

**ANSC 446 / IB 416 Population Genetics  
Problem Set 3**

1. Assume that the following mtDNA sequences were found in four different individuals.

AATCGAGACTTTAGC  
ATTCCAGATTTAAGC  
ATTCCAGATTTAAGC  
AATCGAGACTTTAGC

- a) How many homologous nucleotide sites are aligned?

Answer: **15**

- b) How many sites are segregating?

Answer: **4**

- c) What is the  $p$  distance between the first and second sequences?

Answer:  $4/15 = .2667$

- d) How many pairwise comparisons are possible for this sample?

Using unique sequences:

$$\text{Answer: } \frac{S(S-1)}{2} = \frac{2(2-1)}{2} = 1$$

Using all (even non-unique sequences):

$$\text{Answer: } \frac{S(S-1)}{2} = \frac{4(4-1)}{2} = 6$$

- e) Estimate nucleotide diversity.

Designate two unique sequences:  $A_1$  (sequences 1, 4),  $A_2$  (sequences 2, 3)

Frequencies  $P(A_1) = 0.5$ ,  $P(A_2) = 0.5$

$\pi_{ij}$  for comparisons:  $A_1-A_2$  is  $4/15 = 0.2667$

$$\pi\text{-hat (estimate)} = \frac{(N/(N-1)) \sum (p_i \times p_j) \times \pi_{ij}}{2} \\ = 4/3 \times [(2 \times 0.5 \times 0.5 \times 0.2667)] = \mathbf{0.1778}$$

Warning: this is the simplest case (2 unique sequences); see old exam for more difficult examples. Hint: **draw a unit square** to see why the “2” appears in the equation.

Note: an alternative formula not in the book is to compare all sequences, unique or not:

$$\pi = \left( \frac{2}{s(s-1)} \right) \left( \frac{1}{N} \right) [\pi_{12} + \pi_{13} + \pi_{14} + \pi_{23} + \pi_{24} + \pi_{34}]$$

$$\pi = \left( \frac{1}{6} \right) \left( \frac{1}{15} \right) [4 + 4 + 0 + 0 + 4 + 4] = 16 / (6)(15) = .1778$$

- f) How many transitions are present between the first and the second sequences?

Answer: 1

2. An AFLP marker in *Ambystoma* salamanders was found to be heterozygous in 390 salamanders and homozygous in 100 salamanders. Estimate the effective number of alleles at this marker.

$$\text{Answer: } n_e = 1 / 1 - H = 1 / 1 - 390/490 = 1/.2040 = 4.9 \sim 5 \text{ alleles}$$

where H = heterozygosity.

3. Two populations of lions were sampled and found to have the following allele frequencies for two SNP sites:

	Site 1		Site 2	
	G	C	G	A
Ten African lions	.10	.90	.30	.70
Forty Indian lions	.45	.55	.35	.65

- a) Estimate the mean allele frequency of G at Site 1.

$$\text{Answer } \bar{p} = \frac{10(.10) + 40(.45)}{50} = 19/50 = .38$$

- b) Estimate the weighted variance of allele frequency of G at Site 1.

$$\begin{aligned} \text{Answer: } \hat{V(P)} &= [10/50](.10)^2 + [40/50](.45)^2 - (.38)^2 \\ &= .002 + .162 - .1444 = .0196 \end{aligned}$$

- c) Calculate the  $X^2$  value for the two alleles at Site 1 (to be used towards determining if the two populations are different).

$$\text{Answer: } X^2 = \frac{2(50)(.0196)}{(.38)(.62)} = 8.319$$

d) Calculate genetic identify and its three components for Site 1.

$$\text{Answer: } J_{xy} = \sum_{i=1}^n P_{ix} \cdot P_{iy} = (.10)(.45) + (.90)(.55) = .540$$

$$J_x = \sum_{i=1}^n P_{ix}^2 = (.1)^2 + (.9)^2 = .820$$

$$J_y = \sum_{i=1}^n P_{iy}^2 = (.45)^2 + (.55)^2 = .505$$

$$I = \frac{J_{xy}}{(J_x J_y)^{1/2}} = \frac{.54}{((.82)(.505))^{1/2}} = \frac{.54}{.6435} = .839153 \sim .839$$

e) Estimate the standard genetic distance between the two lion populations at Site 1.

$$\text{Answer: } D = -\ln(I) = -\ln(.839153) = .175362 \sim .175$$

4. Coat color in horses is determined by multiple alleles. A complete black horse (C1 black horse with black mane and tail) is dominant to a bay horse (C2 brown horse with black legs, mane and tail) and a mahogany bay. A Bay is dominant to a mahogany bay (C3 brown horse with black roots, legs, mane, and tail). Your sample has 8000 horses (Black, bay, and mahogany bay).

Color	Observed Number
Black	4000
Bay	2400
Mahogany Bay	1600

a) Estimate the allele frequencies for C1, C2, and C3.

$$\text{Answer: } P_{C3} = \sqrt{1600/8000} = \sqrt{.2} = .447$$

$$P_{C2} = \left[ \frac{(2400 + 1600)}{8000} \right]^{1/2} - (.2)^{1/2} = .260$$

$$P_{C1} = 1 - (.447214 + .259893) = .293$$

**Hint: draw a unit square**

5. A population of caracals was sampled to determine the weight of adult males. The arithmetic mean weight of the sampled caracals was 15 kg, with a standard deviation of 3 kg.

a) Assuming a normal distribution, within what weight range would 95% of adult male caracals be expected to fall?

Hint: as in Hedrick's Figure 1.3, draw a bell shaped curve of a normal distribution showing that 68% of samples fall within one standard deviation of the mean, and 95% fall within 1.96 standard deviations of the mean.

Answer: mean  $\pm$  (1.96 X standard deviation) = 15  $\pm$  (1.96X3)  
 = 15  $\pm$  5.88 kg. **Weight range is 9.12 to 20.88 kg**

b) If 300 adult males were sampled, how many would be expected to weigh between 12 and 15 kg?

Answer: Since the mean is 15 kg, and standard deviation is 3 kg, then 12 kg is one standard deviation below the mean. Given a normal distribution, 68% of samples fall within 1 sd of the mean, with one half of them below the mean, and one half of them above the mean. Therefore 34% would weigh 12-15 kg, and  $0.34 \times 300$  samples = **102 adult male caracals**