

Bioinformatics Quiz
Aug' 2023

Roll no.:

Total Marks:

Note: Pl answer in the space provided

1. What is the Difference between sequencing gap and physical gap. Which of the two was recently addressed resulting in the completion of the reference human genome? [3]

2 ✓ Physical gap is the actual gap in the genomic sequence, whereas sequencing gap is the gap we encounter while trying to obtain the genomic sequence of an organism. Recently, sequencing gaps were addressed, which meant that whole human genome has been sequenced.

2. Give a simple statistic to find the expression level of a gene.

[1]

Total Marks:

Note: Pl answer in the space provided

1. What is the Difference between sequencing gap and physical gap. Which of the two was recently addressed resulting in the completion of the reference human genome? [3]

sequencing gap refers to gaps caused by frame shift errors while physical gaps refer to ~~gaps~~ to the difference in location of the sequence.

Physical gap was recently addressed.

2. Give a simple statistic to find the expression level of a gene. [1]

The statistic is CAI, we measure the probability of each codon in a sample (p_k) & its frequency in the expressed genes (q_k) & the geometric mean of $\frac{p_k}{q_k}$ is a measure of the expression lvl of the sample.

3. Which method is better – hierarchical shotgun sequencing approach or whole genome shotgun sequencing approach for sequencing the genome of a (i) novel virus, (ii) higher eukaryotic organism? Give reasons to support your answer. [3]

Ans. ~~Both hierarchical and shotgun sequencing can work in case of a~~

i) Whole genome shotgun sequencing is a better approach for sequencing the genome of a novel virus because.

- We don't have any prior reference about the segments of a DNA.
- Also the no. of fragments synthesized are manageable in case of novel-virus as the genome is not too long.

ii) Hierarchical shotgun sequencing is a better approach in case of higher eukaryotic organism :-

- In case of whole genome sequencing, the no. of fragments are too large to be able to reconstruct the ~~the~~ entire genome perfectly.
- In case of Hierarchical as we break the genome into chunks and sequence those chunks, it's much more manageable.

4. Give one important application of analyzing the distribution of codons and the distribution of restriction recognition sites in a genome. [2]

Ans. Application of analyzing distribution of codons gives us

- Gives us a ^{hint}~~hint~~_{insite} to the distribution of protein the gene is containing. which _{can} helps identify the gene for a given protein.
- It can help in identification of no. of start codons \Rightarrow no. of genes, poly A-tails and promoters.

Analyzing distribution of RE sites in a genome.

- ~~Give~~ Provides insites into no. of fragments that will be obtained in an experiment, cleaving the DNA with that endonuclease.
- Knowledge of distribution of RE sites is used in cloning and selection of vectors, sequencing of sequence to be cloned.

5. Give the expression of affine scoring scheme for penalizing gaps in a pairwise sequence alignment. Why is it more sensitive than linear scoring scheme?

Ans.

Affine gap penalty = $-d - e(\varphi - 1)$, where $e, d \geq 0$
and φ is the no. of gaps.

It is more sensitive than linear scoring scheme because it captures the difference between a single gap and a gap ~~more~~ ^{being} ~~of~~ ~~more~~ than one, in terms of penalty difference.

Essentially affine penalty gap has two degrees of freedom compared to one degree of freedom of linear penalty gap.

6. The recognition site for Sau3A I is GATC and is contained in the recognition site of BamH I, GGATCC. Which of the following is true? (Tick the correct option). [1]
- (i) The two REs give the same number of fragments.
 - ☒ (ii) SauA I will give larger number of fragments.
 - (iii) BamH I will give larger number of fragments.
 - (iv) Cannot say, will depend on the DNA sequence.
7. Cloning can be used for ----- of an unknown DNA sequence. [1]
- (i) storing
 - (ii) amplification
 - ☒ (iii) expression
 - (iv) all three
 - ☒ (v) storing and amplification
8. Is clone-based sequencing approach necessary for re-sequencing a genome? [1]
- (i) YES
 - ☒ (ii) NO
9. Average number of genetic differences between individuals gives an indication of population size. TRUE/FALSE. Give reasons to support answer. [3]

Ans.

~~TRUE~~
~~FALSE~~, the.

- Larger average genetic difference means the average of no. of genes ^{that} vary ^{is} ~~larger~~ ^{completely larger} b/t two individuals.

Mathematically: Let X_1, X_2, \dots, X_n be ^{all} the genes ~~that vary~~.

$$X_i = \begin{cases} 0, & \text{if gene } X_i \text{ doesn't vary} \\ 1, & \text{if gene } X_i \text{ varies} \end{cases}$$

We know $X_1 + X_2 + \dots + X_n = m$ $\&$ = average no. of genes varying.

\Rightarrow We have nC_m possible sequences of genes.

Since nC_m is an increasing function with increasing m .

\therefore Larger no. of individuals are present, hence larger ^{no.} of mutations and reproduction. Therefore, overall larger population.

10. Give an application for using 'distance-based' scoring function and 'similarity-based' scoring function for pairwise alignment? [2]

Distance-based scoring function is used for phylogenetic analysis of multiple species.

✓ Similarity-based scoring function is used for identifying the closest species to an unknown organism.

11. Importance of using mitochondrial DNA in species identification. [2]

✓ Mitochondrial DNA has least number of mutations across the same species.