USAMA RIAZ

M2 Bioinformatics | Genomics • Workflow Development • Data Science

Paris, France | +33 7 45 27 55 90 | usamariaz145@gmail.com | linkedin.com/in/usamariaz145

RESEARCH EXPERIENCE

Bioinformatics Research Assistant (Stagiaire) — Phagos (Biotech Company) **Jan 2026 – July 2026**

Paris, France

- **SaPhari:** Developed a two-layer phage-host susceptibility prediction framework combining phylogenetic Tree-KNN/IDW with Random Forest and XGBoost, integrating bacterial defence systems, serotype and RBP similarity to rank candidate phages for wet-lab testing. Implemented model evaluation, hyperparameter tuning, diagnostics and prediction workflows for untested strain-phage pairs; tuned Random Forest improved over the IDW baseline (**F1 0.682; AUC-ROC 0.880**).
- **BACPHLIP:** Updated and retrained the phage lifestyle prediction model using **Phagos-curated phage data** to classify phages as **virulent or temperate**, expanding the HMM/domain feature set from 206 to 273 and achieving **97.9% test accuracy and 0.981 F1 score**.

Research Intern — CEA-Genoscope & Université Paris-Saclay **May 2025 – Jul 2025**

Evry, France

Project: The Dynamics of Microbial Community Assembly in Wastewater: Temporal Analysis and Interaction Networks

- Analyzed temporal microbial-community data to investigate ecological succession patterns and interspecies interactions.
- Constructed microbial co-occurrence and interaction networks and integrated sequence-analysis workflows with statistical modelling to investigate environmental drivers.

Research Intern — Ayub Agricultural Research Institute (AARI) **2021 – 2023**

Faisalabad, Pakistan

- Oilseed Research Institute (Aug 2022 – Feb 2023): analyzed crop genetic and agronomic datasets using preprocessing, statistical analysis and bioinformatics approaches to support genetic-diversity and crop-improvement research.
- Vegetable Research Institute (Sep 2021 – Feb 2022): supported experimental design, computational analysis, data handling and visualization for genomic and agronomic research.

EDUCATION

Master's Degree in Bioinformatics — GENIOMHE-AI — Université Paris-Saclay **Sep 2025 – Expected Aug 2027**

Évry, France

Status: M1 completed | M2 in progress (2026–2027)

Relevant coursework: Genomics & NGS Analysis, Machine Learning & Deep Learning, Statistical Genomics, Bioinformatics Algorithms, Databases, Systems Biology, Advanced Functional Genomics, Comparative Genomics & Metagenomics.

Bachelor's (BS) in Bioinformatics — Government College University Faisalabad **Sep 2019 – Aug 2023**

Faisalabad, Pakistan

Relevant coursework: Genomics, Genome Annotation, GWAS, Structural Genomics, Crop Genetics, Data Analysis & Visualization, Computational Algorithms for Biology and NGS.

RESEARCH PROJECTS

Data-driven Approach to Multi-factorial Etiology of Epilepsy **Jan 2023 – Aug 2023**

- Integrated biological datasets to investigate genetic and environmental contributions to epilepsy using computational modelling and statistical correlation.

GWAS of the SBP-Box Gene Family in Cajanus cajan (Pigeon Pea) **Mar 2022 – Jul 2022**

- Processed genomic data, performed population-structure and linkage-disequilibrium analyses, and identified genetic associations relevant to crop improvement.

RESEARCH & TECHNICAL SKILLS

Programming & workflows: Python, C (basic), R (Bioconductor), Bash/Shell, SQL/PostgreSQL, Nextflow and object-oriented programming.

Genomics & Microbial Bioinformatics: NGS analysis, 16S rRNA amplicon analysis, PICRUSt2-SC functional inference, phylogenetics, comparative genomics, genome/gene/protein annotation, GWAS, population structure, linkage disequilibrium

Phage Bioinformatics: Phage-host interaction modelling, host-range prediction, phylogenetic distance modelling, RBP similarity analysis, phage lifestyle prediction, HMM/domain feature analysis, FASTANI/ANI genome-similarity analysis

Data science & machine learning: NumPy, Pandas, scikit-learn, XGBoost, PyTorch, TensorFlow; Random Forest, classification, regression, PCA, clustering, supervised/unsupervised learning, hyperparameter tuning, feature selection and regularization

Tools & environments: Linux/Unix, Bash, Git, GitHub, GitLab, Docker, Jupyter Notebook and RStudio.

Visualization: Matplotlib, Seaborn and Plotly.

SCIENTIFIC ENGAGEMENT & CONFERENCES

Judge — iGEM Grand Jamboree 2026 **13 – 16 Nov 2026**

Paris, France

Selected to serve as a judge at the iGEM Grand Jamboree, an international synthetic biology event.

Presenter — Paris-Saclay Junior Congress

Paris, France **28 May 2026**

Individual video presentation (V8): temporal dynamics of the core antibiotic resistome in a French wastewater treatment plant; a 236-day study using PICRUSt2-SC functional inference.

Poster Co-author — RECOMB 2026 Conference

Greece

SaPhari: Predicting Phage-Host Efficacy through Phylogenetic Distance (Phagos).

LANGUAGES

English — Fluent | French — Beginner | Urdu — Native

REFERENCES

Prof. Abdelghani SGHIR
Professor, Department of Biology — Université d'Évry Paris-Saclay • Institut de Biologie François Jacob • CEA-Genoscope, France
sghir@genoscope.cns.fr

Andrew Holtz
Head of Bioinformatics — Phagos
andrew@phagos.org

Andras Asboth
Bioinformatics Infrastructure Engineer — Phagos
andris@phagos.org