

COVID-19 Cluster Analysis

April 22, 2025
Abbey Marye

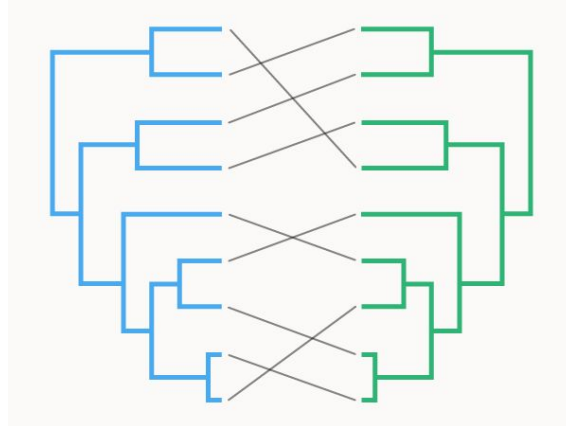


Primary Questions

- Do genomic sequencing and contact tracing cover different populations in Utah? Are there areas of the state where contact tracing revealed clusters, but genomic sequencing was insufficient?
- Does genomic clustering identify plausible transmission clusters that may have been missed by traditional contact tracing alone?
- Are there consistent genomic links within epidemiologically-identified clusters?

Purpose

- During the COVID-19 pandemic, epidemiological case investigation helped to define disease clusters
- SARS-CoV-2 became the most sequenced virus, this information could provide greater resolution for case clustering
- Understanding our sequence coverage across the state can guide us for future use of genomic data



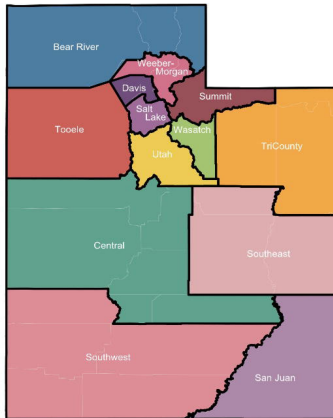
Methods: Study Design

- Retrospective analysis utilizing data from March 1, 2020 - December 31, 2024
 - Alpha (March 2020 - June 2021)
 - Delta (June 2021 - December 2021)
 - Omicron (December 2021 - September 2023)
- Epidemiological and demographic data extracted from EpiTrax
- Genomic data collected from NCBI Genbank and GISAID
- Cases defined as “Interviewed”: interviewed over the phone, on-site at the testing location, interaction with ACTS online survey.
- Cases defined as “Sequenced”: Have valid WGS collection date



