



الجامعة الوطنية - السودان
National University - Sudan

M.Sc. Bioinformatics



Introduction

Bioinformatics is where computer science and biology join to address and solve fundamental problems associated with human health, the environment, and agriculture. Advances in sequencing technology, structural biology, and computational methods have produced an unprecedented volume of biological data, spanning genomic, transcriptomic, proteomic, and single-cell measurements, and the tools required to interpret this data continue to evolve rapidly. With the advancement of computing, machine learning, and information science, a wide range of databases, software, and predictive models is now available to researchers worldwide. This Master's Program equips learners with the practical and theoretical foundation needed to develop and apply modern bioinformatics tools. It integrates disciplines as varied as molecular biology, biochemistry, and statistics with programming, machine learning, and algorithm design. The training provided to graduates prepares them for careers across medical research, drug discovery, veterinary and agricultural research, food safety, agricultural biotechnology, and biodiversity studies, as well as in data-intensive computational roles within biotechnology and pharmaceutical industries.

General Objective

Introduce students to the interdisciplinary nature of bioinformatics and computational biology, and build computational and statistical skills to address expanding biological information.

Specific Objectives

The program qualifies candidates to:

- Build familiarity with computational and statistical methods for addressing problems in molecular biology.
- Develop competence in the storage, retrieval, sharing, and use of biological information across core areas of bioinformatics.
- Master sequence alignment, phylogenetics, genomics, proteomics, structural biology, and data-driven modelling using

machine learning and deep learning.

- Develop applied skills in immunoinformatics, computational drug design, and discovery.
- Identify, retrieve, manage, and exchange research information in the biological and biomedical sciences.
- Apply reproducible, version-controlled computational practices consistent with current research and open-science standards.

Expected Learning Outcomes

By the end of the program, the candidate should be able to:

- Identify strategies and apply appropriate computational tools in bioinformatics.
- Identify topical problems drawn from ongoing research and assess their applications.
- Interpret data relating to nucleotide and amino acid sequences, protein domains, and protein structure.
- Apply next-generation sequencing (NGS) data analysis workflows, including quality control, read alignment, and variant calling.
- Use modern protein structure prediction tools, including AlphaFold2/ColabFold, and apply predicted structures in downstream modelling and docking workflows.
- Apply foundational machine learning and deep learning methods to biological sequence and omics data.
- Develop and implement tools that enable efficient access to and management of biological data, including programmatic access to public databases.
- Conduct reproducible, version-controlled bioinformatics analyses using Git and literate programming tools such as Jupyter notebooks or R Markdown.
- Apply FAIR data principles and open-science practices in the design of bioinformatics research.
- Acquire and apply essential programming skills in Python and R for biological data analysis.

Admission Requirements

- Applicant must satisfy the general regulations set by the Faculty of Graduate Studies and Scientific Research of National University for registration for master degrees.
- Eligible applicants are holders of a Bachelor's degree from the National University or from any other accredited university or college in the following fields:
 - B.Sc. (Honours) with Grade Good (Second Class) minimum or cGPA ≥ 2.5 in: Biochemistry, Physics, Zoology, Botany, Chemistry, Mathematics, Biotechnology, Agriculture, Medicine, Veterinary Medicine, Pharmacy, Medical Laboratory Sciences, Dentistry, Information Technology, Animal Sciences, Ecology, Radiology and Radiological Imaging, Health Informatics, or Computer Sciences.
 - B.Sc. (Honours) with Grade Pass (Third Class), plus at least one qualifying semester in: Biology or related fields, Molecular Medicine, Molecular Genetics, Research Methodology, Computer Science or related fields, Statistics, Information Technology, Health Informatics, or Mathematics.
 - B.Sc. (General) with Grade Good at least (Second Class), plus one qualifying year in the fields listed above.
 - B.Sc. (General) with Grade Good/Pass (Third Class), plus a Post-graduate Diploma in the fields listed above.
 - All eligible candidates should pass an interview and an entry examination.

Study Program

Semester One

Code	Course	Credit Hours	Contact Hours/ Week (Theory)	Contact Hours/ Week (Practical)
BIO-611	Basic Molecular Biology	3(2+1)	2	2
BIO-612	Introduction to Proteomics	2(2+0)	2	0

BIO-613	Programming for Bioinformatics	3(2+1)	2	2
BIO-614	Database Management Systems	2(1+1)	1	2
BIO-615	Fundamentals of Algorithms	2(1+1)	1	2

Semester Two

Code	Course	Credit Hours	Contact Hours/Week (Theory)	Contact Hours/Week (Practical)
BIO-621	Fundamentals of Bioinformatics	4(2+2)	2	4
BIO-622	Computational Structural Biology	4(2+2)	2	4
BIO-623	Alignment, Sequences and Genomics	3(2+1)	2	2
BIO-624	Machine Learning and Deep Learning for Bioinformatics	4(2+2)	2	4
BIO-625	Biostatistics	3(2+1)	2	4

Semester Three

Code	Course	Credit Hours	Contact Hours/Week (Theory)	Contact Hours/Week (Practical)
BIO-631	Immunoinformatics	4(2+2)	2	4

BIO-632	Molecular Modelling and Drug Design	4(2+2)	2	4
BIO-633	Research Methodology	3(3+0)	3	0
BIO-634	Research Paper	4(0+8)	0	8

Courses Contents

BIO-611 Basic Molecular Biology

Theory: Genome organization; DNA structure, replication, repair, and recombination; Prokaryotic and eukaryotic transcription; Post-transcriptional modifications; Translation and transport; Mutations: oncogenes and tumor suppressor genes; Epigenetic regulation: DNA methylation, histone modification, and chromatin accessibility.

