ECOL 453/553 Midterm exam #1 2009 Name______________________________

1. (a) What is the C-value paradox?
   (b) What is the major cause for variation in genome size?

a) C-value paradox is the lack of a strong correlation between genome size and overall organismal complexity.

b) Variation in the amount of non-coding DNA, mostly due to variation in the amount of repetitive DNA.

2. (a) Outline briefly one way by which mutation rates can be estimated indirectly.
   (b) Outline briefly one method by which mutation rates can be measured experimentally.

a) From divergence between species in cases where divergence time is known. For neutral mutations, $D = 2\mu t + 4Ne_{\mu}$ when D is the observed sequence divergence, T is the time in generations since the species split, and Ne is the ancestral population size.

Alternatively from frequencies of diseases in mutation – selection balance: $q = \sqrt{\mu / s}$, where q is the allele frequency and s is the selection coefficient for recessive mutations.

b) From direct counts of mutant progeny in large crosses, usually based on phenotypic screens.

3. Mutation rates vary among regions of the human genome. What are some of the major causes of this variation? In answering, consider both variation among chromosomes and among sites.

Mutation rates are higher in males than females due to the greater number of germ line cell divisions in spermatogenesis than in oogenesis. This results in greater mutation rates on the Y, with autosomes next, and X lowest. Variation in base composition is also correlated with differences in μ. High GC content is associated with higher μ.

On a per-site basis, transitions are more frequent than transversions, and CpG dinucleotides are hotspots for mutation due to deamination of 5-methyl cytosine.

4. (a) What is the neutral theory of molecular evolution?
(b) Outline the major ideas and <u>three</u> predictions of the neutral theory.

a) The neutral theory is the idea that most mutations underlying molecular evolution have no effect on fitness. Proposed by Kimura in the late 1960's.

b) Major ideas: most mutations are deleterious and are eliminated by selection; of those that remain, the vast majority are neutral and are the ones that contribute to both polymorphism and divergence. The fate of these mutations is governed by genetic drift; a very small fraction of new mutations are advantageous and these contribute little to either polymorphism or divergence.

Some predictions:
- Rate of evolution = mutation rate
- Amount of variation depends on population size
- Ratio of polymorphism to divergence is the same for different genes
- Time to fixation for neutral mutations = 4N generations
- Probability of fixation for neutral mutations = allele frequency

5. Does the rate of evolution depend on population size? If so, why? If not, why not? When answering this question, please consider advantageous, deleterious, and neutral mutations. Be as explicit mathematically as you can, but also describe the results in words.

<u>Neutral</u>

Rate of evolution is independent of population size

$$k = 2N \cdot \mu\left(\frac{1}{2N}\right) = \mu$$

($2N\mu$ is the number of new mutations each generation, and $\frac{1}{2N}$ is the probability of fixation for each).

<u>Advantageous</u>

Rate of evolution is <u>positively correlated</u> with population size.

$k = 2N\mu(2s) = 4Ne\mu s$,

where 2s is the probability of fixation for advantageous mutations .

<u>Deleterious</u>

Rate of evolution is negatively correlated with population size since the probability of fixation increases with smaller N (due to drift).

6. Outline <u>two</u> ways by which you might look for evidence of positive selection in a genome. Explain the rationale and assumptions behind each approach, the kind of data that would be needed, and the patterns that would be consistent with positive directional selection.

1) HKA test. Neutral prediction is that the ratio of polymorphism to divergence is the same for different genomic regions. DNA sequences (either coding or non-coding) from multiple individuals within a species and at least one individual from another species can be used to test this prediction. Positive selection can lead to reduced polymorphism and thus a lower ratio of polymorphism to divergence.

2) Tajima's D. Neutral prediction is that SNP frequencies will follow a particular distribution, with most SNPs at low frequency and fewer at increasingly higher frequencies. DNA sequences from multiple individuals within a species are needed to test this. Positive selection may lead to a greater proportion of low-frequency variants. For example, after a selective sweep, variation will be eliminated and will recover from the input of new mutations, which are always present initially at low frequencies. In such situations, π will be less than θ, and Tajima's D (which is based on the differences between π and θ) will take on negative values. Sequences from multiple genes are useful for disentangling the effects of selection from those of demography.

7. (a) How do recombination rates vary across mammalian genomes? Outline briefly (in one sentence for each) three methods by which recombination rates can be measured or inferred.
(b) In what ways do these different approaches give consistent results?

a) Recombination rates vary on a chromosomal scale, and recombination is suppressed near centromeres of most chromosomes

Recombination rates also vary on a smaller scale, such as hotspots over less than 1 kb.

Recombination rates can be measured (1) directly by sperm typing (males only), (2) directly by comparing genetic and physical map distances, and (3) indirectly from patterns of linkage disequilibrium.

b) These methods are very consistent in revealing position and rate of hotspots. For example, hotspots that are identified from sperm typing correspond to breaks between regions of high LD (haplotype blocks).

8. Define linkage mapping. Define linkage disequilibrium mapping. Describe the similarities and differences between them. When would you use each? What are the limitations of each approach?

Linkage mapping refers to genetic mapping based on crosses or pedigrees. Co-inheritance results in associations between markers and traits; these are broken down by recombination over the duration of the experiment (usually 1-3 generations). Thus, associations extend over long genomic distances.

Linkage disequilibrium mapping refers to genetic mapping based on linkage disequilibrium in natural populations. Associations exist between markers and traits because of linkage disequilibirum (i.e. haplotype blocks); these associations are broken down by recombination during the evolutionary history of a sample (on the order of 4N generations). Thus, associations extend over short genomic distances.

Linkage mapping is useful for localizing causative genes to general chromosomal regions but is less useful for fine-scale mapping. It is efficient for finding genes of large effect, including rare mutations, provided that they are segregating in a cross or pedigree.

Linkage disequilibrium mapping is useful for localizing the causative mutation to a small genomic segment. It is also useful for finding genes of both large and small effect, provided that caustaive alleles are not rare in natural populations.

9. Imagine that you are interested in identifying the genes underlying a particular quantitative trait (for example, wing size in Drosophila melanogaster). Describe one reverse-genetic approach that would help you test the effects of particular genes.

Many answers possible, including

a) Transposon mutagenesis
b) RNAi experiments
c) Gene knockouts
d) Site-directed mutagenesis to detect effects of a particular mutation

(with descriptions of each).