Bioinformatics: Through the Lens of COVID-19

ND EPSCoR Lesson Plan

Lesson Title: Bioinformatics: Through the Lens of COVID-19

Lesson Overview: In this lesson, the concept, technique and algorithm of bioinformatics will be introduced through a case study of COVID-19, which includes setting up a Linux operating system, downloading data and installing software in Linux commands, assembling SARS-Cov-2 genome.

Lesson Objectives: After these lessons, the participants are supposed to

1) Understand the concept and significance of bioinformatics
2) Have a basic understanding of *Linux* operating system
3) Have a basic understating of *Linux* commands
4) Understand de novo genome assembly

NSF Subject Classification: Biology and Computer Science

National Next Gen Standards:

HS-LS3-1 Heredity: Inheritance and Variation of Traits
Ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristic traits passed from parents to offspring.

HS-LS3-2 Heredity: Inheritance and Variation of Traits
Make and defend a claim based on evidence that inheritable genetic variations may result from (1) new genetic combinations through meiosis, (2) viable errors occurring during replication, and/or (3) mutations caused by environmental factors.

North Dakota Standards:

HS-LS3-1 Heredity: Inheritance and Variation of Traits
Ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristic traits passed from parents to offspring.

HS-LS3-2 Heredity: Inheritance and Variation of Traits
Make and defend a claim based on evidence that inheritable genetic variations may result from (1) new genetic combinations through meiosis, (2) viable errors occurring during replication, and/or (3) mutations caused by environmental factors.

Grade or Grade Band: K12: 9-12 Grades (High School)

Time Needed (estimate): 3-5 50 minute class periods

Lesson Author: Jessi Kjemhus
Teaches 7-12 science at Northwood Public School since 2015 and 9-12 science at North Border Public School prior. She graduated in 2012 with a Composite Chemistry degree from Mayville State University and then with her Master's from Valley City State University in 2015. She can be contacted at jessi.kjemhus@northwoodk12.com

Scientist/K12 Collaborator & University: Lu Liu, PhD, Department of Computer Science, North Dakota State University (email: lu.liu.2@ndsu.edu)

Scientist Bio/Research: Dr. Lu Liu holds a Ph.D. in Computer Science from The University of Texas at San Antonio (2016), M.S. and B.S. in Computer Science from Beijing University of Posts & Telecommunications (2011 and 2008 respectively). His research interests lie in the broad area of bioinformatics, data mining, machine learning and computational biology. More specifically, his current research focuses on two areas: (1) machine learning methods for modeling and analyzing biological networks; (2) data mining approaches to biomarker discovery and disease prediction via integrated analysis of heterogonous "omics" data.

Hello, my name is Lu Liu and I am an assistant professor at North Dakota State University where I teach computer science. Most of my students want to become programmers, software engineers, hardware engineers, system administrators, network administrators, database administrators, product managers, system architects, UX/UI designers, data scientists and bioinformaticians. Some want to become professors. I grew up in a rural area and we did not have access to a computer because computers were expensive. When I went to university, my elder sister picked computer science as my major because I was confused about so many majors and not sure which one to pick and she saw the opportunities in this subject. The Freshman Year of university was really a struggle for me because I found out my classmates from urban areas had learned programming in high school while I didn't even know how to start a computer. I felt programming was hard because I met a lot of unexpected problems when I started to learn programming. Most of the time, my program crashed. But I just hung in there. I asked for help from my classmates and teachers and searched for answers on the Internet. Until my Senior Year in university I realized that picking computer science was the right choice because during the study I not only mastered the skills of self-learning with the Internet resources but also entered into a field with abundant job opportunities. After I obtained my bachelor's degree, I decided to get a master's degree because I felt I was really getting into the subject. After I obtained my master's degree, I did some internships for a job and decided to pursue a doctorate degree in computer science in the USA. After I obtained my doctorate, I found a faculty job because I preferred to stay in academia. I started my own lab and taught machine learning, data mining and bioinformatics classes. My lab moved to North Dakota in 2018 and my current research lies in applying computer science technologies in analyzing big biological data such as genomic data.

