COWW-Covid in WasteWater Analysis

Daniel Paiva Agustinho, Baylor College of Medicine (Team Leader)

Sushant Potdar, University of Arkansas

Alexis Quintero, National University of Colombia

Andrew Miller, University of Utah

John Arnn, Utah Public Health Laboratory

Rong Guo, University of Maryland, College Park

Rationale

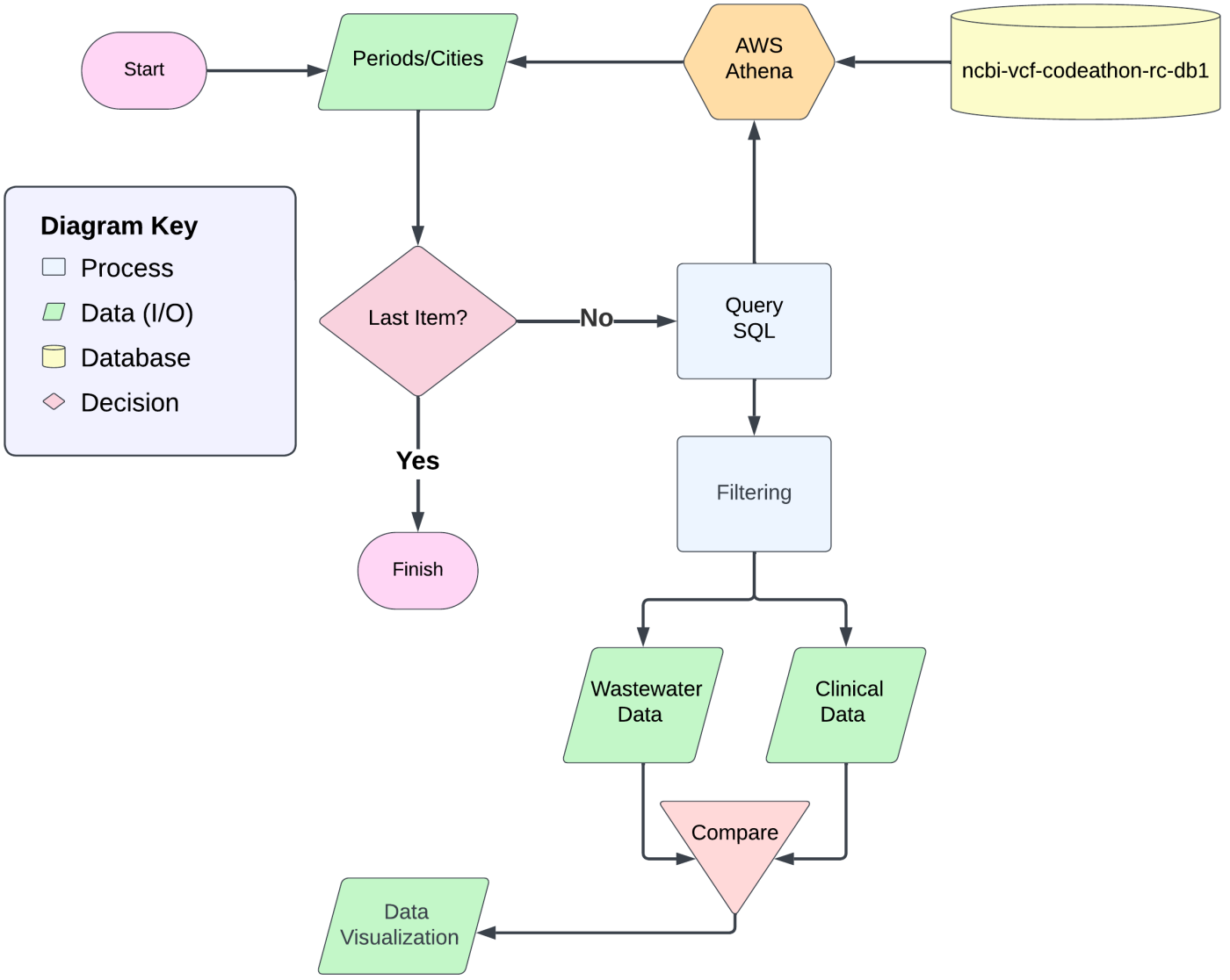
- Clinical isolate limitations:
  - Primarily represents severe cases, not fully capturing milder infections.
- Wastewater sequencing advantages:
  - Broad representation of circulating strains, including asymptomatic and mild cases.
  - Avoids bias towards severe cases.
- Wastewater sequencing challenges:
  - RNA degradation in wastewater may underestimate minor alleles or variants.
  - Includes non-infectious viral RNA, requiring careful interpretation.

