Syllabus

Speciation and adaptation genomics: a how-to guide

Daniel Hooper
dhooper@amnh.org – 520-780-4415 – Bioinformatics Core (Sec. 9, 6th floor, Rm 2)
Office hours by request

Dates: January 8-12, 2024

Meeting Time and Location: M-F 10:00am -12:00pm & 1:00pm – 3:00pm; RGGS Classroom.

Course Description: This course is an introduction to population genomic analyses related to understanding the processes of speciation and adaptation. Students will work through a complete bioinformatic pipeline from raw read data to reviewer-ready results while discussing seminal works from the literature. The course will include five modules: (1) read processing, mapping, and variant calling/filtering; (2) population genetic summary statistics and geographic/genomic clinal analysis, (3) genome-wide association mapping and statistical confidence, (4) scanning for selective sweeps, and (5) structural variant mapping. Over the span of each of these five modules, students will learn the skills required for data manipulation and the making of publication ready figures in R (*Tidyverse*). The workshop will also showcase the benefits of linked-read over short-read sequencing data for population genetics and structural variant discovery. Students will be assigned at-home tutorials for further practice and have a final project to demonstrate skill aptitude.

Credits: 1 course credit

Duration: One week (five days total)

- Number of lectures/meetings per week: Five total
  - Length of lectures/meetings per session: 120 min (10 hours total)
- Number of labs per week: Five total
  - Length of lab per session: 120 min (10 hours total)

Instructional hours: 10 hours lecture; 10 hours lab (5 hours instructor-directed lab, 5 hours in-class exercises to be considered as part of supplemental assignments)

Supplemental assignments: 5 hours in-class exercises; 10 hours tutorials online; 15 hours final problem set.

[*NYS requires 750 minutes/12.5 hours of classroom instruction and 30 hours of supplemental assignments per credit.*]

Learning Objectives:
Module 1: Genomic read processing, mapping, and variant calling/filtering. Understanding the basics of R syntax, importing and exporting data, package installation, data structures, and manipulating data frames.
Module 2: Population genetic summary statistics and geographic/genomic clinal analysis. Basic knowledge of data manipulation and summarization with *Tidyverse* packages
Module 3: Genome-wide association mapping and ascertaining statistical confidence. Basic knowledge of performing statistical tests in R.
Module 4: Scanning for selective sweeps using haplotype-based summary statistics. Basic knowledge of benefits of linked-read sequence data (haplotagging) versus short-read sequence data for population genetics.
Module 5: Structural variant discovery and mapping. Basics of linkage-disequilibrium and reference-mapping based approaches. Focus on benefits of linked-read data over short-read sequence data.

Grading/Evaluation Basis: Participation and attendance at workshop and hands-on exercises completed in class (50%); daily online tutorials (25%); final practical exam/problem set (25%)

Final Exam/Project: For the final practical exam, students will be assigned an open book, take home set of analytical questions to complete. The question set will utilize a real comparative genomic dataset from a recent paper that will require students to demonstrate proficiency in each of the topic modules outlined in the learning objectives of the course.

Course Evaluations: Each student is required to complete an anonymous course evaluation at the end of the workshop. The course evaluation is a tool for faculty and administrators to improve the student learning and instructor teaching experience.