The Tree of Life

Science Film Festival
Knowledge Through Entertainment

Science Film Festival Film

▶ **Out of the Cradle: The Origins of Humanity**

Introduction

In this project, you'll use a Web-based computer program to help make a "human family tree." Think about or draw out your family tree adding aunts, uncles, and cousins. (If you don't have siblings or cousins just draw a big family tree from your imagination.) Based on your family tree, you can see that you are more closely related to your sister (or brother) than you are to your cousin; that is there are fewer "branches" separating you from your sister than there are separating you and your cousin.
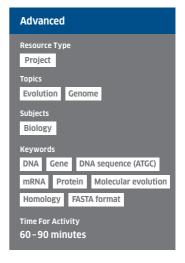
Now imagine that a biologist arrived at a big family reunion and had no idea who were sisters, cousins, aunts, uncles, etc. but tried to sort it out by how all of you look. Just based on how you look, would s/he be able to guess which of the two kids standing next to you is your sister and which is your cousin? In many families, the biologist may be able to make a pretty good guess based on your visible features (called your morphology), like number of arms/legs/eyes, hair color, nose shape, etc. (Notice that some of these morphological features are shared by all humans but that other features can be used to distinguish you from one another.) But this is not a fail-safe approach to determining familial relationships–as some people look more like their cousin than their sister, right? You could just use morphology to make a good guess.

So what is the best way to determine how related you are to one another (besides just asking -- but stick with me here)? The biologist would have to look at your DNA! You get half of your DNA from your mother and half from your father. Both of those "halves" are very similar to one another–with one difference about every 1000 base pairs (but out of three billion total letters–that's three million differences!). And your mother and father got their DNA from their parents and so on up the family tree. Your DNA should be MUCH more similar to your sister's than your cousin's because you and your sister both got your DNA from the same parents, whereas there are many more branches in the tree (and thus many more matings and DNA base pair differences entering the tree) between you and your cousin. That is, you are much more similar genetically to your sister because you have more recent common ancestors than you and your cousin.

So how does all of this apply to biology? For centuries, scientists have been trying to draw the family tree that reflects the history and evolution of all animals on the earth. This tree would show which species are more closely related to one another, like the case where you are "closer" to your sister on your family tree than you are to your cousin. For example, humans are more closely related to chimpanzees than to dolphins, so chimps and humans would have fewer branches between them on the "animal family tree."

How do scientists make this family tree? For many years, scientists relied on comparisons of morphological characteristics (like hair, teeth, limbs, fins, hearts, livers, eyes, etc.) to try to figure out who was more closely related to whom. These kinds of comparisons are often accurate, but as you saw in the example of a human family, these physical characteristics can sometimes be misleading. Evidence of this concept is that different scientists would come up with different trees/relationships by using different sets of morphological information! So which tree is "right?"

To think about how to identify the "right" tree, we have to think about how these animals became different from one another throughout evolution. All heritable morphological changes (those changes that can be passed down to the next generation) are a result of changes (mutations) in an organism's DNA. This mutation can lead to a change in a protein

Guiding Questions

1

Which of these apes do you think is most closely related to humans? Orangutan, chimpanzee, or gorilla?

2

Why do you think this is the case? (Based on how they look? Which parts helped you decide? Nose shape, arm length, amount of hair?)

The Tree of Life

sequence or a change in when, where or how much of the protein gets made. That's it! One or a couple of these changes can lead to big a difference in morphology and/or the way a single cell in the organism can function. So over billions of years of evolution, a slow accumulation of DNA sequence (and thus some protein sequence) changes has led to the existence of all of the earth's different species -- with some more closely related to one another than others. This whole process is called molecular evolution.

So, as we saw with the family reunion example, the best way to see how related two organisms are is to compare their DNA or protein sequences. (Remember that a protein's sequence is encoded in its gene's DNA - so the only way to get a protein sequence change is to get a change in the DNA that codes for it.) Those organisms with the most similar DNA/protein sequence are almost surely more closely related than those with less similar DNA/protein sequences.

