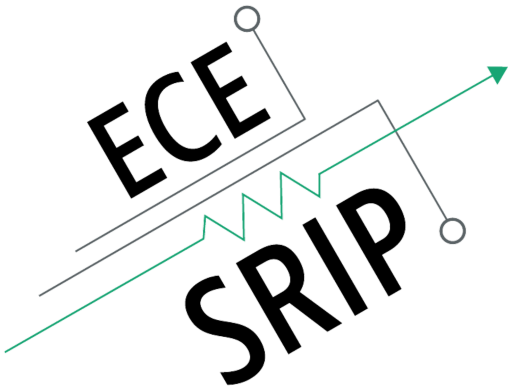


FACULTY MENTOR

Siavash Mirarab

PROJECT TITLE

Deep Learning for (Meta)-Genomic Analyses

PROJECT DESCRIPTION

In this project, we develop methods using deep learning for analyzing metagenomic data using evolutionary principles. The project will involve the design and implementation of new algorithms and testing them on extensive benchmarking datasets.

This project will be in person.

INTERNS NEEDED

2 Students

PREREQUISITES

- Familiarity with machine learning in general and python-based PyTorch platform in particular.

FACULTY MENTOR
Siavash Mirarab

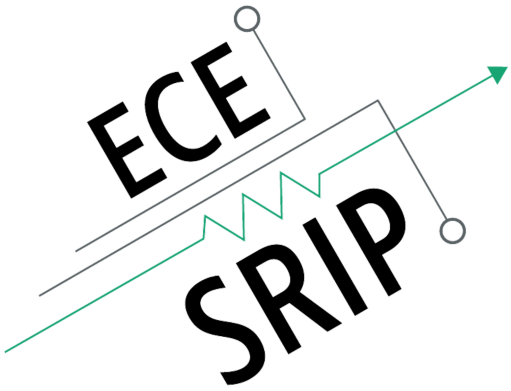
PROJECT TITLE
Statistical methods of phylodynamic reconstruction

PROJECT DESCRIPTION
In this project, we will build on our recent advances in building maximum likelihood methods (ML) for inferring the timing of evolutionary events. As biological entities, including viruses, evolve, they leave a trace of their evolution in their genome. However, figuring out at what *time* major evolutionary events happened (e.g., when SARS-Cov2 crossed into humans) requires sophisticated statistical models that combine various sources of information. We have recently created ML methods for such inferences. However, our methods lack many important features (e.g., uncertainty in input timing information). This project will build such methods.

This project can accommodate both remote and in-person students.

INTERNS NEEDED
2 Students

PREREQUISITES
- A better than basic understanding of statistical inference (e.g., graduate level or very advanced undergraduate level).
- Ability to understand and implement optimization-based methods and statistical inference algorithms.
- The interest/ability to test methods on data for validation



FACULTY MENTOR

Siavash Mirarab

PROJECT TITLE

Algorithms for analyzing biodiversity using genomic data

PROJECT DESCRIPTION

Biologists trying to understand the decline in biodiversity use genetic data. Typically, they use data from a specific gene (COI), which is easy to sequence. However, they get limited signals from a single gene. The cost has now dropped to the extent that we can try to use genome-wide data, which gives us many orders of magnitude more data and hence better signal. Analyzing these large data comes with its own set of challenges, some of which will be addressed here.

This project will be in person.

INTERNS NEEDED

2 Students

PREREQUISITES

- Comfortable with python programming.
- Some level of understanding of optimization.
- Comfort with UNIX and large-scale data analysis.