

9. vaja: Naloge in odgovori

V09-01

Pripravite prileganje aminokislinskih zaporedij človeškega proteina TLR4 (Toll-like receptor 4) s TLR4 iz naslednjih organizmov: miš (*Mus musculus*), podgana (*Rattus norvegicus*), krava (*Bos taurus*), prašič (*Sus scrofa*), vinska mušica (*Drosophila melanogaster*), opica bonobo (*Pan paniscus*). Narišite filogenetsko drevo ter komentirajte (vprašanja spodaj).

Potek dela:

1. Pripravite tekstno datoteko (ali pa si vhodne podatke prilepite v okno spodaj) v formatu FASTA ter jo ustrezno oblikujte (krajša imena zaporedij, ime organizma).
2. Pripravite prileganje s programoma Clustal Omega in T-COFFEE.
3. Oglejte si prileganji ter ju komentirajte in primerjajte.
4. Na osnovi obeh prileganj s programom ClustalW2 Phylogeny pripravite filogenetsko drevo. Se drevesi razlikujeta? Zakaj?
5. Zakaj mislite, da smo v prileganje vključili zaporedje iz vinske mušice?

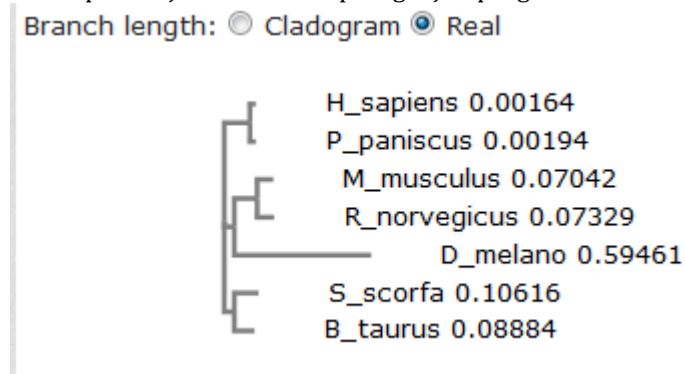
Odgovor

V bazi podatkov (npr. UniProt) poiščemo zaporedja proteina TLR4 iz različnih organizmov in si jih shranimo v .txt datoteko, ki jo preimenujemo v .fasta. Datoteko naložimo (ali pa zgolj skopiramo zaporedja) na spletna strežnika Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) in T-COFFEE (<https://www.ebi.ac.uk/Tools/msa/tcoffee/>).

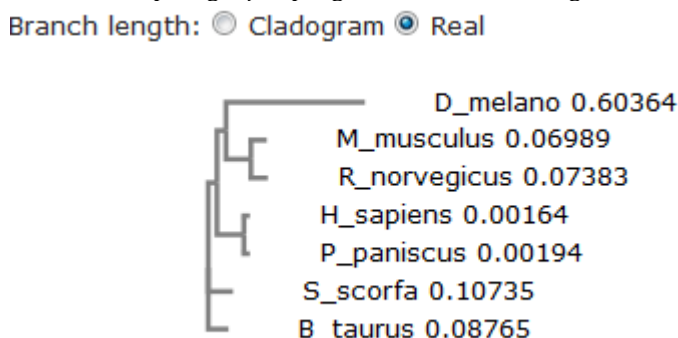
V obeh primerih opazimo, da se protein vinske mušice najbolj razlikuje od ostalih, ki pa so si med seboj bolj podobni. Opazimo, da je razlika med obema prileganja najbolj očitna pri prileganju proteina vinske mušice z ostalimi, saj opazimo različne položaje vrzeli.

Primerjava dreves:

Drevo pridobljeno na osnovi prileganja s programom T-COFFEE



in na osnovi prileganja s programom Clustal Omega



Prvo drevo je iz vidika evolucijske oddaljenosti med organizmi bolj smiselno. Skupaj namreč postavi tako človeka in opico, kot podgano in miš ter kravo in prašiča. V primeru drugega drevesa pa izgleda, kot da sta si mišji in človeški protein bolj sorodna kot mišji in kravji, saj sta kravji in prašičji na svojih vejah in ločena od ostalih.

Zanimivo v obeh primerih izgleda, kot da bi se protein iz vinske mušice in proteina iz miši in podgane razvila iz skupnega prednika, ki pa ni skupen ostalim obravnavanim organizmom. To ne odraža dejanskega poteka

9. vaja – Prileganje zaporedij in filogenetska drevesa (Miha Pavšič / maj 2014)

evolucije, čeprav lahko iz dolžine vej razberemo, da je protein iz vinske mušice od vseh najbolj evolucijsko oddaljen.

