ANSC 446 / IB 416
Population Genetics
Final Exam, December 17, 2009

Name _____

(5 pages) Please underline or indicate your answer. If rounding, use 3 significant digits. Show your work or describe your logic to earn partial credit for incomplete answers.

(6) 1. An AFLP marker in xantusid lizards was found to be heterozygous in 40 lizards and homozygous in 20 lizards.

(3) a. Estimate the effective number of alleles at this marker.

AE = 1 / (1-(40/60)) = 1/(1-0.666) = 1/.333 = **3 alleles**

(3) b. What does this estimate represent or mean?

This level of heterozygosity would also be produced by 3 alleles each present at equal frequencies.

(9) 2. Coat color in horses is determined by multiple alleles. A complete Black horse (C1 black horse with black mane and tail) is dominant to a Bay horse (C2 brown horse with black legs, mane and tail) and a Mahogany Bay. A Bay is dominant to a Mahogany Bay (C3 brown horse with black roots, legs, mane, and tail). Your sample has 3000 horses (Black, Bay, and Mahogany Bay).

Color	Observed Number
Black	480
Bay	1470
Mahogany Bay	1050

Estimate the allele frequencies for C1, C2, and C3.

Answer: $P_{C3} = \sqrt{1050/3000} = \sqrt{.35} =$ **0.592**

$$P_{C2} = \left[\frac{(1470 + 1050)}{3000}\right]^{1/2} - (.35)^{1/2} = (.84)^{1/2} - (.35)^{1/2} =$$ **0.324**

$P_{C1} = 1 - (.59160798 + .32490716) =$ **0.0835**

(6) 3. Four babies were born in a hospital on the same night, and their blood groups were later found to be O, A, B and AB. The four pairs of biological parents were:

O and O – baby O
A and A – baby A
A and B – baby AB
O and B – baby B

Assign the four babies to their correct parents. Indicated next to parents

(9) 4. A survey in one state finds that the proportion of newborn infants who have the recessive autosomal disease cystic fibrosis is 1 in 2500.

(3) a) What is the estimated frequency of this disease allele?

q = P(a) = square root of 1/2500 = **0.02**

(3) b) What proportion of individuals would be carriers of the disease?

P(carriers) = 2pq = 2 X P(A) X P(a) = 2 X (1-0.02) X (0.02) = **0.0392**

(3) c) Assuming a random-mating population, what proportion of matings would be between two carriers?

(P(carriers))^2 = 0.0392^2 = **0.00154**

(6) 5. For the addax antelope, a population size of 1600 is maintained in zoos. Assume that this population size is constant across generations.

(3) a. Assuming that the effective population size equals the census size of 1600, and assuming no selection or migration. How long will it take on average for a new neutral mutant allele to become fixed in this population?

Answer: $T1(p) = 4 N_e = 4(1600) =$ **6400 generations**

(3) b. For a new neutral mutation in this population, what is the probability that it will become fixed in the population?

Answer: 1/(2N) = 1/3200 = **0.0003125**

(8) 6. What are the similarities and differences between the multi-regional and the "out of Africa" hypotheses of modern human origins? Which is favored by genetic evidence?

Both hypotheses agree that the human genus left Africa ca 2 million years ago and that modern humans first evolved in Africa ca 200,000 years ago. The **multi-regional** hypothesis argues that modern humans also evolved outside Africa with genetic contributions from local pre-existing populations of the genus *Homo* (e.g., Neanderthals contributed to the modern human lineage in Europe; "Java Man" contributed in Australasia, etc.). The **out-of-Africa** hypothesis argues that when modern humans left Africa, they replaced the previous non-African regional populations, which did not contribute to the modern human gene pool. Genetic evidence overwhelmingly favors the out-of-Africa theory.

(5) 7. What does Kimura's Neutral Theory claim? What does Ohta's Nearly Neutral Model say about selection and drift in large versus small populations?

Kimura's theory states that genetic variation is primarily influenced by mutation generating neutral variation and genetic drift eliminating variation. The "Nearly Neutral" model also recognizes that variants subject to selection are effectively neutral when the selection coefficient **$s < 1/(2N)$**. Thus in a large population fewer loci are effectively neutral than in a small one, and selection plays a larger role in large populations while drift plays a larger role in small populations.

(5) 8. Assume the probability of a child being born a boy is 0.52. What is the probability that a family with four children would have two boys and two girls?

