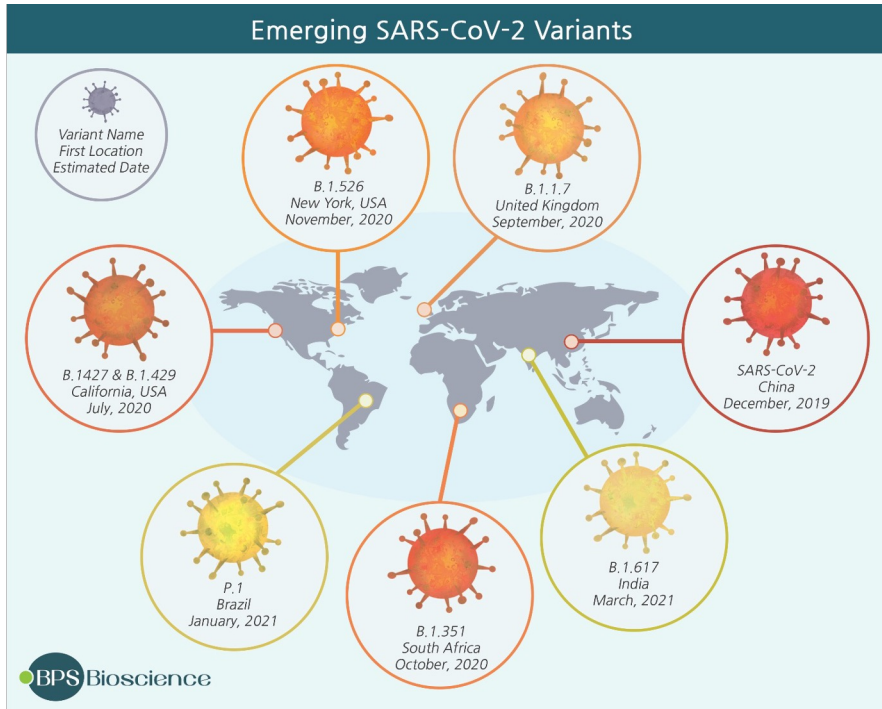


Mapping Mutations in the SARS-CoV-2 Genome



<https://bpsbioscience.com/sars-cov-2-variants>

Rachael Barry

Assistant Professor of Teaching
Department of Molecular Biology &
Biochemistry
UC Irvine

April Williams

Director
The Razavi Newman Integrative Genomics &
Bioinformatics Core
The Salk Institute for Biological Studies

Files: <https://tinyurl.com/SARSiR>
or QR code

Key Players

M115: Allied Health Microbiology

UCI IRB (Self-Exempt) #3001

Barry Lab Independent Study Students



Darren Fan



Micheal Atiya



April Williams (Salk)

Please feel free to use any of these materials, but do not post anything publicly!

Session Outline



Lab Activity Institutional Context (<5 min)



Lab Learning Objectives and Overview (5 min)



Data Visualization Exercise (~30 min)

UNIVERSITY OF CALIFORNIA



8,966
Bachelor's Degrees



1,815
Master's Degrees



393
Doctoral Degrees



46%
First-generation Students
earning bachelor's degrees



2,611
CCC Transfer Students
earning bachelor's degrees



3,636
Federal Pell Grant recipients
earning bachelor's degrees

Total degrees 11,174

Hispanic-Serving Institution (HSI) & Asian American and Native American Pacific Islander-Serving Institution (AANAPISI)

UCI Biological Sciences Context

- 1000-1200 undergraduate students (majors) per year
 - Most are pre-health (medical, dental, PA)
 - Upper Division lab courses (4th year)
 - **Microbiology**, Biochemistry, Physiology, etc.



BIOSCI
2023/2024 Academic Year

Where does this fit into AH Microbiology?

- 35-45 BS in Nursing Science majors per year
 - Allied Health Microbiology (2nd year)
- Lecture (2 x 80 minutes)
- Lab (1 x 3 hours)



Where does this fit into AH Microbiology?

- 8 groups x 5-6 per group
- Week 4, Lecture 2: Structure & Function of Viruses
- Week 5 Lab: this lab!



