TEACHER SOLUTIONS
You've Caught Them All...Now What?
Using Pokémon to Teach Bioinformatics

BEFORE THE WORKSHOP:

1. Give students the pre-assessment
2. Load **Pokemon_sequence** files onto computers that will be used

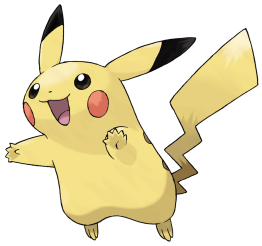
Timeline: 45-60 min class period

Activity	Time
Pre-class Preparation	1.5 hours
Pre-assessment	15 min
Part I: Morphology Matters	7-10 min
Part II: Bioinformatics is a BLAST	10-12 min
Part III: Aligning Sequences – hands on or hands off?	7-10 min
Part IV: Computing Relationships	10-12 min
Part V: Tree of Life	10-12 min
Post-assessment	15 min

PART I. MORPHOLOGY MATTERS

The physical characteristics, or morphological traits, of organisms give us clues about how closely related they are.

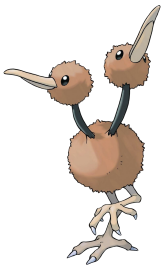
Q: How are the seven Pokémon shown below related to one another?
Fill in the table for each organism. If the organism has a trait, enter 1, if it does not have that trait, enter 0.



Pikachu

Vileplume

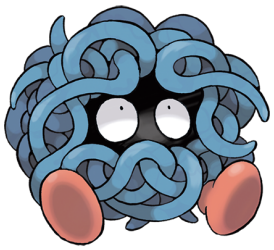
Poliwag



Doduo

Oddish

Exeggutor



Tangela

	Leaves/Flowers	Mouth present	Four appendages	Multiple heads/faces
Pikachu	0	1	1	0
Vileplume	1	1	1	0
Poliwag	0	1	0	0
Doduo	0	1	0	1
Oddish	1	1	0	0
Exeggutor	1	1	0	1
Tangela	0	0	0	0

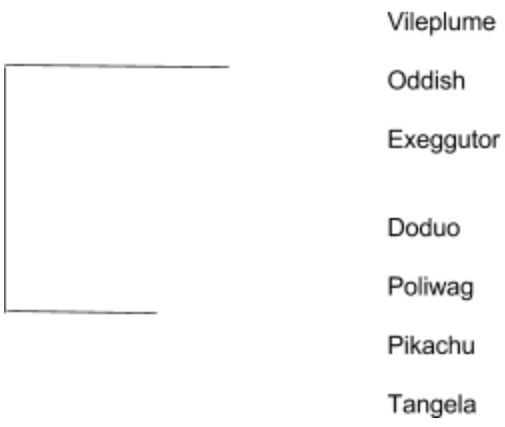
Together, we will draw a tree based on this information that will serve as our hypothesis for the relationships - write it down here:

Teacher Information:
****Key point & transition: highlight that this is a hypothesis that we can test using additional information, like DNA sequences…**

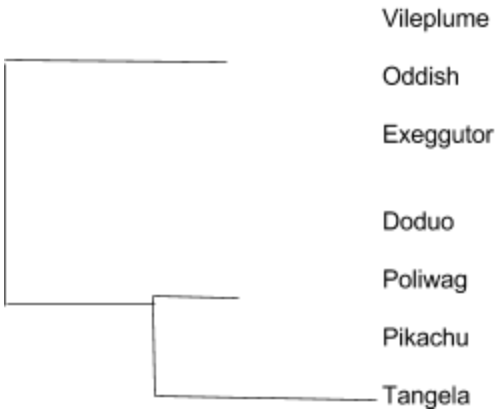
We recommend drawing this tree on the board and leaving it up for the duration of the lesson. Emphasize that this tree is a hypothesis based on morphological data that we will test using molecular data.

Tree to draw: using each trait from left to right, group the individuals. For example, for the first trait, you will end up with two groups, organisms with leaves/flowers (Vileplume, Oddish, Exeggutor) and organisms without leaves/flowers (Pikachu, Poliwag, Doduo, Tangela). For the second trait, you will end up with three groups: organisms with leaves/flowers that have a mouth, organisms without leaves/flowers that have a mouth, and organisms without leaves/flowers that do not have a mouth (Tangela). Below, you will find how the tree should look as each trait is added.

