Powered by **GENOMAC HUB**

*We envision inspiring and empowering life scientists to leverage **GENOMICS AND BIOINFORMATICS** to tackle critical challenges, drive innovation, and promote sustainable progress globally.*

Research Domain: Therapeutics R&D (Multi-Omics)
Research Focus: Anti-Malaria R&D
Research Topic: Investigating Microbial and Plant-Derived Peptides for Malaria Treatment: A Multi-Omics and Machine Learning Approach to Novel Drug Development
Research Aim: To systematically investigate and optimize peptides derived from microbial and plant sources for the development of targeted and effective anti-malarial drugs, addressing the urgent need for novel strategies against malaria and drug-resistant strains.
Research Objectives:

- Evaluate the efficacy of microbial and plant-derived peptides through multi-omics analyses to understand their molecular mechanisms and identify potential targets for anti-malarial drug development.
- Implement machine learning algorithms to analyze large-scale biological data and predict optimal combinations of peptides, facilitating the design and optimization of novel anti-malarial drugs with enhanced efficacy and reduced risk of drug resistance.

LEARNING OBJECTIVES

- **Comprehensive Understanding of Microbial and Plant-Derived Peptides:** Acquire in-depth knowledge of the structural and functional characteristics of peptides derived from microbial and plant sources, including their mechanisms of action and potential interactions with the malaria parasite.
- **Proficiency in Multi-Omics Techniques:** Develop the skills to employ multi-omics approaches, including genomics, proteomics, and metabolomics, to systematically analyze the biological pathways and molecular interactions associated with the selected peptides, providing a holistic understanding of their potential as anti-malarial agents.

- **Application of Machine Learning in Drug Development:** Gain practical expertise in utilizing machine learning algorithms for the analysis of complex biological datasets, enabling the identification of patterns, correlations, and predictive models to guide the rational design and optimization of anti-malarial peptides.
- **In Silico Analysis and Docking Techniques**: Gain proficiency in in silico analyses of designed peptides, including the prediction of phytochemical properties, structure, and function. Develop skills in molecular docking techniques, exploring various software tools for drug target identification, ligand selection, and evaluation of results.
- **Craft Research Papers for Publication:** Learn how to synthesize and present your findings coherently, culminating in the preparation of research papers suitable for publication, contributing to the broader understanding of the therapeutic research and development of anti-malaria peptides.

EXPECTATIONS WHILE UNDERTAKING THIS INTERNSHIP PROGRAM:

- **Knowledge of Genomics and Bioinformatics:** Develop a solid foundation in genomics and bioinformatics, including an understanding of key concepts, methodologies, and technologies used in the program
- **Proficiency in Data Analysis:** Gain proficiency in analyzing genomic data using bioinformatics tools and software. This includes skills in data preprocessing, quality control, data visualization, and statistical analysis.
- **Research Skills:** Acquire research skills necessary for conducting genomics and bioinformatics studies. This includes formulating research questions, designing experiments, collecting and analyzing data, and interpreting research findings.
- **Critical Thinking and Problem-Solving:** Develop critical thinking skills to analyze complex genomic and bioinformatics problems and propose creative solutions. You would be able to evaluate scientific literature, identify research gaps, and contribute to the advancement of knowledge in the field.
- **Computational Skills:** Gain proficiency in software and applications commonly used in bioinformatics, such as Geneious software, web servers etc. to analyze genomics data and interpret results
- **Communication Skills:** You would be able to effectively communicate your research findings and scientific concepts to both technical and non-technical audiences. This includes writing scientific reports, presenting research orally, and participating in scientific discussions and collaborations.
- **Collaboration and Teamwork:** Be able to develop skills in collaborating with peers and professionals in multidisciplinary research teams. This includes effective communication, teamwork, and the ability to contribute constructively to group projects.
- **Professional Development:** You would be able to develop a professional mindset, including skills in time management, organization, and project management. They should also be aware of current trends and advancements in genomics and bioinformatics, and actively seek opportunities for professional growth and development.
- **Publication and Dissemination:** Contribute to the scientific community by publishing their research findings in peer-reviewed journals

