PBJ-1603220001030100 Seat No. _____

B. Sc. Bioinformatics (Sem. III) Examination
November / December - 2018
BI - 301 : Bioinformatics Databases & Sequence Analysis
(New Course)

Time : $2\frac{1}{2}$ Hours]   [Total Marks : **70**

Instructions :   (1)   All questions are compulsory.
                       (2)   The right-side figure indicates total
                             marks of the question.

1   Attempt the following :   **14**
  (a)   Answer the following short questions : (all compulsory)   **4**
        (1)   What is data warehousing ?
        (2)   An advantage of the _____ is that it permits more
              complex, hierarchical (tree-like) structures to be
              represented.
        (3)   What are the two challenges of data integration ?
        (4)   What is data discovery ?

  (b)   Answer any **one** of the following questions.   **2**
        (1)   What are the system requirements for data
              management ?
        (2)   What are Data, Metadata and Application Flow ?

  (c)   Answer any **one** of the following questions :   **3**
        (1)   Explain data standardization.
        (2)   What is data warehousing and its requirements ?

  (d)   Answer any **one** of the following questions.   **5**
        (1)   Explain challenges of big data.
        (2)   What is machine learning and its role
              in biology ?

2 Attempt the following : 14

   (a) Answer the following short questions : 4

      (all compulsory)

      (1) Name some RNA sequence database.

      (2) Define transcription factor.

      (3) Name some databases of protein family, domains and active sites.

      (4) _____ database integrates manually curated hidden Markov models for many domains with a powerful web-based interface and visualizing tools.

   (b) Answer any **one** of the following questions : 2

      (1) Explain about CyBase database.

      (2) Explain Carbohydrate Structure Database.

   (c) Answer any **one** of the following questions : 3

      (1) Explain JUNC DB.

      (2) Give a brief note on sequence motifs and active sites database.

   (d) Answer any **one** of the following questions : 5

      (1) Explain in detail PubChem.

      (2) Explain in detail protein sequence database.

3 Attempt the following : 14

   (a) Answer the following short questions : 4

      (all compulsory)

      (1) _____ is a field of biological research in which the genomic features of different organisms are compared.

      (2) List of protein-protein interactions databases.

      (3) IVDB hosts complete genome sequences of influenza B virus. (True/False)

      (4) _____ is a comprehensive resource database for Ecoli K-12.

(b) Answer any **one** of the following questions :   **2**
   (1) What is protein interaction and its importance ?
   (2) Explain any human disease database.

(c) Answer any **one** of the following questions :   **3**
   (1) Explain HMDB.
   (2) Explain any prokaryotic genome database.

(d) Answer any **one** of the following questions :   **5**
   (1) Explain a database of a Genome annotation, ontologies and nomenclature.
   (2) Write detailed note on UCSC genome browser.

4 Attempt the following :   **14**
(a) Answer the following short questions :   **4**
   (all compulsory)
   (1) Name some immunological databases.
   (2) What are techniques used to analyse gene expression ?
   (3) Expand BARC.
   (4) Name some proteomic data resources.

(b) Answer any **one** of the following questions :   **2**
   (1) What is TAIR ?
   (2) What is PubMed ?

(c) Answer any **one** of the following questions :   **3**
   (1) dbPTM. Explain.
   (2) Mitochondrial gene and protein database. Explain.

(d) Answer any **one** of the following questions :   **5**
   (1) Explain Microarray database.
   (2) Explain DrugBank.

5   Attempt the following :   14

    (a)   Answer the following short questions :   4

       (all compulsory)

       (1)   Scoring a pairwise alignment requires a _____ and _____.

       (2)   The Blosum62 matrix is the best for detecting the majority of high protein similarities (true/false)

       (3)   PHI and PSI Blast.

       (4)   _____ is more suitable for studying quite distant proteins, ____ is for more conserved proteins of domains.

    (b)   Answer any **one** of the following questions :   2

       (1)   What is multiple sequence alignment ?

       (2)   What is log odd score ?

    (c)   Answer any **one** of the following questions :   3

       (1)   Explain FASTA.

       (2)   What is sequence alignment and its importance ?

    (d)   Answer any **one** of the following questions :   5

       (1)   Explain PAM and BLOSUM.

       (2)   Explain Database searching for sequence similarity.

_____