

Michael G. Nute

11 Pine Terrace W, Short Hills, NJ 07078 • 607-227-5694 (mobile) • mn56@rice.edu

Education

Ph.D., Statistics, University of Illinois at Urbana-Champaign

August 2019

Advisor: Prof. Tandy Warnow

Dissertation: Statistical Estimation Problems in Phylogenomics and Applications in Microbial Ecology

B.A., Mathematics, Economics, Cornell University

May 2005

Postdoctoral Experience

Treangen Lab, Department of Computer Science, Rice University

July 2020 - Present

Research Summary

Graduate and postdoctoral research focused on multiple sequence alignment (MSA) and phylogeny estimation, with applications particularly to metagenomics:

- Extending Bayesian algorithms for joint alignment-phylogeny estimation to scale effectively to input with over 1,000 sequences, and identifying performance differences between simulated and biological benchmarks.
- Improved methods for analysis of metagenomic data, including novel data visualization techniques and improved statistical methods for predicting phenotypic association based on metagenome.
- Identification of novel microbes in environmental samples through phylogenetic association.
- Algorithms for taxonomic-identification effective for long, high-error amplicon sequences.
- New methods to identify structural variations in microbial genomes that associate with host phenotype.

Awards & Honors

- Awarded postdoctoral fellowship by National Library of Medicine Training Program in Bioinformatics and Data Science through the Gulf Coast Consortia, 2020-2021
- Awarded doctoral fellowship by the CompGen initiative at UIUC, 2016-2017
- Research fellowship from Caterpillar Corp., 2014-2015
- Named to "List of Teachers Ranked as Excellent by Their Students," Fall 2012

Publications

(**) denotes sole first-author, (*) denotes co-first or equal contributor

- (**) M. Nute, K. Yarlagadda and R. Stumpf. "PICAN-PI: A Graphical Schema to Visualize Microbial Biodiversity" (in preparation)
- (**) M. Nute, T. Savidge, T. Treangen "Comparative Genomics of Hypervirulent *C. difficile* Ribotype 027 Reveals Unique Patterns of Mutation with Putative Functional Importance" (in preparation)
- N. Sapoval, A. Aghazadeh, D. Antunes, A. Balaji, R. Baraniuk, C.J. Barberan, et al. "A Practical Guide to Deep Learning across Computational Biology: Recent Progress, Current Limitations, and Future Perspectives" in revision at *Nature Communications*, November 2021
- B. Killie, A. Balaji, F. Sedlazeck, M. Nute, T. Treangen "Multiple Genome Alignment in the T2T Genome Era" submitted to *Genome Biology*, October 2021
- K. Curry, M. Nute, T. Treangen "It takes guts to learn: machine learning techniques for disease detection from the gut microbiome" (2021) *Emerging Topics in Life Sciences*. DOI: 10.1042/ETLS20210213
- Q. Wang, M. Nute, T. Treangen "Bakdrive: Identifying the Minimum Set of Bacterial Driver Species across Multiple Microbial Communities" (2021) *bioRxiv*, 10.1101/2021.09.24.461746
- S. Soriano, K. Curry, S. Sadrameli, Q. Wang, M. Nute, E. Reeves, R. Kabir, et al. "Alterations to the gut microbiome after sport-related concussion and subconcussive impacts in a collegiate football players cohort." (2021) *medRxiv*
- A. Balaji, N. Sapoval, C. Seto, R.A.L. Elworth, M. Nute, T. Savidge, S. Segarra, T. Treangen "KOMB: Graph-Based Characterization of Genome Dynamics in Microbial Communities" (2021) *bioRxiv*, 10.1101/2020.05.21.109587
- KD Curry, Q Wang, MG Nute, A Tyshaieva, E Reeves, S Soriano, et al. "Emu: Species-Level Microbial Community Profiling for Full-Length Nanopore 16S Reads" (2021) *bioRxiv*, 10.1101/2021.05.02.442339
- B Kille, Y Liu, N Sapoval, M Nute, L Rauchwerger, N Amato, TJ Treangen "Accelerating SARS-CoV-2 low frequency variant calling on ultra deep sequencing datasets" (2021) *20th IEEE International Workshop on High Performance Computational Biology*
- N. Shah, M. Nute, T. Warnow, and M. Pop. (2019) "Misunderstood parameter of NCBI BLAST impacts the correctness of bioinformatics workflows." *Bioinformatics*, 35(9),

