1. Het basis NCBI BLAST programma wordt in verschillende varianten aangeboden: blastn, blastp, tblastn, blastx, tblastx   
   Elke daarvan heeft een eigen bestaansrecht; je moet de variant gebruiken die bij je vraag en de data die je hebt (je query) aansluit.   
   Als je een DNA query hebt, en je wil aantonen welk eiwit in de NR eiwit-database het meest lijkt op het gen in je DNA sequentie, welke BLAST variant gebruik je dan?
   1. Blastn
   2. Blastp
   3. Tblastn
   4. Blastx
   5. Tblastx
2. Pairwise sequence alignment gebruikt om een alignment tussen eiwitsequenties te beoordelen een scorematrix; we hebben twee typen besproken, PAM en BLOSUM. Waarom gebruiken we zo'n scorematrix in plaats van gewoon penalties voor mismatches tussen de aminozuren in het alignment?
   1. Omdat aminozuren, anders dan nucleotides, biochemisch op elkaar lijken en elkaars functie overnemen. Scoringsmatrix geeft de kans op verandering
   2. Enkele penalty voor mismatches gezien grote aantal verschillende aminozuren (20) vergeleken met nucleotides (4) geeft slechte resultaten
   3. Eiwitsequenties zijn beter geconservererd dan nucleotidesequenties
   4. Geen scoringmatrix en veel mismatches leidt tot een algoritme die vastloopt
3. Je hebt twee ver van elkaar afstaande eiwitten die je met pairwise sequence alignment wilt vergelijken.   
   Welke BLOSUM of PAM matrix zou je gebruiken om ze te vergelijken?
   1. BLOSUM45 / PAM250
   2. BLOSUM45 / PAM1
   3. BLOSUM80 / PAM250
   4. BLOSUM80 / PAM1
4. Als je met PSI-BLAST een goedwerkende PSSM (Position Specific Score Matrix) hebt gevonden, kun je die PSSM opslaan in de PFAM database. Later kun je deze hergebruiken door een specifieke BLAST variant te gebruiken; deze zoekt een query eiwit tegen een database met al deze PSSMs. Hoe heet deze BLAST variant?
   1. PHI-BLAST
   2. RPDS-BLAST
   3. HMM-BLAST
   4. PFAM-BLAST
5. Wat zijn de voordelen van Oxford Nanopore sequencing ten opzichte van Ilumina sequencing?
   1. MinION Nanopore Sequencing leest individuele moleculen DNA uit, dus variatie per cel is goed te zien
   2. MNS leeset kort DNA uit (readsize groot), gelimiteerd door hoe voorichtig DNA is geconserveerd
   3. Goedkoper dan Illumina per sequence run
   4. Veel sneller dan Illumina voor zelfde aantal nucleotides
6. Het BLAST algoritme is sneller omdat er een afweging wordt gemaakt tussen snelheid en de mogelijkheid van vals negatieven; het garandeert niet dat de gerapporteerde alignments de allerbest mogelijke zijn. Welke van de "trucs" die BLAST toepast om die versnelling te maken, zorgt ervoor dat je mogelijk goede alignments mist (vals negatieven)?
   1. Threshold parameter
   2. Gebruik van words in index
   3. Alignment threshold waardoor alleen goede local alignments
   4. Het feit dat BLAST alleen LOCAL geeft
7. Welke van de volgende zinnen beschrijft het best het verschil tussen local en global alignment?
   1. Global alignment wordt meestal gebruik voor DNA sequenties; local alignment wordt meestal gebruikt voor eiwitsequenties.
   2. Global alignment staat gaps toe terwijl local alignment dit niet doet.
   3. Global alignment vindt een globaal maximale alignment score; local alignment vindt een lokaal maximale alignment score.
   4. Global alignment maakt een alignment met de gehelle sequenties; local alignment vindt het beste deel (hoogst scorende deel) van de sequenties.
8. Als je BLASTP hit een E-value van 0,0 heeft weet je zeker dat er sprake is van homologie tussen het query eiwit en hit in de eiwitdatabase. (Ja/Nee)
9. Illumina sequencing werkt net als Sanger sequencing met het uitvoeren van een …, die stap voor stap met per nucleotide uitleest. Dit uitlezen gebeurt met een … label dat door een camera wordt geregistreerd. Maar, anders dan met Sanger sequencing, is Illumina sequencing helemaal geminiaturiseerd op een glasplaatje; de … . Omdat alle reacties daarop plaatsvinden, en de reagentia op een gegeven moment op zijn, zijn de reads van Illumina sequencing van kleinere … dan bij Sanger sequencing. Door middel van library prep kunnen er meerdere samples op een laantje van de glasplaat gemultiplexed worden; dan wordt er tijdens de prep een … aan het DNA gehangen.   
   Gebruik de woorden: barcode, flowcell, lengte, PCR, fluorescerend
10. Pairwise alignment is een voorbeeld van … programming. Het zet de twee sequenties die je wilt alignen in een matrix tegen elkaar uit: sequentie aan de … kant en een aan de … . Het lost het probleem van te veel paden door de alignment matrix op door om te beginnen naar slechts 1 positie te kijken; helemaal links … . In dit hokje kijk je naar de score (of gap penalty) komende van een hokje links, boven of … . Je kiest dan de hoogste score (of minst negatieve en markeert dat. Dit doe je voor alle hokjes, diagonaal naar de rechteronderhoek toe. Om dan het beste pad uit te lezen, en dus het hoogst-scorende alignment, loop je van … naar … terug. Je volgt steeds per vakje de richting naar het volgende vakje met de hoogste score.  
    Gebruik de woorden: linksboven, dynamisch, diagonaal, rechtsonder, linker, boven, bovenkant
11. Leg de begrippen “homologie”, “global pairwise alignment”, “orthologie” en “paralogie” uit en hoe aan elkaar gerelateerd zijn.
12. Elke BLAST hit (HSP, high-scoring segment pair) heeft twee scores: een S en een E score. Leg uit wat deze twee scores zijn, hoe ze verschillen, en hoe je ze samen gebruikt om een BLAST hit te beoordelen op "echte" homologie.

|  |
| --- |
|  |

1. De active site is vaak het best geconserveerde deel van een eiwit. Leg uit waarom een PSSM gemaakt met PSI-BLAST beter is in het vinden van de active site van een query in andere eiwitten in de nr database dan normale BLASTP.
2. Leg uit hoe Illumina ("NGS") sequencing van Sanger sequencing verschilt, qua proces en qua sequenties, en waarom het daarom zo belanrijk is bij Illumina sequencing kwaliteitsinformatie over de gevonden sequenties mee te leveren.

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | | |  | | --- | |  | | |

1. Er zijn een aantal voorgekookte databases met sequenties beschikbaar om tegen te BLASTen. Twee daarvan zijn de nr database en de RefSeq database (voor zowel nucleotide als eiwitsequenties). Leg uit wat het verschil tussen deze twee type databases isorthologie