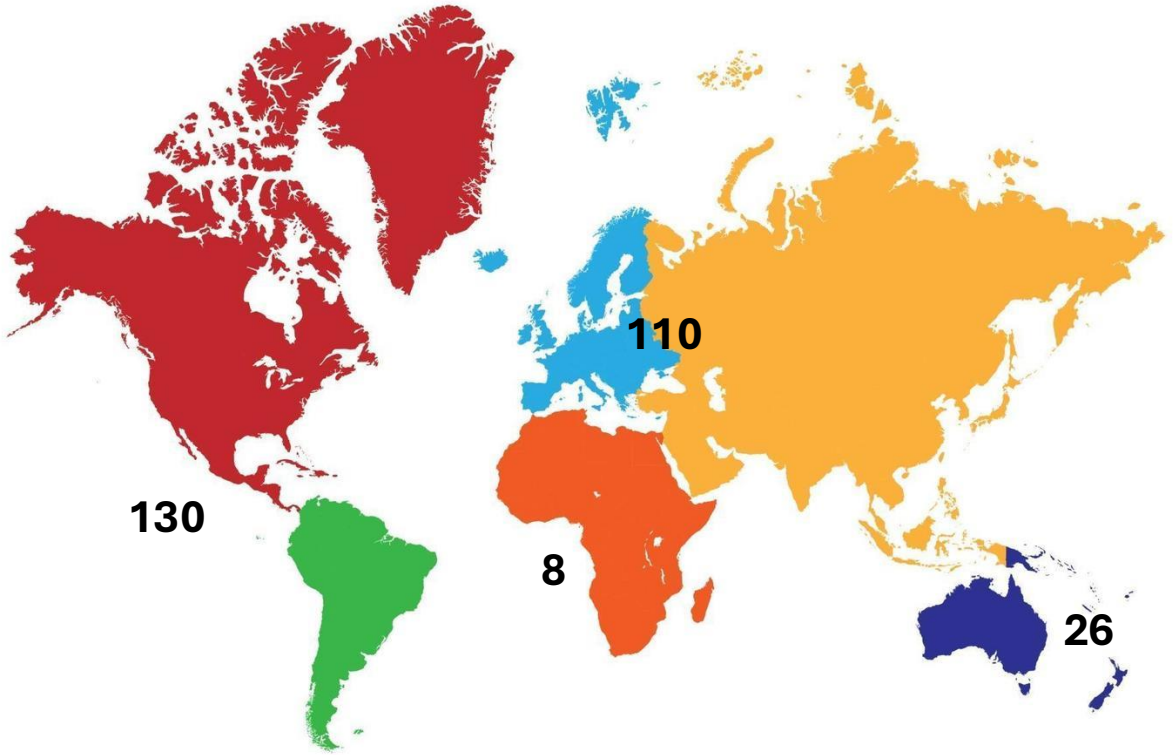
2026 International Statistical Genetics Workshop

Prof Loic Yengo
The University of Queensland
(Zoom)

300 (~200 Session A – ~100 Session B)

Tutors (Session A)
Anais Thijssen
Xiaotong (Mark) Wang
Patrick Turley
Yuna Zhang
Neil Davies
Scott Vrieze
Chris Beam
Gunn-Helen Moen
Aysu Okbay
Xuemin Wang

Tutors (Session B)
Valentin Hivert
Tunde Olasege
Julia Sidorenko
Tian Lin



Credit map: Vecteezy

Outline

(1) Welcome message

(2) assignment of breakout rooms + ice breaking activity for 5 min within groups

(3) A quick overview of the lecture material + how to access the instructions for the practical (HTML file)

(4) Breakout rooms for 30 min to complete **Practical 1** (Pop Gen)

(5) 10 min break

(6) Brief discussion around Practical 1 + Q&A.

(7) Breakout rooms for 30-40 min to complete **Practical 2** (Quant Gen)

(8) Brief discussion around Practical 2 + Q&A.

(9) Concluding remarks.

(*10) Re-pen breakout rooms for those who want to continue to chat or finish the practical.*

Outline

(**1) Welcome message**

(2) assignment of breakout rooms + ice breaking activity for 5 min within groups

(3) A quick overview of the lecture material + how to access the instructions for the practical (HTML file)

(4) Breakout rooms for 30 min to complete **Practical 1** (Pop Gen)

(5) 10 min break

(6) Brief discussion around Practical 1 + Q&A.

(7) Breakout rooms for 30-40 min to complete **Practical 2** (Quant Gen)

(8) Brief discussion around Practical 2 + Q&A.

(9) Concluding remarks.

(*10) Re-pen breakout rooms for those who want to continue to chat or finish the practical.*

Ice breaking questions

- What is your name, education background and what do you expect to learn in this workshop? [1 min]

- Which would you rather lose: your raw data, your scripts, or your notes? [1 min]

- What genetics/genomics buzzword will we all be tired of hearing in five years? [1 min]

Overview of lecture material

➔ https://www.colorado.edu/ibg/workshop-2026

Course Information

- Syllabus This is the place to find the videos to watch ahead of each day. The new videos recorded for the 2026 course will be released at midnight May 18, UTC. Lectures recorded for prior courses are already available. Most days contain a mix of old and new lectures.
- Financial aid application Discounted registrations can be provided to students who demonstrate a financial need.
- Official Announcement

Syllabus

Each day of the course will focus on one topic, or several closely related topics, which will be presented by one or more of the Workshop faculty members.

Depending on session and timezone, meetings may be the day after the listed date for some students.

