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

Tid: Fredagen den 14 mars 2003 kl 8.45-12.45.
Lärare och Jour: Serik Sagitov, tel. 772-5351, mob.tel. 073-6907613, 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. (7 marks)
   - (a) A researcher buys 500 homozygous winged and 500 homozygous wingless flies from the Jackson Labs and puts them into a population cage. Winglessness is recessive. The researcher's fly food is too sticky; wingless flies have a 20% chance of becoming stuck in the food and dying.
     - i. What are the allele and genotype frequencies of the initial population? Calculate the inbreeding coefficient and explain the result.
     - ii. What are selection coefficient s and degree of dominance h for the population cage system?
     - iii. What will be the allele and genotype frequencies of the adult flies of the next generation? (Assume that all death due to stickiness happens before the flies are fully adult.) Calculate the inbreeding coefficient and explain the change from the previous generation inbreeding coefficient.
     - iv. Sketch the likely long-term allele frequencies of this population cage.
   - (b) The researcher's supervisor explains about sticky food, and the researcher corrects her food formula at a point where the frequency of the wingless allele is 0.1. Wingless flies now survive normally.
     - i. What are s and h for the non-sticky situation?
     - ii. What will be the allele frequencies and genotype frequencies in the adult flies of the next generation?
     - iii. Sketch the likely long-term allele frequencies of this population cage. What is the probability that all flies in the cage eventually become wingless?

2. (6 marks) The data in the table are configurations of synonimous polymorphisms in five alleles of the *boss* gene of *D. melanogaster*. In addition to the 15 polymorphic sites there were 433 monomorphic synonimous sites.

T C T C C T G T A G T G C C G
T C C T C C A C G A G G C T G
C A T T C C A C A A T T T C G
C A T C T C A C A G T T C T A
C A T C T C A C A G T T T C A

   (a) Estimate the value of θ from the nucleotide polymorphism S, and give the estimated standard error of the θ estimate assuming no recombination. How this estimate depends on the four nucleotide frequencies in the polymorphic sites?

   (b) Estimate the value of θ from the nucleotide diversity π, and give the estimated standard error of the θ estimate assuming no recombination. How this estimate depends on the four nucleotide frequencies in the polymorphic sites?

   (c) Calculate Tajima's D statistic, and interpret the result.

3. (4 marks) Two highly inbred lines of a plant are crossed to produce an F_1 generation. The F_1 individuals are selfed to produce an F_2. Individuals if the F_2 are then backcrossed to the both inbred lines. What is the inbreeding coefficient of the offspring of these two backcrosses?

4. (4 marks) Using the coalescent theory explain the following signatures in the genealogies left by changes in population size.

   (a) Small populations will have short trees. Large populations will have long trees.

   (b) Growing populations will have more coalescences at the bottom of the genealogy, the tip-ward coalescences are less condensed.

   (c) Structured populations will have unusually long branches near the root of the genealogy.

5. (4 marks) Describe a method of heritability estimation performed under the random mating assumption using resemblance between a parent and an offspring.

6. (5 marks) The data set 1 given on a separate page presents the 11-locus genotypes for two populations (Ghana and N'goye) of the mosquito *Aedes aegypti*.

   (a) Some of the genotypes in locus 11 are missing due to sequencing errors. How would you fill up the missing data? Explain.

(b) Turn to the locus 6 in the Ghana population. List different genotypes, find their frequencies and calculate allele frequencies.

(c) For the same locus 6 in the Ghana population compute the genotype and allele frequencies after collapsing the data to allele “1” and “not-1”. Do these frequencies agree with the HWE model?

(d) The lack of heterozygotes involving allele “1“ of the locus 6 in the Ghana population might be explained in different ways. One possibility is that those heterozygotes have lower fitnesses and are supressed by selection. Would you accept this explanation? Why?

Tables and data supplied

1. Normal distibution table
2. Chi-square distibution table
3. Data set 1

Partial answers and solutions are also welcome. Good luck!

