

Classifying Animals by Appearance versus DNA Sequence

Megan E. Rokop

Hello. My name is Megan Rokop and I work here on MIT Campus at the Broad Institute of MIT and Harvard. Today we'll be doing a lesson about how you classify mammals based on how closely related they are to each other. But before we do that I just wanted to show you a few pictures of mammals and ask you how you think that they're related. So on your worksheet and also on this screen right here, I've shown you ten pictures of mammals. And what I'd like you to do is look through those pictures and pick the two mammals that you think are most closely related to each other. So take a moment to pick those two mammals and I'll see you when you get back!

Hi! Welcome back. I hope you've taken the time to pick out the two animals that you think are most closely related to each other. I'll actually be revealing the answer to that question later in the lesson.

But first I wanted to introduce to you the tool that we'll be using throughout the remainder of the lesson, which is a tool that shows you how closely related groups of organisms are to each other. And I'm going to be showing you this tool on the following slide. Here you can see the tool that we'll be using, which is called a phylogenetic tree. A phylogenetic tree is called a tree because it looks like a tree. It has a trunk and it has branches. Here you can see that this tree has a trunk and it has three branches. At the end of the three branches are three groups of organisms, and those groups of organisms are bacteria, archaea, and eukaryotes. These are actually the three groups of organisms that all known organisms in life fall into.

You've probably heard of bacteria before because they are organisms that are known to cause disease. But actually most bacteria on earth don't cause disease. They do things like live in the soil and decompose matter, or live in your intestines and help you digest food.

The second group are archaea and you may or may not have heard of them. They are prokaryotes just like bacteria, which means that they don't have a nucleus. Because they are prokaryotes they look a lot like bacteria.

The third group are the eukaryotes and that's the group that do have a nucleus. Do you know what the organisms keep in their nucleus? They keep their DNA. We are eukaryotes because as humans we have nuclei in our cells where our DNA is kept.

So let me explain to you what this tree shows you about how bacteria, archaea and eukaryotes are all related to each other. If you look up at the tree what you see is that the branches have different lengths. The branch length between archaea and eukaryotes can be shown if you trace the line between archaea, from archaea around to eukaryotes. The length between archaea and bacteria are shown if you trace the line between archaea, around the side of the tree and back to bacteria. So what you want to do is look at these branch lengths and ask yourself, "Which one is shorter and which one is longer?" And the shorter distance was the one between archaea and eukaryotes. But the distance between archaea and bacteria was much longer. And what that is telling you is that archaea and eukaryotes are more closely related to each other, whereas archaea and bacteria are less closely related. So that's how you use one of these trees. This tree shows you how closely related groups of organisms are to each other because the length of the branch between the two groups is shorter if they're more closely related and longer if they're less closely related.

Now the tree I just showed you only had three branches but you can actually make phylogenetic trees with lots of branches. So in my next slide I'll show you a tree that has a lot of branches. Here's a phylogenetic tree that has many more branches, but you still see the trunk down at the bottom and the three groups of organisms. You can still see that bacteria are on the left, archaea are on the middle and the eukaryotes are on the right. But now what you see is that these organisms, each organism, for instance bacteria, has many different groups shown over here.

The same is true for eukaryotes. There are many different groups of eukaryotes. If you look at this tree you might be able to find where the mammals are. The mammals are actually in this group right here, under animals. And that's what we'll be working with today are all different mammals which fall into that group. But I just wanted to show you this tree to you remind you that there are many other forms of life on the planet and actually there are many more different kinds of bacteria for instance as shown on this tree than there are kinds of animals.

So now that I've shown you this tree and I've explained to you how to read the tree, I want to show you how to make the tree because that's actually how we're going to be doing the lesson today is that you're going to be making phylogenetic trees, and so I'd like to explain to you how we make these trees. Phylogenetic trees can be made in two ways and that's shown on my next slide. The two ways phylogenetic trees can be made are using the physical characteristics of organisms, or using their DNA sequence. And I would like to show you the first way first, which is based on physical characteristics. To do that I'd like to show you three pictures of animals. Here are the three pictures right here. There's a rat and then a frog and then a mouse. And what I want to ask you is which of these organisms do you think are more closely related to each other? And the way that you're going to show me which you think are most closely related is by filling in the blank tree shown on the slide now. This blank tree has a trunk which is right over here, and it also has three empty branches, one, two and three. I want you to write into this tree rat, mouse and frog, where you think that those organisms go. I'm going to give you a moment to draw down a blank copy of this tree and write in rat, frog and mouse into the tree the way that you think that they're related. And then I'll see you in a moment to go over the answer.

