Email: tran.986@buckeyemail.osu.edu | Phone: (330) 361-1477 |
GitHub: https://github.com/tran986   LinkedIn:www.linkedin.com/in/nghi-nia-tran-160966ngk

Nia (Nghi) Tran | Ph.D. Candidate in Bioinformatics & Biostatistics

Summary: PhD-trained bioinformatician and biostatistician with expertise in high-dimensional omics, statistical modeling, and computational pipeline development. Experienced in cross-disciplinary collaboration with clinicians and bench scientists.

Education

Ohio State University
PhD in Microbiology   GPA: 3.9   2022 - 2027 (expected)
Bio-informatics and Bio-statistics
The College of Wooster
B.A. in Biochemistry and Molecular Biology   GPA: 3.5   2017 - 2021

Skills

- **Programming Language**: Python, R, Bash, Git, Snakemake, Docker, HTML, CSS
- **Statistical tests:** Bayesian, linear and mixed models, early stopping, normalization, linear algebra, permutation, simulation, and machine learning
- **Data Analysis:** Metagenomics, transcriptomics, multi-omics integration, metabolic modeling
- **Platforms:** Linux/UNIX, HPC clusters, RStudio, VSCode, Github
- **Soft Skills:** Collaboration, cross-disciplinary communication, curiosity-driven problem solving.

Research Experiences

Graduate Research Associate | The Bradley Lab, The Ohio State University   2022 - Present
- Develop bioinformatic pipelines to analyze genomic microbiome data
- Use remote clusters/HPC for data automation via Batch and Snakemake workflow
- Simulate data for hypothesis testings and method evaluations
- Use R or Python statistical and machine-learning packages for identification of associations/trends and visualization
- Collaborated with a multidisciplinary team of statisticians and bench scientists to validate computational predictions.

QA/QC Microbiology Intern | Ho Chi Minh Institute of Drug Quality Control   `   2020
- Conducted antibiotic resistance testing and microbial analysis following national Drug Quality Control protocols.
- Ensured data standardization and reproducibility, contributing to high-throughput testing workflows

Undergraduate Researcher | The West Lab, College of Wooster   2019 - 2021
- Conducted protein assays, mutagenesis, and spectroscopy to investigate impacts of active-site mutations in yeast proteins.
- Performed computational structure–function analysis to interpret biochemical results, co-authoring in publication.

Projects

1. Analyze Metagenomic Sequences in Colon Cancer Microbiome | https://github.com/tran986/colon_cancer_metaG
- Developed an optimized Python workflow for metagenomic analysis of disease-associated microbes and network
- The pipeline is reproducible across HPC and local setups with significant optimization in time and memory used

2. Identify Potential Risk Factors for Skin and Soft Tissue Infection | https://github.com/tran986/biostat-py-repo
- Applied Python packages for descriptive and inferential statistics to reveal 3 factors significantly linked to infection

3. Build PERMANOVA-based Tool for Pathway Enrichment | https://github.com/tran986/PERMANOVA-based-aggregation
- Applied PERMANOVA and PERMDISP on gene count across sample groups to ID enriched pathway in diabetes
- Built a pipeline to conduct z-score standardization, PCA, non-parametric PERMANOVA, GSEA (gene set enrichment analysis), and (ORA) Over-represented Analysis for gene count/gene expression associated with diseases
- Used machine learning and data simulation in R to validate the tool, which results in better sensitivity and lower false-positive rates than older approaches

4. Phylogenetic Regression and Gene–Function Prediction | https://github.com/pbradleylab/phylogenize
- Developed an R-based regression tool linking microbial gene families to evolutionary and functional traits across taxa. It resolves the non-independence problem among species in traditional linear regression and reduces false positive hits
- Develop a R-based package that combines data permutation, simulations, early stopping to fit a robust regression model and calculate empirical p-value. This approach optimizes computing resources, and significantly reduces false discovery rates/recall, resolves overfitting problems from outliers.

5. Meta-analytic Pipeline for Strain-specific Approach in Metabolic Diseases | https://github.com/tran986/Strain_analyis_pipeline
- Processed raw sequence data, applied quality controls and differential abundance analysis (generalized linear model) on species counts.
- Troubleshoot confounders, noise, multi-colinearity problems in data (using mixed effect model and adaptive shrinkage)
- Used Linux-based cluster to sequence mapping and identified SNPs substitution associated with type 2 diabetes
- Built scalable Snakemake pipelines integrating data across disease cohorts for strain-function/diseases inference.

Leadership Experiences

Center of Microbiome Science Graduate Student Union | Co-founder & Treasurer   2025-Present
- Brainstorm ideas and hold events to help students to network and strive in their scientific career.
- Plan for expenses and successfully request funds/reimbursement for multiple events on and off-campus
- Handle administration and legal paper works to operate the organization

Honors, Publications, Voluntary Works

- Tri-Beta – National Biology Honor Society, *Member*   2019-Present
- Magna Cum Laude - College of Wooster   2021
- Phi Kappa Phi - Honor Society at Ohio State University   2022-Present
- *"Broader than expected tolerance for substitution in the WCGPCK catalytic motif of yeast thioredoxin 2"* 2022
- ½ Marathon runner - Nationwide Children's Hospital, OhioHealth and OSU 4 Miller   2025-Present