Data set 1

Ghana

1	2	3	4	5	6	7	8	9	10	11
11	11	11	12	11	34	12	11	11	11	11
11	11	11	12	11	22	12	11	11	11	11
11	11	11	12	11	23	11	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	12	11	34	12	11	11	11	
11	11	11	11	11	22	11	11	11	11	
11	11	11	11	11	11	11	11	11	11	
11	11	11	12	11	24	12	11	11	11	
11	11	11	11	12	11	11	11	11	11	
11	11	11	11	11	12	11	11	11	11	11
11	11	11	12	11	24	11	11	11	11	11
11	11	11	12	11	24	12	11	11	11	11
11	11	11	12	11	34	12	11	11	11	11
11	11	11	11	11	11	11	11	11	11	
11	11	11	12	11	22	12	11	11	11	
11	11	11	12	11	34	11	11	11	11	11
11	11	11	12	11	23	12	11	11	11	11
11	11	11	12	11	24	11	11	11	11	11
11	11	11	12	11	22	11	11	11	11	11
11	11	11	12	11	34	11	11	11	11	11
11	11	11	11	12	11	11	11	11	11	11
11	11	11	12	11	23	12	11	11	11	11
11	11	11	12	11	34	11	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	34	11	11	11	11	11
11	11	11	11	11	24	12	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	34	12	11	11	11	11
11	11	11	11	11	24	12	11	11	11	11
11	11	11	11	11	24	11	11	11	11	11
11	11	11	11	11	23	11	11	11	11	11
11	11	11	11	11	22	11	11	11	11	11
11	11	11	11	11	23	11	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	34	12	11	11	11	11
11	11	11	11	11	24	11	11	11	11	11
11	11	11	11	11	34	11	11	11	11	11
11	11	11	11	11	23	12	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	23	11	11	11	11	11

Data set 1 (cont)

N'goye

1	2	3	4	5	6	7	8	9	10	11
11	11	11	11	11	11	13	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	33	11	11	11	11	11
11	11	11	11	11	24	11	11	11	11	11
11	11	11	11	12	11	13	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	14	12	11	11	11	11
11	11	11	11	11	14	13	11	11	11	11
11	11	11	12	11	13	23	11	11	11	11
11	11	11	11	11	14	12	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	11	33	11	11	11	11
11	11	11	11	11	11	12	11	11	11	11
11	11	11	11	11	24	13	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	12	11	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	12	11	11	11	11	11
11	11	11	11	11	11	33	11	11	11	11
11	11	11	11	11	13	12	11	11	11	11
11	11	11	11	11	12	11	11	11	11	11
11	11	11	11	11	11	22	11	11	11	11
11	11	11	11	11	11	13	11	11	11	11
11	11	11	11	11	14	13	11	11	11	11
11	11	11	11	11	12	13	11	11	11	11
11	11	11	11	11	11	13	11	11	11	11
11	11	11	11	11	14	11	11	11	11	11
11	11	11	11	11	11	12	11	11	11	11
11	11	11	11	11	14	12	11	11	11	11
11	11	11	12	11	12	12	11	11	11	11
11	11	11	11	11	11	13	11	11	11	11
11	11	11	12	11	12	11	11	11	11	11
11	11	11	11	12	12	11	11	11	11	11
11	11	11	12	11	11	11	11	11	11	11
11	11	11	11	11	12	11	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	12	11	11	11	11	11
11	11	11	11	11	12	11	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	11	33	11	11	11	11

Data set 1 (cont)

N'goye

1	2	3	4	5	6	7	8	9	10	11
11	11	11	11	11	24	11	11	11	11	11
11	11	11	11	11	12	11	11	11	11	11
11	11	11	11	11	11	23	11	11	11	11
11	11	11	11	11	12	13	11	11	11	11
11	11	11	11	11	11	11	11	11	11	11
11	11	11	11	11	14	11	11	11	11	11
11	11	11	12	11	34	12	11	11	11	11
11	11	11	22	11	23	11	11	11	11	11
11	11	11	11	11	11	33	11	11	11	11
11	11	11	11	11	12	11	11	11	11	11

Answers

1. (a) Assume 1 locus 2 alleles model: AA and Aa are winged and Aa are wingless genotypes. Diploid population size $N = 1000$.
    - i. Initial genotype frequencies $D_0 = 0.5$, $H_0 = 0$, $R_0 = 0.5$, allele frequencies $p_0 = 0.5$ and $q_0 = 0.5$. Inbreeding coefficient $F_0 = 1$ says that the current population consists of two unrelated subpopulations.
    - ii. Assume that all death due to stickiness happens before the flies are fully adult. Relative fitnesses $w_{AA} = w_{Aa} = 1$, $w_{aa} = 0.8$. Selection coefficient $s = 0.2$ and degree of dominance $h = 0$.
    - iii. Assume random mating and neglect RGD effects. Average fitness $\bar{w} = 0.95$. Next generation frequencies of the adult flies $D_1 = 0.25/0.95 = 0.263$, $H_1 = 0.5/0.95 = 0.526$, $R_1 = 0.25 \cdot 0.8/0.95 = 0.211$, $p_1 = 0.526$ and $q_1 = 0.474$. Inbreeding coefficient $F_1 = -0.055$ reveals the excess of heterozygots due to fully dominant selection. The dramatic change in F is due to random mating.
    - iv. Favorable allele A is fully dominant, therefore the allele frequency p_t first will rapidly increase but then the fixation process will slow down because the recessive allele will be hidden in heterozygotes. At this point RGD effects will be more visible as random fluctuations which eventually will lead to the full fixation of allele A.