Lahko bi rekli, da je program T-COFFE nalogo opravil bolje, kar ni presenetljivo, saj je posebej prilagojen za delo z manjšim številom zaporedij.

Zaporedje iz vinske mušice smo vključili zato, da imamo vsaj en protein, ki je evolucijsko bolj oddaljen od ostalih. To zaporedje nam na nek način služi tudi kot kontrola pri interpretaciji rezultatov.

V09-02

Poiščite 30 homologov človeškega proteina EpCAM iz različnih organizmov, pripravite prileganje zaporedij (31) s programom Clustal Omega ter narišite filogenetsko drevo. Bodite pozorni, nekateri organizmi imajo zraven EpCAM še njegov paralog, imenovan Trop2. Na osnovi prileganja pripravite podatke za izris filogenetskega drevesa z upoštevanjem korekcije zaradi večkratnih substitucij, drevo pa narišite s programom PhyloDendron (tip drevesa: phenogram).

1. Se razvejitev drevesa skladajo s taksonomijo organizmov?
2. Kako je z ohranjenostjo cisteinskih ostankov? Se to sklada z njihovo vlogo v proteinih?
3. So katere regije še posebej ohranjene? Diskutirajte, kje v strukturi bi se lahko nahajale oz. kakšna bi lahko bila njihova vloga.

Odgovor

Ker potrebujemo večje število homologov, kot v prejšnji nalogi, se je tokrat iskanja le the bolj smiselno lotiti s programom Blast. V bazi UniProt poiščemo zapis za človeški EpCAM in s tem zapisom iščemo s programo blastp. Zapise izberemo tako, da iz posameznega organizma vzamemo po največ en zapis in da izberemo vsaj nekaj organizmov, ki imajo manj podoben protein (npr. ne samo opic, ampak tudi kakšen od človeka evolucijsko bolj oddaljen organizem). Naloženo datoteko z zaporedji nato ustrezno uredimo, tako da bodo imela zaporedja za nas smiselna imena. Pripravljeno izhodno datoteko nato uporabimo za pripravo prileganja s programom Clustal Omega in narišemo filogenetsko drevo z ukazom "Send to ClustalW2_Phylogeny". Tako pripravljeno drevo si nato narišemo s programom PhyloDendron. Pozorni smo na posebne zahteve, ki so navedene v navodilu te naloge!

Rezultati, ki jih boste dobili, se bodo med seboj nekoliko razlikovali, saj ni nujno, da ste vsi uporabili ista zaporedja. Bolj ali manj bo drevo odražalo evolucijsko oddaljenost organizmov, seveda pa boste lahko opazili tudi kakšne anomalije.

Cisteinski ostanki so zelo ohranjeni skozi vse organizme. To se sklada z njihovo vlogo v proteinih, saj tvorijo disulfidne vezi, ki pogosto igrajo zelo pomembno vlogo pri stabilizaciji določenega dela proteina in tako vplivajo na zvitje.

Med ohranjenimi regijami zagotovo opazimo tiroglobulinsko domeno, ki se nahaja nekje okoli aminokislinskih ostankov 70-140. Če smo zelo pozorni zelo ohranjeno regijo opazimo tudi med 210 in 220 aminokislinskim ostankom v človeškem proteinu. Ohranjene regije imajo v proteinih pogosto pomembno vlogo za samo strukturo in/ali funkcijo proteina. V strukturi se pogosto nahajajo v notranjosti proteina, saj je ohranjenost proteinskega jedra za strukturo proteina bolj pomembna, kot ohranjenost zank na površini.

V09-03

Pripravite prileganje tiroglobulinskih domen tipa I (type-1) človeških in mišjih proteinov, ki vežejo inzulinu podoben rastni faktor (IGFBP, insulin-like growth factor binding protein). Kaj je skupna značilnost teh domen? Komentirajte na podlagi zapisa v bazi Prosite.

