Lecture 1: Introduction

Population genetic PCB4553/6685 **1/14/2025**

- Juannan Zhou
- juannanzhou@ufl.edu
- Research interests: Evolutionary biology, synthetic biology, genetics of complex traits, machine learning, yeast

- Office hours: Thursday 11:30am - 1:25pmB
- 122 Bartram Hall

Textbooks

- **Graham Coop. Population and Quantitative Genetics (MC). Online. 3rd Release. minicoop.pdf**
- Sean Rice. Evolutionary Theory: Mathematical and Conceptual Foundations
- John H. Gillespie: Population Genetics - A Concise Guide 2nd Edition
- Charlesworth&Charlesworth: Elements of Evolutionary Genetics 1st Edition

Course schedule

- Course website: https://github.com/juannanzhou/popgen2025S
- Week 1-11: Lectures
- Week 12-15: Paper Discussion
- Week 16: Project presentations

Grading

- Midterm Exam (100 points, 40% of final grade):
  - Weekend-long take-home exam
- Class Participation (20 points, 20% of final grade): Each student will be responsible for leading an in-class discussion on one of the assigned readings.
- Homework will not be graded, but completion strongly recommended

Final Project

- 80 points, 40% of final grade
- PCB4553:
  - The final project is a presentation on the subsequent influence (i.e., the "paper trail") of the paper you presented in class.
- PCB6685:
  - Graduate students in the class are required to identify a research topic and complete a course project where they apply certain population genetic analyses to their own datasets.
  - In cases where a student has yet to generate their own experimental data, publicly available datasets can also be used (e.g. from NCBI).
  - The final project will be an in-class presentation of the findings.

Informal prerequisites

- Familiarity with basic transmission genetics ("**Mendelian genetics**") and mastery of basic algebra and **probability**.
- Some knowledge of linear algebra, calculus and elementary probability theory is a plus

What is population genetics

- “Population genetics is the study of the genetic composition of natural populations and its evolutionary causes and consequences” -Graham Coop

- “Population genetics is a subfield of genetics that deals with genetic differences within and between populations, and is a part of evolutionary biology. Studies in this branch of biology examine such phenomena as adaptation, speciation, and population structure” -Wikipedia

- ME: Popgen is just fancy Mendelian inheritance in a population worked out using probability theory

What is quantitative genetics

- Quantitative genetics deals with *quantitative traits* (phenotypes that vary continuously, e.g. height or mass)

The fundamental forces of evolution

- Mutation
- Recombination
- Genetic drift
- Natural selection
- Migration

Why study population genetics

- Fundamental to understanding evolution
- Applications of PopGen
  - Explaining levels of genetic variation
  - Detecting natural selection
  - Predicting evolution (what is the likelihood a mutation in SARS-CoV-2 becomes fixed in the population?)
  - Inferring demographic history (population size, spatial population structure, migration)
- Stepping stone to theoretical biology

