Tentamentsskrivning i **TMS106: Population genetics, 5p.**

Tid: Torsdagen den 14 mars 2002 kl 8.45-12.45.
Lärare och Jour: Serik Sagitov, tel. 772-5351, rum MC 1420.
Hjälpmedel: Räknedosa utan manualer och med tömda minnen, egen formelsamlingen fyra A4 sidor, utdelade tabeller

Grading system (CTH):

marks	0-11	12-16	17-21	22-30
grade	U	3	4	5

Grading system (GU):

marks	0-11	12-19	20-30
grade	U	G	VG

1. (6 marks) The number of abdominal bristles in *Drosophila melanogaster* is a classic object of experimental study in quantitative genetics. Suppose that there are three genotypes with the following characteristics:

Genotype	A_1A_1	A_1A_2	A_2A_2
Mean no. of abdominal bristles	120	111	80

Assuming that the frequency of the A_1 allele is 0.4 and the total phenotypic variance in bristle number is 434.8, answer the following questions:

   (a) What is the mean phenotype in the population?

   (b) What is the genotypic variance for bristle number in the population? What is the broad sense heritability of bristle number?

   (c) What is the narrow sense heritability of bristle number?

   (d) Suppose we now apply selection to this population so that the mean number of abdominal bristles among reproductives is 115. What will be the mean number of bristles among progeny in the next generation?

2. (3 marks) For a sex-linked gene, the following genotype frequencies are found: males: dominant (.7), recessive (.3); females: dominant homozygote (.5), heterozygotes (.4), recessive (.1).

   Is this population in Hardy-Weinberg equilibrium?

3. (6 marks) In 1995 a researcher collected 250 DNA samples from pink salmon taken from Auke Creek in Southeastern Alaska. The alcohol dehydrogenase (ADH) gene was sequenced from each sample, and 5 alleles were seen. The numbers of chromosomes with each allele were:

Allele	ADH^1	ADH^2	ADH^3	ADH^4	ADH^5
Number	312	93	42	31	22

(a) Assuming these are selectively neutral alleles that mutate at the rate of 10^{-4} per generation, and the population is in equilibrium, what is the effective population size?

(b) The same researcher returned to Auke Creek in 1996 and collected another sample of similar size. In this case, there were 215 ADH^1 chromosomes, and 285 of the other four alleles combined. Pooling the samples, the researcher found the gene frequency of ADH^1 to be 0.527, and estimated that the frequency of ADH^1/ADH^1 homozygotes to be 0.2777.

However, the researcher did not realize that pink salmon breed on a strict 2-year cycle. Therefore, the even-year breeders and the odd-year breeders are distinct populations.

What is the true frequency of ADH^1/ADH^1 homozygotes in Auke Creek.

(c) What is F_{ST}? (Hint: Since only the ADH^1 count is given for the 1996 sample, you may treat the other four alleles as though they were a single allele.)

4. (4 marks) A population is polymorphic for an EcoRI restriction site in a defined chromosomal region. 75% of the chromosomes have the cut site, and 25% do not. A closely linked mutation then occurs that causes a HindIII cut site to appear. This mutation occurs on a chromosome without the EcoRI cut site.

With subsiquent evolution, the frequency of the HindIII cut site evolves to a current value of 0.1, although the frequency of the EcoRI cut site remains at 0.75. During this time, no recombination events occur between the EcoRI and HindIII sites.

(a) Calculate the linkage disequilibrium between the two restriction sites in the current population.

(b) Suppose that the initial and current frequencies of the EcoRI cut site had both been 0.2 instead of 0.75, but all else is as described above. Calculate the linkage disequilibrium between the two restriction sites in the current population.

(c) Compare these two results.

5. (4 marks) Assume that the mutation rate from A to a in a population is 5×10^{-6} and that selection acts against the recessive genotype (aa) with $s = 0.1$.

(a) What is the expected frequency of a at equilibrium if genotype aa is fully recessive in fitness?

(b) If there is partial dominance with respect to fitness such that the degree of dominance $h = 0.5$, what is the expected frequency of the recessive gene a at equilibrium?

(c) What is the mutational load L in these populations? Explain the meaning of the parameter L.

