GUJARAT TECHNOLOGY UNIVERSITY

SCHOOL OF APPLIED SCIENCES & TECHNOLOGY

Post Graduate Diploma
in
Bioinformatics
(1 Year Full Time PG Diploma)

Effective from A.Y. 2021

Curriculum Structure

S. No.	**Title**	**Credit**
Core	**SEMESTER- I**	
1	Introduction to Bioinformatics	4
2	Biomathematics & Biostatistics	4
3	Cloud Computing, Data Mining & Visualization	4
4	Linux and Python for Bioinformatics	4
5	Laboratory I: Bioinformatics	4
6	Laboratory II: Biomathematics & Biostatistics	4
7	Laboratory III: Linux & Python	4
8	Laboratory IV: Cloud Computing, data mining	4
9	Laboratory V: Elective 1/2/3	9
	Electives **(Choose any 3 for total 9 credits)**	
1	Introduction to R and Bioconductor	3
2	Database Management System and web programming	3
3	System Biology: Modelling biological networks	3
4	Computer Aided Drug Designing	3
5	Bioethics & IPR	3
	Total	50
	SEMESTER- II	
1	Project in Collaboration with Industry	30
2	Presentation of Project Completion	20
	Total	**100**

Program Overview

Bioinformatics as a field attempts to build computational models of the biological systems and mechanisms. More specifically, bioinformatics involves creating algorithms, databases, systems, and web applications to solve problems in molecular biology. Due to the drop in sequencing costs we are awash in DNA, RNA, and protein sequences. Massive genomics and metagenomics efforts are opening new horizons in variation analysis. The past few years of structural genomics efforts have produced a crystal structure representative of almost every protein family. Microarray technologies allow simultaneous studies of expression of thousands of genes on a single chip. The improvements keep on coming – more information, higher resolution.

The PG diploma will provide expert training in bioinformatics to subject students. Students upon completion of the diploma can be accommodated in any national/multinational drug designing pharmaceutical/biotechnology/software industry. This course will provide common basic knowledge of Bioinformatics including Bioinformatics, Biomathematics, Biostatistics, and Computer Science.

Highlights

Course Level	**Post- Graduation**
Full- Form	Post –Graduation Diploma in Bioinformatics
Duration	1 year
Examination Type	Semester
Eligibility	Bachelor’s degree in Science/Engineering/ Pharmacy/Agriculture/Dairy/Veterinary or any equivalent qualification with a minimum 50 % (45 % for SC/ ST/ SEBC/ EWSs candidates) at the qualifying examination.
Admission Process	Merit based
Course Fee	INR 25000 (1 Year)
Top Recruiting Sectors	Bioinformatics, Biotechnology, Biomedical Sciences, Pharmaceutical, Agricultural Sectors, Healthcare, IPR, Knowledge base etc.
Job Position	Research Associate, Bioinformatics Trainer, Software Engineer Trainee, Program Developer, Business Development executive, Data Analyst etc.
Student Intake	15

SYLLABUS

SEMESTER 1

Introduction to Bioinformatics

Unit 1: Biology, Biological Data & Bioinformatics

- Introduction to Bioinformatics, history, evolution and its applications in Biological sciences.
- Genome organization, Central Dogma of cell biology, Genome rearrangement, types of variations, Basic concept of Linkage and association.
- Biological Data Types (gene, protein, mRNA, whole genome, exom, gene expression, Metabolic Pathways, network, metabolites).
- Introduction to biological data producing technologies (Sequencing Sanger, NGS. Microarray and its types like NAPPA, Real Time-PCR, Exom sequencing, mRNA expression, Mass spectrometry, Restriction Fragment Length Polymorphism, (RFLP), single nucleotide polymorphism(SNP).)
- Categories of biological databases, primary, secondary, tertiary, nucleotide databases, Protein Databases, Specialized Databases, Database search and retrieval system.

Unit 2: Bioinformatics Method: Sequence analysis, Phylogeny and Evolution

- Introduction to sequence analysis, Local and Global Sequence alignment, Pairwise & Multiple Sequence Alignment using CLUSTALW
- Genome Mapping, Types of maps: Cytogenetic, Transcript map, Comparative map, integrated map. Completing Maps and Sequences, Annotation, Enrichment Analysis, Data Visualization clinical feature association with expression profiles patient stratification by existing advanced computational tools integrating multi-omics data.