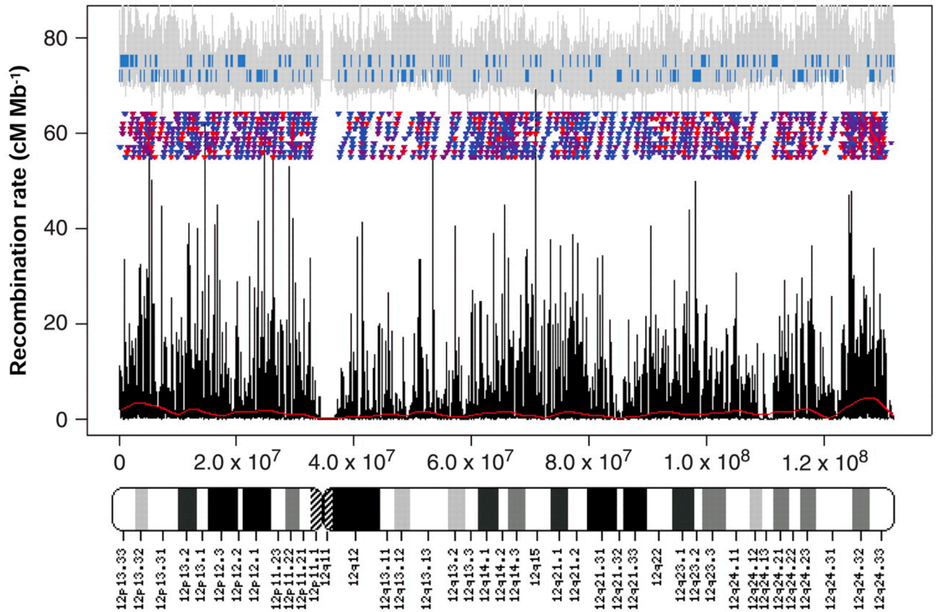
Case study: transposable elements

Barbara McClintock

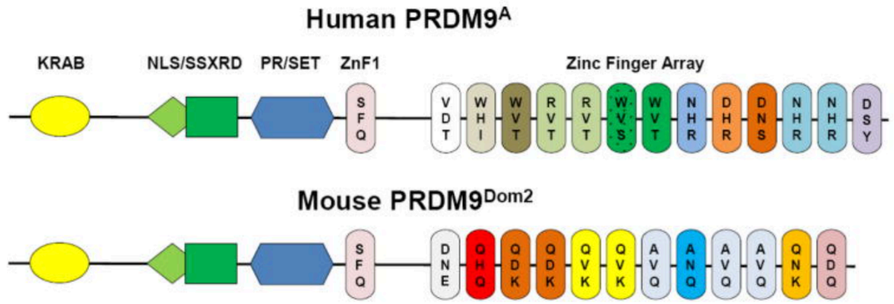
- Transposable elements (TEs): DNA segments capable of making new copies that insert elsewhere in the genome
- Why they are present in the genome?
  - TEs are maintained because they confer benefits on the host by producing adaptively useful mutations
  - TEs are maintained by their ability to replicate within the genome despite potentially deleterious fitness effects of TE insertions
- Work by Charlesworth, Langley, and colleagues in 1989 found that
  - Frequency of TEs in *Drosophila melanogaster* consistent with selection against TEs, thus proving the genomic parasite hypothesis

Case study: discovery of PRDM9 as recombination promoter

- Identification of recombination hotspots using population genetic tools
  - Linkage disequilibrium analysis
- Sequence motif shared by large number of hotspots
- Linked to the DNA-binding domain of zinc finger protein PDRM9



