# Bioinformatics - Encyclopedia Academia



## Bioinformatics

| Bioinformatics | |
|---|---|
| Richting | Informatica |
| Jaar | MINF |
| Studiepunten | 6 |

## Examenvragen

### Academiejaar 2022-2023 (juni)

Note: the images used are mainly from the slides. During the exam, other images with othere values were used.

1. Given the PAM250 matrix below, are following statements true or false?

Bestand:PAM250.png

1.
    1. Amino acid W has a higher mutation rate than amino acid I.
    2. A mutation from I to K is less likely to occur than a mutation from K to I.
    3. If the mutation score is negative, then the substitution doesn't occur.
    4. A BLOSUM matrix is mainly used for local alignment.
    5. A BLOSUM matrix is for more recent data than a PAM matrix.
    6. A substitution from I to M is less likely to occur than by chance.
    7. A BLOSUM matrix assumes a Markov process.
    8. Each PAM matrix is genrated from matrix multiplications of a single PAM matrix.
    9. A PAM12 matrix is for more recent comparisons than a PAM250 matrix.

1. Given the four sequences below:

- A: TCCGAG - B: AAGTAC - C: AAGTTG - D: ACCGAG

Create a phologenetic tree using max. parsimony. Explain why it is the most optimal tree.

1. Given the following alignment matrix.

Bestand:Alignment.png

1.
    1. Which algorithm is used to create this matrix?
    2. What is the used gap penalty?
    3. Writ down the alignment of the sequences. Be sure to put a '-' when there is a gap and a '|' when there is a match.
    4. What is the score of the alignment?

Bestand:Alignmentx.png

1. Given the following incomplete alignment matrix.
    1. What is the gap penalty?
    2. What is the score at position x?

1. A Markov Matrix is created based on the sequences below.

- Id: 123456 - S1: ACT-GA - S2: AC--GA - S3: TCT-GA - S4: TCT-GA - S5: AGTCCA - Ma: **  **

The match states are marked with a '*'.

Bestand:Markov.png

1.
    1. What is the path for sequence S3?
    2. What are the emission probabilities for given match states:
        1. M1
        2. M3
        3. I2
    3. What are the transition probabilities for all transitions originating in following states?
        1. M2
        2. I2
        3. M4

1. A BLAST search gave a hit with a normalized score of 143 and an E-value of 10−2110−21.
    1. Would you trust this hit? Why?
    2. What is the meaning of an E-value?
    3. Can you compare this score when searching the same sequence against a different database?

1. Given the network below

Bestand:Network.png

1.
    1. Which node(s) has the highest degree centrality? What is the value?
    2. Which node(s) has the highest betweenness centrality? What is the value?
    3. Which node(s) has the highest eccentric centrality? What is the value? (we had to calculate this for a way too large graph on the exam)
    4. What is the clustering coefficient of node 7?
    5. What is the diameter of the network?

1. You sequenced a gene. After this, you resequenced the same gene with a higher sequuence error rate. What is the effect on the % identity and the E-value? (select the right option)

- % identity will be higher and E-value will be lower - % identity will be lower and E-value will be higher - % identity will be lower and E-value will be lower - % identity will be higher and E-value will be higher - % identity will be the same and E-value will be the same

1. I performed a BLAST search on Sequence_A and sequence_B was outputted. Which BLAST variant was used?

   Sequence_A

pctkaewwtagtfiqikgyqmnwcklyitpcnppscinqwmifyscshseecsnmsgwslycfrrptecpdrigetephveynqwgfkc

   Sequence_B

accactaatcatgtcaggaccgcctccgccggtggcttcataactttccgtagcagtggg
ggcaagtcgggtttctgggaccggaggagtatcagggtcgaccatcgccctgtggttggg
gcacatatgaagcaatgtcacgtattggagagcttctgacgctgtcttccggtggcgtga
ccgaatccgcctccagaagaggactatagcctggtctcattaactctgagatccgcacgt
tgatgtacacgcgcacgtatgtcgttagggccgcgagttatgtaaacggcctgatccaag
atcctacactcattcgccatacgccgatgttgattctgagtgccgagttgccttgcaagg

1. How would an increased word size parameter influence the results of a BLAST search?

1. Assume sequence-1 evolved from sequence-2. Given following dot plot.

Bestand:Dotplot.png
(note: on the exam, we got a dotplot where two events occured)

1.
    1. Did the length of the sequences remain the same?
    2. Which event occurred that changed the DNA?
    3. How do you find repeated parts in a sequence using a dot plot? Explain in detail.
    4. Sequence C and D are both 100 nucleotide. Sequence D evolved from sequence C by an inversion of nucleotide 50-80. What would the dot plot look like?

1. Which of the following databases do you consult to get information about the secondary structure of a protein?

- Protein Data Bank (PDB) - ProtScale - GenBank - VariantDB - dbSNP - none of the above

1. The CTFR gene can have pathogenic and protective mutations. Explain the tools and databases you would use to identify the biological processes the protein interacts with and check whether a specific sequenced gene has a pathogenic mutation.

1. Which do you need to do before creating a motif visualization of sequences that don't have the same length?

- Log transformation - Multiple Sequence Alignment - BLAST - adjust bit score

1. Given the ASCII table below and the outputted file of an Illumnia sequencer (with a +33 offset).

Bestand:Ascii.png
Bestand:Illumina.jpg

1.
    1. What is the probability the indicated nucleotide is correct?
    2. How many nucleotides have a Phred score between 32 and 36 (both excluded)?

1. Given following output in IGV. What does this data show us? And do we have enough evidence to accept this?

Bestand:Igv.jpg

1. Given following FastQC output. Which machine is this most likely sequenced with? What is the next step?

Bestand:Fastqc.png

1. Given following 4 cladograms. True or False?

Bestand:Cladogram.png

1.
    1. Tree A and B are equivalent
    2. Tree B and C are equivalent
    3. Frogs and crocodiles are paralogs
    4. Frogs and salamander are orthologs
    5. Birds are closer related to lizards than to turtles

1. Given the tree below, is a frog closer related to a fish or to a human? Explain why.

Bestand:Frogtree.png

1. Given the cladogram with scores below. How much support is there that species 1, 2 and 3 belong to the same clade? What are the values called and how are they calculated?

Bestand:Bootstrap.jpg

1. Given the hydrophobicity and transmembrane graphs.

Bestand:Hphob.jpg
Bestand:Transmem.jpg

1.
    1. Do the graphs belong to the same protein? Explain why (not).
    2. Which tool was used to create the graph?
    3. Is the primary or secondary protein structure used to create this graph?