Syllabus for Computational Biology PhD programme at IMSc

All core and any one elective required. Total credits: 60

Sr. No.	Course Code	Course Name	Course Type	Credit
1.	10-LIFE24-601-C	Basic Biology	Core	8
2.	10-LIFE24-603-C	Mathematics and Statistics for Biologists	Core	8
3.	10-LIFE24-604-C	Physical Methods for Biologists	Core	8
4.	10-LIFE24-606-C	Biological Sequence Analysis	Core	8
5.	10-LIFE24-607-C	Systems Biology	Core	8
6.	10-LIFE24-608-C	Research Methodology	Core	4
7.	10-LIFE24-609-C	Project	Core	14
8.	10-LIFE24-601-E	Biophysics of Macromolecular Structures	Elective	6
9.	10-LIFE24-602-E	Simulation Techniques in Biology	Elective	6
10.	10-LIFE24-603-E	Population Biology, Ecology and Evolution	Elective	6
11.	10-LIFE24-604-E	Computational Neuroscience	Elective	6
12.	10-LIFE24-605-E	Modeling of Infectious Diseases	Elective	6
13.	10-LIFE24-606-E	Next-generation sequencing: Technologies, algorithms, applications	Elective	6
14.	10-LIFE24-607-E	Machine Learning	Elective	6

- In addition, courses offered by other CIs/OCCs/NPTEL/Swayam and Institutes and Universities having MoUs with HBNI may be opted.

Detailed syllabus

10-LIFE24-601-C **Basic Biology (81 lecture-hours), 100 marks**				
1	Basic molecular biology	B N	Biomolecules, DNA, RNA, proteins; genetic code; "central dogma"; gene transcription, translation; packaging of DNA in eukaryotes; Transcriptional regulation, miRNA and RNAi, introns/exons, splicing DNA packaging, heterochromatin and euchromatin,	14 lecture-hours 15 marks

		methylation, histone tail modifications, noncoding RNA and gene regulation	
2	Cell biology	Cellular metabolism, cell motility, cytoskeleton, intracellular transport, membrane transport, channels, receptors, signalling, cell cycle	10 lecture-hours 10 marks
3	Genetics	Mendelian genetics, definitions (genes, loci, alleles), dominance; replication, mitosis/meiosis, linkage/crossover	7 lecture-hours 10 marks
4	Developmental biology	differentiation, early development of drosophila via gradients, gap and pair-rule genes, role of hox genes, Williston's "law", other organisms, Waddington's canalization	10 lecture-hours 10 marks
5	Evolutionary biology	Molecular evolution, evolution of DNA, genes, proteins and regulation. Molecular mechanisms of evolution -- mutation, recombination, duplication, mobile elements	8 lecture-hours 10 marks
6	Basics of neuroscience and ecology	Neurons, synapses, neural architecture in various organisms, action potential, Hodgkin-Huxley equation, Ecology and evolution, ecosystems, food webs	8 lecture-hours 10 marks
7	Experimental techniques	PCR, southern/northern/western blots, chromatin immunoprecipitation, microarrays, high-throughput sequencing, ChIP-chip and ChIP-seq, high-resolution microscopy (fluorescence imaging, confocal, FRET, PALM etc), GFP and reporter gene assays	10 lecture-hours 15 marks
8	Protein Structure	Taxonomy: Primary, Secondary and tertiary structure, fold types Protein folding: The Anfinsen experiments, Protein database (PDB), Helix-helix packing in globular proteins, Beta-sheet packing, Folding pathways, thermal denaturation, partially folded intermediates, misfolding and aggregation	6 lecture-hours 10 marks
9.	Immunology	Innate and adaptive immunity - Key players of the immune system: T cells, B cells et al - specificity and cross-reactivity in ligand-receptor (antigen-antibody) interactions - kinetic proofreading for antigenic discrimination - Cell-cell communication and cytokinesigaling - Cell fate: hematopoiesis differentiation tree - Immune repertoire: thymic selection and clonal size distribution - Population dynamics: intracellular (lymphocytes) and species-evel(host-pathogen)	4 lecture hours 5 marks
10.	Other topics	Basics of: Intercellular communication, epidemiology, physiology	4 lecture-hours 5 marks
10-LIFE24-603-C **Mathematics and statistics for biologists (81 lecture-hours, 100 marks)**			
1	Differential equations	introduction to ODEs and PDEs, linear and non-linear, properties, how to solve analytically and numerically; examples -- Hodgkin-Huxley, reaction-diffusion equations, Volterra equations	15 lecture-hours 16 marks
2	Essentials of linear algebra	vectors, matrices, eigenvalues and eigenvectors; orthogonal bases of functions, Sturm-Liouville theory and differential equations; Fourier series and Fourier transforms	15 lecture-hours 16 marks