Project Goals

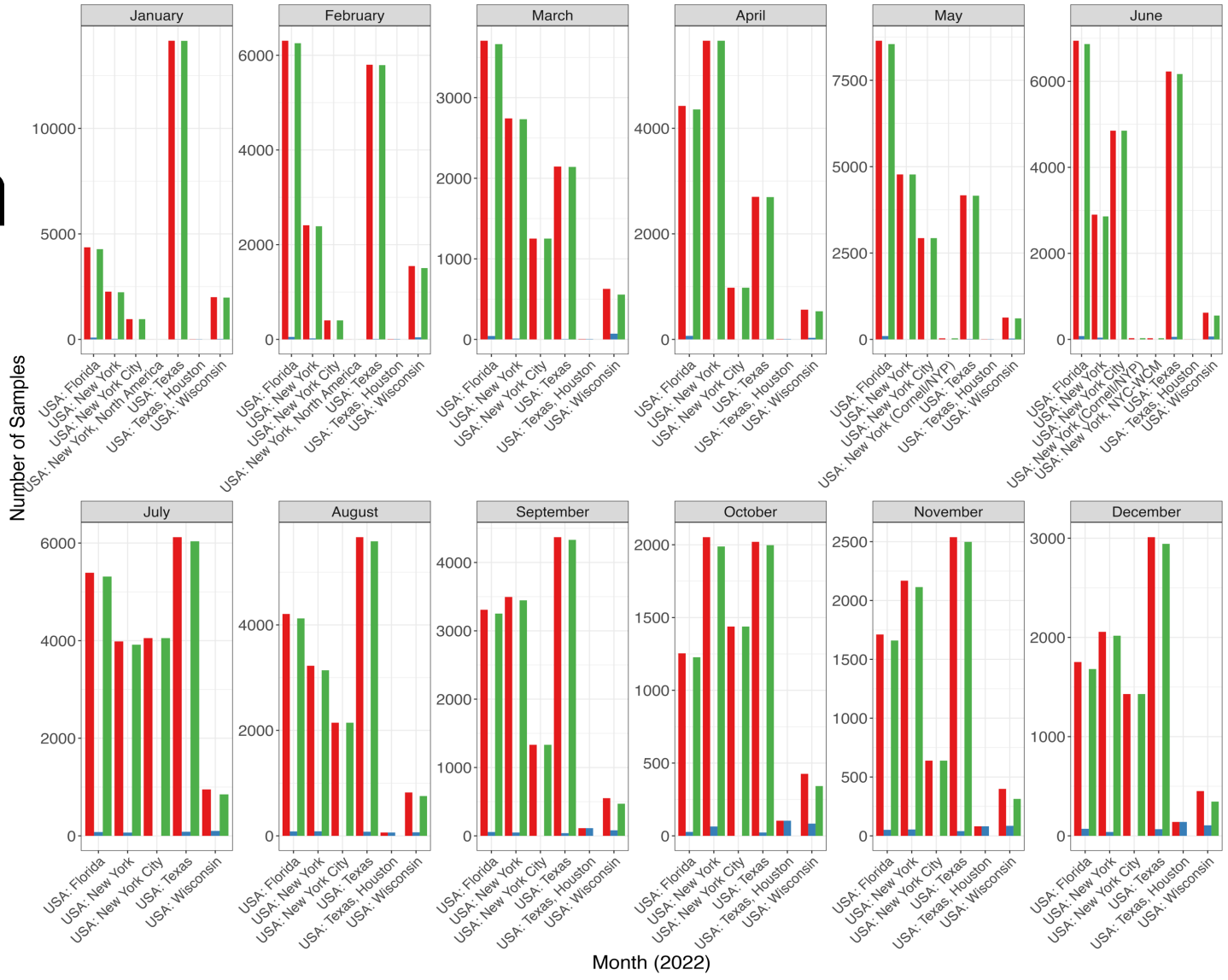
- Compare wastewater sequencing with clinical isolates of SARS-CoV-2 at specific geographical locations and time points.
- Evaluate representativeness of each approach for circulating viral strains.
- Gain insights into diversity and prevalence of SARS-CoV-2 in specific subsets.
- Understand strengths and limitations of wastewater and clinical isolate approaches.

