



# ASM Microbe 2024

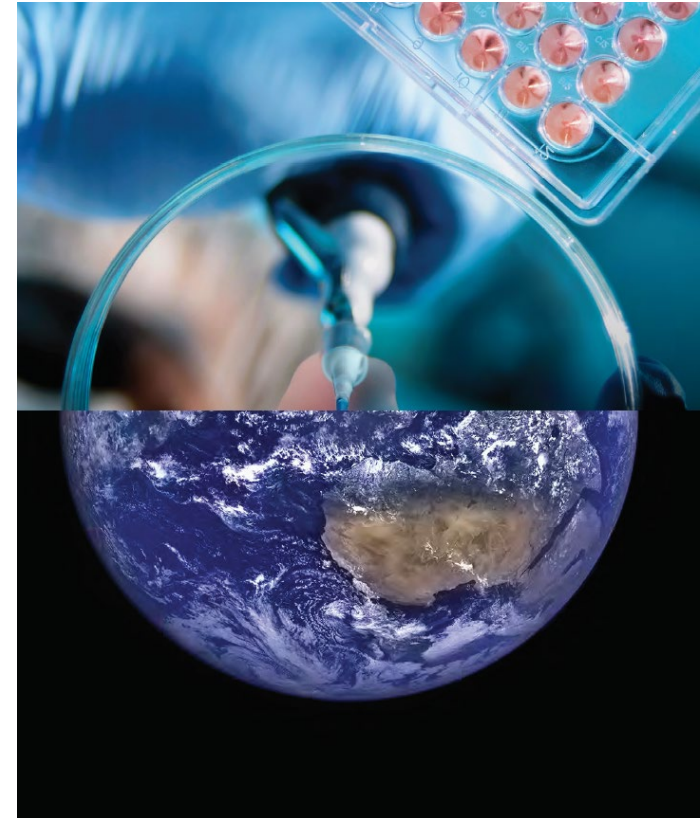
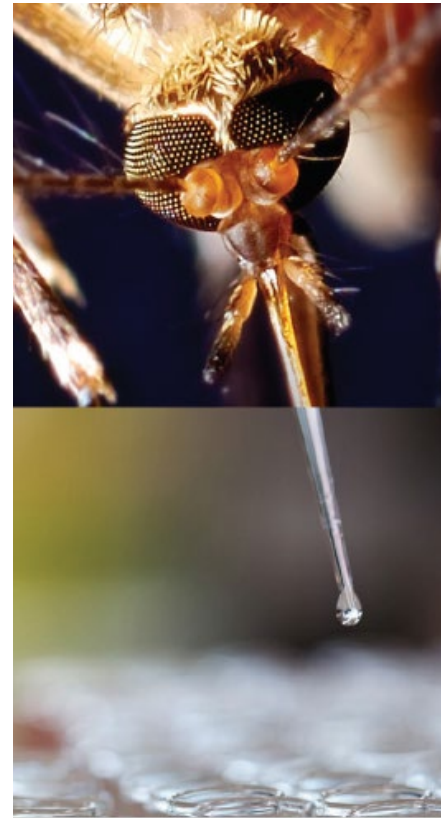
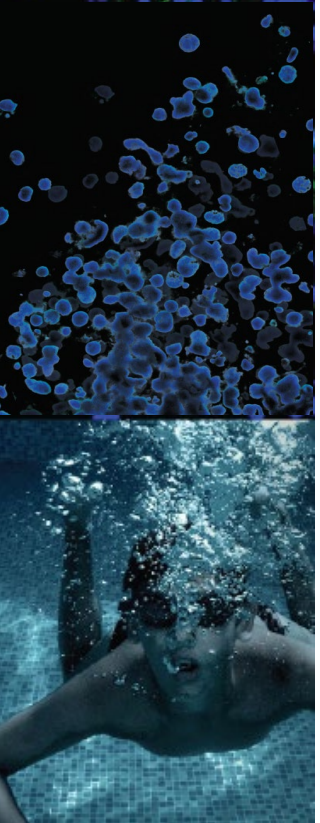
---

