Tentamentsskrivning i **TMS106/MSN560: Population genetics, 5p.**

Tid: Onsdagen den 16 mars 2005 kl 8.30-12.30.
Lärare och Jour: Serik Sagitov, tel. 772-5351, mob.tel. 073-6907613, rum MC 1421.
Hjälpmedel: Räknedosa utan manualer och med tömda minnen, egen formelsamlingen fyra A4 sidor, utdelade tabeller

---

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. (6 marks)

   (a) Inbreeding, mating between relatives, decreases population heterozygosity H. Explain this effect using a simple "one locus - two alleles" population model.

   (b) For the Wright-Fisher model without mutation the expected heterozygosity H_t at generation t satisfies the formula $H_t = H_0(1-\frac{1}{2N})^t$. Here H_0 is the initial heterozygosity at generation zero and N is the diploid population size. Explain this decrease in heterozygosity despite the random mating assumption and despite the fact that genetic drift doesn't change the average allele frequency.

   (c) Recall the example from the book based on an experiment involving 107 actual populations of Drosophila, each of size $N = 16$. In the table below you see the average heterozygosities across 107 populations observed in the first 7 generations.

t	0	1	2	3	4	5	6
H_t	1.00	0.514	0.464	0.504	0.456	0.448	0.428

   Explain how the effective size $N_e = 9$ was calculated from a larger table dealing with 20 generations.

2. (5 marks) The african swallowtail butterfly, Papilio dardanus, mimics several different inedible models. Intermediate phenotypes do not look like any model, and so they are selected against. Genes for the polymorphism are close together on the chromosome forming a tight linkage group - a "supergene".

   (a) Figure 1 presents two different types of selection by plotting the incremental frequency Δp against the current allele frequency p. Copy the graphs to your answer sheet and mark three equilibrium frequencies in each case. Afterwards indicate which of them are stable and which are unstable equilibriums. Using arrows on the curves describe the direction of changes in the allelic frequency for different values of p. Explain.

(b) Which of the two graphs gives the right qualitative picture of the allele frequency dynamics for the supergene described above? Explain.

(c) Describe a selection example corresponding to the other graph.

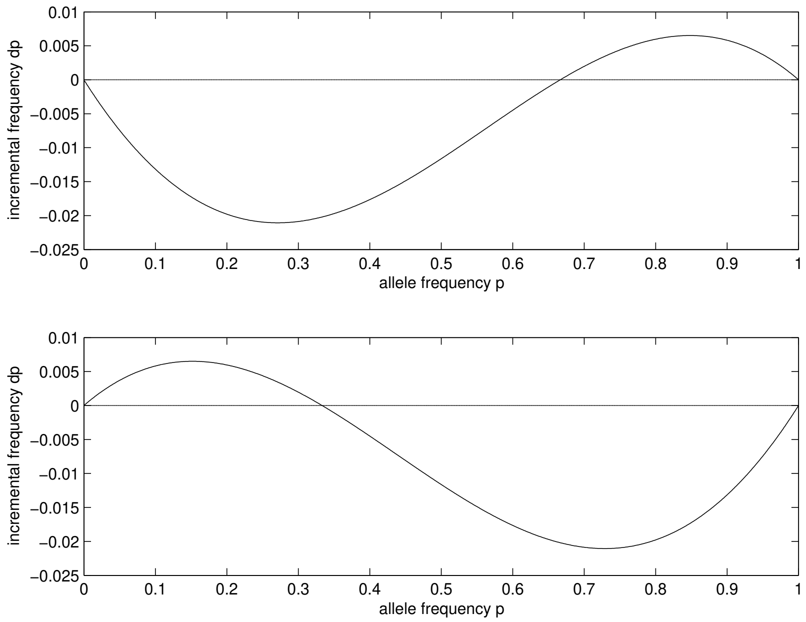


Figure 1: *Change in allelic frequency*

3. (6 marks)

A random sample of 100 individuals from a certain population gave the following genotype counts with respect to a pair of genes:

$AB/AB = 45,\ AB/Ab = 10,\ AB/ab = 6, Ab/Ab = 9,$

$Ab/ab = 26,\ aB/ab = 2,\ ab/ab = 2.$

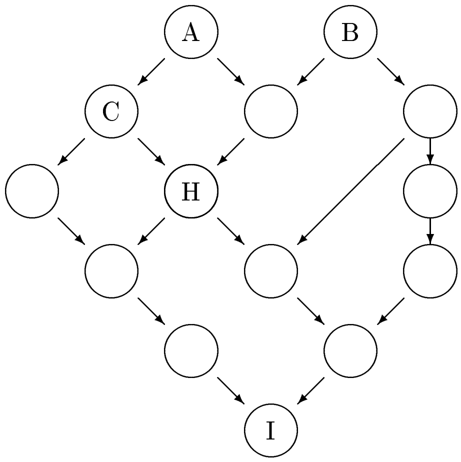
(a) Analyze the marginal genotype frequences separately for two loci using an appropriate statistical test. Suggest a mating system explaining your test results.

