Practica alineament per parelles

Exercici 1

El Gen / Proteïna "Leghemoglobin-1" de *Lupinus luteus* (Llobí groc) Facilita la difusió de l’oxigen als bacteris simbiòtics per tal de promoure la fixació de nitrogen. La subunitat beta de la Hemoglobina humana està implicada en el transport d’oxigen des del pulmó fins als diversos teixits perifèrics. Estan relacionats evolutivament aquests gens / proteïnes?

Anem a comparar les dues proteïnes i els DNA que les codifiquen. Ho farem usant un algorisme global (Needleman–Wunsch) i un local (Smith–Waterman).

1. Descarrega del ncbi en format fasta les seqüència de:
   1. La seqüència de DNA, que codifica per la proteïna, de l'entrada "Y00401" de la DB "nucleotides". I la seqüència peptídica de la proteïna codificada. Instruccions per fer-ho:
      1. Cerca al ncbi dins la DB"nucleotides" i troba l'entrada "Y00401" en format gb
      2. Cerca el CDS on està codificada la proteïna leghemoglobin i fixa't en les coordenades.
      3. Clica a la "FEATURE" CDS (subratllat i en blau) de la proteïna, veuràs que s'il·lumina part de la seqüència i apareixen opcions a baix a la dreta
      4. A baix a la dreta clica sobre Display FASTA.
      5. Quan s'ha obert la seqüència en fasta fixa't en la capçalera del fasta i en la URL, fixa't com inclou les coordenades. No estàs veien tota la seqüència, només la que hi ha en aquestes coordenades
      6. Descarrega-la usant el menu "Send to" -> "File"->Format FASTA (posa-li un nom que tingui sentit pe: leg\_Y00401-dna.fasta)
      7. A la capçalera dels fasta afegeix un espai entre el codi ncbi i els ":" que hi ha abans de les coordenades
      8. Torna a l'entrada original en format gb
      9. En el CDS de la proteïna clica sobre el codi de la proteïna codificada per aquest gen en el camp /protein\_id (CAA68462.1) per saltar a la DB protein
      10. Descarrega't la seqüència de la proteïna en format fasta "Send to" -> "File"->Format FASTA (posa-li un nom que tingui sentit pe: leg\_CAA68462-pep.fasta)
   2. La seqüència proteica de l'entrada "AAA35597" de la DB "proteins" del ncbi. Instruccions per fer-ho:
      1. Cerca al ncbi dins la DB"proteins" i troba l'entrada "AAA35597" en format gb
      2. Descarrega-la usant el menu "Send to" -> "File"->Format FASTA (posa-li un nom que tingui sentit pe: bet\_AAA35597-pep.fasta)
      3. En el cap "DBSOURCE" clica sobre el codi de la base de dades nucleotides "M25079.1". Aquesta és l'entrada a nucleotides de la qual s'ha traduït la seqüència peptídica que has descarregat anteriorment.
      4. Reprodueix els punts iii a viii del apartat a. per descarregar-te la seqüència del DNA codificant (posa-li un nom que tingui sentit pe: bet\_M25079-dna.fasta)
2. Ja tens totes les seqüències que usarem
3. Primer que res compararem les seqüències de DNA usant l'algorisme d'alineament global "Needleman–Wunsch" . Usarem la implementació feta per l'ncbi dins la interfície web del blast.
   1. Cerca a google ncbi blast needleman. O ves directament a aquesta URL: <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&PROG_DEFAULTS=on&BLAST_INIT=GlobalAln&BLAST_SPEC=GlobalAln&BLAST_PROGRAMS=blastn>
   2. Selecciona Nucleotide
   3. Clica "Reset page"
   4. Enganxa en format fasta la seqüència Y00401 descarregada al primer requadre i M25079 al segon. NOTA 1: Les que t'has descarregat prèviament. NOTA 2: En format FASTA! (incloent ">" i tota la capçalera)
   5. Marca "Show results in new window" i clica BLAST
   6. A la nova pàgina fixa't en la capçalera que conté dades sobre el què acabes d'executar. Sobretot amb Qui és la query? Qui és el subject? Quina mida té la query? Quina mida té el subject?
   7. Clica sobre la pestanya "Alignments" i observa els resultats numèrics de l'alineament (Identitat i gaps)
   8. Observa també com s'han alineat aquestes dues seqüències de DNA. Fixa't que al ser un algorisme d'alineament global tota la seqüència "query" ha quedat alineada amb la seqüència "subject"
   9. Descarrega't l'alineament en format "text" (així anomena el blast al seu format). Sembla que hi ha alguna mena de problema al generar aquest fitxer i s'inclouen etiquetes html que no hi haurien de ser, tingues-ho en compta a l'hora de llegir-lo (Sí, a vegades els programes tenen errors i hem d'aprendre com sobreviure'ls). (posa-li un nom que tingui sentit pe: leg\_vs\_bet-dna-needle.blast.txt)