- Molecular phylogenetics, phylogenetic trees, trees and distances. Measuring genetic change, genetic distance. Methods of phylogeny reconstruction Distance matrix methods, Maximum parsimony methods, Maximum likelihood methods,
- Evolution, Models of Molecular evolution, Functional constraints and the rate of substitution patterns of codon usage and base composition. Evolutionary clocks, Neutral theory, Genetic variation within species, Natural selection
- Introduction to Study design, workflow development, management and execution in OMICS studies including comparative Genomics, Proteomics, Metagenomics, Transcriptomics, Molbiome studies.

Unit 3: Bioinformatics Algorithms: Techniques & Applications

- General algorithmic techniques, including dynamic programming, graph-theoretical methods, hidden Markov models, the fast Fourier transform, seeding, and approximation algorithms, word distributions and occurrences, Scoring Matrices
- Algorithms and tools for genome and sequence analysis, including formal and approximate models for gene clusters, advanced algorithms for non-overlapping local alignments and genome tiling, multiplex PCR primer set selection, and sequence/network motif finding.
- Probabilistic models for analysis of gene sequences and predictors. Microarray design and analysis, including algorithms for microarray physical design, missing value imputation, and meta-analysis of gene expression data
- Algorithmic issues arising in the analysis of genetic variation across human population, including computational inference of haplotypes from genotype data and disease association search in case/control epidemiologic studies

- Algorithmic approaches in structural and systems biology, including topological and structural classification in biochemistry, and prediction of protein-protein and domain-domain interactions.

Unit 4: Structural Bioinformatics

- Understanding structural basis for biological phenomena– challenges in structural bioinformatics – integration of structural data with other data.
- Conformational Analysis of proteins– Forces that determine protein structure – polypeptide chain geometries – Ramachandran Map – potential energy calculations.
- Conformational analysis of nucleic acids and carbohydrates – general characteristics of nucleic acid structure – geometries, glycosidic bond – rotational isomers and ribose puckering - forces stabilizing ordered forms – base pairing – base stacking.
- Structure Prediction Methods – Homology Modelling – Fold Recognition Methods – ab initio methods, Molecular Docking
- MOLECULAR INTERACTIONS Interactomes – Macromolecular interactions – Protein-DNA interactions – Protein-Ligand interactions– Interactions database – ProNIT - Docking – principles and methods

REFERENCES:

- Baxevanis A.D., Davison D.B., Page R. D. M. & Petsko G.A. Current Protocols in Bioinformatics. New York, John Wiley & Sons Inc., 2004. ISBN: 0555015254.
- V.R.Srinivas “Bioinformatics: Sequences and genomics analysis” cold Spring Harbor Laboratory press,Cold Spring harbor,USA, 2005.
- N.Gautham, “Bioinformatocs” Narosa Publishing Company, New Delhi, 2006.
- D.Higgins and W.Taylor, ’Bioinformatics: Sequences, Structures and databanks’oxford University Press, Oxford, UK, 2008.
- David W. Mount, “Bioinformatics Sequence and Genome Analysis”, Cold Spring Harbor Laboratory Press.
- Bioinformatics Algorithms: Techniques and Applications Ion Mandau, Alexander Zelikovsky.
- C.R.Cantor&P.R.Schimmel, Biophysical Chemistry Part - I, W.H. Freeman & Co., in San Fransisco, 1980.
- C. Branden and J. Tooze, Introduction to Protein Structure, 2nd Edition, Garland Publishing Inc., New York, 1999.
- P.E. Bourne and H. Weissig (Eds.) Structural Bioinformatics, John-Wiley and Sons,2003.
- David W. Mount, Bioinformatics – Sequence and Genome analysis, Cold Spring Harbor Laboratory Press, New York, 2001. 5. J.Pevsner, Bioinformatics and Functional Genomics, John-Wiley and Sons, 2009.

