

## **FACULTY MENTOR**

Mir Arabbaygi, Siavash

## **PROJECT TITLE**

New integrated algorithms for understanding microbiome

## **PROJECT DESCRIPTION**

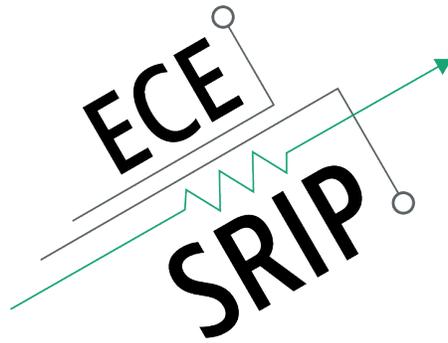
Microbiome is the collection of microorganisms that live in an environment, such as the human gut. Understanding gut microbiome is essential for many downstream applications in health. Essential in understanding microbiome is dissecting its composition, a task that requires modeling the evolutionary relationship between species. Our lab has in the past several years developed several tools for just that task. These tools add new microbiome data to reference datasets and find their best placement in the microbial tree of life. In this project, the student will integrate several of these past algorithms to develop an end-to-end pipeline. They will also work on optimizing the efficiency of the tools, and if interested, can develop methods to further improve their accuracy.

## **INTERNS NEEDED**

1 BS Student, 1 MS Student

## **PREREQUISITES**

- Strong in python, familiar with multi-threaded programming,
- A desire for dealing with big data (often using supercomputers),
- A course in data structures and algorithms.



## **FACULTY MENTOR**

Mir Arabbaygi, Siavash

## **PROJECT TITLE**

Solving the mystery of how human segmental duplications evolved

## **PROJECT DESCRIPTION**

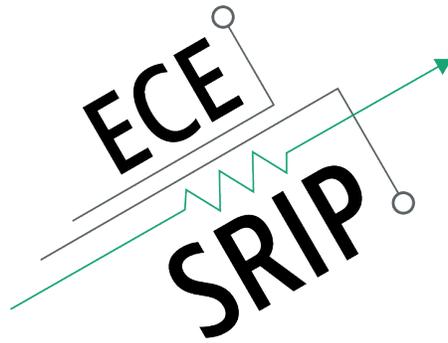
Human genome is known for being notably repetitive. Segmental duplications (SD) are one kind of these repetitive structures in DNA, and are known to be one of main differences that separates Human genome from other primates: Our genomes has the largest fraction of SDs among the sequenced primate genomes! Thanks to the recent advancements in genome sequencing technologies and better algorithms, we have now necessary tools to unveil the evolutionary history of SDs and potentially reveal intricate mechanisms which led to SDs populating our genomes. Now the only thing we need is new algorithms.

## **INTERNS NEEDED**

1 BS Student and 1 MS Student

## **PREREQUISITES**

- You like dealing with big data (often using supercomputers),
- You like discrete math and solving puzzles
- You are familiar with algorithms and graph theory,
- You can code in at least one language (not Matlab)



## **FACULTY MENTOR**

Mir Arabbaygi, Siavash

## **PROJECT TITLE**

Noise reduction techniques for inferring evolutionary trees from genomic data

## **PROJECT DESCRIPTION**

Reconstructing the evolutionary history of organisms (species trees) from genomic data is a long-standing goal of biologists and a prerequisite for many downstream applications. This computational task is often done in a two-step approach where first trees are built from individual parts of the genome, and then the trees are combined to infer a final species tree. The first inference, usually done using maximum likelihood in noisy and the noise propagates to the second step. Several techniques for noise reduction (e.g., Bayesian estimation, co-estimation, binning, etc.) exist, but a comprehensive study of these approaches do not exist. The student(s) will work with a Ph.D. student to study the accuracy and scalability of existing methods and/or to design, implement, and test new algorithms.

## **INTERNS NEEDED**

1 BS Student and 1 MS Student

## **PREREQUISITES**

- A desire for dealing with big data (often using supercomputers),
- Interest in discrete math, familiarity with algorithms, and graph theory,
- Ability to code, in at least one language, and to build data analysis pipelines (e.g., in python or R). Matlab is not enough.