

# Population genetics (*PCB4553/6685*) Midterm take-home exam

**Due: 3/6/2023 11:59PM**

**Please show your calculations**

# Q1 5pts

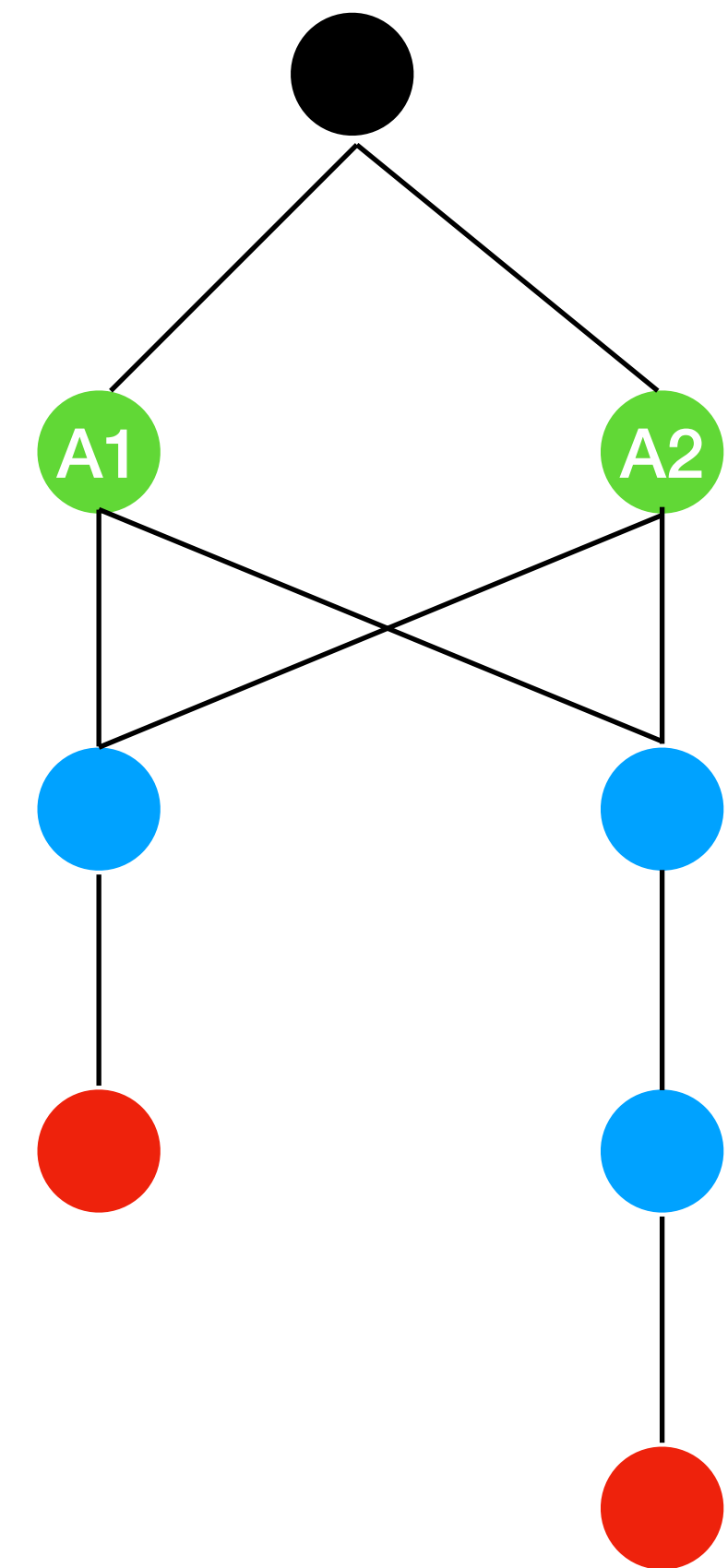
- Suppose we have a hypothetical plant population
  - There is a locus that controls the phenotype of flower color with two alleles  $A$ ,  $a$
  - $AA$  produces yellow flowers,  $aa$  produces white flowers
  - $A$  is dominant over  $a$
  - Frequency of  $A$  is 0.9
  - Assume random mating
- **What is the probability that a pair of half-sibs both produce yellow flowers?**