   10. Clica a la pestanya Dot plot (Exercici autònom: A la plana del moodle d'aquest subtema tens un vídeo de 2 minuts on s'explica què és el dot plot. Amb aquesta ajuda i amb el què saps de la query i de la subject, interpreta el què veus)
   11. No tanquis aquesta pàgina.
4. Ara alinearem les seqüències peptídiques d'aquestes proteïnes usant el mateix mètode d'alineament global "Needleman–Wunsch".
   1. Torna a obrir l'eina del ncbi blast Needleman–Wunsch.
   2. Selecciona protein i clica "Reset page"
   3. Al primer "textbox" enganxa, en format fasta, la seqüència peptídica de la leghemoglobin I de *Lupinus luteus* (CAA68462) i al segon la de la beta-globina humana (AAA35597).
   4. Marca "Show results in new window" i clica BLAST
   5. A la nova pàgina fixa't en la capçalera que conté dades sobre el què acabes d'executar. Sobretot amb Qui és la query? Qui és el subject? Quina mida té la query? Quina mida té el subject?
   6. Clica sobre la pestanya "Alignments" i observa els resultats numèrics de l'alineament (score, Identitat, "positives" AKA similitud, gaps)
   7. Observa també com s'han alineat aquestes dues seqüències peptidiques. Fixa't que al ser un algorisme d'alineament global tota la seqüència "query" ha quedat alineada amb la seqüència "subject"
   8. Descarrega't l'alineament en format "text" (així anomena el blast al seu format). Sembla que hi ha alguna mena de problema al generar aquest fitxer i s'inclouen etiquetes html que no hi haurien de ser, tingues-ho en compta a l'hora de llegir-lo (Sí, a vegades els programes tenen errors i hem d'aprendre com sobreviure'ls). (posa-li un nom que tingui sentit pe: leg\_vs\_bet-pep-needle.blast.txt)
   9. Clica a la pestanya Dot plot (Exercici autònom: A la plana del moodle d'aquest subtema tens un vídeo de 2 minuts on s'explica què és el dot plot. Amb aquesta ajuda i amb el què saps de la query i de la subject, interpreta el què veus)
   10. No tanquis aquesta pàgina.
5. Compara l'alineament de DNA amb l'alineament de proteïna. Creus que aquestes dues seqüències estan relacionades? Els dos alineaments estan representant el mateix, tenint en compte el seu significat biològic: l'evolució d'aquest gen/proteïna des d'un ancestre llunyà comú. Creus que estan indicant el mateix? (recorda les seqüències de proteïna són una traducció directa de la seqüència de l gen, per tant el què li passa al gen li ha de passar a la proteïna. És així?)
6. Ara ho repetirem tot però usant un algorisme d'alineament local (Smith–Waterman). Començarem pel DNA.
   1. Cerca a google emboss water o ves directament a la URL <https://www.ebi.ac.uk/jdispatcher/psa/emboss_water>
   2. Selecciona DNA
   3. Enganxa al primer "textbox", en format fasta, la seqüència Y00401 descarregada i M25079 al segon.
   4. Observa els paràmetres que s'han usat per a córrer el programa
   5. Observa els resultats numèrics de l'alineament (Identitat i gaps)
   6. Observa l'alineament. Tota la seqüència d'ambdós gens ha quedat alineada?
   7. Compara aquest alineament amb el què has obtingut usant el mètode d'alineaments globals Needleman–Wunsch. Són iguals? De nou pensa en el significat biològic de l'alineament poden ser correctes els dos a la vegada?
   8. Descarrega els resultats del alineament al apartat "Result Files". (posa-li un nom que tingui sentit pe: leg\_vs\_bet-dna-water.pair.txt)
   9. No tanquis la pàgina
7. Per finalitzar alinearem les seqüències peptídiques d'aquestes proteïnes usant el mateix mètode d'alineament local: Smith–Waterman.
   1. Obra en una nova pestanya l'emboss water.
   2. Selecciona "Protein"
   3. Al primer "textbox" enganxa,en format fasta, la seqüència peptídica de la leghemoglobin I de *Lupinus luteus* (CAA68462) i al segon la de la beta-globina humana (AAA35597)
   4. Observa els paràmetres que s'han usat per a córrer el programa.
   5. Observa els resultats numèrics de l'alineament (score, Identitat, similitud, gaps)
   6. Observa l'alineament. Tota la seqüència d'ambdues proteïnes ha quedat alineada?
   7. Compara aquest alineament amb el de DNA que has obtingut amb la mateix mètode. De nou els dos alineanemts estàn representat el mateix procés biològic, però en canvi el qué mostres és diferent.
   8. Ara compara aquest alinemant de proteïnes amb el què has obtingut usant el mètode d'alineaments globals Needleman–Wunsch i de nou fes-te el plantejament del significat biològic
8. Reflexió: Tenim 4 alineaments diferents que tots representen exactament el mateix procés biològic. Un és més correcta que els altres?