Odgovor

Do zelenih zaporedij lahko pridemo na več načinov. Eden izmed pristopov je da po UniProt-u iščemo inzolinu podobne rastne faktorje med človeškimi proteini in za vsakega izpišemo zaporedje, ki predstavlja tiroglobulinsko domeno tipa 1. Enako nato ponovimo še za mišje proteine in zaporedja združimo v isto datoteko ter ustrezno poimenujemo. S tako pripravljeno datoteko naredimo prileganje. Ker imamo malo zaporedij, je smiselno uporabiti T-COFFE.

```

Human_IGFBP1  KEFCRIELYRVVESLAKAQETSG----EEISKFYLPNCIKNGFYHSRQCE
Human_IGFBP3  YGFCPREMEDTLNHLKFLNVL----SPR---GVHIPNCIKKGFYKKKQCA
Human_IGFBP4  QGSCQSELHRALERLAASQSRTH----EDLYIIPINCIIRNGNFHPKQCF
Human_IGFBP6  MGFCPRHLDSVLQQLQTEVYR---GAQ---TLYVPCIDHRGFYRKRQCF
Mouse_IGFBP2  RTPCQQLDQVLERISTMRLPDDRGPLEHLYSLHIPNCIKHGRYNLQKCF
Mouse_IGFBP4  QGSCQSELHRALERLAASQSRTH----EDLFIIPINCIIRNGNFHPKQCF
Mouse_IGFBP5  QGFCPRHMEASLQEFKASPRM----VPR---AVYLEPCIDRKGFYRKRQCF
Mouse_IGFBP6  MGFCPRHLDSVLQQLQTEVFRG---GAR---GLYVPCIDLRGFYRKRQCF
      *  . :  : :  :  .  . : * * * * . * : . : *

```



```

Human_IGFBP1  TSMDEAGLCTCYYPWNGKRIPGSP-EIRGDPNC
Human_IGFBP3  PSKGRKRGFCTCVDK-YGQPLPGYTTKGGEDVHC
Human_IGFBP4  PALDGQRGKCTCVDRKTGVKLPGGL-EPKGELDC
Human_IGFBP6  SSQGQRRGKCTCVDR-MGKSLPGSP-DGNGSSSC
Mouse_IGFBP2  MSLNGQRGKCTCVNPNTGKPIQGAP-TIRGDPEC
Mouse_IGFBP4  PALDGQRGKCTCVDRKTGVKLPGGL-EPKGELDC
Mouse_IGFBP5  PSRGRKRKGICTCVDK-YGMKLPGME-YVDGDFQC
Mouse_IGFBP6  SSQGNRRGKCTCVDP-MGQPLPVSP-DGQGSTQC
      : . . * * * * * * :

```

V bazi Prosite najdemo podatke o lastnostih tiroglobulinske domene tipa 1 (<http://prosite.expasy.org/PDOC00377>) . Ugotovimo, da je za tiroglobulinske domene tipa 1 značilna ohranjenost 6ih cisteionov, ki tvorijo tri disulfidne vezi, ki so pomembne za strukturo tiroglobulinske domene tipa 1. Po pregledu rezultatov našega prileganja ugotovimo, da med aminokislinskimi ostanki, ki se prilegajo v vseh zaporedij, dejansko najdemo ohranjenih 6 cisteinov.

V09-04

Ugotovite, ali so aminokislinski ostanki, ki direktno določajo specifičnost tripsina, ohranjeni pri različnih organizmih. Kaj bi lahko bil drug razlog za enako/podobno specifičnost v primeru njihove neohranjenosti?

Odgovor

Najprej moramo ugotoviti, za katere aminokislinske ostanke gre. V bazi UniProt poiščemo človeški tripsin 1 (id: P07477). V zapisu poiščemo informacije o posebnih aminokislinskih ostankih.

Sites

<input type="checkbox"/>	Active site	63	1	Charge relay system
<input type="checkbox"/>	Active site	107	1	Charge relay system
<input type="checkbox"/>	Active site	200	1	Charge relay system
<input type="checkbox"/>	Metal binding	75	1	Calcium
<input type="checkbox"/>	Metal binding	77	1	Calcium; via carbonyl oxygen
<input type="checkbox"/>	Metal binding	80	1	Calcium; via carbonyl oxygen
<input type="checkbox"/>	Metal binding	85	1	Calcium
<input type="checkbox"/>	Site	194	1	Required for specificity <small>(By similarity)</small>

Ugotovimo, da je za specifičnost pomemben aminokislinski ostanek na mestu 194. V primeru človeškega tripsina 1 je to aspartat. Zapise za tripsine iz ostalih organizmov pridobimo iz baze UniProt, pri čemer pazimo, da pomotoma ne vključimo tudi zapisov za inhibitorje tripsina.

Prileganje zaporedij nam pokaže, da Asp194 ohranjen praktično pri vseh organizmih (odvisno od izbire le the, če boste izbrali tudi evolucijsko zelo oddaljene organizme, boste morda opazili razlike). V primeru neohranjenosti bi bil razlog za enako/podobno specifičnost najverjetneje mutacija v kakšno podobno aminokislino na primer glutamat.