PROGRAM OUTLINE AND SCHEDULE

CLASSES	TOPICS/FOCUS	SCHEDULE & DELIVERABLES
General Classes	Overview of genomics, bioinformatics, and their applications in various fields	**WEEK 1**
	Understanding the central dogma of molecular biology	
	Introduction to genomics technologies and data generation	
	Data formats in Genomics and Bioinformatics (Practical)	
	Internet tools and Databases (Practical on data retrieval, Blast etc.)	
	Introduction to software tools and their installation, web servers, and pipeline tools (Practical), Basic Linux Command Line Interface	
	Genomics Data and its Analysis using cutting-edge tools (Practical DNA, RNA and Protein samples)	
Specialized Classes	Introduction to Therapeutic Peptide R&D **(Anti-Malaria)**	**WEEK 2**
	The experimental application of each of these in your field of study	
	Problem identification relative to the above area in the healthcare, industrial, and other life science research space	
	The use of critical thinking and problem-solving tools to design a hypothesis in solving identified problems	
PRACTICAL SESSIONS		
GENOMICS AND BIOINFORMATICS ANALYSIS		
PHASE ONE		
Data Collection and Preprocessing	**Raw Data Retrieval:** Gather diverse raw data samples, encompassing microbial metagenomics and other pertinent natural sources, to initiate comprehensive genomics and bioinformatics analyses crucial for the discovery and optimization of anti-malaria peptides.	
	Table 1: Construction of General Sequence Properties: via data table based on genome information which includes accession number, raw data size, sources, geographical regions platform, genome type, layout, file types, etc.	**Deliverable:** (Materials and Methods)
	Quality Control: Assess data quality, perform trimming, and filter out low-quality reads to ensure reliable results. **Genome Assembly:** Assemble the whole genome sequence of the diverse data samples recovered from different data sources.	**Deliverable:** (Results)
	Write Up 1 and 2: Reads Processing and Genome Assembly	
Comprehensive Genome Analysis	**Functional Genome Annotation:** Gene prediction, Protein features, Specialty features, Chromosomal properties, and Circus-view, among others.	**WEEK 3**
	Write Up 3: Functional Genome Annotation/Protein Identification	

		Deliverable: (Materials and Methods)
	Table 2: Construction of Chromosomal Genome Properties: CDS, Genes, RNA, Hypothetical Protein, Functional Protein, Go assignments, PGfam, Cripsr, etc.	**Deliverable:** (Results)
	PHASE TWO	
	Prediction of Biosynthetic Genes/Secondary Metabolites Using Bagel4 or Antismash: Identify potential biosynthetic genes and secondary metabolites in microbial, plant, or animal assemble sequence data relevant to malaria targets • Utilize genomics and a bioinformatics tool called Antismash for predictive analysis. • Apply these tools to diverse datasets to identify potential bioactive compounds that may serve as targets for anti-malaria Conduct functional analysis to understand the potential roles and functions of the identified genes and metabolites in the context of anti-malaria therapeutic peptide discovery	**Deliverable:** (Results)
	Write Up 4: Biosynthetic Gene Prediction or Prediction of Secondary Metabolites	**Deliverable:** (Materials and Methods)
Genomics and Predictive Metabolite Analysis (Identification of Biomarkers)	**Post-Prediction Result Analysis and Documentation** **Table 4:** Table Construction for Predicted Biosynthetic Genes/Secondary Metabolites that includes the following • Relevant features such as gene names, associated pathways, predicted metabolite structures, and any relevant scores or probabilities • Metadata details in the table, such as the source organism, sample type (microbial, plant, or animal), and specific datasets used • **Figure 1:** Figure showing predicted biosynthetic genes/secondary metabolites with relevant figures.	**WEEK 4**
	PHASE THREE	**WEEK 5**
	Machine Learning-Based Therapeutic Peptide Prediction and Design: Each step in the machine learning-based therapeutic peptide prediction and design process contributes to the specificity, diversity, and customization of the designed peptides, fostering a more efficient and targeted drug development approach.	
	STAGE 1: AMP PREDICTION **CAMPR3 for AMP Collection:** CAMPR3 serves as a specialized machine-learning model for collecting of AMPs from biosynthetic genes, enhancing the specificity of peptide selection for further analysis • **Prediction of One 99-Length AMP:** Predict a single 99-length AMP with the highest probability across all models • **Breakdown to Twenty 20-length AMPs:** Break down the prediction to identify twenty 20-length AMPs with the highest probability across the majority of models.	**Deliverable:** (Materials and Methods)