Why didn't scientists use DNA sequences to build the trees 100 years ago? First, it has only been about 50 years since the discovery that DNA is actually the genetic material that gets passed on through

generations. Second, DNA and protein sequencing technologies have only recently gotten efficient enough that DNA/protein sequence data is available from many different kinds of animals. With all of this new information, scientists are working hard to build the "true" animal family tree. And there have been cases where the tree built using DNA sequence data differs from those built using morphological data!

Key Objectives
• To be able to use a computer program on the web to compare a DNA sequence from several human genes with the corresponding genes in other animals.

Materials
• Computer with Internet access
• Lab notebook

Tasks/Steps

On a separate page in this activity sheet, we have copied partial DNA sequences from four different genes, and for each one we have included the human version as well as the same gene for several other animals. For each gene you should make a hypothesis about which animal is most closely related to humans, then use the computer program described below to analyze the DNA sequence to see if your hypothesis was correct.

For example, "Gene 1" below on this page is for humans as well as several different ape species. Humans are very closely related to ape species. Which of these apes do you think is most closely related to humans? Orangutan, chimpanzee, or gorilla? Why do you think this is the case? (Based on how they look? Which parts helped you decide? Nose shape, arm length, amount of hair?) Make your hypothesis, then follow these steps:

1 Go to the NCBI website:

→ www.ncbi.nlm.nih.gov/

2 On the menu bar on the right-hand side, click on "BLAST." BLAST stands for Basic Local Alignment Search Tool. It is a powerful Web-based tool for sequence alignment. It has several features, but we will use it to compare two genes, showing us where the genes are similar and where they are different.

3 The computer will now display a page with general information about BLAST. If you scroll down, you get to a section with a "Specialized searches" heading. Click on "Global Align (Compare two sequences across their entire span (Needleman-Wunsch))".

4 On the next page, you will see two blank boxes, in which you can enter the two sequences that you want to compare. Leave all the numbers and settings at their default values.

5 From the genes displayed lower on this page, copy and paste the sequence you wish to compare the other sequences to (the human sequence) into the upper blank box that says "Enter Query Sequence". A sequence looks like this, with a ">" symbol and gene name on the first line with several lines of letters underneath.

The Tree of Life

This way of writing out a DNA sequence is called the FASTA format.

Copy and paste the second sequence to compare into the blank box where it says "Enter Subject Sequence". Then click on the button that says "Align."

For your first trial, use the ">human_CFTR" (be sure to include all the 729 letters in the sequence below) as Query Sequence and ">orangutan_CFTR" (with its 729 letters) as the Subject Sequence.

It's very important to use your computer's copy and paste ability for putting your sequences into BLAST. It would be almost impossible to accurately type all those letters one-by-one!

6) **Reading the data. The computer will display "BLAST Results." Scroll down to the strings of letters. The top row of letters (labeled "Query") corresponds to the sequence you pasted into the "Query Sequence" box. The bottom row of letters (labeled "Sbjct" or "Subject") corresponds to the sequence you pasted into the "Subject Sequence" box. Where the base/letters at a particular position are the same (e.g. "C" in both sequences), a vertical line will connect the two lines of letters. Where there is a mismatch (e.g. a "C" at position 10 in the Query Sequence and a "G" at position 10 in the Subject Sequence), the letters will NOT be connected by a vertical line. If you look carefully, you will see on the same page that BLAST also adds up the count of matches as a fraction of**

the total number of letters. Can you find it under "Identities"?

7) **Analysis & Counting. Use BLAST to compare the human gene to the one for each animal: human vs. orangutan, human vs. chimp, human vs. gorilla. Count up the # of differences between each of the test species and human (as a raw number and a fraction/percentage of the total). The species with the fewest differences between itself and human is likely to be the most closely related to humans.**

8) **How does the data fit with your hypothesis? What did you learn about using morphology vs. sequence comparison to infer evolutionary relationships?**

9) **Do the same analysis on the other three genes below. Keep in mind that in this experiment you must compare similar genes to get meaningful results. If you compare a pig sequence for gene BMP7 and a gorilla gene CFTR, you'll get garbage output (but you won't hurt anything)!**

The Tree of Life

Gene 1

Here is a partial DNA sequence from humans, chimp, gorilla, and orangutan for the Cystic Fibrosis gene (CFTR). In the body this gene's product is involved in making sure mucous doesn't build up in the lungs and that the pancreas secretes the right enzymes to help you digest your food. If this gene is damaged, a patient gets Cystic Fibrosis. Since all of the animals listed have lungs and a pancreas, it makes sense they would have a similar CFTR gene sequence that would provide a similar function.