## RAISIN

A Pipeline Intended to Better Characterize Variants on the Amino Acid Level

Nikhita Puthuveetil, MS  
Senior Bioinformatician, ATCC®  
June 14, 2024

Credible Leads to Incredible™

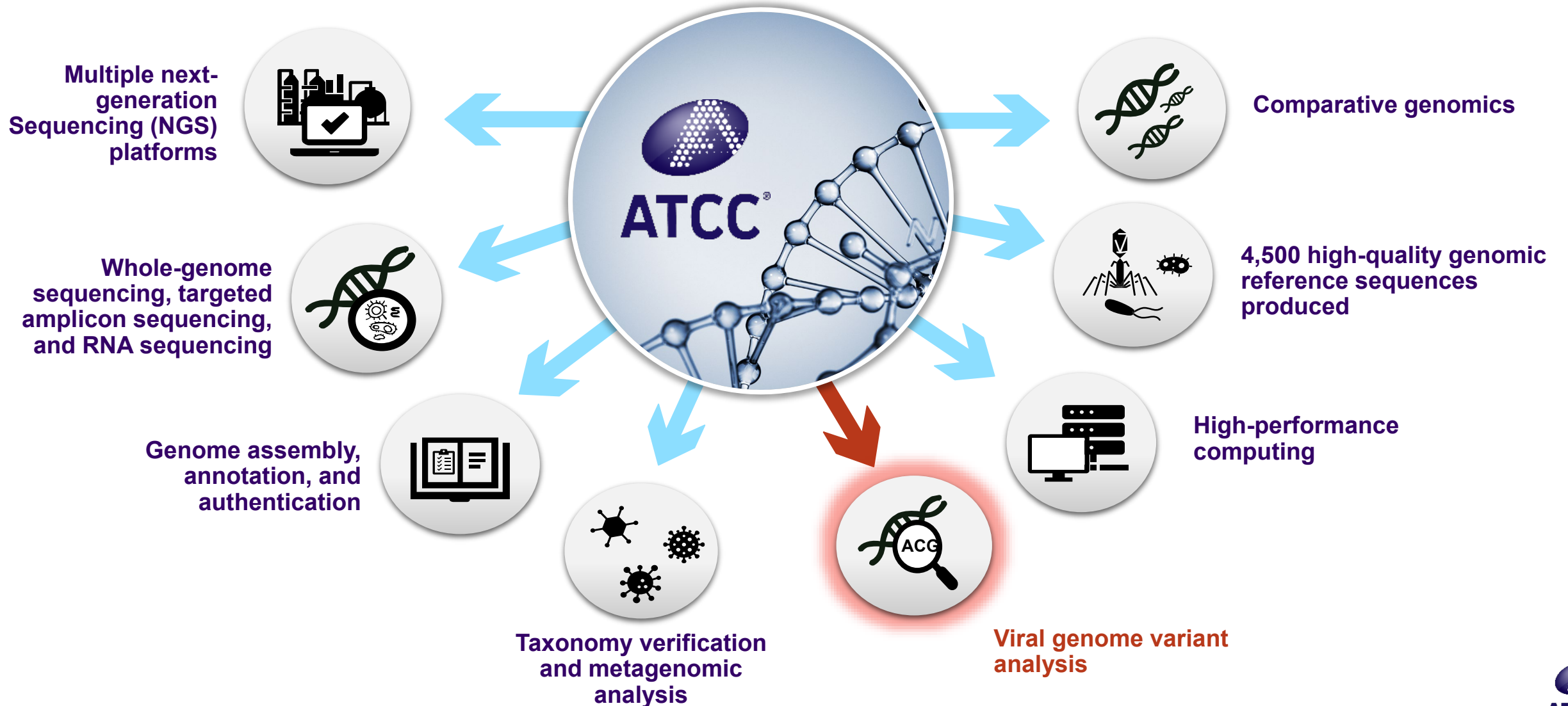


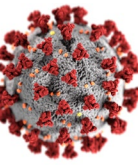
Visit us at booth #912

# About ATCC®

- Founded in 1925, ATCC® is a non-profit organization with HQ in Manassas, VA, and an R&D and Services center in Gaithersburg, MD
- World's premier biological materials resource and standards development organization
  - 5,000 cell lines
  - 80,000 microorganisms
  - Genomic & synthetic nucleic acids
  - Media/reagents
- ATCC® collaborates with and supports the scientific community with industry-standard biological products and innovative solutions
- Growing portfolio of products and services
- Sales and distribution in 150 countries, 19 international distributors
- Talented team of 600+ employees, over one-third with advanced degrees

# ATCC'S Sequencing & Bioinformatics Center

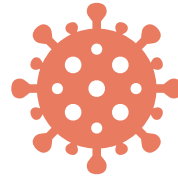
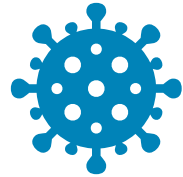
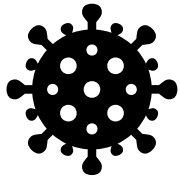




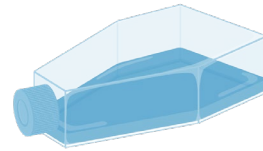
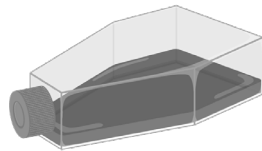
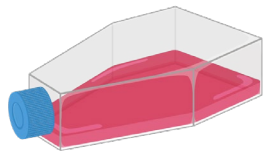
# Tracking variance in SARS-CoV-2 samples

*Ensuring product authenticity and creating a standard for SARS-CoV-2*

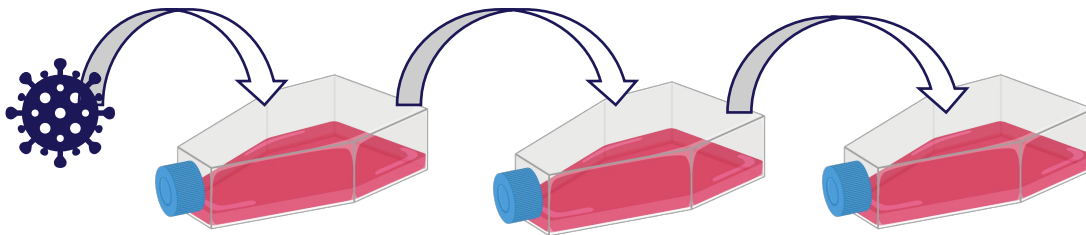
- A need to track genomic variance between SARS-CoV-2 samples quickly and effectively
  - Between strains