Biomathematics & Biostatistics

Unit 1: Calculus

- Introduction to biomathematics & its application in bioinformatics,
- Graphs and functions: Linear, Quadric, Exponential, Periodic function, combination of simple function with example from biological questions, Logarithmic function Slope of curves, Idea of derivative.
- Calculation of derivatives: Derivatives of simple functions, Derivative of exponential function, Derivative of sum of two functions,
- Differentiation and its application in Biology: Product rule in differentiation, Derivatives of Sine and Cosine functions, Plotting derivatives, Differential calculus to understand actin polymerization, Enthalpy and Entropy of a chemical reaction, Growth curve, Idea of curvature, Curvature, Free energy, Energy of spring-like protein, Maxima and Minima of a function, Force and energy, DNA unzipping, Plotting mathematical functions,
- Integration: definite integrals, integration of simple functions, Integral as "anti-derivative", Definite integrals, Integral as area under a curve, Integration by parts, finding derivative and integral given a set of data points.

Unit2: Differential Equations

- Simple differential equations, first order differential equations, Examples: Polymerizing and depolymerizing filaments, Cell growth: Concentration gradient, Second order differential equations. Motion of an object under external force : Newton's equations

Unit3: Vectors

- Physical quantities like position and force as vectors, Attracting and repelling charges, Vector addition, Calculation of forces in a system of charges, Calculation of magnitude and direction of a vector, Unit vectors, Calculation of resultant force, Dot product and cross product, Polar coordinate system, Gradient of a scalar.

Unit4: Biostatistics

- Mean & Variance: Introduction to statistics, Mean/Average, Variance, Standard deviation, Median, Mode. Hypothesis testing, Non pragmatics, paramatics test
- Distribution function: Introduction to distribution functions, Normal distribution, Examples from biology: End to-end vector distribution of DNA, Concentration distribution, Confidence Interval.
- Normal Distribution: Gaussian function, Peak as average of normal distribution, Width of a Gaussian and standard deviation.
- Fitting a function to experimental data: Linear fit, Least-square fit- errors
- Uniform and Poisson distributions; Knudson's analysis Uniform distribution, Poisson distribution, Knudson's analysis of retinoblastoma patients, Poisson statistics and tumor.
- Statistical tests: t test, f test, ANOVA, Chi Square test, Regression Analysis

Unit5: Statistical thermodynamics of biological systems

- Temperature, Energy and Entropy: Definition of temperature, Definition of internal energy, Definition of entropy, Calculation of entropy, entropy of a flexible protein.
- Partition function, Free energy: Definitions, Calculation of partition function, Calculation of Free energy, Thermal Equilibrium, Bending of DNA

- Bending fluctuations of DNA and spring-like proteins: Worm-like chain model, Partition function, Gibbs free energy
- Force-extension and looping of DNA: Force extension relation of single stranded DNA, Persistence length, Looping of DNA
- Thermodynamics of protein organization along DNA: Proteins binding on DNA, Calculation of energy, entropy and free energy.

REFERENCES:

- Matthew He, Sergey Petoukhov, “Mathematics of Bioinformatics: Theory, Practice, and Applications”, Wiley Series of Bioinformatics, 2010.
- Jae K. Lee, “Statistical Bioinformatics: A Guide for Life and Biomedical Science Researchers”, Wiley-Blackwell.

Cloud Computing, Data Science & Visualization

Unit 1: Cloud Computing

- Data collection - data wrangling, cleaning, and sampling to get a suitable data set;
- Data management - accessing data quickly and reliably;
- Exploratory data analysis – generating hypotheses and building intuition; distributed problem solving.
- Prediction or statistical learning; and
- Communication – summarizing results through visualization, stories, and interpretable summaries.
- Bioinformatics workflow/pipeline generation using Snakemake, Workflow automation, version control github, and unit testing

UNIT 2: Data Mining & Visualization

- Data Mining – Data ware housing-OLAP-Data Pre-processing, Classification Techniques: Decision Tree, ID3,K-Nearest Neighbour Classifier
- Data mining concepts and tasks. Forms of knowledge representation. Pre-processing and data transformation. Algorithmic methods: classification and association rules, decision trees, grouping, K Nearest Neighbour Classifier, Naive Bayes- near Neighbour Search and attribute selection. Study and applications with biological data.
- Study of AI and machine learning algorithms in computational biology, combining theory with practice. Techniques and advances in the field of computational biology, analysing sets of biological data generated by high-throughput technologies.