Day 0: Before the start of the course

- Topics: Accessing the ISG Workshop cloud environment, ISGW Forum

Day 1: June 1, 2026; Background

- Lead: Loïc Yengo
- Topics: biometrical modeling; population genetics; data sources; ethics and histo

Lectures

Introduction to Quantitative Genetics Theory: Estimation of heritability from individual-level data (95 minutes)

Introduction to Population Genetics Theory: Hard-Weinberg Equilibrium, Linkage Disequilibrium (72 minutes)

Access Practical

➔ https://www.colorado.edu/ibg/workshop-2026

Course Information

- Syllabus This is the place to find the videos to watch ahead of each day. The new videos recorded for the 2026 course will be released at midnight May 18, UTC. Lectures recorded for prior courses are already available. Most days contain a mix of old and new lectures.
- Financial aid application Discounted registrations can be provided to students who demonstrate a financial need.
- Official Announcement

Syllabus

Each day of the course will focus on one topic, or several closely related topics, which will be presented by one or more of the Workshop faculty members.

Depending on session and timezone, meetings may be the day after the listed date for some students.

Day 0: Before the start of the course

- Topics: Accessing the ISG Workshop cloud environment, ISGW Forum

Day 1: June 1, 2026; Background

- Lead: Loïc Yengo
- Topics: biometrical modeling; population genetics; data sources; ethics and

Practicals

Information about the practical is in the part 3 of the introduction to population genetics lecture series.

The practical can be viewed at https://ibg.colorado.edu/workshop2026/practicals/pop_and_quant/Practical_Pop_and_Quant_Gen.html

Practical looks like this…

Introduction to Population and Quantitative Genetics

International Statistical Genetics Workshop

First login to RStudio

```
https://workshop-ood.colorado.edu/
```

Part 0. Preparation

Copy the Practical document into your home directory using the following `R` commands.

```
setwd("~/.")
system("mkdir -p Day1")
system("cp /home/loic/2026/Practical_Pop_and_Quant_Gen.html ~/Day1/.")
system("cd ~/Day1")
setwd("Day1")
```

Population Genetics

[PG] Part 1: Visualize changes in allele frequencies

The `R` commands below define two functions: (1: `generateNextGeneration()`) generates genotypes for the next generation from a set of genotypes in the current population, and (2: `sim`), which uses (1) to simulate the evolution over g generations of a population with a constant sample size (N), a certain number m of SNPs segregating in the founding (or ancestral) population with a frequency (p). The `sim` function then plot the frequency of each of the `m` alleles over the course of `g` generations (each curve corresponds to a SNP).

Population Genetics

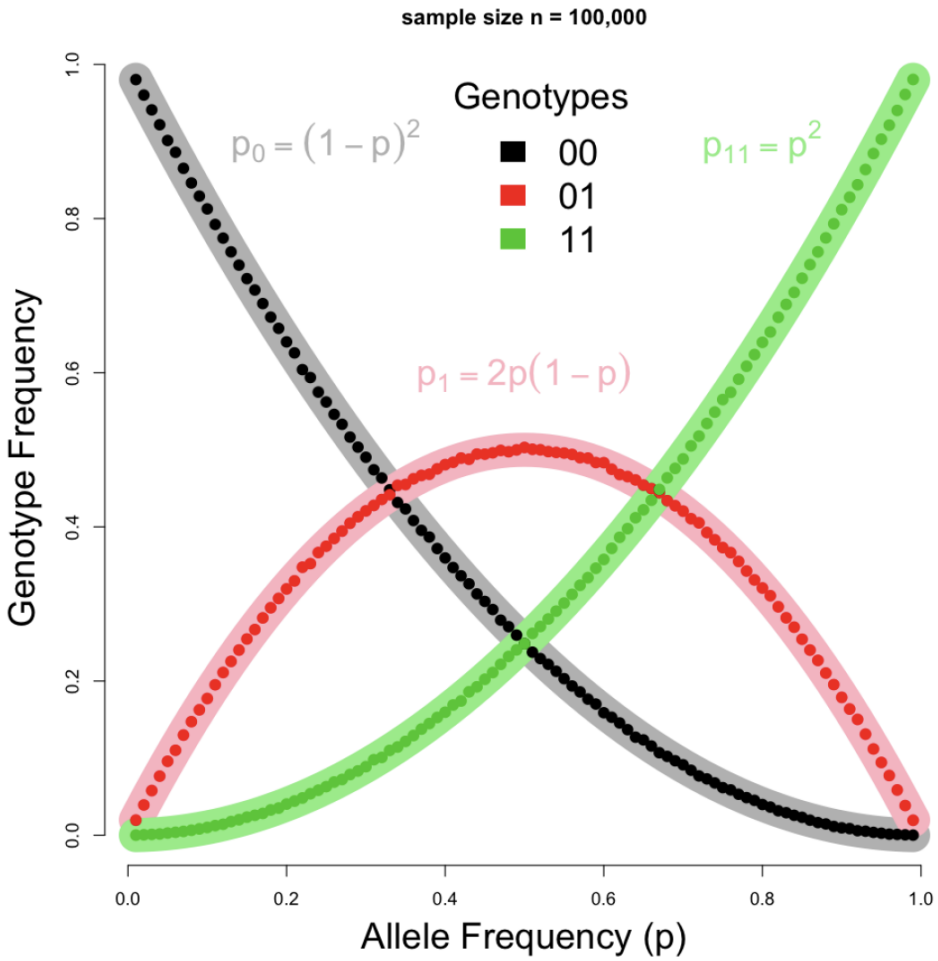
Population Genetics Theory is concerned with characterizing and quantifying **genetic variation** within and between **populations**.

Population Genetics Theory is the theory of **alleles and genotypes frequencies** within and between **groups of individuals**.