Preparation/Materials

Background knowledge students must have to be successful: Basic understanding in DNA and how each species has their own genome and that has been a mission of scientists to understand the genome. For this lesson, students will need a background in computer operation.

Essential Terminology:

- Bioinformatics- interdisciplinary field (such as biology, computer science, information engineering, mathematics and statistics) that develops methods and software tools for understanding biological data.
- DNA- a self-replicating material which is present in nearly all living organisms as the main constituent of chromosomes. It is the carrier of genetic information.
- Nucleotide- a compound consisting of a nucleoside linked to a phosphate group. Nucleotides form the basic structural unit of nucleic acids such as DNA.
- Operating System- is system software that manages computer hardware, software resources, and provides common services for computer programs.
- Genome - the complete set of genes or genetic material present in a cell or organism.
- Genome assembly- aligning and merging fragments from a longer DNA sequence in order to reconstruct the original sequence

Resources:

- Ubuntu walk through
- What is Bioinformatics video
- EPSCoR Website
- Gizmo

Websites:

- https://ubuntu.com/tutorials/create-a-usb-stick-on-ubuntu#1-overview
- https://www.youtube.com/watch?v=K9MqyU298uE
- https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/
- https://www.ndepscor.ndus.edu/ndep/nature/sunday-academy/stem-module-topics/
- https://www.explorelearning.com/

Materials needed:

- USB sticks- OPTIONAL: all downloaded and set up (Lesson 2/Activity 1)
- Jigsaw puzzles for each group
- Computer
- Projector
- Bubbles or Nerf balls
- Spoons for each student
- Ping pong balls

PowerPoint – found as separate attachment

Procedure/Activities

Recommend to set up the USB, Activity 1, before hand so it is ready to go if you have a group who is not strong with computers or the time for it. It will allow students to jump right into the lesson.

Lesson 1: There are numerous options for this lesson for the teacher to find what works best for their class. Have groups made prior as teacher knows students best. Try and pair up students who are strong with computers with those who aren't.

- Opening Pre-Lesson Games (10-15 minutes)
    - Information about each games and materials needed found in resources portion of the document
- Optional: Disease Spread Lab simulation Gizmo (takes about 20 minutes)
- Slide 2
    - Hand out Pre-Lesson Activity questions to partners or groups you have assigned
        - The answer key is located after the student copy in this document
        - Talk about the answers. The activity is not graded just to help introduce the topic
- Go over nucleotide, DNA, genome, genome assembly, and bioinformatics vocab words
- Slides 3-5
- Slide 6
    - Video for slide 6

      https://www.youtube.com/watch?v=K9MqyU298uE
- Have students research and come up with a list of bioinformatics applications
- Slide 7 and 8
    - disease diagnosis, personalized treatment
    - food fermentation (pickle, beer, cheese)
    - forensic (Golden State Killer)
    - select plants with desired properties (high yield, nutrition, insect resistance, environment friendly (use less water))

Lesson 2: Setting up a Linux operating system (30 minutes) *IF TEACHER DOES NOT ALREADY DO THIS OTHERWISE JUMP TO LESSON 3

- Go over operating system vocab
- Slide 9
    - Make a list of popular software students use in their lives
- Slide 10
    - Different forms of computers: there are supercomputers, desktops, laptops, etc.
- Slide 11
    - Go through the picture on the slide about the different components
- Slides 12-14

- Slide 15 Activity 1, hand out brief Activity 1 sheet
    - This website has step by step instructions for the students to follow
    - **https://ubuntu.com/tutorials/tutorial-create-a-usb-stick-on-windows#1-overview**
- Goal is to create a bootable USB stick which will take about 30-40 minutes. This will prepare for the following lessons

Lesson 3:

- Downloading data and installing bioinformatics tools in Linux commands (45 minutes). In this lesson, we will use Linux commands to download SARS-Cov-2 sequence data and install bioinformatics tools, such as SRA toolkit.
- Slide 17
- Slide 17-21
- Slide 22 Activity 2 Sheet
- Slide 23 Extension for students who finished early
    - https://ubuntu.com/tutorials/command-line-for-beginners#1-overview

Lesson 4:

- Downloading data and installing bioinformatics tools in Linux commands (45 minutes)

  In this lesson, we will learn how to assemble SARS-Cov-2 genome.
- Slide 24 recap trial of Linux commands
- Slide 25 and 26 Activity 3
    - Had out puzzle pieces without a picture and with some missing pieces
- Slide 27
    - Look at SARS-CoV 2 genome
        - https://genome.ucsc.edu/cgi-bin/hgGateway for other organisms genome
    - Go through discussion questions
- Slide 27

Extensions for above average students:

- In lesson 3, more simple Linux commands will be given to students who wants to learn more about the Linux operating system.
    - https://ubuntu.com/tutorials/command-line-for-beginners#1-overview
- In lesson 4, participants will be instructed to download more SARS-Cov-2 sequence data and carry out a simple analysis of downloaded data. In lesson 4, participants will be instructed to assemble more SARS-Cov-2 sequence data.
- Can have students research about different diseases or pandemics in the past. What about the viruses or bacteria make them so deadly, what do we know about their genome, what is done to treat and prevent these diseases? A powerpoint can be made or just use it for class discussion.

Mediation/Support for students that need it:

- Pair with a student who has stronger background in computer science
- Could always get jump drive ready and save that process
- Frequent check-ins with students to see their progress and if they have any questiosn

Lesson Outline (for research-based lessons)

1) Observe Phenomena
2) What questions should we be able to answer?
3) Write a Hypothesis
4) Come up with a Research Plan
5) Carry out investigation
6) Revisit the Background Research
7) Construct Explanations. (TASKS-Publish/Communicate Findings)

Standards Alignment

ND Science Standard(s):

HS-LS3-1 Heredity: Inheritance and Variation of Traits

Ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristic traits passed from parents to offspring.

Performance Expectation

HS-LS3-2 Heredity: Inheritance and Variation of Traits

Make and defend a claim based on evidence that inheritable genetic variations may result from (1) new genetic combinations through meiosis, (2) viable errors occurring during replication, and/or (3) mutations caused by environmental factors.

Disciplinary Core Idea: e.g. Application of how technology is used to expand what we know about biology, in this case genome of a specific virus

North Dakota DPI Standards:

Computer Science Standards

- Computational Thinking 9-12
  - Problem Solving & Algorithms Strategies for understanding and solving problems
  - Data Creation & Analysis Data can be collected, used, and presented with computing devices or digital tools

ND ELA

- W.7 Conduct short as well as more sustained research projects to answer questions (including self-generated questions) or solve problems.
- a. Develop a research question.
- b. Narrow or broaden the inquiry when appropriate.
- c. Synthesize multiple source

Next Gen Standards:

HS-LS3-1 Heredity: Inheritance and Variation of Traits

Ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristic traits passed from parents to offspring.

Performance Expectation

HS-LS3-2 Heredity: Inheritance and Variation of Traits

Make and defend a claim based on evidence that inheritable genetic variations may result from (1) new genetic combinations through meiosis, (2) viable errors occurring during replication, and/or (3) mutations caused by environmental factors.

Cross Cutting Concepts

- Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects.

Unit Objectives

- Students will be able to carry out Linux commands to analyze SARS-Cov2 genome and understand importance of knowing a organism's or virus' genome

Assessment

- Students will be able to analyze SARS-Cov2 genome and explain why it is essential to study its genome along with other organisms.