UNIT 3: Data Visualization

- Overview of data visualization - Data Abstraction -Analysis: Four Levels for Validation- Task Abstraction - Analysis: Four Levels for Validation

REFERENCES:

- Taylor, R.C.: An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics. BMC Bioinf. **11**(Suppl 12), S1 (2010).
- Statistical Modelling and Machine Learning Principles for Bioinformatics Techniques, Tools, and Applications (Algorithms for Intelligent Systems) 1st ed. 2020 Edition

Linux & Python for Bioinformatics

UNIT 1: Linux

Linux file system layout, Server connection and basic Linux commands, File permission layout, path, introduction to command line, Tabular data manipulation, shell redirections, pipes, bash scripting, biology pipline and data formats.

UNIT 2: Python

Introduction to Programming, Types of programming, Introduction to Python, Editing code, types of error, error handling, Text manipulation

UNIT 3: Conditions in Python, Exception Handling

UNIT 4: List & Loops in python

Unit 5: Functions, paired data and Dictionary in python

REFERENCES

- Jason Kinser, “Python for Bioinformatics”, Jones and Bartlett Publishers, Sudbury, Massachusetts 2009
- Martin J., “Python for Biologist: A complete programming course for beginners

ELECTIVES

Introduction to R & Bioconductor

UNIT 1: Introduction to R, History, Analysis: Functions, Classes, Methods

UNIT 2: Bioconductors: Introduction to Bioconductor and case studies, Packages, Objects, High throughput sequence flow, Preprocessing and Quality Assessment, Differential Expression, Machine Learning, Gene Set Enrichment Analysis, Short Reads.

REFERENCES:

- Cotton R, "Learning R: A Step-by-Step Function Guide to Data Analysis 1st Edition", O Relliye Publishers.

Web Resources

- https://intro2r.com/
- https://rstudio-education.github.io/hopr/

Database Management System

Unit 1: Introduction to Visual Basic

- Introduction to Visual Basic: IDE, working with forms, developing an application, variables, datatypes and modules, procedures and control structures, arrays in VB. Working with VBControls: Creating and using controls, working with control arrays - ODBC and Data Access Objects.

Unit 2: User interfaces

- Menus Events and Dialog Boxes: Menu and Events definition, Event model in VB, Menu Interfaces, Mouse Events, Dialog Boxes: Definition, Types of Dialog Boxes, Applying dialog. Graphics, MDI and FlexGrid: Graphics for application, Multiple Document Interface and Using the FlexGrid Control

Unit 3: VB Classes and Objects

- Classes: Definition, advantages of classes, class methods. Objects: Definition and methodology. Introduction to VB Classes and Objects, Creating various forms, Objects and projects. Working with objects, Classes and class modules, Creating VB objects.

Unit 4: Introduction to DBMS

- Advantages and Components of a Database Management Systems - Feasibility Study – Class Diagrams - Data Types - Events - Normal Forms - Integrity - Converting Class Diagrams to Normalized Tables - Data Dictionary. Query Basics - Computation Using Queries – Subtotals and GROUP BY Command - Queries with Multiple Tables Subqueries – Joins, Testing Queries.

Unit 5: Introduction to Oracle 9

- ORACLE - Introduction to Oracle, Data definition languages - Data Manipulation language, Data Control Language, Data types in Oracle. Constraints in Oracle, Data and String Functions, Union and Intersect operator, Sub queries, Introduction to PL / SQL, Simple PL / SQL programs.

REFERENCES

- Steven Holzner, “Visual Basic 6 Programming: Black Book”, Dreamtech Press, 2000.
- ISBN:13: 9788177220537
- C. J. Date, A. Kannan, “Database Systems”, Pearson Education Publication, 2006
- Noel Jerke, “Visual Basic 6: The Complete Reference”, Tata McGraw Hill, 1999. ISBN:139780074636664
- Kevin Loney, George Kuch, “Oracle – The complete Reference”, Tata McGraw Hill Publication, 2005 3. C. J. Date, “Database Systems”, Addison Wesley Publication, 1990.

System Biology

UNIT 1: Introduction to cell biology

- Molecules of life: Genes and proteins, Transcription networks, Regulation of gene transcription

UNIT 2: Biological Network

- Types of biological networks Network motifs in transcription regulation, Network motifs in signaling networks, Origins of biological robustness, Optimal gene circuits Week 9: Kinetic modeling of biochemical reactions Week 10: Kinetic modeling of large scale biomolecular networks , network analysis in systems biology

UNIT3: Integration of regulatory and metabolic networks

- Gene set Enrichment and Network Analyses In the 'Gene Set Enrichment and Network Analyses' module the emphasis is on tools developed by the Ma'ayan Laboratory to analyze gene sets. Several tools will be discussed including: Enrichr, GEO2Enrichr, Expression2Kinases and DrugPairSeeker.

