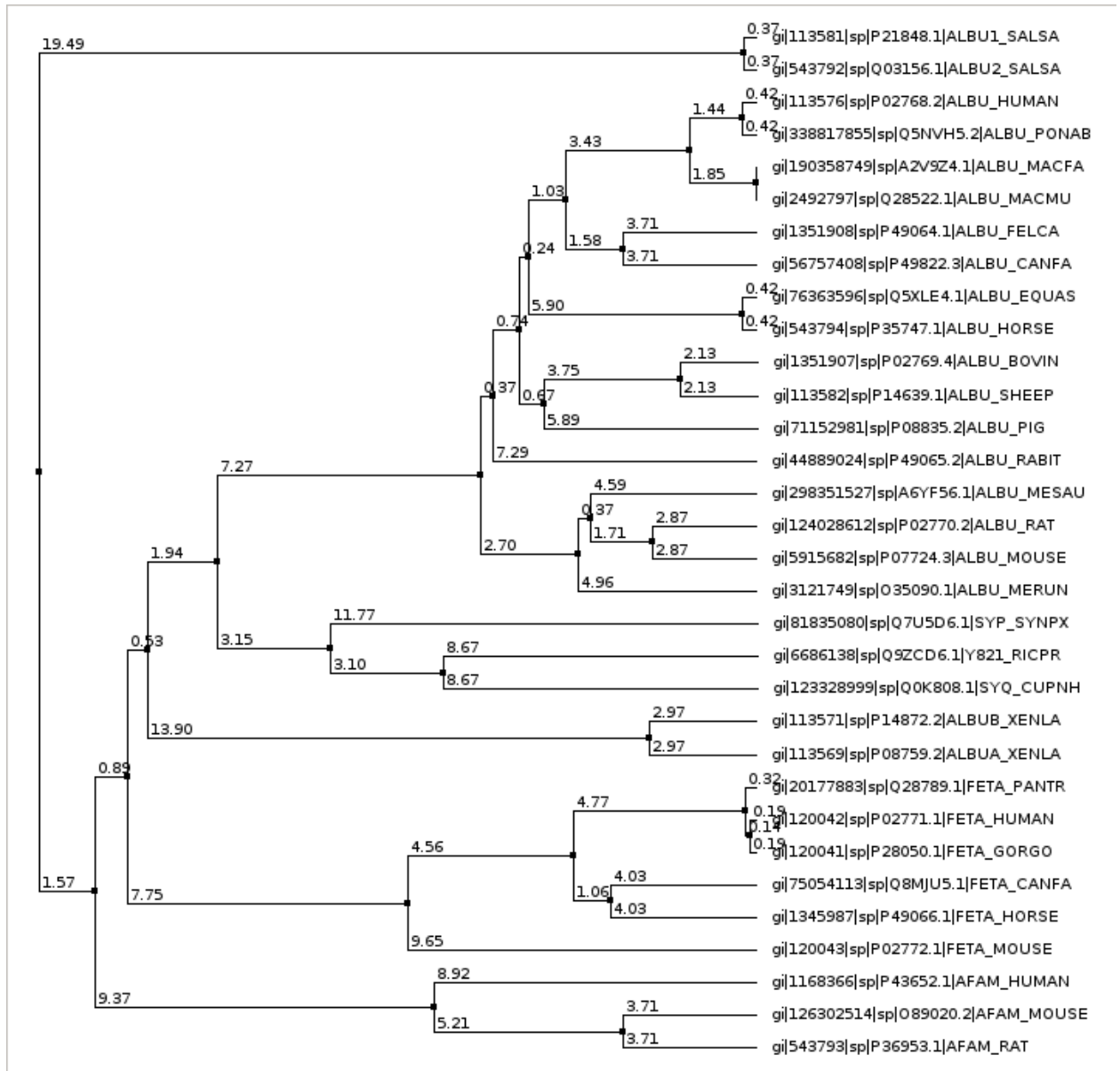
b) Surinkite visų atitikmenų sekas fasta formatu į vieną failą
(naudokite "Get selected sequences" opciją NCBI blast) ir išsaugokite tolimesniam naudojimui.