6. (3 marks) A plant is known to have an inbreeding coefficient of 0.15. What is the inbreeding coefficient of the offspring if:

   (a) the plant is self-fertilized?

   (b) it is crossed with an unrelated plant?

7. (4 marks) *Jingwei* is a gene of unknown function that evolved relatively recently in *Drosophila teissieri* and *Drosophila yakuba* after coopting some of the exons of *Adh* gene.

   A set of 10 *jingwei* DNA sequences in *D. teissieri* and 20 *jingwei* DNA sequences in *D. yakuba* has 67 segregating sites with 24 beeing polymorphic within species and 43 sites beeing variable across species but fixed within species.

   21 of the 24 polymorphic sites are synonimous and 3 are nonsynonimous. Among the fixed differences 16 are synonimous replacements and 27 are nonsynonimous.

   (a) Put together a 2x2 contingency table using the data. The ratio of synonimous to nonsynonimous fixed differences is 16:27. Is this kind of ratio one would expect under the neutral mutation model? Explain why.

   (b) Consider the 16:27 ratio for fixed changes as a substitution pattern typical for the *Jingwei* gene. What ratio for polymorphic sites one would expect under the neutral mutation model? Under the natural selection model? Explain why.

   (c) Set a relevant null hypothesis and test it at 1% significance level. What is the P-value of the test?

Partial answers and solutions are also welcome. Good luck!

ANSWERS

1. Assuming HWE

Genotype	A_1A_1	A_1A_2	A_2A_2
Mean no. of abdominal bristles	120	111	80
Frequency	0.16	0.48	0.36

   (a) Phenotypic mean $\mu = 120 \cdot 0.16 + 111 \cdot 0.48 + 80 \cdot 0.36 = 101.28$

   (b) Genotypic variance $\sigma_g^2 = (120 - \mu)^2 \cdot 0.16 + (111 - \mu)^2 \cdot 0.48 + (80 - \mu)^2 \cdot 0.36 = 264.44$, the broad sense heritability of bristle number $H^2 = \frac{264.44}{434.8} = 0.61$.

   (c) Average effect of gene substitution $\alpha = 22.2$, the additive variance $\sigma_a^2 = 236.56$, the narrow sense heritability of bristle number $h^2 = \frac{236.56}{434.8} = 0.54$.

   (d) Response to selection $\mu' - \mu = 0.54(115 - 101.28) = 7.41$, the mean number of bristles among progeny in the next generation $\mu' = 108.69$.

2. X-linked HWE: $D_f = p^2$, $H_f = 2pq$, $R_f = q^2$, $p_m = p_f = p$. Given $\hat{D}_f = 0.5$, $\hat{H}_f = 0.4$, $\hat{R}_f = 0.1$, $\hat{p}_m = 0.7$ we see that $\hat{p}_m = \hat{p}_f = 0.7$.

   We can not test H_0: HWE

   using the chi-square test of goodness of fit witout a given sample size n. The larger is n the further is the discrepancy from the H_0. Let us check for different values of n. If $n = 500$

Genotype	A_1A_1	A_1A_2	A_2A_2
Observed counts	250	200	50
Expected counts	245	210	45

   then observed test statistic $X^2 = 1.13$, df=3-1-1=1, p-value of the test $P = 2(1 - \Phi(\sqrt{1.13})) = 2(1 - \Phi(1.06)) = 0.29$ is too high to reject H_0.

   If $n = 5000$

Genotype	A_1A_1	A_1A_2	A_2A_2
Observed counts	2500	2000	500
Expected counts	2450	2100	450

   then observed test statistic $X^2 = 11.34$. The p-value of the test $P = 2(1 - \Phi(3.37)) = 0.001$ and we reject H_0.

   Conclusion: assuming that the sample size is moderately large we conclude that the population is in HWE.

3. (a) 1995 allele frequencies:

Allele	ADH^1	ADH^2	ADH^3	ADH^4	ADH^5
frequency	0.624	0.186	0.084	0.062	0.044

Assuming infinite alleles model, $F = 1 - H = 0.624^2 + 0.186^2 + 0.084^2 + 0.062^2 + 0.044^2 = 0.437$. Mutation - RGD equilibium $F = \frac{1}{1+\theta}$, where $\theta = 4N_e\mu$. Thus $N_e = \frac{\theta}{4\mu} = \frac{1-F}{4F\mu} = 3223$.