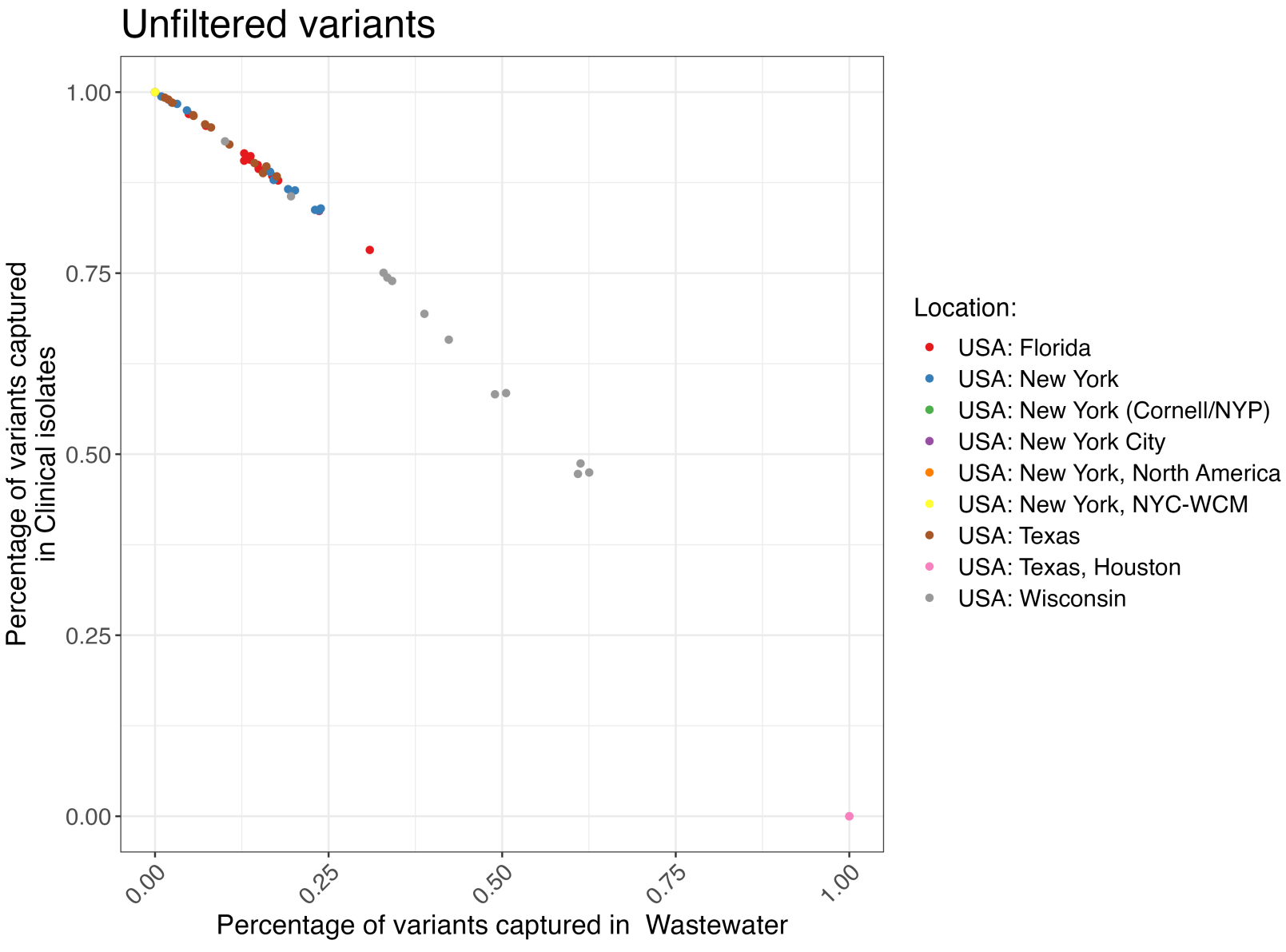
Approach

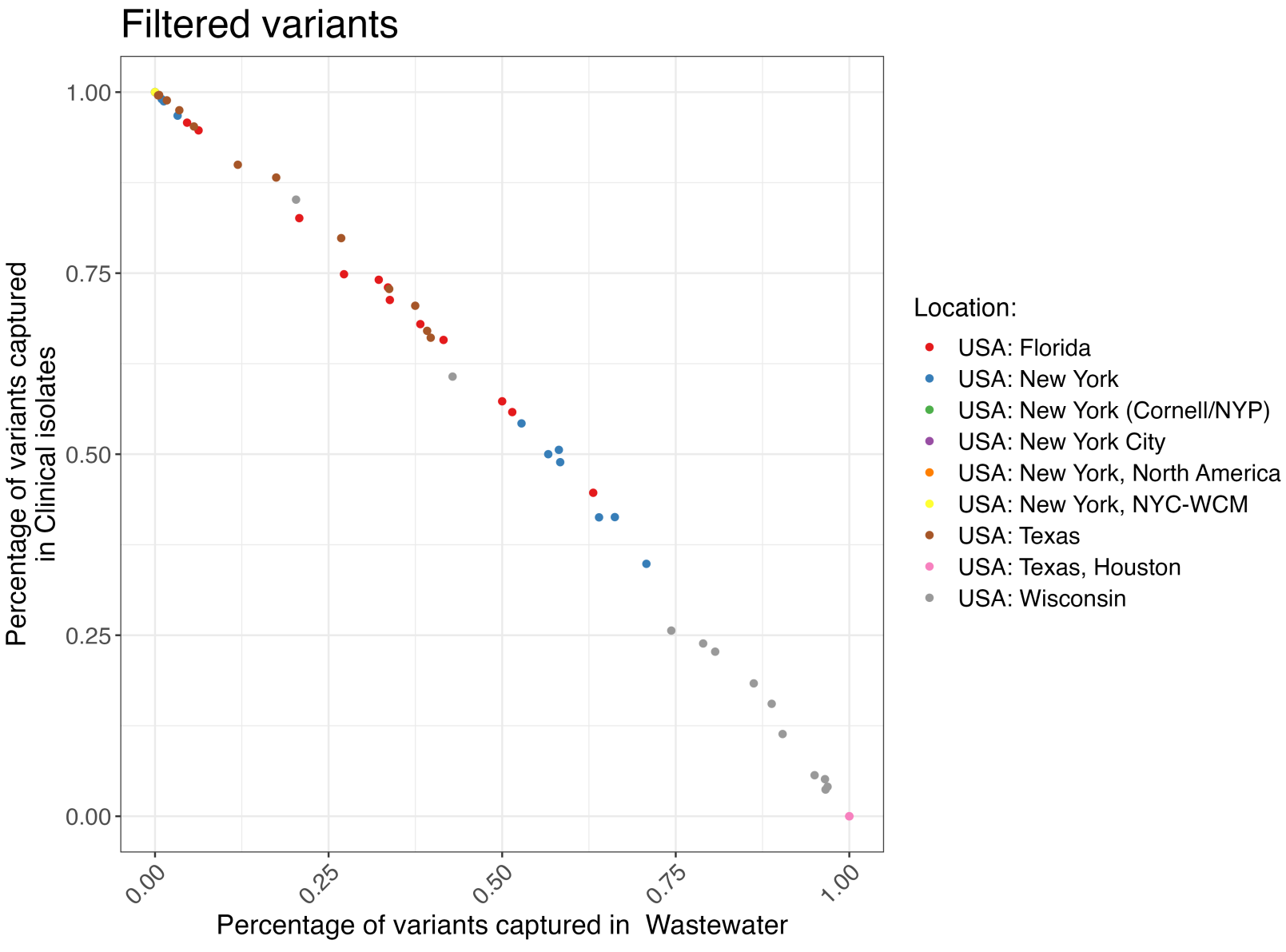
- Filters used
  - Synonymous and intergenic variants were removed.
  - Allele frequency cut-off ≥ 0.4



Sample Distribution




Unfiltered variants
Percentage of variants captured in Clinical isolates
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Percentage of variants captured in Wastewater
Location:
USA: Florida
USA: New York
USA: New York (Cornell/NYP)
USA: New York City
USA: New York, North America
USA: New York, NYC-WCM
USA: Texas
USA: Texas, Houston
USA: Wisconsin


Filtered variants
Percentage of variants captured in Clinical isolates
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Percentage of variants captured in Wastewater
Location:
USA: Florida
USA: New York
USA: New York (Cornell/NYP)
USA: New York City
USA: New York, North America
USA: New York, NYC-WCM
USA: Texas
USA: Texas, Houston
USA: Wisconsin

Conclusions

- When analyzing filtered variants, Wasterwater and Clinical data seems to have little overlap in our comparison.
- A possible cause may be the short time period of the analysis (1 month).
- For most locations, under sampling of Wastewater could be an issue.

Future directions

- Use different time windows.
- Understand the strengths and limitations of wastewater and clinical isolate approaches in terms of temporal alignment.
- https://github.com/NCBI-Codeathons/vcf-4-population-genomics-team-agustinho/tree/main