Daily Plans and Assessments

Learning Target for each day/activity

After these lessons, the participants are supposed to

- Lesson 1: Understand the concept and significance of bioinformatics with viruses today
- Lesson 2: Master the skills of setting up a Linux operating system
- Lesson 3: Master basic linux commands
- Lesson 4: Understand de novo genome assembly

Criteria for Success/Assessment for each activity

After these lessons, the participants are supposed to

- Lesson 1: Solve questions on the worksheet about human genome and bioinformatics and understand how easy diseses can be spread
- Lesson 2: Create a bootable USB stick Hands-on experience of creating bootable USB sticks
- Lesson 3: Successfully download SARS-Cov-2 sequence data and install bioinformatics tools with Linux commands
- Lesson 4: Successfully assemble SARS-Cov-2 sequence data with Linux commands

Additional Lesson Resources / Materials

References:

Liu, L. "Bioinformatics through the lens of COVID19". 29 July, 2020.

Saskatchewan Science Centre. "Classroom Activity- Disease Transmission". 29 July, 2020. https://static1.squarespace.com/static/563a8427e4b02d05f44d829d/t/564ce8d2e4b0e4c59118f59a/1447880914071/classroom+activity%3B+science+1+-4+Heath+disease+transmission.pdf

Websites for purchasing materials

16 GB USB Stick

https://www.amazon.com/SanDisk-Flash-Cruzer-Glide-SDCZ60-016G-B35/dp/B007YX9O9O/ref=sr_1_3?dchild=1&keywords=usb+stick&qid=1596135559&sr=8-3

If need general supplies:

Nasco

https://www.enasco.com/c/Education-Supplies/Science

Flinn

https://www.flinnsci.com/

Carolina

https://www.carolina.com/lab-supplies-and-equipment/science-lab-supplies/science-lab-classroom-supplies/10300.ct

School Specialty

https://www.schoolspecialty.com/science-supplies-and-products

Amazon

www.amazon.com

Pre-Lesson Activity
Name:

Please try to finish all the questions and when you are done let the instructor know.

Multiple choice

1. DNA is made of basic building blocks, called nucleotides. How many different types of nucleotides are there in human DNA? (1 point)

A. 4 B. 5 C. 6 D. 7

2. Approximately how many nucleotides are there in a brain cell of human? (1 point)

A. 6,000,000 B. 6,000,000,000 C. 600,000,000 D. 60,000,000

2. How many chromosomes are there in a brain cell of human? (1 point)

A. 22 B. 23 C. 44 D. 46

3. Hard disk, USB drive and other storage media are used to store data. Each storage medium can only store data within its capacity, which can be measured with the following data storage units, TB, MB, KB and GB. (For example, a USB drive of 4GB) Among these four data storage units, which relationship between them is correct? (1 point)

A. 1MB < 1KB < 1GB < 1TB B. 1TB < 1GB < 1KB < 1MB

C. 1KB < 1MB < 1GB < 1TB D. 1GB < 1KB < 1MB < 1TB

Pre-Lesson Activity Answer Key
Name:

Please try to finish all the questions and when you are done let the instructor know.

Multiple choice

1. DNA is made of basic building blocks, called nucleotides. How many different types of nucleotides are there in human DNA? (1 point)

A. 4 B. 5 C. 6 D. 7

2. Approximately how many nucleotides are there in a brain cell of human? (1 point)

A. 6,000,000 B. 6,000,000,000 C. 600,000,000 D. 60,000,000

2. How many chromosomes are there in a brain cell of human? (1 point)

A. 22 B. 23 C. 44 D. 46

3. Hard disk, USB drive and other storage media are used to store data. Each storage medium can only store data within its capacity, which can be measured with the following data storage units, TB, MB, KB and GB. (For example, a USB drive of 4GB) Among these four data storage units, which relationship between them is correct? (1 point)

