# Introduction to phylogenetic trees

Phylogenetic trees are scientific hypotheses regarding the relationships of species, individuals, genes, etc. The tree is a graph that should be "read" (interpreted) from tip to root or root to tip. This dimension represents time (arrow), with the direction toward tips representing recent time, and toward nodes and root representing the past. Each species or individual or gene in a tree is represented by a tip, or terminal branch. An ancestor common to two species is represented by a node and the (internal) branch below. An ancestor and its descendents make up a clade (monophyletic group).
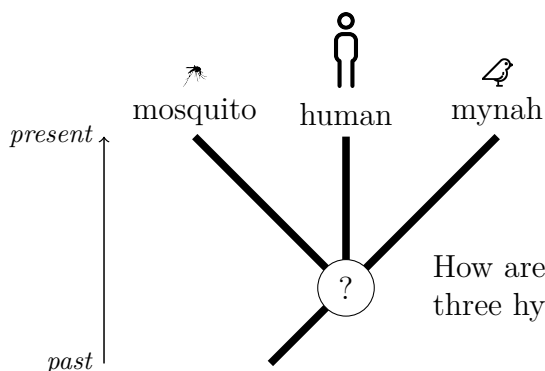
How are these three species related to each other? These are three hypotheses of relationships among these species.

1. One hypothesis is shown on the left. Here, the human and mynah are most closely related to each other. They share a most recent common ancestor, **B** (node and branch below), which was a vertebrate. The mosquito, human and mynah share the common ancestor, **A** (node and branch below), an animal.
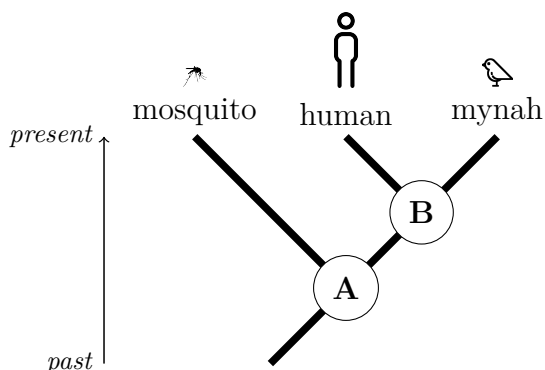
   (a) Which are the most closely related species in this tree?

   (b) Is the mosquito more closely related to the mynah or to the human?
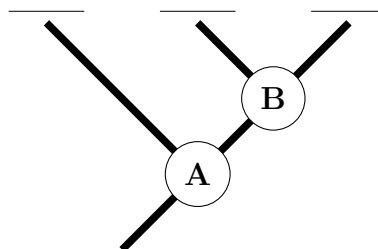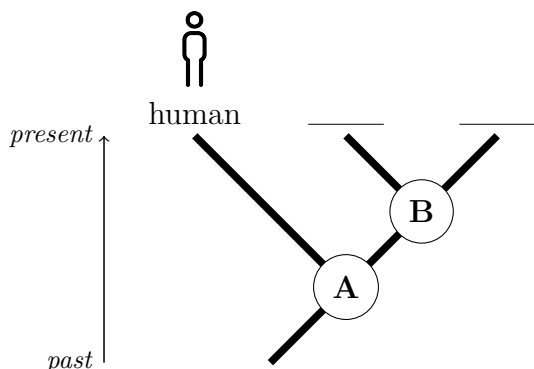
2. Complete labelling taxa in the other two trees and describe these hypotheses in words.

3. Which do you think is the 'true' tree? Why?

4. Each node in this tree represents a speciation event, or splitting of an ancestral species to form two new species. How many such speciation events are there in each tree?

Icons: freepik.com

Scientists try to estimate the best tree by using different kinds of information. Think of other types of data to address this question – for instance, other features of these animals you might know. Gene sequences are most commonly used to find out which tree best fits the data.

