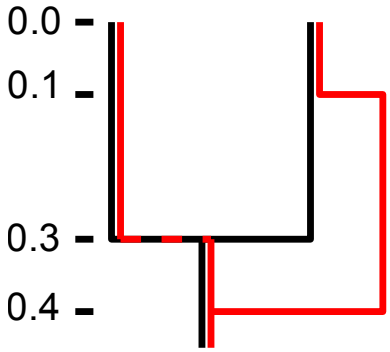
COMPUTATIONAL POPULATION GENETICS — EXERCISE SHEET 8

1. Ten haplotye sequences have been sampled from a diploid population. Four polymorhic sites a, b, c and d have been found (in this order on a chromosome), and the recombination rates ($\rho = 4N_e r$) between them are known to be 1.2 between a and b, 0.8 between b and c and 1.4 between c and d.

   We consider now the first three sequences in the sample and begin with the first two, which are ATAT and GCTT (giving only the data for the four polymorphic sites). Assume that the ancestral recombination graph (ARG) of the first two sequences is as given in the figure with time in units of $2N_e$ generation, and recombination happened between loci a and b. That is, the black tree applies to site a and the red tree applies to b, c and d.



   (a) Which of the information given above is used when the ARG is approximated according to Li and Stephen's PAC model?

   (b) In addition to the information given above assume that in sites a and b the third lineage coalesces with the black tree before it coalesces with the red tree, that is it coalesces with the ancestral sequence of sequence 2 for locus a. Conditioned on this information, calculate the probability according to the PAC model that sequence 3 has...

   - an A in site c.
   - an A in site c and also an A in site d.

   (c) Given the ancestral recombination graph in the figure and the information that the third sequence at loci a and b coalesces with the black tree before it coalesces with the red tree but neglecting all sequence information, calculate for the ARG model the probability that at locus c on sequence 3 coalesces with the red tree before it coalesces with the black tree.

2. Five diploid SNPs have been sequenced for 5 indivuals sampled from a population. The SNPs are in the same genomic region but all their distances are greater than the read length of the sequencing method, such that we have only the following genotype information for each site:

```
indiv.  1:  AT, CG, CG, TT, CC
indiv.  2:  AT, GG, CG, TT, CC
indiv.  3:  AA, GG, CG, TA, GG
indiv.  4:  AA, GG, CC, TT, CG
indiv.  5:  AT, GG, GG, AT, GG
```

   Apply Clark's algorithm to phase the data.