There should be 729 bases/letters for each sequence.

>human_CFTR

```
ATATTTGAAAGCTGTGTCTGTAAACTGATGGCTAACAAAACTAGGATTTTGGTCACTTC
TAAAATGGAACATTTAAAGAAAGCTGACAAAATATTAATTTTGAATGAAGGTAGCAGCT
ATTTTTATGGGACATTTTCAGAACTCCAAAATCTACAGCCAGACTTTAGCTCAAAACTC
ATGGGATGTGATTCTTTCGACCAATTTAGTGCAGAAAGAAGAAATTCAATCCTAACTGA
GACCTTACACCGTTTCTCATTAGAAGGAGATGCTCCTGTCTCCTGGACAGAAACCAATC
TTTTAAACAGACTGGAGAGTTTGGGGAAAAAAGGAAGAATTCTATTCTCAATCCAATCA
ACTCTATACGAAAATTTTCCATTGTGCAAAAGACTCCCTTACAAATGAATGGCATCGAA
GAGGATTCTGATGAGCCTTTAGAGAGAAGGCTGTCCTTAGTACCAGATTCTGAGCAGGG
AGAGGCGATACTGCCTCGCATCAGCGTGATCAGCACTGGCCCCACGCTTCAGGCACGAA
GGAGGCAGTCTGTCCTGAACCTGATGACACACTCAGTTAACCAAGGTCAGAACATTCAC
CGAAAGACAACAGCATCCACACGAAAAGTGTCACTGGCCCCTCAGGCAAACTTGACTGA
ACTGGATATATATTCAAGAAGGTTATCTCAAGAAACTGGCTTGGAAATAAGTGAAGAAA
TTAACGAAGAAGACTTAAAGG
```

>orangutan_CFTR

```
ATATCTTAAAGCTGTGTCTGTAAACTGATGGCTAACAAAACTAGGATTTTGGTCACTTC
TAAAATGGAACATTTAAAGAAAGCTGACAAAATTTTAATTTTACATGAAGGTAGCAGCT
ATTTTTATGGGACATTTTCAGAACTCCAAAATCTACGGCCAGACTTTAGCTCAAAACTC
ATGGGATGTGATTCTTTCGACCAATTTAGTGCAGAAAGAAGAAATTCAATCCTAACTGA
GACTTTACGCCGTTTCTCATTAGAAGGAGATGCTCCTGTCTCCTGGACAGAAACCAACC
TTTTAAACAGACTGGAGAGTTTGGGGAAAAAAGGAAGAATTCTATTCTCAATCCAATCA
ACTCTATACGAAAATTTTCCATTGTACAAAAGACTCCCTTACAAATGAATGGCATCGAA
GAGGATTCTGATGAGCCTTTCGAGAGAAGGGTGTCCTTAGTTCCAGATTCTGAGCAGGG
AGAGGCGATACTGCCTCGCATCAGCGTGATCAGCACTGGCCCCATGCTTCAGGCACGAA
GGAGGCAGTCTGTTCTGAACCTGATGACACAGTCAGTTAACCAAGGTCAGAACATTCAC
CGAAAGACAACAGCATCCACACGAAAAGTGTCACTGGCCCCTCAGGCAAACTTGACTGA
ATTGGATATATATTCAAGAAGGTTATCTCAAGAAACTGGCTTGGAAATAAGTGAAGAAA
TTAATGAAGAAGACTTAAAGG
```

