



University of  
New Hampshire

# Fall 2022 Environmental Sciences Seminar Series

Wednesday, October 5  
2:30-3:30pm  
James Hall, G46

## *Deep Learning for Population Genetic Data*



### **Dr. Sara Mathieson**

Assistant Professor  
Computer Science Department  
Haverford College

In the early 2010's, machine learning approaches for population genetic data began to emerge, and recently deep learning has been used to estimate detailed evolutionary parameters. Convolutional neural networks (CNNs) in particular have been used to infer recombination rates, natural selection, and population sizes without the need for summary statistics. All these methods rely on simulated training data, but it is often challenging to create realistic simulations for populations with unique evolutionary histories, such as *Anopheles* mosquitoes. In general, the role of simulated data in population genetics is difficult to overstate, as it is used for method validation and comparison. Here I will discuss a new generative adversarial network (GAN) method that automatically creates simulated data for any population. Our method, called pg-gan, works by training a parametric generator and CNN discriminator in concert, until there is a close match between real and simulated data. We apply pg-gan to study demography and selection in several human populations, as well as improve GAN training with a Wasserstein loss. Our approach has proven useful in evaluating and strengthening simulated data, especially for understudied populations that deviate from broad geographic groups.

Hosted by Joshua Tremblay & the Ecological Genomics Working Group. Series sponsored by the Natural Resources and Earth Systems Science (NRESS) Ph.D. Program, in partnership with the Earth Systems Research Center, and the Natural Resources and the Environment and Earth Sciences Departments.

**Free—All are welcome! (No RSVP necessary)**