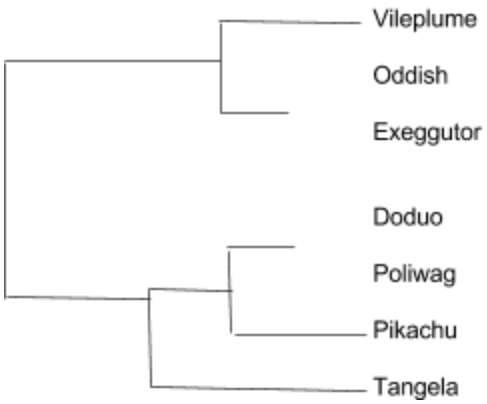
1. Flowers/leaves or not



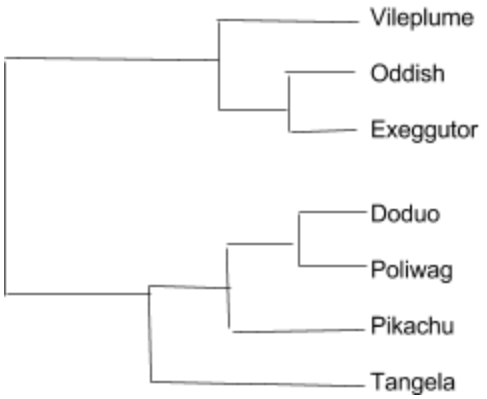
2. Mouth or not



3. Four appendages or not



4. Multiple heads/faces or not



PART II. BIOINFORMATICS IS A BLAST

Teacher information: **Key point to make: the BLAST database is an enormous resource compiled by scientists from around the world that can be used to answer lots of different types of questions.

GOAL: Identify sequences from close relatives of your organism and what those relatives are.

Source File

1. Open the sequence file for your organism

BLAST

1. In a web browser, navigate to: https://ncbi.nlm.nih.gov

2. Click on BLAST on the right side of the screen.

3. BLAST stands for Basic Local Alignment Search Tool, and is a database of DNA, RNA, and protein sequences, which are submitted to the database by scientists from around the world working on many different organisms.

4. Since we are working on DNA, click on Nucleotide BLAST.

5. Copy (Command-C or Ctrl-C) and paste (Command-V or Ctrl-V) the DNA sequence from your organism into the box below “Enter Query Sequence”. Be sure you only include the DNA sequence and not the name of your organism!

6. Make sure the correct database is selected – it should be the Nucleotide collection (nr/nt). If it is not selected, click on the dropdown menu and select this database, otherwise do not change any other settings on the page.

7. Scroll down to the bottom of the page and click BLAST

8. Wait for the next page to load – when it does, it should look similar to the example. It will not be exactly the same because you are looking at a different gene from a different organism.

9. Scroll down. Under the heading “Graphic Summary”, you will see a graphical depiction of your sequence, or *query*, at the top in blue. The red and pink lines represent other sequences, or *hits*, from the BLAST database that most closely resemble the sequence of your organism. Some of these hits may match only part of the sequence you submitted, and some may match the whole sequence.

10. Scroll down. Under the heading “Descriptions” find the list of “Sequences producing significant alignments.” These are the links to the same hits that were represented graphically above – do not click on the links! The description of the sequence usually begins with the genus and species of the organism it came from, for example, in *Homo sapiens*, *Homo* is the genus and *sapiens* is the species.

11. Check the boxes next to the first 10 *different* genus names. If your list contains only a few genus names, click on as many as you can find, and then select different species within the same genus. It is okay if you don’t know what they are yet.

12. Make sure the number of checked/selected boxes is 10, as in the picture.

13. Click on download. Select FASTA (aligned sequences), and click Continue. This will download a file of the 10 selected DNA sequences to your computer – note where the file will be saved. The file will be called ‘seqdump’.

14. **Do not open the file in Microsoft Word**. If you are on a Mac, open the ‘seqdump’ file with TextEdit (pre-installed in your applications folder). If you are on a PC, open the file with WordPad. Look over the sequence, but do not type/delete anything. We will use this file as the starting point for the next step of our project.

15. If you have extra time, try googling the names of the organisms in your list to learn more about them.

PART III: ALIGNING SEQUENCES - HANDS ON OR HANDS OFF?

Teacher information: **Key point to bring in from prior knowledge: differences in the sequences including "gaps" arise from mutations in DNA replication process.