(b) Could these two genes be on different chromosomes? Explain using an appropriate statistical test.

(c) How would you explain the obvious deficit of aB gametes in the sample?

4. (6 marks)

   In the breeding of dogs there are often complicated patterns of inbreeding like the one shown in the figure. Here the circles represent individuals of either sex.



   (a) Identify on the graph five loops connecting the individual I to the ancestors A, B, C, and H.

   (b) Compute the inbreeding coefficient of I by summing up the contributions from the five loops.

   (c) Justify the one-loop formula that you are supposed to use in the previous task.

5. (7 marks)

   Kimura's two-parameter model of nucleotide substitution suggests the following estimate of the actual number of nucleotide differences per site for two homologous sequences

$$\hat{k} = -\frac{1}{2}\ln(1-2P-Q) - \frac{1}{4}\ln(1-2Q).$$

   Here P is the number of differences per site due to transitions, and Q is the number of differences per site due to transversions.

Sequence 1	AAC	GAA	CGC	TGG	CGG	CGT	GCC	CAA
Sequence 2	ATT	GAA	CGC	TGG	CGG	CAT	GCC	CAC
Sequence 3	GTT	GAT	CGC	TGG	CGG	CGT	GCA	CAC

Sequence 1	TAA	CAC	ATG	CAA	GTC	GAG	CGG	G
Sequence 2	TTA	CAC	ATG	CAA	GTC	GAA	CGG	A
Sequence 3	TAA	CAC	ATG	CAA	GTC	GAA	CGG	T

(a) Compute the Kimura distance $\hat{k}$ between all three pairs of sequences and compare them to the observed number of differences per site $d = P + Q$. Draw a tree connecting the three sequences so that the pairwise distances give the corresponding branch lengths.

(b) Briefly describe the Kimura model (compare it to the Jukes-Cantor model) and explain why $\hat{k}$ is usually larger than d.

(c) Estimate the effective size of the population from which the sequences are taken. Use two different methods based on the infinite sites mutation model. Assume that the mutation rate per site per generation is in the range $10^{-8} - 10^{-9}$.

(d) Describe the infinite sites mutation model. Is this sequence matix really compatible with the infinite sites mutation model?

Tables and data supplied

1. Normal distibution table
2. Chi-square distibution table

Partial answers and solutions are also welcome. Good luck!

Numerical answers

Problem 2b. First graph. Underdominance.

Problem 2c. Second graph describes overdominance - sicle cell anemia example.

Problem 3. A two loci genotype table summarizes the data. Here the gamete total is counted like $46 = 15 + 9 \times 2 + 0 + 13$.

	AB	Ab	aB	ab	Gamete total
AB	45	10	0	6	106
Ab		9	0	26	54
aB			0	2	2
ab				2	38

3a. The next two tables separate the two loci.

Genotype	AA	Aa	aa	Total
Count	64	32	4	100
Expected	64	32	4	100

BB	Bb	bb	Total
45	18	37	100
29	50	21	100

It is clear that the first locus is in HWE while the second has a significant deficit of heterozygotes: the chi-square test statistic is 41.5 with df=1. Plausible explanation for HWE on one locus and deficit of heterozygotes on the other is positive assortative mating acting on gene B but not on gene A.