Where does this fit into AH Microbiology?

MINAH:

Pathogens are continuously evolving and virulence is not a static property. Understanding mechanisms that impact pathogen evolution (i.e. vertical and horizontal genetic variation, mutations, recombination, etc.) is central to limiting pathogen evolution.

Microbiology in Nursing and Allied Health (MINAH) Undergraduate Curriculum Guidelines

Lourdes Norman-McKay and the ASM MINAH Undergraduate
Curriculum Guidelines Committee
<https://doi.org/10.1128/jmbe.v19i1.1524>

These recommended curriculum guidelines for undergraduate microbiology in nursing and allied health curricula are aligned to ASM's 2012 Curriculum Guidelines for undergraduate microbiology (<https://www.asm.org/index.php/guidelines/curriculum-guidelines>) as well as the test plan for the National Council Licensure Examination for Registered Nurses published in 2016 (NCLEX-RN; https://www.ncsbn.org/RN_Test_Plan_2016_Final.pdf).

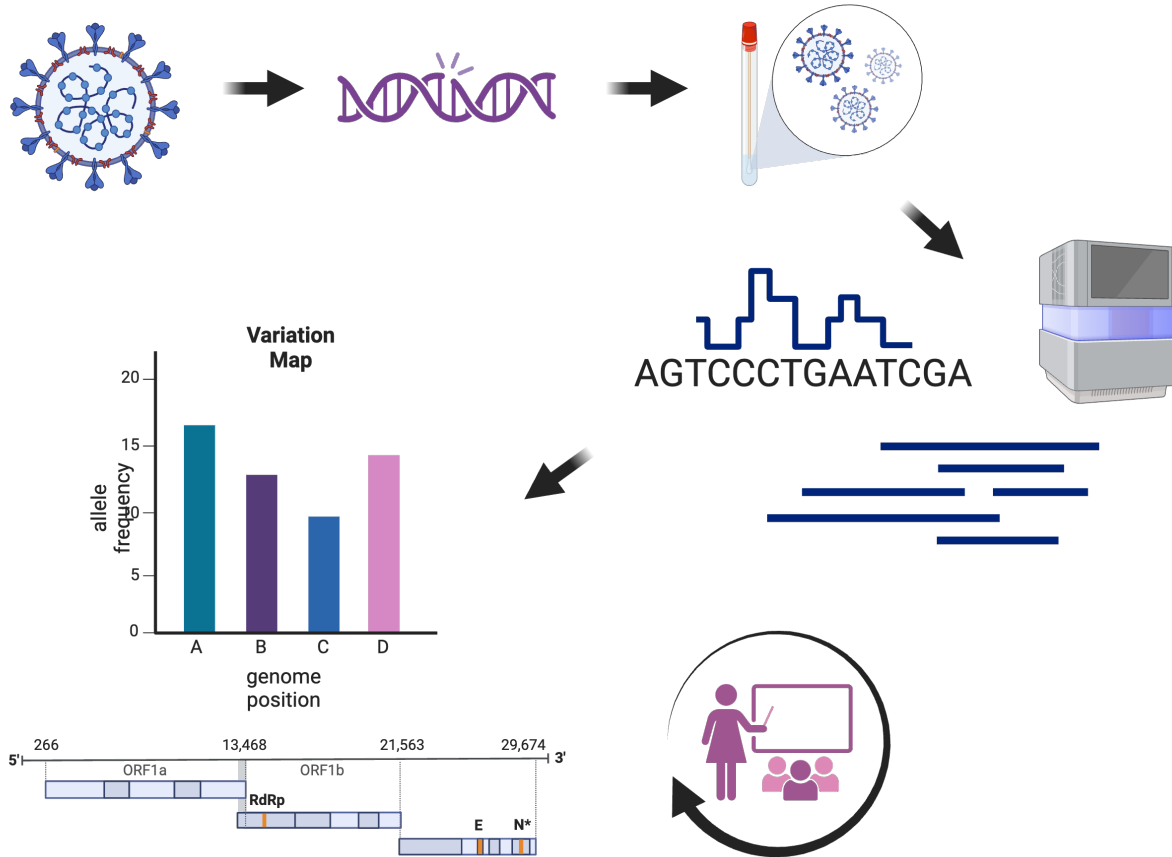
Recommended Curriculum Guidelines for Undergraduate Microbiology Education

