TATHAGATA DEBNATH

PhD Candidate in Computer Science

Phone · Email · GitHub · LinkedIn · Google Scholar · ORCID · Website

SUMMARY Computational biologist designing fast algorithms for genomic data analysis. Published in IEEE TPAMI and IEEE/ACM TCBB (81 citations) with 2 CRAN R packages for circular genome clustering. Research spans RNA-seq alternative splicing, metagenomics, proteomics, and ML-based disease classification from microbiome profiles. Seeking bioinformatics internship for summer 2026.

PUBLICATIONS

- **T. Debnath**, M. Song. "Fast Optimal Circular Clustering and Applications on Round Genomes." *IEEE/ACM TCBB*, 18(6), 2021. [First author; CRAN: OptCirClust]
- **Y. Chen, T. Debnath**, A. Cai, M. Song. "Circular Silhouette and a Fast Algorithm." *IEEE TPAMI*, 45(11), 2023. [CRAN: CircularSilhouette]
- **H. Bendzus-Mendoza, A. Rodriguez, T. Debnath**, et al. "Radiation exposure induces genome-wide alternative splicing events in Aedes aegypti." *Scientific Reports*, 2025.
- **A.D. Lopez, T. Debnath**, M. Pinch, I. Hansen. "Phosphoproteomics Analyses of Aedes aegypti fat body reveals blood meal-induced Signaling." 2024.
- Plus 6 additional publications (10 total; 81+ citations) — full list at Google Scholar

SOFTWARE & TOOLS

- **OptCirClust** (CRAN) — $O(KN \log^2 N)$ circular clustering for DNA, RNA, and bacterial/mitochondrial/chloroplast genomes. CRAN
- **CircularSilhouette** (CRAN) — Linear-time silhouette for circular cluster quality, 3000 *times* faster than standard. Applied to solar cycle and genome segmentation. CRAN
- **Pohori** (in development) — Model-free alternative splicing detection for short-read and long-read RNA-seq
- **ancestSeq** (GitHub, C++) — Ancestral genomic sequence reconstruction from present-day genomes

RESEARCH PROJECTS

Gut Microbiome Disease Classifier *[16S rRNA, CLR, Metagenomics]*

- Classified Crohn's Disease, T2D, and Healthy (0.907 AUC) from 16S rRNA OTU profiles using CLR compositional transformation addressing zero-inflation and compositionality constraints.
- Identified biomarkers via SHAP: Faecalibacterium depletion in Crohn's, Roseburia enrichment in healthy gut, Ruminiclostridium elevation in T2D.
- Designed cross-platform validation (759 samples, 16S amplicon and shotgun WGS) characterizing batch effects between sequencing platforms.

Scientific Literature Retrieval System *[RAG, Hybrid Search, Bioinformatics]*

- Built a scientific literature retrieval system for nutrition research using hybrid search (FAISS + BM25 + RRF) over peer-reviewed papers.
- Generated cited answers with source document and page attribution, enabling traceable evidence-based Q&A over research corpora.
- Demonstrated practical tool for scientists to query domain-specific literature with machine-readable provenance.

TECHNICAL SKILLS

- **Languages:** R, Python, C/C++, SQL, Bash
- **Bioinformatics:** RNA-seq, Alternative Splicing (MATS), 16S rRNA, Shotgun Metagenomics, BLAST, Sequence Alignment, Phylogenetics, Bioconductor, curatedMetagenomicData
- **ML/Statistics:** scikit-learn, XGBoost, SHAP, CLR Transformation, Nested CV, Compositional Data Analysis
- **Tools:** CRAN Package Development, Docker, Git, LaTeX, FastAPI

EDUCATION

- **Ph.D. in Computer Science** (Expected 2026)
  New Mexico State University, Advisor: Dr. Joe Song
- **M.Tech in Computer Science and Engineering** (2020)
  Indian Institute of Technology (BHU), Varanasi