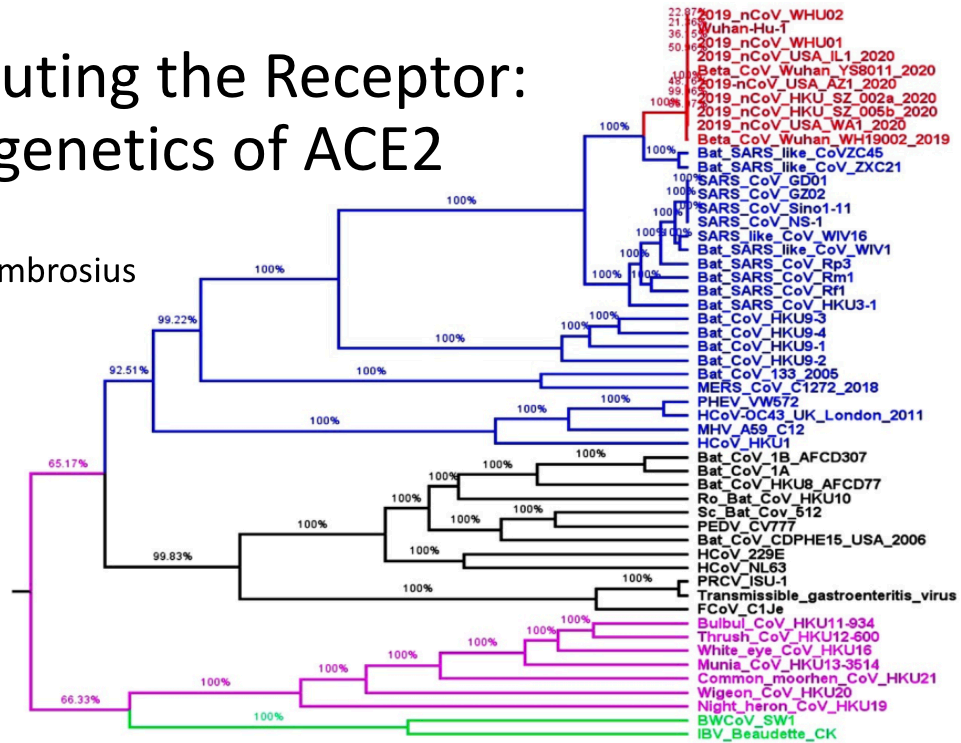


Computing the Receptor: Phylogenetics of ACE2

By Em Ambrosius



Computing CoVID-19: A Presentation of the North Carolina School of Science and Math, Summer 2020

COMPUTING CoVID-19: SUMMER 2020

COMPUTING THE RECEPTOR:

PHYLOGENETICS OF THE ACE2 GENE

Developers:

Em Ambrosius and Robert Gotwals

June 8, 2020

COMPUTING CoVID-19. COPYRIGHT HELD BY THE NORTH CAROLINA SCHOOL OF
SCIENCE AND MATH, MAY 1, 2020. ALL RIGHTS RESERVED. PARTS OF THIS LAB MODIFIED
FROM GOTWALS. [1]

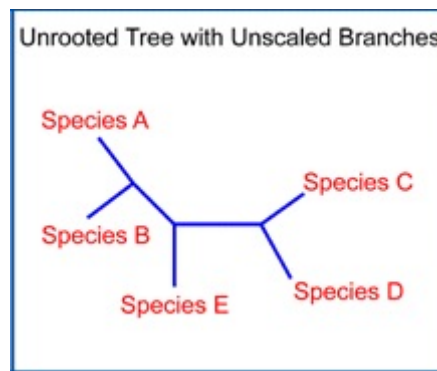


Figure 2: Example of an unrooted tree

A clade is a group of organisms with a common ancestor. Figure 3 shows which groupings of organisms on a phylogenetic tree would be considered clades and which are not clades because they do not contain only organisms from one single common ancestor. [4]