Practical: DNA extraction by different methods; RNA extraction; Polymerase Chain Reaction (PCR); Real-Time PCR; RFLP; Gel electrophoresis; DNA sequencing principles, including a comparative overview of Sanger and next-generation sequencing chemistry.

BIO-612 Introduction to Proteomics

Theory: Protein structure and function; amino acids and their properties; Amino acids form polypeptides; Protein structure: four levels of organization; Cellular functions performed by proteins; Protein folding and misfolding; Protein-protein interaction; Mass-spectrometry-based proteomics: sample preparation, peptide identification, and database searching.

Practical: None.

BIO-613 Programming for Bioinformatics

Theory and Practical: Python fundamentals and scripting for bioinformatics; Scientific computing with Python: NumPy, pandas, and BioPython; Data visualization with Matplotlib; Introduction to R for biological data analysis; Bio-

conductor basics: importing and manipulating omics data structures; Version control with Git/GitHub for reproducible scripts and collaborative work; Writing reproducible analyses using Jupyter notebooks; Data clustering, classification, searching, and sorting in applied bioinformatics contexts.

BIO-614 Database Management Systems

Theory and Practical: Introduction to database management systems; Data models; Structured Query Language (SQL); Relational database design and storage; Programmatic access to biological databases: NCBI Entrez Programming Utilities (E-utilities) and the Ensembl REST API.

BIO-615 Fundamentals of Algorithms

Theory and Practical: Computing algorithms; Sorting, searching, and string matching; Graphs; Trees; Algorithm design and analysis.

BIO-621 Fundamentals of Bioinformatics

Theory: Introduction to bioinformatics and bioinformatics resources; Open-access bibliographic resources and literature databases; Biological databases; Bioinformatics software and tools; Sequence file formats; Protein and nucleic acid sequence; Submitting DNA sequences to public databases; Taxonomy and phylogeny; Sequence patterns and profiles.

Practical: Entrez and literature searches: PubMed, PubMed Central, OMIM, OMIA, Citation Matcher; Biological database systems: nucleotide/genome databases, protein sequence databases, structure databases, protein pattern databases; File format conversion: FmtSeq, ReadSeq; Sequence file formats: GenBank, FASTA, FASTQ; Protein and nucleic acid property analysis: ExPASy server tools, EMBOSS; Sequence and alignment software: BioEdit, MEGA, MEME; Introduction to command-line workflow environments (e.g., Galaxy) as a foundation for reproducible pipelines.

BIO-622 Computational Structural Biology

Theory: Gene prediction; Primer design; Proteomics; X-ray crystallography;

Predicting protein structure and function from sequence; Structure prediction strategies: secondary structure prediction and the deep-learning-based approach to tertiary structure prediction (AlphaFold2 and related models); Homology modelling concepts; classification and comparison of protein 3D structures; Protein–protein interaction; Basic concepts in the identification of disease genes (SNPs).

Practical: Proteomics tools at the ExPASy server; Primer analysis using OLIGO and Primer3; SNP database (dbSNP) and related software; ORF-finding databases and software; Secondary structure prediction tools; Structure visualization using ChimeraX and PyMOL; Protein structure prediction using AlphaFold2/ColabFold; Homology modelling using MODELLER; Active-site identification and ligand prediction software.

BIO-623 Alignment, Sequences and Genomics

Theory: Sequence analysis; Scoring matrices; Pairwise sequence alignments; Multiple sequence alignments (MSA); Comparative genomics; Sequence patterns and profiles; Algorithms for derivation and searching of sequence patterns; Taxonomy and phylogeny; Next-generation sequencing (NGS): sequencing platforms, FASTQ data, read quality, and the general short-read analysis workflow.

Practical: Sequence analysis packages: EMBOSS, NCBI Toolkit, SMS; Pairwise alignment; Database search tools: BLAST, FASTA; Multiple sequence alignment: MAFFT, MUSCLE, Clustal Omega; Sequence patterns and profiles: generation of sequence profiles, PSI-BLAST; Derivation and searching of sequence patterns: MEME/MAST, PHI-BLAST, ScanProsite, PRATT; Protein motif and domain analysis: MEME/MAST, eMotif, InterProScan, ProSite, ProDom, Pfam; Phylogenetic analysis using MEGA, IQ-TREE, and RAXML; NGS practical: quality control with FastQC, adapter trimming, reference-based read alignment (BWA or Bowtie2), and an introduction to variant calling.