Recombination rate variation along chromosome 12
Myers et al. 2005. Science



A sneak peek

Equation, ref.	Equation, ref.
Relatedness (Inbreeding) coefficient $F_{ij} = 0 \times r_0 + (1/4)r_1 + (1/2)r_2$, (2.4)	Generalized HWE $(1-F)p^2 + Fp,\ (1-F)2pq,\ (1-F)q^2 + Fq$, (2.5)
Inbreeding coefficient $F = \sum_{i=1}^{K} \frac{1}{2^{n_i}}\left(1 + f_{A_i}\right)$, (2.11)	F statistics $F_{\mathrm{IT}} = 1 - \frac{H_I}{H_T},\ F_{\mathrm{IS}} = 1 - \frac{H_I}{H_S},\ F_{\mathrm{ST}} = 1 - \frac{H_S}{H_T}$, (3.1)–(3.3).
Relationship among F statistics $(1 - F_{\mathrm{IT}}) = (1 - F_{\mathrm{IS}})(1 - F_{\mathrm{ST}})$, (3.4)	Linkage disequilibrium (LD) $D = p_{AB} - p_A p_B$, (3.15)
Decay of LD $D_t = (1-c)^t D_0$, (3.22)	Decay of Heterozygosity $H_t = \left(1 - \frac{1}{2N_e}\right)^t H_0$, (4.2)
Equilibrium level of neutral heterozygosity $H = \frac{4N_e\mu}{1+4N_e\mu} \approx 4N_e\mu$, (4.13)	Coalescent time and time to MRCA $\mathbb{E}[T_k] = \frac{2N_e}{\binom{k}{2}}$, $\mathbb{E}[T_{MRCA}] = 4N_e(1-1/n)$, (4.32) (4.36)
Number pairwise diffs. & segregating sites $\mathbb{E}[\pi] = 4N_e\mu$, $\mathbb{E}[S] = 4N_e\mu \sum_{k=n}^{2} \frac{1}{k-1}$, (4.24), (4.39)	Expectation of d_N/d_S $d_N/d_S = (1 - C - B) + 2NBf_B$, (5.7)
Model-based F_{ST} expectations. $F_{\mathrm{ST}} = \frac{T}{T+4N_e}$, $F_{IM} = \frac{1}{1+4N_I m}$, (6.4), (6.7)	Phenotypic covar. between relatives (i & j) $Cov(X_1, X_2) = 2F_{1,2}V_A + r_2 V_D$, (7.17), (7.33)
Cross trait (1 & 2) covar. between relatives $Cov(X_{1,i}, X_{2,j}) = 2F_{i,j}V_{A,1,2}$, (7.23)	Breeder's equation $R = h^2 S = V_A\beta = \frac{V_A}{\overline{w}}\frac{\partial \overline{w}}{\partial \overline{x}}$, (8.4), (8.16), (8.19)
Multi-variate breeders equation $\mathbf{R} = \mathbf{G}\mathbf{V}^{-1}\mathbf{S} = \mathbf{G}\boldsymbol{\beta}$, (9.2)	Hamilton's Rule $2F_{i,j}B > C$, (9.13)
Frequency next generation (haploid & diploid). $p_{t+1} = \frac{w_1}{\overline{w}}p_t,\ p_{t+1} = \frac{w_{11}p_t^2 + w_{12}p_t q_t}{\overline{w}}$, (10.4), (10.19)	Frequency change $\Delta p_t = \frac{(\overline{w}_1 - \overline{w}_2)}{\overline{w}} p_t q_t = \frac{1}{2}\frac{p_t q_t}{\overline{w}}\frac{d\overline{w}}{dp}$, (10.23), (10.24)
Haploid cumulative change (use $s/2$ for diploid case) $p_\tau \approx \frac{p_0}{p_0 + q_0 e^{-s\tau}}$, $\tau \approx \frac{1}{s}\log\left(\frac{p_\tau q_0}{q_\tau p_0}\right)$, (10.9) (10.11)	Heterozygote advantage equilibrium $p_e = \frac{s_2}{s_1 + s_2}$, (10.31)
Diploid mutation-selection equilibrium $q_e = q_t = \frac{\mu}{hs}$, $q_e = \sqrt{\frac{\mu}{s}}$ (if $h = 0$) , (11.6), (11.7)	Migration-selection equil. & cline width. $q_{e,1} = \frac{m}{hs}$, $0.6\sigma/\sqrt{s}$, (11.12), (11.13)
Selected prob. fixation (haploid & diploid) $p_F(1/2N) = 2s$, $P_F(1/2N) \approx 2hs$, , $Ns \gg 1$, (12.7), (12.8)	Prob. fixation for weakly selected alleles ($h = 1/2$) $P_F\left(\frac{1}{2N}\right) = \frac{1-e^{-s}}{1-e^{-2Ns}}$, $s < 0$ for deleterious allele. , (12.12)

Lecture	Week	Date		Topic	Book chapter
1	1	1/14/2025	Intro	Introduction to course	MC1 Intro, Appendix
2	1	1/16/2025	Math background	Mathematical background	MC Appendix
3	2	1/21/2025	Descriptive statistics	Allele and Genotype Frequencies	MC2
4	2	1/23/2025	Neutral theory	Loss of heterozygosity due to drift	MC4.1
5	3	1/28/2025	Neutral theory	The Coalescent and patterns of neutral diversity	MC4.2
6	3	1/30/2025	Neutral theory	The coalescent process of a sample of alleles; Diffusion Approximation	MC4.3