19BT344 BIOPERL

Hours Per Week :

L	T	P	C
3	-	2	4

Total Hours :

L	T	P
45	-	30

WA/RA	SSH/HSH	CS	SA	S	BS
15	45	30	8	2	2

COURSE DESCRIPTION AND OBJECTIVES:

The course aims to familiarize with the basic programming concepts in Perl with biology and also shell scripting and programming with Unix which is used for file processing. The overall objective of the course is to apply the Perl for solving the biological problems and also to handle and store the DNA and protein sequences.

COURSE OUTCOMES:

Upon completion of the course, the student will be able to achieve the following outcomes:

COs	Course Outcomes	POs
1	Familiarize with the basic programming concepts, usage of tools and utilities of UNIX.	1,2
2	Apply Perl concepts to handle file formats and stores the DNA and protein sequences.	1,2, 3,5
3	Develop Perl programs using procedures and functions to solve biological problems.	2,3
4	Create Perl programs for research project purpose to analyze the biological data.	1,5,12

SKILLS:

- ✓ *Find applications of scripting in teaching.*
- ✓ *Develop scripts for solving simple biology problems.*
- ✓ *Design scripts that can handle, store the DNA and protein sequences.*
- ✓ *Develop scripts to handle sequences in FASTA format.*

UNIT - I   **L-9**

BIOLOGY AND COMPUTER SCIENCE: The organization of gene structure, DNA and protein sequences, Getting started with perl - a low and long learning curve, Perl's benefits, Installing perl on your computer, How to run Perl programs in command line, Types of text editors and finding help from user manuals.

UNIT - II   **L-9**

THE ART OF PROGRAMMING SEQUENCES AND STRINGS: The art of programming - individual approaches to programming, Edit-Run-Revise (and Save), An environment of programs, Programming strategies, The programming process, Sequences, Types of operators, Variables, Representing sequence data, A program to store a DNA sequence, Concatenating DNA fragments transcription - DNA to RNA, Using the Perl documentation, Calculating the reverse complement in Perl, Files and arrays, Reading proteins from files, Scalar variables, List context.

UNIT - III   **L-9**

MOTIFS, LOOPS, SUBROUTINES AND BUGS: Motifs and loops - flow control, code layout, finding Motifs, counting nucleotides, exploding strings into arrays, operating on strings writing to files.
Subroutines and Bugs: Subroutines, Scoping and subroutines, Command - line arguments and arrays, passing data to subroutines, modules and libraries of subroutines, fixing bugs in your code.

UNIT - IV   **L-9**

MUTATIONS AND RANDOMIZATION: Random number generators, A program using randomization, A program to simulate DNA mutation, Generating random DNA, Analyzing DNA.

UNIT - V   **L-9**

THE GENETIC CODE: Hashes, Data structures and algorithms for Biology, The genetic code, Translating DNA into proteins, Reading DNA from files in FASTA format, Reading frames.

LABORATORY EXPERIMENTS

LIST OF EXPERIMENTS   **TOTAL HOURS: 30**

1.   Chop, chomp based simple Perl program.
2.   Program based on control structures - dowhile, foreach and with control flow statements- redo, next, goto, etc.
3.   Subroutines.
4.   Writing a sequence to a file.
5.   Retrieving sequence file and searching for a pattern.
6.   Obtaining basic sequence statistics.
7.   Comparing files. Combining and extracting data from different files using modules.
8.   MSA using Perl and conserved domain identification and hast table creation.
9.   Blast using Bioperl.
10. CGI- Perl Programs for developing MSA.

TEXT BOOKS:

1.   James Tisdall, "Beginning perl for bioinformatics", O'REILLY, ISBN: 0-596-00080-4.
2.   S.Sai Giridhar and S.Krupanidhi "Introductory Workbook on Perl for Biology Students", Published by Biology-Online.org, 2009.
3.   Martin C. Brown, "The complete reference of Perl", Osborne, McGraw-Hill Publishers, ISBN:978-0072121421.