>chimpanzee_CFTR

```
ATATTTGAAAGCTGTGTCTGTAAACTGATGGCTAACAAAACTAGGATTTTGGTCACTTC
TAAAATGGAACATTTAAAGAAAGCTGACAAAATATTAATTTTGCATGAAGGTAGCAGCT
ATTTTTATGGGACATTTTCAGAACTCCAAAATCTACGGCCAGACTTTAGCTCAAAACTC
ATGGGATGTGATTCTTTCGACCAATTTAGTGCAGAAAGAAGAAATTCAATCCTAACTGA
GACCTTACGCCGTTTCTCATTAGAAGGAGATGCTCCTGTCTCCTGGACAGAAACCAATC
TTTTAAACAGACTGGAGAGTTTGGGGAAAAAAGGAAGAATTCTATTCTCAATCCAATCA
ACTCTATACGAAAATTTTCCATTGTGCAAAAGACTCCCTTACAAATGAATGGCATCGAA
GAGGATTCTGATGAGCCTTTAGAGAGAAGGCTGTCCTTAGTACCAGATTCTGAGCAGGG
AGAGGCGATACTGCCTCGCATCAGCGTGATCAGCACTGGCCCCACGCTTCAGGCACGAA
GGAGGCAGTCTGTTCTGAACCTGATGACACACTCAGTTAACCAAGGTCAGAACATTCAC
CGAAAGACAACAGCATCCACACGAAAAGTGTCACTGGCCCCTCAGGCAAACTTGACTGA
ACTGGATATATATTCAAGAAGGTTATCTCAAGAAACTGGCTTGGAAATAAGTGAAGAAA
TTAACGAAGAAGACTTAAAGG
```

>gorilla_CFTR

ATATCTTAAAGCTGTGTCTGTAAACTGATGGCTAACAAAACTAGGATTTTGGTCACTTC
TAAAATGGAACATTTAAAGAAAGCTGACAAAATATTAATTTTGCATGAAGGTAGCAGCT
ATTTTTATGGGACATTTTCAGAACTCCAAAATCTACGGCCAGACTTTAGCTCAAAACTC
ATGGGATGTGATTCTTTCGACCAATTTAGTGCAGAAAGAAGAAATTCAATCCTAACTGA
GACCTTACGCCGTTTCTCATTAGAAGGAGATGCTCCTGTCTCCTGGACAGAAACCAATC
TTTTAAACAGACTGGAGAGTTTGGGGAAAAAAGGAAGAATTCTATTCTCAATCCAATCA
ACTCTATACGAAAATTTTCCATTGTACAAAAGACTCCCTTACAAATGAATGGCATCGAA
GAGGATTCTGATGAGCCTTTAGAGAGAAGGCTGTCCTTAGTACCAGATTCTGAGCAGGG
AGAGGCGATACTGCCTCGCATCAGCGTGATCAGCACTGGCCCCACGCTTCAGGCACGAA
GGAGGCAGTCTGTTCTGAACCTGATGACACACTCAGTTAACCAAGGTCAGAACATTCAC
CGAAAGACAACAGCATCCACACGAAAAGTGTCACTGGCCCCTCAGGCAAACTTGACTGA
ACTGGATATATATTCAAGAAGGTTATCTCAAGAAACTGGCTTGGAAATAAGTGAAGAAA
TTAACGAAGAAGACTTAAAGG

Gene 2

Here is a partial DNA sequence from humans, pig, rabbit, and sheep for the Bone Morphogenetic Protein 7 gene (BMP7). Bone Morphogenetic Proteins represent signals found in the body that help induce bone growth.

There should be 196 bases/letters for each sequence.

>human_BMP7

AGAACCGCTCCAAGACGCCCAAGAACCAGGAAGCCCTGCGGATGGCCAACGTGGCAGAG
AACAGCAGCAGCGACCAGAGGCAGGCCTGTAAGAAGCACGAGCTGTATGTCAGCTTCCG
AGACCTGGGCTGGCAGGACTGGATCATCGCGCCTGAAGGCTACGCCGCCTACTACTGTG
AGGGGGAGTGTGCCTTCCC