Hi! Welcome back! Hopefully you've had a chance to fill in the blank tree that I showed you with the three organisms, rat, frog and mouse that I showed you pictures of in my last segment. Now I'd like to show you what that answer is. And the answer is in this slide right here. So as you can see, rat and mouse are at the top of the tree, whereas frog is at the bottom. And this is because if you look at these distances here, the branch between rat and mouse is very short, and so that shows that rat and mouse are very closely related. In contrast, if you trace the distance between rat all the way around to frog, that would be a very long distance, showing you that these organisms are less closely related. I also just want to point out that whether you put rat and mouse on the top or mouse and then rat, it doesn't matter. Both are correct because the branch length between rat and mouse on this slide here are actually symmetrical. So it's OK if you put rat and then mouse, or mouse then rat. Well great! Now you've learned how to fill in phylogenetic trees using physical characteristics.

So now I want to go back to all ten of our mammals and have you fill in that tree. So what I want you to do is look at the worksheet that you have with the ten pictures of mammals, which is entitled, "Classification of mammals by physical characteristics." This worksheet has the ten pictures of mammals but it also has a blank phylogenetic tree. So I'll show you that picture in my next slide. Remember, here are the pictures with the ten animals and here is the blank phylogenetic tree. Now it's not actually blank because I filled in three spots for you with three other mammals that aren't in your list of ten. Those three mammals that I filled in for you are platypus, all the way up at the top, cow, which is here in the middle, and macaque which is down here at the bottom. So you can use those three animals that I filled in for you as reference points. What I'd like you to do now is look at the ten pictures of mammals that I gave you and fill in this blank phylogenetic tree with the ten mammals that you have pictures of. Remember you can use the platypus, cow and macaque as reference points for thinking about where you want to place the organisms in the tree. Also remember that earlier in the lesson I asked you to pick out the two organisms that you thought were most closely related to each other. And so you can think about where you would want to put those on the tree. Would you like them to be close together or far apart? So take a moment to fill in your blank tree with your ten animals and I'll see you in a moment.

Hi! Welcome back! Hopefully you've taken a moment to take the ten pictures of the mammals that I've shown you and fill in your phylogenetic tree based on physical characteristics. But now I'd like to show you how to make a tree the other way, which is using DNA sequence. Remember, as I'll show you on this slide, there are two ways of making phylogenetic trees. You've done one, but now I'd like to show you how to do the second one, which is using DNA sequence. What I'd like to do first is show you the type of DNA sequences that you'll be using to do this exercise, so that you know what a DNA sequence looks like when we write it with scientific notation. So here in this slide what I'm showing you is first of all the structure of DNA. You probably remember the shape of DNA as a double helix and that's shown here on the left. Well, we can write that double helix out in scientific notation by writing the sequence of each of its two strands. And that's shown here right next to the double helix. Remember DNA is a double helix so it has two strands. So we write one strand on the left and one strand on the right.

The other things you should remember about DNA for this lesson are that it's made up of four letters, A, T, G and C. And whenever you find an A on one strand, there's a T on the other strand in that position. And whenever you find a G on one strand, there's a C on the other strand in that position. And so that's how we write a DNA sequence. Now in the slide on the left what I'm showing you is a DNA sequence written vertically, but most of the time scientists actually show this written horizontally. So here's a picture of DNA written horizontally. But sometimes we actually only write one strand. And so here at the end of my slide I'm showing you that you can write just the top strand of DNA because you know the base pairing rules, so you can always write the second strand and infer it from the top strand. So for now on, all the DNA sequences I'll be showing you are just the top strand of the sequence.

So now let me show you three DNA sequences and we're going to make a tree from them. So my next slide shows three DNA sequences. I want you to think about these DNA sequences as three different organisms. It doesn't matter what they look like because this time

we're going to be making the tree based on DNA sequence. I'm going to call these organisms first, second and third. So now on my slide you'll see that they've been named first, second and third. And I also have a new blank phylogenetic tree for you with three empty branches. What I'd like you to do is now fill in this blank tree using the organisms first, second and third, and just look at their DNA sequences just like you looked at rat, frog and mouse, and think about which ones look the most similar to each other and put those close together. So I'll give you a moment to fill in this tree based on DNA sequence and I'll see you in a moment.

Hi! Welcome back! Now that you've filled in the blank tree using DNA sequence, I actually want to tell you something else about making trees using DNA sequence before I show you the answer. The thing I'd like to tell you is that you can actually use quantitative measures to measure the differences of DNA between organisms in order to make these trees. So what you just did was qualitative. You looked at the three DNA sequences and thought about which ones looked like each other. But now I want to show you a slide where you can count the differences between DNA from each of these three organisms.