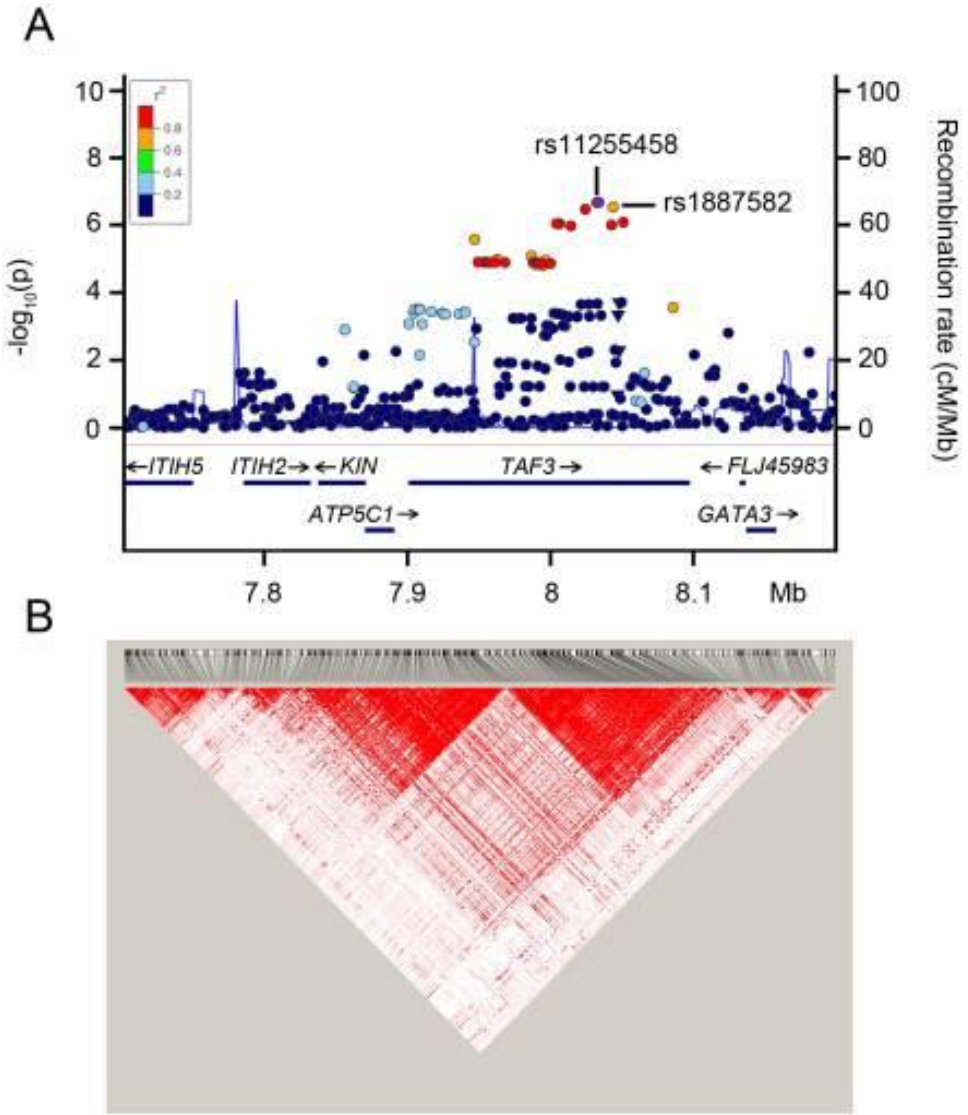
Key concepts

- Hardy-Weinberg Equilibrium
- Linkage Disequilibrium
- Fixation index (F_{ST})

Hardy-Weinberg Equilibrium

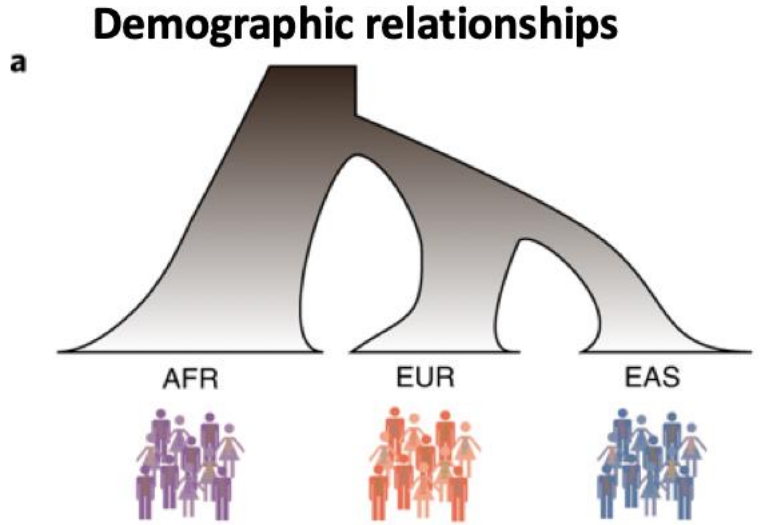


Linkage Disequilibrium



Pistis et al. (2013) Plos One

Population structure = frequency differences between populations



Martin et. al., *Nature Genetics*, 2019

Ancestral population

p_0

➔ F_{ST}

p_1

p_2

p_K

Derived population 1

Derived population 2

Derived population K

Time (drift)

Quantitative Genetics

Estimation of heritability from individual-level data

Heritability (h^2) quantifies the degree to which inter-individual differences and resemblance in the population are due to genetic factors.

Chial, H. (2008) Polygenic inheritance and gene mapping. Nature Education 1(1):17

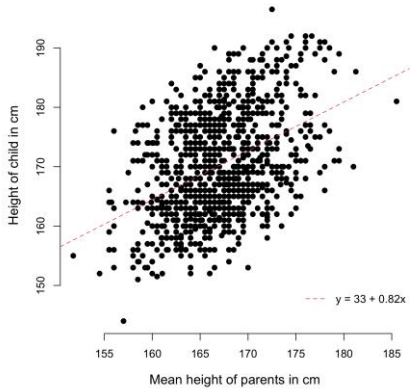
Definitions

Heritability (h^2) quantifies the degree to which inter-individual differences and resemblance in the populations are due to genetic factors.