BIO-624 Machine Learning and Deep Learning for Bioinformatics

Theory and Practical: Clustering methods (k-means, hierarchical clustering)

applied to biological data; Classification and prediction (decision trees, random forests, support vector machines); Fundamentals of neural networks and deep learning; Practical deep learning using PyTorch or TensorFlow/Keras for biological sequence and omics data; Survey of biological foundation models and their applications, including protein language models (e.g., ESM) and genomic and single-cell language models (e.g., scGPT); Model evaluation, overfitting, and validation strategies for high-dimensional biological data.

BIO-625 Biostatistics

Theory and Practical: Introduction to biostatistics: significance and basic concepts; Descriptive statistics: measures of central tendency and dispersion; Introduction to inference; sampling distribution and confidence intervals; Sampling variability, confidence intervals, and p-values for means and differences in means; An introduction to hypothesis testing: the paired t-test; Comparing means among two or more independent populations; Measures of association, sampling variability, confidence intervals, and p-values for binary outcomes; Simple linear regression; Comparing proportions between two independent populations; Non-parametric alternatives; analysis of variance; Analyzing categorical data: goodness-of-fit and contingency tables; Statistical considerations for high-dimensional omics data, including multiple-testing correction (Bonferroni, FDR/Benjamini–Hochberg).

BIO-631 Immunoinformatics

Theory: Introduction to immunology; Antigens and antibodies; Major histocompatibility complex; Disorders of the human immune system; Introduction to immunoinformatics and immunological databases; Computational vaccinology.

Practical: Designing vaccines using online tools: VaxiJen, IEDB, PATCHDOCK; Epitope prediction tools, including NetMHCpan and IEDB tools, alongside SOSUI for membrane-protein analysis; Offline tools: MODELLER, ChimeraX, RasMol, CLC.

BIO-632 Molecular Modelling and Drug Design

Theory: Introduction to drug discovery and development; Drug targets; Lead identification and modification; Computer-aided drug design; Drug delivery; Pre-clinical and clinical testing; AI-assisted virtual screening in modern lead identification.

Practical: Molecular visualization software: PyMOL and RasMol; Geometry optimization; Binding-site identification; Structure-based drug design (molecular docking), using protein structures generated in BIO-622, including AlphaFold-predicted models where experimental structures are unavailable; Ligand-based drug design: QSAR.

BIO-633 Research Methodology

Theory: Foundations of research methodology and introduction to different types of research; What is research?; Finding research gaps and research topics; Proposal writing; Pilot studies; Ethical considerations in research; Ways of data collection; Obstacles in research; Scientific writing; FAIR data principles (Findable, Accessible, Interoperable, Reusable) and open-science and reproducibility practices in computational research.

BIO-634 Research Paper

Practical: Seminar in Bioinformatics I: a variable-content course engaging participants in current bioinformatics research and developing oral presentation and scientific communication skills. Senior Project in Bioinformatics II: the first part of a two-part guided research project, in which the student selects a research topic and writes a detailed proposal outlining its goals and objectives. Senior Project in Bioinformatics III: the second part of the two-part guided research project, in which the student carries out the proposed research and presents the results in the form of a scientific research paper.

Human Resources and Facilities

Teaching Staff

Not finalized yet.

Facilities

- **Lecture Rooms:** One Lecture Room: 150 seats.
- **Laboratories:** One Teaching Lab: 50 seats.
- **Computer Labs:** One Computer Laboratory: 125 computers.
- **Libraries:** University main library: 400 seats; E-Library: 250 seats.
- **Duration of the Program:** 12 months (48 weeks / 3 Semesters).
- **Teaching Modules:** Lectures, practical sessions, seminars, tutorials, presentations, assignments, and computational modeling projects.
- **Teaching Language:** English.

Examination Regulations

- Abide by the examinations rules of the general regulations of the graduate studies of the National University-Sudan.
- Examinations are run within the University and assessed according to the set regulations. The written and practical examinations are run according to the requirements of the program.
- The pass mark for any course is 60%, and students must succeed in both the theoretical and practical components of each course separately.
- A student missing an examination without an acceptable reason will be assigned a failure in the concerned course(s).
- A student has the right to sit for supplementary examinations in no more than one-third of the number of courses, or to re-sit examinations in all courses if they failed in more than one-third to one-half of the number of courses.

- A student failing more than 50% of the courses will be discontinued.
- A student failing any supplementary examination shall be discontinued.
- Supplementary examinations convene after the final examinations of each semester.
- Results of the written examinations are approved by the Faculty Council, upon a recommendation from the Program Committee.

Assessment Methods:

- Courses with no practical sessions: Continuous assessment 30% , Final examination 70%.
- Courses involving practical sessions: Continuous assessment 30% , Final examinations: Theory (30%), Practical (40%).

Grading System: A+ (90-100), A (80-89), B+ (75-79), B (70-74), C+ (65-69), C (60-64), F (Below 60).

Award of the Degree

The Scientific Council of the National University of Sudan, based on the recommendation of the Board of the Faculty of Graduate Studies and Scientific Research, shall award the successful candidate:

M.Sc. in Bioinformatics



NU.EDU.SD



Better Education
Better **World**

www.nu.edu.sd