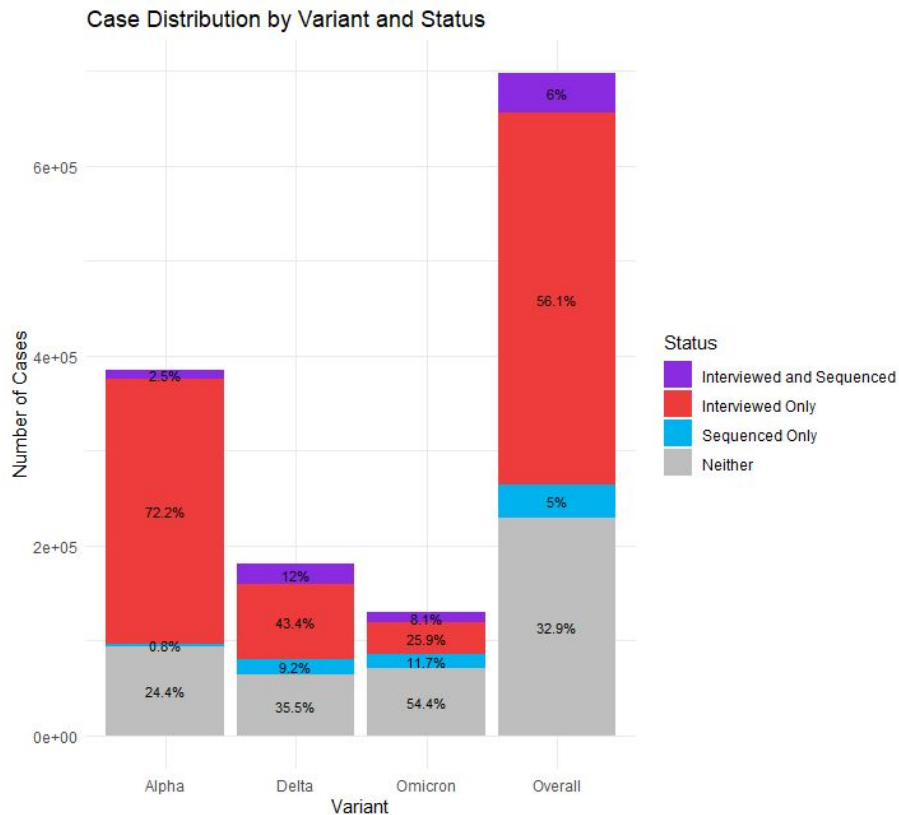
Methods: Question 1

- Descriptive statistics of overall population
- Binomial logistic regression to assess likelihood of a case being sequenced in each local health jurisdiction
- Multinomial logistic regression to determine likelihood of being sequenced vs. interviewed

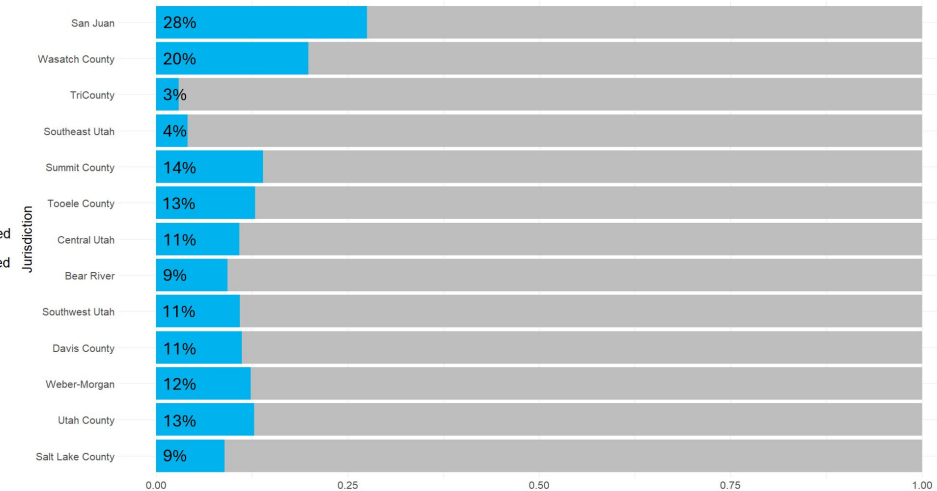
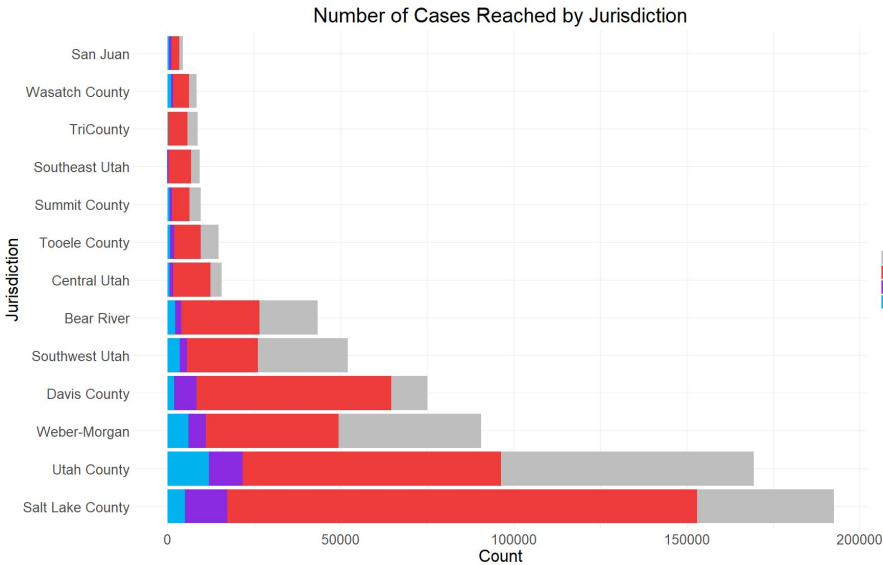


