

NIDHI SHAH

READING, MA. Phone: (240)-708-9426, Email: nidhirshah1992@gmail.com

SUMMARY

Accomplished researcher and software engineer with more than 4 years of experience in developing algorithms and software for analyzing genomic and healthcare data. Expertise in computational analysis. Strong background in software engineering, quantitative science, genomics, and metagenomics. Strongly believes in sharing ideas, collaborative work, and communication.

EDUCATION

MAY 2021

PH.D. IN COMPUTER SCIENCE, UNIVERSITY OF MARYLAND – COLLEGE PARK, USA

- Developed algorithms and software for analyzing metagenomic datasets
- Dissertation work resulted in 6 journal and conference publications and 3 open-source software

DECEMBER 2018

M.S. IN COMPUTER SCIENCE, UNIVERSITY OF MARYLAND – COLLEGE PARK, USA

- **Computer Science:** Algorithms, Machine Learning, Natural Language Processing, Distributed Systems, Database, Privacy in big data, Network Analysis, Genomics, and Systems Biology

MAY 2014

B.TECH. IN INFORMATION TECHNOLOGY, VEERMATA JIJABAI TECHNOLOGICAL INSTITUTE, INDIA

- GPA 9.1 / 10 – 2nd rank in the class of 2014
- Tuition supported graciously by Narotam Sekhsaria Foundation and Maharashtra Government

PROFESSIONAL EXPERIENCE

University of Maryland – COLLEGE PARK, USA

SUMMER 2016 - PRESENT

Research assistant for Professor Mihai Pop

- Designed and developed software for taxonomic annotation and abundance estimation in metagenomic datasets
- Mentored high-school students
- Conducted software analysis hands-on tutorials
- Led and managed multi-institution collaborative project

Bioinformatics Exchange of Students and Teachers (BEST) Summer School – HEILIGKREUZTAL, GERMANY

JUNE 2017

Student Researcher

- Summary: BEST is a joint effort of the University of Tuebingen and the University of Maryland to provide a weeklong intensive collaboration between students and professors at both schools.
- Analyzed gene catalogs, their efficacy, and created best practices for designing new catalogs

Princeton University - Summer Program in Algorithmic and Combinatorial Thinking (PACT) –

JUNE 2015 - AUGUST 2015

PRINCETON, USA

Student and Mentor

- Attended advanced classes on approximation algorithms and wide range of computer science topics presented by invited speakers
- Presented lectures and mentored high school students studying discrete mathematics

Samsung Research Institute – BANGALORE, INDIA

JULY 2014-APRIL 2016

Software Development Engineer

Developed prototypes for Smart home projects

SKILLS

- **Software** – Experience in using **Python, R,** and **shell scripting** for NGS workflows and analytics using packages such as **scikit, scipy, pandas, numpy,** and **ggplot.** Software development using **C++** and **Java.** Version control using **git.** Experience with HPC environment such as **SLURM** and **PBS.**
- **Communication** – Strong written and oral communication skills demonstrated through research publications and over many oral presentations to scientific community. Excelled at cross-team communications in graduate school.
- **Leadership** – Demonstrated the ability to drive bioinformatics software development projects. Executed project plan and drove decision making to meet deadlines.
- **Management** – Exhibited good time management and prioritization skills to simultaneously manage multiple tasks.

SELECTED RESEARCH PROJECTS

Graph partitioning based taxonomic annotation

- ⇒ Designed and developed an outlier detection algorithm to identify the best alignments (database hits) for short DNA sequences, typically from 16S rRNA gene.
- ⇒ Developed a novel algorithm for taxonomic annotation that uses these significant outliers within database search results to cluster database sequences into partitions.
- ⇒ Improves precision and has a unique ability to relate sequences in a taxonomy-agnostic manner.

Publications:

- **Shah, N.**, Altschul, S.F. and Pop, M., 2018. Outlier detection in BLAST hits. *Algorithms for Molecular Biology*, 13(1), pp.1-9.
- **Shah, N.**, Meisel, J.S. and Pop, M., 2019. Embracing ambiguity in the taxonomic classification of microbiome sequencing data. *Frontiers in genetics*, 10, p.1022.

Code: <https://github.com/marbl/ATLAS> (Implemented in Python and bash)

Phylogeny guided taxonomic abundance profile estimation

- ⇒ Designed and created new database of marker genes by mining over 170,000 genomes from the RefSeq repository.
- ⇒ Build alignment-tree pairs and a new set of reference packages for the TIPP pipeline.
- ⇒ Developed TIPP2 with accurate abundance estimation.

Publications:

- **Shah, N.**, Nute, M.G., Warnow, T. and Pop, M., 2019. Misunderstood parameter of NCBI BLAST impacts the correctness of bioinformatics workflows. *Bioinformatics*, 35(9), pp.1613-1614.
- **Shah, N.** TIPP2: metagenomic taxonomic profiling using phylogenetic markers with Erin Molloy, Mihai Pop, and Tandy Warnow (accepted at *Bioinformatics*).

Code: https://github.com/shahnidhi/TIPP_reference_package and <https://github.com/smirarab/sepp/tree/tipp2>

Critical analysis of gene catalogs and its use in metagenomic analysis

- ⇒ Investigate how gene catalogs are constructed and used in metagenomic analyses and how it impacts results in downstream tasks.
- ⇒ Critically evaluate one such catalog, the integrated gene catalog (IGC), as a case study.
- ⇒ Propose a set of best practices for creating and using gene catalogs in microbiome studies.

Publication:

- **Shah, N.** A critical assessment of gene catalogs for metagenomic analyses with Seth Commichaux, Jay Ghurye, Alexander Stoppel, Jessica Goodheart, Guillermo Luque, Michael Cummings and Mihai Pop (under review at *Bioinformatics*).

Changepoint detection based metagenomic sequence clustering

- ⇒ Designed and developed an algorithm, Binnacle, for clustering metagenome sequences (contigs) into bins.
- ⇒ Developed tools to estimate span and coverage of graph scaffolds and identify locations of misjoins in scaffold using changepoint detection algorithm.

Publication:

- **Shah, N.** Binnacle: using scaffolds to improve the contiguity and quality of metagenomic bins with Harihara Subramaniam, Jacquelyn Meisel, and Mihai Pop (under review at *Frontiers in microbiology*).

Code: <https://github.com/marbl/binnacle> (Implemented in Python and bash)

HONORS AND AWARDS

- Deans Fellowship, Department of Computer Science, University of Maryland
- Awarded WiSE Grant fellowship to attend Grace Hopper Conference 2015
- Winner of Topcoder contest conducted at Samsung, India in 2014
- Awarded Navrotam Sekhsaria foundation fellowship, covering tuition fee for undergraduate education

TEACHING AND OUTREACH

- Organized and conducted hands-on “Taxonomic Identification workshop” at the Mid-atlantic microbiome meetup 2019
- Conducted guest lectures in CMSC423: Bioinformatic Algorithms, Databases, and Tools for Fall 2018 and Fall 2019 classes
- Teaching assistant at the STAMPS course (Strategies and Techniques for Analyzing Microbial Community Population Structures) at the Marine Biological Laboratory 2019
- Organized activities and setup a booth educating the public about “Microbiome and its role in human health” at the Maryland Day, 2019
- UMD CS student volunteer: Reviewed applications and supporting visit day and orientation for newly admitted students - April 2019
- Sub-reviewer for Workshop on Algorithms in Bioinformatics (WABI) – 2019