Žr. duomenys-2b.fa

c) Paleiskite JalView programą (<http://www.jalview.org/>)

1 Žmogaus serum albumin

download.html). Prisimindami praėjusias pratybas, gaukite su MAFFT programa palygininį bei filogenetinį medį. Pateikite ataskaitoje gauto medžio vaizdą. Kokio organizmo serumo albuminas yra panašiausias į žmogaus. Sekų suliginimo apačioje atkreipkite dėmesį į konservatyvumo grafiką.

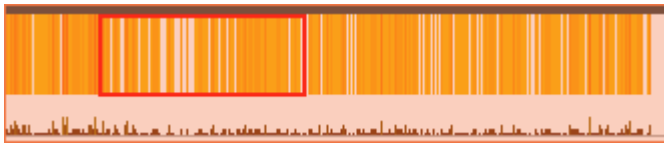


Panašiausias: 338817855 | sp | Q5NVH5.2 | ALBU_PONAB

3. Parinkite 20 aminorūgščių fragmentą iš ŽSA sekos, kuri:

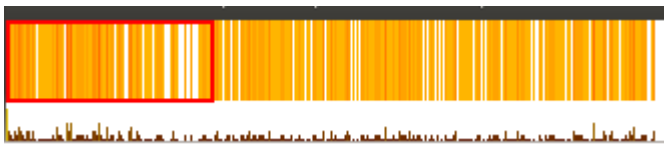
a) būtų kiek įmanomai skirtinga nuo kitų žinduolių sekų - Ji būtų naudojama gaminant antikūnus atskirti žmogaus kraujui nuo kitų žinduolių kraujo.

DDNPNLPRLVRPEVDVMCTA



b) būtų kiek įmanoma panaši į kitų žinduolių serumo albumino sekas (iš labiausiai konservatyvios palyginio dalies srities) – Ji būtų naudojama gaminti antikūnus leidžiančių identifikuoti, kad tiriamas skystis yra apskritai žinduolio kraujas. Sekos parinkimą argumentuokite ir iliustruokite

MKWVTFISLLFLFSSAYSRG



4. Naudodami BioPython (rekomenduoju –<http://biopython.org/DIST/docs/tutorial/Tutorial.html> ar BioJava bibliotekas parašykite programą, kuri atliktų 2-rą bei 3-čią užduotis t.y.:

- surinktu panašias žinduolių sekas duotai žmogaus baltymo sekai. (nuotolinis Blast paieškos vykdymas, gauto autputo parsinimas gaunant palyginį fasta formatu)
- parinktų 15 aminorūgščių fragmentą iš žmogaus baltymo sekos, kuris būtų
 - originaliausias žmogui ir labiausiai skirtųsi nuo kitų sekų.
 - būtų labiausiai panašus į kitas žinduolių sekas.

5. Ataskaitoje trumpai aprašykite taikyto algoritmo principus. Kaip atlikote fragmento paiešką? Kokį įvertį vertindami fragmento panašumą tarp žmogaus baltymo ir kitose baltymo sekose naudojote. Siųsdami ataskaitą kartu pridėkite ir veikiančios programos kodą. Ataskaitų laukiau iki Lapkričio pabaigos. Programa bus plėtojama vykdant likusią trečią užduotį.

- Pasigaminama sekų matrica (eilutės – sekos, stulpeliai – simboliai)
- Kiekvienam stulpeliui apskaičiuojamas pirmos (ŽSA sekos) ir likusių eilučių panašumas
 - Panašumas įvertinamas to paties simbolio pasikartojimų skaičiumi
- Suskaičiuojamos nurodyto ilgio panašumų sumos
 - Paprastas sumavimo optimizavimas: turėti eilės sumą, kuri „stumiamą“ prie vieno

galo pridedant elementą, o iš kito galo atimant

4. Sumavimo metu ieškoma didžiausia ir mažiausia suma, jų pozicijos įsimenamos ir naudojamos galutinam rezultatui (secai) atspausdinti