UNIT4: Deep Sequencing Data Processing and Analysis

- Basic steps and popular pipelines to analyze RNA-seq and ChIP-seq data going from the raw data to gene lists to figures. These lectures also cover UNIX/Linux commands. Graph theoretic analysis of biological networks, Biological networks and drug development.

REFERENCES

- Introduction to Systems Biology. By Sangdun Choi. Published by Humana Press, 2007, ISBN 9781588297068
- Computational Systems Biology. By Andres Kriete, Roland Eils. Published by Academic Press, 2005, ISBN 012088786X
- Systems Biology in Practice: Concepts, Implementation and Application. By Edda Klipp, Ralf Herwig, Axel Kowald, Christoph Wierling, Hans Lehrach., Published by Wiley-VCH, 2005, ISBN: 978-3-527-60488-3
- System Biology: Computational Systems Biology (Hardcover) by Andres Kriete (Editor), Roland Eils (Editor)
- Microarray Data Analysis: Gene Expression Data Analysis. A Beginner's Guide By: Helen Causton (Imperial College), J Quackenbush and AlvisBrazma (The European Bioinformatics Institute)
- A Practical Approach to Microarray Data Analysis (Hardcover) by Daniel P. Berrar (Editor), Werner Dubitzky (Editor), Martin Granzow (Editor)

Computer Aided Drug Design

Unit 1: Introduction to Computer Aided Drug Design (CADD)

- History, different technique sand applications Quantitative Structure Activity Relationships: Basics History and development of QSAR: Physicochemical parameters and methods to calculate physicochemical parameters: Hammett equation and electronic parameters (sigma), lipophilicity effects and parameters (log P, substituent constant), steric effects (Taft steric and MR parameters) Experimental and theoretical approaches for the determination of these physicochemical parameters

Unit 2: Quantitative Structure Activity Relationships

- Applications Hansch analysis, Free Wilson analysis and relationship between them, Advantages and disadvantages; Deriving 2D-QSAR equations 3D-QSARapproaches and contour map analysis Statistical methods used in QSAR analysis and importance of statistical parameters

Unit3: Molecular Modelling and Docking

- Molecular and Quantum Mechanics in drug design, Energy Minimization Methods: comparison between global minimum conformation and bioactive conformation, Molecular docking and drug receptor interactions: Rigid docking, flexible docking and extra-precision docking.

Unit4: Molecular Properties and Drug Design

- Prediction and analysis of ADMET properties of new molecules and its importance in drug design. De novo drug design: Receptor/enzyme-interaction and its analysis, Receptor/enzyme cavity size prediction, predicting the functional components of cavities, Fragment based drug design. Homology modeling and generation of 3D-structure of protein

Unit 5: Pharmacophore Mapping and Virtual Screening

- Concept of pharmacophore, pharmacophore mapping, identification of Pharmacophore features and Pharmacophore modeling; Conformational search used in pharmacophore mapping In Silico Drug Design and Virtual Screening Techniques Similarity based methods and Pharmacophore based screening, structure based In-silico virtual screening protocols.

REFERENCES

- Computational and structural approaches to drug discovery, Robert M StroudandJanet.F Moore,RCS Publishers.
- Introduction to Quantitative Drug Design by Y.C. Martin, CRC Press, Taylor&Francisgroup..
- Drug Design by Ariens Volume 1 to 10,Academic Press, 1975, Elsevier Publishers.
- Principles of Drug Design by Smith andWilliams, CRC Press, Taylor & Francis.
- The Organic Chemistry of the Drug Design and Drug action by Richard B. Silverman, Elsevier Publishers.
- Medicinal Chemistry by Burger, Wiley Publishing Co
- An Introduction to Medicinal Chemistry –Graham L. Patrick, Oxford University Press.
- Wilson and Gisvold’s Text book of Organic Medicinal and Pharmaceutical Chemistry, Ippincott Williams & Wilkins.
- Comprehensive Medicinal Chemistry – Corwin and Hansch, Pergamon Publishers.
- Computational and structural approaches to drug design edited byRobert M Stroudand Janet. F Moore

Bioethics & IPR

Unit 1: Introduction

- Bioethics and IPR introduction and application with case studies. Types of IPR, Patent Law: rights and ownership

Unit 2: General Principals

- Transparency, Consent, Equitable sharing of benefits, Pluralism: conciliating different value systems.

