Teaching Tips for the Molecular Phylogeny Simulation

In this file, you will find additional details and suggestions for running the activity that did not fit in the printed article in *The Science Teacher.*

When to use:
Students need to have a basic understanding of DNA and point mutations in order to understand what is going on. If you are teaching evolution at the beginning of the year, you will need to check how much they learned in previous science classes. Personally, I prefer to teach the evolution unit last, or at least after molecular biology and genetics, because of the importance of molecular and genetic evidence in evolution. Teaching evolution last, as the capstone, allows me to remind students of what they have learned all year, pulling all the threads together, after they have learned how to entertain and evaluate alternate hypotheses, the focus of this activity. As stated in the article, this activity can be used before or after discussion of speciation and phylogeny. Since the data generation takes less than half a period, I will often use the other half of the period to introduce the definition of species and speciation, then continue with a deeper explanation of phylogeny after the activity.

Data sheet:
I strongly encourage you to use the data sheet provided. It took me a lot of time to get it right; why reinvent the wheel? If you choose to make your own data sheet, make sure you use a non-proportional font (such as Courier). This will ensure that when you enter the data in the spreadsheet the columns will all line up, regardless of what letter (or space) is in a particular position. You will need to print the data sheet on legal size paper; resizing it to fit on letter size will give a very small font and is likely to cause data entry errors. In the first line, 'N:' has two meanings: On the one hand, it represents the base number; on the other, it provides a place for students to put their name.

Introducing the activity:
If you use Figure 2 to introduce the activity, you may find it useful to also use Figure 2D, which shows what the data sheets would look like for the six students after the three rounds of mutation and splitting. All students have the first mutation. The two groups formed by the first split are also clear. After the third split, student E has formed a separate group, but will continue to mutate each time independently. This will also give them an idea of the sort of groupings they will be looking for when they do their own analysis.

Timekeeper:
You may either assign one student to be timekeeper, to say *mutate,* or do that yourself. I prefer to do it myself, so that I can allow a little more time at the beginning, to check that everyone is recording the data correctly, then speed up later—especially once most have been divided out into groups of one. But if you have a student who you know will not be able to record the data accurately, it may be better to give that student a special job.

Dividing the group:
It is better to divide the group yourself, so you can control how large the groups are. Try to get some large, some smaller groups, some who are separated from everyone else early, others late. This is both more realistic and will lead to more interesting data analysis.

Classes over 15 students:
If you have more than 15–20 students, you will want to split the class into two groups so each group does not have too much data to analyze. The easiest way to do this is to split the group into two equal parts after the first mutation. Then allow two or three mutations before you do any more splitting. When you compile the data, you will clearly see the two large groups, based on the mutations shared by half the class. Split the raw data into these two groups, and ask half the class to analyze each half of the data.

Compiling the data:
I have provided an Excel spreadsheet, formatted and with the original DNA sequence and letters A–Z entered. If you have more than 26 students, you can decide what additional characters to use. Mix the data sheets you have collected so the order will be different from the original circle. Write the code letter on the data sheet so you can go back to check it if there are any questions about the data later. You will not need more than five keystrokes per cell to enter the data. Simply type a space or letter until you have entered all the mutations in a cell and leave the rest blank. For example, for mutation 52 A, go to the box showing bases 51–55 and enter _A, then tab to the next cell. Data entry will take less than one minute per student, and will speed up as you gain confidence, so the whole thing will take about 20 minutes for a class of 30. If you happen to be lucky and have a teaching assistant, this is a simple task to assign.

Using the data provided:
As stated in the article, the raw data used for the example in the article is available online, but I strongly encourage you to generate your own data for each class. Students gain a much better understanding of the underlying concepts if they have experienced the splitting of groups themselves. It will also make it impossible for students to copy results from previous students or other classes, since each data set will be unique.

Determining parsimony:
One of the most important parts of the activity is not just forming a tree, but justifying it. I will show how this is done, with reference to the compiled data in Figure 3 in the print article, moving from examples that are quite clear to ones that are less clear.

- Position 52: Four students have the same mutation and might be considered to have been in the same group at that time. But A and J share eight other mutations, only two of which are also shared with C and F, the two from the time the whole class was together. Furthermore, C and F share 14 other mutations that are not shared with A andJ. It is very clear that the two are from very different groups, and that position 52 must be due to "parallel evolution." Position 54 has a similar explanation, as the two appear to have been separated at the first division, based on other data.

- Position 45: G and I have this one mutation in common with E, H, and B. But G and I have four connections with C and F, while E, H, and B have five similar mutations not shared with G and I or C and F, so the best explanation is that this is also parallel evolution. A similar reasoning holds for position 100. These both show groups separated at the second division, that subsequently have one mutation in common.

- Positions 51 and 68 both show the same mutation for A and D, while A and J have the same mutation at two positions, 30 and 52. In this case, two explanations would be equally parsimonious. But if A were grouped with D instead of J, this would lead to a three way convergence at site 52, which is highly unlikely, so based on that, the tree shown appears to provide the most parsimonious explanation.

- Position 53: Three-way convergence does happen occasionally, however. Here A and C appear to have been separated in the first division, based on the five mutations A shares with J and D that are not shared with C or I. Similarly, C shares seven mutations with F, and I shares six mutations with G, whereas C and I have only the one pair in common.

- It is interesting that most of the questions occur in one block, from 51–55. This is real data, so it was not planned, but if you have something like this occur, you can use it to point out that some sections of DNA seem prone to mutation, while others are more highly constrained, mirroring what actually occurs in nature.

Justification of model:
In the analysis section of the report, students are asked to provide a justification of their model. Most will be able to describe the reason for the larger groupings. Fewer will be able to produce an argument following the lines just used above to describe exactly why their model is the most parsimonious. This is a fairly advanced concept and fairly abstract. Do not spend too much time on it.

Drawing the tree:
It is far easier to draw a tree by hand than by computer, unless you have access to a program that allows full control of placement of elements on the page. Students would have to create the background lines, the branch lines, and text boxes for the mutations. It can take an hour to draw the tree on the computer, or five minutes by hand. Realistically, unless they have access to and experience with design software, it is not worth the effort.

Comparing trees:
If you decide to compare the different trees produced by the different groups (which I strongly encourage), one of the best ways to do it is to have two or three groups discuss differences between their models in conference. In the process, they often discover the principle of equivalence of structures with the same branching pattern but different order. They will also learn how to judge parsimony of models as they have to go back to the data to determine how many mutations each group has in common. Then when you display the best three or four, students will be much more active in the process of evaluation. As I wrote in the article, it may be better to use a low-tech option than a high-tech option if it facilitates side-by-side comparison.