Tentamentsskrivning i **TMS106/MSA610: Population genetics, 7.5 hp**

Tid: Måndagen den 25 maj 2009 kl 8.30-12.30.

Lärare och Jour: Serik Sagitov, tel. 772-5351, mob.tel. 073-6907613, rum MV H3026.

Hjälpmedel: Räknedosa utan manualer och med tömda minnen, EGEN formelsamlingen fyra A4 sidor.

Grading system (CTH):	marks	0-11	12-17	18-23	24-30
	grade	U	3	4	5

Grading system (GU):	marks	0-11	12-20	21-30
	grade	U	G	VG

1. (5 marks) A male S has recently married G. His son R from a previous marriage has met A, a younger sister of G. If both couples S-G and R-A produce daughters, how genetically close would be the girls? As close as half-sisters (same father different mothers)? Explain by comparing relevant coefficients.

2. (5 marks) A population at steady state in the infinite-alleles neutral model has a homozygosity F equal to 12.5%.

   (a) With random mating, how many equally frequent alleles would be required to produce the same level of homozygosity?

   (b) Given the mutation rate per site per generation $\mu = 3 \times 10^{-8}$ estimate the effective population size. Assume that the gene in question consists of 500 nucleotides.

3. (5 marks) Some subpopulations of *Drosophila melanogaster* show an altitudinal gradient in the allozymes of alcohol dehydrogenase in which the frequency of the *Adh-F* allele increases with altitude.

   The data in the accompanying table are estimates of the allele frequency of *Adh-F* in seven samples of adult flies captured either in the mountains, in the foothills, or on the plains of the Caucasus Mountains of the former Soviet Union.

Elevation	Allele frequency
Mountain	0.321
Mountain	0.226
Foothill	0.131
Foothill	0.109
Plain	0.088
Plain	0.082
Plain	0.035

Each allele frequency is based on electrophoresis of approximately 300 adult flies.

(a) Calculate the F statistics F_{SE} (subpopulations within elevations), F_{ET} (elevations within the total), and F_{ST} (subpopulations relative to the total).

(b) What do the magnitudes of the F statistics suggest regarding genetic differentiation among subpopulations in the frequency of *Adh-F* with respect to altitude?

4. (5 marks) In the ABO blood group system, there are two functional alleles, A and B. Alleles A and B control transferase enzymes that connect the proper sugar molecule (glucosamine or n-acetyl glucosamine) to a common precursor substance.

   Most likely, B was the result of a rare mutation of the A allele. O is a nonfunctional allele that recognizes no substrate, and no sugar molecule is transferred, leaving the precursor unchanged. In the ABO system, O is now the most frequent allele. If there is no selective advantage, O should continue to increase at the expense of A and B. Explain why.

5. (5 marks) In some studies, married couples have higher correlation coefficients for intelligence than do siblings. In modern western culture, we tend to marry someone who is about our own intelligence, although this is probably an over simplification.

   If intelligence were controlled by a single genetic locus with two alleles, S for smart and D for dumb, then three phenotypes would be possible, SS for smart persons, SD for persons with average intelligence, and DD for persons who are mentally challenged.

   Of course, we know that intelligence is a multifactorial trait and not a single gene trait, but it is interesting to see what happens if it were a single gene trait with assortive mating where smart persons were only allowed to mate with smart persons, average persons with average persons, and mentally challenged only with mentally challenged.

   (a) Are the gene frequencies going to change?

   (b) What would happen to the genotype frequencies after many generations?

6. (5 marks) Two pure lines of corn have the following mean and standard deviations as their ear lengths are measured in cm:

   Tom Thumb (P1): $\mu = 16.80$, $\sigma = 0.817$
   BMS (P2): $\mu = 6.63$, $\sigma = 1.887$

   Generation F1 is formed by crossing these two pure lines, and afterwards randomly mating F1 individuals produce generation F2:

generation F1: $\mu = 12.12$, $\sigma = 1.519$
generation F2: $\mu = 12.89$, $\sigma = 2.252$

(a) What are the major sources of variation within each of the two pure lines? Draw two curves over the same pair of axes summarizing phenotypic variation in P1 and P2.

(b) Discuss the relationship between the means and standard deviations for F1 on one hand, and P1, P2 on the other hand.

(c) Why it is not surprising that phenotypic variance in F2 is larger than in F1. Estimate heritability of the trait value and explain your calculations.

Partial answers and solutions are also welcome. Good luck!

Short answers

Problem 1. The coancestry coefficient - the IBD probability for a theoretical offspring for the daughters in question, is equal to $\left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 = \frac{1}{8}$. Since this number is exactly the coancestry coefficient for half-sisters, we conclude that these girls are genetically as closely related as half-sisters.

Problem 2a. According to the HWE formula for k alleles with frequencies $\frac{1}{k}$

$$0.125 = \left(\frac{1}{k}\right)^2 + \ldots + \left(\frac{1}{k}\right)^2 = \frac{1}{k}$$

and the answer is $k = 8$ alleles.

Problem 2b. From $F = \frac{1}{1+\theta}$ we obtain $\theta = 7$. Here $\theta = 4N_e\mu$, with $\mu = 3 \times 10^{-8} \times 500 = 1.5 \times 10^{-5}$ being the mutation rate per gene per generation. It follows
$$N_e = \frac{7}{6} \times 10^5 \approx 120000.$$

Problem 3a. (See page 284 of the 4th edition.) Three heterozygosity frequencies $H_S = 0.2266$, $H_E = 0.2285$, and $H_T = 0.2433$ give three fixation indices $F_{SE} = 0.008$, $F_{ET} = 0.061$, $F_{ST} = 0.068$,

Problem 3b. We see that most of the differentiation among subpopulations is correlated with altitude; there is very little genetic differentiation among subpopulations at each elevation.

Problem 4. This is a case of irreversible mutation, because the nonfunctional allele O has negligible probability to mutate back to a functional version.

Problem 5. If the current generation frequencies are D_0 for SS, H_0 for SD, and R_0 for DD, then the assortative mating $SS \times SS$, $SD \times SD$, and $DD \times DD$ leads to the following frequencies in the next generation

$$D_1 = D_0 + H_0/4, \ H_1 = H_0/2, \ R_1 = R_0 + H_0/4.$$

5a. The allele frequencies do not change, since $p_1 = D_1 + H_1/2 = D_0 + H_0/2 = p_0$.

5b. Obviously, in the long run $H_n \to 0$ and we end up with two genetically isolated subpopulations with genotype frequencies p_0 for SS and $1 - p_0$ for DD.

Problem 6a. Environmental variation.

Problem 6b. Both the mean and the standard deviation for F1 lie between those of P1 and P2. If typical genotypes of P1 and P2 are respectively AA

and aa, then the F1 genotype is Aa implying the absence of genetic variation. Clearly, the allele A gives higher trait value compared to a. Notice that all three variances are different even though each of them is caused solely by environmental factors. This indicates about possible interaction between genotypic and environmental deviations.

Problem 6c. The F2 variation should be higher because besides the environmental variation there is present genotypic variation. The broad sense heritability can be computed after averaging the three environmental variances

$$\sigma_e^2 = \frac{0.817^2 + 1.887^2 + 1.519^2}{3} = 2.179.$$

using the formula

$$H^2 = 1 - \frac{\sigma_e^2}{\sigma_p^2} = 1 - \frac{2.179}{2.252^2} = 0.43.$$