   (b) Initial allele frequencies $p_0 = 0.9$ and $q_0 = 0.1$, genotype frequencies $D_0 = 0.81$, $H_0 = 0.18$, $R_0 = 0.01$.
    - i. No selection $s = 0$.
    - ii. Pretty close to the initial frequencies.
    - iii. In the absence of selection the RGD produces random fluctuations in the allele frequency p_t curve. Fixation probability of the allele a is $q_0 = 0.1$.

2. Configuration of sites in $n = 5$ aligned sequences of length $l = 448$

Configuration	(5,0)	(4,1)	(3,2)	Total
Number	433	6	9	448
Number of pairwise mm	0	4	6	
Total number of pairwise mm	0	24	54	78

   (a) Nucleotide polymorphism $S = 15/448 = 0.0335$, $\hat{\theta} = S/a_1 = 0.0161$, where $a_1 = 2.083$. Estimated standard error $s_{\hat{\theta}} = 0.0101$ is the square root of $\frac{\hat{\theta}}{la_1} + \frac{a_2\hat{\theta}^2}{a_1^2} = 0.000102$, where $a_2 = 1.424$.

The contribution of a polymorphic site to the value of S does not depend much on the site configuration if the sample size n is large enough.

(b) Nucleotide diversity $\pi = 78/4480 = 0.0174$. Estimated standard error $s_\pi = 0.0114$ is the square root of $\frac{b_1\pi}{l} + b_2\pi^2 = 0.000131$, where $b_1 = 0.5$, $b_2 = 0.367$.

The contribution of a polymorphic site to the value of π is larger if the four nucleotide proportions at this site are almost equal and vice versa.

(c) Tajima's statistic $D = \frac{\pi - S/a_1}{\sqrt{e_1 S + e_2 S(S-1/l)}} = \frac{0.0013}{0.0180} = 0.072$, where $e_1 = 0.0096$, $e_2 = 0.0040$. Positive value of D indicates that polymorphic nucleotides are a little more frequent than expected under the hypothesis of neutrality. This might be due to balancing selection or diversifying selection.

We can not say how significant is this result since we do not have the null distribution table.

3. The inbreeding coefficient of an offspring to parents coming from two different backcrosses is

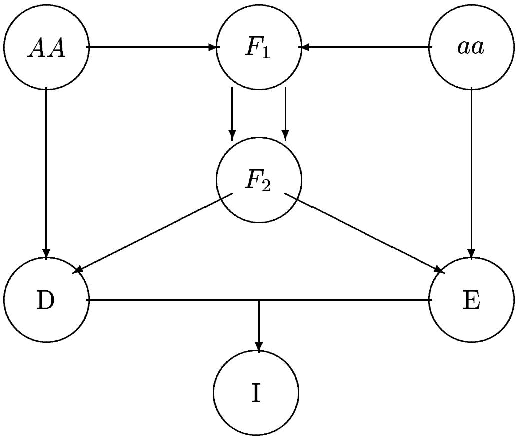
$$\frac{1}{4} \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{7}{16}$$

First term: alleles of I are inherited by D and E directly from different inbred lines (assumed to be unrelated).

Second term: one allele of I comes from an inbred line, and another allele of I comes from the F_2 individual.

Third term: two alleles of I come from the same chromosome of the F_2 individual.

Fourth term: two alleles of I come from different chromosomes of the F_2 individual.



4.

5.

6. (a) Locus 11 seems to be fixed for allele “1“ in both populations. Therefore the empty spots might be filled in with genotypes “11“.

   (b) Locus 6 frequencies in the Ghana population

Allele	“1“	“2“	“3“	“4“
Observed count	19	26	17	18
Frequency	0.238	0.325	0.213	0.225

Genotype	“11“	“12“	“22“	“23“	“24“	“34“
Observed count	9	1	5	7	8	10
Frequency	0.225	0.025	0.125	0.175	0.200	0.250

   (c) Grouped frequencies

Genotype group	“11“	“1X“	“XX“
Observed count	9	1	30
Expected count	2.27	14.51	23.23

   The observed chi-square test statistic is $X^2 = 34.5$. The approximate null distribution is the chi-square distribution with one degree of freedom. The square root of $X^2 = 34.5$ is 5.87 which according to the three-sigma rule gives a very small P-value. We reject the null hypothesis of HWE.

   (d) Locus 6 frequencies in the N’goye population

Genotype	“11“	“12“	“13“	“14“	“23“	“24“	“33“	“34“
Observed count	22	14	2	6	1	3	1	1

   show that the heterozygotes involving allele “1“ are not supressed by selection.