3	Probability theory and statistics	basic concepts -- random variables, mean, variance, moments; conditional probabilities, hypothesis and data, likelihood, Bayes' theorem; probability distributions -- binomial, multinomial, Poisson, normal; the central limit theorem; hypothesis testing, significance testing (orthodox and Bayesian methods); parameter estimation	20 lecture-hours 30 marks
4	Simulations	Introduction to Markov Chain Monte Carlo for exploring space of hypotheses: ergodicity, detailed balance, convergence. Metropolis and Gibbs sampling	9 lecture-hours 8 marks
5	Machine learning	decision tree learning, artificial neural networks, support vector machines, Bayesian learning and Bayesian networks	12 lecture-hours 15 marks
6	Other topics	Game theory, applications to evolutionary biology, agent-based modelling of complex systems	10 lecture-hours 15 marks
10-LIFE24-604-C **Physical Methods for Biologists (81 lecture-hours, 100 marks)**			
1	Basic physics of soft matter	What is soft matter, length scales and time scales, biological matter as soft matter, self-organization and self assembly, illustrative examples - DNA, microtubules and/or actin and lipid membranes, coarse-grained representations, interactions and bonding in soft matter systems (including van der Waals forces, hydrogen bonding, electrostatics and screening), what can be measured, energy scales	10 lecture-hours 12 marks
2	Thermodynamics and statistical mechanics	Thermal equilibrium, the idea of entropy, laws of thermodynamics, free energies, Legendre transformations, different ensembles and relation to computational biology examples, Boltzmann distribution, harmonic oscillator, equipartition theorem, virial theorem, thermodynamics of self assembly, simple ideas of phase transitions, Poisson-Boltzmann theory, dealing with electrostatics	12 lecture-hours 15 marks
3	Noise, diffusion and drift	Thermal fluctuations and noise, random walk, diffusion equation as continuum limit of the random walk, probability density, continuity equation, Fick's law, drift-diffusion equation, Stokes-Einstein formula, example of receptor clustering	10 lecture-hours 10 marks
4	Mechanics of continuous media	Elasticity of isotropic solids, estimates for elastic constants of biological materials, fluids in biology, basics of fluid mechanics, Pascal's law, Euler's equation, viscosity, Reynolds number, Navier-Stokes equation, flow through narrow pipes, dimensionless groups, swimming of microorganisms, hydrodynamic interactions, rheology of biological matter, introduction to viscoelasticity, Maxwell model	15 lecture-hours 18 marks
5	Polymers, membranes and gels	Simple ideas of polymers and membranes, polymer elasticity, polymer dynamics (Rouse and Zimm model) qualitative discussion, scaling ideas in polymers, semi-flexibility, membrane elasticity, membrane fluctuations, passive gels	15 lecture-hours 18 marks
6	Proteins, enzymes, catalysis	Biological catalysts, Gibbs free energy, transition state complex, substrate, products, active sites, activation energy barrier, induced-fit hypothesis, cofactors, coenzymes,	5 lecture hours 8 marks

