



Mansoura University
Faculty of Computers and Information Sciences
Department of Computer Science
Final Examination – Second Semester- 2016-2017
Course: Bioinformatics (UNV 310)



Grade: Third Year (Agriculture Biotechnology) Exam Date: June 8, 2017
Time Allowed: 2 Hours Total Marks: 60 Questions in 4 Papers

Answer All the following Questions

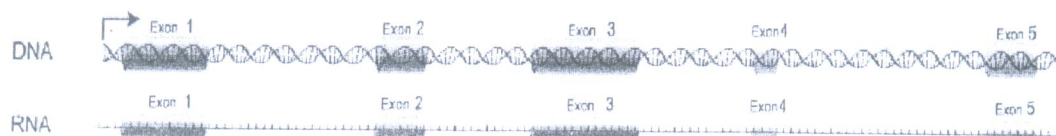
Question 1. [25 Marks]: Select the Correct Word

1. The alignment procedure that tries to align regions with high level of matches without considering the alignment of rest of the sequences is:			
a) multiple sequence alignment	b) pair-wise alignment	c) global alignment	d) local alignment
2. Which of the following is a sequence alignment tool provided by NCBI:			
a) Chime	b) BLAST	c) FASTA	d) Clustal-W
3. The procedure of aligning many sequences simultaneously is called:			
a) multiple sequence alignment	b) pair wise alignment	c) global alignment	d) local alignment
4. Sequence alignment helps scientists:			
a) to trace out evolutionary relationships	b) to infer the functions of newly synthesized genes	c) to predict new members of gene families	d) all of these
5. The transcription process can be defined by the following equation			
a) DNA → DNA	b) DNA → Protein	c) DNA → RNA	d) RNA → Protein
6. The direction of the sense strand is:			
a) 3' → 3'	b) 3' → 5'	c) 5' → 3'	d) 5' → 5'
7. The following organisms are considered Eukaryotes except:			
a) human	b) plants	c) archea	d) fungi
8. The basic unit of heredity is:			
a) Genome	b) Gene	c) Transcriptome	d) Proteomics
9. "AUGACACUGU" is a sequence from:			
a) DNA	b) Protein	c) RNA	d) none of these
10. RNA polymerase binds to a			
a) promoter	b) gene	c) DNA	d) protein
11. A chain of amino acids is called:			
a) ribosome	b) RNA	c) gene	d) polypeptide
12. The two polypeptide chains of Hemoglobin molecule are:			
a) γ and δ	b) α and β	c) α and φ	d) α and π
13. Which of the following RNA's carries amino acids to ribosomes:			
a) rRNA	b) tRNA	c) miRNA	d) lncRNA
14. The analysis of mixture of DNA sequences extracted from different organisms is called:			
a) Metagenomics	b) Transcriptomics	c) Proteomics	d) Genomics
15. Which of the following is responsible for synthesizing a complementary DNA strand:			
a) RNA polymerase	b) Ribosome	c) DNA polymerase	d) miRNA

16. The name of FASTQ read starts with:			
a) @	b) #	c) >	d) +
17. The name of FASTA read starts with:			
a) <	b) >	c) +	d) \$
18. Base quality of the FASTQ read usually encoded in:			
a) Phred+44	b) Phred+22	c) Phred+33	d) none of these
19. The computer program that stitches the short reads together into longer sequences is called:			
a) Assembler	b) Aligner	c) Sequencer	d) Gene finder
20. My PhD thesis was in:			
a) Sequence Assembly	b) Sequence Alignment	c) DNA Sequencing	d) none of these
21. Contigs are assembled further into:			
a) Reads	b) Scaffolds	c) Chromosomes	d) none of these
22. My brother name is AMR, the best alignment of his name to my name is:			
a) -AMR- SA-RA	b) AMR- SARA	c) AMR----- ---SARA	d) none of these
23. The most significant alignment reported by the following E-values is:			
a) $2e^{-6}$	b) $2e^{-3}$	c) $2e^{-1}$	d) $2e^{-20}$
24. The BLAST tool that is used to compare protein sequence with sequences in one of protein databases is:			
a) blastb	b) blastn	c) blastx	d) tblastn
25. If the Affine gap penalty model (match: 1, mismatch: -1, gap opening: -2, gap extension: -1) is used to score the following alignment,			
<div style="text-align: center;"> -AMR- SA-RA The total similarity score is </div>			
a) -5	b) 0	c) -1	d) none of these

