

Bioloģisko sekvenču (virkņu) analīzes pamatmetodes

____. variants

DARBA UZDEVUMI

1. Iepazīties ar bioloģiskās informācijas datubāzēm internetā un informācijas meklēšanas pamatprincipiem.
2. Noteikt laboratorijas darbā dotās DNS sekvences identitāti, analizēt gēna intronu – eksonu struktūru.
3. Identificēt proteīnu kodējošo DNS sekvences rajonu, translēt DNS sekvences par aminoskābju sekvencēm.
4. Iepazīties ar sekvenču salīdzināšanas principiem.
5. Filoģenētiskā analīze.

TEORIJA

Viens no svarīgākajiem bioinformātikas uzdevumiem ir DNS un proteīnu sekvenču analīze. Pētījumu rezultātā noteiktās DNS un proteīnu sekvences tiek ievietotas starptautiskās publiski pieejamās datu bāzēs. Pētniekiem visā pasaulē ir pieejamas datubāzes, kas dod iespēju salīdzināt savu pētījumu rezultātus ar datu bāzēs atrodamo. Bioloģiskās informācijas apjoms ir ļoti liels un, lai datubāzēs atrastu interesējošo informāciju, nepieciešamas efektīvas meklēšanas sistēmas. Ir divi galvenie informācijas meklēšanas veidi: 1) teksta meklēšana izmantojot atslēgas vārdus vai konkrētus datubāzes ierakstu numurus (Accession Number, Gene Identifier u.c.); 2) homoloģijas meklēšana salīdzinot jūs interesējošo DNS vai proteīna aminoskābju sekvenci ar visu datubāzi. Viens no populārākajiem homoloģijas meklēšanas rīkiem ir BLAST, kas ļauj meklēt līdzību starp DNS – DNS (BLASTN, TBLASTX), proteīnu – proteīnu (BLASTP), DNS – proteīnu (BLASTX) un proteīnu – DNS (TBLASTN) secībām. Homoloģijas meklēšanas pamatā ir pieņēmums, ka sekvencēm ir līdzīga izcelsme, t.i., ka novērotā līdzība starp sekvencēm norāda uz to izcelsmi no kopīga senča evolūcijas procesā. Filoģenētiskā analīze (phylogeny) ir sugas (vai citas taksonomiskas vienības) evolucionārā vēsture. Filoģenētiskās analīzes pamatā ir ideja, ka visi dzīvie organismi ir cēlušies no kopīga senča. Evolucionārās attiecības starp taksonomiskajām vienībām parasti tiek attēlotas filoģenētiskā koka (dendrogrammas) veidā. Filoģenētiskie koki var tikt konstruēti gan atsevišķiem gēniem, gan veseliem genomiem.

REAKTĪVI (MATERIĀLI, VAI PROGRAMMAS)

Laboratorijas darbā tiek izmantotas sekojošas DNS sekvences:

1. http://priede.bf.lu.lv/grozs/Biologija_nebiologiem/Ievads_molekularaja_un_sunas_biologija_%28Biol2046%29%2F4_Tema%2FLab_Darbi%2FSekvences/
2. Darbā ir doti sekvenču pāri – genomiskā (gDNS) un komplementārā (cDNS, mRNS kopija), piemēram, Sek1_cDNS.txt un Sek1_gDNS.txt
3. Kopā ir 5 sekvenču pāri, kas atbilst 5 laboratorijas darba variantiem. Katram studentam ir jāizpilda viens laboratorijas darba variants.

Laboratorijas darbā tiek izmantotas sekojošas interneta vietnes un datubāzes:

1. <http://www.ncbi.nlm.nih.gov/> - ASV Nacionālā biotehnoloģijas informācijas centra mājas lapa un tajā atrodamā datubāze GenBank, kā arī homoloģijas meklēšanas rīks BLAST.
2. <http://www.ebi.ac.uk/> - Eiropas bioinformātikas institūta mājas lapa un tajā atrodamais rīks Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

EKSPERIMENTĀLĀ DAĻA

1. Nosaka doto cDNS un gDNS sekvenču identitāti izmantojot BLASTN homoloģijas meklēšanas rīku (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). gDNS sekvenču līdzību meklē pret cilvēka RefSeq Genomic DNS datubāzi. cDNS sekvenču līdzību meklē pret nr DNS datubāzi.

2. Nosaka genomiskās DNS intronu - eksonu struktūru salīdzinot genomisko un cDNS sekvenču ar rīku SPIDEY (<http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/>).

3. Nosaka atvērto nolasīšanas rāmi (Open Reading Frame) izmantojot rīku ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) un veic garākā ORF analīzi ar BLASTP rīku.

4. Nosaka ORF aminoskābju sekvenču vislīdzīgāko aminoskābju sekvenču NCBI proteīnu datu bāzē. Atlasa 10 aminoskābju sekvenču starp BLASTP rezultātiem, kuras nav identiskas ORF sekvenču. Saglabā sekvenču datora cietajā diskā, kā teksta failu „FASTA (complete sequence)” formātā.

5. Veic atlasīto aminoskābju sekvenču daudzkārtējo salīdzinājumu izmantojot Eiropas bioinformātikas institūta mājas lapā atrodamo rīku Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). Apskatās iegūto daudzkārtējo sekvenču salīdzinājumu un identitātes matricu, novērtē iegūto dendrogrammu, atrod līdzīgākos un atšķirīgākos aminoskābju sekvenču pārus.

REZULTĀTU APSTRĀDE

1. BLASTN homologijas meklēšana norāda, ka darbā dotā genomiskās DNS sekvenca atbilst GenBank atrodamai RefSeq sekvencai ar Accession Number _____, bet cDNS sekvenca atbilst sekvencai ar Accession Number _____. Genomiskā un cDNS satur _____ gēnu.
2. SPIDEY rezultāti liecina, ka genomiskā DNS satur _____ eksonus un _____ intronus. Garākais eksons ir _____ nukleotīdus garš, bet īsākais _____ n garš.
3. ORF Finder rezultāti atrod _____ atvērtos nolasīšanas rāmjus. Garākais nolasīšanas rāmis ir _____ nukleotīdus vai _____ aminoskābes garš.
4. BLASTP analīze liecina, ka garākā aminoskābju sekvenca ir identiska _____ proteīnam.

Tālākajam daudzkārtējam sekvenču salīdzinājumam tiek atlasītas sekojošas sekvences:

N.p.k.	Proteīna nosaukums	Sugas nosaukums	„Accession“ numurs
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			

5. Izmantojot Clustal Omega rīku iegūst daudzkārtējo sekvenču salīdzinājumu. Identiskas aminoskābes salīdzinājumā apzīmē ar simbolu ____, līdzīgas aminoskābes ar simboliem ___ vai ____, pārtraukumus salīdzinājumā apzīmē ar _____. Identitātes matrica liecina, ka līdzīgākās ir sekvences _____ un _____, kuras ir _____% identiskas, bet atšķirīgākās ir sekvences

Vārds, Uzvārds. _____, _____ . Stud. apl. Nr. _____.

Ievads molekulārajā un šūnas bioloģijā – Laboratorijas darbs Nr. 16

_____ un _____, kuras ir _____% identiskas. Zemāk uzskicē analīzes rezultātā iegūto filoģenētisko koku, zariem pieraksta klāt sugu nosaukumus.

Darba veikšanas datums _____.

Vārds, Uzvārds. _____, _____ . Stud, apl. Nr. _____.