So on this slide what you'll see is that I've now put in a place for us to write down the number of differences between the organisms first and third, first and second, and second and third. So what I'd like you to do is think about and count how many differences are there between first and third and I'm going to write that right here. Then you're going to count the differences between first and second and I'm going to write that number right there. Finally you'll count the differences between second and third and we'll write that number right there. So I'd like you to take a minute to count those differences and I'll see you when you're done.

Hi! Welcome back! Now that you've counted the differences between the organisms first, second and third, I'd like to show the answers for how many differences are there. So in my slide I will now record the number of differences between these groups of organisms. So here you can see that first and third had only one difference, so I've written the number one right there. However first and second had six differences and second and third had six differences. This means that first and third are really closely related, whereas first and second are less closely related and second and third are less closely related. So now we can use this quantitative information to fill in the tree. Remember the blank tree I showed you? Well, let's trace the distance of the lines between the different organisms and see how they're related. When you filled in the tree qualitatively it should have looked like this on my next slide. Third and first were up at the top, and second was down at the bottom. If you think about the distance between third and first, that would be this distance right here, and that's a very short distance. But think about tracing the distance between for instance first, all the way around the tree to second. That distance is much longer. And so now let's think about the number of differences we counted. Think about the one difference between third and first as being the distance that you measure between those organisms on the tree. It's a very small number so you get a very small distance. But you measured six differences between second and first, which means that the length of the line between them should be six times as long. And that's how we make these trees using DNA sequence quantitatively.

So now what I'd like you to do is to start to make this tree using DNA sequence, but instead of the organisms that I've given you, using the organisms that are on your worksheet. Remember the ten mammals? We're going to fill in that DNA tree based on DNA sequence now. So let me show you what we're going to do. If you turn to the first page of your worksheet, you'll see DNA sequences from six of the organisms that I've shown you already. And I'll show you that on my next slide. What you can see on the front page of your worksheet entitled, "Classification of Mammals based on DNA Sequence" are these six DNA sequences from cow, dog, hedgehog, horse, mouse and shrew. What I'd like you to do is count the number of differences between each pair of organisms on this list. And I'd like you to record your answers in the chart that I've given you on the front page of that worksheet. I'll show you that chart in my next slide.

That chart looks like this. As you can see, some of the boxes are shaded, but some of them are unshaded. You only need to see the unshaded ones to fill in the unshaded ones because this chart is actually symmetrical, so the shaded boxes right here are the same as the unshaded boxes. So you don't need to fill in both halves of the chart, you only need to fill in the unshaded boxes right here. I've done an example for you to show you how to do this. If you see the box right here that says number three, that box is at the intersection of the dog column and the hedgehog row. And what that's telling you is that dog and hedgehog DNA sequences have three differences between them. So I counted the differences and I put the number three in that box. So what I'd like you do now is count the differences between the remaining pairs of animals and then place the number of differences in each of the unshaded boxes in this chart. So you should count the differences between the DNA sequences now and I'll see you when you're done.

Welcome back! By now you should have filled in each of the unshaded boxes on your chart to show the number of DNA sequence differences between these organisms. So let me show you what you got when you filled in the chart. On this slide you see that your chart was filled in to contain these numbers. So the first column should have read 3,5,3,8,5. The second column reads 3,1,8,3. The third column reads 3,7,1. The fourth column reads 8,3. And then finally the last box should have said 7. So now that you have this chart filled in, we're going to learn how to use this information to fill in the blank phylogenetic tree for these mammals. So in this worksheet after the chart, the next picture you'll see is a blank phylogenetic tree where you can fill the mammals in that you just counted the differences between. And I'll show you what that tree looks like on my next slide.

On this slide you can see that this time the tree is not actually completely blank. There is only five blank branches left for you to fill in. Remember you counted the DNA sequence differences between six mammals. The one that is on the tree already is cow, the first one. The other five are the animals that you're going to fill in. So on this tree there's five blank branches, two above cow, two below cow in a pair, and one right here. So you're going to take the remaining five mammals whose DNA sequences you have and fill them in on the tree around cow. On your worksheet I've written a few suggestions for you about you can use these numbers to fill in the tree, but my main advice is to first figure out which of those organisms are most closely related to each other and think about putting them close on the tree. And then think about how the organisms are related to cow. Whatever is most closely related to cow you should put closest on the tree. Then place the next most closely related organism to cow on the tree, and

finally, end with the organism that is the most distantly related from cow, and place it furthest from cow on the tree. What I'm going to give you now is some time to fill in the tree with the remaining five mammals and I'll see you when you're done.

