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

Tid: Fredagen den 12 mars 2004 kl 8.45-12.45.
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	12-16	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) Think of genetic variation in a locus of length 4 kb in the framework of a diploid population of effective size 30,000 individuals. Assume equilibrium between random genetic drift at the locus and neutral mutation of rate $2 \cdot 10^{-8}$ per nucleotide per generation.

   (a) Compute the expected heterozygosity and the effective number of alleles at the locus.

   (b) How many alleles would you expect to find in a random sample of 10 chromosomes containing the locus of interest? Compare it with the effective number of alleles, and explain the difference.

   (c) What is the expected number of polymorphic nucleotide sites in the 10 sampled copies of the gene?

   (d) How many pairwise nucleotide differences would you expect to find in such a sample?

2. (5 marks)

   A population with an assortative mating system is scored for three loci. The genotypic frequency results are as follows:

Locus i	A_iA_i	A_ia_i	a_ia_i
1	0.09	0.42	0.49
2	0.36	0.48	0.16
3	0.11	0.38	0.51

   (a) Compute three inbreeding coefficients for these three loci.

   (b) Explain the differences in inbreeding coefficients across the loci.

   (c) What results would you expect with a similar analysis in the case of consanguineous mating? (Consanguineous mating, or inbreeding, is the sexual union of closely related individuals, such as brothers and sisters, or cousins.)

3. (4 marks) The individuals from a population of lizards has a mean adult weigth of 40 grams. However, of the lizards that actually bred during a particular year, the mean weight is 50 grams. The mean adult weight of the resulting offspring is 42 grams. Assuming that the parents and offspring experienced identical environmental conditions, what is the heritability of adult weight?

4. (6 marks)

   Two populations were studied in two environments with the following results (treat the two populations as independent evolutionary units):

   Environment 1

Genotype	AA	AB	BB
Number of newborns	160	480	360
Number of adults	90	450	300

   Environment 2

Genotype	AA	AB	BB
Number of newborns	160	480	360
Number of adults	150	400	340

   (a) Estimate the relative fitnesses of the three genotypes in each environment using the convention that the fitness of the AB heterozygote is 1.

   (b) Assuming random mating and ignoring genetic drift, figure out what happens to the allele frequences in the long run.

   (c) Give two examples illustrating each of these two types of environment.

5. (5 marks)

   (a) Compute the coefficient of coancestry F_{XY} between uncle X and nephew Y. Draw the corresponding mating diagram with (A, B) standing for X's parents and (C, D) for Y's parents.

   (b) The covariance between phenotypic values of X and Y is computed as $2F_{XY}\sigma_a^2 + (F_{AC}F_{BD} + F_{AD}F_{BC})\sigma_d^2$. Express the slope of the regression line for uncle-nephew values in terms of the heritability coefficients h^2 and H^2.

6. (3 marks)

   In the human immunodeficiency virus HIV the rate of nucleotide evolution has been estimated at about 0.01 substitutions per synonimous site per year. Two viruses isolated in 1999 differ in approximately 40% of their synonimous sites. Estimate the year in which the isolates last shared a common ancestor.

Partial answers and solutions are also welcome. Good luck!

Answers

1. Mutation rate per nucleotide per generation $\mu = 2 \cdot 10^{-8}$, and mutation rate per gene per generation $M = 4000 \times \mu = 8 \cdot 10^{-5}$.

   (a) Assuming IAM
   expected heterozygosity is $H = \frac{\Theta}{1+\Theta} = 0.91$, where $\Theta = 4N_eM = 9.6$, effective number of alleles $k_e = 1 + 9.6 = 10.6$.

   (b) Ewens sampling formula: expected number of alleles in 10 copies of the gene $\mathrm{E}(K) = 1 + \frac{\Theta}{\Theta+1} + \ldots + \frac{\Theta}{\Theta+9} = 7.11$.
   Effective number of alleles $k_e = 10.6$ is an under-estimate of the total number of alleles, while $\mathrm{E}(K) = 7.11$ estimates the number of alleles in a sample of size 10.

   (c) Assuming ISM
   expected number of polymorphic sites $\mathrm{E}(S) = 4000 \cdot \theta \cdot a_1 = 27.2$, where $\theta = 4N_e\mu = 0.0024$ and $a_1 = 1 + \frac{1}{2} + \ldots + \frac{1}{9} = 2.83$.

   (d) Expected number of pairwise mismatches $4000\binom{10}{2}\theta = 432$.

2.

   (a)

Locus i	A_iA_i	A_ia_i	a_ia_i	p_i	F_i
1	0.09	0.42	0.49	0.3	0
2	0.36	0.48	0.16	0.6	0
3	0.11	0.38	0.51	0.3	0.095

   (b) Assortative mating means mating preferences for a similar phenotype. Obviously the phenotype in question is effected by the third locus but not by the first two.

   (c) In contrast to assortative mating, consanguineous mating brings equal inbreeding values $F_i > 0$ across all loci i.

3. Truncation selection $h^2 = \frac{42-40}{50-40} = 0.2$.

4.

   (a)

	Environment 1		
Genotype	AA	AB	BB
Number of newborns	160	480	360
Number of adults	90	450	300
P(Survival to maturity)	$\frac{90}{160} = 0.56$	$\frac{450}{480} = 0.94$	$\frac{300}{360} = 0.83$
Relative fitnesses	$\frac{0.56}{0.94} = 0.60$	1	$\frac{0.83}{0.94} = 0.89$

Environment 2			
Genotype	AA	AB	BB
Number of newborns	160	480	360
Number of adults	150	400	340
P(Survival to maturity)	$\frac{150}{160} = 0.938$	$\frac{450}{480} = 0.833$	$\frac{300}{360} = 0.944$
Relative fitnesses	$\frac{0.938}{0.833} = 1.126$	1	$\frac{0.944}{0.833} = 1.133$

(b) Current allele frequencies: $p_1 = p_2 = \frac{160+240}{160+480+360} = 0.4$.

Environment 1: overdominance with $s_1 = 0.40$ and $s_2 = 0.11$. Stable equilibrium frequency $\hat{p}_1 = \frac{s_2}{s_1+s_2} = 0.22$. In the long run the current $p_1 = 0.4$ goes down toward $\hat{p}_1 = 0.22$.

Environment 2: underdominance with $s_1 = 0.126$ and $s_2 = 0.133$. Unstable equilibrium frequency $\hat{p}_2 = \frac{s_2}{s_1+s_2} = 0.51$. In the long run the current $p_2 = 0.4$ (being lower than $\hat{p}_2 = 0.51$) goes down to zero.

(c) Environment 1 example: sicle cell anemia.

Environment 2 example by Olov Lindqvist: Bugs on a Swedish flag. Three colors of bugs: blue AA, green Aa, and yellow aa. Birds can spot green bugs everywhere, while blue and yeallow bugs can hide in the blue and yellow firelds of the Swedish flag.

Depending on the initial allele frequencies all bugs become of one color: either all be blue or all be yellow.

5. $F_{XY} = 0.125$, $b = 0.25h^2$.

6. Jukes-Cantor genetic distance $\frac{3}{4}\ln(\frac{3}{3-4\cdot 0.4}) = 0.57$. Estimated divergence time $\frac{0.57}{2\cdot 0.01} = 28.5$ years. Estimated year of divergence: 1970.