# Answer

- For half-sibs
  - Probability sharing 1 allele =  $1/2$
  - Probability sharing 0 allele =  $1/2$
- If sharing 0 allele:
  - $P(\text{yellow flowers} \mid \text{sharing 0 allele}) = P(\text{not aa aa} \mid \text{sharing 0 allele}) = 1 - .1^4 = 0.9999$
- If sharing 1 allele:
  - $P(\text{yellow flowers} \mid \text{sharing 1 allele}) = P(\text{not aa aa} \mid \text{sharing 1 allele}) = 1 - .1^3 = 0.999$
- Putting together:
- $P(\text{yellow flower} \mid \text{half-sibs}) = 1/2 * 0.9999 + 1/2 * 0.999 = 0.99945$

# Q2

1. Explain in math or plain language why  $1 + F_{A_i}$  appears in Eq. 2.11 in the book for the calculation of the kinship coefficient.
2. Given the genealogy on the right, assuming  $F_{A_1} = 0.3$  and  $F_{A_2} = 0.2$  for the two ancestors colored green, calculate the kinship coefficient for the two individuals colored red.
3. Do we need to consider the ancestor colored black in the calculation? Why?



# Answer

- $F_A$  accounts for the probability that two alleles, if both traced back to *different* alleles in the common ancestor coalescing back in time due to the inbreeding of the common ancestor
- $(1/2)^6 \times (1+.3) + (1/2)^6 \times (1+.2) = 0.0390625$
- No, we don't since the  $F_A$  for both more recent common ancestors are given.

# Q3

- Using the genotype frequencies for the two populations on the right, answer:
  - Are the two populations in linkage disequilibrium?
  - If the two populations are mixed in equal proportions, what is the initial LD measured in  $D$ ?
  - Assuming the recombination rate between the two loci is  $c = 0.1$  per generation, how many generations will it take for LD to be reduced to  $1/10$  of the current value?

Population	P_AB	P_Ab	P_aB	P_ab
1	0.02	0.18	0.08	0.72
2	0.12	0.28	0.18	0.42

# Answer

- 1. The two populations are in linkage equilibrium
- 2.
  - $P_A = .3$ ;  $P_a = .7$ ;
  - $P_B = .2$ ;  $P_b = .8$ ;
- $D = P_A \times P_B - P_{AB} = .3 \times .2 - .07 = .01$
- 3. Use Eq. 3.21
  - $D^t = (1 - c)^t D^0$
  - $D^t / D^0 = (1 - c)^t$
  - $.1 = (1 - c)^t$
  - $.1 = 0.9^t$
  - $t \approx 22$

# Q4

- Show why  $F_{ST}$  is also related to the correlation between alleles.



- See derivation on whiteboard

# Q5

- The average  $\pi$  for the **X**-chromosome of a mammalian species is 0.004 per base pair. Assuming a mutation rate of  $10^{-8}$  per basepair, what is your estimate for the effective population size  $N_e$ ?

- $\pi = 3N_e\mu$

- $4 \times 10^{-3} = 3N_e 10^{-8}$

- $N_e \approx 1.3 \times 10^5$

# Q6

- Use the McDonald-Kreitman table on the right

1. Calculate the McDonald-Kreitman test score

2. What is your conclusion about if this gene is under selection?

3. Why can we use the MK test to detect selection? Please provide your intuition.

4. Use the appropriate statistical test to see if this result is significant (feel free to use softwares like *R*)

	no. polymorphism	no. fixed substitutions
Non synonymous	45	100
Synonymous	60	400

# Q6

- Use the McDonald-Kreitman table on the right
  1.  $DN/DS < PN/PS$
  2. This gene is under negative selection
  3. N/S ratio is expected to be one under neutrality. If  $DN/DS$  is lower than  $PN/PS$ , this indicates that polymorphisms are being fixed at a lower rate than expected, providing evidence for negative selection. The intuition is similar for positive selection.
  4. Fisher exact test ( $p < 0.001$ )

	no. polymorphism	no. fixed substitutions
Non synonymous	45	100
Synonymous	60	400
Ratio	0.75	0.25

# Q7

- We have genotyped 5 loci for a mainland and an island population of some bird species
- Assuming the migrate rate from the mainland to the island is 0.01 per generation, what is the estimation of the effective population size on the island?

