

المستوي الثالث / برنامج البيولوجيا الجزيئية
مقرر: المعلوماتية الحيوية (307 ح)

Date: 13-06-2024

Time: 2 Hours

Marks: 90

الإمتحان في 3 صفحات:

Answer ALL the following questions:

QUESTION (1):

(20 Marks)

(A) Which of the following statements are probably true (✓) and which are probably false (×).

DO NOT copy the sentences to your answer sheet

- 1- Compared to command-line tools, web-based tools are more appropriate for analyzing large-scale datasets.
- 2- Each sequence read in SRA is typically 800 base pairs.
- 3- Every two homologous molecules must share statistically significant amino acid (or nucleotide) identity.
- 4- It is estimated that humans and rodents diverged about 90 thousand years ago.
- 5- Orthologs and paralogs must have the same function.
- 6- The percent homology of two protein sequences is the sum of both identical and similar matches.
- 7- There is more than a RefSeq entry for each splice variant.
- 8- WGS sequences have been available since 2002 and they are considered part of the GenBank/EMBLBank/DDBJ releases.
- 9- A GenBank accession number refers to the most recent version of a given sequence.
- 10- Each sequence read in SRA is typically 50–400 base pairs.

(B) Choose the best answer:

Only the letter, DO NOT copy the choices to your answer sheet

- 1- Which one of the following does not have the proper format of an accession number?
(a) rs41341344; (b) J03093; (c) 1PBO; (d) NT_030059; or
(e) all of these have proper formats.
- 2- Which of the following databases is derived from mRNA information?
(a) PBD; (b) OMIM; or (c) HTGS. (d) dbEST;
- 3- Which of the following databases can be used to access text information about human diseases?
(a) EST (b) PBD (c) HTGS (d) OMIM
- 4- What is the difference between RefSeq and GenBank?
(a) RefSeq includes publicly available DNA sequences submitted from individual laboratories and sequencing projects.
(b) GenBank provides nonredundant curated data.
(c) GenBank sequences are derived from RefSeq.
(d) RefSeq sequences are derived from GenBank and provide nonredundant curated data.
- 5- Orthologs are defined as:

- (a) Homologous sequences in the same species which have similar and often redundant functions.
- (b) Homologous sequences that share little amino acid identity but share great structural similarity.
- (c) Homologous sequences in the same species that arose through gene duplication.
- (d) Homologous sequences in different species that share an ancestral gene
- 6-In a database search or in a pairwise alignment, sensitivity is defined as:
- (a) The ability of a search algorithm to find true positives (i.e., homologous sequences) and to avoid false positives (i.e., unrelated sequences having high similarity scores).
- (b) The ability of a search algorithm to find true positives (i.e., homologous sequences) and to avoid false positives (i.e., homologous sequences that are not reported).
- (c) The ability of a search algorithm to find true positives (i.e., homologous sequences) and to avoid false negatives (i.e., unrelated sequences having high similarity scores).
- (d) The ability of a search algorithm to find true positives (i.e., homologous sequences) and to avoid false negatives (i.e., homologous sequences that are not reported).
- 7-You have a reasonably short, typical, double-stranded DNA sequence. Basically, how many proteins can it potentially encode?
- (a) 1 (b) 2 (c) 3 (d) 6
- 8-Which of the following amino acids is Asparagine?
- (a) Q (b) W (c) Y (d) N
- 9-Which of the following amino acids is Phenylalanine?
- (a) P (b) H (c) F (d) L
- 10-If you want literature information, what is the best website to visit?
- (a) OMIM (b) Entrez (c) PubMed (d) PROSITE

QUESTION (2):

(20 Marks)

Complete the following with suitable words:

DO NOT copy the sentences to your answer sheet

- _____(1)_____ was the first protein to have its structure resolved by X-ray crystallography.
- A _____(2)_____ refers to an assembly in which DNA sequence is collected and arranged to reflect the sequence of each chromosome.
- _____(3)_____ are homologous sequences that arose by a mechanism such as gene duplication.
- The _____(4)_____ browser automatically analyzes and annotates genome data and present genomic data via its web browser.
- Statistical significance and _____(5)_____ analyses are performed to assess the hypothesis that 2 proteins are homologous.
- While homology is a qualitative inference that reflect whether sequences are homologous or not, identity and similarity are _____(6)_____ inferences.
- _____(7)_____ sequences are defined as the sequences that contain stretches of amino acids (or nucleotides) with limited information content.
- Each sequence read in SRA is relatively short; typically _____(8)_____ base pairs.
- NCBI assigns unique sequence identification numbers that apply to the individual sequences within a record; these numbers are called _____(9)_____ numbers.
- There are many millions of species alive today, and they can be grouped into the 3 major branches: bacteria, _____(10)_____, and eukaryotes.

- Homology is often manifested by significant similarity in nucleotide or amino acid sequence and almost always manifested in _____(11)_____.
- β globin gene is located on chromosome 11, while α globin gene is located on chromosome _____(12)_____.
- _____(13)_____ are large segments of DNA that are used to clone large amount of DNA into yeast.
- _____(14)_____ is authoritative and the most comprehensive, centralized protein sequence database.
- _____(15)_____ is a main EBI-situated worldwide repository for gene expression data.
- _____(16)_____ is a string of about 4-12 numbers and/or alphabetic characters that are associated with a molecular sequence record.
- A _____(17)_____ pairwise alignment includes all residues of both sequences.
- In a typical scoring scheme there are 2 gap penalties; _____(18)_____ and _____(19)_____.
- _____(20)_____ is a physical chemist who dedicated her career to applying the computational technologies to support advances in biology and medicine, most notably the creation of protein and nucleic acid databases.

QUESTION (3):

(25 Marks)

Answer the following:

- (A) Define "BLAST", list its types, and summarize the general steps of performing BLAST. (15 marks)
- (B) Define multiple sequence alignment and list its typical uses. (10 marks)

QUESTION (4):

(25 Marks)

Answer the following:

- (A) What are the consequences of a faulty alignment? (5 marks)
- (B) What are the uses of BLAST? (5 marks)
- (C) Define "GSSs", "HTGS", and "EST". (6 marks)
- (D) What are the prominent resources (databases) in NCBI? (9 marks)

Best Wishes

Examiner: Prof. Ahmed Ghoneim