Welcome back! Now you should have filled in your tree of mammals based on DNA sequence. And I'd like to show you what the tree should look like now. So on my slide I'll show the blank tree that you just had, and now the organisms will appear that you just filled in using the DNA sequence information. So here's the tree with the animals you just filled in shown in blue. And the chart with your differences is shown on the right for your reference. Let me explain to you how you can use the numbers in these charts to have filled in the tree.

What I would suggest you have done first was remember you were to look for the organisms that are most closely related to each other. In your chart, that would be the number that's the smallest because the smallest number of differences would be most closely related. If you look on the chart there's only one difference, right here, between dog and horse. So dog and horse had to be on the tree very close to each other. But there's actually another pair of organisms that also gave the number one and that's right here, at the intersection of hedgehog and shrew. So hedgehog and shrew also had to be in a pair close together. Now on the tree there were two places around cow for pairs of organisms. So how do you know which pair goes in which place? There was a pair of organisms above cow and also a pair of organisms below cow. Well the answer is you need to determine which ones are more closely related to cow. So let's look at the cow column of the chart right here. In the cow column dog and horse gave only three differences, so they should be closer to cow, whereas hedgehog and shrew gave five differences, so they should be the pair furthest from cow. The only organism we have left to place is mouse. And you remember mouse, if you look in the mouse row right here, and the mouse column right here, it always gives the highest numbers, 7's and 8's. And that's why mouse ended up furthest away from cow on the tree. So if you look at the tree, mouse can be seen right here. And if you trace the distance to cow, it's the furthest organism from cow of the five that you filled in.

Well now we've filled in phylogenetic trees in two ways, based on physical characteristics and based on DNA sequence. So what I'd like you to do now is take those two trees, look at them, and answer the following questions for me. These questions are on your worksheet but I'll also show them to you on my next slide. The questions I want you to think about the answers to now are:

- Do everyone's trees based on physical characteristics look the same?
- Why or why not?
- And do everyone's DNA based trees look the same?
- Why or why not?

So I'd like you to take a moment to think about the answers to these questions with your classmates and I'll see you when you've discussed the answers.

Welcome back! I hope you enjoyed discussing the answers to these questions with your classmates. Now I'd like to go over the answers to these questions with you. The first question was, Do everyone's trees based on physical characteristics look the same? I'll bet that a lot of the

trees in your class looked very different from each other. This is because making trees based on physical characteristics is subjective. For instance I might look at the ten organisms and think, "Mouse and shrew are very closely related to each other because they both have very long tails." But another person may look at mouse and possum and think that they're very closely related to each other because they're both very small. Because people use different differences between these organisms to classify them, making trees based on physical characteristics can be subjective.

In contrast, when you make phylogenetic trees based on DNA sequence, there's two characteristics of making those trees that way that make them much more reliable. One characteristic is that they are objective. You are using DNA sequence to determine how similar the organisms are and DNA sequence is just a string of letters, A's, T's, G's, and C's, that everyone interprets the same way. So that's why it's more objective to make trees based on DNA sequence. The other thing is that making trees based on DNA sequence is quantitative because you're counting the number of differences between the letters. This means that you can actually calculate a number of differences. And so everyone calculates the same number when they count the differences and that's why the trees turn out the same, when, for instance, I would make a tree based on DNA sequence versus another scientist.

So now that we've discussed the answers to this question, I would like to move on to another question that I'd like you to discuss. This question is shown on my next slide. The question is: Which method do you think is used to generate currently scientifically accepted phylogenetic trees? Do you think scientists now use physical characteristics or a DNA based method? I'll give you a moment to discuss the answer to this question with your classmates and I'll see you when you have an answer.

Welcome back! I hope you enjoyed discussing the answer to this question with your classmates and now we'll go over it. The question we were discussing is: What method do you think is used to generate current, scientifically accepted phylogenetic trees? Well the answer to this question has actually changed fairly recently and that's because in the 1970's a technique was invented called DNA sequencing. And DNA sequencing allowed scientists to view the DNA sequence of organisms. DNA sequencing means that you can determine the order of the letters A's, T's, G's and C's in any organism's DNA. Before the 1970's this technique was not available and so scientists used physical characteristics to make these trees. But after the 1970's when scientists were able to sequence the DNA of organisms, scientists began making phylogenetic trees using DNA sequence. This is because making trees using DNA sequence is both more objective and more quantitative than making trees based on physical characteristics. So now that we know how scientists make phylogenetic trees these days you might ask, "Is that any different than the way we made phylogenetic trees today based on DNA sequence?" And the answer is that there's really only one main difference. And that difference can be demonstrated in my next slide.