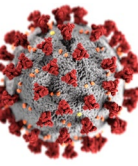


- Between cell lines



- Between passages

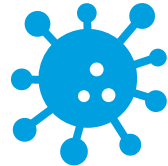
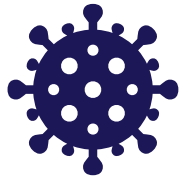




# Could we do this with any virus?

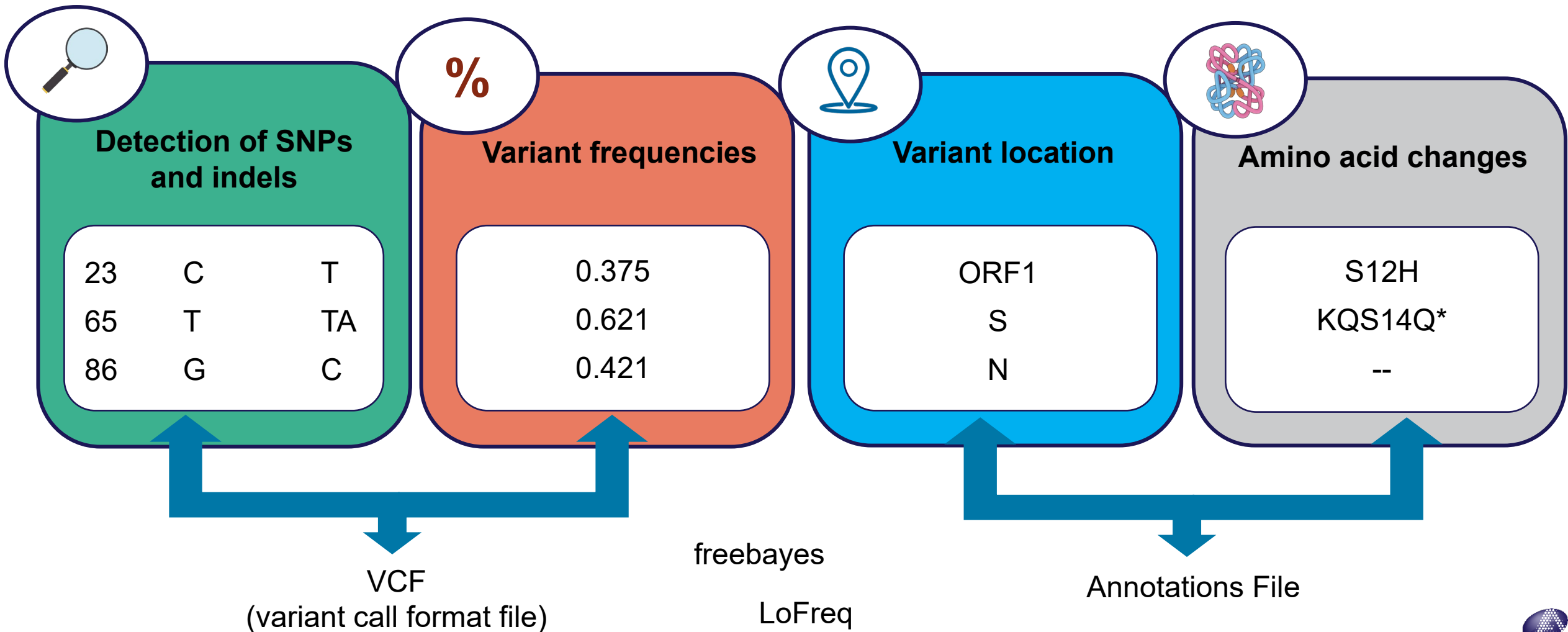
*Extending SARS-CoV-2 analysis to any virus sample*

- A need to track genomic variance for **any virus** sample quickly and effectively
- Ensure product authenticity for any strain