		Michaelis-Menten enzyme kinetics	
7	Electrostatics in biology	Continuum methods, solvation and ions, implicit solvent models, Poisson equation, Poisson-Boltzmann equation, solvation free energy	5 lecture hours 8 marks
6	Out of equilibrium	Active matter, simple examples, what do we need to model them, polymerization forces, cell streaming, molecular motors and models, active gels	5 lecture-hours 6 marks
7	Other topics	Interfacial tension in biological systems, Laplace pressure, wetting and spreading, osmotic effects, capillary effects in biology, micro-rheology for biological systems	4 lecture-hours 5 marks
10-LIFE24-606-C **Biological sequence analysis (81 lecture-hours, 100 marks)**			
1	Biomolecules	Basics (DNA, RNA, proteins)	1 lecture 0 marks
2	Probability theory	Basic laws -- joint probabilities, conditional probabilities, likelihood, Bayes' theorem	5 lecture-hours 8 marks
3	String algorithms	finding common substrings and subsequences: Boyer-Moore algorithm, suffix trees, finding strings with mismatches	10 lecture-hours 14 marks
4	Sequence alignment	algorithms for pairwise and multiple sequence alignment -- scoring model, Needleman-Wunsch and Smith-Waterman algorithms, BLAST and other heuristic algorithms, significance of scores, structural alignment	12 lecture-hours 18 marks
5	Sequence assembly	assembling short reads, with and without scaffold; ChIP-seq algorithms	10 lecture-hours 10 marks
6	Markov models	Markov chains, hidden Markov models, Baum-Welch and Viterbi algorithms, profile HMMs and software (HMMer, etc)	15 lecture-hours 15 marks
7	Transcriptional regulation	Transcription factor binding sites, position weight matrices, sequence logos, motif-finding via expectation maximisation (MEME) and Gibbs sampling	8 lecture-hours 10 marks
8	Phylogenetic trees	building a tree from pairwise distances, neighbour-joining, parsimony	10 lecture-hours 10 marks
9	Proteins: Structural characterization	x-ray crystallization, circular dichroism, spectroscopy, NMR, single molecule experiments	5 lecture-hours 7 marks
10	Proteins: Homology modelling	Homology Modeling, Visualization	5 lecture-hours 8 marks
10-LIFE24-607-C **Systems Biology (81 lecture hours, 100 marks)**			
1	Networks in biology	The diversity of networks across space and time in biological systems Intra-cellular networks: The gene network and protein-protein interaction network Intra-cellular networks: The metabolic network Intra-cellular networks: signaling networks - pathways and enzyme-substrate reaction cascades The signaling network coordination of immune response to infection	27 lecture-hours 35 marks

		Reconstructing biological networks from lab experiments Structural analysis of networks: Global properties Structural analysis of networks: Motifs and Modules Dynamics on biological networks: Modeling signaling pathways Inter-cellular networks: Neuronal networks Inter-organism networks: Contact structure and contagion propagation Inter-species networks: Stability-instability of food webs	
2	Patterns in Biology	Temporal patterns: Biological clocks and circadian rhythms Oscillatory activity in Pancreatic beta cells and insulin secretion Pattern formation during development Development in Drosophila Development of the vertebrate body plan Modeling developmental patterns: Reaction-diffusion models and Turing Patterns Spatial patterns: Linear stability analysis and Fourier modes Autocatalysis and lateral inhibition: Gierer-Meinhardt and related pattern generation mechanisms in biosystems, center-surround principle in retina and cortex Modeling genesis of functional patterns: Ocular dominance columns Development of plants and L-systems modeling Cell differentiation and Random NK Boolean Networks Morphogenesis Fractals in biology: Examples (1/f noise, circulation system), characterization Fractals in biology: Generation mechanisms	27 lecture-hours 35 marks
3	Waves in biology	Importance of waves in biology for communication and coordination Intra-cellular waves: Calcium waves, targets and spirals Inter-cellular waves: Waves in the brain, heart and uterus Excitable media models of physiological systems Ionic basis of excitation: Hodgkin-Huxley formalism Simple and complex models of excitability Excitability, Oscillatory and Bistability regimes of systems Wave propagation through inter-cellular gap junctions: Diffusion approximation Genesis and dynamics of spiral waves: kinematic approach Nonlinear dynamical aspects of spiral waves: Restitution and dispersion Excitation-contraction coupling and the role of organ structure in wave dynamics Bidomain models of biological electrical activity Waves in single populations: Fisher waves Waves in interacting populations: Propagating epidemics, spiral waves in host-parasite spatial dynamics	27 lecture-hours 30 marks

