

**II Esercitazione di Bioinformatica**  
***“Similarity search and alignments”***

Seguire il percorso indicato nel testo sottostante e stilare una breve relazione con i contenuti recuperati e elaborati man mano, usando diverse risorse informatiche.

A) Gli allineamenti multipli di proteine sono estremamente utili per ricavare informazioni sulla filogenesi, sulla struttura secondaria, sull'individuazione di domini strutturali e funzionali e sul folding proteico. Nel corso di questa esercitazione cercheremo di utilizzare la ricerca di similarità per identificare una proteina umana a partire da un frammento della sua sequenza e di evidenziare la sua eventuale appartenenza ad una famiglia di proteine. Le sequenze di alcuni altri membri della famiglia (proteine umane o di altri organismi, con elevatissima similarità con la proteina in analisi) verranno allineate con Clustal Omega. L'allineamento multiplo e il cladogramma ottenuti saranno analizzati per interpretare i risultati.

**Links utili:**

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

<https://www.ebi.ac.uk/Tools/msa/clustalo/>

1 Collegarsi all'home page del programma Blast dell'NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>).

2 Scegliere Protein-protein BLAST (blastp).

3 Incollare la sequenza seguente (cut and paste):

```
VQRPGRIP TSSRCGNTIPDDDNQVVLSLSPGSRVYVLP RPVCFEKG TNYTVRLELPQYTSSDSDVESPYTLIDSLVLMPYC  
KSLDIF TVGGSGDGVV TNSAWETFQRYRCLENSRSVVKTPMTDVCRNII FSI SALLHQ TGLACECDPQGSLSVCDPNGG  
QCQCRPNVVGRTCNRCAPGTFGFGPSGCKPCECHLQG
```

4 Scegliere il database Uniprot/SwissProt. Inviare la ricerca.

6 Analizzare il risultato della ricerca di similarità, per capire a quale proteina umana corrisponde il frammento di sequenza in analisi, **recuperarne la sequenza di riferimento, salvandola in un file di testo e seguire il collegamento a EntrezGene per conoscerne funzione, localizzazione e quello a OMIM per eventuale coinvolgimento in malattie genetiche.**

7 Ritornare a Protein-protein BLAST (blastp), usare come query la sequenza appena recuperata, limitando la ricerca a Reference sequences (RefSeq) di proteine del gruppo tassonomico Homo/Pan/Gorilla (sottofamiglia Homininae dei primati). Inviare la ricerca.

8 Analizzare il risultato della ricerca di similarità'.

9 **Selezionare le sequenze di uomo, bonobo (P. paniscus) e gorilla della famiglia delle laminine beta (1, 2 e 3), recuperarle in formato FASTA e copiarle in un file di testo.**

10 Modificare gli headers delle sequenze in modo da renderli brevi (max 8 caratteri), informativi ed univoci (ad es. Hs1, Hs2, Pp1, ...).

11 Collegarsi all'home page dell'EBI dedicata agli allineamenti multipli (<http://www.ebi.ac.uk/Tools/msa/>) e scegliere Clustal Omega

12 Incollare le sequenze dal file di testo, avviare l'analisi con i parametri di default. **Analizzare l'allineamento, il cladogramma e l'albero filogenetico associato.**