$(4!/(2!2!))X(0.52)^2X(0.48)^2$ = ((4X3X2)/(2X2))(.2704)(.2304) = **0.3738** or 37.38%

(9) 9. What's *different* between a cladogram, a phylogram and a chronogram?

In a **cladogram**, the branch lengths are not informative, and only the evolutionary relationships are shown with no indication of the degree of difference separating the OTUs. By contrast, a **phylogram** has branch lengths that correspond to the degree of evolutionary change that occurred along each branch. A **chronogram** is a tree where the branch lengths represent the amount of time (eg millions of years) that passed since the taxa diverged.

(12) 10. Consider two loci (A and B) with two alleles each (A_1 and A_2; B_1 and B_2). A human population currently has the following gametic frequencies:

0.2 A_1B_1, 0.2 A_1B_2, 0.1 A_2B_1, and 0.5 A_2B_2.

(3) a) Estimate the current linkage (or gametic) disequilibrium.

Answer: $D = (.2)(.5) – (.2)(.1) = .10 - .02 =$ **0.08**

(3) b) If the A and B loci are linked with recombination equal to .3, how much gametic disequilibrium is expected after 4 more generations of random matings?

Answer: $D_4 = (1 – c)^t \times D_0 = (1 – .3)^4 (.08) =$ **0.0192**

(3) c. How far apart are loci A and B in map units (i.e., in centiMorgans)?

30 cM or map units

(3) d. How far apart would you expect them to be in terms of base pairs?

On average, 30 cM is approximately **30 million base pairs**

(18) 11. Which of the following generally apply to the following methods: Neighbor Joining (NJ), minimum evolution (ME), maximum parsimony (MP) or maximum likelihood (ML)? Write NJ, ME, MP and/or ML next to the phrase if the statement applies to the method.

(3) a. May be run using the Jukes-Cantor model NJ ME ML

(3) b. Uses a distance matrix NJ ME

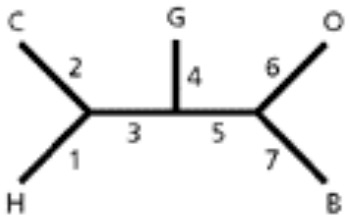
(3) c. Uses a clustering algorithm NJ

(3) d. Places operational taxonomic units at external nodes NJ ME MP ML

(3) e. Can be used with mtDNA sequences NJ ME MP ML

(3) f. Bootstrap support may be calculated using pseudoreplicates NJ ME MP ML

(33) 12. Consider the unrooted cladogram on the right: How many of the following elements are present on the tree?



(3) a. How many OTUs? 5

(3) b. How many internal branches? 2

(3) c. What numbers are next to internal branches? 3, 5

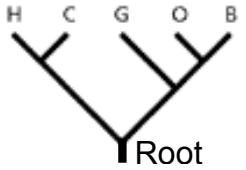
(3) d. How many branches total? 7

(3) e. How many terminal nodes? 5

(3) f. How many internal nodes? 3

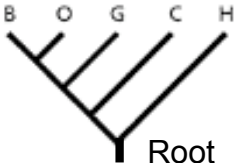
(3) g. Name one number or letter at a terminal node. C (also G,O,H,B)

(4) h. If the root were at position **3**, what would be the tree's topology (indicate the root)?

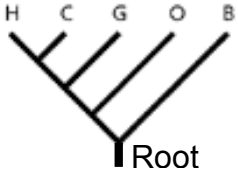


Note: since this is a **cladogram**, the relative branch lengths are not important

(4) i. If the root were at position **1**, what would be the tree's topology (indicate the root)?



(4) j. If in the tree at the top of the page, the letters B, C, G, H and O signify baboon, chimpanzee, gorilla, human and orang-utan, respectively, then what would a rooted tree look like that showed the true relationships among these species?



Note: for all the trees, there are other ways to draw the branching pattern, just as long as the relationships among the taxa are maintained.

(16) 13. Assume that the fitness values are 0.8, 1.0 and 0.7 for genotypes A_1A_1, A_1A_2, and A_2A_2, respectively.

(3) a. What are the values for the selection coefficients s_1 and s_2?

$s_1 = 0.2$ and $s_2 = 0.3$ (for A_1A_1 and A_2A_2, respectively)

(3) b. What is this type of fitness relationship among genotypes called?

