Bayesian Phylogenetics   *Find and work with a partner*

The substitution matrix below shows the single generation probability of each substitution.

	A	C	G	T
A	0.7	0.1	0.1	0.1
C	0.1	0.7	0.1	0.1
G	0.1	0.1	0.7	0.1
T	0.1	0.1	0.1	0.7

1. Case 1 shows the relationship between an ancestor (Anc) and descendant (Der) node. Given the substitution matrix above, what is $P(Der = A | Anc = A)$? Note that you have to consider the probabilities that $A \rightarrow A$; $A \rightarrow C \rightarrow A$; $A \rightarrow G \rightarrow A$ and $A \rightarrow T \rightarrow A$.

2. Case 2 shows the relationship between an ancestor (Anc) and two descendant (D1 and D2) node. Given the substitution matrix above, what is $P(D1 = A, D2 = T | Anc = A)$?

Bootstrapping

Below are Y-chromosome SNP data from three extant baboon species, and a 6000 year old baboon from southern Africa (likely a *P. ursinus* individual).

Site	1	2	3	4	5	6
P. ursinus	1	0	0	1	0	1
Ancient	1	0	0	1	1	1
P. cynocephalus	0	1	0	1	1	1
P. anubis	0	1	1	0	0	0

Using the provided random number generator, generate a single bootstrap sample from these data (6 sites).

Sample	
P. ursinus	
Ancient	
P. cynocephalus	
P. anubis	

Now, compute the pairwise distance matrix from your bootstrap sample (i.e. for each pair of baboons, compute the number of differences between the resampled data rows). Finally, construct the UPGMA tree relating the baboons in your bootstrap sample. If you have equal values, use the die to choose randomly between them.

D	Pu	An	Pc	Pa
Pu				
An				
Pc				
Pa				

Bonus: build the UPGMA tree from the original data. If you have lots of time, build the NJ tree.