Results: Sample Characterization

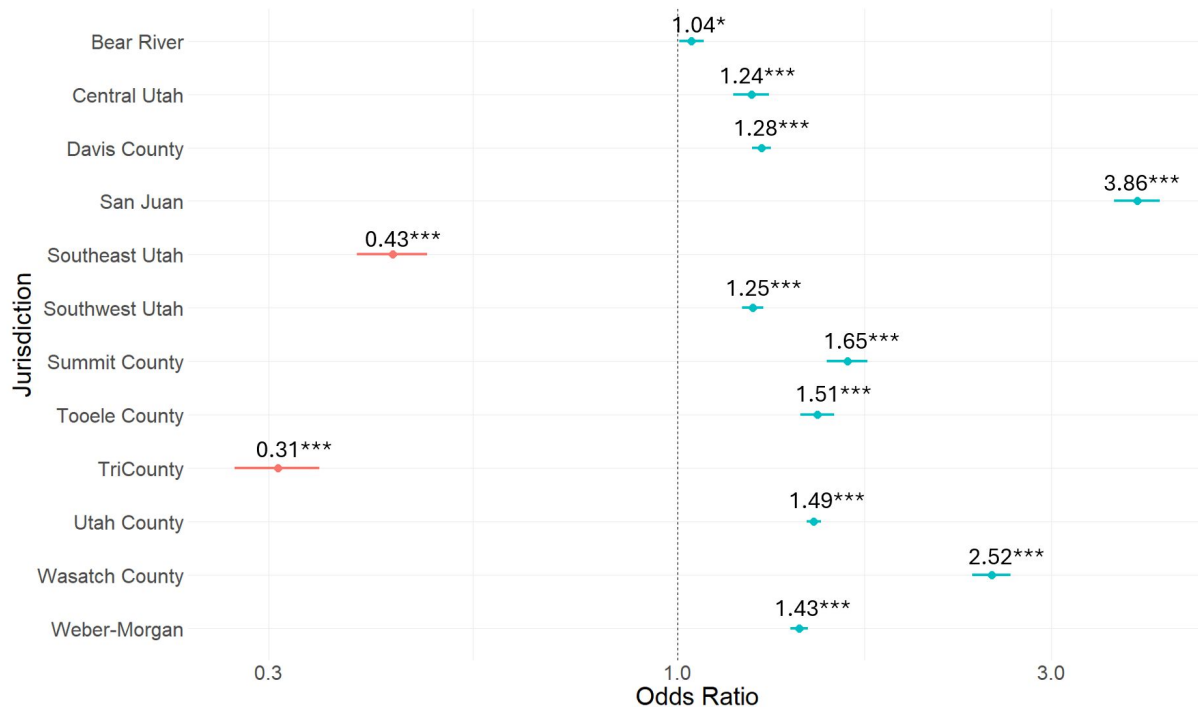
- Average case age: 35.1
- 51.1% female
- 71.8% White, non-Hispanic
- 27.6% Salt Lake County