Student Learning Objectives

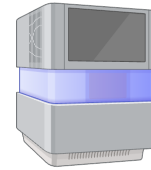
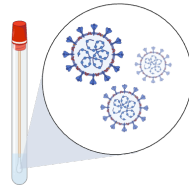
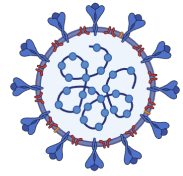
Students will be able to work collaboratively to...

- define and create examples of different types of mutations
- interpret data about SARS-CoV-2 mutations across the genome and within individual genes
- use maps of mutations to predict information about parts of genes and proteins

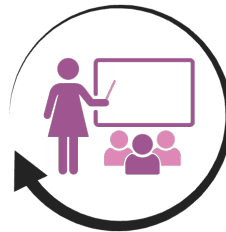
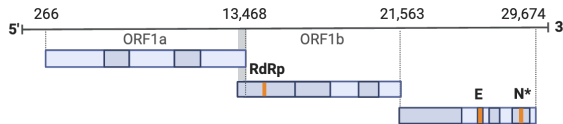
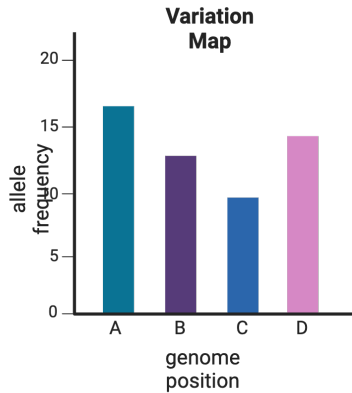
Activity Overview



Activity Overview

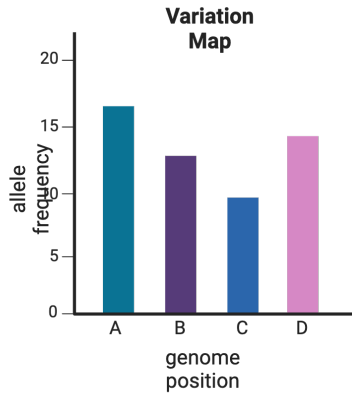
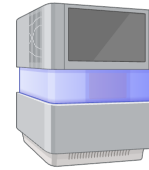
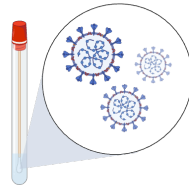
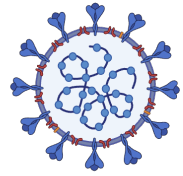


AGTCCCTGAATCGA

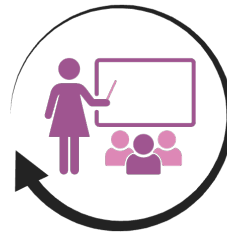


- UPenn SARS-CoV-2 variant database
- June 2020
- International data – South African, China, Europe

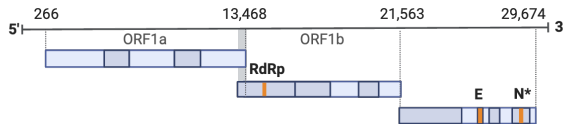
Activity Overview



AGTCCCTGAATCGA



- Use R to plot variant data by:
 - Mutation type
 - Genome position
- Class discussion



Assessment

- Students turn in pre-work which is checked for completion
 - Screenshot of Posit Cloud account sign up + project set up
- Group Google worksheet
 - 1 grade per group
 - Rubric posted in workshop materials

Basic Lesson Outline (pre-class)