Two versions of this activity have been provided:

One version includes the table with Gene 1 already filled out across the top. Students will then align the rest of the genes to what they have for Gene 1. This activity will likely result in most students having the same solution. We recommend this version when there is a large time constraint.

Another version contains a blank table. We recommend this version when more time is available. Starting with a blank table allows students to more fully experience the work that needs to be done by an alignment software and why we must rely on computational tools for help. This version will also result in a wider variety of solutions, which is also realistic when comparing different alignment softwares. Students may get frustrated here. That's good! It illustrates the need for a better method - using computers.

It is important to emphasize that there may not be one "correct" answer and that is okay - it is something real scientists still struggle with. The table below shows two of many possible solutions that students may come up with.

In order for us to build a phylogeny, we must first align our sequences.

Consider the following DNA sequences:

>Gene 1
AATGCTGACC
>Gene 2
AATGCTACC
>Gene 3
ACTGCTGACC
>Gene 4
ATGCTGC
>Gene 5
TTGTGACC

Align these sequences in the table below. As with the morphology task, you may approach this by thinking of the presence or absence of a trait. However, in this case, the trait is a nucleotide. Start with the first nucleotide in ‘Gene 1’. Where there is information present (i.e. there is a matching nucleotide), write the letter. Where there is missing information (i.e. a missing nucleotide), fill in the space with a gap. Gaps are represented by a hyphen (-). **There is no single correct answer.** Multiple solutions are possible, so it may not be necessary to fill in all of the spaces given. Similarly, if you find you need extra spaces, you may add some to the end of the table.

Gene 1	A	A	-	T	-	G	C	T	G	A	C	C		
Gene 2	A	A	-	T	-	G	C	T	-	A	C	C		
Gene 3	A	-	C	T	-	G	C	T	G	A	C	C		
Gene 4	A	-	-	T	-	G	C	T	G	-	C	-		
Gene 5	-	-	-	T	T	G	-	T	G	A	C	C		

Gene 1	A	A	T	G	C	T	G	A	C	C				
Gene 2	A	A	T	G	C	T	-	A	C	C				
Gene 3	A	C	T	G	C	T	G	A	C	C				
Gene 4	A	-	T	G	C	T	G	-	G	C				
Gene 5	-	T	T	G	-	T	G	A	C	C				

Based on what you know about DNA, what do you think causes gaps in sequence alignments?

Short Answer: Mutations.

Expanded Answer: During DNA replication, there are mistakes made called ‘mutations.’ Some mutations might be point mutations, where the character (in this case, a specific nucleotide) gets changed. Other mutations are indels, where

nucleotides can be added to or removed from the original sequence. These mutation-caused differences between sequences are often represented as gaps.

Why do you think aligning sequences is important before creating a phylogeny?

Answer: Alignments can provide a visual of how similar or different sequences are, which will correspond to how closely or distantly the sequences are related.

Additionally, consider the sequences you retrieved from BLAST. Would you be able to align them by hand? How can we make this task easier for ourselves?

Answer: No (or yes, but it could take years). The easiest way to align sequences is with a computer.

PART IV: COMPUTING RELATIONSHIPS

GOAL:

MAFFT

1. In a new browser tab, navigate to http://mafft.cbrc.jp/alignment/server/. MAFFT is a program that will help us build a phylogenetic tree from the close relatives of our species.

2. In the seqdump file, select all of the text (Command-A or Ctrl-A), and copy.

3. In the MAFFT window, paste the sequences.

4. Scroll down. Note that there are many different settings and parameters that can be changed, but do not change any for now.

5. Click on the submit button part way down the page.

6. Your results will appear on the screen. The first step that MAFFT has done is to align the DNA sequences – this is a complex mathematical process because the sequences will not be identical. This is because errors in DNA replication lead to changes in the base pair itself and can also lead to deletions or insertions of DNA – over long periods of time, these mutations accumulate and can make it difficult to determine how two sequences match up.

7. Scroll down. One way you can visualize the alignment is in this way. Hyphens (-) indicate ‘gaps’ in the alignment, which can come from missing data or deletions in the DNA sequence itself.

8. Scroll back up and click on View.

9. Click on Start MSAViewer in this window.

10. Look at the alignment – scroll over using the scroll bar at the top. Each base pair (A, G, T, or C) is highlighted in a different color, making it easy to see where there are differences between sequences.

