

Fylogenetická analýza

Vyberte si 28 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)

1) Připravte si data (zeditujte názvy), tak aby byla jednoznačně kratce a smysluplně pojmenována (např: vymazte "sp|gi|P31415626|QNX_Y_CJAPONa nahradte slovem rice, R. solanacearum, atd ..., ze které sekvence pochází.

2) Udělejte vícenásobné sekvenční přiložení (ideálně ClustalW2 (už není dostupný na EBI, dá se stáhnout jako stand-alone z clustal.org či je dostupný na stránkách simgene.com/ClustalW 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

4) Popište výsledné přiložení 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 evolece_?

(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 přiložení a vygenerujte jeho sekvenční logo (např. <http://weblogo.berkeley.edu/>)

Postranlační modifikace

Zvolte si jednu z eukaryotních proteinových sekvencí (alespoň 350 aminokyselin) a za využití nástrojů pro predikci postranlačních modifikací v sekvenci vyznačte graficky potenciální modifikovaná místa. Uveďte i negativní výsledky... Jelikož programy pracují s různými predikčními algoritmy, zkuste použít i více programů pro podobnou predikci...

je možné využít následující predikční programy, ale výčet je limitován a lze nalézt spoustu dalších:

INTERPRO, balíky na http://www.expasy.org/proteomics/post-translational_modification

Predikce glykosylace program GPP

<http://comp.chem.nottingham.ac.uk/cgi-bin/glyco/bin/getparams.cgi>

Predictions for O-GlcNAc sites YinOYang 1.2

<http://www.cbs.dtu.dk/services/YinOYang/>

predictions for O- β -GlcNAc attachment sites in eukaryotic protein sequences, Intracellular O-glycosylation is characterised by the addition of N-acetylglucosamine, in a beta anomeric linkage, to Serine and Threonine residues in a protein.

NetPhos 2.0 Server

<http://www.cbs.dtu.dk/services/NetPhos/>

The NetPhos 2.0 server produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins.

NetPhosK 1.0 Server

<http://www.cbs.dtu.dk/services/NetPhosK/>

The NetPhosK 1.0 server produces neural network predictions of kinase specific eukaryotic protein phosphorylation sites. Currently NetPhosK covers the following kinases: PKA, PKC, PKG, CKII, Cdc2, CaM-II, ATM, DNA PK, Cdk5, p38 MAPK, GSK3, CKI, PKB, RSK, INSR, EGFR and Src.

NetAcet 1.0 Server

<http://www.cbs.dtu.dk/services/NetAcet/>

NetAcet 1.0 server predicts substrates of N-acetyltransferase A (NatA). The method was trained on yeast data but, as mentioned in the article describing the method, it obtains similar performance values on mammalian substrates acetylated by NatA orthologs.

Terminator

<http://www.isv.cnrs-gif.fr/terminator3/index.html>

Terminator predicts N-terminal methionine excision, N-terminal acetylation, N-terminal myristoylation and S-palmitoylation of either prokaryotic or eukaryotic proteins originating from organellar or nuclear genomes

SulfoSite

<http://sulfosite.mbc.nctu.edu.tw/>

is to computationally predict sulfation sites within given protein sequences

Sulfinator

<http://web.expasy.org/sulfinator/>

The Sulfinator predicts tyrosine sulfation sites in protein sequences. Tyrosine sulfation is an important post-translational modification of proteins that go through the secretory pathway.

PrePS - Prenylation Prediction Suite

<http://mendel.imp.ac.at/sat/PrePS/index.html>

Myristoylator

<http://web.expasy.org/myristoylator/>

Myristoylator predicts N-terminal myristoylation of proteins by neural networks. Only N-terminal glycines are myristoylated (leading methionines are cleaved prior to myristoylation).

GPI-SOM: Identification of GPI-anchor signals by a Kohonen Self Organizing Map

<http://gpi.unibe.ch/>

Predikce sumoylace

<http://www.abgent.com/sumoplot>

The SUMOplot™ Analysis Program predicts and scores sumoylation sites in your protein. The presence of this post-translational modification may help explain larger MWs than expected on SDS gels due to attachment of SUMO protein (11kDa) at multiple positions in your protein.

SUMO-1 (small ubiquitin-related modifier; also known as PIC1, UBL1, Sentrin, GMP1, and Smt3) is a member of the ubiquitin and ubiquitin-like superfamily. Most SUMO-modified proteins contain the tetrapeptide motif B-K-x-D/E where B is a hydrophobic residue, K is the lysine conjugated to SUMO, x is any amino acid (aa), D or E is an acidic residue.

The SUMOplot™ Analysis Program predicts the probability for the SUMO consensus sequence (SUMO-CS) to be engaged in SUMO attachment. The SUMOplot™ score system is based on two criteria: direct amino acid match to SUMO-CS. substitution of the consensus amino acid residues with amino acid residues exhibiting similar hydrophobicity

CSS-Palm

<http://csspalm.biocuckoo.org/>