In this slide you can see 2,000 letters of DNA. Now this may look like a lot of DNA to you because you're used to looking at segments of DNA in this lesson that are only about sixteen letters long. Well, we used segments of DNA that are sixteen letters long to make our trees today based on DNA. But the difference between how we did it today and how scientists do it is that scientists use a few thousand letters of DNA to make these trees, like what's shown here on the

slide. Now it's OK that scientists use only segments of DNA instead of all of the DNA in an organism, because the amount of DNA in an organism is actually enormous. In this slide when you look at 2,000 letters of DNA it looks like a lot, but the amount of DNA in a mammal is about 3 billion letters of DNA. So it's computationally extremely hard to compare 3 billion letters of DNA from different organisms. So instead, scientists use small segments of DNA, a few thousand letters long, and compare those segments to compare organisms and make these trees. Today we used small segments that were about sixteen letters long in our lesson, but that was to demonstrate the principle that scientists use small segments of DNA, even though the scientist's segments are a little bit longer.

Now when you think about segments of DNA that are thousands of letters long, think about doing our lesson today and how you would have to count the pair-wise differences between the different organism's DNA sequence. That would probably take a long time! And that's why scientists have written computer programs that take these thousands of letters of DNA from different organisms and count the pair-wise differences between them. So the only main difference between how scientists make these trees and how you made them is that scientists use thousands of letters from each organism and they use computer programs to count the pair-wise differences and then the computer programs generate the phylogenetic trees using the number of differences between each organism. And that's how scientists make these trees today.

Now that we've answered that question, I want to leave you with one other question about your tree and that's shown on my next slide. The question is: Which animal's placement on the DNA based tree surprised you the most and why? I'd like you to take a moment to think about your answer to this question and when we come back I'll tell you my answer to this question.

Welcome back! I hope you've taken a moment to think about which of the organisms' placement on your DNA tree surprised you the most. This was obviously an opinion question and so everyone may have come up with a different answer, but I thought I'd share my answer with you since there's an organism on this tree that surprised me a lot. The organism that surprised me the most is the tenrec. And so to talk about a tenrec I'd first like to remind you of the pictures of the ten mammals that are from this exercise. And I'll show you those now. When I first started working at the Broad Institute I had actually never heard of a tenrec, which is right down here on the bottom right corner of the slide. I had never heard of a tenrec but I looked at this picture here and I thought that the tenrec looks just like a hedgehog. I thought it looked like a hedgehog because it has a small pointed nose and small eyes and it has quills all over its body. In addition it has small feet that seem to have sharp claws. So I looked at this picture of the tenrec and the picture next to it, the picture of the hedgehog, and I assumed that since they looked so similar they must be really close on the tree.

Well now let me show you the tree so that you can see where tenrec and hedgehog end up. Here's the tree now all filled in. And I want you to look for tenrec and hedgehog. I'll point out to you that tenrec is up at the top and the closest organism to it is actually the elephant. Meanwhile, hedgehog is more in the middle and the closest related organism to it is the shrew. Now tenrecs and elephants don't look anything like each other but they're the organisms that are closest together on this tree. And based on DNA that shows you that the tenrec's closest relative that we've sequenced so far is the elephant. Well this surprised me a lot because I think that we

often think that things that look very similar are very closely related. But in this case, hedgehog looks like tenrec but tenrec is more closely related to elephant. So that surprised me a lot.

If you had asked me at the beginning of this lesson like I asked you to pick out two organisms from this group of ten that looked most like each other, I would have picked out tenrec and hedgehog. But seeing this tree based on DNA, I now know that tenrec has DNA more similar to elephants. So I'd like you to think about when I asked you that question at the beginning of the exercise, what two organisms did you pick that you thought were most closely related? And I'll give you some time to look back at the DNA based tree and to look at where they actually are. Now you can do that once this lesson is over, but for now I would just like to say that I had a lot of fun giving this lesson and I really hope that you enjoyed learning about phylogenetic trees, about how they show you how closely related organisms are and how you can make them, either based on physical characteristics or based on DNA sequence. Thanks for joining me!

