EXERCISES
Session 2: Population Genetics and Genomics

Exercise 1 [part 1]
Population geneticists analyze the extent and patterns of genetic variation within populations, the distribution of genotypes, and the influence of evolutionary forces like natural selection and genetic drift. These factors drive evolutionary change and speciation. Based on your current understanding, how would you address the following questions?

i. How can we measure genetic variation within a population?
- Genetic variation is assessed using methods such as DNA sequencing, microsatellites, and single nucleotide polymorphism (SNP) analysis.
- Heterozygosity levels and allele frequencies are key indicators of variation.
- Techniques like gel electrophoresis, restriction fragment length polymorphism (RFLP), and whole-genome sequencing help quantify diversity.

ii. How can we determine whether a population's genetic structure is changing over time?
- By comparing genetic data across multiple generations using longitudinal studies.
- Analyzing Hardy-Weinberg equilibrium deviations can reveal evolutionary forces acting on the population.
- Methods such as F-statistics (e.g., F_ST) help measure genetic differentiation between subpopulations.
- Molecular markers and phylogenetic analyses track changes in allele frequencies.

Exercise 2 [part 1]
Which term refers to the complete set of genetic information shared by all individuals within a population?

1. Gene pool
2. Genome
3. Chromosome complement
4. Breeding unit
5. Race

Answer: 1

Various mechanisms help preserve genetic diversity within a population. Why is maintaining this diversity beneficial?

1. Homozygosity provides an evolutionary advantage.
2. Increased diversity promotes inbreeding benefits.
3. Genetic diversity enhances a population's ability to adapt to environmental changes.
4. Greater genetic diversity increases the likelihood of haploidy.

5. Genetic diversity prevents diploidy.

Answer: 3

Exercise 3 [part 1]

Human serum contains a protein called haptoglobin, which binds to hemoglobin following the lysis of erythrocytes. In some populations, three electrophoretic haptoglobin variants are observed, determined by three autosomal co-dominant alleles: A, B, and C. Among 500 individuals, the following genotype counts were recorded:

Genotypes						$\sum$
AA	AB	AC	BB	BC	CC	
109	123	128	33	70	37	500

i. Compute the genotype frequencies within the sample.

$P(AA) = \frac{109}{500} = 0.218$

$P(AB) = \frac{123}{500} = 0.246$

$P(AC) = \frac{128}{500} = 0.256$

$P(BB) = \frac{33}{500} = 0.066$

$P(BC) = \frac{70}{500} = 0.14$

$P(CC) = \frac{37}{500} = 0.074$

ii. Compute the allele frequency within the sample.

$P(A) = (2 * 109 + 123 + 128)/1000 = 0.469$
$P(B) = (2 * 33 + 123 + 70)/1000 = 0.259$
$P(C) = (2 * 37 + 128 + 70)/1000 = 0.272$

Exercise 4 [part 1]
Cystic fibrosis (CF) is an autosomal recessive disease. In individuals who are homozygous for the recessive allele, a malfunction in salt transport occurs in certain cell types, such as those in the lungs. This defect leads to thick mucus buildup, increasing the risk of bacterial lung infections.

Historically, only a small number of individuals with CF survived into adulthood. In Europe, the birth prevalence of CF has been approximately 1 in 2,500 newborns. This frequency is maintained by a balance between mutation and natural selection.

In a European sample, cystic fibrosis (CF) carrier status was analyzed based on three genotypes: 16,890 individuals were homozygous for the normal allele (RR), 855 individuals were heterozygous carriers (Rr), and 18 individuals were homozygous for the recessive CF allele (rr).

Calculate the frequency of each allele in the sample.
Number of individuals = 16,890+855 + 18 = 17,763
Number of alleles=2*(17,763) = 35,526
P(R) = (2*16,890+855)/35,526 = 0.975
P(r) = (855+2*18)/35,526 = 0.025

Exercise 5 [part 1]
A geneticist studying a human population discovers that a recessive allele (s) associated with sickle cell disease is present in a community of 1,024 individuals. Individuals carrying two copies of the recessive allele (ss) develop sickle cell disease, while those with at least one dominant allele (S) have normal red blood cells. Within the population, 1 in 256 individuals has sickle cell disease, indicating they are homozygous recessive (ss).

i) Assuming random mating and no selection for this gene, how many individuals with normal red blood cells are carriers of the recessive allele?
Thus, it is the heterozygous, Ss we are interested in
P(s) = q
P(S) = p
$q^2 = 1/256 \Rightarrow q = 0.0625$ (p = 1 – 0.0625)
P(Ss) = 2pq = 0.117
heterozygotes: 0.117 * 1024 = 120

Exercise 6 [part 1] – in R

Exercise 7 [part 1] – in R

Exercise 8 [part 2]
In a study of 10,000 individuals of Danish ancestry, researchers analyzed two genetic loci, including the CFTR gene (associated with cystic fibrosis) and the APOE gene (linked to Alzheimer's disease risk). These loci were chosen due to their well-established roles in disease susceptibility and health outcomes, providing insights into the genetic factors influencing these conditions in the population.