Heritability can be approached in terms of

- Differences between people in the population: h^2 = var(G) / var(Y),

- Resemblance between relatives (in families): corr(Y_i,Y_j) = h^2R_{ij} + Residual



Standard GRM estimator

$$\hat{\pi}_{jk} = \frac{1}{m}\sum_i \frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1-p_i)}$$

where, x_{ij} and x_{ik} are the minor allele count (x_{ij}, x_{ik} = 0,1 or 2) at SNP i for individuals j and k respectively, p_i the minor allele frequency (MAF) of SNP I and m the number of SNPs used to calculate the GRM.

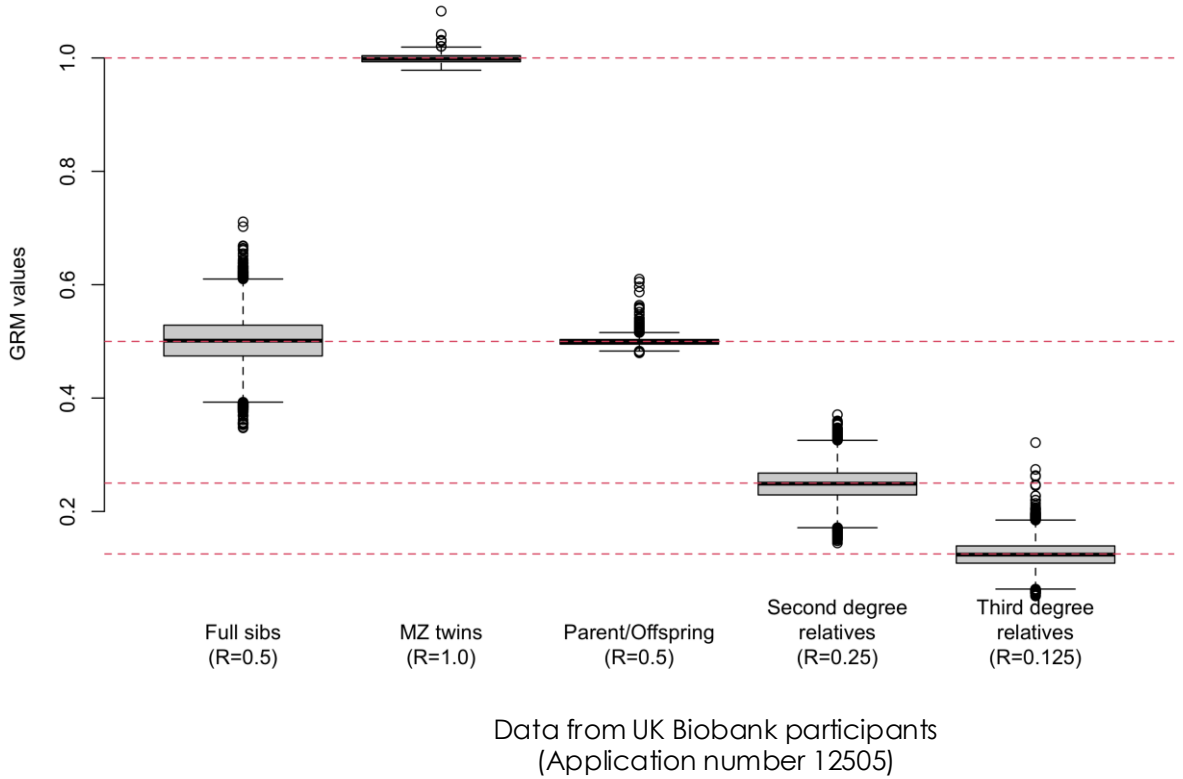
Example of GRM between *N*=3 individuals (over m=1000 SNPs)

```
[$bash] zless myGRM.grm.gz
1 1 1000  0.99
1 2 1000 -0.01
1 3 1000  0.01
2 2 1000  1.03
2 3 1000  0.03
3 3 1000  1.01
```

Distribution of GRM values

The expectation (over a large sample of relatives) of the $\hat{\pi}_{jk}$ is exactly R_{jk}.

Observed relatedness may be still vary within a type of pedigree relationship.



Data from UK Biobank participants
(Application number 12505)

Summary and next part

Observing simultaneously corr(Y_i,Y_j) and R_{ij} is key to estimating h^2

GRMs can be quantified using actual SNP data (show more variation than expected genetic relatedness)

Software **GCTA** can calculate GRM and use them for estimating h^2 (Part 3)

Two approaches for estimating heritability from SNP data

- Estimation using Haseman-Elston (HE) regression

- Estimation using Genome-based REstricted Maximum Likelihood (GREML)

GREML estimation with GCTA

Run GCTA to estimate heritability of trait 1 using GREML

gcta64 --grm myData --pheno phenotype.txt --mpheno 1 **--reml** --out myGREML_estimates

[generates 2 files: myGREML_estimates.log, myGREML_estimates.hsq]

Source	Variance	SE
V(G)	0.398550	0.023990
V(e)	0.578277	0.019175
Vp	0.976827	0.019107
V(G)/Vp	0.408004	0.020539
logL	-2722.000	
logL0	-2932.909	
LRT	421.817	
df	1	
Pval	0.0000e+00	
n	6000	

LDMS method

Step 1: Calculate SNP attributes: MAF (p_i) and LD score (ℓ_i)

$\ell_i = \sum_{k=1}^{m} r_{ik}^2$ (r_{ik}^2: squared correlation of allele counts between SNP i and SNP k)

Step 2: Groups SNPs based on MAF and LD

e.g., 6 MAF groups:]1%-5%],]5%-10%],]10%-20%],]20%-30%],]30%-40%] and]40%-50%]
+ 4 LD score groups (quartile) with each MAF group => K=6x4=24 LDMS groups.