Hi! Welcome to the Teacher's Guide for my Blossoms lesson on "Classifying Mammals" both by physical characteristics and by DNA sequence. Today I'll be reviewing with you the learning objectives for this lesson. The main learning objectives are that the students learn how to make phylogenetic trees, both based on DNA sequence and based on physical characteristics. The students will be working with a group of ten animals whose pictures are provided in the lesson, and they will first look at those pictures and make a tree based on how the organisms look. And then we will provide them with DNA sequence from those organisms and a new blank phylogenetic tree which they will fill in based on DNA sequence.

The prerequisites for knowledge of this lesson are mainly that the students have some understanding of what DNA is, that DNA is a genetic information that codes for organisms and that DNA is made up four letters, A's, T's, G's and C's. It's helpful if the students know the base pairing rules for A pairing with T and G pairing with C, and it's also helpful if the students know how to write a DNA sequence and for instance label the ends of the DNA sequence with their 5' and 3' ends. We do review this information in DNA briefly during the lesson, but it helps if the students have seen it before. It's also helpful if the students know that each species on earth has a different characteristic DNA sequence and that's why organisms of different species are different because they have different instructions in their DNA.

There are two handouts that are provided with this lesson. One is entitled Classification of Mammals by physical characteristics and that's the first one that the students will be using. The other is Classification of Mammals by DNA sequence and that's the second worksheet that the students will be using. The only other supplies necessary are blank paper in case the students need scratch paper, and writing utensils. The timing of the lesson is that there's about thirty minutes of video footage from the lesson, leaving about 30 minutes for the students to work on the activities between each segment. I'll go over those activities now.

After segment one, the students will be looking at the group of ten mammals and their pictures in the worksheet, the first worksheet that they'll use. And I'll ask the students to pick two organisms from that list that they think are very closely related to each other. So that's what the students should be doing after segment one is picking two organisms out of the list that, based on their pictures, they think are very closely related to each other.

After segment two I will teach the students how to make a phylogenetic tree based on physical characteristics using a very small example with only three animals, a rat, a frog and a mouse. I'll give the students a blank phylogenetic tree that has three blank branches and after segment two the students will be asked to fill in that blank tree using the rat, the frog and the mouse and placing them where they think they'll go on the tree. So after segment two the students will copy down the blank tree and write in rat, frog and mouse where they think they'll go on that tree just based on looking at the pictures of those organisms.

After segment three the students will be filling in the first blank phylogenetic tree that they get with the ten mammals. That's the tree on Classification of Mammals by physical characteristics. They'll use the ten pictures of the animals on that worksheet and the blank phylogenetic tree to fill in the tree based on their own ideas about how these organisms are related. It's totally fine that the different students in the class won't fill in this tree in the same way because they'll actually see later that because filling in trees based on physical characteristics is subjective, that it actually makes sense if they all get different answers in this part.

Segment four is about learning how to make trees based on DNA sequence and so after segment four I'll ask the students to fill in a small blank phylogenetic tree with only three blank branches with three organisms which I'll just call first, second and third. The way they'll be doing that is looking at DNA sequences from the organisms first, second and third, and just looking at those sequences and thinking about which ones look the most similar to each other. So after segment four the students will be looking at their trees, at these DNA sequences and filling in their small trees based on those three sequences.

When we return in segment five, I'll explain to the students that we can actually quantitatively count the number of differences between the DNA sequences of these three organisms, first, second and third. So after segment five the students will be counting letters of DNA to figure out how many differences are between first and second, second and third and first and third.

In segment six they'll translate these new skills to the second worksheet which is Classification of Mammals based on DNA sequence. In that worksheet they'll get DNA sequences from the mammals that we're working with and they'll also have a blank chart that allows them to fill in that chart with the data that they've recorded about the number of sequence differences between each of the mammals. So again, after segment six the students will be counting DNA sequence differences and filling in a chart to record those numbers.

After segment seven the students will then use the numbers in that chart to fill in a new blank phylogenetic tree, this time based on DNA sequence. That second tree actually isn't completely blank, there's only five blank branches. So the students will be filling in that tree using the DNA sequence data from the five organisms whose DNA sequences they've just counted. I have some suggestions written on that worksheet for how the students will translate the numbers in the chart into the blank phylogenetic tree because this section can be a bit challenging for students. So those instructions will be there in written format above their blank phylogenetic tree based on DNA, although it may also be helpful to review those suggestions with them while they're learning how to translate the numbers in that chart into the blank phylogenetic tree.

The next thing that the students will do after segment eight is actually compare the two trees they've made. They'll now have two trees, one based on DNA sequence and one based on physical characteristics. And I'll pose two questions for discussion for the students so that they can discuss whether their classmates' physical characteristics trees and their classmates' DNA based trees look similar to each other. The students can discuss these either individually, in pairs, or as a class as a whole.

