ANSC 446 / IB 416
Population Genetics
Final Exam, December 15, 2008

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 50 lizards and homozygous in 10 lizards.

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

AE = 1 / (1-(50/60)) = 1/(1-0.8333) = 1/.1666 = 6 alleles

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

This level of heterozygosity would also be produced by 6 alleles each present at equal frequencies.

(14) 2. Assume that the following mtDNA sequences were found in four different individuals sampled from a population.

```
CATCGAGACTTGAGT
CTTCCAGATTTAAGC
CTTCCAGATTTAAGC
CATGGAGACTTGAGT
```

(3) a) How many sites are segregating? 6

(3) b) What proportion of nucleotide sites differ between the first and second sequences? 5/15 = 0.3333

(3) c) How many transversions are present between the first and second sequences? 2

(5) d) Estimate the population nucleotide diversity from this sample.

Three unique sequences: A (sequence 1), B (sequences 2, 3), C (sequence 4)
Frequencies P(A) = 0.25, P(B) = 0.5, P(C) = 0.25
πij for comparisons: A-B 5/15 = 0.3333, A-C 1/15 = 0.06666, B-C 6/15 = 0.4
π-hat (estimate) = $(N/(N-1)) \Sigma$ (pi X pj) X πij
= 4/3 X [(2 X 0.25 X 0.5 X 0.3333) + (2 X 0.25 X 0.25 X 0.06666) + (2 X 0.5 X 0.25 X 0.4)] = 4/3 X (.08333 + 0.008333 + 0.1)] = **0.2555**

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

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

(15) 4. A population of 400 cattle is surveyed at a locus with two codominant alleles. The genotype A_1A_1 is found to be present in 60 cattle, A_1A_2 is present in 40 cattle, and A_2A_2 is present in 300 cattle.

(3) a. What are the genotype frequencies?

$Pr(A_1A_1) = 60/400 =$ **0.15** $Pr(A_1A_2) = 40/400 =$ **0.10** $Pr(A_2A_2) = 300/400 =$ **0.75**

(3) b. What are the allele frequencies?

$Pr(A_1) = p = (60 + 40/2) / 400 =$ **0.20** $Pr(A_2) = q = (300 + 40/2) / 200 =$ **0.80**

(3) c. What would be the genotype frequencies if the population were at Hardy-Weinberg equilibrium?

$Pr(A_1A_1) = 0.2^2 =$ **0.04** $Pr(A_1A_2) = 2(0.2)(0.8) =$ **0.32** $Pr(A_2A_2) = 0.8^2 =$ **0.64**

(3) d. Estimate the level of inbreeding (inbreeding coefficient) in the population.

$f = 1 - H/2pq = 1 - ((40/400)/(2 \times 0.2 \times 0.8)) = 1 - (0.1/0.32) =$ **0.6875**

Hint: verify using f = 0.6875 on unit square

(3) e. Assuming the population has equal numbers of males and females, what is the effective population size of this inbred population?

$Ne = N / (1+f) = 400/(1+0.6875) =$ **237**

(4) 5. At a biallelic autosomal locus, a deleterious allele with a dominance level of 0.2 and selection coefficient of 0.3 is found at a frequency of 0.1. Assuming random mating, what will be the frequency of the allele in the next generation?

$$q_1 = \frac{q_0[1-s(hp_0+q_0)]}{1-2hsp_0q_0-sq_0^2} = \frac{(0.1(1-(.3((.2)(.9)+.1))))}{(1-((2)(.2)(.3)(.1)(.9))-(.3)(.1)(.1))} = \frac{0.0916}{0.9862} =$$ **0.0929**

(6) 6. For the Arabian oryx, a population size of 734 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 734, 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: $T1(p) = 4\ N_e = 4(734) =$ **2936 generations**

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

Answer: $1/(2N) = 1/1468 =$ **0.000681**

(4) 7. What is the difference between linkage disequilibrium and meiotic drive?

LD refers to loci that are close to each other on a chromosome and thus their alleles on the same haplotype segregate together. Meiotic drive refers to segregation distortion in which heterozygotes do not produce equal proportions of their two alleles in gametes due to "selfish" DNA, e.g. the *t* locus in the mouse.

(5) 8. 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.

(4) 9. What are two proposed disadvantages and two proposed advantages for the evolution of sexual reproduction (and recombination).