Step 3: Calculate a GRM for each group of SNPs.

Step 4: Estimate jointly the "σ_g^2" for each SNP group.

$\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \sum_{k=1}^{K} \sigma_{g,k}^2 \mathbf{GRM}_k + \sigma_e^2 I_n)$ => $\sigma_g^2 = \sum_{k=1}^{K} \sigma_{g,k}^2$

Outline

(**1) Welcome message**

(2) assignment of breakout rooms + ice breaking activity for 5 min within groups

(3) A quick overview of the lecture material + how to access the instructions for the practical (HTML file)

(4) Breakout rooms for 30 min to complete **Practical 1** (Pop Gen)

(5) 10 min break

(6) Brief discussion around Practical 1 + Q&A.

(7) Breakout rooms for 30-40 min to complete **Practical 2** (Quant Gen)

(8) Brief discussion around Practical 2 + Q&A.

(9) Concluding remarks.

(*10) Re-pen breakout rooms for those who want to continue to chat or finish the practical.*

Practical looks like this…

- You can run the first part on your own RStudio
- Encouraged to work together (sharing screen)

Introduction to Population and Quantitative Genetics

International Statistical Genetics Workshop

First login to RStudio

```
https://workshop-ood.colorado.edu/
```

Part 0. Preparation

Copy the Practical document into your home directory using the following `R` commands.

```
setwd("~/.")
system("mkdir -p Day1")
system("cp /home/loic/2026/Practical_Pop_and_Quant_Gen.html ~/Day1/.")
system("cd ~/Day1")
setwd("Day1")
```

Population Genetics

[PG] Part 1: Visualize changes in allele frequencies

The `R` commands below define two functions: (1: `generateNextGeneration()`) generates genotypes for the next generation from a set of genotypes in the current population, and (2: `sim`), which uses (1) to simulate the evolution over g generations of a population with a constant sample size (N), a certain number m of SNPs segregating in the founding (or ancestral) population with a frequency (p). The `sim` function then plot the frequency of each of the `m` alleles over the course of `g` generations (each curve corresponds to a SNP).

Population Genetics

[PG] Part 1: Visualize changes in allele frequencies

The `R` commands below define two functions: (1: `generateNextGeneration()`) generates genotypes for the next generation from a set of genotypes in the current population, and (2: `sim`), which uses (1) to simulate the evolution over g generations of a population with a constant sample size (N), a certain number m of SNPs segregating in the founding (or ancestral) population with a frequency (p). The `sim` function then plot the frequency of each of the `m` alleles over the course of `g` generations (each curve corresponds to a SNP).

```
generateNextGeneration <- function(x,n){
  m  <- ncol(x)
  N  <- nrow(x)
  iM  <- sample(1:N,n,replace = TRUE)
  iF  <- sample(1:N,n,replace = TRUE)
  xm  <- x[iM,]
  xf  <- x[iF,]
  xa  <- t(sapply(1:n,function(k) rbinom(m,1,prob=0.5*xm[k,])))
  xb  <- t(sapply(1:n,function(k) rbinom(m,1,prob=0.5*xf[k,])))
  xn  <- xa + xb
  return(xn)
}

sim <- function(N,  # sample size
```

Outline

(**1) Welcome message**

(2) assignment of breakout rooms + ice breaking activity for 5 min within groups

(3) A quick overview of the lecture material + how to access the instructions for the practical (HTML file)

(4) Breakout rooms for 30 min to complete Practical 1 (Pop Gen)

(5) 10 min break

(6) Brief discussion around Practical 1 + Q&A.

(7) Breakout rooms for 30-40 min to complete **Practical 2** (Quant Gen)

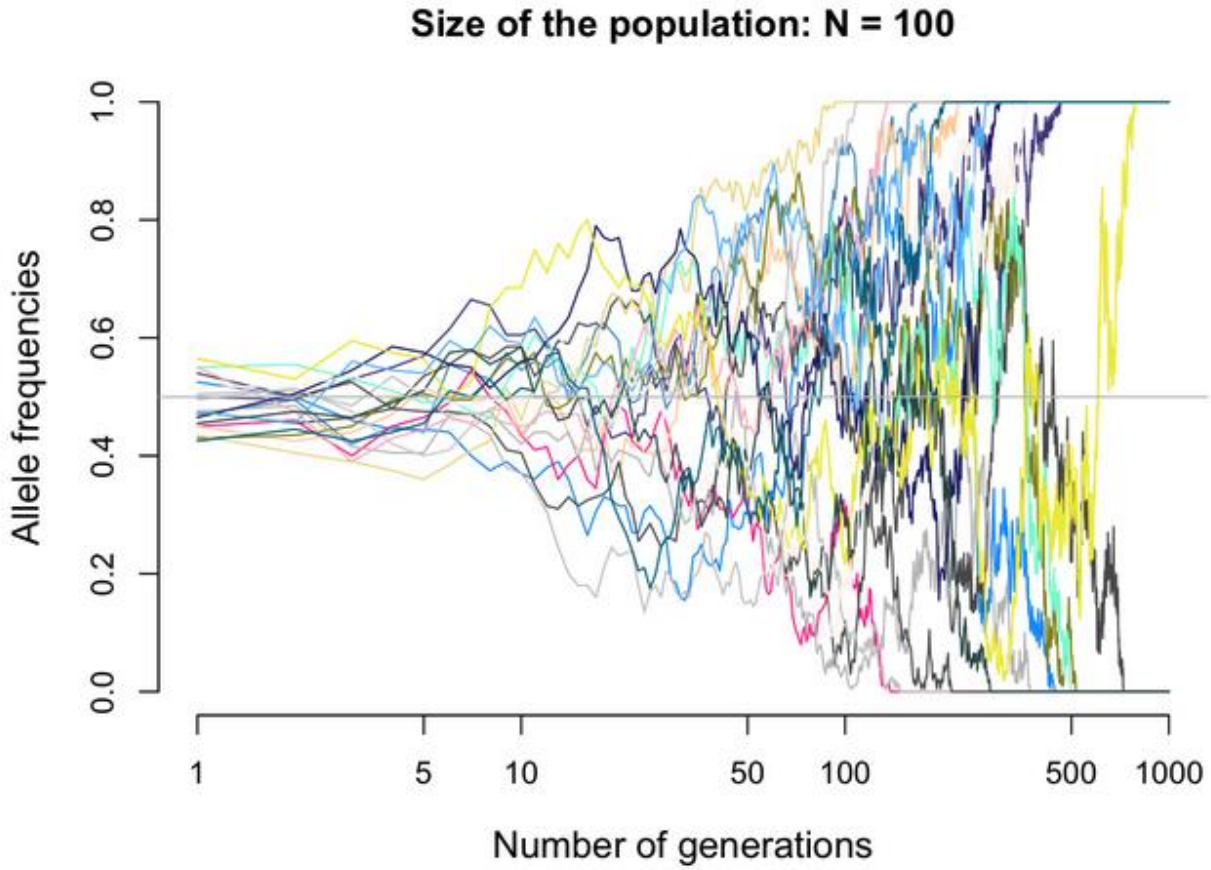
(8) Brief discussion around Practical 2 + Q&A.