The tree below is a tree of relationships among different forms of the SARS-CoV virus gene sequences. Two of the sequences are from bats, (Bat SARS-CoV-b, Bat SARS-CoV-a) and one is from a human (SARS-CoV-2).
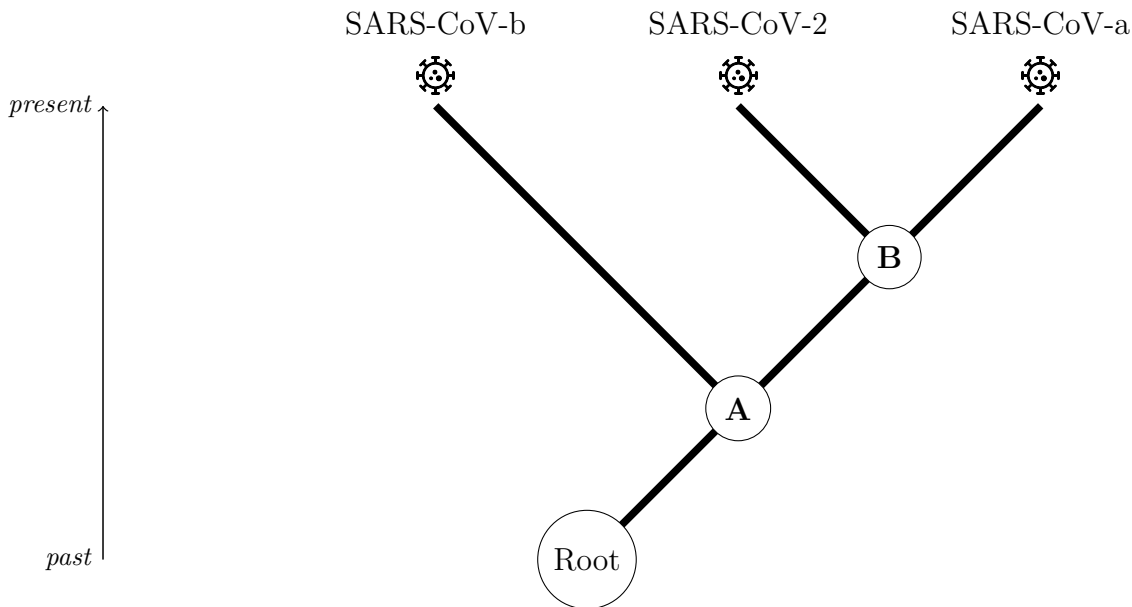
1. Which is the closest relative of the SARS-CoV-2 sequence and why?

2. What do internal branches represent?

3. What are nodes **A** and **B**?

4. True or false?

   (a) Ancestor **B** is the common ancestor of Bat SARS-CoV-a and b
       <u>True / False</u>

   (b) Bat SARS-CoV-a is more closely related to SARS-CoV-2 than to Bat SARS-CoV-b
       <u>True / False</u>

   (c) SARS-CoV-2 is more closely related to Bat SARS-CoV-a than to SARS-CoV-b
       <u>True / False</u>

   (d) Ancestor **A** arose after ancestor **B**
       <u>True / False</u>

Icons: freepik.com

In the tree below, the hosts of the viruses at the present day are also given. We will try to understand whether or not the viruses switched hosts in the part of their evolutionary history given here.
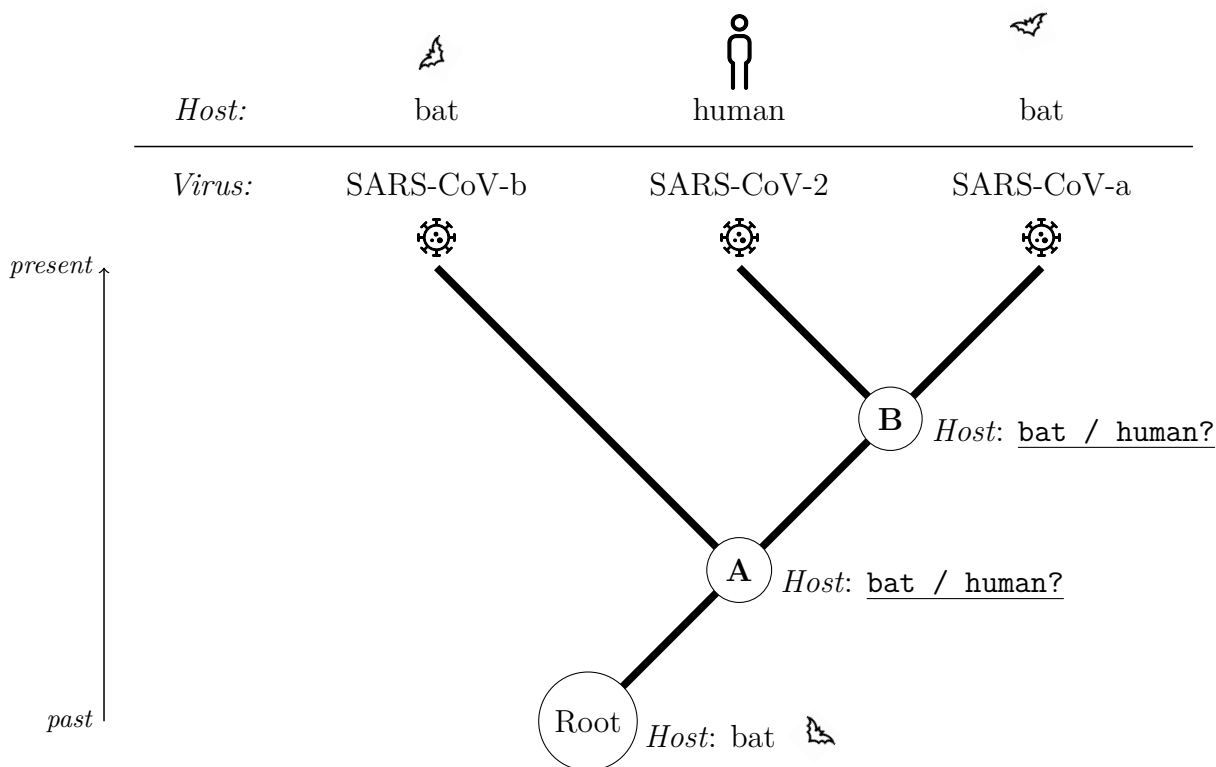