- Register for Posit Cloud Account (5 minutes)
- Read/view background info (25 minutes)

 posit Cloud

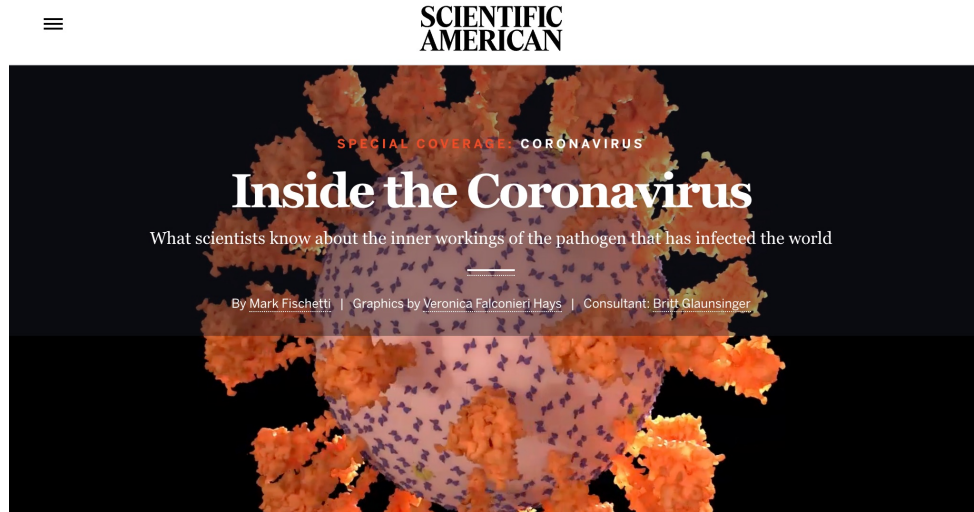
Friction free data science

Posit Cloud (formerly RStudio Cloud) lets you access Posit's powerful set of data science tools right in your browser – no installation or complex configuration required.

GET STARTED

ALREADY A USER? LOG IN

If you already have a shinyapps.io account, you can log in using your existing credentials.



Register for posit.cloud

Formerly
RStudioCloud
(also Jupyter
notebooks –
Python)



	Cloud Free	Cloud Basic	Cloud Standard	Cloud Instructor	Cloud Student
	\$0 / forever	\$25 / month	\$75 / month	\$15 / month <small>With three options available for covering student costs</small>	\$5 / month
	Learn more	Learn more	Learn more	Learn more	Learn more
SHARED SPACES	1 (with member & project limits)	Unlimited	Unlimited	Unlimited	1 (with member & project limits)
PROJECTS	25	Unlimited	Unlimited	Unlimited	25
OUTPUTS	25	50	100	100	25
COMPUTE HOURS	25 hours per month	150 hours included per month additional @ 10¢ per hour	500 hours included per month additional @ 10¢ per hour	300 hours included per month additional @ 10¢ per hour <small>Unlimited hours with a fixed price per user plans are also available</small>	75 hours included per month additional @ 10¢ per hour

FOR QUALIFYING EDUCATORS AND STUDENTS

Basic Lesson Outline (in class)

Introduction by instructor (15 minutes)

Group Discussion 1: Mutations Definitions and Examples (15 minutes)

Data Visualization 1: Whole SARS-CoV-2 Genome (15 minutes)

Group Discussion 2: Whole Genome Data (5 minutes)