Then, in segment nine, we'll come back and talk about the answers to that question and then I'll pose another question which is how the students think that scientists currently make these trees. And so then after segment nine the students will again have time to have a discussion about the way that scientists now make these trees, such that they generate the current scientifically accepted versions of these trees.

When we come back in segment ten I'll tell the students how scientists currently use the DNA sequence to make these trees and then I'll pose another question to the students which is that which animal on their DNA based tree surprised them the most in terms of where it was placed on the tree. Now this is an opinion question so obviously the different students in the class can have different opinions and that's totally fine. But when we come back for segment eleven,

I'll just give my answer to that question as an example, although it may be different or similar to your students' answers.

In segment eleven I'll reveal that the tenrec was the organism that surprised me the most based on its placement on the DNA tree and I'll explain a little bit why that is. Finally, I'll pose one last question for the students that they can do after the completion of segment eleven which is the end of the lesson. This question is: Where were the two organisms that at the very beginning of the lesson in segment one the students picked as being the two most closely related to each other. They can now look back at their final and complete, filled in DNA based tree, find those two organisms and see out of curiosity whether those two organisms are close to each other on the tree or far away.

So that's the last thing they can do in the lesson after segment eleven is done. I hope these instructions have been useful to you and thank you so much for your interest in our lesson based on Classifying Mammals on DNA sequence versus physical characteristics.

Hi! Welcome back to the Teacher's Guide Part Two for the lesson on Classification of Mammals on physical characteristics and on DNA sequence. I wanted to tell you a little bit about an extension to the lesson that we've done today in case you'd like to use it with your students. The necessary equipment for this lesson are computers and an internet connection. So if you and your students have access to computers and an internet connection, then the software I'll be showing you today are all freely accessible on the web, and so you and your students can do this exercise with just internet access.

The lesson relates to the lesson that we've done already, making phylogenetic trees based on physical characteristics and DNA sequence. The difference is that in this part of the lesson the students can actually use the computer program that scientists have written that allows scientists to analyze hundreds or thousands of letters of DNA sequence from mammals such that they can have the computer count the pair-wise differences and make the phylogenetic tree. The instructions for how to use this software are all contained in the two additional handouts that are part of this lesson's extension.

There are two additional parts and they can be used together or separately. But in general the first part is all about how to find a sequence of DNA from a mammal on the internet, how you can find that sequence and use it figure out how similar different mammals are to each other on DNA sequence. The second part is where the students will use the computer software on the internet that allows them to actually generate a phylogenetic tree using hundreds of letters of DNA sequence from seven different mammals. So let me show you how this exercise works.

It starts with the slide I'm displaying now. This slide is a screen shot from an internet program from NCBI. This program allows you to find a sequence of any gene from any organism by just typing it into this search database. So this is what the page will look like when you go on the internet following the instructions on the extension handout. What you'll see is that there's a window at the top of this handout where you can type in the name of any organism and any gene and find a sequence. So you just go to the pull down bar under nucleotide which is highlighted here and then in the search box you type in the name of an organism and the name of a gene and it will find that sequence for you.

In the example I've given you type in the Latin name for possum and then type in a gene name that is associated with cholesterol. It's actually the gene for the receptor for bad cholesterol which is called LDL. So you type in the Latin name for possum in the search box which is in the handout and the LDL receptor gene name, also in the handout, and then you just click enter and the search finds you the sequence. What you'll see when you do this search is on the next slide.

When you do the search you'll get this result. *Monodelphis domestica* is the Latin name for possum and so you see that it says *Monodelphis domestica* at the top and then LDL receptor which is the name of the gene that I'm using in this example that has to do with bad cholesterol. Then what you can see here are some different characteristics about this gene. But if you scroll down to the very bottom of this website, when you get to the bottom you actually see this information right here. And this information at the bottom here is the DNA sequence of this gene in possum. So in this way you can type into this search database any name of any organism and any gene name, and then the database will provide you with the DNA sequence. So this one is of the DNA sequence of the gene I'm using in this example from possum.

Well now what can your students do with this sequence? What they can do is highlight and copy this sequence and then paste it into one more search database and I'll show you that database on the next page. This database is called Blast and it is also part of NCBI and what you're

students can do is click on the link that is shown right here that is called nucleotide Blast. This link takes you to a search database that allows you to search all known DNA sequences, or all known nucleotide sequences in the database that contain all sequences that scientists have generated. So what your students will do is click on that link and the next page that will come up is shown in my next slide.