A. 1MB < 1KB < 1GB < 1TB B. 1TB < 1GB < 1KB < 1MB

C. 1KB < 1MB < 1GB < 1TB D. 1GB < 1KB < 1MB < 1TB

Pre-Lesson Games

These three activities show how germs are passed during outbreaks

Materials:
Spoons for each student
Ping pong balls or something similar
Bubbles or Nerf balls

Activity 1: Disease Spread through Direct Contact

In this game, students are asked to wander around the room shaking hands with various students
.
1. Select one student to secretly pretend to be infected by a disease. Instruct the student to squeeze the other participants hand while shaking hands.
2. Once your hand is squeezed, you are also infected with the disease and must squeeze others hands when you shake their hand.
3. This can be done quickly, even with a 30 second or one minute time limit. After about 30 seconds, ask the students if they had their hand squeezed, if so they have been infected by the disease.
4. To conclude, relate this game to other situations such as public places where washing hands is important or on farms with animals. Animals on a farm may be contained in close quarters and can spread disease through direct contact. This occurs even when an animal does not yet show symptoms. If the animal showed symptoms, it would be separated from the herd or flock for treatment.

Activity 2: Pathogens Spread through Indirect Contact

In this game the children race with ping pong balls on spoons to show diseases can be shared with indirect contact to the disease.

1. Divide into two or more groups. Designate a race course, perhaps around the classroom.
2. Give each child a spoon. Each group must race around the outlined course one at a time in a relay with their ball on the spoon. If the ball falls they may pick it up, but at that point they are considered “sick” and must walk slowly the rest of the course.
3. Once the student gets to the end of the outlined course, they must transfer the ball to the next team member’s spoon without directly touching the ball or they will be “sick”. If this occurs, they will become “sick” and need to walk slowly. Once each person has completed the course they are done. Depending on how many became infected the game will take different amounts of time.

This is an example of pathogens being transferred from one animal to another through indirect contact to the pathogen. The students did not have to directly touch one another in order for the pathogen to be transferred (the ping pong ball is the pathogen). Not all of the disease

carriers displayed symptoms. Diseases or pathogens can be transferred without animals or people coming into contact directly, they can be carried on equipment, surfaces, boots etc.

Activity 3: Airborne Diseases

In this activity the leader/teacher will blow bubbles at the group of students to demonstrate how some diseases can be airborne.

1. Tell the students that each time a bubble lands on them they have become infected with an airborne pathogen. This shows the students how some diseases are passed through the air.
2. A healthy living element can be added by giving each student a tissue and have them stop the bubbles from touching them using the tissues. This is a great way to encourage use of tissues when sneezing.
3. Another version of this game can be played as a simple dodge ball game using soft Nerf balls as the disease spreading pathogen. Once someone is hit with the ball they have been infected with the pathogen.

Name: _____ Date: _____

Student Exploration: Disease Spread

Vocabulary: disease, epidemic, infect, infectious disease, pathogen

Prior Knowledge Questions (Do these BEFORE using the Gizmo.)

1. Why do you think it is important to cover your mouth when you cough? _____

   _____

   _____

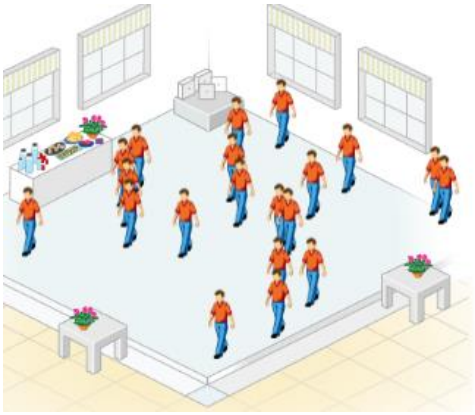
2. Why should you always wash your hands before you eat? _____

   _____

   _____

Gizmo Warm-up

When a person has a **disease**, his or her normal body functions are disrupted. Some diseases, such as diabetes and most cancers, are not spread from one person to another. But other diseases, such as the flu and strep throat, can be spread. These diseases are known as **infectious diseases**. Infectious diseases are caused by viruses, bacteria, and other agents known as **pathogens**.



