Project: Time since onset of selection

Input: Multiple SNP matrices, with n haplotypes. All but one of the SNP matrices represent regions evolving neutrally. One SNP matrix represents a region under selection, with constant selection pressure s for \tao generations. Since the onset of selection, individuals carrying the favored allele are (1 + s) more likely to be selected as parent than individuals not carrying the favored allele. The mutation, and recombination rates are the same in all SNP matrix samples.

Output: Provide an estimate of s, \tao .

Project resources: You will not be receiving the input matrices. Instead, use **msms** to generate SNP matrices under selection and evolving neutrally. While you can use parameters \theta, s, N, n, \tao, \pho to simulate, none of these can be used in the program. All parameters must be estimated from the data.

Project goals: Write a project report describing your methodology/algorithm, and results on simulated data. The results must include the following:

- Produce likelihood curves for every data-set for s and \tao. In other words, the x-axis is the value of the parameter, and the y-axis is a likelihood estimate, or some objective function that you are optimizing to estimate the parameter.
- Show the dependence of the strength of your prediction on the values of n, \theta, \pho, s, \tao.

As the project requires extensive simulations, do build the simulation framework quickly, and plan ahead so you can at least complete some simulations on time. You are allowed to work in groups of 2. However, each person must take responsibility for a distinct set of simulations while working together on the core algorithm.

- **Week 1.** Form teams, plan and launch forward simulations to generate your data-sets. At a minimum, you should consider a range of values of s and \tao. Choose value of \tao such that you cover near fixation, and also post-fixation scenarios.

- **Week 2.** Write and submit a report on the algorithm or the methodology you will beusing. This will not be independently graded, so you have the opportunity to change your method, but doing it upfront will help you plan.
- **Week 3.** Run your method on simulated data, and produce charts showing the results.
- **Week 4.** Prepare a presentation, and a report.

Skills: Python, basic human genetics. No more than 2 people per project.