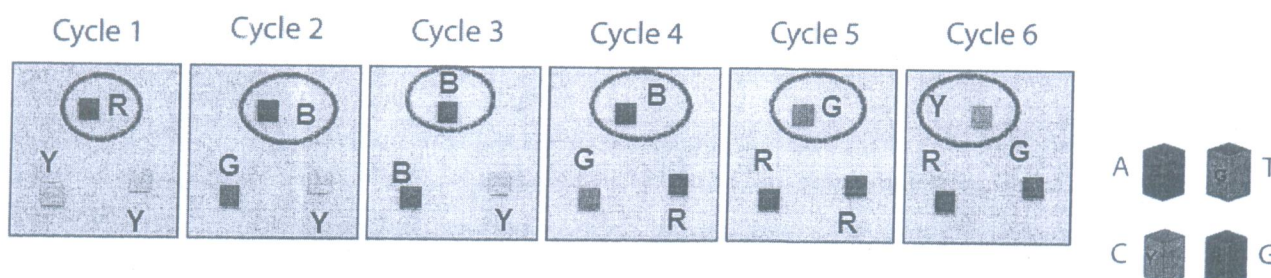
Question 2.[14 Marks]:Answer the Following:

- a. Suppose you have the following DNA, and RNA strands, give me three examples of alternative splicing. [3 Marks]



- b. Suppose you have the following set of pictures taken by a camera during sequencing by synthesis process. Give me the result of the sequencing process for the colors indicated by circles. [3 Marks]

Notes: R: (RED, G), B: (Blue, A), G: (Green, T), Y: (Yellow, C).



- c. Suppose you are working as Bioinformatics engineer in one of the leading Bio-Technology companies. You have a set of sequencing reads produced by Illumina platform (e.g. 100 sequences). Your team-leader asked you to do the following tasks:

1. Analyze the evolutionary relationships among the sequences and build a Phylogenetic tree.
2. Find the regions of similarity among the sequencing reads to predict the conserved patterns.
3. Combine the sequencing read together into one sequence and you have a reference genome.
4. Combine the sequencing read together into one sequence and you do not have a reference genome.

Describe your plan or your steps to complete these tasks, including the algorithms/tools you will use and why. [8 Marks]

Question 3.[8 Marks]:Answer the Following:

- a. The following is tree data produced by MUSCLE, Draw the corresponding Phylogenetic tree. [2 Marks]

```
(
(
DRR001336.843735:0.30064,
DRR001336.843734:0.22436)
:0.04354,
DRR001336.843732:0.11117,
DRR001336.843733:0.13473);
```

- b. Suppose you have three sequencing reads R1, R2, and R3. The overlap lengths among the sequencing reads are described by the following: $R_1 \rightarrow R_2 : 79$, $R_1 \rightarrow R_3 : 10$, $R_2 \rightarrow R_3 : 2$, $R_3 \rightarrow R_1 : 150$ [2 Marks]

1. Draw the overlap graph.

2. Find the order of visiting nodes according to greedy algorithm.

- c. You have a set of assembled contigs with the following lengths: Contig 1 is 2Kb, Contig 2 is 5Kb, Contig 3 is 20Kb, Contig 4 is 10 Kb, Contig 5 is 15 Kb, Calculate the N50 Score. [2 Marks]
- d. Suppose you have the following distance matrix, Draw the star diagram described by this matrix [2 Marks]

	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆
S ₁		-2	7	-1	0	-3
S ₂			5	-1	0	-4
S ₃				-1	-5	-5
S ₄					7	-1
S ₅						-3
S ₆						

Question 4.[8 Marks]: Based on the completed dynamic programming alignment matrix below, answer the following questions:

- a. Is this a Needleman-Wunsch (global) or Smith-Waterman (local) alignment? [1 Mark]

- What was the nucleotide match score parameter? Assume that any match (e.g., A-A, C-C, G-G, or T-T) receives the same score. [2 Marks]
- What was the mismatch score? This should be a negative number. [2 Marks]
- What was the gap penalty? Assume we used a fixed (linear) gap penalty. This should also be a negative number. [2 Marks]
- What is the score of the highest-scoring local alignment? [1 Mark]

		T	C	T	C	A	C
	0	0	0	0	0	0	0
T	0	4	2	4	2	0	0
C	0	2	8	6	8	6	4
C	0	0	6	7	10	8	10
A	0	0	4	5	8	14	12

Question 5.[5 Marks]: Write your opinion about the content of Bioinformatics course for this year, including the advantages and disadvantages, mid-term exams and the final exam, your suggestions to improve the course content in the future. (You can answer this question in Arabic)

End of Exam
My Best Wishes