

Fylogenetická analýza

Vyberte si 18 sekvencí **různých** proteinů, které sdílejí alespoň 30% identitu. Zvolte sekvence z různých organismů. Zvolte i dvě sekvence, které mají identitu nižší (ale jsou nalezeny jako potenciální podobné sekvence) - nevolte ale totálně odlišné (především v délce)

1) Připravte si data (zeditujte nazvy), tak aby byla jednoznačně krátce a smysluplně pojmenována (napr: vymazte "sp|gi|P31415626|QNX_Y_CJAPONa nahradte slovem rice, R. solanacearum, myš, atd ..., ze které sekvence pochází. Bude to následně přehlednější.

2) Udělejte vícenásobné sekvenční příložení (ideálně ClustalW2 (už není dostupný na EBI, dá se stáhnout jako stand-alone z clustal.org či je dostupný na stránkách <https://www.genome.jp/tools-bin/clustalw>), Clustal Omega či jiným softwate. **Pro jednodušší práci je mít sekvence obdobné délky**, ale není nutné

3) Proveďte dvě fylogenetické analýzy s použitím Neighbor Joining metodou a pak UPGMA (možno na: https://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/), ale klidně vyzkoušejte jiné stránky

4) Popište výsledné příložení a zahrňte grafické obrázky alignmentu i fylogeneze.

(5) Proveďte totéž na úrovni (c)DNA..

(6) Mají stromy stejnou topologii?

(7) Mají stromy stejnou délku větví?

?

(8) Pokud 6 a 7 není pravda, popište rozdíly a zkuste navrhnout, proč se tyto stromy od sebe liší.

(9) Ukazují stromy náznaky paralogní evoluce_?

(10) Které uzly jsou ortologní a které paralogní?

Pro znázornění a formátování stromů můžete využít např. <http://itol.embl.de> (upload) či jakýkoli jiný program.

Do protokolu zahrňte i Vaše sekvence ve FASTA formátu

(11) vyberte si nejkonzervovanější část příložení a vygenerujte jeho "sekvenční logo" (např. <http://weblogo.berkeley.edu/>)