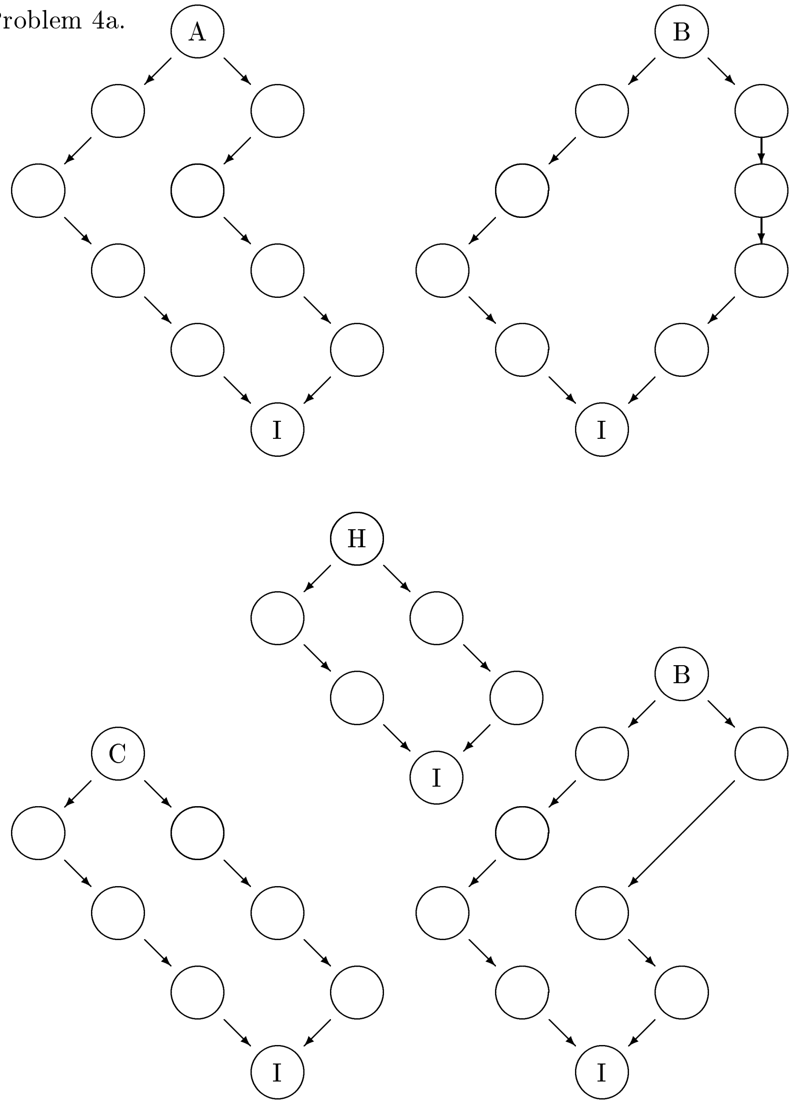
3b. To test for linkage equilibrium put together a table for gamete counts

	B	b	Total
A	106(86)	54(74)	160
a	2(22)	38(18)	40
Total	108	92	200

The chi-square test statistic is 50.5 with df=1 making linkage disequilibrium significant. We conclude that these two genes must be located close to each other on the same chromosome.

3c. A possible explanation of the number of aB gametes being below the expected under the linkage equilibrium assumption is that allele a is a recent mutation which appeared on a chromosome with allele b on the neighbor locus. The few observed gametes aB have resulted from rear recombination events since the mutation.

Problem 4a.



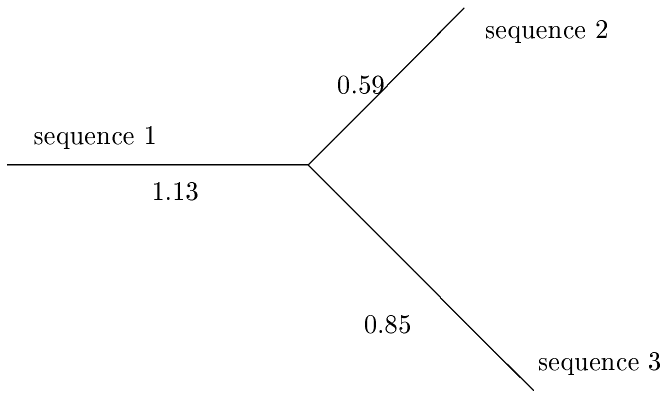
4b. Since $F_H = (1/2)^3 = 1/8$, we have

$$F_I = (1/2)^9 + (1/2)^9 + (1/2)^5(1 + 1/8) + (1/2)^7 + (1/2)^8 = 0.05.$$

Problem 5a. Observed differences among 46 nucleotides

ssequence pair	1-2	1-3	2-3
transitions	4	3	2
transversions	3	5	4
$\hat{k}$	.172	.198	.144
d	.152	.174	.130

These three pairwise Kimura distances can be summarized in a tree form as follows



5c. Two estimates of $\theta = 4N_e\mu$: $S/a_1 = \frac{10}{46}/(1+\frac{1}{2}) = 0.145$ and $\pi = (9 \times 2 + 3)/(3 \times 46) = 0.152$. Therefore the effective size lies in the range $N_e = 3.7 \times 10^6 - 3.7 \times 10^7$.

5d. The last site variation contradicts the infinite sites mutation model.