CFTR			APOE		
WT|WT	WT|F508del	F508del|F508del	ε3 | ε3	ε3 | ε4	E4 | ε4
3156	4997	1847	6837	3065	98

i. Compute the allele- and genotype frequencies for the two alleles in both the CFTR and APOE-locus

p(A) = (2*3156+4997) / (2*10000) = 0.565
p(a) = (2*1847+4997) / (2*10000) = 0.435

p(AA) = 3156 /10000 = 0.3156
p(Aa) = 4997 /10000 = 0.4997
p(aa) = 1847 /10000 = 0.1847

p(B) = (2*6837+3065) / (2*10000) = 0.837
p(b) = (2*98+3065) / (2*10000) = 0.163

p(BB) = 6837 /10000 = 0.6837
p(Bb) = 3065 /10000 = 0.3065
p(bb) = 98 /10000 = 0.0098

ii. Perform the necessary calculations to determine whether the allele frequencies at the two loci follow Hardy-Weinberg equilibrium proportions.

	AA	**Aa**	**aa**	**BB**	**Bb**	**bb**
Obs.	3156	4997	1847	6837	3065	98
Exp. frekvens	$p(A)^2$ $=(0.565)^2$ $=0.3197$	$2p(A)\,p(a)$ $=2\times0.565\times0.435$ $=0.4914$	$p(a)^2$ $=(0.435)^2$ $=0.1888$	$p(B)^2$ $=(0.837)^2$ $= 0.7005$	$2p(B)\,p(b)$ $=2\times0.837\times0.163$ $=0.2729$	$p(b)^2$ $=(0.163)^2$ $=0.02658$
Exp. number	3197.337	4914.326	1888.337	7004.853	2729.294	265.853
(Obs.-Exp.)²/Exp.	0.5344	1.3908	0.9048	4.0222	41.2922	105.9782
χ^2(1 df)		2.8302; p=0.0925			151.2926; p=9.05e^{-35}	

Exercise 9 [Part 2]

Cystic fibrosis (CF) is an autosomal, recessive disease. Homozygotes individuals (rr) have reduced function of the salt transporters in some cell types, including the lungs. This entails mucus in the airways which increases the risk of bacterial infections.

Until recently, only few individuals survived to adulthood. In Europe, the frequency of newborns with CF has historical been 1:2500. This rate is maintained by a balance between selection and mutation.

In a study blood samples were collected to determine the genotype frequency of a certain CF mutation. The disease-causing allele is called r, while the non-disease-causing allele is called R. The genotypes of the enrolled participants are shown in the table below.

WT|WT	WT|F508del	F508del|F508del
16890	855	18

i) Compute the allele frequency in the sample

Alleler total =2*(17763) =35526

Allelfrekvens R = (2*16890+855)/35526 = 0,975

Allelfrekvens r = (855+2*18)/35526 = 0,025

ii) How many individuals do you expect to be born with each genotype if the sample is in Hardy-Weinberg proportions?

RR = p^2 = 0,975²*17763 =16886

Rr = $2pq$ =2*0,975*0,025*17763= 866

rr = q^2 =0,024²*17763 = 11

iii) Is the sample in Hardy-Weinberg proportions?

iv) Compute what the mutation rate (μ) must be, when there in the population through many generations has been maintained a frequency of 1:2500 among newborns and when the fitness (w=1-s) of CF-individuals is 0%.

Mutationsrate, $\mu = q^2 * s$

Frekvens af q^2 = 1/2500 = 0,0004

Fitness = 1-s = 0 % => s = 1, selektionskoefficient

Mutationsrate, $\mu = q^2 * s$ = 0,0004 * 1 = 0,0004

Assume that a new medical treatment for CF has been developed such that patients with CF live longer and many of the CF-carriers also get children. This entails that the fitness of individuals with CF is changed to 40% compared to persons without CF.

v) At equilibrium, what will the new frequency be of the CF-allele r, and what will the frequency be of newborns with CF?