Unit3: Introduction to biosafety

- Overview of biosafety, Risk assessment, Cartagena protocol on Biosafety, GMOs: Concerns and challenges Transgenic technology, Future opportunities and challenges

Unit 3: Key Aspects of Bioethics & IPR

- Ethical aspects of technology as such, Ethical aspects of granting exclusive IP rights over a technology Ethical aspects of seeking exclusive IP rights over a technology, Ethical aspects of exercising exclusive rights over a technology.

REFRENCES:

- IPR,Biosafety and Bioethics by Deepa Goel andShomini Parashar, Pearson publisher
- Basics of Patenting published by GTU Reference Books:
- Biotechnology in the Welfare of Mankind – Ali Khan
- Sasson A., Biotechnologies in developing countries present and future, UNESCO Publishers, 1993
- Singh K., Intellectual Property Rights on Biotechnology, BCIL, New Delhi.
- Biotechnology and Genomics, P.K.Gupta,Rastogi Publications.

Laboratory I: Bioinformatics

- Introduction to major online resources and interpreting search results NCBI, GenBank/ DDBJ/ EMBL, EMBOSS, Ensembl, PubMed, NR, UniProt, InterPro, OMIM, Enzymes, KEGG, Expassy, SWISS-PROT, PROSITE, PFam, PIR, InerPro, PSSM, TGCA. PDBPARAM, SCOPE, CATH, BioGRID.
- BLAST, standalone BLAST, PSI-BLAST, Genome Annotation, UCSC Genome Annotations, ENCODE, GWAS catalog, Galaxy, MEGA, PAML, Network, HyPhy, Data Monkey, gNOMO, Artemis/ACT, MLST typing, ResFinder, SeqSero, Serotype finder tool, Mobile element finder, GRIMM.

- DeepViewer, MODELLER, Auto Dock, Pymol, Rosetta, CASP.

- Genomics and proteomics analysis using online tool Galaxy

- Aligner base caller programs, RNA seq analysis, Assembly programs, synteny with circos

Laboratory II: Biomathematics & Biostatistics

MATLAB

Descriptors/Topics: Introduction to MATLAB

- Definition of MATLAB, Getting started, MATLAB working environment: Installing and running simple MATLAB programs, MATLAB key features.
- Variable names, Numerical variables and operations, Logical variables and operators, String variables, Vectors and Matrices, Plotting basics, MATLAB built-in functions, Flow control: if, else, if else, for and while loops. Solving linear equations, solving ordinary differential equations.
- Descriptors/Topics: SimBiology Toolbox: Getting started with SimBiology, Modeling: Import, build, and export mechanistic or PKPD representation of system dynamics, Running simulation: Simulate responses to biological variability or different dosing conditions, scan parameter ranges, calculate sensitivities
- Descriptors/Topics: Optimization and Biological Applications: Optimization with constraints, Bioinformatics Toolbox, Biological data analysis pipeline or work flows using Microarray data, Mass-spectrometry data and NGS data, Image Processing

1. The MATLAB interface
2. Write script files in MATLAB
3. Write simple functions
4. Import data and plot graphs
5. Using in-built functions
6. Creating your own functions
7. Use MATLAB to perform mathematical calculations
8. Running simulations using SimBiology Toolbox
9. Analyzing biological data using the MATLAB Bioinformatics Toolbox

Laboratory III: Linux & Python

Linux commands for server connection

File permission, pipeline

Python programs

To gain basic knowledge on programming skills in Python.

To know how to calculate the length of the sequence using perl program.

To perform the reverse of the given sequence using Rev List.

To perform concatenation of the given sequence by using dot operator.

To find the complement and reverse complement of the given sequence To know the GC content of the given DNA sequence in order to evaluate the stability of DNA. .

To convert DNA to protein sequence by using perl program and be used in the translation process.

To parse PDB and FASTA file

Retrieving DNA Sequence from the Database and Translating it into Protein Sequence

Laboratory IV: Cloud Computing

Shell Scripting

Introduction to shebang, hash, declaration of variables, file directory management, High performance computing using shell.

