Maliha Aziz

P (480) 735-1085

aziz.maliha@gmail.com

Highly analytical and vision-driven computational biologist with over a decade of experience applying AI/ML and multi-omics approaches to biomarker discovery, host-pathogen interactions, and translational research. Experienced in building infrastructure, algorithms, and collaborative teams that drive innovation at scale.

CORE SKILLS

- Bioinformatics & Computational Biology
- Multi-Omics Data Integration
- Machine Learning
- LLM (Ollama--llama3.2:3b, deepseek)
- DNA based LLM – DNABERT2
- Protein based LLM -- Evolutionary Scale Modeling (ESM)
- Predictive Analytics
- Biomarker Discovery
- Software & Pipeline Development (HPC, LIMS, APIs)

PROFESSIONAL EMPLOYMENT

Antibiotic Resistance Action Center, George Washington University   November 2012-Current
Washington, DC
Research Scientist

Technical Leadership & Diagnostics Development

- Designed high-throughput pipelines for genomic, metagenomic, microbiome analysis and molecular diagnostics including host-microbe qPCR panels and ML-driven marker prioritization.
- Applied AI/ML approaches to integrate multi-modal data (genomic, metagenomic, clinical) for biomarker discovery and host-pathogen dynamics.
- Developed qPCR assays for 15 probiotic species using unsupervised learning of unique core-genome targets from ~200 genomes.
- Identified secreted protein markers for 8 epidemiologically relevant infectious disease bacterial and fungal organisms to inform point-of-care test development.
- Used Elastic net regression, PCA, EFA, and network analysis to identify zoonotic E. coli using accessory element signatures from >8000 genomes.
- Led the technology efforts for COVID-19 surveillance for GWU Re-Opening Campus efforts, including design of a high-throughput LIMS and vendor coordination.
- Built infrastructure for custom databases (e.g., MDAP) and represented the center on institutional computing boards.
- Recruited and supervised bioinformatics staff and interns; managed interdisciplinary collaborations.

Translational Microbiome Research

- Developed a HPC based 16S rRNA microbiome processing pipeline that performed OTU based analysis. Later refactored the bash script to move away from OTU-based analysis to an ASV- based DADA2 method.
- Performed analytics on >30 longitudinal and/or cross-sectional nasal, gut and penile microbiome-based studies that included 100+ participants.

o Conducted comparative genomic analysis on 50+ publicly available genomes of the 5 Bacteria Associated with HIV Seroconversion, Inflammation and Immune Cells (BASIC) to identify key genes and functions that may play a role in this association.

Translational Genomics Research Institute (TGen),   10/08-10/12
Flagstaff, AZ
Bioinformatics Analyst

Technical Lead
o Developed the prototype version of a next generation data processing pipeline – NASP for microbial data. The pipeline aligns Illumina short-reads to reference genome and calls SNPs using publicly available tools. The final version has now been cited 251 times.
o Developed a canonical-Single Nucleotide Polymorphism (SNP) detection pipeline. The pipeline identifies SNPs characterizing a clade in a phylogenetic tree. The identified SNPs can be converted to diagnostic assays to be used in clinical settings.

Diagnostics Development
o Performed the in-silico analysis that led to the development of BactQuant assay using APIs, mysql database and set operations written in Java. The assay targets the 16S gene and provides expanded bacterial taxonomic detection, coverage and quantification.

ADDITIONAL EXPERIENCE

Research Assistant at Scientific Data Management Lab, Arizona State University, Tempe, Arizona
Software Engineer at Millennium Software Ltd, Karachi, Pakistan

SELECTED AI/ML PROJECTS

Probabilistic Host Attribution with Mobile Genetic Elements (2023–2025):
  Developed and validated a Bayesian latent class model integrating accessory genome features and sample similarity priors to infer zoonotic transmission routes of E. coli.

Deep Learning for Fungal Diagnostic Imaging (2025):
  Designed a CNN-based image classification model for black Aspergillus spp. with augmentation strategies to overcome limited sample sizes. Integrated growth metadata to evaluate performance improvements.