# Components of variant analysis

Example data



# A need for a variant annotation tool

*If you only need a VCF + annotations file, can't anyone do this?*



**Variant callers that provide amino-acid translations are rare**



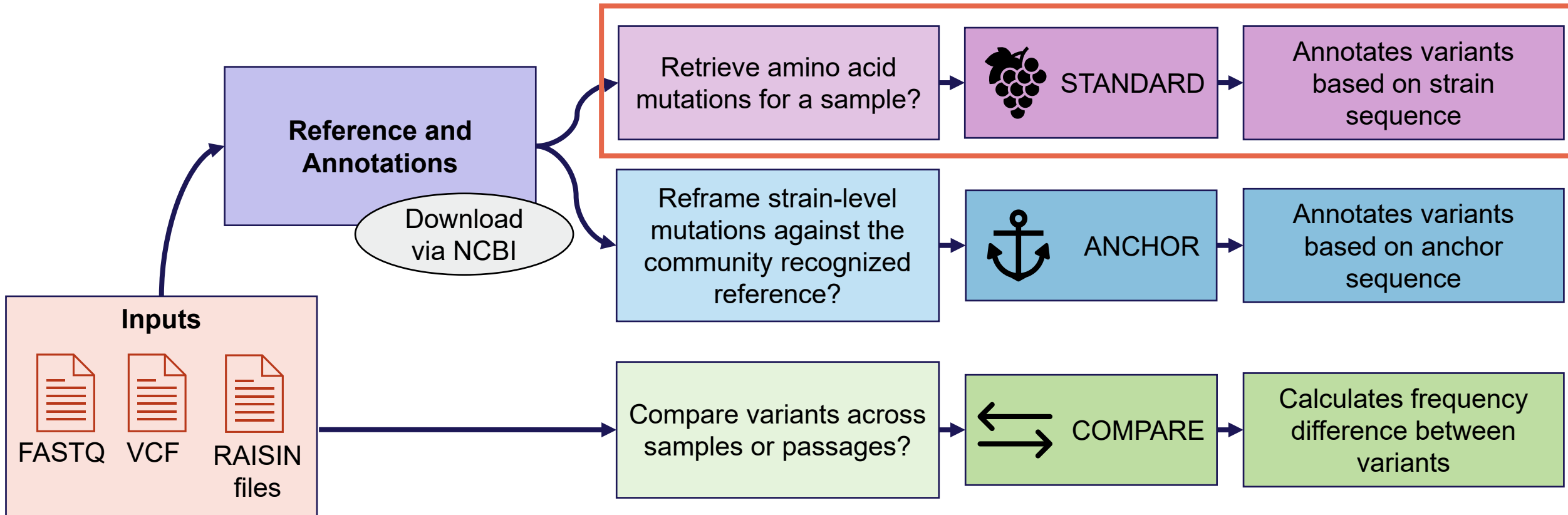
**Available annotations databases aren't comprehensive**



**Manually translating amino acid variants is time-consuming**

# Overview of RAISIN

Retrieving Amino acid Implications from Sequencing IterationNs (RAISIN)