Fitness= 40 % => selektionskoefficienten *s* er lig 0,6
Ny ligevægt, allelfrekvenser:
q=√(μ /s) => √(0,0004/0,6) =0,0258
Ny frekvens af nyfødte med CF: q^2 = 0,0006667 dvs ca 1:1500

Exercise 10 [part 2]

Out of 2,400 births at a UK hospital, 6 newborn deaths were recorded, all of which were attributed to a colon defect caused by a recessive lethal allele (co) that follows an autosomal inheritance pattern. Assuming Hardy-Weinberg equilibrium, answer the following:

i. What is the frequency of the co allele in the newborn population?
P(*co co*)=6/2400=0.0025
P(*co*)=0.0025^0.5=0.05

ii. What is the frequency of heterozygous carriers in both newborns and adults?
New-born: P(co+ co) = 2pq = 2*0.95*0.05=0.095
Adults: P(co+ co) = 2pq / (1-q^2) = 0.095 / (1-0.0025)=0.0952

Exercise 11 [part 2]

Red-green colour blindness is caused by a recessive, X-linked gene (f). In a large population, the frequency of the f allele is denoted as q in both males and females. Assuming random mating and no selection, answer the following:

i. What is the expected frequency of colour-blind females in the population?

		Females		
		p^2	$2pq$	q^2
Males	p			
	q			

Thus, there must be q^2 females that are colour-blind.

ii. What is the expected frequency of marriages between a colour-blind male and a non-colour-blind female?

		Females		
		p^2	$2pq$	q^2
Males	p			
	q			

$P(c\text{-}man\ c\ nc\text{-}women) = qp^2 + q(2pq) = qp^2 + 2pq^2$
$= q(1-q)^2 + 2(1-q)q^2$
$= q(1^2 + q^2 - 2x1q) + 2(1-q)q^2$
$= q + q^3 - 2q^2 + 2q^2 - 2q^3$
$= q - q^3 = q(1-q^2)$

q is the proportion of colour-blind males, and $(1-q^2)$ is the proportion of non-colour-blind females.

iii. What is the expected frequency of colour-blind boys born from a colour-blind father and a non-colour-blind mother?

		Females		
		p^2	$2pq$	q^2
Males	p			
	q			

How many colour-blind boys out of all boys = ½ x $q(2pq)=pq^2$
Colour-blindness among boys from that specific marriage = $pq^2 / q(1-q^2) = pq/1-q$

Exercise 12 [part 2] – In R

Exercise 13 [part 3]

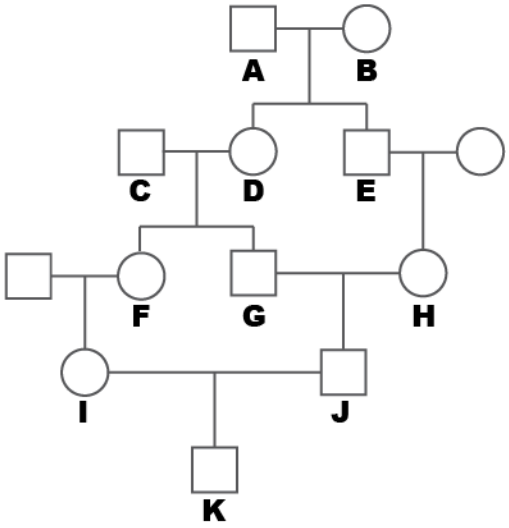
1. Given a population with an average inbreeding coefficient (F) and allele frequencies p (for allele A) and q (for allele a), what are the expected genotype frequencies at this locus? Additionally, how does inbreeding influence both allele frequencies and genotype frequencies?

A population in Hardy-Weinberg proportions with average inbreeding of *F*, there will be excess of homozygotes of *pqF*, and *2pqF* fewer heterozygotes.

AA: p^2+pqF
Aa: $2pq-2pqF$
aa: q^2+pqF

With inbreeding the allele frequencies are not affected.