(9) Concluding remarks.

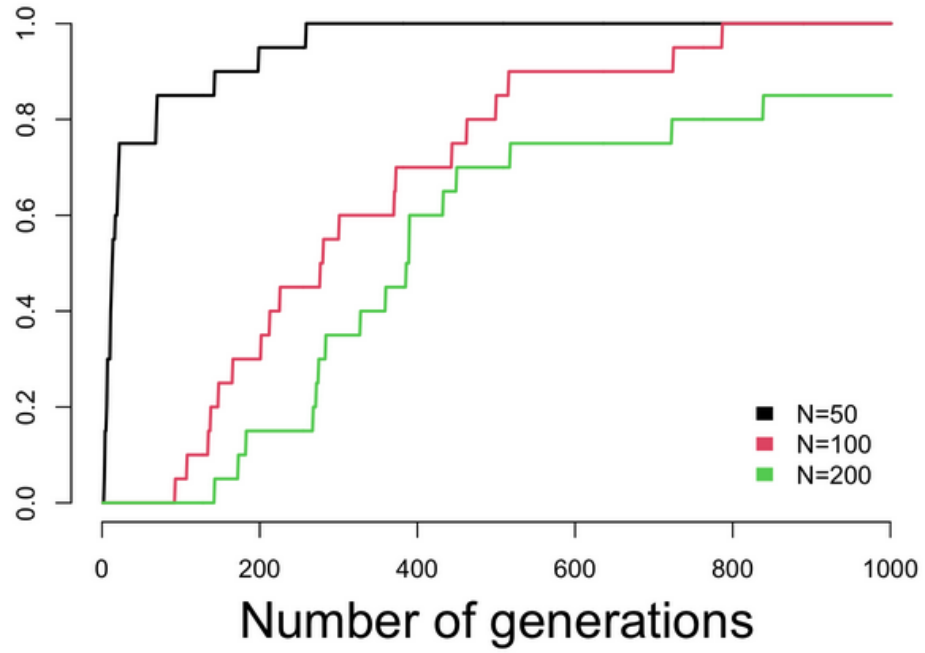
(*10) Re-pen breakout rooms for those who want to continue to chat or finish the practical.*

Feedback on Prac – Part 1

Genetic drift can be a powerful evolutionary force

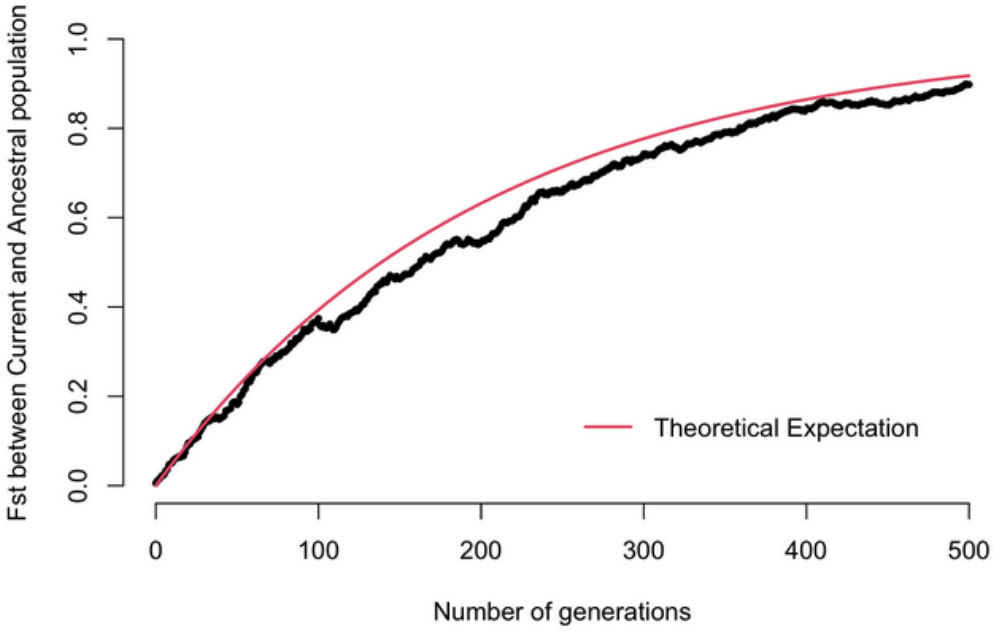


All genetic variation is ***fixated*** if we wait long enough (in the absence of new mutations)