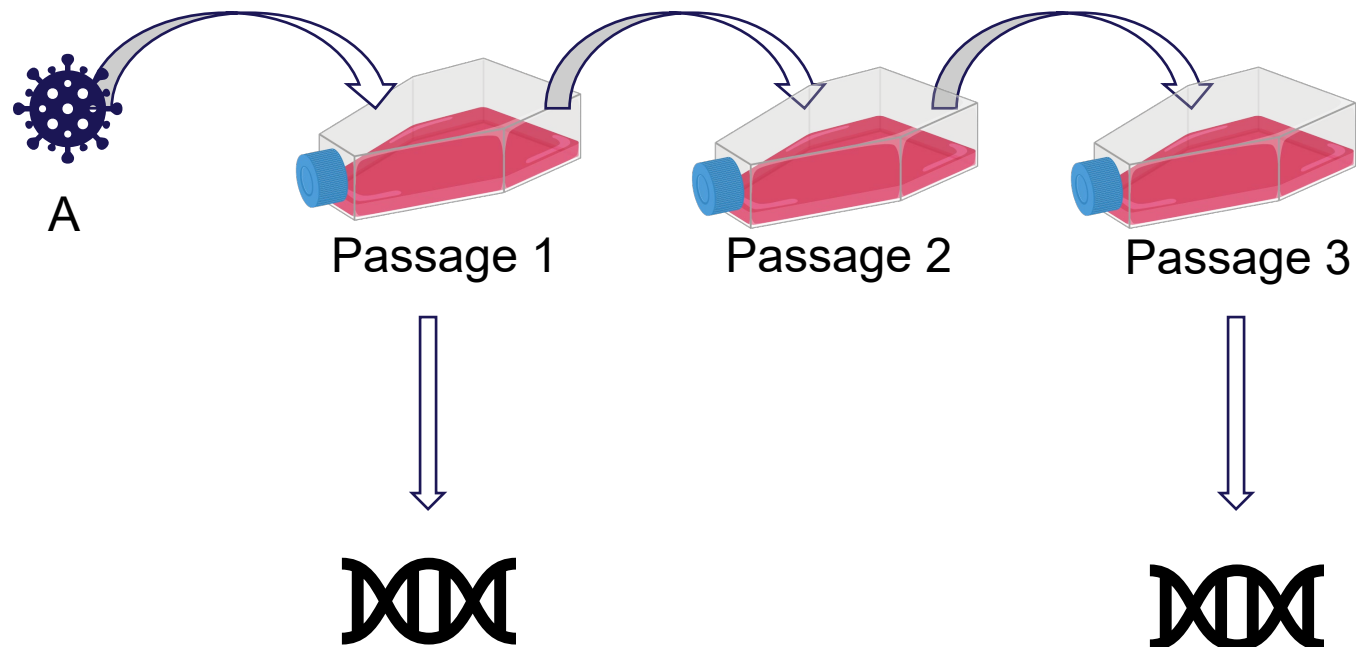




# Case Study

How we used RAISIN's STANDARD mode

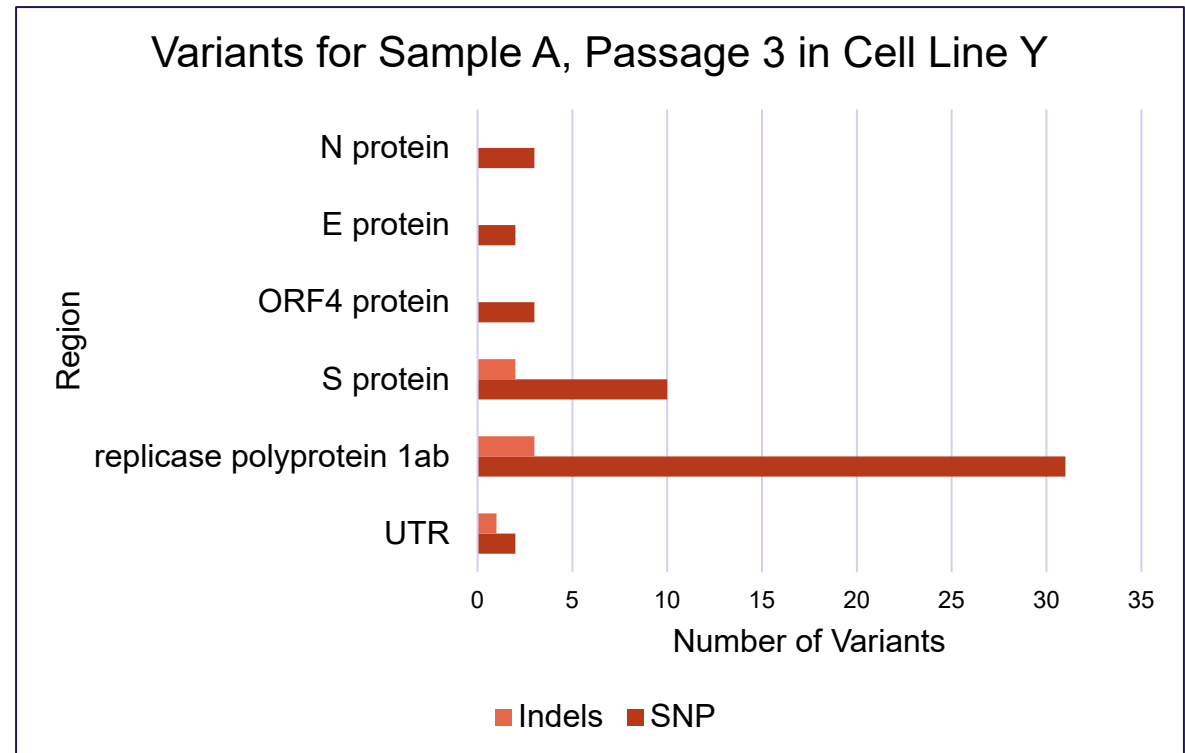
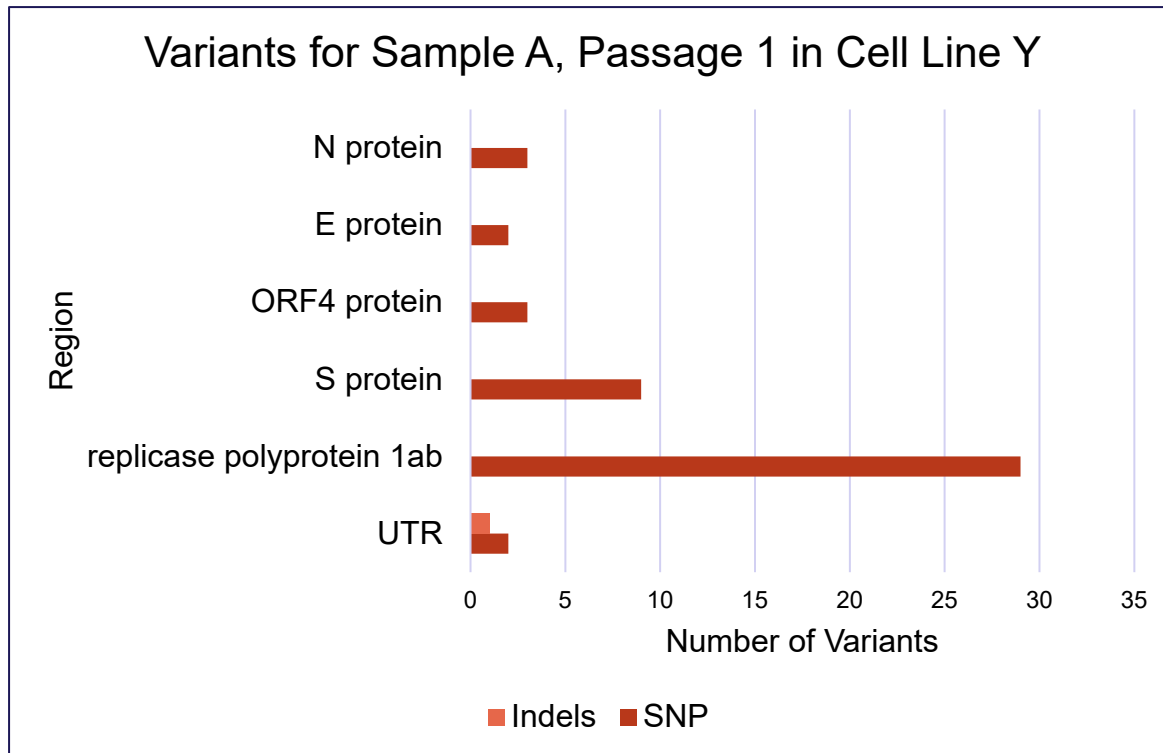
- In one of our previous human coronavirus studies, sample A was serially passaged in cell line Y to determine if Y was an adequate host for this sample
- The 1<sup>st</sup> and 3<sup>rd</sup> passages were sequenced and then run through RAISIN STANDARD.