Results: Surveillance Methods by Demographics and Jurisdiction

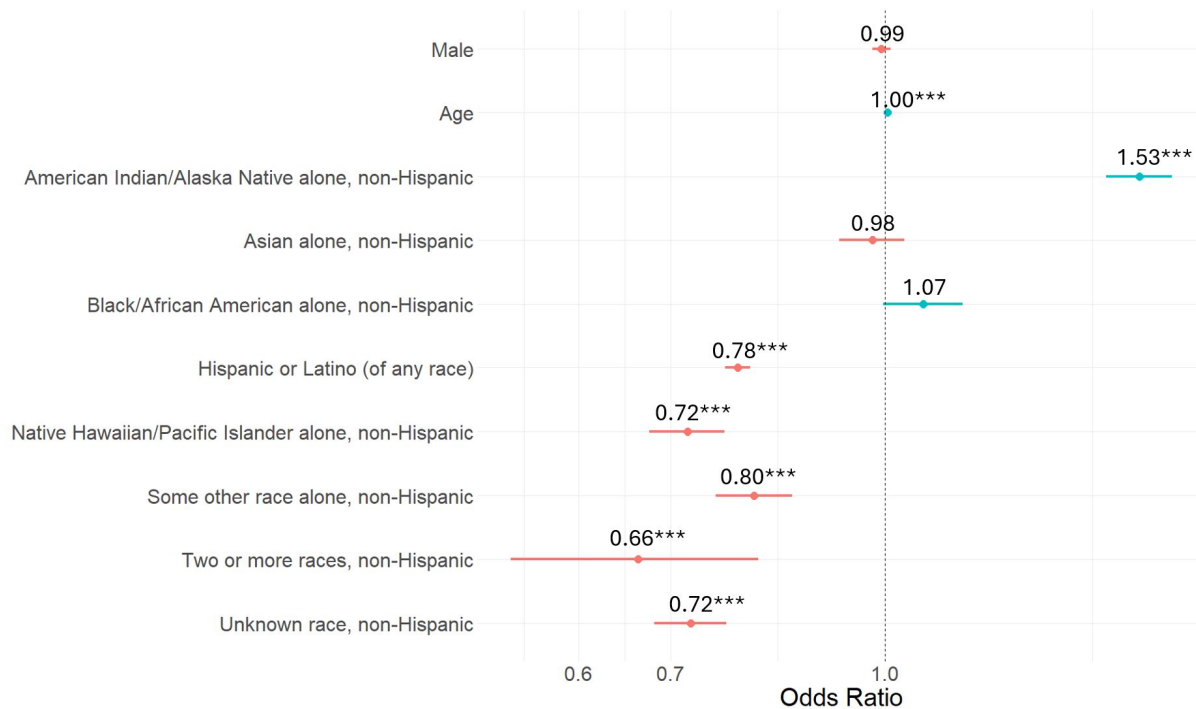


Results: Binomial logistic regression



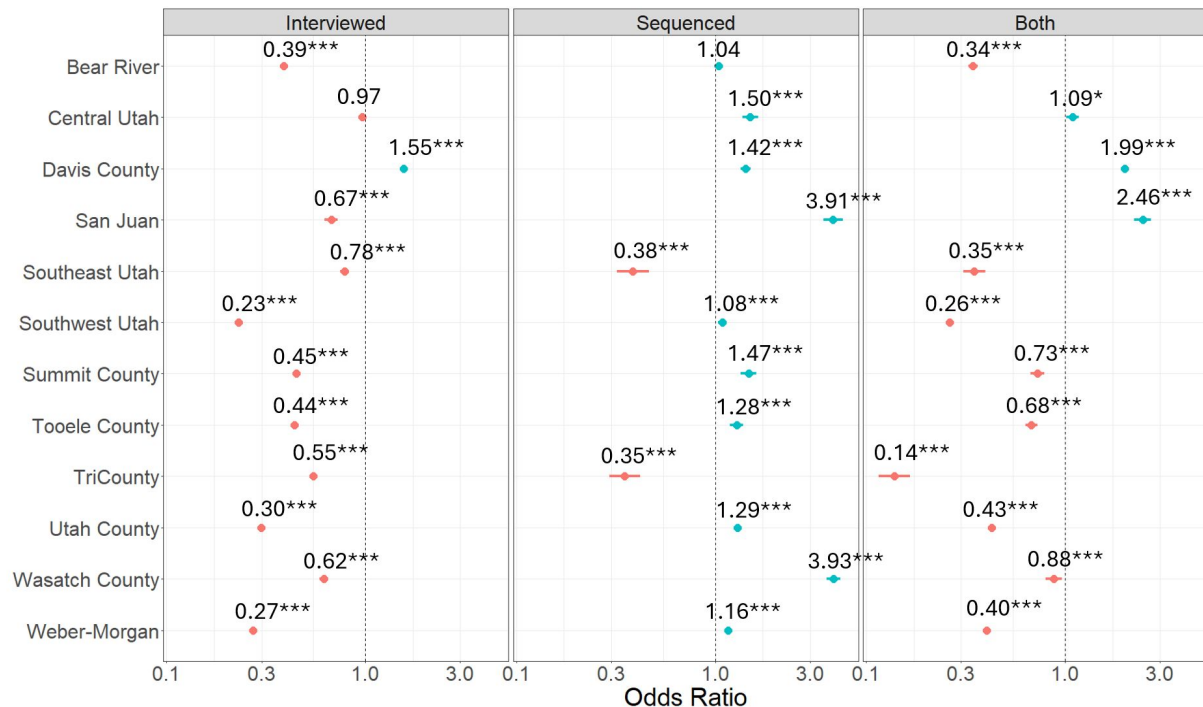