(b) 1995 frequency of ADH^1/ADH^1 homozygotes is $0.624^2 = 0.389$ assuming HWE. While 1996 frequency of ADH^1/ADH^1 homozygotes is $0.43^2 = 0.185$ assuming HWE.

The overall frequency of ADH^1/ADH^1 assuming equal subpopulation sizes is $\frac{0.389+0.185}{2} = 0.287$.

(c) Odd-year $H = 0.469$, even-year $H = 0.490$. Observed heterozygosity (assuming equal subpopulation sizes): $H_S = 0.480$.

Expected heterozygosity under HWE: $H_T = 2 \cdot 0.527 \cdot 473 = 0.499$.

Fixation index $F_{ST} = 1 - \frac{H_S}{H_T} = 0.038$.

4. (a) Gamete frequencies

	EcoRI cut site	EcoRI no cut	Total
HindIII cut site	0.00	0.10	0.10
HindIII no cut	0.75	0.15	0.90
Total	0.75	0.25	1.00

The linkage disequilibrium $d = -0.075$.

(b) Gamete frequencies

	EcoRI cut site	EcoRI no cut	Total
HindIII cut site	0.00	0.10	0.10
HindIII no cut	0.20	0.70	0.90
Total	0.20	0.80	1.00

The linkage disequilibrium $d = -0.020$.

(c) To compare these two results calculate two proportions

(a) $\frac{d}{d_{min}} = 1$ and

(b) $\frac{d}{d_{min}} = 1$.

This equality of two measures of disequilibrium reflect the fact that in both cases no recombinations have occured.

5. (a) Assuming mutation-selection equilibrium the expected frequency of a is $\hat{q} = \sqrt{\mu/s} = 0.0071$ or 0.71%.

(b) Assuming mutation-selection equilibrium the expected frequency of a is $\hat{q} = 2\mu/s = 0.0001$ or 0.01%.

(c) According to the Haldane-Muller principle the mutational load L in these populations are

(a) $L = 0.000005$ and

(b) $L = 0.00001$.

The mutation load is the reduction in average fitness due to harmful genes present in the population under the mutation pressure. In the additive case (b) the mutation load causes the "genetic death" of 0.001% of newborns. The load is twice lower in the fully reccesive case (a) because the death of a mutant homozygote removes two genes whereas the death of a heterozygote removes only one mutant gene.

The load is not necessarily a real detriment to the population, because most species produce more offspring than the resources can support.

6. (a) A plant is known to have an inbreeding coefficient of $F_0 = 0.15$. If the plant is self-fertilized its offspring has the inbreeding coefficient $F_1 = \frac{1}{2} + \frac{1}{2}F_0 = 0.575$.

   (b) If it is crossed with an unrelated plant its offspring has the inbreeding coefficient $F_1 = 0$.

7. (a) Cross classification of differences

	Fixed	Polymorphic	Total
Synonimous	16	21	37
Replacement	27	3	30
Total	43	24	67

In the absence of selection forces each site has the same probability of fixation. The number of synonimous sites is calculated as the number of four-fold synonimous sites plus one-third of two-fold synonimous sites. Therefore the ratio of synonimous to nonsynonimous fixed differences of 16:27 does not contradict the neutral mutation model.

   (b) We consider the 16:27 ratio for fixed changes as a substitution pattern typical for the *Jingwei* gene. Then under the neutral mutation model we expect the same ratio for polymorphic sites. If the gene evolution is under adaptative selection, then we expect fewer nonsynonimous polymorphisms because they are either quickly removed by negative selection or get fixed by positive selection.

   (c) Test H_0: "*Jingwei* gene is selectively neutral" using the McDonald-Kreitman test. Expected counts

	Fixed	Polymorphic	Total
Synonimous	23.75	13.25	37
Replacement	19.25	10.75	30
Total	43	24	67

Observed test statistic $X^2 = 2.52 + 4.53 + 3.12 + 5.59 = 15.76$, df=1. The P-value of the test $P = 2(1 - \Phi(\sqrt{15.76})) = 2(1 - \Phi(3.97))$ is smaller than 0.05%. Reject H_0 at 1% significance level.