FACULTY MENTOR
Siavash Mirarab

PROJECT TITLE
Deconvolving DNA Mixtures

PROJECT DESCRIPTION
Description: Modern study of ecology to quantify biodiversity and help with species conservation

efforts require analyzing DNA sequence data captured from mixtures of species. These mixture data can be modeled as strings of A, C, G, and Ts, with each string having length ~150 and coming from one of several species. The goal is to detect which species, populations, or even individuals are present in the environment where the sample was taken. These kinds of "metagenomic" analyses are routinely performed for microbiome. However, much less is known about algorithmic techniques to perform similar analyses on vertebrate genomes. In this project, we will test existing methods. developed for microbes to see if they can handle much larger vertebrate genomes. Anticipating the answer to be no, we will go on to either change existing methods or develop new ones that will make sure analyses possible.
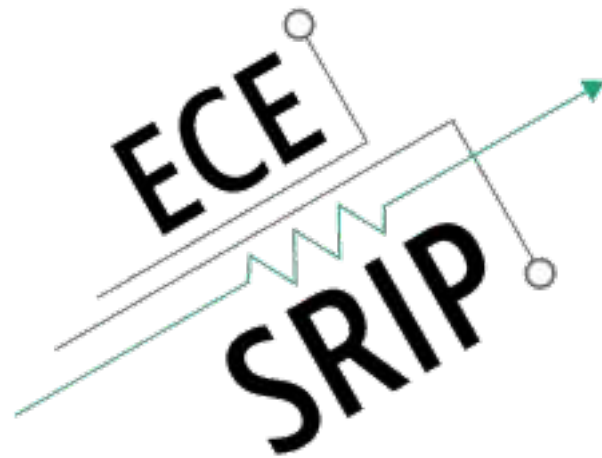
This project can accommodate both remote and in-person students

INTERNS NEEDED
2 Students

PREREQUISITES
1. Ability to program (ideally in Python)
2. Interest in working with large biological data
3. some familiarity with Linux and bash is a plus
4. No Background in Biology needed

FACULTY MENTOR
Siavash Mirarab

PROJECT TITLE
New Species Sampling Algorithms for Faster Statistical Inference

PROJECT DESCRIPTION
Description: Inferring evolutionary relationships between organisms is a fundamental question in biology, but is hampered by scalability of existing methods. There are methods that are accurate but not scalable, and methods that are scalable but not accurate. In this project, we will take a method developed in our lab that is accurate but not very scalable and will make it scalable. We will explore a sampling strategy, where, the base (slow) algorithm is applied on subsets of the data, sampled either randomly or using a targeted approach. The results from these smaller inference problems will be combined at the end. We will seek to find sampling strategies that minimize the number of samples while maximizing accuracy. We will think about the question from both practical and theoretical aspects.
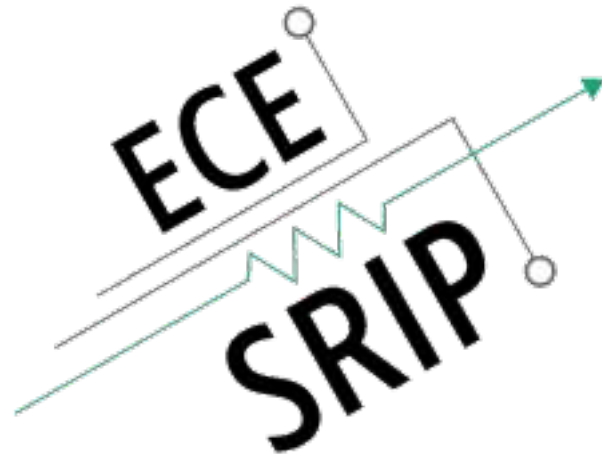
This project can accommodate both remote and in-person students

INTERNS NEEDED
1 Student

PREREQUISITES
1. Ability to program in some language (ideally in python and/or C++)
2. Interest in working with large biological data (no background in biology needed)
3. Some familiarity with Linux and bash is a plus

FACULTY MENTOR
Siavash Mirarab

PROJECT TITLE
Machine Learning Methods for Reconstructing Evolutionary Past

PROJECT DESCRIPTION
Description: Machine learning has revolutionized many areas of computational biology, but has made little inroads in reconstructing evolutionary past. Recently, our lab has pioneered some methods that start to make deep learning useful for this fundamental aspect of biology. However, machine learning, while competitive with existing methods, is still not better than them in terms of accuracy. In this project, we will explore ways to improve our recent deep learning method, DEPP, and to further apply it solve challenges that the traditional methods cannot address.

This project can accommodate both remote and in-person students

INTERNS NEEDED
2 Students

PREREQUISITES
1. Ability to program in python
2. Some familiarity with deep learning (the more the better)
3. Being comfortable working in a linux environment and handling large datasets.
4. No knowledge of biology is needed.