	Locus1	Locus2	Locus3	Locus4	Locus5
Island	0.1	0.2	0.08	0.9	0.4
Mainland	0.5	0.7	0.6	0.4	0.3

# Relating $F_{ST}$ to population history - island model

- probability that our lineages coalesce either allele migrates back to mainland:

- $\frac{1/2N_I}{1/2N_I + 2m}$

- Level of heterozygosity on the island is:

- $H_I = \left(1 - \frac{1/(2N_I)}{1/(2N_I) + 2m}\right) H_M$

- Reduction of heterozygosity on the island

- $F_{IM} = 1 - \frac{H_I}{H_M} = \frac{1/(2N_I)}{1/(2N_I) + 2m} = \frac{1}{1 + 4N_I m}$

- considering the island as our sub-pop,  $F_{IM} = F_{ST}$

- $$F_{IM} = 1 - \frac{H_I}{H_M} = \frac{1/(2N_I)}{1/(2N_I) + 2m} = \frac{1}{1 + 4N_I m}.$$

- $$F_{IM} = 1 - 0.26/0.46 = 0.43$$

- $$0.43 = \frac{1}{1 + 4N_I \times 0.01}$$

- $$N_I \approx 33.1$$

	Locus1	Locus2	Locus3	Locus4	Locus5	H
Island	0.1	0.2	0.08	0.9	0.4	
H	0.18	0.32	0.1472	0.18	0.48	0.26144
Mainland	0.5	0.7	0.6	0.4	0.3	
H	0.5	0.42	0.48	0.48	0.42	0.46



# Q8

- On the right, we have the survival probability for the three genotypes in a large diploid population
  - What type of selection is acting on the population?
  - Re-express the survival probabilities in terms of relative fitness suitable for this type of selection
  - Assuming  $P_A = 0.01$ , calculate the generations it takes for  $A$  to reach a frequency of 0.1
  - If instead the survival probability of  $Aa$  is 0.6, what type of selection do we have? And what is the equilibrium frequency for allele  $A$ ?

Genotype	Survival probability
AA	0.5
Aa	0.3
aa	0.1

- Generations to reach certain frequency in diploids

- $$\tau = -\log \left( \frac{p_\tau q_0}{q_\tau p_0} \right) / \log(1 - s)$$

- $$\tau \approx \frac{1}{s} \log \left( \frac{p_\tau q_0}{q_\tau p_0} \right)$$

- Starting frequency
  - $p_0 = 1/(2N)$
- Near fixation frequency
  - $p_\tau = 1 - 1/(2N)$
- $$\tau \approx \frac{4}{s} \log(2N)$$
- Compared with the haploid time:
  - $$\tau \approx \frac{2}{s} \log(N)$$

# Q8

- On the right, we have the survival probability for the three genotypes in a large diploid population

1.What type of selection is acting on the population?

- Directional selection

2.Re-express the survival probabilities in terms of relative fitness suitable for this type of selection

