CURREnT sOlUTiOn FOR BiF 501 ( Finals)

MCQs:

1. Lcs scoring mechanism is for \_\_\_\_\_\_\_\_\_ sequence similarity analysis  
2. Physiological base approach \_\_\_\_\_\_\_\_\_\_ disease centric  
3. An association rule has two parts \_\_\_\_\_\_\_ if and then  
4. Transfer of gene among organism is called \_\_\_\_\_\_\_\_\_ horizontal gene transfer  
5. Eco R1 break down DNA at \_\_\_\_\_\_\_\_\_\_\_ GAATTC  
6. Programming language used for modeler \_\_\_\_\_\_\_\_ Python  
7. It is to compute the prefix function \_\_\_\_\_\_\_\_ O(m)  
8. Prob used by sequencing in hybridization have size \_\_\_\_\_\_\_\_ 8-32  
9. Which is correct \_\_\_\_\_\_\_ homology vs paralogy + orthology  
10. \_\_\_\_\_\_\_ helps in finding global alignment \_\_\_\_\_\_ Needleman wunsch  
11. Brute force algorithm use every possible ……….. to find the solution \_\_\_\_\_ l mer  
12. LCS scoring award 1 for \_\_\_\_\_\_\_\_ match  
13. …….. programming approach does not require prior sequence alignment\_\_\_\_\_\_\_\_\_\_ dynamic  
programming  
14. Brute force is a \_\_\_\_\_\_\_\_\_ motif finding problem  
15. DNA array is also known as \_\_\_\_\_\_\_\_ DNA chip  
16. GCG stand for \_\_\_\_\_\_\_\_\_ Genetics Computer Group  
17. Gene mapping is determining the …… on chromosome \_\_\_\_\_\_\_\_ location  
18. MS2 genom has \_\_\_\_\_\_\_ amino acid \_\_\_\_\_\_129  
19. Method use fr lesd validation\_\_\_\_\_\_\_\_ virtual screening  
20. Use for embryonic development \_\_\_\_\_\_\_\_\_\_ Homeobox genes  
21. concerned with the effect of genetic factors on reactions to drugs \_\_\_\_\_\_\_\_ Pharmacogenetics  
22. Sequence in EMBL formate start with \_\_\_\_\_\_\_\_\_\_ SQ  
23. an ordered sequence of symbols \_\_\_\_\_\_\_\_\_\_ pattern  
24. deletion in alignment in Edit graph shows \_\_\_\_\_ east ( horizontal arrow)  
25. how many methods for lead validation? \_\_\_\_\_\_\_\_\_\_\_ 5  
26. Chou fasman algorithm helps to find \_\_\_\_\_\_\_\_ 2ndary structure of protein

27. Plant mitochondria contain genes \_\_\_\_\_\_\_\_\_\_ 150 – 200kb  
28. Brut force is an ……. Pattern finding algorithm \_\_\_\_\_\_\_\_ Approximation algorithm  
29. Which identify the structure with minimum free energy \_\_\_\_\_\_\_\_\_\_ Ab initio  
30. Length of each line for FASTA algorithm should be \_\_\_\_\_\_ shorter than 80 or ( 60)  
31. Genome rearrangement results in a change of \_\_\_\_\_\_\_\_\_\_ gene ordering  
32. Human genome consist of repeats \_\_\_\_\_\_\_ 20 %  
33. Responsible for protein function \_\_\_\_\_\_\_\_ protein structure  
34. …… pattern matches with ***k*** mismatches or differences \_\_\_\_\_\_\_ approximate  
35. We look for molecule which need to be selected as target and those molecule that cause  
disease \_\_\_\_\_\_\_\_ target discovery  
36. GEO for \_\_\_\_\_\_\_ gene expression omnibus  
37. Intron start from5 end is ……… ans 3´ end is ……… \_\_\_\_\_\_\_\_\_\_\_ GT , AG  
38. SINE are ……… bp long \_\_\_\_\_\_80 – 300  
39. Needle man publicsh in \_\_\_\_\_\_\_\_\_ 1970  
40. Chou fasman algorithm was 1st used in \_\_\_\_\_\_\_\_\_\_ 1974 – 1978  
41. The notion of spectral similarity with shared peaks count \_\_\_\_\_\_\_\_\_ P1, P2-S1 and S2  
42. Human genome project \_\_\_\_\_\_\_\_\_ 1988  
43. TCGGGATTTCC is in drosophila for activate \_\_\_\_\_\_\_ immunity  
44. BLOSUM 6 is good for \_\_\_\_\_\_ un gaped alignment  
45. Disease that are hereditary \_\_\_\_\_\_\_\_ 3000 – 4000  
46. Computational technique use for protein not present in database \_\_\_\_\_\_\_\_\_\_\_ Denovo