Disadvantages: 1. Parent only passes half of alleles to offspring; 2. Recombination can break up favorable allele combinations between two loci
Advantages: 1. Potential for allele combinations with higher fitness states; 2. Avoidance of Muller's ratchet.

(6) 10. 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) 11. Consider two loci (A and B) with two alleles each (A_1, A_2, B_1, and B_2). A human population currently has the following gametic frequencies:

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

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

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

(3) b) If the A and B loci are linked with recombination equal to .2, how much gametic disequilibrium is expected after 4 more generations of random matings?

Answer: $D_4 = (1 – c)^t \times D_0 = (1 – .2)^4 (.08) =$ **0.0328**

(3) c. How far apart are loci A and B in terms of map units?

20 cM or map units

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

On average, 20 cM is approximately **20 million base pairs**

(24) 12. 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.

(4) a. Requires an accurate alignment of sequences NJ ME MP ML

(4) b. Uses a distance matrix NJ ME

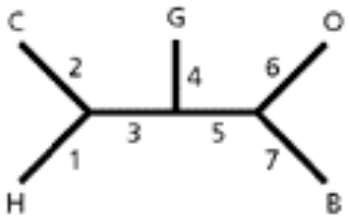
(4) c. Uses a clustering algorithm NJ

(4) d. Compares optimality criteria across trees ME MP ML

(4) e. May conduct a heuristic search ME MP ML

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

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



(2) a. How many OTUs? 5

(2) b. How many internal branches? 2

(2) c. How many external branches? 5

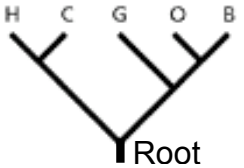
(2) d. How many branches total? 7

(2) e. How many terminal nodes? 5

(2) f. How many internal nodes? 3

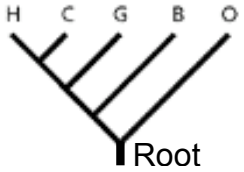
(2) g. How many nodes total? 8

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

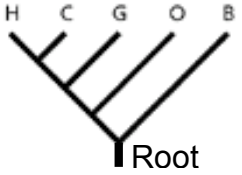


Note: since this is a **cladogram**, the relative branch lengths are not important

(4) i. If the root were at position **6**, 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.

(12) 14. Assume that the fitnesses are 0.8, 1.0 and 0.7 for alleles A_1A_1, A_1A_2, and A_2A_2, respectively.

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

Overdominance or heterozygote advantage

(3) b. Calculate the mean fitness when q = 0.1.

w-bar = $p^2w11 + 2pqw12 + q^2w22$ = (0.81 X 0.8) + (2 X 0.9 X 0.1 X 1) + (0.01 X 0.7) = (0.648 + 0.18 + 0.007) = **0.835**

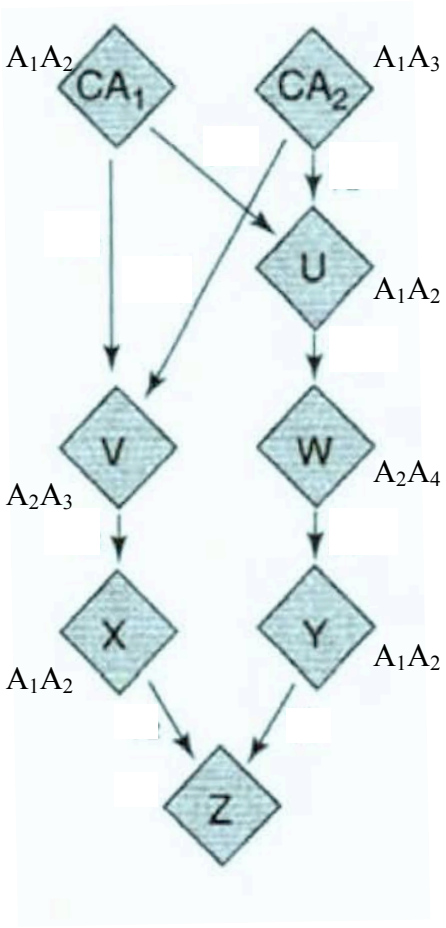
(3) c. Is the mean fitness higher for any other allele frequency? Why or why not?