So this screenshot is now of the program that will allow you to type in your DNA sequence and search the database for any similar DNA sequences that are in the database. So the large box shown at the top right here is where the DNA sequence will be pasted. And then all you do is scroll down to the bottom of the page and click on a big blue button that says “blast” and then you’re searching the database of all known DNA sequences. This is just like doing a Google search for instance where you type in a word and you hit the Google button and you find all sorts of different matches to your search. The only difference is that what you’re searching for is a DNA sequence and so the database you search is of all known DNA sequences. What you get as a result of this search is shown on my next slide.

The results of this search are shown here on the website that will pop up when you hit the blast button. And what you actually want to do is scroll down past this box of red lines and what you’ll see is my next slide. So what you’ll see when you scroll down past the red lines is this information here. This is the result of your search. For every search it gives you a species name and a score. And the score is how well that sequence of that species matched to the one you inserted into the search. The one you inserted into the search was the possum gene and so this allows you to see how a lot of different organisms are related to possum. So let me show you that slide again but this time I’m going to put boxes around the different names of the species that come up when you do this search.

So here’s the screenshot again from the website, but now I put boxes around the top seven animals that show up when you do this search. The first one that shows up is the Latin name for possum and that one gets the highest score. Well that makes sense because the sequence you inserted into the database was the possum sequence. But if you look at the results, the next one that shows up is actually the Latin name for platypus, shown here. The next one is *Sus scrofa*, which is pig. And then you get cow and then horse and then human and the last one that you get, which is all the way down at the bottom of the screen is macaque. These are some of the different animals that you saw earlier in the exercise where all of the ten pictures of the mammals were shown.

Next to the names of the mammals there is also the score column and the scores go down as the screen goes down. This is because the best match is shown first, so it gets the highest score, and then the scores decrease as the matches get worse and worse throughout the page. So the best match is at the top with the highest score, and the worst match is at the bottom with the lowest score. So this allows students to see that as the scores are higher, it means that that organism is more related to possum because possum is the one you put in the database.

Well now what can we do with these sequences because now we have DNA sequences from seven different mammals. It’s all sequences from the same gene, remember that gene that is associated with bad cholesterol. But now we have it from possum, platypus, pig, cow, horse, human and macaque. Well, we can take these seven mammals’ sequences and use the computer program that makes phylogenetic trees from these sequences. And that’s what the second worksheet that’s part of the extension does. And this uses a different piece of software.

So let's move on the second worksheet and I'll show you what that looks like. When you go to this worksheet you'll see a website that looks like this and that website is called ClustalW. ClustalW is one of the programs that scientists have written that allow you to plug in DNA sequences and make a phylogenetic tree. Down here at the bottom of this web page is a search box and in the search box your students can plug in the DNA sequences from the seven different mammals into that search box. And then all they do is click the button at the bottom of the page that says "Run." And then what this program does is that it takes the seven different DNA sequences and counts the pair-wise differences between them and it makes the phylogenetic tree. So when you hit run it takes two or three minutes, and after two or three minutes a web page pops up that looks like my next slide.

This web page looks like this and this is the result from the ClustalW program. If you scroll all the way down to the bottom of this page, you'll then see this picture here and that is the phylogenetic tree that this program has created for you based on the seven sequences that you inserted into the program. So you can see here that this tree has seven branches with possum, platypus, pig, cow, horse, human and macaque. And it shows on the tree how those organisms are all related to each other. Now that you have a tree you can also actually go and look back at the DNA sequences of the seven different mammals all lined up to show you how similar they are. This is on the same ClustalW website that you saw your tree on, it's just further up on the page. So right now if you're looking at the phylogenetic tree on that web page, you just have to scroll up a little bit and you'll see what's called a sequence alignment, and I'll show you that on my next slide.

So this sequence alignment is just up a little bit further on the ClustalW page and it shows the seven sequences from human, macaque, boar, cow, horse, possum and platypus all lined up with each other. And whenever there is a DNA letter that's the same in all seven animals it gives a star. So for instance this letter right here is a T in all seven animals and so that's marked with a star at the bottom of the column. So this is what a sequence alignment is. It takes the DNA sequences from all seven mammals and it lines them up and then it puts a star next to any single letter that's the same in all seven mammals.

So this is an extension to the activity that you can do if you have computers and internet access. It shows you how to find the DNA sequence of any animal by just typing in the name of that animal and the name of the gene you want into the database and pulling up that sequence. It also shows you how to use a computer program to take that same gene sequence from seven different mammals, plug it into a computer program, and obtain a phylogenetic tree that shows you how those seven animals are related.

I hope you enjoyed this lesson extension and thank you for joining me!