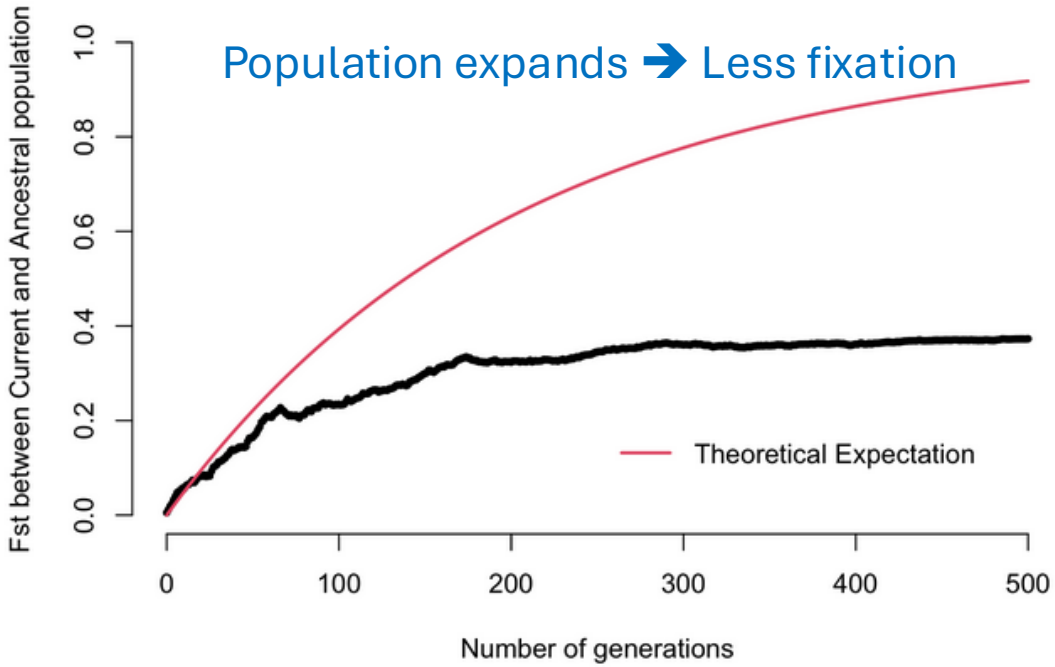


Fixation happens faster in smaller populations

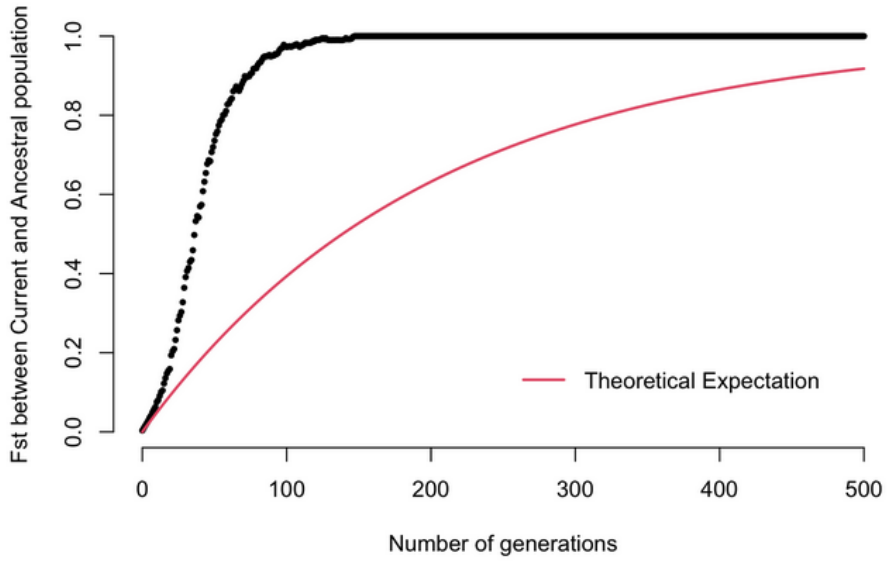
Feedback on Prac – Part 1



Population expands ➔ Less fixation



Population shrinks ➔ More fixation



Take home

- Genetic drift can induce large changes in allele frequencies especially when population size is small
- Fixation is inevitable unless new genetic variation enters the population (mutation, migration)
- Feel free to reuse the code for strengthening intuition

Outline

(**1) Welcome message**

(2) assignment of breakout rooms + ice breaking activity for 5 min within groups

(3) A quick overview of the lecture material + how to access the instructions for the practical (HTML file)

(4) Breakout rooms for 30 min to complete Practical 1 (Pop Gen)

(5) 10 min break

(6) Brief discussion around Practical 1 + Q&A.

(7) Breakout rooms for 30-40 min to complete Practical 2 (Quant Gen)

(8) Brief discussion around Practical 2 + Q&A.

(9) Concluding remarks.

(10) Re-pen breakout rooms for those who want to continue to chat or finish the practical.

