Title: Branching Out: Using NCBI, BLAST, and MEGA to construct a phylogenetic tree.

Author(s):
Carla Dee Reedy and Jennie Rankin
North Marion High and Episcopal School of Jacksonville
carla.reedy@marion.k12.fl.us and rankinj@esj.org

Abstract:
In this lesson, there are two different levels available for students: a basic application (part 1) and a more difficult extension (part 2). In part 1, students will use a data set provided (FASTA formatted nucleotides) to BLAST in the NCBI database to find homologs (homologous organisms) to make a phylogenetic tree using blast. In part one, students may choose to do an extension where they find the common name and images of the organisms to add to the tree. In part 2, students will do an extension where they locate a different common gene and find the nucleotide sequence to BLAST and make a new phylogenetic tree.

Subject, Grade, Level:
This lesson is appropriate for biology, biotechnology, and environmental courses at both the middle school and high school level.

Learning objectives: Students will be able to:

- Recognize that all organisms share similar genes with other organisms
- Recognize that gene diversity is due to mutations and these mutations can be used to chart relationships
- Use a nucleotide sequence to build a phylogenetic tree
- Use the scientific name of organisms to look up common names
- Use applications to align homologs (homologous genes from other organisms)
- Use NCBI website to find a different nucleotide sequence of a gene of interest (extension)
- Complete a new phylogenetic tree using new information

Timeframe:
50 minutes part 1, 50 minutes for extension (part 2)

List of materials:

Computer access, protocol sheets

Procedure and general instructions (for instructor). REQUIRED.
Prerequisite knowledge: Students should know the basic structure of DNA and replication including nucleotide structure. The students should be fairly familiar with computers and how

to use them. Students who are not familiar may need help with some of the activities such as searching google and how to copy and paste. I usually put a few tech savvy students in charge of helping other students or team less knowledgeable students with capable ones.

Prior to the lesson, print the student protocols and put the student data into an electronic source the students can access (ie. google docs)

Introduce the topic by reminding them that genes are a sequence of nucleotides that code for specific proteins. Discuss the first two objectives to clarify why this activity is useful and how it is used to chart genetic relationships. **(Timing: 10 to 15 minutes)**

Tell the students to read and follow the directions for part 1 and instruct them as to what they need to do with their product (print, save, etc.) (**Timing: 25 minutes for the more capable, up to 35 for the less tech savvy**)

For part two, tell the students they will find two nucleotide sequences to use for their BLAST. They will look for the HOXA1 gene and then choose a gene they want to research. Again, tell them what you want them to do with their product. (**Timing: 50 minutes**)

Procedure and general instructions (for students).
NCBI BLAST Protocol

1. Obtain nucleotide sequence for Human Hemoglobin B from Data File from teacher
2. Copy all the data given (highlight then Ctrl C) this is in FASTA format
3. Open https://blast.ncbi.nlm.nih.gov/Blast.cgi
4. Click "Nucleotide Blast" box
5. On the next screen, under "Enter Sequence Query" PASTE the copied data as is
6. Enter job title as "Human Hemoglobin B"
7. Make sure "Database" reads "Nucleotide collection (nr/nt)"
8. Leave all other settings as default
9. Click "BLAST" on lower left corner
10. WAIT PATIENTLY- BLAST results will automatically pop up when ready
11. When results pop up, scroll down to the "Sequences producing significant alignments"
12. You will now click the box next to each homolog you want to make a tree with
13. Click only any one "Homo sapien" option that must include "HBB" in the name , scroll to the bottom of the list and click seven OTHER species (they must all be different/no repeats). The first word/initial is the Genus name, the second word is the species name
14. Go back to the top of the "Descriptions" label and click the "Distance tree of results"
15. When the tree pops up, on the third box with a drop down, titled "Sequence label", click the drop down and choose "Taxonomic name (if available)"
16. You should see the 8 selected homologs and then 1 labeled "unknown/query" in a tree, don't worry about this one
17. Capture the image in a format that can be saved, edited and printed later, if you do not have the means to capture the image see below

18. To capture your tree image in a pdf: above the tree, in the upper right corner, click "tools", from drop down, select "download" in the next frame, slide down to "pdf"
19. In the next window, click "Create PDF-file" then press "save" and your pdf will download at the bottom left of your screen titled as "undefined.pdf"
20. Open that pdf, right click on the tree image and save it to your desktop as "HBB Homolog Tree"
21. Then save again in an editable format and Google the common name for each of the eight genus and species on the tree. Write/type the common name near the genus and species for each
22. Then, search each organism and find an image of that organism to add to the tree as well.
23. Turn this document in to your teacher as instructed (i.e.,electronically, paper, poster, presentation)

Part 2 EXTENSION: How to find a gene/protein sequence on NCBI

1. Google "Human nucleotide sequence for HOXA1 NCBI"
2. Look for an article about HOXA1 that has "NCBI" in the title
3. Open that article and it should take you to PubMed article in NCBI
4. Look on the right side of the screen for "Table of Contents"
5. Scroll down Table of Contents until you see "NCBI Reference Sequence", click that link
6. Under Genomic, #1, Download, click "FASTA"
7. Highlight and Ctrl C to copy the nucleotide sequence from the greater than (>) symbol down to the end of the nucleotide sequence
8. Follow the steps of Part 1 from #3 to #23, make appropriate title changes
9. Repeat this process with one other human gene by doing a search on google to find a different gene.
10. Follow the directions to produce a new tree with at least 5 different species. At the end, you should have 3 different phylogenetic trees: HBB, HOXA1, and your chosen gene.

Student data - Nucleotide Sequence for HBB students need access to copy (from ">NG" to end of nucleotides) and paste

>NG_059281.1:5001-6608 Homo sapiens hemoglobin subunit beta (HBB), RefSeqGene on chromosome 11

ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGA

GGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGC

AGGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAG

ACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACCCTTAGGCTGCTGG

TGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGG

CAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGAC

AACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAACT

TCAGGGTGAGTCTATGGGACGCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATAG

GAAGGGGATAAGTAACAGGGTACAGTTTAGAATGGGAAACAGACGAATGATTGCATCAGTGTGGAAGTCT

CAGGATCGTTTTAGTTTCTTTTATTTGCTGTTCATAACAATTGTTTTCTTTTGTTTAATTCTTGCTTTCT

TTTTTTTTCTTCTCCGCAATTTTTACTATTATACTTAATGCCTTAACATTGTGTATAACAAAAGGAAATA

TCTCTGAGATACATTAAGTAACTTAAAAAAAAACTTTACACAGTCTGCCTAGTACATTACTATTTGGAAT

ATATGTGTGCTTATTTGCATATTCATAATCTCCCTACTTTATTTTCTTTTATTTTTAATTGATACATAAT

CATTATACATATTTATGGGTTAAAGTGTAATGTTTTAATATGTGTACACATATTGACCAAATCAGGGTAA

TTTTGCATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTTTATCTTATTTCTAATA

CTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATGTATCATGCCTCTTTGCACCATTCTAAAG

AATAACAGTGATAATTTCTGGGTTAAGGCAATAGCAATATCTCTGCATATAAATATTTCTGCATATAAAT

TGTAACTGATGTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTT

ATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCATGTTCATACCTCTT

ATCTTCCTCCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTC
A

CCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCACAAGTAT
CA

CTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACTAAACT

GGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTGCAA