

Järjestused: <https://tinyurl.com/2jrdpyzu>

OSA I BLAST analüüs

a. Avage nimetatud lingil asuv FASTA formaadis järjestuste fail ja valige endale meelepärane järjestus: fail **sequences.fas**

b. Minge aadressile: <https://tinyurl.com/y2sgusys> Sinna saate vastavatesse tulpadesse ülesande vastused lisada. Palun sisestage ka oma nimi ja **valitud järjestuse** nimi.

c. Kopeeri esimesest failist valgujärjestus ja otsi sobiva NCBI BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) programmi abil parim joendus DNA andmebaasist 'RefSeq Genome Database'. Kasuta otsingus **ainult inimesega** (Homo sapiens) seotud järjestusi!

d. Vasta uuritava valgu kohta järgmistele küsimustele (Pane tähele, et otsitakse terve järjestuse koordinaate!):

- ❖ Kromosoomi number, algus- ja lõppkoordinaat inimese genoomis?
- ❖ Parima vaste andnud kromosoomi täielik pikkus andmebaasis?
- ❖ Parima vaste täielik bitskoor, katvuse protsent ja E-väärtus?

e. Vali d. punktis saadud joenduste hulgast **päringjärjestuse** koordinaatide järgi **esimese eksoni** valgu järjestus (andmebaasi ehk *Subject* järjestus). Jooksuta antud lõik sobiva NCBI BLAST programmiga valkude andmebaasi '**UniProtKB/SwissProt**' vastu.

f. Tulemuste lehel vajuta **MSA Viewer** lingile. Avanev lehekülg näitab tulemusi mitme järjestuse joendusena (*multiple sequence alignment*).

- ❖ Mitu erinevat liiki on tulemustes näha?
- ❖ Mitu imetaja liiki on tulemustes näha?

- ❖ Milline liik on väikseima katvusega?
- ❖ Kas peale inimese oli veel mõnel liigil katvus 100%?
- ❖ Mitu konserveerunud ala antud järjestuste võrdluses leiti?
- ❖ Teise mitte-konserveerunud ala algus- ja lõppkoordinaat antud joonduses?

Lisaülesanne: Arvuta mõlema joonduse jaoks skoor, identsusprotsent ja katvuse protsent. Skoori arvutamisel annab paardumine 2 ja mittepaardumine -3 ja vahe (gap) -5.

Joondus A

GCTTGGG-CGCGAAACC

TTAGGTCGGGACACC

Joondus B

PGPHGARRHNQV

PGNHG-RLHN

OSA II Praimeri disain

Uuritavaks objektiks on **inimese** genoomis asuva geeni **homeostatic iron regulator (HFE1) 2. ekson** ja selles asuvad mutatsioonid.

Eesmärk: Disainida unikaalne PCR praimeritepaar, millega saaks PCR-i abil üles amplifitseerida konkreetse eksoni [ENSE00003637434] järjestus.

Kasutatavad programmid:

SNPmasker: <https://bioinfo.ut.ee/snpmasker/>

Primer3: <https://primer3.ut.ee/>

GenomeTester: <https://bioinfo.ut.ee/genometester/>

Järjestus ise on leitav failist: **sequence_for_PCR.fasta**

NB! Antud järjestus sisaldab **eksoni järjestust (264 nt) + külgnevaid alasid (200 nt)**!

Nõuded PCR praimerite jaoks:

- a. praimerite vahele peab jääma uuritava ekson ja praimerid **ei tohi** kattuda **eksoni järjestusega (Vihje: "Targets" parameeter Primer3 programmis)**
- b. praimerid **ei tohi** seonduda **SNP-de** asukohtadele (**Vihje: SNPmasker programmiga SNP-de ka korduste maskeerimine**) ja oma **3' otsaga korduste peale (Vihje: "Lowercase masking" parameeter Primer3 programmis)**
- c. praimerid peavad andma genoomis ainult **ühe PCR produkti (Vihje: GenomeTester programmiga praimerite kontrollimine)**
- d. kahe praimerid maksimaalne sulamistemperatuuri erinevus: **2 °C**
- e. PCR produkti pikkuse vahemik: **300-600 nt**
- f. seonduvate oligote kontsentratsioon lahuses: **20 nM**
- g. praimerid minimaalne sulamistemperatuur: **55 °C**
- h. praimerid maksimaalne pikkus: **21 nt**
- i. K+ kontsentratsioon PCR puhvris: **20 mM**

Salvestada tulemusfaili parima praimeripaari **vasak (sense)** ja **parem (antisense) järjestus** ning nende **seonduvate arvud** inimese genoomis + **produktide arv**.

Võtke **GenomeTester** programmi tulemuste lehelt lahti "**Multiplex input file**" HTML kujul ja kopeerige sealt oma tulemusfaili PCR produkti järjestus.

Minge **NCBI BLAST** koduleheküljele, valige **sobiv BLAST programm** ja joondage eelnevalt saadud PCR produkti DNA järjestus "**Protein Data Bank proteins (pdb)**" valkude andmebaasi vastu.

Leidke tulemuste hulgast vastused järgmistele küsimustele:

- a. kui sorteerida tulemused katvuse järgi, siis milline on parima joonduse **liigi nimetus, täielik skoor, E-väärtus** ja **identsuse protsent**?

