# Lesson activities

▼ (1) Explore Databases

**Databases** 1. We will be using the National Center for Biotechnology Information (NCBI) Genbank Database in our next activity: - NCBI: https://www.ncbi.nlm.nih.gov/

1. **Interactive Question**: What are other public databases that contain genetic information or metadata?

   ○ Answer after question prompt #1 on the shared google doc.

2. How to use NCBI's Genbank database:

   ○ Introduction to NCBI Basic Local Alignment Search Tool ( `BLAST` ) `BLAST` is a program that can detect sequence similarity between a Query sequence and sequences within a database.
   ○ There are great existing tutorials and slides that have been shared to better understand the NCBI's `BLAST` alignment algorithm.

      ■ http://www.shodor.org/media/content//petascale/materials/dataIntensive/BLAST/BLAST_Intro_pdf.pdf
      ■ https://community.gep.wustl.edu/wiki/images/2/28/2011_8b_BLASTrv7_rev.pdf

▼ (2) Mystery Species Activity

**Let's get started:**

1. Open the mystery species data file: mystery_species.fasta.txt
2. Navigate to NCBI Blast
3. Choose any DNA sequence from the file and copy the sequence.
4. Paste the sequence in the "Enter Query Sequence" window.
5. Search all "Standard databases (nr etc.)" for "Highly similar sequences (megablast)"
6. Investigate the top match. Use a search engine to retrieve a photo of your mystery species!

▼ (3) Comand Line Basics

**Follow UNIX Setup instructions used in The Carpentries** https://carpentries.org/ :

- Window's users install Git for Windows
  - https://gitforwindows.org/

**Useful Links**

- Command Line for Mac and Windows
    - Lesson 1: https://swcarpentry.github.io/shell-novice/01-intro/index.html
    - Lesson 2: https://swcarpentry.github.io/shell-novice/02-filedir/index.html
    - Lesson 3: https://swcarpentry.github.io/shell-novice/03-create/index.html

▼ (4) Data Analysis & Data Visualization (usegalaxy.org)

**Sign up and Tutorial Steps**

1. Sign-up for a Galaxy account at usegalaxy.org
2. Upload sequences file to UseGalaxy.org instance. (Zika data from Next Strain.org; https://nextstrain.org/zika)
3. Run Multiple Sequence Alignment using MAFFT.
4. Visualize the alignment.
    - Integrated MSA Viewer vs IGV.
    - Interpret the results.

5. Tree building (zika virus; https://nextstrain.org/zika)
    - Generate a Phylogeny (study evolution) using IQtree.
    - Visualize the phylogeny using the integrated "Phylogenetic Tree Visualization" tool.

**Useful Links**

- Introduction to Galaxy
    - https://galaxyproject.org/learn/
    - https://training.galaxyproject.org/training-material/topics/introduction/

▼ (ADVANCED) Nextstrain.org Zika Tutorial

1. Follow the Next Strain tutorial on how to set up a visualization of Zika virus data: https://nextstrain.org/docs/tutorials/zika.
2. Follow Installation/Setup Instructions.
3. Helpful Tips:
    - Install latest version of Node.js : https://nodejs.org/en/
    - Best to Install Augur & Auspice with Conda.

4. Follow the commands to analyze the data provided.

## Note: Update to last command in Zika Tutorial

- "Explore the Results" to work with the latest version of augur.

```
augur export v2 \
  --tree results/tree.nwk \
  --metadata data/metadata.tsv \
  --node-data results/branch_lengths.json \
             results/traits.json \
             results/nt_muts.json \
             results/aa_muts.json \
  --colors config/colors.tsv \
  --lat-longs config/lat_longs.tsv \
  --auspice-config config/auspice_config.json \
  --output auspice/zika.json \
```