Biostatistics 666
Problem Set 2
Due January 28, 2003

1. Consider a sample of $n = 2$ DNA sequences from a population of size $N = 250$ diploid individuals (i.e., $2N = 500$ sequences).
   
   a) What is the expected coalescence time for the two sequences?
   
   b) What is the probability that coalescence occurs at generation 500?

2. Consider a sample of $n = 10$ DNA sequences from a population of size $N = 2500$ diploid individuals.
   
   a) What is the expected time to the 1$^{st}$, 2$^{nd}$, 3$^{rd}$ … 9$^{th}$ coalescence event?
   
   b) What is the expected coalescence time to the MRCA of all 10 sequences?
   
   c) What is the expected number of polymorphisms in the sample? Assume that the sequences are 1,000 base pairs long and that the mutation rate is $10^{-8}$ per base-pair per generation.

3. Consider a sample of $n = 4$ DNA sequences. Assume that the diploid population size $N$ and the mutation rate $\mu$ are known.
   
   a) Sketch out possible shapes of the genealogy for this sample. How often do you expect each type of genealogy to occur?
   
   b) What is the expected time to the MRCA?
   
   c) What proportion of mutations do you expect to occur in exactly two descendant sequences?