

Title: Scalable algorithms for rapidly advancing DNA sequencing technologies

Speaker: Dr. Jain is currently a postdoctoral fellow with Dr. Adam Phillippy in the Genome Informatics Section at the National Institutes of Health (NIH), USA.

Abstract:

Genomics continues to have an immense impact in life sciences, from elucidating fundamental genetic mechanisms to providing new keys to understand diseases. Catalyzed by breakthroughs in genomic technologies, high-throughput DNA sequencing has become a major generator of data, catching up with the most prominent big-data applications such as astronomy and social media. As a result, it has become challenging to design scalable DNA sequence algorithms with suitable accuracy guarantees.

In the first part of my talk, I will formulate the problem of aligning a biological sequence to a ‘pan-genome’ graph. The graph-based representation has been touted as the new standard for human genome reference, as it can succinctly capture human genetic variation across the globe. I will present new complexity results and algorithms for computing an optimal sequence alignment using such graphs. In the second part, I will present an approximate sketch-based algorithm for mapping long noisy sequences. The proposed algorithm is able to map sequences to the entire NCBI genome database (totaling ~1 Tbp) for the first time, enabling real-time genomic analysis. I will also discuss generalization of the proposed algorithms to a wider set of problems, and their impact in tackling long-standing biological questions.

Bio:

Dr. Jain is currently a postdoctoral fellow with Dr. Adam Phillippy in the Genome Informatics Section at the National Institutes of Health (NIH), USA. He got his doctorate in Computational Science at Georgia Tech under the supervision of Prof. Srinivas Aluru, and bachelors in Computer Science from IIT Delhi. He is a recipient of the Best GPU Paper Award at IEEE HiPC’14, Best Poster awards at RECOMB’19, CRIDC’19, IITD-OpenHouse’14, and Reproducibility-Initiative Award at ACM Supercomputing’16.