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

Tid: Måndagen den 24 maj 2010 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+
	grade	U	3	4	5

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

1. (4 marks) For rare deleterious conditions, even though they are rare there are a lot of carriers in the population, and the ratio of carriers to affected individuals is high.

   Most of these deleterious alleles are in the carrier form to eliminate them you would have to get rid of a large portion of the human population since so many people are carriers. Everyone is a carrier for a few deleterious alleles. So genetically cleansing a population of deleterious alleles is not something that can be done realistically.

   Question. Suggest a formula for the ratio mentioned in the first sentence. Carefully explain how this formula is obtained.

2. (5 marks) A population is maintained at effective population size of 10,000 for a very long time. A set of 30 diploid individuals from that population is taken into the lab, and this lab population is maintained for a total of 8 generations at effective population size 20. At the end of this process, how much heterozygosity should there be on average at a locus with mutation rate 10^{-6}? What assumptions do you make?

3. (4 marks) In a population of the flour beetle Tribolium the mean weight of pupae is 200 mg. The phenotypic variance is 400 mg^2 and the additive genetic variance is 100 mg^2. If individuals with a mean pupa weight two phenotypic standard deviations above the mean are selected, what is the expected average pupa weight among the progeny?

4. (4 marks) What is the equilibrium frequency of a recessive gene arising with mutation rate of $4 \cdot 10^{-6}$ and a reproductive fitness in homozygotes of 0.8? What would it be if the gene were partially dominant with $h = 0.05$?

5. (3 marks) Most protein-coding genes have a forward mutation rate (normal to mutant) that is at least an order of magnitude greater than the reverse mutation rate (mutant back to normal). Why should this be the case?

6. (4 marks) Below comes the abstract of the paper

   Flint-Garcia et al. Structure of linkage disequilibrium in plants. Annu Rev Plant Biol. 2003;54:357-74.

   > Future advances in plant genomics will make it possible to scan a genome for polymorphisms associated with qualitative and quantitative traits. Before this potential can be realized, we must understand the nature of linkage disequilibrium (LD) within a genome. LD, the nonrandom association of alleles at different loci, plays an integral role in association mapping, and determines the resolution of an association study. Recently, association mapping has been exploited to dissect quantitative trait loci (QTL). With the exception of maize and Arabidopsis, little research has been conducted on LD in plants. The mating system of the species (selfing versus outcrossing), and phenomena such as population structure and recombination hot spots, can strongly influence patterns of LD. The basic patterns of LD in plants will be better understood as more species are analyzed.

   Explain how selfing can influence the patterns of LD in plants.

7. (6 marks) The nucleotide sequences shown here are the complete sets of polymorphic nucleotides found in a region of 5 kb in a sample of six chromosomes from corn *Zea mays*. The polymorphic nucleotides are not adjacent, as shown here, but scattered throughout the 5 kb.

chromosome 1	GCCTT	TATGG	CCTGT	ATGAG
chromosome 2	ACTAT	TAAGG	CTTGT	TTGAT
chromosome 3	ACCAC	TGTCG	CCCGT	ACGCG
chromosome 4	GTCAT	TGTGG	TCCTC	TTGAG
chromosome 5	GCTTT	TATGA	CCTTT	ATAAG
chromosome 6	ACCAT	CATGA	CCTTT	ATAAT

   (a) Do the data fit the infinitely many sites model?

   (b) Estimate $\theta = 4N_e\mu$ from both the number of segregating sites and the average number of pairwise mismatches. What does μ stand for in this case?

   (c) Do these estimates seem to be consistent with one another? What might it suggest about the selective forces acting on the sequence polymorphisms?

Partial answers and solutions are also welcome. Good luck!

Short answers

Problem 1. Assuming Hardy-Weinberg equilibrium, a rare deleterious disease can be modelled by three genotypes AA, Aa, aa with genotype frequences (for newborns) $D = p^2$, $H = 2pq$ and $R = q^2$, where the allele frequency q is small. The ration of carriers to affected individuals then becomes

$$\frac{H}{R} = \frac{2p}{q} \approx \frac{2}{q}.$$

Problem 2. Use the formula (assuming infinitely many allaeles model)

$$H_t = 1 - F_t = (1 - \frac{1}{2N_e})^t (1 - F_0)$$

where $N_e = 20$, $t = 8$.

The $H_0 = 1 - F_0$ is computed from $H_0 = \frac{\theta}{1+\theta}$ where $\theta = 4N_e\mu = 0.04$ with $N_e = 10000$ and $\mu = 10^{-6}$ (assuming genetic drift-mutation equilibrium).

The answer is

$$(1 - \frac{1}{40})^8 \frac{0.04}{1.04} = 0.03.$$

Problem 3. Ofspring phenotypic value P_o is related to the parent phenotypic value P through a linear regression model

$$P_o = \mu + \beta(P - \mu) + \epsilon$$

with the slope given by

$$\beta = \frac{h^2}{2} = \frac{100}{2 \cdot 400} = 0.125.$$

If $P - \mu = 2 \cdot \sqrt{400} = 40$, then the predicted P_o is (in mg)

$$P_o = 200 + 0.125 \cdot 40 = 205.$$

Problem 4. Mutation-selection equilibrium for a recessive gene is described by the formula $q = \sqrt{\frac{\mu}{s}} = \sqrt{\frac{4 \cdot 10^{-6}}{0.2}} = 0.45\%$.

Even with a small degree of dominance $h = 0.05$ the answer becomes drastically different $q = \frac{\mu}{sh} = \frac{4 \cdot 10^{-6}}{0.2 \cdot 0.05} = 0.04\%$.

Problem 5. There are many ways to mutate a gene so that it doesn't work so well. But for every one of those mutations there are only a few ways to fix it to return it to wild type efficiency.

Problem 6. Under selfing the deficit of heterozygotes will result in genotypes of the form Ab/Ab, AB/AB, aB/aB, and ab/ab. Recombination will not be effective here in increasing the linkage equilibrium.

Problem 7a. The data does NOT fit the IMS mutation model. Take a pair of informative sites (whose configuration is different from (5,1)) like 1 and 3 in the data set. Each of them split the 6 sequences in two groups (1,4,5)(3,3,6) and (1,3,4,6)(2,5) respectively. It is impossible to draw a tree which will result in such two splittings due to two mutations.

Problem 7b. Two estimates of θ

$$\hat{\theta} = \frac{S}{a_1} = \frac{20/5000}{1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}} = 0.00175$$

$$\pi = \frac{132}{5000 \cdot 15} = 0.00176.$$

Here μ is the number of mutations per nucleotide site per generation.

Problem 7c. Good agreement between these two estimates is in favor of the null hypothesis of neutrality, which claims that the observed genetic variation is mainly due to the equilibrium between the random genetic drift and neutral mutations. Thus the role of the selective forces acting on the sequence polymorphisms is not significant.