What's in a Mutt: an intro to genomics through dog DNA analysis, EN.601.147.13

Instructor: Rachel Sherman
Room: Hodson 315
Piazza: https://piazza.com/jhu/spring2019/en60114713

Intersession 2019, M/W/F, 3:00-5:00pm

Course Description:
Have you ever looked at a cute pup and wondered what mix it is? If you've then proceeded to take your best guess, chances are your guess is wrong. Recent research shows that we're terrible at guessing a mutt's makeup by eye – in fact, the average person is wrong about 75% of the time. So how can we figure out a dog's breed? As DNA sequencing becomes commonplace, more people are turning to sequencing to discover the breed of their best friends. And while knowing a dog's breed is often just a matter of curiosity, this knowledge can also help inform vet's diagnoses and treatments, make shelter dogs more adoptable, and teach us new things about our canine companions.

This course will provide an introduction to genome sequencing techniques by examining how commercial dog DNA sequencing works and how that compares to state-of-the-art sequencing techniques for both dogs and humans. The course will be a mixture of lectures and hands-on analysis of real dog DNA data. Students will learn the computational techniques used to determine a dog's breed, and will be given a mutt's DNA sequence to try to determine its genetic makeup.

Prerequisites: Students should have basic programming skills; no specific programming language is required.

Course Aims:
By the end of this course, students will:
1. Be familiar with the latest commercial and state-of-the-art genome sequencing technologies.
    a. Understand how commercial genome sequencing kits work, and what their caveats are.
    b. Be able to discuss the pros and cons of using state-of-the-art genome sequencing techniques vs standard commercial approaches.
2. Understand how dog breed is determined from DNA sequence
3. Be able to use existing software and statistical methods to determine the most likely dog breed makeup from DNA sequence
4. Be able to apply the methods used in the class to DNA sequencing data from other species

Course Topics:
- Why sequence your dog?

- How commercial DNA sequencing works (for dogs and people!)
- What other sequencing technologies are available and how they work
- Compare commercial vs state-of-the-art methods and their pros and cons
- How is dog breed determined from DNA?
- Project: Determine a dog's breed from its DNA sequence

Assessments:
Students will be assessed based on attendance and completion of the project. The project will be done primarily during in-class lab sessions, with anything unfinished being completed at home. The final project submission will be a short report of the most likely dog breed(s) of the sample analyzed, and a short description of the techniques used to obtain their results.

Schedule (subject to change):

Date	Topic
M Jan 7	Why sequence our dogs? Defining dog breeds, intro to the Mutt Mix Project Overview of commercial dog DNA sequencing via SNP arrays
W Jan 9	SNP array sequencing (continued) SNP analysis techniques: clustering and feature selection *Discussion:* Ethics of commercial canine sequencing
F Jan 11	SNP biology: Linkage, recombination, and phasing *In-class lab:* Clustering breeds using SNP data
M Jan 15	SNP analysis techniques: Hidden Markov Models *In-class lab:* Exploring genotype signatures in mutts vs purebreds
W Jan 17	Sequencing alternatives to SNP arrays *In-class lab:* Determining percentage mixtures of a mutt
F Jan 19	Alternative analyses: Alignment and assembly *In-class lab:* Determining percentage mixtures of a mutt, continued
M Jan 22	**MLK Day, no class**
W Jan 24	Using GWAS to determine SNPs of interest *In-class lab:* Exploring whole genome sequencing data in IGV browser
F Jan 26	Cutting edge dog genomics research *In-class activity*: TBD; reserved for catch up on assignments if needed