2. Compute the inbreeding coefficient for individual K in the pedigree below.



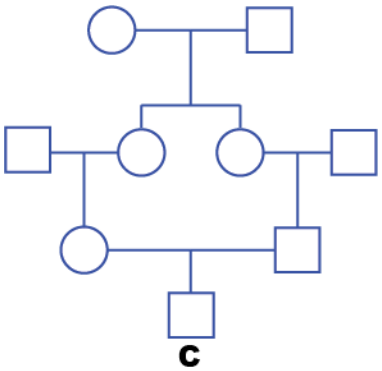
The inbreeding coefficient F is calculated as: $F_X = \sum_A \left(\frac{1}{2}\right)^{n_A} (1 + F_A)$

In the pedigree we can see the following chains where the ancestor is underlined:

IF<u>C</u>GJ	5 individuals	(1/2)^5	=1/32	=4/128
IF<u>D</u>GJ	5 individuals	(1/2)^5	=1/32	=4/128
IFD<u>A</u>EHJ	7 individuals	(1/2)^7		=1/128
IFD<u>B</u>EHJ	7 individuals	(1/2)^7		=1/128

Thus, in inbreeding coefficient becomes: $F_K = \frac{10}{128} = 5/64$

3. It turns out that individual C in the pedigree above is the offspring of a cousin marriage. Redraw the pedigree focusing on individual C, and use this updated information to calculate the inbreeding coefficient (F) for individual K.



IFCGJ	5 individuals	$(1/2)^5(1+1/16)$	=1/32(1+2/32)	=34/1024=17/512
IFDGJ	5 individuals	(1/2)^5	=1/32	=16/512
IFDAEHJ	7 individuals	(1/2)^7		=1/128=4/512
IFDBEHJ	7 individuals	(1/2)^7		=1/128=4/512

$$F_K = \frac{17}{512} + \frac{16}{512} + \frac{4}{512} + \frac{4}{512} = 41/512$$

Without accounting for individual C the inbreeding coefficient was 5/64=40/512, thus a minimal increase in inbreeding.

Exercise 14

The ability to taste the bitter compound phenylthiocarbamide (PTC) is controlled by a dominant allele (T), meaning individuals with at least one T allele can perceive its bitterness, while tt homozygotes cannot. PTC tasting is linked to variation in the *TAS2R38* gene, which encodes a bitter taste receptor on the tongue.

PTC itself is not found in foods but is chemically similar to naturally occurring bitter compounds in certain vegetables, such as broccoli, Brussels sprouts, kale, and cabbage. The ability to taste these compounds may have evolutionary significance, as it could help individuals avoid toxic plants that contain bitter-tasting alkaloids.

i) In a study of 100 students, 62 individuals had the TT genotype (strong tasters), 32 had the Tt genotype (moderate tasters), and 6 had the tt genotype (non-tasters). Calculate the allele frequencies of T and t in this population.

Svar A: Der er 100 studerende, hvorfor der er 200 alleller. Hver homozygot bidrager med 2 og hver heterozygot med en:

$$f_T = \frac{2 \cdot \text{Antal TT} + \text{Antal Tt}}{200} = \frac{2 \cdot 62 + 32}{200} = \frac{156}{200} = 78\%$$

$$f_t = \frac{2 \cdot \text{Antal tt} + \text{Antal Tt}}{200} = \frac{2 \cdot 6 + 32}{200} = \frac{44}{200} = 22\%.$$

Dette er ren optælling og kræver ingen særlige antagelser.

ii) Using the observed genotype frequencies, assess whether the population is in Hardy-Weinberg equilibrium by comparing expected and observed proportions.

Svar B: Er der H-W vil $P(TT) = P(T)^2$, $P(Tt) = 2P(T)P(t)$ og $P(tt) = P(t)^2$, så vi kan sammenligne med de observerede: Vi ser, der er rimelig overensstemmelse. Men vi burde foretage et statistisk test (særligt interesserede kan se nedenfor).

Dvs. vi kan ikke afvise, der er H-W-*proportioner*, og dermed kan vi ikke afvise, der er H-W-ligevægt. Men vi kan faktisk ikke konkludere, at der er HWE, for vi ved ikke, om der er genetiske kræfter på spil, eller om der er tilfældig paring.

Vi kan lave et χ^2-test for H-W-proportioner, men det er ikke pensum:

$$X^2 = \sum_i \frac{(O-E)^2}{E} = \frac{(62-60.8)^2}{60.8} + \frac{(32-34.3)}{34.3} + \frac{(6-4.8)^2}{4.8}$$
$$\approx 0.022 + 0.157 + 0.278 = 0.46.$$

Eftersom dette er mindre end 3.84, som er grænsen for, hvornår et χ^2-test med 1 frihedsgrad er signifikant, så kan vi ikke afvise, at der er H-W-proportioner.