1. Use the information in this tree to list the names of the three viruses (tips).

2. Do we know the hosts that they infect? Name them. How do we know this?

3. Do you think there was a host switch during the evolution of these three viruses? Why?

4. What was the host of the common ancestor of SARS-CoV-2 and Bat SARS-CoV-a (ancestor 2)? Can you guess what it might have been?

5. Can you guess the host of the common ancestor of Bat SARS-CoV-b, SARS-CoV-2 and Bat SARS-CoV-a)?

6. The root viral ancestor's host was a bat. We know this because all closest relatives of these viruses infect bats. Does this knowledge change your answer to Q4?

7. Can we determine who ancestor **B**'s host was? How?

Icons: freepik.com

Check if you got this tree:
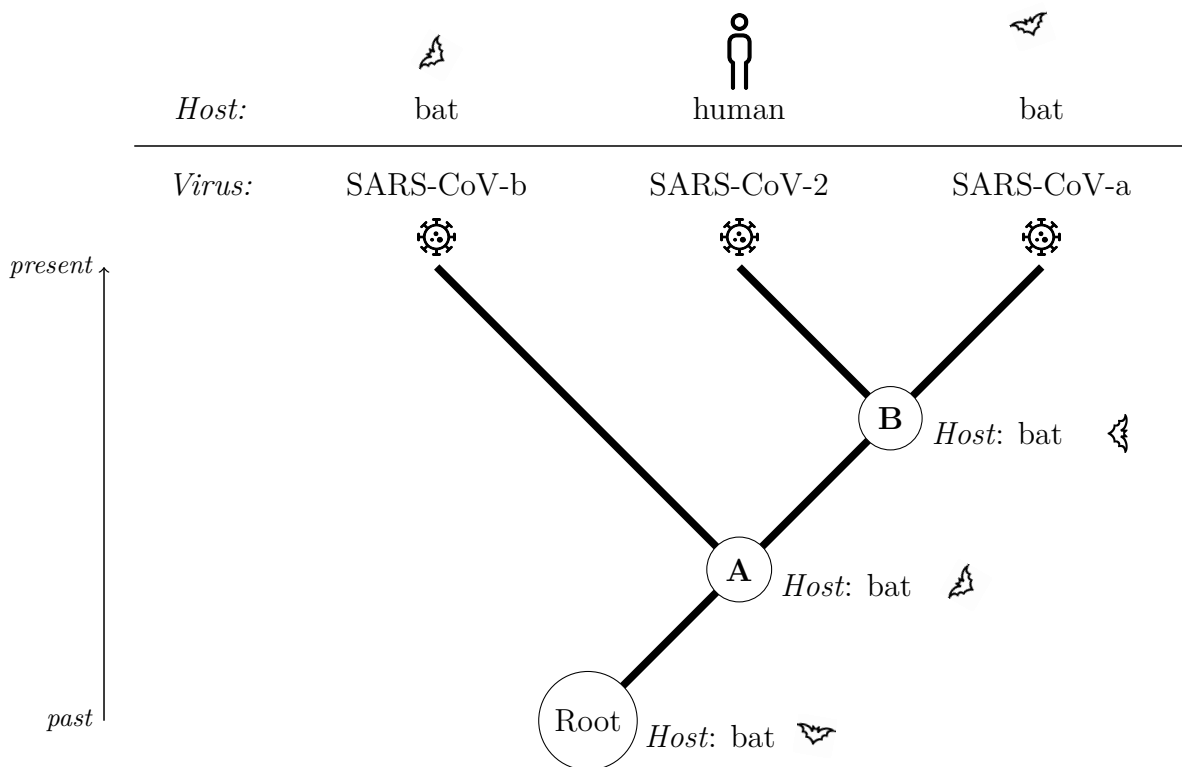
Summarize the information we have so far:

1. Names of the three viruses:

2. Names of their hosts:

3. The root viral ancestor's host was:

4. Viral ancestor **B**'s host was:

5. Viral ancestor **A**'s host was:

6. Viral ancestor **A** had two descendants.

   (a) Name them:
   (b) Name their hosts:

7. Is the host of any descendent virus different from that of its virus ancestor? If so, that descendent virus switched hosts. The switch occurred in virus _____ from host _____ to host _____

8. How many switches occurred in the history of these three viruses?

9. How might your inference have changed if the host of the root viral ancestry had been human?