

FACULTY MENTOR

Yatish Turakhia

PROJECT TITLE

Hardware Acceleration of Computational Genomics

PROJECT DESCRIPTION

Description: Turakhia Lab (<http://turakhia.ucsd.edu/>) is looking for exceptionally talented and highly motivated ECE undergraduate and master's students to work on high-impact research projects at the intersection of computer systems and computational genomics. Genomic data is one of the fastest growing data types on the planet and is far outpacing Moore's law. From personalized medicine to species conservation, genomic data has far-reaching applications, but computational costs are posing ever greater challenges to exploit the full potential of this data. We are exploring novel algorithms and hardware (GPU/FPGA/ASIC) acceleration approaches to speed up computational genomics tasks, such as genome assembly, read alignment and whole-genome alignments, by orders of magnitude. In this project, the student(s) will work closely with the PI and other lab members and collaborators to:

1. Study and design new hardware-friendly algorithms and data structures for expensive computational genomics tasks.
2. Implement the design/algorithm and compare its performance to existing methods.
3. Publish and present results at top-tier conferences and workshops.

Students are expected to be skilled in C++ and CUDA/Verilog programming and have a good understanding of data structures, algorithms and GPU/FPGA architecture. Students who have taken ECE 284, ECE 208, CSE 208A or related courses at UCSD will be prioritized. Spring quarter students may also earn ECE 199/299 credits for this work. Relevant reading:

- Dally et al. "Domain-specific hardware accelerators", Communications of the ACM 2020, <https://dl.acm.org/doi/10.1145/3361682>.
- 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>
- Goenka et al. "SegAlign: A Scalable GPU-Based Whole Genome Aligner", (SC 2020). <https://doi.org/10.1109/SC41405.2020.00043>

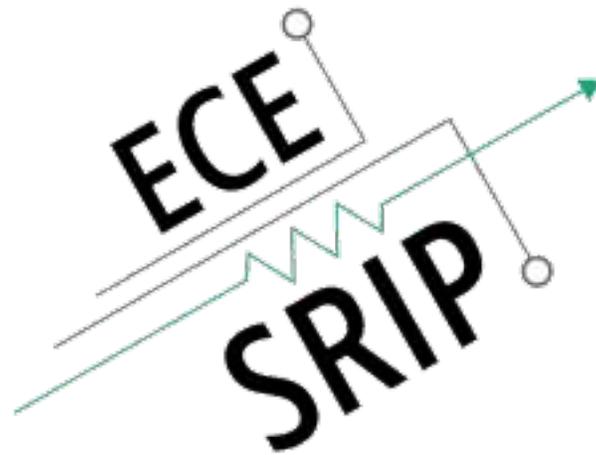
This project will be in person.

INTERNS NEEDED

2 Students

PREREQUISITES

1. C++ programming, CUDA/Verilog programming, Data structures and Algorithms, GPU/FPGA architecture



FACULTY MENTOR

Yatish Turakhia

PROJECT TITLE

Large-scale COVID-19 Phylogenetics

PROJECT DESCRIPTION

Description: Turakhia Lab (<http://turakhia.ucsd.edu/>) is looking for exceptionally talented and highly motivated ECE undergraduate and master's students to work on high-impact research projects on large-scale COVID-19 phylogenetics. With over 4 million (and counting) whole SARS-CoV-2 genomes sequences already, the COVID-19 pathogen is the most sequenced pathogen in history. Phylogenetic analysis using these genomes has played a vital role in tracking the virus evolution and transmission, but is posing major computational challenges. Our lab has been involved in maintaining and refining a comprehensive phylogenetic tree consisting of all SARS-CoV-2 genome available through online databases, and are now working on further improving our tools for speed and accuracy. In this project, the student(s) will work closely with the PI and other lab members and collaborators to:

1. Develop and improve existing tools for large-scale COVID-19 phylogenetics.
2. Compare developed tools with related tools for speed and accuracy.
3. Publish and present results at top-tier conferences and workshops.

Relevant reading:

1. Turakhia et al. "Ultrafast Sample Placement on Existing Trees (UShER) Empowers RealTime Phylogenetics for the SARS-CoV-2 Pandemic", Nature Genetics 2021, <https://doi.org/10.1038/s41588-021-00862-7>.
2. McBroome et al., "A Daily-Updated Database and Tools for Comprehensive SARS-CoV-2 Mutation-Annotated Trees", <https://doi.org/10.1093/molbev/msab264>.
3. Hodcroft et al., "Want to track pandemic variants faster? Fix the bioinformatics bottleneck", Nature 2020, <https://doi.org/10.1038/d41586-021-00525-x>.

This project will be in person.

INTERNS NEEDED

2 Students

PREREQUISITES

1. C++ programming, Data structures and Algorithms