Mathematical Genetics, Fall 2017, HW 5

Instructions: Do all problems and show your work.

1. Consider the following table of SNP data (with sequence labels on the left and site labels across the top):

	1	**2**	**3**	**4**	**5**	**6**
a	G	C	G	T	T	A
b	G	C	G	T	T	A
c	C	C	G	T	T	T
d	G	G	C	A	A	T
e	G	G	C	T	T	T

Assume that the **ancestral sequence** is GGGTTT and that we have infinite sites mutation.

(a) Compute Watterson's estimate of θ.

(b) Compute Tajima's estimate of θ.

(c) Compute the (unfolded) site frequency spectrum.

(d) Compute the numerators for Tajima's D statistic and Fu and Li's D^*.

(e) Draw a (properly labeled) coalescent tree that is consistent with this data set.

2. Consider a WF model with fast fluctuations in population size, where there are 3 sizes: $N_1 = N$, $N_2 = 4N$, $N_3 = N/2$, and the proportions of time spent at these sizes is 0.2, 0.3, and 0.5, respectively.

(a) In the Kingman coalescent obtained by letting $N \to \infty$, what is the pairwise coalescence rate?

(b) What is the mean time until the first coalescence in a sample of size 10?