Predictive Modeling of Host-Specific E. coli Populations (2023–2024):
  Applied dimensionality reduction and feature engineering on >3000 E. coli genomes to detect accessory gene patterns associated with host species; models used for zoonotic source tracing.

Protein Language Model Embeddings for Host Attribution (2025):
        Explored how deep latent representations of protein sequences can complement traditional GWAS markers and guide integration into hybrid predictive frameworks. Extracted per-gene protein sequences from E. coli pangenomes and generated embeddings with ESM2 models; applied DBSCAN and Leiden clustering alongside cosine and Hausdorff similarity analyses. Identified embedding-derived features that highlight host-associated patterns across human, chicken, turkey, pork, and beef isolates, informing the development of combined statistical–LLM methodologies.

EDUCATION

PhD in Health Data science (In progress)   George Washington University
Master's Degree in Computational Biosciences   Arizona State University, Tempe, Arizona
Bachelor of Computer Science   National University of Computer and Emerging Sciences

– FAST, Karachi, Pakistan

AWARDS & FELLOWSHIPS

2024 - 2025   NSF Co-Design of Trustworthy AI Systems (DTAIS) fellowship (Award Number: 2125677)

KEY PUBLICATIONS (h-index=29)

Peer-reviewed Method Development Manuscripts

1. **Aziz, M*.**, Palmer, A.*, Iversen, S. *et al.* Design and validation of *Dolosigranulum pigrum* specific PCR primers using the bacterial core genome. Scientific Reports **13**, 6110 (2023). ** denotes equal contribution*

2. Li MB, Park DE, **Aziz M**. *et al.* Integrating Sample Similarities into Latent Class Models: A Tree-Structured Shrinkage Approach. Biometrics. *(2021)*

Peer-reviewed Environmental Health and Food Manuscripts

3. **Aziz M*,** Park DE*, Vanessa . *et al.* Zoonotic *Escherichia coli* and Urinary Tract Infections in Southern California. * *denotes equal contribution* (Accepted at mBio)

4. **Aziz M**, Davis GS, Park DE, Idris AH, Sariya S, Wang Y, Zerbonne S, Nordstrom L, Weaver B, Statham S, John son TJ, Campos J, Castro-Nallar E, Crandall KA, Wu Z, Liu CM, DeBiasi RL, Price LB. Pediatric urinary tract infections caused by poultry-associated *Escherichia coli*. Microbiology Spectrum (2024)

5. Park DE, **Aziz M**, Koch BJ, Roach K, Clabots C, Johnson JR, Price LB, Liu CM. Gut microbiome predictors of Escherichia coli sequence type 131 colonization and loss. eBioMedicine. (2024)

6. Liu C, **Aziz M**, Park D, Wu Z, Stegger M, Li M, Wang Y, Schmidlin K, Johnson TJ, Koch B, Hungate B, Nordstrom L, Gauld L, Weaver B, Rolland D, Statham S, Hall B, Sariya S, Davis G, Johnson JR, Keim P, Price LB. Using source-associated mobile genetic elements to identify zoonotic extraintestinal *E. coli* infections. One Health (2023)

FULL LIST OF PUBLICATIONS (Total = 76)

https://scholar.google.com/citations?hl=en&user=hzeHIYAAAAAJ&view_op=list_works&sortby=pubdate

PATENT

1. U.S. Provisional Application No US 63/430,607, Methods for detecting Dolosigranulum pigrum.

2. U.S. Provisional Application No. US 63/387,912, Methods of pathogen biomarker detection.

PRESENTATIONS

2025   Poster - **Maliha Aziz**, Daniel E. Park, Edward H. Sung, Yashan Wang, Zhenke Wu, Cindy M. Liu, and Lance B. Price. *Latent hosts: Probabilistic host attribution using mobile genetic elements through Bayesian latent class analysis.* GWU TRAILS conference "AI at Work: Building and Evaluating Trust". 2025

2024   Talk - **Maliha Aziz**, Daniel E. Park, Zhenke Wu, Cindy M. Liu and Lance B. Price. *A novel statistical-genomic approach for inferring host-origins of Escherichia coli.* International Plant and Animal Genome Conference (PAG) 2024