Overdominance or heterozygote advantage

(3) c. What is the equilibrium frequency for A_2?

$q_e = s1/(s1 + s2) = 0.2 / (0.2 + 0.3) =$ **0.4**

(3) d. For the A locus, at what allele frequencies will the genetic load be minimized?

At q = 0.4 and p = 0.6 (i.e., at equilibrium)

(4) e. Four isolated populations have initial allele frequencies for A_2 of 0.0, 0.45, 0.50 and 1.0. At what frequency will allele A_2 stabilize in each of these four populations?

Initial A_2 of q = 0.0 will remain at 0.0
Initial A_2 of q = 0.45 will go to 0.4
Initial A_2 of q = 0.5 will go to 0.4
Initial A_2 of q = 1.0 will remain at 1.0

(6) 14. At a biallelic autosomal locus, a deleterious additive allele with a selection coefficient of 0.4 is found at a frequency of 0.3. Assuming random mating, what will be the frequency of the allele in the next generation?

$$q_1 = \frac{q_0[1-s(hp_0+q_0)]}{1-2hsp_0q_0-sq_0^2} = \frac{(0.3(1-(.4((.5)(.7)+.3))))}{(1-((2)(.5)(.4)(.3)(.7))-(.4)(.3)(.3))} = \frac{0.222}{0.844} = \mathbf{0.252}$$

(24) 15. Give the best definition for the following terms (2 points each):

a. Coefficient of inbreeding, **f**, is the probability that 2 homologous alleles are identical by descent (IBD)

b. Genetic drift: chance allele frequency changes due to finite population sizes.

c. Homoplasy: characters are similar, but their states have evolved independently, due to **convergent** or **parallel evolution**.

d. Fisher's fundamental theorem of natural selection: rate of increase in fitness of any species at any time equals its genetic variance in fitness at that time.

e. Wahlund effect: subpopulation structure leads to reduction in overall heterozygosity, even if each subpopulation is in Hardy-Weinberg equilibrium.

f. Heuristic method: A shortcut used by phylogenetic methods that rely on optimality criteria, for finding the best tree without examining or scoring every possible tree.

g. Operational taxonomic unit, **OTU**: an individual or taxon or sequence at a terminal node of a tree

h. Lineage sorting: the fixation of ancestral polymorphisms following the phylogenetic divergence of species. Lineage sorting may occur in a way that leads the topology of some **gene trees** to be incongruent with the topology of the **species tree**.

i. Vicariance: the separation of a previously continuous organismal range by past geological or environmental events

j. Epistasis: the interaction of fitness values at different loci

k. Genetic hitchhiking: when a neutral allele associated (in linkage disequilibrium) with an allele at a different locus is "carried along" and increases in frequency because of the selective advantage of the associated allele.

l. Selective sweep: reduction of heterozygosity and molecular variation in the region around a positively selected allele, due to genetic hitchhiking; e.g., selective sweeps with a selection coefficient of s = 0.01 can reduce variation 10 kb away from selected locus.

(14) 16. Name an **example** of each of the following (2 points each):

a. Darwinian (positive) selection: dark moths increased following pollution

b. Outgroup: a human sequence is an outgroup to a clade of chimp mtDNAs

c. Founder effect: out of Africa, peopling of the Americas

d. Cline: incidence of sickle cell between coast and interior of West Africa

e. Quantitative trait: height among humans

f. Heterozygote advantage: the allele that causes sickle-cell disease

g. Test of neutrality (include description): Non-synonymous vs. silent substitutions: within species, are used to identify purifying and Darwinian selection. Non-synonymous will be less common after purifying selection, but more common after Darwinian selection.

(14) 17. In the pedigree on the right, CA1 and CA2 are outbred.

(3) a. What is the inbreeding coefficient for individual X?

***f* = 0**

(3) b. What is the inbreeding coefficient for individual Z?

f = (0.5)^6 + (0.5)^6 = 1/64 + 1/64 = 2/64 = **0.03125**

(8) c. Genotypes are shown for the A locus for individuals included in the pedigree. What are the possible genotypes for individual Z to have at the A locus? Which of these would be identical by descent, and which have identity in state?

Four possible genotypes for Z:
A_1A_2 the two alleles are not identical
A_2A_2 identical by descent
A_1A_3 the two alleles are not identical
A_2A_3 the two alleles are not identical