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

Overview of what you will do today:

1. Hypothesize evolutionary relationships based on morphology
2. Use DNA sequences from an organism to identify its closest relatives
3. Align DNA sequences from different organisms
4. Use DNA sequences from different organisms to determine how they are related to one another
5. Place the organism you are studying in the larger tree of life
6. Learn about other questions scientists can answer using these same techniques

Vocabulary:

Morphology: the form and structure of an organism or any of its parts

Genotype: all or part of the entire set of genes in a cell, an organism, or an individual

Phylogeny: the evolutionary history of a genetically related group of organisms

Bioinformatics: the collection, classification, storage, and analysis of biochemical and biological information using computers especially as applied to molecular genetics and genomics

Tree of Life: the complete evolutionary history of all of the living organisms in the world

Alignment: arrangement of DNA sequences in correct or appropriate relative positions

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.
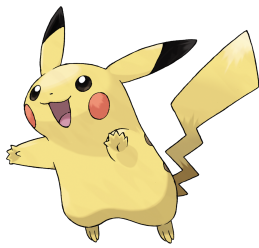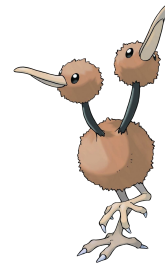

Pikachu   Vileplume   Poliwag   Doduo
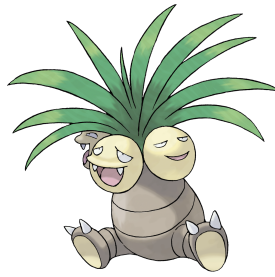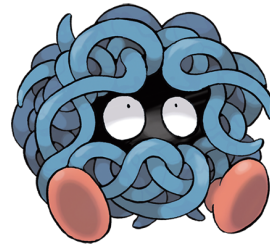
Oddish   Exeggutor   Tangela

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

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

What are some of the problems with drawing a tree based on morphological data like the one we just did?

PART II. BIOINFORMATICS IS A BLAST

GOAL: Identify the close relatives of your organism

Source File
1. Open the sequence file for your organisms

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. In the file called **Pokemon_sequence**, 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 or Notepad**. 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. In the **Pokemon_sequence** file, copy the sequence of your organism, then paste it at the beginning of the **seqdump** file. Make sure that the sequence for each organism, which begins with a '>' starts on its own line.

16. 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?

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, 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														
Gene 3														
Gene 4														
Gene 5														

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

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, 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														
Gene 2														
Gene 3														
Gene 4														
Gene 5														

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

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

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?

PART IV: COMPUTING RELATIONSHIPS

GOAL: To create a DNA sequence alignment and phylogenetic tree of this species and its closest wild relatives.

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? Go up to the board and write it! Also, draw the final tree below.

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

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

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?