Uwaise Ibna Islam

PhD Student in Computer Science (Bioinformatics)
954-666-2083 | uwaiseibna03@gmail.com | Google Scholar | GitHub

Research Summary

PhD candidate specializing in scalable computational genomics with focus on population-scale haplotype analysis and algorithm development. Proven track record developing high-performance bioinformatics tools (PBWT indexing achieving 305% speedup), publishing in top-tier venues, and building production-ready software. Expertise spans sequence alignment algorithms, variant analysis pipelines, machine learning for genomic data, and cloud-based HPC workflows. Seeking bioinformatics internship to apply computational skills to real-world genomic challenges.

Education

University of Florida | PhD in Computer Science (Bioinformatics), Expected May 2028

- Research focus: Scalable algorithms for population genomics, PBWT-based haplotype matching, variant analysis
- Advisor: Dr. Christina Boucher | GPA: 4.00

Florida Atlantic University | MS in Computer Science, GPA: 3.97, Aug 2022 – May 2024

- Research: Cross-species genotype-phenotype association prediction using deep learning
- Presidential Fellowship recipient; Academic Excellence Award (2023, 2024)

Chittagong University of Engineering & Technology | BS in Computer Science, Feb 2016 – Jun 2021

Technical Skills

- **Genomics & Variant Analysis**: PBWT/PBML algorithms, Haplotype phasing, Variant calling (GATK, BCFtools), Population genetics (PLINK, VCFtools), Sequence alignment (BWA, SAMtools), Phylogenetics, SMEM computation
- **Programming**: C++ (primary), Python, R, Bash/Shell, SQL; Experience with performance optimization, multi-threading, parallel computing
- **Machine Learning**: PyTorch, TensorFlow, Keras, Scikit-learn; Neural networks, CNNs, supervised/unsupervised learning
- **Data Structures & Libraries**: SDSL (succinct data structures), HTSlib, run-length encoding, LCP/LCS arrays, sparse indexing
- **Infrastructure & DevOps**: HPC clusters (SLURM), Docker, Git/GitHub, Linux/UNIX, GCP, R/Bioconductor package development
- **Bioinformatics Workflows**: SHAPEIT, BEDtools, Detectron2, DensePose, .bcf/.vcf processing, genomic benchmarking

Research Experience

Graduate Research Assistant | University of Florida, Boucher Lab *Jul 2025 – Present*

- Designed and implemented PBML, a novel algorithm for scalable PBWT indexing enabling efficient haplotype matching in population-scale genomic datasets (1000 Genomes, BIG Initiative biobank).
- Integrated Boyer-Moore-Li skipping strategy with forward-backward PBWT traversal to accelerate SMEM (Set Maximal Exact Match) queries with configurable minimum length thresholds (L = 1,000–10,000).
- Engineered bidirectional PBWT implementation supporting LCP/LCS computation, run-based haplotype retrieval, and multi-threaded .bcf query processing using SDSL and HTSlib libraries.
- Achieved 38% faster index construction, 62% faster single-threaded queries, and 306% multi-threaded speedup on 1000 Genomes data (2,504 samples, 84.4M variants) compared to existing PBWT tools.
- Optimized long haplotype tract queries ($L \geq 10{,}000$): reduced runtime by 73% and memory by 20%, reporting 1.15M+ genome-wide matches with 95–99.8% query coverage.
- Validated on real-world biobank data (BIG Initiative): demonstrated 38% query time reduction and up to 93% speedup for long-range haplotype identification, critical for IBD detection and rare variant studies.

Graduate Research Assistant | Florida Atlantic University, Assis Group *Aug 2022 – May 2025*

- Developed deep learning framework predicting genotype-phenotype associations from multi-species genomic alignments, achieving 100% cross-validated accuracy across diverse phenotypes (dietary preferences, environmental adaptations).
- Authored R/Bioconductor package implementing scalable genomic analysis pipelines with parallel computing, supporting 59 species and integrating neural network models for evolutionary genomics research.
- Optimized large-scale sequence processing workflows, reducing runtime by 7x through algorithmic improvements and HPC utilization.
- Published methodology in *Briefings in Bioinformatics* (IF: 9.5); package adopted by collaborating research groups.

PROJECTS

High-Performance PBWT Toolkit (C++/SDSL) *C++/SDSL*

- Implemented run-length encoded PBWT with support for SMEM thresholds, efficient column construction, and backward/forward navigation.
- Benchmarked performance against existing PBWT libraries, achieving noticeable speedups in memory-sensitive tasks.

Multi-Species Genotype-Phenotype Association Models *Python, R, Machine Learning*

- Developed models leveraging incomplete multi-species alignments and evolutionary gaps to infer phenotype associations.
- Designed simulation pipelines to generate synthetic sequence-phenotype datasets for model validation.
- Achieved 100% cross-validated prediction accuracy on simulated and real datasets, highlighting model robustness.

Human Body-Weight Estimator | PyTorch, Detectron2, CNN, DensePose *June 2024*

- Designed a pipeline to estimate weight from 2400 real-world selfie images using DensePose for segmenting body regions.
- Trained a 7-layer CNN model to predict weights based on segmented density maps, achieving strong prediction performance.

Private Image Repository | Flask, Google Cloud, SQL *Fall 2023*

- Built a user-authenticated image repository using Flask and Google Cloud Run with CI/CD pipelines.
- Integrated SQL-based authentication, Google Secret Manager, and serverless deployment for scalability and security.
- Designed modular architecture enabling easy addition of new features and efficient handling of user requests, demonstrating strong software engineering practices.

PUBLICATIONS

- **Islam U.I.**, Cozzi D., Gagie T., Bonizzoni P., Prezza N., Rossi M., Scalable PBWT Queries with Minimum-Length SMEM Constraints, *Under review, RECOMB 2026*
- **Islam U.I.**, Campelo dos Santos AL, Kanjilal R, Assis R., Learning genotype-phenotype associations from gaps in multi-species sequence alignments, *Briefings in Bioinformatics*, 2025 [IF: 9.5]
- **Islam U.I.**, Sarker I.H., Haque E., Hoque M.M., Machine Learning Model for Predicting Individual Substance Abuse, *Annals of Data Science*, 2022

INDUSTRY EXPERIENCE

Software QA Engineer | Enosis Solutions *Aug 2021 – Jul 2022*

- Automated end-to-end testing workflows using Selenium and Python, improving QA coverage to 63% and reducing regression bugs by 40%.
- Performed JMeter-based performance testing on production APIs; identified bottlenecks leading to 82% improvement in response times.
- Collaborated with development teams on continuous integration pipelines, ensuring code quality in agile environment.

HONORS & CERTIFICATIONS

- **FAU Presidential Fellowship** (2022–2024) | **FAU Academic Excellence Award** (2023, 2024)
- IBM Data Science Professional Certificate | Linux Foundation Certified