**Exercise using the Hudson animator**

Download and open the Hudson from [www.coalescent.dk](http://www.coalescent.dk/). It is developed by Anders M. Mikkelsen, Jotun Hein and Mikkel Schierup as a tool for the visualization of the following continuous time processes.

* The basic coalescent and the Coalescent with recombination
* Coalescent with exponential growth
* Coalescent with migration

Please consult the manual before doing the exercises, it can be found under **help** at the start page. It briefly describes how to control the applet.

**The basic Coalescent without recombination**

Choose *coalescent with recombination* (we will set recombination to zero)

1. The basic coalescent. Choose **n**=5 sequences and **rho**=0 (no recombination). Press **recalc** and the animation starts. The speed can be controlled with speed. Details regarding each node can be seen in the small window at the right when moving the mouse over the node.
   1. What is the time to the first coalescence event? Write it down.
   2. What is the time to the most recent common ancestor? Write it down.
   3. Repeat a and b 5 times (by pressing recalc). How does the time to first coalescence and time to most recent common ancestor vary?
2. Try with 10 and 20 sequences. What are the times to the first coalescence and the most recent common ancestor in these cases? Write it down for 5 different runs.

**Coalescent with recombination**

The recombination rate is determined by **rho**=2**Nc**. In the animation, recombination events are marked as blue nodes (in contrast to the green nodes of coalescent events).

1. Set **n**=5, and **rho**=1. Press **recalc** and study the animation. Does any recombination events occur? If not, try again. How many recombination events do you expect to see? Try a total of 25 times and write down the number of recombination events in each case. What does the distribution look like?

Look at a couple of simulations in more detail. Study where in the sequence, recombination events occur

* 1. Can you find examples that different part of the sequence have different most recent common ancestors (marked in green) at different time points?.
  2. Can you find any examples of “trapped” non-ancestral material (coloured white) between blocks of ancestral material?
  3. Try to press the **trees** window pane. Here it is possible to study each of the different trees over the sequence. How many different trees are there and how does this relate to the number of recombination events? Try to find examples of each of the following recombination types
     + - Recombination changing the topology of the tree
       - Recombination changing the branch length but not the topology of the tree
       - Recombination that does not change the tree

What is the total number of each of the tree types during the 5 replicate runs? How does that match with theory.

1. Try setting **rho**=2 and **n**=10
   1. How many recombinations occur now?
   2. At which time do different parts of the sequence find a most recent common ancestor?
   3. What is the time until the first part of the sequence finds a most recent common ancestor (calculate the average over 5 replicates). At which time have all the different parts of the sequence found a most recent common ancestor (average over 5 replicates)? Compare this time with the time to most recent common ancestor without recombination.
2. Try setting **rho**=5 and **n**=20, how many recombination events occur now? How many different most recent common ancestors are there over the sequence now?

**Coalescent with exponential growth**

Now choose *coalescent with exponential growth*. This is controlled with the parameter **exp,** which is equal to *Nb.* This parameter measure how many times the present population is larger than the population 2**N** (**N**=size of present day population) ago. In studies of human mitochondria (there is no recombination in mitochondria) all estimates suggest that **exp**>100.

1. Try to simulate **n**=10 sequences and different runs with **exp**=1, 10, 100, 1000
   1. How does the shape of the genealogical tree depend on the value of **exp**?
   2. How can that be?
   3. How would this altered shape be visible in a set of sequences evolving on the tree? Would there be fewer or more “singletons”? How would Tajima’s D be affected?

Hint: If you push **trees** the same tree will appear without any crossing branches

**Coalescent with migration**

Now choose *coalescent with migration* It is only possible to simulate two populations, but the number of individuals and the migration rates between the populations can be freely chosen. **M1** is the migration rate from population 1 to population 2. The separation between populations is marked with a dotted line and migration events are shown as blue nodes.

1. Try to set **n**=10, **M1**=0.5 and **M2**=0.5 and run the animation
   1. How many migration events occur in total? Note it down.
   2. How many of these happen from there are 10 until there are two sequences left?
   3. How many migration events occur during the time interval when there are only two ancestors left?
   4. Repeat the animation with the same parameters three times – how large is the variance?
2. Try to push **trees**. Here you can observe the tree without migration events and without crossing branches, but you can see which population each individual came from.
   1. How many migration events can you infer must have happened?
   2. How does this correspond with the true number?
   3. Repeat this exercise a couple of times
3. Try to investigate different migration rates and number of sequences.
   1. How is the shape of the tree affected (e.g. if **M1**=**M2**=0.1)?
   2. How would this change of tree shape be expected to be visible in a set of real sequences (would there be more or fewer singletons?).