Odds of sequencing a case across the sampling period, using Salt Lake County as a reference

Results: Binomial logistic regression



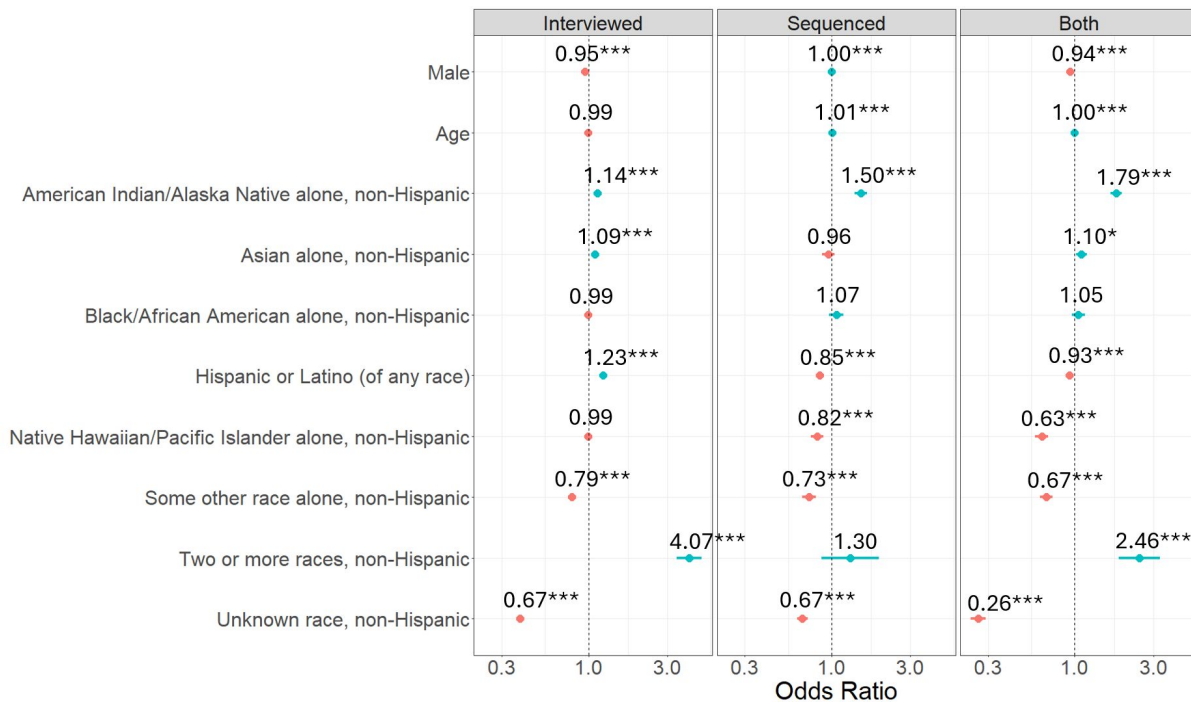
Adjusted odds of sequencing a case across the sampling period.