**SUBJECTIVE:**

**Q. What is the run time of brute force algorithm?**Ans. Brute force algorithm to solve the Motif Finding problem  
running time of *O*(*l · n*t) . Cannot runs it on biological samples. It is a Faster greedy techniquenot correct but have good performance. It is an Approximation algorithm , Running time of this  
algorithm is *O*(*ln2 +lnt*), which is vastly better than the *O*(*lnt*) of IMPLEMOTIFSEARCH or  
even the *O*(4*lnt*) of BRUTEFORCEMEDIANSTRING

**Q. What is the difference between protein sequence and DNA sequence?**  
Created by: SABA ALWI Page 3  
Ans:

|  |  |
| --- | --- |
| **DNA** | **Protein** |
| A DNA sequence has a chain of nucleotides | A protein sequence has chain of amino acids |
| The nucleotides are A, C, G and T for the DNA and A, C, G and U for RNA. | There are around 20 amino acids for protein |
| Nucleotides are joined together through phosphodiester linkage | Amino acids are join together by peptide bonds |

**Q. What are substitution matrices? Also mention its types.**Ans: **Scoring or Substitution Matrices:**It is use for scoring amino acid substitutions in pairwise alignments. They reflect substitution  
rates that are originated by evolutionary events. For this we align the sequence and count the  
mutation at each position.  
**Types:**Some of the substitution matrices to compute sequence alignments are:  
PAM: Point Accepted mutations  
BLOSUM: BLOCK Substitution Matrix

**Q. Define suffix trees?**Ans: **Suffix Trees:**It is a compressed tree containing all the suffixes and allows many problems on strings to be

solved quickly.

**Q. What is the difference between smith water man algorithm and needle man wunshman  
algorithm in term of matrix filling?**Ans: There ar two main differences between them,  
**Needleman-Wunsch Algorithm**1. It performs global alignment on two sequences.  
Created by: SABA ALWI Page 4  
2. The final step in the algorithm is the trace back for the best alignment  
Start at the bottom-right corner. Follow where maximum value comes from  
**Smith-Waterman Algorithm**1. It is use to finds the best local alignment between two subsequences.  
2. Traceback starts with maximum value in the matrix and then go backwards,

**Q. What is MATE and MATE strategy?**Ans: During fragment assembly, the reads of Sequences whose suffix and prefix are  
overlapping and they are likely to be consecutives are called as mate fragments or simply mates.  
**Mates strategy:  
*Overlap***: Finding potentially overlapping reads  
***Layout***: Finding the order of reads along DNA  
***Consensus***: Deriving the DNA sequence from layout Best match between the suffix of one read  
and the prefix of another. By using this strategy we can sequence the whole fragment into  
single genome. Large number of reads to ensure that experimental errors are reduced to minor  
noise.

**Q. Which algorithm avoid the back tracking?**Ans**: KMP Algorithm:**A linear time algorithm for string matching which does not involve backtracking of strings  
Is called KMP algorithm.

**Q. What would be the lexicographic order of, (TAT,ATG,TGG,GGT,GTG,TGC) ?**Ans: Lexicographic order of that spectrum is , **ATG,GGT,GTG, TAT, TGC, TGG**

**Q. l=3, TATGGTGC find the spectrum?**Ans: If *l* = 3 and *s* = **TATGGTGC**,  
then its spectrum will be as follow,  
***Spectrum(s, l*)** = {**TAT, ATG, TGG, GGT, GTG, TGC**}  
Created by: SABA ALWI Page 5  
and its Lexicographic order will be,  
Lexicographic order, like **ATG,GGT,GTG, TAT, TGC, TGG**

**Q. What is denovo protein sequencing?**Ans: It is used for the protein that are not present in any data base, So its sequence is not known.  
Or if it is known then it may be modified by some mutation or by some chemical modification.