**Summer Institute in Statistical Genetics**

**Module 7: Population Genetics Applications**

**July 13-15, 2022**

**Instructors:** Ryan Hernandez (he/him/his) [ryan.hernandez@ucsf.edu](mailto:ryan.hernandez@ucsf.edu)

Timothy O’Connor (he/him/his) [timothydoconnor@gmail.com](mailto:timothydoconnor@gmail.com)

**Teaching Assistants:** Jacqueline Robinson [Jacqueline.Robinson@ucsf.edu](mailto:Jacqueline.Robinson@ucsf.edu)

**Zoom:**  <https://ucsf.zoom.us/j/94036621565?pwd=clRybGZNVGJ3TU1YTUQxV3J6R21YQT09>

**Slack link:** TBD

**Recorded lectures:** TBD

**Module Overview:** In this module we will discuss population genetics ideas and methods, with a focus on applications to data. Lectures will be given live, with recordings posted after the lecture. Attendance is crucial, as there will be multiple “hands-on” software components as well as small-group discussions.

**Establishing our virtual culture:** The goal of SISG is to strengthen the statistical and genetic proficiency and career preparation of scholars from all backgrounds, especially those from groups historically underrepresented in STEM. Within this workshop, we re-emphasize this by striving to create an environment that values integrity, diversity, collaboration, and respect. You are welcome to turn on your video camera if you please, but be mindful of what is in your background to maintain your own privacy. Please keep your microphones muted in order to minimize background noise. Within the Slack channel, feel free to chat and ask questions. Within Zoom, use the chat only if necessary (or you forget). Please use respectful screen names, including pronouns if you are comfortable doing so. Any disrespectful activity should be reported to the Instructors or TAs. Questions are encouraged throughout this module, but please use the Raise Hand feature in the Participants tab to be called on.

**Software required:** Rstudio, Java, MSMSplay, Selscan (please see module website).

**Module Schedule (all times in Pacific/Eastern time):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Day** | **Start (P/E)** | **End (P/E)** | **Topic** | **Objectives** | **Instructor** |
| **Wednesday** | 11:30/2:30 | 11:45/2:45 | Introduction and course overview | How we plan to use zoom, and other fun things. |  |
| 11:45/2:45 | 1:00/4:00 | Intro to Population Genetics | -Describe the key evolutionary forces: Drift, Mutation (Part 1), Migration, and Selection, and how they function in concert (Part 2) -Learn about how the Site Frequency Spectrum (SFS) can be influenced by these forces (Part 3) | O’Connor |
| **1:00/4:00** | **1:30/4:30** | **BREAK** |  |  |
| 1:30/4:30 | 2:30/5:30 | Coalescent Simulations | -MSMSplay and the impact of demography on the SFS | O’Connor |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Day** | **Start (P/E)** | **End (P/E)** | **Topic** | **Objectives** | **Instructor** |
| **Thursday** | 8:30/11:30 | 9:45/12:45 | Neutrality and some of its deviations | - Learn about the population genetics view of the life cycle | Hernandez |
| - Introduce a few Pop Gen summary statistics |
| - Revisit Hardy-Weinberg Equilibrium: Assumptions & violations |
| **9:45/12:45** | **10:00/1:00** | **BREAK** |  |  |
| 10:00/1:00 | 11:00/2:00 | Intro to  Wright-Fisher Simulations | - Simulate the standard neutral model, demographic effects, and natural selection | Hernandez |
| 11:00/2:00 | 11:45/2:45 | Small group discussion | LINK |  |
| **11:45/2:45** | **12:15/3:15** | **BREAK** |  |  |
| 12:15/3:15 | 1:30/4:30 | Population structure analysis | -Methods to identify global estimates of population structure: Principal Component Analysis (PCA) and Structure/Admixture | O’Connor |
| -Local ancestry can identify segments of the genome corresponding to different ancestries. |
| Local ancestry can be applied in a number of different ways, including demographic modeling, selection, refining PCA signals, and association analyses. |
| 1:30/4:30 | 2:30/5:30 | Running PCA and Admixture | -Understand how to run Principal Component Analysis on genetic data and what it can tell you. | O’Connor |
| -Run the program ADMIXTURE and identify the ancestral composition of a genome. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Day** | **Start (P/E)** | **End (P/E)** | **Topic** | **Objectives** | **Instructor** |
| **Friday** | 8:30/11:30 | 9:45/12:45 | Natural Selection | -Understanding different types of natural selection | Hernandez |
| -What are the signatures of natural selection (positive and negative) |
| -How does selection affect linked sites |
| -Genetic load |
| **9:45/12:45** | **10:00/1:00** | **BREAK** |  |  |
| 10:00/1:00 | 11:00/2:00 | Searching for Selection Signatures | Using selscan to identify signatures of selection | Hernandez |
| 11:00/2:00 | 11:45/2:45 | Small group discussion | LINK |  |
| **11:45/2:45** | **12:15/3:15** | **BREAK** |  |  |
| 12:15/3:15 | 1:15/4:15 | Cryptic Relatedness and Fine-Scale Population Structure | - Define fine-scale population structure and cryptic relatedness | O’Connor |
| - How is it identified |
| - Identity-by-descent • Rare variation |
| - Why it can be important for association analyses, especially of rare variants |
| 1:15/4:15 | 2:15/5:15 | Pop Gen Meets Quant Gen | - Evolutionary models of complex traits | Hernandez |
| - How demography and selection affect genetic architecture |
| 2:15/5:15 | 2:30/5:30 | Discussion/recap |  |  |