- S. Roch, M. Nute, T. Warnow. (2019) “Long-branch attraction in species tree estimation: inconsistency of partitioned likelihood and topology-based summary methods” *Systematic Biology*, vol. 68, iss. 2
- (**) M. Nute, E. Saleh, and T. Warnow. (2018) “Evaluating Statistical Multiple Sequence Alignment in Comparison to Other Alignment Methods on Protein Data Sets” *Systematic Biology*, vol. 68, iss. 3
- (*) K. Amato, J. Sanders, S. J. Song, M. Nute, J. Metcalf, L. Thompson, J. Morton, A. Amir, V. McKenzie, G. Humphrey, G. Gogul, J. Gaffney, A. Baden, G. Britton, F. Cuozzo, A. Di Fiore, N. Dominy, T. Goldberg, A. Gomez, M. Kowaleski, R. Lewis, A. Link, M. Sauther, S. Tecot, B. White, K. Nelson, R. Stumpf, R. Knight, and S. Leigh. (2018) “Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut microbiomes” *ISME J*
- (**) M. Nute, E. Molloy, J. Chou, and T. Warnow. (2018) “The Performance of Coalescent-Based Species Tree Estimation Methods under Models of Missing Data.” *BMC Genomics*, 19(Suppl 5):286
- (**) M. Nute and T. Warnow (2016). “Scaling statistical multiple sequence alignment to large datasets” *BMC Genomics* 17 (Suppl 10): 764, special issue for RECOMB-CG
- (*) N. Nguyen, M. Nute, S. Mirarab, and T. Warnow (2016). “HIPPI: Highly accurate protein family classification with ensembles of HMMs” *BMC Genomics* 17 (Suppl 10):765, special issue for RECOMB-CG.
- J. Chou, A. Gupta, S. Yaduvanshi, R. Davidson, M. Nute, S. Mirarab, and T. Warnow (2015). “A comparative study of SVDquartets and other coalescent-based species tree estimation methods” *BMC Genomics*, 16(Suppl 10), S2, special issue for RECOMB-CG

Invited Talks

- “Benchmarking BAli-Phy on Biological vs. Simulated Data” presented at Phylogenomics Software Symposium on Large-Scale Multiple Sequence Alignment, Montpellier, France (Aug 2018)
- “SEPP and TIPP Tutorial” presented at 2018 & 2019 Strategies and Techniques for Analyzing Microbial Population Structure (STAMPS) at the Marine Biological Laboratory, Woods Hole, MA (Aug 2018, Aug 2019)
- “The Performance of Coalescent-Based Species Tree Estimation Methods under Models of Missing Data” presented at RECOMB satellite meeting on Comparative Genomics (Oct 2017).
- “HIPPI: Highly Accurate Protein Family Classification with Ensembles of HMMs” presented at 2017 Great Lakes Bioinformatics Conference (May 2017)
- “HIPPI: Highly Accurate Protein Family Classification with Ensembles of HMMs” presented at 2017 UIUC Coordinated Science Lab Student Conference (Feb 2017)
- “Scaling Statistical Multiple Sequence Alignment Methods to Large Datasets” *2016 Phylogenomics Symposium and Software School*, Austin, TX (Jun. 2016)
- “Large-scale Multiple Sequence Alignment with PASTA and UPP” Software tutorial for *2016 Phylogenomics Symposium and Software School* (Jun. 2016)
- “Large-scale Multiple Sequence Alignment with PASTA and UPP” Invited tutorial for Patricia Babbit’s lab, University of California at San Francisco (Jan. 2016)

Posters

- “Unique Patterns of Genetic Mutation in *C. difficile* Hypervirulent Strain RT027” presented at *31st Keck Annual Research Conference* (Houston, TX. October 2021)
- “Mining Clinical *C. difficile* Isolates for Genomic Markers of Infection Progression” presented at *National Library of Medicine Informatics Training Conference* (Virtual, June 2021)
- “Taxon Labelling for Unidentified OTUs with Visual Phylogenetic Placement” Poster presented at *Mid-Atlantic Microbiome Meetup*, (College Park, MD. Nov. 2016)
- “Taxon Labelling for Unidentified OTUs with Visual Phylogenetic Placement” Poster presented at *Mid-Atlantic Microbiome Meetup*, (College Park, MD. Nov. 2016)
- “Boosting Bayesian MCMC Multiple Sequence Alignment with Phylogeny-Aware Divide-and-Conquer” Poster presented at *Pacific Symposium on Biocomputing*, (Waimea, HI. Jan. 2016)

Prior Work Experience

Caterpillar, Champaign, IL

May 2014 – May 2015

The Hanover Insurance Group, Worcester, MA

March 2007 – August 2012

Monitor Group, Cambridge, MA

June 2005 – March 2007

Other:

- Coach, University of Illinois Rowing Team, Spring 2013-Fall 2014
- Operations Planner, Head of the Charles Regatta, October 2007-Present
- Github: MGNute