- See table

Genotype	w
AA	1
Aa	0.6
aa	0.2

# Q8

3. Assuming  $P_A = 0.01$ , calculate the generations it takes for  $A$  to reach a frequency of 0.1

- $\tau = -\log \left( \frac{p_\tau q_0}{q_\tau p_0} \right) / \log(1 - s)$

- $s = 0.8$

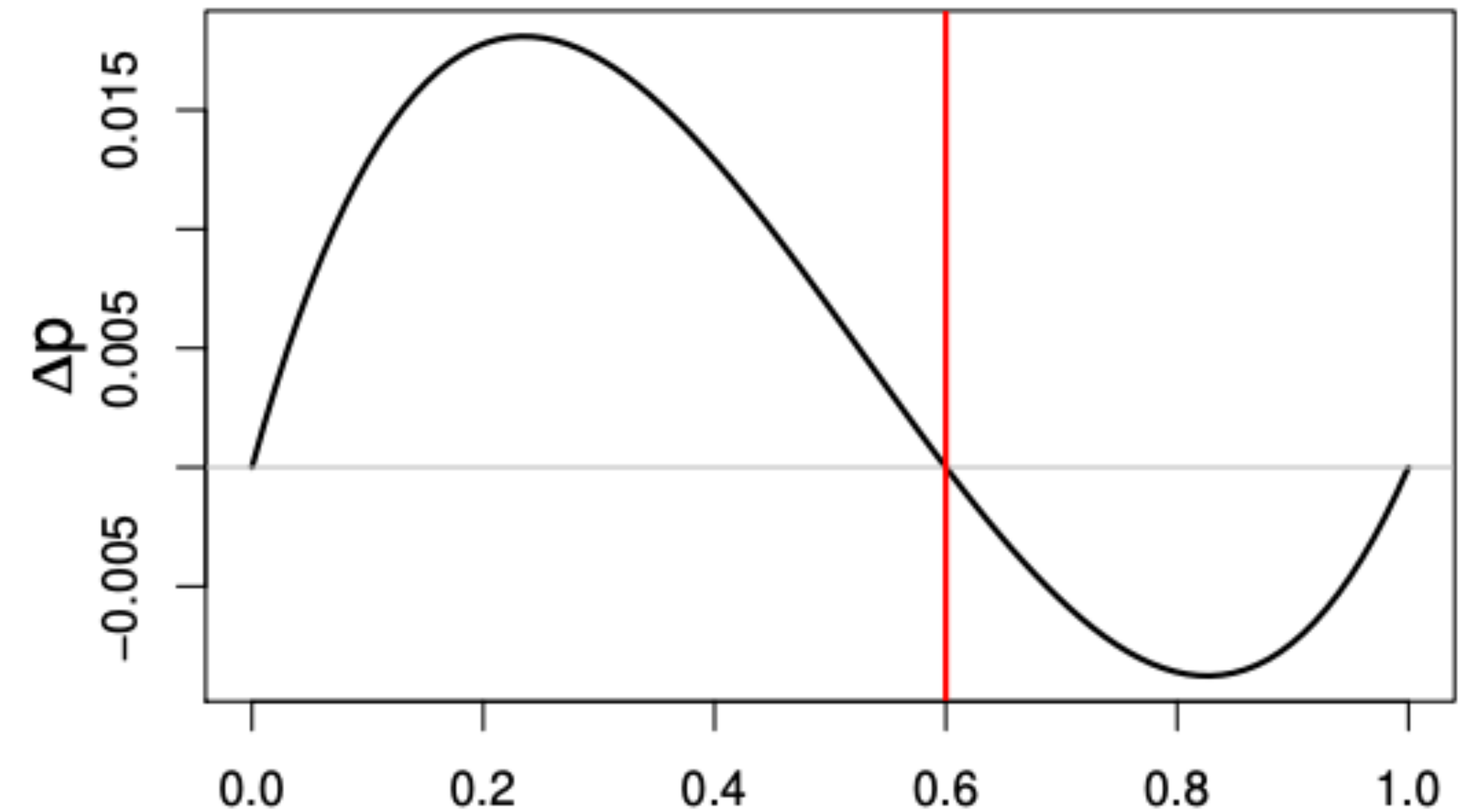
- $\tau = -\frac{2}{\log(1 - 0.8)} \log\left(\frac{0.1 \times 0.99}{0.9 \times 0.01}\right) \approx 3$

4. If instead the survival probability of  $Aa$  is 0.6, what type of selection do we have? And what is the equilibrium frequency for allele  $A$ ?

Genotype	w
AA	1
Aa	0.6
aa	0.2

# Review: Balancing selection with heterozygote advantage

- equilibrium frequency by setting  $\Delta p_t = 0$
- $p_t q_t (w_1 - w_2) = 0$
- Equilibria at
- $p = 0$  or  $q = 0$
- but also
- $p_e = \frac{s_2}{s_1 + s_2}$
- Stable equilibrium!



# Q8

4.If instead the survival probability of  $Aa$  is 0.6, what type of selection do we have? And what is the equilibrium frequency for allele  $A$ ?

$$\bullet s_1 = 0.17$$

$$\bullet s_2 = 0.83$$

$$\bullet p_e = \frac{s_2}{s_1 + s_2}$$

$$\bullet p_e = 0.83$$

Genotype	Survival probability	Relative fitness
AA	0.5	0.83
Aa	0.6	1.00
aa	0.1	0.17

Genotype	Relative fitness
AA	0.5
Aa	0.6
aa	0.1