Shan P Thomas

shanptom@gmail.com | +1 (567) 901-2394 | Toledo, OH, USA
shanptom.github.io | linkedin.com/in/shanptom | github.com/shanptom

Professional Summary

Computational Biologist & Bioinformatics Engineer

Pipeline builder and statistical methods developer with 8+ years of experience spanning computational genomics, clinical NGS operations, and microbial ecology. I combine production-grade pipeline engineering in Nextflow and R/Python with hands-on sequencing expertise across six Illumina instruments in CAP-accredited labs—offering a rare, full-stack perspective that seamlessly bridges the gap between wet-lab reality and dry-lab analysis. Seeking a bioinformatics scientist or computational biology role in European biotech, diagnostics, or research.

Technical Skills

Pipeline Development: Nextflow, Snakemake, Bash scripting, workflow orchestration on HPC/SLURM
Programming & Analysis: R, Python, BASH, Empirical Bayes methods, XGBoost
Genomics Tools: QIIME2, DADA2, variant calling workflows, multi-omics integration, ecological network analysis
Infrastructure: Docker, Singularity, AWS, GCP, VPS, HPC, SLURM, Git/GitHub
Platforms: Illumina (MiSeq, NextSeq 550, HiSeq 2500/4000/X), Oxford Nanopore, MGI, Sanger

Deployed Software & Tools

MicrobeStudio — Interactive web application for microbial community analysis: taxonomic visualization, diversity metrics, and XGBoost-based indicator species discovery. App

CodeSparkR — R package integrating 400+ LLMs via OpenRouter for AI-powered code assistance in bioinformatics workflows. GitHub

PolishGBIF — Cloud-deployed (Docker + GCP Cloud Run) Shiny dashboard for biodiversity data exploration and interactive visualization. App

Experience

Computational Research Lead (Graduate Research Assistant)

University of Toledo · Toledo, USA · Jan 2022 – Present

- Designed and deployed lab-wide Nextflow and Snakemake pipelines on SLURM HPC clusters for 16S/18S metabarcoding and shotgun metagenomics, **reducing processing time by 60%** across 5+ concurrent projects.

- Developing an Empirical Bayes statistical framework to distinguish authentic ancient DNA from modern contamination — directly applicable to false-positive reduction in pathogen detection and biosurveillance.
- Built ecological co-occurrence network analyses and multi-omics integration workflows (genomic, geochemical, environmental) for US Government-funded coastal microbiome research.
- Created MicrobeStudio (R Shiny + XGBoost), an interactive web application for indicator species discovery, adopted by collaborators across multiple institutions.

Technical Consultant — NGS Platforms

Specialized Scientific Solutions · Remote · Dec 2020 – May 2021

- Provided workflow optimization consultation for Illumina and MGI sequencing platforms to clinical and research clients.

Assistant Manager — NGS Operations

LifeCell International (CAP-accredited) · Chennai, India · Apr 2020 – Nov 2020

- Managed computational and laboratory workflows for WES, WGS, and HLA typing on Illumina NextSeq 550, coordinating secondary analysis pipelines and QC metrics.
- Ensured CAP compliance through LIMS integration and clinical-grade workflow documentation.

Research Associate — Clinical NGS

MedGenome Labs (CAP-accredited) · Bangalore & Cochin, India · Dec 2017 – Mar 2020

- Operated high-throughput clinical NGS across Illumina platforms (MiSeq, NextSeq 550, HiSeq 2500/4000/X), achieving **>95% first-pass success rate** on >5,000 clinical libraries. Supported variant calling and secondary analysis for diagnostic reporting.

Researcher — Environmental Metagenomics

Geobiotechnology Lab · Trichy, India · Nov 2015 – Nov 2017

- Led metabarcoding projects on paleoenvironmental DNA; developed early bioinformatics analysis pipelines. 2 first-author publications.

Senior Research Fellow

Indian Agricultural Research Institute · New Delhi, India · Dec 2013 – Oct 2015

- Molecular barcoding and Sanger sequencing for biodiversity projects.

Education

PhD in Biology (Expected 2026) — University of Toledo, OH, USA | GPA: 3.9 Focus: Microbial ecology, Bayesian statistical methods, computational pipeline development

M.Sc. in Biotechnology — Bharathidasan University, Trichy, India | 2012–2014

Selected Publications

1. **Thomas, S. P.**, et al. (2019). Legacy of a Pleistocene bacterial community. *Microbiological Research*, 226, 65–73. (IF: 6.1)
2. **Thomas, S. P.**, et al. (2018). Metabarcoding of PalEnDNA as an efficient tool to recover ancient bacterial diversity. *Geomicrobiology Journal*, 35(9), 798–803. (IF: 2.2)
3. Stetten, L., ..., **Thomas, S. P.**, et al. (2025). Biogeochemical controls on iron speciation across coastal soils. *Sci. Total Environ.* (IF: 8.2)
4. Machado-Silva, F., ..., **Thomas, S. P.**, et al. (2024). Short-term groundwater fluctuations drive subsurface redox variability. *Environ. Sci. Technol.*, 58(33). (IF: 11.4)

6 publications · 108 citations · h-index: 4

Presentations

- AquaEcOmics, France, March 2025 (Oral): Co-occurrence networks in aquatic microbial communities
- PAGES, Spain, May 2017 (Oral): Spatio-temporal dynamics of bacterial communities