

SSR PG COLLEGE, NIZAMABAD

M,SC BIOTECHNOLOGY

Semester – III

PAPER – II, INTERNAL - I

BIOINFORMATICS

Credit I: Foundations to Bioinformatics

1. The primary goal of bioinformatics is to:

- a) Synthesize new biological molecules in the lab
- b) Analyze and interpret biological data using computational tools
- c) Design wet-lab experiments exclusively
- d) Replace traditional biological research

Answer: b) Analyze and interpret biological data using computational tools

2. Which of the following is NOT considered a core type of bioinformatics data?

- a) Nucleic acid sequence
- b) Protein structure
- c) Laboratory glassware inventory
- d) Metabolomic information

Answer: c) Laboratory glassware inventory

3. A database like GenBank is a primary example of a:

- a) Protein structure database
- b) Metabolic pathway database
- c) Nucleic acid sequence database
- d) Proteomic database

Answer: c) Nucleic acid sequence database

4. The Protein Data Bank (PDB) is a central repository for:

- a) DNA sequences
- b) 3D structural data of proteins and nucleic acids
- c) Gene expression profiles
- d) Metabolic compounds

Answer: b) 3D structural data of proteins and nucleic acids

5. Which file format is commonly used for storing multiple sequence alignments?

- a) PDB
- b) FASTA
- c) JPEG
- d) DOCX

Answer: b) FASTA

6. The National Center for Biotechnology Information (NCBI) is a major bioinformatics portal. It is maintained by:

- a) The European Molecular Biology Laboratory (EMBL)
- b) The US National Institutes of Health (NIH)
- c) The DNA Data Bank of Japan (DDBJ)

d) The Swiss Institute of Bioinformatics (SIB)

Answer: b) The US National Institutes of Health (NIH)

7. A tool like BLAST, which is freely accessible online, is best categorized as:

- a) A downloadable free tool
- b) A commercial software package
- c) A free online tool
- d) A bioinformatics journal

Answer: c) A free online tool

8. Which of the following is a key role of the internet in bioinformatics?

- a) Providing physical laboratory space
- b) Enabling global access to databases and computational tools
- c) Manufacturing laboratory equipment
- d) Conducting automated experiments

Answer: b) Enabling global access to databases and computational tools

9. UniProt is a comprehensive resource for:

- a) Genomic data
- b) Protein sequence and functional information
- c) Metabolomic profiles
- d) DNA barcoding

Answer: b) Protein sequence and functional information

10. The European counterpart to the NCBI's GenBank is:

- a) PDB
- b) UniProt
- c) EMBL-Bank (maintained by EBI)
- d) DDBJ

Answer: c) EMBL-Bank (maintained by EBI)

11. Which resource would you use to find a published 3D structure of a protein?

- a) GenBank
- b) Protein Data Bank (PDB)
- c) KEGG
- d) PubMed

Answer: b) Protein Data Bank (PDB)

12. The field of bioinformatics emerged significantly due to the advent of:

- a) The microscope
- b) Large-scale DNA sequencing projects
- c) The discovery of penicillin
- d) Centrifugation techniques

Answer: b) Large-scale DNA sequencing projects

13. KEGG and MetaCyc are examples of databases that primarily contain information on:

- a) Protein sequences
- b) Metabolic pathways
- c) Gene expression data
- d) DNA markers

Answer: b) Metabolic pathways

14. Which of the following is a downloadable, free bioinformatics tool for sequence analysis?

- a) NCBI BLAST web interface
- b) EMBOSS
- c) Microsoft Excel
- d) Adobe Acrobat

Answer: b) EMBOSS

15. The design of a bioinformatics database is crucial for:

- a) The aesthetic appearance of the website
- b) Efficient data storage, retrieval, and integration
- c) Increasing the cost of access
- d) Limiting user access

Answer: b) Efficient data storage, retrieval, and integration

16. A "web portal" in bioinformatics, like the NCBI, serves as a:

- a) Single point of access to a wide variety of resources and tools
- b) Firewall to block data access
- c) Social media platform for scientists
- d) Online marketplace for laboratory reagents

Answer: a) Single point of access to a wide variety of resources and tools

17. Metabolomic data in bioinformatics refers to information about:

- a) The complete set of DNA in an organism
- b) The complete set of proteins in an organism
- c) The complete set of small-molecule metabolites in a biological system
- d) The complete set of RNA transcripts

Answer: c) The complete set of small-molecule metabolites in a biological system

18. Which of these is a primary literature resource accessed via bioinformatics portals?

- a) GenBank
- b) PDB
- c) PubMed
- d) BLAST

Answer: c) PubMed

19. The FASTA format is characterized by:

- a) A header line starting with a ">" followed by sequence data
- b) Complex binary code for efficient storage
- c) A detailed 3D coordinate list
- d) A spreadsheet with multiple columns

Answer: a) A header line starting with a ">" followed by sequence data

20. The historical development of bioinformatics was heavily influenced by the need to manage data from:

- a) The invention of the telescope
- b) The Human Genome Project
- c) Classical Mendelian genetics
- d) The development of the pH meter

