

**CSE 4253: Fundamentals of Genomics and Proteomics**  
**Assignments for Lab**

**Lab Assignment Answering Guideline**  
**Submission deadline: Saturday (29-05-2021)**

Put all your code (preferably in Java) in a single file named (**roll\_firstname.pdf**) and upload it in the classroom.

**Write program code for the following:**

1. Consider partial digest  $L = \{1, 1, 2, 2, 2, 3, 3, 4, 4, 5, 5, 5, 6, 7, 7, 7, 8, 9, 10, 11, 12\}$   
Implement an algorithm to solve Partial Digest problem for L (i.e. find X such that  $\Delta X = L$ ).
2. Implement any dynamic programming algorithm for pair-wise global alignment of sequences  $S1 = \{CTCGCAGC\}$  and  $S2 = \{CATTCAG\}$ . Give the final alignment table given that  $E(a, -) = -2$ ,  $E(-, b) = -2$  and  $E(a, b) = 5$  for match  $E(a, b) = -2$  for mismatch.
3. Consider the distance of 5 sequences. Implement unweighted pair group method with arithmetic mean algorithm and construct the Phylogenetic tree.

	S1	S2	S3	S4	S5
S1	-	22	39	39	41
S2	-	-	41	41	43
S2	-	-	-	18	20
S4	-	-	-	-	10
S4	-	-	-	-	-

4. Implement an algorithm to find a de Bruijn graph for the sequence **GACTTACGTACT** with  $k = 3$  and generate the corresponding Eulerian walk.
5. Implement agglomerative algorithm with single link distance measure and produce a dendrogram tree for the following single continuous feature.

	a	b	c	d	e
feature	1	2	5	6	8