In the *Disease Spread* Gizmo, you will be able to observe how various pathogens can spread through a group of people. Click **Play** ( ) and observe.

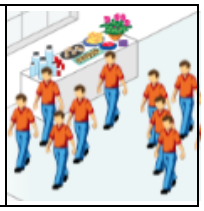
1. Describe what happened on the SIMULATION pane: _____

   _____

   _____

2. Look at the color key on the bottom right of the Gizmo. What is happening when a person changes color? _____

_____

Activity A: **Person-to-person transmission**	Get the Gizmo ready: • Click **Reset** ( ). • On the CONTROLS tab under **Active Diseases**, turn off **Foodborne** and turn on **Person to person**. • Set the **Number of people** to 5.	

Question: What factors affect how quickly a pathogen spreads from person to person?

1. Predict: Some pathogens are spread directly from one person to another. This can happen when people come into direct contact or share items, such as drinking glasses. What do you think might affect how quickly a pathogen is spread from person to person?

   _____

   _____

2. Identify: Select the SIMULATION tab on the left and the TABLE tab on the right. (You will want the table tab open to answer question C.)
   A. What does the purple person represent? _____

   B. Click **Play**, and observe the simulation for a while. What must happen for the disease to spread from one person to another? _____

   C. How long did it take to **infect** five people? _____

3. Experiment: Click **Reset**. Change the **Number of people** to 15. Click **Play**, and record how long it takes to infect five people. Repeat this four times for a total of 5 trials, and calculate the mean time. Repeat the experiment when there are 25 people and 35 people in the room.

Number of people	Time to infect five people, 5 trials (h)	Mean time (h)
15		
25		
35		

4. Interpret: Study the data you collected. What trend do you see in the data, and how would you explain it?

_____

_____

_____

5. Experiment: Not all pathogens are equally infectious. Click **Reset**. Set the **Number of people** to 20. Under **Probability of transmion**, select **Low** for **Person to person**.

On the SIMULATION tab, click **Play**. Record the time it takes to infect five people for five trials, and find the mean. Then repeat the experiment with a medium and high probability of transmission. (Note: For the "Medium" setting, move the slider half-way between the **Low** and **High** positions.)

Transmission probability	Time to infect five people, 5 trials (h)	Mean time (h)
Low		
Medium		
High		

6. Interpret: Study the data you collected in the table above. What trend do you see in the data, and how would you explain it? _____

_____

_____

7. Analyze: On the CONTROLS tab, place the **Probability of transmission** slider under **Person to person** half-way between **Low** and **High**. Select the SIMULATION and GRAPH tabs. Click **Play**.

   A. At what time did the disease spread most slowly? Most quickly? _____

   _____

   B. How could you explain this change in the rate of the disease's spread? _____

   _____

   _____

   _____

8. Apply: An **epidemic** is the rapid spread of an infectious disease. How do you think a government could try to prevent an epidemic of a dangerous person-to-person pathogen?

_____

Activity B: **Foodborne and airborne transmission**	Get the Gizmo ready: • Click **Reset**. • On the CONTROLS tab under **Active diseases**, turn off **Person to person** and turn on **Foodborne**.	

Question: How do foodborne and airborne pathogens spread?

1. Predict: How do you expect the spread of a foodborne disease to be similar to and different from the spread of a person-to-person disease? _____

   _____

   _____

2. Observe: Select the SIMULATION tab. Click **Play** and closely watch the people moving around the room.

   A. What does each person do just before becoming infected? _____

   _____

   B. How are foodborne pathogens transmitted? _____

   _____

   C. If a person in the simulation never eats or drinks anything from the buffet table, is it possible for them to become sick with the foodborne disease? Explain your answer.