Results: Multinomial logistic regression



Odds of a case successfully undergoing sequencing only, interview only, or both, as compared to neither interview nor sequencing, across the sampling period.

Results: Multinomial logistic regression

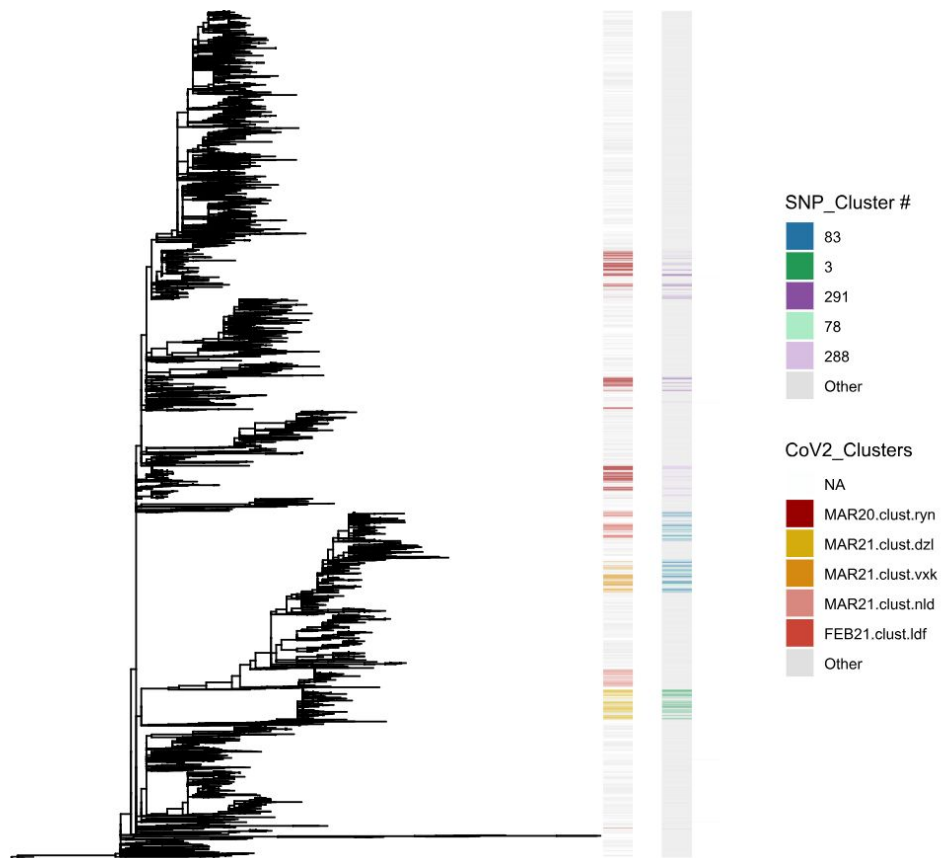


Adjusted odds of a case successfully undergoing sequencing only, interview only, or both, as compared to neither interview nor sequencing, across the sampling period.