Answer: b) The Human Genome Project

Credit II: Comparison Methods in Bioinformatics

1. In sequence alignment, two sequences that share a common evolutionary origin are said to be:

- a) Analogous
- b) Homologous
- c) Identical
- d) Paralogous

Answer: b) Homologous

2. A gap in a sequence alignment is introduced to account for:

- a) Point mutations
- b) Conserved regions
- c) Insertions or deletions (indels) during evolution
- d) Sequencing errors only

Answer: c) Insertions or deletions (indels) during evolution

3. Genes in different species that evolved from a common ancestral gene by speciation are called:

- a) Paralogs
- b) Orthologs
- c) Xenologs
- d) Analogs

Answer: b) Orthologs

4. Which scoring matrix is derived from closely related sequences and is used for building phylogenetic trees?

- a) BLOSUM 45
- b) BLOSUM 80
- c) PAM 1
- d) PAM 250

Answer: c) PAM 1

5. The BLOSUM62 matrix is most commonly used for:

- a) Aligning very distantly related sequences
- b) Detecting sequencing errors
- c) Standard protein sequence alignment
- d) DNA sequence alignment

Answer: c) Standard protein sequence alignment

6. The Needleman-Wunsch algorithm is designed to find the:

- a) Local similarity between two sequences
- b) Best global alignment between two sequences
- c) Best alignment for multiple sequences
- d) Conserved domains in a single sequence

Answer: b) Best global alignment between two sequences

7. The Smith-Waterman algorithm is used for:

- a) Global alignment
- b) Local alignment
- c) Multiple sequence alignment
- d) Database searching only

Answer: b) Local alignment

8. The BLAST algorithm is a heuristic method, meaning it:

- a) Always finds the mathematically optimal alignment
- b) Uses a slow but exhaustive search
- c) Uses shortcuts to rapidly find significant matches, sacrificing guaranteed optimality
- d) Is only used for DNA sequences

Answer: c) Uses shortcuts to rapidly find significant matches, sacrificing guaranteed optimality

9. Which program would you use to quickly find similar nucleotide sequences in a database?

- a) BLASTP
- b) BLASTN
- c) BLASTX
- d) TBLASTN

Answer: b) BLASTN

10. The FASTA algorithm often proceeds by first looking for:

- a) Gapped alignments
- b) Short exact matches (ktups) to identify regions of potential similarity
- c) Global alignments using dynamic programming
- d) Structural similarities

Answer: b) Short exact matches (ktups) to identify regions of potential similarity

11. The main goal of Multiple Sequence Alignment (MSA) is to:

- a) Identify the longest sequence in a set
- b) Infer evolutionary relationships and identify conserved regions
- c) Calculate the molecular weight of proteins
- d) Find the best pairwise alignment

Answer: b) Infer evolutionary relationships and identify conserved regions

12. The ClustalW tool is a classic example of which MSA algorithm type?

- a) Iterative
- b) Progressive
- c) Stochastic
- d) Structural-based

Answer: b) Progressive

13. A Position-Specific Scoring Matrix (PSSM) is generated from a multiple sequence alignment and is used to:

- a) Store the raw sequence data
- b) Describe the conservation of amino acids at each position in a protein family
- c) Calculate the pI of a protein
- d) Perform global pairwise alignment

Answer: b) Describe the conservation of amino acids at each position in a protein family

14. Which BLAST variant uses a PSSM to search a database?

- a) BLASTN
- b) BLASTP
- c) PSI-BLAST
- d) PHI-BLAST

Answer: c) PSI-BLAST

15. A consensus sequence derived from an MSA represents:

- a) The longest sequence in the alignment
- b) The most frequent amino acid or base at each position
- c) A randomly chosen sequence
- d) The sequence with the fewest gaps

Answer: b) The most frequent amino acid or base at each position

16. A gap penalty is used in alignment algorithms to:

- a) Reward the insertion of gaps
- b) Penalize the creation of gaps, reflecting their lower evolutionary likelihood compared to substitutions
- c) Ignore gaps in the final score
- d) Increase the speed of the algorithm

Answer: b) Penalize the creation of gaps, reflecting their lower evolutionary likelihood compared to substitutions

17. Protein sequences are often preferred over DNA sequences for distant homology detection because:

- a) Proteins are easier to sequence
- b) The genetic code is degenerate, and protein sequences conserve functional information better over long evolutionary distances
- c) DNA sequences are longer
- d) Protein alignment algorithms are faster

Answer: b) The genetic code is degenerate, and protein sequences conserve functional information better over long evolutionary distances

18. In a dot-matrix plot for sequence comparison, a diagonal line indicates:

- a) A region of mismatch
- b) A region of insertion
- c) A region of similarity between the two sequences
- d) The end of the sequence

Answer: c) A region of similarity between the two sequences

19. Genes that arise from a gene duplication event within a genome are called:

- a) Orthologs
- b) Xenologs
- c) Paralogs
- d) Homologs

Answer: c) Paralogs

20. The E-value in a BLAST report represents:

- a) The exact score of the alignment
- b) The percentage identity between the query and hit sequence
- c) The number of expected random hits with a score equal to or better than the observed score
- d) The evolutionary distance between the sequences

Answer: c) The number of expected random hits with a score equal to or better than the observed score

SHORT ANSWER QUESTIONS

1. Define EXPASY
2. What is Homology
3. Define Gap penalty
4. Mention any two Bioinformatics web portals
5. Define Match and Mismatch