

Beantwoordt de vragen zo volledig mogelijk, laat je kennis zien, dan kan ik je ook bij een incompleet antwoord zoveel mogelijk punten geven! Antwoorden mogen in het Engels of Nederlands.

- 1) Leg de begrippen “homologie”, “global pairwise alignment”, “orthologie” en “paralogie” uit en hoe aan elkaar gerelateerd zijn. (8)
- 2) BLAST helpt met het ontdekken van conservering en homologie van een query (dna of eiwit) met een hele grote database.
 - a. Waarom zijn aminozuur-sequenties beter geconserveerd dan nucleotide-sequenties? (2)
 - b. Welke BLAST-variant moet je gebruiken als je een gen (dna) op wilt zoeken in een database met eiwitten? (2)
 - c. Waarom zou je zo’n vertaalde query willen doen? (2)
 - d. Waarom heb je voor een nucleotide alignment geen BLOSUM of PAM matrix nodig? (2)
- 3) Het BLAST algoritme staat alignment tussen een query (vraag) sequentie en een enorme database van andere sequenties toe.
 - a) Wat is de “wordsize” van een BLAST query? (2)
 - b) Als ik de “wordsize” van 11 naar 22 vergroot, wat voor effect heeft dat op de tijd die mijn BLAST query neemt? Welk ander effect heeft deze parameter vergroten? (4)
 - c) Welke andere parameter beïnvloedt de snelheid van het zoeken in de database met een word? (2)
- 4) Multiple choice: omcirkel de letter van het juiste antwoord.
Een Position Specific Score Matrix (PSSM) heeft 20 kolommen (voor elk aminozuur 1) en heeft een aantal rijen corresponderend aan de lengte van het query eiwit. Waarop zijn de individuele scores in de matrix gebaseerd? (4)
 - a. Een PAM of BLOSUM matrix.
 - b. De frequentie van voorkomen in een Multiple Sequence Alignment.
 - c. De score van de aminozuren links, linksboven en boven in de matrix.
 - d. De achtergrondfrequentie van het voorkomen van dat aminozuur in alle eiwitten.
- 5) BLAST is een heuristisch algoritme; de resultaten van een query zijn niet gegarandeerd juist maar er hangt een kans aan dat de gevonden alignment klopt; de E-value. De formule is;

$$E = K m n e^{(-\lambda S)}$$

- a) Kun je de waarden K en λ (lambda) aanpassen per query? Waarom wel, of niet? [3]
 - b) Als je de database van sequenties waartegen de query gedaan wordt verdubbelt, hoe verandert de E-value dan bij gelijke alignment score S ? [3]
 - c) De E-value formule is een zogenaamde *negatieve exponentiële* functie. Wat betekent dit voor de E-value van een hele goede (hoge) alignment score's S ? [4]
- 6) BLOSUM62 is gekozen als de default scorematrix voor BLAST. Waarom werkt een BLOSUM matrix beter bij het beoordelen van homologie in een alignment dan een PAM matrix? (5)

- 7) Multiple choice: omcirkel de letter van het juiste antwoord.
Scorematrices zijn er in verschillende mate van strengheid; bijvoorbeeld PAM10, BLOSUM45, BLOSUM90. Hoe kies je een matrix met een bepaald getal uit voor een pairwise alignment?(4)
- a. Het getal geeft de snelheid aan waarmee de matrix een score levert.
 - b. Het getal geeft de mate van overeenkomst tussen de sequenties.
 - c. Het getal geeft de penalty van een gap opening, naast de score per aminozuur.
 - d. Het getal geeft de gemiddelde score van een aminozuur-match in de matrix.
- 8) 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 is en wanneer je voor de ene of de ander zou kiezen. (5)
- 9) De active site is vaak het best geconserveerde deel van een eiwit.
- a. 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. (3)
 - b. Als je eenmaal een goede PSSM hebt gemaakt, kan die gesaved worden en opgeslagen in een database, zodat je hem weer kan gebruiken. Hoe heet de database van NCBI met eerder gemaakt PSSM's? (2)
 - c. Welke BLAST gebruik je als je deze PSSM's wilt hergebruiken? (2)
- 10) Multiple choice: omcirkel de letter van het juiste antwoord.
Welke output van een BLAST search geeft een schatting van het aantal vals positieven die de search in zijn alignments heeft? (4)
- a. Bit score
 - b. Percentage identity
 - c. E-value
 - d. Alignment score S