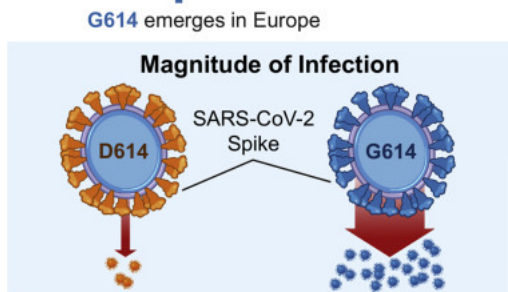
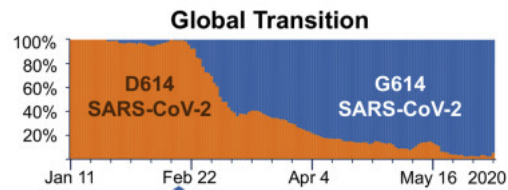
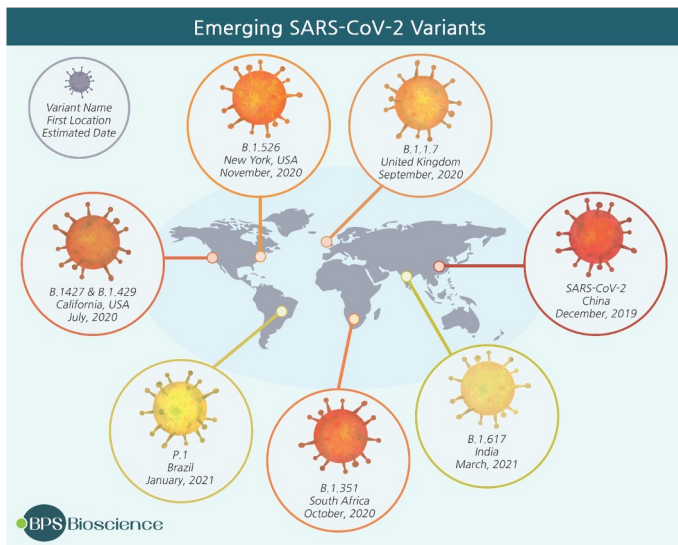
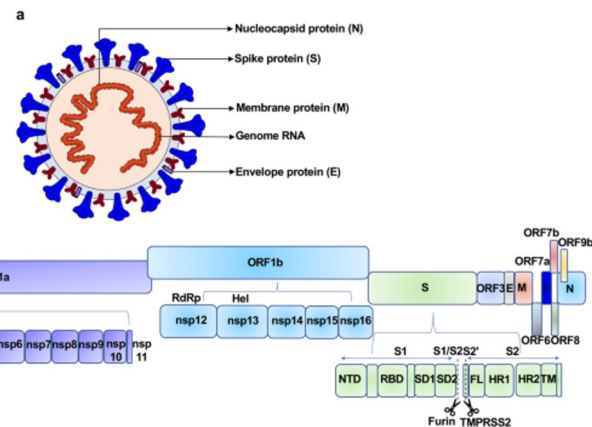
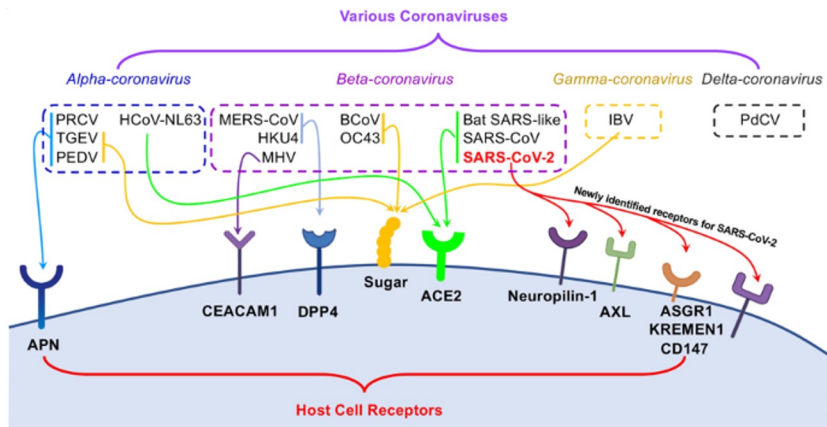
Data Visualization 2: Spike Protein (10 minutes)

Group Discussion 3: Spike Protein (5 minutes)

Data visualization 3: Gene of Choice (10 minutes)

Group Discussion 4: Students Share Out Gene of Choice Data (15 minutes)

Total: 1.5-2 hours



The Delta variant is more contagious than previous strains—it may cause more than 2x as many infections

ORIGINAL COVID-19 STRAIN vs **DELTA VARIANT**

Vaccines protect you from hospitalization, severe infections, and death

cdc.gov/coronavirus

CDC

CS 32201-AA 08/02/2021

Basic Lesson Outline (in class)

Introduction by instructor (15 minutes)

Group Discussion 1: Mutations Definitions and Examples (15 minutes)

Data Visualization 1: Whole SARS-CoV-2 Genome (15 minutes)

Group Discussion 2: Whole Genome Data (5 minutes)

Data Visualization 2: Spike Protein (10 minutes)

Group Discussion 3: Spike Protein (5 minutes)

Data visualization 3: Gene of Choice (10 minutes)

Group Discussion 4: Students Share Out Gene of Choice Data (15 minutes)

Total: 1.5-2 hours

Before we start plotting data...