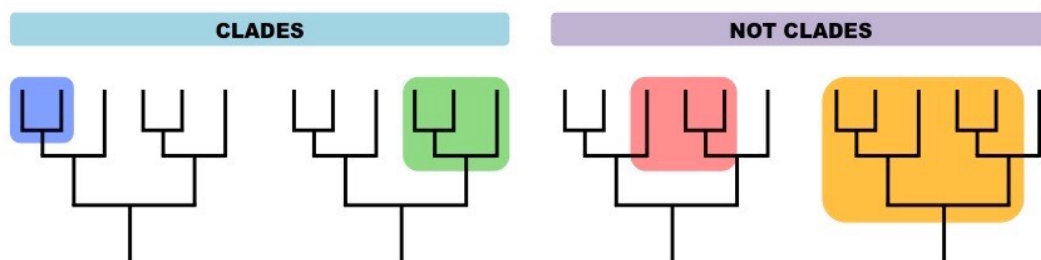


Figure 3: Examples of clades and groupings that are not clades. [4]

There are numerous tools that can be used to create phylogenetic trees. Trees are built from alignments of two or more sequences (either nucleotides/base pairs or amino acids) of organisms. A tree with two organisms certainly would not be very interesting – in most cases, trees have 10 or more different organisms.

For this lab, you will use a tool called "UniProt" to find and align genomic sequences for the ACE2 protein of some common organisms. Once aligned, UniProt will construct a phylogenetic tree for you. We will also ask you to construct a tree by hand, using a mathematical algorithm known as the *unweighted pair group method with arithmetic mean* (UPGMA) method.

Figure 4 shows a sample flowchart for the UPGMA method. Yes, it looks complicated, but we will walk you through doing this in the webinar.

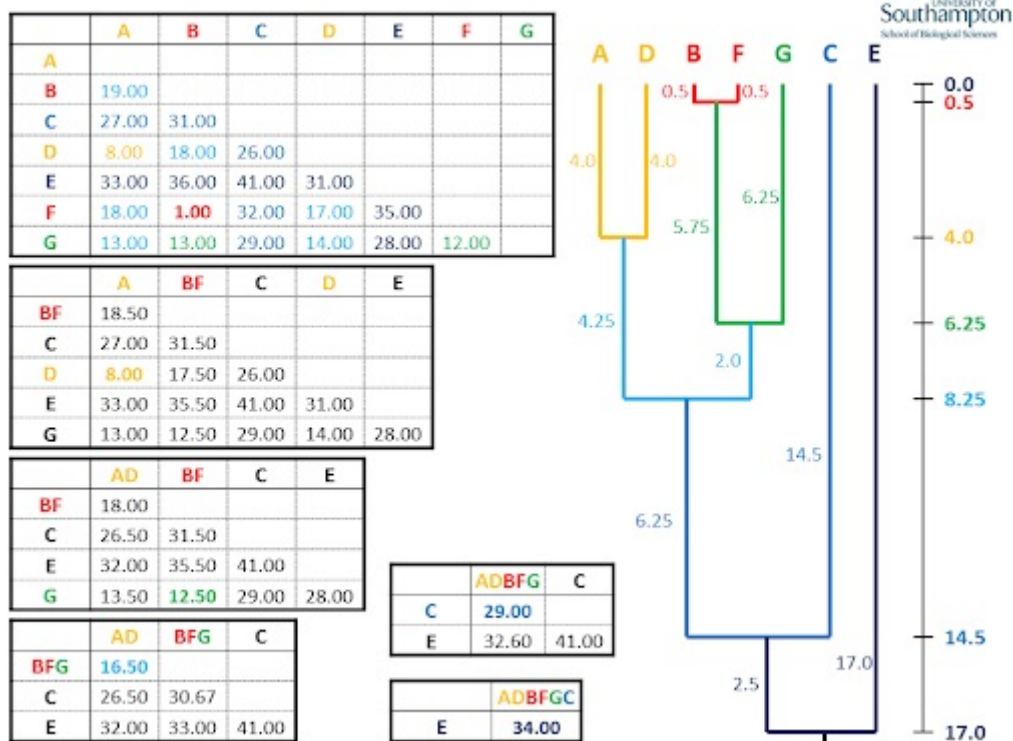


Figure 4: Example of the UPGMA phylogenetic tree method (Source: University of South Hampton)

Once you have done that, you will very much appreciate the capabilities of a computer!

2

STUDENT ACTIVITY

In this activity, we are going to find the sequences for the ACE2 protein for seven (7) organisms. You will be asked to prepare the data, build a matrix of the data using Excel, and then apply the UPGMA method to determine the mathematics of your tree and then draw your tree by hand.

2.1 PART 1 – PREPARING THE STRAINS

This section of the lab will reference the Google Sheet found [here](#).