- b. kui tulemuste hulgast leidub joondusi, mille **identsus on suurem kui 45%**, siis mitmele **erinevale liigile** vasteid leiti? (**Vihje: kasutage tulemuste lehe filtreerimise võimalusi (Filter Results)**)

- c. kui filtreerida tulemused jättes välja kõik **imetajad** (*mammals (taxid:40674), exclude*) **mitu joondust** on suurema **maksimaalse skooriga** kui **60** ja milline on parima vaste andnud **liigi eestikeelne** nimetus?
-

OSA III Geenide ennustamine prokarüootses genoomis

Sekveneeriti ja pandi kokku bakteriliigi *Helicobacter pylori* uue tüve, millel on ka plasmiid, genoomijärjestus <https://tinyurl.com/5n7y9chv>. Teie ülesandeks on ennustada selles bakteritüves olevate valke kodeerivate geenide asukohad. Laadige genoomijärjestuse fail alla ning kasutage geeniennustuseks programmi **GeneMark.hmm** prokarüootide jaoks mõeldud versiooni <http://exon.gatech.edu/GeneMark/gmhmp.cgi>

Ennustust aitab täpsemaks muuta, kui programmile anda ette uuritavale võimalikult lähedane liik (eelistatult sama) - **Select species** rippmenüüst. Siis kasutatakse sellele liigile sobivat geenistruktuurimudelit. Märkige programmi väljunditeks nii valgusjärjestused kui nukleotiidsed järjestused - **Output options: Protein sequence, Gene nucleotide sequence**.

Options

Output format for gene prediction	Output options	Optional: results by E-mail
<input checked="" type="radio"/> LST <input type="radio"/> GFF	<input checked="" type="checkbox"/> Protein sequence <input checked="" type="checkbox"/> Gene nucleotide sequence Coding potential graph (not for multi FASTA) <input type="checkbox"/> PDF <input type="checkbox"/> PostScript	E-mail <input type="text"/> Subject GeneMark.hmm prokar <input type="checkbox"/> Compress files

Programmi käivitamiseks vajutage Start GeneMark.hmm nuppu.

Action Start GeneMark.hmm

Kui programm on töö lõpetanud, on tulemuseks kolm erinevat faili:

output	link
Coordinates of predicted genes	gmhmp.out
Protein sequences	gmhmp.out.faa
Nucleotide sequence	gmhmp.out.fnn

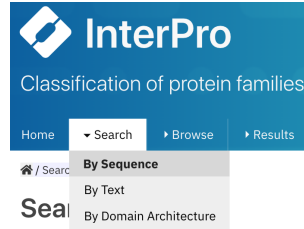
Kasutades infot nendes tulemusfailides, vastake järgnevatele küsimustele:

- Mitu valku kodeerivat geeni programm sellesse bakterisse kokku ennustas?
- Mitut valku kodeerivat geeni sisaldab selles tüves esinev plasmiidijärjestus?
- Kui pikk on plasmiidis
 - ❖ kõige pikem valku kodeeriv geen?

- ❖ kõige lühem valku kodeeriv geen?
- d. Kui pikka **valku** need geenid vastavalt kodeerivad?
- e. Kas need geenid algavad alternatiivse või standardse startkoodoniga?
- f. Kas kõige pikem ja kõige lühem geen plasmiidis paiknevad mõlemad samasuunalisel DNA ahelal või mitte?
- g. Leidke nii kõige pikema kui ka kõige lühema **plasmiidis** kodeeritava **valgu funktsioon** rakus, kasutades **sobivat BLASTi alamprogrammi ja Non-redundant protein sequences (nr) andmebaasi**.
Otsingu kiirendamiseks piirake otsingut ainult liigiga *Helicobacter pylori* (*taxid:210*) (**Organism**).
- ❖ Kõige pikema plasmiidis oleva valgu funktsioon:
 - ❖ Kõige lühema plasmiidis oleva valgu funktsioon:

OSA IV Funktsionaalsete domeenide ennustamine valgus

1. Kasutage analüüsis eelmises ülesandes tuvastatud kõige pikemat plasmiidis esinevat valku. Leidke selles valgujärjestuses esinevad funktsionaalsed alamregioonid, kasutades valguperekondade andmebaasi InterPro <https://www.ebi.ac.uk/interpro/> otsingut.



a. Millised osad sellest valgust paiknevad...

	koordinaadid
tsütoplasmas	
ekstratsellulaarselt	
rakumembraanis	

Klõpsake InterPro tulemuse lehel valguperekonna nimele, kuhu see järjestus otsingu tulemusena määrati (**Protein family membership**) ning avage järgneval lehel vasakul serval paiknevast menüüst link **Taxonomy**.

Overview
Proteins
Domain Architectures
Taxonomy
Proteomes
Structures
AlphaFold
Genome3D

b. Kas sarnase funktsiooniga valke leidub lisaks bakteritele ka viirustes või eukarüootides?

2. Analüüsige hiirest pärit tundmatu funktsiooniga valgujärjestust kasutades **InterPro andmebaasi**. Järjestus asub <https://tinyurl.com/3m3u9jms>

a. Millisesse valguperekonda kuuluva valguga on tegemist?

- b. Mitu erinevat funktsionaalset alamregiooni (domeeni) selles valgus esineb?
- c. Millistel koordinaatidel uuritavas valgus paikneb **DNA-ga seonduv domeen**?

Uurige **DNA-ga seondumise eest vastutavat domeeni** lähemalt, klikkides joonise paremal servas vastavale InterPro andmebaasi ID-le (algab tähisega IPR...).

- d. Millised järgnevatest väidetest on tõesed?
- ❖ Seda domeeni esineb imetajate valkudes
 - ❖ Seda domeeni esineb nii loomade kui bakterite valkudes, kuid ei esine viirustel
 - ❖ Seda domeeni esineb loomades, kuid ei esine taimedes
 - ❖ Seda domeeni on leitud kahes kaheksajala (*Octopus vulgaris*) valgus

Lisaülesanne: Mitmest eksonist koosneb uuritud hiirest pärit valgu homoloogi (sama funktsiooniga valku) kodeeriv geen inimese genoomis? Millises kromosoomis see geen inimese genoomis paikneb?