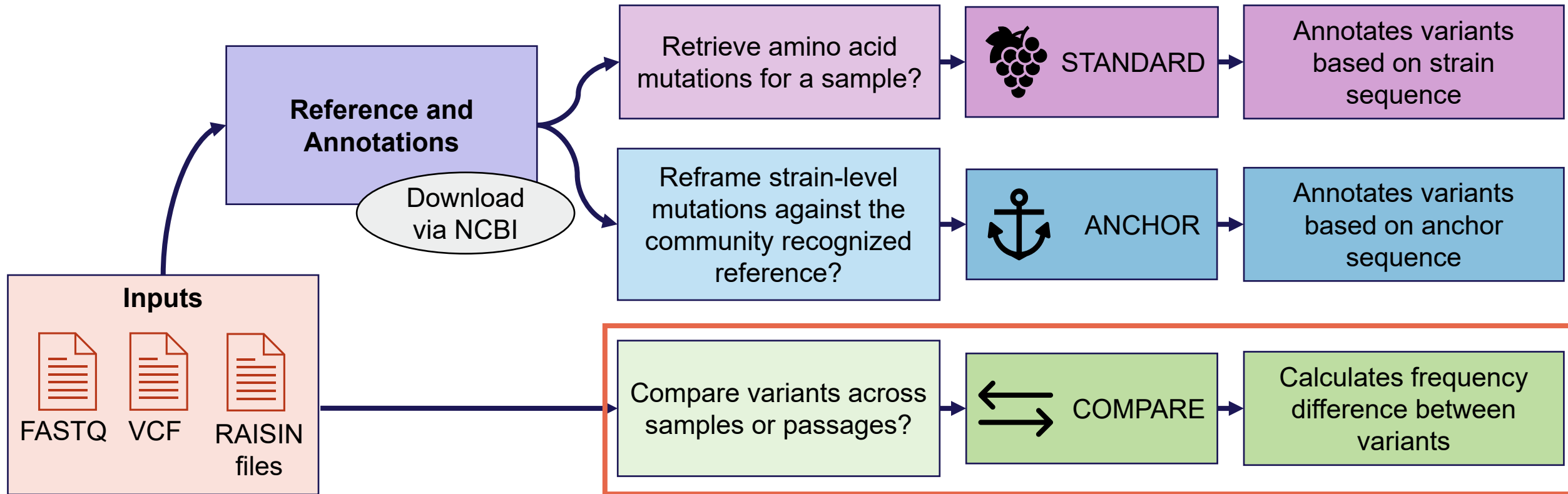
# Increase in variants in S and replicase regions

Passage 3 shows more variants in the S and replicase regions



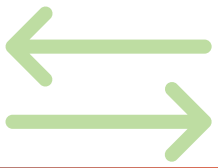
# Overview of RAISIN: Compare mode

Retrieving Amino acid Implications from Sequencing IteratioNs (RAISIN)