>pig_BMP7

AGAACCGCTCCAAGACGCCCAAGAACCAGGAAGCCCTGCGGGTGGCCAACGTCGCAGAG
AACAGCAGCAGTGACCAGCGGCAGGCCTGTAAGAAGCATGAGCTCTACGTCAGCTTCCG
GGACCTGGGCTGGCAAGACTGGATCATCGCGCCCGAAGGCTATGCCGCCTACTACTGCG
AGGGGGAGTGCGCCTTCCC

>rabbit_BMP7

AGAACCGCTCCAAGGCACCCAAGAACCAAGAGGCGCTGCGAGTGGCCAACGTGGCAGAA
AACAGCAGCAGTGACCAGCGGCAGGCGTGCAAGAAACACGAACTGTACGTCAGCTTCCG
CGACCTGGGCTGGCAGGATTGGATCATTGCCCCGGAAGGCTACGCCGCCTACTACTGCG
AGGGAGAGTGCGCCTTCCC

>sheep_BMP7

AGAATCGCTCCAAGGCGCCCAAGAACCAAGAAGCCCTGCGGGTGGCCAACGTCGCAGAA
AACAGCAGCAGTGACCAGAGGCAGGCATGTAAGAAGCACGAGCTATACGTCAGCTTCCG
GGACCTGGGCTGGCAGGATTGGATCATCGCCACCCGAAGGCTATGCCGCCTACTACTGCG
AGGGGGAGTGCGCCTTCCC

Gene 3

Here is a partial DNA sequence from humans, cow, dog, and horse for Leptin (LEP), a signal found in the body that tells your brain how much fat you have stored away. Leptin may help regulate how hungry you feel.

There should be 426 bases/letters for each sequence.

>human_LEPTIN

```
TGTGGCTTTGGCCCTATCTTTTCTATGTCCAAGCTGTGCCCATCCAAAAAGTCCAAGAT
GACACCAAAACCCTCATCAAGACAATTGTCACCAGGATCAATGACATTTCACACACGCA
GTCAGTCTCCTCCAAACAGAAAGTCACCGGTTTGGACTTCATTCCTGGGCTCCACCCCA
TCCTGACCTTATCCAAGATGGACCAGACACTGGCAGTCTACCAACAGATCCTCACCAGT
ATGCCTTCCAGAAACGTGATCCAAATATCCAACGACCTGGAGAACCTCCGGGATCTTCT
TCACGTGCTGGCCTTCTCTAAGAGCTGCCACTTGCCCTGGGCCAGTGGCCTGGAGACCT
TGGACAGCCTGGGGGGGTGTCCTGGAAGCTTCAGGCTACTCCACAGAGGTGGTGGCCCTG
AGCAGGCTGCAGG
```

>cow_LEPTIN

```
TGTGGCTTTGGCCCTATCTGTCTTACGTGGAGGCTGTGCCCATCCGCAAGGTCCAGGAT
GACACCAAAACCCTCATTAAGACAATTGTCACCAGGATCAATGACATCTCACACACGCA
GTCCGTCTCCTCCAAACAGAGGGTCACTGGTTTGGACTTCATCCCTGGGCTCCACCCTC
TCCTGAGTTTGTCCAAGATGGACCAGACATTGGCGATCTACCAACAGATCCTCACCAGT
CTGCCTTCCAGAAATGTGGTCCAAATATCCAATGACCTGGAGAACCTCCGGGACCTTCT
CCACCTGCTGGCCGCCTCCAAGAGCTGCCCCTTGCCGCAGGTCAGGGCCCTGGAGAGCT
TGGAGAGCTTGGGCGTTGTCCTGGAAGCTTCCCTCTACTCCACCGAGGTGGTGGCCCTG
AGCCGGCTGCAGG
```

>dog_LEPTIN

```
TGTGGCTCTGGCCCTATCTGTCCTGTGTTGAAGCTGTGCCAATCCGAAAAGTCCAGGAC
GACACCAAACCCTCATCAAGACGATTGTCGCCAGGATCAATGACATTTCACACACTCA
GTCTGTCTCCTCCCAACAGAGGGTCGCTGGTCTGGACTTCATTCCTGGGCTCCAACCAG
TCCTGAGTTTGTCCAGGATGGGCCAGACGTTGGCCATATACCAACAGATCCTCAACAGT
CTGCATTCCAGAAATGTGGTCCAAATATCTAATGACCTGGAGAACCTCCGGGACCTTCT
CCACCTGCTGGCCTCCTCCAAGAGCTGCCCCTTGCCCCGGGCCAGGGGCCTGGAGACCT
TTGAGAGCGTGGGCGGCGTCCTGGAAGCCTCACTCTACTCCACAGAAGTGGTGGCTCTG
AACAGACTGCAGG
```