ELECTIVE COURSES (semester flexible)

10-LIFE24-601-E Biophysics of Macromolecular Structures (48 lecture hours, 100 marks, 6 credits)

(i) Structure and Biophysics of Biomolecules

Introduction to macromolecular chemistry, building blocks for macromolecular structures, biophysical methods for structure analysis, nucleic acid structure, protein-nucleic acid interactions, membrane proteins, microtubules and other supramolecular assemblies, investigative methods from the atomic to cellular levels, including X-ray crystallography, NMR spectroscopy, molecular dynamics, electron and light microscopy, AFM, single molecule techniques and simulations

1. Kinetics

Chemical kinetics and application to dynamical processes in proteins, self assembly processes, classical kinetics, transition state theory, unimolecular decomposition, potential energy surfaces, scattering processes and photodissociation processes, enzyme kinetics

III. Biophysical approaches to Biopolymers

Basics of polymers, protein folding problem, protein aggregation, DNA, DNA electrostatics, DNA force extension relations, RNA folding, polymerization, polymerization forces, dynamic instability, tread-milling and their physical description

1. Biophysical Approaches to Membranes

Lipids and Membranes: Structure of various cell membranes, surface tension and curvature energies, Helfrich theory, clustering, phase separation, nanoscale structures i.e. rafts, multicomponent membranes

<u>SPECIAL TOPICS</u>

V. Kinetics and statistical mechanics of helix coil transitions; physical approaches to the refolding and assembly of multi-subunit proteins; fluorescence spectroscopic studies of macromolecules, molecular basis of enzyme catalysis, antibody structure and function, virus structure and assembly

10-LIFE24-602-E Simulation Techniques in Biology (48 lecture hours, 100 marks, 6 credits)

I. Molecular Dynamics

Introduction to MD and applications in biology and drug design; Basic Statistical mechanics: Basic thermodynamics, Ensembles (microcanonical, canonical, grand canonical, isothermal-isobaric), Virial theorem, Nose-Hoover chains; Forcefields and interaction potential: Many body potentials, Born-Oppenheimer approximation, electrostatic interactions including Ewald sum, interaction potential for organic molecules; popular forcefields: AMBER, CHARMM, OPLS etc.,;Integration methods and Liouville time operators Phase space concepts, Liouville theorem, Equilibrium solution of Liouville equation, Trotter factorization; Integration algorithms: Verlet, Velocity-Verlet, Gear-Predictor, multiple-time step algorithm, holonomic constraints (RATTLE/SHAKE)

II. Monte Carlo Simulations

Importance Sampling, Random variables and stochastic processes,lattice models, Random walks, Gibbs sampling, sampling errors, configurational-bias Monte Carlo method, Markov chain Monte Carlo, Advanced Monte Carlo methods: Parallel tempering, simulated annealing

III. Reaction Diffusion

Predator Prey Models, Reaction Kinetics, diffusion-limited reactions, Population dynamics, Reaction-diffusion Equations

IV. Brownian/Stochastic simulations

Stochastic reaction-diffusion models:Compartment-based reaction-diffusion algorithm, reaction-diffusion master equation, pattern formation; Diffusion: Brownian motion, On/Off-Lattice models, diffusion to adsorbing surfaces, reactive boundary conditions, Einstein-Smoluchowski relation; Stochastic models of transport processes in cells: Fokker Planck Equations, Brownian ratchet models, Chapman - Kolmogorov equation, Gillespie algorithm, chemical master equation

SPECIAL TOPICS

V. Free energy methods

Potential of mean force, umbrella sampling, Adaptive bias force method, thermodynamic integration

VI. Binding and Docking

Enzyme-substrate recognition process, Search Algorithms (simulated annealing, steepest descent, genetic algorithms), Scoring Functions, Applications of Docking, Softwares for docking

10-LIFE24-601-E Population Biology, Ecology and Evolution (48 lecture hours, 100 marks, 6 credits)