	NOTE: Predicting one 99-length AMP offers insights into the characteristics of longer peptides while breaking down twenty 20-length AMPs ensures diversity and practicality for experimental validation.	
	Write Up 5: Machine Learning-Based Therapeutic Peptide Prediction and Design	
	Table 5: Containing the predicted AMPs for each category (99 and 20)	
PREDICTION AND DESIGNING OF THERAPEUTIC PEPTIDES WITH MACHINE LEARNING	**STAGE 2: ACTIVITY-BASED PEPTIDE RECONSTRUCTION AND FUNCTIONAL PREDICTION** • Reconstructed predicted AMPs from CAMPR3, emphasizing antimalarial activity, to generate novel peptides with enhanced therapeutic potential • Prioritized parameters such as SVM scores, N & C terminus preferences, and other details during the redesign process.	**WEEK 6**
	Write Up 6: Activity-based Peptide Reconstruction and Functional Prediction	**Deliverable:** (Materials and Methods)
	Table 6: Contained the reconstructed peptides and parameters of the activities of each peptide	**Deliverable:** (Results)
	PHASE FOUR	**WEEK 7**
In Silico Characterization of Designed Therapeutic Peptides	**Phytochemical properties of designed therapeutic Peptides, in silico prediction of structure & function of the peptide.** • Hydrophobicity (numbers or ratio) • Number of G and p • Negative or Positive Net charge • Molecular formula • Boman Index • **Experimental Verification of peptide using BLAST** • Estimated half-life • **Predicted peptide toxicity** • Cell penetration and others • Instability Index, Aliphatic index, Grand average of hydropathicity (GRAVY)	**Deliverable:** (Results)
	Write Up: Phytochemical Properties and In Silico Prediction of Peptide Structure and Functional	**Deliverable:** (Materials and Methods) **Deliverable:** (Results)
	Visualization of all Peptide/Phytochemical Results (Tables and Figures for All or Most of the Phytochemical Properties)	
	PHASE FIVE	
Computational Drug Design/Molecular Docking	**The Application of drug design and discovering therapeutic peptide Molecular Docking:** • Drug target identification and retrieval • Ligand selection of leads hits, and retrieval • Pharmacokinetics analysis and Lipinski rule of 5 • Exploring software used for docking, their peculiarities, similarities, and differences • Protein preparation and ligand preparation process	**WEEK 8 & 9**

	• Exploring docking types based on protein structure: rigid-rigid, rigid-flexible, protein-ligand, protein-protein, protein-DNA • Docking analysis using different software like Pymol, Ds visualizer, etc.	
	Write Up: Drug Design and Molecular Docking	**Deliverable:** (Materials and Methods)
	Evaluation and Visualization of Results	**Deliverable:** (Results)

RESEARCH PROJECT OUTLINE FOR PUBLICATION	
Introduction	**BTW WEEK 6 -7**
Materials and Methods	**WEEK 9**
Results	**WEEK 10 and 11**
Discussion	**WEEK 12**
Conclusion	**WEEK 12**
References	**WEEK 13**
Certification and Recommendation Letter	
Follow up and Publication	