>horse_LEPTIN

```
TGTGGCTTTGGCCCTATCTGTTCTTCATTGAAGCTGTGCCCATCCGAAAAGTCCAGGAT
GACACCAAAACCCTCATCAAGACGATTGTCACCAGGATCAATGACATTTCACACACGCA
GTCAGTCTCCTCCAAACAGAGGGTCACTGGTTTGGACTTCATTCCTGGGCTTCACCCTG
TCCTGAGTTTGTCCAAGATGGACCAGACATTGGCAATCTACCAACAGATCCTTACAAGT
CTGCCTTCCAGAAATGTGATCCAGATATCTAATGACCTGGAGAACCTCCGGGACCTTCT
CCACCTGCTGGCCTCCTCCAAGAGTTGCCCCTTGCCCCAGGCCAGGGGTCTGGAGACCT
TGGCGAGCCTGGGCGGTGTCCTGGAAGCTTCACTCTACTCCACAGAGGTGGTAGCCCTG
AGCAGGCTGCAGG
```

The Tree of Life

Gene 4

Here is a partial DNA sequence from humans, mouse, and rat for Opsin1 (OPS1MW) Opsins are involved in providing color vision in the eye. Changes in the function of an opsin protein can lead to color-blindness.

There should be 776 bases/letters for each sequence.

>human_OPSIN

```
CCCTTCGAAGGCCCGAATTACCACATCGCTCCCAGATGGGTGTACCACCTCACCAGTGT
CTGGATGATCTTTGTGGTCATTGCATCCGTTTTCACAAATGGGCTTGTGCTGGCGGCCA
CCATGAAGTTCAAGAAGCTGCGCCACCCGCTGAACTGGATCCTGGTGAACCTGGCGGTC
GCTGACCTGGCAGAGACCGTCATCGCCAGCACTATCAGCGTTGTGAACCAGGTCTATGG
CTACTTCGTGCTGGGCCACCCTATGTGTGTCCTGGAGGGCTACACCGTCTCCCTGTGTG
GGATCACAGGTCTCTGGTCTCTGGCCATCATTTCCTGGGAGAGATGGGATGGTGGTCTGC
AAGCCCTTTGGCAATGTGAGATTTGATGCCAAGCTGGCCATCGTGGGCATTGCCTTCTC
CTGGATCTGGGCTGCTGTGTGGACAGCCCGCCCATCTTTGGTTGGAGCAGGTACTGGC
CCCACGGCCTGAAGACTTCATGCGGCCCAGACGTGTTCAGCGGCAGCTCGTACCCCGGG
GTGCAGTCTTACATGATTGTCCTCATGGTCACCTGCTGCATCACCCCACTCAGCATCAT
CGTGCTCTGCTACCTCCAAGTGTGGCTGGCCATCCGAGCGGTGGCAAAGCAGCAGAAAG
AGTCTGAATCCACCCAGAAGGCAGAGAAGGAAGTGACGCGCATGGTGGTGGTGATGGTC
CTGGCATTCTGCTTCTGCTGGGGACCATACGCCTTCTTCGCATGCTTTGCTGCTGCCAA
CCCTGGCTA
```