1. Single species population

Continuous and discrete-time models of population growth (Logistic and related models)
Models of age-structured populations
Population dynamics in the presence of noise
Time-series analysis of data
Flies: Model experimental organism for studying population dynamics
Modeling migration of populations
Territorial behavior
Fundamentals of game theory
Evolution of cooperation between individuals
Spatial dynamics of strategies (Example: Spatial Prisoner's Dilemma)

2. Interaction between multiple populations

Introduction to food webs and ecological interactions between species
Predator-prey interactions: Lotka-Volterra and related models
Functional response
Competition
Cooperation
Multiple prey and predators: Generalized Lotka-Volterra and related models
Stability vs complexity in ecosystems: Single trophic level
Stability vs complexity in ecosystems: Multiple trophic levels
Experimental techniques for studying impact of diversity on stability
The robustness of complex ecological networks

3. Evolution and population genetics

Fundamentals of population genetics: Random mating and Hardy-Weinberg principle
Classical mathematical genetics: Single locus with multiple alleles
Classical mathematical genetics: Multiple loci
X-linked genes; Linkage and its distribution

The molecular basis of classical genetics
Fitness landscapes and mathematical models of evolution
The major transitions in evolution
Mutation and natural selection
Random genetic drift
Neutral theory of evolution
Coevolution and evolutionary game theory
Evolutionary ecology

10-LIFE24-604-E Computational Neuroscience (48 lecture hours, 100 marks, 6 credits)

1. Neurons, Synapses, Gap Junctions and Small Circuits
   - Introduction to the biological components of the nervous system
   - Types of Neurons and Glial cells
   - Neuronal activity: Action potential and Graded potential
   - Ion channels and electrical activity of neurons
   - Dynamics of graded potential neurons (Example: retina)
   - Dynamics of action potential neurons, spikes and spike trains
   - Dynamics of inter-neuron communication: Synaptic transmission
   - Dynamics of inter-neuron communication: Gap junctions
   - Introduction to GENESIS/NEURON simulation platforms
   - Neuron-Glial interaction
   - Small neuronal circuits and motifs

2. Systems Neuroscience
   - Introduction to the computational perspective for studying the brain
   - Introduction to Neural Network Models: McCulloch-Pitts paradigm
   - Associative Memory and the Hopfield Network
   - Storage capacity and stability of memories in Hopfield Network: Mean-field theory
   - Learning: Donald Hebb's Hypothesis, Long-Term Potentiation and STDP
   - Perceptron and related models: learning to generalize
   - Dynamics of Learning: Hebbian and Competitive principles
   - Information theory and neuro-communication
   - Development of the nervous system in a growing organism
   - Evolution of the nervous system: from single cells to the brain
   - Invertebrate neuroscience: C. elegans as a model organism
   - Modeling the nervous system of invertebrates
   - Sensory-motor integration in the nervous system

3. Vision and cognitive neuroscience
   - Introduction to Sensory Processing in the Nervous System
   - Components of the Visual System
   - Dynamics of Early Visual Processing at Retina
   - Receptive fields and centre-surround principle (Mach bands, etc.)
   - Processing at the Primary Visual Cortex and Higher Brain Areas
   - Modeling edge detection, shape from texture and motion detection
   - Visual binding: Synchronization of neuronal activity
   - Optical illusions as tool for studying vision
   - Information theory of vision

Introduction to cognitive neuroscience
Experimental tools of cognitive neuroscience: fMRI, PET, etc.
Linguistic ability: A model system for cognitive neuroscience

10-LIFE24-605-E Modeling of Infectious Diseases (48 lecture hours, 100 marks, 6 credits)

1. Genomics & evolutionary biology of pathogens
   Dynamics of molecular evolution
   Vertical and horizontal gene transfer
   Genomic landscape of pathogens, vectors and humans (Example: malaria); Coevolution and Red queen hypothesis
   Gene regulation, pathogenesis and immune response
   Evolution of virulence

2. The biology and modeling of host-pathogen interactions
   The immune system: design, phylogeny and ontogeny
   The functional anatomy of immune response
   Analysis of idiotypic network interactions
   Systems biology principles for intra-cellular signaling in immune response
   Systems-level modeling of Mycobacterium tuberculosis host-parasite protein-protein interactions
   Micro-epidemiology: population dynamics of viruses and host cells, May-Nowak and related models; application to HIV

