1. In a Wright-Fisher model of $M$ haplotypes, offspring in generation $k+1$ choose their parents at random from generation $k$. Let $Z = (Z_1, \ldots, Z_M)$ be the number of offspring respectively of the $M$ haplotypes labelled $1, 2, \ldots, M$ in generation $k$. Show that

$$P(Z = z) = \frac{M!}{z_1! \cdots z_M!} M^{-M}, \quad z_1 + \cdots + z_M = M.$$ 

(a) What is the probability distribution of the number of offspring of a particular haplotype in generation $k$?

(b) Show that the mean number of haplotypes in generation $k$ which have no offspring in generation $k+1$ is $M \left( 1 - M^{-1} \right)^M$ by using (a).

2. In a coalescent tree of three haplotypes, let $T_3, T_2$ be the times respectively while there are three and two ancestors of the haplotypes. Derive a formula for the density of the time to the most recent common ancestor $W = T_3 + T_2$.

3. Four haplotypes can have two possible different unlabelled coalescent trees. Sketch the two trees and work out the respective probabilities of their occurrence.

4. (Optional) Simulate five coalescent trees of ten haplotypes and sketch them. If $T_{10}, \ldots, T_2$ are times while there are $10, 9, \ldots, 2$ edges in the coalescent tree, then a distribution identity useful for simulation is

$$T_j = -\left( \frac{j}{2} \right)^{-1} \log U_j, \quad j = 10, \ldots, 2,$$

where $U_{10}, U_9, \ldots, U_2$ are independent uniform random variables on (0,1).

5. In a coalescent tree of $r$ haplotypes let $(Z, r - Z)$ be the number of haplotypes subtended in the sample by the two edges just before the most recent common ancestor (MRCA) of the sample. Find the probability distribution of $Z$.

6*. A sub-sample of $n$ haplotypes is taken from $r \geq n$ haplotypes. Let $(Z, r - Z)$ be the number of haplotypes subtended in the sample by the two edges just before the MRCA.

(a) Show that, conditional on $Z$, the probability that the $n$ haplotypes have the same MRCA as the $r$ haplotypes is

$$1 - \frac{\binom{Z}{n} + \binom{r-Z}{n}}{\binom{r}{n}}$$

(b) Find the unconditional probability $q_{nr}$ that the sub-sample and sample have the same MRCA and show that

$$\lim_{r \to \infty} q_{nr} = \frac{n-1}{n+1}$$

Thus the probability of a sample of $n$ haplotypes having the same MRCA as the population of haplotypes is $(n-1)/(n+1)$.

Questions with a * are more challenging.