>mouse_OPSIN

```
CCCTTTGAAGGCCCCAATTATCACATTGCTCCCAGGTGGGTGTACCACCTCACCAGCAC
CTGGATGATTCTTGTGGTCGTTGCATCTGTCTTCACTAATGGACTTGTGCTGGCAGCCA
CCATGAGATTCAAGAAGCTGCGCCATCCACTGAACTGGATTCTGGTGAACTTGGCAGTT
GCTGACCTAGCAGAGACCATTATTGCCAGCACTATCAGTGTTGTGAACCAAATCTATGG
CTACTTCGTTCTGGGACACCCTCTGTGTGTCATTGAAGGCTACATTGTCTCATTGTGTG
GAATCACAGGCCTCTGGTCCCTGGCCATCATTTCCTGGGAGAGATGGCTGGTGGTCTGC
AAGCCCTTTGGCAATGTGAGATTTGATGCTAAGCTGGCCACTGTGGGAATCGTCTTCTC
CTGGGTCTGGGCTGCTATATGGACGGCCCCACCAATCTTTGGTTGGAGCAGGTACTGGC
CTTATGGCCTGAAGACATCCTGTGGCCCAGACGTGTTCAGCGGTACCTCGTACCCCGGG
GTTCAGTCTTATATGATGGTCCTCATGGTCACGTGCTGCATCTTCCCACTCAGCATCAT
CGTGCTCTGCTACCTCCAAGTGTGGCTGGCCATCCGAGCAGTGGCAAAGCAACAGAAAG
AATCTGAGTCCACTCAGAAGGCCGAGAAGGAGGTGACACGCATGGTGGTGGTGATGGTC
TTCGCATACTGCCTCTGCTGGGGACCCTATACTTTCTTTGCATGCTTTGCTACTGCCCA
CCCTGGCTA
```

>rat_OPSIN

```
CCCTTTGAAGGTCCCAATTATCACATTGCTCCAAGGTGGGTGTACCACCTCACCAGCAC
CTGGATGATTCTTGTGGTCATTGCATCTGTCTTCACAAATGGACTCGTGCTGGCAGCCA
CCATGAGGTTCAAGAAGCTGCGTCATCCTCTGAACTGGATTCTAGTGAACTTGGCAGTT
GCTGACCTAGCAGAGACCATTATTGCCAGCACTATCAGTGTTGTGAACCAAATCTATGG
CTACTTTGTGCTGGGCCACCCTCTGTGTGTCATAGAAGGCTACATTGTCTCACTATGTG
GGATCACAGGCCTCTGGTCCTTGGCCATCATTTCCTGGGAGAGATGGCTGGTGGTCTGC
AAGCCCTTTGGCAATGTGAGATTTGATGCTAAACTGGCCACTGTGGGAATCGTCTTCTC
CTGGGTCTGGGCTGCTGTATGGACGGCCCCACCAATCTTTGGTTGGAGCAGGTACTGGC
CTTATGGCCTGAAGACATCGTGGTCCAGACGTGTTCAGCGGTACCTCGTATCCTGGG
GTTCAGTCTTATATGATGGTCCTCATGGTCACGTGCTGCATCTTCCCACTCAGCATCAT
CGTGCTCTGCTACCTCCAAGTGTGGCTGGCCATCCGAGCAGTGGCAAAGCAACAGAAAG
AATCTGAGTCCACCCAGAAGGCTGAGAAGGAGGTGACACGCATGGTGGTGGTGATGGTC
TTCGCATACTGCCTCTGCTGGGGGCCCTATACTTTCTTTGCATGCTTTGCTACTGCCCA
TCCTGGCTA
```

You can also use a multiple sequence alignment program (like T-Coffee or CLUSTAL W) instead of BLAST to do a "multiple sequence alignment," comparing sequences from multiple species all at one time. Your input file should be a list of FASTA formatted sequences representing the same gene in different organisms (the same format as the genes above).

Here are two multiple sequence (DNA/RNA/ protein) alignment tools:

T-coffee:

→ http://tcoffee.crg.cat/apps/tcoffee/ index.html

has useful colored output (click "html" under Scores column of results page). Note that sequences must be in FASTA format (see help file).

ClustalW:

→ www.clustal.org/clustal2/

Don't worry about all the optional parameters – simply paste in all of your sequences in FASTA format.

Authors/Source

Science Buddies Staff. (2017, July 28). The Tree of Life – I (basic). Retrieved from

→ https://www.sciencebuddies.org/science-fair-projects/project-ideas/Genom_p004/genetics-genomics/tree-of-life-basic