- 1) Provide a definition for each of the following common types of mutations. In addition, provide an example of each type of mutation. If the mutation changes an amino acid, show how it does so in your example (you can find a codon table in the “Resources” section below).
 - a) Missense mutation:
 - i) Definition:
 - ii) Example:

 - b) Nonsense mutation:
 - i) Definition:
 - ii) Example:

 - c) Silent mutation:
 - i) Definition:
 - ii) Example:

 - d) Frameshift mutation:
 - i) Definition:
 - ii) Example:

- 2) Which kind of mutation (of the 4 we just defined) do you think is most commonly found in a wild population, point mutations or frameshift mutations? Why?

- 3) Which kind of point mutations do you think are most common? Why?

Basic Lesson Outline (in class)

Introduction by instructor (15 minutes)

Group Discussion 1: Mutations Definitions and Examples (15 minutes)

Data Visualization 1: Whole SARS-CoV-2 Genome (15 minutes)

Group Discussion 2: Whole Genome Data (5 minutes)

Data Visualization 2: Spike Protein (10 minutes)

Group Discussion 3: Spike Protein (5 minutes)

Data visualization 3: Gene of Choice (10 minutes)

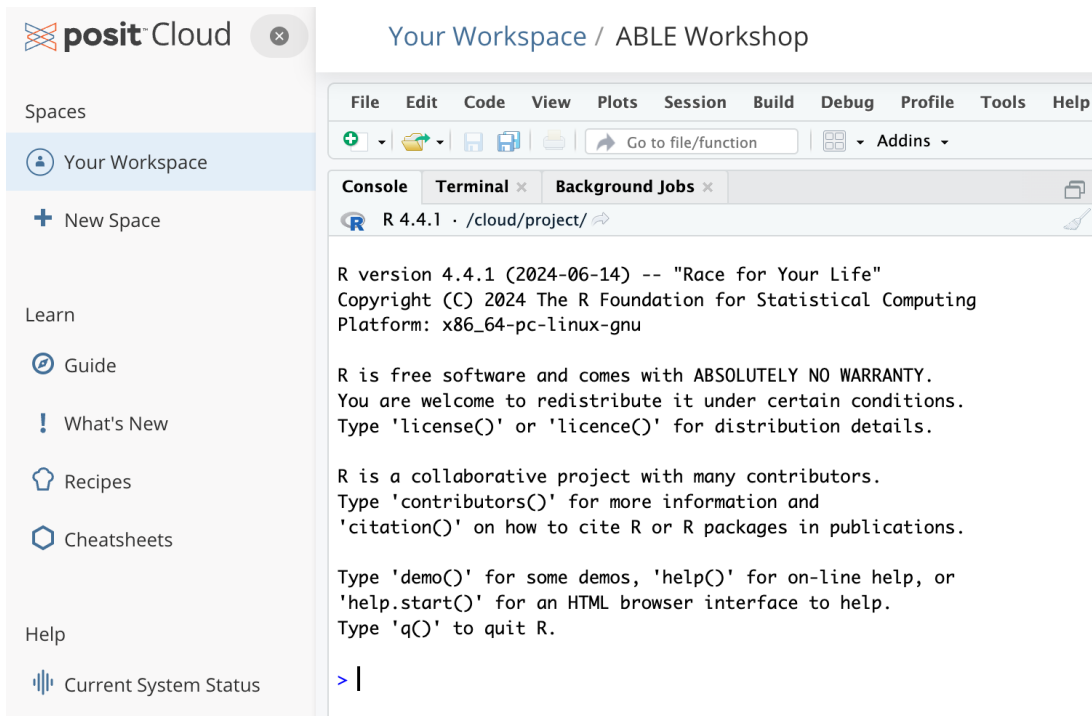
Group Discussion 4: Students Share Out Gene of Choice Data (15 minutes)

Total: 1.5-2 hours

Our Data

- UPenn SARS-CoV-2 variant database
- Mid-June 2020
- International data - primarily South African data with some Chinese and European data