3. Epidemiology: data analysis and mathematical modeling
   Epidemics: Dynamics and basic reproductive ratio R0
   Estimation of R0 from data - statistical techniques
   Immunization and other public health intervention strategies
   SIR model of epidemics: derivation and solution
   Variants of SIR model: SEIR, SIS and SIRS
   Modeling vector-borne diseases
   Host-parasite models (example: Nicholson-Bailey model)
   Cellular automata models
   Eco-epidemiological models
   Contact network: structure and dynamics
   Agent-based models of infection propagation

10-LIFE24-606-E Next-generation sequencing: Technologies, algorithms, applications (48 lecture hours,100 marks, 6 credits)

Basics: Next-generation sequencing technologies -- Illumina, SoLiD, Ion Torrent: basics, operations, limitations of each Illumina platform: read quality checking, in detail; Single-end vs paired-end reads

Assays, overview: High-throughput assays based on NGS: RNA-seq, ChIP-seq, ChIP-exo, bisulfite-sequencing; DNA accessibility (DNAse-seq, Faire-seq, Atac-seq); chromatin conformation (Hi-C and related, Chia-pet and similar); and basic algorithms for analysis of above

Genome assembly algorithms: overlap-layout-consensus algorithms, de Bruijn graph based algorithms, difficulties, contigs and supercontigs, how to "finish" genomes, future perspectives

Basic string algorithms: substring-matching -- Knuth-Morris-Pratt and Boyer-Moore (for familiarity), suffix trees, suffix arrays, Burrows-Wheeler transform and FM index

Read mapping programs: using suffix trees (RNA-star), using BWT (hisat2, bowtie, BWA)
Transcript assembly and quantification: Basic concepts in assembling, estimating expression/relative expression; tools: Cufflinks, pseudoalignment-based algorithms (Salmon, Kallisto)
Graph-based genome alignment: Representing a "pan-genome" as a graph, BWT on a graph, HISAT2

Peak-calling: in ChIP-seq, ChIP-exo; motif-finding in ChIP-seq data
Analysis of actual RNA-seq datasets (using hisat2, cufflinks), and ChIP-seq datasets (using bwa/macs/gem)
Single-cell RNA-seq, atac-seq: techniques and challenges
New technologies and assays: Nanopore (long reads); STARR-seq and STAP-seq
Original literature discussion, student seminars

10-LIFE24-607-E Machine Learning (48 lecture hours,100 marks, 6 credits)

While the theory is general, applications will focus on biomedical problems.
Basics of probability theory and statistics (univariate and multivariate models); discrete and continuous probability distributions; Bayesian probability theory, priors, conjugate priors and posteriors for discrete distributions; probabilistic models, parameters and hyperparameters

Basics of machine learning: supervised, unsupervised, reinforcement learning, examples

Unsupervised ML: clustering (agglomerative clustering, k-means, Gaussian mixture models, t-SNE, etc)

Review of linear algebra
Dimensionality reduction: principal component analysis,t-SNE, UMAP

Supervised learning, classification problems: Linear models: linear regression, logistic regression, ridge and lasso penalties, generalized linear models

Model selection: AIC, BIC, Bayesian Occam razor; dangers of overfitting; using training/validation/testing sets, bias-variance trade-off

Other supervised learning methods: naive Bayes, decision trees, random forests and gradient boosting, support vector machines

Introduction to neural networks and deep networks; reinforcement learning, generative adversarial networks

Other topics: Feature selection, missing data information

Applications to biology and medicine: transcriptomics and clustering, sequence analysis and regulatory site prediction, protein structure prediction, image analysis and biomedical applications; literature; hands-on exercises based on scikit-learn library

References:
Tom Mitchell, “Machine Learning: An Artificial Intelligence Approach” (1997)

James, Witten, Hastie, Tibshirani, "An introduction to statistical learning" (2nd ed, 2021)
Kevin Murphy, "Probabilistic Machine Learning: An Introduction" (2022)
SciKit-Learn library documentation (online)
Literature applications