11. Now, we will make a phylogenetic tree based on these DNA sequences. Click on Tree at the top of your screen.

12. Under the heading ‘Method’, make sure the option “Average linkage (UPGMA)” is selected. Then, click ‘Go’

13. When the next page loads, click ‘View tree on Phylo.io’. This will open a new window, which you can make bigger, as necessary.

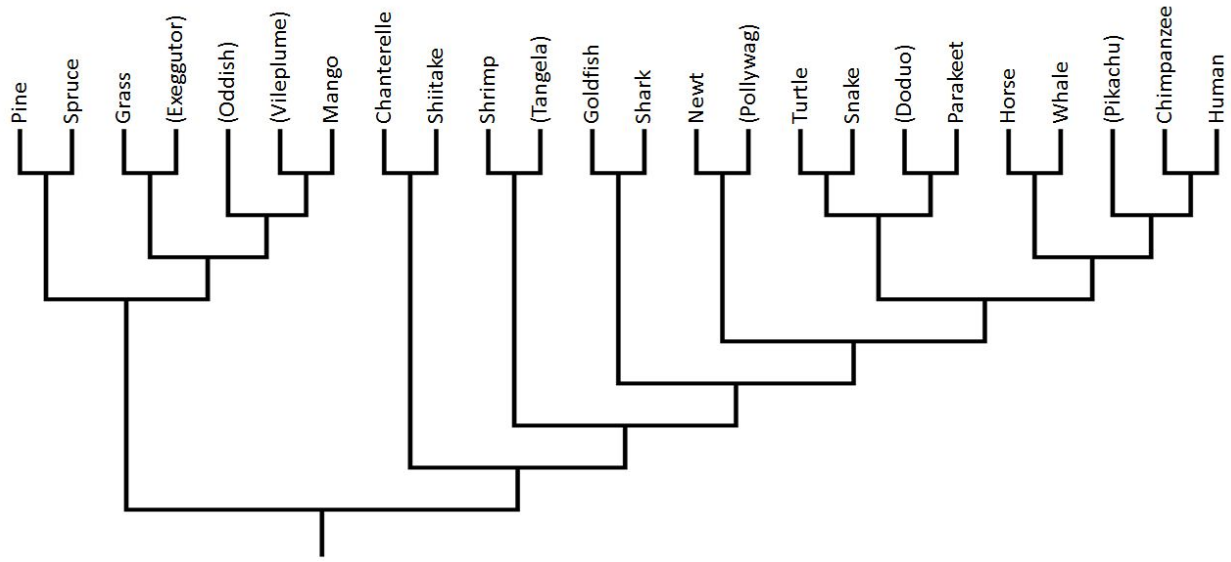
14. You have successfully built a phylogenetic tree! Now, let’s see what these things are, and where they fall under the larger tree of life…

PART V: TREE OF LIFE

1. Use google to search for the names of the organisms in your tree. List each scientific name (*Genus species*) and the major group it falls under, out of the following: Monocot, Eudicot, Mammal (See the example).

Scientific Name	**Major Group**
Homo sapiens	Mammals
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	

2. Where do these organisms fall out on the tree of life that is on the board?



Teacher information: When the tree of life is drawn out on the board, make sure to leave the names within the parentheses blank. These are the spots that the students will be filling in at the end of the lesson.

3. How do the relationships between all of the individuals compare to the relationships we hypothesized based on morphology?

   Answer: Very different. Relationships have changed in a lot of cases. For instance, the morphological tree shows Oddish and Exeggutor as being more closely related to each other than to Vileplume. However, the molecular phylogeny reveals that Oddish is actually more closely related to Vileplume, with Exeggutor falling as the outgroup.

4. Based on these exercises, do you think that molecular or morphological data are better for inferring evolutionary relationships? Why?

   Answer: Molecular because it considers more data (*hundreds* of base pairs instead of *four* morphological characteristics). At a deeper level, the molecular data also gives a better idea of character order: a DNA sequence has a set order, the first base is first; the morphological traits could be considered in any order.

5. Can you think of an example where molecular data would be better for inferring evolutionary relationships? Can you think of an example where morphological data would be better for inferring evolutionary relationships?

   Answer: Molecular can be better in instances where morphology is damaged or very difficult to interpret. It is also more powerful, providing more information with, relatively, less effort. Morphological data may be more appropriate when dealing with fossils or organisms for whom DNA extraction is difficult (very small, very hard to catch, etc.).