Feedback on Prac – Part 2 (GCTA) – Relatedness

Including relatives

```
Summary result of REML analysis:
Source  Variance        SE
V(G)    0.387651        0.023493
V(e)    0.591308        0.019178
Vp      0.978959        0.019038
V(G)/Vp 0.395983        0.020233
```

Excluding relatives

```
Summary result of REML analysis:
Source  Variance        SE
V(G)    0.261986        0.027990
V(e)    0.734552        0.027633
Vp      0.996538        0.020777
V(G)/Vp 0.262896        0.026446
```

Difference = any factor shared by close relatives that influences the trait (environmental or genetic)

Feedback on Prac – Part 2 (GCTA) – Effect of allele frequency

Single GRM analysis

Summary result of REML analysis:

Source	Variance	SE
V(G)	0.548611	0.031984
V(e)	0.467182	0.023087
Vp	1.015792	0.022572
V(G)/Vp	0.540082	0.024745

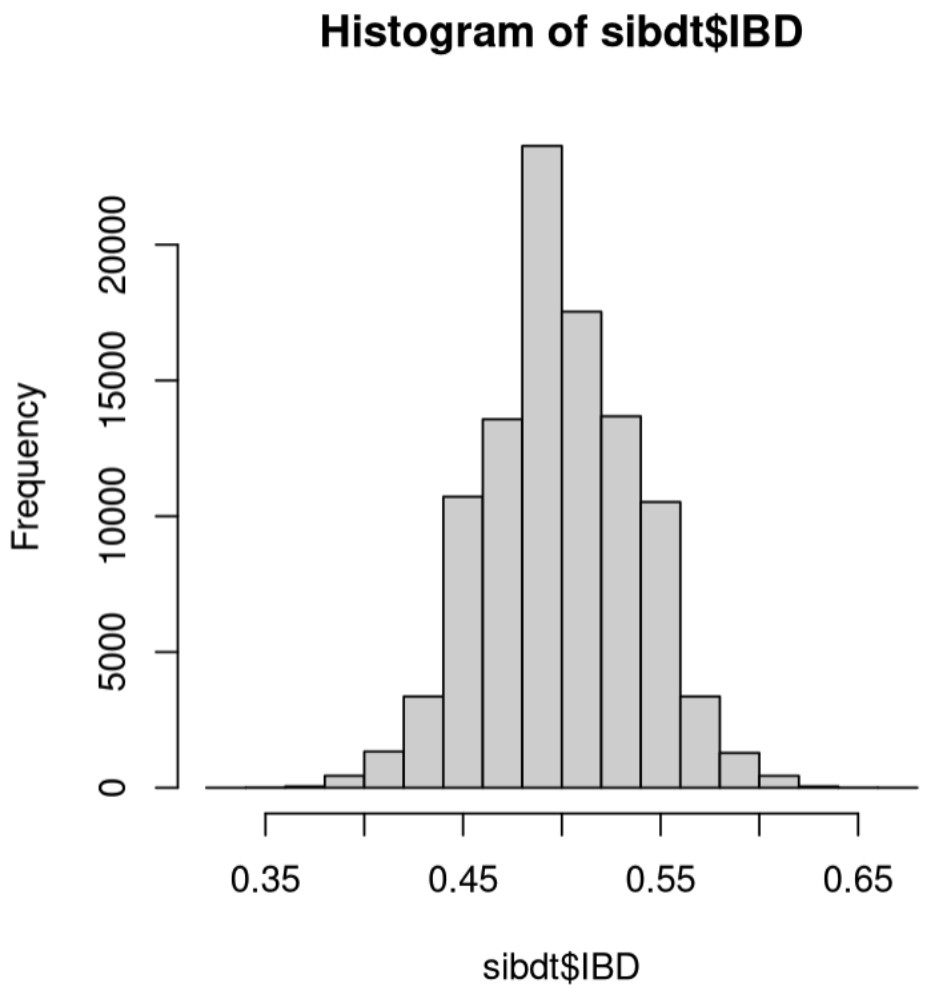
MAF-stratified analysis

Summary result of REML analysis:

Source	Variance	SE
V(G1)	0.605476	0.034624
V(G2)	0.026344	0.011579
V(e)	0.376139	0.013897
Vp	1.007958	0.035287
V(G1)/Vp	0.600695	0.015007
V(G2)/Vp	0.026136	0.011510
Sum of V(G)/Vp	0.626831	0.017852

Assumption in single GRM analysis = all SNP contribute equally to heritability
Stratified analysis gives more freedom in your analysis and also produce less biased estimates

Feedback on Prac – Part 2 (GCTA) – Sibling analysis



Using within-family IBD variation can help overcoming many biases in heritability estimates

medRxiv

THE PREPRINT SERVER FOR HEALTH SCIENCES

Follow this preprint

Within-family heritability estimates for behavioural and disease phenotypes from 500,000 sibling pairs of diverse ancestries

Loic Yengo, Yanyu Liang, 23andMe Research Team, Xin Wang, Julie M. Granka, David M. Evans, Julia Sidorenko, Peter M. Visscher

doi: https://doi.org/10.1101/2025.09.17.25336022

Take home

- Heritability from close relatives can be different from that of distant relatives – not always a bias (missing heritability)
- MAF/LD Stratification can minimize biases. Bayesian methods are even more flexible (Day 5 - PRS)
- Within-family is the gold standard but require quite large sample sizes.

Concluding remarks

Link with Day 2/4: GCTA syntax is inspired by PLINK
Link with Day 3: Refreshed on IBD and variance estimation
Link with Day 5: Heritability as an upper bound for PRS accuracy
Link with Day 6: Application to GWAS summary statistics.

nature

Explore content | About the journal | Publish with us

nature > articles > article

Article | Open access | Published: 12 November 2025

Estimation and mapping of the missing heritability of human phenotypes

Pierrick Wainschtein, Yuanxiang Zhang, Jeremy Schwartzentruber, Irfahan Kassam, Julia Sidorenko, Petko P. Fiziev, Huanwei Wang, Jeremy McRae, Richard Border, Noah Zaitlen, Sriram Sankararaman, Michael E. Goddard, Jian Zeng, Peter M. Visscher, Kyle Kai-How Farh & Loic Yengo

Nature **649**, 1219–1227 (2026) | Cite this article

83k Accesses | **30** Citations | **370** Altmetric | Metrics

Paper to check out.