

Biostatistics 666
Problem Set 3
Due February 8, 2006

Modeling Events in the Coalescent

1. Consider the genealogy of a set of three sequences from a population of $N = 2500$ diploid individuals. Assume that the recombination rate per sequence per generation is $r = 10^{-5}$ and the mutation rate is $\mu = 10^{-5}$ per sequence per generation.

When we move backwards in time by one generation, what is the probability of...

- a) ... a coalescence event?
- b) ... a mutation occurring?
- c) ... a recombination event occurring?
- d) ... one of these three events occurring?

In addition, calculate the following quantities ...

- e) Number of generations do you expect to pass before a), b) or c) occurs?
- f) When one of these events occurs, what is the probability that it is a recombination?

Finally, ...

- g) Assume that the following events occur in order as we trace the ancestry of the three sequences: mutation (event closest to the present), coalescence, recombination, mutation, coalescence, coalescence (most distant event). Sketch out a genealogy for this sample.

The E-M Algorithm

1. In a study of the ACE gene, an *Alu* insertion deletion (I/D) and a nearby C/T single nucleotide polymorphism were genotyped in randomly sampled individuals. The following genotypic counts are found:

	C/C	C/T	T/T
I/I	25	0	0
I/D	40	10	0
D/D	15	10	0

- a) Describe the E-M procedure for estimating two marker haplotype frequencies, including relevant equations.
- b) Use the E-M algorithm to estimate haplotype frequencies p_{IC} , p_{IT} , p_{DC} and p_{DT} for the two markers.
- c) Compare the likelihood of the observed counts assuming linkage equilibrium between markers and using the haplotype frequencies estimated by E-M. Comment on your result.
- d) Estimate D' and Δ^2 for this marker pair.
- e) Is there evidence for ancestral recombination between the I/D polymorphism and the C/T mutation in this population? Would the I/D polymorphism be a suitable surrogate for the C/T SNP in an association study?