# COMPARE mode output

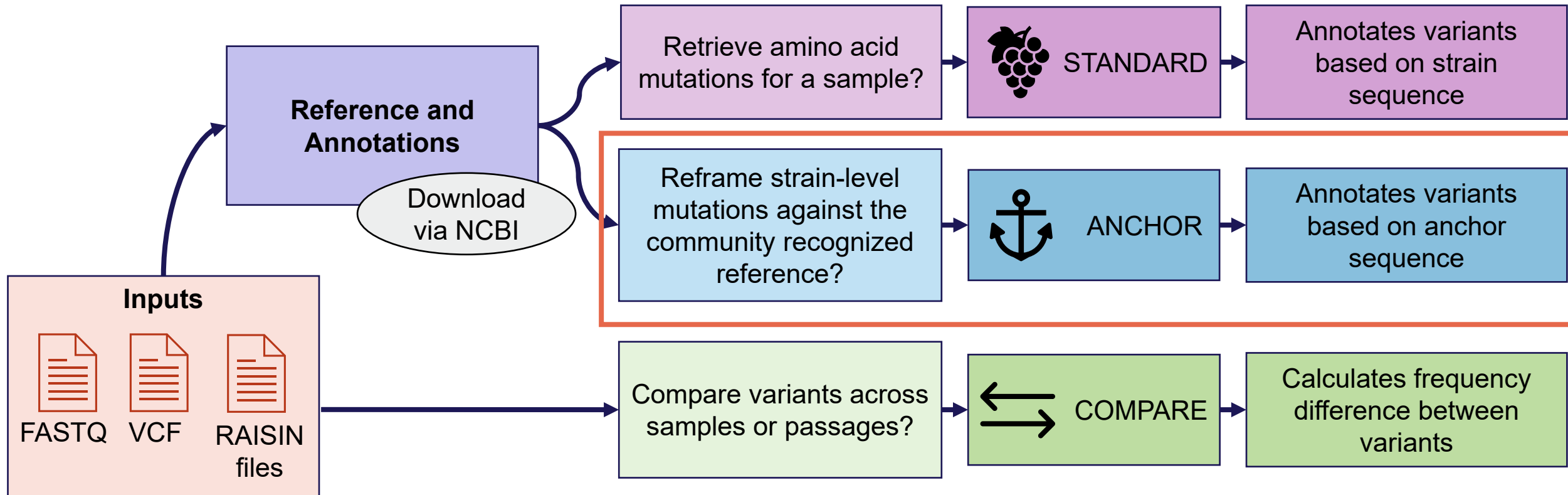
Compare variants across samples or passages



| Position | Reference allele | Alternate allele | Protein                   | Frequency Difference |
|----------|------------------|------------------|---------------------------|----------------------|
| 55       | C                | A                | UTR                       | 9.09%                |
| 60       | A                | G                | UTR                       | -1.87%               |
| 432      | G                | A                | replicase polyprotein 1ab | 0%                   |
| 1779     | A                | G                | replicase polyprotein 1ab | 0.69%                |
| 20300    | T                | C                | replicase polyprotein 1ab | 0%                   |
| 20676    | C                | T                | S protein                 | -4.54%               |
| 20842    | T                | G                | S protein                 | 1.38%                |
| 21164    | G                | A                | S protein                 | -3.22%               |

# Overview of RAISIN: ANCHOR mode

Retrieving Amino acid Implications from Sequencing IteratiONs (RAISIN)

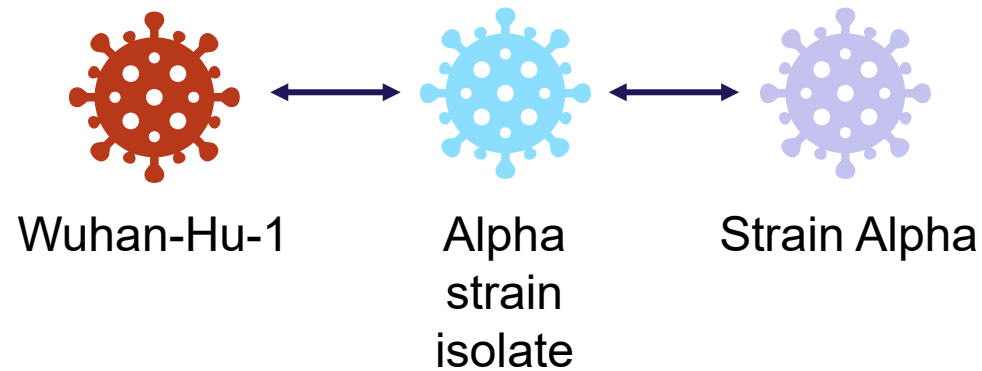




# Why the need for ANCHOR mode?

*Looking back to SARS-CoV-2 strain authentication*

- Comparison of sample strain to other strains



- Three-way comparison of sample to strain reference and Wuhan-Hu-1 reference
  - How has the sample deviated from Wuhan-Hu-1?
  - How has the sample deviated from the strain of interest?



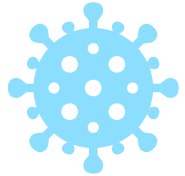
# ANCHOR mode localizes variants

*Variants are under one coordinate system*

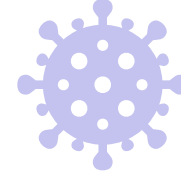
Sample should have a deletion at position 50

Position 50 is based on the coordinates of Wuhan-Hu-1

Where does that correspond to in the alpha strain?



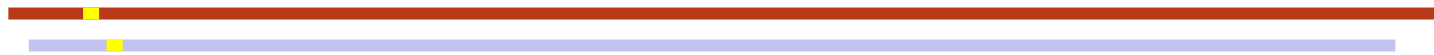
ATAA



Alpha strain isolate

Wuhan-Hu-1 reference  
29,903 bp

Alpha strain reference  
29,893 bp



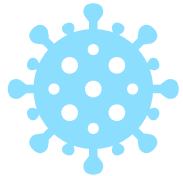
Deletion is at position 45 based on alpha strain



# ANCHOR mode localizes variants

*Variants are under one coordinate system*

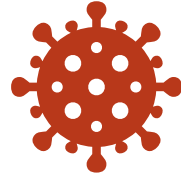
Sample should have a deletion at position 50



ATAA

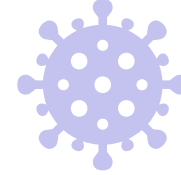
Alpha strain isolate

Position 50 is based on the coordinates of Wuhan-Hu-1



Wuhan-Hu-1 reference  
29,903 bp

Where does that correspond to in the alpha strain?