   _____

   _____

3. Analyze: Select the GRAPH tab, and wait for every person to become infected.

   A. At what time did the disease spread most slowly? Most quickly? _____

   _____

   B. How could you explain this change in the rate of the disease's spread? _____

   _____

   _____

   _____

4. Compare: How does the spread of a foodborne pathogen compare to the spread of the person-to-person pathogen you studied in activity A? _____

_____

5. Predict: How would you expect the spread of an airborne disease to be similar to and different from the spread of a foodborne disease and a person-to-person disease?

_____

_____

_____

_____

6. Experiment: Run a few simulations with the airborne pathogen.

   A. What patterns do you notice in how the airborne pathogen spreads? _____

   _____

   _____

   B. How does the spread of an airborne pathogen compare to the spread of foodborne and person-to-person pathogens? _____

   _____

   _____

7. Think about it: Suppose there is an infectious disease at a party. How could doctors tell if the disease was foodborne, airborne, or transmitted person to person?

_____

_____

_____

_____

Activity 1: Creating a Bootable USB Stick

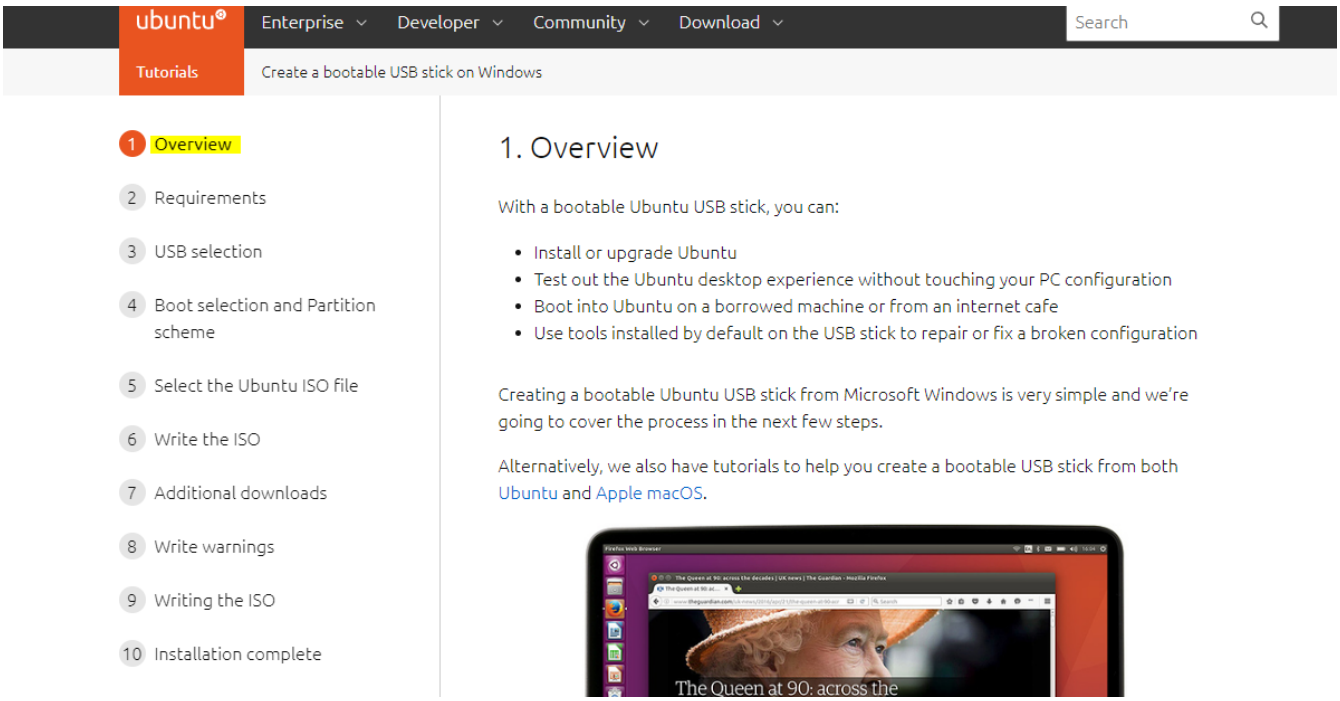
In this activity you will create a bootable USB stick to carry out further activities leading you to SARS-CoV 2 genome.

