



Grundlagen der Bioinformatik

SoSe 2018

Assignment 03

Submit electronically in Ilias by 7.5.2018, 10h

1 Keyword tree definition (1 point)

Please provide an exact definition of a *keyword tree*.

2 Keyword tree example (2 points)

Please **draw** the *keyword tree* for the following words: **first, wire, win, fit, firmly and firmer.**

3 NCBI BLAST Search (2 points)

Using the NCBI website, please **determine** the **most likely gene** and **organism** assignment for this DNA sequence:

```
>Read666
GGTTCGAGGATGATCGGCTTGGCAGCAGGAATCGCCTCGATGGTCGCTTTCTTGGCGGCCATGAAAAAGGCGATGTCCTTG
CCGTCGACCGGATGTGACTTGCCATCATGCACGGTGACGCGCAGATCCTCCACCGGGAAGCCGGCAACGACGCCGTCAGCC
AGCGCCTGACGCACGCCCTTCTCGACTGCCGCCATGAAGACTCCGGGAATCAGCCGCCCTTGACGATATCGACGAACTCG
AAGCCGGCGCCCGTTTCGAGCGGCTCGTCCCGCAACATCACTTCTCC
```

4 FASTA header and BLAST match (5 points)

Determine in which project the following DNA sequencing read was produced:

```
>gn1|SRA|SRR172902.1 USI-EAS376:1:1:1:1204
CGGTCGTA CTGCGCTGGGCCNCGCCAGCGCCAGCCGAGTNGATTTCTAACGCCTGC
CGGGANGCATGGCCG
```

Please explain each line of the following BLAST match that was obtained for the above sequence:

```
>WP_104992252.1 MULTISPECIES: integrase [Deinococcus] PTA66341.1 integrase [Deinococcus sp. OD32]
Length=236
```

```
Score = 44.7 bits (104), Expect = 3e-04, Method: Compositional matrix adjust.
Identities = 21/24 (88%), Positives = 22/24 (92%), Gaps = 0/24 (0%)
Frame = -2
```

```
Query 74 GHASRQALEIXSRLALGXAREYD 3
        GHASRQ+LEI SRLALG AQREYD
Sbjct 205 GHASRQSLEIYSRLALGEAQREYD 228
```

In addition, which flavor of BLAST (i.e. BLASTN, BLASTX, BLASTP) was used here? Which reference database was most likely used?