|  |  |
| --- | --- |
| GBIO0002 | 1 |
|  |  |

**Homework 2**

Genetics and bioinformatics

**Important dates:**

* Submit report /presentation files before 10 December 2018,
* The presentation will be on 11 December 2018

**Marks:**

* Part 1- 15 marks (select either Q&A or literature-based homework)
* Part 2- 15 marks (select either Q&A or literature-based homework)

**Evaluation:**

* For the Q&A homework your work will be evaluated based on the accuracy and completeness of answers. Write your answers in the form of a report.
* For the literature-style homework, evaluated based on the completeness of your slides, your presentation skills, your understanding of the matter, and your answers to questions in class (see the information given during the introductory class). No report is needed. Recall the document “critical evaluation of a report/paper”.

**Instruction:**

* Form a group of 2-3 persons and complete the homework in both parts (1&2). Make sure there are about 8 groups maximum.
* For the presentation, you have to present your selected paper within 15 minutes. Everybody in your group should present something.
* A report / slides presentation (whatever is applicable – see above) needs to be submitted in electronic format via the website by the deadline. Please note that the submission system will be closed automatically.
* Compress all files into ONE zip file before submission.

**Part 1:** Genetics and DNA sequencing.

**This part will be uploaded by today afternoon.**

**Part 2:** Genome-wide association studies - analytics.

**Q&A homework**

**Question: I**

(I) Perform BLAST of following sequence using the NCBI blastn server against NR (No redundant) database. Based on BLAST results, estimate to which species these sequences belong.

**Copy the BLAST output on which your assertion is based.**

(a) Human (b) Mouse (c) none of the above

>sequence 1

GACATCGTGATGACCCAGTCTCCTTCCACCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCCAGTCAGAGTATTAGTAGCTGGTTGGCCTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGTTCCTGATCTATAAGGCGTCTAGTTTAGAAAGTGGGGTCCCATCAAGGTTCAGCGGCAGTGGATCTGGGACAGAATTCACTCTCACCATCAGCAGCCTGCAGCCTGATGATTTTGCAACTTATTACTGCCAACAGTATAATACGTTCGGCCAAGGGACCAAGCTGGAGATCAAAC

>sequence2

CAGGTGCAGCTGGTGCAGTCTGGGGCTGAGGTGAAGAACCCTGGGGCCTCAGTGAAGGTCTCCTGCAAGGCTTCTGGATACACCTTCACCGGCTACTATATGCACTGGGTGCGACAGGCCCCTGGACAAGGACTTGAGTGGATGGGCTGGATCAACCCTAACAGTGGTGGCACAAACTATGCACAGAAGTTTCGGACTGGGTCACCATGACCAGGGACACGTCCATCAGCACAGCCTATATGGAGCTGAGCAGGCTGAGATCTGACGACACGGCCGTTTATTACTGTGCGAGAAATATAGCAACAACTGGTGATGCTTTTGATATTGGGGCCAAGGGACAATGGTCACCGTCTCCTCAG

**Question: 2**

Please align globally using Needleman–Wunsch algorithm the following DNA sequences.

Use the following scoring rules:

***a) gap -5;***

***b) match between two nucleotides +5;***

***c) mismatch between two nucleotides +3;***

**Sequence A**: CTTGAA

**Sequence B:** CTT

* Show the alignment matrix (template given below)
* Show the trace-back path giving the best alignment (i.e. arrows)
* Give all possible alignment (if multiple)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **C** | **T** | **T** | **G** | **A** | **A** |
|  |  |  |  |  |  |  |  |
| **C** |  |  |  |  |  |  |  |
| **T** |  |  |  |  |  |  |  |
| **T** |  |  |  |  |  |  |  |

2) Align locally the following sequences using the following scoring rules:

***a) gap -2;***

***b) mismatch -1;***

***c) match +2;***

Sequence A: GGTATACC

; Sequence B: TATA

* Show the alignment matrix and trace-back path

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **G** | **G** | **T** | **A** | **T** | **A** | **C** | **C** |
|  |  |  |  |  |  |  |  |  |  |
| **T** |  |  |  |  |  |  |  |  |  |
| **A** |  |  |  |  |  |  |  |  |  |
| **T** |  |  |  |  |  |  |  |  |  |
| **A** |  |  |  |  |  |  |  |  |  |

**Literature-style homework**

1. Dubchak, I. and Frazer, K., 2003. Multi-species sequence comparison: the next frontier in genome annotation. Genome biology, 4(12), p.122.
2. Boyle, E.A., Li, Y.I. and Pritchard, J.K., 2017. An expanded view of complex traits: from polygenic to omnigenic. Cell, 169(7), pp.1177-1186.
3. Frazer, K.A., Elnitski, L., Church, D.M., Dubchak, I. and Hardison, R.C., 2003. Cross-species sequence comparisons: a review of methods and available resources. Genome research, 13(1), pp.1-12.