Yes. The highest level of fitness for an overdominant system will be at equilibrium, which is when qe = s1/(s1 + s2) = 0.2 / (0.2 + 0.3) = 0.4

(3) d. What types of genes (or what region of the genome) would you expect to display this type of fitness relationship between homozygote and heterozygote genotypes?

Genes in the major histocompatibility complex (MHC), involved in immunity

(12) 15. In the pedigree on the right, CA1 and CA2 are outbred.



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

f = 0

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

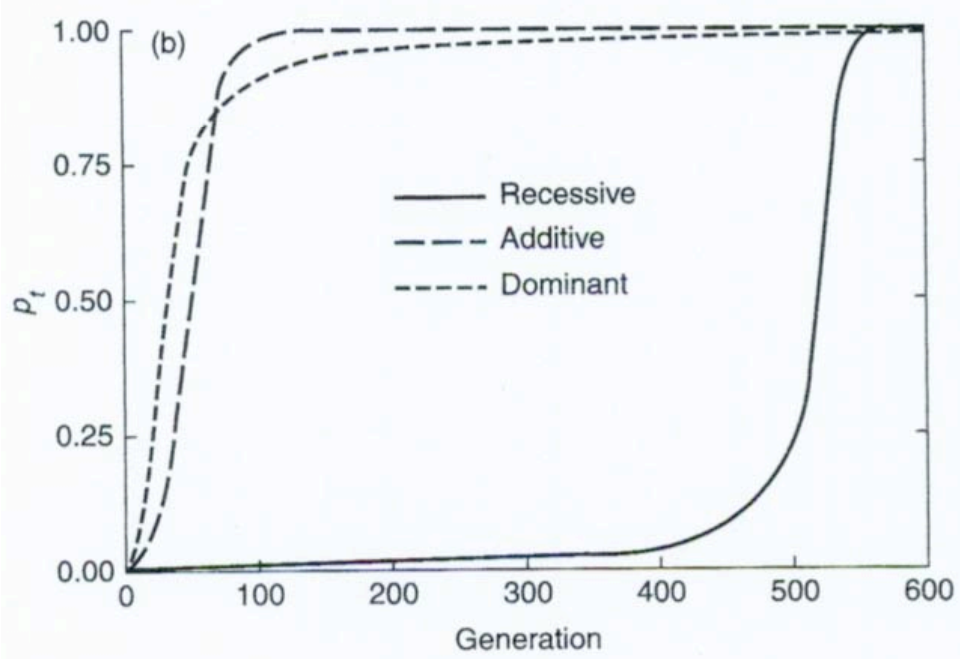
f = (0.5)^6 + (0.5)^6 = 1/64 + 1/64 = 2/64 = 0.03125

(6) 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?

Three possible genotypes for Z:

A_1A_1 identity in state
A_2A_2 identical by descent
A_1A_2 neither

(6) 16. The following Figure from Hedrick's text shows the rise in frequency for an allele at an initially low frequency (p_0 = 0.01) undergoing positive Darwinian selection (s = 0.1), in cases where the allele is recessive, dominant, or additive:



Given that the fitness value for an additive allele is exactly half way between the fitness of a dominant allele and the fitness of a recessive allele, why doesn't the curve showing increase in an additive allele fall exactly intermediate between the curve for a dominant and the curve for a recessive allele?

At low frequencies an allele is found almost entirely in heterozygous genotypes. Selection does not affect the heterozygous state if the allele is recessive, but the heterozygotes are subject to selection if the allele is additive. Hence at low frequencies an additive (like a dominant) allele increases quickly in frequency due to positive selection acting on the heterozygotes, while a recessive allele does not.

(24) 17. 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. Continent-island model of migration: **unidirectional** gene flow due to migration. Allele frequencies shift only on the "island" towards that of the migrants.

d. General model of migration: A population is divided into k subpopulations, with gene flow possible in all directions for all subpopulations. Allele frequencies shift

towards a common frequency.

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

f. Cline: a directional change in allele frequencies across (geographic) space or between subpopulations (or between species), potentially due to selection or substructure. May be stable or transient.

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

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) 18. 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. Plesiomorphy: ectothermy in reptiles

d. Autapomorphy: bipedalism in humans

e. Quantitative trait: height among humans

f. paraphyly: relationship of reptiles and birds on a tree

g. polyphyly: "marine mammals" (seals, whales etc.)