Demonstration: Whole Genome Variation Mapping



The screenshot shows the Posit Cloud RStudio interface. The top bar displays "Your Workspace / ABLE Workshop". The left sidebar contains navigation options: Spaces (Your Workspace, New Space), Learn (Guide, What's New, Recipes, Cheatsheets), and Help (Current System Status). The main console window shows the R version 4.4.1 (2024-06-14) and the following text:

```
R version 4.4.1 (2024-06-14) -- "Race for Your Life"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

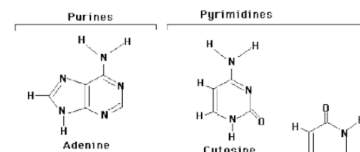
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

Now we'll start working with the whole genome data...

- 4) How many mutations are in the UPenn database? Hint: take the number of rows of data and subtract 1 (for the header of the table).
- 5) Based on the data presented, which point mutation class is most common?
- 6) There are two general types of nucleotide substitutions, transitions and transversions. What is the difference between the two?
- 7) What is the GC content of SARS-CoV-2? You'll have to look this up. An approximate number is ok.
- 8) How does the SARS-CoV-2 genome GC content compare to the GC content of a human genome?
- 9) Why do you think there are so many C → T nucleotide substitutions? Consider the fact that deamination of cytosine is not uncommon in cells.



Demonstration (time permitting): Spike Variation Mapping

Introduction by instructor (15 minutes)

Group Discussion 1: Mutations Definitions and Examples (15 minutes)

Data Visualization 1: Whole SARS-CoV-2 Genome (15 minutes)

Group Discussion 2: Whole Genome Data (5 minutes)

Data Visualization 2: Spike Protein (10 minutes)

Group Discussion 3: Spike Protein (5 minutes)

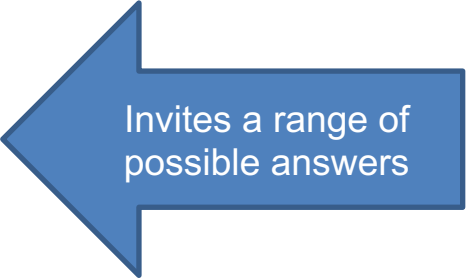
Data visualization 3: Gene of Choice (10 minutes)

Group Discussion 4: Students Share Out Gene of Choice Data (15 minutes)

Total: 1.5-2 hours

12) What do you see when you look at the distribution of mutations across the whole genome?

13) What else would you like to know about these mutations or the virus they came from?



Invites a range of possible answers

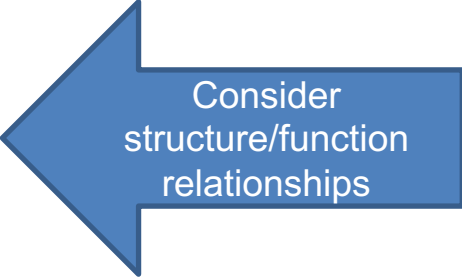
Now let's look at individual genes...

S gene (Spike)

14) What is the function of the S gene?

15) What do you see when you look at the distribution of mutations across the S gene?

16) Does this give you any ideas about what parts of the S gene are more important than others?

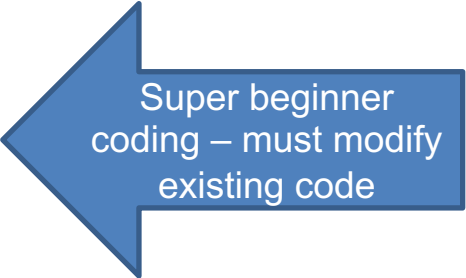


Consider structure/function relationships

If there is time (E gene [Envelope], M gene [Matrix], or N gene [nucleocapsid]):

17) What is the function of your chosen gene?

18) Consider questions 15 and 16, but now in the context of your chosen gene, not S. Do your answers change? Why or why not?



Super beginner coding – must modify existing code

Questions?



Mini Workshop Files:

<https://tinyurl.com/SARSinR>



Please fill out an evaluation for these materials:

<https://tinyurl.com/Barry>

[MiniEval](#)