Materials:

- Flash drive
- Computer

1. Go to https://ubuntu.com/tutorials/create-a-usb-stick-on-windows#1-overview

    a. Brings you to a website as seen below. You will go through each of the left tabs in order. Be sure to read the information and steps CAREFULLY



    b. Once installation is complete (step 10) inform your teacher you are done.

Practice Linux Command Sheet

- Please type the following commands (underlined) in the Terminal. After each command, type **Enter key**. What after # are comments, which are used to explain the commands.
  - whoami # short for “Who am I?”
  - man whoami # man short for “manual”, “man whoami” means # check the manual of the command whoami. If you # want to exit form the manual, type q
  - pwd # short for “print working directory”
  - man pwd
  - ls # short for “list” directory content
  - man ls
  - ls -a # -a is an option, short for “all”
  - ls -l # -l is an option, short for “long”
  - mkdir test # short for “make directory”, to create a directory # named test
  - ls
  - man mkdir
  - cd ./test # short for ”change directory”
  - man cd
  - cd .. # short for “change directory”, to upper directory
  - rmdir ./test # short for “remove directory”
  - ls
  - man rmdir
  - wget www.ndsu.edu # download from a url
  - ls
  - wc index.html # short for “word counting”
  - man wc
  - cat index.html # show the content in the file
  - man cat
  - head index.html # show first 10 lines of the file
  - man head
  - tail index.html # show last 10 lines of the file
  - man tail
  - tar cvfz index.tar.gz index.html # create a compressed file
  - man tar
  - ls
  - rm index.html # short for “remove”
  - man rm
  - ls
  - tar zvxf index.tar.gz # extract from the compressed file

Activity 2 Downloading Data and Bioinformatics Tools

Please execute the following commands and let the instructor know when you are done:

wget https://github.com/mfq426/ND_Sunday_Academy/raw/master/simple_read1.fastq

wget https://github.com/mfq426/ND_Sunday_Academy/raw/master/simple_read2.fastq

wget https://github.com/voutcn/megahit/releases/download/v1.2.9/MEGAHIT-1.2.9-Linux-x86_64-static.tar.gz

tar zvxf MEGAHIT-1.2.9-Linux-x86_64-static.tar.gz

cd ./MEGAHIT-1.2.9-Linux-x86_64-static/bin/

./megahit –test

Once finished, and you want to learn more:please check out the following link:
https://ubuntu.com/tutorials/command-line-for-beginners#1-overview

Activity 3 De Nova Assembly

First activity:

1. This game is a simple simulation of de novo assembly. You can think each piece is a read and the whole picture is a genome. If you successfully assemble all pieces then you get a genome.
2. I have taken away edge pieces because in de novo assembly you can't tell a read is an edge or not. In a regular jigsaw puzzle, you can. So I need take them away.
3. I will not show you the whole picture beforehand because in de nova assembly the whole genome is unknown to us. You need figure it out by yourself.
4. When you are done, please let me know.

Next,

- Please execute the following commands to assemble reads
    - ./megahit -1 simple_read1.fastq -2 simple_read2.fastq -o result
    - ls
    - cd ./result
    - ls
    - wc final.contigs.fa
    - head final.contigs.fa

Questions:

What surprised you once you saw the SARS-Cov 2 genome sequence?

Why is it important for scientists to figure out SARS-Cov 2 genome sequence?

How could knowing the genome of one virus help us combat another virus?

Extension: Can have students research about different diseases or pandemics in the past. What about the viruses or bacteria make them so deadly, what do we know about their genome, what is done to treat and prevent these diseases? A powerpoint can be made or prepare to share with class to discuss.