FACULTY MENTOR
Turakhia, Yatish

PROJECT TITLE
A GPU-accelerated library for aligning genomic long reads

PROJECT DESCRIPTION
"Turakhia Lab (http://turakhia.ucsd.edu/) is looking for exceptionally talented and highly motivated ECE undergraduate and master's students to work on the following high-impact research project at the intersection of computer systems and computational genomics during the Spring and Summer quarters as part of the ECE Spring/Summer Research Internship program (SRIP).

The volume of long read genome sequencing data has been growing at a rapid pace and its vast potential of discovering a broad spectrum of genomic mutations and generating a highly contiguous genome assembly is only starting to be realized. But assembling the genome using long reads is expensive and will be aggravated much further once we replace linear references with graph-based references consisting of a massive volume of haplotype data. In this project, the student(s) will work closely with the PI and other collaborators to:

1. Study and design new GPU-friendly algorithms and data structures for aligning long reads to linear and graph sequences.
2. Develop a user-friendly GPU library based on the above and compare the performance to previous works.
3. Publish and present results at top-tier conferences and workshops.

Students are expected to be skilled at C++ and CUDA programming and have a good understanding of data structures, algorithms and GPU hardware architecture. Spring quarter students may also earn ECE 199/299 credits for this work.

Relevant reading:
1. Turakhia et al. "Darwin: A Genomics Co-processor Provides up to 15,000X Acceleration on Long Read Assembly", (ASPLOS 2018). https://doi.org/10.1145/3173162.3173193
2. Goenka et al. "SegAlign: A Scalable GPU-Based Whole Genome Aligner", (SC 2020). https://doi.org/10.1109/SC41405.2020.00043
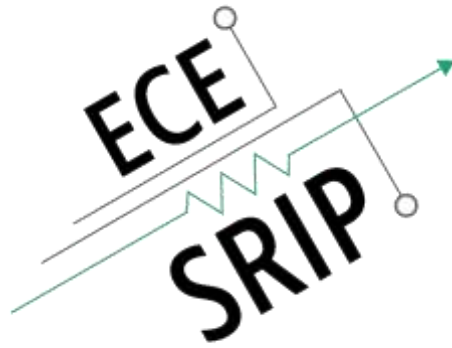
This project is in-person.

INTERNS NEEDED
2

PREREQUISITES
C++ programming, CUDA programming, Data structures and Algorithms, GPU architecture

FACULTY MENTOR
Turakhia, Yatish

PROJECT TITLE
Real-time SARS-CoV-2 phylogenetics

PROJECT DESCRIPTION
Turakhia Lab (http://turakhia.ucsd.edu/) is looking for exceptionally talented and highly motivated ECE undergraduate and master's students to work on the following high-impact research project at the intersection of computer systems and computational genomics during the Spring and Summer quarters as part of the ECE Spring/Summer Research Internship program (SRIP).

Scientists and public health practitioners have been sequencing the genome of the SARS-CoV-2 virus from the early onset of the COVID-19 pandemic last December and over 200K virus samples have been sequenced so far. This is unprecedented – we have never been able to track the evolution of a pathogen so closely in any previous outbreak and this has enabled a wide range of "genomic contact tracing" applications but has also pushed the current bioinformatics tools to beyond their limits. We're currently working with an international group of collaborators to enable phylogenetic analysis of SARS-CoV-2 to be performed in real time. In this project, the student(s) would contribute to the project as follows:

1.   Extend our current real-time phylogenetic placement tool, UShER [1], to handle maximum likelihood computations in addition to maximum parsimony computations.
2.   Design and implement optimization strategies to enable this heavy computation to meet real time constraints. The student will also explore the possibility of using existing GPU libraries to speed up this computation.
3.   Co-author papers based on this work to be published in top-tier journals.

Relevant reading:
1.   Turakhia et al. "Ultrafast Sample Placement on Existing Trees (UShER) Empowers Real-Time Phylogenetics for the SARS-CoV-2 Pandemic", bioRxiv 2020 https://doi.org/10.1101/2020.09.26.314971
2.   Turakhia et al. "Stability of the SARS-CoV-2 phylogenies", PLOS Genetics 2020 https://doi.org/10.1371/journal.pgen.1009175

This project is in-person.

INTERNS NEEDED
2

PREREQUISITES
C++ programming, Data structures and Algorithms