Hadoop

Introduction to Apache Hadoop, Hadoop Distributed File System (HDFS) and associated open source software projects like cloud Aligner, cloudburst, cross bow, Myrna, BioDoop etc

Map Reducer

Introduction to MapReducer, and associated open source software projects like cloud Aligner, MrsRF etc

Laboratory V: Electives

Elective: Database Management System

Creating Simple application forms in Visual Basic.

a) Creating a form for simple Arithmetic Calculations

b) Creating a form for simple Biological applications.

ORACLE, PL/SQL

Creation of student information records containing Roll number, Name, Subject Code Marks etc.,

Finding the total and average marks, result for each student table.

Record Manipulations such as Deletion, Modification, Addition and Counting the Record.

Creating table that demonstrates simple biological applications.

Creating table to demonstrate applications with biological sequences.

Create a database that demonstrates "Library Information System" with VB forms and

Query language (User Interface with VB).

Create a database that stores and retrieves simple biological applications. (User Interface with VB)

Create a database for "Railway Reservation System". (User Interface with VB) **2**

Create a database that stores and retrieves biological sequences and to find the similarities between two sequences. (User Interface with VB)

Elective: Introduction to R and Bioconductor

R installation, objects, classes and Methods

SNP Associations : snpMatrix

Differential Expression of Tag Counts : baySeq, DEGseq, edgeR

Peak Analysis : ChIPpeakAnno, chipseq, ChIPseqR

Alignment I/O & QA : Rsamtools, ShortRead

Solexa Base Calling & QA : Rolexa

Read Simulation : ChIPsim

Annotation I/O : biomaRt, rtracklayer

Annotation Packages : BSgenome.*, GenomicFeatures.*, org.*.db, SNPlocs.*
Annotation Infrastructure : AnnotatinDbi, BSgenome, GenomicFeatures
Sequence Visualizations : GenomeGraphs, HilbertVis, HilbertVisGUI

Sequence Infrastructure I : Biostrings, IRanges

Sequence Infrastructure II & Analysis : Genominator

Interval Alternative : genomeIntervals

Elective: Computer aided drug designing

Pharmacophore preparation

Chemical Library preparation based on Pharmacophore., Insilico Screening (ADME), Toxicity screening, QSAR (Quantitative structure-activity relationship) Model

Introduction to Structure based Drug Design and process

Layout of Docking, Data mining, literature study and acquisition of target structure, Comparative modelling of protein (Homology modelling) *target structure not available, Server based –PHYRE, RaptorX, SWISSMODEL, I-TASSER, etc., Protein structure validation(ProSA), Ramachandran plot, assessment(RAMPAGE,Pdbsum,Procheck), Active site/ Pocket identification – MetaPocket,CastP,Active site identification using PyMol,-Molecular Docking *using AutoDock vina/AutoDock Tools/PyRx (For docking of multiple ligands), Protein and ligand preparation, Setting grid parameters and Docking parameters -Docking analysis (based on binding energy, Hydrogen bond interactions, electrostatic interactions, hydrophobic interactions, etc.), BINANA (BINding ANAlyser), Pdbsum for visualising protein-ligand interactions, Building protein-ligand complex and visualization(publication standard), Report construction

Introduction to Molecular Dynamics Simulations

Basic Principles of Molecular Dynamics Simulations(Non-bonded interactions,Bonding Potentials, Force fields), Molecular Dynamics Simulation using GROMACS/VMD-NAMD, -Installation of softwares required for simulations, Definition of molecular structures, Force field and topology,Preparation of files, Simulation.

Elective: System Biology

1. Exploring systems biology-based databases: KEGG, Reactome, BRENDA, BioCyc, BioModels, SABIO-RK, MetaNetX, JWS Online, BiGG, RHEA, GEM, STRING, APID, ModelSeed, DAVID
2. Understanding SBML file format, libSBML, MIRIAM, SBGN
3. Importing, studying and constructing toy/published models in CellDesigner/Cell Collective
4. Simulating networks in COPASI
5. Constructing, annotating and analyzing network in Cytoscape and its plugins
6. Basics of MATLAB
7. Exploration and Flux analysis of metabolic networks using COBRA Toolbox/RAVEN Toolbox/COBRApy
8. Metabolomics data Analysis using MetaboLights, MetaboAnalyst