Methods: Clustering

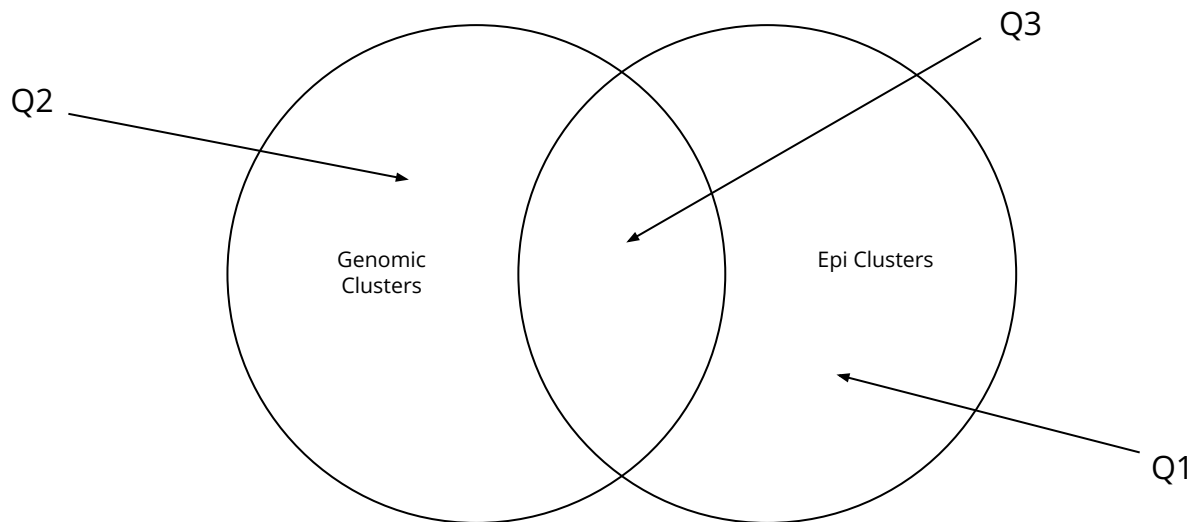
- Epidemiological Clustering
 - Clustering cases based on “outbreak” tag in EpiTrax
- Genomic Clustering
 - SNP Calling and alignment - snippy
 - SNP Matrix generation - snpdists
 - ML Tree generation - MAPLE
 - Clustering:
 - Cov2clusters on ML tree
 - Hclust on SNP matrix

Results: Comparing Genomic Clustering Methods



Methods: Cluster Similarity

- Q1/Q2: Determine Frequency of non-overlapping clusters
- Q3: Traditional cluster comparison methods (measures of similarity)
 - Jaccard Index
 - Adjusted Rand Index



Methods: Question 2 and 3

Thoughts?

References

1. Sobkowiak B, Haghmaram P, Prystajec N, Zlosnik JEA, Tyson J, Hoang LMN, et al. The utility of SARS-CoV-2 genomic data for informative clustering under different epidemiological scenarios and sampling. *Infect Genet Evol.* 2023 Sep;113:105484.
2. Snell LB, Fisher CL, Taj U, Stirrup O, Merrick B, Alcolea-Medina A, et al. Combined epidemiological and genomic analysis of nosocomial SARS-CoV-2 infection early in the pandemic and the role of unidentified cases in transmission. *Clin Microbiol Infect.* 2022 Jan;28(1):93–100.
3. Suster CJE, Arnott A, Blackwell G, Gall M, Draper J, Martinez E, et al. Guiding the design of SARS-CoV-2 genomic surveillance by estimating the resolution of outbreak detection. *Front Public Health.* 2022 Oct 5;10:1004201.
4. Stirrup O, Hughes J, Parker M, Partridge DG, Shepherd JG, Blackstone J, et al. Rapid feedback on hospital onset SARS-CoV-2 infections combining epidemiological and sequencing data. *eLife.* 2021 Jun 29;10:e65828.
5. Sobkowiak B, Kamelian K, Zlosnik JEA, Tyson J, Silva AGD, Hoang LMN, et al. Cov2clusters: genomic clustering of SARS-CoV-2 sequences. *BMC Genomics.* 2022 Oct 19;23(1):710.