1. Search for ACE2 in UniProt[5]. Select the following organisms by clicking the check boxes.

ACE2_HUMAN

ACE2_MOUSE

ACE2_RAT

ACE2_PAGLA

ACE2_FELCA

ACE2_BOVIN

ACE2_PONAB

2. Record the "Entry" Codes for each animal and record them in the given table at the top of the Google Sheets page.
3. Now click the "Align" tab at the top of the table. When the alignment is done (it might take a few seconds), you will see the ACE2 amino sequences for all seven organisms lined up together. As you can probably imagine, it would be very difficult and time intensive to count the numbers of differences in these sequences between each pair of organisms! Luckily, there are tools that can do this for you.
4. Click the "Download" tab at the top of the table. Change the format setting to "FASTA" and "Uncompressed" and then click "Go." A new window will open.
5. Highlight and copy all of the results in this new window.
6. Go to the "Ident and Sim" tool[6]. The tool calculates the identity and similarity of sequences for each pair of organisms.
7. Delete all of the sample data from both input boxes. Then paste your results into the first input box. (The second input box should be empty.) Click "Submit."
8. Another new window will open. This will show the results for each individual comparison between organisms. The "Alignment length" tells you the total number of amino acids being compared and the "Identical residues" tells you how many of these amino acids the two organisms have in common. Given these two pieces of information, how could you figure out the number of amino acids that differ between the two organisms?

2.2 PART 2 – COMPLETING THE MATRIX

1. Now that you know how to determine the number of amino acid differences between organisms for this sequences, you will calculate and organize your data in the Google Sheet.
2. You will now program the matrix using the values you got from the “Ident and Sim” tool. Look across the matrix to column C (Mouse) and then go down to row 12 (Human). This cell (C12) is where you will enter a formula to calculate the number of differences in the ACE2 amino sequences between these two organisms.
3. You will recall from an earlier step that the Ident and Sim tool examined an alignment length of 805 amino acids and found 661 identical residues. So, to find the number of differences, you will enter the following formula:
$$=805-661$$
4. Repeat this process for the remaining cells in your matrix. You may find it helpful to look back at the Entry codes you recorded. The codes will help you interpret the output from the Ident and Sim tool.

Be able to answer the following questions:

2.2.1 QUESTIONS

1. Multiple cells in the matrix contain zeros. Why do you think the values of these cells are zero?
2. Where do you see the greatest number of distances between two organisms?
3. Where do you see the fewest number of differences between two organisms?
4. How do you think this matrix can help you determine the relatedness of these seven organisms?

2.3 PART 3 – CALCULATING DISTANCES AND PHYLOGENETIC TREE CONSTRUCTION

Now you will build the UPGMA phylogenetic tree. A more detailed example and explanation of the steps to complete this method is listed with the assignment.^[7]

1. Have a piece of paper or drawing software to keep track of distances while building the phylogenetic tree. Start by drawing a line off to the side which will serve as the measure of evolution. The top of the page should be notated as 0.0 and represents present time.
2. From the matrix you completed in Part 2 find the smallest number of amino acid differences between any two organisms.
3. Make a clade name with these two identified organisms and note the number of amino acid differences between them. Then this number will be divided by two automatically to calculate the distance back in units of evolution time that separate these two organisms. Mark this distance and first clade on your phylogenetic tree.
4. In the blue boxes of the next matrix, fill in the clade just created and the remaining organisms. For all amino acid differences between two individual organisms, copy those values from the first matrix. All relations of the individual organisms and the clade are then calculated by taking the average number of amino acid differences between the two organisms that make up the clade. Ex. If organism X and organism Y make up a clade. The number of amino acid differences between XY and organism B is the differences between B and X plus the differences between B and Y divided by two.
5. Continue this process of creating clades, finding the distances between them, and calculating their averages until all boxes in the Google Sheet are filled and the total length of the phylogentic tree is calculated.

From this phylogenetic tree, the distances between organisms and common ancestors should be understood before taking the quiz.

REFERENCES

- [1] Robert Gotwals. Building phylogenetic trees using genetic data student activity sheet 3a.
- [2] Wikipedia Contributors. Phylogenetic tree, May 2020.
- [3] Phylogenetic trees | evolutionary tree (article).
- [4] Clades.
- [5] The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1):D506–D515, nov 2018.
- [6] Sequence manipulation suite: Ident and sim, Nov 2017.
- [7] Richard Edwards. Upgma animated example.