Review: Deep Learning for Population Genetic Inference (Sheehan & Song)
*R*eviewed by: Kartik Sawhney, Patrick O'Grady, Rishab Mehra, Thomas Liu & Alexandra Bourdillon

Introduction
Population genetics seeks to bridge nuanced theories of evolution and natural selection. In the case of statistical inference, this field has made considerable strides in developing models for genetic variation and populations of alleles. However, precise algorithms for identifying evolutionary relationships and ancestral inference is computationally complex because of numerous demographic and biological variables.

Prior to this paper, existing literature in population genomics was yet fully explore the relationship between demography and natural selection. With the growing prevalence of large-scale genomic data and increasingly complex genetic models, this issue is particularly important since demography and selection often confound results by producing similar genomic markers. While related works have primarily focused on the versatility of selection estimates to differing demographic situations, this paper instead aims to examine demographic and selective factors jointly.

Methods
The Sheehan & Song model as described in this paper puts forth a novel approach to applying Deep Learning pipelines for the purpose of identifying selection and demographic features within a genome. Their initial model was developed on simulated data from a population of *Drosophila melanogaster* genomes from Zambia.

The first method they tried was the Approximate Bayesian Computation (ABC), which is a likelihood-free inference method. This study implemented a rejection-based approximate Bayesian inference by setting a tolerance threshold with which simulations had to meet in order to be incorporated into the dataset (training = 75%, testing = 25%).

The researchers then tried a 4 layer neural network, with a linear activation function to predict the demographic, and a softmax classifier for selection. For initializing the weights of the neural net, their model relied on auto-encoders. This model performed better than the ABC model.

Discussion
In undertaking this research, the study's researchers demonstrated the potential of deep learning in the field of population genomic analysis which is so appealing because it allows for the inferred solution to problems that statistical inference could not solve for both computational and theoretical reasons. Furthermore, the deep learning approach allows researchers to easily distinguish uninformative and informative summary statistics.

Due to the success of the research, there are many future paths of development. These include but are not limited to learning how various summary statistics relate to parameters, incorporating a wider array of simulations both by the parameters and methods used, using deep learning to prepare data for ABC, using deep learning for continuous parameter inference, and combining "black-box" models with existing theory in the area of research. Not to mention, the solution developed in this paper could be used alongside existing methods of population genomics to compare results of differing approximations.