Alpha strain reference  
29,893 bp

- ANCHOR mode takes care of the math and reports all variants based on the coordinate position of the anchor reference
- Reporting variants under one coordinate system is convenient








# ANCHOR mode classifies variants based on differences



Compares each allele and assigns a classification (I-IV)

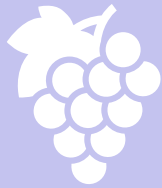
| Variant Classification |                         | Example Scenarios |               |               |
|------------------------|-------------------------|-------------------|---------------|---------------|
|                        |                         | Anchor Allele     | Strain Allele | Sample Allele |
| I                      | Anchor allele different | <b>T</b>          | G             | G             |
| II                     | Sample allele different | T                 | T             | <b>C</b>      |
| III                    | Strain allele different | T                 | <b>A</b>      | T             |
| IV                     | All alleles different   | <b>C</b>          | <b>G</b>      | <b>T</b>      |

## Legend:

-  **ANCHOR:** a scientific community recognized or experiment-specific sequence
-  **STRAIN:** a sequence that is more closely related to the sample
-  **SAMPLE:** a consensus sequence of the sample reads mapped to the strain sequence

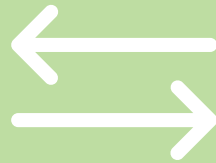
# Summary of RAISIN modes

*With each mode of RAISIN, viral variants are characterized under a different lens*



## STANDARD

**Retrieves amino acid mutations for viral variants**



## COMPARE

**Ideal for detecting differences in variants between passages**



## ANCHOR

**Reframes variants against a community recognized reference**

- Future developments:
  - Expanding to other microbes
  - Addition of a predictive viral annotation step

Visit us at [genomes.atcc.org](https://genomes.atcc.org)



**Nikhita Puthuveetil, MS**

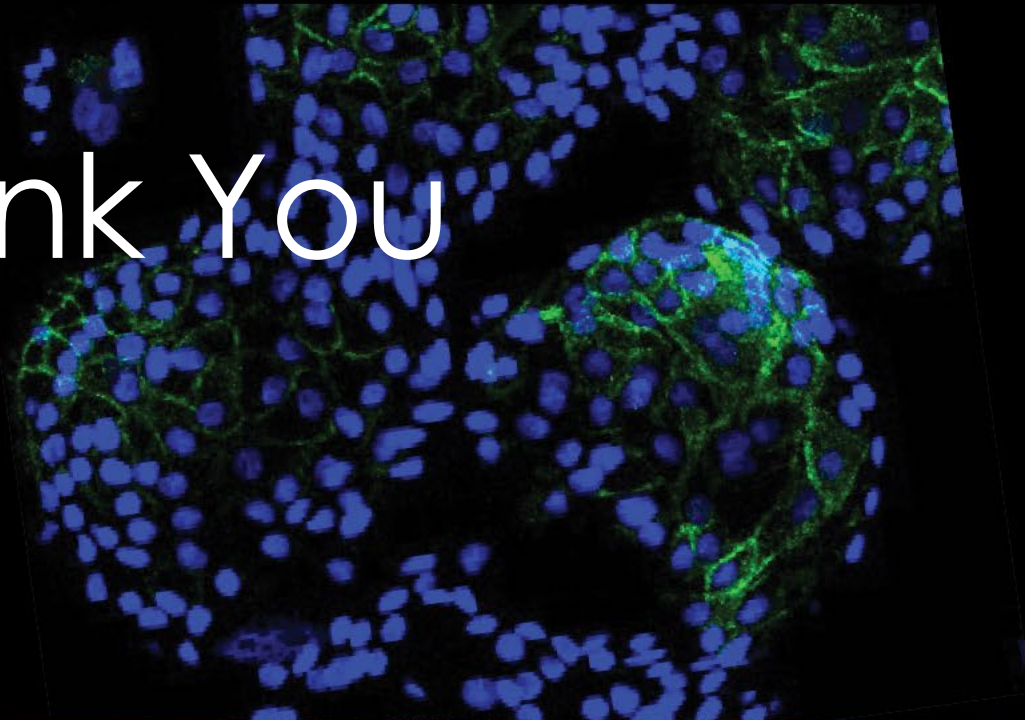
Senior Bioinformatician

ATCC

[nputhuveetil@atcc.org](mailto:nputhuveetil@atcc.org)

Visit us at booth #912  ATCC®

Thank You



**CREDIBLE**  
**MODELS**

**INCREDIBLE**  
**OUTCOMES**



ATCC® | CREDIBLE LEADS TO INCREDIBLE