Bachelor's / Master's / Semester Project

Exploring new algorithms and hardware architectures for Genomic Sequence Alignment

Our understanding of human genomes today is affected by the ability of modern computing technology to quickly and accurately determine an individual's entire genome. However, timely analysis of genomic data remains a challenge. One of the most fundamental computational steps in most bioinformatics analyses is *genomic sequence alignment.* The execution time of this step constitutes the **main performance bottleneck** in genomic data analysis.

Increasing the number of CPUs used for genome analysis decreases the overall analysis time, but significantly escalates the cost of building, maintaining, and cooling such computing cluster, as well as the power/energy consumed by the cluster. This is a critical shortcoming with respect to both energy production and environmental friendliness. Cloud computing platforms can be used as an alternative to distribute the work load, but transferring the data between the clinic and the cloud poses new privacy and legal concerns. Therefore, our goal is to build **new, fast, and accurate methods to quickly analyze such data locally**.

In our research group, we developed several efficient hardware architectures and algorithmic solutions to tackle this problem. You will work with other researchers to design and analyze new algorithms and ideas. You will also implement and evaluate these new algorithms using real genomic data.

Requirements
- Outstanding programming skills (C/C++)
- Good knowledge of design and analysis of algorithms and data structures
- An interest in developing and evaluating new ideas
- Strong work ethic

For **example studies** you may perform please see:
- "GateKeeper: A New Hardware Architecture for Accelerating Pre-Alignment in DNA Short Read Mapping", *Bioinformatics*, 2017.
- "Shouji: Fast and Efficient Computation of Banded Sequence Alignment", *arXiv preprint* 2018.
- "GRIM-Filter: Fast Seed Location Filtering in DNA Read Mapping Using Processing-in-Memory Technologies", *BMC Genomics*, 2018.
- "Shifted Hamming Distance: A Fast and Accurate SIMD-friendly Filter to Accelerate Alignment Verification in Read Mapping", *Bioinformatics* 2015.
- "Accelerating Read Mapping with FastHASH", *BMC Genomics* 2013.

If you are interested, please email: **Professor Onur Mutlu**, omutlu@gmail.com
and **Can Firtina,** canfirtina@gmail.com
https://people.inf.ethz.ch/omutlu/
https://safari.ethz.ch/work-with-us/projects/